A Note on Distribution-Free Estimation of Maximum Linear Separation of Two Multivariate Distributions [1]

By Albert Vexler, Aiyi Liu, Enrique F. Schisterman and Chengqing Wu [2]

Division of Epidemiology, Statistics and Prevention Research, NICHD, NIH, U.S.A.

SUMMARY

We consider linear separation of two continuous multivariate distributions. Under mild conditions, the optimal linear separation uniquely exists. A kernel smoothed approach is proposed to estimate the optimal linear combination and the corresponding separation measure. The proposed method yields consistent estimators allowing construction of confidence intervals.

1 Introduction

The quantity $P\{X > Y\}$, where X and Y are two random variables with continuous distribution functions, is a popular measure of separation between the two distributions, and has found application in many areas. In reliability context it is known as the stress-strength coefficient, with X being the stress at failure of a component and Y the stress affecting the component (e.g. [4]). In diagnostic medicine literature, the quantity is the area under the so-called receiver operating characteristic (ROC) curve,

---

[1] AMS 2000 *subject classifications.* Primary: 62G99; Secondary: 62H30, 62G20.

Key words and phrases: The nonparametric Behrens-Fisher problem; measure of separation; Kernel estimation; Stress-strength problem; Receiver operating characteristic curves; Linear combination.

[2] *Correspondence Address:* Albert Vexler, Biometry and Mathematical Statistics Branch, National Institute of Child Health and Human Development, Department of Health and Human Services, 6100 Executive Blvd., Bethesda, MD 20892, USA. Email: vexlera@mail.nih.gov

an import measure of diagnostic accuracy of biomarkers. Here, X are the measurements of the biomarker taken from the population having certain disease, and Y are the measurements taken from the population that do not have such disease. For a chosen cut point c, a subject will be classified as having the disease if the marker's value is larger than c, and not having the disease if otherwise. Plotting the true positive rate, $P\{X > c\}$, versus the false positive rate, $P\{Y > c\}$, for every possible c then yields the ROC curve of the biomarker; see e.g. [3]. Indeed, the quantity is the focus of the non-parametric Behrens-Fisher hypothesis testing (e.g. [14]).

In many studies, multivariate observations occur with measurements from a number of correlated variables collected from each subject. In this situation, we also need a measure of separation between two multivariate distributions F and G, in order to tell how different these two populations are. While there seem to be no unique ways to extend the measure of separation clearly defined in the univariate case to the multivariate case, one popular extension is the measure of maximum linear separation (MLS). Let $\mathbf{X}$ and $\mathbf{Y}$ be two p-dimensional random (column) vectors distributed as F and G, respectively. Then the MLS measure is defined as

$$\rho = \max_{\mathbf{a}} P\{\mathbf{a}'\mathbf{X} > \mathbf{a}'\mathbf{Y}\}, \tag{1.1}$$

where $\mathbf{a}$ is a non-random vector with the same dimension. Hence, ρ is the maximum separation among all linear transformations of the two random vectors.

The above extension has applications in diagnostic medicine [5, 9, 10, 13]. If $\mathbf{X} = (X_1, \ldots, X_p)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_p)$ are the measurements of p diagnostic biomarkers from diseased and non-diseased population, respectively, then ρ achieves the maximum ROC area among linear combinations of the biomarkers. Consequently, linear combination can improve diagnostic accuracy, since $\rho \geq P\{X_i > Y_i\}$ for any $i = 1, \ldots, p$; the latter, as a special linear combination of the markers with the ith element in $\mathbf{a}$ being 1 and others being 0, is the ROC area of the ith biomarker.

Issues remain, however, with regard to the MLS measure. Does the optimal linear combination $\mathbf{a}_o$ exist such that

$$P\{\mathbf{a}_o'\mathbf{X} > \mathbf{a}_o'\mathbf{Y}\} = \max_{\mathbf{a}} P\{\mathbf{a}'\mathbf{X} > \mathbf{a}'\mathbf{Y}\} = \rho\,? \tag{1.2}$$

If such optimal linear combination $\mathbf{a}_o$ exists, how should one estimate it and make inference on the combination and the MLS measure ρ when observations from $\mathbf{X}$ and $\mathbf{Y}$ are available? Most of these issues have been well addressed in the literature (e.g. [5, 10, 13]) when both $\mathbf{X}$ and $\mathbf{Y}$ are assumed to follow multivariate normal distributions, but remained unsolved when only general assumptions are made on distributions.

In the present paper, we derive conditions for the optimal combinations to exist. We then propose approaches to estimate the linear combination and the corresponding MLS. We show that the proposed methods yield consistent estimation for the two quantities, and along the way construct confidence intervals. In order to display several special aspects of the considered issue, Section 4 presents results of Monte Carlo simulations. In Section 5 we demonstrate the proposed methods using a real data set from a study of biomarkers related to coronary heart disease. We give some concluding remarks in Section 6.

2 Existence of the MLS measure

For simplicity, we consider $p = 2$. (For higher dimension, see Remark 2.1 below). We make the following assumptions.

(A0) *Continuity of distributions:* The distributions F and G of $\mathbf{X} = (X_1, X_2)'$ and $\mathbf{Y} = (Y_1, Y_2)$ are both absolutely continuous;

(A1) *Superiority of a marker:* $\rho_1 > \rho_2$, where $\rho_i = P\{X_i > Y_i\}$, $i = 1, 2$;

(A2) *Distinguishability of a marker:* $\rho_2 > 1/2$.

The assumptions (A1) and (A2) can be interpreted in the context of diagnostic medicine as follows. Assume that values of X_i from diseased subjects are in general larger than values of Y_i from non-diseased subjects, then the requirement that the two separation measure be larger than 1/2 indicates that both biomarkers have discriminatory ability, since equality to 1/2 merely means the marker performs no better than flipping a coin to decide the disease status of a subject. The condition that $\rho_1 > \rho_2$ states that the first biomarker has better discriminatory ability (larger ROC area) than the second one, and hence the question becomes whether the diagnostic accuracy can be improved by adding the second marker to the first one; see also below for technical arguments.

Now consider finding an optimal $\mathbf{a} = (a_1, a_2)'$ such that the probability $P\{a_1X_1 + a_2X_2 > a_1Y_1 + a_2Y_2\}$ is maximized. Note that assumptions (A1) and (A2) rule out a_1 being 0 for the solution. Therefore, the optimization problem reduces to find a λ^{opt}, such that

$$\lambda^{opt} = \arg\max_{\lambda} \rho(\lambda), \tag{2.3}$$

where $\rho(\lambda) = P\{X(\lambda) > Y(\lambda)\}$ with $X(\lambda) = X_1 + \lambda X_2$ and $Y(\lambda) = Y_1 + \lambda Y_2$. If such λ^{opt} exists, then $\rho(\lambda^{opt})$ yields the MLS measure ρ.

We now obtain conditions for the existence of a finite "largest peak" of $\rho(\lambda)$.

Proposition 2.1 *Suppose conditions* (A0) - (A2) *hold. Then, there exists a finite* $\lambda^{opt} \in (-\infty, \infty)$ *such that*

$$\lambda^{opt} = \arg \max_{-\infty \le \lambda \le \infty} \rho(\lambda).$$

Proof. By definition (2.3), we have

$$\begin{aligned} \rho(\lambda) &= P\left\{\lambda(X_2 - Y_2) + (X_1 - Y_1) > 0, (X_2 - Y_2) > 0\right\} \\ &\quad + P\left\{\lambda(X_2 - Y_2) + (X_1 - Y_1) > 0, (X_2 - Y_2) \le 0\right\}. \end{aligned} \tag{2.4}$$

Therefore,

$$\lim_{\lambda\to\infty}\rho(\lambda) = P\left\{(X_2 - Y_2) > 0\right\} = \rho_2 \quad \text{and} \quad \lim_{\lambda\to-\infty}\rho(\lambda) = P\left\{(X_2 - Y_2) \leq 0\right\} = 1 - \rho_2.$$

It is clear that, if there exists some finite λ_0 such that $\rho(\lambda_0) > \max(\rho_2, 1-\rho_2)$, then at least one finite maximum of $\rho(\lambda)$ exists. It now follows from Conditions (A1) and (A2) that

$$\rho(0) \equiv \rho_1 \quad > \quad \rho_2 = \max(\rho_2, 1-\rho_2).$$

The proof is complete.

Solutions to (2.3) may not exist if conditions (A0)-(A2) do not hold. For a counter-example, let us assume that $X_1 - Y_1$ is normally distributed and $X_2 - Y_2$ exponentially distributed random variables. Hence, condition (A1) is not satisfied. At this rate, $\max\rho(\lambda) = 1$ at $\lambda = \infty$ and for all $\lambda < \infty$, $\rho(\lambda) < 1$. Similarly, if, for example, $X_1 - Y_1 \sim N(0,1)$ and is independent of $X_2 - Y_2 \sim N(-1,1)$, thus condition (A2) is not satisfied, then $\max\rho(\lambda) = P\{X_2 < Y_2\} \simeq 0.841$, at $\lambda = -\infty$.

Remark 2.1. In order to evaluate the multivariate case of this issue, in a simple form, we assume existence of $p = 3$ markers. Suppose, by applying Proposition 2.1, we conclude that there are finite $\lambda_{10}^{opt} = \arg\max_{-\infty\leq\lambda\leq\infty} P\left\{X_1 + \lambda X_2 > Y_1 + \lambda Y_2\right\}$ and $\lambda_{20}^{opt} = \arg\max_{-\infty\leq\lambda\leq\infty} P\left\{X_2 + \lambda X_3 > Y_2 + \lambda Y_3\right\}$. Thus, inequalities $P\{X_1 > Y_1\} > P\left\{X_2 > Y_2\right\}$ and $P\{X_2 > Y_2\} > P\left\{X_3 > Y_3\right\}$ are presumed. Let the markers be ordered with the distinguishability, i.e.

$$P\left\{X_1 + \lambda_{10}^{opt}X_2 > Y_1 + \lambda_{10}^{opt}Y_2\right\} > P\left\{X_2 + \lambda_{20}^{opt}X_3 > Y_2 + \lambda_{20}^{opt}Y_3\right\} > 1/2.$$

Then, there exist finite

$$\begin{aligned}\left\{\lambda_1^{opt}, \lambda_2^{opt}\right\} &= \arg\max_{-\infty\leq\lambda_1,\lambda_2\leq\infty}\rho(\lambda_1,\lambda_2),\\ \rho(\lambda_1,\lambda_2) &= P\left\{X_1 + \lambda_1 X_2 + \lambda_2 X_3 > Y_1 + \lambda_1 Y_2 + \lambda_3 X_3\right\}.\end{aligned}$$

The demonstration of this fact is similar to the proof of Proposition 2.1. For example, to estimate possible values of $\rho(\infty,\infty)$, we have to analyze situations where $\lambda_2/\lambda_1 \to c$ (c is a positive constant), $\lambda_2/\lambda_1 \to 0$ or $\lambda_1/\lambda_2 \to 0$ as $\lambda_1 \to \infty$ and $\lambda_2 \to \infty$. If $\lambda_2/\lambda_1 \to c$, $\lambda_2/\lambda_1 \to 0$ or $\lambda_1/\lambda_2 \to 0$ then

$$\begin{aligned}
\rho(\infty,\infty) &= P\left\{X_2 + cX_3 > Y_2 + cY_3\right\} < P\left\{X_2 + \lambda_{20}^{opt} X_3 > Y_2 + \lambda_{20}^{opt} Y_3\right\} \\
&< P\left\{X_1 + \lambda_{10}^{opt} X_2 > Y_1 + \lambda_{10}^{opt} Y_2\right\} = \rho(\lambda_{10}^{opt}, 0), \\
\rho(\infty,\infty) &= P\left\{X_2 > Y_2\right\} < P\left\{X_1 > Y_1\right\} < \rho(\lambda_{10}^{opt}, 0), \\
\rho(\infty,\infty) &= P\left\{X_3 > Y_3\right\} < P\left\{X_1 > Y_1\right\} < \rho(\lambda_{10}^{opt}, 0),
\end{aligned}$$

respectively.

Remark 2.2. Note that, if the conditions (A0)-(A2) hold and in addition,

(A3) $\rho''(s) \equiv d^2\rho(s)/ds^2$ exists for any $s \in [a,b]$ and $\rho''(s) \neq 0$,

where the constants a and b are given (we allow that $a = -\infty$ and/or $b = \infty$), then by applying directly the usual Taylor quadratic expansion around the point $\lambda = \lambda_0 \equiv \arg\max_{\lambda\in[a,b]} \rho(\lambda)$ to the function $\rho(\lambda)$, we conclude that there exists unique finite $\lambda^{opt} \in [a,b]$ such that

$$\lambda^{opt} = \arg\max_{a\le\lambda\le b} \rho(\lambda).$$

Note that, by definition (2.3),

$$\rho(\lambda) = P\left\{X_2 < Y_2\right\} + P\left\{X_2 > Y_2\right\} F_{1XY}(\lambda) - P\left\{X_2 < Y_2\right\} F_{2XY}(\lambda), \quad (2.5)$$

where

$$F_{1XY}(\lambda) = P\left\{\frac{Y_1 - X_1}{X_2 - Y_2} < \lambda \middle| X_2 > Y_2\right\}, \; F_{2XY}(\lambda) = P\left\{\frac{Y_1 - X_1}{X_2 - Y_2} < \lambda \middle| X_2 < Y_2\right\}.$$

Therefore the condition on existence of $\rho''(s)$ lays restriction on conditional distributions F_{1XY} and F_{2XY}.

3 Nonparametric estimation of λ^{opt} and ρ

We suppose that random samples, $\{(X_{1j}, X_{2j}),\ 1 \le j \le n\}$ and $\{(Y_{1k}, Y_{2k}),\ 1 \le k \le m\}$, of $\mathbf{X}$ and $\mathbf{Y}$ are available; the bivariate outcomes (X_{1j}, X_{2j}) and (Y_{1k}, Y_{2k}) are independent for all $1 \le j \le n$ and $1 \le k \le m$. Let $X_j(\lambda) = X_{1j} + \lambda X_{2j},\ j = 1, \dots, n$ be independent random variables identically distributed as the random variable $X(\lambda)$. Similarly $Y_k(\lambda) = Y_{1k} + \lambda Y_{2k},\ k = 1, \dots, m$ are independent identically distributed random variables, with distribution function $P\{Y(\lambda) < y\}$. A natural class of estimators of $\rho(\lambda)$ is the U-process based on smooth/empirical Wilcoxon-Mann-Whitney statistics:

$$\hat{\rho}_{nm}(\lambda) = \sum_{j=1}^{n} \sum_{k=1}^{m} w_{jk} G\left(X_j(\lambda), Y_k(\lambda)\right), \tag{3.6}$$

where $0 \le G(.,.) \le 1$ is a real nonnegative Borel function, such that $G(X_j(\lambda), Y_k(\lambda))$ is a reasonable estimator of $P\{X(\lambda) - Y(\lambda) > 0\}$ based upon observations $X_j(\lambda)$ and $Y_k(\lambda)$; $\{w_{jk}\}$ is a sequence of positive numbers such that $\sum_{j=1}^{n} \sum_{k=1}^{m} w_{jk} = 1$, e.g. $w_{jk} = 1/(nm)$. This class of estimators covers several special cases considered by various authors. For example, following [6, 8], we have a kernel-smooth form of estimator (3.6), where $w_{jk} = 1/(nm)$ and $G(X_j(\lambda), Y_k(\lambda)) = \Phi\left(\frac{X_j(\lambda) - Y_k(\lambda)}{h}\right)$, $\Phi = \int \phi$, the kernel ϕ is a mean 0 density function, often standard normal and the parameter h controls the amount of smoothing (λ is fixed); following [3, 9], we have an empirical form of estimator (3.6), where $w_{jk} = 1/(nm)$ and $G(X_j(\lambda), Y_k(\lambda)) = I\left\{X_j(\lambda) > Y_k(\lambda)\right\}$, $I\{B\}$ is the indicator function of a set B. Note that, since

$$\begin{aligned} \rho(\lambda) &= \int_{-\infty}^{\infty} P\left\{Y(\lambda) < u\right\} \frac{dP\left\{X(\lambda) < u\right\}}{du} du \\ &= \int_{-\infty}^{\infty} P\left\{X(\lambda) > u\right\} \frac{dP\left\{Y(\lambda) < u\right\}}{du} du, \end{aligned} \tag{3.7}$$

following [15] we can define function G in the kernel-empirical form

$$G(X_j(\lambda), Y_k(\lambda)) = \frac{1}{h_{nm}} \int_{u\in(Y_k(\lambda),\infty)} K_X\left(\frac{u - X_j(\lambda)}{h_{nm}}\right) du \qquad (3.8)$$
$$(or) = \frac{1}{h_{nm}} \int_{u\in(-\infty,X_j(\lambda))} K_Y\left(\frac{u - Y_k(\lambda)}{h_{nm}}\right) du,$$

where the kernels K_X and K_Y are bounded measurable absolutely integrable functions ($\int K_X = \int K_Y = 1$) and the bandwidth h_{nm} is a positive constant; one practical meaning of using (3.8) is, if the empirical density of $X(\lambda)$ or $Y(\lambda)$ for some λs is close to some theoretical function then we can chose a special form of K_X or K_Y respectively.

3.1 Consistency of estimators of ρ.

Consider estimators $P_n^{X(\lambda)}(u)$ and $P_m^{Y(\lambda)}(u)$ of the distribution functions $P^{X(\lambda)}(u) \equiv P\{X(\lambda) < u\}$ and $P^{Y(\lambda)}(u) \equiv P\{Y(\lambda) < u\}$ respectively. Following (3.7), we define

$$\hat{\rho}_{nm}(\lambda) = \int_{-\infty}^{\infty} P_m^{Y(\lambda)} dP_n^{X(\lambda)}. \qquad (3.9)$$

Hence

$$\begin{aligned}\rho(\lambda) - \hat{\rho}_{nm}(\lambda) &= \int_{-\infty}^{\infty} P^{Y(\lambda)} d(P^{X(\lambda)} - P_n^{X(\lambda)}) + \int_{-\infty}^{\infty} (P^{Y(\lambda)} - P_m^{Y(\lambda)}) dP_n^{X(\lambda)} \\ &= \int_{-\infty}^{\infty} (P_n^{X(\lambda)} - P^{X(\lambda)}) dP^{Y(\lambda)} + \int_{-\infty}^{\infty} (P^{Y(\lambda)} - P_m^{Y(\lambda)}) dP_n^{X(\lambda)} \\ &\leq D_n^- + D_m^+, \qquad (3.10)\end{aligned}$$

where

$$D_n^- = \sup_{-\infty<s<\infty} \left| P_n^{X(\lambda)}(s) - P^{X(\lambda)}(s) \right|,$$
$$D_m^+ = \sup_{-\infty<s<\infty} \left| P^{Y(\lambda)}(s) - P_m^{Y(\lambda)}(s) \right|.$$

Let $P_n^{eX(\lambda)}(u)$ and $P_m^{eY(\lambda)}(u)$ be the empirical distribution functions corresponding to these samples and $P_n^{X(\lambda)}(u) = P_n^{eX(\lambda)}(u)$, $P_m^{Y(\lambda)}(u) = P_m^{eY(\lambda)}(u)$. From Birnbaum and

McCarty [1], D_n^- and D_m^+ have distribution functions which depend on the sample size n, m, but not on $P^{X(\lambda)}$ and $P^{Y(\lambda)}$. It follows from this fact and (3.10) that for every $\epsilon > 0$

$$P\left\{\sup_{-\infty<\lambda<\infty} |\rho(\lambda) - \hat{\rho}_{nm}(\lambda)| > \epsilon\right\} \to 0, \text{ as } n, m \to \infty. \tag{3.11}$$

Consequently, the estimator $\sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda)$ of ρ is consistent

$$P\left\{\left|\rho - \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda)\right| > \epsilon\right\} \;\to\; 0, \text{ as } n, m \to \infty, \tag{3.12}$$

since

$$\left|\sup_{-\infty<\lambda<\infty} \rho(\lambda) - \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda)\right| \le \sup_{-\infty<\lambda<\infty} |\rho(\lambda) - \hat{\rho}_{nm}(\lambda)|\,. \tag{3.13}$$

Adaptations of the kernel-smooth and the kernel-empirical estimators of ρ lead to the same asymptotic conclusion. For example, by applying the first definition from (3.8) and (3.9), we obtain

$$P_n^{X(\lambda)}(u) = \int \int_{-\infty}^{\frac{u-x}{h_{nm}}} K_X(z)\, dz dP_n^{eX(\lambda)}(x) \text{ and } P_m^{Y(\lambda)}(u) = P_m^{eY(\lambda)}(u).$$

Suppose, in addition, that for all $s > 0$ and u, $\left|K_X\left(\frac{u}{s}\right)\frac{u}{s^2}\right| \le c$, where $c > 0$ is a constant. Then by applying Taylor theorem to the function $r(h) = \int_{-\infty}^{\frac{u-X(\lambda)}{h}} K_X(z)\, dz$ around the point $h = 0$ we have, under condition (A0),

$$\begin{aligned}
\left|\int_{-\infty}^{\frac{u-X(\lambda)}{h_{nm}}} K_X(z)\, dz - 1\right| I\{X(\lambda) < u\} &= h_{nm}\left|K_X\left(\frac{u-X(\lambda)}{\theta_1 h_{nm}}\right)\frac{u-X(\lambda)}{(\theta_1 h_{nm})^2}\right| \\
&\quad \times I\{X(\lambda) < u\} \le h_{nm}c, \quad 0 < \theta_1 < 1, \\
\int_{-\infty}^{\frac{u-X(\lambda)}{h_{nm}}} K_X(z)\, dz I\{X(\lambda) \ge u\} &= h_{nm} K_X\left(\frac{u-X(\lambda)}{\theta_2 h_{nm}}\right)\frac{X(\lambda)-u}{(\theta_2 h_{nm})^2} \\
&\quad \times I\{X(\lambda) \ge u\} \le h_{nm}c, \quad 0 < \theta_2 < 1,
\end{aligned}$$

and hence

$$\begin{aligned}\left|\int_{-\infty}^{\frac{u-X(\lambda)}{h_{nm}}} K_X(z)\,dz - I\{X(\lambda)<u\}\right| &= \left|\int_{-\infty}^{\frac{u-X(\lambda)}{h_{nm}}} K_X(z)\,dz - 1\right| I\{X(\lambda)<u\} \\ &\quad + \int_{-\infty}^{\frac{u-X(\lambda)}{h_{nm}}} K_X(z)\,dz I\{X(\lambda)\geq u\} \\ &\leq 2h_{nm}c.\end{aligned}$$

Therefore, D_n^- by (3.10) is no greater than

$$\begin{aligned}&\sup_{-\infty<s<\infty}\left|P_n^{eX(\lambda)}(s) - P^{X(\lambda)}(s)\right| + \sup_{-\infty<s<\infty}\left|P_n^{X(\lambda)}(s) - P_n^{eX(\lambda)}(s)\right| \\ &\leq \sup_{-\infty<s<\infty}\left|P_n^{eX(\lambda)}(s) - P^{X(\lambda)}(s)\right| + 2ch_{nm}\end{aligned}$$

and, consequently, uniform consistency of (3.11), (3.12) holds (as $n, m \to \infty$ and $h_{nm} \to 0$) in this rate, where $\hat{\rho}_{nm}$ is absolutely continuous function.

Now, it is clear that the consideration of estimator $\hat{\rho}_{nm}$ in the full kernel form (3.6), where $w_{jk} = 1/(nm)$ and

$$G(X_j(\lambda), Y_k(\lambda)) = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{u} K_Y\left(\frac{z-Y_k(\lambda)}{h_{nm}}\right)dz\right) d\left(\int_{-\infty}^{u} K_X\left(\frac{z-X_j(\lambda)}{h_{nm}}\right)dz\right),$$

lead us to the same asymptotic property of the estimator.

3.2 Consistency of estimators of λ^{opt}

Denote for all $a_{nm} \leq b_{nm}$

$$\lambda_{nm}^{opt}(a_{nm}, b_{nm}) = \inf\left\{\lambda_0 \geq a_{nm} :\ \hat{\rho}_{nm}(\lambda_0) = \sup_{a_{nm}\leq\lambda\leq b_{nm}} \hat{\rho}_{nm}(\lambda)\right\}. \tag{3.14}$$

Note that, obviously, $\lambda_{nm}^{opt}(a_{nm}, b_{nm}) \neq \sup\left\{\lambda_0 \geq a_{nm} :\ \hat{\rho}_{nm}(\lambda_0) = \sup_{a_{nm}\leq\lambda\leq b_{nm}} \hat{\rho}_{nm}(\lambda)\right\}$ if function G in (3.6) is $I\{X_j(\lambda) > Y_k(\lambda)\}$. Assume that $\rho(\lambda)$ is a uniformly continuous function and

$$\sup_{x:|x-\lambda^{opt}|>\eta} \rho(x) < \rho(\lambda^{opt}) \quad \text{for all} \ \ \eta > 0. \tag{3.15}$$

(This condition defines $\rho(\lambda^{opt})$ as "unique largest peak" of ρ). From this assumption it follows that, to prove $\lambda^{opt}_{nm}(-\infty,\infty) \to \lambda^{opt}$ in probability, it suffices to prove that $\rho\left(\lambda^{opt}_{nm}(-\infty,\infty)\right) \to \rho\left(\lambda^{opt}\right)$ in probability, as $n, m \to \infty$. Now

$$
\begin{aligned}
&\left|\rho\left(\lambda^{opt}_{nm}(-\infty,\infty)\right) - \rho\left(\lambda^{opt}\right)\right| \qquad (3.16)\\
&\leq \left|\rho\left(\lambda^{opt}_{nm}(-\infty,\infty)\right) - \hat{\rho}_{nm}\left(\lambda^{opt}_{nm}(-\infty,\infty)\right)\right| + \left|\hat{\rho}_{nm}\left(\lambda^{opt}_{nm}(-\infty,\infty)\right) - \rho\left(\lambda^{opt}\right)\right|\\
&\leq 2 \sup_{-\infty<\lambda<\infty} \left|\rho(\lambda) - \hat{\rho}_{nm}(\lambda)\right|,
\end{aligned}
$$

Which, along with (3.11), yields $\lambda^{opt}_{nm}(-\infty,\infty) \to \lambda^{opt}$ in probability.

Remark 3.1. The condition (3.15) holds if function ρ satisfies the conditions (A0)-(A3) considered in Section 2. Suppose for fixed data sets $\{(X_{1j}, X_{2j})\}_{j=1}^{n}$ and $\{(Y_{1j}, Y_{2j})\}_{j=1}^{m}$ there exist

$$
\begin{aligned}
&\lim_{\lambda\to\infty} \hat{\rho}_{nm}(\lambda) \geq \lim_{\lambda\to-\infty} \hat{\rho}_{nm}(\lambda) \quad \text{a.s.; some } s,\ \hat{\rho}_{nm}(s) > \lim_{\lambda\to\infty} \hat{\rho}_{nm}(\lambda) \text{ a.s.}\\
&\text{and } \ \hat{\rho}''_{nm}(u) \neq 0 \text{ for all } u \ \text{ a.s.}
\end{aligned}
$$

Then, by applying directly the proof scheme of propositions 2.1 and Remark 2.2 we conclude that $\lambda^{opt}_{nm}(-\infty,\infty)$ is "unique largest peak" of $\hat{\rho}_{nm}$.

3.3 Upper confidence bounds for ρ.

Following [1, 7], by the inequalities (3.13) and (3.10), we can obtain a distribution-free Birnbaum-McCarty type upper confidence bound for the MLS measure ρ. Note that

$$
\left|\rho - \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda)\right| \leq \sup_{-\infty<\lambda<\infty} \left|\rho(\lambda) - \hat{\rho}_{nm}(\lambda)\right| \leq D_n^- + D_m^+,
$$

where D_n^- and D_m^+ by (3.10). Therefore,

$$
P\left\{\rho \leq \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda) + \epsilon\right\} \geq P^D_{nm}(\epsilon), \qquad (3.17)
$$

where P_{nm}^D is the convolution of the distribution functions of Kolmogorov-Smirnov statistics D_n^- and D_m^+. To determine ϵ, note that from [1], we have

$$\lim_{m,n\to\infty} \left| P_{nm}^D(\epsilon) - Q_{mn}(\epsilon) \right| = 0,$$

where

$$\begin{aligned} Q_{mn}(\epsilon) &= 1 - \frac{n}{m+n} e^{-2m\epsilon^2} - \frac{m}{m+n} e^{-2n\epsilon^2} \\ &\quad - \frac{2^{3/2}\pi^{1/2} mn\epsilon^2}{(m+n)^{3/2}} e^{-2mn\epsilon^2/(m+n)} \int_{-2\pi n\epsilon/(m+n)^{1/2}}^{2\pi m\epsilon/(m+n)^{1/2}} e^{-z^2} dz. \end{aligned}$$

Thus an approximation of ϵ can be obtained by solving the equation $Q_{mn}(\epsilon) = 1 - \alpha$, where α is the upper confidence bound for ρ (e.g. [1, 7]).

Although, even in the case of $\rho(0)$-estimation, the nonparametric upper bound is reasonable (e.g. [12]), we suspect that the Birnbaum-McCarty confidence intervals are quite conservative (e.g. [4, p.153]). It is natural to improve the upper bound by applying asymptotic normality (e.g. [2], [4, p.154-155]). That is, by virtue of inequality $\sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda) \geq \hat{\rho}_{nm}(\lambda^{opt})$ (under assumptions of Proposition 2.1), we obtain

$$P\left\{\rho \leq \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda) + \frac{z_\alpha}{W_{nm}}\right\} \geq P\left\{\rho \leq \hat{\rho}_{nm}(\lambda^{opt}) + \frac{z_\alpha}{W_{nm}}\right\}, \tag{3.18}$$

where

$$\begin{aligned} W_{nm}^{-2} &= Var(\hat{\rho}_{nm}(\lambda^{opt})) = \frac{1}{n+m}\left(\rho(\lambda^{opt}) + (n-1)E\left(P^{X(\lambda^{opt})}(Y(\lambda^{opt}))\right)^2\right. \\ &\quad \left. +(m-1)E\left(1 - P^{Y(\lambda^{opt})}(X(\lambda^{opt}))\right)^2 - (n+m-1)\rho(\lambda^{opt})^2\right), \end{aligned}$$

$W_{nm}(\rho - \hat{\rho}_{nm}(\lambda^{opt}))$ is asymptotically $\mathbf{N}(0,1)$ and z_α is the $(1-\alpha)$ percentile of the standard normal distribution (e.g. [4, p.154]). In a similar manner to inequality (5.43) of [4, p.154], by applying $\min(n,m)Var(\hat{\rho}_{nm}(\lambda^{opt})) \leq 1/4$ to (3.18), we can avoid the evaluation of W_{nm} , and therefore asymptotically (as $n, m \to \infty$)

$$P\left\{\rho \leq \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda) + \frac{z_\alpha}{2\min(n,m)^{1/2}}\right\} \geq 1 - \alpha. \tag{3.19}$$

This gives an alternative upper bound for ρ.

Remark 4.1. It is clear that we also have a lower bound for ρ:

$$P\left\{\rho \geq \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda) + \frac{z_\alpha}{W_{nm}}\right\} \leq P\left\{\rho \geq \hat{\rho}_{nm}(\lambda^{opt}) + \frac{z_\alpha}{W_{nm}}\right\} \approx \alpha. \qquad (3.20)$$

4 Monte Carlo simulations

The results reported in this section focus on the performance of two models

$$X_1 \sim \mathbf{N}(1,1),\ X_2 = 2X_1 + \xi,\ \xi \sim \mathbf{N}(0,4), \qquad Y_1 \sim \mathbf{N}(0,1),\ Y_2 \sim \mathbf{N}(0,1) \qquad (4.21)$$

$$\text{and} \qquad X_1 \sim \mathbf{N}(1,4),\ X_2 \sim \mathbf{N}(1,3), \qquad Y_1 \sim \mathbf{N}(0,2),\ Y_2 \sim \mathbf{N}(0,7). \qquad (4.22)$$

Certainly, in either case, conditions (A0) - (A2) are satisfied, and hence, function $\rho(\lambda)$, which corresponds to (4.21) or (4.22), is maximized at finite values of λ. We depict these functions in Figure 1.

Figure 1

Thus, in cases of (4.21) and (4.22), we have $\lambda^{opt} = 0.400$, $\rho = 0.789$ and $\lambda^{opt} = 0.600$, $\rho = 0.697$, respectively. By utilizing 10000 repetitions of generated data sets, which satisfy (4.21) and (4.22), we evaluate estimators (3.6) and (3.14). Table 1 presents results of the simulations, where the empirical and kernel forms of (3.6) and (3.14) are considered in the first and the second line related to same (n, m), respectively.

The formal notations of Table 1 are

$$
\begin{aligned}
\text{empirical} \quad \hat{\rho}_{nm}(\lambda) &= \frac{1}{nm}\sum_{j=1}^{n}\sum_{k=1}^{m} I\left\{X_j(\lambda) > Y_k(\lambda)\right\},\\
\lambda_{nm}^{opt} &= \inf\left\{\lambda_0 \geq -25 :\ \hat{\rho}_{nm}(\lambda_0) = \sup_{-25\leq\lambda\leq 25}\hat{\rho}_{nm}(\lambda)\right\};\\
\text{kernel} \quad \hat{\rho}_{nm}(\lambda;\alpha) &= \frac{1}{nm}\sum_{j=1}^{n}\sum_{k=1}^{m}\Phi\left(\frac{X_j(\lambda)-Y_k(\lambda)}{h_{nm}(\alpha)}\right),\ h_{nm}(\alpha) = (nm)^{-\alpha},\\
\lambda_{nm}^{opt}(\alpha) &= \arg\max_{\lambda_0}\hat{\rho}_{nm}(\lambda_0;\alpha);
\end{aligned}
$$

$\overline{T}$ is the Monte Carlo mean of a statistic T;

$$
\begin{aligned}
&\alpha_\lambda = \arg\min_{0<\alpha<3}\overline{\left(\lambda_{nm}^{opt}(\alpha)-\lambda^{opt}\right)^2},\ \alpha_\rho = \arg\min_{0<\alpha<3}\overline{\left(\max_\lambda\hat{\rho}_{nm}(\lambda;\alpha)-\rho\right)^2};\\
&\hat{\rho}_{nm}^{(I)} = \sup_{-25\leq\lambda\leq 25}\hat{\rho}_{nm}(\lambda), \quad \hat{\rho}_{nm}^{(K)} = \max_\lambda\hat{\rho}_{nm}(\lambda;\alpha_\rho);\\
&\Lambda^{(I)} = \overline{\left(\lambda_{nm}^{opt}-\lambda^{opt}\right)^2}, \quad \Lambda^{(K)} = \overline{\left(\lambda_{nm}^{opt}(\alpha_\lambda)-\lambda^{opt}\right)^2};\\
&\text{and} \quad \Delta^{(I)} = \overline{\left(\sup_{-25\leq\lambda\leq 25}\hat{\rho}_{nm}(\lambda)-\rho\right)^2}, \quad \Delta^{(K)} = \overline{\left(\max_\lambda\hat{\rho}_{nm}(\lambda;\alpha_\rho)-\rho\right)^2}.
\end{aligned}
$$

Table 1

Therefore, in the context of the kernel estimation, the simulations indicate that optimal values of bandwidth h_{nm} tend to be different with respect to the estimators of λ^{opt} and ρ, as well as the marker distributions. However, in accord with Table 1, in the context of the kernel ρ-estimation, $h_{nm} = (nm)^{-0.1}$ is the observed optimal bandwidth for the considered cases. The estimators of λ^{opt} converge more slowly than the estimators of ρ. In terms of Figure 2, which represents the histograms of the empirical and kernel estimators of λ^{opt} for different (n, m) under model (4.21), we conclude that assuming asymptotic normal distributions of these estimators is not indubitable.

Figure 2

Consider the upper confidence bounds (3.17), (3.18) and (3.19) for ρ, under (4.21). The values of ϵ, which correspond to $Q_{nm}(\epsilon) = 0.95$, are $0.481\,(n = m = 15)$, $0.288\,(n = m = 35)$, $0.180\,(n = m = 75)$, $0.180\,(n = m = 75)$, $0.151\,(n = m = 100)$ and $0.119\,(n = m = 150)$. The obtained Monte Carlo respective estimators of $P\left\{\rho \le \sup_{-\infty<\lambda<\infty} \hat{\rho}_{nm}(\lambda) + \epsilon\right\}$ are 0.997, 0.998, 1, 0.987 and 0.980. However, the Monte Carlo evaluations of (3.18) (and (3.19)) with $\alpha = 0.05$ are 0.903(0.951), 0.988(0.989), 0.973(0.985), 0.959(0.965) and 0.956(0.963) with respect to $n = m = 15$, $n = m = 35$, $n = m = 75$, $n = m = 75$, $n = m = 100$ and $n = m = 150$.

Note that, since the generated data sets are normally distributed, assuming the knowledge of this fact, we can use the parametric technique of [13] to evaluate the optimal combination of the markers. At this rate, for example, by basing on (4.21), we observed approximately the same Monte Carlo variances of the parametric and nonparametric estimators of ρ, starting with $n = m \ge 35$. However, the Monte Carlo variance of the nonparametric λ^{opt}-estimator is about 7 times as that of the Monte Carlo variance of the relative parametric estimator. This is, perhaps, partly because of positive devaluations $\lambda_{nm}^{opt} - \lambda$ do not lead to "big" values related to $|\rho_{nm}(\lambda_{nm}^{opt}) - \rho|$ (see, Figures 1 and 2).

Due to the continuation of the kernel estimator of $\rho(\lambda)$ we can maximize this estimator for $\lambda \in (-\infty, \infty)$, whereas in the case of the empirical estimator we should use the finite bounds for the maximization, i.e. $\lambda \in (-25, 25)$.

5 An example

This section demonstrates the performance of the proposed linear combinations using data from a study evaluating biomarkers related to atherosclerotic coronary heart disease. Free radicals have been implicated in the atherosclerotic coronary heart disease process. Well-developed laboratory methods may grant an ample number of biomarkers of individual oxidative stress and antioxidant status. These markers quantify different phases of the oxidative stress and antioxidant status of an individual.

A population-based sample of randomly selected residents of Erie and Niagara counties of the state of New York, U.S.A., $35 - 79$ years of age, was the focus of this study. The New York State Department of Motor Vehicles drivers' license rolls were utilized as the sampling frame for adults between ages of 35 and 65; where the elderly sample (age $65 - 79$) was randomly selected from the Health Care Financing Administration database. A cohort comprised of 942 men and women were selected for the analyses, yielding $n = 143$ cases (individuals with myocardial infarction) and $m = 799$ controls. We illustrate the proposed method with two biomarkers, serum cholesterol (X_1 for case; Y_1 for control) and serum glucose (X_2 for case; Y_2 for control), both are believed to be risk factors of atherosclerotic coronary heart disease, see [11]. Note that, the nonparametric estimators of the areas under the ROC curves with respect to biomarkers X_1, Y_1 and X_2, Y_2 are $\hat{P}\{X_1 > Y_1\} = 0.668$ and $\hat{P}\{X_2 > Y_2\} = 0.667$. We now demonstrate that joint use of the two biomarkers, in the context of the ROC area, has a stronger discriminatory ability for the case and control populations.

In order to utilize the optimal linear combination of the biomarkers, we first attempt to apply the parametric method by [13], which requires multivariate normality. Due to the skewness of the original data, the log-transformation was implemented in order to bring the data distribution closer to normality. The empirical characteristics of the transformed data are $\overline{\ln(X_1)} = 5.627$, $\overline{\ln(X_2)} = 4.747$, $v\hat{a}r(\ln(X_1)) = 0.031$, $v\hat{a}r(\ln(X_2)) = 0.073$, $c\hat{o}r(\ln(X_1), \ln(X_2)) = 0.076$, $\overline{\ln(Y_1)} = 5.473$, $\overline{\ln(Y_2)} = 4.630$, $v\hat{a}r(\ln(Y_1)) = 0.089$, $v\hat{a}r(\ln(Y_2)) = 0.049$ and $c\hat{o}r(\ln(Y_1), \ln(Y_2)) = -0.020$. Thus, by [13], we obtain that $\hat{\rho} = 0.678$ is the parametric estimator of the area under the ROC curve based on the best linear combination. Hence, the profit of the consideration of the two biomarkers is not substantial. Consider the proposed nonparametric estimators based on the transformed data. The empirical form of ρ and λ^{opt}-estimators provides with $\hat{\rho} = 0.725$ and $\hat{\lambda}^{opt} = 1.548$ (the kernel estimators for different bandwidths are about the same).

In order to investigate the robustness of the parametric technique, we also apply both methods to the original non-transformed data, where $\overline{X_1} = 282.315$, $\overline{X_2} = 120.469$, $v\hat{a}r(X_1) = 2810.231$, $v\hat{a}r(X_2) = 2025.307$, $c\hat{o}r(X_1, X_2) = 0.064$, $\overline{Y_1} = 248.208$, $\overline{Y_2} = 105.660$, $v\hat{a}r(Y_1) = 4926.610$, $v\hat{a}r(Y_2) = 1103.846$ and $c\hat{o}r(Y_1, Y_2) = -0.037$. Subsequently, we estimate $\hat{\rho} = 0.680$, by [13]. Now consider the proposed nonparametric method. By applying the empirical form of the estimator of the best linear combination of the biomarkers, we calculate $\hat{\rho} = 0.735$ with $\hat{\lambda}^{opt} = 4.446$ (the kernel estimators for different bandwidths are about the same). In this case, in accord with Section 3.3, we obtain an upper bound with 95% coverage probability for the MLS measure of $0.735 + 0.104 = 0.839\,(P\{\rho < \hat{\rho} + 0.104\} \geq 0.95)$ and $0.735 + 0.068 = 0.803\,(P\{\rho < \hat{\rho} + 0.068\} \geq 0.95)$, respectively from (3.17) and (3.19). The obtained upper bounds for the MLS give the best performance that the optimal linear combination of the biomarkers can achieve.

Therefore, using the parametrical method requires log-transformation of the data. However, applying the proposed method can be based on the non transformed data and yields an estimated area, which is greater than the area under the ROC curve based on the transformed data.

6 Conclusion and discussion

In this paper we have shown, through analytic derivations, simulations, and a real data example, that a non-parametric linear combination of several individual variables improves the distribution-separation measure. Both the empirical and kernel-smoothed approaches lead to consistent estimation of the linear combination and the separation measure. However the kernel forms of the estimator facilitate the practical maximization of the target estimator.

In a subsequent paper we will address asymptotic normality of the proposed estimators of the linear combination and the MLS and optimal bandwidth selections, which needs substantial mathematical details. Our limited simulation results have shown that asymptotically the nonparametric estimator of the MLS tends to be normally distributed while the nonparametric estimator of the optimal linear combination is likely to have a limiting exponential distribution.

Acknowledgments. The authors would like to thank the Associate Editor and two referees for suggestions which helped very much to improve this paper.

References

[1] Birnbaum, Z.W. and McCarty, R.C. (1958). A distribution-free upper confidence bound for $\Pr\{Y < X\}$, based on independent samples of X and Y. *Ann. Math. Statist.*, **29**, 558-562.

[2] Govindarajulu, Z. (1968). Distribution-free confidence bounds for $P(X < Y)$. *Ann. Inst. Stat. Math.*, **20**, 229-238.

[3] Hsieh, F. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Statist.,* **24**, 25-40.

[4] Kotz, S., Lumelskii, Y. and Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. World Scientific Publishing, London.

[5] Liu, A., Schisterman, E.F. and Zhu, Y. (2005). On linear combinations of biomarkers to improve diagnostic accuracy. *Stat. Med.,* **24**, 37-47.

[6] Lloyd, C.J. (1998). Using smoothed receiver operating characteristic to summarize and compare diagnostic systems. *J. Amer. Statist. Assoc.,* **93**, 1356-1364.

[7] Owen, D.B., Craswell, K.S and Hanson, D.L. (1964). Nonparametric upper confidence bounds for $\Pr\{Y < X\}$ and confidence limits for $\Pr\{Y < X\}$ when X and Y are normal. *J. Amer. Statist. Assoc.,* **59**, 906-924.

[8] Peng, L. and Zhou, X. (2004). Local linear smoothing of receiver operating characteristic (ROC) curves. *J. Stat. Plann. Inf.,* **118**, 129-143.

[9] Pepe, M.S. and Thompson, M.L. (2000). Combining diagnostic test results to increase accuracy. *Biostatistics,* **1-2**, 123-140.

[10] Reiser, B. and Faraggi, D. (1997). Confidence intervals for the generalized ROC criterion. *Biometrics*, **53**, 644-52.

[11] Schisterman, E.F., Faraggi, D., Browne, R., Freudenheim, J., Dorn, J., Muti, P., Armstrong, D., Reiser, B. and Trevisan, M. (2001). TBARS and cardiovascular disease in a population-based sample. *Journal of Cardiovascular Risk,* **8**, 1–7.

[12] Sprent, P. (1989). *Applied Nonparametric Statistical Methods.* Chapman & Hall, London.

[13] Su, J.Q. and Liu, J.S. (1993). Linear combinations of multiple diagnostic markers. *J. Amer. Statist. Assoc.,* **88**, 1350-1355.

[14] Troendle, J. F. (2002). A likelihood ratio test for the nonparametric Behrens-Fisher problem. *Biometrical Journal,* **44**, 813-824.

[15] Yang, S.S. (1985). A smooth nonparametric estimator of a quantile function. *J. Amer. Statist. Assoc.,* **80**, 1004-1011.

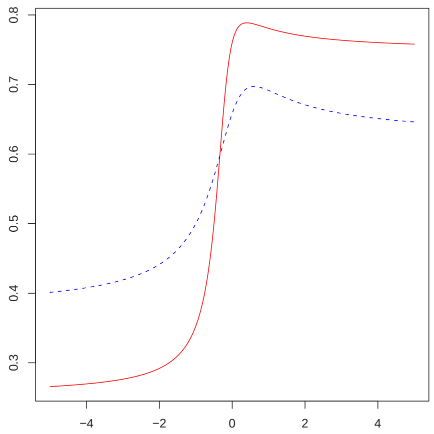

Figure 1: Curves (—) and (- - -) of function $\rho(\lambda)$ corresponding to (4.21) and (4.22), respectively.

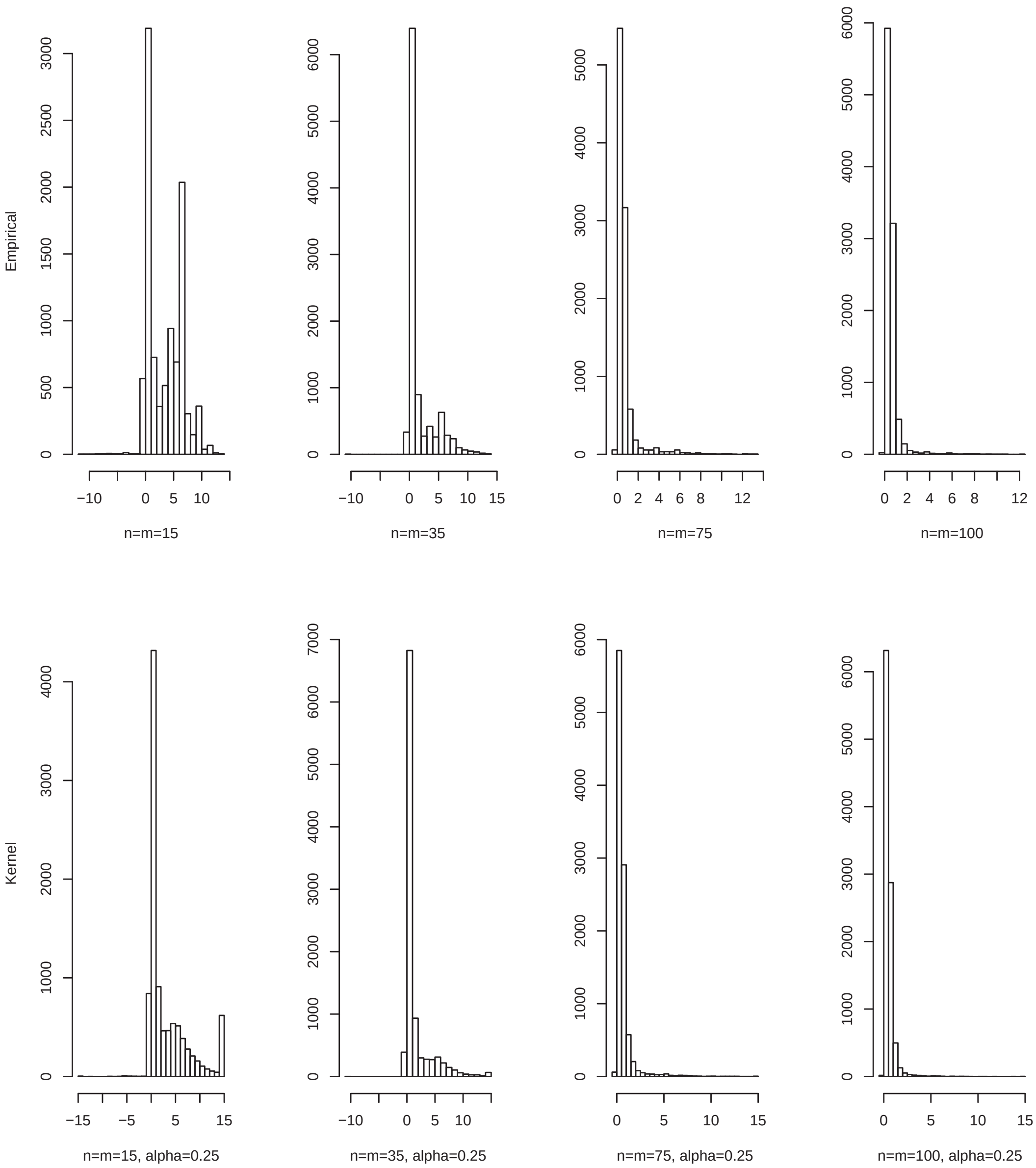

Figure 2: The histograms of the empirical and kernel estimators of λ^{opt} for different (n, m) under model (4.21).

Table 1: Monte Carlo simulation results.

$m = n$	α_λ	$\overline{\lambda_{nm}^{opt}}$ / $\overline{\lambda_{nm}^{opt}(\alpha_\lambda)}$	$\Lambda^{(I)}$ / $\Lambda^{(K)}$	α_ρ	$\overline{\hat{\rho}_{nm}^{(I)}}$ / $\overline{\hat{\rho}_{nm}^{(K)}}$	$\Delta^{(I)}$ / $\Delta^{(K)}$
			Under model (4.21)			
15		3.4543	19.0295		0.7979	0.0086
15	1.0286	3.3153	18.7196	0.1000	0.8007	0.0070
35		1.7015	7.5830		0.7962	0.0034
35	0.2857	1.4687	7.2225	0.1000	0.7915	0.0031
75		0.7280	1.3359		0.7938	0.0014
75	0.2208	0.6650	1.1930	0.1000	0.7891	0.0014
100		0.5654	0.4478		0.7923	0.0011
100	0.1000	0.5734	0.3554	0.1000	0.7924	0.0011
150		0.4823	0.1175		0.7912	0.0007
150	0.2208	0.4635	0.0926	0.1000	0.7878	0.0007
			Under model (4.22)			
15		2.3713	15.6403		0.7182	0.0096
15	1.8333	2.3682	15.6223	0.1000	0.7249	0.0085
35		1.4290	4.9566		0.7095	0.0042
35	0.3364	1.3431	4.8359	0.1000	0.7077	0.0037
75		0.8034	0.7008		0.7050	0.0018
75	0.5727	0.7996	0.6665	0.1000	0.7026	0.0017
100		0.7181	0.2906		0.7025	0.0013
100	0.1000	0.6953	0.2664	0.1000	0.7003	0.0013