

Research Article

Soft Roc Curves

Xin Huang¹, Narayanaswamy Balakrishnan^{2,3} and Yixin Fang^{4*}

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, USA

²Department of Mathematics and Statistics, McMaster University, Canada

³King Saud University, Saudi Arabia

⁴Division of Biostatistics, Department of Population Health, New York University, USA

*Corresponding author: Yixin Fang, Division of Biostatistics, Department of Population Health, New York University, New York, NY 10016, USA

Received: August 25, 2014; Accepted: October 13, 2014; Published: November 18, 2014

Abstract

Receiver operating characteristic (ROC) curves are a popular tool for evaluating continuous diagnostic tests. However, the traditional definition of ROC curves incorporates implicitly the idea of “hard” thresholding, which cannot encompass the situation when some intermediate classes are introduced between test result positive and negative, and also results in the empirical curves being step functions. For this reason, we introduce here the definition of soft ROC curves, which incorporates the idea of “soft” thresholding. The softness of a soft ROC curve is controlled by a regularization parameter that can be selected suitably by a cross-validation procedure. A byproduct of the soft ROC curves is that the corresponding empirical curves are smooth. The methods developed here are then examined through some simulation studies as well as a real illustrative example.

Keywords: Cross-validation; Diagnostic test; Intermediate class; Regularization parameter; Thresholding

Introduction

Receiver Operating Characteristic (ROC) curves is a popular tool for evaluating continuous diagnostic tests; see, for example, Pepe [1]. However, the traditional definition of ROC curves incorporates implicitly the idea of “hard” thresholding. To be specific, let T be the outcome of a continuous diagnostic test and D be the disease status. Given a threshold c , the hard thresholding scheme defines a subject as diseased ($\hat{D}=1$) if the test result $T = t$ exceeds c , and as non-diseased ($\hat{D}=0$) otherwise. It thus results in a binary classifier,

$$(H) \quad I(t-c) = \begin{cases} 1, & t-c \geq 0, \\ 0, & t-c < 0. \end{cases}$$

The ROC curve is then a graphical plot of true positives, $E\{I(T-c)|D=1\}$, versus false positives, $E\{I(T-c)|D=0\}$, for $-\infty < c < \infty$. It can be expressed as

$$R(p) = 1 - G[F^{-1}(1-p)], \quad 0 < p < 1,$$

where $F(\cdot)$ and $G(\cdot)$ are the distributions of T , given $D = 0$ and $D = 1$, respectively.

However, from the medical practitioners point of view, if the test result is close to the given threshold c , then one may be indecisive about the status of diseases. This is a common situation for tests with ambiguous thresholds (e.g., prostate-specific antigen, which is shown to be not a dichotomous marker [2]). Thus, practitioners tend to implement an intermediate class between the negative and positive [3], within which patients are diagnosed as diseased or non-diseased according to some probability model. Hozo and Djulbegovic [4] provide a definition of acceptable regret threshold to explain such phenomenon. They demonstrate that different practitioners might adapt different acceptable regret thresholds for withholding treatment even when the diagnostic tests exceed the pre-defined threshold. Unfortunately, the existing hard-thresholding scheme does not incorporate such intermediate classes. Furthermore, there are other disadvantages in the hard thresholding scheme. In

particular, the discontinuity of the binary classifier results in the corresponding estimated ROC curve being a step function, while the underlying ROC curve is likely to be smooth. Consequently, due to the discontinuity in the step function, the variability of the estimated ROC curve becomes large.

To overcome these disadvantages, we consider soft-thresholding scheme,

$$(S) \quad I_{\delta}(t-c) = \begin{cases} 1, & t-c \geq \delta, \\ ?, & -\delta < t-c < \delta, \\ 0, & t-c < -\delta, \end{cases}$$

where the value ? is between 0 and 1 and will be discussed in the next section, and δ is a regularization parameter controlling the softness. In particular, when $\delta=0$, the soft thresholding simply becomes the hard thresholding. When decision-making rule I_{δ} is applied with threshold c , the sensitivity (a.k.a. the true positive probability) equals $E\{I_{\delta}(T-c)|D=1\}$ and the specificity (a.k.a. true negative probability) equals $E\{1-I_{\delta}(T-c)|D=0\}$.

The rationale of this soft-thresholding scheme is that if the test result is close to the given threshold c , then one may be indecisive about the status of the disease. Hence, we refer to $I_{\delta}(\cdot)$ as the indecisive function. And the probability model within the intermediate class can be formulated by? In the indecisive function. We will show that different indecisive functions will result in different soft ROC curves. The idea used here is similar in principle to the one used in designing randomization tests to achieve a given significance level in hypothesis testing [5]. The indecisive function has been considered in the literature of ROC analysis. Many authors have used smooth functions to approximate the indicator function, which can also be considered as indecisive functions. For example, Liu et al. [6] and Liu and Tan [7] used an S-type function to approximate the indicator function for the empirical False Positive Rate (FPR) and True Positive Rate (TPR). Huang et al. [8], Wang et al. [9], and Ma and Huang [10,11] used the sigmoid function to approximate the indicator function in the empirical estimate of the Area Under the ROC Curve (AUC).

Instead of looking for an approximation, in this work, we examine the definition of ROC curves directly and introduce the soft ROC curves based on the soft-thresholding. More importantly, we build a bridge between the approximation of an ROC curve and the approximation of its AUC. Moreover, continuity of the proposed soft ROC curves is a promising byproduct, although it is not our primary goal. We should point out that in the literature of ROC; many authors have discussed methods to smooth ROC curves. For example, Zou et al. [12] proposed a non-parametric estimator from kernel estimates of the distribution functions F and G . Peng and Zhou [13] proposed a local linear regression for the ROC curve, while Ren et al. [14] proposed a Penalized Spline Linear Mixed-Effects model (PSLME). In this paper, we demonstrate that the proposed soft ROC method not only has similar performances when compared to the local linear regression and the PSLME methods in terms of smoothing, but also has a clearer explanation to the smoothing parameter and much easier implementation.

The remainder of this paper is organized as follows. In Section 2, we define the soft ROC curve, and derive some of its properties. In Section 3, we propose methods to choose the regularization parameter δ . In Section 4, the proposed methods are examined through some simulation studies and a real data example. Finally, some discussion is made in Section 5, and all technical details are relegated to the Appendix.

Soft ROC Curves

When an indecisive function I_δ is applied with threshold c , we can define a soft ROC curve as follows.

Definition 1: A plot of true positives, $E\{I_\delta(T-c)|D=1\}$, versus false positives, $E\{I_\delta(T-c)|D=0\}$, for all possible values of c , is called the soft ROC curve with respect to the indecisive function I_δ .

Assume that a test is performed on m non-diseased subjects, yielding testing outcomes X_i , and on n diseased subjects, yielding outcomes Y_j . Then, an empirical estimate of the soft ROC curve w.r.t. I_δ is

$$\widehat{R}_\delta(p) = 1 - \widehat{G}_\delta \left[\widehat{F}_\delta^{-1}(1-p) \right], \quad p \in (0,1), \tag{1}$$

where $\widehat{G}_\delta(c) = \frac{1}{n} \sum_{j=1}^n I_\delta(Y_j - c)$ and $\widehat{F}_\delta(c) = \frac{1}{m} \sum_{i=1}^m I_\delta(X_i - c)$. The area under the soft ROC curve w.r.t. I_δ , denoted by AUC_δ , is derived in the following theorem, and its proof is presented in the Appendix A.

Theorem 1: For the soft ROC curve w.r.t. to the indecisive function $I_\delta(\cdot)$, we have

$$AUC_\delta = E\{K_\delta(Y-X)\},$$

where $X \sim F(\cdot)$, $Y \sim G(\cdot)$, $K_\delta(Y-X) = \int_{-\infty}^{\infty} I_\delta(Y-c) I_\delta(X-c) dc$, and I_δ is the derivative of F_δ .

We remark that for functions with piecewise constant, the derivative is defined by using Dirac Delta function. From Theorem 1, we see that an unbiased estimate of AUC_δ is given by

$$\widehat{AUC}_\delta = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n K_\delta(Y_j - X_i). \tag{2}$$

It is worth emphasizing if the hard-thresholding decision rule (H) is applied, then we use the classical ROC curve to evaluate its performance, Whereas if the soft-thresholding decision rule (S) is applied, then we use the newly proposed soft ROC curve to evaluate

its performance. In other words, which type of ROC curves is used for evaluation depends on the underlying decision rule that is applied. Actually, it is not necessary to define a new ROC curve for any new decision rule. However, we define soft curves for at least three reasons. First, the soft-thresholding decision rule is simple and appropriate. Second, the resulted empirical soft ROC curve is continuous. Third, the relationship between K_δ and I_δ is mathematically beautiful.

Two-sided soft ROC curves

We can categorize indecisive functions and soft ROC curves into one-sided and two-sided according to the following definition.

Definition 2: If $I_\delta(t-c)=0$ for $t < c$, I_δ and the corresponding soft ROC curve are said to be one-sided. Otherwise, they are said to be two-sided.

We now present some examples of indecisive functions I_δ and their corresponding K_δ , which are all displayed in Figure 1. The corresponding detailed calculations are presented in the Appendix B.

Example 1: Order 0 two-sided indecisive function is given by

$$I_\delta(t-c) = \frac{1}{2} 1\{-\delta \leq t-c < \delta\} + 1\{t-c \geq \delta\},$$

where $1\{\cdot\}$ is an indicator function. This implies that the disease status is totally indecisive (the chance of being diagnosed as diseased is 50%) when t is within δ of threshold c . The corresponding K_δ is

$$K_\delta(s) = \frac{1}{4} 1\{-2\delta \leq s < 0\} + \frac{3}{4} 1\{0 \leq s < 2\delta\} + 1\{s \geq 2\delta\}.$$

Example 2: Order 1 two-sided indecisive function is given by Appendix C

$$I_\delta(t-c) = \left[\frac{1}{2} + \frac{1}{2\delta}(t-c) \right] 1\{-\delta \leq t-c < \delta\} + 1\{t-c \geq \delta\}.$$

This indecisive function is continuous, and it implies that the probability of being diagnosed as diseased is linearly increasing in $t-c$ when t is within δ of threshold c . The corresponding K_δ is

$$K_\delta(s) = \left[\frac{1}{2} + \frac{5s}{4\delta} \text{sign}(s) \frac{s^2}{2\delta^2} \right] 1\{-2\delta \leq s < 2\delta\} + 1\{s \geq 2\delta\},$$

where $\text{sign}(\cdot)$ is the sign function.

Example 3: Order ∞ two-sided (Sigmoid) indecisive function is given by Appendix D

$$I_\delta(t-c) = \frac{1}{1 + e^{-\delta(t-c)}}.$$

An appealing property of the sigmoid function is that it has infinite derivatives. The corresponding K_δ is

$$K_\delta(s) = -e^{\delta s} \left[\frac{1}{1 - e^{\delta s}} + \frac{1}{(1 - e^{\delta s})^2} \delta s \right].$$

This K_δ also enjoys the property of having infinite derivatives.

One-sided indecisive functions

In this subsection, we present two examples of one-sided indecisive functions I_δ and their corresponding K_δ , which are displayed in Figure 2. The indecisive functions are similar to the ones in Examples 1 and 2, but the order 1 one-sided K_δ takes on a reasonable form, unlike its two-sided counterpart. The corresponding detailed calculations are

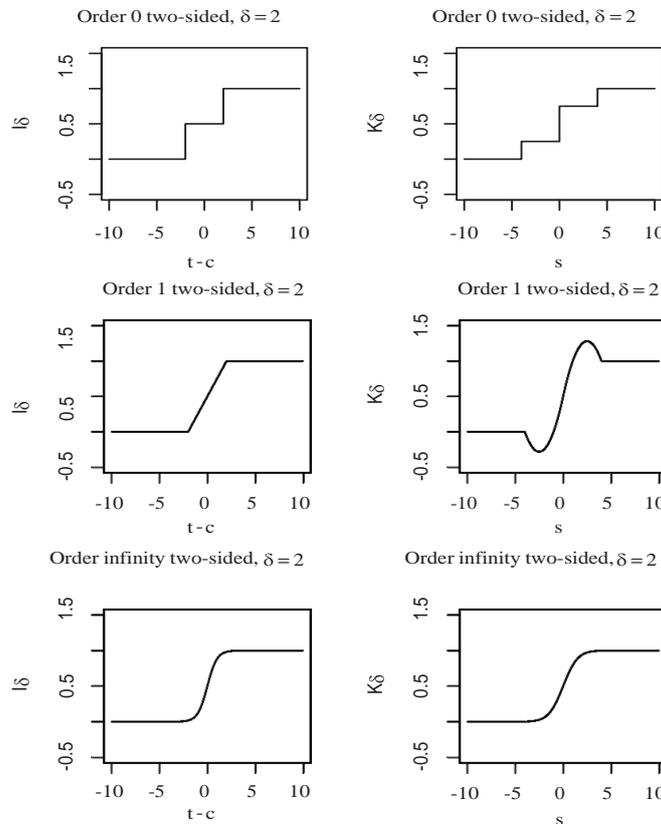


Figure 1: Two-sided I_δ and their corresponding K_δ .

presented in the Appendix E.

Example 4: Order 0 one-sided indecisive function is given by

$$I_\delta(t - c) = \frac{1}{2}1\{0 \leq t - c < \delta\} + 1\{t - c \geq \delta\}.$$

This implies that the disease status is totally indecisive when t is in the interval $[c, c + \delta)$. The corresponding K_δ is

$$K_\delta(s) = \frac{1}{4}1\{-\delta \leq s < 0\} + \frac{3}{4}1\{0 \leq s < \delta\} + 1\{s \geq \delta\}.$$

Example 5: Order 1 one-sided indecisive function is given by Appendix F

$$I_\delta(t - c) = \frac{1}{\delta}(t - c)1\{0 \leq t - c < \delta\} + 1\{t - c \geq \delta\}$$

This implies that the probability of being diagnosed as diseased is linearly increasing in $t - c$ when t is in the interval $[c, c + \delta)$. The corresponding K_δ is

$$K_\delta(s) = \left[\frac{1}{2} + \frac{s}{\delta} - \text{sign}(s) \frac{s^2}{2\delta^2} \right] 1\{-\delta \leq s < \delta\} + 1\{s \geq \delta\}$$

Surprisingly, the minor change in this indecisive function from its two-sided counterpart results in a big change in the corresponding K_δ , and K_δ has a continuous derivative. In what follows, we will focus on this indecisive function. Of course, the procedures developed here for this indecisive function can also be applied to other indecisive functions.

Selection of Regularization Parameter

Method based on softness

The regularization parameter δ controls the softness of a soft ROC curve. The bigger the δ is, the softer the ROC curve is. When δ is taken as zero, it becomes the traditional ROC curve as mentioned earlier. Hence, it is important to select an appropriate regularization parameter δ . First, we define the softness of a soft ROC curve as follows.

Definition 3: For a soft ROC curve with a regularization parameter δ , the softness is defined as

$$\alpha = 1 - \frac{P(Y - X > \delta)}{P(Y - X > 0)},$$

where $X \sim F(\cdot)$ and $Y \sim G(\cdot)$. The hardness is then naturally defined as $1 - \alpha$.

The softness α indirectly controls the form of the empirical soft ROC curve estimated from (1). For example, if the order 1 one-sided indecisive function is used, the softness ranges from 0 (when $\delta = 0$ in which case the soft ROC curve becomes a step function) to 1 (when $\delta = \infty$ in which case the soft ROC curve becomes a diagonal line). As mentioned before, the idea of soft-thresholding is similar to the one used in designing randomization tests in hypothesis testing. In this regard, the softness defined above is analogous to significance level in the setting of randomization tests.

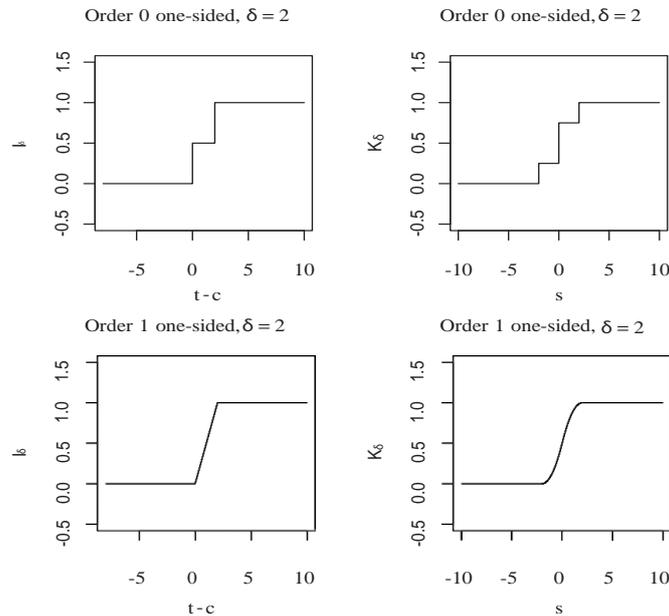


Figure 2: One-sided $l\delta$ and their corresponding $K\delta$.

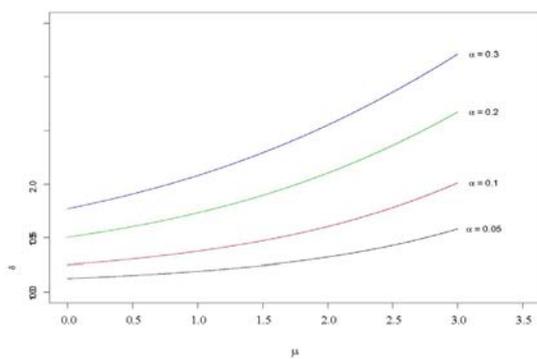


Figure 3: Plots of δ versus mean difference μ for some given α .

Figure 3 shows the plots of δ versus the differences of means of diseased and non-diseased populations for some choices of α . Here, we denote $\mu = E\{Y\} - E\{X\}$ and assume that the two populations are normal with unit standard deviation.

Evidently, a non-parametric estimate of softness is given by

$$\alpha = 1 - \frac{\sum_{i=1}^m \sum_{j=1}^n I(Y_j - X_i - \delta)}{\sum_{i=1}^m \sum_{j=1}^n I(Y_j - X_i)} \tag{3}$$

For a pre-specified α , we can choose a regularization parameter δ . But the determination of α is quite subjective. Recall that the same issue is present in hypothesis testing wherein the significance level is usually taken to be 5%. From the limited simulation studies we have carried out, we would suggest considering softness between 0.1 and 0.3. In the next subsection, we propose a cross-validation procedure for selecting an appropriate δ without prefixing α .

Method based on cross-validation

In this subsection, we propose a Cross-Validation (CV) procedure for selecting δ by minimizing the Average Mean Squared Error (AMSE) [14],

$$AMSE(\delta) = E \left\{ \frac{1}{K} \sum_{k=1}^K \left[\widehat{R}_\delta(p_k) - R(p_k) \right]^2 \right\} \tag{4}$$

where p_k is in a fine grid of $(0,1), k=1, \dots, K$.

For this purpose, we randomly split the sample into two parts, or we randomly split the diseased and non-diseased samples into two parts each. For each random split, we treat one part as a training sample and the other as a validation sample. Based on the training sample, we construct the soft ROC curve and obtain the estimate $\widehat{R}_\delta^{(1)}$, and based on the validation sample, we construct the regular ROC curve and obtain the estimate $\widehat{R}^{(2)}$. By repeating this random split many times, we obtain the following cross-validation estimate of the AMSE:

$$CV_\delta = \frac{1}{H} \frac{1}{K} \sum_{h=1}^H \sum_{k=1}^K \left[\widehat{R}_\delta^{(1,h)}(p_k) - \widehat{R}^{(2,h)}(p_k) \right]^2, \tag{5}$$

where H is the number of random splits. Then, δ is chosen as the one that minimizes CV_δ in (5).

The split ratio (training/validation) can be chosen to be either 1:1 or 2:1. From our limited simulation studies, we observe that the results are not sensitive to the split ratio. Such an idea of cross-validation has been considered by many authors including Bickel and Levina [15]. In theory, Shao [16] examined the consistency of cross-validation procedures with different split ratios in linear regressions.

Numerical Results

In this section, the proposed methods are examined through Monte Carlo simulation studies and a real data example. The R codes are available from the authors upon request.

Simulation Study

We investigated the performance of the CV procedure through two simulation studies. Let Y be the diseased population and X be the

non-diseased population. We considered two types of distributions: (1) both X and Y follows normal distributions; and (2) both X and Y follows double exponential distributions. We considered four settings of means: (i) $(\mu_y, \mu_x) = (1, 0)$; (ii) $(\mu_y, \mu_x) = (1.5, 0)$; (iii) $(\mu_y, \mu_x) = (2, 0)$; and (iv) $(\mu_y, \mu_x) = (2.5, 0)$. We considered two settings of variances: (A) $Var(Y) = 1$ and $Var(X) = 1$; and (B) $Var(Y) = 2$ and $Var(X) = 1$. Therefore, there were $2 \times 4 \times 2 = 16$ data generating settings. The sample sizes were taken as $m = n = 50$ or $m = n = 100$, and the split ratio was set as 2:1. For each data generating setting, 300 replications were performed to calculate the efficiency measure

$$\frac{AMSE(\hat{\delta}^{cv})}{AMSE(0)} \tag{6}$$

and the efficacy measure,

$$\frac{AMSE(\hat{\delta}^{cv})}{\min_{\delta} AMSE(\delta)} \tag{7}$$

where $\hat{\delta}^{cv}$ is the δ chosen by the CV procedure. The simulation results so obtained are summarized in Table 1. These results show that all efficiencies are less than 1, while efficacies are all close to 1, which indicates the optimality of $\hat{\delta}^{cv}$. From this table, we also observe that $\hat{\delta}^{cv}$ is decreasing when the difference $\mu_y - \mu_x$ increases. In fact, when Y and X are well-distinguished, the indecisive interval vanishes.

We also compared the smoothed empirical ROC curves; a promising byproduct of the proposed soft ROC method with the local linear regression method [13] and PSLME model [14]. In this study, four datasets were generated from two normal distributions with unit standard deviation and means being $(\mu_y, \mu_x) = (1.5, 0)$. And the sample size were taken as $m = n = 20$, $m = n = 50$, $m = n = 100$, and $m = n = 500$. The results are presented in Figure 4, which show that if the goal is to smooth the empirical ROC curve, all the methods perform similarly.

Pancreatic cancer serum biomarkers example

The dataset comes from a case-control study at Mayo Clinic which included 90 patients with pancreatic cancer and 51 subjects with pancreatitis. These data were originally analyzed by Wieand et al. [17]. Two continuous positive scale serum biomarkers were available to diagnose a patient with pancreatic cancer: CA-125, a cancer antigen, and CA-19-9, a carbohydrate antigen. We applied the CV method to select regularization parameters for CA-125 and CA-19-9, which turn out to be 0.04 and 0.115, respectively. The corresponding empirical ROC, soft ROC curves, and smoothed ROC curves by local linear regression and PSLME model are displayed in Figure 5 and Figure 6. Again, for the overall performance, we observe that the smoothed ROC curves estimated from the soft ROC method are similar to the smoothed curves from existing smoothing methods.

Discussion

Many authors have considered using the sigmoid function to approximate the indicator function when calculating the AUC, but without clear reasoning. In this paper, by introducing soft ROC curves, we have provided a connection between the approximation to ROC curve and the approximation to the corresponding AUC. This explains in some way as to why we can use some function to approximate the indicator function while calculating the AUC.

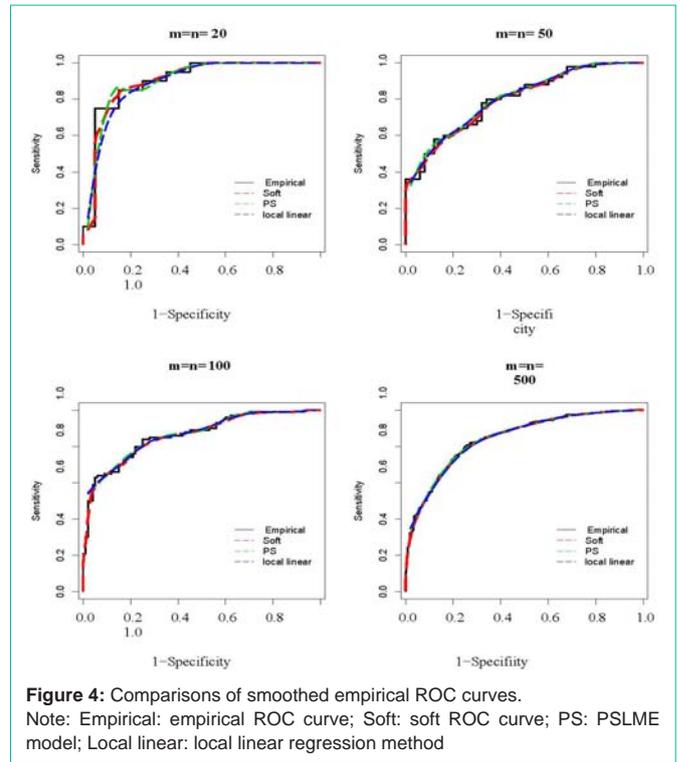


Figure 4: Comparisons of smoothed empirical ROC curves. Note: Empirical: empirical ROC curve; Soft: soft ROC curve; PS: PSLME model; Local linear: local linear regression method

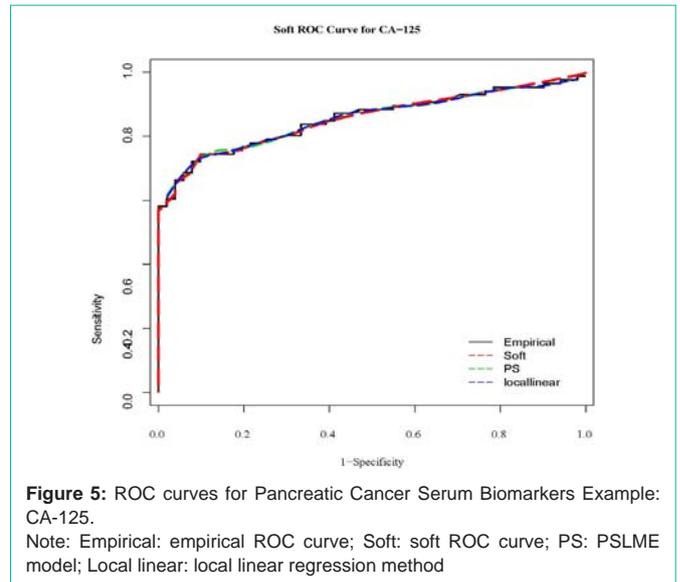
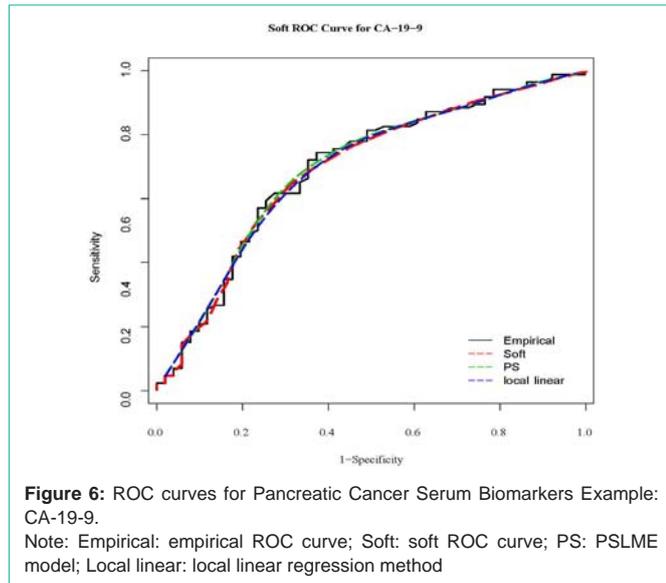


Figure 5: ROC curves for Pancreatic Cancer Serum Biomarkers Example: CA-125. Note: Empirical: empirical ROC curve; Soft: soft ROC curve; PS: PSLME model; Local linear: local linear regression method

The selection of the regularization parameter in a soft ROC curve is a critical issue. The application of the proposed cross-validation procedure is straightforward. Since the cross-validation is one of the most popular methods for model selection, we have examined it in the present context, by means of Monte Carlo simulation studies and a real example, and have shown that it performs well. However, the consistency of the proposed cross-validation procedure remains as an open problem.

An asymptotic estimate of the variability of the estimated soft ROC curve can be easily developed by following arguments similar to those of by Pepe [1]. However, if the asymptotic variance is not derivable, one can also use the bootstrap.



The choice of the indecisive function varies in different practical scenario. For example, if one wants to model the decision behaviors of the medical practitioners, he/she can use the order 0 two-sided indecisive function, where the chance of being diagnosed as diseased is 50% when the test result is within the soft threshold, or the order 1 one-sided indecisive function, where the probability of being diagnosed as diseased is linearly increasing when the test result is within the soft threshold. Moreover, if one focuses on the computational issue of the ROC curves, the order infinity two-sided function (sigmoid function) can be used [8,11].

In the future study, we will also generalize the indecisive function by allowing the regularization parameter to vary across different test results. Under appropriate generalizations, the indecisive function becomes a risk type function; the corresponding soft ROC and AUC thus have direct applications on clinical trials.

Acknowledgment

We are grateful to Dr. Hua Liang from University of Rochester for his smoothing ROC programs for PSLME model.

References

1. Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press: Oxford. 2004.

2. Thompson IM. PSA: a biomarker for disease. A biomarker for clinical trials. How useful is it? *J Nutr.* 2006; 136: 2704S.
3. Body R, Foex B. On the philosophy of diagnosis: is doing more good than harm better than "primum non nocere"? *Emerg Med J.* 2009; 26: 238-240.
4. Hozo I, Djulbegovic B. When is diagnostic testing inappropriate or irrational? Acceptable regret approach. *Med Decis Making.* 2008; 28: 540-553.
5. Lehmann EL. Testing Statistical Hypotheses, 2nd edn. Springer: New York. 1997.
6. Liu Z, Tan M, Jiang F. Regularized F-measure maximization for feature selection and classification. *J Biomed Biotechnol.* 2009; 617946.
7. Liu Z, Tan M. ROC-based utility function maximization for feature selection and classification with applications to high-dimensional protease data. *Biometrics.* 2008; 64: 1155-1161.
8. Huang X, Qin G, Fang Y. Optimal Combinations of Diagnostic Tests Based on AUC. *Biometrics.* 2011; 67: 568-576.
9. Wang Z, Chang YI, Ying Z, Zhu L, Yang Y. A parsimonious thresholding-independent protein feature selection method through the area under receiver operating characteristic curve. *Bioinformatics.* 2007; 23: 2788-2794.
10. Ma S, Huang J. Regularized ROC method for disease classification and biomarker selection with microarray data. *Bioinformatics.* 2005; 21: 4356-4362.
11. Ma S, Huang J. Combining multiple markers for classification using ROC. *Biometrics.* 2007; 63: 751-757.
12. Zou KH, Hall WJ, Shapiro DE. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Stat Med.* 1997; 16: 2143-2156.
13. Peng L, Zhou XH. Local linear smoothing of receiver operating characteristic (ROC) curves. *Journal of Statistical Planning and Inference.* 2004; 118: 129-143.
14. Ren H, Zhou XH, Liang H. A flexible method for estimating the ROC curve. *Journal of Applied Statistics.* 2004; 31: 773-784.
15. Bickel PJ, Levina E. Regularized estimation of large covariance matrices. *Annals of Statistics.* 2008; 36: 199-227.
16. Shao J. Linear model selection by cross-validation. *Journal of the American Statistical Association.* 1993; 88: 486-495.
17. Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika.* 1989; 76: 585-592.