

**NONPARAMETRIC MAXIMUM LIKELIHOOD
ESTIMATION OF A CONCAVE RECEIVER
OPERATING CHARACTERISTIC CURVE
VIA GEOMETRIC PROGRAMMING**

KYEONG EUN LEE AND JOHAN LIM

ABSTRACT. A receiver operating characteristic (ROC) curve plots the true positive rate of a classifier against its false positive rate, both of which are accuracy measures of the classifier. The ROC curve has several interesting geometrical properties, including concavity which is a necessary condition for a classifier to be optimal. In this paper, we study the nonparametric maximum likelihood estimator (NPMLE) of a concave ROC curve and its modification to reduce bias. We characterize the NPMLE as a solution to a geometric programming, a special type of a mathematical optimization problem. We find that the NPMLE is close to the convex hull of the empirical ROC curve and, thus, has smaller variance but positive bias at a given false positive rate. To reduce the bias, we propose a modification of the NPMLE which minimizes the \mathcal{L}_1 distance from the empirical ROC curve. We numerically compare the finite sample performance of three estimators, the empirical ROC curve, the NPMLE, and the modified NPMLE. Finally, we apply the estimators to estimating the optimal ROC curve of the variance-threshold classifier to segment a low depth of field image and to finding a diagnostic tool with multiple tests for detection of hemophilia A carrier.

1. Introduction

A receiver operating characteristic (ROC) curve is a graphical representation of two measures relating to the accuracy of diagnostics, false positive rate (FPR) and true positive rate (TPR). FPR is the probability of making a false positive decision, whereas TPR is that of making a correct positive decision. To be specific, suppose we use a continuous variable X to diagnose a certain

Received September 16, 2009.

2010 *Mathematics Subject Classification.* 62G05.

Key words and phrases. concavity, geometric programming, receiver operating characteristic curve.

This paper was issued from the BIEN2009 conference: INWES Asian Network held from August 27 to 29 in Busan, Korea, hosted by the Association of Korean Women Scientists and Engineers (KWSE).

disease; if the value of X is larger than a critical value c , we classify the subject into a disease (positive) group; otherwise, we classify it to a non-disease (negative) group. Let S_1 and S_2 be survival functions of X of the true negative and the true positive group, respectively. The FPR and the TPR of a given classifier is

$$(1) \quad \begin{aligned} \text{FPR}(c) &= S_1(c) \\ \text{TPR}(c) &= S_2(c). \end{aligned}$$

The ROC curve plots $\text{TPR}(c)$ against $\text{FPR}(c)$ for different values of c . Thus, it is defined, for $p \in [0, 1]$, as:

$$(2) \quad R(p) = S_2\{S_1^{-1}(p)\}.$$

Concavity is a necessary condition for the optimal ROC curve, which is a plot of the highest achievable TPR for each FPR level by a given classifier. Suppose an ROC curve is not concave on an interval $[a, b] \subset [0, 1]$, where a and b correspond to the sensitivity at critical values c_1 and c_2 respectively. Then, for the diagnosis with critical values $c \in [c_2, c_1]$, we find a better diagnostic system which has the same sensitivity but low specificity by randomly choosing between two diagnoses with critical values c_1 and c_2 . Thus, the concave upper envelope (least concave majorant) of the observed ROC curve represents the potential optimal ROC curve.

Several procedures has been proposed to estimate the optimal concave ROC curve. The most common one in the literature assumes a parametric error distribution which provides a concave ROC curve (Campbell and Ratnaparkhi [3]; Dorfman et al. [4]; Pan and Metz [21]; Metz and Pan [17]; Lloyd [13]). Lloyd [14] proposes a procedure to compute the maximum likelihood estimator (MLE) of a concave ROC curve when we know the probabilities of true positive and true negative. Also in a machine learning literature, Flach and Wu [6] propose a mirroring method which computes a ROC curve by reflexing the column and row of the confusion matrix of the FPR and TPR.

In this paper, we are interested in the nonparametric maximum likelihood estimator (NPMLE) of a concave ROC curve. We characterize it as a solution to a geometric program (GP), a special type of mathematical optimization problem. The GP can easily be transformed to a convex optimization problem, and then solved globally and efficiently using interior-point methods; we refer the reader to the paper “*A Tutorial on Geometric Programming* (Boyd et al. [2])” for more on geometric programming.

The numerical study based on GP shows that the NPMLE is close to the convex hull of the empirical ROC curve and has positive bias at a given FPR, although its variance is much smaller than that of the empirical ROC curve. To this reason, we further propose a modification of the NPMLE to minimize the \mathcal{L}_1 distance from the empirical ROC curve within a certain class of concave ROC functions.

The paper is organized as follows. In Section 2, we introduce the NPMLE and the main problem which needs to be solved in order to obtain it. In Section 3, we briefly describe the GP and characterize the main problem as a GP. In Section 4, we implement a numerical study to investigate the performance of the NPMLE and show that the NPMLE has positive bias, although it has smaller variance than the empirical ROC curve. In Section 5, we introduce a modification of the NPMLE to reduce a bias and numerically compare its bias and MSE to those of the NPMLE and the empirical ROC curve. In Section 6, we apply the modified NPMLE to estimating the ROC curves for two real examples. In the first example, we estimate the optimal ROC curves of the local variance classifier (LVC) in segmenting a low depth of field image. In the second example, we compare several classifiers to detect hemophilia A carriers. In Section 7, we briefly summarize the paper.

2. NPMLE

2.1. Main problem

Let us consider independent random samples from two populations, a disease (positive group, $i = 2$) and a non-disease (negative group, $i = 1$) population. Let S_i be the survival function (1 - cdf) of the i th population, and let x_{i1}, \dots, x_{in_i} be the observed classification scores for $i = 1, 2$. Complete observations occur on a subset of scores $x_1 < x_2 < \dots < x_m$, where $\{x_1, \dots, x_m\}$ is the union of all observed scores x_{i1}, \dots, x_{in_i} for $i = 1, 2$. For the i th population, let d_{ij} denote the number of observations in the interval $[x_j, x_{j+1})$, then $n_{ij} = \sum_{r=j}^m d_{ir}$ is the number of observations whose values are larger than or equal to x_j .

We consider a binary classifier responsible for assigning a subject to group 2 if its value x is larger than a constant γ , and to group 1 otherwise. The false positive rate (FPR) is $P(X > \gamma | i = 1) = S_1(\gamma)$ and the true positive rate (TPR) is $P(X > \gamma | i = 2) = S_2(\gamma)$. The ROC curve of the classifier is the parametric curve $(S_1(\gamma), S_2(\gamma))$, the plot of the TPR against the FPR. In other words, the ROC curve $R(p)$ is that in (2).

We are interested in estimating the ROC curve under the concavity constraint from observations x_{i1}, \dots, x_{in_i} for $i = 1, 2$. The NPMLE of the ROC curve without concavity constraint is

$$\widehat{R}_n(p) = \widehat{S}_{2n} \{ \widehat{S}_{1n}^{-1}(p) \} \quad \text{for } p \in (0, 1),$$

where \widehat{S}_{1n} and \widehat{S}_{2n} are the unconstrained MLEs of S_1 and S_2 . They are equal to empirical survival curves of populations 1 and 2, and maximize

$$(3) \quad \mathcal{L}(S_1, S_2) = \prod_{i=1}^2 \prod_{j=1}^m \{ S_i(x_{j-1}) - S_i(x_j) \}^{d_{ij}}$$

over all pairs of piecewise constant functions which are right continuous with jumps only at x_1, \dots, x_m (Kaplan and Meier [10]; Johansen [8]; Feltz and Dykstra [5]).

The NPMLEs $\widehat{S}_1^{\text{con}}$ and $\widehat{S}_2^{\text{con}}$ under concavity constraint is the solution to

$$(4) \quad \begin{aligned} &\text{maximize} && \mathcal{L}(S_1, S_2) = \prod_{i=1}^2 \prod_{j=1}^m \left\{ S_i(x_j -) - S_i(x_j) \right\}^{d_{ij}} \\ &\text{subject to} && R(p) = S_2(S_1^{-1}(p)) \text{ being concave in } p. \end{aligned}$$

The solution to (4) has all mass at the observed points x_1, \dots, x_m as in an unconstrained NPMLE. If some estimators $\widetilde{S}_1^{\text{con}}$ and $\widetilde{S}_2^{\text{con}}$ do not have such a property, we could find an alternative solution satisfying concavity constraint, whose likelihood value is larger than that of $(\widetilde{S}_1^{\text{con}}, \widetilde{S}_2^{\text{con}})$. Thus, the problem (4) is equivalent to

$$(5) \quad \begin{aligned} &\text{maximize} && \mathcal{L}(S_1, S_2) = \prod_{i=1}^2 \prod_{j=1}^m \left\{ S_i(x_{j-1}) - S_i(x_j) \right\}^{d_{ij}} \\ &\text{subject to} && \begin{aligned} &(S_2(x_j) - S_2(x_{j-1})) / (S_1(x_j) - S_1(x_{j-1})) \\ &\leq (S_2(x_{j+1}) - S_2(x_j)) / (S_1(x_{j+1}) - S_1(x_j)). \end{aligned} \end{aligned}$$

The likelihood function (5) could be rewritten with variables

$$p_{ij} = S_i(x_j) / S_i(x_{j-1}), \quad q_{ij} = 1 - p_{ij}, \quad i = 1, 2, \quad j = 1, \dots, m.$$

The variables of the problem (5) can be expressed with p_{ij} s as $S_i(t_j) = \prod_{r=1}^j p_{ir}$. So the objective can be written as

$$\begin{aligned} \mathcal{L}(S_1, S_2) &= \prod_{i=1}^2 \prod_{j=1}^m \left\{ S_i(x_{j-1}) - S_i(x_j) \right\}^{d_{ij}} \\ &= \prod_{i=1}^2 \prod_{j=1}^m \left\{ 1 - \frac{S_i(x_j)}{S_i(x_{j-1})} \right\}^{d_{ij}} \left\{ \frac{S_i(x_{j-1})}{S_i(x_{j-2})} \cdots S_i(x_1) \right\}^{d_{ij}} \\ &= \prod_{i=1}^2 \prod_{j=1}^m (1 - p_{ij})^{d_{ij}} \prod_{r < j} p_{ir}^{d_{ij}}. \end{aligned}$$

The concavity constraints are also written with p_{ij} s for $i = 1, 2$, and $j = 1, 2, \dots, m$. As addressed above, the MLEs of S_1 and S_2 are specified only up to finite points $\{x_1, x_2, \dots, x_m\}$, and, hence, we are only concerned with the concavity on these points. Thus, the constraints are

$$\frac{S_2(x_j) - S_2(x_{j-1})}{S_1(x_j) - S_1(x_{j-1})} \leq \frac{S_2(x_{j+1}) - S_2(x_j)}{S_1(x_{j+1}) - S_1(x_j)},$$

(note that S_1 and S_2 are nonincreasing in x_j) which are also written using p_{ij} s as

$$\left(\frac{p_{1j}}{1 - p_{1j}} \right) \leq \left(\frac{p_{2j}}{1 - p_{2j}} \right) \left(\frac{1 - p_{2(j+1)}}{1 - p_{1(j+1)}} \right)$$

for $j = 1, 3, \dots, m - 1$. Finally, the problem to be solved is

$$(6) \quad \begin{aligned} & \text{maximize} && \mathcal{L}(\{p_{1j}\}, \{p_{2j}\}) = \prod_{i=1}^2 \prod_{j=1}^m (1 - p_{ij})^{d_{ij}} \prod_{r < j} p_{ir}^{d_{ij}} \\ & \text{subject to} && \left(\frac{p_{1j}}{1 - p_{1j}} \right) \left(\frac{1 - p_{2j}}{p_{2j}} \right) \left(\frac{1 - p_{1(j+1)}}{1 - p_{2(j+1)}} \right) \leq 1 \quad \text{for } j = 2, 3, \dots, m - 1. \end{aligned}$$

3. Computation via geometric program

First we will introduce two classes of functions, *monomials* and *posynomials*, which define a geometric program. Let x_1, \dots, x_n denote n real positive variables, and $x = (x_1, \dots, x_n)$ represent a vector with components x_i . A *monomial* is a real valued function f of x with the form

$$(7) \quad f(x) = cx_1^{a_1} x_2^{a_2} \cdots x_n^{a_n},$$

where $c > 0$ and $a_i \in \mathbb{R}$. A *posynomial* is the sum of one or more monomials, namely

$$(8) \quad f(x) = \sum_{k=1}^K c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}},$$

where each c_k is positive.

A *geometric program* is an optimization problem of the form

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 1, \quad i = 1, \dots, m, \\ & && g_j(x) = 1, \quad j = 1, \dots, p, \end{aligned}$$

where f_i are posynomials, g_j are monomials, and $x_i > 0$ are the optimization variables. Here, the objective must be a posynomial and must be minimized, the equality constraints must be a monomial equal to one, and the inequality constraints must be a posynomial less than or equal to one. See the tutorial by Boyd, et al. [2] for more on geometric programming.

We can solve a GP efficiently by converting it to a *non-linear convex optimization problem*, *i.e.*, a problem with convex objective and inequality constraint functions, and linear equality constraints. The converting is achieved by performing a logarithmic transformation in the sense that, in place of the original variables x_i , we use their logarithms, $y_i = \log x_i$ (so $x_i = e^{y_i}$). The transformed version of a GP is a convex optimization problem for which efficient solution methods are well developed (Boyd and Vandenberghe [2]; Nesterov and Nemirovsky [19]; Norcedal and Wright [20]). In particular, interior methods for GPs are very efficient and robust. Furthermore, high-quality implementation of a primal-dual interior-point method for GPs is already available from solvers. Some examples of methods, which have a simple interface that recognizes and solves GPs, include CVX (Grant, Boyd, and Ye [7]), GGPLAB (Mutapcic, Koh, Kim, and Boyd [18]), and YALMIP (Löfberg [15]).

The GP-based methods have been used for solving problems that arise in estimating multinomial probabilities under constraints. Alldredge and Armstrong (1974) consider the problem of estimating overlap sizes created by interlocking sampling schemes. Mazumdar and Jefferson (1983) consider the monotonicity in Bernoulli probabilities when sums of k Bernoulli random variables are observed. Briker, Kortanek, and Xu (1997) consider the constraints on the local odd ratios when the frequencies are observed along with an $r \times c$ rectangular array. The problems considered in the aforementioned papers involve maximizing the likelihood function of multinomial observations under certain constraints and can be reformulated as GPs. Recently, Lim et al. [12] and Lim et al. [11] studied the application of a GP to estimating current status data and stochastically ordered survival functions.

We rewrite the main problem (6) using variables $q_{ij} = 1 - p_{ij}$ for $i = 1, 2$, and $j = 1, 2, \dots, m$:

$$(9) \quad \begin{aligned} & \text{maximize} && \mathcal{L}(\{(p_{ij}, q_{ij})\}) = \prod_{i=1}^2 \prod_{j=1}^m q_{ij}^{d_{ij}} \prod_{r < j} p_{ir}^{d_{ir}} \\ & \text{subject to} && \begin{aligned} & \left(\frac{p_{1j}}{q_{1j}}\right) \left(\frac{q_{2j}}{p_{2j}}\right) \left(\frac{q_{1j}}{q_{2(j+1)}}\right) \leq 1 \quad \text{for } j = 2, 3, \dots, m-2, \\ & p_{ij} + q_{ij} = 1 \quad \text{for } i = 1, 2, \quad j = 1, 2, \dots, m. \end{aligned} \end{aligned}$$

The objective of the problem (9) is a monomial of the variables. As shown in the above, in a GP, the only equality constraints allowed involve monomials, so the problem (9) is not a GP, since the equality constraints are posynomial equalities. We replace those posynomial equalities with inequalities to obtain the problem

$$(10) \quad \begin{aligned} & \text{maximize} && \mathcal{L}(\{(p_{ij}, q_{ij})\}) \\ & \text{subject to} && \begin{aligned} & \left(\frac{p_{1j}}{q_{1j}}\right) \left(\frac{q_{2j}}{p_{2j}}\right) \left(\frac{q_{1j}}{q_{2(j+1)}}\right) \leq 1 \quad \text{for } j = 2, 3, \dots, m-2, \\ & p_{ij} + q_{ij} \leq 1 \quad \text{for } i = 1, 2, \quad j = 1, 2, \dots, m, \end{aligned} \end{aligned}$$

which is a GP.

We establish the equivalence between (9) and (10). Since the posynomials $p_{ij} + q_{ij}$ are *monotone increasing* in either p_{ij} or q_{ij} , the objective is decreasing in the variables, and so we can see that at the optimal point, the inequality constraints $p_{ij} + q_{ij} \leq 1$ must be tight. In other words, if for some i and j , $\bar{p}_{ij} + \bar{q}_{ij} < 1$ is strict, then $(\bar{p}_{ij}, \bar{q}_{ij})$ cannot be optimal since $(\hat{p}_{ij}, \hat{q}_{ij})$ with $\hat{p}_{ij} = \bar{p}_{ij}$ for $j = 1, 2, \dots, n-1$ and $\hat{q}_{ij} = 1 - \bar{p}_{ij} > \bar{q}_{ij}$ are feasible for (10) but

$$\begin{aligned} & \mathcal{L}(\{(\bar{p}_{kl}, \bar{q}_{kl}), (k, l) \neq (i, j)\}, (\hat{p}_{ij}, \hat{q}_{ij})) \\ & > \mathcal{L}(\{(\bar{p}_{kl}, \bar{q}_{kl}), (k, l) \neq (i, j)\}, (\bar{p}_{ij}, \bar{q}_{ij})). \end{aligned}$$

Hence, both problems (9) and (10) are equivalent, in the sense that their optimal values and solutions are the same, and the solution of one problem is

readily obtained from that of the other, and vice versa. In summary, we can solve the original problem (9), through the tractable GP (10).

4. Numerical study

Next, we implement a simulation study to investigate the performance of the NPMLE.

We consider a two class classification problem using a single variable x . We consider a classifier which assigns a subject into group 1 (resp. group 2) if $x \leq c$ (resp. if $x > c$). We consider two errors distributions, the normal distribution, and t-distribution with 5 degrees of freedom. For normal errors, the data set in group 1 is generated from the normal distribution with mean 0 and variance 1, whereas the data in group 2 is generated from the normal distribution with mean of δ and same variance. Thus, its true ROC curve is

$$(11) \quad R(p) = 1 - \Phi(-\delta + \Phi^{-1}(1-p)),$$

which is a concave function on $[0, 1]$; $\Phi(x)$ is the cumulative distribution function (cdf) of the standard normal distribution. For t-error distributions, the data set in group 1 is generated from a t-distribution with mean 0, whereas the data in group 2 is generated from one with a mean of δ . The degrees of freedom of the t-distribution is set to be 5. Thus, its true ROC curve again has the same shape as (11) which only differs in using the cdf of the t-distribution with mean 0 and 5 degrees of freedom instead of Φ . In both normal and t cases, δ is chosen as 0.5 or 1. The sample sizes (n_1, n_2) are chosen as either equal sample sizes $n_1 = n_2$ or unequal sample sizes $n_1 = 2n_2 (\neq n_2)$. The total sample size $n = n_1 + n_2$ is chosen as $n = 50, 100, 200, 300$ for equal sample size cases, and as $n = 60, 120, 210, 300$ for unequal sample size cases.

We generate 100 data sets for each case we consider. In each data set, we compute the empirical ROC and the NPMLE of the concave ROC curve. We approximate the bias and the mean squared error (MSE) of the estimated ROC curves at grid points $p \in [0, 1]$. In Figures 1-2, we plot the results of the cases involving normal error distribution with $n_1 = n_2$, and those for t-error distribution with $n_1 \neq n_2$. The results obtained for other cases are similar to the reported ones.

The figures show that the NPMLE has larger positive bias, but smaller variance than the empirical ROC curve. A necessary condition for the NPMLE to be consistent is that the number of observations d_{ij} in the interval $[x_j, x_{j+1})$ increases for every j th interval and the i th population. However, this will not be true unless the data is interval censored.

5. Modified NPMLE

Section 4 shows that the constrained NPMLE has smaller variance but has larger positive bias than the empirical ROC. For this reason, its mean squared error (MSE) is similar to that of the empirical ROC. In this section, we propose

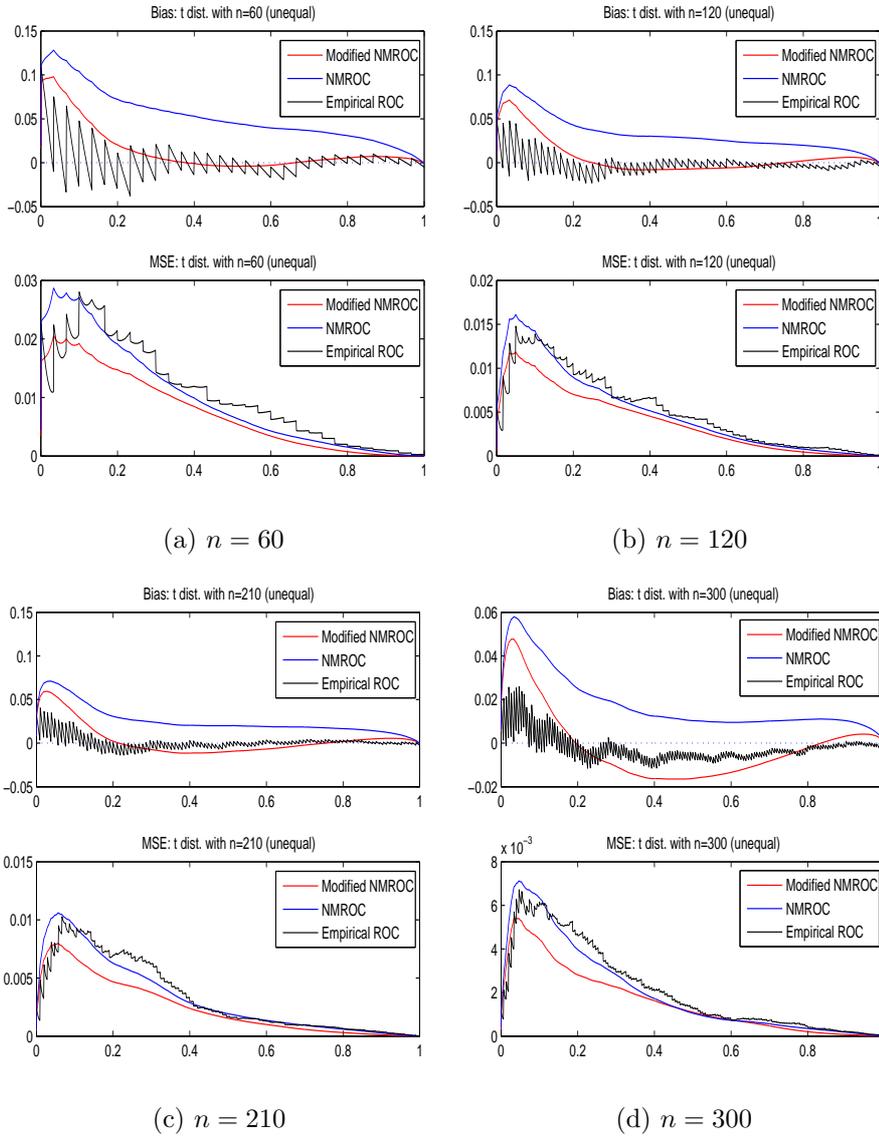


FIGURE 2. t-error distribution with unequal sample sizes. Here, $n_1 = 2 \cdot n_2$ and $n = n_1 + n_2$. The modified NMPLE is introduced in the next section.

a function to minimize its \mathcal{L}_1 distance to the empirical ROC curve in a given specific class of concave ROC curves.

We consider the class of concave ROC curves which consists of mixtures of the NPMLE and a concave lower envelope. Let $R^{\text{emp}}(p)$ and $R^{\text{mle}}(p)$ be the empirical ROC curve and the NPMLE, respectively. We define the concave lower envelope $R^{\text{lower}}(p)$ as

$$R^{\text{lower}}(p) = \max(R_1(p), p),$$

where $R_1(p)$ is the piecewise linear line connecting points $(0, 0)$, $(p^*, R^{\text{emp}}(p^*))$, and $(1, 1)$ and $p^* = \operatorname{argmax}_p \{R^{\text{mle}} - R^{\text{emp}}\}$. To be specific, the class of ROC curves we are considering is

$$\mathcal{R} = \{\lambda R^{\text{mle}} + (1 - \lambda)R^{\text{lower}}, \lambda \in [0, 1]\}.$$

Finally, we can define the modified NPMLE as

$$R^{\text{m.mle}}(p) = \operatorname{argmin}_{R \in \mathcal{R}} \int_0^1 |R(p) - R^{\text{emp}}(p)| dp,$$

where the integration is approximated by a grid point method.

We further numerically compare $R^{\text{m.mle}}(p)$ and two other estimators, $R^{\text{mle}}(p)$ and $R^{\text{emp}}(p)$ using exactly the same simulation setting as that used in the previous section. By computing the bias and the MSE of the modified NPMLE and then plotting them together in Figures 1 and 2, we found that the bias of the modified NPMLE is significantly smaller than the NPMLE. The modified NPMLE introduces an extra variance, however, its overall MSE is smaller than those of the NPMLE and the empirical ROC curve.

6. Examples

In this section, we illustrate the proposed method with two real examples; one is the segmentation of the region of interest (ROI) in a low depth of field (DOF) image; the other is the detection of hemophilia A carriers using two measurements in blood samples.

6.1. Low depth of field

In this subsection, we consider a segmentation problem in low depth of field (DOF) images. In DOF images, only the object of interest (OOI) is focused and the other regions defocused and blurred. As a result, the focused regions normally have more high frequency components than the defocused ones. The classifier to extract OOI from a DOF image often relies on the detection of the high frequency areas in the image (Wang et al. [24]; Yim and Bovik [25]).

We consider a classifier which is based on a local variance σ_t^2 , a measure of the high-frequency components at a pixel t :

$$\sigma_t^2 = \frac{1}{|\xi_t|} \sum_{s \in \xi_t} (z_s - \bar{z}_t)^2,$$

where ξ_t is the set of pixels including pixel t and its neighborhood pixels, $|\xi_t|$ is the number of pixels in ξ_t , and z_s is the intensity of images between 0 and 255

at pixel s . The local variance based classifier (LVC), denoted by $\gamma(c)$, it assigns pixel z_t to focused regions if $\log(\sigma_t^2) \geq c$, otherwise, assigns it to defocused regions.

We estimate the optimal ROC curve of the LVC in segmenting an DOF image. Here, the optimal ROC curve implies that the curve plots the highest TPRs that the classifier can achieve at given FPRs. As stated in the introduction, the optimal ROC curve should be concave and we estimate concave ROC curve from the observed data.

Figure 3 (a) and (b) plots the flower image to be segmented into focused and defocused regions and its hand-labeled true segments. We obtained empirical FPRs and TPRs by moving the threshold value c from 1 to 11 by 0.1. We apply three estimators, empirical ROC curve, NPMLE, and modified NPMLE to estimating the ROC curve. The results are plotted in Figure 4.

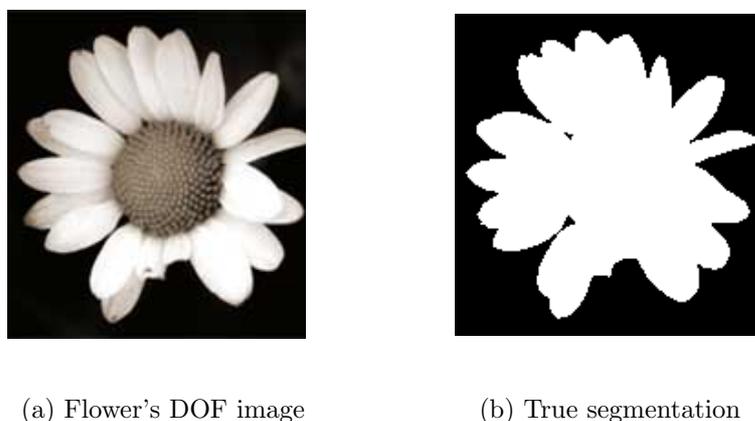


FIGURE 3. Flower images and its manually segmentation results.

The results show that the empirical ROC curve is not concave and the lower concave envelope used to modify the NPMLE becomes a line connecting $(0, 0)$ and $(1, 1)$. We find the modified NPMLE among mixtures of NPMLE and the lower concave envelope to minimize the \mathcal{L}_1 distance from the empirical ROC curve. The estimators are plotted in Figure 4, where we further notice that the NPMLE is close to, but not equal to, the convex hull of the empirical ROC curve.

6.2. Hemophilia A carrier

The second example is borrowed from Johnson and Wichern [9]. The data is related to detecting potential hemophilia A carriers using two blood sample

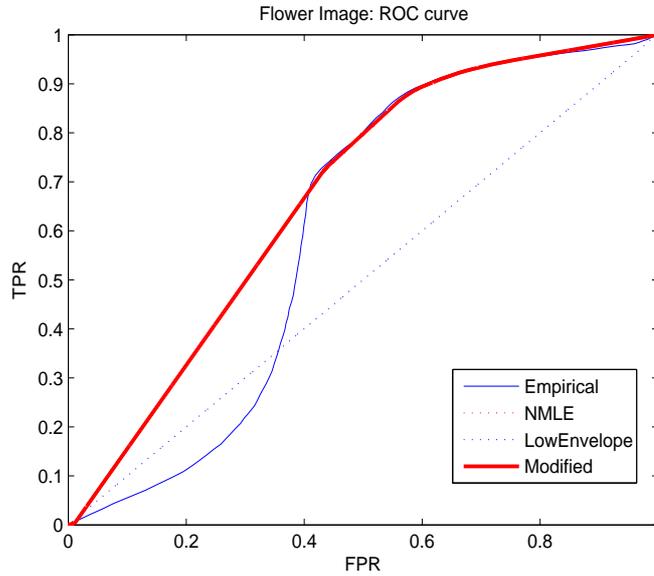


FIGURE 4. Estimated ROC curves – the empirical ROC curve, the NPMLE, and the modified NPMLE– of LVC (local variance classifier) to segment the flower image.

measurements of two groups of women, normal subject group (Group 1 : 30 women) and hemophilia A carrier group (Group 2: 45 women). The blood samples were analyzed and AHF activity and AHF-like antigen measurements were taken. We consider several linear classifiers of two variables,

$$X_1 = \log_{10}(\text{AHF activity}), \quad \text{and} \quad X_2 = \log_{10}(\text{AHF-like antigen}),$$

to detect potential hemophilia A carriers: $L_1 = -X_1$, $L_2 = X_2$, $L_3 = \text{PCA}_1 = -(X_1 + X_2)/\sqrt{2}$, $L_4 = \text{PCA}_2 = -(X_1 - X_2)/\sqrt{2}$, and $L_5 = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$, where PCA_1 and PCA_2 are the first and the second principal scores in the principal component analysis (PCA), and $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$ are the regression estimates from the fit of the logistic regression model between group indicator Y (“ $Y = 0$ ” indicates Group 1 and “ $Y = 1$ ” indicates Group 2) and covariates X_1 and X_2 . Here, L_1 is based on the negative AHF activity due to the negative association between AHF activity and hemophilia. Figure 5 plots the modified NPMLE of concave ROC curves of the aforementioned 5 methods. It shows the ROC curve of L_5 has the highest TPR at any given FPR. However, it is not fair to compare L_5 to other methods since it uses observed data to estimate the parameters in the linear classifier. L_4 , the linear classifier based on the second principal score perform almost as good as the logistic classifier L_5 and outperforms all other classifiers. The second principal component is the average

between $-X_1$ and X_2 both of which are positively associated with the indicator of the hemophilia A carrier.

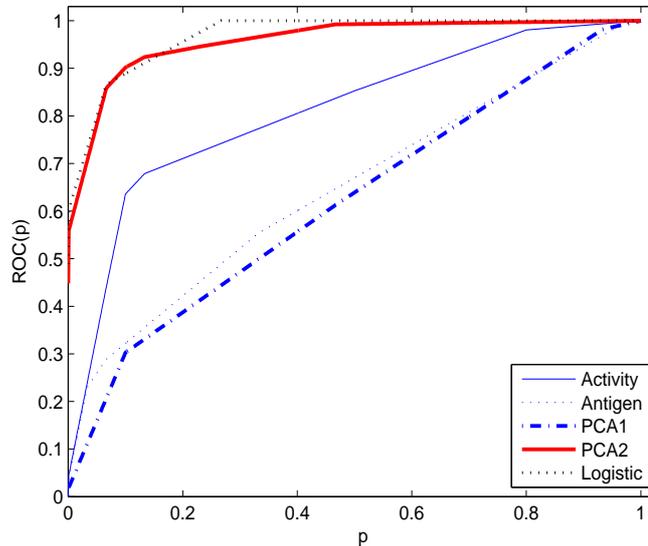


FIGURE 5. Estimated ROC curves of the given five linear classifiers.

7. Conclusion

The concave ROC curve plays an important role in searching for optimal classifiers or comparing the performance of different classifiers. In this paper, we study the NPMLE of a concave ROC curve and find that it is close to the convex hull of an empirical ROC curve. Thus, it is positively biased but has smaller variance. We introduce a modification of the NPMLE to reduce the bias by minimizing the uniform distance from the empirical ROC curve, which is an unbiased estimator, among a given class of concave ROC curves. We numerically show that the modified NPMLE has larger variance than the NPMLE, but smaller bias and MSE than the NPMLE. We finally apply the modified NPMLE to estimating the optimal ROC curve of the LVC in segmenting a flower image into focused and defocused regions and to finding a diagnostic tool with multiple tests for detection of hemophilia A carrier.

References

- [1] S. Boyd, S. Kim, L. Vandenberghe, and A. Hassibi, *A tutorial on geometric programming*, *Optim. Eng.* **8** (2007), no. 1, 67–127.

- [2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.
- [3] G. Campbell and M. V. Ratnaparkhi, *An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis*, *Communication in Statistics* **22** (1993), 1681–1697.
- [4] D. D. Dorfman, K. S. Berbaum, C. E. Metz, R. V. Length, J. A. Hanlet, and H. A. Dagga, *Proper receiver operating characteristic analysis: The bigamma model*, *Academic Radiology* **4** (1996), 138–140.
- [5] C. Feltz and R. Dykstra, *Maximum likelihood estimation of the survival functions of N stochastically ordered random variables*, *J. Amer. Statist. Assoc.* **80** (1985), no. 392, 1012–1019.
- [6] P. A. Flach and S. Wu, *Repairing concavities in ROC curves*, *Proceedings of 2003 UK Workshop on Computational Intelligence*, 38–44, 2003.
- [7] M. Grant, S. Boyd, and Y. Ye, *CVX: Matlab software for disciplined convex programming*, 2005; Available from <http://www.stanford.edu/~boyd/cvx/>.
- [8] S. Johansen, *The product limit estimator as maximum likelihood estimator*, *Scand. J. Statist.* **5** (1978), no. 4, 195–199.
- [9] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Third edition. Prentice Hall, Inc., Englewood Cliffs, NJ, 1992.
- [10] E. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, *J. Amer. Statist. Assoc.* **53** (1958), 457–481.
- [11] J. Lim, S.-J. Kim, and X. Wang, *Estimation of stochastically ordered survival functions via geometric programming*, To appear in *Journal of Computational and Graphical Statistics*; Available from http://www.stat.snu.ac.kr/johanlim/sorder_v3.pdf.
- [12] J. Lim, X. Wang, and W. Choi, *Matimum likelihood estimation of ordered multinomial parameters by geometric programing*, *Compuational Statistics and Data Analysis* **53** (2009), 889–893.
- [13] C. Llyod, *Regression models for convex ROC curves*, *Biometrics* **56** (2000), 862–867.
- [14] ———, *Estimation of a convex ROC curve*, *Statist. Probab. Lett.* **59** (2002), no. 1, 99–111.
- [15] J. Löfberg, *YALMIP. Yet Another LMI Parser*, Version 2.4, 2003; Available from <http://control.ee.ethz.ch/~joloef/yalmip.php>.
- [16] M. W. McIntosh and M. S. Pepe, *Combining several screening tests: optimality of the risk score*, *Biometrics* **58** (2002), no. 3, 657–664.
- [17] C. E. Metz and X. Pan, *“Proper” binormal ROC curves: theory and maximum-likelihood estimation*, *J. Math. Psych.* **43** (1999), no. 1, 1–33.
- [18] A. Mutapcic, K. Koh, S. Kim, and S. Boyd, *ggplab: A Matlab toolbox for geometric programming*, 2006; Available from <http://www.stanford.edu/~boyd/ggplab/>.
- [19] Y. Nesterov and A. Nemirovsky, *Interior-Point Polynomial Methods in Convex Programming*, volume 13 of *Studies in Applied Mathematics*, SIAM Philadelphia, PA, 1994.
- [20] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research. Springer-Verlag, New York, 1999.
- [21] X. Pan and C. E. Metz, *The “proper” binomial model: Parametric receiver operating characteristic curve estimation with degenerate data*, *Academic Radiology* **4** (1997), 380–389.
- [22] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, Oxford, 2003.
- [23] F. Preparata and S. J. Hong, *Convex hulls of finite sets of points in two and three dimensions*, *Comm. ACM* **20** (1977), no. 2, 87–93.
- [24] J. Z. Wang, J. Li, R. M. Gray, and G. Wiederhold, *Unsupervised multiresolution segmentation for images with low depth of field*, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **23** (2001), 85–90.

- [25] C. Yim and A. C. Bovik, *Multiresolution 3-D range segmentation using focused cues*, IEEE Transaction on Image Processing **7** (1998), 1283–1299.

KYEONG EUN LEE
DEPARTMENT OF STATISTICS
KYUNGPOOK NATIONAL UNIVERSITY
DAEGU 702-701, KOREA
E-mail address: `artlee@knu.ac.kr`

JOHAN LIM
DEPARTMENT OF STATISTICS
SEOUL NATIONAL UNIVERSITY
SEOUL 151-742, KOREA
E-mail address: `johanlim@snu.ac.kr`