REVSTAT – Statistical Journal
Volume 16, Number 1, January 2018, 23–43

FINDING THE OPTIMAL THRESHOLD OF A PARAMETRIC ROC CURVE UNDER A CONTINUOUS DIAGNOSTIC MEASUREMENT

Authors: Yi-Ting Hwang
– Department of Statistics, National Taipei University, Taipei, Taiwan
hwangyt@gm.ntpu.edu.tw

Yu-Han Hung
– Department of Statistics, National Taipei University, Taipei, Taiwan
lalamomok0914@hotmail.com

Chun Chao Wang
– Department of Statistics, National Taipei University, Taipei, Taiwan
ccw@gm.ntpu.edu.tw

Harn-Jing Terng
– Advpharma, Inc., New Taipei city, Taiwan
ternghj@advpharma.com.tw

Received: March 2015 Revised: July 2016 Accepted: July 2016

Abstract:

- The accuracy of a binary diagnostic test can easily be assessed by comparing the sensitivity and specificity with the status of respondents. When the result of a diagnostic test is continuous, the assessment of accuracy depends on a specified threshold. The receiver operating characteristic (ROC) curve, which includes all possible combinations of sensitivity and specificity, provides an appropriate measure for evaluating the overall accuracy of the diagnostic test. Nevertheless, in practice, a cutoff value is still required to make easier its clinical usage easier. The determination of a proper cutoff value depends on how important the practitioner views the specificity and sensitivity. Given particular values of specificity and sensitivity, this paper derives the optimal cutoff value under two parametric assumptions on the outcomes of the diagnostic test. Because the optimal cutoff value does not have a closed form, the numerical results are tabulated for some parameter settings to find the optimal cutoff value. Finally, real data are employed to illustrate the use of the proposed method.

Key-Words:

- *bilogistic model; binormal model; optimal threshold; sensitivity; specificity.*

AMS Subject Classification:

- 62C05.

1. INTRODUCTION

A diagnostic test that results in a continuous value is often evaluated using the receiver operating characteristic (ROC) curve. Let TP, FP, FN and TN denote the true positive decision, false positive decision, false negative decision and true negative decision, respectively. The following table provides 4 possible diagnostic test decisions:

True status	Test result	
	Positive	Negative
Case	TP	FN
Normal	FP	TN

Let $P[\text{TP}]$ be the probability that a true positive decision is made, and let $P[\text{TN}]$, $P[\text{FP}]$ and $P[\text{FN}]$ be defined similarly. The true positive rate (TPR) and the true negative rate (TNR) can be derived from $P[\text{TP}]$, $P[\text{TN}]$, $P[\text{FP}]$ and $P[\text{FN}]$ as

$$\text{TPR} = \frac{P[\text{TP}]}{P[D+]}, \tag{1.1}$$

$$\text{TNR} = \frac{P[\text{TN}]}{P[D-]}, \tag{1.2}$$

where $P[D+] = P[\text{TP}] + P[\text{FN}]$ denotes the prevalence of a disease and $P[D-] = P[\text{TN}] + P[\text{FP}] = 1 - P[D+]$.

A ROC curve is constructed from different values for the TPR and FPR. The determination of the TPR and FPR requires a cutoff value to classify the normal and diseased populations when the outcome is continuous. The ROC curve is then formed using TPRs and FPRs derived from all possible cutoff values. However, for practical use, the continuous outcome has to be dichotomized such that the investigator or practitioner can easily use it to discriminate the disease status. Nevertheless, the ROC curve does not provide direct information on how to determine such a cutoff value. It is thus important to find an optimal cutoff value (OCV) such that the probabilities of correct decisions are maximized.

Let S_D and S_N denote the outcome of the diagnostic measure for the disease group and the normal group, respectively, and let F_D and F_N denote the corresponding distribution functions. The ROC curve can be represented as

$$\text{ROC}(t) = \bar{F}_D(\bar{F}_N^{-1}(t)),$$

where $t \in (0, 1)$, $\bar{F}_D(t) = 1 - F_D(t)$ is the survival function of $F_D(t)$ and $\bar{F}_N(t)$ is

defined similarly. Because the FPR and TPR are functions of $\bar{F}_D$ and $\bar{F}_N$ as

$$\begin{aligned}\text{FPR}(c) &= P[S_N > c|N] = 1 - F_N(c) = \bar{F}_N(c),\\ \text{TPR}(c) &= P[S_D > c|D] = 1 - F_D(c) = \bar{F}_D(c),\end{aligned}$$

for a given cutoff $c \in (-\infty, \infty)$, the ROC curve can be represented in terms of the TPR and FPR.

To derive the OCV, an additional objective function is required. Three objectives have been discussed in the literature to find the OCV (Akobeng [1]; Kumar [5]). The first objective function is defined as the distance from the ROC curve to the point (0,1), that is,

$$C_1(c) = \sqrt{(1 - \text{TPR}(c))^2 + (\text{FPR}(c))^2} \tag{1.3}$$

and the OCV is the point at which $C_1(c)$ has the minimum. The second objective function proposed by Youden [9] is the vertical distance from the line of equality to the point on the ROC curve, which is

$$C_2(c) = \text{TPR}(c) + \text{TNR}(c) - 1, \tag{1.4}$$

and the OCV is the point that maximizes C_2. $C_2(c)$ is known as the Youden index. An alternative and equivalent representation of $C_2(c)$ is

$$|\text{TPR}(c) - (1 - \text{TNR}(c))|$$

expressed by Lee [6] and Krzanowski and Hand [4]. The third objective function is a weighted function of the probability of four diagnostic decisions, defined by Metz [8] as

$$C_3(c) = C_0 + C_{\text{TP}}P[\text{TP}] + C_{\text{TN}}P[\text{TN}] + C_{\text{FP}}P[\text{FP}] + C_{\text{FN}}P[\text{FN}], \tag{1.5}$$

where C_0 is the overhead cost, C_{TP} represents the average cost of the medical consequences of a true positive decision, and the remainder of the costs are defined similarly. Based on (1.1) and (1.2), expression (1.5) can be rewritten as

$$\begin{aligned}C_3(c) =& \{C_0 + C_{\text{FP}} \times P[D-] + C_{\text{FN}} \times P[D+]\}\\ &+ \{[C_{\text{FN}} - C_{\text{TP}}] \times P[D+]\} \times \text{TPR}(c)\\ &+ \{[C_{\text{TN}} - C_{\text{FP}}] \times P[D-]\} \times \text{TNR}(c)\end{aligned} \tag{1.6}$$

In particular, the first term on the right-hand side of (1.6) includes only the three costs and the prevalence, which do not depend on the decision of a diagnostic test. Because the determination of the OCV is not related to this term, it is neglected in the following discussion. Thus, in terms of (1.6), the best cutoff value is the one that minimizes C_3. The critical value occurs at

$$\frac{\partial \text{TPR(c)}}{\partial \text{TNR(c)}} = -\frac{(C_{\text{TN}} - C_{\text{FP}}) \times P[D-]}{(C_{\text{FN}} - C_{\text{TP}}) \times P[D+]},$$

which is the slope of a line of isoutility or the tangent line in the ROC space. Metz [8] concluded that the OCV on a ROC curve must be tangent to the highest line of isoutility that intersects with the ROC curve.

The OCV derived from the first and second objective functions is determined empirically (Kumar [5]). Under the binormal model and assuming that the slope of the tangent line to the ROC curve equals η, an explicit form for the OCV under $C_3(c)$ is derived and is referred to as P252 in Halperm *et al.* [3]. However, the third objective function uses not only the cost for each decision but also the prevalence of the disease. The latter can possibly be obtained empirically using the existing data, whereas the cost of the medical consequences is difficult to obtain. Thus, it is rarely used in the medical literature (Kumar [5]).

For a practitioner, sensitivity and specificity, which correspond to the TPR and TNR, are commonly used measures, and the importance of these two measures depends on the purpose of the diagnostic test. Thus, rather than the equal weight setting for the TPR and TNR as in (1.3) and (1.4), in this paper, we suggest using a more general objective function,

$$C(c) = \alpha \times \mathrm{TPR}(c) + \beta \times \mathrm{TNR}(c), \tag{1.7}$$

where $0 < \alpha, \beta < 1$ and $\alpha + \beta = 1$, to derive the OCV. The weight α can be regarded as the relative cost for an additional cost of classifying a TP compared to an additional cost of classifying a TN. Assuming the location and scale parametric assumption, the OCV can be then obtained under $C(c)$. In particular, when $\alpha = 0$, the objective function in (1.7) is the usual criterion for finding the OCV by minimizing the FPR or maximizing the specificity. Conversely, when $\beta = 0$, the objective function is the usual criterion for finding the OCV by maximizing the sensitivity. Section 2 describes the basic definition of the ROC curve and the derivation for the OCV. Section 3 presents the numerical results. Sections 4 and 5 provide a real application and discussions, respectively.

2. METHOD

Assume that F_D and F_N belong to a location and scale family. In other words, both distributions can be expressed by a standard form, say F, with different location and scale parameters. Let (μ_D, γ_D) and (μ_N, γ_N) denote the parameters for F_D and F_N, respectively. The FPR and TPR can be represented in terms of F as

$$\mathrm{TPR}(c) = P\Big[\frac{S_D - \mu_D}{\gamma_D} > \frac{c - \mu_D}{\gamma_D}\Big] = F\Big(\frac{\mu_D - c}{\gamma_D}\Big) \tag{2.1}$$

$$\mathrm{FPR}(c) = P\Big[\frac{S_N - \mu_N}{\gamma_N} > \frac{c - \mu_N}{\gamma_N}\Big] = F\Big(\frac{\mu_N - c}{\gamma_N}\Big). \tag{2.2}$$

Let t_p denote the critical value of F, i.e., $1 - F(t_p) = p$. Given $\mathrm{FPR}(c)$, the following relationship is obtained:

$$t_{\mathrm{FPR}} = F_N^{-1}(\mathrm{FPR}(c)) = -\frac{c - \mu_N}{\gamma_N},$$

and

$$c = \mu_N - \gamma_N \times t_{\mathrm{FPR}}. \tag{2.3}$$

Additionally, given $\mathrm{TPR}(c)$, we have

$$t_{\mathrm{TPR}} = F_D^{-1}(\mathrm{TPR}(c)) = -\frac{c - \mu_D}{\gamma_D},$$

and

$$c = \mu_D - \gamma_D \times t_{\mathrm{TPR}}. \tag{2.4}$$

Given FPR and TPR, (2.3) and (2.4) provide the relationship between two critical values as

$$t_{\mathrm{TPR}} = \frac{\mu_D - \mu_N}{\gamma_D} + \frac{\gamma_N}{\gamma_D} t_{\mathrm{FPR}} = a + b t_{\mathrm{FPR}}, \tag{2.5}$$

where $a = (\mu_D - \mu_N)/\gamma_D$ and $b = \gamma_N/\gamma_D$. From (2.5), a linear relationship exists between two critical values of F_D and F_N, where a is the intercept and b is the slope. Given $\mathrm{FPR}(c)$, the ROC curve can be represented as

$$\mathrm{ROC}(c) = P[S_D > c] = F\Big(\frac{\mu_D - c}{\gamma_D}\Big). \tag{2.6}$$

Substituting the value of c defined in (2.3) into (2.6) yields

$$\mathrm{ROC}(c) = P[S_D > c] = F\Big(\frac{\mu_D - \mu_N + \gamma_N \times t_{\mathrm{FPR}}}{\gamma_D}\Big) = F(a + b t_{\mathrm{FPR}}).$$

Under the location and scale family as defined in (2.1), (2.2) and (2.5), (1.7) becomes

$$C(c) = \alpha F\Big(a + b\Big(\frac{\mu_N - c}{\gamma_N}\Big)\Big) + \beta F\Big(\frac{c - \mu_N}{\gamma_N}\Big).$$

The OCV can then be determined by finding the critical value of $\frac{dC}{dc} = 0$, where

$$\frac{dC(c)}{dc} = \alpha f\Big(a + b\Big(\frac{\mu_N - c}{\gamma_N}\Big)\Big) \times \Big(-\frac{b}{\gamma_N}\Big) + \beta f\Big(\frac{c - \mu_N}{\gamma_N}\Big) \times \Big(\frac{1}{\gamma_N}\Big) \tag{2.7}$$

and $f(\cdot)$ is the density function of $F(\cdot)$. The following theorem discusses two location and scale families. The proof for Theorem 2.1 is provided in the Appendix, and the proof for Theorem 2.2 is similar.

Theorem 2.1. *Assume that $F(\cdot) = \Phi(\cdot)$ is a standard normal distribution function. To be consistent with the conventional notation, the scale parameters are denoted by σ_D and σ_N. Then,*

1. *When $b = 1$, we obtain*

$$OCV = \mu_N + \frac{a}{2}\sigma_N - \frac{\sigma_N}{a}\log\Big(\frac{1-\beta}{\beta}\Big). \tag{2.8}$$

2. *When $b \neq 1$, we obtain*

$$OCV = \frac{T \pm \sqrt{T^2 - 2(1-b^2)R/\sigma_N^2}}{(1-b^2)/\sigma_N^2}, \tag{2.9}$$

where

$$R = \frac{\mu_N^2 - (a\sigma_N + b\mu_N)^2}{2\sigma_N^2} + \log(\frac{\alpha b}{\beta}), \tag{2.10}$$

$$T = \frac{\mu_N - ab\sigma_N - b^2\mu_N}{\sigma_N^2}, \tag{2.11}$$

and R and T have to satisfy the condition $T^2 - 2(1-b^2)R/\sigma_N^2 > 0$.

Theorem 2.2. *Assume that $F(\cdot)$ is a standard logistic distribution function, i.e.,*

$$F(x) = [1 + \exp(-x)]^{-1}.$$

Then,

1. *When $b = 1$, we obtain a closed form for the OCV as*

$$OCV = -\sigma_D \log(q), \tag{2.12}$$

where

$$q = \frac{(\alpha-\beta) \pm \sqrt{\alpha\beta(\exp(a) + \exp(-a) - 2)}}{\Big[\beta\exp\Big(-\frac{\mu_N}{\gamma_N}\Big) - \alpha\exp\Big(-\frac{\mu_D}{\gamma_N}\Big)\Big]\exp\Big(\frac{\mu_D+\mu_N}{\gamma_N}\Big)}. \tag{2.13}$$

2. *When $b \neq 1$, the OCV is found numerically by solving the following nonlinear equation*

$$\begin{aligned}&\frac{\beta}{\gamma_N}\exp\Big(\frac{\mu_N}{\gamma_N}\Big)k^{-\frac{1}{\gamma_N}}\Big(\exp\Big(\frac{b\mu_N + a\gamma_N}{\gamma_N}\Big)k^{-\frac{1}{\gamma_D}} + 1\Big)^2 \\ &\quad = \frac{\alpha b}{\gamma_N}\exp\Big(\frac{(b\mu_N + a\gamma_N}{\sigma_N}\Big)k^{-\frac{1}{\gamma_D}}\Big(\exp\Big(\frac{\mu_N}{\gamma_N}\Big)k^{-\frac{1}{\gamma_N}} + 1\Big)^2,\end{aligned} \tag{2.14}$$

where $k = e^c$.

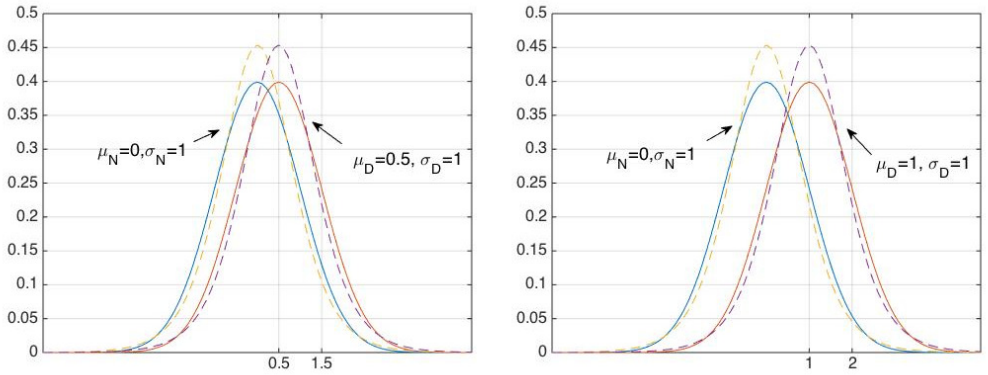
2.1. Relationship between the objective function and cutoff values

As c increases, the TPR decreases and the TNR increases. Because we assume that a case has a higher test value, the relative change in the TPR with respect to c is more rapid than that in the TNR. Furthermore, as expected, increasing μ_D means a smaller overlapping area in the densities for the normal and diseased populations and results in an increase in the TPR. When μ_D is fixed, the influence of σ_D on the TPR depends on c. When c is closer to μ_D, increasing σ_D reduces the TPR.

To understand how the parametric assumption influences the relationship between the objective function and the OCV, the basic features for the binormal and bilogistic models are discussed in the following. The common feature is that both distributions are symmetric about the location parameter. Nevertheless, the scale parameter in the normal distribution is the standard deviation, whereas the scale parameter in the logistic distribution is equal to the standard deviation times $\sqrt{3}/\pi$. Finally, the kurtosis of the normal distribution equals 3, whereas that of the logistic distribution equals 4.2.

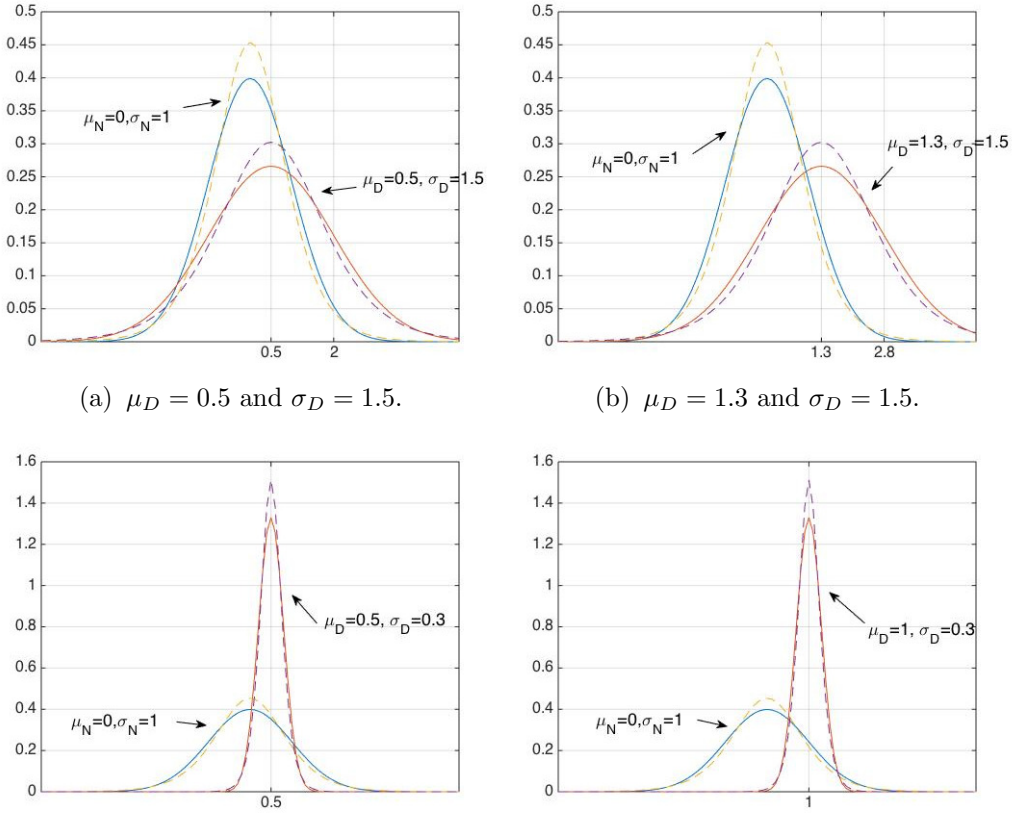
Assuming that $\mu_N = 0$ and $\sigma_N = 1$, Figures 1(a)–1(b) display the normal and logistic density functions for the normal and diseased populations when $b = 1$, and Figures 2(a)–2(d) display the situations when $b \neq 1$, where the solid line represents the normal distribution and the dashed line represents the logistic distribution and the left curve is for the control population and the right curve is for the diseased population. Under the same settings of μ_D and σ_D, the tail probability for the logistic distribution is slightly larger than that for the normal distribution. Furthermore, the mode of the logistic distribution is higher than that of the normal distribution because it has a larger kurtosis. These distinct features influence the TPR and TNR as shown in Table 1. Furthermore, due to a more concentrated feature for the logistic distribution, under the considered situation, the TNR of the logistic distribution is slightly larger than that of the normal distribution when c is closer to the μ_N, whereas for larger c, the TNR of the logistic distribution is slightly smaller. Thus, under the assumption that $\mu_N < \mu_D$, to have a higher TPR, the cutoff value for the logistic distribution is smaller than that for the normal distribution. In contrast, when investigating the TNR, the cutoff values for the logistic distribution might not be smaller.

The proposed objective function is a weighted function of the TPR and TNR. Figures 3(a)–3(b) show the relationship between the objective function C and the cutoff value c for various βs assuming that $\mu_N = 0$, $\sigma_N = 1$ and $\mu_D = 1$, $\sigma_D = 1$. For the binormal assumption, Figure 3(a) shows that when $\beta = 0.5$ and OCV=0.5, we obtain $C(\text{OCV}) = 0.6915$. When $\beta = 0.7$, that is, the specificity is more important than the sensitivity, we obtain OCV=1.3473 and $C(\text{OCV}) = 0.7470$.



(a) $\mu_D = 0.5$ and $\sigma_D = 1$. (b) $\mu_D = 1$ and $\sigma_D = 1$.

Figure 1: The probability density functions for normal distribution and logistic distributions for $\mu_N = 0$, $\sigma_N = 1$ and $b = 1$, where the solid line represents the normal curve and the dashed line represents the logistic curve.



(a) $\mu_D = 0.5$ and $\sigma_D = 1.5$. (b) $\mu_D = 1.3$ and $\sigma_D = 1.5$.

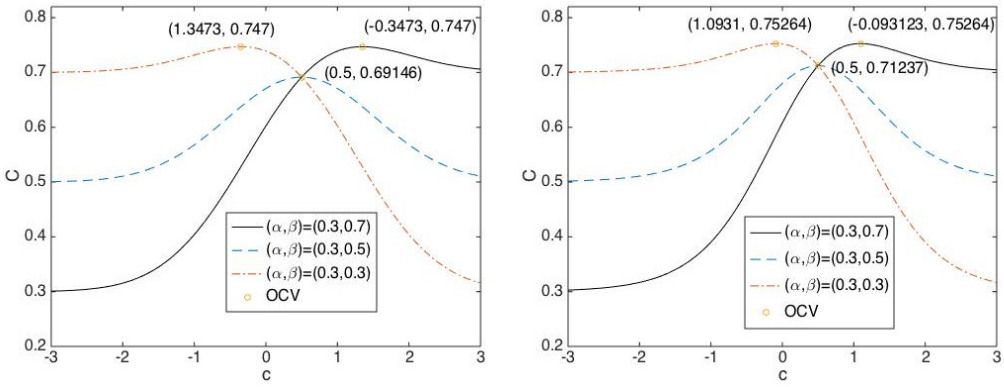
(c) $\mu_D = 0.5$ and $\sigma_D = 0.3$. (d) $\mu_D = 1$ and $\sigma_D = 0.3$.

Figure 2: The probability density functions for the normal distribution and logistic distribution for $\mu_N = 0$, $\sigma_N = 1$ and $b \neq 1$, where the solid line represents the normal curve and the dashed line represents the logistic curve.

Table 1: TPR and TNR under $c = 0.5, 1.5, 2$ for the binormal model and bilogistic model assuming $\mu_N = 0$ and $\sigma_N = 1$.

μ_D	σ_D	c	Normal distribution		Logistic distribution	
			TPR	TNR	TPR	TNR
0.5	1	0.5	0.5000	0.6915	0.5000	0.7124
		2	0.0668	0.9772	0.0618	0.9741
1	1	0.5	0.6915	0.6915	0.7124	0.7124
		2	0.1587	0.9772	0.1402	0.9741
0.5	1.5	0.5	0.5000	0.6915	0.5000	0.7124
		1.5	0.2525	0.9331	0.2298	0.9382
1	1.5	0.5	0.6306	0.6915	0.6467	0.7124
		1.5	0.3694	0.9332	0.3533	0.9382
0.5	0.3	0.5	0.5000	0.6915	0.5000	0.7124
		1.5	0.0004	0.9332	0.0024	0.9382
1	0.3	0.5	0.9522	0.6915	0.9536	0.7124
		1.5	0.0478	0.9332	0.0464	0.9382

Conversely, when $\beta = 0.3$, that is, the sensitivity is more important than the specificity, we obtain OCV=-0.3473 and $C(\text{OCV}) = 0.7470$. Figure 3(b) shows a similar pattern for when the bilogistic model is considered, but $C(\text{OCV})$ is slightly larger and the OCV is moving towards small values. This result arises from a larger kurtosis for the logistic distribution.



(a) Binormal model. (b) Bilogistic model.

Figure 3: Relationship between cutoff values and C under the binormal model and bilogistic model under various combinations of (α, β), where $\circ$ indicates the point at (OCV, C(OCV)).

2.2. Special cases

Depending on the purpose of the test, the investigator might be more interested in the specificity as long as the sensitivity reaches a specific limit, or vice versa. That is, an investigator might want to have a diagnostic test in which the sensitivity is at least larger than a pre-specified value L, where $0 < L < 1$. Then, the OCV is obtained by maximizing the specificity under the constraint that the sensitivity is larger than L, i.e., $\text{TPR} \geq L$. Likewise, the OCV can be obtained by maximizing the sensitivity under the constraint that the specificity is larger than L, i.e., $\text{TNR} \geq L$. The following derives the boundary for the TPR and TNR under the binormal and bilogistic models. The following proofs can be obtained in a straightforward manner.

Theorem 2.3. *Assume that $F(\cdot)$ is a standard normal distribution function and that $L > 0$ is a pre-specified constant. Then,*

1. *When $L \leq$ TPR, upper bounds of c and the TNR are*

$$c \leq \mu_D - \sigma_N \Phi^{-1}(L),$$
$$\text{TNR} \leq \Phi\Big(\frac{\mu_D - \mu_N - \sigma_N \Phi^{-1}(L)}{\sigma_N}\Big).$$

Thus, the OCV equals $\mu_D - \sigma_N \Phi^{-1}(L)$.

2. *When $L \leq$ TNR, a lower bound of c and an upper bound of the TNR are given as*

$$c \geq \mu_N - \sigma_N \Phi^{-1}(1 - L),$$
$$\text{TNR} \leq \Phi\Big(\frac{\mu_D - \mu_N + \sigma_N \Phi^{-1}(1 - L)}{\sigma_N}\Big).$$

Thus, the OCV equals $\mu_N - \sigma_N \Phi^{-1}(1 - L)$.

Theorem 2.4. *Assume that $F(\cdot)$ is a bilogistic distribution function and that $L > 0$ is a pre-specified constant. Then,*

1. *When $L \leq$ TPR, upper bounds of c and the TNR are*

$$c \leq \mu_D - \gamma_N \log(\frac{L}{1-L}),$$
$$\text{TNR} \leq \frac{1}{1 + exp\Big(\frac{\mu_N - \mu_D + \gamma_N \log\left(\frac{L}{1-L}\right)}{\gamma_N}\Big)}.$$

Thus, the OCV equals $\mu_D - \gamma_N \log(\frac{L}{1-L})$.

2. *When $L \leq$ TNR, a lower bound of c and an upper bound of the TNR are given as*

$$c \geq \mu_N - \gamma_N \log\Big(\frac{1-L}{L}\Big),$$
$$\text{TNR} \leq \frac{\exp(\frac{\mu_D-\mu_N+\gamma_N\log(\frac{1-L}{L})}{\gamma_N})}{1+\exp\Big(\frac{\mu_D-\mu_N+\gamma_N\log(\frac{1-L}{L})}{\sigma_N}\Big)}.$$

Thus, the OCV equals $\mu_N - \gamma_N \log\Big(\frac{1-L}{L}\Big)$.

3. NUMERICAL RESULTS

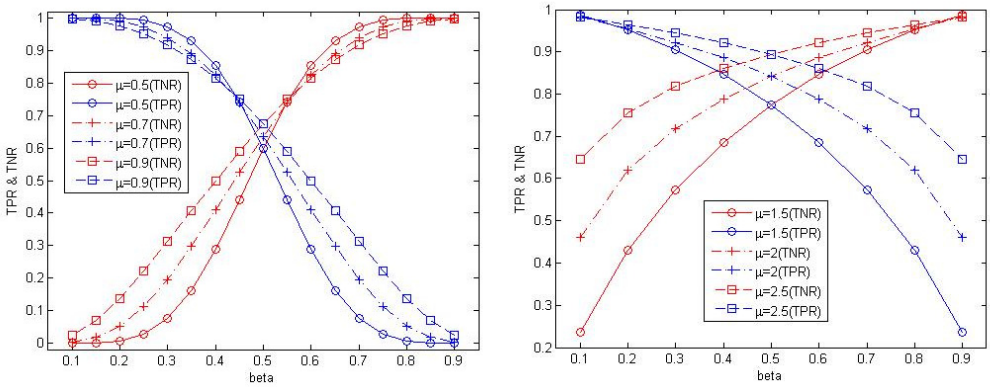
Based on the objective function defined in (1.6), Section 2 derives the OCV under the binormal and bilogistic models. When the binormal model is assumed, the OCV can be obtained explicitly, whereas under the bilogistic model, the OCV can be obtained explicitly only when b=1. The following discusses the OCV, TPR, and TNR under various settings for β and the location and scale parameters.

For simplicity, the standard normal distribution is assumed for the control population, i.e., $\mu_N = 0$ and $\sigma_N = 1$. Because the formula for determining the OCV varies with b, the following discussion considers $b = 1$ and $b \neq 1$ separately. For each scenario, the parameter setting is classified into two situations. The first scenario considers different values of μ_D given σ_D. The second scenario considers different values of σ_D given μ_D. Furthermore, the settings for μ_D and σ_D are discussed according to the effect size ES $= \mu_D/\sigma_D$. Additionally, μ_D is assumed to be larger than μ_N. Moreover, because $\beta = 0$ and $\beta = 1$ correspond to special cases discussed in Section 2.2, the numerical results only consider $0.1 \leq \beta \leq 0.9$. Similar results for the bilogistic model are given in the Supplement.

3.1. Situation I when σ_D is fixed and μ_D is varied

The first situation discusses the numerical results when σ_D is fixed and ES is varied. For ES < 1, μ_D equals 0.5, 0.7 and 0.9, whereas for $1 <$ ES, μ_D equals 1.5, 2 and 2.5. Figures 4(a)–4(b) display the relationship between TPR and TNR with respect to β when μ_D is varied and $\sigma_D = 1$. When β increases, the investigator is more interested in the TNR. As expected, the TNR increases while the TPR decreases. Increasing μ_D means that the difference in the testing result between two groups becomes more evident. Furthermore, for a fixed β

and σ_D, the OCV is a function of μ_D, as given in (2.8). Thus, as μ_D increases, the OCV increases, which corresponds to an increase in the TNR and a decrease in the TPR. Furthermore, due to a symmetric property, the OCV is located at TPR=TNR when $\beta = 0.5$. Table 2 presents the OCV, TPR and TPR for each scenario.



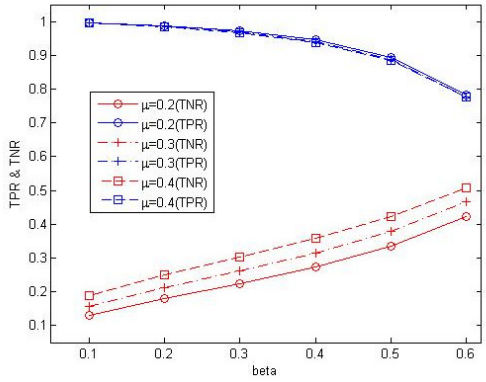
(a) ES < 1. (b) $1 <$ ES.

Figure 4: TNR and TPR at the OCV for various combinations of μ_D, β and ES under the binormal model and $b = 1$.

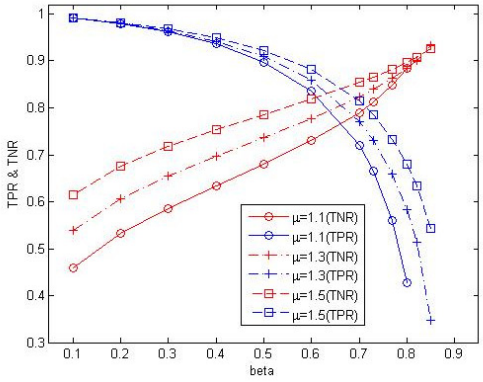
Table 2: Numerical results for TNR, TPR and OCV under the binormal model with various μ_Ds and $\sigma_D = 1$.

ES	μ_D	σ_D	Measures	β								
				0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.5	0.5	1	OCV	−4.1444	−2.5226	−1.4446	−0.5609	0.2500	1.0609	1.9446	3.0226	4.6444
			TPR	1.0000	0.9987	0.9741	0.8556	0.5987	0.2874	0.0743	0.0058	0.0000
			TNR	0.0000	0.0058	0.0743	0.2874	0.5987	0.8556	0.9741	0.9987	1.0000
0.7	0.7	1	OCV	−2.7889	−1.6304	−0.8604	−0.2292	0.3500	0.9292	1.5604	2.3304	3.4889
			TPR	0.9998	0.9901	0.9407	0.8236	0.6368	0.4093	0.1948	0.0515	0.0026
			TNR	0.0026	0.0515	0.1948	0.4093	0.6368	0.8236	0.9407	0.9901	0.9998
0.9	0.9	1	OCV	−1.9914	−1.0903	−0.4914	−0.0005	0.4500	0.9005	1.3914	1.9903	2.8914
			TPR	0.9981	0.9767	0.9180	0.8161	0.6736	0.4998	0.3116	0.1378	0.0232
			TNR	0.0232	0.1378	0.3116	0.4998	0.6736	0.8161	0.9180	0.9767	0.9981
1.5	1.5	1	OCV	−0.7148	−0.1742	0.1851	0.4797	0.7500	1.0203	1.3149	1.6742	2.2148
			TPR	0.9866	0.9530	0.9057	0.8462	0.7734	0.6843	0.5734	0.4309	0.2374
			TNR	0.2374	0.4309	0.5734	0.6843	0.7734	0.8462	0.9057	0.9530	0.9866
2	2	1	OCV	−0.0986	0.3069	0.5764	0.7973	1.0000	1.2027	1.4236	1.6931	2.0986
			TPR	0.9821	0.9548	0.9227	0.8855	0.8413	0.7874	0.7178	0.6205	0.4607
			TNR	0.4607	0.6205	0.7178	0.7874	0.8413	0.8855	0.9227	0.9548	0.9821
2.5	2.5	1	OCV	0.3711	0.6955	0.9111	1.0878	1.2500	1.4122	1.5889	1.8045	2.1289
			TPR	0.9834	0.9644	0.9440	0.9211	0.8944	0.8617	0.8189	0.7566	0.6447
			TNR	0.6447	0.7566	0.8189	0.8617	0.8944	0.9211	0.9440	0.9644	0.9834

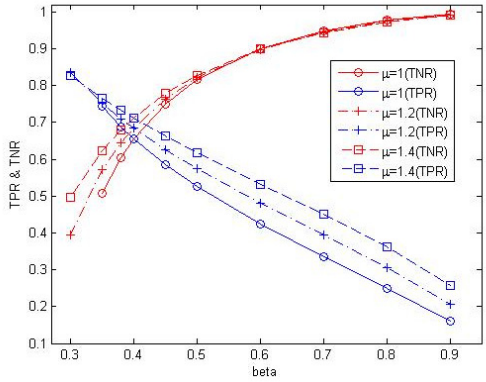
Figures 5(a)–5(d) display the TPR and TNR at the OCV when β is varied and $\sigma_D \neq 1$. The pattern for the TPR and TNR with respect to β is no longer symmetric. Similar to $\sigma_D = 1$, as β increases, the TPR decreases and the TNR increases. However, the relationship between the TPR and TNR depends on σ_D, ES and β. When ES < 1 and $\sigma_D = 0.5$, the TPR is always larger than the TNR regardless of β. This is because $\sigma_D = 0.5$ means that the result obtained from the diseased group is more homogeneous, and the diagnostic test has a higher ability to detect a case even if ES < 1. However, when ES < 1 and $\sigma_D = 1.5$, the TPR is larger than the TNR only if $\beta < 0.4$. Furthermore, when ES > 1, the TPR is larger than the TNR only for some βs.



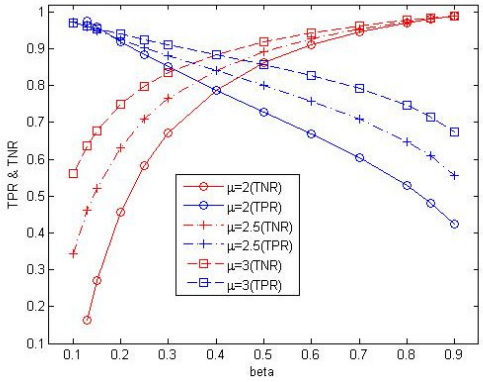
(a) ES < 1 & $\sigma_D = 0.5$.



(b) $1 <$ ES & $\sigma_D = 0.5$.



(c) ES < 1 & $\sigma_D = 1.5$.



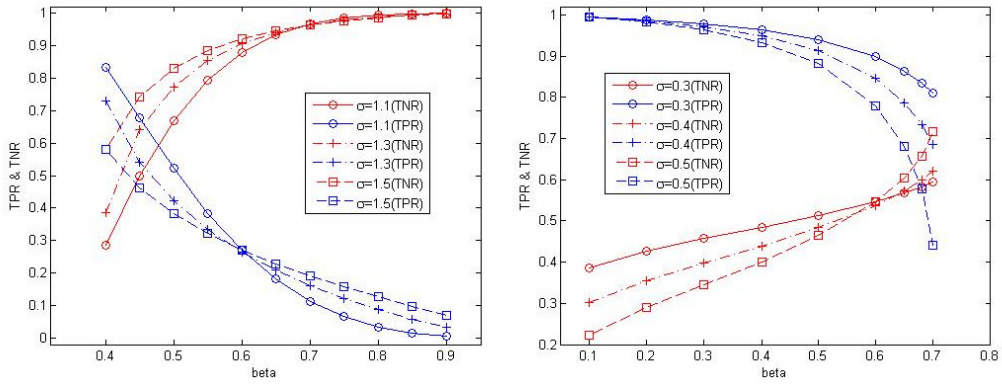
(d) $1 <$ ES & $\sigma_D = 1.5$.

Figure 5: TNR and TPR at the OCV when μ_D, σ_D, β and ES are varied, $b \neq 1$ and the binormal model are assumed.

3.2. Situation II when σ_D is varied and μ_D is fixed

Situation II provides numerical results for OCV, TPR and TNR when $\mu_D = 0.5$ and σ_D is varied. When $\mu_D = 0.5$, ES < 1 means that σ_D is larger than $\sigma_N = 1$, which means that it is easier to conclude a FN. Figure 6(a) shows the relationship between the TPR and TNR at the OCV with respect to β when σ_D is varied and ES < 1. The pattern of change for the TPR with respect to σ_D is related to β. When β increases, TPR expectedly decreases because β is the weight for the TNR. Nevertheless, when $0.5 < \beta$, the TPR becomes very small and slightly increases as σ_D increases. In addition, the TNR is large as long as $0.6 < \beta$, as listed in Figure 6(a).

When $\mu_D = 0.5$, $1 <$ ES means that σ_D is smaller than $\sigma_N = 1$, which indicates that it is easier to conclude a TP. Figure 6(b) displays the relationship between the TPR and TNR with respect to β when σ_D is varied and $1 <$ ES. Expectedly, as σ_D increases, the TPR decreases regardless of β. Unlike ES < 1, the relationship between the TNR and σ_D depends on β. When $\beta < 0.6$, the TNR decreases as σ_D increases, whereas when $0.6 < \beta$, the TNR increases as σ_D increases.



(a) ES < 1. (b) $1 <$ ES.

Figure 6: TNR and TPR at the OCV for various combinations of σ_D, β and ES under the binormal model and $\mu_D = 0.5$.

As β increases, the TNR is more important and results in a larger OCV. Table 3 demonstrates this trend. The impact of σ_D on the OCV is related to ES. When ES < 1, as σ_D increases, the OCV increases. Nevertheless, when ES > 1, the trend reverses.

Table 3: The relationship among OCV, TNR and TPR when the binormal model is assumed, $\mu_D = 0.5$ and σ_D is varied.

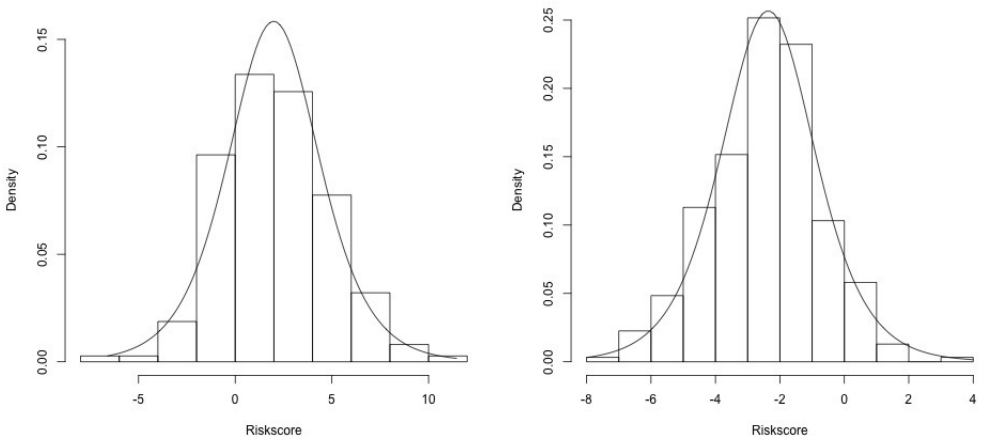
ES	μ_D	σ_D	Measures	β								
				0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.45	0.5	1.1	OCV	—	—	—	−0.5684	0.4400	1.1730	1.8288	2.5112	3.3878
			TPR	—	—	—	0.8343	0.5218	0.2703	0.1135	0.0337	0.0043
			TNR	—	—	—	0.2849	0.6700	0.8796	0.9663	0.9940	0.9996
0.38	0.5	1.3	OCV	—	—	—	−0.2929	0.7493	1.3147	1.7900	2.2693	2.8720
			TPR	—	—	—	0.7290	0.4239	0.2654	0.1605	0.0868	0.0340
			TNR	—	—	—	0.3848	0.7732	0.9057	0.9633	0.9884	0.9980
0.33	0.5	1.5	OCV	—	—	—	0.2000	0.9490	1.4109	1.8068	2.2097	2.7192
			TPR	—	—	—	0.5793	0.3824	0.2718	0.1918	0.1272	0.0695
			TNR	—	—	—	0.5793	0.8287	0.9209	0.9646	0.9864	0.9967
1.67	0.5	0.3	OCV	−0.2872	−0.1851	−0.1085	−0.0384	0.0344	0.1192	0.2368	—	—
			TPR	0.9957	0.9888	0.9787	0.9636	0.9397	0.8978	0.8098	—	—
			TNR	0.3870	0.4266	0.4568	0.4847	0.5137	0.5474	0.5936	—	—
1.25	0.5	0.4	OCV	−0.5196	−0.3711	−0.2583	−0.1532	−0.0417	0.0940	0.3072	—	—
			TPR	0.9946	0.9853	0.9710	0.9488	0.9122	0.8450	0.6851	—	—
			TNR	0.3017	0.3553	0.3981	0.4391	0.4833	0.5374	0.6207	—	—
1	0.5	0.5	OCV	−0.7609	−0.5570	−0.4001	−0.2518	−0.0904	0.1163	0.5753	—	—
			TPR	0.9942	0.9827	0.9641	0.9336	0.8812	0.7786	0.4401	—	—
			TNR	0.2234	0.2888	0.3445	0.4006	0.4640	0.5463	0.7175	—	—

— Numerical data are not available.

4. CASE STUDY

Early detection may improve the survival of patients with lung cancer. Chian *et al.* (2015) investigated peripheral blood mononuclear cell (PBMC)-derived gene expression signatures for their potential in the early detection of non-small cell lung cancer (NSCLC). PBMCs were obtained from 187 patients with NSCLC and from 310 non-cancer controls based on an age- and gender-matched case-control study. Controlling for gender, age and smoking status, 15 NSCLC-associated molecular markers were used to construct a risk score to distinguish subjects with lung cancer from controls. Detailed markers and the model construction are presented in Chian *et al.* (2016).

From the preventive perspective in health management, a higher sensitivity is preferred such that the disease can be detected earlier. Thus, β might be set to be smaller than 0.5. Nonetheless, cancer-specific clinicians often examine highly suspicious subjects. Thus, they may wish to have a higher specificity test. Figure 7 presents the histograms of the risk scores for the case and control groups for the PBMC data. The bilogistic model appears to be appropriate for these data. The maximum likelihood estimators of μ and γ are obtained for each group. The corresponding estimates of μ and γ for the case are 1.9911 and 1.5782 and those for the control are -2.3620 and 0.9739. Based on these estimates, the logistic density curves are plotted on top of the histogram in Figure 7.



(a) Case. (b) Control.

Figure 7: Histograms for risk scores for case and control groups for PBMC data, where the solid curve is the logistic density curve.

Under the bilogistic assumption, Table 4 lists the OCVs for β ranging from 0.1 to 0.9 for the risk score derived from the PBMC data. Figure 8 presents the corresponding TPR and TNR. For instance, when $\beta = 0.4$, the OCV equals -0.634. The test would expect to have equal chances at approximately 0.85 to identify a true positive or a true negative. Nevertheless, when $\beta = 0.6$, the test would have a higher chance to find a true negative.

Table 4: OCV for the PBMC data.

β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
OCV	-2.864	-1.565	-1.027	-0.634	-0.291	0.044	0.409	0.861	1.581

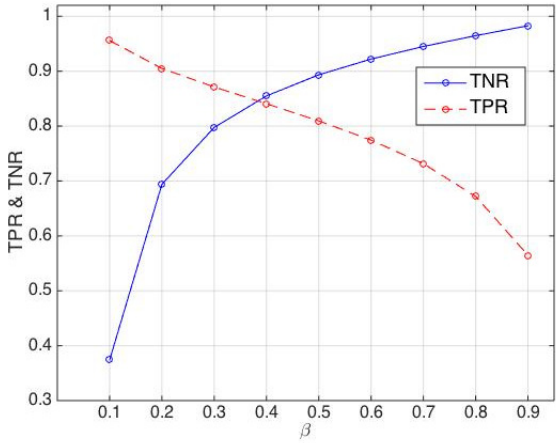


Figure 8: TPR and TNR under various βs for the PBMC data.

5. DISCUSSION AND CONCLUSION

The determination of the cutoff value is practically important. Because the ROC curve includes two important measures, TPR and TNR, to obtain the optimal operating point (OOP) or OCV, an additional objective function is required. One of two existing criteria can be regarded as the special case of the proposed criterion. The objective function C_3 requires information about the cost for the incorrect decision, which cannot be easily obtained. Furthermore, the OCV for this criterion is determined by setting the slope of the tangent line to the ROC curve to a pre-specified value (Halperm [3]). Because the slope is a function of the prevalence of the disease and costs, it is difficult to explain clinically (Kumar [5]).

The OCV is often obtained empirically (Kumar [5]). This paper derives the closed form for the OCV under the location and scale family. The binormal model is the most commonly used parametric assumption for the ROC curve. Under such an assumption, this paper provides exact formulas for the OCV. Furthermore, numerical results are presented under various scenarios. When $b = 1$, the TPR and TNR are related to the weight (β). In particular, increasing β means increasing the TNR. Nevertheless, when $b \neq 1$, regardless of β, the TNR might not be higher than 0.5. In particular, when the binomial model is violated, this paper provides another parametric choice, the bilogistic model. However, there is no closed form for the OCV. This paper provides a nonlinear equation for determining the OCV. In addition to discussing the OCV for the bilogistic model, the difference between these two parametric models is also addressed. The result of this paper can provide guidance for practitioners to choose the OCV.

Rather than choosing the OCV based on the sensitivity and specificity, Linnet *et al.* [7] suggested using the likelihood ratio

$$LR(c) = \frac{f(\frac{\mu_D - c}{\gamma_D})}{f(\frac{\mu_N - c}{\gamma_N})} \tag{5.1}$$

as an alternative for interpreting the test result. If (5.1) exceeds 1, then the relative frequency of the distribution of diseased individuals exceeds that of the normal individuals. In other words, given the index test result c, a respondent is more likely to have the disease. Their result can also be extended to the location and scale family.

APPENDIX: Proof of Theorem 2.1

Assume that F is the standard normal distribution function. To be consistent with the conventional notation, γ_D and γ_N are replaced by σ_D and σ_N, respectively. Therefore, (2.7) becomes

$$\text{(A.1)} \quad \frac{\partial C(c)}{\partial c} = \frac{-\alpha b}{\sqrt{2\pi}\sigma_N} \exp\Big(-\frac{[a + \frac{b(\mu_N - c)}{\sigma_N}]^2}{2}\Big) + \frac{\beta}{\sqrt{2\pi b}\sigma_N} \exp\Big(-\frac{b^2\big(c - \mu_N\big)^2}{2\sigma_N^2}\Big)$$

and set $\frac{\partial C(c)}{\partial c} = 0$ to obtain the OCV. An explicit formula for OCV can be determined and is dependent on b.

When $b = 1$, i.e., $\sigma_N^2 = \sigma_D^2$, the objective function and the corresponding derivative with respect to c are

$$\text{(A.2)} \qquad C = \alpha\Phi\Big(a + \frac{\mu_N - c}{\sigma_D}\Big) + \beta\Phi(\frac{c - \mu_N}{\sigma_D})$$

and

$$\text{(A.3)} \quad \frac{\partial C}{\partial c} = \frac{-\alpha}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{1}{2}\Big[a + \frac{\mu_N - c}{\sigma_D}\Big]^2\right) + \frac{\beta}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{1}{2}\Big(\frac{c - \mu_N}{\sigma_D}\Big)^2\right).$$

Let $\frac{\partial C}{\partial c} = 0$. We have

$$-\alpha b \exp\left(-\frac{[a\sigma_D + \mu_N - c]^2}{2\sigma_N^2}\right) + \beta \exp\left(-\frac{(c - \mu_N)^2}{2\sigma_D^2}\right) = 0,$$

which implies

$$\text{(A.4)} \qquad \log\Big(\frac{\alpha}{\beta}\Big) - \frac{[a\sigma_D + (\mu_N - c)]^2}{2\sigma_D^2} + \frac{(c - \mu_N)^2}{2\sigma_D^2} = 0.$$

After simplifying the preceding equation, we obtain

$$\frac{2(\mu_D - \mu_N)c + \mu_N^2 - \mu_D^2}{2\sigma_D^2} + \log(\frac{\alpha}{\beta}) = 0$$

and the OCV as given in (2.8).

When $b \neq 1$, the objective function and the corresponding derivative with respect to c are

$$C = \alpha\Phi\Big(a + b(\frac{\mu_N - c}{\sigma_N})\Big) + \beta\Phi\Big(\frac{c - \mu_N}{\sigma_N}\Big)$$

and

$$\frac{\partial C}{\partial c} = \frac{-\alpha b}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{1}{2}\left[a + b\left(\frac{\mu_N - c}{\sigma_N}\right)\right]^2\right) + \frac{\beta}{\sqrt{2\pi}\sigma_N} \exp\left(-\frac{1}{2}\left[\frac{c - \mu_N}{\sigma_N}\right]^2\right).$$

Let $\frac{\partial C}{\partial c} = 0$. We obtain

$$-\alpha b \exp\left(-\frac{[a\sigma_N + b(\mu_N - c)]^2}{2\sigma_N^2}\right) + \beta \exp\left[-\frac{(c-\mu_N)^2}{2\sigma_N^2}\right] = 0,$$

which implies

$$\log(\frac{\alpha b}{\beta}) - \frac{[a\sigma_N + b(\mu_N - c)]^2}{2\sigma_N^2} + \frac{(c-\mu_N)^2}{2\sigma_N^2} = 0. \tag{A.5}$$

Rearranging (A.5), we obtain

$$\frac{(1-b^2)}{2\sigma_N^2}c^2 - \frac{(\mu_N - ab\sigma_N - b^2\mu_N)}{\sigma_N^2}c + \frac{\mu_N^2 - (a\sigma_N + b\mu_N)^2}{2\sigma_N^2} + \log(\frac{\alpha b}{\beta}) = 0$$

and the OCV is equal to

$$c = \frac{T \pm \sqrt{T^2 - 2(1-b^2)R/\sigma_N^2}}{(1-b^2)/\sigma_N^2} \tag{A.6}$$

where R and T are defined in (2.10) and (2.11), respectively.

ACKNOWLEDGMENTS

This work has been supported by NSC102-2118-M-305-002-MY2 from the Ministry of Science and Technology, Taiwan. We also acknowledge the valuable suggestions from the referees.

REFERENCES

[1] Akobeng, A.K. (2007). Understanding diagnostic tests 3: receiver operating characteristic curves, *Acta Padiatrica*, **96**, 644–647.

[2] Chian, C.F.; Hwang, Y.T.; Terng, H.J. *et al.* (2016). Panels of tumor-derived RNA markers in peripheral blood of patients with non-small cell lung cancer: their dependence on age, gender and clinical stages. To appear in *Oncotarget*.

[3] Halperm, E.T.; Albert, M.; Krieger, A.M.; Metz, C.E. and Maidment, A.D. (1996). Comparison of receiver operating characteristic curves on the basis of optimal operating points, *Acad. Radiol.*, **3**, 245–253.

[4] Krzanowski, W.J. and Hand, D.J. (2009). *ROC Curves for Continuous Data*, CRC Press, New York.

[5] Kumar, R. and Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers, *Indian Pediatrics*, **48**, 277–287.

[6] Lee, C.T. (2006). A solution for the most basic optimization problem associated with an ROC curve, *Statistical Methods for Medical Research*, **15**, 571–584.

[7] Linnet, K.; Bossuyt, P.M.M.; Moons, K.G.M. and Reitsma, J.B. (2012). Quantifying the accuracy of a diagnostic test or marker, *Clinical Chemistry*, **58**(9), 1292–1301.

[8] Metz, C.E. (1978). Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, **8**(4), 283–298.

[9] Youden, W.J. (1950). Index for rating diagnostic tests, *Cancer*, **3**(1), 32–35.