



Published in final edited form as:

*Stat Med.* 2016 May 30; 35(12): 2031–2057. doi:10.1002/sim.6816.

## Equivalence of binormal likelihood-ratio and bi-chi-squared ROC curve models

**Stephen L. Hillis, Ph.D.\***

Departments of Radiology and Biostatistics, The University of Iowa, 3710 Medical Laboratories, 200 Hawkins Drive, Iowa City, IA 52242-1077, U.S.A.

### Abstract

A basic assumption for a meaningful diagnostic decision variable is that there is a monotone relationship between it and its likelihood ratio. This relationship, however, generally does not hold for a decision variable that results in a binormal ROC curve. As a result, receiver operating characteristic (ROC) curve estimation based on the assumption of a binormal ROC-curve model produces improper ROC curves that have “hooks,” are not concave over the entire domain, and cross the chance line. Although in practice this “improperness” is usually not noticeable, sometimes it is evident and problematic. To avoid this problem, Metz and Pan proposed basing ROC-curve estimation on the assumption of a binormal likelihood-ratio (binormal-LR) model, which states that the decision variable is an increasing transformation of the likelihood-ratio function of a random variable having normal conditional diseased and nondiseased distributions. However, their development is not easy to follow. I show that the binormal-LR model is equivalent to a bi-chi-squared model in the sense that the families of corresponding ROC curves are the same. The bi-chi-squared formulation provides an easier-to-follow development of the binormal-LR ROC curve and its properties in terms of well-known distributions.

### Keywords

Receiver operating characteristic (ROC) curve; diagnostic radiology; binormal likelihood ratio; bi-chi-squared; PROPROC

### 1. Introduction

Receiver operating characteristic (ROC) curve analysis is a well-established method for evaluating and comparing the performance of diagnostic tests [1, 2, 3]. Throughout I assume that the purpose of a diagnostic test is to classify subjects as diseased or not diseased, based on whether a decision variable (DV) exceeds a threshold value. The performance of a diagnostic test as a function of the threshold value is typically described by an ROC curve, which is a plot of sensitivity versus 1–specificity for all possible threshold values. Sensitivity and 1–specificity are often referred to as true positive fraction (tpf) and false positive

\* steve-hillis@uiowa.edu.

#### Disclaimer

The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

fraction (fpf), respectively. A commonly reported ROC-curve summary measure is the area under the ROC curve (AUC).

A parametric estimate of the ROC curve is often based on the assumption of a *binormal ROC curve*. A binormal ROC curve can be defined either in terms of its functional form or in terms of the distribution of a DV: (1) An ROC curve that plots as a straight line in probit space (i.e., a plot of  $\Phi^{-1}$  (tpf) vs.  $\Phi^{-1}$  (fpf)) is a binormal ROC curve with *binormal parameters*  $a$  and  $b$ , where  $a$  is the intercept and  $b$  is the slope. (2) Equivalently, a binormal ROC curve corresponds to a DV having a *binormal distribution*, i.e., the DV is normally distributed for nondiseased subjects as well as for diseased subjects. These two definitions are related by the equations  $a = (\mu_2 - \mu_1)/\sigma_2$  and  $b = \sigma_1/\sigma_2$ , where  $(\mu_2, \sigma_2^2)$  and  $(\mu_1, \sigma_1^2)$  denote the mean and variance parameter pairs for the nondiseased and diseased distributions for a binormal DV. In this paper I primarily utilize the second definition of the binormal ROC curve. Because an ROC curve is invariate to an increasing transformation of the DV, more generally a binormal ROC curve corresponds to a DV that follows the *binormal model*, which states that there exists an increasing transformation such that the transformed DV has a binormal distribution. For large samples the assumption of a binormal ROC curve has been shown to perform well for DV distributions that can vary greatly from the binormal distribution [4, 5, 6, 7].

For example, consider a study where a radiologist is asked to assign likelihood-of-disease confidence levels to images using a discrete five-level ordinal integer scale (e.g., 1 = “definitely not diseased”, . . . , 5 = “definitely diseased”); for this situation the investigator might assume that the observed ordinal ratings represent the binning of values of a latent (i.e., unobserved) continuous DV that represents the reader's likelihood-of-disease perception. A binormal ROC curve is appropriate if we assume that the latent DV has a binormal distribution or is an increasing transformation of a binormal DV.

In most situations a meaningful DV should be an increasing transformation of the likelihood ratio (likelihood of being diseased divided by likelihood of not being diseased) [2].

According to Egan [8, pp 19, 37], a DV having this property and its corresponding ROC curve are said to be *proper*; otherwise, they are said to be *improper*. A proper ROC curve is concave (down) everywhere and never crosses the chance line [2]. For a given test result, a proper DV provides the optimal criterion for classifying subjects in the sense that the resulting proper ROC curve is uniformly above all other ROC curves based on the test result [2].

The binormal ROC curve is improper unless the binormal DV conditional diseased and nondiseased distribution variances are equal. When it is improper, there is a single fpf value such that the ROC curve is concave on one side and convex (concave up) on the other side; there also is a single fpf value where the ROC curve crosses the chance line [2], resulting in a “hooked” ROC curve. For example, rating data [9] from a radiologist that result in a hooked estimated binormal ROC curve estimate are presented in Table 1, with the hooked ROC curve displayed in Figure 1a (solid line).

For the typical situation where the binormal DV diseased-distribution variance is larger, the binormal ROC curve is convex for large  $\text{fpf}$  values and exhibits a hook in the upper right corner, as in Figure 1a, resulting in a threshold below which the likelihood ratio increases as the DV decreases. In practical terms, the convexity implies that below this threshold, the more “normal” an image appears the more likely it is to be diseased. Although for many situations the region of convexity and the hook and chance-line crossing are not noticeable, for some situations they can be noticeable enough to call into question the validity of conclusions based on the fitted curve. In addition, for some data sets, which Metz [10] refers to as *degenerate* data sets, the binormal likelihood is maximized when one or more of the parameter estimates is on the border of the parameter space, resulting in binormal ROC curves which have unacceptable zigzag shapes.

To circumvent these problems, Metz and Pan [11, 12] proposed estimating the ROC curve under the assumption that the DV is an increasing transformation of the likelihood-ratio transformation of a binormal random variable. The resulting ROC curve is always proper. Although they refer to their model as the *proper binormal model*, I find this name misleading since the underlying assumed decision-variable model is *not* binormal; thus, I will instead use the terminology of Hillis and Berbaum [13] and refer to their model as the *binormal likelihood-ratio model* (binormal-LR model).

Today this approach is often used for ROC analysis of radiologic imaging data. If the maximum-likelihood binormal ROC curve does not exhibit a noticeable hook, then typically the corresponding maximum-likelihood binormal-LR ROC curve is virtually identical to it. On the other hand, when the binormal hook is noticeable, the binormal-LR ROC curve provides a more acceptable curve because it is concave and hence has no hook. Thus this approach is appealing for researchers familiar with the binormal-model approach, but who want to avoid its problems.

However, the Metz and Pan [12] development is not easy to follow, especially for researchers that do not have extensive statistical training, and contains several important omissions: they do not provide an equation expressing  $\text{tpf}$  as an analytic function of  $\text{fpf}$  and vice versa and do not provide proofs for area-under-the-ROC curve (AUC) and partial area-under-the-ROC curve (pAUC) formulas. I show that the binormal-LR model is equivalent to a bi-chi-squared model in the sense that both models result in the same family of ROC curves. The bi-chi-squared model approach allows for specification of  $\text{tpf}$  and  $\text{fpf}$  as analytic functions of each other, straightforward derivation of AUC and pAUC formulas and other properties, and an easier-to-follow development of the binormal-LR ROC curve.

An outline of this paper is as follows. I define the binormal-LR ROC curve and corresponding notation in Section 2. I show in Section 3 that, provided  $b > 1$ , the family of binormal-LR ROC curves is equivalent to the family based on a proper bi-chi-squared model. Formulas for the binormal-LR ROC curve, AUC, and pAUC are derived in Section 4 using the bi-chi-squared model. I compare the bi-chi-squared and Metz and Pan approaches in Section 5, and discuss estimation using the bi-chi-squared approach in Section 6. Examples are presented in Section 7, followed by concluding remarks.

## 2. Binormal likelihood-ratio ROC curve definition and notation

Let  $ROC_{LR;a,b}$  denote a binormal-LR ROC curve based on the likelihood ratio transformation of a binormal random variable having binormal parameters  $a$  and  $b$ , where  $-\infty < a < \infty$ ,  $b > 0$ . For example, let  $Y$  be a binormal random variable with conditional distribution  $(Y|D=0) \sim N(0, 1)$ , where  $D$  is an indicator of disease status, with  $D=1$  denoting diseased and  $D=0$  denoting nondiseased. It follows that  $(Y|D=1) \sim N(a/b, 1/b^2)$ . Let  $LR_Y(\cdot)$  denote the likelihood ratio function for  $Y$ ; i.e.

$$LR_Y(y) = \frac{f_Y(y|D=1)}{f_Y(y|D=0)} = \frac{b\phi\left(\frac{y-a/b}{1/b}\right)}{\phi(y)} = \frac{b\phi(by-a)}{\phi(y)}$$

where  $f_Y(\cdot|D=d)$  denotes the density function for  $Y$  conditional on disease status  $d$  and  $\phi(\cdot)$  is the standard normal density function. Then  $ROC_{LR;a,b}$  is based on the DV  $LR_Y(Y)$ . As noted in the previous section,  $ROC_{LR;a,b}$  is proper and hence also concave.

Because the ROC curve corresponding to  $LR_Y(Y)$  is invariant to increasing transformations of  $LR_Y(Y)$ ,  $ROC_{LR;a,b}$  can also be defined by

$$U = \ln [LR_Y(Y)] = \ln(b) + \frac{1}{2} \left[ (1-b^2)Y^2 + 2abY - a^2 \right] \quad (1)$$

More generally, if  $\tilde{Y}$  is a binomial random variable with binormal parameters  $a$  and  $b$  and conditional distributions  $(\tilde{Y}|D=0) \sim N(\mu_1, \sigma_1^2)$  and  $(\tilde{Y}|D=1) \sim N(\mu_2, \sigma_2^2)$ , then it is straightforward to show that the log-likelihood function can be written in the form

$$\ln [LR_{\tilde{Y}}(\tilde{Y})] = \ln(b) + \frac{1}{2} \left[ (1-b^2) \left( \frac{\tilde{Y} - \mu_1}{\sigma_1} \right)^2 + 2ab \left( \frac{\tilde{Y} - \mu_1}{\sigma_1} \right) - a^2 \right]$$

which has the same form as (1) but with  $Y$  replaced by  $(\tilde{Y} - \mu_1)/\sigma_1$ . It is easy to show that  $((\tilde{Y} - \mu_1)/\sigma_1|D=0) \sim N(0, 1)$  and  $((\tilde{Y} - \mu_1)/\sigma_1|D=1) \sim N(a/b, 1/b^2)$ . It follows that  $\ln [LR_{\tilde{Y}}(\tilde{Y})]$  and  $\ln [LR_Y(Y)]$  have the same conditional distributions, and hence define the same binormal-LR ROC curve. Thus we see that  $ROC_{LR;a,b}$  is uniquely determined by  $a$  and  $b$ .

When estimating the binormal-LR ROC curve, Metz and Pan assume that the DV has the same distribution as the likelihood-ratio transformation of a binormal random variable (or more generally, as an increasing transformation of the likelihood-ratio transformation of a binormal random variable). Although each binormal-LR ROC curve is defined by a pair of  $a$  and  $b$  parameters of a corresponding binormal random variable, the role of the corresponding binormal random variable is only to define the distribution of the binormal-LR DV and of itself is generally not useful for practical purposes. A common misunderstanding among researchers is that Metz and Pan estimate binormal  $a$  and  $b$  parameters under the assumption that the DV has a binormal distribution and then use the

corresponding binormal-LR ROC curve (i.e., having the same  $a$  and  $b$  values) as the estimate – this is not what they do. Instead, they estimate the binormal-LR ROC curve  $a$  and  $b$  parameters under the assumption that the DV has a binormal-LR distribution. Thus for a given data set, the estimated binormal-LR ROC curve  $a$  and  $b$  values will in general differ from the estimated binormal ROC curve  $a$  and  $b$  values.

### 3. Relationship between binormal-LR and bi-chi-squared ROC curves

In this section I show for  $b \neq 1$  that  $\text{ROC}_{\text{LR};a,b}$  can be defined by a DV having a noncentral bi-chi-squared distribution with 1 degree of freedom, up to a scale factor. For  $b = 1$  I show that  $\text{ROC}_{\text{LR};a,b}$  is the same as the binormal ROC curve. Throughout I assume that  $Y$  is a binormal random variable with binormal parameters  $a$  and  $b$ . Because  $a$  and  $b$  uniquely determine  $\text{ROC}_{\text{LR};a,b}$  I will without loss of generality assume  $Y$  has conditional distributions  $(Y|D=0) \sim N(0, 1)$  and  $(Y|D=1) \sim N(a/b, 1/b^2)$ .

#### 3.1. Results for $b \neq 1$

It is straightforward to show that  $U(1)$  can be written in the form

$$U = g(a, b) + \frac{(1 - b^2)}{2} (Y - c_1)^2 \quad (2)$$

where

$$c_1 = -\frac{ab}{(1 - b^2)} \quad (3)$$

For  $b < 1$  it follows from (2) that  $u = \ln [\text{LR}_Y(y)]$  is quadratic in  $y$  and attains its minimum value at  $y = c_1$ . Hence

$$Y^* \equiv (Y - c_1)^2 = \left( Y + \frac{ab}{(1 - b^2)} \right)^2$$

is an increasing transformation of  $U$  and  $\text{ROC}_{\text{LR};a,b}$  can be equivalently defined based on  $Y^*$  as the DV. (Note that here and elsewhere, the term *increasing transformation* refers to a *strictly* increasing transformation.)

Figure 1b displays the conditional distributions for  $Y$ , where  $Y$  has a binormal distribution with parameters  $a = 1.06$ ,  $b = 0.464$ ; this is the binormal distribution corresponding to the hooked binormal ROC curve displayed in Figure 1a. The relationship between  $U$  and  $Y$  is displayed in Figure 1c.

Let  $\chi^2_{1;\nu}$  denote a noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter  $\nu$ . It is well known that  $\chi^2_{1;\nu}$  is the distribution of a squared normal

random variable having unit variance and mean  $\sqrt{\nu}$ . Because  $(Y - c_1|D=0) \sim N\left(\frac{ab}{(1-b^2)}, 1\right)$  and  $(Y - c_1|D=1) \sim N\left(\frac{a}{b(1-b^2)}, 1/b^2\right) = \frac{1}{b}N\left(\frac{a}{(1-b^2)}, 1\right)$ , it follows that

$$\begin{aligned} Y^* | (D=0) &\sim \chi^2_{1; \frac{a^2 b^2}{(1-b^2)^2}} \\ Y^* | (D=1) &\sim \frac{1}{b^2} \chi^2_{1; \frac{a^2}{(1-b^2)^2}} \end{aligned} \quad (4)$$

A parameterization that I will use for derivations and computational formulas is

$$\begin{aligned} Y^* | (D=0) &\sim \chi^2_{1; \theta} \\ Y^* | (D=1) &\sim \lambda \chi^2_{1; \lambda \theta} \end{aligned} \quad (5)$$

where the unconstrained parameter space is given by

$$\lambda > 0, \theta \geq 0$$

From (4-5) it follows that

$$\theta = \frac{a^2 b^2}{(1-b^2)^2}, \lambda = \frac{1}{b^2} \quad (6)$$

Thus the  $(\lambda, \theta)$  parameter space corresponding to  $\{(a, b) : 0 < b < 1\}$  is

$$\lambda > 1, \theta \geq 0$$

From (3) and (6) it follows that

$$\theta = c_1^2$$

and hence  $\theta$  can be interpreted, in terms of  $Y$ , as the squared difference between the nondiseased mean and the threshold where the likelihood ratio attains its minimum.

The conditional distributions for  $Y^*$  are displayed in Figure 1d and the binormal-LR ROC curve defined by  $Y^*$  is displayed as the dashed line in Figure 1a. To create the plot of the ROC curve in Figure 1a,  $\theta$  and  $\lambda$  were computed with  $a = 1.06$ ,  $b = 0.464$  using (6), and then tpf was computed as a function of fpf using formula (13), derived in Section 4.1.

Using well-known results for the noncentral chi-square distribution, the conditional means and variances are given by

$$\begin{aligned}
 E [Y^* | (D=0)] &= 1 + \theta \\
 E [Y^* | (D=1)] &= \lambda (1 + \lambda \theta) \\
 \text{var} [Y^* | (D=0)] &= 2 + 4\theta \\
 \text{var} [Y^* | (D=1)] &= \lambda^2 (2 + 4\lambda \theta) \quad (7)
 \end{aligned}$$

In summary I have shown for  $b < 1$  that  $ROC_{LR;a,b}$  can be defined by  $Y^*$ , which has conditional noncentral chi-square distributions with 1 degree of freedom (d.f.), up to a scale factor. In addition,  $Y^*$  is proper because it is an increasing transformation of the likelihood ratio function of  $Y$ . Hence  $Y^*$  has a *proper bi-chi-squared distribution*.

If  $b > 1$  then it follows from (2) that  $u = \ln [LR_Y(y)]$  is quadratic in  $y$ , attains its maximum value at  $y = c_1$ , and  $-Y^* = -(Y - c_1)^2$  is an increasing transformation of  $U$ . Thus  $-Y^*$  is proper and  $ROC_{LR;a,b}$  can be equivalently defined based on  $-Y^*$  as the DV. The conditional distributions for  $Y^*$  are again given by (5) with equations (6-7) still applicable, but now with

$$0 < \lambda < 1, \theta \geq 0$$

Hence  $-Y^*$  has a *proper negative bi-chi-squared distribution*.

Figure S1a (available in the online Supporting Materials) displays the ROC curve defined by a binormal random variable  $Y$  with parameters  $a = 8.65$ ,  $b = 4.40$ . For this curve the hook occurs in the lower left corner of the graph and is barely discernible. The conditional distributions for  $Y$ , the relationship between  $U$  and  $Y$ , and the conditional distributions of  $-Y^*$  are displayed in Figures S1b, S1c, and S1d, respectively. The corresponding binormal-LR ROC curve, defined by  $-Y^*$ , is displayed as the dashed line in Figure S1a.

No constraints were placed on  $a$  in deriving the distribution of  $Y^*$ . From (4) we see that the conditional distributions of  $Y^*$  are invariant to the sign of  $a$ . It follows that the binormal-LR curve is invariant to the sign of  $a$ :

$$ROC_{LR;a,b} = ROC_{LR;-a,b} \quad (8)$$

Metz and Pan [12, p. 5] state that “Without loss of generality, we assume that  $a \geq 0$  and  $b > 0$  . . .” but do not give any particular reason for this assumption; in particular, they do not note that the binormal-LR curve is invariant to the sign of  $a$ . It is natural to assume  $a \geq 0$  for performing binormal ROC curve estimation because  $a < 0$  represents the unlikely situation that a DV performs worse than chance. However, because the binormal-LR ROC curve corresponds to a proper DV, it cannot depict a DV that performs worse than chance, and hence it would at first appear that we should not constrain  $a \geq 0$  in order to obtain the most inclusive set of binormal-LR curves. However, the invariance result (8) shows that the set of binormal-LR ROC curves remains unchanged by the constraint  $a \geq 0$ .

### 3.2. Summary

I now formally summarize the results from Section 3.1.

**Result 1**—The family of binormal-LR ROC curves for  $-\infty < a < \infty, b \neq 1$ , is given by

$$\{ROC_{LR;a,b} : a \geq 0, b > 0, b \neq 1\}$$

Result 1 follows from the invariance property (8).

Motivated by the results from Section 3.1, I now formally define the particular one-degree-of-freedom two-parameter bi-chi-squared distribution that is utilized in this paper. Although a bi-chi-squared distribution could more generally have more than one degree of freedom and more or less than two parameters, for simplicity I refer to this particular form as the *bi-chi-squared distribution*.

**Definition 1**—A decision variable  $X$  has a **bi-chi-squared distribution** with parameters  $\lambda > 0, \theta \geq 0$  if its conditional distributions are given by

$$\begin{aligned} X | (D=0) &\sim \chi^2_{1;\theta} \\ X | (D=1) &\sim \lambda \chi^2_{1;\lambda\theta} \end{aligned}$$

Note that the noncentrality parameter  $\theta$  functions as a shape parameter for the nondiseased-case distribution. For the diseased-case distribution,  $\lambda$  functions in the dual role of both a shape (through its effect on the noncentrality parameter) and scale parameter. As  $\lambda$  increases for a given value of  $\theta$ , the mean and variance of the diseased distribution increase, with  $\lambda = 1$  corresponding to equal conditional distributions.

**Definition 2**—The **bi-chi-squared model** states that the decision variable is an increasing transformation of a decision variable having either a positive (with  $\lambda > 1$ ) or negative (with  $\lambda < 1$ ) bi-chi-squared distribution. That is, the decision variable is an increasing function of  $W$ , defined by

$$W = \begin{cases} Y^*, & \lambda \geq 1 \\ -Y^*, & \lambda < 1 \end{cases} \quad (9)$$

where  $Y^*$  has a bi-chi-squared distribution with parameters  $\theta$  and  $\lambda$ . The corresponding **bi-chi-squared ROC curve** is defined by  $W$  and will be denoted by  $ROC_{\chi^2;\lambda,\theta}$ .

Theorem 1 below states that a binormal-LR ROC curve (with  $b \neq 1$ ) can be defined by a bi-chi-squared DV, as was shown in Section 3.1.

**Theorem 1**—Fix  $a \geq 0, b > 0$  with  $b \neq 1$ . Define  $\theta$  and  $\lambda$  by (6). Then



$$ROC_{LR;a,b} = ROC_{\chi^2;\lambda,\theta} \quad (10)$$

Because (6) defines a one-to-one mapping from  $\{(a, b) : a > 0, b > 0, b \neq 1\}$  to  $\{(\lambda, \theta) : \theta > 0, \lambda > 0, \lambda \neq 1\}$  with

$$a = \sqrt{\frac{\theta(\lambda - 1)^2}{\lambda}}, b = \frac{1}{\sqrt{\lambda}} \quad (11)$$

it follows that the family of binormal-LR ROC curves with  $b \neq 1$  is equal to the family of bi-chi-squared ROC curves with  $\lambda \neq 1$ , as formally stated in the following Corollary. Thus the binormal-LR model and bi-chi-squared model are equivalent in the sense that they result in the same families of ROC curves.

**Corollary 1—**

$$\{ROC_{LR;a,b} : a > 0, b > 0, b \neq 1\} = \{ROC_{\chi^2;\lambda,\theta} : \theta > 0, \lambda > 0, \lambda \neq 1\}$$

Table 2 provides a summary of the main results from Sections 3.1-2

**3.3. Results for b = 1**

**3.3.1. Case: a ≠ 0—**The condition  $b = 1$  implies that  $Y$  has conditional distributions  $(Y|D = 0) \sim N(0, 1)$  and  $(Y|D = 1) \sim N(a, 1)$ . Then from (1),  $U = \ln [LR_Y(Y)]$  takes the form

$$U = aY - \frac{1}{2}a^2$$

Let  $ROC_{Bin;a,b}$  denote a binormal ROC curve with binormal parameters  $a$  and  $b$ . It follows that if  $a > 0$  then  $U$  is an increasing transformation of  $Y$ ; hence  $ROC_{LR;a,b=1} = ROC_{Bin;a,b=1}$ . If  $a < 0$  then  $U$  is an increasing transformation of  $-Y$ ; because  $-Y$  is binormal with binormal parameters  $|a|$  and  $b = 1$ , it follows that  $ROC_{LR;a<0,b=1} = ROC_{Bin;|a|,b=1}$ . Thus, as was true for  $b \neq 1$ , the binormal-LR ROC curve is invariant to the sign of  $a$ .

**3.3.2. Case: a = 0—**For  $a = 0$  and  $b = 1$  the conditional distributions of  $Y$  are the same and  $U = \ln [LR_Y(Y)] = 0$ , which defines a discrete ROC curve with only two points,  $(0, 0)$  and  $(1, 1)$ ; hence depending on the threshold chosen, all subjects are classified either as diseased or nondiseased. In contrast, the ROC curve based on  $Y$  is the chance line  $tpf = fpf$ . Noting that the binormal-LR curve converges to the chance line as  $a$  approaches 0 for  $b = 1$ , I arbitrarily define  $ROC_{LR;a=0,b=1}$  to be the chance line, i.e.,

$$ROC_{LR;a=0,b=1} \equiv ROC_{Bin;a=0,b=1}$$

This situation is not discussed in the Metz and Pan development.

**3.3.3. Inclusion of  $\lambda = 1$  in the bi-chi-squared parameter space**—Setting  $\lambda = 1$  for any value of  $\theta$  in (5) results in both conditional distributions of  $Y^*$  (5) having  $\chi^2_{1;\theta}$  distributions and a chance-line ROC curve. It follows that

$$ROC_{\chi^2;\lambda=1,\theta} = ROC_{LR;a=0,b=1} = ROC_{Bin;a=0,b=1}$$

Thus the chance line is the only concave binormal ROC curve (i.e.,  $b = 1$ ) that can be exactly represented by a bi-chi-squared ROC curve.

## 4. Formulas

In this section I derive basic formulas for the bi-chi-squared curve. Results from this section are presented in Table 3.

### 4.1. Tpf for fixed fpf

**4.1.1. Derivation and results for  $\lambda > 1$** —Letting  $F_{\nu}$  denote the distribution function for a random variable having a chi-squared distribution with one degree of freedom and noncentrality parameter  $\nu$ , it follows from (9) that

$$ROC_{\chi^2;\lambda,\theta} = \{(1 - F_{\theta}(c), 1 - F_{\lambda\theta}(c/\lambda)) : c \geq 0\} \quad (12)$$

From (12) we can express  $c$  in terms of fpf:  $c = F_{\theta}^{-1}(1 - fpf)$ . Substituting into (12) yields

$$ROC_{\chi^2;\lambda,\theta} = \left\{ \left( fpf, 1 - F_{\lambda\theta} \left( \frac{1}{\lambda} F_{\theta}^{-1}(1 - fpf) \right) \right) : 0 < fpf < 1 \right\}$$

Thus

$$tpf(fpf) = 1 - F_{\lambda\theta} \left( \frac{1}{\lambda} F_{\theta}^{-1}(1 - fpf) \right), 0 < fpf < 1 \quad (13)$$

**4.1.2. Derivation and results for  $\lambda < 1$** —It follows from (9) that

$$ROC_{\chi^2;\lambda,\theta} = \{(F_{\theta}(c), F_{\lambda\theta}(c/\lambda)) : c \geq 0\} \quad (14)$$

Expressing  $c$  in (14) in terms of fpf results in the equivalent expression

$$ROC_{\chi^2;\lambda,\theta} = \left\{ \left( fpf, F_{\lambda\theta} \left( \frac{1}{\lambda} F_{\theta}^{-1}(fpf) \right) \right) : 0 < fpf < 1 \right\}$$

Thus

$$tpf(fpf) = F_{\lambda\theta} \left( \frac{1}{\lambda} F_{\theta}^{-1}(fpf) \right)$$

Formulas expressing  $fpf$  as a function of  $tpf$ , which are included in Table 3, can be derived from the above results.

## 4.2. AUC

**4.2.1. Derivation and results for  $\lambda > 1$** —Fix  $\lambda > 1$ . Let  $AUC_{\chi^2}$  denote the area under the bi-chi-squared ROC curve. Let  $Y_D^*$  and  $Y_{\bar{D}}^*$  denote independent random variables with the same distributions as  $Y^*|(D=0)$  and  $Y^*|(D=1)$ , respectively. That is,

$$Y_{\bar{D}}^* \sim \chi_{1;\theta}^2 \quad \text{and} \quad Y_D^* \sim \lambda\chi_{1;\lambda\theta}^2$$

Bamber [14] shows that the probability of a continuous diseased test result exceeding a continuous nondiseased test result is equal to the AUC. It follows from (9) that

$$AUC_{\chi^2} = Pr \left( Y_D^* > Y_{\bar{D}}^* \right)$$

Let  $X_{\bar{D}}$  and  $X_D$  be independent random variables such that

$$X_{\bar{D}} \sim N \left( \sqrt{\theta}, 1 \right)$$

$$X_D \sim \sqrt{\lambda} N \left( \sqrt{\lambda\theta}, 1 \right) = N \left( \lambda\sqrt{\theta}, \lambda \right) \quad (15)$$

Then

$$X_{\bar{D}}^2 \sim \chi_{1;\theta}^2 \quad \text{and} \quad X_D^2 \sim \lambda\chi_{1;\lambda\theta}^2$$

Thus

$$\begin{aligned} AUC_{\chi^2} &= Pr \left( Y_D^* > Y_{\bar{D}}^* \right) \\ &= Pr \left( X_D^2 > X_{\bar{D}}^2 \right) \\ &= Pr \left( X_D > X_{\bar{D}}, X_{\bar{D}} > 0 \right) + Pr \left( X_D < X_{\bar{D}}, X_{\bar{D}} < 0 \right) + Pr \left( X_D < -X_{\bar{D}}, X_{\bar{D}} > 0 \right) + Pr \left( X_D > -X_{\bar{D}}, X_{\bar{D}} < 0 \right) \end{aligned}$$

It follows that

$$AUC_{\chi^2} = Pr(A_1) + Pr(A_2)$$

where the mutually exclusive events  $A_1$  and  $A_2$  are defined by

$$A_1 = \left\{ X_D > X_{\bar{D}}, X_D > -X_{\bar{D}} \right\}$$

$$A_2 = Pr \left\{ X_D < X_{\bar{D}}, X_D < -X_{\bar{D}} \right\}$$

The contour plot of the bivariate  $(X_{\bar{D}}, X_D)$  density function and subspaces corresponding to  $A_1$  and  $A_2$  are graphically displayed in Figure 2a. Note that  $X_{\bar{D}}$  and  $X_D$  are notated as  $X_1$  and  $X_2$ , respectively, in the figure. Also displayed are the subspaces corresponding to  $A_3$  and  $A_4$ , defined by

$$A_3 = Pr \left\{ X_D > X_{\bar{D}}, X_D < -X_{\bar{D}} \right\}$$

$$A_4 = Pr \left( X_D < X_{\bar{D}}, X_D > -X_{\bar{D}} \right)$$

Let  $F_{BVN}(\cdot, \cdot, \rho)$  denote the distribution function for a standardized bivariate normal distribution with correlation  $\rho$ , and let  $MVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \sigma_{12})$  denote a bivariate normal distribution with means  $\mu_1, \mu_2$ , variances  $\sigma_1^2, \sigma_2^2$ , and covariance  $\sigma_{12}$ . Writing  $A_1$  and  $A_2$  in the form

$$A_1 = \left\{ -\left(X_D - X_{\bar{D}}\right) < 0, -\left(X_D + X_{\bar{D}}\right) < 0 \right\}$$

$$A_2 = \left\{ X_D - X_{\bar{D}} < 0, X_D + X_{\bar{D}} < 0 \right\}$$

and noting that

$$\left( X_D - X_{\bar{D}}, X_D + X_{\bar{D}} \right) \sim MVN \left( \sqrt{\theta}(\lambda - 1), \sqrt{\theta}(\lambda + 1), \lambda + 1, \lambda + 1, \lambda - 1 \right)$$

it follows that

$$Pr(A_1) = F_{BVN} \left( \frac{\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}, \frac{\sqrt{\theta}(\lambda + 1)}{\sqrt{\lambda + 1}}, \frac{\lambda - 1}{\lambda + 1} \right)$$

$$Pr(A_2) = F_{BVN} \left( -\frac{\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}, \frac{\sqrt{\theta}(\lambda + 1)}{\sqrt{\lambda + 1}}, \frac{\lambda - 1}{\lambda + 1} \right)$$

Hence

$$AUC_{\chi^2} = F_{BVN}(u_1, u_2; \rho) + F_{BVN}(-u_1, -u_2; \rho) \quad (16)$$

where

$$u_1 = \frac{\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}, u_2 = \sqrt{\theta} \sqrt{(\lambda + 1)}, \rho = \frac{\lambda - 1}{\lambda + 1} \quad (17)$$

**4.2.2. Derivation and results for  $\lambda < 1$** —Fix  $\lambda < 1$ . From (9) and Bamber's result [14] we have

$$AUC_{\chi^2} = Pr\left(-Y_D^* > -Y_{\bar{D}}^*\right) = Pr\left(Y_D^* < Y_{\bar{D}}^*\right)$$

It follows that

$$AUC_{\chi^2} = 1 - Pr\left(Y_D^* > Y_{\bar{D}}^*\right) \quad (18)$$

Since the derivation for  $Pr(Y_D^* > Y_{\bar{D}}^*)$  in Section 4.2.1 did not depend on the value of  $\lambda$ , it follows that the formula for  $Pr(Y_D^* > Y_{\bar{D}}^*)$  for  $\lambda < 1$  is the same as for  $\lambda > 1$ . Hence

$$AUC_{\chi^2} = 1 - [F_{BVN}(u_1, u_2; \rho) + F_{BVN}(-u_1, -u_2; \rho)] \quad (19)$$

where  $u_1$ ,  $u_2$ , and  $\rho$  are given by (17). Because  $A_1$ ,  $A_2$ ,  $A_3$  and  $A_4$  are mutually exclusive with  $\sum_{i=1}^4 P(A_i) = 1$ , it also follows from (18) that

$$AUC_{\chi^2} = Pr(A_3) + Pr(A_4) \quad (20)$$

Although a formula similar to (16) could be derived from this expression, for our purposes (20) will be used for deriving the alternative AUC expression (discussed below) in Appendix A.

**4.2.3. Alternative AUC expression**—Metz and Pan [11, 12] present an alternative expression for  $AUC_{\chi^2}$ , expressing it as the sum of the AUC of the corresponding binormal ROC curve plus a nonnegative term. However, they do not provide a proof for their results. For completeness I derive their result in Appendix A. I show in Section 5 that their result, in terms of the bi-chi-squared parameterization, is given by

$$AUC_{\chi^2} = \Phi\left(\frac{\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}\right) + 2F_{BVN}\left(-\frac{\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}, 0; -\frac{2\sqrt{\lambda}}{\lambda + 1}\right), \lambda \neq 1 \quad (21)$$

The first term on the right side is the AUC of the corresponding binormal ROC curve.

### 4.3. Partial AUC

Two different partial areas under the ROC curve have been proposed [15, 16, 17]: the area under the ROC curve over an fpf interval ( $fpf_1 < fpf_2$ ), which I denote by  $pAUC_{fpf}(fpf_1, fpf_2)$ , and the area to the right of the ROC curve over a tpf interval ( $tpf_1 < tpf_2$ ), which I denote by  $pAUC_{tpf}(tpf_1, tpf_2)$ . Both partial areas are often normalized by dividing by the length of the interval, which allows the partial area to be interpreted as the average value of tpf or fpf over their respective intervals. Furthermore, it is easy to geometrically show

$$pAUC_{tpf}(tpf_0, 1) = AUC - pAUC_{fpf}(0, fpf(tpf_0)) - tpf_0 \times [1 - fpf(tpf_0)]$$

Because  $pAUC_{fpf}(fpf_1, fpf_2) = pAUC_{fpf}(0, fpf_2) - pAUC_{fpf}(0, fpf_1)$  and  $pAUC_{tpf}(tpf_1, tpf_2) = pAUC_{tpf}(tpf_1, 1) - pAUC_{tpf}(tpf_2, 1)$ , it suffices to derive only the formula for  $pAUC_{fpf}(0, fpf_0)$ , which is derived in Appendix B and presented in Table 3.

Pan and Metz [11] provide partial AUC formulas but do not provide proofs. Although their formulas are different from the ones in this paper, I was able to empirically demonstrate equality of their partial AUC formulas with the formulas derived in this paper by comparing partial AUC results from PROPROC with those computed using the formulas in this paper. I do not include analytical proofs showing equivalence because their formulas have no clear conceptual advantage and require the use of numerical methods.

## 5. Comparison with Metz and Pan approach

### 5.1. Metz and Pan approach

Metz and Pan [12] derive results for  $ROC_{LR;a,b}$  based on the DV

$$V = \frac{(b+1)Y}{a + \sqrt{a^2 + 2(1-b^2)Y}} \quad (22)$$

which is an increasing transformation of  $LR_Y(Y)$ . They use the  $(d_a, c)$  parameterization for  $V$ , given by

$$d_a = \frac{\sqrt{2}a}{\sqrt{1+b^2}}, 0 \leq d_a < \infty \quad (23)$$

$$c = \frac{b-1}{b+1}, -1 < c < 1 \quad (24)$$

The positive density support of  $V$  is given by

$$\begin{aligned} \frac{d_a}{4c} \sqrt{1+c^2} \leq V \leq \infty, & \text{ if } c < 0 \\ -\infty \leq V \leq \infty, & \text{ if } c = 0 \\ -\infty \leq V \leq \frac{d_a}{4c} \sqrt{1+c^2}, & \text{ if } c > 0 \end{aligned}$$

Metz and Pan [11, 12] do not write tpf as an analytic function of fpf or vice versa. Instead, for  $c \neq 0$  they give the following formulas for fpf and tpf corresponding to threshold  $v_c$ :

$$fpf(v_c) = \Phi \left[ -(1-c)v_c - \frac{d_a}{2} \sqrt{1+c^2} \right] + \left\{ \Phi \left[ -(1-c)v_c + \frac{d_a}{2c} \sqrt{1+c^2} \right] - H(c) \right\} \quad (25)$$

$$tpf(v_c) = \Phi \left[ -(1+c)v_c + \frac{d_a}{2} \sqrt{1+c^2} \right] + \left\{ \Phi \left[ -(1+c)v_c + \frac{d_a}{2c} \sqrt{1+c^2} \right] - H(c) \right\} \quad (26)$$

where  $H(c) = 0$  if  $c < 0$  and  $H(c) = 1$  if  $c > 0$ .

Metz and Pan [11, 12] give the following AUC formula without proof:

$$AUC = \Phi \left( \frac{d_a}{\sqrt{2}} \right) + 2F_{BVN} \left( -\frac{d_a}{\sqrt{2}}, 0; -\frac{1-c^2}{1+c^2} \right) \quad (27)$$

Using the relationships between the bi-chi-squared  $(\lambda, \theta)$  and Metz and Pan  $(d_a, c)$  parameterizations discussed below, in terms of the  $(\lambda, \theta)$  parameterization (27) takes the form of equation (21).

### 5.2. Parameterization relationships

It follows from (23-24) that

$$a = d_a \sqrt{\frac{1+c^2}{(1-c)^2}} \quad \text{and} \quad b = \frac{c+1}{1-c} \quad (28)$$

Using (28) and (11) it is straightforward to derive the relationships between the bi-chi-squared  $(\lambda, \theta)$  and Metz and Pan  $(d_a, c)$  parameterizations,  $c \neq 0$  and  $\lambda \neq 1$ :

$$\lambda = \frac{(1-c)^2}{(c+1)^2}, \theta = \frac{1}{16c^2} d_a^2 (c^2+1) (c+1)^2$$

and

$$c = \frac{1 - \sqrt{\lambda}}{1 + \sqrt{\lambda}}, d_a = \frac{\sqrt{2\theta(\lambda - 1)^2}}{\sqrt{(\lambda+1)}}$$

It follows that  $c$  is a strictly decreasing function of  $\lambda$  and  $d_a$  is a strictly increasing function of  $\theta$  for fixed  $\lambda$ .

According to Metz and Pan [12, p. 12], an important computational reason they preferred the  $(d_a, c)$  parameterization was because for fixed  $d_a$  there is a one-to-one relationship between AUC and  $c$  in each of the  $\{c < 0\}$  and  $\{c > 0\}$  regions; specifically, AUC is an increasing (decreasing) function of  $c$  if  $c < 0$  ( $c > 0$ ). Similarly, for fixed  $\theta$ , AUC is an increasing (decreasing) function of  $\lambda$  if  $\lambda < 1$  ( $\lambda > 1$ ); a proof is provided in Appendix C. It can also be shown for fixed  $\lambda > 1$  that AUC is an increasing function of  $\theta$ . Figure 3 illustrates how the AUC changes with increasing or decreasing theta or lambda, given the value of the other parameter.

### 5.3. Advantages of the bi-chi-squared approach

The bi-chi-squared approach has two important advantages over the Metz and Pan approach. First, it is conceptually easier to explain and understand: because the log likelihood-ratio function of a binormal random variable  $Y$  is symmetric about  $c_1$ , it follows that either  $Y^* = (Y - c_1)^2$  or  $-Y^*$  is an increasing transformation of  $LR_Y(Y)$  and hence can be used as the DV for defining the binormal-LR ROC curve. In contrast, it is not immediately obvious that  $V(22)$  is an increasing transformation of  $LR_Y(Y)$ . The second advantage is that the conditional distributions of  $Y^*$  are familiar chi-squared distributions, which allows for easy derivation of statistical properties with no further need to consider the relationship between  $Y^*$  and  $Y$ . In contrast, the conditional distributions of  $V$  are not familiar and have support boundaries that are functions of the distribution parameters, making derivations more tedious. The conditional densities of  $V$  are not described analytically or graphically by Metz and Pan. For example, Metz and Pan derive results for tpf and fpf by determining the conditional probabilities of regions of  $Y$  that map into  $(V > v_c)$ , rather than working directly with the conditional distributions of  $V$ .

## 6. Estimation

Software for estimating the binormal-LR ROC curve has been extensively developed by Charles Metz and colleagues [11, 12, 18] at the University of Chicago, with much of the work undertaken in collaboration with Kevin Berbaum and colleagues at the University of Iowa. This software was originally called PROPROC, but recently has been renamed PBM; however, the program has remained essentially the same (personal communications with



Charles E. Metz, June 3, 2012, and Lorenzo L. Pesce, March 13, 2014). The algorithm is extremely fast and has undergone extensive testing. It can be downloaded in various formats from websites at either university (<http://metz-roc.uchicago.edu/> and <http://perception.radiology.uiowa.edu/>). PROPROC is used with the stand-alone multireader ROC software *OR-DBM MPMC 2.5* [19] as well as the program *OR-DBM MPMC 3.0 for SAS* [20], both freely available at <http://perception.radiology.uiowa.edu/>.

In this section I show that the bi-chi-squared model makes it possible to estimate ROC-curve parameters for discrete rating data using standard statistical software to maximize the likelihood function; I refer to such an algorithm as a *bi-chi-squared algorithm*. Specifically, I describe a bi-chi-squared algorithm that uses the SAS procedure NLMIXED to maximize the likelihood.

I emphasize that the bi-chi-squared algorithm developed in this section is not intended for general use, but rather for empirically validating the bi-chi-squared approach. In Example 1 in Section 7 I show that AUC estimates for a real data set obtained using the bi-chi-squared algorithm are virtually identical to those obtained using the PROPROC algorithm.

### 6.1. Assumptions

The basic assumption when estimating the binormal-LR ROC curve is that the DV, explicit or implicit, has the same distribution as the likelihood-ratio transformation of a binormal random variable (or more generally, as an increasing transformation of the likelihood-ratio transformation of a binormal random variable); this is the approach taken by Metz and Pan. As discussed in Section 2, it is important to note that they do not assume the DV has a binormal distribution. Similarly, for estimation purposes using the bi-chi-squared approach, I assume the DV follows the bi-chi-squared model, not the binormal model.

### 6.2. Bi-chi-squared algorithm

**6.2.1. Likelihood function**—I assume that a reader's ordinal likelihood-of-disease ratings represent the binning of values of a latent bi-chi-squared model DV; i.e., the DV has either a positive (with  $\lambda \geq 1$ ) or negative (with  $\lambda < 1$ ) bi-chi-squared distribution. Specifically, letting  $R$  denote an observed rating variable having  $p$  distinct values  $r_1 < \dots < r_p$  and letting  $L$  denote the latent DV, it is assumed that  $\Pr (R = r_j) = \Pr (c_{j-1} < L < c_j)$ , where  $-\infty = c_0 < c_1 < \dots < c_{p-1} < c_p = \infty$ ; i.e.,  $(c_{j-1}, c_j)$  is the latent bin corresponding to  $R = r_j$ .

Furthermore, I assume that the reader assigns likelihood-of-disease ratings independently to  $n_0$  nondiseased and  $n_1$  diseased case images. Let  $n_{0i}$  and  $n_{1i}$  denote the number of

nondiseased and diseased cases assigned rating  $r_i$ , respectively; thus  $\sum_{i=1}^p n_{0i} = n_0$  and

$\sum_{i=1}^p n_{1i} = n_1$ . The log likelihood, after deleting terms that do not depend on model parameters, is given by

$$ll = \sum_{i=1}^p n_{0i} \ln \Pr (R=r_i|D=0) + \sum_{i=1}^p n_{1i} \ln \Pr (R=r_i|D=1)$$

Define

$$\begin{aligned} fpf_i &\equiv Pr(R \geq r_i | D=0) = Pr(L > c_{i-1} | D=1) \\ tpf_i &\equiv Pr(R \geq r_i | D=0) = Pr(L > c_{i-1} | D=1) \end{aligned}$$

$i = 1, \dots, p$  and

$$fpf_{p+1} = tpf_{p+1} \equiv 0$$

Then

$$ll = \sum_{i=1}^p n_{0i} \ln (fpf_i - fpf_{i+1}) + \sum_{i=1}^p n_{1i} \ln (tpf_i - tpf_{i+1}) \tag{29}$$

A bi-chi-squared algorithm maximizes  $ll$  with respect to  $\{\lambda, \theta, c_1, \dots, c_{p-1}\}$ , or alternatively with respect to another parameterization such as  $\{\lambda, \theta, fpf_1, \dots, fpf_p\}$  or  $\{\lambda, \theta, \Delta_1, \dots, \Delta_p\}$ , where  $\Delta_i = fpf_i - fpf_{i+1}$ . Here the bi-chi-squared parameter space is  $\lambda < 0, \theta > 0$ . Note that I have included  $\lambda = 1$  in the parameter space, which for any value of  $\theta$  corresponds to the chance-line ROC curve as discussed in Section 3.3.3.

With this approach, maximum likelihood estimation excludes the subset of proper binormal ROC curves. This approach is justified because any proper binormal ROC curve (thus  $b = 1$ ) can be approximated by a bi-chi-squared curve with arbitrary precision. I note that PROPROC and PBM take a similar approach (Lorenzo L. Pesce, personal communication, April 18, 2014.)

**6.2.2. Maximizing the likelihood using PROC NLMIXED**—Using the

parameterization  $\{\lambda, \theta, \Delta_1, \dots, \Delta_p\}$ , where  $\Delta_i = fpf_i - fpf_{i+1}$  (and thus  $\sum_{i=1}^p \Delta_i = 1$ ), I use the SAS procedure NLMIXED to compute parameter estimates that maximize  $ll$  (29) for the data set in Example 1 in Section 7. There are five possible ratings:  $r_i = i, i = 1, \dots, 5$ . For initial 1 in Section set  $\hat{\Delta}_{i;\text{initial}} = \max(\hat{\Delta}_i, .001), i = 1, 2, 3, 4$ , where  $\hat{\Delta}_i$  is the empirical estimate (i.e., proportion of nondiseased subjects assigned rating  $i$ ). The algorithm sets  $\hat{\Delta}_{5;\text{initial}} = 1 - \sum_{i=1}^4 \hat{\Delta}_{i;\text{initial}}$ . To avoid  $\hat{\Delta}_{5;\text{initial}} < .0001$ , if  $\sum_{i=1}^4 \hat{\Delta}_{i;\text{initial}} < 0.999$  then I modify  $\hat{\Delta}_{i;\text{initial}}, i = 1, 2, 3, 4$  by subtracting a small constant from one of them to ensure that  $\sum_{i=1}^4 \hat{\Delta}_{i;\text{initial}} \leq 0.999$ . This process results in initial  $\Delta_i = fpf_i - fpf_{i+1}$  estimates close to the empirical estimates, but with no estimates less than .001 for each of the ten modality-reader combinations. Initial estimates for  $\lambda$  and  $\theta$  are computed using (6) with  $a$  and  $b$  replaced by binormal-model estimates. The log-likelihood function is computed using equation (29) with  $tpf_j$  replaced by

$$tpf_i = \begin{cases} 1 - F_{\lambda\theta} \left[ \frac{1}{\lambda} F_{\theta}^{-1} (1 - f_{pf_i}) \right], & \lambda \geq 1 \\ F_{\lambda\theta} \left[ \frac{1}{\lambda} F_{\theta}^{-1} (f_{pf_i}) \right], & \lambda < 1 \end{cases} \quad (30)$$

where the  $f_{pf_i}$  are computed from the  $f_i$ . Note that (30) combines the expressions for  $tpf(f_{pf})$  for  $\lambda < 1$  and  $\lambda > 1$  from Table 3b and also includes  $\lambda = 1$  for which the expression reduces to  $tpf_i = f_{pf_i}$ .

The SAS code used for computing the ROC parameter estimates for Example 1 is provided in Appendices S1 and S2, available in the online Supporting Materials. Two versions of the codes are provided: a simple version without macros in Appendix S1 that estimates the bi-chi-squared ROC curve for one reader-treatment combination, and a more extensive version with macros in Appendix S2 that does the complete Example 1 analysis. The code is somewhat similar to that provided by Gönen [21, Chapter 5], who discusses estimating the binormal ROC curve using the SAS NLMIXED procedure. The main difference is that the bi-chi-squared code involves deltas rather than thresholds in the parameterization; this seemed to be a more natural approach because of the need to accommodate both the positive and negative bi-chi-squared distributions. Gönen [21, Appendix] provides a short introduction to PROC NLMIXED, which I recommend for the reader who is not familiar with PROC NLMIXED.

Although the code in Appendix S2 works well for the Example 1 data in the next section, again it should be noted that it is not intended for general use, but rather for validation of the bi-chi-squared approach. In particular, a program for general use would need a more rigorous approach for selecting starting values and need to be more thoroughly tested. For example, Pesce *et al.* [18] describe a rigorous method for selecting initial estimates for PROPROC.

### 6.3. Continuous data

For continuous data, Metz *et al.* [22] propose defining categories of the observed data that correspond to vertical or horizontal jumps in the empirical ROC curve and then applying the approach used for ordinal discrete data to the defined categories. Specifically, they state that “ML estimation of an ROC curve from continuously-distributed data is equivalent to ML estimation from ordinal category data if truth state runs in the rank-ordered data (described next) are interpreted as categorical data.” This method of categorization is referred to as the LABROC4 categorization scheme and is discussed in the context of fitting binormal-LR models by Metz and Pan [12] and by Pesce and Metz [18]. Similarly, the bi-chi-squared algorithm for discrete data can be applied to continuous data.

### 6.4. Relationship between bi-chi-squared and PROPROC/PBM algorithms

Both the bi-chi-squared and PROPROC/PBM estimation algorithms maximize the same log-likelihood function (29), but with different formulas used for  $f_{pf}$  and  $tpf$ . Bi-chi-squared uses (30), derived from the bi-chi-squared model, to estimate  $tpf$  as a function of  $f_{pf}$ , whereas PROPROC/PBM uses (25) and (26), derived from the relationship between  $V$  (22) and  $Y$ , to

compute fpf and tpf as functions of a threshold. Of course, the bi-chi-squared algorithm could also have computed tpf and fpf as functions of a threshold. Thus the log-likelihood functions are the same but parameterized differently.

From a computational perspective neither parameterization appears to have a substantive advantage if similar starting values and likelihood maximization techniques are used. This conjecture is based on the similarities between the parameterizations as discussed in Section 5.2, but further study is needed. Furthermore, it appears that the initial estimates and optimization approach discussed by Pesce and Metz [18] could be adapted for the bi-chi-squared parameterization. From my perspective, the main advantage of the bi-chi-squared approach is not computational, but as previously discussed, easier conceptual understanding and derivation of properties.

## 7. Examples

### 7.1. Example 1: Comparison of binormal, bi-chi-squared, PROPROC algorithm results

In this example I estimate the AUC for ten ROC curves using the PROPROC and bi-chi-squared algorithms under the assumption of a binormal-LR/bi-chi-squared model for the DV. The PROPROC algorithm estimates are computed using *OR/DBM MRMC 3.0 for SAS* [20] and the bi-chi-squared estimates are computed using the bi-chi-squared algorithm discussed in Section 6.2, as implemented by SAS code provided in Appendix S2 in the online Supporting Materials. Binormal estimates of  $a$  and  $b$ , computed using *OR/DBM MRMC 3.0 for SAS* [20], are used to compute initial  $\lambda$  and  $\theta$  estimates for the bi-chi-squared algorithm using relationship (6).

The data are from a study [9] that compares the relative performance of single spin-echo magnetic resonance imaging (MRI) and cine MRI in detecting thoracic aortic dissection. There are 114 patients, 45 with an aortic dissection and 69 without. Five radiologists independently interpret all of the images using a five-point ordinal scale. Data for one of the reader-modality combinations (reader 5, cine MRI) are presented in Table 1.

Table 4 presents the AUC estimates for both algorithms and parameter estimates resulting from the bi-chi-squared algorithm. The AUC values in Table 4 are the same for both algorithms except for reader 3 and modality 1, for which the AUCs are 0.908 (PROPROC) and 0.929 (bi-chi-squared) with the PROPROC solution having the higher likelihood. This discrepancy for reader 3 and modality 1 was resolved by running the bi-chi-squared algorithm again but with several other starting values, which resulted in a value of AUC equal to the PROPROC value; the revised bi-chi-squared values are shown within parentheses.

### 7.2. Example 2: Comparison of estimated binormal and bi-chi-squared ROC curves and latent distributions

Figure 4a shows the binormal and bi-chi-squared maximum-likelihood estimated (MLE) ROC curves for the data in Table 1a. The corresponding binormal and bi-chi-squared latent distributions are displayed in Figures 4b and 4c. (Note that the binormal MLE ROC curve and its latent DV distribution are also shown in Figures 1a and 1b, respectively.)

Figure 4b includes the binormal MLE estimates:  $a = 1.06$  and  $b = 0.46$ . Figure 4c includes the bi-chi-squared ( $\lambda, \theta$ ) estimates and the corresponding estimates  $a = 0.67$  and  $b = 0.33$  computed using (11). Thus the  $a$  and  $b$  estimates for the binormal MLE curve are different than those for the bi-chi-squared MLE curve, as is generally the case. In contrast, in Figure 1a the binormal and bi-chi-squared (or binormal-LR) ROC curves have the same ( $a = 1.06, b = 0.46$ ) parameter values because only the binormal ROC curve is estimated from the data; the purpose of Figure 1a is to display the binormal-LR ROC curve that is defined by that same  $a$  and  $b$  values. The distinction between these two graphs is important: although each possible binormal ROC curve defines a corresponding binormal-LR ROC curve having the same  $a$  and  $b$  values, as illustrated by Figure 1a, if binormal and binormal-LR ROC curves are estimated from the same data then the parameter values will typically differ, as illustrated in Figure 4 and previously discussed in Section 2.

Both ROC curves in Figure 4a are similar for fpf values less than that of the last empirical operating point. However, for greater fpf values the ROC curves are discernibly different: the binormal curve has a hook in the upper right corner in contrast to the binormal-LR curve, which has no hook because it is concave. As a result, the binormal-LR curve has the higher AUC (0.841 vs. 0.833).

### 7.3. Example 3: Interpretation of bi-chi-squared parameters in terms of similarity with corresponding binormal ROC curve

When presented with bi-chi-squared estimation results such as in Table 4, it is natural to ask if it is possible to know which of the bi-chi-squared ROC curves is approximated by its corresponding binormal curve from an examination of the ( $\lambda, \theta$ ) values. One way to answer this question is to determine if the corresponding binormal ROC curve has discernible improperness; if it does not, then the bi-chi-squared curve will be closely approximated by it. Along these lines, Metz and Pan [12, pp 15-16] write “For a given pair of curve parameter values, the corresponding proper and conventional binormal ROC curves are indistinguishable if and only if no ‘hook’ is evident in the conventional binormal ROC ... Empirically, this occurs if and only if  $d_a \leq 6|c|$ .”

The degree of improperness of the corresponding binormal ROC curve can be measured by the binormal mean-to-sigma ratio, defined by

$$r = \frac{a}{1 - b} \quad (31)$$

Hillis and Berbaum [13] note that  $\Phi(r)$  is the unique fpf where the binormal ROC curve crosses the chance line. Based on this result, they classify the improperness of a binormal ROC curve as *indiscernible* if  $|r| \geq 3$ , *noticeable* if  $|r| \geq 2$ , and *slight* if  $2 < |r| < 3$ . From (11) and (31) it follows that

$$r = \begin{cases} \sqrt{\theta} (\sqrt{\lambda} + 1), & \lambda > 1 \\ -\sqrt{\theta} (\sqrt{\lambda} + 1) & \lambda < 1 \end{cases}$$

and hence

$$|r| = \sqrt{\theta} \left( \sqrt{\lambda+1} \right), \lambda \neq 1$$

Figure 5 displays a contour map of  $|r|$  in the bi-chi-squared parameter space. This figure shows the region where the bi-chi-squared ROC curves are similar to their corresponding binormal curves. For example, we see that if  $\lambda < 1$  and  $\theta \approx 2.25$  then the two curves are similar because  $|r| \approx 3$ .

Alternatively, I could have rewritten the Metz and Pan inequality  $d_a \leq |c|$  in terms of the bi-chi-squared parameterization. However, the advantage of the proposed method is that it is based on an interpretable index of improperness for the corresponding binormal ROC curve.

## 8. Discussion

The binormal-LR model assumption is attractive to researchers because it always results in proper estimated ROC curves, including curves similar to “acceptable” binormal ROC curves, i.e., binormal curves that are not noticeably nonconcave. For this reason binormal-LR ROC curve estimation has become common in radiologic diagnostic studies.

Metz and Pan [11, 12] provide a detailed development of the binormal-LR ROC curve; however, their development is not easy to follow, especially for researchers that do not have extensive statistical training. The main contribution of this paper is the recognition that the binormal-LR model (with equal-variance ROC curves excluded) is equivalent to a bi-chi-squared model. The bi-chi-squared approach is easier to explain and understand, and allows for easier derivation of statistical properties. This paper develops the bi-chi-squared approach and uses it to fill in some of the omissions in the Metz and Pan papers, such as writing tpf and fpf as analytic functions of each other and providing proofs for the AUC and partial AUC formulas; furthermore, it illustrates how the approach makes it possible to fit ROC curves using standard statistical software. It is important to note the bi-chi-squared model discussed in this paper involves only one-degree-of-freedom noncentral bi-chi-squared distributions.

In my opinion the terms *proper binormal ROC curve* and *proper binormal model* used by Metz and Pan are misnomers because the only way an ROC curve can be both proper and binormal is for it to be an equal-variance ( $b = 1$ ) binormal curve. Thus I recommend that these terms not be used, but instead suggest the terms *binormal-LR ROC curve* and *binormal-LR model*. Although it might seem reasonable to alternatively use the terms *bi-chi-squared ROC curve* and *bi-chi-squared model*, in general I do not recommend these terms because they involve a particular form of the bi-chi-squared distribution and hence are not as specific as *binormal-LR ROC curve* and *binormal-LR model*. However, I do recommend these terms when it is important to indicate that the bi-chi-squared distribution was used for conceptualization, estimation or derivation of results.

Because a bi-chi-squared distribution looks much different than a binormal distribution, it might at first seem that the assumption of a bi-chi-squared model is a major change from the

assumption of a binormal model in terms of the underlying DV. In this paper I derived results by showing that a binormal-LR curve with parameters  $(a, b < 1)$  can be based on the bi-chi-squared random variable  $Y^* = (Y - c_1)^2$  where  $Y$  is binormal with parameters  $(a, b)$  and  $c_1$  is the point where its likelihood ratio attains its minimum. Alternatively, because  $Y^*$  is an increasing transformation of  $|Y - c_1|$ , I could have derived results based on  $|Y - c_1|$  as the DV, which has a bi-folded-normal distribution [23]. When the binormal DV is not visibly improper, the shape of the distribution of  $|Y - c_1|$  is similar to that of the corresponding binormal DV, except for the left tail where both distributions end at zero. Thus we can consider the DVs of the bi-chi-squared and binormal models to be similar after an increasing transformation when improperness is not visible in the corresponding binormal model. The reason I chose the bi-chi-squared approach over the bi-folded-normal approach was because the chi-squared distribution is more familiar.

Other two-parameter proper ROC models that have been proposed include the two-parameter bigamma model proposed by Dorfman *et al.* [24], a contaminated binormal model [25], a weighted power-function ROC curve model [26], and more generally a method for constructing pairs of diseased and nondiseased distributions that have a specified monotonic likelihood ratio [27]. Although these models have been extensively compared with the binormal model, there has been little study devoted to comparing them with each other; this is an area for future research.

Because a central chi-squared distribution is a particular form of a gamma distribution, it might seem that the proposed bi-chi-squared model would be closely related to the bigamma model proposed by Dorfman *et al.* [24]. For their model the two parameters are the conventional shape and scale parameters. The conditional nondiseased and diseased distributions have gamma distributions with the same shape parameter but different scale parameters. Their model includes bi-chi-squared (now using the term more generally) models where the two conditional distributions have central chi-squared distributions with the same degrees of freedom (which can be any positive integer) but different scale parameters. In contrast, the binormal-LR bi-chi-squared conditional distributions are noncentral chi-squared with one degree of freedom, but do not have the same shape (unless  $\lambda = 1$  which results in the chance-line ROC curve). The bigamma model assumes that the diseased distribution variance is not less than that of the nondiseased distribution, but the bi-chi-squared model in this paper does not have this limitation. Comparison of these two approaches is an area for future research.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This research was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under Award Number R01EB013667. I thank Carolyn Van Dyke, M.D. for sharing her data set. I thank two reviewers for their very helpful suggestions for revising this paper.

### A. Alternative AUC formula derivation

In this section I derive the alternative AUC formula (21), which I restate below:

$$AUC_{\chi^2} = \Phi\left(\frac{\sqrt{\theta}(\lambda-1)}{\sqrt{\lambda+1}}\right) + 2F_{BVN}\left(-\frac{\sqrt{\theta}(\lambda-1)}{\sqrt{\lambda+1}}, 0; -\frac{2\sqrt{\lambda}}{\lambda+1}\right), \lambda \neq 1$$

Notation and definitions are the same as in Section 4.2. Recall that  $X_{\bar{D}}$  and  $X_D$  are independent random variables such that  $X_{\bar{D}} \sim N(\sqrt{\theta}, 1)$  and  $X_D \sim N(\lambda\sqrt{\theta}, \lambda)$ , and  $A_1, A_2, A_3$  and  $A_4$  are mutually exclusive events defined by

$$\begin{aligned} A_1 &= \left\{ X_D > X_{\bar{D}}, X_D > -X_{\bar{D}} \right\}, A_2 = \left\{ X_D < X_{\bar{D}}, X_D < -X_{\bar{D}} \right\} \\ A_3 &= \left\{ X_D > X_{\bar{D}}, X_D < -X_{\bar{D}} \right\}, A_4 = \left\{ X_D < X_{\bar{D}}, X_D > -X_{\bar{D}} \right\} \end{aligned} \quad (A1)$$

with their corresponding  $(X_{\bar{D}}, X_D)$  subspaces graphically displayed in Figure 2a. In Section (4.2.1) I showed that

$$AUC_{\chi^2} = \begin{cases} Pr(A_1) + Pr(A_2), & \lambda > 1 \\ Pr(A_3) + Pr(A_4), & \lambda < 1 \end{cases} \quad (A2)$$

Define  $X = X_{\bar{D}}$  if  $D = 0$ ,  $X = X_D$  if  $D = 1$ . It is easy to show that  $X$  has a binormal distribution with  $a$  and  $b$  defined by (11), i.e.,  $X$  is a binormal DV that corresponds to a bi-chi-squared ROC curve with parameters  $\lambda$  and  $\theta$ . The binormal AUC for  $X$  is given by

$$\begin{aligned} AUC_{bin} &= Pr\left(X_D > X_{\bar{D}}\right) \\ &= \Phi\left(\frac{\sqrt{\theta}(\lambda-1)}{\sqrt{1+\lambda}}\right) \\ &= Pr(A_1) + Pr(A_3) \end{aligned} \quad (A3)$$

It follows from (A2) and (A3) that

$$AUC_{\chi^2} = \begin{cases} AUC_{bin} + Pr(A_2) - Pr(A_3), & \lambda > 1 \\ AUC_{bin} + Pr(A_4) - Pr(A_1), & \lambda < 1 \end{cases}$$

Thus to prove (21) I only need to show

$$2F_{BVN}\left(-\frac{\sqrt{\theta}(\lambda-1)}{\sqrt{\lambda+1}}, 0; -\frac{2\sqrt{\lambda}}{\lambda+1}\right) = \begin{cases} Pr(A_2) - Pr(A_3), & \lambda > 1 \\ Pr(A_4) - Pr(A_1), & \lambda < 1 \end{cases} \quad (A4)$$



To prove (A4) I first define the transformation

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = P' \Sigma^{-\frac{1}{2}} \begin{pmatrix} X_{\bar{D}} \\ X_D \end{pmatrix} \quad (\text{A5})$$

where

$$P = \frac{\sqrt{\lambda}}{\sqrt{1+\lambda}} \begin{pmatrix} \frac{1}{\sqrt{\lambda}} & -1 \\ 1 & \frac{1}{\sqrt{\lambda}} \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}, \Sigma^{-\frac{1}{2}} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\sqrt{\lambda}} \end{pmatrix}$$

Note that  $\Sigma$  is the covariance matrix of  $(X_{\bar{D}}, X_D)$  and  $\Sigma^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} = \Sigma^{-1}$ .

The transformation defined by (A5) consists of two steps: (1) The product  $\Sigma^{-\frac{1}{2}} (X_{\bar{D}}, X_D)'$  results in a rescaling of  $X_{\bar{D}}$  and  $X_D$ , resulting in independent standard normal random variables  $V_1 = X_{\bar{D}}$  and  $V_2 = X_D / \sqrt{\lambda}$ . This transformation is illustrated in Figure 2b, which shows the contour lines for the  $(V_1, V_2)$  density function; the subspaces corresponding to events  $A_1, A_2, A_3$  and  $A_4$ ; and the orthogonal column vectors of the  $P$  matrix (up to a scale constant), denoted by  $w_1$  and  $w_2$ . Note that  $w_2$  coincides with the line  $V_2 = V_1 / \sqrt{\lambda}$ , which is the transformed  $X_D = X_{\bar{D}}$  line from Figure 2a. (2) Multiplying by  $P'$  results in a counterclockwise rotation of the  $(V_1, V_2)$  axes such that  $W_1$  and  $W_2$  are the components of the projections onto the  $w_1$  and  $w_2$  vectors (after unit-length normalization), respectively, in Figure 2b; i.e.,  $W_1 = \frac{\sqrt{\lambda}}{\sqrt{1+\lambda}} (V_1 / \sqrt{\lambda} + V_2)$  and  $W_2 = \frac{\sqrt{\lambda}}{\sqrt{1+\lambda}} (-V_1 + V_2 / \sqrt{\lambda})$ .

It follows from (A5) that

$$(W_1, W_2) \sim MVN \left( \sqrt{\theta} \sqrt{\lambda+1}, 0, 1, 1, 0 \right) \quad (\text{A6})$$

The independence of  $W_1$  and  $W_2$  and  $E(W_2) = 0$  imply that the density of  $(W_1, W_2)$ , denoted by  $f(W_1, W_2)$ , is symmetric about the line  $w_2 = 0$ ; i.e.,  $f(w_1, w_2) = f(w_1, w_2) (w_1, -w_2)$  for all  $(w_1, w_2)$ .

### A.1. Derivation for $\lambda > 1$

Figure 2c shows the contour lines for the  $(W_1, W_2)$  density function and the subspaces corresponding to the events  $A_1, A_2, A_3$  and  $A_4$  in the  $(W_1, W_2)$  space. Comparing Figure 2c to Figure 2a shows that the  $x_2 = x_1$  and  $x_2 = -x_1$  lines in Figure 2a correspond to the  $w_2 = 0$  and  $w_2 = -.5 [(\lambda - 1) / \sqrt{\lambda}] w_1$  lines in Figure 2c. In Figure 2c,  $A_2$  has been partitioned into two mutually exclusive subsets,  $A_{21}$  and  $A_{22}$  with the added line  $w_2 = .5 [(\lambda - 1) / \sqrt{\lambda}] w_1$  separating the two subsets:

$A_{21}=A_2 \cap \left\{ W_2 \leq .5 \left[ (\lambda - 1) / \sqrt{\lambda} \right] W_1 \right\}$  and  $A_{22}=A_2 \cap \left\{ W_2 > .5 \left[ (\lambda - 1) / \sqrt{\lambda} \right] W_1 \right\}$   
 with  $\Pr (A_1) = \Pr (A_{11}) + \Pr (A_{12})$ .

It follows from the symmetry of  $f_{(W_1, W_2)}$  about the line  $w_2 = 0$  that

$$\Pr (A_{21}) = \Pr (A_3)$$

Thus

$$\begin{aligned} \Pr (A_2) - \Pr (A_3) &= \Pr (A_2) - \Pr (A_{21}) \\ &= \Pr (A_{22}) \\ &= \Pr \left( W_2 < -\frac{(\lambda-1)}{2\sqrt{\lambda}} W_1, W_2 > \frac{(\lambda-1)}{2\sqrt{\lambda}} W_1 \right) \\ &= 2 \Pr \left( W_2 < -\frac{(\lambda-1)}{2\sqrt{\lambda}} W_1, W_2 > 0 \right) \end{aligned}$$

and hence we can write

$$\Pr (A_2) - \Pr (A_3) = 2 \Pr \left( W_2 + \frac{(\lambda - 1)}{2\sqrt{\lambda}} W_1 < 0, -W_2 < 0 \right) \tag{A7}$$

From (A6) it follows that

$$\left( -W_2, W_2 + \frac{(\lambda - 1)}{2\sqrt{\lambda}} W_1 \right) \sim MVN \left( 0, \frac{1}{2} \frac{\sqrt{\theta}}{\sqrt{\lambda}} (\lambda - 1) \sqrt{\lambda + 1}, 1, \frac{(1 + \lambda)^2}{4\lambda}, -1 \right) \tag{A8}$$

From (A7) and (A8) it follows that

$$\Pr (A_2) - \Pr (A_3) = 2F_{BVN} \left( -\frac{\sqrt{\theta} (\lambda - 1)}{\sqrt{\lambda + 1}}, 0; -\frac{2\sqrt{\lambda}}{\lambda + 1} \right)$$

### A.2. Derivation for $\lambda < 1$

Figure 2d is similar to Figure 2c, with the difference being that  $\lambda = 4$  for Figure 2c and  $\lambda = .$

25 for Figure 2d. For this reason, the line  $w_2 = - .5 \left[ (\lambda - 1) / \sqrt{\lambda} \right] w_1$  has a positive slope in Figure 2d, whereas it had a negative slope in Figure 2c. In Figure 2d,  $A_4$  has been partitioned into two mutually exclusive subsets,  $A_{41}$  and  $A_{42}$  with the added line

$w_2 = .5 \left[ (\lambda - 1) / \sqrt{\lambda} \right] w_1$  separating the two subsets:

$A_{41}=A_4 \cap \left\{ W_2 \leq .5 \left[ (\lambda - 1) / \sqrt{\lambda} \right] W_1 \right\}$  and  $A_{42}=A_4 \cap \left\{ W_2 > .5 \left[ (\lambda - 1) / \sqrt{\lambda} \right] W_1 \right\}$ ,  
 with  $\Pr (A_4) = \Pr (A_{41}) + \Pr (A_{42})$ .

It follows from the symmetry of  $f_{(W_1, W_2)}$  about the line  $w_2 = 0$  that

$$Pr(A_{41}) = Pr(A_1)$$

Thus

$$\begin{aligned} Pr(A_4) - Pr(A_1) &= Pr(A_4) - Pr(A_{41}) \\ &= Pr(A_{42}) \\ &= Pr\left(W_2 < -\frac{(\lambda-1)}{2\sqrt{\lambda}}W_1, W_2 > \frac{(\lambda-1)}{2\sqrt{\lambda}}W_1\right) \\ &= 2Pr\left(W_2 < -\frac{(\lambda-1)}{2\sqrt{\lambda}}W_1, W_2 > 0\right) \end{aligned}$$

and hence we can write

$$Pr(A_4) - Pr(A_1) = 2Pr\left(W_2 + \frac{(\lambda-1)}{2\sqrt{\lambda}}W_1 < 0, -W_2 < 0\right) \tag{A9}$$

Note that the right side of (A9) is identical in form to the right side of (A7). From (A9) and (A8) it follows that

$$Pr(A_4) - Pr(A_1) = 2F_{BVN}\left(-\frac{\sqrt{\theta}(\lambda-1)}{\sqrt{\lambda+1}}, 0; -\frac{2\sqrt{\lambda}}{\lambda+1}\right)$$

## B. Partial AUC derivation

In this section I derive the formula for the partial area under the bi-chi-squared partial ROC curve,  $pAUC(0, f_{pf_0})$ , where  $0 < f_{pf_0} < 1$ . For the derivations I use the following result [2, 28].

### Result 2

Let  $Y$  denote a decision variable having continuous conditional distributions, and let  $Y_{\bar{D}}$  and  $Y_D$  denote independent random variables having the same distributions as the nondiseased and diseased conditional distributions of  $Y$ , respectively. Let  $\xi_{f_{pf_0}}$  denote the threshold corresponding to  $f_{pf_0}$ ; *i.e.*,  $Pr(Y_{\bar{D}} > \xi_{f_{pf_0}}) = f_{pf_0}$ . Then

$$pAUC(0, f_{pf_0}) = Pr\left[Y_D > Y_{\bar{D}}, Y_{\bar{D}} > \xi_{f_{pf_0}}\right] \tag{B1}$$

Let  $\rho, Y_D^*, Y_{\bar{D}}^*, X_{\bar{D}}, X_D$  be defined as in Section 4.2.1, and again let  $F_\nu$  denote the distribution function for a random variable having a  $\chi_{1;\nu}^2$  distribution.

**B.1. Derivation and results for  $\lambda > 1$**

Define  $\xi_{fpf_0}$  by  $Pr\left(\frac{Y_D^*}{D} > \xi_{fpf_0}\right) = fpf_0$ , or equivalently

$$\xi_{fpf_0} = F_{\theta}^{-1}(1 - fpf_0)$$

It follows from (15) and (B1) that the bi-chi-squared partial AUC is given by

$$\begin{aligned} pAUC_{fpf}(0, fpf_0) &= Pr\left[Y_D^* > Y_{\bar{D}}^*, Y_{\bar{D}}^* > \xi_{fpf_0}\right] \\ &= Pr\left(X_D^2 > X_{\bar{D}}^2, X_{\bar{D}}^2 > \xi_{fpf_0}\right) \\ &= Pr\left(X_D > X_{\bar{D}}, X_{\bar{D}} > \sqrt{\xi_{fpf_0}}\right) + Pr\left(X_D < X_{\bar{D}}, X_{\bar{D}} < -\sqrt{\xi_{fpf_0}}\right) + Pr\left(X_D < -X_{\bar{D}}, X_{\bar{D}} > \sqrt{\xi_{fpf_0}}\right) + Pr\left(X_D > -X_{\bar{D}}, X_{\bar{D}} < -\sqrt{\xi_{fpf_0}}\right) \end{aligned}$$

or equivalently,

$$\begin{aligned} pAUC_{fpf}(0, fpf_0) &= Pr\left(-\left(X_D - X_{\bar{D}}\right) < 0, -X_{\bar{D}} < -\sqrt{\xi_{fpf_0}}\right) \\ &+ Pr\left(X_D - X_{\bar{D}} < 0, X_{\bar{D}} < -\sqrt{\xi_{fpf_0}}\right) \\ &+ Pr\left(X_D + X_{\bar{D}} < 0, -X_{\bar{D}} < -\sqrt{\xi_{fpf_0}}\right) \\ &+ Pr\left(-\left(X_D + X_{\bar{D}}\right) < 0, X_{\bar{D}} < -\sqrt{\xi_{fpf_0}}\right) \end{aligned} \tag{B2}$$

Noting that

$$\begin{aligned} \left(X_D + X_{\bar{D}}, X_{\bar{D}}\right) &\sim MVN\left(\sqrt{\theta}(\lambda + 1), \sqrt{\theta}, \lambda + 1, 1, 1\right) \\ \left(X_D - X_{\bar{D}}, X_{\bar{D}}\right) &\sim MVN\left(\sqrt{\theta}(\lambda - 1), \sqrt{\theta}, \lambda + 1, 1, -1\right) \end{aligned}$$

and defining

$$\rho_1 = \frac{-1}{\sqrt{\lambda + 1}}$$

it follows from (B2) that

$$\begin{aligned}
 pAUC_{f_{pf}}(0, f_{pf_0}) &= F_{BVN} \left( \frac{\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}, \sqrt{\theta} - \sqrt{\xi_{f_{pf_0}; \rho_1}} \right) \\
 &+ F_{BVN} \left( \frac{-\sqrt{\theta}(\lambda - 1)}{\sqrt{\lambda + 1}}, -\sqrt{\theta} - \sqrt{\xi_{f_{pf_0}; \rho_1}} \right) \\
 &+ F_{BVN} \left( \frac{-\sqrt{\theta}(\lambda + 1)}{\sqrt{\lambda + 1}}, \sqrt{\theta} - \sqrt{\xi_{f_{pf_0}; \rho_1}} \right) \\
 &+ F_{BVN} \left( \frac{\sqrt{\theta}(\lambda + 1)}{\sqrt{\lambda + 1}}, -\sqrt{\theta} - \sqrt{\xi_{f_{pf_0}; \rho_1}} \right) \tag{B3}
 \end{aligned}$$

Let  $u_1$  and  $u_2$  be defined by (17) and define

$$\begin{aligned}
 y_1 &= \sqrt{\theta} - \sqrt{\xi_{f_{pf_0}}} = \sqrt{\theta} - \sqrt{F_{\theta}^{-1}(1 - f_{pf_0})} \\
 y_2 &= \sqrt{\theta} + \sqrt{\xi_{f_{pf_0}}} = \sqrt{\theta} + \sqrt{F_{\theta}^{-1}(1 - f_{pf_0})} \tag{B4}
 \end{aligned}$$

From (B3) and (B4) it follows that

$$\begin{aligned}
 pAUC_{f_{pf}}(0, f_{pf_0}) &= F_{BVN}(u_1, y_1; \rho_1) + F_{BVN}(-u_1, -y_2; \rho_1) + F_{BVN}(-u_2, y_1; \rho_1) + F_{BVN}(u_2, -y_2; \rho_1) \\
 &\tag{B5}
 \end{aligned}$$

## B.2. Derivation and results for $\lambda < 1$

It follows from (B1) that

$$pAUC_{f_{pf}}(0, f_{pf_0}) = Pr \left[ -Y_D^* > -Y_{\frac{D}{\lambda}}^*, -Y_{\frac{D}{\lambda}}^* > \tilde{\xi}_{f_{pf_0}} \right]$$

where  $\tilde{\xi}_{f_{pf_0}}$  is such that  $Pr \left( -Y_{\frac{D}{\lambda}}^* > \tilde{\xi}_{f_{pf_0}} \right) = f_{pf}$ . It follows that

$$Pr \left( Y_{\frac{D}{\lambda}}^* < -\tilde{\xi}_{f_{pf_0}} \right) = f_{pf_0}$$

and

$$-\tilde{\xi}_{f_{pf_0}} = \xi_{1-f_{pf_0}}$$

We have

$$\begin{aligned}
 pAUC_{fpf}(0, fpf_0) &= Pr \left[ -Y_D^* > -Y_{\bar{D}}^*, -Y_{\bar{D}}^* > \tilde{\xi}_{fpf_0} \right] \\
 &= Pr \left[ Y_D^* < Y_{\bar{D}}^*, Y_{\bar{D}}^* < -\tilde{\xi}_{fpf_0} \right] \\
 &= Pr \left( Y_{\bar{D}}^* < -\tilde{\xi}_{fpf_0} \right) - Pr \left( Y_D^* > Y_{\bar{D}}^*, Y_{\bar{D}}^* < -\tilde{\xi}_{fpf_0} \right) \\
 &= fpf_0 - Pr \left( Y_D^* > Y_{\bar{D}}^*, Y_{\bar{D}}^* < -\tilde{\xi}_{fpf_0} = \xi_{1-fpf_0} \right) \\
 &= fpf_0 - \left[ Pr \left( Y_D^* > Y_{\bar{D}}^* \right) - Pr \left( Y_D^* > Y_{\bar{D}}^*, Y_{\bar{D}}^* > \xi_{1-fpf_0} \right) \right] \quad (B6)
 \end{aligned}$$

Because the derivations for  $Pr \left( Y_D^* > Y_{\bar{D}}^* \right)$  and  $Pr \left( Y_D^* > Y_{\bar{D}}^*, Y_{\bar{D}}^* > \xi_{fpf_0} \right)$  for  $\lambda > 1$  in Sections 4.2.1 and B.1, respectively, do not depend on the value of  $\lambda$ , it follows that

$Pr \left( Y_D^* > Y_{\bar{D}}^* \right)$  and  $Pr \left( Y_D^* > Y_{\bar{D}}^*, Y_{\bar{D}}^* > \xi_{1-fpf_0} \right)$  in (B6) can be replaced by the right side of equations (16) and (B5), respectively, with  $fpf_0$  replaced by  $1 - fpf_0$  in the definitions of  $y_1$  and  $y_2$  (B4). It follows that

$$\begin{aligned}
 pAUC_{fpf}(0, fpf_0) &= fpf_0 - [F_{BVN}(u_1, u_2; \rho) + F_{BVN}(-u_1, -u_2; \rho)] \\
 &\quad + [F_{BVN}(u_1, y_3; \rho_1) + F_{BVN}(-u_1, -y_4; \rho_1) + F_{BVN}(-u_2, y_3; \rho_1) + F_{BVN}(u_2, -y_4; \rho_1)]
 \end{aligned} \quad (B7)$$

where  $u_1$  and  $u_2$  are defined by (17) and

$$\begin{aligned}
 y_3 &= \sqrt{\theta} - \sqrt{\xi_{1-fpf_0}} = \sqrt{\theta} - \sqrt{F_{\theta}^{-1}(fpf_0)} \\
 y_4 &= \sqrt{\theta} + \sqrt{\xi_{1-fpf_0}} = \sqrt{\theta} + \sqrt{F_{\theta}^{-1}(fpf_0)}
 \end{aligned}$$

### C. Relationships between $\lambda$ , $\theta$ and AUC for the bi-chi-squared ROC curve

In this section I show for fixed  $\theta$  that AUC is an increasing (decreasing) function of  $\lambda$  if  $\lambda > 1$  ( $\lambda < 1$ ).

To show this result for  $\lambda > 1$ , fix  $\theta > 0$  and let  $\lambda_1$  and  $\lambda_2$  be such that  $0 < \lambda_1 < \lambda_2$ . Let  $Y_1$  and  $Y_2$  denote bi-chi-squared DVs having parameter pairs  $(\theta, \lambda_1)$  and  $(\theta, \lambda_2)$ , respectively, and let  $AUC_1$  and  $AUC_2$  denote their corresponding AUCs. Let  $c > 0$  denote an arbitrary threshold and let  $fpf_j(c)$  and  $tpf_j(c)$  denote the corresponding fpf and tpf values:  $fpf_j(c) = Pr$

$(Y_j > c | D = 0)$ ,  $\text{tpf}_j(c) = \Pr(Y_j > c | D = 1)$  for  $i = 1, 2$ . Note that  $\text{fpf}_1(c) = \text{fpf}_2(c)$  because  $(Y_i | D = 0) \sim \chi_{1;\theta}^2, i = 1, 2$ . Thus  $\text{tpf}_2(c) > \text{tpf}_1(c)$  for all  $c < 0$  implies  $\text{AUC}_2 > \text{AUC}_1$ , and hence it suffices to show  $\text{tpf}_2(c) > \text{tpf}_1(c)$ .

Let  $V_{\mathcal{E}}$  denote a random variable having a  $\chi_{1;\xi}^2$  distribution. It is straightforward to show

$$\Pr(\lambda_2 V_{\lambda_2 \theta} > c) > \Pr(\lambda_1 V_{\lambda_2 \theta} > c) \quad (\text{C1})$$

and

$$\Pr(V_{\lambda_2 \theta} > c / \lambda_1) > \Pr(V_{\lambda_1 \theta} > c / \lambda_1) \quad (\text{C2})$$

To show (C2), note that  $\Pr(V_{\lambda_i \theta} > c / \lambda_1) = 1 - \Pr(|Z + \sqrt{\lambda_i \theta}| < \sqrt{c / \lambda_1})$ ,  $i = 1, 2$ , where  $Z$  is a standard normal random variable. Using (C1) and (C2) we can write  $\text{tpf}_2(c) = \Pr(Y_2 > c | D = 1) = \Pr(\lambda_2 V_{\lambda_2 \theta} > c) > \Pr(\lambda_1 V_{\lambda_2 \theta} > c) = \Pr(V_{\lambda_2 \theta} > c / \lambda_1) > \Pr(V_{\lambda_1 \theta} > c / \lambda_1) = \Pr(\lambda_1 V_{\lambda_1 \theta} > c) = \Pr(Y_1 > c | D = 1) = \text{tpf}_1(c)$ ; thus we have shown  $\text{tpf}_2(c) > \text{tpf}_1(c)$  for arbitrary  $c$ . The proof for  $\lambda < 1$  is similar.

It can similarly be shown for fixed  $\lambda = 1$  that AUC is an increasing function of  $\theta$ .

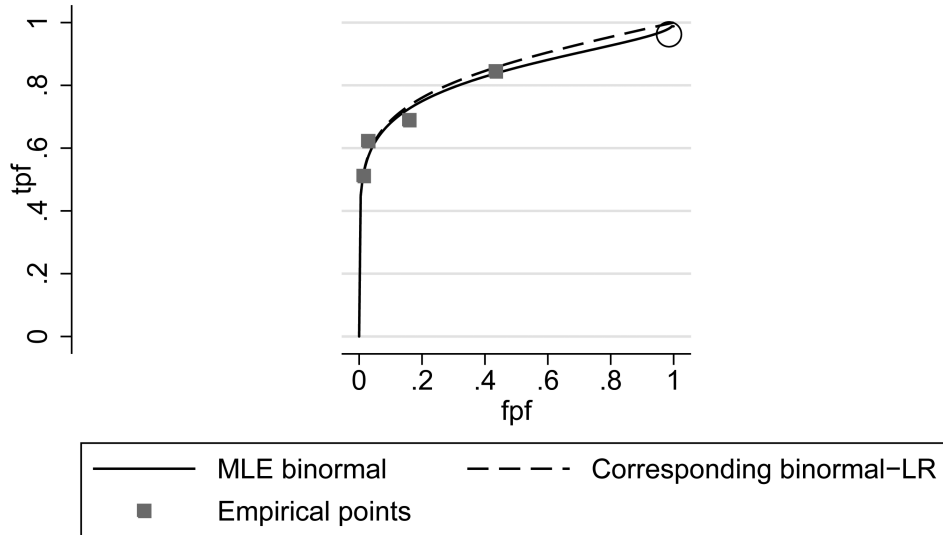
## References

1. Metz CE. ROC methodology in radiologic Imaging. *Investigative Radiology*. 1986; 21:720–733. [PubMed: 3095258]
2. Pepe, M. The statistical evaluation of medical tests for classification and prediction. Oxford; New York: 2003.
3. Zhou, XH., Obuchowski, NA., McClish, DK. Statistical methods in diagnostic medicine. Wiley; Hoboken: 2011.
4. Hanley JA. The robustness of the binormal assumptions used in fitting ROC curves. *Medical Decision Making*. 1988; 8:197–203. [PubMed: 3398748]
5. Hanley JA. The use of the 'binormal' model for parametric ROC analysis of quantitative diagnostic tests. *Statistics in Medicine*. 1996; 15:1575–1585. [PubMed: 8855483]
6. Hajian-Tilaki KO, Hanley JA, Joseph L, Collet JP. A comparison of parametric and nonparametric approaches to ROC analysis of quantitative diagnostic tests. *Medical Decision Making*. 1997; 17:94–102. [PubMed: 8994156]
7. Swets JA. Form of empirical ROCS in discrimination and diagnostic tasks: implications for theory and measurement of performance. *Psychological Bulletin*. 1986; 99:181–198. [PubMed: 3515382]
8. Egan, JP. Signal detection theory and ROC analysis. Academic Press; New York: 1975.
9. Van Dyke, CW., White, RD., Obuchowski, NA., Geisinger, MA., Lorig, RJ., Meziare, MA. Cine MRI in the diagnosis of thoracic aortic dissection.. 79th RSNA Meetings; Chicago, IL. November 28 - December 3, 1993;
10. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. *Investigative Radiology*. 1989; 24:234–245. [PubMed: 2753640]
11. Pan XC, Metz CE. The “proper” binormal model: parametric receiver operating characteristic curve estimation with degenerate data. *Academic Radiology*. 1997; 4:380–389. [PubMed: 9156236]

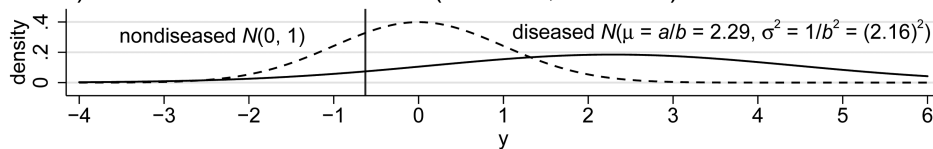
12. Metz CE, Pan XC. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *Journal of Mathematical Psychology*. 1999; 43:1–33. [PubMed: 10069933]
13. Hillis SL, Berbaum KS. Using the mean-to-sigma ratio as a measure of the improperness of binormal ROC curves. *Academic Radiology*. 2011; 18:143–154. DOI: 10.1016/j.acra.2010.09.002. [PubMed: 21232682]
14. Bamber D. Area above ordinal dominance graph and area below receiver operating characteristic graph. *Journal of Mathematical Psychology*. 1975; 12:387–415.
15. Thompson ML, Zucchini W. On the statistical analysis of ROC curves. *Statistics in Medicine*. 1989; 8:1277–1290. DOI: 10.1002/sim.4780081011. [PubMed: 2814075]
16. McClish DK. Analyzing a portion of the ROC curve. *Medical Decision Making*. 1989; 9:190–195. [PubMed: 2668680]
17. Jiang YL, Metz CE, Nishikawa RM. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology*. 1996; 201:745–750. [PubMed: 8939225]
18. Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. *Academic Radiology*. 2007; 14:814–829. DOI: 10.1016/j.acra.2007.03.012. [PubMed: 17574132]
19. Schartz, KM., Hillis, SL., Pesce, LL., Berbaum, KS. [July 22, 2015] OR-DBM MRMC (Version 2.5) [Computer software]. Available for download from <http://perception.radiology.uiowa.edu>.
20. Hillis, SL., Schartz, KM., Berbaum, KS. [July 22, 2015] OR/DBM MRMC for SAS (Version 3.0) [Computer software]. Available for download from <http://perception.radiology.uiowa.edu>.
21. Gönen, M. Analyzing receiver operating characteristic curves with SAS. SAS Institute; 2007.
22. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*. 1998; 17(9):1033–1053. [PubMed: 9612889]
23. Leone FC, Nelson LS, Nottingham RB. The folded normal distribution. *Technometrics*. 1961; 3:543–550.
24. Dorfman DD, Berbaum KS, Metz CE, Lenth RV, Hanley JA, AbuDagga H. Proper receiver operating characteristic analysis: The bigamma model. *Academic Radiology*. 1997; 4:138–149. [PubMed: 9061087]
25. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data - Part II. A formal model. *Academic Radiology*. 2000; 7:427–437. [PubMed: 10845402]
26. Mossman D, Peng H. Constructing "proper" ROCs from ordinal response data using weighted power functions. *Medical Decision Making*. 2014; 34(4):523–535. [PubMed: 24029820]
27. Samuelson FW. Two-sample models with monotonic likelihood ratios for ordinal regression. *Journal of Mathematical Psychology*. 2011; 55:223–228.
28. Dodd LE, Pepe MS. Partial AUC estimation and regression. *Biometrics*. 2003; 59:614–623. [PubMed: 14601762]



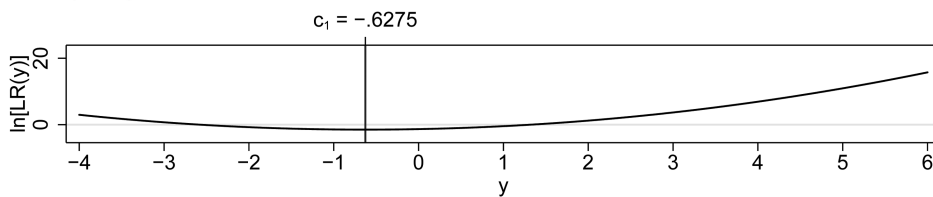
a) Binormal ( $a = 1.06, b = 0.46$ ) and corresponding binormal-LR ROC curves



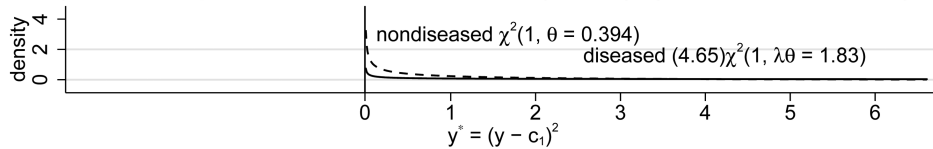
b) Binormal distribution for  $Y$  ( $a = 1.06, b = 0.46$ )



c) Log-likelihood ratio for  $Y$



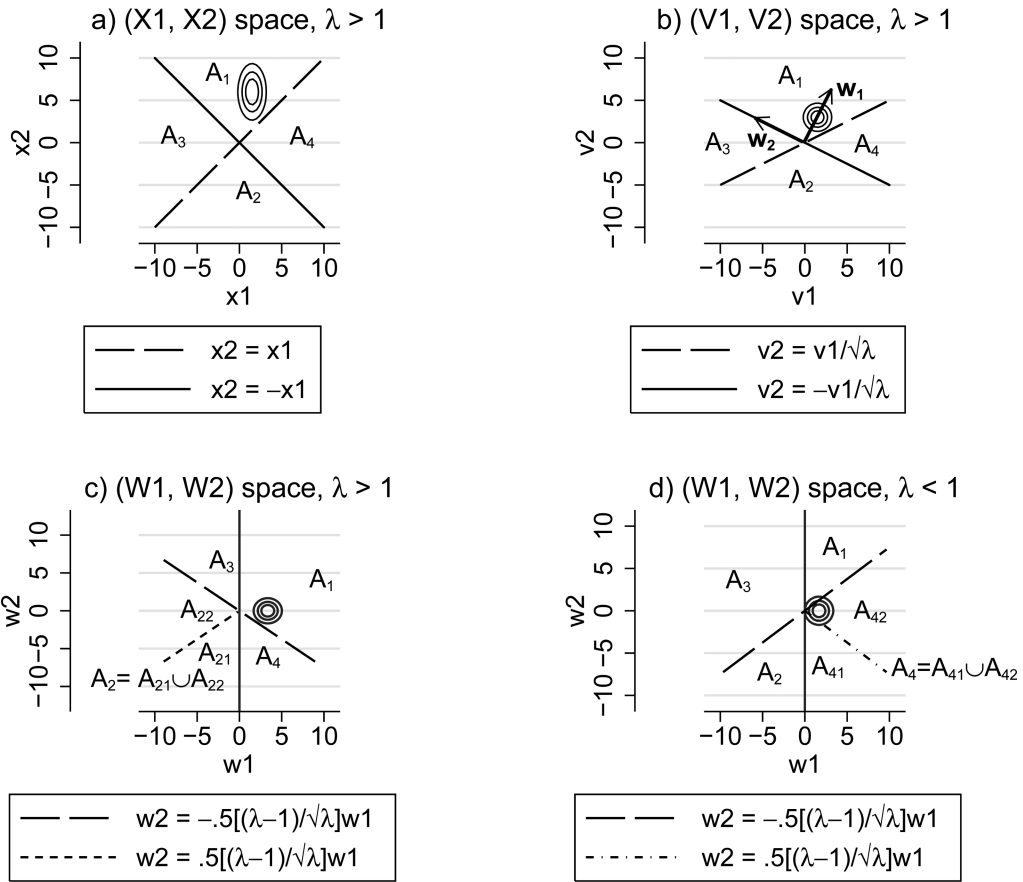
d) Bi-chi-squared distribution for  $Y^* = (Y - c_1)^2$  ( $\theta = 0.394, \lambda = 4.65$ )



**Figure 1.**

Binormal ROC curve (with  $b < 1$ ) and related plots. The binormal ROC curve (solid line) in (a) is the maximum likelihood estimate for the data given in Table 1; the circle indicates the hook in it. For the corresponding binormal random variable  $Y$ , the binormal conditional densities and the log-likelihood ratio are displayed in (b) and (c), respectively, where  $c_1 = -ab/(1 - b^2)$  is the threshold where the log-likelihood ratio attains its minimum. The bi-chi-squared conditional densities for  $Y^* = (Y - c_1)^2$  are displayed in (d), and the binormal-LR ROC curve, based on  $Y^*$ , is the dashed line in (a). Notes: the nondiseased and diseased

densities are denoted by the dashed and solid lines, respectively, in (b) and (d);  $\chi^2(1, \nu)$  denotes a chi-squared distribution with one degree of freedom and noncentrality parameter  $\nu$ ; LR: likelihood ratio function;  $\lambda = 1/b^2$ ;  $\theta = a^2 b^2 / (1 - b^2)^2$ .



**Figure 2.**

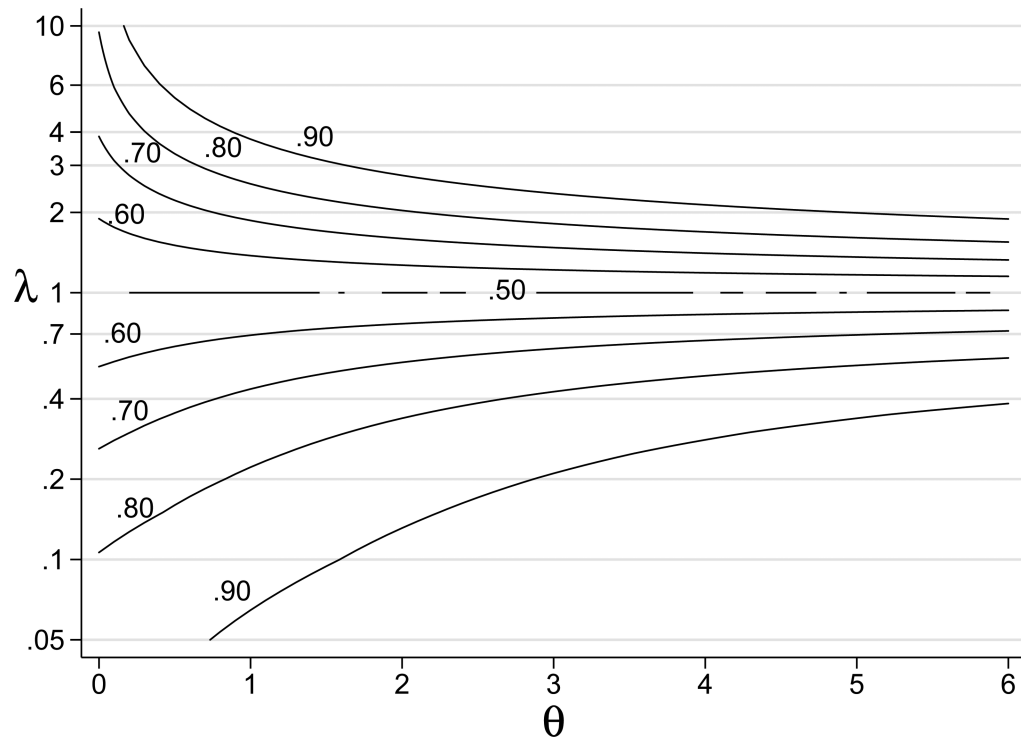
Plots related to AUC proofs.  $X_1$  and  $X_2$  are independent random variables such that

$$X_1 \sim N(\sqrt{\theta}, 1) \text{ and } X_2 \sim N(\lambda\sqrt{\theta}, \lambda); \text{ thus } X_1^2 \sim \chi_{1;\theta}^2 \text{ and } X_2^2 \sim \lambda\chi_{1;\lambda\theta}^2.$$

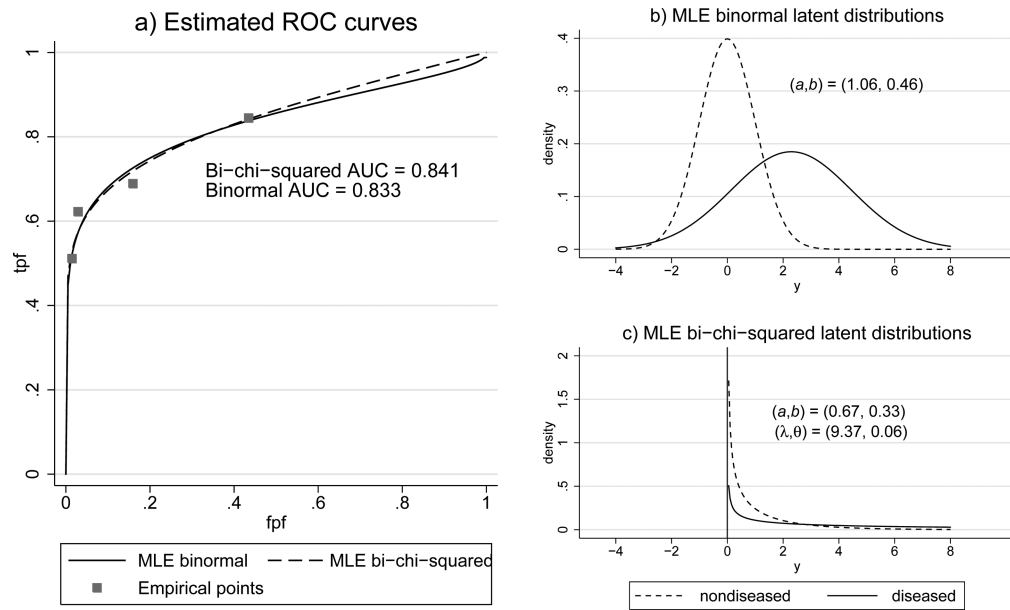
For plots (a), (b) and (c),  $\theta = 2.25$ ,  $\lambda = 4$ ; for plot (d)  $\theta = 2.25$ ,  $\lambda = 0.25$ . Plot (a) shows the contour plot of the  $(X_1, X_2)$  density function and the subspaces corresponding to events  $A_1, A_2, A_3$  and  $A_4$ . Plot (b) shows the transformed contour plot and subspaces in the  $(V_1, V_2)$  space, where  $V_1 = X_1$  and  $V_2 = X_2/\sqrt{\lambda}$ . Plot (b) also shows the column vectors of the orthogonal matrix

$$P = \frac{\sqrt{\lambda}}{\sqrt{1+\lambda}} \left[ \left( \frac{1}{\sqrt{\lambda}}, 1 \right)', \left( -1, \frac{1}{\sqrt{\lambda}} \right)' \right],$$

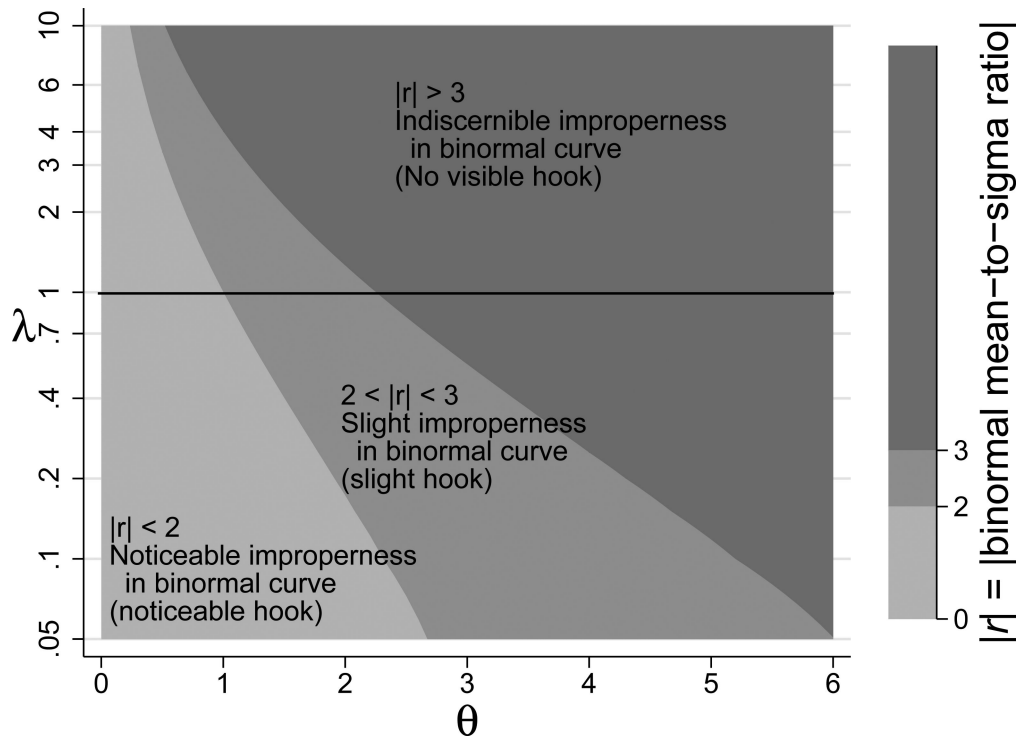
denoted by  $w_1$  and  $w_2$ , up to a scaling constant. Plot (c) shows the transformed contour plot and subspaces in the  $(W_1, W_2)$  space, where  $(W_1, W_2)' = P'(V_1, V_2)'$ ; i.e.  $W_1$  and  $W_2$  are the components of the projections onto the normalized  $w_1$  and  $w_2$  vectors in (b). Plot (d) is similar to (c), except that  $\lambda = 0.25$ .



**Figure 3.** Contour map of AUC in the bi-chi-squared parameter space. There is a one-to-one relationship between the values of  $\lambda$  and AUC for each value of  $\theta$  in the region  $\lambda > 1$  and also in the region  $\lambda < 1$ . Notes: for  $\lambda = 1$  the ROC curve is the chance line with AUC = 0.5; the  $\lambda$ -axis is scaled logarithmically.



**Figure 4.** Maximum-likelihood estimated binormal and bi-chi-squared ROC curves with corresponding latent decision variable distributions for Table 1a data (reader 5, modality 1). Note: MLE: maximum likelihood estimated.



**Figure 5.** Improperness of the binormal model corresponding to a bi-chi-squared model. Improperness categories are defined in terms of  $|r|$ , where  $r = a/(1 - b)$  is the mean-to-sigma ratio for the corresponding binormal model, as discussed by Hillis and Berbaum [13]. Notes:  $a$  and  $b$  define the binormal model corresponding to the bi-chi-squared model defined by  $(\theta, \lambda)$ , i.e.,  $a = \sqrt{\theta(\lambda - 1)^2/\lambda}$  and  $b = 1/\sqrt{\lambda}$ ; the  $\lambda$ -axis is scaled logarithmically; values of  $r$  for  $\lambda = 1$  are not included because  $r$  is not defined if  $\lambda = 1$ .

**Table 1**

ROC rating data for a radiologist [9].

Status	Rating					Total
	1	2	3	4	5	
Nondiseased	39	19	9	1	1	69
Diseased	7	7	3	5	23	45

The estimated binormal ROC curve is displayed in Figure 1a.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Binormal likelihood-ratio ROC curve: original definition and corresponding bi-chi-squared formulation.

a) Original definition of  $ROC_{LR;a,b}$  (based on log-likelihood-ratio):

$$U = \ln[LR_Y(Y)] = g(a, b) + \frac{(1-b^2)}{2}(Y - c_1)^2$$

$$ROC_{LR;a,b} = \{(\Pr(U > c|D=0), \Pr(U > c|D=1)) : -\infty < c < \infty\}$$

b) Bi-chi-square formulation:

$$Y^* \equiv (Y - c_1)^2$$

$$Y^* | (D = 0) \sim \chi^2_{1;\theta}$$

$$Y^* | (D = 1) \sim \lambda \chi^2_{1;\lambda\theta}$$

$$ROC_{\chi^2;\lambda,\theta} = \begin{cases} \{\Pr(Y^* \geq c | D = 0), \Pr(Y^* \geq c | D = 1) : c \geq 0\}, & \lambda \geq 1 \\ \{\Pr(-Y^* \geq c | D = 0), \Pr(-Y^* \geq c | D = 1) : c \leq 0\}, & \lambda < 1 \end{cases}$$

or equivalently,

$$ROC_{\chi^2;\lambda,\theta} = \{(\Pr(W > c|D=0), \Pr(W > c|D=1)) : -\infty < c < \infty\}$$

where

$$W = \begin{cases} Y^*, & \lambda \geq 1 \\ -Y^*, & \lambda < 1 \end{cases}$$

c) Relationship ( $b > 1$ ):

$$ROC_{LR;a,b} = ROC_{\chi^2;\lambda,\theta}$$

where

$$\theta = \frac{a^2 b^2}{(1 - b^2)^2}$$

$$\lambda = \frac{1}{b^2}$$

Notes:  $Y$  is binormal with conditional distributions  $Y | (D=0) \sim N(0, 1)$ ,  $Y | (D=1) \sim N(ab, 1/b^2)$ ;  $LR_Y(\cdot)$  is the likelihood ratio function for  $Y$ ;  $c_1 = -ab/(1 - b^2)$  is the threshold where  $LR_Y(\cdot)$  attains its minimum ( $b < 1$ ) or maximum ( $b > 1$ );  $ROC_{LR;a,b}$  is the binormal likelihood-ratio ROC curve with parameters  $a$  and  $b$ ;  $ROC_{\chi^2;\lambda,\theta}$  is the bi-chi-squared ROC curve with parameters  $\lambda > 0$  and  $\theta > 0$ .



**Table 3**

Tpf, fpf, AUC, and pAUC for the bi-chi-squared ROC curve.

a) Bi-chi-squared ROC curve:

$$ROC_{x, \lambda, \theta}^2 = \begin{cases} \{1 - F_{\theta}(c), 1 - F_{\lambda\theta}(c/\lambda); c \geq 0\}, & \lambda > 1 \\ \{F_{\theta}(c), F_{\lambda\theta}(c/\lambda); c \geq 0\}, & \lambda < 1 \end{cases}$$

b) Tpf for fixed fpf,  $0 < \text{fpf} < 1$ :

$$\text{tpf}(\text{fpf}) = \begin{cases} 1 - F_{\lambda\theta} \left[ \frac{1}{\lambda} F_{\theta}^{-1}(1 - \text{fpf}) \right], & \lambda > 1 \\ F_{\lambda\theta} \left[ \frac{1}{\lambda} F_{\theta}^{-1}(\text{fpf}) \right], & \lambda < 1 \end{cases}$$

c) Fpf for fixed tpf,  $0 < \text{tpf} < 1$ :

$$\text{fpf}(\text{tpf}) = \begin{cases} 1 - F_{\theta} \left[ \lambda F_{\lambda\theta}^{-1}(1 - \text{tpf}) \right], & \lambda > 1 \\ F_{\theta} \left[ \lambda F_{\lambda\theta}^{-1}(\text{tpf}) \right], & \lambda < 1 \end{cases}$$

d) AUC:

$$\text{AUC} = \begin{cases} F_{\text{BVN}}(u_1, \gamma_2; \rho) + F_{\text{BVN}}(-u_1, -u_2; \rho), & \lambda > 1 \\ 1 - [F_{\text{BVN}}(u_1, u_2; \rho) + F_{\text{BVN}}(-u_1, -u_2; \rho)], & \lambda < 1 \end{cases}$$

where

$$u_1 = \frac{\sqrt{\theta(\lambda-1)}}{\sqrt{\lambda+1}}, u_2 = \sqrt{\theta\sqrt{\lambda+1}}, \rho = -\frac{\lambda-1}{\lambda+1}$$

e) AUC alternative expression, proposed by Metz & Pan [11, 12]:

$$\text{AUC} = \Phi \left( \frac{\sqrt{\theta(\lambda-1)}}{\sqrt{\lambda+1}} \right) + 2F_{\text{BVN}} \left( -\frac{\sqrt{\theta(\lambda-1)}}{\sqrt{\lambda+1}}, 0; -\frac{2\sqrt{\lambda}}{\lambda+1} \right), \lambda \neq 1$$

f) Partial AUC:

$$\begin{aligned} \text{pAUC}_{\text{tpf}}(0, \text{fpf}_0) &= \begin{cases} F_{\text{BVN}}(u_1, \gamma_1; \rho_1) + F_{\text{BVN}}(-u_1, -\gamma_2; \rho_1) + F_{\text{BVN}}(-u_2, \gamma_1; \rho_1) + F_{\text{BVN}}(u_2, -\gamma_2; \rho_1), & \lambda > 1 \\ \text{fpf}_0 - F_{\text{BVN}}(u_1, u_2; \rho) - F_{\text{BVN}}(-u_1, -u_2; \rho) + F_{\text{BVN}}(u_1, \gamma_3; \rho_1) + F_{\text{BVN}}(-u_1, -\gamma_4; \rho_1) + F_{\text{BVN}}(-u_2, \gamma_3; \rho_1) + F_{\text{BVN}}(u_2, -\gamma_4; \rho_1), & \lambda < 1 \end{cases} \\ \text{pAUC}_{\text{tpf}}(\text{tpf}_0, 1) &= \text{AUC} - \text{pAUC}_{\text{tpf}}(0, \text{fpf}(\text{tpf}_0)) - \text{tpf}_0 \times [1 - \text{fpf}(\text{tpf}_0)] \end{aligned}$$

where  $\rho, u_1$  and  $u_2$  are defined in (d) and

$$\rho_1 = \frac{-1}{\sqrt{\lambda+1}}$$



**Table 4**

Maximum likelihood parameter estimates for binormal-LR ROC curves for Van Dyke [9] data for comparing spin-echo and cine MRI in detecting thoracic aortic dissection. For the bi-chi-squared algorithm, initial estimates of  $\lambda$  and  $\theta$  are functions of binomial parameter estimates and initial delta estimates are empirically estimated.

Modality	Reader	AUC		Bi-Chi parameter estimates	
		PROPROC	Bi-Chi	$\lambda$	$\theta$
1	1	0.934	0.934	3.418921	1.706011
1	2	0.891	0.891	3.172872	1.324854
1	3	0.908	0.929 (0.908)	2.532216 (46.929408)	3.239197 (0.006334)
1	4	0.977	0.977	786.713272	0.000017
1	5	0.841	0.841	9.366031	0.059426
2	1	0.952	0.952	3.788983	1.697356
2	2	0.926	0.926	73.205625	0.000024
2	3	0.930	0.930	3.940212	1.234458
2	4	1.000	1.000	1.283937	780.544368
2	5	0.943	0.943	12.075745	0.217397

Notes: Modality 1 = spin-echo MRI, 2 = cine MRI; PROPROC and Bi-Chi are the PROPROC and bi-chi-squared algorithms, respectively. The discrepancy between the PROPROC and bi-chi-squared AUC values for reader 3, modality 1 was resolved by rerunning the bi-chi-squared algorithm using several other starting values, which resulted in the values shown in parentheses.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript