Published in final edited form as:

Acad Radiol. 2011 February ; 18(2): 143–154. doi:10.1016/j.acra.2010.09.002.

Using the Mean-to-Sigma Ratio as a Measure of the Improperness of Binormal ROC Curves

Stephen L. Hillis, Ph.D. and

Center for Research in the Implementation of Innovative Strategies in Practice (CRIISP) Iowa City VA Medical Center, Iowa City, IA, U.S.A. Department of Biostatistics University of Iowa, Iowa City, IA, U.S.A

Kevin S. Berbaum, Ph.D.

Department of Radiology University of Iowa, Iowa City, IA, U.S.A

Abstract

Rationale and Objectives—A basic assumption for a meaningful diagnostic decision variable is that there is a monotone relationship between the decision variable and the likelihood of disease. This relationship, however, generally does not hold for the binormal model. As a result, ROC-curve estimation based on the binormal model produces improper ROC curves that are not concave over the entire domain and cross the chance line. Although in practice the “improperness” is typically not noticeable, there are situations where the improperness is evident. Presently, standard statistical software does not provide diagnostics for assessing the magnitude of the improperness.

Materials and Methods—We show how the mean-to-sigma ratio can be a useful, easy-to-understand and easy-to-use measure for assessing the magnitude of the improperness of a binormal ROC curve by showing how it is related to the chance-line crossing. We suggest an improperness criterion based on the mean-to-sigma ratio.

Results—Using a real-data example we illustrate how the mean-to-sigma ratio can be used to assess the improperness of binormal ROC curves, compare the binormal method with an alternative proper method, and describe uncertainty in a fitted ROC curve with respect to improperness.

Conclusions—By providing a quantitative and easily computable improperness measure, the mean-to-sigma ratio provides an easy way to identify improper binormal ROC curves and facilitates comparison of analysis strategies according to improperness categories in simulation and real-data studies.

Keywords

receiver operating characteristic (ROC) curve; diagnostic radiology; mean-to-sigma ratio; binormal model; proper ROC model

Corresponding author information: Stephen L. Hillis, Ph.D., Senior Biostatistician, Center for Research in the Implementation of Innovative Strategies in Practice (CRIISP), Department of Veterans Affairs (VA) Iowa City Medical Center (152), 601 Highway 6 West, Iowa City, IA 52246-2208, Ph: 319-338-0581 x7680, steve-hillis@uiowa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Introduction

For diagnostic studies that evaluate and compare medical imaging modalities (e.g., mammography) that require a human reader (typically a radiologist) to interpret generated images with respect to disease likelihood or severity, a commonly used method for estimating a receiver operating characteristic (ROC) curve is to use maximum likelihood estimation based on the assumption of a latent binormal model [1–4]; we refer to this method as the *binormal method*. The latent binormal model assumption states that there exists a monotone transformation that, when applied to the decision variable of interest, results in a latent decision variable that is normally distributed for nondiseased cases as well as for diseased cases, with the means and variances allowed to differ for the two distributions. 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 it is typical to assume that these ratings represent the binning of values of a latent (i.e., unobserved) continuous decision variable representing the reader’s likelihood-of-disease perception. Often the ROC-curve summary measure of interest is the area under the curve (AUC).

For large samples the binormal method has been shown to perform well for decision variable distributions that can vary greatly from the binormal distribution [5–8]. We refer to the ROC curve corresponding to a latent binormal model decision variable as the *binormal ROC curve*. Throughout we assume that the decision variable of interest is continuous and that larger values of it are more indicative of disease.

In most practical situations a meaningful decision variable should be an increasing function of the likelihood ratio (likelihood of being diseased divided by likelihood of not being diseased)[9, p 94]. A decision variable model having this property and its corresponding ROC curve are said to be *proper*[10, p 37]. A function whose first derivative is decreasing throughout an open interval is called *concave* or *concave downward* in that interval, and a function whose first derivative is increasing throughout an open interval is called *convex* or *concave upward* in that interval [11, pp 144–145]. Since the slope of an ROC curve for a continuous decision variable is equal to the likelihood ratio at the corresponding threshold, it follows that the slope of a proper ROC curve decreases as the false positive fraction (fpf) increases, that is, a proper ROC curve will be concave everywhere ($0 \leq \text{fpf} \leq 1$) [9, pp 70–71]. If the decision variable is not an increasing function of the likelihood function, then its model and corresponding ROC curve are said to be *improper*[10, p 37].

The latent binormal model is improper if the nondiseased and diseased distribution variances differ; furthermore, there is a single fpf value such that the ROC curve is concave on one side and convex on the other side[9, p 83]. In addition, as we show later, there is a single fpf value where the ROC curve crosses the chance line, implying that for a range of fpf values the decision variable performs worse than guessing.

Although binormal model ROC curves are improper unless the diseased and nondiseased variances are equal, in practice the “improperness” is so small that it is not apparent when looking at the ROC curve. However, there are situations when the improperness is apparent, with the ROC curve visibly crossing below the chance line and having an obvious “hook”. For these situations we deem the ROC curve and its corresponding binormal model to be *noticeably or slightly improper*, depending on how easily the improperness can be seen. Pan and Metz [12, p 381] note that “because ROC curves do not show shapes of this kind when they are estimated from reliable data sets, hooks and degeneracy can be considered artifacts of the conventional binormal ROC model.”

Presently, researchers often ignore or do not check for improperness in fitted binormal ROC curves, even though there can be situations where the magnitude of the improperness is large enough to make the validity of conclusions based on the improper ROC curve questionable. Furthermore, standard statistical software packages do not provide any diagnostics for assessing the magnitude of the improperness; thus the researcher can only know the extent of the improperness from visually examining ROC-curve plots, which often is not done when the researcher is primarily interested in an ROC-curve summary index, such as the AUC.

There is not general agreement on an appropriate analysis strategy for ROC data that will satisfactorily account for the inherent improperness of binormal ROC curves. At one end of the spectrum is the strategy of using the binormal method and ignoring any improperness in resulting ROC curves, and at the other end is the strategy of always using a proper method that never results in improper ROC curves. In between are other strategies, such as using a proper method only when the binormal method produces a clearly visible improper ROC curve. Although improperness can be visually assessed from graphs, a discussion of the different analysis strategies requires a quantitative improperness measure that is easy to compute and interpret. Our purpose is to investigate the properties of the mean-to-sigma ratio as a quantitative measure of improperness. However, we do not attempt to discuss which analysis strategy should be used, since that would require separate treatment.

In summary, our main purpose is to show how the mean-to-sigma ratio can be a useful, easy-to-understand, and easy-to-use measure for assessing the magnitude of the improperness of binormal ROC curves. The outline of the paper is as follows. We illustrate the inherent improperness of the binormal model with an example, show how the mean-to-sigma ratio can be used as a measure for assessing the degree of improperness, and discuss alternative proper models. Using data from a multi-reader multi-modality study, we illustrate the usefulness of the mean-to-sigma ratio for assessing improperness and for comparing the binormal method with an alternative method based on a proper model.

Materials and Methods

Example of a noticeably improper binormal ROC curve

To illustrate the inherent improperness in binormal ROC curve estimation, consider Table 1 which shows the rating data for one reader from a study [13] that will be described in more detail in the Results section. Figure 1 shows the corresponding fitted binormal ROC curve; note that there is a visible hook and chance-line crossing near the upper right-hand corner of the unit square. In Figure 1 the ROC curve crosses the chance line at the point $(\text{fpf}, \text{tpf}) = (0.976, 0.976)$, where tpf stands for true positive fraction, shown by the intersection of the “crossing” reference line with the ROC curve. Furthermore, this ROC curve is concave for $\text{fpf} < 0.735$, but is convex for $\text{fpf} > 0.735$. Letting $\text{ROC}(t)$ denote the tpf corresponding to $\text{fpf} = t$, the point $(.735, \text{ROC}(.735))$ on the ROC curve separates the concave and convex portions of the curve and thus is an *inflection point*; this point is shown by the intersection of the “inflection” reference line and the ROC curve in Figure 1.

In general we will denote the $\text{fpf} = \text{tpf}$ value where the ROC curve crosses the chance line by t_0 and refer to t_0 as the *chance-line crossing fpf*; similarly, we will denote the fpf value where the ROC curve changes from convex to concave (for either increasing or decreasing fpf) by t_1 and refer to t_1 as the *inflection-point fpf*. (More precisely, t_1 is the fpf -coordinate of the inflection point.) For example, in Figure 1 the chance-line crossing fpf is $t_0 = .976$ and the inflection-point fpf is $t_1 = .735$.

Figure 2 shows the latent nondiseased and diseased distribution densities that yield the ROC curve in Figure 1, with y denoting the latent binormal-model decision variable with the standard parameter constraints imposed: $\mu_1 = 0$ and $\sigma_1 = 1$. The log-likelihood ratio (diseased divided by nondiseased density) is shown in the upper part of the figure. The crossing and inflection reference lines in Figure 2 mark the decision variable thresholds, $c_0 = -1.98$ and $c_1 = -0.628$, that correspond to $t_0 = 0.976$ and $t_1 = 0.735$, respectively, in Figure 1. We refer to c_0 and c_1 as the *chance-line crossing threshold* and the *inflection-point threshold*, respectively. In Figure 2, $\Pr(Y > c_0)$ is the same for both the nondiseased and diseased distributions; correspondingly, the ROC curve crosses the chance line at (t_0, t_0) in Figure 1, with $t_0 = \Pr(Y > c_0 | \text{nondiseased}) = 1 - \Phi(c_0) = \Phi(-c_0)$, where $\Phi(\cdot)$ is the cumulative standard normal distribution function. In Figure 2 the log-likelihood ratio slope is zero at $y = c_1$, with the log-likelihood ratio decreasing with increasing y for $y \leq c_1$ and increasing with increasing y for $y \geq c_1$; correspondingly, the ROC curve is concave in $(0, t_1)$ and convex in $(t_1, 1)$, where $t_1 = \Phi(-c_1)$.

Although it is clear from Figure 1 that the ROC curve is not concave everywhere, it is not possible to visually identify the point where the ROC curve ceases to be concave without the reference line. In general, one cannot discern the point where the ROC curve changes from concave to convex, even for visibly improper curves. However, it is possible without the reference line to see where the ROC curve crosses the chance line for visibly improper curves. For this reason we believe that specifying the chance-line crossing fpf conveys an easier-to-interpret assessment of the improperness of the binormal ROC curve than specifying the inflection-point fpf.

Mean-to-sigma ratio as a measure of improperness

The main point of this paper is that the degree of improperness of a binormal ROC curve can easily be assessed by the mean-to-sigma ratio, which we denote by r . The mean-to-sigma ratio is defined by

$$r = \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1} \quad (1)$$

where μ_2 and σ_2 are the mean and standard deviation of the diseased latent decision variable distribution, and μ_1 and σ_1 are the corresponding parameters for the nondiseased latent decision variable distribution. Defining $\Delta\mu = \mu_2 - \mu_1$ and $\Delta\sigma = \sigma_2 - \sigma_1$, we can write

$$r = \frac{\Delta\mu}{\Delta\sigma}$$

Letting a and b denote the usual parameters of the binormal model, i.e., $a = (\mu_2 - \mu_1)/\sigma_2$ and $b = \sigma_1/\sigma_2$, it follows from Equation (1) that

$$r = \frac{a}{1 - b} \quad (2)$$

For $\sigma_1 - \sigma_2 = 0$ we define $r = \infty$ if $\mu_2 - \mu_1 > 0$ and $r = -\infty$ if $\mu_2 - \mu_1 < 0$. If $\sigma_2 - \sigma_1 = 0$ and $\mu_2 - \mu_1 = 0$, then r is undefined.

The mean-to-sigma ratio was first introduced by Swets et al [14], who noticed that it seemed to be approximately constant for a variety of experiments. Some support for this conclusion was provided by later analyses [8,15,16]. For example, Green and Swets [16, p 95] note that $r \approx 4$ describes the relationship between the latent means and standard deviations for many studies. These references only discuss using r to describe the relationship between the latent diseased and nondiseased distribution means and standard deviations.

However, a useful characteristic of the mean-to-sigma ratio that has received virtually no attention is its relationship to the binormal ROC-curve chance-line crossing fpf. Specifically, we show in the Appendix (available online at www.academicradiology.org) that for a binormal ROC curve with $\sigma_1 \neq \sigma_2$ there exists a unique chance-line crossing fpf given by

$$t_0 = \Phi(r) \quad (3)$$

with corresponding chance-line crossing threshold

$$c_0 = \frac{\sigma_2 \mu_1 - \sigma_1 \mu_2}{\sigma_2 - \sigma_1} \quad (4)$$

It follows from Equation (4) that for $\mu_1 = 0$ and $\sigma_1 = 1$ the corresponding threshold can be expressed in terms of the mean-to-sigma ratio:

$$c_0 = -r \quad (5)$$

The only previous reference to relationship (3) of which we are aware is a brief mention by Dorfman et al [17, p 141], but they do not prove the result or suggest using r as a measure for assessing improperness. We also show in the Appendix that for increasing fpf the ROC curve crosses the chance-line from above if $b < 1$ and from below if $b > 1$. These results are summarized in part (a) of Table 2.

For example, the latent binormal distribution in Figure 2 is based on the parameter estimates for the Table 1 data, given by $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 2.29337$, and $\sigma_2 = 2.15743$ ($a = 1.06301$, $b = 0.46351$). From Equations (2), (3), and (5) we have $r = a/(1-b) = 1.9814$, $t_0 = \Phi(r) = .976$, and $c_0 = -r = -1.9814$; hence it follows that the chance line crossing occurs at (.976, .976), as indicated in Figure 1, and $\Pr(Y > -1.9814) = .976$ for both nondiseased and diseased distributions, as indicated in Figure 2.

Table 3 shows the chance-line crossing fpf t_0 for various values of r , computed using Equation (3). For example, we see that for $r = 1$ the chance line cross occurs at (.841, .841) and for $r = -1$ it occurs at (.159, .159); for $r = 2$ it occurs at (.977, .977) and for $r = -2$ it occurs at (0.023, 0.023). Note from Equation (3) that $t_0 > .5$ if $r > 0$ and $t_0 < .5$ if $r < 0$.

Figure 3 shows ROC curves for combinations of $\mu_2 - \mu_1 = 1, 2, 3$ and $r = .5, 1, 1.5, 2, 2.5, 3, 3.5, 4$, with $\sigma_1 = 1$. Vertical crossing reference lines that intersect the fpf axis at $(t_0, 0)$ have been added to clearly show the chance-line crossing fpfs. The value of σ_2 (not shown) varies for each combination; from Equation (1) it is given by

$$\sigma_2 = \sigma_1 + (\mu_2 - \mu_1)/r \quad (6)$$

Figure 4 shows ROC curves for combinations of $\mu_2 - \mu_1 = .25, .75, 1.25$ with $r = -.5, -1, -1.5, -2, -2.5, -3, -3.5, -4$, with $\sigma_1 = 1$. As in Figure 3, the value of σ_2 varies for each combination according to Equation (6). Since $r < 0$ and $a = (\mu_2 - \mu_1)/\sigma_2 > 0$, it follows from Equation (2) that $b > 1$. Writing Equation (1) in the form

$$r = \frac{\mu_2 - \mu_1}{\sigma_1} \frac{\sigma_1}{\sigma_2 - \sigma_1} = \frac{\mu_2 - \mu_1}{\sigma_1} \frac{b}{1 - b}$$

shows that

$$|r| > (\mu_2 - \mu_1)/\sigma_1 \quad (7)$$

for the Figure 4 combinations, since $b > 1$ and hence $|b/(1 - b)| > 1$. Thus there is only one ROC curve in Figure 4 for $r = -.5$, which corresponds to $\mu_2 - \mu_1 = .25$, since r cannot be equal to $-.5$ if $\mu_2 - \mu_1 = .75$ or $\mu_2 - \mu_1 = 1.25$ due to constraint (7). Similarly, there are only two curves in Figure 4 for $r = -1.0$ since r cannot be equal to -1.0 if $\mu_2 - \mu_1 = 1.25$.

Figures 3 and 4 illustrate that, assuming $\mu_2 - \mu_1 > 0$, the ROC curve will cross the chance line from above in the upper right quadrant for increasing fpf if $0 < r < \infty$, while if $-\infty < r < 0$ the ROC curve will cross the chance line from below in the lower left quadrant. For example, in Figure 3 the crossing must be in the upper right quadrant since $r > 0$ implies $t_0 = \Phi(r) > .5$, and the crossing must be from above since $b < 1$ ($\sigma_2 - \sigma_1 > 0$ follows from $r > 0$, $\mu_2 - \mu_1 > 0$ and Equation (1)). Similarly, in Figure 4 the ROC curves cross from below in the lower left quadrant since $r < 0$ and $b > 1$.

What is important to note from Figures 3 and 4 is that we can see the crossing of the chance line if $|r| = 2$, we can barely see it if $|r| = 2.5$, and it is not discernable for $|r| = 3$. From these observations we classify the improperness of a binormal ROC curve as *indiscernible* if $|r| \geq 3$, *noticeable* if $|r| \leq 2$, and *slight* if $2 < |r| < 3$, as summarized in Table 4. However, we note that these boundaries are somewhat arbitrary.

For the special case of the equal variance binormal model, if $\mu_2 - \mu_1 > 0$ then $r = \infty$ and if $\mu_2 - \mu_1 < 0$ then $r = -\infty$. For either situation the ROC curve intersects the chance line only at the right and left sides of the unit square; hence the ROC curve does not “cross” the chance line.

Inflection-point fpf and threshold

We show in the Appendix the following results for the inflection-point fpf and threshold. The unique inflection-point fpf is given by

$$t_1 = \Phi\left(\frac{\sigma_1(\mu_2 - \mu_1)}{\sigma_2^2 - \sigma_1^2}\right) = \Phi\left(\frac{r}{1 + b^{-1}}\right) \quad (8)$$

and the corresponding inflection-point threshold is given by

$$c_1 = - \frac{\sigma_1^2 \mu_2 - \sigma_2^2 \mu_1}{\sigma_2^2 - \sigma_1^2}$$

If $\mu_1 = 0$ and $\sigma_1 = 1$ then c_1 can be written as

$$c_1 = \frac{-\mu_2}{\sigma_2^2 - 1} = \frac{-r}{1+b^{-1}} \quad (9)$$

The derivative of the likelihood ratio and log-likelihood ratio function is zero only at c_1 ; otherwise it is nonzero, having the same sign for all thresholds less than c_1 and the opposite sign for all thresholds greater than c_1 . If $b < 1$ then the binormal ROC curve is concave over $(0, t_1)$ and convex over $(t_1, 1)$; if $b > 1$ then the binormal ROC curve is convex over $(0, t_1)$ and concave over $(t_1, 1)$. These results are listed in part (b) of Table 2. Formulas (8) and (9) were used to compute the inflection-point threshold and fpf in Figures 1 and 2.

As mentioned earlier, we believe that the mean-to-sigma ratio is a more meaningful measure of improperness than the infection-point fpf since, unlike the inflection-point fpf, the chance-line crossing fpf can be visually discerned for noticeably or slightly improper ROC curves.

Proper methods

We refer to an estimation method that always produces a proper ROC-curve estimate as a *proper method*. One approach that results in a proper method is maximum likelihood estimation based on a proper model. Some examples of proper models are the equal variance binormal model, the gamma model [5,10, pp 195–201, 17], the contaminated binormal model [18–20], and the model proposed by Pan and Metz [12]. Although proper methods have the advantage of always producing a proper estimated ROC curve, the binormal method remains the standard method for estimating ROC curves, with proper methods used only occasionally

Of these proper models, the model proposed by Pan and Metz [12] is most closely related to the binormal model; for this reason we use it in Example 2 in the Results section to illustrate the usefulness of the mean-to-sigma ratio for comparing the performance of the binormal method with a proper method. Briefly, for every binormal model a corresponding proper model is defined by using the likelihood ratio of the binormal model as the new decision variable; the resulting model is always proper. Although References [12,21] refer to this as the “‘proper’ binormal model”, we prefer to call it the *binormal likelihood-ratio model* (binormal-LR model) because we believe this name better describes it. Although the binormal-LR model is derived from the binormal model, we note that it is not a binormal model because its decision variable is not a monotone function of the decision variable of a binormal model (except for the case when the binormal-LR model is the same as an equal-variance binormal model). The binormal-LR model can be estimated using the PROC procedure [21, p 13].

Improperness classification rule proposed by Pan and Metz

Pan and Metz [12, pp 1–2] define a hook as “the proportion of an ROC curve that lies below the +45° diagonal ‘guessing line’ of the ROC plot.” Metz and Pan [21] state that empirically the “hook” in a binormal ROC curve is evident if

$$\frac{d_a}{|c|} \leq 6 \quad (10)$$

with

$$d_a = \frac{\sqrt{2}a}{\sqrt{1+b^2}} \text{ and } c = \frac{b-1}{b+1} \quad (11)$$

where a and b are the binormal distribution parameters, and d_a and c represent a parameterization of the corresponding binormal-LR model. (Actually Metz and Pan use “ $\lesssim$ ” instead of “ $\leq$ ”; we have substituted the latter symbol to facilitate comparison with r .) We assume $b \neq 1$ to avoid a zero denominator in $d_a/|c|$. We show in the Appendix that

$$\frac{d_a}{|c|} = |r|M \quad (12)$$

where

$$M = \sqrt{\frac{2(b+1)^2}{1+b^2}}$$

From equation (12) it follows that condition (10) is equivalent to

$$|r| \leq 6/M$$

It can be shown that

$$M \leq 2 \quad (13)$$

with $M = 2$ if and only if $b = 1$. It follows from Equations (12) and (13) that

$$\text{if } |r| < 3 \text{ then } \frac{d_a}{|c|} \leq 6$$

but the reverse implication does not hold; that is, our rule is more conservative in classifying a binormal curve as having an evident hook. For typical values of b , M is close to 2; in particular, if $.5 \leq b \leq 2$ then $1.90 \leq M \leq 2$. For example, for $b = 0.5$ or 2.0 , $M = 1.90$ and hence $d_a/|c| \leq 6$ if and only if $|r| \leq 3.16$; more generally, for $.5 \leq b \leq 2$ the boundary of 6 in condition (10) corresponds to a boundary between 3 and 3.16 for $|r|$ for a given b . Thus we see that for typical values of b the boundary $d_a/|c| = 6$ is similar to our proposed boundary of $|r| = 3$ separating slight and indiscernible improperness. On the other hand, if $b = 0.1$ or 10 ,

then $M = 1.548$ and hence $d_a/c \leq 6$ if and only if $|r| \leq 3.88$. For instance, if $a = 3.0$ and $b = 1$, then $d_a/c = 5.16$ (from Eq (11)) and $r = 3.33$; here the hook should be evident according to rule (10) but our rule says it is not evident, which we believe is a reasonable assertion since the chance line crossing occurs at the fpf value $\Phi(3.33) = .99957$. This last example illustrates an important advantage of the mean-to-sigma ratio – it has a clear interpretation of improperness in terms of the chance line crossing.

Metz and Pan [21, p 13] note that the binormal and corresponding binormal-LR ROC curves “are indistinguishable if and only if no ‘hook’ is evident in the conventional binormal ROC.” Using our rule this statement suggests that a binormal ROC curve having indiscernible improperness ($|r| \geq 3$) is indistinguishable from the corresponding binormal-LR ROC curve.

Simulation study implications

For a simulation study that simulates ROC data using the binormal model, we recommend that the mean-to-sigma ratio be computed for each binormal model and that only models that are indiscernibly improper ($|r| \geq 3$) be included in the study.

Results

Example 1: Assessing the degree of improperness using the mean-to-sigma ratio

The data for this example were provided by Carolyn Van Dyke, MD, who had obtained them in a study [13] that compared the relative performance of single spin-echo magnetic resonance imaging (MRI) and cine MRI in detecting thoracic aortic dissection. There were 45 patients with an aortic dissection and 69 patients without a dissection imaged with both spin-echo and cine MRI. Five radiologists independently interpreted all of the images using a five-point ordinal scale. We will refer to the cine modality and reader 1 combination as CINE1, the spin-echo modality and reader 1 combination as SPIN1, etc.

For each of the ten modality-reader combinations, Table 5 presents the parameter estimates for the latent binormal model and the corresponding mean-to-sigma ratio, chance-line crossing fpf, and AUC; the binormal-LR AUC computed using the PROPROC method is also included. From Table 4 we see that CINE5 is the only combination with a noticeably improper ROC curve, with $r = 1.98$. This is the curve previously discussed and displayed in Figure 1. Two of the curves have slight improperness: CINE4 ($r = 2.41$) and SPIN2 ($r = 2.99$). The other seven curves have indiscernible improperness, with $3.00 < r < 4.76$ for six of them and $r = 33.19$ for the other.

The ten binormal ROC curves are displayed in Figure 5. The improperness of the CINE5 ROC curve ($r = 1.98$) is easy to discern while that of the CINE4 curve ($r = 2.41$) can be seen upon close examination. On the other hand, the improperness of the other curves ($r \geq 2.99$) is not visible.

Example 2: Comparing the binormal and binormal-LR methods using the mean-to-sigma ratio

In this example we use the same data as in Example 1, but now use the mean-to-sigma ratio as a tool for comparing the binormal and binormal-LR methods. Figure 6 shows the binormal and binormal-LR ROC curves for five of the modality-reader combinations.

Figures 7(a) and 7(b) compare the methods for the two combinations having the lowest mean-to-sigma ratios: CINE5 ($r = 1.98$) and CINE4 ($r = 2.41$). Here we see that the binormal and binormal-LR ROC curves are similar except in the region defined by fpf values larger

than the highest fpf of any of the operating points (i.e., the empirical (fpf, tpf) points). In this region we see that each binormal-LR ROC curve, unlike the binormal ROC curve, does not have a hook or chance-line crossing. Thus the binormal-LR AUC is higher than the binormal AUC for both combinations: 0.841 versus 0.833 for CINE5, and 0.977 versus 0.970 for CINE4.

Figures 7(c) and 7(d) compare the methods for two indiscernibly improper combinations, SPIN5 ($r = 3.00$) and SPIN3 ($r = 3.17$), that have similar ROC curves with approximately equal AUCs: 0.943 (binormal-LR) versus 0.945 for SPIN5, and 0.930 (binormal-LR) versus 0.928 for SPIN3. Note that for both combinations the binormal and binormal-LR curves are virtually indistinguishable. These results are not surprising, considering that the curves are indiscernibly improper and the binormal-LR family of curves contains curves very similar to any of the indiscernibly improper binormal curves, as previously discussed.

Figure 6(e) compares the methods for an indiscernibly improper combination, CINE3 ($r = 4.76$), where the binormal-LR and binormal curves are notably different and have moderately different AUCs: 0.908 (binormal-LR) versus 0.929. We see from the plot that the binormal-LR curve visibly passes closer to the first 3 operating points than the binormal curve, while the binormal curve passes closer to the last operating point. The important point to note from Figure 6(e) is that the binormal-LR method, as implemented by the PROPROC procedure, can produce a visibly different curve than the binormal method, even when the binormal method produces a curve that is indiscernibly improper. We note that this finding contradicts the assertion by Metz and Pan [21, p 29] that the binormal LR model provides fitted curves that are virtually identical to those of the binormal model when no hooks are evident with the binormal model fitted curves.

In this example the mean-to-sigma ratio has enabled us to compare the binormal and binormal-LR procedures for different levels of binormal improperness. We found that the two methods not only can produce visibly different curves when the binormal curves are noticeably or slightly improper, but also when the binormal curves are indiscernibly improper.

Example 3: Assessing improperness uncertainty

In this example we give a 90% confidence interval for the degree of improperness by constructing a confidence interval for the absolute mean-to-sigma ratio. Mean-to-sigma ratios were computed for each of 1000 bootstrap samples from the data set used for the previous examples, with each bootstrap sample having the same number of diseased and nondiseased cases as the original data. The minimum and 5th and 95th percentiles of the bootstrap absolute-value mean-to-sigma ratios are reported in Table 5 (the 5th and 95th percentiles define the 90% confidence interval). The median of the ten minimum values is 1.11, with eight of them less than 1.33. The median of the ten 5th percentiles is 1.75, with eight of them less than 1.95. The 95th percentiles all exceed 4.35. Thus both noticeably improper and indiscernibly improper curves are included in the confidence intervals for all of the test-reader combinations. This example shows how the mean-to-sigma ratio can be useful for describing uncertainty in a fitted ROC curve with respect to improperness.

The ROC curves, corresponding summary indices and data for creating ROC-curve plots for the first two examples, as well as the mean-to-sigma ratios for the third example, were obtained using a SAS (SAS for Windows, Version 9.2, Copyright © 2002–2008 by SAS Institute Inc., Cary, NC, USA) macro that is available to the public [22]. Freely available stand-alone software that performs the same functions is also available for download in the multi-reader program *DBM MPMC* 2.2[23,24] and in the single reader programs *PROPROC* and *ROCKIT*, written by Charles Metz and colleagues and available at

<http://xray.bsd.uchicago.edu/krl/index.htm>. Graphical displays were created using STATA/IC 10.1 for Windows (Copyright © 1985–2009, College Station, TX, USA).

Discussion

Although the standard method for estimating an ROC curve is maximum likelihood estimation based on the latent binormal model, the binormal model implies that the decision variable is not a monotone function of the likelihood ratio; hence this method produces improper ROC curves that are not concave everywhere and cross the chance line, implying that the test performs worse than chance for a range of fpf values. Although in most situations the degree of improperness is so small that it cannot be seen, it is important to be able to easily identify those ROC curves where the improperness is visible. We have shown how the mean-to-sigma ratio, by having a clear interpretation in term of the chance-line crossing fpf , provides an easy way to assess the degree of improperness in binormal ROC curves.

We recommend that ROC analysis software print out the value of the mean-to-sigma ratio for each computed binormal ROC curve, along with a warning message when the improperness is slight or noticeable. While we have suggested classification categories for assessing improperness based on the mean-to-ratio, we emphasize that these categories are somewhat arbitrary and encourage researchers to formulate their own guidelines.

Although improperness can be visually assessed from plots, the mean-to-sigma ratio provides an objective measurement of improperness that is not dependent on individual subjectivity. Furthermore, it is useful when plots are not easily obtained from software, for reporting results in journal articles when space requirements restrict the use of plots, and when the number of ROC curves renders examination of plots impractical (e.g., a multi-reader study with many readers and modalities or a simulation study.) On the other hand, the mean-to-sigma ratio measures only one aspect of improperness; e.g., it does not assess the area of the hook. Thus we recommend also looking at ROC curves, when feasible, to visually assess improperness.

In addition to allowing easy assessment of the degree of improperness for binormal ROC curves, we saw in an example how the mean-to-sigma can help clarify differences between the binormal and other methods by allowing for comparison of the methods according to different levels of binormal improperness. In particular, we saw that the binormal-LR method, as implemented by the PROPROC procedure, can produce visibly different ROC curves regardless of the level of improperness of the binormal curve. Furthermore, we saw how it could be useful in describing uncertainty in a fitted binormal ROC curve with respect to improperness.

Since we interpret the mean-to-sigma ratio in terms of the chance-line crossing, we could have alternatively used the chance-line crossing fpf as a measure of improperness instead of the mean-to-sigma ratio. However, there were several reasons why we chose the mean-to-sigma ratio: (1) Our interest was primarily in assessing improperness for the binormal model, for which the mean-go-sigma ratio is easy to compute from the model parameters. (2) In general, an improper ROC curve does not necessarily cross the chance line; for such curves the chance-line crossing fpf would be misleading. (3) The mean-to-sigma ratio is an established measure for assessing the relationship between the means and standard deviations; viewing it as a measure of improperness gives it a dual function.

We have not attempted to discuss the advantages and disadvantages of various strategies to circumvent the binormal improperness problem since that would require separate treatment. What we have shown, however, is how the mean-to-sigma ratio, by providing a quantitative

and easily computable improperness measure, facilitates such a discussion and allows for the comparison of the analysis methods according to improperness categories in simulation and real-data studies.

Acknowledgments

Grant support: This research was supported by the National Institutes of Health, grant R01EB000863.

Acknowledgements and Disclaimer

The authors thank Carolyn Van Dyke, M.D. for sharing her data set. The authors also thank the reviewers for their very helpful comments and suggestions. This research was supported by the National Institutes of Health, grant R01EB000863. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

1. Dorfman DD, Alf E Jr. Maximum likelihood estimation of parameters of signal-detection theory and determination of confidence intervals: rating method data. *Journal of Mathematical Psychology* 1969;6:487–496.
2. Dorfman, DD. RSCORE II. In: Swets, JA.; Pickett, RM., editors. *Evaluation of diagnostic systems: methods from signal detection theory*. Academic Press; San Diego, CA: 1982. p. 212–232.
3. Dorfman DD, Berbaum KS. Degeneracy and discrete receiver operating characteristic rating data. *Academic Radiology* 1995;2:907–915. [PubMed: 9419659]
4. Metz CE, Herman BA, Shen JH. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine* 1998;17:1033–1053. [PubMed: 9612889]
5. Hanley JA. The robustness of the binormal assumptions used in fitting ROC curves. *Medical Decision Making* 1988;8:197–203. [PubMed: 3398748]
6. 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]
7. HajianTilaki 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]
8. 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]
9. Pepe, M. *The statistical evaluation of medical tests for classification and prediction*. Oxford; New York: 2003.
10. Egan, JP. *Signal detection theory and ROC analysis*. Academic; New York: 1975.
11. Stein, SK. *Calculus and analytic geometry*. 4. McGraw-Hill; New York: 1987.
12. 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]
13. Van Dyke, CW.; White, RD.; Obuchowski, NA.; Geisinger, MA.; Lorig, RJ.; Meziene, MA. Cine MRI in the diagnosis of thoracic aortic dissection. 79th RSNA Meetings; Chicago, IL. November 28–December 3, 1993;
14. Swets JA, Tanner WP, Birdsall TG. Decision processes in perception. *Psychological Review* 1961;68:301–340. [PubMed: 13774292]
15. Swets JA. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psychological Bulletin* 1986;99:100–117. [PubMed: 3704032]
16. Green, DM.; Swets, JA. *Signal detection theory and psychophysics*. Peninsula Publishing; Los Altos: 1988. Original work: Green DM, Swets JA. *Signal detection theory and psychophysics*. New York: Wiley, 1966
17. 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]

18. Dorfman DD, Berbaum KS, Brandser EA. A contaminated binormal model for ROC data - Part I. Some interesting examples of binormal degeneracy. *Academic Radiology* 2000;7:420–426. [PubMed: 10845401]
19. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data-Part II. A formal model. *Academic Radiology* 2000;7:427–437. [PubMed: 10845402]
20. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data -Part III. Initial evaluation with detection ROC data. *Academic Radiology* 2000;7:438–447. [PubMed: 10845403]
21. Metz CE, Pan XC. “Proper” binormal ROC curves: theory and maximum-likelihood estimation. *Journal of Mathematical Psychology* 1999;43:1–33. [PubMed: 10069933]
22. Hillis, SL.; Schartz, KM.; Pesce, LL.; Berbaum, KS.; Metz, CE. DBM MRMC procedure for SAS (computer software). [Accessed August 1, 2009.]. Available for download from <http://perception.radiology.uiowa.edu>
23. Berbaum, KS.; Schartz, KM.; Pesce, LL.; Hillis, SL. DBM MRMC 2.2 (computer software). [Accessed August 1, 2009.]. Available for download from <http://perception.radiology.uiowa.edu>
24. Berbaum, KS.; Metz, CE.; Pesce, LL.; Schartz, KM. DBM MRMC 2.1 User’s Guide (software manual). [Accessed August 1, 2009.]. Available for download from <http://perception.radiology.uiowa.edu>
25. Stein, SK. *Calculus in the first three dimensions*. McGraw-Hill; New York: 1967.

Equation Section (Next) Appendix

For completeness we include in this section proofs of the chance-line crossing and inflection-point results for the binormal model summarized in Table 2, as well as showing the relationship between d_a/c and r . Let Y denote a binormal decision variable corresponding to a nondiseased population with mean μ_1 and standard deviation σ_1 , and a diseased population with mean μ_2 and standard deviation σ_2 , with $\sigma_1 \neq \sigma_2$. Let D denote disease status, with $D = 1$ denoting diseased and $D = 0$ denoting nondiseased. Thus $(Y|D=0) \sim N(\mu_1, \sigma_1^2)$ and $(Y|D=1) \sim N(\mu_2, \sigma_2^2)$. Let r denote the mean-to-sigma ratio: $r = (\mu_2 - \mu_1)/(\sigma_2 - \sigma_1)$. Let $\text{ROC}(\cdot)$ denote the ROC curve function for Y ; i.e., for a given fpf t , the corresponding tpf is given by $\text{ROC}(t)$. Let $\text{LR}(\cdot)$ denote the likelihood -ratio function for Y given by

$$\text{LR}(y) = \frac{\frac{1}{\sigma_2} \varphi\left(\frac{y-\mu_2}{\sigma_2}\right)}{\frac{1}{\sigma_1} \varphi\left(\frac{y-\mu_1}{\sigma_1}\right)} \quad (\text{A1})$$

where $\varphi(\cdot)$ is the standard normal density function.

Chance-Line Crossing Theorem

Theorem 1

There exists a unique chance-line crossing fpf, $t_0 = \Phi(r)$, with corresponding unique chance-line crossing threshold, $c_0 = (\sigma_2\mu_1 - \sigma_1\mu_2)/(\sigma_2 - \sigma_1)$. That is, the ROC curve corresponding to Y crosses the chance line once and only once at the point (t_0, t_0) and c_0 is the corresponding threshold such that $\Pr(Y > c_0 | D = 0) = t_0$.

Proof

Let c denote a threshold corresponding to a chance-line crossing fpf. Then

$$\Pr(Y > c | D=0) = \Pr(Y > c | D=1)$$

or equivalently,

$$\Pr(Y < c | D=0) = \Pr(Y < c | D=1)$$

It follows that

$$\Phi\left(\frac{c - \mu_1}{\sigma_1}\right) = \Phi\left(\frac{c - \mu_2}{\sigma_2}\right) \quad (\text{A2})$$

Solving Equation (A2) for c yields the unique solution

$$c_0 = \frac{\sigma_2 \mu_1 - \sigma_1 \mu_2}{\sigma_2 - \sigma_1} \quad (\text{A3})$$

Note that c_0 is defined since $\sigma_1 \neq \sigma_2$.

Let t_0 denote the chance-line crossing fpf corresponding to c_0 . Then

$$t_0 = \Pr(Y > c_0 | D=0) = 1 - \Phi\left(\frac{c_0 - \mu_1}{\sigma_1}\right) = \Phi\left(\frac{\mu_1 - c_0}{\sigma_1}\right) = \Phi\left(\frac{\mu_1 - \frac{\sigma_2 \mu_1 - \sigma_1 \mu_2}{\sigma_2 - \sigma_1}}{\sigma_1}\right) = \Phi\left(\frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}\right)$$

Thus

$$t_0 = \Phi\left(\frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}\right) = \Phi(r)$$

Uniqueness of t_0 follows from the uniqueness of c_0 .

Corollary

If $b < 1$ ($b > 1$) then the ROC curve crosses from above (below) the chance line for increasing fpf.

Proof

Since the slope of the chance line is unity, then for increasing fpf the ROC curve will cross the chance line from above if its derivative at t_0 is less than unity, and from below if its derivative is greater than unity. The derivative of the ROC curve evaluated at t_0 is equal to the likelihood ratio evaluated at c_0 [9, p 70]; thus

$$\left. \frac{\partial \text{ROC}(t)}{\partial t} \right|_{t=t_0} = \text{LR}(c_0). \quad (\text{A4})$$

Substitution into the right-hand side of Equation (A4) using Equations (A1) and (A3) yields

$$\left. \frac{\partial \text{ROC}(t)}{\partial t} \right|_{t=t_0} = \frac{\sigma_1}{\sigma_2} = b$$

and the corollary follows.

Inflection-Point Theorem

Theorem 2

If $b < 1$ the ROC curve is concave over $(0, t_1)$ and convex over $(t_1, 1)$, where $t_1 = \Phi \left[\sigma_1(\mu_2 - \mu_1) / (\sigma_2^2 - \sigma_1^2) \right]$. If $b > 1$ the ROC curve is convex over $(0, t_1)$ and concave over $(t_1, 1)$.

Proof

The ROC curve for a binormal decision variable is twice differentiable. From basic calculus results concerning concave functions it follows that the binormal ROC curve is concave (convex) over an open interval if its second derivative is negative (positive) throughout the interval [25, pp 282–283]. Our approach is to show that the second derivative of the binormal ROC curve is negative throughout $(0, t_1)$ and positive throughout $(t_1, 1)$ if $b < 1$, and positive throughout $(0, t_1)$ and negative throughout $(t_1, 1)$ if $b > 1$.

Let t denote an fpf with corresponding threshold c . The derivative of the ROC curve evaluated at t is equal to the likelihood ratio evaluated at c , i.e.,

$$\frac{\partial \text{ROC}(t)}{\partial t} = \text{LR}(c)$$

It follows, using the chain rule, that

$$\frac{\partial^2 \text{ROC}(t)}{\partial t^2} = \frac{\partial \text{LR}(c)}{\partial c} \frac{\partial c}{\partial t} \quad (\text{A5})$$

Since

$$t = \Pr(Y > c | D=0) = 1 - \Pr(Y < c | D=0) = 1 - \Phi \left(\frac{c - \mu_1}{\sigma_1} \right) \quad (\text{A6})$$

then t is a strictly decreasing function of c and $\frac{\partial t}{\partial c} = -\frac{1}{\sigma_1} \varphi\left(\frac{c - \mu_1}{\sigma_1}\right)$. It follows that Equation (A5) is equivalent to

$$\frac{\partial^2 \text{ROC}(t)}{\partial t^2} = \frac{\partial \text{LR}(c)}{\partial c} \left[-\frac{1}{\sigma_1} \varphi\left(\frac{c - \mu_1}{\sigma_1}\right) \right]^{-1} \quad (\text{A7})$$

Since $\varphi[(c - \mu_1)/\sigma_1] > 0$, it follows from Equation (A7) that the second derivative of the ROC curve and the derivative of the likelihood ratio have opposite signs when evaluated at t and c , respectively. That is,

$$\text{sign}\left(\frac{\partial^2 \text{ROC}(t)}{\partial t^2}\right) = -\text{sign}\left(\frac{\partial \text{LR}(c)}{\partial c}\right)$$

where

$$\text{sign}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$$

Since the logarithmic function is strictly increasing, the likelihood ratio and log-likelihood ratio derivatives have the same sign; hence it follows that

$$\text{sign}\left(\frac{\partial^2 \text{ROC}(t)}{\partial t^2}\right) = -\text{sign}\left(\frac{\partial \ln \text{LR}(c)}{\partial c}\right) \quad (\text{A8})$$

It is straightforward to show that

$$\ln \text{LR}(c) = \ln(\sigma_1) - \ln(\sigma_2) + \frac{1}{2} \left[\frac{-(c - \mu_2)^2}{\sigma_2^2} + \frac{(c - \mu_1)^2}{\sigma_1^2} \right]$$

and

$$\frac{\partial}{\partial c} \ln \text{LR}(c) = \frac{-\sigma_1^2(c - \mu_2) + \sigma_2^2(c - \mu_1)}{\sigma_1^2 \sigma_2^2} \quad (\text{A9})$$

It follows from Equation (A9) that the derivative of the log-likelihood ratio evaluated at $c + \Delta$, where $c_1 = (\sigma_2^2 \mu_1 - \sigma_1^2 \mu_2) / (\sigma_2^2 - \sigma_1^2)$, is given by

$$\left. \frac{\partial}{\partial c} \ln LR(c) \right|_{c=c_1+\Delta} = \frac{\Delta(\sigma_2^2 - \sigma_1^2)}{\sigma_1^2 \sigma_2^2} \quad (A10)$$

Using Equation (A6) it is easy to show that c_1 is the threshold corresponding to t_1 .

From Equation (A10) it follows that for $b < 1$ (hence $\sigma_2^2 - \sigma_1^2 > 0$) the derivative of the log-likelihood ratio evaluated at c is positive for $c > c_1$ and negative for $c < c_1$; and for $b > 1$ (hence $\sigma_2^2 - \sigma_1^2 < 0$) the derivative of the log-likelihood ratio evaluated at c is negative for $c > c_1$ and positive for $c < c_1$. Since the derivative of the log-likelihood ratio has the opposite sign of the second derivative of the ROC curve evaluated at the corresponding fpf (A8), and since thresholds less than c_1 correspond to fpfs greater than t_1 and vice versa, it follows that for $b < 1$ the second derivative of the ROC curve evaluated at t is positive for $t_1 < t < 1$ and negative for $0 < t < t_1$; and for $b > 1$ the second derivative of ROC curve evaluated at t is negative for $t_1 < t < 1$ and positive for $0 < t < t_1$.

Corollary 1

The fpf value $t_1 = \Phi \left[\sigma_1(\mu_2 - \mu_1) / (\sigma_2^2 - \sigma_1^2) \right]$ is the unique inflection-point fpf, and $c_1 = (\sigma_2^2 \mu_1 - \sigma_1^2 \mu_2) / (\sigma_2^2 - \sigma_1^2)$ is its corresponding inflection-point threshold.

Proof

It follows immediately from Result 2 that t_1 is an inflection-point fpf, and uniqueness follows from the Theorem 2 result that the ROC curve does not change from concave to convex at any other point. That c_1 is the corresponding inflection-point threshold is easy to show using Equation (A6).

Corollary 2

The derivative of the likelihood ratio and log-likelihood ratio function is zero only at the inflection-point threshold c_1 ; otherwise it is positive for $c > c_1$ and negative for $c < c_1$ if $b < 1$, and negative for $c > c_1$ and positive for $c < c_1$ if $b > 1$.

Proof

From the previous corollary the inflection-point threshold is given by

$c_1 = (\sigma_2^2 \mu_1 - \sigma_1^2 \mu_2) / (\sigma_2^2 - \sigma_1^2)$. Using Equation (A9) we have

$$\left. \frac{\partial}{\partial c} \ln LR(c) \right|_{c=c_1} = \frac{-\sigma_1^2 \left(\frac{\sigma_2^2 \mu_1 - \sigma_1^2 \mu_2}{\sigma_2^2 - \sigma_1^2} - \mu_2 \right) + \sigma_2^2 \left(\frac{\sigma_2^2 \mu_1 - \sigma_1^2 \mu_2}{\sigma_2^2 - \sigma_1^2} - \mu_1 \right)}{\sigma_1^2 \sigma_2^2} = 0$$

Since the log-likelihood ratio function is a strictly increasing function of the likelihood ratio function, then it follows that

$$\left. \frac{\partial}{\partial c} LR(c) \right|_{c=c_1} = 0$$

The last part of the corollary was shown in the proof for Theorem 2.

Relationship between $d_a/|c|$ and r

Theorem 3

Let

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

where a and b are binormal distribution parameters, d_a and c represent a parameterization of the corresponding binormal-LR model, $a \geq 0$ and $b \neq 1$. Then

$$\frac{d_a}{|c|} = |r|M$$

where

$$M = \sqrt{\frac{2(b+1)^2}{1+b^2}}$$

Proof

Suppose $b > 1$. Then $|c| = -(b-1)/(b+1)$ and it follows that

$$\frac{d_a}{|c|} = \frac{a}{b-1} \frac{\sqrt{2}(b+1)}{\sqrt{1+b^2}} = -r \sqrt{\frac{2(b+1)^2}{1+b^2}} = |r|M$$

since r is negative if $b > 1$ by Equation (2). Similarly, if $b < 1$ then $|c| = -(b-1)/(b+1)$ and it follows that

$$\frac{d_a}{|c|} = \frac{a}{1-b} \frac{\sqrt{2}(b+1)}{\sqrt{1+b^2}} = r \sqrt{\frac{2(b+1)^2}{1+b^2}} = |r|M$$

since r is positive if $b < 1$.

Figure 1. Binormal ROC curve for the data in Table 1. The curve is concave to the left of the inflection reference line ($\text{fpf} < t_1$) and convex to the right ($\text{fpf} > t_1$), and drops below the chance line to the right of the crossing reference line ($\text{fpf} > t_0$). Thus t_0 is the chance-line crossing fpf and t_1 is the inflection-point fpf .

Figure 2. The latent binormal distribution that yields the ROC curve in Figure 1 and its corresponding log-likelihood ratio. The binormal distribution parameters are $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_2 = 2.29337$, $\sigma_2 = 2.15743$ ($a = 1.06301$, $b = 0.46351$). The inflection and crossing reference lines correspond to those in Figure 1, with the thresholds c_0 and c_1 corresponding to the fpf values t_0 and t_1 , respectively, in Figure 1. Thus c_0 is the chance-line crossing threshold and c_1 is the inflection-point threshold.

Figure 3. ROC curves for positive values of the mean-to-sigma ratio r and $\mu_2 - \mu_1 = 1, 2, 3$, with $\sigma_1 = 1$. The intersection of the ROC curve with the chance line occurs at (t_0, t_0) , where $t_0 = \Phi(r)$. The vertical crossing reference lines intersect the fpf-axis at $\text{fpf} = t_0$. The value of σ_2 (not shown) varies for each combination and is given by $\sigma_2 = \sigma_1 + (\mu_2 - \mu_1)/r$.

Figure 4. ROC curves for negative values of the mean-to-sigma ratio r and $\mu_2 - \mu_1 = 1, 2, 3$, with $\sigma_1 = 1$. The intersection of the ROC curve with the chance line occurs at (t_0, t_0) , where $t_0 = \Phi(r)$. The vertical crossing reference lines intersect the fpf axis at $\text{fpf} = t_0$. The value of σ_2 (not shown) varies for each combination and is given by $\sigma_2 = 1 + (\mu_2 - \mu_1)/r$.

Figure 5. Binormal ROC curves for the Van Dyke et al [13] data by reader.

Figure 6. Binormal and binormal likelihood-ratio (binormal-LR) ROC curves for the Van Dyke et al [13] data. Figures (a) and (b) compare the curves for the two modality-reader combinations having the lowest mean-to-sigma ratios, figures (c) and (d) show two indiscernibly improper combinations where the two methods give similar results, and figure (e) is an indiscernibly improper combination where the two methods give visibly different results. Notes: “AUC(binormal)” and “AUC(binormal-LR)” indicate binormal and binormal-LR AUCs, respectively; “CINE5” indicates the cine modality and reader 5 combination, etc.

Table 1

Rating data for a radiologist from Van Dyke et al [13].

	Rating				
	1	2	3	4	5
Normals:	39	19	9	1	1
Diseased:	7	7	3	5	23
	46	26	12	6	24
					114

Table 2

Summary of chance-line crossing and inflection-point results for the binormal model with $\sigma_1 \neq \sigma_2$. The results are proved in the Appendix.

a) Chance-line crossing results

fpf:

$$t_0 = \Phi(r)$$

threshold:

$$c_0 = \begin{cases} \frac{\sigma_2 \mu_1 - \sigma_1 \mu_2}{\sigma_2 - \sigma_1} & \text{general formula} \\ -r & \text{if } \mu_1 = 0, \sigma_1 = 1 \end{cases}$$

direction:

$$\begin{cases} \text{from above for increasing fpf} & \text{if } b < 1 \\ \text{from below for increasing fpf} & \text{if } b > 1 \end{cases}$$

b) Inflection-point results

fpf:

$$t_1 = \Phi\left(\frac{\sigma_1(\mu_2 - \mu_1)}{\sigma_2^2 - \sigma_1^2}\right) = \Phi\left(\frac{r}{1 + b^{-1}}\right)$$

threshold:

$$c_1 = \begin{cases} -\frac{\sigma_1^2 \mu_2 - \sigma_2^2 \mu_1}{\sigma_2^2 - \sigma_1^2} & \text{general formula} \\ \frac{-\mu_2}{\sigma_2^2 - 1} = \frac{-r}{1 + b^{-1}} & \text{if } \mu_1 = 0, \sigma_1 = 1 \end{cases}$$

Concave and convex segments of ROC curve:

$$\begin{cases} \text{concave over } (0, t_1) \text{ and convex over } (t_1, 1) & \text{if } b < 1 \\ \text{convex over } (0, t_1) \text{ and concave over } (t_1, 1) & \text{if } b > 1 \end{cases}$$

Slope of the likelihood and log-likelihood ratio functions:

$$\begin{cases} \text{zero} & \text{if } c = c_1 \\ \text{positive} & \text{if } c > c_1, b < 1 \text{ or } c < c_1, b > 1 \\ \text{negative} & \text{if } c > c_1, b > 1 \text{ or } c < c_1, b < 1 \end{cases}$$

Note: $r = (\mu_2 - \mu_1)/(\sigma_2 - \sigma_1)$ is the mean-to-sigma ratio.

Table 3

Mean-to-sigma ratio r and the corresponding chance-line crossing $\text{fpf } t_0$ for the binormal model.

r	t_0
-4.0	0.00003
-3.5	0.00023
-3.0	0.00135
-2.5	0.00621
-2.0	0.02275
-1.5	0.06681
-1.0	0.15866
-0.5	0.30854
0.0	0.50000
0.5	0.69146
1.0	0.84134
1.5	0.93319
2.0	0.97725
2.5	0.99379
3.0	0.99865
3.5	0.99977
4.0	0.99997

Table 4

Improperness classification of binormal ROC curves based on the mean-to-sigma ratio r .

Criteria	Improperness Classification
$ r \leq 2$	Noticeable
$2 < r < 3$	Slight
$ r \geq 3$	Indiscernible

VanDyke et al [13] data parameter estimates. All parameter estimates are based on the latent binormal model except for the PROPROC AUC, i.e., the AUC for the binormal likelihood-ratio model estimated using the PROPROC procedure.

Table 5

MODALITY	READER	μ_2	σ_2	r	t_0	AUC		$ r $ bootstrap results		
						Binormal	PROPROC	Min	5 th pct	95 th pct
Cine	1	3.17	1.86	3.67	0.9999	0.933	0.934	0.82	1.30	28.43
	2	2.50	1.78	3.19	0.9993	0.890	0.891	0.53	0.88	29.41
	3	2.74	1.58	4.76	1.0000	0.929	0.908	2.29	2.95	15.03
	4	9.56	4.96	2.41	0.9921	0.970	0.977	1.00	1.43	16.45
	5	2.29	2.16	1.98	0.9762	0.833	0.841	0.47	1.02	5.60
Spin echo	1	3.68	1.99	3.72	0.9999	0.951	0.952	1.33	1.95	42.69
	2	3.70	2.24	2.99	0.9986	0.935	0.926	1.05	2.38	4.35
	3	3.32	2.05	3.17	0.9992	0.928	0.930	1.17	1.68	17.57
	4	6.93	1.21	33.19	1.0000	1.000	1.000	3.89	10.07	23.89
	5	4.11	2.37	3.00	0.9986	0.945	0.943	1.32	1.82	8.04

Notes: μ_2 and σ_2 are the mean and variance for the latent diseased distribution, $\mu_1 = 0$ and $\sigma_1 = 1$ for the nondiseased distribution, $r = (\mu_2 - \mu_1)/(\sigma_2 - \sigma_1)$ is the mean-to-sigma ratio, and $t_0 = \Phi(r)$ is the chance-line crossing fpr; Min, 5th pct, and 95th pct are the minimum, 5th and 9th percentile values for $|r|$ based on 1000 bootstrap samples.