Published in final edited form as:

J Math Psychol. 2012 August 1; 56(4): 256–273. doi:10.1016/j.jmp.2012.05.003.

The three-class ideal observer for univariate normal data: Decision variable and ROC surface properties

Darrin C. Edwards and Charles E. Metz

The Department of Radiology, the University of Chicago, Chicago, IL, USA

Darrin C. Edwards: d-edwards@uchicago.edu

Abstract

Although a fully general extension of ROC analysis to classification tasks with more than two classes has yet to be developed, the potential benefits to be gained from a practical performance evaluation methodology for classification tasks with three classes have motivated a number of research groups to propose methods based on constrained or simplified observer or data models. Here we consider an ideal observer in a task with underlying data drawn from three univariate normal distributions. We investigate the behavior of the resulting ideal observer's decision variables and ROC surface. In particular, we show that the pair of ideal observer decision variables is constrained to a parametric curve in two-dimensional likelihood ratio space, and that the decision boundary line segments used by the ideal observer can intersect this curve in at most six places. From this, we further show that the resulting ROC surface has at most four degrees of freedom at any point, and not the five that would be required, in general, for a surface in a six-dimensional space to be non-degenerate. In light of the difficulties we have previously pointed out in generalizing the well-known area under the ROC curve performance metric to tasks with three or more classes, the problem of developing a suitable and fully general performance metric for classification tasks with three or more classes remains unsolved.

Keywords

ROC analysis; three-class classification; ideal observer decision rules

1. Introduction

We are working to extend the well-known observer performance evaluation methodology of receiver operating characteristic (ROC) analysis [1, 2] to classification tasks with three or more classes. This could conceivably be of benefit, for example, in a medical decision-making task in which a region of a patient image must be characterized as containing a malignant lesion, a benign lesion, or only normal tissue [3, 4]; or in a task in which multiple types of abnormality or defect must be distinguished from each other and from benign or normal cases [5].

Unfortunately, a fully general extension of ROC analysis to classification tasks with more than two classes has yet to be developed, for the following reasons. It is known that the performance of an observer in a classification task with N classes ($N \geq 2$) can be completely

© 2012 Elsevier Inc. All rights reserved.

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.

described by a set of $N^2 - N$ conditional error probabilities [6], and that the performance of the ideal observer (that which minimizes Bayes risk [6]) is completely characterized by an ROC hypersurface in which these conditional error probabilities depend on a set of $N^2 - N - 1$ decision criteria [7]. Although analytic expressions for the ideal observer's conditional error probabilities given reasonable models for the underlying observational data have been worked out in the two-class case [8], this has not yet been accomplished in a fully general manner for tasks with three or more classes. Furthermore, we have shown that an obvious generalization of the area under the ROC curve (AUC) does not in fact yield a useful performance metric in tasks with three or more classes [9]. More recently, we showed that complicated constraining relationships exist among the decision boundaries themselves for the ideal observer [10]. These constraining relationships appear to imply that it is highly unlikely that analytical expressions for the conditional error probabilities in terms of the decision criteria can be developed which are as simple to interpret as those for the two-class task.

Despite the difficulties just described, the potential benefits to be gained from a practical performance evaluation methodology for classification tasks with three classes have motivated a number of research groups to propose such methods. These practical methods reduce the number of degrees of freedom used to describe the observer's performance, either by implicitly leaving the remaining degrees of freedom out of the analysis, or by explicitly imposing restrictions on the form of the observer's decision rule or on the set of decision criteria used by the observer. An example of the former would be the work of Mossman [11] and of Dreiseitl et al. [12], which proposed reporting performance only in terms of the three "sensitivities" (conditional classification rates for correctly identifying the three classes) in a three-class task; while examples of the latter would include the works of Scurfield [13, 14], of Chan et al. [3, 15], of He et al. [5, 16], and of Schubert et al. [17]. (Recently, He et al. have pointed out [18] that, given complete analytic knowledge of a three-class ideal observer's "sensitivity" ROC surface, whose coordinates are the three correct classification probabilities, one can recover the probability density functions (pdfs) of the pair of likelihood ratios used by that observer. These pdfs, in turn, can be used to recover a complete description of the observer's performance, including the misclassification ROC surface. However, the practical consequences of this theoretical result have yet to be fully investigated; and for the present, we are concerned here with a situation in which the underlying data pdfs are already assumed known, obviating the need for such an intermediate step via measurement of the sensitivity surface.)

The first work by Scurfield mentioned above [13], and that of Schubert et al. [17], are of particular interest because they consider models which assume not only a particular form for the observer's decision rule, but also that the underlying observational data, for which decisions are being made, are univariate. Such a univariate model is worth exploring for both theoretical and pragmatic reasons. Since a univariate model will be more tractable, in general, than a multivariate model, the former can serve as an important step in the development of a fully general classification and performance evaluation methodology; although the development of a suitable performance metric and evaluation methodology is beyond the scope of the present work, this is our long-term goal. More pragmatically, a variety of medical decision-making tasks can be envisioned in which the underlying data consist of a single variable: the result of a particular blood test; or the size of a lesion in a context where other features of the lesion can be ignored, such as a skin test for allergies or tuberculosis. The psychophysical task considered by Scurfield was the ability of an observer to distinguish monotonic auditory stimuli into "low tone," "medium tone," and "high tone" classes based on perceived frequency; the clinical task considered by Schubert et al. was to distinguish chronic allograft nephropathy from normal kidney function and proteinuria based on protein markers present in patient urine.

The decision rule considered in both papers just mentioned is defined by a pair of thresholds on the underlying data; observations less than the lower threshold are assigned to one class, those between the two thresholds to a second class, and those above the higher threshold to a third class. This is a straightforward generalization of two-class classification models (such as the conventional [19] and proper [8] binormal models; recall that the latter is motivated by an ideal observer approach to the underlying data model of the former) in which a single threshold is used to separate univariate data into two classes. While more complicated observer behavior is certainly conceivable, it is perhaps surprising that the three-class ideal observer itself is not limited to such a decision strategy. For normally distributed underlying data, in particular, we show here that, depending on the values of its five decision criteria and the distributional parameters of the underlying data, the three-class ideal observer operates by dividing the underlying data axis into as many as seven regions (i. e., by dividing the underlying data axis with as many as six thresholds). This is by no means intended as a criticism of the works already cited, which do not directly address the ideal observer; the particular observer models developed in those works were explicitly assumed to apply to arbitrary decision strategies, as opposed to specifically ideal observer decision strategies.

A univariate decision variable model in a task with more than two classes was investigated also by Kijewski et al. [20]. However, evaluation was performed only using standard two-class ROC analysis on pairs of classes.

In Sec. 2, we review the theory of the three-class ideal observer, and develop explicit expressions for the ideal observer's likelihood ratio decision variables in the case of normally distributed underlying data. In particular, because the two decision variables are functions of univariate data, it is shown that one of those decision variables can be expressed as a relation of the other (a curve, open or closed, in the decision variable space). Because the form of this relation between the decision variables is determined by the relative means and variances of the three pdfs of the underlying data, conditional on membership in each of the three classes, we consider three distinct situations: in Sec. 3, the case of all three variances being equal; in Sec. 4, the case in which two of the three variances are equal, and the third is distinct from them; and in Sec. 5, the case in which all three variances are distinct. In Sec. 6, these results are used to develop closed-form expressions for the ROC operating points of the ideal observer under consideration. In Sec. 7, we summarize some general properties of parametric surfaces that will be of use in subsequent sections. In Sec. 8, we apply these properties to the ROC surface at hand to obtain constraints on the number of degrees of freedom of that surface. In Sec. 9, we discuss the most salient qualitative features of the preceding results; and finally, the conclusions of the present work are given in Sec. 10.

2. General Theory

We define the actual class (the "truth") to which an observation belongs as $\mathbf{t}$, and the class to which it is assigned (the "decision") as $\mathbf{d}$, where $\mathbf{t}$ and $\mathbf{d}$ can take on any of the values $\pi_1, \dots, \pi_i, \dots, \pi_N$, the labels of the various classes. (We use boldface type to denote statistically variable quantities.) For simplicity, we will also write π_k to denote the event $\mathbf{t} = \pi_k$, as in the a priori probability $P(\pi_k)$ and the conditional pdfs $\pi(\vec{x}|\pi_k)$. It can be shown [6, 7] that an N -class ideal observer makes decisions regarding statistically variable observations $\vec{x}$ by maximizing the expected utility of those decisions. This is done by constructing a set of $N-1$ likelihood ratio decision variables, given by

$$LR_k \equiv \frac{p_{\vec{x}}(\vec{x}|\mathbf{t}=\pi_k)}{p_{\vec{x}}(\vec{x}|\mathbf{t}=\pi_N)} \quad (1)$$

for $k < N$; and then partitioning this likelihood ratio decision variable space, where the boundaries of the partitions are given by segments of hyperplanes:

$$\begin{aligned} &\text{decide } d=\pi_i \text{ iff} \\ &\sum_{k=1}^{N-1} (U_{ik}-U_{jk})P(\mathbf{t}=\pi_k)LR_k > (U_{jN}-U_{iN})P(\mathbf{t}=\pi_N). \end{aligned} \quad (2)$$

(Note that, for continuously distributed data such as we are considering, situations in which

$\sum_{k=1}^{N-1} (U_{ik}-U_{jk})P(\mathbf{t}=\pi_k)LR_k = (U_{jN}-U_{iN})P(\mathbf{t}=\pi_N)$ can be decided arbitrarily without affecting expected utility, because these events will have probability zero.) Here the numbers U_{ij} are the utilities of deciding an observation is from class π_j given that it is actually from class π_i .

Thus in a three-class classification task, the behavior of a three-class ideal observer is completely determined by segments of the three decision boundary lines

$$\gamma_{121}LR_1 - \gamma_{212}LR_2 = \gamma_{313} - \gamma_{323} \quad (3)$$

$$\gamma_{131}LR_1 + (\gamma_{232} - \gamma_{212})LR_2 = \gamma_{313} \quad (4)$$

$$(\gamma_{131} - \gamma_{121})LR_1 + \gamma_{232}LR_2 = \gamma_{323}, \quad (5)$$

which we call, respectively, the “1-vs.-2” boundary, the “1-vs.-3” boundary, and the “2-vs.-3” boundary. Here γ_{jij} is defined as the quantity $(U_{ij} - U_{ji})P(\pi_j)$, which is non-negative if correct classifications are never worse than incorrect ones. These three equations can be all multiplied by a single non-zero constant without changing the decision boundaries; this means the six γ_{jij} can be “normalized” to yield five independent decision criteria. These five decision criteria, directly dependent on the utilities U_{ij} , are analogous to the single decision criterion used by the two-class ideal observer. (That is, a particular combination of the five decision criteria settings yields a particular point on the ideal observer’s ROC hypersurface, just as a particular decision threshold setting yields a particular point on the two-class ideal observer’s ROC curve.)

We now impose the assumption that the pdfs from which the underlying data $\vec{x}$ are drawn are univariate normal distributions. In principle, this would imply a dependence on six underlying parameters, the means and variances of the three distributions. However, note that since the actual labels π_k of the classes are arbitrary, we are free to choose the class with the largest variance as corresponding to class π_3 , the next largest as belonging to class π_2 , and that with the smallest variance as belonging to class π_1 . (Situations in which some of the variances are equal will be handled explicitly in the next few sections.) Furthermore, note that we are free to linearly transform the underlying data variable x so that the π_3 class distribution is a standard normal function. This gives us

$$p(x|\pi_1) = N(x; \mu_1, \sigma_1^2), \quad (6)$$

$$p(x|\pi_2)=N(x;\mu_2,\sigma_2^2), \quad (7)$$

$$p(x|\pi_3)=N(x;0,1), \quad (8)$$

with $\sigma_1^2 \leq \sigma_2^2 \leq 1$, where $N(x;\mu,\sigma^2)$ is a normal (“Gaussian”) function of x with mean μ and variance σ^2 .

The likelihood ratios defined in Eq. 1 can now be written explicitly as

$$LR_1 = \frac{1}{\sigma_1} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x-\mu_1}{\sigma_1} \right)^2 - x^2 \right] \right\} \quad (9)$$

and

$$LR_2 = \frac{1}{\sigma_2} \exp \left\{ -\frac{1}{2} \left[\left(\frac{x-\mu_2}{\sigma_2} \right)^2 - x^2 \right] \right\}. \quad (10)$$

Since both LR_1 and LR_2 are functions of a single variable x , we can of course regard the above expressions as defining a parametric curve $[LR_1(x), LR_2(x)]$, or equivalently $LR_2(LR_1)$, in the likelihood ratio decision variable plane. In the next three sections, we will examine the properties of such curves, and their relationships with the ideal observer’s decision strategy of Eqs. 3–5, for given values of the underlying data parameters ($\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$). We will refer to an example of such a parametric curve in a given context as a “likelihood ratio curve.”

Note that the expressions in Eqs. 9 and 10 are valid for both $\sigma_i^2=1$ and $\sigma_i^2<1$. For the former case, the expressions immediately simplify to

$$LR_i = \exp \left\{ \mu_i x - \frac{\mu_i^2}{2} \right\}; \quad (11)$$

note that the likelihood ratio in this case is monotonic in x , and thus has no extremum with respect to x . (This is mathematically equivalent to the situation in two-class classification in which both classes have equal variances, and the single likelihood ratio is a monotonic function of the single underlying data variable.)

For the case $\sigma_i^2<1$, it will prove convenient to reexpress the expressions above as

$$LR_i = \frac{1}{\sigma_i} \exp \left\{ -\frac{1}{2} \left(\frac{1-\sigma_i^2}{\sigma_i^2} \right) \left[x - \left(\frac{\mu_i}{1-\sigma_i^2} \right) \right]^2 + \frac{\mu_i^2}{2(1-\sigma_i^2)} \right\}; \quad (12)$$

although this appears more cumbersome, it will prove advantageous for the work of the next few sections. For example, it is immediately evident from the above equation that LR_i has an extremum at the value

$$x_{\max} = \frac{\mu_i}{1 - \sigma_i^2}, \quad (13)$$

and that this value in fact corresponds to a maximum of LR_i (since $\sigma_i^2 < 1$). We can carry this line of reasoning further to obtain

$$\begin{aligned} \frac{dLR_2}{dLR_1} &= \frac{\frac{dLR_2}{dx}}{\frac{dLR_1}{dx}} \\ &= \frac{\left(\frac{1-\sigma_2^2}{\sigma_2^2}\right) \left[x - \frac{\mu_2}{1-\sigma_2^2}\right] LR_2}{\left(\frac{1-\sigma_1^2}{\sigma_1^2}\right) \left[x - \frac{\mu_1}{1-\sigma_1^2}\right] LR_1} \quad (14) \end{aligned}$$

and

$$\begin{aligned} \frac{d^2LR_2}{dLR_1^2} &= \frac{\frac{d}{dx} \left(\frac{dLR_2}{dLR_1} \right)}{\frac{dLR_1}{dx}} \\ &= \frac{\frac{dLR_1}{dx} \frac{d^2LR_2}{dx^2} - \frac{dLR_2}{dx} \frac{d^2LR_1}{dx^2}}{\left(\frac{dLR_1}{dx} \right)^3} \\ &= \frac{-1}{\left(\frac{1-\sigma_1^2}{\sigma_1^2} \right)^3 \left(x - \frac{\mu_1}{1-\sigma_1^2} \right)^3 LR_1^3} \left\{ \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(x - \frac{\mu_1}{1-\sigma_1^2} \right) \right. \\ &\quad \times \left(\frac{1-\sigma_2^2}{\sigma_2^2} \right) \left(x - \frac{\mu_2}{1-\sigma_2^2} \right) \\ &\quad \times \left[\left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(x - \frac{\mu_1}{1-\sigma_1^2} \right) - \left(\frac{1-\sigma_2^2}{\sigma_2^2} \right) \left(x - \frac{\mu_2}{1-\sigma_2^2} \right) \right] \\ &\quad \left. - \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(\frac{1-\sigma_2^2}{\sigma_2^2} \right) \left(\frac{\mu_1}{1-\sigma_1^2} - \frac{\mu_2}{1-\sigma_2^2} \right) \right\} LR_1 LR_2. \quad (15) \end{aligned}$$

This expression will prove useful in analyzing in detail the shapes of the curves in LR space defined by Eqs. 9 and 10, in particular in Secs. 4.2 and 5. Since it will turn out to be the case that the likelihood ratio curves of interest in those two cases are closed loops, it will also be useful to consider the *curvature* [21] of such curves, which is similar in spirit to the second derivative of a function but defined more generally for parametric curves as

$$\kappa \equiv \frac{\frac{dLR_1}{dx} \frac{d^2LR_2}{dx^2} - \frac{dLR_2}{dx} \frac{d^2LR_1}{dx^2}}{\left[\left(\frac{dLR_1}{dx} \right)^2 + \left(\frac{dLR_2}{dx} \right)^2 \right]^{3/2}}; \quad (16)$$

note that this has the same numerator as the second derivative in Eq. 15. In particular, the sign of the curvature is determined entirely by the same cubic polynomial appearing in braces in the numerator of Eq. 15:

$$\kappa_p \equiv \left\{ \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(x - \frac{\mu_1}{1-\sigma_1^2} \right) \left(\frac{1-\sigma_2^2}{\sigma_2^2} \right) \left(x - \frac{\mu_2}{1-\sigma_2^2} \right) \right. \\ \times \left[\left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(x - \frac{\mu_1}{1-\sigma_1^2} \right) - \left(\frac{1-\sigma_2^2}{\sigma_2^2} \right) \left(x - \frac{\mu_2}{1-\sigma_2^2} \right) \right] \\ \left. - \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(\frac{1-\sigma_2^2}{\sigma_2^2} \right) \left(\frac{\mu_1}{1-\sigma_1^2} - \frac{\mu_2}{1-\sigma_2^2} \right) \right\}; \quad (17)$$

we will make extensive use of this quantity κ_p in the next three sections.

3. Case I: Three Equal Variances

If all three variances are equal, then by the arguments preceding Eq. 6, we can assume without loss of generality that they are all unit variances. For the purpose of labeling the classes, however, it is more useful in this particular case to select the labels of the classes such that $\mu_2 < \mu_3 (= 0) < \mu_1$. We can safely assume here that no two means are equal, because otherwise at least two of the normal distributions of the underlying data would have both equal means and equal variances; those resulting distributions would thus be identical, observations drawn from those two distributions would be indistinguishable even in principle, and the problem would be at most a two-class classification problem. (This is equivalent to the case in the conventional [19] and proper [8] binormal models in which the curve parameters are $a = 0$ and $b = 1$, and the only possible classification method is equivalent to random “guessing.”)

From Eq. 11, we have

$$\text{LR}_1 = \exp \left\{ \mu_1 x - \frac{\mu_1^2}{2} \right\} \quad (18)$$

and

$$\text{LR}_2 = \exp \left\{ \mu_2 x - \frac{\mu_2^2}{2} \right\}, \quad (19)$$

which can readily be combined to yield

$$\text{LR}_2 = \exp \left(-\frac{\mu_2^2}{2} \right) \left[\text{LR}_1 \exp \left(+\frac{\mu_1^2}{2} \right) \right]^{\mu_2/\mu_1}, \quad (20)$$

a power law relation between LR_2 and LR_1 with negative exponent (due to the signs assumed here for μ_1 and μ_2). Given the sign convention for the means we have adopted, as x increases from $-\infty$ to ∞ , LR_2 decreases from ∞ to 0, while LR_1 increases from 0 to ∞ . The slope of this curve, consistent with Eq. 14, is

$$\frac{d\text{LR}_2}{d\text{LR}_1} = \frac{\mu_2 \text{LR}_2}{\mu_1 \text{LR}_1}, \quad (21)$$

which is everywhere negative by inspection. The second derivative, again consistent with Eq. 15, is

$$\frac{d^2 LR_2}{dLR_1^2} = \frac{\mu_1 \mu_2 (\mu_2 - \mu_1) LR_1 LR_2}{(\mu_1 LR_1)^3}, \quad (22)$$

which is everywhere non-negative by inspection. An example of this case is shown in Fig. 1 with $\mu_1 = 1$ and $\mu_2 = -1$, and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = LR_1$, $LR_1 = 1$, and $LR_2 = 1$, respectively. This set of γ_{jj} can be obtained by choosing, e. g., $U_{1|1} = 1$, $U_{2|2} = P(\pi_1)/P(\pi_2)$, $U_{3|3} = P(\pi_1)/P(\pi_3)$, and $U_{ij} = 0$ for $i \neq j$. (It must be emphasized that this choice of decision criteria, and thus of decision boundaries, is by no means unique; nor is it a consequence of the choice of parameters of the underlying data distributions, any more than the choice of decision criterion is dependent on the a and b parameters of the two-class binormal model. This choice of decision criteria is adopted here only to provide an especially simple set of boundary lines.)

Note that the points in Fig. 1 where the decision boundaries intersect the likelihood ratio curve represent points where the decision being made changes from one class to another. Furthermore, since Eqs. 18 and 19 are both monotonic relations, these points each correspond to a specific value of x with this same property, the specific value determinable by inverting the relations in Eqs. 18 and 19. Thus we have a pair of values of x , let us say x_1 and x_2 , such that our observer decides π_2 if $x < x_1$, decides π_3 if $x_1 \leq x \leq x_2$, and decides π_1 if $x > x_2$. It is clear from this “data space” representation of the ideal observer’s decision rule that it is indeed valid to consider, in a fashion similar to two-class classification, the behavior of the ideal observer decision strategy to be a simple partitioning of the underlying data space via a set of thresholds.

The distinction made here between the decision criteria, γ_{jj} , and the data space thresholds, x_k , is important but potentially confusing. In the two-class case, the decision rule obtained from Eq. 2 is to decide π_1 iff $LR_1 > \gamma_{212}/\gamma_{121}$; that is, the decision criterion (after normalization) is exactly the same as the threshold dividing the decision variable space into the two regions where classes π_1 and π_2 are decided respectively. With three or more classes, however, the decision strategy of Eq. 2, governed by the criteria γ_{jj} , can no longer be described in terms of “thresholds” on the decision variables in any useful way; but because the underlying data x are in the present case univariate, the region of (LR_1, LR_2) space for which a class π_i is decided can always be mapped to a set of segments of the x axis. The data space “thresholds” to which we refer throughout this paper are then simply the boundaries of these segments, but it must be emphasized that they are secondary quantities derived from the γ_{jj} criteria governing the decision boundaries in (LR_1, LR_2) space.

For the **example of** Fig. 1, finding the relevant data space thresholds could be done analytically; in the more complicated cases of subsequent sections, it can in principle always be done numerically. Since we are concerned in this work, however, with developing a broad understanding of the behavior of an ideal observer acting on three-class univariate normal data, we will leave aside the details of such calculations, taking as given that they can indeed be carried out (as was of course done to generate the figures in this paper). We emphasize here that these data space thresholds are dependent on, and thus less fundamental than, the decision criteria γ_{jj} with which the ideal observer makes decisions. Although the choice of values for these criteria in a particular task is an important issue, it is beyond the scope of the present work. Here we seek to characterize the range of behavior available to

the three-class ideal observer (much as, in a two-class task, the ROC curve represents the range of behavior available to the observer, and the choice of the appropriate operating point on this curve must be made based on factors “external” to the curve itself, i. e., the appropriate choice of γ_{jj} and thus of the utilities U_{jj}).

In the present example it is possible to exploit the simple form of the decision boundaries, noting that the boundary $LR_2 = 1$ corresponds to an x such that $p(x|\pi_2) = p(x|\pi_3)$, and similarly $LR_1 = 1$ implies $p(x|\pi_1) = p(x|\pi_3)$; that is, at the values of x where the pdfs “cross” one another. It then follows from Eqs. 18 and 19 that the thresholds in question are $x_1 = -1/2$ and $x_2 = 1/2$, respectively. These thresholds are plotted with the underlying data pdfs in Fig. 2.

Figure 2 is similar to those presented by other researchers when considering the three-class classification of univariate data (see Fig. 1 of Scurfield [13], to name one example). Partitioning the data space into three regions with two thresholds is a straightforward and intuitive extension of the behavior of a typical two-class observer, which partitions the data space into two regions with a single threshold. Even in the two-class case, however, intuition can occasionally be misleading; it is known, for example, that the behavior of the two-class ideal observer, even for univariate normal data, will (except in the special case of equal variances of the two classes) partition the data space with a *pair* of thresholds, with one class being decided for observations between the thresholds, and the other decided for observations outside them [8]. One of the basic purposes of this paper is to demonstrate that the behavior of the three-class ideal observer also involves more complex situations than implied by Fig. 2. Perhaps counterintuitively, this can be shown even in the case of equal variances; consider the same underlying data as before ($\mu_1 = 1, \mu_2 = -1$), but now let the a priori class membership probabilities and decision utilities, and hence the corresponding decision criteria γ_{jj} , be such that the decision boundaries are $LR_2 = 2LR_1 - 4$ (“1-vs.-2”) and $LR_2 = -2LR_1 + 2$ (“2-vs.-3”; it can be shown that for these values, the “1-vs.-3” line is immaterial to the decision strategy [10]). This situation is shown in Fig. 3(a). Although less intuitive than the decision boundaries shown in Fig. 1, such a choice of decision criteria does in fact represent valid behavior of the three-class ideal observer [10]. (Briefly, the decision criteria γ_{jj} are still non-negative, meaning that incorrect decisions are never assigned a greater utility U_{jj} than that of the corresponding correct decision U_{jj} .) This set of γ_{jj} can be obtained by choosing, e. g., $U_{1|1} = 1, U_{2|2} = P(\pi_1)/(2P(\pi_2)), U_{3|3} = 3P(\pi_1)/(2P(\pi_3)), U_{1|2} = U_{3|2} = U_{2|1} = 0, U_{3|1} = -1, U_{1|3} = -3P(\pi_1)/(2P(\pi_3)),$ and $U_{2|3} = P(\pi_1)/(2P(\pi_3)).$

It is seen in Fig. 3(a) that the ideal observer now makes use of *three data space* thresholds, which we can call x_1, x_2 , and x_3 , and decides π_2 if $x < x_1$ or if $x_2 < x < x_3$; decides π_3 if $x_1 < x < x_2$; and decides π_1 if $x_3 < x$. This data space representation of the given decision rule is shown in Fig. 3(b). (We note in passing that if the LR_2 -intercept of the decision boundary separating classes π_2 and π_3 is allowed to decrease sufficiently from 2, then one will eventually arrive at a situation in which that boundary does not make any contact with the likelihood ratio curve at all. All observations will then be classified as either π_1 or π_2 , and none as π_3 , based on comparison with a single threshold; in this sense, the “intuitive” situation in which two thresholds separate the data into three classes is neither the most “complex,” nor the most “simple,” behavior achievable by the ideal observer.)

4. Case II: Two Equal Variances

Recall from the arguments preceding Eqs. 6–8 that we have arbitrarily, but without any loss of generality, adopted the convention that $\sigma_1^2 \leq \sigma_2^2 \leq \sigma_3^2 = 1$. This leaves us with only two

ways in which exactly two of the variances can be equal: $\sigma_1^2 < \sigma_2^2 = \sigma_3^2 = 1$, which we address in Sec. 4.1; and $\sigma_1^2 = \sigma_2^2 < \sigma_3^2 = 1$, which we address in Sec. 4.2.

4.1. Case IIa: $\sigma_2^2 = \sigma_3^2 = 1$

Because the variances σ_2^2 and σ_3^2 are equal, we can conclude as at the beginning of Sec. 3 that the means μ_2 and $\mu_3 (= 0)$ are different; otherwise, the distributions of classes π_2 and π_3 would be identical, and we would in principle have a task equivalent to a two-class problem. Furthermore, it will prove convenient within this section to adopt the additional convention that $\mu_2 < 0$, by transforming the underlying data variable to $-x$ if necessary.

This gives us an expression for LR_1 which is identical to that of Eq. 12:

$$LR_1 = \frac{1}{\sigma_1} \exp \left\{ -\frac{1}{2} \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left[x - \left(\frac{\mu_1}{1-\sigma_1^2} \right) \right]^2 + \frac{\mu_1^2}{2(1-\sigma_1^2)} \right\}; \quad (23)$$

the expression for LR_2 is unchanged from Eq. 19. Rather than attempt to present an explicit relation between LR_2 and LR_1 analogous to Eq. 20, which becomes increasingly less informative as the cases we consider in this paper become more complicated, we note that

$$\frac{dLR_2}{dLR_1} = \frac{-\mu_2 LR_2}{\left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left[x - \frac{\mu_1}{1-\sigma_1^2} \right] LR_1}, \quad (24)$$

and the curvature polynomial κ_p of Eq. 17 is

$$\kappa_p = \left\{ -\left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(x - \frac{\mu_1}{1-\sigma_1^2} \right) \mu_2 \times \left[\left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \left(x - \frac{\mu_1}{1-\sigma_1^2} \right) + \mu_2 \right] + \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right) \mu_2 \right\}. \quad (25)$$

Note that κ_p is now quadratic rather than cubic as in the general case. The roots at which $\kappa_p = 0$ are

$$x_{\text{infl}} = \frac{\mu_1}{1-\sigma_1^2} + \left(\frac{-\mu_2 \pm \sqrt{\mu_2^2 + 4 \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right)}}{2 \left(\frac{1-\sigma_1^2}{\sigma_1^2} \right)} \right); \quad (26)$$

since the quantity within the square root is positive, there will always be two distinct roots, corresponding to inflection points of the likelihood ratio curve. Moreover, since the quantity in the square root sign must be greater than μ_2^2 , the two inflection points will “bracket” the quantity $\mu_1/(1-\sigma_1^2)$, the value of x which maximizes LR_1 from Eq. 13.

With the conventions we have adopted, as x increases from $-\infty$ to ∞ , LR_2 will decrease from ∞ to 0, while LR_1 increases from 0 to its maximum value and then decreases to 0 again. Also, since the leading (quadratic) coefficient in Eq. 25 is the positive quantity $-\mu_2[(1-\sigma_1^2)/\sigma_1^2]^2$, we can say that the curvature changes from positive to negative to positive while the likelihood ratio curve is traversed in the same orientation (i. e., that of increasing x). An example of this situation is shown in Fig. 4(a) with $\mu_1 = -1/3$, $\sigma_1^2 = 1/3$, μ_2

$= -1/2$, and of course $\sigma_2^2=1$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = 2LR_1 - 1.75$, $LR_1 = 1$, and $LR_2 = -3LR_1 + 3.25$, respectively. Unlike the curves of Sec. 3, this likelihood ratio curve has two inflection points (where $\kappa_p = 0$ and changes sign), indicated in Fig. 4(a) with circles; as expected, one of these occurs “before” the maximum of LR_1 (in the direction of increasing x), and the other “after” it.

In Fig. 4(a), it is evident that the given decision boundaries intersect the likelihood ratio curve in five places (although in general, depending on the location of the intersection point and slopes of the decision boundaries, they could of course intersect in fewer than five places). As in the previous section, each such intersection point corresponds to a unique value of x . (Although LR_1 is no longer monotonic in x , it is monotonic in a quadratic function of x ; a given value of LR_1 will thus correspond to two values of x , which can readily be shown to correspond to distinct values of LR_2 .) We can thus represent the given decision strategy in terms of five values of x , namely $x_1 \dots x_5$, such that our observer decides π_2 if $x < x_1$, or if $x_2 < x < x_3$, or if $x_4 < x < x_5$; decides π_3 if $x_1 < x < x_2$ or if $x_5 < x < x_4$; and decides π_1 if $x_3 < x < x_4$. This data space representation of the given decision rule is shown in Fig. 4(b).

4.2. Case IIb: $\sigma_1^2 = \sigma_2^2 < 1$

Because the variances σ_1^2 and σ_2^2 are equal, we can refer to both as simply σ^2 , and **again** conclude that the means μ_1 and μ_2 must be different; otherwise, the distributions of classes π_1 and π_2 would be identical, and we would in principle have a task equivalent to a two-class problem. Furthermore, it will prove convenient within this section to adopt the additional convention that $\mu_2/(1 - \sigma^2) < \mu_1/(1 - \sigma^2)$, or more simply $\mu_2 < \mu_1$, by transforming the underlying data variable to $-x$ if necessary. (This is consistent with the more general convention $\mu_2/(1 - \sigma_2^2) < \mu_1/(1 - \sigma_1^2)$ which we will adopt in the next section; this fixes the “orientation” of the (closed) likelihood ratio curve so that as x increases, the curve is traversed in a clockwise direction.)

This gives us expressions for LR_1 and LR_2 which are essentially identical with those of Eq. 12, with the additional constraint of having a common variance σ^2 :

$$LR_1 = \frac{1}{\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{1 - \sigma^2}{\sigma^2} \right) \left[x - \left(\frac{\mu_1}{1 - \sigma^2} \right) \right]^2 + \frac{\mu_1^2}{2(1 - \sigma^2)} \right\} \quad (27)$$

$$LR_2 = \frac{1}{\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{1 - \sigma^2}{\sigma^2} \right) \left[x - \left(\frac{\mu_2}{1 - \sigma^2} \right) \right]^2 + \frac{\mu_2^2}{2(1 - \sigma^2)} \right\}. \quad (28)$$

As in the preceding section, we claim that an explicit form for the curve $LR_2(LR_1)$, although certainly achievable, is not much more informative than expressions for the first and second derivatives of the likelihood ratio curve, and the (numerically obtained) graphs of particular instances of such curves.

The first derivative of the likelihood ratio curve, again starting from the general form of Eq. 14, is

$$\frac{dLR_2}{dLR_1} = \frac{\left[x - \frac{\mu_2}{1-\sigma^2} \right] LR_2}{\left[x - \frac{\mu_1}{1-\sigma^2} \right] LR_1}, \quad (29)$$

while the curvature polynomial is

$$\kappa_p = \left\{ \left(\frac{1-\sigma^2}{\sigma^2} \right) \left(x - \frac{\mu_1}{1-\sigma^2} \right) \left(\frac{1-\sigma^2}{\sigma^2} \right) \left(x - \frac{\mu_2}{1-\sigma^2} \right) \times \left(\frac{1-\sigma^2}{\sigma^2} \right) \left[\frac{\mu_2}{1-\sigma^2} - \frac{\mu_1}{1-\sigma^2} \right] - \left(\frac{1-\sigma^2}{\sigma^2} \right)^2 \left(\frac{\mu_1}{1-\sigma^2} - \frac{\mu_2}{1-\sigma^2} \right) \right\} \quad (30)$$

$$= \left(\frac{1-\sigma^2}{\sigma^2} \right)^3 \left[\frac{\mu_2}{1-\sigma^2} - \frac{\mu_1}{1-\sigma^2} \right] \times \left\{ \left(x - \frac{\mu_1}{1-\sigma^2} \right) \left(x - \frac{\mu_2}{1-\sigma^2} \right) + \frac{\sigma^2}{1-\sigma^2} \right\}; \quad (31)$$

as in Sec. 4.1, this is quadratic, rather than cubic as in the general case. The quantity outside the braces is strictly negative (given the constraints we have adopted for this section, namely $\sigma^2 < 1$ and $\mu_2 < \mu_1$), and so the roots at which $\kappa_p = 0$ can be shown to be

$$x_{\text{infl}} = \frac{\mu_1 + \mu_2}{2(1-\sigma^2)} \pm \sqrt{\left(\frac{\mu_1 - \mu_2}{2(1-\sigma^2)} \right)^2 - \frac{\sigma^2}{1-\sigma^2}}. \quad (32)$$

If $[(\mu_1 - \mu_2)/2]^2 < \sigma^2(1 - \sigma^2)$, then clearly κ_p has no real roots, and the likelihood ratio curve has negative curvature everywhere. If $[(\mu_1 - \mu_2)/2]^2 = \sigma^2(1 - \sigma^2)$, then κ_p has a single root at $x_{\text{infl}} = (\mu_1 + \mu_2)/[2(1 - \sigma^2)]$. This point is between the two values of x from Eq. 13 where LR_1 and LR_2 take their maxima, but since the sign of the curvature does not change here, it is not an inflection point, and the curve is qualitatively the same as the case with no root. An example of this situation is shown in Fig 5(a) with $\mu_1 = 1/2$, $\mu_2 = -1/2$, and $\sigma^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = LR_1$, $LR_2 = 2LR_1 - 2$, and $LR_2 = LR_1/2 + 1$, respectively. The point at which $\kappa_p = 0$ is indicated with a circle; as just described, it is not a true inflection point, and it lies between the maxima of LR_1 and LR_2 .

In Fig. 5(a), it is evident that the given decision boundaries intersect the likelihood ratio curve in four places (although in general, depending on the location of the intersection point and slopes of the decision boundaries, they could of course intersect in fewer than four places). We note here in passing that although it might appear from this figure that two of those four intersections lie along the LR_i axes, this is in fact not the case; from Eqs. 27 and 28, it must be the case that LR_1 is greater than zero if LR_2 is, and vice versa (the likelihood ratio curve does of course approach these axes asymptotically). As in previous sections, each such intersection point corresponds to a unique value of x . We can thus represent the given decision strategy in terms of four values of x , namely $x_1 \dots x_4$, such that our observer decides π_3 if $\mathbf{x} < x_1$, or if $x_2 < \mathbf{x} < x_3$, or if $x_4 < \mathbf{x}$; decides π_2 if $x_1 < \mathbf{x} < x_2$; and decides π_1 if $x_3 < \mathbf{x} < x_4$. This data space representation of the given decision rule is shown in Fig. 5(b).

If on the other hand $[(\mu_1 - \mu_2)/2]^2 > \sigma^2(1 - \sigma^2)$, then κ_p has two distinct roots; and since the sign of κ_p must change at each, these are clearly inflection points. Furthermore, note that the value of κ_p when x is $\mu_1/(1 - \sigma^2)$ or $\mu_2/(1 - \sigma^2)$, namely the maxima of LR_1 and LR_2 , is the negative quantity

$$\left(\frac{1-\sigma^2}{\sigma^2}\right)^2 \left[\frac{\mu_2}{1-\sigma^2} - \frac{\mu_1}{1-\sigma^2} \right]; \quad (33)$$

since κ_p tends to $-\infty$ as x tends to $\pm\infty$, we can conclude that the roots of κ_p (the inflection points of the likelihood ratio curve) both lie between the maxima of LR_1 and LR_2 . An example of this situation is shown in Fig 6(a) with $\mu_1 = 1$, $\mu_2 = -1/2$, and $\sigma^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = LR_1$, $LR_2 = -LR_1/2 + 1.55$, and $LR_2 = 1.5$, respectively. The points at which $\kappa_p = 0$ are indicated with circles; as just described, these correspond to inflection points and lie between the maxima of LR_1 and LR_2 .

In Fig. 6(a), it is evident that the given decision boundaries intersect the likelihood ratio curve in five places (although in general, depending on the location of the intersection point and slopes of the decision boundaries, they could of course intersect in fewer than five places). As in previous sections, each such intersection point corresponds to a unique value of x . We can thus represent the given decision strategy in terms of five values of x , namely $x_1 \dots x_5$, such that our observer decides π_3 if $x < x_1$, or if $x_3 < x < x_4$, or if $x_5 < x$; decides π_2 if $x_1 < x < x_2$; and decides π_1 if $x_2 < x < x_3$ or if $x_4 < x < x_5$. This data space representation of the given decision rule is shown in Fig. 6(b).

5. Case III: Three Different Variances

We turn at last to the most general case, in which $\sigma_1^2 < \sigma_2^2 < \sigma_3^2 = 1$. The parametric expressions for the likelihood ratios, their first derivatives, and the curvature polynomial are as expressed in Eqs. 12, 14, and 17.

In the more general case in which $\mu_1/(1-\sigma_1^2)$ and $\mu_2/(1-\sigma_2^2)$ are distinct, we adopt the convention $\mu_2/(1-\sigma_2^2) < \mu_1/(1-\sigma_1^2)$ (by transforming the underlying data variable to $-x$ if necessary); as mentioned in Sec. 4.2, this has the effect of causing the loop formed by the likelihood ratio curve to be traversed in a clockwise direction as x increases. (We will consider the special case in which $\mu_1/(1-\sigma_1^2) = \mu_2/(1-\sigma_2^2)$ at the end of this section.) **Because** $\sigma_1^2 < \sigma_2^2$, κ_p of Eq. 17 is seen to have a (leading) cubic term with the strictly positive coefficient

$$\left(\frac{1-\sigma_1^2}{\sigma_1^2}\right) \left(\frac{1-\sigma_2^2}{\sigma_2^2}\right) \left(\frac{1-\sigma_1^2}{\sigma_1^2} - \frac{1-\sigma_2^2}{\sigma_2^2}\right); \quad (34)$$

intuitively, the curvature is initially negative as x increases from $-\infty$, and is positive as x increases to $+\infty$. Further algebraic manipulation reveals that κ_p is the sum of the strictly negative constant term

$$-\left(\frac{1-\sigma_1^2}{\sigma_1^2}\right) \left(\frac{1-\sigma_2^2}{\sigma_2^2}\right) \left(\frac{\mu_1}{1-\sigma_1^2} - \frac{\mu_2}{1-\sigma_2^2}\right) \quad (35)$$

and a cubic polynomial with the three distinct roots $\mu_1/(1-\sigma_1^2)$, $\mu_2/(1-\sigma_2^2)$, and

$$\frac{\left(\frac{1-\sigma_1^2}{\sigma_1^2}\right)\left(\frac{\mu_1}{1-\sigma_1^2}\right) - \left(\frac{1-\sigma_2^2}{\sigma_2^2}\right)\left(\frac{\mu_2}{1-\sigma_2^2}\right)}{\frac{1-\sigma_1^2}{\sigma_1^2} - \frac{1-\sigma_2^2}{\sigma_2^2}}. \quad (36)$$

(This third root is distinct because it can be shown to be strictly greater than $\mu_1/(1-\sigma_1^2)$.) A cubic polynomial with three distinct roots, of course, has a local maximum and a local minimum; since adding a constant term does not change the shape of the curve, this statement is true of κ_p itself as well. Furthermore, since the constant term being added is strictly negative, we need distinguish only two possible situations. If the local maximum of κ_p is zero, then κ_p has two roots: one at this local maximum, which lies between $\mu_2/(1-\sigma_2^2)$ and $\mu_1/(1-\sigma_1^2)$ and is not an inflection point (since the sign of κ_p does not change here), and an inflection point which lies at some value of x greater than $\mu_1/(1-\sigma_1^2)$. Similarly, if the local maximum of κ_p is negative, then κ_p has only a single root, an inflection point which again lies at some value of x greater than $\mu_1/(1-\sigma_1^2)$. An example of this situation is shown in Fig. 7(a) with $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -1/4$, and $\sigma_2^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = LR_1$, $LR_2 = 2LR_1 - 2$, and $LR_2 = (7/8)LR_1 + 1/4$, respectively. The point at which $\kappa_p = 0$ is indicated with a circle; as expected, it is an inflection point, and lies “after” the maximum of LR_1 (in the direction of increasing x). The data space representation of this decision rule, again involving six threshold values of x , is shown in Fig. 7(b) (although the thresholds of the third region are so close together as to be difficult to distinguish in the graph).

On the other hand, if the local maximum of κ_p is positive, then κ_p has three distinct roots, all inflection points: two lying between $\mu_2/(1-\sigma_2^2)$ and $\mu_1/(1-\sigma_1^2)$ (cf. the situation described in Sec. 4.2), and a third lying at some value of x greater than $\mu_1/(1-\sigma_1^2)$. An example of this situation is shown in Fig. 8(a) with $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -2/3$, and $\sigma_2^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = (1/4)LR_1$, $LR_2 = (3/10)LR_1 - 3/20$, and $LR_2 = -(3/8)LR_1 + 17/8$, respectively. The points at which $\kappa_p = 0$ are indicated with circles; as expected, all three are inflection points, two of which lie between the maxima of LR_1 and LR_2 , and one of which lies “after” the maximum of LR_1 (in the direction of increasing x). The data space representation of this decision rule, again involving six threshold values of x , is shown in Fig. 8(b). Because the graph in Fig. 8(a) bears some qualitative similarities to Fig. 4(a), we also show an expanded view of a small region of that figure near the maximum of LR_2 in Fig. 9, which demonstrates that LR_2 does indeed increase from zero to this maximum, that the first inflection point occurs not at this maximum but at a slightly higher value of x , and that the likelihood ratio curve has a finite curvature at this maximum (despite what appears, in Fig. 8(a), to be a “cusp” at that point due to the resolution of that figure).

In the special case $\mu_1/(1-\sigma_1^2) = \mu_2/(1-\sigma_2^2)$ mentioned at the beginning of this section (i. e., the case in which the maxima of LR_1 and LR_2 occur at the same point on the likelihood ratio curve), κ_p is still cubic, but has a triple root at this common value of $x = \mu_i/(1-\sigma_i^2)$. Since the sign of the curvature changes here, this point can be said to be an inflection point. However, the likelihood ratio curve in this case forms a loop only in a degenerate sense, because LR_2 is actually a single-valued function of LR_1 ; as a parametric function of x , LR_1 and LR_2 will increase from zero to their maxima, and then decrease back to zero while tracing along the same curve. This follows from the fact that both are monotonic functions of quadratic forms

with maxima at the same location; more intuitively, if one reexpresses Eq. 12 in terms of $x = \mu_i / (1 - \sigma_i^2) + \Delta x$, both likelihood ratios are monotonic functions of $(\Delta x)^2$. An example of this situation is shown in Fig. 10(a) with $\mu_1 = 1/10$, $\sigma_1^2 = 1/10$, $\mu_2 = 1/18$, and $\sigma_2^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = 2LR_1 - 2$, $LR_2 = -2LR_1 + 6$, and $LR_2 = (3/4)LR_1 + 1/2$, respectively. The point at which $\kappa_p = 0$ is indicated with a circle; as expected, it is technically an inflection point, and lies at the common maximum of LR_1 and LR_2 .

In Fig. 10(a), it is evident that the given decision boundaries intersect the likelihood ratio curve in three places (although in general, depending on the location of the intersection point and slopes of the decision boundaries, they could of course intersect in fewer than three places). Unlike the situations presented in previous sections, each such intersection point corresponds to a pair of values of x (the $\mu_i / (1 - \sigma_i^2) \pm \Delta x$ just mentioned). We can thus represent the given decision strategy in terms of six distinct values of x , namely $x_1 \dots x_6$, such that our observer decides π_3 if $x < x_1$, or if $x_2 < x < x_3$, or if $x_4 < x < x_5$, or if $x_6 < x$; decides π_2 if $x_1 < x < x_2$ or if $x_5 < x < x_6$; and decides π_1 if $x_3 < x < x_4$. This data space representation of the given decision rule is shown in Fig. 10(b).

6. ROC Operating Points

An example of a likelihood ratio curve, as described in the preceding sections, is shown in Fig. 8(a) with $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -2/3$, and $\sigma_2^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = (1/4)LR_1$, $LR_2 = (3/10)LR_1 - 3/20$, and $LR_2 = -(3/8)LR_1 + 17/8$, respectively. It is shown in Sec. 5 that one can represent **this particular** decision strategy in terms of six distinct values of x , namely $x_1 \dots x_6$, such that our observer decides π_3 if $x < x_1$, or if $x_2 < x < x_3$, or if $x_4 < x < x_5$, or if $x_6 < x$; decides π_2 if $x_1 < x < x_2$ or if $x_3 < x < x_6$; and decides π_1 if $x_5 < x < x_6$. This data space representation of the given decision rule is shown in Fig. 8(b). In general, for different choices of underlying data parameters and decision criteria, one can have fewer than six data space thresholds x_i .

For a given set of decision criteria, one can calculate or estimate the nine classification probabilities $P(\mathbf{d} = \pi_i | \mathbf{t} = \pi_j)$, where $\mathbf{t}$ is the actual class to which an observation belongs, and $\mathbf{d}$ is the class to which it is assigned; we write this quantity more succinctly as P_{ij} . Since these are conditional probabilities, we can eliminate three of them; we have shown previously [9, 22] that the performance of any observer in a three-class classification task can be completely described using the six possible misclassification probabilities $P_{ij} (i \neq j)$. Other researchers have proposed simplifying the description of observer performance by reporting performance only in terms of the three sensitivities P_{ij} [11, 5] or in terms of the three probabilities P_{3j} [3, 15]; however, our long-term goal is a general performance evaluation method not subject to such simplifications.

As the ideal observer's decision criteria γ_{ij} are swept through their possible values, the resulting operating points sweep out a surface in the ROC space whose coordinates are the six misclassification probabilities. Moreover, since the set of decision criteria can be multiplied by any constant scale factor without changing the decision boundary line segments of Eqs. 3–5, there are only five degrees of freedom amongst the γ_{ij} , and thus at most five degrees of freedom along such a “misclassification” ROC surface in the most general case [7].

Based on the preceding paragraphs, it follows that, for the example underlying data and decision criteria choices of Fig. 8(a),

$$P_{12}=P(x_5 < x < x_6 | \pi_2). \quad (37)$$

Combining this with Eq. 7 gives

$$P_{12}=\Phi\left(\frac{x_6-\mu_2}{\sigma_2}\right)-\Phi\left(\frac{x_5-\mu_2}{\sigma_2}\right), \quad (38)$$

where

$$\Phi(x) \equiv \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x'^2}{2}\right\} dx' \quad (39)$$

is the standard normal cumulative probability distribution function (cdf).

Similarly, it is readily shown that

$$P_{13}=\Phi(x_6)-\Phi(x_5) \quad (40)$$

$$P_{2j}=\Phi\left(\frac{x_2-\mu_j}{\sigma_j}\right)-\Phi\left(\frac{x_1-\mu_j}{\sigma_j}\right)+\Phi\left(\frac{x_4-\mu_j}{\sigma_j}\right)-\Phi\left(\frac{x_3-\mu_j}{\sigma_j}\right) \quad (41)$$

$$P_{3j}=1-\Phi\left(\frac{x_6-\mu_j}{\sigma_j}\right)+\Phi\left(\frac{x_5-\mu_j}{\sigma_j}\right)-\Phi\left(\frac{x_4-\mu_j}{\sigma_j}\right)+\Phi\left(\frac{x_3-\mu_j}{\sigma_j}\right)-\Phi\left(\frac{x_2-\mu_j}{\sigma_j}\right)+\Phi\left(\frac{x_1-\mu_j}{\sigma_j}\right). \quad (42)$$

Although the details of the above expressions depend strongly on the parameters of the underlying data $(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2)$ and on the particular choices of decision criteria γ_{jj} , the same general form (sums of differences of standard normal cdfs) will be obtained for all such choices. For a given set of decision criteria γ_{jj} , the intersection points $x_1 \dots x_6$ can be found numerically, as stated in Sec. 3. The ROC operating point coordinates corresponding to this choice of decision criteria is then given by equations of the form of Eqs. 38–42. Of course, as shown in Secs. 3–5, it may be the case that, for a given set of values of the underlying data parameters and a particular set of decision criteria, there may actually be fewer than six such intersection points of the likelihood ratio curve and the decision boundary line segments. In that case, the expressions for the P_{ij} will also be simpler, but will still follow the basic pattern shown above, i. e., sums of differences of values of $\Phi(\cdot)$.

We claimed at the end of Sec. 3 that the behavior of the ideal observer can be shown to be quantitatively different from that of the “intuitive” observer which partitions the data space into three regions with two thresholds [13, 17]. We are now in a position to verify this explicitly, for the first example considered by Schubert et al. The data model of that example, after transformation to the conventions we adopt in Secs. 3–5 (namely $\mu_3 = 0$, $\sigma_3^2 = 1$, $-\sigma_1^2 < \sigma_2^2$, and $\mu_2/(1-\sigma_2^2) < \mu_1/(1-\sigma_1^2)$), is given by $\mu_1 = \sqrt{2}$, $\sigma_1^2 = 1/16$, $\mu_2 = \sqrt{2}/2$, and $\sigma_2^2 = 1/4$. The misclassification costs c_{ij} (our $-U_{ij}$) are uniform, and the priors $P(\pi_k)$ are all $1/3$; this yields uniform $\gamma_{jj} = 1/3$, with a resulting expected utility [22] of

$$E\{U\} = - \sum_{i,j \neq i} \gamma_{ij} P_{ij}; \quad (43)$$

the negative of this is the Bayes cost $q(\mathbb{A})$ evaluated by Schubert et al. [17] This situation is shown in Fig. 11(a), and the corresponding data space representation is shown in Fig. 11(b); the latter is essentially identical (given the coordinate transformation mentioned above) to Fig. 2 of Schubert et al. [17], except for the presence of the third threshold x_3 . The ideal observer operating point is given by

$$P_{1j} = \Phi\left(\frac{x_3 - \mu_j}{\sigma_j}\right) - \Phi\left(\frac{x_2 - \mu_j}{\sigma_j}\right) \quad (44)$$

$$P_{2j} = \Phi\left(\frac{x_2 - \mu_j}{\sigma_j}\right) - \Phi\left(\frac{x_1 - \mu_j}{\sigma_j}\right) \quad (45)$$

$$P_{3j} = 1 - \Phi\left(\frac{x_3 - \mu_j}{\sigma_j}\right) + \Phi\left(\frac{x_1 - \mu_j}{\sigma_j}\right) \quad (46)$$

These evaluate to $P_{12} = 0.232$, $P_{13} = 0.124$, $P_{21} = 0.084$, $P_{23} = 0.311$, $P_{31} = 0.004$, and $P_{32} = 0.121$, which gives a cost of 0.292. This cost is less than the 0.297 achieved by the observer considered by Schubert et al.; this is consistent with the definition of the ideal observer as that which optimizes expected utility.

7. Parametric Surfaces and Jacobians

Consider an n -dimensional vector space with elements $\vec{v} \equiv (v_1, v_2, \dots, v_n)$. A function such as

$$v_n = f(v_1, v_2, \dots, v_{n-1}) \quad (47)$$

defines a surface in this space. If we now take a point $\vec{v}$ on this surface, and a closely neighboring point $\vec{v} + \Delta \vec{v}$ also assumed to lie on the surface, then the difference between those points will approach a tangent vector to the surface in the limit as $\Delta \vec{v}$ approaches zero

in magnitude. In particular, the $n-1$ vectors $(0, 0, \dots, \frac{1}{(i)}, 0, \dots, \partial f / \partial v_i)^\dagger$ form a linearly independent set of such tangent vectors. (The dagger symbol indicates the matrix transpose; we are typically concerned with column vectors here.) This is the sense in which we say the surface has $n-1$ degrees of freedom.

Now suppose we have a parametric surface defined by the set of n functions

$$v_i = g_i(u_1, u_2, \dots, u_{n-1}). \quad (48)$$

(Later we will generalize this to include situations in which the number of parameters u_j is different from the number of dependent variables v_i , as explained in more detail at the end of this section. We begin, however, with the straightforward case in which these numbers are equal.) By the implicit function theorem, we can again obtain v_n as a function of $(v_1, \dots, v_{n-1})$ as in Eq. 47 [21]; however, we are no longer assuming that this function has $n-1$ linearly independent tangent vectors at every point (take, e. g., the case of a curve in a three-

dimensional space). We will from here on refer to the number of linearly independent tangent vectors to a surface at a point on that surface as the number of degrees of freedom of the surface at that point.

If we now take a point $\vec{u}$ in the parameter space, and a closely neighboring point $\vec{u} + \Delta\vec{u}$, then the difference between the corresponding points $\vec{v}$ will approach a tangent vector to the surface in the limit as $\Delta\vec{u}$ approaches zero in magnitude. In particular, if we consider a $\Delta\vec{u}$ of the form $(0, 0, \dots, \Delta u_j, \dots, 0)^\dagger$, the resulting tangent vector is

$$\Delta u_j (\partial v_1 / \partial u_j, \partial v_2 / \partial u_j, \dots, \partial v_{n-1} / \partial u_j, \partial v_n / \partial u_j)^\dagger. \quad (49)$$

To determine if the $n - 1$ tangent vectors of this form are linearly independent or not, we can form an $n \times (n - 1)$ matrix with these vectors as columns, and then ask if the resulting matrix is full rank or not. Note that the n th row of this matrix consists of elements of the form

$$\frac{\partial v_n}{\partial u_j} = \sum_i \frac{\partial f}{\partial v_i} \frac{\partial v_i}{\partial u_j}, \quad (50)$$

meaning that the n th row is a linear combination of the first $n - 1$ rows of the matrix. Consequently, the rank of the matrix in question is just the rank of the $(n - 1) \times (n - 1)$ matrix consisting of the first $n - 1$ rows of the original matrix, i. e., a matrix whose ij th element is $\partial v_i / \partial u_j$. But this matrix in turn is just the Jacobian of the coordinate transformation determined by the first $n - 1$ equations of Eq. 48 [21].

Thus, in our case, the number of degrees of freedom of the ROC surface

$$P_{32} = R(P_{12}, P_{13}, P_{21}, P_{23}, P_{31}) \quad (51)$$

at any point is found by constructing the Jacobian of the transformation from decision criteria γ_{jij} (only five of which are independent, as explained in Sec. 6) to misclassification probabilities P_{ij} (not counting the dependent variable P_{32}), and then finding the rank of this Jacobian. That is, the misclassification probabilities P_{ij} are analogous to the v_i of the above general example, and the decision criteria γ_{jij} are analogous to the u_j . In the past, we have employed a normalization convention such that the six γ_{jij} form a unit vector [23]; here, however, it will prove more convenient to adopt a normalization convention such that $\gamma_{232} = 1$. With this convention, the Jacobian in question is **a matrix with components**

$$[J_{\gamma \rightarrow P}]_{ij} = \frac{\partial P_i}{\partial \gamma_j} \quad (52)$$

where P_i is the i th component of the vector $(P_{12}, P_{13}, P_{21}, P_{23}, P_{31})^\dagger$, and γ_j is the j th component of the vector $(\gamma_{212}, \gamma_{313}, \gamma_{121}, \gamma_{323}, \gamma_{131})^\dagger$.

These γ_{jij} appearing in Eqs. 3–5 have a fundamental theoretical motivation, but in practice, it tends to be more intuitive to parametrize the decision boundary line segments using their common intersection point ($LR_1 = a$, $LR_2 = b$), and the slopes of these three decision boundaries m_{12} , m_{13} , and m_{23} , respectively. To show that these also form a “good” set of coordinates, we could in principle form the Jacobian $J_{m \rightarrow \gamma}$, and then show that this Jacobian is invertible everywhere that either γ or $\vec{m}$ are defined. (We denote the vector $(a, b, m_{12},$

$m_{13}, m_{23})^\dagger$ by $\vec{m}$ because its components are mostly slopes.) It is a bit simpler, however, just to show that there is a unique $\vec{m}$ for a given γ and vice versa, and that the transformation is thus one-to-one. We must, of course, assume that the $\vec{m}$ in question is “valid”, i. e., that the resulting set of decision boundary line segments satisfies the properties that they must have if derived from a set of γ [10].

For a given $\vec{\gamma}$, it is clear from Eq. 3 that $m_{12} = \gamma_{121}/\gamma_{212}$; similarly, $m_{13} = \gamma_{131}/(\gamma_{212} - 1)$, and $m_{23} = (\gamma_{121} - \gamma_{131})$. If the “1-vs.-2” and “1-vs.-3” decision boundaries intersect at a point (a, b) , then the “2-vs.-3” decision boundary must also pass through this point since Eq. 5 is just the difference of Eqs. 4 and 3; similarly, if the first two are parallel, the third will also be parallel to those two (making (a, b) effectively at infinity). In either case, it is clear that the components of $\vec{m}$ are uniquely determined by those of γ .

Working in the opposite direction, if we have a given $\vec{m}$, then we can immediately write

$$\alpha_1 \left(\frac{LR_2 - b}{LR_1 - a} \right) = \alpha_1 m_{12} \quad (53)$$

$$\alpha_2 \left(\frac{LR_2 - b}{LR_1 - a} \right) = \alpha_2 m_{13} \quad (54)$$

$$\alpha_3 \left(\frac{LR_2 - b}{LR_1 - a} \right) = \alpha_3 m_{23}, \quad (55)$$

where the scale factors α_i are introduced here because they do not change the location of the lines, but could in principle change the values of the resulting γ . These equations can be rearranged to more closely correspond to Eqs. 3–5:

$$\alpha_1 m_{12} LR_1 - \alpha_1 LR_2 = \alpha_1 (a m_{12} - b) \quad (56)$$

$$\alpha_2 m_{13} LR_1 - \alpha_2 LR_2 = \alpha_2 (a m_{13} - b) \quad (57)$$

$$\alpha_3 m_{23} LR_1 - \alpha_3 LR_2 = \alpha_3 (a m_{23} - b). \quad (58)$$

Comparing these equations with Eqs. 3–5, we find that our normalization condition $\gamma_{232} = 1$ implies $\alpha_3 = -1$; and the requirement that Eq. 4 minus Eq. 3 must yield Eq. 5 implies that $\alpha_2 = \alpha_1 - 1$ and $\alpha_1 = (m_{13} - m_{23})/(m_{13} - m_{12})$, leaving no remaining degrees of freedom. The resulting values of γ can then be read off from these equations; for example, $\gamma_{121} = \alpha_1 m_{12} = m_{12}(m_{13} - m_{23})/(m_{13} - m_{12})$. The relation between γ and $\vec{m}$ is thus one-to-one, which means that, instead of the Jacobian of Eq. 52, we are free to consider the Jacobian **defined** by

$$[J_{m \rightarrow \gamma}]_{ij} = \frac{\partial P_i}{\partial m_j}, \quad (59)$$

where m_j is the j th component of the vector $(a, b, m_{12}, m_{13}, m_{23})^\dagger$, as will be done in the next section.

It will also prove convenient to be able to generalize the concept of the Jacobian to include situations in which the number of coordinates in the parametric spaces of interest are not necessarily the same. That is, we will refer to a matrix of the above form, with elements P_{ij} for some set of **parameters** u_j , as a Jacobian, even if the resulting matrix fails to be square. In our situation, if there are more than five such u_j , then the transformation can be regarded as a form of projection (from a higher-dimensionality space to a lower); while if there are fewer than five such u_j , the resulting parametric “surface” in the ROC space with coordinates P_{ij} will be degenerate (having fewer than five degrees of freedom, analogous to a curve in a three-dimensional space).

8. Degrees of Freedom of ROC Surfaces

In Secs. 3–5 and Appendix A, we show that, for univariate trinormal data, the decision boundary line segments used by the ideal observer can intersect the likelihood ratio curve in at most six places; an example of such a situation is shown in Fig. 8(a). More generally, the number of such intersections depends in a complicated way on the parameters of the underlying data (the means and variances of the data pdfs) and on the choice of decision criteria governing the decision boundaries (the γ_{jj} of Eq. 3–5 or, equivalently, the boundary slopes and intersection point $\vec{m}$ determined by these parameters). We show in this section that the rank of the Jacobian of Eq. 59 depends primarily on the number of intersections of the decision boundaries and the likelihood ratio curve. More particularly, the behavior of $J_{m \rightarrow p}$ falls into three categories: one when there are six such intersection points; another when there are four or fewer intersection points; and a third when there are five intersection points.

It will prove useful to note first that a given set of decision boundary parameters $\vec{m}$ completely determines the set of data-space thresholds x_k corresponding to the intersection points of those decision boundaries with the likelihood ratio curve. These thresholds x_k , in turn, completely determine the values of the P_{ij} , as was shown in Eq. 38 and Eqs. 40–42. This means that the Jacobian of Eq. 59 can be factored into a pair of Jacobians, by application of the chain rule to each element:

$$\frac{\partial P_{ij}}{\partial m_k} = \frac{\partial P_{ij}}{\partial x_l} \frac{\partial x_l}{\partial m_k}, \quad (60)$$

yielding

$$J_{m \rightarrow p} = J_{x \rightarrow p} J_{m \rightarrow x}. \quad (61)$$

(If we restricted ourselves to situations in which there are exactly five such x_k , then we could regard this simply as a succession of changes of coordinate representations. As described at the end of the preceding section, however, we wish to allow for more general situations, analogous to projections and degenerate “surfaces.”)

8.1. Six intersection points

Consider the case in which there are six intersection points of the decision boundaries with the likelihood ratio curve, an example of which is given in Fig. 8(a). In Appendix A we show that in such a case, one of the decision boundary line segments must fail to intersect the likelihood ratio curve at all. It follows that the slope of this decision boundary m_k can be varied without changing any of the ROC coordinates P_{ij} . This in turn means that at least one of the last three columns of $J_{m \rightarrow p}$ as given in Eq. 59 will be a zero vector. The Jacobian $J_{m \rightarrow p}$ is not full rank in this case, and the corresponding ROC surface has no more than four

linearly independent tangent vectors, i. e., degrees of freedom, at this point (i. e., the operating point determined by the decision criteria corresponding to $\vec{m}$).

8.2. Four or fewer intersection points

In considering the case in which there are four or fewer intersection points of the decision boundaries with the likelihood ratio curve, we make use of the factorization of the Jacobian $J_{m \rightarrow p}$ described in Eqs. 60 and 61. For example, for exactly four such intersection points, we have that the first matrix on the right hand side of Eq. 61, namely $J_{x \rightarrow p}$ has only 4 columns; while the second matrix, namely $J_{m \rightarrow x}$, is a 4×5 matrix. The rank of each is thus necessarily four or less [21]. It follows that the rank of the product, $J_{m \rightarrow p}$, is also four or less. The same argument applies if we have three or fewer intersection points ($J_{m \rightarrow p}$ is the product of a $5 \times k$ and a $k \times 5$ matrix, and thus has rank less than or equal to k with $k \leq 3$). We again conclude that the Jacobian $J_{m \rightarrow p}$ is not full rank in this case, and the corresponding ROC surface has less than five linearly independent tangent vectors, i. e., degrees of freedom, at this point.

8.3. Five intersection points

In the situation in which there are exactly five intersection points of the decision boundaries with the likelihood ratio curve, the 5×5 Jacobian $J_{m \rightarrow p}$ factors into two 5×5 matrices, namely $J_{x \rightarrow p}$ and $J_{m \rightarrow x}$. Note that if the common intersection of the three decision boundary lines lies outside the likelihood ratio curve, this will again result in a situation, similar to that described in Sec. 8.1, in which one of the decision boundary line segments fails to intersect the likelihood ratio curve, resulting in a Jacobian $J_{m \rightarrow p}$ of less than full rank. We will thus consider here only situations in which the intersection of the three decision boundaries lies inside the likelihood ratio curve, an example of which is shown in Fig. 12 with $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -2/3$, and $\sigma_2^2 = 1/2$; and with an arbitrary choice of the decision criteria of Eqs. 3–5 such that the decision boundaries are segments of the lines $LR_2 = (1/4)LR_1 - 1/16$, $LR_2 = (3/10)LR_1 - 3/40$, and $LR_2 = -(4/9)LR_1 + 31/36$, respectively. In Appendix B, we show that for this special case, although it is true that $J_{x \rightarrow p}$ is full rank, the Jacobian $J_{m \rightarrow x}$ is not, and therefore $J_{m \rightarrow p}$ also fails to be full rank.

It is, furthermore, not difficult to see that this special case, in fact, adequately represents the general situation when there are exactly five intersection points of the decision boundaries and the likelihood ratio curve. As explained in Sec. 8.1, if the intersection point of the three decision boundaries lies outside the likelihood ratio curve, then $J_{m \rightarrow p}$ will fail to be full rank. On the other hand, if the intersection point of the three decision boundaries lies inside the likelihood ratio curve, then it must be the case that one of the decision boundaries intersects the likelihood ratio curve in three places, and the other two boundaries intersect in one place each. (The five intersection points cannot be distributed with two on two of the boundaries and one on the remainder, because then the intersection point of the three decision boundaries cannot consistently be defined as either inside or outside the likelihood ratio curve.) For such a situation, careful examination of the argument of Appendix B will show that $J_{m \rightarrow x}$ is always less than full rank; although the particular points x_j associated with a given decision boundary line (the one with slope m_j) will of course depend on the details of the shape of the likelihood ratio curve and the locations of the decision boundary lines, the behavior of the determinant of the resulting Jacobian will be the same as that found in the example of Appendix B: two columns with four zeros each (corresponding to the decision boundaries which only intersect the likelihood ratio curve once each), and a remaining 3×3 sub-matrix with determinant zero (because the $\frac{1}{a}$ column is $-m_j$ times the $\frac{1}{b}$ column due to Eq. B.7).

A more geometric, and thus perhaps more intuitive, way to see this is to consider the decision boundary line segment which intersects the likelihood ratio curve three times; call its slope m_1 . We can consider “shifting” the decision boundaries by first moving their intersection point (a, b) by some small amount $(\Delta a, \Delta b)$ such that $\Delta b/\Delta a = m_1$; note that the three intersection points lying on this decision boundary will be unchanged as a result of such a shift. This shift will, as a side effect, alter the intersection points of the other two decision boundaries; however, we are free to impose independently changes in the slopes of these other two decision boundaries Δm_2 and Δm_3 to obtain the original intersection points. The practical outcome of these machinations is that we will have found a non-zero $\Delta \vec{m} = (\Delta a, m_1 \Delta a, 0, \Delta m_2, \Delta m_3)$ which leaves unchanged the intersection points of the decision boundaries with the likelihood ratio curve, and thus the coordinates P_{ij} of the corresponding ROC operating point via equations of the form of Eqs. 38–42. By the arguments of Sec. 7, the resulting ROC surface must have fewer than five tangent vectors, i. e., degrees of freedom, at this operating point. Noting the generality of this result (i. e., its applicability to any set of decision criteria such that the decision boundaries intersect the likelihood ratio curve in five places), and combining it with the results of Secs. 8.1 and 8.2, we are forced to conclude that the ROC surface of a three-class ideal observer for univariate trinormal data must everywhere have strictly fewer than five degrees of freedom.

9. Discussion

As implied in Sec. 1, the behavior of the ideal observer even for a simple, restricted underlying data model can be quite complicated, particularly when viewed in terms of that underlying data. However, we would like to reiterate the point made in Sec. 4.1, that one of the goals of this paper was not to focus on the process of determining the numerical values of the data space thresholds on x , but rather to develop a more qualitative but broad understanding of the ideal observer’s behavior under this data model. Again, the analogy made in Sec. 3 with the two-class proper binormal model [8] seems apt: although the determination of the pair of thresholds on x for a given decision criterion is an important practical issue, it is of less theoretical interest than the observation that the ideal observer (in the simpler representation) acts by comparing a single decision variable, the likelihood ratio, with that single decision criterion.

Thus the determination of the particular data space thresholds (the “b” subfigures of this paper) for a given set of decision criteria, and of the relations among those thresholds, though daunting, should be regarded primarily as a computational issue, of considerably less importance than the behavior of the ideal observer (the “a” subfigures)—that is, the choice of decision criteria γ_{jij} —which leads to those sets of thresholds. In Appendix A, we prove that the maximum number of such intersection points, and thus thresholds on x , is indeed six as seen in Sec. 5.

In that vein, we note one qualitative feature of interest regarding the likelihood ratio curve not mentioned in the preceding sections. In Figs. 7(a) and 8(a), one can see a “hook” in the likelihood ratio curve as it approaches the (LR_1, LR_2) origin as x approaches $+\infty$ (i. e., beyond the largest inflection point). That is, as one traverses the curve in the direction of increasing x , the slope dLR_2/dLR_1 , which is decreasing between the maximum of LR_1 and the largest inflection point, begins increasing again, tending toward $+\infty$ as the curve approaches the origin. This of course makes sense given that $\sigma_1^2 < \sigma_2^2$, which means that as x tends to $\pm\infty$, $\pi(x|\pi_2)$ will dominate $\pi(x|\pi_1)$, and thus LR_2 will dominate LR_1 . Similarly, no such hook is seen in Figs. 5(a) or 6(a), because $\sigma_1^2 = \sigma_2^2$; given our convention in Sec. 4.2 that $\mu_2 < \mu_1$, it is straightforward to see that LR_2 dominates LR_1 as x approaches $-\infty$, but that LR_1 dominates LR_2 as x approaches $+\infty$. Thus the curve has infinite slope in the former case and zero slope in the latter near the origin, and no hook is present.

This proved confusing when graphs we produced during our preliminary investigations appeared to have no such hook, despite the variances of the underlying data distributions being different. Further investigation soon revealed that a hook was indeed present, but was simply too small to be easily seen in the initial version of the graphs in question. This situation can be compared with the well-known observation that a “hook” is always present in the ROC curve of an observer under the conventional binormal model as long as the variance of the two underlying data distributions are different; however, when the difference between the variances is small enough, the hook may become difficult to observe [8]. Although this analogy is appropriate in the sense that in both cases, the presence or absence of the hook in question is a result of variances of the underlying data distributions being unequal or equal, it is important to note that the “hook” under discussion here is a characteristic of the behavior of the ideal observer itself, and not of a comparison with non-ideal observer behavior as in the two-class ROC curve’s “hook.”

One of the numerous obstacles to the successful extension of ROC analysis to classification tasks with more than two classes has been the difficulty in developing a suitable performance metric for evaluating observer performance in such tasks. Since our discovery that the obvious generalization of the AUC performance metric, the hypervolume under the ROC surface, fails to be a useful performance metric for tasks with more than two classes [9], we have invested much effort in developing such a suitable performance metric [23, 24, 25]. Although we have argued that there are theoretical grounds for expecting our proposed performance metric to address some of the drawbacks inherent in the hypervolume performance metric, our proposed metric nevertheless relies on the assumption that the ROC surface is non-degenerate, i. e., possesses five degrees of freedom, over some non-empty region of support in the domain of the ROC space.

Recently, however, the conjecture has been raised that the ROC hypersurface of an arbitrary observer might fail to have five degrees of freedom, even if the observer in question is an ideal observer acting on multivariate data [26]. It is certainly the case that we have been heretofore unable to calculate explicitly ROC operating points for an arbitrary three-class ideal observer (i. e., one making decisions for underlying data with arbitrary pdfs), let alone tangent vectors along such an ROC surface. In this sense, the present work represents a significant step towards addressing the issue of the number of degrees of freedom of ideal observer ROC surfaces in general. Despite the outcome being in this case a negative one—in the sense that we have indeed shown that for univariate trinormal underlying data, the ideal observer’s ROC surface has fewer than five degrees of freedom—we believe the insights we have developed in the course of this work will allow us to make further progress in addressing the more general issue.

It is worth pointing out that an objection in the opposite direction could also be raised: the present study involved only univariate trinormal underlying data; is it not possible that univariate but not normally distributed data could be found such that the ideal observer’s ROC surface has five degrees of freedom? We consider this possibility unlikely. The arguments of Sec. 8.2, for example, are clearly independent of the underlying data pdfs involved; if there are fewer than five intersection points for a given choice of decision criteria, then no local manipulation of those decision criteria will be able to yield five linearly independent tangent vectors at that operating point. Similarly, if there are exactly five intersection points, then it should still be the case as discussed in Sec. 8.3 that three of them lie on one decision boundary, and one on each of the other two; although the details would be different from the trinormal case, the resulting Jacobian should still fail to be full rank for the reasons stated. Finally, it is at least in principle conceivable that a set of underlying data exists such that the resulting likelihood ratio curve is sufficiently convoluted to allow six or more intersection points with the decision boundary lines without the

intersection of the decision boundaries lying outside the likelihood ratio curve. However, in a practical situation, the available observational data would have to justify the additional complexity entailed in such a model as compared with the univariate trinormal model.

10. Conclusion

Although a fully general three-class extension to ROC analysis has yet to be developed, it is to be hoped that insights obtained in the consideration and analysis of constrained or simplified models will prove useful in the development of such fully general methodology. In the present paper, we have investigated the behavior of the three-class ideal observer under a particular constrained underlying data model, a univariate trinormal model. By categorizing the various possible situations given the values of the parameters (the means and variances) of this underlying data model, we were able to “catalog” the possible behaviors of the ideal observer’s decision variables under this model.

This in turn allows one to make statements regarding the behavior of the ideal observer itself; in particular, we showed that given the constraints on the possible shapes of the likelihood ratio curve, the decision boundaries used by the ideal observer can intersect that curve in at most six places. **The ideal observer nature of this model was demonstrated by comparison with a non-ideal decision rule investigated by other researchers; as expected, the ideal observer attains a lower cost (greater expected utility) for a given set of decision criteria and data model parameters.**

We anticipate that these results will prove of use in the further development of a fully general extension of ROC analysis to classification tasks with three or more classes. In particular, extending the underlying data model from the univariate trinormal one considered here to a bivariate or multivariate trinormal model should provide sufficient flexibility to model the behavior of observers, such as algorithmic or human observers, more relevant to typical medical decision-making tasks, including the medical imaging tasks mentioned in the introduction.

We have also shown that, for the underlying data model considered here, the ROC surface must always be degenerate; that is, one always has fewer than five linearly independent tangent vectors to the ROC surface, and thus the ROC surface has fewer than five degrees of freedom, at any operating point. Although this implies that the development of a suitable performance metric for classification tasks with three or more classes may prove an even greater challenge than we had previously demonstrated, we believe the insights we have gained through the work reported here will prove invaluable in addressing this more general issue.

Acknowledgments

The authors would like to thank Craig Abbey, whose questions and comments on this subject inspired us to investigate it more thoroughly than we had previously.

This work was supported by grant R01 EB000863 from the National Institutes of Health (Kevin Berbaum, the University of Iowa, principal investigator) through a subcontract from the University of Iowa to the University of Chicago (Charles Metz, subcontract principal investigator).

References

1. Egan, JP. Signal Detection Theory and ROC Analysis. Academic Press; New York: 1975.
2. Metz CE. Basic principles of ROC analysis. Seminars in Nuclear Medicine. 1978; VIII:283–298. [PubMed: 112681]

3. Chan, H-P.; Sahiner, B.; Hadjiiski, LM.; Petrick, N.; Zhou, C. Design of three-class classifiers in computer-aided diagnosis: Monte Carlo simulation study. In: Sonka, Milan; Michael Fitzpatrick, J., editors. Proc SPIE Vol 5032 Medical Imaging 2003: Image Processing. SPIE; Bellingham, WA: p. 567-578.
4. Edwards DC, Lan L, Metz CE, Giger ML, Nishikawa RM. Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions. *Med Phys*. 2004; 31:81–90. [PubMed: 14761024]
5. He X, Metz CE, Tsui BMW, Links JM, Frey EC. Three-class ROC analysis — A decision theoretic approach under the ideal observer framework. *IEEE Trans Med Imag*. 2006; 25:571–581.
6. Van Trees, HL. Detection, Estimation and Modulation Theory: Part I. John Wiley & Sons; New York: 1968.
7. Edwards DC, Metz CE, Kupinski MA. Ideal observers and optimal ROC hypersurfaces in N -class classification. *IEEE Trans Med Imag*. 2004; 23:891–895.
8. Metz CE, Pan X. 'Proper' binormal ROC curves: Theory and maximum-likelihood estimation. *J Math Psychol*. 1999; 43:1–33. [PubMed: 10069933]
9. Edwards DC, Metz CE, Nishikawa RM. The hypervolume under the ROC hypersurface of 'near-guessing' and 'near-perfect' observers in N -class classification tasks. *IEEE Trans Med Imag*. 2005; 24:293–299.
10. Edwards DC, Metz CE. Restrictions on the three-class ideal observer's decision boundary lines. *IEEE Trans Med Imag*. 2005; 24:1566–1573.
11. Mossman D. Three-way ROCs. *Med Decis Making*. 1999; 19:78–89. [PubMed: 9917023]
12. Dreiseitl S, Ohno-Machado L, Binder M. Comparing three-class diagnostic tests by three-way ROC analysis. *Med Decis Making*. 2000; 20:323–331. [PubMed: 10929855]
13. Scurfield BK. Multiple-event forced-choice tasks in the theory of signal detectability. *J Math Psychol*. 1996; 40:253–269. [PubMed: 8979976]
14. Scurfield BK. Generalization of the theory of signal detectability to n -event m -dimensional forced-choice tasks. *J Math Psychol*. 1998; 42:5–31. [PubMed: 9606159]
15. Sahiner, B.; Chan, H-P.; Hadjiiski, LM. Performance analysis of 3-class classifiers: Properties of the 3D ROC surface and the normalized volume under the surface. In: Jiang, Yulei; Eckstein, Miguel P., editors. Proc SPIE Vol 6146 Medical Imaging 2006: Image Perception, Observer Performance, and Technology Assessment. SPIE; Bellingham, WA: p. 61460C1-61460C7.
16. He X, Frey EC. Three-class ROC analysis — The equal error utility assumption and the optimality of three-class ROC surface using the ideal observer. *IEEE Trans Med Imag*. 2006; 25:979–986.
17. Schubert CM, Thorsen SN, Oxley ME. The ROC manifold for classification systems. *Pattern Recognit*. 2011; 44:350–362.
18. He X, Gallas BD, Frey EC. Three-class ROC analysis — toward a general decision theoretic solution. *IEEE Trans Med Imag*. 2010; 29:206–215.
19. Metz CE, Herman BA, Shen J-H. Maximum-likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statist Med*. 1998; 17:1033–1053.
20. Kijewski MF, Swensson RG, Judy PF. Analysis of rating data from multiple-alternative tasks. *J Math Psychol*. 1989; 33:428–451.
21. Grossman, SI. Multivariable Calculus, Linear Algebra, and Differential Equations. 2. Harcourt Brace Jovanovich; San Diego, CA: 1986.
22. Edwards DC, Metz CE. Optimization of restricted ROC surfaces in three-class classification tasks. *IEEE Trans Med Imag*. 2007; 26:1345–1356.
23. Edwards, DC.; Metz, CE. A utility-based performance metric for ROC analysis of N -class classification tasks. In: Jiang, Yulei; Sahiner, Berkman, editors. Proc SPIE Vol 6515 Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment. SPIE; Bellingham, WA: p. 6515031-65150310.
24. Edwards, DC.; Metz, CE. Optimality of a utility-based performance metric for ROC analysis. In: Sahiner, Berkman; Manning, David J., editors. Proc SPIE Vol 6917 Medical Imaging 2008: Image Perception, Observer Performance, and Technology Assessment. SPIE; Bellingham, WA: p. 69170F1-68170F6.

25. Edwards, DC.; Metz, CE. Comparing the performance of two observers using a novel utility-based performance metric for ROC analysis. In: Sahiner, Berkman; Manning, David J., editors. Proc SPIE Vol 7263 Medical Imaging 2009: Image Perception, Observer Performance, and Technology Assessment. SPIE; Bellingham, WA: p. 72630W1-72630W7.
26. He, X.; Frey, E. Revisiting the general three-class ROC theory: A decision theoretic approach. Medical Image Perception Conference XIII; Santa Barbara, CA. 2009.

Appendix A. Curvature and Inflection Points

As was mentioned in Sec. 1, one obstacle to the development of a fully general extension of ROC analysis to the three-class classification task has been the difficulty in deriving explicit expressions for the various conditional error probabilities as functions of the decision criteria γ_{ij} of Eqs. 3–5. This difficulty is evident even when one adopts such a simple model as the univariate trinormal model investigated here; although one can certainly work out the decision boundaries for a *given* set of decision criteria, and from that the appropriate set of thresholds on x and the resulting class assignments, there seems to be no obvious way to generalize from one such choice of decision criteria to the next. Indeed, the even-numbered figures show that even the ordering (as a function of increasing x) of the class assignments themselves can be different from one such choice to the next.

That being said, are there any useful generalizations that can be made, based on the categorization, developed in the preceding sections, of the possible shapes of the likelihood ratio curve? Can we say with any certainty, for example, that the maximum number of decision thresholds on x is six, the number seen in the figures of Sec. 5? In fact, this can be shown to be the case, based on the behavior of the curvature polynomial κ_p which, as has been demonstrated in the preceding sections, determines the properties of the inflection points of the likelihood ratio curve.

More explicitly, we wish to characterize the number of times the likelihood ratio curve can be intersected by any one of the decision boundary line segments. This is a special case of the more general problem of determining how many times a given function $f(x)$ can be intersected by a line $l(x) = mx + b$, or equivalently, how many zeros the function $g(x) \equiv f(x) - l(x)$ can have. Since the second derivative $f''(x)$ vanishes everywhere, we clearly have $g''(x) = f''(x)$. Now if $g(x)$ has n zeros in a given interval, its first derivative $g'(x)$ will in general have at least $n - 1$ zeros: one between each pair of zeros of $g(x)$, and possibly others if a zero of $g(x)$ also corresponds to a zero of $g'(x)$. (Exceptions to this would basically correspond to either discontinuities in $f(x)$, or regions in which $f(x)$ is constant, and it should be clear from the form of Eqs. 12–15 that such special cases do not apply to our situation of interest.) Similarly, we would expect to find at least $n - 2$ zeros of $g''(x)$, and thus of $f''(x)$, in the same interval. Working backwards, we can conclude that if $f''(x)$ has a given number of zeros in a given interval, then $g(x)$ has at most two more than this number of zeros in the same interval. Since the number of zeros of the second derivative (or of the curvature, per Eqs. 15 and 16) sets an upper bound on the number of intersections of the likelihood ratio curve with a decision boundary line segment, it makes sense to consider the most complicated situation first: the last example of Sec. 5, in which the likelihood ratio curve forms a closed curve with three distinct inflection points.

Because $LR_2(x)$ takes two values for each value of $LR_1(x)$ in the case under consideration, it will prove convenient to break the curve into two parts, one for $x \leq \mu_1/(1-\sigma_1^2)$, and one for $x > \mu_1/(1-\sigma_1^2)$ (recall that this is the value at which LR_1 takes its maximum), so that LR_2 is a (single-valued) function of LR_1 on each segment. We will refer to the former as the “upper” part and the latter as the “lower” part (see Fig. 8(a)). We showed in Sec. 5 that there are two inflection points on the upper part, specifically in the interval $\mu_2/(1-\sigma_2^2) < x < \mu_1/(1-\sigma_1^2)$; by

the arguments of the preceding paragraph, this implies that there can be at most four intersections of a decision boundary line segment with this part of the likelihood ratio curve.

Supposing a particular decision boundary line segment does achieve such a maximum of four intersections with the upper part of a given likelihood ratio curve, we can then conclude that there must be at least three points where the slope of this line segment is equal to the slope of the likelihood ratio curve, i. e., where $g'(x) = 0$. But then an inflection point will occur between each such pair of points, and since the inflection points occur at values of x greater than $\mu_2/(1 - \sigma^2)$, at least two of these points must occur for $x > \mu_2/(1 - \sigma^2)$. Furthermore, since the slope of the likelihood ratio curve is negative in the interval $\mu_2/(1 - \sigma^2) < x < \mu_1/(1 - \sigma_1^2)$ (see Eq. 14), this implies that the slope of the decision boundary line segment in question is also negative. (This argument holds true even if the decision boundary line segment intersects the upper part of the likelihood ratio curve in three places; i. e., the slope of that decision boundary must be negative.)

This immediately rules out the possibility that the “1-vs.-2” decision boundary line segment could intersect the upper part of the likelihood ratio curve in four places, because that boundary must have positive slope [10]. It also implies that the LR_2 -intercept of the decision boundary line segment is positive (otherwise, the decision boundary would not traverse the first quadrant, in which the likelihood ratio curve lies, at all). The “2-vs.-3” decision boundary of Fig. 8(a) is a suitable example of this situation; furthermore, inspection of that figure demonstrates that the intersection point of the three decision boundary lines (in the case where one of them intersects the upper part of the likelihood ratio curve four times) must lie *outside* the likelihood ratio curve. This can be seen by traversing the decision boundary starting at its LR_2 -intercept, outside of the likelihood ratio curve; after intersecting the curve once, the decision boundary must be inside the curve, and after the fourth such intersection, it must again be outside.

If the “1-vs.-3” decision boundary line segment intersects the upper part of the likelihood ratio curve in four places, then the constraint that the intersection point of the three decision boundaries must lie outside the likelihood ratio curve means that this decision boundary intersection point must lie to the left of the likelihood ratio curve (very close to the LR_2 axis), or else in the second quadrant (see Fig. 2 of Ref. [10]). This in turn implies that the “2-vs.-3” line (even if defined, which it is not in the latter case) cannot intersect the likelihood ratio curve at all, and the “1-vs.-2” line can only intersect the curve in two places, both on the upper part, for a total of six intersections.

On the other hand, if it is the “2-vs.-3” decision boundary which intersects the upper part of the likelihood ratio curve in four places, the constraint that the intersection point of the three decision boundaries must lie outside the likelihood ratio curve means that this decision boundary intersection point must lie to the right of the likelihood ratio curve, and thus the “1-vs.-2” decision boundary cannot intersect the curve at all, while the “1-vs.-3” decision boundary can only intersect the lower part (essentially as depicted in Fig. 8(a)). By the above arguments, a decision boundary line segment can intersect the lower part of the likelihood ratio curve in at most three places, because it possesses a single inflection point (for some $x > \mu_1/(1 - \sigma_1^2)$). There will also be up to two points where the slope of the decision boundary equals the slope of the likelihood ratio curve; since the slope of the likelihood ratio curve is positive in the interval $x > \mu_1/(1 - \sigma_1^2)$ (see Eq. 14), so is the slope of the decision boundary line segment. We also claim that the LR_2 -intercept of this decision boundary is non-negative: if it were negative, then $g(x)$ would decrease from a positive value to 0, then to a negative local minimum, and then increase back to a positive local maximum; that is, $g''(x)$ would go from positive to negative. But our curve behaves the opposite way:

$d^2\text{LR}_2/d\text{LR}_1^2$ goes from negative (at $\text{LR}_1 = 0$) to positive as LR_1 increases, as seen in Figs. 7(a) and 8(a). (The curvature polynomial κ_p of Eq. 17 is positive as x tends to $+\infty$, but in this portion of the curve, increasing x corresponds to decreasing LR_1 .)

But a line with positive slope and non-negative y -intercept must have a non-positive x -intercept; this means that the “1-vs.-3” line cannot intersect the lower part of the likelihood ratio curve in three places, because that line must have a positive LR_1 -intercept [10]. Therefore, if the “2-vs.-3” decision boundary line segment intersects the upper part of the likelihood ratio curve in four places, the “1-vs.-3” line can intersect the lower part in at most two places, again for a maximum total of six intersections (and thus decision thresholds in the underlying data space). From the results of Secs. 3 to 4.2, it should be clear that this maximum cannot be exceeded in situations where the likelihood ratio curve has been shown to have fewer than three inflection points (thus reducing the number of possible intersections of line segments with such curves).

Appendix B. Five Intersections

We consider in detail here the example given in Sec. 8.3 and illustrated in Fig. 12, for which $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -2/3$, and $\sigma_2^2 = 1/2$; and for which the decision criteria of Eqs. 3–5 are chosen such that the decision boundaries are segments of the lines $\text{LR}_2 = (1/4)\text{LR}_1 - 1/16$, $\text{LR}_2 = (3/10)\text{LR}_1 - 3/40$, and $\text{LR}_2 = -(4/9)\text{LR}_1 + 31/36$, respectively.

For this choice of underlying data parameters and decision criteria, the first matrix on the right side of Eq. 61 can be shown to be

$$J_{x \rightarrow p} = \begin{bmatrix} 0 & 0 & 0 & \frac{-1}{\sigma_2} \varphi\left(\frac{x_4 - \mu_2}{\sigma_2}\right) & \frac{1}{\sigma_2} \varphi\left(\frac{x_5 - \mu_2}{\sigma_2}\right) \\ 0 & 0 & 0 & -\varphi(x_4) & \varphi(x_5) \\ \frac{-1}{\sigma_1} \varphi\left(\frac{x_1 - \mu_1}{\sigma_1}\right) & \frac{1}{\sigma_1} \varphi\left(\frac{x_2 - \mu_1}{\sigma_1}\right) & \frac{-1}{\sigma_1} \varphi\left(\frac{x_3 - \mu_1}{\sigma_1}\right) & \frac{1}{\sigma_1} \varphi\left(\frac{x_4 - \mu_1}{\sigma_1}\right) & 0 \\ -\varphi(x_1) & \varphi(x_2) & -\varphi(x_3) & \varphi(x_4) & 0 \\ \frac{1}{\sigma_1} \varphi\left(\frac{x_1 - \mu_1}{\sigma_1}\right) & \frac{-1}{\sigma_1} \varphi\left(\frac{x_2 - \mu_1}{\sigma_1}\right) & \frac{1}{\sigma_1} \varphi\left(\frac{x_3 - \mu_1}{\sigma_1}\right) & 0 & \frac{-1}{\sigma_1} \varphi\left(\frac{x_5 - \mu_1}{\sigma_1}\right) \end{bmatrix} \quad (\text{B.1})$$

where

$$\varphi(x) \equiv \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} \quad (\text{B.2})$$

is the standard normal pdf (the derivative of $\Phi(\cdot)$). From the location of the zeros in this matrix, and the fact that $\varphi(\cdot)$ is a non-linear function, $J_{x \rightarrow p}$ can be shown to be full rank.

Evaluating the second matrix on the right side of Eq. 61, or $J_{m \rightarrow x}$, is somewhat more involved. In principle, it is straightforward enough to imagine changing one of the decision boundary parameters, b for instance, by some small amount; the resulting value $b + \Delta b$ will result in a given intersection point, and hence the corresponding x_j , being altered in turn to $x_j + \Delta x_j$; taking the limit as Δb tends to zero will then give x_j/b . For notational convenience, we will define $\text{LR}_{j1} \equiv \text{LR}_1(x_j)$ and $\text{LR}_{j2} \equiv \text{LR}_2(x_j)$, the coordinates of the intersection point corresponding to x_j ; and

$$\frac{\partial \text{LR}_{jk}}{\partial x_j} \equiv \left. \frac{\partial \text{LR}_k}{\partial x} \right|_{x_j}. \quad (\text{B.3})$$

Now suppose that x_j corresponds to the point in LR space at which a decision boundary line with slope m_j intersects the likelihood ratio curve. We now consider changing b to $b + \Delta b$ while leaving the other decision boundary parameters (a , m_{12} , m_{13} , m_{23}) unchanged. In particular, m_j is unchanged (whichever decision boundary it may correspond to); since we originally have

$$m_j = \frac{LR_{j2} - b}{LR_{j1} - a}, \quad (B.4)$$

the result after changing b is

$$\begin{aligned} m_j &= \frac{LR_{j2} + \frac{\partial LR_{j2}}{\partial x_j} \Delta x_j - (b + \Delta b)}{LR_{j1} + \frac{\partial LR_{j1}}{\partial x_j} \Delta x_j - a} \\ &= \frac{m_j(LR_{j1} - a) + \frac{\partial LR_{j2}}{\partial x_j} \Delta x_j - \Delta b}{LR_{j1} + \frac{\partial LR_{j1}}{\partial x_j} \Delta x_j - a} \quad (B.5) \\ m_j \frac{\partial LR_{j1}}{\partial x_j} \Delta x_j &= \frac{\partial LR_{j2}}{\partial x_j} \Delta x_j - \Delta b \\ \frac{\partial x_j}{\partial b} &= \frac{1}{\frac{\partial LR_{j2}}{\partial x_j} - m_j \frac{\partial LR_{j1}}{\partial x_j}}, \end{aligned}$$

where in the last step we have taken the limit as Δb tends to zero.

Similarly, we can change a to $a + \Delta a$ while leaving b and all three slopes unchanged, to obtain

$$\begin{aligned} m_j &= \frac{LR_{j2} + \frac{\partial LR_{j2}}{\partial x_j} \Delta x_j - b}{LR_{j1} + \frac{\partial LR_{j1}}{\partial x_j} \Delta x_j - (a + \Delta a)} \\ &= \frac{m_j(LR_{j1} - a) + \frac{\partial LR_{j2}}{\partial x_j} \Delta x_j}{LR_{j1} - a + \frac{\partial LR_{j1}}{\partial x_j} \Delta x_j - \Delta a} \quad (B.6) \\ m_j \left(\frac{\partial LR_{j1}}{\partial x_j} \Delta x_j - \Delta a \right) &= \frac{\partial LR_{j2}}{\partial x_j} \Delta x_j \\ \frac{\partial x_j}{\partial a} &= \frac{-m_j}{\frac{\partial LR_{j2}}{\partial x_j} - m_j \frac{\partial LR_{j1}}{\partial x_j}}, \end{aligned}$$

once again having, in the last step, taken the limit as Δa tends to zero. Comparing Eqs. B.5 and B.6 yields the relation

$$\frac{\partial x_j}{\partial a} = -m_j \frac{\partial x_j}{\partial b}. \quad (B.7)$$

Finally, we can consider changing m_j to $m_j + \Delta m_j$ while leaving a , b , and the other two slopes unchanged:

$$\begin{aligned}
m_j + \Delta m_j &= \frac{\text{LR}_{j2} + \frac{\partial \text{LR}_{j2}}{\partial x_j} \Delta x_j - b}{\text{LR}_{j1} + \frac{\partial \text{LR}_{j1}}{\partial x_j} \Delta x_j - a} \\
&= \frac{m_j(\text{LR}_{j1} - a) + \frac{\partial \text{LR}_{j2}}{\partial x_j} \Delta x_j}{\text{LR}_{j1} - a + \frac{\partial \text{LR}_{j1}}{\partial x_j} \Delta x_j} \quad (\text{B.8}) \\
\Delta m_j(\text{LR}_{j1} - a) + (m_j + \Delta m_j) \frac{\partial \text{LR}_{j1}}{\partial x_j} \Delta x_j &= \frac{\partial \text{LR}_{j2}}{\partial x_j} \Delta x_j \\
\frac{\partial x_j}{\partial m_j} &= \frac{\text{LR}_{j1} - a}{\frac{\partial \text{LR}_{j2}}{\partial x_j} - m_j \frac{\partial \text{LR}_{j1}}{\partial x_j}},
\end{aligned}$$

where in the last step we have retained only linear terms in Δm_j and taken the limit as Δm_j tends to zero. Comparing Eqs. B.5 and B.8 yields the relation

$$\frac{\partial x_j}{\partial m_j} = (\text{LR}_{j1} - a) \frac{\partial x_j}{\partial b}. \quad (\text{B.9})$$

By applying Eqs. B.7 and B.9 we can, again for the specific example of Fig. 12, write the second matrix on the right side of Eq. 61 as

$$J_{m \rightarrow x} = \begin{bmatrix} -m_{23} \frac{\partial x_1}{\partial b} & \frac{\partial x_1}{\partial b} & 0 & 0 & (\text{LR}_{11} - a) \frac{\partial x_1}{\partial b} \\ -m_{23} \frac{\partial x_2}{\partial b} & \frac{\partial x_2}{\partial b} & 0 & 0 & (\text{LR}_{21} - a) \frac{\partial x_2}{\partial b} \\ -m_{23} \frac{\partial x_3}{\partial b} & \frac{\partial x_3}{\partial b} & 0 & 0 & (\text{LR}_{31} - a) \frac{\partial x_3}{\partial b} \\ -m_{12} \frac{\partial x_4}{\partial b} & \frac{\partial x_4}{\partial b} & (\text{LR}_{41} - a) \frac{\partial x_4}{\partial b} & 0 & 0 \\ -m_{13} \frac{\partial x_5}{\partial b} & \frac{\partial x_5}{\partial b} & 0 & (\text{LR}_{51} - a) \frac{\partial x_5}{\partial b} & 0 \end{bmatrix}. \quad (\text{B.10})$$

By exploiting the symmetry properties of determinants [21], the determinant of the above matrix can be shown to be $(\text{LR}_{51} - a) \frac{\partial x_5}{\partial b}$ times $(\text{LR}_{41} - a) \frac{\partial x_4}{\partial b}$ times the determinant of the 3×3 matrix

$$\begin{bmatrix} -m_{23} \frac{\partial x_1}{\partial b} & \frac{\partial x_1}{\partial b} & (\text{LR}_{11} - a) \frac{\partial x_1}{\partial b} \\ -m_{23} \frac{\partial x_2}{\partial b} & \frac{\partial x_2}{\partial b} & (\text{LR}_{21} - a) \frac{\partial x_2}{\partial b} \\ -m_{23} \frac{\partial x_3}{\partial b} & \frac{\partial x_3}{\partial b} & (\text{LR}_{31} - a) \frac{\partial x_3}{\partial b} \end{bmatrix}, \quad (\text{B.11})$$

which in turn is zero, because the first column of this sub-matrix is a constant multiple of the second. The Jacobian $J_{m \rightarrow x}$, and thus the Jacobian $J_{m \rightarrow P}$ of Eq. 61, cannot be full rank.

- Extension of ROC analysis to tasks with more than two classes is difficult
- We examine the three-class ideal observer for univariate trinormal data
- This choice allows explicit analytic calculation of ideal observer behavior
- We hope the insights gained will enable us to analyze more realistic data models

\$watermark-text

\$watermark-text

\$watermark-text

Figure 1.

The likelihood ratio curve when the underlying data are drawn from unit-variance normal distributions with parameters $\mu_1 = 1$, $\mu_2 = -1$, and $\mu_3 = 0$, and where the decision boundary line segments are shown with dashed lines. An expression such as " π_1 " indicates a region where a particular decision is made.

Figure 2.

The data space thresholds, shown as dashed lines, corresponding to the decision boundaries of Fig. 1, plotted against the underlying data pdfs ($\mu_2 = -1$, $\mu_3 = 0$, $\mu_1 = 1$, all variances equal to one). An expression such as " π_1 " indicates a region where a particular decision is made.

Figure 3.

The likelihood ratio curve, (a), when the underlying data are drawn from unit-variance normal distributions with parameters $\mu_1 = 1$, $\mu_2 = -1$, and $\mu_3 = 0$, and where the decision boundary line segments are shown with dashed lines. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as " π_1 " indicates a region where a particular decision is made.

Figure 4.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = -1/3$, $\sigma_1^2 = 1/3$, $\mu_2 = -1/2$, $\sigma_2^2 = 1$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. Points where the curvature vanishes are shown with circles. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as “ π_1 ” indicates a region where a particular decision is made.

Figure 5.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = 1/2$, $\sigma_1^2 = \sigma_2^2 = 1/2$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. The point where the curvature vanishes is shown with a circle. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as " π_1 " indicates a region where a particular decision is made.

Figure 6.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = 1$, $\mu_2 = -1/2$, $\sigma_1^2 = \sigma_2^2 = 1/2$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. The points where the curvature vanishes are shown with circles. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as “ π_1 ” indicates a region where a particular decision is made.

Figure 7.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -1/4$, $\sigma_2^2 = 1/2$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. The point where the curvature vanishes is shown with a circle. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as “ π_1 ” indicates a region where a particular decision is made.

Figure 8.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -2/3$, $\sigma_2^2 = 1/2$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. The points where the curvature vanishes are shown with circles. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as “ π_1 ” indicates a region where a particular decision is made.

Figure 9.

The portion of the likelihood ratio curve of Fig. 8(a) near the maximum of LR_2 . The first point along that curve where the curvature vanishes is visible here, and is shown with a circle.

Figure 10.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = 1/10$, $\sigma_1^2 = 1/10$, $\mu_2 = 1/18$, $\sigma_2^2 = 1/2$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. The point where the curvature vanishes is shown with a circle. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as " π_1 " indicates a region where a particular decision is made.

Figure 11.

The likelihood ratio curve, (a), when the underlying data are drawn from normal distributions with parameters $\mu_1 = \sqrt{2}$, $\sigma_1^2 = 1/16$, $\mu_2 = \sqrt{2}/2$, and $\sigma_2^2 = 1/4$; and where the decision boundary line segments are shown with dashed lines. Points where the curvature vanishes are shown with circles. The data space thresholds, shown as dashed lines, corresponding to those decision boundaries, are plotted against the underlying data pdfs in (b). An expression such as “ π_1 ” indicates a region where a particular decision is made.

Figure 12.

The likelihood ratio curve when the underlying data are drawn from normal distributions with parameters $\mu_1 = 1/8$, $\sigma_1^2 = 1/8$, $\mu_2 = -2/3$, $\sigma_2^2 = 1/2$, $\mu_3 = 0$, and $\sigma_3^2 = 1$; and where the decision boundary line segments are shown with dashed lines. An expression such as " π_1 " indicates a region where a particular decision is made.