Optimization of an ROC hypersurface constructed only from an observer's within-class sensitivities

Darrin C. Edwards* and Charles E. Metz

Department of Radiology, The University of Chicago, Chicago, IL 60637

ABSTRACT

We have shown in previous work that an ideal observer in a classification task with N classes achieves the optimal receiver operating characteristic (ROC) hypersurface in a Neyman-Pearson sense. That is, the hypersurface obtained by taking one of the ideal observer's misclassification probabilities as a function of the other $N^2 - N - 1$ misclassification probabilities is never above the corresponding hypersurface obtained by any other observer. Due to the inherent complexity of evaluating observer performance in an N-class classification task with $N > 2$, some researchers have suggested a generally incomplete but more tractable evaluation in terms of a hypersurface plotting only the N "sensitivities" (the probabilities of correctly classifying observations in the various classes). An N-class observer generally has up to $N^2 - N - 1$ degrees of freedom, so a given sensitivity will still vary when the other $N - 1$ are held fixed; a well-defined hypersurface can be constructed by considering only the maximum possible value of one sensitivity for each achievable value of the other $N - 1$. We show that optimal performance in terms of this generally incomplete performance descriptor, in a Neyman-Pearson sense, is still achieved by the N-class ideal observer. That is, the hypersurface obtained by taking the maximal value of one of the ideal observer's correct classification probabilities as a function of the other $N - 1$ is never below the corresponding hypersurface obtained by any other observer.

Keywords: ROC analysis, three-class classification, ideal observer decision rules

1. INTRODUCTION

We are attempting to extend the well-known observer performance evaluation methodology of receiver operating characteristic (ROC) analysis[1,2] to classification tasks with three 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]

Unfortunately, a fully general but tractable extension of ROC analysis has yet to be developed. It is known that the performance of an observer in a classification task with N classes ($N \geq 2$) can be completely described by a set of $N^2 - N$ conditional error probabilities,[4,5] and that the performance of the ideal observer (that which minimizes Bayes risk[4]) is completely characterized by an ROC hypersurface in which these conditional error probabilities depend on a set of $N^2 - N - 1$ decision criteria.[5] Although analytic expressions for the ideal observer's conditional error probabilities given reasonable models for the underlying observational date have been worked out in the two-class case,[6] 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.[7] More recently, we showed that complicated constraining relationships exist among the decision criteria themselves for the ideal observer.[8] 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.[6]

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 required to describe the observer's performance, either by implicitly leaving the remaining degrees of freedom out of the analysis, or

*Correspondence: E-mail: d-edwards@uchicago.edu; Telephone: 773 834 5094; Fax: 773 702 0371

by explicitly imposing restrictions on the form of the observer's decision rule or on the set of decision criteria used by the observer.

Scurfield evaluated an observer which used a specified decision rule with only two degrees of freedom (as opposed to the five decision criteria used by the general three-class ideal observer) by plotting a set of six (two-dimensional) surfaces in three-dimensional ROC spaces.[9] Mossman proposed plotting the surface formed only from the set of three "sensitivities" (conditional probabilities of correctly classifying observations) for an observer with two degrees of freedom, and applied this method to an observer with a specified decision rule.[10] Chan *et al.* began with an ideal observer model, and reduced the number of decision criteria from five to two by imposing explicit assumptions on the observer's decision utilities; the observer's performance was then plotted as a surface in a three-dimensional ROC space, the axes of which are the probabilities of deciding an observation to be malignant conditional on each of the three actual class memberships.[11] He *et al.* investigated an ideal observer model in which the decision rule is restricted to a form similar to that proposed by Scurfield; the nature of the restrictions is such that performance evaluation in terms of only the three sensitivities provides a complete description of this observer's performance.[12]

A common theme among these remarkably diverse methods is the idea of an "ROC surface," *i.e.*, a surface with two degrees of freedom in a three-dimensional ROC space. An appealing feature of such a construct is its visualizability: it can be plotted as readily as any elevation map, for example, in stark contrast to the fully general three-class classification task involving a hypersurface with five degrees of freedom in a six-dimensional ROC space as mentioned above. While it is true that not all of the proposed methods described in the preceding paragraph involve a "sensitivity" ROC surface, the general division of an N-class observer's conditional decision probabilities into a set of N sensitivities and a set of $N^2 - N$ misclassification rates[5] makes this particular construct a natural candidate for further analysis.

On the other hand, it can be argued that measurement of performance in terms of only N conditional classification rates must be an incomplete description of observer performance in a classification task with more than two classes, which requires $N^2 - N$ such classification rates as stated above. Acknowledging this incompleteness, we would like to ask whether there is any sense in which such an incomplete performance metric is at least well-defined. In particular, is there any observer decision rule, dependent on only $N - 1$ (rather than $N^2 - N - 1$) decision criteria, for which the observer's sensitivity ROC hypersurface is always above the corresponding hypersurface obtained for any other observer? If so, what form does this decision rule take?

In the next section, we show that the three-class observer which optimizes performance only in terms of the sensitivity surface is in fact the three-class ideal observer, with its decision utilities constrained in a particular way (reducing its degrees of freedom from five to two as necessary). Additionally, the form of the constraints on the ideal observer's behavior are identical to those considered by He *et al.*.[12] In Sec. 3, we extend this result to the general case of an N-class observer, showing that the observer which attains the optimal sensitivity hypersurface is a restricted form of the N-class ideal observer, and in particular a straightforward generalization of the three-class observer considered by He *et al.*[12] to N classes. Our conclusions are stated in Sec. 4.

2. THREE-CLASS OBSERVERS

We have shown[5] that the N-class ideal observer — that observer which minimizes Bayes risk — also achieves optimal performance in an ROC sense, by virtue of satisfying the Neyman-Pearson criterion. This was the same argument used by Van Trees[4] to show that the two-class ideal observer achieves the optimal ROC curve for a given two-class classification task. This technique of satisfying the Neyman-Pearson criterion, essentially an application of an integral form of the method of Lagrange multipliers,[13] is straightforward (conceptually, if not notationally) and flexible, and we apply it in this section to answer the question of what observer optimizes performance in terms of only the three observer sensitivities.

We denote by P_{ij} the conditional probability of a given observer deciding an observation is drawn from the ith class, conditional on it actually being drawn from the jth class. Thus, the three sensitivities are P_{11}, P_{22}, and P_{33}. Decisions are assumed to be made based on statistically variable observational data; in particular,

$$P_{ij} \equiv \int_{Z_i} p(\vec{x}|\pi_j)\, d^m\vec{x}, \tag{1}$$

where Z_i is the region for which observations $\vec{\mathbf{x}}$ (of dimension m) are decided to belong to the class labeled π_i ($1 \le i \le 3$).

Without loss of generality, we seek to maximize P_{33} subject to the constraints $P_{11} = \alpha_{11}$ and $P_{22} = \alpha_{22}$ where $0 \le \alpha_{11} \le 1$ and $0 \le \alpha_{22} \le 1$. We define the function

$$F \equiv P_{33} + \lambda_{11}(P_{11} - \alpha_{11}), +\lambda_{22}(P_{22} - \alpha_{22}) \tag{2}$$

where λ_{11} and λ_{22} are the so-called Lagrange multipliers. Note that if we can find a decision rule (a partitioning of the domain of $\vec{\mathbf{x}}$ into Z_1, Z_2, and Z_3) that maximizes F for arbitrary values of λ_{11} and λ_{22}, then this will be equivalent to maximizing P_{33} at the point at which the constrain equations are satisfied (*i.e.*, at the point $P_{11} = \alpha_{11}, P_{22} = \alpha_{22}$).

We first rewrite F by applying rules for conditional probabilities:

$$\begin{aligned} F &= -\lambda_{11}\alpha_{11} - \lambda_{22}\alpha_{22} + (1 - P_{13} - P_{23}) + \lambda_{11}(1 - P_{21} - P_{31}) + \lambda_{22}(1 - P_{12} - P_{32}) \\ &= 1 + \lambda_{11}(1 - \alpha_{11}) + \lambda_{22}(1 - \alpha_{22}) - \{\lambda_{22}P_{12} + P_{13} + \lambda_{11}P_{21} + P_{23} + \lambda_{11}P_{31} + \lambda_{22}P_{32}\} \\ &= 1 + \lambda_{11}(1 - \alpha_{11}) + \lambda_{22}(1 - \alpha_{22}) - \left\{ \int_{Z_1} \lambda_{22}p(\vec{x}|\pi_2) + p(\vec{x}|\pi_3)\, d^m\vec{x} \right. \\ &\quad \left. + \int_{Z_2} \lambda_{11}p(\vec{x}|\pi_1) + p(\vec{x}|\pi_3)\, d^m\vec{x} + \int_{Z_3} \lambda_{11}p(\vec{x}|\pi_1) + \lambda_{22}p(\vec{x}|\pi_2)\, d^m\vec{x} \right\}. \end{aligned} \tag{3}$$

For a given set of values of the parameters λ_{11} and λ_{22}, F is maximized when the quantity in braces is minimized. This quantity, in turn, can be minimized by assigning a given $\vec{x}$ to the region Z_i such that the ith integrand (from among the integrals in braces in Eq. 3) is minimized. (Situations in which two or more of the integrands yield the same minimal value for a given $\vec{x}$ can be decided in an arbitrary but consistent fashion.)

That is,

$$\text{decide} \quad \pi_1 \quad \text{iff} \quad \lambda_{22}p(\vec{x}|\pi_2) < \lambda_{11}p(\vec{x}|\pi_1) \quad \text{and} \quad p(\vec{x}|\pi_3) < \lambda_{11}p(\vec{x}|\pi_1) \tag{4}$$

$$\text{decide} \quad \pi_2 \quad \text{iff} \quad \lambda_{11}p(\vec{x}|\pi_1) \le \lambda_{22}p(\vec{x}|\pi_2) \quad \text{and} \quad p(\vec{x}|\pi_3) < \lambda_{22}p(\vec{x}|\pi_2) \tag{5}$$

$$\text{decide} \quad \pi_3 \quad \text{iff} \quad \lambda_{11}p(\vec{x}|\pi_1) \le p(\vec{x}|\pi_3) \quad \text{and} \quad \lambda_{22}p(\vec{x}|\pi_2) \le p(\vec{x}|\pi_3). \tag{6}$$

We can divide these relations by $p(\vec{x}|\pi_3)$ to obtain

$$\text{decide} \quad \pi_1 \quad \text{iff} \quad \lambda_{11}\text{LR}_1 - \lambda_{22}\text{LR}_2 > 0 \quad \text{and} \quad \lambda_{11}\text{LR}_1 > 1 \tag{7}$$

$$\text{decide} \quad \pi_2 \quad \text{iff} \quad \lambda_{11}\text{LR}_1 - \lambda_{22}\text{LR}_2 \le 0 \quad \text{and} \quad \lambda_{22}\text{LR}_2 > 1 \tag{8}$$

$$\text{decide} \quad \pi_3 \quad \text{iff} \quad \lambda_{11}\text{LR}_1 \le 1 \quad \text{and} \quad \lambda_{22}\text{LR}_2 \le 1, \tag{9}$$

where $\text{LR}_i \equiv p(\vec{x}|\pi_i)/p(\vec{x}|\pi_3)$ are the likelihood ratio decision variables used by the ideal observer.[4,5] The decision boundary lines which partition the $(\text{LR}_1, \text{LR}_2)$ decision plane into the regions Z_1, Z_2, and Z_3 are thus

$$\lambda_{11}\text{LR}_1 - \lambda_{22}\text{LR}_2 = 0 \tag{10}$$

$$\lambda_{11}\text{LR}_1 = 1 \tag{11}$$

$$\lambda_{22}\text{LR}_2 = 1. \tag{12}$$

Note that Eq. 12 is just the difference between Eqs. 10 and 11. If we require λ_{11} and λ_{22} to be positive, the decision rule is an ideal observer decision rule.[5] Since neither the decision variables nor the form of the decision rule depend on the particular choices of α_{11} and α_{22}, we can conclude that the three-class sensitivity ROC surface, obtained by allowing λ_{11} and λ_{22} to take on all possible positive values, is optimal for the observer defined in Eqs. 10–12, in the sense that no other observer can achieve a higher sensitivity surface (*i.e.*, a surface with a greater value of P_{33} at a given value of (P_{11}, P_{22})). The optimal observer for this performance metric is seen to be the three-class ideal observer, with its decision criteria constrained so that the line separating classes π_1 and π_3 is vertical, the line separating classes π_2 and π_3 is horizontal, and the line separating classes π_1 and

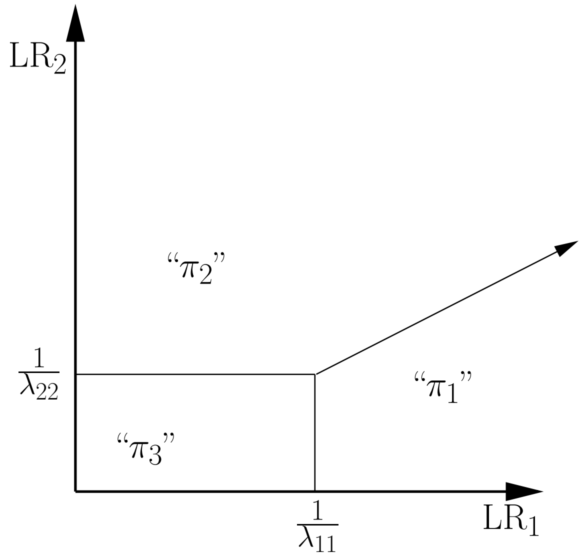

Figure 1. The decision rule which is found to be optimal in the sense of maximizing the ROC surface composed of only the observer sensitivities. The decision variables are the likelihood ratios used by the general three-class ideal observer, and the number of decision criteria is reduced from five (for the general three-class ideal observer) to two.

π_2 passes through the origin with slope $\lambda_{11}/\lambda_{22}$ (and thus intersects the other two lines as required). Note that the number of free decision criteria has been reduced from five (for the general three-class ideal observer) to two (as expected for a surface in a three-dimensional ROC space).

This decision rule is shown in Fig. 1. It is interesting to note that this observer is identical to the special case of the ideal observer evaluated by He *et al.*,[12] which we have shown[14, 15] to be a special case of the decision rule proposed by Scurfield.[9]

3. *N*-CLASS OBSERVERS

The results of the preceding section can be generalized to tasks with N classes for any $N > 2$. We now have a set of N^2 conditional classification probabilities P_{ij}, with N sensitivities P_{ii}. Equation 1 remains unchanged, except that there are of course now N regions Z_i into which the domain of $\vec{x}$ is partitioned (*i.e.*, classes into which the observations are classified), and the observations are drawn from N distributions of the form $p(\vec{x}|\pi_j)$.

Without loss of generality, we seek to maximize P_{NN} subject to the constraints $P_{ii} = \alpha_{ii}$ for $1 \leq i \leq N-1$, where $0 \leq \alpha_{ii} \leq 1$. We define the function

$$F \equiv P_{NN} + \sum_{i=1}^{N-1} \lambda_{ii}(P_{ii} - \alpha_{ii}), \tag{13}$$

where the λ_{ii} are the Lagrange multipliers. Note that if we can find a decision rule (a partitioning of the domain of $\vec{\mathbf{x}}$ into Z_i $\{1 \leq i \leq N\}$) that maximizes F for arbitrary values of the λ_{ii}, then this will be equivalent to maximizing P_{NN} at the point at which the constrain equations are satisfied (*i.e.*, at the point $P_{ii} = \alpha_{ii}$ $\{1 \leq i \leq N-1\}$).

As in the preceding section, we rewrite F by applying rules for conditional probabilities to obtain:

$$F = -\sum_{i=1}^{N-1} \lambda_{ii}\alpha_{ii} + \left(1 - \sum_{i=1}^{N-1} P_{iN}\right) + \sum_{i=1}^{N-1} \lambda_{ii} \left(1 - \sum_{\substack{j=1 \\ j \neq i}}^{N} P_{ji}\right)$$

$$
\begin{aligned}
&= 1 + \sum_{i=1}^{N-1} \lambda_{ii}(1-\alpha_{ii}) - \left\{ \left[\sum_{i=1}^{N-1} \left(\sum_{\substack{j=1\\j\neq i}}^{N} \lambda_{jj} P_{ij} \right) + P_{iN} \right] + \left[\sum_{i=1}^{N-1} \lambda_{ii} P_{Ni} \right] \right\} \\
&= 1 + \sum_{i=2}^{N} \lambda_{ii}(1-\alpha_{ii}) \\
&\quad - \left\{ \sum_{i=1}^{N-1} \int_{Z_i} \left[\sum_{\substack{j=1\\j\neq i}}^{N} \lambda_{jj} p(\vec{x}|\pi_j) \right] + p(\vec{x}|\pi_N)\, d^m\vec{x} + \int_{Z_N} \sum_{i=1}^{N-1} \lambda_{ii} p(\vec{x}|\pi_i)\, d^m\vec{x} \right\}. \qquad (14)
\end{aligned}
$$

For a given set of values of the parameters λ_{ii} $\{1 \leq i \leq N-1\}$, F is maximized when the quantity in braces is minimized. This quantity, in turn, can be minimized by assigning choosing the regions Z_i such that a given $\vec{x}$ to the region Z_i such that the ith integrand (from among the integrals in braces in Eq. 14) is minimized. (Situations in which two or more of the integrands yield the same minimal value for a given $\vec{x}$ can be decided in an arbitrary but consistent fashion.)

That is,

$$
\begin{aligned}
\text{decide} \quad \pi_i\{i<N\} \quad \text{iff} \quad & \lambda_{jj} p(\vec{x}|\pi_j) < \lambda_{ii} p(\vec{x}|\pi_i) \qquad \{i<j<N\} \\
\text{and} \quad & p(\vec{x}|\pi_N) < \lambda_{ii} p(\vec{x}|\pi_i) \\
\text{and} \quad & \lambda_{jj} p(\vec{x}|\pi_j) \leq \lambda_{ii} p(\vec{x}|\pi_i) \qquad \{j<i<N\} \qquad (15) \\
\text{decide} \quad \pi_N \quad \text{iff} \quad & \lambda_{jj} p(\vec{x}|\pi_j) \leq p(\vec{x}|\pi_N) \qquad \{j<N\}. \qquad (16)
\end{aligned}
$$

We can divide these relations by $p(\vec{x}|\pi_N)$ to obtain

$$
\begin{aligned}
\text{decide} \quad \pi_i\{i<N\} \quad \text{iff} \quad & \lambda_{ii}\mathrm{LR}_i - \lambda_{jj}\mathrm{LR}_j > 0 \qquad \{i<j<N\} \\
& \text{and} \lambda_{ii}\mathrm{LR}_i > 1 \\
& \text{and} \lambda_{jj}\mathrm{LR}_j - \lambda_{ii}\mathrm{LR}_i \leq 0 \qquad \{j<i<N\} \qquad (17) \\
\text{decide} \quad \pi_N \quad \text{iff} \quad & \lambda_{jj}\mathrm{LR}_j \leq 1 \qquad \{j<N\}, \qquad (18)
\end{aligned}
$$

where $\mathrm{LR}_i \equiv p(\vec{x}|\pi_i)/p(\vec{x}|\pi_N)$ are the likelihood ratio decision variables used by the ideal observer.[4,5] The decision boundary hyperplanes which partition the $\vec{\mathrm{LR}} \equiv (\mathrm{LR}_1, \ldots, \mathrm{LR}_{N-1})$ decision space into the regions Z_i are thus

$$
\begin{aligned}
\lambda_{ii}\mathrm{LR}_i - \lambda_{jj}\mathrm{LR}_j &= 0 \qquad \{i<j<N\} \qquad (19) \\
\lambda_{ii}\mathrm{LR}_i &= 1 \qquad \{i<N\}. \qquad (20)
\end{aligned}
$$

Note that any of these equations, for example that defining part of the boundary between classes π_j and π_k , can be expressed as the difference of two other such equations (in this example, those defining boundaries between classes π_i and π_j, and between classes pi_i and π_k). If we require the λ_{ii} to be positive, the resulting decision rule is an ideal observer decision rule.[5] Since neither the decision variables nor the form of the decision rule depend on the particular choices of α_{ii}, we can conclude that the N-class sensitivity ROC hypersurface, obtained by allowing the λ_{ii} to take on all possible positive values, is optimal for the observer defined in Eqs. 19 and 20, in the sense that no other observer can achieve a higher sensitivity hypersurface (*i.e.*, one with a greater value of P_{NN} at a given value of $(P_{11}, \ldots, P_{(N-1)(N-1)})$). The optimal observer for this performance metric is seen to be the N-class ideal observer, with its decision criteria constrained so that the boundary separating classes π_i and π_N is a hyperplane defined by $\mathrm{LR}_i = 1/\lambda_{ii}$, while the boundary separating classes π_i and π_j is a hyperplane defined by $\lambda_{ii}\mathrm{LR}_i = \lambda_{jj}\mathrm{LR}_j$.

Although an intuitive geometric understanding of this decision rule is more elusive than in the three-class case, it is at least evident that the boundaries intersect as expected; that is, the boundary separating classes π_i and π_j intersects the boundary separating classes π_i and π_k, and also intersects the boundary separating classes π_j and π_k. Note also that the number of free decision criteria has been reduced from $N^2 - N - 1$ (for the general N-class ideal observer) to $N - 1$ (as expected for a hypersurface in an N-dimensional ROC space). More importantly, comparison of Eqs. 19 and 20 with Eqs. 10–12 reveals this N-class observer to be an obvious extension from three to N classes of the observer described in the preceding section.

4. CONCLUSIONS

A fully general performance evaluation methodology for the three-class classification task has yet to be developed, a frustrating state of affairs given the great success and wide application of ROC analysis to two-class classification tasks. A primary reason for the difficulty in developing a fully general extension of ROC analysis to the three-class classification task is the rapid increase in the number of performance measurement variables and decision criteria necessary to characterize observer (in particular, ideal observer) performance. Specifically, the number of sensitivities or misclassification rates needed increases from two to six (and to $N^2 - N$ in the general case), while the number of decision criteria increases from a single decision variable threshold to a set of five mutually constrained[8] criteria (and to $N^2 - N - 1$ in the general case). In short, the complexity of the problem increases not linearly with the number of classes, but quadratically.

The motivation for the numerous proposed methods, outlined in Sec. 1, for evaluating the performance of a three-class classifier in terms of two-dimensional surfaces in three-dimensional ROC spaces (rather than the five-dimensional hypersurfaces in six-dimensional ROC spaces required by the theory) is thus quite clear. We currently lack a theoretical framework with which to judge the appropriateness of any of the proposed methods to any particular classification task. However, even if one chooses to adopt a performance evaluation metric known to provide an incomplete description of observer performance, it is still reasonable to ask what observer, if any, will achieve optimal performance with respect to that metric.

We have addressed that question in regard to measurement of an observer's performance in terms of only its sensitivities (the probabilities of correctly classifying the three, or in general N, classes of observations). Theoretically, this is clearly an incomplete measure of performance (another set of three, or in general $N^2 - 2N$, misclassification rates are necessary). Conceding this point, we consider it a nontrivial observation, derived in the preceding sections, that the observer which optimizes this limited performance metric is not one unrelated to the general ideal observer, nor an arcane special case of the ideal observer, but a special case of the ideal observer which is in a subjective sense quite simple, and which has been independently evaluated from very different perspectives by other researchers.[9,12] We find these results at once reassuring and encouraging, and hope that research into this thorny problem will continue to bear unexpected fruit.

ACKNOWLEDGMENTS

This work was supported by grant W81XWH-04-1-0495 from the US Army Medical Research and Materiel Command (D. C. Edwards, principal investigator). Charles E. Metz is a shareholder in R2 Technology, Inc. (Sunnyvale, CA).

REFERENCES

1. J. P. Egan, *Signal Detection Theory and ROC Analysis*, Academic Press, New York, 1975.
2. C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine* **VIII**(4), pp. 283–298, 1978.
3. D. C. Edwards, L. Lan, C. E. Metz, M. L. Giger, and R. M. Nishikawa, "Estimating three-class ideal observer decision variables for computerized detection and classification of mammographic mass lesions," *Med. Phys.* **31**, pp. 81–90, 2004.
4. H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*, John Wiley & Sons, New York, 1968.
5. D. C. Edwards, C. E. Metz, and M. A. Kupinski, "Ideal observers and optimal ROC hypersurfaces in N-class classification," *IEEE Trans. Med. Imag.* **23**, pp. 891–895, 2004.
6. C. E. Metz and X. Pan, "'Proper' binormal ROC curves: Theory and maximum-likelihood estimation," *J. Math. Psychol.* **43**, pp. 1–33, 1999.

7. D. C. Edwards, C. E. Metz, and R. M. Nishikawa, "The hypervolume under the ROC hypersurface of 'near-guessing' and 'near-perfect' observers in N-class classification tasks," *IEEE Trans. Med. Imag.* **24**, pp. 293–299, 2005.
8. D. C. Edwards and C. E. Metz, "Restrictions on the three-class ideal observer's decision boundary lines," *IEEE Trans. Med. Imag.* **24**, pp. 1566–1573, 2005.
9. B. K. Scurfield, "Generalization of the theory of signal detectability to n-event m-dimensional forced-choice tasks," *J. Math Psychol.* **42**, pp. 5–31, 1998.
10. D. Mossman, "Three-way ROCs," *Med. Decis. Making* **19**, pp. 78–89, 1999.
11. H.-P. Chan, B. Sahiner, L. M. Hadjiiski, N. Petrick, and C. Zhou, "Design of three-class classifiers in computer-aided diagnosis: Monte carlo simulation study," in Proc. SPIE Vol. 5032 *Medical Imaging 2003: Image Processing*, Milan Sonka and J. Michael Fitzpatrick, eds., pp. 567–578, (SPIE, Bellingham, WA), 2003.
12. X. He, C. E. Metz, B. M. W. Tsui, J. M. Links, and E. C. Frey, "Three-class ROC analysis — I. A decision theoretic approach," *IEEE Trans. Med. Imag.*, 2005. (in review).
13. S. I. Grossman, *Multivariable Calculus, Linear Algebra, and Differential Equations: Second Edition*, Harcourt Brace Jovanovich, San Diego, CA, 1986.
14. D. C. Edwards and C. E. Metz, "Review of several proposed three-class classification decision rules and their relation to the ideal observer decision rule," in Proc. SPIE Vol. 5749 *Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment*, Miguel P. Eckstein and Yulei Jiang, eds., pp. 128–137, (SPIE, Bellingham, WA), 2005.
15. D. C. Edwards and C. E. Metz, "Analysis of proposed three-class classification decision rules in terms of the ideal observer decision rule," *J. Math. Psychol.*, 2005. (in review).