Hypervolume under the ROC hypersurface of a "near-guessing" ideal observer in a three-class classification task

Darrin C. Edwards*, Charles E. Metz, Robert M. Nishikawa

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

ABSTRACT

We express the performance of the three-class "guessing" observer in terms of the six probabilities which make up a three-class receiver operating characteristic (ROC) space, in a formulation in which "sensitivities" are eliminated in constructing the ROC space (equivalent to using false-negative fraction and false-positive fraction in a two-class task). We then show that the "guessing" observer's performance in terms of these conditional probabilities is completely described by a degenerate hypersurface with only two degrees of freedom (as opposed to the five required, in general, to achieve a true hypersurface in such a ROC space). It readily follows that the hypervolume under such a degenerate hypersurface must be zero. We then consider a "near-guessing" task; that is, a task in which the three underlying data probability density functions (PDFs) are nearly identical, controlled by two parameters which may vary continuously to zero (at which point the PDFs become identical). The hypervolume under the ROC hypersurface of an observer in the three-class classification task tends continuously to zero as the underlying data PDFs converge continuously to identity (a "guessing" task). The hypervolume under the ROC hypersurface of a "perfect" ideal observer (a task in which the three data PDFs never overlap) is also found to be zero in the ROC space formulation under consideration. This suggests that hypervolume may not be a useful performance metric in three-class classification tasks, despite the utility of the area under the ROC curve for two-class tasks.

Keywords: ROC analysis, three-class classification, ROC performance metrics

1. INTRODUCTION

We are attempting to develop a fully automated mass lesion classification scheme for computer-aided diagnosis (CAD) in mammography. This scheme will combine two schemes developed at the University of Chicago: one for automatically detecting mass lesions in mammograms,[1–5] and one for classifying known lesions as malignant or benign.[6–10] Combining these two types of CAD scheme is inherently difficult, because the output of the detection scheme will necessarily include false-positive (FP) computer detections in addition to the malignant and benign lesions to be classified. These FP computer detections correspond to objects which were by design not included in the training sample of the classification scheme, because they are not members of the data population (benign and malignant mass breast lesions) for which the classification scheme was created. It is clear then that the detection scheme's output cannot be used unmodified as the input to the classification scheme.

Our approach has been to treat this problem explicitly as a three-class classification task. That is, the outputs of the detection scheme should be classified as malignant lesions, benign lesions, and non-lesions (FP computer detections), and the classifier to be estimated is the ideal observer decision function for this task. Such an approach presents considerable difficulties of its own. On the one hand, decision functions, in particular ideal observer decision functions, increase rapidly with the number of classes involved. On the other hand, fully general performance evaluation methods, in particular a three-class extension of receiver operating characteristic (ROC) analysis have yet to be developed for such a task.

Although we have had preliminary success in using Bayesian artificial neural networks (BANNs)[11,12] to estimate three-class ideal-observer-related decision variables,[13,14] the task of developing a three-class extension to ROC analysis has proved somewhat more daunting. Our initial efforts in this direction have thus been more

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

theoretical than practical so far.[15] One issue we began to investigate recently was the calculation of an obvious generalization of the well-known area under the ROC curve (AUC) performance metric, a quantity we are calling the "hypervolume under the ROC hypersurface." Detailed consideration of the integrals involved in calculating this quantity led us to the counterintuitive conclusion that, despite the great success and utility of the AUC performance metric in two-class classification tasks, the hypervolume under the ROC hypersurface does not appear to be a useful performance metric in three-class classification tasks. (We also have reason to believe that this result applies to N-class classification tasks in general when $N > 2$.[16]) The proof of this claim is arrived at by considering observer performance in two extremes: the "guessing" observer and the "perfect" observer.

2. THE ROC HYPERSURFACE OF THE THREE-CLASS "GUESSING" OBSERVER

The performance of an observer in a three-class classification task is completely determined by a hypersurface with five degrees of freedom in a six-dimensional ROC space.[17] Without loss of generality, we can specify any point in the ROC space by a vector of the misclassification probabilities $[P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2), P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3), P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1), P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3), P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2), P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1)]^\dagger$.[15] Here the three classes are denoted by π_1, π_2, and π_3, $\mathbf{d}$ denotes the class to which an observation is assigned (the "decision"), and $\mathbf{t}$ is the class to which it actually belongs (the "truth"); we use boldface type to denote statistically variable quantities. We can also, again without loss of generality, consider the ROC hypersurface to be given by $P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1)$ considered as a function of the other five misclassification probabilities.[15]

Consider the performance of an observer which makes decisions by "guessing", that is, in a random fashion unrelated to the actual class $\mathbf{t}$ from which a given observation is drawn. (Note that this corresponds to the performance of the ideal observer when the probability density functions (PDFs) of the observational data are identical, *i. e.*, $p(\vec{\mathbf{x}}|\pi_1) = p(\vec{\mathbf{x}}|\pi_2) = p(\vec{\mathbf{x}}|\pi_3)$.) In this case, we clearly must have

$$P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2) = P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3), \tag{1}$$

$$P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1) = P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3), \tag{2}$$

$$P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2) = P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1). \tag{3}$$

Defining $\alpha \equiv P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2)$, $\beta \equiv P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1)$, and $\gamma \equiv P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2)$, we see that the performance of the "guessing" observer is given by a locus of vectors of the form

$$\begin{bmatrix} \alpha \\ \alpha \\ \beta \\ \beta \\ \gamma \\ \gamma \end{bmatrix}, \tag{4}$$

where each of α, β, γ is restricted to the range $[0, 1]$. Furthermore, note that

$$\begin{aligned} P(\mathbf{d}=\pi_1) &= P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_1)P(\mathbf{t}=\pi_1) + P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2)P(\mathbf{t}=\pi_2) \\ &\quad + P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3)P(\mathbf{t}=\pi_3) \\ &= \alpha P(\mathbf{t}=\pi_1) + \alpha P(\mathbf{t}=\pi_2) + \alpha P(\mathbf{t}=\pi_3) \\ &= \alpha, \end{aligned} \tag{5}$$

and similarly $P(\mathbf{d}=\pi_2) = \beta$, $P(\mathbf{d}=\pi_3) = \gamma$. This immediately gives $\gamma = 1 - \alpha - \beta$. Thus the performance of the guessing observer is given by

$$\begin{bmatrix} P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1) \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha \\ \beta \\ \beta \\ 1-\alpha-\beta \\ 1-\alpha-\beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ -1 \\ -1 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}. \tag{6}$$

This is the parametric equation for a two-dimensional plane in a six-dimensional space; the actual performance of the "guessing" observer will of course be further restricted to a triangle within this plane such that $0 \le \alpha \le 1, 0 \le \beta \le 1, 0 \le 1 - \alpha - \beta \le 1$.

3. THE ROC HYPERSURFACE OF A THREE-CLASS "NEAR-GUESSING" OBSERVER

Consider observational data $\vec{\mathbf{x}}$ drawn from three PDFs:

$$\begin{aligned} p(\vec{x}|\mathbf{t} = \pi_1) &= p(\vec{x}|\mathbf{t} = \pi_3) + \delta h_1(\vec{x}), & (7)\\ p(\vec{x}|\mathbf{t} = \pi_2) &= p(\vec{x}|\mathbf{t} = \pi_3) + \epsilon h_2(\vec{x}), & (8)\\ p(\vec{x}|\mathbf{t} = \pi_3), & & (9) \end{aligned}$$

where $0 \le \delta \le 1$, $\int h_1(\vec{x}) d^n\vec{x} = 0$, and $|h_1(\vec{x})| \le p(\vec{x}|\mathbf{t} = \pi_3)$; and similarly, $0 \le \epsilon \le 1$, $\int h_2(\vec{x}) d^n\vec{x} = 0$, and $|h_2(\vec{x})| \le p(\vec{x}|\mathbf{t} = \pi_3)$. In the limit as δ and ϵ both approach zero, we expect the performance of any observer for this task to converge smoothly to that of the "guessing" observer.

Decisions are made by partitioning the decision variable space into three regions, determined by a total of five parameters; we denote these parameters by the components of a vector $\vec{\gamma}$. An observer which uses more than five parameters for a three-class classification task can always be replaced by a simplified observer, such that the "excess" parameters are eliminated by the requirement that $P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_1)$ be minimized, thereby collapsing the dimensionality of the parameter space to five. On the other hand, an observer which uses fewer than five decision parameters will fail to generate a true ROC hypersurface — *i. e.*, one with five degrees of freedom in the six-dimensional ROC space. Such "degenerate" observers will not be considered here (apart from the "guessing" observer itself).

We can thus define three regions which partition the original data space, given particular values of the parameters $\vec{\gamma}$, by:

$$\begin{aligned} \mathcal{D}_1(\vec{\gamma}) &\equiv \{\vec{x} : \mathbf{d} = \pi_1 \quad \text{given} \quad \vec{\gamma}\}, & (10)\\ \mathcal{D}_2(\vec{\gamma}) &\equiv \{\vec{x} : \mathbf{d} = \pi_2 \quad \text{given} \quad \vec{\gamma}\}, & (11)\\ \mathcal{D}_3(\vec{\gamma}) &\equiv \{\vec{x} : \mathbf{d} = \pi_3 \quad \text{given} \quad \vec{\gamma}\}. & (12) \end{aligned}$$

For a non-random observer, the $\mathcal{D}_i$ can be expected to depend implicitly on the PDFs in Eqs. 7–9, and therefore on δ and ϵ. The misclassification probabilities which define the ROC hypersurface are then given by

$$\begin{bmatrix} P(\mathbf{d} = \pi_1|\mathbf{t} = \pi_2) \\ P(\mathbf{d} = \pi_1|\mathbf{t} = \pi_3) \\ P(\mathbf{d} = \pi_2|\mathbf{t} = \pi_1) \\ P(\mathbf{d} = \pi_2|\mathbf{t} = \pi_3) \\ P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_2) \\ P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_1) \end{bmatrix} = \begin{bmatrix} \int_{\mathcal{D}_1} p(\vec{x}|\mathbf{t} = \pi_2)\ d^n\vec{x} \\ \int_{\mathcal{D}_1} p(\vec{x}|\mathbf{t} = \pi_3)\ d^n\vec{x} \\ \int_{\mathcal{D}_2} p(\vec{x}|\mathbf{t} = \pi_1)\ d^n\vec{x} \\ \int_{\mathcal{D}_2} p(\vec{x}|\mathbf{t} = \pi_3)\ d^n\vec{x} \\ \int_{\mathcal{D}_3} p(\vec{x}|\mathbf{t} = \pi_2)\ d^n\vec{x} \\ \int_{\mathcal{D}_3} p(\vec{x}|\mathbf{t} = \pi_1)\ d^n\vec{x} \end{bmatrix}. \tag{13}$$

Using Eqs. 7 and 8, we can rewrite this as

$$\begin{bmatrix} P(\mathbf{d} = \pi_1|\mathbf{t} = \pi_2) \\ P(\mathbf{d} = \pi_1|\mathbf{t} = \pi_3) \\ P(\mathbf{d} = \pi_2|\mathbf{t} = \pi_1) \\ P(\mathbf{d} = \pi_2|\mathbf{t} = \pi_3) \\ P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_2) \\ P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_1) \end{bmatrix} = \begin{bmatrix} P(\mathbf{d} = \pi_1|\mathbf{t} = \pi_3) + \epsilon \int_{\mathcal{D}_1} h_2(\vec{x})\ d^n\vec{x} \\ P(\mathbf{d} = \pi_1|\mathbf{t} = \pi_3) \\ P(\mathbf{d} = \pi_2|\mathbf{t} = \pi_3) + \delta \int_{\mathcal{D}_2} h_1(\vec{x})\ d^n\vec{x} \\ P(\mathbf{d} = \pi_2|\mathbf{t} = \pi_3) \\ P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_3) + \epsilon \int_{\mathcal{D}_3} h_2(\vec{x})\ d^n\vec{x} \\ P(\mathbf{d} = \pi_3|\mathbf{t} = \pi_3) + \delta \int_{\mathcal{D}_3} h_1(\vec{x})\ d^n\vec{x} \end{bmatrix}. \tag{14}$$

Defining the functions $H_{ij} \equiv \int_{\mathcal{D}_i} h_j(\vec{x})\; d^n\vec{x}$ allows us to simplify the notation slightly:

$$\begin{bmatrix} P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1) \end{bmatrix} = \begin{bmatrix} P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3)+\epsilon H_{12} \\ P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3)+\delta H_{21} \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_3)+\epsilon H_{32} \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_3)+\delta H_{31} \end{bmatrix}. \tag{15}$$

Now of course $P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_3) = 1 - P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3) - P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3)$; for simplicity, we will write $\alpha \equiv P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3)$ and $\beta \equiv P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3)$. Equation 15 can now be written as

$$\begin{bmatrix} P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1) \end{bmatrix} = \begin{bmatrix} \alpha+\epsilon H_{12} \\ \alpha \\ \beta+\delta H_{21} \\ \beta \\ 1-\alpha-\beta+\epsilon H_{32} \\ 1-\alpha-\beta+\delta H_{31} \end{bmatrix}, \tag{16}$$

which further simplifies to

$$\begin{bmatrix} P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_1|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_1) \\ P(\mathbf{d}=\pi_2|\mathbf{t}=\pi_3) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_2) \\ P(\mathbf{d}=\pi_3|\mathbf{t}=\pi_1) \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha \\ \beta \\ \beta \\ 1-\alpha-\beta \\ 1-\alpha-\beta \end{bmatrix} + \delta \begin{bmatrix} 0 \\ 0 \\ H_{21} \\ 0 \\ 0 \\ H_{31} \end{bmatrix} + \epsilon \begin{bmatrix} H_{12} \\ 0 \\ 0 \\ 0 \\ H_{32} \\ 0 \end{bmatrix}. \tag{17}$$

The first term on the righthand side of this equation is just the expression for the "guessing" observer (*cf.* the lefthand side of Eq. 6). The other two terms on the righthand side of the equation tend to zero as δ and ϵ tend to zero. Note that the H_{ij} may in general depend on δ and ϵ *via* Eqs. 10–12, but

$$\begin{aligned} |H_{ij}| &= \left| \int_{\mathcal{D}_i} h_j(\vec{x})\, d^n\vec{x} \right| \\ &\le \int_{\mathcal{D}_i} |h_j(\vec{x})|\, d^n\vec{x} \\ &\le \int_{\mathcal{D}_i} p(\vec{x}|\mathbf{t}=\pi_3)\, d^n\vec{x} \\ &= P(\mathbf{d}=\pi_i|\mathbf{t}=\pi_3) \\ &\le 1. \end{aligned} \tag{18}$$

Thus the H_{ij} are bounded, and will possess Taylor expansions in δ and ϵ (*i. e.*, will not depend on terms of the form δ^{-k}, ϵ^{-k} for positive integers k). Therefore, operating points on the ROC hypersurface of the "near-guessing" observer tend continuously toward points on the ROC hypersurface of the "guessing" observer. Note that the six terms α, β, δH_{21}, δH_{31}, ϵH_{12}, ϵH_{32} are not all independent, since they each depend implicitly for fixed δ and ϵ on the five decision parameters $\vec{\gamma}$. That is, the ROC hypersurface given by Eq. 17 possesses only five degrees of freedom.

4. THE HYPERVOLUME UNDER THE ROC HYPERSURFACE OF A THREE-CLASS "NEAR-GUESSING" OBSERVER

In the preceding section, it was shown that the ROC hypersurface of a "near-guessing" observer tends continuously to the ROC hypersurface of a "guessing" observer as the PDFs of the observational data tend arbitrarily toward identical distributions. Intuitively, one would expect that the hypervolumes under these hypersurfaces should also tend toward each other. Since intuition can occasionally be an unreliable guide in analyzing three-class classification tasks, it would be reassuring if the results of the preceding section could be applied directly to the calculation of the relevant hypervolumes.

For this section, we will write $P(\mathbf{d} = \pi_i | \mathbf{t} = \pi_j)$ as $P_{ij}(\vec{\gamma})$, emphasizing that it is a function of the decision parameters chosen. We thus rewrite Eq. 15 to obtain

$$\begin{bmatrix} P_{12}(\vec{\gamma}) \\ P_{13}(\vec{\gamma}) \\ P_{21}(\vec{\gamma}) \\ P_{23}(\vec{\gamma}) \\ P_{32}(\vec{\gamma}) \\ P_{31}(\vec{\gamma}) \end{bmatrix} = \begin{bmatrix} P_{13}(\vec{\gamma}) + \epsilon H_{12}(\vec{\gamma}) \\ P_{13}(\vec{\gamma}) \\ P_{23}(\vec{\gamma}) + \delta H_{21}(\vec{\gamma}) \\ P_{23}(\vec{\gamma}) \\ P_{32}(\vec{\gamma}) \\ P_{32}(\vec{\gamma}) + \delta H_{31} - \epsilon H_{32} \end{bmatrix}. \tag{19}$$

To find the hypervolume under the ROC surface given by P_{31} considered as a function of $(P_{12}, P_{13}, P_{21}, P_{23}, P_{32})$, one must evaluate the integral

$$\int\!\!\int\!\!\int\!\!\int\!\!\int P_{31}\, dP_{12}\, dP_{13}\, dP_{21}\, dP_{23}\, dP_{32}. \tag{20}$$

(The domain of the integral is simply the set of all P_{ij} such that P_{31} is defined.) Note that, for the "guessing" observer, we expect this integral to be zero due to dimensionality considerations — the ROC hypersurface has only two degrees of freedom (*cf.* Eq. 6), not the five required in this six-dimensional ROC space. To see this explicitly, one can rearrange the order of integration to obtain

$$\int\!\!\int\!\!\int\!\!\int\!\!\int P_{31}\, dP_{32}\, dP_{12}\, dP_{13}\, dP_{21}\, dP_{23}, \tag{21}$$

and then consider the innermost integral $\int P_{31}\, dP_{32}$ for fixed values of $P_{12} = P_{13}$ and $P_{21} = P_{23}$. Then the limits of integration of this innermost definite integral become, again by Eq. 6,

$$\int_{1-P_{13}-P_{23}}^{1-P_{13}-P_{23}} P_{31}\, dP_{32}, \tag{22}$$

which is zero by inspection.

We now return to the general case of a "near-guessing" observer. One way to evaluate the integral in Eq. 20 is to reexpress it explicitly in terms of the decision parameters $\vec{\gamma}$, *via* the Jacobian

$$J \equiv \begin{vmatrix} \frac{\partial P_{12}}{\partial \gamma_1} & \frac{\partial P_{12}}{\partial \gamma_2} & \frac{\partial P_{12}}{\partial \gamma_3} & \frac{\partial P_{12}}{\partial \gamma_4} & \frac{\partial P_{12}}{\partial \gamma_5} \\ \frac{\partial P_{13}}{\partial \gamma_1} & \frac{\partial P_{13}}{\partial \gamma_2} & \frac{\partial P_{13}}{\partial \gamma_3} & \frac{\partial P_{13}}{\partial \gamma_4} & \frac{\partial P_{13}}{\partial \gamma_5} \\ \frac{\partial P_{21}}{\partial \gamma_1} & \frac{\partial P_{21}}{\partial \gamma_2} & \frac{\partial P_{21}}{\partial \gamma_3} & \frac{\partial P_{21}}{\partial \gamma_4} & \frac{\partial P_{21}}{\partial \gamma_5} \\ \frac{\partial P_{23}}{\partial \gamma_1} & \frac{\partial P_{23}}{\partial \gamma_2} & \frac{\partial P_{23}}{\partial \gamma_3} & \frac{\partial P_{23}}{\partial \gamma_4} & \frac{\partial P_{23}}{\partial \gamma_5} \\ \frac{\partial P_{32}}{\partial \gamma_1} & \frac{\partial P_{32}}{\partial \gamma_2} & \frac{\partial P_{32}}{\partial \gamma_3} & \frac{\partial P_{32}}{\partial \gamma_4} & \frac{\partial P_{32}}{\partial \gamma_5} \end{vmatrix} \tag{23}$$

where the vertical bars indicate that the determinant of the enclosed matrix is to be taken, and where γ_i denotes the ith component of $\vec{\gamma}$. (We assume that indices of the parameters $\vec{\gamma}$ have been chosen appropriately so that no negative sign is introduced, *i. e.*, volumes remain positive.)

For the "guessing" observer, this reduces to

$$J_{\text{guessing}} \equiv \begin{vmatrix} \frac{\partial P_{13}}{\partial \gamma_1} & \frac{\partial P_{13}}{\partial \gamma_2} & \frac{\partial P_{13}}{\partial \gamma_3} & \frac{\partial P_{13}}{\partial \gamma_4} & \frac{\partial P_{13}}{\partial \gamma_5} \\ \frac{\partial P_{13}}{\partial \gamma_1} & \frac{\partial P_{13}}{\partial \gamma_2} & \frac{\partial P_{13}}{\partial \gamma_3} & \frac{\partial P_{13}}{\partial \gamma_4} & \frac{\partial P_{13}}{\partial \gamma_5} \\ \frac{\partial P_{23}}{\partial \gamma_1} & \frac{\partial P_{23}}{\partial \gamma_2} & \frac{\partial P_{23}}{\partial \gamma_3} & \frac{\partial P_{23}}{\partial \gamma_4} & \frac{\partial P_{23}}{\partial \gamma_5} \\ \frac{\partial P_{23}}{\partial \gamma_1} & \frac{\partial P_{23}}{\partial \gamma_2} & \frac{\partial P_{23}}{\partial \gamma_3} & \frac{\partial P_{23}}{\partial \gamma_4} & \frac{\partial P_{23}}{\partial \gamma_5} \\ \frac{\partial P_{32}}{\partial \gamma_1} & \frac{\partial P_{32}}{\partial \gamma_2} & \frac{\partial P_{32}}{\partial \gamma_3} & \frac{\partial P_{32}}{\partial \gamma_4} & \frac{\partial P_{32}}{\partial \gamma_5} \end{vmatrix} \tag{24}$$

where $P_{32} = P_{33} = 1 - P_{13} - P_{23}$. For a "near-guessing" observer, we combine Eqs. 19 and 23 to obtain

$$J_{\text{near}} \equiv \begin{vmatrix} \frac{\partial(P_{13}+\epsilon H_{12})}{\partial\gamma_1} & \frac{\partial(P_{13}+\epsilon H_{12})}{\partial\gamma_2} & \frac{\partial(P_{13}+\epsilon H_{12})}{\partial\gamma_3} & \frac{\partial(P_{13}+\epsilon H_{12})}{\partial\gamma_4} & \frac{\partial(P_{13}+\epsilon H_{12})}{\partial\gamma_5} \\ \frac{\partial P_{13}}{\partial\gamma_1} & \frac{\partial P_{13}}{\partial\gamma_2} & \frac{\partial P_{13}}{\partial\gamma_3} & \frac{\partial P_{13}}{\partial\gamma_4} & \frac{\partial P_{13}}{\partial\gamma_5} \\ \frac{\partial(P_{23}+\delta H_{21})}{\partial\gamma_1} & \frac{\partial(P_{23}+\delta H_{21})}{\partial\gamma_2} & \frac{\partial(P_{23}+\delta H_{21})}{\partial\gamma_3} & \frac{\partial(P_{23}+\delta H_{21})}{\partial\gamma_4} & \frac{\partial(P_{23}+\delta H_{21})}{\partial\gamma_5} \\ \frac{\partial P_{23}}{\partial\gamma_1} & \frac{\partial P_{23}}{\partial\gamma_2} & \frac{\partial P_{23}}{\partial\gamma_3} & \frac{\partial P_{23}}{\partial\gamma_4} & \frac{\partial P_{23}}{\partial\gamma_5} \\ \frac{\partial P_{32}}{\partial\gamma_1} & \frac{\partial P_{32}}{\partial\gamma_2} & \frac{\partial P_{32}}{\partial\gamma_3} & \frac{\partial P_{32}}{\partial\gamma_4} & \frac{\partial P_{32}}{\partial\gamma_5} \end{vmatrix}. \quad (25)$$

From the properties of determinants,[18] it can be shown that

$$J_{\text{near}} = J_{\text{guessing}} + \delta J_1 + \epsilon J_2 + \delta\epsilon J_3, \quad (26)$$

where

$$J_1 = \begin{vmatrix} \frac{\partial P_{13}}{\partial\gamma_1} & \frac{\partial P_{13}}{\partial\gamma_2} & \frac{\partial P_{13}}{\partial\gamma_3} & \frac{\partial P_{13}}{\partial\gamma_4} & \frac{\partial P_{13}}{\partial\gamma_5} \\ \frac{\partial P_{13}}{\partial\gamma_1} & \frac{\partial P_{13}}{\partial\gamma_2} & \frac{\partial P_{13}}{\partial\gamma_3} & \frac{\partial P_{13}}{\partial\gamma_4} & \frac{\partial P_{13}}{\partial\gamma_5} \\ \frac{\partial H_{21}}{\partial\gamma_1} & \frac{\partial H_{21}}{\partial\gamma_2} & \frac{\partial H_{21}}{\partial\gamma_3} & \frac{\partial H_{21}}{\partial\gamma_4} & \frac{\partial H_{21}}{\partial\gamma_5} \\ \frac{\partial P_{23}}{\partial\gamma_1} & \frac{\partial P_{23}}{\partial\gamma_2} & \frac{\partial P_{23}}{\partial\gamma_3} & \frac{\partial P_{23}}{\partial\gamma_4} & \frac{\partial P_{23}}{\partial\gamma_5} \\ \frac{\partial P_{32}}{\partial\gamma_1} & \frac{\partial P_{32}}{\partial\gamma_2} & \frac{\partial P_{32}}{\partial\gamma_3} & \frac{\partial P_{32}}{\partial\gamma_4} & \frac{\partial P_{32}}{\partial\gamma_5} \end{vmatrix}, \quad (27)$$

$$J_2 = \begin{vmatrix} \frac{\partial H_{12}}{\partial\gamma_1} & \frac{\partial H_{12}}{\partial\gamma_2} & \frac{\partial H_{12}}{\partial\gamma_3} & \frac{\partial H_{12}}{\partial\gamma_4} & \frac{\partial H_{12}}{\partial\gamma_5} \\ \frac{\partial P_{13}}{\partial\gamma_1} & \frac{\partial P_{13}}{\partial\gamma_2} & \frac{\partial P_{13}}{\partial\gamma_3} & \frac{\partial P_{13}}{\partial\gamma_4} & \frac{\partial P_{13}}{\partial\gamma_5} \\ \frac{\partial P_{23}}{\partial\gamma_1} & \frac{\partial P_{23}}{\partial\gamma_2} & \frac{\partial P_{23}}{\partial\gamma_3} & \frac{\partial P_{23}}{\partial\gamma_4} & \frac{\partial P_{23}}{\partial\gamma_5} \\ \frac{\partial P_{23}}{\partial\gamma_1} & \frac{\partial P_{23}}{\partial\gamma_2} & \frac{\partial P_{23}}{\partial\gamma_3} & \frac{\partial P_{23}}{\partial\gamma_4} & \frac{\partial P_{23}}{\partial\gamma_5} \\ \frac{\partial P_{32}}{\partial\gamma_1} & \frac{\partial P_{32}}{\partial\gamma_2} & \frac{\partial P_{32}}{\partial\gamma_3} & \frac{\partial P_{32}}{\partial\gamma_4} & \frac{\partial P_{32}}{\partial\gamma_5} \end{vmatrix}, \quad (28)$$

and

$$J_3 = \begin{vmatrix} \frac{\partial H_{12}}{\partial\gamma_1} & \frac{\partial H_{12}}{\partial\gamma_2} & \frac{\partial H_{12}}{\partial\gamma_3} & \frac{\partial H_{12}}{\partial\gamma_4} & \frac{\partial H_{12}}{\partial\gamma_5} \\ \frac{\partial P_{13}}{\partial\gamma_1} & \frac{\partial P_{13}}{\partial\gamma_2} & \frac{\partial P_{13}}{\partial\gamma_3} & \frac{\partial P_{13}}{\partial\gamma_4} & \frac{\partial P_{13}}{\partial\gamma_5} \\ \frac{\partial H_{21}}{\partial\gamma_1} & \frac{\partial H_{21}}{\partial\gamma_2} & \frac{\partial H_{21}}{\partial\gamma_3} & \frac{\partial H_{21}}{\partial\gamma_4} & \frac{\partial H_{21}}{\partial\gamma_5} \\ \frac{\partial P_{23}}{\partial\gamma_1} & \frac{\partial P_{23}}{\partial\gamma_2} & \frac{\partial P_{23}}{\partial\gamma_3} & \frac{\partial P_{23}}{\partial\gamma_4} & \frac{\partial P_{23}}{\partial\gamma_5} \\ \frac{\partial P_{32}}{\partial\gamma_1} & \frac{\partial P_{32}}{\partial\gamma_2} & \frac{\partial P_{32}}{\partial\gamma_3} & \frac{\partial P_{32}}{\partial\gamma_4} & \frac{\partial P_{32}}{\partial\gamma_5} \end{vmatrix}. \quad (29)$$

If we denote the hypervolume under the ROC hypersurface of the "guessing" observer by

$$\begin{aligned} I_{\text{guessing}} &= \int\!\!\int\!\!\int\!\!\int\!\!\int P_{31}\, dP_{12}\, dP_{13}\, dP_{21}\, dP_{23}\, dP_{32} \\ &= \int\!\!\int\!\!\int\!\!\int\!\!\int P_{32}(\vec{\gamma}) J_{\text{guessing}}\, d^5\vec{\gamma}, \end{aligned} \quad (30)$$

then the hypervolume under the ROC hypersurface of a "near-guessing" observer becomes

$$I_{\text{near}} = \int\!\!\int\!\!\int\!\!\int\!\!\int [P_{32}(\vec{\gamma}) + \delta H_{31} - \epsilon H_{32}] \left[J_{\text{guessing}} + \delta J_1 + \epsilon J_2 + \delta\epsilon J_3\right] d^5\vec{\gamma} \quad (31)$$

$$= I_{\text{guessing}} + \delta I_1 + \epsilon I_2 + \delta^2 I_3 + \epsilon^2 I_4 + \delta\epsilon I_5 + \delta^2\epsilon I_6 + \delta\epsilon^2 I_7, \quad (32)$$

where the integrals $I_1 \ldots I_7$ are finite (*i. e.*, they may depend on higher integral powers of δ and ϵ, but not on δ^{-k} or ϵ^{-k} for positive integers k). That is, in the limit as δ and ϵ tend toward zero, I_{near} tends toward I_{guessing} in a continuous fashion.

5. THE HYPERVOLUME UNDER THE ROC HYPERSURFACE OF A THREE-CLASS "NEAR-PERFECT" OBSERVER

In the preceding sections, we established that the hypervolume under the ROC hypersurface of a "guessing" observer is zero, and furthermore that this result is not singular: an observer in a "near-guessing" task will achieve a ROC hypersurface with hypervolume approaching zero continuously as the data PDFs approach identity. An ideal observer in a "perfect" task — *i. e.*, in which the data PDFs never overlap — will also achieve a ROC hypersurface with zero hypervolume, because it can achieve the operating point $\vec{0}$, and thus will not, for any rational decision rule, achieve points interior to the unit hypercube defining ROC space. It is reasonable to ask whether "near-perfect" observers, performing tasks for which the overlap in the underlying data PDFs is nearly negligible, behave similarly to "near-guessing" observers, in the sense that the hypervolume under the ROC hypersurface of such an observer will approach zero in a continuous fashion.

Consider observational data $\vec{\mathbf{x}}$ drawn from three PDFs $p(\vec{x}|\mathbf{t} = \pi_j)$ where $1 \leq j \leq 3$. We denote the mean of $p(\vec{x}|\mathbf{t} = \pi_j)$ by $\vec{\mu}_j$ and note that, without loss of generality, the mean of $p(\vec{x}|\mathbf{t} = \pi_3)$ can be taken to be $\vec{0}$. Furthermore, note that we can apply a linear transformation to the data $\vec{\mathbf{x}}$, and thus effectively to the $\vec{\mu}_j$, such that each of the resulting $\vec{\mu}_j$ is either (a) mutually orthogonal to, or (b) a scalar multiple of, any of the other $\vec{\mu}_i$. Because the transformation applied is linear, the ideal observer for this task will remain the same, and hence the task itself can be considered essentially unchanged.

Let us consider now an observer for this task which is generally not ideal; in fact, we will consider only a single operating point achieved by this observer. The observer decides $d = \pi_i$ for a given observation $\vec{x}$ if

$$(\vec{x} - \vec{\mu}_i) \cdot \frac{(\vec{\mu}_j - \vec{\mu}_i)}{|\vec{\mu}_j - \vec{\mu}_i|} < \frac{1}{2}|\vec{\mu}_j - \vec{\mu}_i| \qquad \{j : 1 \leq j \leq 3, j \neq i\}, \tag{33}$$

with equality for any such relation between two classes being decided in an arbitrary but consistent manner. That is, the observer places hyperplanes between the means of any two classes when attempting to decide between those classes (rather than placing those hyperplanes in the likelihood ratio decision variable space, as would the ideal observer).

Now suppose the task is made slightly "easier," while the observer itself remains unchanged. That is, consider the mean of one PDF, say $\vec{\mu}_i$ for $i \neq 3$, being increased by a factor $1 + \delta$ for $0 \leq \delta \leq 1$, while the location of the decision hyperplanes does not change, except in the special case where $\vec{\mu}_j = \alpha\vec{\mu}_i$ for some other PDF (again with $j \neq 3$). In this latter case we increase both means ($\vec{\mu}'_j = (1 + \delta)\vec{\mu}_j$, $\vec{\mu}'_i = (1 + \delta)\vec{\mu}_i$), and the location of the corresponding decision hyperplane shifts accordingly.

Note that $\vec{\mu}'_i$ is now further away from each decision hyperplane relevant to $\mathbf{d} = \pi_i$ in Eq. 33. In the case $\vec{\mu}_j = \alpha\vec{\mu}_i$, the decision hyperplane is now a distance of $|\frac{\vec{\mu}'_j - \vec{\mu}'_i}{2}| = (1 + \delta)|\frac{\vec{\mu}_j - \vec{\mu}_i}{2}|$ from $\vec{\mu}'_i$. For non-collinear $\vec{\mu}_j$, the direction from $\vec{\mu}'_i$ to the decision hyperplane is given by $\vec{\mu}_j - \vec{\mu}_i$, and since $\vec{\mu}_j$ and $\vec{\mu}'_i$ are orthogonal, $(\vec{\mu}'_i - \vec{\mu}_i) \cdot (\vec{\mu}_j - \vec{\mu}_i) = -\delta|\vec{\mu}_i|^2$; since this quantity is negative, it follows that $\vec{\mu}'_i$ is further from that decision plane than $\vec{\mu}_i$.

It immediately follows from this that none of the misclassification probabilities making up the coordinates of the observer's operating point can increase when moving from the old task to the new one. To see this, consider a change of coordinates in the data space such that $\vec{\mu}'_i$ is now the origin. All of the decision hyperplanes separating this class from the others are effectively moving away from the center of its PDF; since the hyperplanes are translating without rotating, we see immediately that the probability $P(\mathbf{d} = \pi_i|\mathbf{t} = \pi_i)$ cannot decrease (and will increase in general), while the other probabilities $P(\mathbf{d} = \pi_j|\mathbf{t} = \pi_i)$ $(j \neq i)$ cannot increase (and will decrease in general).

Note that any PDF $p(\vec{x})$ must decrease more rapidly than $|\vec{x}|^{-n}$ for sufficiently large $|\vec{x}|$, where n is the dimensionality of $\vec{x}$. This allows us to state qualitatively the sense in which the observer under consideration is "near-perfect": we hypothesize that the $|\vec{\mu}_i|$ are all sufficiently large that this limiting condition is met. Given this condition, the only situation in which an error probability $P(\mathbf{d} = \pi_j|\mathbf{t} = \pi_i)$ $(j \neq i)$ will fail to decrease is if this probability is already zero. By allowing all of the $|\vec{\mu}_i|$ to increase in the manner described above, we can

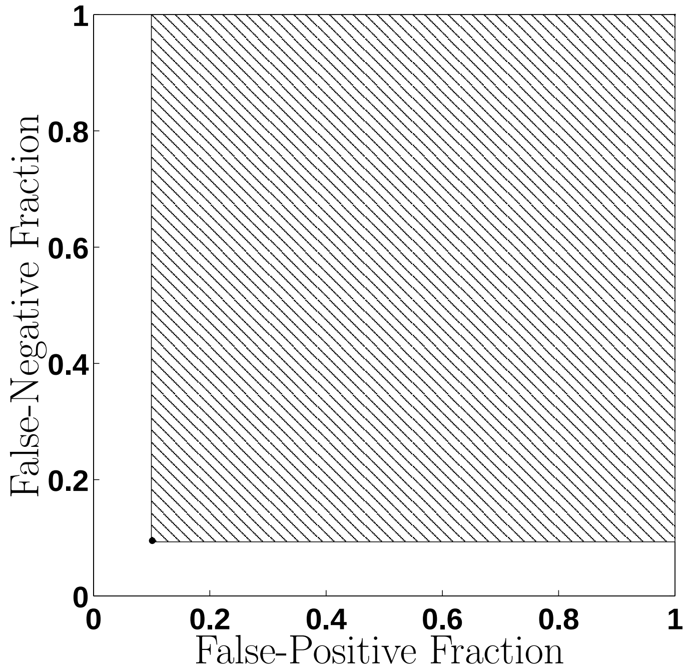

Figure 1. Operating point of an observer in a two-class classification task with coordinates $(\mathrm{FPF}_0, \mathrm{FNF}_0)$, denoted by the point at the lower left corner of the crosshatched region. Since no rational observer will achieve points in the crosshatched region, the area under this observer's ROC curve cannot be greater than $1 - (1 - \mathrm{FPF}_0)(1 - \mathrm{FNF}_0)$.

clearly obtain in general a situation in which each of the misclassification probabilities is either decreasing, or equal to zero.

This implies that the hypervolume under the ROC hypersurfaces of the observers under consideration (however we chose to define their decision rules for operating points other than those described above) must also decrease as the task is made "easier" as described above. To see this, note that if a given observer achieves an operating point $\vec{P}$ on its ROC hypersurface, it cannot achieve another point $\vec{P}'$ such that the components of these points satisfy $P_i' > P_i$ $(1 \leq i \leq 6)$ (because such an observer could be replaced by an observer which achieved $\vec{P}$ for all such points by using the original decision rule for the point $\vec{P}$, thereby achieving unambiguously better performance at those points). Thus, knowing that a given observer achieves an operating point of $\vec{P}$ implies that that observer's ROC hypersurface must have a hypervolume under it of no greater than $1 - \prod_{i=1}^{6}(1 - P_i)$; as the (non-zero) P_i decrease, this upper limit on the hypervolume must also decrease to zero. This point is illustrated in Fig. 1 for the two-class case; here the observer's false-negative fraction, FNF_0, corresponds to $P(\mathbf{d} = \pi_2 | \mathbf{t} = \pi_1)$, and the false-positive fraction, FPF_0, corresponds to $P(\mathbf{d} = \pi_1 | \mathbf{t} = \pi_2)$.

To summarize, we have shown that the known operating point of our simple observer will move closer to the origin for arbitrary data PDFs as those PDFs are moved further apart (*i. e.*, as the underlying task is made "easier"), implying that the hypervolume under its ROC hypersurface will also converge to zero. In fact, reasoning as above, one can see that the ideal observer will also be unable to achieve operating points within the region $P_i' > P_i$ $(1 \leq i \leq 6)$, since the ideal observer's ROC hypersurface is never above that of any other observer at any given point in the domain of the ROC space.[15] The hypervolume under the ideal observer's ROC hypersurface will thus also converge to zero as the underlying data PDFs are moved apart.

6. CONCLUSIONS

In three-class classification tasks, it can be shown that the hypervolume under the ROC hypersurface of both the "guessing" observer and the "perfect" observer are zero. More importantly, we have shown in each of these performance extremes that the convergence to zero is smooth rather than discontinuous. This convergence can be considered completely general for "near-guessing" observers and generally true for "near-perfect" observers which follow rational decision rules (analogous to false-negative fraction and false-positive fraction being monotonically related in a two-class task); that is, the conclusions appear to hold true for arbitrary underlying data PDFs.

In the two-class classification task, the area under the ROC curve (AUC) is considered a useful performance metric for a variety of reasons. One of the most pleasing and straightforward of these is the simple relationship between AUC and the "separability" of the two underlying data PDFs (*i. e.*, the difficulty of the task). Namely, the AUC (with the two-class ROC defined as a plot of false-negative fraction versus false-positive fraction) of a "perfect" observer is zero, and increases in some sense uniformly as the task is made more difficult, until one arrives at the "guessing" observer with an AUC of 0.5. In a three-class classification task, this straightforward relationship appears to break down, and both "perfect" and "guessing" observers yield ROC hypersurfaces with zero hypervolume. It would appear that, due to this ambiguity, hypervolume under the ROC hypersurface of an three-class observer is not a useful performance metric: Does a hypervolume of 0.005 indicate an observer faced with an exceptionally difficult or exceptionally easy task? One hopes that some other performance metric from two-class classification can be generalized usefully for three-class classification; perhaps a quantity which is equal to AUC in the two-class case has a generalization which is not equal to the hypervolume, but can be shown to be of use for other reasons.

ACKNOWLEDGMENTS

This work was supported in parts by grant R01-CA60187 from the National Cancer Institute (R.M. Nishikawa, principal investigator) and grant R01-GM57622 from the National Institutes of Health (C.E. Metz, principal investigator). Charles E. Metz and Robert M. Nishikawa are shareholders in R2 Technology, Inc. (Sunnyvale, CA). It is the University of Chicago Conflict of Interest Policy that investigators disclose publicly actual or potential significant financial interests which would reasonably appear to be directly and significantly affected by the research activities.

REFERENCES

1. U. Bick, M. L. Giger, R. A. Schmidt, R. M. Nishikawa, D. E. Wolverton, and K. Doi, "Automated segmentation of digitized mammograms," *Acad. Radiol.* **2**, pp. 1–9, 1995.
2. F.-F. Yin, M. L. Giger, K. Doi, C. E. Metz, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images," *Med. Phys.* **18**, pp. 955–963, 1991.
3. F.-F. Yin, M. L. Giger, C. J. Vyborny, K. Doi, and R. A. Schmidt, "Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of mammographic masses," *Invest. Radiol.* **28**, pp. 473–481, 1993.
4. F.-F. Yin, M. L. Giger, K. Doi, C. J. Vyborny, and R. A. Schmidt, "Computerized detection of masses in digital mammograms: Automated alignment of breast images and its effect on bilateral-subtraction technique," *Med. Phys.* **21**, pp. 445–452, 1994.
5. M. A. Kupinski, *Computerized Pattern Classification in Medical Imaging.* Ph.D. thesis, The University of Chicago, Chicago, IL, 2000.
6. Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi, "Automated computerized classification of malignant and benign masses on digitized mammograms," *Acad. Radiol.* **5**, pp. 155–168, 1998.
7. Z. Huo, M. L. Giger, and C. E. Metz, "Effect of dominant features on neural network performance in the classification of mammographic lesions," *Phys. Med. Biol.* **44**, pp. 2579–2595, 1999.
8. Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, and C. E. Metz, "Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness," *Acad. Radiol.* **7**, pp. 1077–1084, 2000.
9. Z. Huo, M. L. Giger, and C. J. Vyborny, "Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis," *IEEE Trans. Med. Imag.* **20**, pp. 1285–1292, 2001.

10. Z. Huo, M. L. Giger, C. J. Vyborny, and C. E. Metz, "Breast cancer: Effectiveness of computer-aided diagnosis — Observer study with independent database of mammograms," *Radiology* **224**, pp. 560–568, 2002.
11. D. J. S. MacKay, *Bayesian Methods for Adaptive Models.* Ph.D. thesis, California Institute of Technology, Pasadena, CA, 1992.
12. M. A. Kupinski, D. C. Edwards, M. L. Giger, and C. E. Metz, "Ideal observer approximation using Bayesian classification neural networks," *IEEE Trans. Med. Imag.* **20**, pp. 886–899, 2001.
13. D. C. Edwards, C. E. Metz, and R. M. Nishikawa, "Estimation of three-class ideal observer decision functions with a Bayesian artificial neural network," in Proc. SPIE Vol. 4686 *Medical Imaging 2002: Image Perception, Observer Performance, and Technology Assessment*, Dev P. Chakraborty and Elizabeth A. Krupinski, eds., pp. 1–12, (SPIE, Bellingham, WA), 2002.
14. 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.*, 2003. (in press).
15. D. C. Edwards, C. E. Metz, and M. A. Kupinski, "Ideal observers and optimal ROC hypersurfaces in N-class classification," *IEEE Trans. Med. Imag.*, 2003. (in review).
16. 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.*, 2003. (in review).
17. H. L. Van Trees, *Detection, Estimation and Modulation Theory: Part I*, John Wiley & Sons, New York, 1968.
18. S. I. Grossman, *Multivariable Calculus, Linear Algebra, and Differential Equations: Second Edition*, Harcourt Brace Jovanovich, San Diego, CA, 1986.