Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions

Harrison H. Barrett, Craig K. Abbey, and Eric Clarkson

Department of Radiology, Optical Sciences Center and Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85724 5067

Received October 30, 1997; accepted January 22, 1998; revised manuscript received February 18, 1998

We continue the theme of previous papers [J. Opt. Soc. Am. A **7**, 1266 (1990); **12**, 834 (1995)] on objective (task-based) assessment of image quality. We concentrate on signal-detection tasks and figures of merit related to the ROC (receiver operating characteristic) curve. Many different expressions for the area under an ROC curve (AUC) are derived for an arbitrary discriminant function, with different assumptions on what information about the discriminant function is available. In particular, it is shown that AUC can be expressed by a principal-value integral that involves the characteristic functions of the discriminant. Then the discussion is specialized to the ideal observer, defined as one who uses the likelihood ratio (or some monotonic transformation of it, such as its logarithm) as the discriminant function. The properties of the ideal observer are examined from first principles. Several strong constraints on the moments of the likelihood ratio or the log likelihood are derived, and it is shown that the probability density functions for these test statistics are intimately related. In particular, some surprising results are presented for the case in which the log likelihood is normally distributed under one hypothesis. To unify these considerations, a new quantity called the likelihood-generating function is defined. It is shown that all moments of both the likelihood and the log likelihood under both hypotheses can be derived from this one function. Moreover, the AUC can be expressed, to an excellent approximation, in terms of the likelihood-generating function evaluated at the origin. This expression is the leading term in an asymptotic expansion of the AUC; it is exact whenever the likelihood-generating function behaves linearly near the origin. It is also shown that the likelihood-generating function at the origin sets a lower bound on the AUC in all cases. [S0740-3232(98)02106-1]

OCIS codes: 110.3000, 110.4280, 110.5010.

1. INTRODUCTION

This is the third in a series of papers[1,2] on the theoretical basis of objective assessment of image quality (OAIQ). Fundamental to this theory is the idea that image quality must be defined by the performance of some observer on some specific task. The tasks of interest can be broadly categorized as classification and estimation. An important special case of classification is binary hypothesis testing. In image analysis it is often necessary to detect some object or abnormality, and the two hypotheses can be described as signal absent and signal present; in that case the task is called signal detection.

Previous papers in this series have considered signal-detection tasks as one possible basis for a definition of image quality. The tasks have, in principle, allowed for a degree of randomness in both signal and background, but the theory has been restricted to linear observers, i.e., ones that calculate a discriminant function that is a linear functional of the data. It is well known that linear observers are suboptimal for detecting random signals, so the theory presented previously does not accurately reflect the maximum achievable performance on such detection tasks. The figure of merit used was a signal-to-noise ratio (SNR) based on first- and second-order statistics of the discriminant function, and there was little discussion of how it related to other possible figures of merit.

This paper also deals with signal detection and other binary hypothesis testing, but the performance assessment is based on a construct from radar theory called the receiver operating characteristic, or ROC curve. The ROC curve depicts the trade-off between the probability of detection (true-positive fraction) and the false-alarm rate (false-positive fraction) as the decision threshold is varied. The area under the ROC curve (AUC) is often advocated as an overall figure of merit for the task.

For a normally distributed discriminant function, the relation between the AUC and the SNR used in previous papers in the series is well understood, but real-world images are seldom normal, and even if we model the image statistics by a multivariate normal (Gaussian) density, with nonlinear discriminants the normality may be lost.

It is well known, and shown in Section 2, that the ideal discriminant function is the likelihood ratio. Since monotonic transformations of the discriminant function do not change a decision, the logarithm of the likelihood ratio, or log likelihood for short, can also be used. A decision strategy based on the likelihood or log likelihood is referred to as the ideal observer. It sets an upper limit to the performance of any observer on a particular detection task and thus gives a measure of the quality of image data that is uncomplicated by consideration of the limitations of humans or other suboptimal observers. On the other hand, the ideal observer must perform nonlinear operations on the image data in almost all cases, so it has been difficult to compute its performance.

In Section 3 we develop a general theory of the ROC curve that is applicable to an arbitrary linear or nonlinear discriminant function. We derive several new expressions for the AUC, depending on what is known about the statistical properties of the discriminant function.

Then in Section 4 we apply this theory to the ideal observer. First we examine the statistical properties of the log likelihood and demonstrate that there are strong constraints on the form of its probability density functions under the two hypotheses. This leads to definition of the likelihood-generating function, a function of one scalar variable from which all moments of both the likelihood and the log likelihood can be derived under both hypotheses. Moreover, the value of the likelihood-generating function at the origin will be shown to play a key role in determining the AUC.

2. BACKGROUND

A. Image Data

A digital image consists of a set of M real numbers, often called gray levels or pixel values. We can arrange these values as an $M \times 1$ column vector $\mathbf{g}$, with the mth component being the gray level associated with pixel m. We regard each g_m as a continuous random variable and thus $\mathbf{g}$ as a continuous random vector. In practice, of course, the gray levels are quantized by analog-to-digital conversion or the discrete number of photons contributing to a pixel value, but these effects are not treated here. In addition, real gray levels may be restricted to positive values, since they correspond to irradiances, but it is nevertheless convenient to take the range of each g_m as $(-\infty, \infty)$; the positivity restriction is then contained in the probability density function. We also assume that the sum of the squares of the elements of $\mathbf{g}$ is finite, so each image is a vector in an M-dimensional Euclidean space, which we refer to as data space.

An image is related to an object by the action of an imaging system, which can include image-forming elements, image sensors, and image-reconstruction algorithms or other computer processing. We do not consider any of these items here in detail. The viewpoint is that the imaging system delivers an image vector $\mathbf{g}$, which we shall use to perform a specific task, and we need only to know the statistical properties of $\mathbf{g}$ to assess the quality of the imaging system.

We have already begun to describe these statistical properties in detail in the first paper in this series (Ref. 1, designated OAIQ I), which dealt with measurement noise and object variability, and in two papers on statistical properties of the expectation–maximization algorithm.[3,4] We shall continue the process in a later paper with a discussion of more general nonlinear reconstruction algorithms and their effects on image quality.

B. Binary Decisions and ROC Curves

For each image we must decide between two hypotheses, H_0 and H_1. For definiteness we refer to these hypotheses as, respectively, signal absent and signal present, but the mathematics applies as well to discrimination between two signals or two object classes. Moreover, we shall allow wide latitude in what is considered to be a signal; the signal is whatever component distinguishes H_1 from H_0.

Initially we impose only two restrictions on the decision strategy: it must be nonrandom (a particular image must always lead to the same decision), and it must not allow equivocation (every image must lead to some decision, either signal present or signal absent). These conditions imply that decision making is equivalent to partitioning the data space. For all data vectors $\mathbf{g}$ in one subspace Γ_1 the decision will be that the signal is present, and for its orthogonal complement Γ_0 the decision will be signal absent. Devising a decision strategy is equivalent to defining Γ_1.[5]

The decision strategy can also be expressed as a two-step process: first compute a discriminant function $t = \theta(\mathbf{g})$; then compare it to a threshold x. If $t \geqslant x$, choose hypothesis H_1; otherwise choose H_0. This process is equivalent to partitioning data space, since the contours $\theta(\mathbf{g}) = x$ are the surfaces dividing Γ_0 from Γ_1. In this view the decision strategy amounts to computing a (generally nonlinear) discriminant function $\theta(\mathbf{g})$ and comparing it to a threshold.

There are four possible outcomes for each individual decision. If the decision is signal present and it really is present, the decision is a true positive (TP), while a decision of signal present for an image with no signal is a false positive (FP). The conditional probability of a positive decision, given that the signal is actually present, is called the true-positive fraction, or TPF. In the medical literature, TPF is called sensitivity; in the radar literature, it is the probability of detection.

True negatives (TN) and false negatives (FN) and associated fractions (TNF and FNF) are defined similarly. In radar, FPF is called the false-alarm rate. In medicine, TNF (which is the same as $1 - \mathrm{FPF}$) is called the specificity.

The TPF at threshold x is given by

$$\mathrm{TPF}(x) = \Pr(t \geqslant x | H_1) = \int_x^\infty \mathrm{d}t\, \mathrm{pr}(t|H_1), \quad (2.1)$$

where $\Pr(t \geqslant x | H_1)$ is the probability that $t \geqslant x$ and $\mathrm{pr}(t|H_1)$ is the probability density function of the continuous random variable t; both of these quantities are conditional on hypothesis H_1 being true or on the signal actually being present. In general, we use $\Pr(\cdot)$ for probabilities and $\mathrm{pr}(\cdot)$ for probability density functions, though other notations are also introduced for the latter as needed.

The FPF at threshold x is given by

$$\mathrm{FPF}(x) = \Pr(t \geqslant x | H_0) = \int_x^\infty \mathrm{d}t\, \mathrm{pr}(t|H_0). \quad (2.2)$$

The threshold x controls the trade-off between TPF and FPF. Graphically, this trade-off is portrayed by the ROC curve, which is a plot of $\mathrm{TPF}(x)$ versus $\mathrm{FPF}(x)$.

C. Optimal Decisions

Most approaches to decision theory start by defining a cost or loss function. For a binary decision problem, the cost is a 2×2 matrix $\mathbf{C}$, where element C_{ij} ($i, j = 0$ or 1) is the cost of making decision D_i when hypothesis H_j is true. Positive costs will usually be assigned to incorrect decisions and zero or negative costs to correct ones.

The risk is the average value of the cost, but Bayesian and frequentist approaches to decision theory[5–7] differ in how this average is computed. The frequentist would average over many realizations of the data for the same true hypothesis, thereby getting separate risks for the signal-absent and signal-present conditions. The pure Bayesian, on the other hand, would average with respect to (possibly subjective) prior probabilities on the hypotheses, saying nothing about other realizations of the data. This procedure leads to the Bayesian expected loss,[6] which is a function of $\mathbf{g}$.

A hybrid approach, used in many books and adopted here, is to average over prior probabilities *and* data, by use of both $\Pr(H_j)$ and $\mathrm{pr}(\mathbf{g}|H_j)$. These probabilities are intended to be objective ones, readily interpretable in a sampling sense. To interpret $\Pr(H_1)$, for example, we envision repeated experiments in which the signal is present in this fraction of the repetitions. Devout Bayesians often object to this interpretation, but it is justified here since we can actually do such experiments, under controlled conditions with specified frequencies of occurrence. Such experiments would form the basis for an empirical measure of image quality, and the quantities computed here should be regarded as long-run, frequentist averages of those empirical measures. In fact, the ROC curve itself is a frequentist concept, even when it applies to a discriminant function derived from a Bayesian viewpoint. The ROC curve is a way of keeping score over repeated trials, and real Bayesians do not keep score.

The double averaging yields a quantity often called the Bayes risk, which can be viewed as a frequentist average of the Bayesian expected loss. The Bayes risk is given by

$$R = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} \Pr(D_i, H_j) = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} \Pr(D_i|H_j) \Pr(H_j). \tag{2.3}$$

Since decision D_i is a certainty for $\mathbf{g}$ in region Γ_i, the Bayes risk can also be written as

$$R = \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} \Pr(H_j) \int_{\Gamma_i} \mathrm{d}^M g\, \mathrm{pr}(\mathbf{g}|H_j). \tag{2.4}$$

The objective is to minimize R through choice of Γ_0 and Γ_1.

Following van Trees,[5] we rewrite Eq. (2.4) as an integral over Γ_1 alone. Since the decision rule does not allow equivocation, Γ_0 and Γ_1 together constitute all of the data space, and we must have

$$\int_{\Gamma_0} \mathrm{d}^M g\, \mathrm{pr}(\mathbf{g}|H_j) + \int_{\Gamma_1} \mathrm{d}^M g\, \mathrm{pr}(\mathbf{g}|H_j) = 1. \tag{2.5}$$

With this condition and a little algebra, Eq. (2.4) becomes

$$\begin{aligned} R = {} & C_{01}\Pr(H_1) + C_{00}\Pr(H_0) \\ & + \int_{\Gamma_1} \mathrm{d}^M g \Big[C_{11}\Pr(H_1)\mathrm{pr}(\mathbf{g}|H_1) \\ & + C_{10}\Pr(H_0)\mathrm{pr}(\mathbf{g}|H_0) - C_{01}\Pr(H_1)\mathrm{pr}(\mathbf{g}|H_1) \\ & - C_{00}\Pr(H_0)\mathrm{pr}(\mathbf{g}|H_0) \Big]. \end{aligned} \tag{2.6}$$

The first two terms are constants, independent of $\mathbf{g}$, and hence do not affect the decision strategy. To minimize the integral, we must choose Γ_1 to include all portions of data space for which the integrand is negative and exclude those for which it is positive. A point $\mathbf{g}$ is then in Γ_1 if

$$(C_{00} - C_{10})\Pr(H_0)\mathrm{pr}(\mathbf{g}|H_0) > (C_{11} - C_{01})\mathrm{pr}(\mathbf{g}|H_1)\Pr(H_1). \tag{2.7}$$

In other words, we make decision D_1 if[8]

$$\frac{\mathrm{pr}(\mathbf{g}|H_1)}{\mathrm{pr}(\mathbf{g}|H_0)} > \frac{(C_{10} - C_{00})\Pr(H_0)}{(C_{01} - C_{11})\Pr(H_1)}. \tag{2.8}$$

The left-hand side of this relation is the likelihood ratio, defined by

$$\Lambda(\mathbf{g}) = \frac{\mathrm{pr}(\mathbf{g}|H_1)}{\mathrm{pr}(\mathbf{g}|H_0)}. \tag{2.9}$$

Thus the optimal discriminant function is the likelihood ratio, and the optimal threshold is

$$x = \frac{(C_{10} - C_{00})\Pr(H_0)}{(C_{01} - C_{11})\Pr(H_1)}. \tag{2.10}$$

We refer to any decision strategy based on computing the likelihood ratio and comparing it to a threshold as the ideal observer. Equivalently, the ideal observer can compute the logarithm of Λ, called the log likelihood λ, and compare it to the logarithm of the threshold. Since the logarithm is a monotonic function, exactly the same decision is reached with either Λ or λ, and the latter is frequently computationally easier.

3. FIGURES OF MERIT

A. General Considerations and Definitions

The Bayes risk is one possible figure of merit for binary decision problems. If the ideal observer, using data from system A, achieves a lower risk on a particular problem than it would using data from system B, system A can be ranked higher, at least for this task and observer. The risk value is, however, difficult to interpret. There are no generally accepted rules for defining costs, or even any standard units of measure for them.

Moreover, Bayes risk is seldom useful for suboptimal observers. For example, linear discriminants are frequently computationally tractable[9] when we do not have sufficient information to compute the Bayesian discriminant function or the Bayes risk. Another rationale for certain linear observers is that they may be good models of human observers.[10] In this case the utility of the observer must be assessed by its correlation with psychophysical data.

An alternative figure of merit, applicable to both optimal and suboptimal observers, is the TPF at some specified FPF (or, in radar terminology, the probability of detection at a specified false-alarm rate). This figure of merit is called the Neyman–Pearson criterion. Its computation requires knowledge of the probability law for the discriminant function under the two hypotheses but not priors or costs.

Both the Bayes risk and TPF at a specified FPF depend not only on the task and the quality of the data but also on the chosen operating point on the ROC curve. The operating point, however, is fairly arbitrary; different users of the image data will assign different costs and priors and hence use different operating points. For this reason, many workers in signal detection and image quality advocate using the entire ROC curve as the quality metric. A common scalar figure of merit is the area under the ROC curve, denoted AUC. AUC varies from 0.5 for a worthless system to 1.0 for a system that allows the task to be performed perfectly.

Since the likelihood ratio is the optimum discriminant function for any particular choice of operating point on the ROC curve, it gives the maximum TPF at any specified FPF, and it can also be shown[11] to maximize the area under the ROC curve. Thus the likelihood observer is ideal with respect to each of these figures of merit.

Another figure of merit, which can be defined for any discriminant function t, is the SNR, given by

$$\mathrm{SNR}_t{}^2 \equiv \frac{[\langle t\rangle_1 - \langle t\rangle_0]^2}{\frac{1}{2}\mathrm{var}_1(t) + \frac{1}{2}\mathrm{var}_0(t)}, \tag{3.1}$$

where $\langle t\rangle_j$ denotes the conditional expectation of t, given that H_j is true and $\mathrm{var}_j(\,\cdot\,)$ denotes the corresponding conditional variance.

If t is normally distributed under both hypotheses, it is well known (and shown in Subsection 3.D) that SNR_t is related to AUC by

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\mathrm{SNR}_t}{2}\right), \tag{3.2}$$

where $\mathrm{erf}(\,\cdot\,)$ is the error function. Thus in this case there is a simple monotonic relation between AUC and SNR_t, so it does not matter which we adopt as a figure of merit.

We can define another SNR simply by inverting Eq. (3.2):

$$\mathrm{SNR(AUC)} \equiv 2\,\mathrm{erf}^{-1}(2\,\mathrm{AUC} - 1), \tag{3.3}$$

where $\mathrm{erf}^{-1}(\,\cdot\,)$ is the inverse of the error function. In the literature, SNR(AUC) is often referred to as d_a or d_A. Of course, knowledge of SNR(AUC) allows exact computation of AUC, but it is difficult to compute SNR(AUC) from first principles without normality assumptions. On the other hand, the results of psychophysical studies of human-observer performance are almost universally reported in terms of AUC or SNR(AUC), so methods of computing these metrics for the ideal observer are needed if we wish to see how nearly the human approximates the ideal.

In the remainder of this section we consider various ways of computing the AUC, depending on what knowledge we have of the statistics of the problem.

B. Discriminant Function with Known Probability Law

Suppose that we are given a discriminant function $t = \theta(\mathbf{g})$ and that we know its densities $\mathrm{pr}(t|H_0)$ and $\mathrm{pr}(t|H_1)$. For notational convenience we define

$$\mathrm{pr}(t|H_j) \equiv p_j(t). \tag{3.4}$$

The area under the ROC curve is given by

$$\mathrm{AUC} = \int_0^1 \mathrm{TPFd(FPF)}, \tag{3.5}$$

where $\mathrm{TPF}(x)$ and $\mathrm{FPF}(x)$ are given by Eqs. (2.1) and (2.2), respectively. Since FPF is a monotonic function of x, we can change the variable of integration from $\mathrm{FPF}(x)$ to x, thus obtaining

$$\mathrm{AUC} = -\int_{-\infty}^{\infty} \mathrm{d}x\ \mathrm{TPF}(x)\,\frac{\mathrm{d}}{\mathrm{d}x}\,\mathrm{FPF}(x) \tag{3.6}$$

where the minus sign arises, since $\mathrm{FPF}(x) \to 1$ as $x \to -\infty$.

From Eq. (2.2) and Leibniz's rule, we have

$$\frac{\mathrm{d}}{\mathrm{d}x}\,\mathrm{FPF}(x) = -p_0(x), \tag{3.7}$$

so

$$\mathrm{AUC} = \int_{-\infty}^{\infty} \mathrm{d}x p_0(x) \int_x^{\infty} \mathrm{d}t p_1(t). \tag{3.8}$$

There are various ways of rewriting this expression. One is to recognize that the cumulative probability distribution function of t under H_1 is given by

$$P_1(x) \equiv \Pr(t < x|H_1) = \int_{-\infty}^{x} \mathrm{d}t p_1(t) = 1 - \int_x^{\infty} \mathrm{d}t p_1(t). \tag{3.9}$$

Thus

$$\mathrm{AUC} = 1 - \int_{-\infty}^{\infty} \mathrm{d}x p_0(x) P_1(x). \tag{3.10}$$

Another form for AUC is obtained by use of the step function to rewrite Eq. (3.8) as

$$\mathrm{AUC} = \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}t p_0(x) p_1(t)\mathrm{step}(t - x). \tag{3.11}$$

With a change of variables $y = t - x$, we obtain

$$\begin{aligned}\mathrm{AUC} &= \int_{-\infty}^{\infty} \mathrm{d}x \int_{-\infty}^{\infty} \mathrm{d}y p_0(x) p_1(y + x)\mathrm{step}(y) \\ &= \int_0^{\infty} \mathrm{d}y [p_0 \otimes p_1](y),\end{aligned} \tag{3.12}$$

where $\otimes$ denotes a one-dimensional correlation integral. Computation of AUC by this formula thus requires cross correlating p_0 and p_1 and then integrating the result from 0 to ∞.

We can also use the signum function, related to the step by

$$\mathrm{step}(x) = \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{sgn}(x), \tag{3.13}$$

so Eq. (3.11) becomes

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}t p_0(x)p_1(t)\mathrm{sgn}(t-x). \tag{3.14}$$

The step function can be expressed in terms of its Fourier transform,

$$\mathscr{F}\{\mathrm{step}(x)\} = \frac{1}{2}\,\delta(\xi) + \mathscr{P}\left\{\frac{1}{2\pi i\xi}\right\}, \tag{3.15}$$

from which we easily find

$$\mathrm{step}(x) = \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\exp(2\pi i\xi x), \tag{3.16}$$

where $\mathscr{P}$ indicates that the singular integral must be interpreted as a Cauchy principal value. With Eq. (3.16), Eq. (3.11) becomes

$$\begin{aligned}\mathrm{AUC} &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}t p_0(x)\\ &\quad\times p_1(t)\exp[2\pi i\xi(t-x)]\\ &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\,\psi_0(\xi)\psi_1^*(\xi),\end{aligned} \tag{3.17}$$

where $\psi_j(\xi)$ is the characteristic function for t under hypothesis H_j. Specifically,

$$\begin{aligned}\psi_j(\xi) &\equiv \langle\exp(-2\pi i\xi t)\rangle_j\\ &= \int_{-\infty}^{\infty}\mathrm{d}t p_j(t)\exp(-2\pi i\xi t)\\ &= \mathscr{F}\{p_j(t)\}.\end{aligned} \tag{3.18}$$

C. Discriminant Function with Known Moments

In the derivation in Subsection 3.B we assumed that the densities $p_j(t)$ were known, but we rarely have that much information about a test statistic. Suppose now that we know only a few low-order moments of t under the two hypotheses.

The moments can, of course, be related to derivatives of the characteristic function, but it is somewhat more convenient to use the moment-generating function, defined by

$$M_j(\beta) = \langle\exp(\beta t)\rangle_j = \int_{-\infty}^{\infty}\mathrm{d}t p_j(t)\exp(\beta t) = \psi_j\left(\frac{i\beta}{2\pi}\right). \tag{3.19}$$

If we think of ξ and β as complex variables, the functions $\psi_j(\xi)$ and $M_j(\beta)$ are related by a 90° rotation in the complex plane and a scaling of the argument by 2π.

Except for an unconventional sign in the exponent, $M_j(\beta)$ is the two-sided Laplace transform[12] of $p_j(t)$. The integral in Eq. (3.19) converges for complex β in a strip parallel to and including the imaginary axis. The strip is defined by $-c_1 < \mathrm{Re}\,\beta < c_2$, where Re denotes real part and c_1 and c_2 are some positive constants. In Appendix A we show by use of the Cauchy–Riemann conditions that $M_j(\beta)$ is an analytic function in this strip.

A related function is the cumulant-generating function, which is simply the logarithm of the moment-generating function:

$$L_j(\beta) = \log[M_j(\beta)]. \tag{3.20}$$

Since $M_j(\beta)$ is analytic at the origin, so is its logarithm, and $L_j(\beta)$ can therefore be expanded in a Taylor series about the origin (Maclaurin series) as

$$L_j(\beta) = \sum_{n=0}^{\infty}\frac{1}{n!}L_j^{(n)}(0)\beta^n, \tag{3.21}$$

where $L_j^{(n)}(\beta)$ denotes the nth derivative of $L_j(\beta)$. Derivatives of $L_j(\beta)$ (known as cumulants) are related to derivatives of $M_j(\beta)$, which in turn are related to moments of t by derivatives of the moment-generating function:

$$\langle t^n\rangle_j = M^{(n)}(0). \tag{3.22}$$

The first five coefficients in Eq. (3.21) are

$$L_j^{(0)}(0) = \log[M_j(0)] = 0, \tag{3.23a}$$

$$L_j^{(1)}(0) = \frac{M_j^{(1)}(0)}{M_j(0)} = \langle t\rangle_j \equiv \bar{t}_j, \tag{3.23b}$$

$$L_j^{(2)}(0) = \langle(t-\bar{t}_j)^2\rangle_j \equiv {\sigma_j}^2, \tag{3.23c}$$

$$L_j^{(3)}(0) = \langle(t-\bar{t}_j)^3\rangle_j \equiv {\sigma_j}^3 S_j, \tag{3.23d}$$

$$L_j^{(4)}(0) = [\langle(t-\bar{t}_j)^4\rangle_j - 3{\langle(t-\bar{t}_j)^2\rangle_j}^2] \equiv {\sigma_j}^4 K_j. \tag{3.23e}$$

Here $\bar{t}_j$, ${\sigma_j}^2$, S_j, and K_j are, respectively, the mean, variance, skewness, and kurtosis of $p_j(t)$. Different definitions of kurtosis appear in the literature, but with the one used here, $K = 0$ for a Gaussian.

With these moments, $\psi_j(\xi)$ can be written as

$$\psi_j(\xi) = M_j(-2\pi i\xi) = \exp\left[-2\pi i\bar{t}_j\xi - 2\pi^2{\sigma_j}^2\xi^2 + i\,\frac{4\pi^3}{3}{\sigma_j}^3 S_j\xi^3 + \frac{2\pi^4}{3}{\sigma_j}^4 K_j\xi^4 + \ldots\right]. \tag{3.24}$$

From Eqs. (3.17) and (3.24), AUC is given by

$$\begin{aligned}\mathrm{AUC} &= \frac{1}{2} + \frac{1}{2\pi}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\\ &\quad\times\sin\left[2\pi(\bar{t}_1-\bar{t}_0)\xi - \frac{4\pi^3}{3}({\sigma_1}^3 S_1 - {\sigma_0}^3 S_0)\xi^3 + \cdots\right]\exp[-2\pi^2({\sigma_0}^2+{\sigma_1}^2)\xi^2\\ &\quad + (2\pi^4/3)({\sigma_1}^4 K_1 + {\sigma_0}^4 K_0)\xi^4 + \ldots].\end{aligned} \tag{3.25}$$

Care must be used in truncating this expansion. For example, suppose that the term in ξ^4 has a negative coefficient but the one in ξ^6 has a positive coefficient; then the integral would converge if terms through ξ^4 were retained but not with terms through ξ^6.

D. Univariate Normal Statistics

If $p_0(t)$ and $p_1(t)$ are both univariate normal (not necessarily with the same variance), then the skewness and kurtosis of each are zero, and all higher terms in the expansion (3.24) vanish identically. In that case, the moment-generating and characteristic functions are given by, respectively,

$$M_j(\beta) = \exp(\overline{t_j}\beta + \tfrac{1}{2}\sigma_j{}^2\beta^2), \tag{3.26}$$

$$\psi_j(\xi) = \exp(-2\pi i\overline{t_j}\xi - 2\pi^2\sigma_j{}^2\xi^2). \tag{3.27}$$

Note that $\psi_j(\xi)$ falls off rapidly as $\mathrm{Re}\,\xi \to \pm\infty$, but $M_j(\beta)$ blows up as $\mathrm{Re}\,\beta \to \pm\infty$.

From Eq. (3.25), we now have

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2\pi}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\sin[2\pi(\overline{t_1} - \overline{t_0})\xi] \times \exp[-2\pi^2(\sigma_0{}^2 + \sigma_1{}^2)\xi^2]. \tag{3.28}$$

The principal value can be implemented as a limit:

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2\pi}\lim_{\epsilon\to 0}\int_{-\infty}^{\infty}\mathrm{d}\xi\,\frac{\xi}{\xi^2 + \epsilon}\sin[2\pi(\overline{t_1} - \overline{t_0})\xi] \exp[-2\pi^2(\sigma_0{}^2 + \sigma_1{}^2)\xi^2]. \tag{3.29}$$

A tabulated integral[13] then yields the error-function relation, Eq. (3.2).

E. Arbitrary Test Statistic, Unknown Probability Law

When t is a complicated function of $\mathbf{g}$, we may know neither its densities nor its moments. We may, however, know $\mathrm{pr}(\mathbf{g}|H_j)$ from the basic physics of the image-forming process and from knowledge of the signal and background.[14] In those cases we can express AUC in terms of integrals over $\mathbf{g}$ rather than over t.

Even though t is a specific function $\theta(\mathbf{g})$, it is useful to regard this function as a probabilistic mapping. As a formal device[15] we can write

$$\mathrm{pr}(t|\mathbf{g}) = \delta[t - \theta(\mathbf{g})], \tag{3.30}$$

so that

$$\begin{aligned}\mathrm{pr}(t|H_j) &= \int_\infty \mathrm{d}^M g\,\mathrm{pr}(t|\mathbf{g})\mathrm{pr}(\mathbf{g}|H_j) \\ &= \int_\infty \mathrm{d}^M g\,\mathrm{pr}(\mathbf{g}|H_j)\delta[t - \theta(\mathbf{g})].\end{aligned} \tag{3.31}$$

A shorthand form for this expression is obtained if we use the notation of Eq. (3.4) and also define

$$q_j(\mathbf{g}) \equiv \mathrm{pr}(\mathbf{g}|H_j). \tag{3.32}$$

Then Eq. (3.31) is

$$p_j(t) = \int_\infty \mathrm{d}^M g\,q_j(\mathbf{g})\delta[t - \theta(\mathbf{g})]. \tag{3.33}$$

The delta function defines an $(M - 1)$-dimensional surface in the M-dimensional data space; all points on this surface have $\theta(\mathbf{g}) = t$ and hence contribute to the probability density on t at the same t.

From Eqs. (3.11) and (3.33) we have

$$\mathrm{AUC} = \int_{-\infty}^{\infty}\mathrm{d}x\int_{-\infty}^{\infty}\mathrm{d}t\int_\infty \mathrm{d}^M g\,q_0(\mathbf{g})\delta[x - \theta(\mathbf{g})] \times \int_\infty \mathrm{d}^M g'\,q_1(\mathbf{g}')\delta[t - \theta(\mathbf{g}')]\mathrm{step}(t - x). \tag{3.34}$$

The delta functions allow us to perform the integrals over t and x, with the result

$$\mathrm{AUC} = \int_\infty \mathrm{d}^M g\int_\infty \mathrm{d}^M g'\,q_0(\mathbf{g})q_1(\mathbf{g}')\mathrm{step}[\theta(\mathbf{g}') - \theta(\mathbf{g})]. \tag{3.35}$$

This expression demonstrates that AUC is unchanged by a monotonic point transformation. If we replace $\theta(\mathbf{g})$ with $\Theta(\mathbf{g}) = h[\theta(\mathbf{g})]$, where $h(x)$ is a monotonically increasing function, then the step function remains unchanged: if $\mathrm{step}[\theta(\mathbf{g}') - \theta(\mathbf{g})] = 1$ for some set of values of $\mathbf{g}'$ and $\mathbf{g}$, then $\mathrm{step}[\Theta(\mathbf{g}') - \Theta(\mathbf{g})] = 1$ for precisely this same set.

As in Subsection 3.B, we can represent the step function by means of its Fourier transform and obtain

$$\begin{aligned}\mathrm{AUC} &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\int_\infty \mathrm{d}^M g\int_\infty \mathrm{d}^M g'\,q_0(\mathbf{g}) \times q_1(\mathbf{g}')\exp\{2\pi i\xi[\theta(\mathbf{g}') - \theta(\mathbf{g})]\} \\ &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\langle\exp[-2\pi i\xi\theta(\mathbf{g})]\rangle_0 \times \langle\exp[2\pi i\xi\theta(\mathbf{g}')]\rangle_1 \\ &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\psi_0(\xi)\psi_1^*(\xi).\end{aligned} \tag{3.36}$$

The final form of Eq. (3.36) is identical to Eq. (3.17); we have taken advantage of the fact that the expectations can be computed from either the probability density on $\mathbf{g}$ or the one on $\theta(\mathbf{g})$.

F. Two-Alternative Forced-Choice Interpretations

Equations (3.11) and (3.35) can be interpreted in terms of two-alternative forced-choice (2AFC) experiments. In a 2AFC experiment, two independent data vectors $\mathbf{g}$ and $\mathbf{g}'$ are generated, with $\mathbf{g}$ drawn from $\mathrm{pr}(\mathbf{g}|H_0)$ and $\mathbf{g}'$ drawn from $\mathrm{pr}(\mathbf{g}'|H_1)$. Two test statistics $\theta(\mathbf{g})$ and $\theta(\mathbf{g}')$ are computed, and the data vector that gives the higher value is assigned to H_1. This assignment is correct if $\theta(\mathbf{g}') > \theta(\mathbf{g})$. Thus the probability of a correct decision is

$$\begin{aligned}\Pr(\mathrm{correct}) &= \Pr[\theta(\mathbf{g}') > \theta(\mathbf{g})] \\ &= \int_\infty \mathrm{d}^M g\int_\infty \mathrm{d}^M g'\,q_0(\mathbf{g})q_1(\mathbf{g}')\mathrm{step}[\theta(\mathbf{g}') - \theta(\mathbf{g})],\end{aligned} \tag{3.37}$$

which, by Eq. (3.35), is AUC.

A similar interpretation applies to Eq. (3.11). If we denote the test statistics by $x = \theta(\mathbf{g})$ and $t = \theta(\mathbf{g}')$, the 2AFC decision is correct if $t > x$ and Eq. (3.11) gives the probability of this event occurring.

G. Linear Discriminants

Now suppose that $t = \theta(\mathbf{g})$ is a linear function of the data, so we can write it as

$$\theta(\mathbf{g}) = \mathbf{w}^t\mathbf{g}, \tag{3.38}$$

where $\mathbf{w}$ is an $M \times 1$ vector and $\mathbf{w}^t\mathbf{g}$ is the scalar product of $\mathbf{w}$ and $\mathbf{g}$.

With this form of the test statistic, Eq. (3.35) becomes

$$\mathrm{AUC}_{\mathrm{lin}} = \int_\infty \mathrm{d}^M g \int_\infty \mathrm{d}^M g' q_0(\mathbf{g}) q_1(\mathbf{g}') \mathrm{step}[\mathbf{w}^t(\mathbf{g}' - \mathbf{g})]. \tag{3.39}$$

The change of variables $\mathbf{g}'' = \mathbf{g}' - \mathbf{g}$ yields

$$\begin{aligned}\mathrm{AUC}_{\mathrm{lin}} &= \int_\infty \mathrm{d}^M g'' \int_\infty \mathrm{d}^M g q_0(\mathbf{g}) q_1(\mathbf{g} + \mathbf{g}'') \mathrm{step}(\mathbf{w}^t\mathbf{g}'') \\ &= \int_\infty \mathrm{d}^M g'' [q_0 \otimes q_1](\mathbf{g}'') \mathrm{step}(\mathbf{w}^t\mathbf{g}''),\end{aligned} \tag{3.40}$$

where $[q_0 \otimes q_1](\mathbf{g}'')$ denotes a multidimensional cross-correlation integral with shift $\mathbf{g}''$. This equation shows that the AUC can be found by cross correlating q_0 and q_1 and then integrating the result over the half-space $\mathbf{w}^t\mathbf{g} > 0$.

The similarity in form between Eqs. (3.12) and (3.40) should be noted; Eq. (3.12) holds for an arbitrary discriminant function (but requires the probability densities for that function), while Eq. (3.40) holds specifically for a linear discriminant and requires knowledge of the data densities.

With a linear discriminant, we can also relate AUC to the multivariate characteristic functions for $\mathbf{g}$, defined by

$$\Psi_j(\boldsymbol{\rho}) = \int_\infty \mathrm{d}^M g p(\mathbf{g}|H_j) \exp(-2\pi i \boldsymbol{\rho}^t \mathbf{g}), \tag{3.41}$$

where $j = 0$ or 1 and $\boldsymbol{\rho}$ is an M-dimensional frequency vector conjugate to the data vector $\mathbf{g}$.

From the definition of $\psi_j(\xi)$ in Eq. (3.18), along with Eq. (3.38), we have

$$\psi_j(\xi) = \int_\infty \mathrm{d}^M g p(\mathbf{g}|H_j) \exp(-2\pi i \xi \mathbf{w}^t \mathbf{g}) = \Psi_j(\mathbf{w}\xi), \tag{3.42}$$

so that Eq. (3.36) becomes

$$\mathrm{AUC}_{\mathrm{lin}} = \frac{1}{2} + \frac{1}{2\pi i} \mathscr{P} \int_{-\infty}^{\infty} \frac{\mathrm{d}\xi}{\xi} \Psi_0(\mathbf{w}\xi) \Psi_1^*(\mathbf{w}\xi). \tag{3.43}$$

Thus all we need to know to compute AUC for a linear discriminant is the behavior of the characteristic functions of the data along a line through the origin and parallel to $\mathbf{w}$ in the M-dimensional Fourier space. This statement is equivalent to saying that all we need are integrals of the data densities q_0 and q_1 over $(M - 1)$-dimensional hyperplanes normal to $\mathbf{w}$. With nonlinear discriminants we need integrals over $(M - 1)$-dimensional surfaces defined by $\theta(\mathbf{g}) =$ constant.

In practice, however, we can often get by with even less information for computing $\mathrm{AUC}_{\mathrm{lin}}$. A linear discriminant $\mathbf{w}^t\mathbf{g}$ is univariate normal if $\mathbf{g}$ is multivariate normal, so SNR_t is easily computed and the conditions for the use of Eq. (3.2) are exactly satisfied. Moreover, even if $\mathbf{g}$ is not normally distributed, we can often appeal to the central-limit theorem and show that $\mathbf{w}^t\mathbf{g}$ is approximately univariate normal anyway. In summary, it is usually safe to compute AUC from SNR by Eq. (3.2) for a linear discriminant.

4. IDEAL OBSERVER

As we saw in Section 2, the discriminant function for the ideal observer can be either the likelihood ratio $\Lambda(\mathbf{g})$ or its logarithm $\lambda(\mathbf{g})$. To compute the ideal-observer AUC with the formalism of Section 3, we need the probability density functions of either Λ or λ under both hypotheses. Equivalently, we can also get AUC from the characteristic functions or moment-generating functions. If these exact specifications prove difficult to get, we can attempt to compute some low-order moments of Λ or λ and relate them approximately to AUC.

With all of these approaches, the ideal observer differs fundamentally from other observers because its discriminant function (whether Λ or λ) already contains all of the relevant statistical information about the task. As we shall see, this fact imposes strong constraints on the forms of the densities, moments, or other statistical descriptors.

We shall illustrate these points first in Subsection 4.A with respect to moments and moment-generating functions and then in Subsection 4.B for the probability density functions. A particularly strong constraint will emerge when we consider normal log likelihoods in Subsection 4.C.

A. Moments of Λ and λ

The likelihood ratio is a ratio of two densities $q_1(\mathbf{g})$ and $q_0(\mathbf{g})$, and the same two densities are the ones needed to compute moments of Λ under two hypotheses. It follows at once that the moments under H_0 are related to those under H_1 by

$$\begin{aligned}\langle \Lambda^{k+1} \rangle_0 &= \int_\infty \mathrm{d}^M g q_0(\mathbf{g}) \left[\frac{q_1(\mathbf{g})}{q_0(\mathbf{g})}\right]^{k+1} \\ &= \int_\infty \mathrm{d}^M g q_1(\mathbf{g}) \left[\frac{q_1(\mathbf{g})}{q_0(\mathbf{g})}\right]^{k} = \langle \Lambda^k \rangle_1.\end{aligned} \tag{4.1}$$

In particular, the mean of Λ under H_0 is always 1, since

$$\langle \Lambda \rangle_0 = \langle \Lambda^0 \rangle_1 = \int_\infty \mathrm{d}^M g q_0(\mathbf{g}) \frac{q_1(\mathbf{g})}{q_0(\mathbf{g})} = \int_\infty \mathrm{d}^M g q_1(\mathbf{g}) = 1, \tag{4.2}$$

and the variance of Λ under H_0 is easily expressed in terms of the mean under H_1:

$$\mathrm{var}_0(\Lambda) = \langle \Lambda^2 \rangle_0 - \langle \Lambda \rangle_0{}^2 = \langle \Lambda \rangle_1 - 1. \tag{4.3}$$

Moreover, since $\Lambda = e^{\lambda}$, we can rewrite Eq. (4.1) as

$$\langle \exp[(k+1)\lambda] \rangle_0 = \langle \exp(k\lambda) \rangle_1 . \tag{4.4}$$

Since Eq. (4.4) holds for arbitrary (even complex) k, we see that

$$M_0(\beta + 1) = M_1(\beta), \tag{4.5}$$

where $M_j(\beta)$ denotes the moment-generating function for λ under H_j. In fact, this property of the moment-generating functions holds only for log-likelihood ratios. Swensson and Green[16] showed that there is no other statistic whose moment-generating functions satisfy Eq. (4.5).

From Eqs. (4.5) and (3.19), the corresponding relation for characteristic functions is

$$\psi_0\left(\xi + \frac{i}{2\pi}\right) = \psi_1(\xi). \tag{4.6}$$

Notice that $M_0(\beta)$ can be used to generate moments of both λ and Λ under both hypotheses. From Eq. (3.22) with $t = \lambda$ we have

$$\langle \lambda^k \rangle_0 = M_0^{(k)}(0), \tag{4.7}$$

and with Eq. (4.5),

$$\langle \lambda^k \rangle_1 = M_1^{(k)}(0) = M_0^{(k)}(1). \tag{4.8}$$

Moments of Λ are found from $M_0(\beta)$ even more simply; moments under H_0 are given by

$$\langle \Lambda^k \rangle_0 = \langle \exp(k\lambda) \rangle_0 = M_0(k) \tag{4.9}$$

and under H_1 by

$$\langle \Lambda^k \rangle_1 = \langle \exp[(k+1)\lambda] \rangle_0 = M_0(k+1). \tag{4.10}$$

B. Probability Density Functions

Since a probability density function is uniquely determined by the corresponding characteristic function,[17] Eq. (4.6) implies that there is a relation between $p_0(\lambda)$ and $p_1(\lambda)$. To derive this relation, we write

$$\begin{aligned} p_1(\lambda) &= \mathscr{F}^{-1}\{\psi_1(\xi)\} = \int_{-\infty}^{\infty} \mathrm{d}\xi \psi_0\left(\xi + \frac{i}{2\pi}\right) \exp(2\pi i \xi \lambda) \\ &= e^{\lambda} \int_{-\infty + i/2\pi}^{\infty + i/2\pi} \mathrm{d}z\, \psi_0(z) \exp(2\pi i z \lambda), \end{aligned} \tag{4.11}$$

where $z = \xi + i/(2\pi)$. If $\psi_0(z)$ is analytic in the strip $0 \leqslant \mathrm{Im}(z) \leqslant 1/(2\pi)$, we can shift the contour and get

$$p_1(\lambda) = e^{\lambda} \int_{-\infty}^{\infty} \mathrm{d}z\, \psi_0(z) \exp(2\pi i z \lambda). \tag{4.12}$$

We show in Appendix A that the shift is allowed as long as $\langle \Lambda \rangle_1$ is finite.

Equation (4.12) shows that we need only multiply $p_0(\lambda)$ by e^{λ} to get $p_1(\lambda)$:

$$p_1(\lambda) = e^{\lambda} p_0(\lambda). \tag{4.13}$$

Of course, both $p_0(\lambda)$ and $p_1(\lambda)$ must be properly normalized to unity, so the only functions that can be densities for the log likelihood under H_0 are ones that remain normalized after multiplication by e^{λ}, i.e.,

$$\int_{-\infty}^{\infty} \mathrm{d}\lambda\; e^{\lambda} p_0(\lambda) = 1. \tag{4.14}$$

If we know that $p_0(\lambda)$ really is the density for a log likelihood under H_0, however, Eq. (4.14) is trivially satisfied, since it is equivalent to Eq. (4.2).

From Eq. (4.13) we readily find a relation between the densities for Λ under the two hypotheses. Since λ and Λ are related by a monotonic transformation, we can write

$$p_j(\lambda) = \frac{\mathrm{pr}(\Lambda | H_j)}{|\mathrm{d}\lambda / \mathrm{d}\Lambda|}. \tag{4.15}$$

The Jacobian $|\mathrm{d}\lambda / \mathrm{d}\Lambda|$ is the same under H_0 and H_1, so Eq. (4.13) becomes

$$\mathrm{pr}(\Lambda | H_1) = e^{\lambda} \mathrm{pr}(\Lambda | H_0) = \Lambda \mathrm{pr}(\Lambda | H_0). \tag{4.16}$$

It is instructive to rewrite this equation as

$$\frac{\mathrm{pr}(\Lambda | H_1)}{\mathrm{pr}(\Lambda | H_0)} = \Lambda. \tag{4.17}$$

In this form, the relation was known to Green and Swets,[18] who described it as follows: "To paraphrase Gertrude Stein, the likelihood ratio of the likelihood ratio is the likelihood ratio." To paraphrase Green and Swets, the likelihood ratio is a sufficient statistic for deciding between H_0 and H_1. If we were given *any* function of the data, $t(\mathbf{g})$ and we wanted to make an optimal decision based only on $t(\mathbf{g})$ and not on the original $\mathbf{g}$, we would form the likelihood ratio $\mathrm{pr}(t(\mathbf{g})|H_1)/\mathrm{pr}(t(\mathbf{g})|H_0)$ and compare it with a threshold. In most cases this strategy, though optimal when only $t(\mathbf{g})$ is available, would be inferior to forming the likelihood ratio from the original data, and there would thus be an information loss inherent in using $t(\mathbf{g})$ in place of $\mathbf{g}$. From Eq. (4.17) we see that there is no such information loss if $t(\mathbf{g})$ is the sufficient statistic $\Lambda(\mathbf{g})$.

C. Normal Log Likelihoods

Much of the literature on the ideal observer has proceeded from the assumption—implicit or explicit—that the log likelihood is normally distributed. One justification for this assumption is that λ is a linear functional of $\mathbf{g}$ for nonrandom signals and Gaussian noise, and it may be approximately linear even with random signals and/or non-Gaussian noise. As noted in Subsection 3.G, the central-limit theorem often leads to normality for a linear discriminant.

Even if λ is not a linear discriminant, however, it may still be approximately or asymptotically normal. It is common in statistics to consider data sets $\mathbf{g}$ that consist of many independent observations $\mathbf{g}_k$. Because of the independence, we can write

$$\mathrm{pr}(\mathbf{g}|H_j) = \prod_{k=1}^{K} \mathrm{pr}(\mathbf{g}_k | H_j). \tag{4.18}$$

The log likelihood is then given by

$$\lambda = \sum_{k=1}^{K} [\log \mathrm{pr}(\mathbf{g}_k | H_1) - \log \mathrm{pr}(\mathbf{g}_k | H_0)]. \tag{4.19}$$

By the central-limit theorem, λ is asymptotically normal as the number of observations K becomes large.

A similar argument can be made in an imaging context. Suppose that the image $\mathbf{g}$ can be divided up into K statistically independent subimages $\mathbf{g}_k$, so that Eq. (4.18) again applies. If N of the subimages are different under the two hypotheses [so that the log probabilities in Eq. (4.19) do not simply cancel], then λ is a sum of N independent random variables; it is therefore normally distributed by the central-limit theorem if N is large.

There are situations in imaging in which normality is not a good approximation, and a detailed discussion of ideal-observer performance in those cases is given in Subsection 5.C, but first we explore some little-recognized consequences of a normality assumption for the log likelihood.

Suppose that λ is normally distributed under H_0 with mean $\bar{\lambda}_0$ and variance $\mathrm{var}_0(\lambda)$. One might expect that $\bar{\lambda}_0$ and $\mathrm{var}_0(\lambda)$ could be specified independently and that two additional independent parameters would be needed to specify the mean and variance under H_1; we shall show that this expectation is incompatible with Eq. (4.13).

The moment-generating functions for a general, normally distributed, discriminant function are given by Eq. (3.26). If, however, the discriminant function is the log likelihood, then Eq. (4.2) requires that $\langle\Lambda\rangle_0 = 1$, or, with Eqs. (3.26) and (4.9),

$$M_0(1) = \exp[\bar{\lambda}_0 + \tfrac{1}{2}\,\mathrm{var}_0(\lambda)] = 1. \tag{4.20}$$

Thus $\bar{\lambda}_0$ and $\sigma_0{}^2$ must be related by

$$\bar{\lambda}_0 = -\tfrac{1}{2}\,\mathrm{var}_0(\lambda), \tag{4.21}$$

and Eq. (3.26) must take the form

$$M_0(\beta) = \exp[\tfrac{1}{2}(\beta^2 - \beta)\mathrm{var}_0(\lambda)] \tag{4.22}$$

if it is to apply to a log likelihood.

For Eq. (4.13) to hold, $p_1(\lambda)$ must also be normal, and we can apply Eq. (4.8) to Eq. (4.22) and determine the mean and variance of λ under H_1. The results are

$$\bar{\lambda}_1 = M_0^{(1)}(1) = \tfrac{1}{2}\,\mathrm{var}_0(\lambda) = -\bar{\lambda}_0, \tag{4.23}$$

$$\mathrm{var}_1(\lambda) = M_0^{(2)}(1) - [M_0^{(1)}(1)]^2 = \mathrm{var}_0(\lambda). \tag{4.24}$$

The moment-generating function under H_1 is then given by

$$M_1(\beta) = M_0(\beta + 1) = \exp[\tfrac{1}{2}(\beta^2 + \beta)\mathrm{var}_0(\lambda)]. \tag{4.25}$$

The probability density functions for a normal log likelihood are thus

$$p_0(\lambda) = \frac{1}{[2\pi\,\mathrm{var}_0(\lambda)]^{1/2}}\exp\left[-\frac{[\lambda + \frac{1}{2}\,\mathrm{var}_0(\lambda)]^2}{2\,\mathrm{var}_0(\lambda)}\right], \tag{4.26}$$

$$p_1(\lambda) = \frac{1}{[2\pi\,\mathrm{var}_0(\lambda)]^{1/2}}\exp\left[-\frac{[\lambda - \frac{1}{2}\,\mathrm{var}_0(\lambda)]^2}{2\,\mathrm{var}_0(\lambda)}\right]. \tag{4.27}$$

It is easy to verify that Eq. (4.13) is satisfied.

Relations (4.23)–(4.27) were also derived by Fukunaga,[8] who started with the assumption that the data are normally distributed with equal covariance, which implies that the log likelihood is normal under both hypotheses. The derivation given above uses only the weaker assumption that λ is normal under H_0. Since exact or approximate normality of λ can occur with decidedly nonnormal data, our approach is much more general.

With this normal model, all statistical properties of λ under both hypotheses are determined by the single parameter $\mathrm{var}_0(\lambda)$. From Eqs. (4.10) and (4.22), this parameter can also be expressed as

$$\mathrm{var}_0(\lambda) = \log\langle\Lambda\rangle_1. \tag{4.28}$$

Therefore, to fully characterize a normal *log* likelihood, we need only calculate the mean of the *likelihood* under H_1.

Next we examine the AUC for this model. Since the discriminant function is normal under both hypotheses, AUC is given exactly by Eq. (3.2), and the relevant SNR is given by

$$\mathrm{SNR}_\lambda{}^2 = \frac{[\langle\lambda\rangle_1 - \langle\lambda\rangle_0]^2}{\frac{1}{2}\,\mathrm{var}_1(\lambda) + \frac{1}{2}\,\mathrm{var}_0(\lambda)} = \mathrm{var}_0(\lambda) = \log\langle\Lambda\rangle_1. \tag{4.29}$$

Hence with Eq. (3.2),

$$\mathrm{AUC} = \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}[\tfrac{1}{2}\sqrt{\log\langle\Lambda\rangle_1}]. \tag{4.30}$$

This expression, like all the others in Subsection 4.C, follows rigorously once we make the initial assumption that λ is normally distributed under H_0; no further assumptions or approximations are required.

5. LIKELIHOOD-GENERATING FUNCTION

A. Definitions and Basic Properties

In view of Eq. (4.13), both $p_0(\lambda)$ and $p_1(\lambda)$ can always be derived from a single nonnegative function $f(\lambda)$ as follows:

$$p_0(\lambda) = \exp(-\tfrac{1}{2}\lambda)f(\lambda), \qquad p_1(\lambda) = \exp(\tfrac{1}{2}\lambda)f(\lambda). \tag{5.1}$$

Though it is easy to find the $f(\lambda)$ associated with the normal densities of Eqs. (4.26) and (4.27), normality is not required for the existence of this function or for any of the other results in this section.

The characteristic functions and moment-generating functions are determined from $f(\lambda)$ by

$$\psi_0(\xi) = F\left(\xi - \frac{i}{4\pi}\right), \qquad \psi_1(\xi) = F\left(\xi + \frac{i}{4\pi}\right), \tag{5.2}$$

$$M_0(\beta) = F_L(\beta - \tfrac{1}{2}), \qquad M_1(\beta) = F_L(\beta + \tfrac{1}{2}), \tag{5.3}$$

where $F(\xi)$ is the Fourier transform of $f(\lambda)$ and $F_L(\beta)$ is its two-sided Laplace transform,

$$F_L(\beta) = \int_{-\infty}^{\infty} \mathrm{d}\lambda\, f(\lambda)\exp(\beta\lambda). \tag{5.4}$$

Normalization of the densities requires that $\psi_j(0) = 0$ and $M_j(0) = 0$, which means that $F(\pm i/4\pi) = 1$ and $F_L(\pm\frac{1}{2}) = 1$. We can enforce these conditions by defining new functions $T(\xi)$ and $G(\beta)$ such that

$$F(\xi) = \exp\left[\left(\xi + \frac{i}{4\pi}\right)\left(\xi - \frac{i}{4\pi}\right)T(\xi)\right], \tag{5.5}$$

$$F_L(\beta) = \exp[(\beta + \tfrac{1}{2})(\beta - \tfrac{1}{2})G(\beta)]. \tag{5.6}$$

The functions $T(\xi)$ and $G(\beta)$ are related to each other by

$$T(\xi) = -4\pi^2 G(-2\pi i\xi). \tag{5.7}$$

If we allow complex arguments, we can derive all statistical properties of the likelihood and log likelihood from either $T(\xi)$ or $G(\beta)$. We choose to work with $G(\beta)$, which we call the likelihood-generating function.

In terms of $G(\beta)$, the characteristic and moment-generating functions for the log likelihood are given by

$$\begin{aligned}\psi_0(\xi) &= \exp\left[\xi\left(\xi - \frac{i}{2\pi}\right)T\left(\xi - \frac{i}{4\pi}\right)\right] \\ &= \exp\left[-4\pi^2\xi\left(\xi - \frac{i}{2\pi}\right)G\left(-2\pi i\xi - \frac{1}{2}\right)\right],\end{aligned} \tag{5.8}$$

$$\begin{aligned}\psi_1(\xi) &= \exp\left[\xi\left(\xi + \frac{i}{2\pi}\right)T\left(\xi + \frac{i}{4\pi}\right)\right] \\ &= \exp\left[-4\pi^2\xi\left(\xi + \frac{i}{2\pi}\right)G\left(-2\pi i\xi + \frac{1}{2}\right)\right],\end{aligned} \tag{5.9}$$

$$M_0(\beta) = \exp[\beta(\beta - 1)G(\beta - \tfrac{1}{2})], \tag{5.10}$$

$$M_1(\beta) = \exp[\beta(\beta + 1)G(\beta + \tfrac{1}{2})]. \tag{5.11}$$

From these equations we see that the basic requirements $\psi_j(0) = 1$ and $M_j(0) = 1$ are satisfied.

Two equivalent expressions for $G(\beta)$ can be derived from Eqs. (5.10) and (5.11) by changing variables:

$$G(\beta) = \frac{\log M_0(\beta + \frac{1}{2})}{(\beta - \frac{1}{2})(\beta + \frac{1}{2})} = \frac{\log M_1(\beta - \frac{1}{2})}{(\beta - \frac{1}{2})(\beta + \frac{1}{2})}. \tag{5.12}$$

Moments of the likelihood ratio are easily expressed in terms of $G(\beta)$:

$$\log\langle\Lambda^k\rangle_1 = \log\langle\Lambda^{k+1}\rangle_0 = k(k + 1)G(k + \tfrac{1}{2}). \tag{5.13}$$

Furthermore, we can compute SNR_λ if we know $G(\beta)$; the requisite moments are given by

$$\bar{\lambda}_0 = -G(-\tfrac{1}{2}), \qquad \bar{\lambda}_1 = G(\tfrac{1}{2}); \tag{5.14}$$

$$\begin{aligned}\mathrm{var}_0(\lambda) &= 2[G(-\tfrac{1}{2}) - G'(-\tfrac{1}{2})], \\ \mathrm{var}_1(\lambda) &= 2[G(\tfrac{1}{2}) + G'(\tfrac{1}{2})],\end{aligned} \tag{5.15}$$

where $G'(\beta)$ is the derivative of $G(\beta)$. Hence

$$\mathrm{SNR}_\lambda{}^2 = \frac{[G(\frac{1}{2}) + G(-\frac{1}{2})]^2}{G(\frac{1}{2}) + G(-\frac{1}{2}) + G'(\frac{1}{2}) - G'(-\frac{1}{2})}. \tag{5.16}$$

We see that $\mathrm{SNR}_\lambda{}^2$ has the structure X^2/X if the derivative of G is approximately the same at $\frac{1}{2}$ and $-\frac{1}{2}$. In that case,

$$\mathrm{SNR}_\lambda{}^2 \simeq \bar{\lambda}_1 - \bar{\lambda}_0 = G(\tfrac{1}{2}) + G(-\tfrac{1}{2}) \simeq 2G(0). \tag{5.17}$$

It is interesting to compare these results with those for a normally distributed log likelihood, as discussed in Subsection 4.C. Comparing Eqs. (5.10) and (5.11) with Eqs. (4.22) and (4.25), we see that the likelihood-generating function for a normal log likelihood is $G(\beta) = \frac{1}{2}\mathrm{var}_0(\lambda) = \text{constant}$. With that observation, Eq. (5.17) is in accord with Eq. (4.29); by Eq. (4.8), $\log M_1(1) = \log \bar{\Lambda}_1 = 2G(\frac{3}{2})$, so Eqs. (4.29) and (5.17) agree exactly in the Gaussian case in which $G(\frac{3}{2}) = G(\frac{1}{2}) = G(-\frac{1}{2})$.

B. Restrictions on the Likelihood-Generating Function

One restriction on the form of $G(\beta)$ is Marcinkiewicz's theorem,[17] which says that if $\exp(P)$ is a characteristic function and P is a polynomial, the order of the polynomial can be at most 2. This means that the only polynomial form for $G(\beta)$ is the one assumed when λ is normally distributed, namely, $G = \text{constant}$. Nevertheless, it may be a useful approximation to treat $G(\beta)$ as a low-order polynomial if we restrict attention to β near the origin.

Another restriction arises from the so-called hermiticity property of Fourier transforms (which has nothing to do with Hermitian operators). Since the function $f(\lambda)$ defined in Eq. (5.1) is real, its Fourier transform must satisfy $F(-\xi) = F^*(\xi)$ for real ξ. In Eq. (5.5) the coefficient of $T(\xi)$ in the exponent is real and even, so hermiticity of $F(\xi)$ implies that $T(-\xi) = T^*(\xi)$, which in turn requires that

$$G(2\pi i\xi) = G^*(-2\pi i\xi), \qquad (\xi \text{ real}). \tag{5.18}$$

Other restrictions on the behavior of $G(\beta)$ arise from fundamental inequalities. Jensen's inequality[1,19] says that $h(\langle x\rangle) \geqslant \langle h(x)\rangle$, where h is any concave function (negative second derivative) and x is a random variable. Applying this inequality to Eq. (5.11) with β real and $h(x) = \log(x)$ yields

$$\beta(\beta + 1)G(\beta + \tfrac{1}{2}) = \log\langle\exp(\beta\lambda)\rangle_1 \geqslant \beta\bar{\lambda}_1 = \beta G(\tfrac{1}{2}). \tag{5.19}$$

With a change of variables, we find

$$(\beta + \tfrac{1}{2})G(\beta) \geqslant G(\tfrac{1}{2}), \qquad \beta \text{ real}, \ \beta \geqslant \tfrac{1}{2}, \tag{5.20}$$

which says that $G(\beta)$ can fall off with increasing β along the real axis but no faster than $1/(\beta + \frac{1}{2})$.

Other useful results can be obtained by applying Jensen's inequality to the expectation of Λ^α (α real and nonnegative):

$$\log\langle\Lambda^\alpha\rangle_j \geqslant \alpha\langle\lambda\rangle_j, \qquad j = 0, 1. \tag{5.21}$$

For $\alpha = 1$ and $j = 0$, we find

$$-G(-\tfrac{1}{2}) = \bar{\lambda}_0 \leq 0, \tag{5.22}$$

and for $\alpha = 1$ and $j = 1$,

$$\bar{\lambda}_1 \leq \log \bar{\Lambda}_1 = 2G(\tfrac{3}{2}). \tag{5.23}$$

C. Relation of the Likelihood-Generating Function to Area under an ROC Curve

The key expression for AUC is Eq. (3.17), which we can rewrite in terms of the likelihood-generating function as

$$\begin{aligned}\text{AUC} &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\exp\left\{-4\pi^2\left(\xi^2 - \frac{i\xi}{2\pi}\right)\right.\\ &\quad\times\left.\left[G\left(-2\pi i\xi - \frac{1}{2}\right) + G\left(2\pi i\xi + \frac{1}{2}\right)\right]\right\}\\ &= \frac{1}{2} + \frac{1}{2\pi i}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\exp\left\{-4\pi^2\left(\xi^2 - \frac{i\xi}{2\pi}\right)H(\xi)\right\},\end{aligned} \tag{5.24}$$

where

$$H(\xi) \equiv G(-2\pi i\xi - \tfrac{1}{2}) + G(2\pi i\xi + \tfrac{1}{2}). \tag{5.25}$$

The AUC is thus determined completely by a particular combination of likelihood-generating functions, denoted $H(\xi)$. From the fact that $\psi_0(\xi)\psi_1^*(\xi)$ is the Fourier transform of a real quantity, we can show that $H^*(\xi) = H(-\xi)$ for real ξ. That means that we can write

$$H(\xi) = H_r(\xi) + iH_i(\xi), \tag{5.26}$$

where $H_r(\xi)$ and $H_i(\xi)$ are both real and

$$H_r(-\xi) = H_r(\xi), \qquad H_i(-\xi) = -H_i(\xi). \tag{5.27}$$

Moreover, from basic properties of characteristic functions[17] we know that $|\psi_j(\xi)| \leq 1$ for real ξ. Hence $|\psi_0(\xi)\psi_1^*(\xi)| \leq 1$, from which we can show that

$$\xi^2 H_r(\xi) + \frac{\xi}{2\pi}H_i(\xi) \geq 0. \tag{5.28}$$

To make use of these properties, we recognize that the factor $1/\xi$ in the AUC integral is odd and the integral is over a symmetric interval, so only the odd part of $\psi_0(\xi)\psi_1^*(\xi)$ contributes to the integral. Thus

$$\begin{aligned}\text{AUC} &= \frac{1}{2} - \frac{1}{2\pi}\mathscr{P}\int_{-\infty}^{\infty}\frac{\mathrm{d}\xi}{\xi}\\ &\quad\times\exp\left\{-4\pi^2\left[\xi^2 H_r(\xi) + \frac{\xi}{2\pi}H_i(\xi)\right]\right\}\\ &\quad\times\sin\left\{4\pi^2\left[\xi^2 H_i(\xi) - \frac{\xi}{2\pi}H_r(\xi)\right]\right\}.\end{aligned} \tag{5.29}$$

Because of relation (5.28), the exponential factor in Eq. (5.29) falls off rapidly as ξ increases unless $H_r(\xi)$ decreases more rapidly than $1/\xi^2$. In addition, the factor $1/\xi$ serves to emphasize small ξ, so the main contribution to the integral comes from ξ very near the origin. It is therefore reasonable to expand $H(\xi)$ in a Taylor series about $\xi = 0$ (Maclaurin series). It is shown in Appendix A that $H(\xi)$ is analytic for all real ξ, so this series is guaranteed to converge.

Since $H(\xi)$ is a combination of likelihood-generating functions, we begin by expanding $G(\beta)$ as

$$G(\beta) = \sum_{n=0}^{\infty}\frac{1}{n!}G^{(n)}(0)\beta^n. \tag{5.30}$$

The derivatives $G^{(n)}(0)$ are all real, since $G(\beta)$ is real for real β.

When we combine Eqs. (5.25) and (5.30) the terms with odd n cancel, so the expansion for $H(\xi)$ is

$$H(\xi) = 2\sum_{k=0}^{\infty}\frac{1}{(2k)!}G^{(2k)}(0)\left(2\pi i\xi + \frac{1}{2}\right)^{2k}. \tag{5.31}$$

As a first approximation we might assume that the expansion for $G(\beta)$ can be truncated after $n = 1$, so that $G(\beta)$ is approximated by a linear function near the origin. If this approximation is valid, then $H(\xi) \simeq 2G(0)$. The AUC integral then has the same structure as Eq. (3.28), and we find readily that

$$\text{AUC} = \tfrac{1}{2} + \tfrac{1}{2}\,\text{erf}[\tfrac{1}{2}\sqrt{2G(0)}]. \tag{5.32}$$

This result is compatible with relations (3.2) and (5.17). If $G(\beta)$ is approximately linear, the derivative terms in the denominator of Eq. (5.16) cancel and $\text{SNR}_\lambda{}^2$ is $2G(0)$. In addition, AUC is determined solely by $G(0)$ in this approximation.

The upshot of this calculation is that we have extended the range of validity of the error-function formula, Eq. (3.2). As originally derived, it held exactly only for Gaussian discriminant functions, which in the case of the ideal observer would mean $G(\beta)$ = constant. Now we see that the formula also holds with non-Gaussian log likelihoods as long as we can approximate $G(\beta)$ by a linear function near the origin.

Though motivated by the approximation that $G(\beta)$ is linear near the origin, Eq. (5.32) is actually much more general. We demonstrate in Appendix B that Eq. (5.32) is the leading term in an asymptotic expansion for AUC. This term alone approaches the true AUC as $G(0)$ increases. As shown in Eq. (B22), the first correction term is proportional to the second derivative $G''(0)$, and it falls off as $[G(0)]^{-3}\exp[-\tfrac{1}{2}G(0)]$. Thus, if the curvature of the likelihood-generating function at the origin is small or if $G(0)$ is moderately large, Eq. (5.32) is an excellent approximation to AUC.

D. Lower Limit on Area under an ROC Curve

We shall now show that $G(0)$ can be used to set a lower limit on AUC with no approximations at all. The starting point is Eq. (B8), which, with the change of variables $\alpha = 2\pi\xi$, becomes

$$\text{AUC} = 1 - \frac{1}{2\pi}\int_0^{\infty}\frac{\mathrm{d}\alpha}{\alpha^2 + \frac{1}{4}}\exp\left[-2\left(\alpha^2 + \frac{1}{4}\right)\text{Re}\,G(i\alpha)\right]. \tag{5.33}$$

From Eq. (5.10) with $\beta = \frac{1}{2} + i\alpha$ and Eq. (4.9), it follows that

$$\exp[-2(\alpha^2 + \tfrac{1}{4})G(i\alpha)] = M_0(\tfrac{1}{2} + i\alpha) = \langle\Lambda^{1/2}\Lambda^{i\alpha}\rangle_0. \tag{5.34}$$

Now we can separate $G(i\alpha)$ into real and imaginary parts and recognize $\exp[-2i(\alpha^2 + \frac{1}{4})\mathrm{Im}\, G(i\alpha)]$ as a pure phase factor, so

$$\exp[-2(\alpha^2 + \tfrac{1}{4})\mathrm{Re}\, G(i\alpha)] = |M_0(\tfrac{1}{2} + i\alpha)| = |\langle \Lambda^{1/2}\Lambda^{i\alpha}\rangle_0|, \qquad (\alpha \text{ real}). \tag{5.35a}$$

Using this equation with

$$|\langle \Lambda^{1/2}\Lambda^{i\alpha}\rangle_0| \leqslant \langle|\Lambda^{1/2}\Lambda^{i\alpha}|\rangle_0 = \langle \Lambda^{1/2}\rangle_0 = \exp[-\tfrac{1}{4}G(0)], \tag{5.35b}$$

we can infer that

$$\exp[-2(\alpha^2 + \tfrac{1}{4})\mathrm{Re}\, G(i\alpha)] \leqslant \exp[-\tfrac{1}{4}G(0)]. \tag{5.36}$$

Returning to the AUC integral in Eq. (5.33), we now have

$$\mathrm{AUC} \geqslant 1 - \tfrac{1}{2}\exp[-\tfrac{1}{2}G(0)]. \tag{5.37}$$

This inequality was derived without any approximations on the form of $G(\beta)$. It provides a lower bound for the AUC obtained by an ideal observer, regardless of the probability laws for the data or for the likelihood ratio. As with the approximate formula (5.32), the bound depends solely on $G(0)$.

E. Some Interpretations and Interrelations

We have just seen that $G(0)$ plays a pivotal role in specifying the performance of an ideal observer. It fully specifies AUC if the linearity approximation holds, and it sets an exact lower bound in all cases. Because of its importance, we now look at some ways of interpreting $G(0)$.

Note from Eqs. (5.1) and (5.6) that

$$G(0) = -4\log[F_L(0)] = -4\log\left[\int_{-\infty}^{\infty} \mathrm{d}\lambda f(\lambda)\right]. \tag{5.38}$$

The nonnegative function $f(\lambda)$ introduced in Eq. (5.1) is not a normalized probability density function, so the logarithm does not vanish. Curiously, when $f(\lambda)$ is multiplied by either $\exp(\frac{1}{2}\lambda)$ or $\exp(-\frac{1}{2}\lambda)$, it *is* a normalized density; but without the exponential factors it is not normalized, and, moreover, the ideal-observer performance is determined (or at least strongly influenced) by just how far away it is from being normalized.

Other forms for $G(0)$ are obtained by expressing $f(\lambda)$ in terms of $p_0(\lambda)$ or $p_1(\lambda)$, yielding

$$\begin{aligned} G(0) &= -4\log\left[\int_{-\infty}^{\infty} \mathrm{d}\lambda p_1(\lambda)\exp(-\tfrac{1}{2}\lambda)\right] \\ &= -4\log\langle\Lambda^{-1/2}\rangle_1 = -4\log\left[\int_{-\infty}^{\infty} \mathrm{d}\lambda p_0(\lambda)\exp(\tfrac{1}{2}\lambda)\right] \\ &= -4\log\langle\Lambda^{1/2}\rangle_0. \end{aligned} \tag{5.39}$$

Thus, within the linear approximation, AUC is fully determined by $\langle\Lambda^{-1/2}\rangle_1$ or $\langle\Lambda^{1/2}\rangle_0$. This conclusion is to be compared with Eq. (4.30), in which we related AUC to $\langle\Lambda\rangle_1$. There is no contradiction in this difference; Eq. (4.30) was derived on the assumption that λ was normally distributed, which means $G(\beta)$ = constant. From Eq. (5.13) we see that $\log\langle\Lambda\rangle_1 = 2G(\frac{3}{2})$, so Eqs. (4.30) and (5.32) agree if $G(\frac{3}{2}) = G(0)$, as it must for a normal log likelihood. If normality fails, Eq. (5.32) will be the better approximation, since it approximates $G(\pm\frac{1}{2})$ by $G(0)$ rather than $G(\frac{3}{2})$.

The average $\langle\Lambda^{1/2}\rangle_0$ can also be expressed as an integral over data space, so Eq. (5.39) can be written as

$$G(0) = -4\log\left[\int_{\infty} \mathrm{d}^M g\,\sqrt{q_0(\mathbf{g})q_1(\mathbf{g})}\right]. \tag{5.40}$$

This form shows that $G(0)$ increases as the overlap between $q_0(\mathbf{g})$ and $q_1(\mathbf{g})$ is reduced.

An interesting way to visualize the effect of $G(0)$ is to plot $M_0(\beta)$ (which is identical to $\langle\Lambda^{\beta}\rangle_0$) versus β for real β (see Fig. 1). From Eq. (5.10) this plot must necessarily go through unity at $\beta = 0$ and $\beta = 1$, but the value at $\beta = \frac{1}{2}$ (which is not necessarily the minimum of the curve) is given by

$$M_0(\tfrac{1}{2}) = \langle\Lambda^{1/2}\rangle_0 = \exp[-\tfrac{1}{4}G(0)] = \int_{-\infty}^{\infty} \mathrm{d}\lambda f(\lambda). \tag{5.41}$$

It can be shown from the Schwarz inequality that $G(0) \geqslant 0$ and hence $M_0(\frac{1}{2}) \leqslant 1$. At $\beta = \frac{3}{2}$, the plot goes through $\langle\Lambda\rangle_1$, which is $\geqslant 1$. From the behavior of the curve in Fig. 1, we see that larger values for $\langle\Lambda\rangle_1$ imply smaller values for $\langle\Lambda^{1/2}\rangle_0$.

In summary, AUC increases as any of the following things occur: $G(0)$ gets larger, the integral of $f(\lambda)$ gets smaller, the expectation of the square root of the likelihood under H_0 gets smaller, the expectation of the likelihood ratio under H_1 get larger, or the overlap between $q_0(\mathbf{g})$ and $q_1(\mathbf{g})$ is reduced. In a signal-detection problem, all of these things occur when the signal contrast is increased or the noise level is reduced. Conversely, in the limit of very small signal contrast or large noise, $f(\lambda)$ is indistinguishable from $p_0(\lambda)$ or $p_1(\lambda)$ and hence normalized to unity, $\langle\Lambda\rangle_1$ is indistinguishable from $\langle\Lambda\rangle_0$ and hence unity, $q_0(\mathbf{g})$ is indistinguishable from $q_1(\mathbf{g})$ so the integral of their geometric mean in Eq. (5.40) is unity, $G(0)$ is zero, and AUC = 1/2. We note that inequality (5.37) becomes an equality in both limits, large and small $G(0)$.

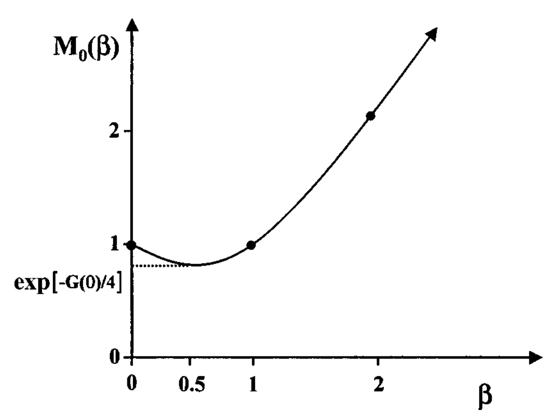

Fig. 1. Illustration of the behavior of the function $M_0(\beta)$.

As a practical matter, we can use (5.39) to evaluate $G(0)$ for any problem in which we have an analytic expression for Λ and a way of generating sample data vectors $\mathbf{g}$ under the null hypothesis. All we have to do is repeatedly generate $\mathbf{g}$, form $[\Lambda(\mathbf{g})]^{1/2}$ for each $\mathbf{g}$, and approximate the ensemble average $\langle \Lambda^{1/2} \rangle_0$ with a sample average. The same approach gives us a way of numerically evaluating $G(\beta)$ for any β, even complex ones.

6. SUMMARY AND CONCLUSIONS

The main purpose of this paper has been to bridge the gap between two segments of the literature on task-based assessment of image quality. One segment, which includes the earlier papers in this series, has mainly expressed performance on detection and classification tasks in terms of simple SNR's determined from the first and second moments of the discriminant function. The other segment, which includes much of the psychophysical literature, has reported the full ROC curve and used the area under it as a summary figure of merit. In the past, contact between these two segments was made mainly by assuming normal statistics and relating AUC to SNR by the error-function formula (3.2). Little attention was paid to establishing the validity of the normality assumption or to investigating the relation between SNR and AUC if normality did not hold.

Here we began in Section 2 with a broad overview of decision theory. After demonstrating that any nonrandomized decision rule could be cast in the form of comparing a discriminant function to a threshold, we derived the well-known result that the optimum discriminant function was the likelihood ratio or its logarithm. The ideal observer was defined as one that used one of these discriminants.

In Section 3 we derived many different expressions for SNR and AUC metrics. Several of these expressions have not appeared previously in the literature. In particular, we demonstrated that AUC for an arbitrary discriminant function could be expressed by a principal-value integral involving the characteristic functions of the discriminant function. This formula was to play a key role in the subsequent discussion.

In Section 4 we examined the properties of the ideal observer from first principles. Several strong constraints on the moments of the likelihood ratio or the log likelihood were derived, and it was shown that the probability density functions for these test statistics were intimately related. In particular, we were able to derive some surprising results for the case in which the log likelihood is normally distributed under one hypothesis. We showed that it is then necessarily normal under the other hypothesis and that the two means and two variances could be expressed in terms of a single parameter. This led to unanticipated new expressions for AUC in the normal case. In particular, we found that AUC was determined exactly by the mean of Λ under H_1.

In Section 5 we attempted to unify these considerations by defining a new quantity called the likelihood-generating function $G(\beta)$. We showed that all moments of both the likelihood and the log likelihood under both hypotheses could be derived from this one function. Moreover, the AUC could be expressed exactly as a principal-value integral involving the likelihood-generating function. Perhaps the most surprising result was that the AUC could be expressed, with one reasonable approximation, in terms of a single value, $G(0)$. Moreover, this same value sets a lower bound to AUC without approximation.

The obvious next step in the exploitation of this new theory is to apply it to practical problems in signal detection and image quality. This will be the theme of a subsequent paper in this series.

APPENDIX A: ANALYTICITY CONSIDERATIONS

A function of a complex variable is analytic if and only if it satisfies the Cauchy–Riemann conditions.[20] To apply this test to the characteristic function $\psi_j(z)$, we write

$$\psi_j(z) \equiv u_j(\xi, \eta) + i v_j(\xi, \eta) = \int_{-\infty}^{\infty} \mathrm{d}\lambda p_j(\lambda) \exp(-2\pi i z \lambda)$$

$$= \int_{-\infty}^{\infty} \mathrm{d}\lambda p_j(\lambda) \exp(-2\pi i \xi \lambda) \exp(2\pi \eta \lambda), \tag{A1}$$

where $z = \xi + i\eta$. The function $\psi_j(z)$ is analytic at point z if and only if

$$\frac{\partial u_j(\xi, \eta)}{\partial \xi} = \frac{\partial v_j(\xi, \eta)}{\partial \eta}, \qquad \frac{\partial v_j(\xi, \eta)}{\partial \xi} = -\frac{\partial u_j(\xi, \eta)}{\partial \eta}. \tag{A2}$$

Since $p_j(\lambda)$ is real, separating real and imaginary parts of Eq. (A1) shows that

$$u_j(\xi, \eta) = \int_{-\infty}^{\infty} \mathrm{d}\lambda p_j(\lambda) \cos(2\pi \xi \lambda) \exp(2\pi \eta \lambda), \tag{A3}$$

$$v_j(\xi, \eta) = -\int_{-\infty}^{\infty} \mathrm{d}\lambda p_j(\lambda) \sin(2\pi \xi \lambda) \exp(2\pi \eta \lambda). \tag{A4}$$

It is easy to show that Eq. (A2) is satisfied if we can differentiate under the integral sign. Lang[21] gives the conditions under which this operation is permissible; in essence, the integral must be absolutely convergent before and after differentiation. The conditions are thus

$$\int_{-\infty}^{\infty} \mathrm{d}\lambda p_j(\lambda) \exp(2\pi \eta \lambda) < \infty, \tag{A5}$$

$$\int_{-\infty}^{\infty} \mathrm{d}\lambda p_j(\lambda) |\lambda| \exp(2\pi \eta \lambda) < \infty. \tag{A6}$$

If $p_j(\lambda)$ is specifically the density for the log likelihood under H_j, relation (A5) reads

$$\langle \Lambda^{2\pi\eta} \rangle_j < \infty. \tag{A7}$$

In addition, the Cauchy–Schwarz inequality shows that relation (A6) is satisfied if

$$\langle \lambda^2 \rangle_j \langle \Lambda^{4\pi\eta} \rangle_j < \infty. \tag{A8}$$

For $\eta = 0$ (i.e., z on the real axis), relation (A7) is trivially satisfied and relation (A8) is satisfied if the second moment of λ exists, so $\psi_j(z)$ is analytic on the real axis under this weak assumption.

If $|\eta| > 0$, however, the exponential factor might cause one of the integrals to diverge. Some special cases are of interest. For example, in some problems the range of values of λ is restricted. If $p_j(\lambda) = 0$ for $\lambda \leqslant a \leqslant 0$, then $\psi_j(z)$ is analytic in the upper half-plane, and if $p_j(\lambda) = 0$ for $\lambda \geqslant b \geqslant 0$, then $\psi_j(z)$ is analytic in the lower half-plane. And if $p_j(\lambda) = 0$ unless $a \leqslant \lambda \leqslant b$, then $\psi_j(z)$ is an entire function (analytic for all finite z); this result is a statement of the Paley–Wiener theorem.[22,23]

In Subsection 4.B we need to assume that $\psi_0(z)$ is analytic in the strip $0 \leqslant \mathrm{Im}(z) \leqslant 1/(2\pi)$. For any point in this strip we can bound the integral in relation (A5) by using Jensen's inequality with the concave function $h(x) = x^{2\pi\eta}$. The result is

$$\int_{-\infty}^{\infty} \mathrm{d}\lambda p_0(\lambda)\exp(2\pi\eta\lambda) \leqslant \int_{-\infty}^{\infty} \mathrm{d}\lambda p_0(\lambda)\exp(\lambda) = \langle \exp(\lambda)\rangle_0 . \tag{A9}$$

This expectation is the same as $\langle \Lambda \rangle_0$, which we know to be unity by Eq. (4.2), so relation (A7) is always satisfied in the strip under consideration. By a similar argument, relation (A8) is satisfied if $\langle \Lambda^2 \rangle_0 = \langle \Lambda \rangle_1 < \infty$. Thus, as long as the mean of Λ is finite under H_1, we are safe in making the assumption required to derive Eq. (4.13).

If we have established a horizontal strip of analyticity for $\psi_j(z)$, we get a vertical strip of analyticity for $M_j(\beta)$ by dint of Eq. (3.19). If $\psi_j(z)$ is analytic for $a \leqslant \mathrm{Im}\, z \leqslant b$, then $M_j(\beta)$ is analytic for $2\pi a \leqslant \mathrm{Re}\, \beta \leqslant 2\pi b$. With the minimal assumptions of the last paragraph, this argument shows that $M_0(\beta)$ is analytic at least over $0 \leqslant \mathrm{Re}\, \beta \leqslant 1$.

Moreover, from the strip for $M_0(\beta)$, we can get a strip of analyticity for $G(\beta)$ by rewriting Eq. (5.10) as

$$G\left(\beta - \frac{1}{2}\right) = \frac{\log M_0(\beta)}{\beta(\beta - 1)} . \tag{A10}$$

Appearances not withstanding, the right-hand side of this equation is analytic in the same vertical strip where $M(\beta)$ is. In spite of the denominator, there are no poles at $\beta = 0$ or 1; $M_0(\beta)$ must be unity at these points for proper normalization of the densities (see Subsection 5.A), so the logarithm vanishes. The multiple-valued nature of the logarithm causes no problems either, since $M_0(\beta)$ does not go to zero in the strip. The branch cut lies outside the strip, so the strip lies entirely in a single branch of the logarithm.

Thus $G(\beta - \frac{1}{2})$ is analytic in the strip $0 \leqslant \mathrm{Re}\, \beta \leqslant 1$, and $G(\beta)$ itself is analytic in the shifted strip $-\frac{1}{2} \leqslant \beta \leqslant \frac{1}{2}$. Specifically, this guarantees that $G(2\pi i \xi \pm \frac{1}{2})$ is analytic for all real ξ, so the function $H(\xi)$ defined in Eq. (5.25) is analytic for all real ξ.

APPENDIX B: AN ASYMPTOTIC EXPANSION FOR THE AREA UNDER AN ROC CURVE OF THE LIKELIHOOD RATIO

Beginning with Eq. (3.17) and using Eq. (4.6) together with $\psi_1^*(\xi) = \psi_1(-\xi)$, we may write

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2\pi i} \mathscr{P} \int_{-\infty}^{\infty} \frac{\mathrm{d}\xi}{\xi}\, \psi_0(\xi)\, \psi_0\left(-\xi + \frac{i}{2\pi}\right). \tag{B1}$$

Let $z = \xi + i\eta$ as in Appendix A. From that appendix we know that $\psi_0(z)$ is analytic for $0 \leqslant \mathrm{Im}(z) \leqslant 1/2\pi$. Therefore $\psi_0(-z + i/2\pi)$ is analytic for $0 \leqslant \mathrm{Im}(-z + i/2\pi) \leqslant 1/2\pi$, which simplifies to $0 \leqslant \mathrm{Im}(z) \leqslant 1/2\pi$ also. Thus the integrand in Eq. (B.1) is analytic in this region except for the simple pole at $\xi = 0$. Since $\psi_0(0) = \psi_1(0) = 1$, the residue at this pole is 1. Let C be the contour that traverses the negative real axis in the z plane from $z = -\infty$ to $z = -\epsilon$, follows the upper semicircle of radius ϵ centered at the origin to $z = \epsilon$, and then continues over the rest of the real axis from $z = \epsilon$ to $z = \infty$. Then, letting $\epsilon \to 0$, we have

$$\mathrm{AUC} = \frac{1}{2} + \frac{1}{2\pi i} \int_C \frac{\mathrm{d}z}{z}\, \psi_0(z) \times \psi_0\left(-z + \frac{i}{2\pi}\right) + \left(\frac{1}{2\pi i}\right)\pi i . \tag{B2}$$

The last term cancels the contribution from the integral over the semicircle, so that what is left is the principal-value integral in Eq. (B.1).

Now let C' be the contour that traverses the horizontal line $\mathrm{Im}(z) = 1/4\pi$ from $\mathrm{Re}(z) = -\infty$ to $\mathrm{Re}(z) = \infty$. Since the integrand is analytic between these two contours, we may replace C with C' in Eq. (B2). This gives

$$\mathrm{AUC} = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\mathrm{d}\xi}{\xi + \frac{i}{4\pi}}\, \psi_0\left(\xi + \frac{i}{4\pi}\right) \times \psi_0\left(-\xi + \frac{i}{4\pi}\right). \tag{B3}$$

By use of Eq. (5.2), this is

$$\mathrm{AUC} = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\mathrm{d}\xi}{\xi + i/(4\pi)} F(\xi)F(-\xi). \tag{B4}$$

From Eq. (5.5) we may express this in terms of $T(\xi)$. Using the fact that $F(-\xi) = F^*(\xi)$, we know that $T(-\xi) = T^*(\xi)$. This means that $T(-\xi) + T(\xi) = 2\,\mathrm{Re}\, T(\xi)$ and that this is an even function of ξ. Now we have

$$\mathrm{AUC} = 1 + \frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{\mathrm{d}\xi}{\xi + i/(4\pi)} \times \exp\left[2\left(\xi^2 + \frac{1}{16\pi^2}\right)\mathrm{Re}\, T(\xi)\right]. \tag{B5}$$

A little algebra gives us

$$\frac{1}{\xi + i/(4\pi)} = \frac{\xi}{\xi^2 + 1/(16\pi^2)} - \frac{1}{4\pi}\frac{i}{\xi^2 + 1/(16\pi^2)} . \tag{B6}$$

The odd part of the integrand drops out, and we are left with

$$\mathrm{AUC} = 1 - \frac{1}{4\pi^2}\int_0^\infty \frac{\mathrm{d}\xi}{\xi^2 + 1/(16\pi^2)} \times \exp\left[2\left(\xi^2 + \frac{1}{16\pi^2}\right)\mathrm{Re}\ T(\xi)\right]. \tag{B7}$$

To express this in terms of $G(z)$, we use Eq. (5.7) and the fact that $\mathrm{Re}\ T(\xi) = \mathrm{Re}\ T(-\xi)$ to get

$$\mathrm{AUC} = 1 - \frac{1}{4\pi^2}\int_0^\infty \frac{\mathrm{d}\xi}{\xi^2 + 1/(16\pi^2)} \times \exp\left[-8\pi^2\left(\xi^2 + \frac{1}{16\pi^2}\right)\mathrm{Re}\ G(2\pi i\xi)\right]. \tag{B8}$$

From Section 5 we know that $G(0)$ is real and positive. Let $\tilde{G}(z) = G(z) - G(0)$. Then we may write

$$\mathrm{AUC} = 1 - I_0 + I, \tag{B9}$$

with

$$I_0 = \frac{1}{4\pi^2}\int_0^\infty \frac{\mathrm{d}\xi}{\xi^2 + 1/16\pi^2} \times \exp\left[-8\pi^2\left(\xi^2 + \frac{1}{16\pi^2}\right)G(0)\right], \tag{B10}$$

$$I = \frac{1}{4\pi^2}\int_0^\infty \mathrm{d}\xi \exp\left[-8\pi^2\left(\xi^2 + \frac{1}{16\pi^2}\right)G(0)\right]b(\xi), \tag{B11}$$

$$b(\xi) = \frac{1 - \exp\{-8\pi^2[\xi^2 + 1/(16\pi^2)]\mathrm{Re}\ \tilde{G}(2\pi i\xi)\}}{\xi^2 + 1/(16\pi^2)}. \tag{B12}$$

If we define

$$h(\xi) = \exp(-8\pi^2\xi^2), \tag{B13}$$

then

$$I = \exp\left[-\frac{G(0)}{2}\right]\int_0^\infty \mathrm{d}\xi h[\sqrt{G(0)}\xi]b(\xi). \tag{B14}$$

Note that $b(0) = 0$ and $b'(0) = 0$. This means that the magnitude of I can be expected to be small if $G(0)$ is not too small and $b''(0)$ is not too large.

The integral I_0 can be computed analytically[22] to give

$$\mathrm{AUC} = \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}[\tfrac{1}{2}\sqrt{2G(0)}] + I \tag{B15}$$

For I there is an asymptotic expansion as $\sqrt{G(0)} \to \infty$.[23] To compute this expansion, we need the Maclaurin series for $b(\xi)$,

$$b(\xi) = \sum_{n=1}^\infty b_n\xi^{2n}, \tag{B16}$$

and the Mellin transform of h at the odd integers,

$$M_h(2n + 1) = \int_0^\infty \frac{\mathrm{d}\xi}{\xi}\ h(\xi)\xi^{2n+1}. \tag{B17}$$

The asymptotic expansion for I is then given by

$$I \sim \frac{1}{4\pi^2}\exp\left[-\frac{G(0)}{2}\right]\sum_{n=1}^\infty b_n M_h(2n + 1)\,[\sqrt{G(0)}]^{-2n-1}. \tag{B18}$$

Assuming that this expansion is valid, we have for any given integer $N \geqslant 1$ positive numbers Δ_N and E_N such that, for $\sqrt{G(0)} > \Delta_N$,

$$\left|4\pi^2 I\exp\left[\frac{G(0)}{2}\right] - \sum_{n=1}^{N-1} b_n M_h(2n + 1)[\sqrt{G(0)}]^{-2n-1}\right| < E_N[\sqrt{G(0)}]^{-2N-1}. \tag{B19}$$

(For $N = 1$ there is no sum on the left). Generally speaking, we expect Δ_N to increase without bound as N increases so that, for a fixed $G(0)$, there are only a finite number of N with $\sqrt{G(0)} > \Delta_N$. This is because asymptotic series do not necessarily converge. They are nevertheless useful approximations in many circumstances.

The $N = 1$ case in relation (B19) gives

$$\left|\mathrm{AUC} - \left(\frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[\frac{1}{2}\sqrt{2G(0)}\right]\right)\right| \leqslant \frac{E_1}{4\pi^2[\sqrt{G(0)}]^3}\exp\left[-\frac{1}{2}G(0)\right] \tag{B20}$$

for $\sqrt{G(0)} > \Delta_1$. The values for E_1 and Δ_1 depend ultimately on the probability density $p_0(\Lambda)$. Clearly, for large enough $G(0)$, the error-function expression provides a good approximation to the AUC.

To compute this asymptotic expansion explicitly, we use[22]

$$M_h(2n + 1) = \frac{1 \times 2 \times \cdots \times (2n - 1)}{4(4\pi)^{2n}}\sqrt{\frac{1}{2\pi}}, \tag{B21}$$

$$\mathrm{Re}\ \tilde{G}(2\pi i\xi) = -2\pi^2 G''(0)\xi^2 + \frac{2\pi^4}{3}G^{(4)}(0)\xi^4 + \ \ldots\ . \tag{B22}$$

The latter expansion may be used to compute the numbers b_n. For example, $b_1 = -16\pi^4 G''(0)$. Using this value in the first term of the asymptotic expansion for I gives us an approximation

$$\mathrm{AUC} \approx \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[\frac{1}{2}\sqrt{2G(0)}\right] - \frac{G''(0)}{8}\exp\left[-\frac{G(0)}{2}\right][\sqrt{G(0)}]^{-3}. \tag{B23}$$

We would expect this to be good for large $G(0)$. There are also indications that the approximation

$$\mathrm{AUC} \approx \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}[\tfrac{1}{2}\sqrt{2G(0)}] \tag{B24}$$

is good for small $G(0)$.

ACKNOWLEDGMENTS

The authors have benefited greatly from discussion with many people. We especially thank Michel Defrise for pointing out the asymptotic properties of log likelihoods;

Jack Denny for calling attention to relevant segments of the statistics literature; Brandon Gallas for advice on Monte Carlo simulation of likelihood ratios; Robert Wagner for convincing us that SNR_λ was more likely to be useful than SNR_Λ and for introducing us to Gertrude Stein; and Kyle Myers, Charles Metz, and Richard Swensson for very careful and critical reading of the manuscript. This work was supported by the National Institutes of Health (NIH) through grant RO1 CA52643, but it does not represent an official position of the NIH.

Address correspondence to Harrison H. Barrett, Department of Radiology, Radiology Research Laboratory, University of Arizona, Tucson, Arizona 85724. E-mail address, barrett@radiology.arizona.edu.

REFERENCES

1. H. H. Barrett, “Objective assessment of image quality: effects of quantum noise and object variability,” J. Opt. Soc. Am. A **7**, 1266–1278 (1990).
2. H. H. Barrett, J. L. Denny, R. F. Wagner, and K. J. Myers, “Objective assessment of image quality. II. Fisher information, Fourier crosstalk, and figures of merit for task performance,” J. Opt. Soc. Am. A **12**, 834–852 (1995).
3. H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, “Noise properties of the EM algorithm I: theory,” Phys. Med. Biol. **39**, 833–846 (1994).
4. D. W. Wilson, B. M. W. Tsui, and H. H. Barrett, “Noise properties of the EM algorithm II: Monte Carlo simulations,” Phys. Med. Biol. **39**, 847–872 (1994).
5. H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), Vol. I.
6. J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).
7. J. L. Melsa and D. L. Cohn, *Decision and Estimation Theory* (McGraw-Hill, New York, 1978).
8. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. (Academic, New York, 1990).
9. H. H. Barrett, T. A. Gooley, K. A. Girodias, J. P. Rolland, T. A. White, and J. Yao, “Linear discriminants and image quality,” Image Vision Comput. **10**, 451–460 (1992).
10. H. H. Barrett, J. Yao, J. Rolland, and K. J. Myers, “Model observers for assessment of image quality,” Proc. Natl. Acad. Sci. USA **90**, 9758–9765 (1993).
11. C. E. Metz, Department of Radiology, University of Chicago, Chicago, Ill. (personal communication, 1998).
12. R. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1965).
13. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, 4th ed., Formula 3.954 (Academic, New York, 1965).
14. H. H. Barrett and C. K. Abbey, “Bayesian detection of random signals on random backgrounds,” in *Information Processing in Medical Imaging*, Proceedings of the 15th International Conference, IPMI 97, Poultney, Vt., June 9–13, 1997, Lecture Notes in Computer Science, J. Duncan and G. Gindi, eds. (Springer-Verlag, Berlin, 1997).
15. D. Middleton, *An Introduction to Statistical Communication Theory* (McGraw-Hill, New York, 1960).
16. R. M. Swensson and D. M. Green, “On the relations between random walk models for two-choice response times,” J. Math. Psychol. **15**, 282–291 (1977).
17. A. N. Shiryayev, *Probability* (Springer-Verlag, New York, 1984).
18. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics* (Wiley, New York, 1966).
19. L. Rade and B. Westergren, *Beta Mathematics Handbook*, 2nd ed. (CRC Press, Boca Raton, Fla., 1990).
20. R. V. Churchill, J. W. Brown, and R. F. Verkey, *Complex Variables and Applications*, 3rd ed. (McGraw-Hill, New York, 1974).
21. S. Lang, *Real and Functional Analysis*, 3rd ed. (Springer-Verlag, New York, 1993).
22. R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (American Mathematical Society, New York, 1934), Vol. 19.
23. R. Strichartz, *A Guide to Distribution Theory and Fourier Transforms* (CRC Press, Boca Raton, Fla., 1994).