Published in final edited form as:

J Opt Soc Am A Opt Image Sci Vis. 2015 July 01; 32(7): 1288–1301.

Shannon information and ROC analysis in imaging

Eric Clarkson and Johnathan B. Cushing*

College of Optical Sciences, The University of Arizona, 1630 East University Boulevard, Tucson, Arizona 85721, USA

Abstract

Shannon information (SI) and the ideal-observer receiver operating characteristic (ROC) curve are two different methods for analyzing the performance of an imaging system for a binary classification task, such as the detection of a variable signal embedded within a random background. In this work we describe a new ROC curve, the Shannon information receiver operator curve (SIROC), that is derived from the SI expression for a binary classification task. We then show that the ideal-observer ROC curve and the SIROC have many properties in common, and are equivalent descriptions of the optimal performance of an observer on the task. This equivalence is described mathematically by an integral transform that maps the ideal-observer ROC curve onto the SIROC. This then leads to an integral transform relating the minimum probability of error, as a function of the odds against a signal, to the conditional entropy, as a function of the same variable. This last relation then gives us the complete mathematical equivalence between ideal-observer ROC analysis and SI analysis of the classification task for a given imaging system. We also find that there is a close relationship between the area under the ideal-observer ROC curve, which is often used as a figure of merit for imaging systems and the area under the SIROC. Finally, we show that the relationships between the two curves result in new inequalities relating SI to ROC quantities for the ideal observer.

1. INTRODUCTION

Shannon information (SI) has been used in the theory of signal detection or binary classification in imaging as a way to quantify the quality of the data being generated by the imaging system for particular detection or classification tasks [1–4]. Another approach to this image quality problem is receiver operating characteristic (ROC) analysis, which also has a long history as a method to analyze imaging data for signal detection and binary classification [5–13]. One way to relate these two quantities is by using Fano's inequality on a 2-alternative forced choice (2AFC) test. Fano's inequality relates SI to the probability of error for the ideal observer in a classification task. The probability of error for the ideal observer in the 2AFC test is one, minus the area under the ideal observer ROC curve (AUC). However, the SI for the 2AFC test is not the same as the SI for the actual classification task. In addition, Fano's inequality gives only a lower bound for the probability of error, and thus an upper bound for the ideal observer AUC. We will show that the SI for the binary

*Corresponding author: jcushing@optics.arizona.edu.

OCIS codes: (110.2960) Image analysis; (110.3000) Image quality assessment

classification task and the ROC curve are much more closely related via an integral transform that converts a probability of error function to a conditional entropy function.

One practical benefit of this relation is that we can quantify the effect on the minimum probability of error of increasing the SI and vice versa. We also believe this will lead to improved method for computing both quantities by transferring computational techniques between information theory and ROC analysis. Additional practical benefits are the inequalities derived below relating information quantities to ROC quantities. We hope this will stimulate further work relating communication theory, where SI is dominant, to imaging, where ROC plays a more prominent role. This desire for increased communication between the two fields is a primary motivation for the work discussed below.

In Section 2, we establish some notation to be used in the remaining sections. In Section 3, we define the ROC curve for the ideal observer in terms of the false negative fraction (FNF) and the false positive fraction (FPF) as functions of the threshold that minimize the probability of error. This threshold depends on the prior probabilities of the two classes and will be treated throughout this work as a variable upon which all information theoretic functions depend. We will show that the ROC curve for the likelihood ratio, the test statistic for the ideal observer, has certain special properties that are not shared with the ROC curves for other observers. In Section 4, we define SI in terms of the likelihood ratio and the threshold variable. We use this description to define the Shannon information receiver operator curve (SIROC), which has many of the same special properties of the ideal ROC curve. In Sections 5 and 6, we derive expressions for the areas under the ROC curve and SIROC in terms of the square integrals of the functions describing the abscissa of the corresponding curves. In Section 7, we show how the ROC curve and SIROC are related, and define the integral transforms that map one on to the other. These results also give us an integral transform relating the minimum probability of error to the conditional entropy of the class variable given the image data. In Section 8, we show how the unitary Mellin transform can be used to derive alternate expressions for the areas under the ROC curve and SIROC which reveal how closely these two quantities are related. We then derive some inequalities connecting SI to ROC analysis in Section 9. In Section 10, we show that the area under the SIROC is also an integral transform, relating the average cost as a function of threshold to the conditional entropy function. In Section 11, we show some simulation results, and in Section 12, we summarize and briefly discuss the extension of this work to multi-class classification problems, which will be the subject of a later paper.

2. SOME NOTATION

We will formulate the binary classification task as follows: Is the data vector $\mathbf{g} \in \mathbb{R}^M$ produced by the imaging system drawn from the signal present PDF $p_{r1}(\mathbf{g})$ or from the signal absent PDF $p_{r0}(\mathbf{g})$? We call these options Hypothesis 1 ($H1$) and Hypothesis 0 ($H0$). We assume these PDFs take into account random variations in the objects being imaged, as well as measurement noise from the imaging system itself. Assume prior probabilities P_1 and $P_0 = 1 - P_1$ for the two hypotheses. We will call the options for the end result of the classification Decision 1 ($D1$) and Decision 0 ($D0$). The classifier with the minimum probability of error computes the likelihood ratio

$$\Lambda(\mathbf{g}) = \frac{pr_1(\mathbf{g})}{pr_0(\mathbf{g})} \quad (1)$$

and a threshold

$$y = \frac{P_0}{P_1} = \frac{1}{P_1} - 1. \quad (2)$$

If $\Lambda(\mathbf{g}) > y$, then the decision is $D1$. Otherwise the decision is $D0$. Notice that y gives the odds against $H1$ for the given prior probabilities. As P_1 varies from 0 to 1, the variable y goes from 0 to ∞ . It will be convenient to use y instead of P_1 to describe the prior probabilities. Note that

$$P_1 = \frac{1}{1+y} \quad \text{and} \quad P_0 = \frac{y}{1+y}. \quad (3)$$

For the expression of a function $f(\Lambda(\mathbf{g}))$ under Hypothesis j , we will use the notation

$$\langle f(\Lambda(\mathbf{g})) \rangle_j = \int_{\mathbb{R}^M} f(\Lambda(\mathbf{g})) pr_j(\mathbf{g}) d^M g. \quad (4)$$

By a slight abuse of notation we will also write

$$\langle f(\Lambda) \rangle_j = \int_0^\infty f(\Lambda) pr_j(\Lambda) d\Lambda, \quad (5)$$

where $pr_j(\Lambda)$ is the PDF for the random variable $\Lambda = \Lambda(\mathbf{g})$ under Hypothesis j . When the function notation pr_j is used with a scalar argument, it is this PDF that we will be referring to. An important relation between the two PDFs for the likelihood ratio is that $pr_1(\Lambda) = \Lambda pr_0(\Lambda)$ [14]. Therefore, we have a relation between expectations: $\langle f(\Lambda) \rangle_1 = \langle \Lambda f(\Lambda) \rangle_0$. We will be using these two relations often.

3. IDEAL OBSERVER ROC CURVE

The minimum probability of error is a function of the prior probabilities. In terms of the threshold variable, this function is given by

$$P_e(y) = \frac{y}{1+y} \langle \text{step}[\Lambda - y] \rangle_0 + \frac{1}{1+y} \langle \text{step}[y - \Lambda] \rangle_1. \quad (6)$$

The first angle bracket is the false negative fraction (FNF) and the second angle bracket is the false positive fraction (FPF). It is important to remember that this equation gives the minimum possible value for the probability of error for the given prior probabilities. No other classifier will result in a smaller probability of error for the given imaging system. The imaging system is considered to be fixed throughout our discussion, so the threshold is the only variable that will appear in all functions.

One way to plot an ROC curve is to plot FNF versus FPF as y varies. The more common presentation of the ROC curve is to plot $\text{TPF} = 1 - \text{FNF}$ versus FPF. However, for our purposes it will be convenient to use the FNF versus FPF version to relate ROC analysis to Shannon information. To discuss this curve we set

$$p(y) = \langle \text{step}[\Lambda - y] \rangle_0 \quad \text{and} \quad q(y) = \langle \text{step}[y - \Lambda] \rangle_1. \quad (7)$$

The ROC curve is then the parametric curve given by $(p(y), q(y))$, and the minimum probability of error is given by

$$P_e(y) = \frac{y}{1+y} p(y) + \frac{1}{1+y} q(y). \quad (8)$$

The two functions $p(y)$ and $q(y)$ are determined by the imaging system.

When $y = 0$, we have one endpoint of the ROC curve at $p(0) = 1$ and $q(0) = 0$. At the other endpoint, as $y \rightarrow \infty$ we have $p(y) \rightarrow 0$ and $q(y) \rightarrow 1$. For derivatives, we have

$$\frac{dp(y)}{dy} = -pr_0(y) \quad \text{and} \quad \frac{dq(y)}{dy} = pr_1(y). \quad (9)$$

These equations imply that the slope of the tangent line at the point $(p(y), q(y))$ on the ROC curve is $-y$, which then implies that the ROC curve is concave upward. Therefore, at the threshold y , the FNF is $q(y)$, the FPF is $p(y)$, and the slope of the ROC curve at that point is the negative threshold. The tangent line at $(p(y), q(y))$ is the line of constant probability of error and is given by the points (p, q) that satisfy

$$P_e(y) = \frac{y}{1+y}p + \frac{1}{1+y}q. \quad (10)$$

The p and q intercepts of this tangent line are given by

$$p_0(y) = \left(\frac{1+y}{y}\right) P_e(y) \quad \text{and} \quad q_0(y) = (1+y)P_e(y). \quad (11)$$

These equations tell us that the p intercept is the minimum probability of error for the given threshold divided by the probability of no signal, and the q intercept is the same minimum probability of error divided by the probability of a signal. Therefore we have

$$\frac{1}{P_e(y)} = \frac{1}{p_0(y)} + \frac{1}{q_0(y)}. \quad (12)$$

Thus, the minimum probability of error $P_e(y)$ can be found from the ROC curve by finding the point where the slope is $-y$, and then drawing the tangent line. In this sense, the ROC curve completely determines the minimum probability of error as a function of the prior probabilities.

4. SHANNON INFORMATION ROC CURVE

The Shannon information for this classification task is given by [15–18]

$$I = P_1 \left\langle \ln \left[\frac{pr_1(\mathbf{g})}{P_1 pr_1(\mathbf{g}) + P_0 pr_0(\mathbf{g})} \right] \right\rangle_1 + P_0 \left\langle \ln \left[\frac{pr_0(\mathbf{g})}{P_1 pr_1(\mathbf{g}) + P_0 pr_0(\mathbf{g})} \right] \right\rangle_0. \quad (13)$$

The expectations are finite since I is bounded by $\ln 2$. As a function of y , this Shannon information is given by

$$I(y) = \frac{1}{1+y} \left\langle \ln \left[\frac{\Lambda(1+y)}{\Lambda+y} \right] \right\rangle_1 + \frac{y}{1+y} \left\langle \ln \left[\frac{1+y}{\Lambda+y} \right] \right\rangle_0. \quad (14)$$

The source or state variable s is a random variable that satisfies $s = 0$ with probability P_0 , and $s = 1$ with probability P_1 . The entropy of the source variable is given by $H = -P_1 \ln P_1 - P_0 \ln P_0$. As a function of y , this entropy is given by

$$H(y) = -\frac{1}{1+y} \ln \left(\frac{1}{1+y} \right) - \frac{y}{1+y} \ln \left(\frac{y}{1+y} \right). \quad (15)$$

The conditional entropy of the source variable is a quantity we will be interested in, and it is given by $C_e(y) = H(y) - I(y)$. Explicitly, we have the formula

$$C_e(y) = -\frac{1}{1+y} \left\langle \ln \left(\frac{\Lambda}{\Lambda+y} \right) \right\rangle_1 - \frac{y}{1+y} \left\langle \ln \left(\frac{y}{\Lambda+y} \right) \right\rangle_0. \quad (16)$$

The notation for conditional entropy here is nonstandard, but useful for our purposes. For example, this notation emphasizes the similarity between the conditional entropy for a detection problem and the minimum probability of error for the same problem. The conditional entropy is always a nonnegative quantity.

To define the SIROC, we set

$$u(y) = -\left\langle \ln \left(\frac{y}{\Lambda+y} \right) \right\rangle_0 \quad v(y) = -\left\langle \ln \left(\frac{\Lambda}{\Lambda+y} \right) \right\rangle_1 \quad (17)$$

and plot the curve $(u(y), v(y))$ as y varies from 0 to ∞ . This is a curve in the first quadrant of (u, v) space and has many properties in common with the ordinary ROC curve. The conditional entropy is now given by

$$C_e(y) = \frac{1}{1+y} v(y) + \frac{y}{1+y} u(y). \quad (18)$$

We can show that the quantity $u(y)$ is the negative expectation of the log of the posterior probability $Pr(H_1|g)$, under hypothesis 1. Similarly $v(y)$ is the negative expectation of the log of the posterior probability $Pr(H_0|g)$, under hypothesis 0.

We will describe some of the basic properties of the SIROC. When $y \rightarrow 0$, we have $u(y) \rightarrow \infty$ and $v(y) \rightarrow 0$. At the other endpoint, as $y \rightarrow \infty$ we have $u(y) \rightarrow 0$ and $v(y) \rightarrow \infty$. For derivatives we have

$$\frac{du(y)}{dy} = -\frac{1}{y} \left\langle \frac{\Lambda}{\Lambda+y} \right\rangle_0 \quad \frac{dv(y)}{dy} = \left\langle \frac{1}{\Lambda+y} \right\rangle_1 = \left\langle \frac{\Lambda}{\Lambda+y} \right\rangle_0. \quad (19)$$

Therefore, the slope of the SIROC at $(u(y), v(y))$ is $-y$. This again implies that the SIROC is concave upward. The tangent line at the point $(u(y), v(y))$ is the line of constant conditional entropy and is given by the points (u, v) that satisfy

$$C_e(y) = \frac{y}{1+y}u + \frac{1}{1+y}v. \quad (20)$$

The u and v intercepts of this tangent line are given by

$$u_0(y) = \left(\frac{1+y}{y}\right) C_e(y) \quad \text{and} \quad v_0(y) = (1+y)C_e(y). \quad (21)$$

These equations tell us that the u intercept is the conditional entropy for the given threshold divided by the probability of no signal, and the v intercept is the same conditional entropy divided by the probability of a signal. Therefore we have

$$\frac{1}{C_e(y)} = \frac{1}{u_0(y)} + \frac{1}{v_0(y)}. \quad (22)$$

Thus the conditional entropy $C_e(y)$ and, hence, the Shannon information $I(y)$ can be recovered from the SIROC by finding the point on the curve where the slope is $-y$ and then drawing the tangent line. In this sense, the SIROC completely determines the Shannon information as a function of the prior probabilities.

5. AREA UNDER THE ROC CURVE

A figure of merit that is independent of the prior probabilities is the area under the ROC curve A_Λ . This area is related to the area under the usual ROC curve, AUC_Λ , by

$$1 - AUC_\Lambda = \int_0^\infty p(y) \frac{dq(y)}{dy} dy \equiv A_\Lambda. \quad (23)$$

The quantity AUC_Λ is also the probability of a correct answer on a 2AFC test, and is therefore directly measurable for any observer. The smaller the integral A_Λ , the better the observer performs in the classification task. Because of the relation between $p(y)$ and $q(y)$ noted above, we can express this integral in terms of $p(y)$ only. The result we arrive at in the last equation in this section was derived in a slightly different way in a previous publication [19]. We present this derivation because we will use a similar method to get a corresponding result for the area under the SIROC in the next section. We start with

$$\frac{dq(y)}{dy} = -y \frac{dp(y)}{dy}, \quad (24)$$

which we derived above. Integrating both sides, and then integrating by parts on the right, gives us

$$q(y) = -yp(y) + \int_0^y p(y') dy'. \quad (25)$$

This relation shows that the function $q(y)$ is completely determined by the function $p(y)$. Thus, the ROC curve for the likelihood ratio is also determined completely by the false positive probability $p(y)$ as a function of threshold. This fact will be important later.

Next note that

$$\begin{aligned} yp(y) &= y \int_y^\infty p r_0(y') dy' \leq \int_y^\infty y' p r_0(y') dy' \\ &= \int_y^\infty p r_1(y') dy', \end{aligned} \quad (26)$$

so that $yp(y) \rightarrow 0$ as $y \rightarrow \infty$. Therefore the function $p(y)$ satisfies the constraint

$$\int_0^\infty p(y) dy = 1. \quad (27)$$

This constraint will also be useful later.

Returning to A_Λ we can integrate by parts again and find that

$$A_\Lambda = -\int_0^\infty yp(y) \frac{dp(y)}{dy} dy = -\frac{1}{2} yp^2(y) \Big|_0^\infty + \frac{1}{2} \int_0^\infty p^2(y) dy. \quad (28)$$

The first term on the right vanishes at both endpoints, so we have the expression

$$A_\Lambda = \frac{1}{2} \int_0^\infty p^2(y) dy. \quad (29)$$

Therefore, we only need to know the FPF as a function of threshold to compute the ideal observer AUC. It is important to note that this relation only applies when the likelihood ratio is being used as the test statistic. The maximum value for A_Λ occurs when the two data

PDFs are the same, so that $\Lambda(\mathbf{g}) = 1$. In this case, $p(y) = \text{step}(1 - y)$ and $A_\Lambda = 1/2$. The minimum value for A_Λ is zero, which can only occur when the supports of the PDFs $p_{r_1}(\mathbf{g})$ and $p_{r_0}(\mathbf{g})$ do not intersect.

For example, if the signal absent pdf has compact support and the signal present pdf is a shifted version of the signal absent pdf, then the supports of the two pdfs will not intersect if the signal is large enough.

6. AREA UNDER THE SIROC

By analogy, with the ROC curve we can define a figure of merit related to Shannon information that is independent of the prior probabilities by computing the area under the SIROC:

$$A_s = \int_0^\infty u(y) \frac{dv(y)}{dy} dy. \quad (30)$$

We will show that this quantity only depends on the function $u(y)$. As with the ROC curve, we start with

$$\frac{dv(y)}{dy} = -y \frac{du(y)}{dy}, \quad (31)$$

which was derived above. By integrating both sides and integrating by parts on the right we arrive at

$$v(y) = -yu(y) + \int_0^y u(y') dy'. \quad (32)$$

We used the fact that

$$\lim_{y \rightarrow 0} yu(y) = -\lim_{y \rightarrow 0} \left\langle y \ln \left(\frac{y}{\Lambda + y} \right) \right\rangle_0 = 0 \quad (33)$$

to eliminate one of the terms from the integration by parts. This fact will also be important later. Next note that

$$\lim_{y \rightarrow \infty} yu(y) = -\lim_{y \rightarrow \infty} \left\langle y \ln \left(\frac{y}{\Lambda + y} \right) \right\rangle_0 = \langle \Lambda \rangle_0 = 1, \quad (34)$$

which implies that $u(y)$ is not an integrable function. This is a difference between the ROC curve and SIROC, since $p(y)$ integrates to 1.

Returning to A_S , we have

$$A_S = -\int_0^\infty y u(y) \frac{du(y)}{dy} dy = -\frac{1}{2} y u^2(y) \Big|_0^\infty + \frac{1}{2} \int_0^\infty u^2(y) dy. \quad (35)$$

To evaluate the first term on the right in this equation, we can write

$$\begin{aligned} y u^2(y) &= \left[\left\langle \sqrt{y} \ln \left(\frac{y}{\Lambda + y} \right) \right\rangle_0 \right]^2 \\ &= \left[\left\langle \sqrt{y} \ln \left(1 + \frac{\Lambda}{y} \right) \right\rangle_0 \right]^2. \end{aligned} \quad (36)$$

The first form can be used to show that $y u^2(y) \rightarrow 0$ as $y \rightarrow 0$. The second form can be used to show that $y u^2(y) \rightarrow 0$ as $y \rightarrow \infty$. Therefore, we have

$$A_S = \frac{1}{2} \int_0^\infty u^2(y) dy. \quad (37)$$

Thus the function $u(y)$ completely determines the Shannon information function $I(y)$ and the area under the SIROC.

The maximum value for A_S occurs when the two data PDFs are the same. In this case,

$$\max A_S = \frac{1}{2} \int_0^\infty \ln^2 \left(1 + \frac{1}{y} \right) dy. \quad (38)$$

To evaluate this integral, we make the substitution $2w = \ln(1 + 1/y)$. This reduces the integral in question to a tabulated integral and we have

$$\max A_S = \frac{1}{2} \int_0^\infty \frac{4w^2}{2 \sinh^2(w)} dw = \frac{\pi^2}{6}. \quad (39)$$

The minimum value for A_S is zero, which can only occur when the supports of the PDFs $p r_1(\mathbf{g})$ and $p r_0(\mathbf{g})$ do not intersect.

7. SHANNON INFORMATION AND ROC ANALYSIS

We have seen in the sections above that the ideal-observer ROC curve and the SIROC have many properties in common. In this section, we will show that the ROC curve completely determines the SIROC, and vice versa. Thus ROC analysis and Shannon information are equivalent descriptions of the optimum performance that we can expect from an imaging system on a detection task.

First we show that the functions $p(x)$ and $u(y)$ are related by an integral transform. Suppose that $y > 0$ and start with the integral

$$\begin{aligned} \int_0^\infty \frac{p(x)}{x+y} dx &= \int_0^\infty \int_x^\infty \frac{p(x)}{x+y} d\Lambda dx \\ &= \int_0^\infty \int_0^\Lambda \frac{p(x)}{x+y} dx d\Lambda. \end{aligned} \quad (40)$$

The first integral is finite since $x p(x) \rightarrow 0$ as $x \rightarrow \infty$. This fact justifies the change in the order of integration by Fubini's theorem. Proceeding further, we have

$$\int_0^\infty \int_0^\Lambda \frac{p(x)}{x+y} dx d\Lambda = \int_0^\infty \ln \left(\frac{\Lambda+y}{y} \right) p(\Lambda) d\Lambda = u(y). \quad (41)$$

Therefore, we have a relation between the ROC curve and the SIROC given by

Integral Transform I

$$u(y) = \int_0^\infty \frac{p(x)}{x+y} dx. \quad (42)$$

Note that $p(x)$ is the ordinate of the unique point on the ideal-observer ROC curve where the slope is $-x$. Similarly, $u(y)$ is the ordinate of the unique point on the SIROC where the slope is $-y$. Therefore, we can see at this point that the ideal-observer ROC curve, which is derived from the function $p(x)$, determines the SIROC, which is derived from the function $u(y)$. Another way of describing this relationship is that the ideal observer FPF, as a function of the threshold determines the expectation under H_1 of the log of $Pr(H_1|g)$ as a function of the prior probability of a signal. We may also write the relation between these two functions as

$$u(y) = \int_0^\infty \frac{p(x)}{1 + \frac{y}{x}} \frac{dx}{x}, \quad (43)$$

which states that the function $u(y)$ is the Mellin convolution of $p(y)$ with $(1+y)^{-1}$. If we define the Mellin transform of $p(y)$ in the usual way,

$$\mathcal{M}p(z) = \int_0^\infty y^{z-1} p(y) dy, \quad (44)$$

then, since $p(0) = 1$, $p'(y) = 0$, and $yp(y) \rightarrow 0$ as $y \rightarrow \infty$, we know that $\mathcal{M}p(z)$ is analytic in the strip $0 < \text{Re} z < 1$. The Mellin transform of $(1+y)^{-1}$ is also analytic in this strip and is given by $\pi \csc(\pi z)$. Therefore, $\mathcal{M}u(z)$ is analytic in this strip and we have $\sin(\pi z) \mathcal{M}u(z) = \pi \mathcal{M}p(z)$ for $0 < \text{Re} z < 1$. This relation shows the complete equivalence of the Shannon information description of the detection problem and the ideal-observer ROC analysis of the problem since the Mellin transform is an invertible integral operator.

The function $\mathcal{M}p(z)$ may be written as

$$\begin{aligned} \mathcal{M}p(z) &= \int_0^\infty \int_y^\infty y^{z-1} p_0(\Lambda) d\Lambda dy \\ &= \int_0^\infty \left[\int_0^\Lambda y^{z-1} dy \right] p_0(\Lambda) d\Lambda. \end{aligned} \quad (45)$$

Performing the inner integral, we may write this function in terms of the Renyi- z divergence $D_z(p_{r1}||p_{r0})$ between the signal present and signal absent PDFs:

$$\mathcal{M}p(z) = \frac{1}{z} \langle \Lambda^z \rangle_0 = \frac{1}{z} \exp[-(1-z) D_z(p_{r1}||p_{r0})]. \quad (46)$$

This equation shows that, if the function $D_z(p_{r1}||p_{r0})$ is known for $0 < z < 1$, then the ROC curve and SIROC are determined. However, to use the formula for the inverse Mellin transform to recover these curves we would need to evaluate $D_z(p_{r1}||p_{r0})$ for complex z . We will see that complex values for z also enter into formulas for the areas under the ROC curve and SIROC below.

Next we will derive an integral transform relation for the function $p(y)$ and the conditional entropy function $C_e(y)$. We start with the equation

$$u(y) = \frac{d}{dy} [(1+y) C_e(y)], \quad (47)$$

which follows from the slope property of the SIROC noted above. Now we can use the fact that $C_e(0) = 0$ to derive

Integral Transform II

$$(1+y)C_e(y)=\int_0^\infty p(x)\ln\left(1+\frac{y}{x}\right)dx. \quad (48)$$

This equation relates the conditional entropy to the ideal observer FPF. It is interesting in its own right because it shows how the ideal observer FPF as a function of threshold determines the conditional information function and, hence, the Shannon information as a function of the prior probability for a signal.

The third integral transform relates $u(y)$ to $P_e(y)$. We start with the relation between the FPF and the minimum probability of error,

$$p(x)=\frac{d}{dx}[(1+x)P_e(x)], \quad (49)$$

which follows from the slope property of the ROC curve. By integrating the integral transform relating $u(y)$ to $p(x)$ by parts we arrive at

Integral Transform III

$$u(y)=\int_0^\infty \frac{(1+x)P_e(x)}{(x+y)^2}dx. \quad (50)$$

This relation shows how the minimum probability of error function $P_e(y)$ determines the function $u(y)$ which, in turn, determines the SIROC and the Shannon information.

By integrating the third last integral transform with respect to y , we get an integral relation between the conditional entropy function $C_e(y)$ and the minimum probability of error function $P_e(y)$:

Integral Transform IV

$$\frac{1+y}{y}C_e(y)=\int_0^\infty \frac{1+x}{x}P_e(x)\frac{1}{x+y}dx. \quad (51)$$

Thus, the conditional entropy at y is a weighted integral of the minimum probability of error at all values for the odds against the signal:

$$C_e(y)=\int_0^\infty \omega(x,y)P_e(x)dx, \quad (52)$$

and the weighting function is

$$\omega(x, y) = \frac{y(1+x)}{x(1+y)} \left(\frac{1}{x+y} \right). \quad (53)$$

Similar integral formulas relating the minimum probability of error to f-divergences appear in [20]. This particular integral transform, however, is new and the relation between Shannon information and ROC analysis is not discussed in [20].

We can compare this relationship between conditional entropy and probability of error to the well-known Fano inequality. In the current setting, we can state this inequality by defining $y_c(y) = [1 - P_e(y)]/P_e(y)$ and using the function $H(y)$ defined above. In this notation, the Fano inequality is $C_e(y) \leq H(y_c(y))$. Thus, the Fano inequality relates the conditional entropy at the point on the SIROC, where the slope is y to the minimum probability of error at the point on the ideal observer ROC curve with the same slope. The price we pay for this point-to-point mapping between the two curves is an inequality instead of the equalities we have in the integral transform equations. This integral transform tells us that a system change that increases the Shannon information will also result in a decrease in the minimum probability of error over at least some interval in the range of thresholds. Therefore, if we optimize an imaging system using an information-based figure of merit, such as SI or the area under the SIROC, then we can infer that the minimum probability of error will also be optimized. In particular, the area under the SIROC is an information figure of merit that is independent of the prior probabilities on the two hypotheses, a property it shares with the ideal observer AUC.

Finally, by combining this last integral relation with the integral relation between $p(y)$ and $u(y)$, we arrive at a relation between the FNF function $q(y)$ and the function $v(y)$:

Integral Transform V

$$\frac{v(y)}{y} = \int_0^\infty \frac{q(x)}{x} \frac{1}{x+y} dx. \quad (54)$$

Another way of describing this relationship is that the ideal observer FNF, as a function of the threshold, determines the expectation under H_0 of the log of $Pr(H_0|g)$ as a function of the prior probability of a signal. Therefore, there is an integral operator that maps the ideal-observer ROC curve onto the SIROC. Note that this is not a point-to-point mapping between the two curves. The integral operator is given by

Integral Transform VI

$$[u(y), v(y)] = \int_0^\infty \left[p(x), \frac{yq(x)}{x} \right] \frac{1}{x+y} dx, \quad (55)$$

so the entire ideal-observer ROC curve is needed to use this equation to plot each point on the SIROC.

Note that integral relations between Shannon information and estimation error have been derived for Gaussian and Poisson noise models, where the variable of integration is a parameter in the noise model [21,22]. At present, we know of no connection between those integral relations and the one presented here between conditional entropy and minimum probability of error.

8. AREAS UNDER CURVES AND THE UNITARY MELLIN TRANSFORMS

We can use the unitary Mellin transform to derive alternate formulas for the areas under the ideal observer ROC curve and the SIROC. These formulas show how the two areas are related to each other. The unitary Mellin transform operates on functions defined on $\mathbb{R}^+$ and gives functions defined on $\mathbb{R}$. This operator is given by the equation

$$\tilde{\mathcal{M}}f(\alpha) = \frac{1}{\sqrt{2\pi}} \int_0^\infty y^{-\frac{1}{2}+i\alpha} f(y) dy. \quad (56)$$

This is a well-defined operator for the functions $p(y)$ and $u(y)$, since their Mellin transforms are defined in the strip $0 < s < 1$. As its name implies, the unitary Mellin transform is a unitary operator between $L^2(\mathbb{R}^+)$ to $L^2(\mathbb{R})$. Explicitly, we have

$$\int_0^\infty f^2(y) dy = \int_{-\infty}^\infty [\tilde{\mathcal{M}}f(\alpha)]^2 d\alpha. \quad (57)$$

Now we use the definition of the ideal observer FPF function $p(y)$ and a standard property of the Mellin transform that relates the transforms of a function and its derivative to find that

$$\tilde{\mathcal{M}}p(\alpha) = \frac{1}{\frac{1}{2}+i\alpha} \langle \Lambda^{\frac{1}{2}+i\alpha} \rangle_0. \quad (58)$$

Therefore, the area under the ideal observer ROC curve is given by

$$A_\Lambda = \frac{1}{4\pi} \int_{-\infty}^\infty \left| \langle \Lambda^{\frac{1}{2}+i\alpha} \rangle_0 \right|^2 \frac{d\alpha}{\frac{1}{4}+\alpha^2}. \quad (59)$$

This expression was also derived in [14] using other means, and used in [21,23].

We can perform a similar calculation for the SIROC. Using the fact that $\sin[\pi(1/2 + i\alpha)] = \cosh(\pi\alpha)$, we have the relation $\pi \tilde{\mathcal{H}} p(\alpha) = \cosh(\pi\alpha) \tilde{\mathcal{H}} u(\alpha)$. The unitary property then gives us

$$A_S = \frac{\pi}{4} \int_{-\infty}^{\infty} \left| \left\langle \Lambda^{\frac{1}{2} + i\alpha} \right\rangle_0 \right|^2 \operatorname{sech}^2(\pi\alpha) \frac{d\alpha}{\frac{1}{4} + \alpha^2}. \quad (60)$$

This relation also gives us an expression for the difference between

$$\pi^2 A_\Lambda - A_S = \frac{\pi}{4} \int_{-\infty}^{\infty} \left| \left\langle \Lambda^{\frac{1}{2} + i\alpha} \right\rangle_0 \right|^2 \tanh^2(\pi\alpha) \frac{d\alpha}{\frac{1}{4} + \alpha^2}. \quad (61)$$

Therefore, if high frequencies (large α) do not contribute very much to the ideal observer AUC, then $\pi^2 A_S$ will give a good approximation to A_Λ .

9. INEQUALITIES

To get an inequality relating conditional entropy to ideal observer AUC, we apply the Schwartz inequality to the integral relation between $C_e(y)$ and $p(y)$. This gives us

$$(1+y)C_e(y) \leq \sqrt{2A_\Lambda \int_0^\infty \ln^2 \left(1 + \frac{y}{x} \right) dx}. \quad (62)$$

The integral has essentially been computed above and is given by

$$\int_0^\infty \ln^2 \left(1 + \frac{y}{x} \right) dx = y \int_0^\infty \ln^2 \left(1 + \frac{1}{x} \right) dx = \left(\frac{\pi^2}{3} \right) y. \quad (63)$$

Therefore, we have an inequality relating the conditional entropy to the ideal observer AUC:

$$(1+y)C_e(y) \leq \pi \sqrt{\frac{2A_\Lambda y}{3}}. \quad (64)$$

In terms of the prior probabilities, this inequality is

Inequality I

$$C_e(y) \leq \pi \sqrt{\frac{2A_\Lambda P_1 P_0}{3}}, \quad (65)$$

where $y = P_0/P_1$ as usual. Therefore, the ideal observer AUC sets an upper bound for the conditional entropy and, thus, a lower bound for the Shannon information. We have previously derived a relationship between the ideal observer AUC and Fisher information for signal parameters when the two classes represent small differences in those parameters [22,24,25]. In this situation, we can therefore also get an inequality that relates the conditional entropy to the Fisher information for the signal parameters.

For the second inequality, we can use the two integral expressions for areas under the two ROC curves in the previous section to derive

Inequality II

$$A_s \leq \pi^2 A_\Lambda. \quad (66)$$

This implies that the area under the SIROC sets an upper bound for the area under the usual TPF versus FPF version of the ideal observer ROC curve and, thus, on the area under any such ROC curve for the given detection task.

For the third inequality, we can use the triangle formed by a tangent line to the SIROC and the two axes to get an inequality for the conditional entropy:

$$\frac{(1+y)^2}{2y} C_e^2(y) < A_s. \quad (67)$$

This inequality also can be written as

Inequality III

$$C_e(y) < \sqrt{2A_s P_1 P_0}. \quad (68)$$

If we combine the second and third inequalities, we get one similar to the first one, but without the 2/3 factor.

For the fourth inequality, a similar geometric argument leads to

$$\frac{(1+y)^2}{2y} P_e^2(y) < A_\Lambda, \quad (69)$$

which also can be written as

Inequality IV

$$P_e(y) < \sqrt{2A_\Lambda P_1 P_0}. \quad (70)$$

We will examine how tight some of these inequalities are for Gaussian statistics in the simulation section below.

10. SHANNON INFORMATION AND AVERAGE COST

When computing the probability of error, we are assigning a cost of 1 to FPF and FNF events, and a cost of 0 to TPF and TNF events. The probability of error as defined above is then the average cost, sometimes called the Bayesian risk. We can extend these results to consider a general cost matrix:

$$\mathbf{C} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \quad (71)$$

which assigns a cost C_{ij} to D_i when H_j is true. The average cost or Bayesian risk is then

$$\bar{C} = \sum_{i,j=0}^1 C_{ij} Pr(D_i | H_j) P_j. \quad (72)$$

This quantity also can be written as

$$\bar{C} = (C_{01} - C_{11}) Pr(D_0 | H_1) P_1 + (C_{10} - C_{00}) Pr(D_1 | H_0) P_0 + C_{11} P_1 + C_{00} P_0. \quad (73)$$

We will assume that $C_{01} - C_{11} > 0$ and $C_{10} - C_{00} > 0$. In other words, we assume that it costs more to make a mistake than to get it right when the signal is present and when it is absent. Therefore, any classifier that minimizes the average cost associated with the original matrix also minimizes the average cost associated with the modified cost matrix,

$$\mathbf{C}' = \begin{bmatrix} 0 & C_{01}-C_{00} \\ C_{10}-C_{11} & 0 \end{bmatrix} = \begin{bmatrix} 0 & C'_{01} \\ C'_{10} & 0 \end{bmatrix}. \quad (74)$$

We will assume that the cost matrix has this form and drop the primes in the following. It is important to note that the process of going from $\mathbf{C}$ to $\mathbf{C}'$ is not reversible, so if the actual average cost is important to know, then we must use the original cost matrix. However, there are applications where there is no cost associated with getting the classification right, in which case $\mathbf{C} = \mathbf{C}'$. In any case, for the purposes of system design, we are interested in minimizing the average cost and therefore we may use $\mathbf{C}'$ in place of $\mathbf{C}$.

The classifier with the minimum average cost computes the likelihood ratio $\Lambda(\mathbf{g})$ and a threshold:

$$x = \frac{C_{10}P_0}{C_{01}P_1} = R \frac{P_0}{P_1}. \quad (75)$$

If $\Lambda(\mathbf{g}) > x$, then the decision is $D1$. Otherwise, the decision is $D0$. As P_0 varies from 0 to 1, the variable x goes from 0 to ∞ . We will call R the error cost ratio. This number is the ratio of the cost of a False Positive response to the cost of a False Negative response.

We will use x as a variable in the average cost and consider the cost matrix and, hence, R to be fixed. Thus, varying x corresponds to changing the prior probabilities. Note that

$$P_1 = \frac{R}{R+x} \quad \text{and} \quad P_0 = \frac{x}{R+x}. \quad (76)$$

In our notation, the average cost at a particular threshold value is a function of x and is given by

$$\overline{C}(x) = C_{10} \frac{x}{R+x} p(x) + C_{01} \frac{R}{R+x} q(x). \quad (77)$$

It now follows from the equations relating the ROC curve and the SIROC that

$$\frac{1+y}{y} C_e(y) = \frac{1}{C_{10}} \int_0^\infty \frac{R+x}{x} \overline{C}(x) \frac{1}{x+y} dx. \quad (78)$$

We also can write this equation as

Integral Transform V

$$\frac{1+y}{y} C_e(y) = \frac{1}{C_{10}C_{01}} \int_0^\infty \frac{C_{10}+C_{01}x}{x} \bar{C}(x) \frac{1}{x+y} dx. \quad (79)$$

Thus, a weighted integral of the minimum average cost as a function of the threshold for the likelihood ratio will also give us the conditional entropy for the detection task. It then follows that a change in the imaging system that reduces the conditional entropy and, therefore, increases the Shannon information, and corresponds to a reduction in the minimum average cost over at least some interval in the range of prior probabilities.

11. SIMULATIONS: ONE-DIMENSIONAL GAUSSIAN

To explore the SIROC and the ROC curve and their related statistics, two one-dimensional Gaussian models are used. Assigning the signal absent parameter a standard normal distribution and the signal present a Gaussian distribution with a mean and variance that can be set as necessary. In other words, we use the two hypotheses $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(m, \sigma^2)$.

A. ROC Curve and SIROC

We first examine the ROC curve and SIROC when $m = 0$ and σ is varied. We hope to see similar behavior for the two curves. These results are shown in Fig. 1. As can be seen in the figure, the SIROC and the ROC curve are altered in the same way as the signal present variance is altered. In particular, the $\sigma = 0.42$ curve crosses both the $\sigma = 1.41$ curve and the $\sigma = 2.0$ curve for ROC and SIROC.

Next we fix $\sigma = 1$ and vary m . These results are shown in Fig. 2. As the distributions for signal present and signal absent approach equality, the curves move outward.

B. A_Λ and A_S Calculations

The A_Λ and A_S are calculated by numerical integration of ROC curve and SIROC, such as those shown in Figs. 1 and 2. In Fig. 3, we show these two areas when $m = 0$ and σ is varied. They both peak at $\sigma = 1$ as expected, although the peak for the SIROC is smoother than the one for the ROC curve.

In Fig. 4, the areas A_Λ and A_S are generated by fixing $\sigma = 1$ and varying m . Both plots peak at $m = 0$ as expected. Again, the SIROC is smoother at the peak than the ROC curve.

Using the data collected for Fig. 4, we can examine Inequality II relating A_Λ to A_S to see how close A_S is to $\pi^2 A_\Lambda$. This is shown in Fig. 5 when $m = 0$ and σ is varied, and when $\sigma = 1$ and m is varied. We see that the inequality is not very sharp, except when both numbers are small.

C. Conditional Entropy and Probability of Error

In Fig. 6, we show a plot of the minimum probability of error $P_e(y)$ for the case $m = 0.001$ and $\sigma = 1.1$. We can show that the peak occurs at the value of y that satisfies $p(y) = q(y)$.

In Fig. 7, we show a plot of the conditional entropy $C_e(y)$ for the case $m = 0.001$ and $\sigma = 1.1$. We can show that the peak occurs at the value of y that satisfies $u(y) = v(y)$.

Using the data collected for conditional entropy and the ideal observer AUC for a set of signal parameters allows us to check Inequality I, relating conditional entropy to A_Λ . These plots are shown in Fig. 8 in the case where $m = 0$ and σ is varied, and in Fig. 9 in the case where $\sigma = 1$ and m is varied.

The conditional entropy can be compared directly with the minimum probability of error through Integral Transform IV described above. To simulate this relation, the integral was discretized over the range $[0.001, 99.9]$ at intervals of length $\Delta x = 0.1$. In Fig. 10, we show plots of both sides of the integral relation for $m = 0.001$ and $\sigma = 1.1$. The plot shows good agreement over a large range of values for y .

12. SIMULATIONS: MULTIVARIATE GAUSSIAN

Results from a single dimension sometimes do not extrapolate well to higher dimensions. To examine the SIROC and ROC relations in higher dimensions, multivariate Gaussian distributions were also simulated. For this simulation, five dimensions were used, with hypotheses defined by $H_0: \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1: \mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{K})$. We chose $\mathbf{m} = [0, m, m, m, 0]^T$ and

$$\mathbf{K}_{ij} = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (i-j)^2 \right]. \quad (80)$$

A. ROC Curve and SIROC Creation

In Fig. 11, we show ROC and SIROC plots for the case where $m = 0$ and σ varies. Note that the relative changes in the curves are similar between ROC and SIROC. In Fig. 12, we have $\sigma = 0$ and m is variable.

B. A_Λ and A_s Calculations

The area can be calculated directly from the data collected to plot the curves of Figs. 11 and 12. By generating several plots of ROC and SIROC for varied signal present parameters, a plot of A_s and A_Λ compared with the input distribution parameters is generated in Figs. 13 and 14.

In Fig. 13, the mean is set to $\mathbf{0}$ so only the difference between the covariance matrices is observed. An asymmetric distribution results in a peak area where the covariances are similar.

Figure 15 shows a symmetric adjustment of area about the indistinguishable point. This is the same relationship shown when using the one-dimensional Gaussian case.

Inequality II can be checked by comparing the data from A_Λ and A_s for the multivariate Gaussian.

C. Conditional Entropy and Probability of Error

To develop the minimum probability of error for higher dimensional problems, the ROC metrics of FPF and FNF can be used with their threshold to compute $P_e(y)$. By sampling the signal present and signal absent distributions and then comparing the ideal observer to thresholds the function $P_e(y)$ can be computed. Figure 16 gives the shape of minimum probability of error for $\mathbf{m} = [00100]^\dagger$ and $\sigma = 1$.

Conditional entropy can be computed from the SIROC coordinate functions. Figure 17 gives the shape of conditional entropy for a given signal present and signal absent for $\mathbf{m} = [00100]^\dagger$ and $\sigma = 1$.

As was done for the conditional entropy case, the validity of Inequality I can be verified using the calculations above. For Figs. 18 and 19, the value of $\mathbf{K}$ remains the same, while the mean vector has returned to $\mathbf{m} = [0mm0]^\dagger$.

The conditional entropy can be compared directly with probability of error through Integral Transform IV. The integral was discretized and several thresholds were evaluated for probability of error. Then the conditional entropy calculation was compared with the sum of probability of error values for a given signal present and signal absent distribution.

In Fig. 20, the approximation to conditional entropy C_a would approach the value of C_e as the discretized integral is more finely sampled.

13. CONCLUSION

We have defined the SIROC for a detection task that makes use of the two components of the conditional entropy in the same way the ideal-observer ROC curve makes use of the two components of the probability of error. We have shown that the two curves have many properties in common, although there are some significant differences. We have derived expressions for the areas under these curves, which can be used as figures of merit for an imaging system, and described the mathematical relationship between these two areas.

Perhaps the most striking result is the integral equation relating the minimum probability of error, as a function of the odds against a signal, to the conditional entropy as a function of the same variable. This expression shows that the value conditional entropy function at a particular value of its argument is a weighted integral of the minimum probability of error function over its entire domain. This integral transform and other similar transforms from the same section show that the ideal-observer ROC curve and the SIROC are equivalent descriptions of system performance, although the map from one curve to the other is nonlocal in the sense that it involves weighted integrals over the whole ideal-observer ROC curve to compute each point on the SIROC. This result also shows that any modification of the imaging system that increases the Shannon information for the detection task, and hence decreases conditional entropy, will result in a decrease in the minimum probability of error over at least some interval of prior probabilities for the signal. Finally, we have shown that this new viewpoint results in some new inequalities relating Shannon information to ROC analysis. The simulations that we have carried out for this work are low in complexity but

serve to illustrate the results and check how tight some of the inequalities are for Gaussian statistics.

We have been able to extend these results to the multi-class classification task. In this situation, we again find that the conditional entropy and the minimum probability of error are related by an integral transform. We also have an SIROC hyper-surface which is closely related to an ROC hyper-surface that has been discussed in the literature. These results will be presented in a forthcoming publication.

Acknowledgments

National Institutes of Health (NIH) (P41-EB002035); U.S. Department of Homeland Security (Department of Homeland Security) (HSHQDC-14-C-BOOIO).

References

1. Neifeld MA, Ashok A, Baheti P. Task specific information for imaging system analysis. *J Opt Soc Am A*. 2007; 24:B25–B41.
2. Ashok A, Baheti P, Neifeld MA. Compressive imaging system design using task-specific information. *Appl Opt*. 2008; 47:4457–4471. [PubMed: 18758516]
3. Ashok A, Neifeld MA. Information-based analysis of simple incoherent imaging systems. *Opt Express*. 2003; 11:2153–2162. [PubMed: 19466103]
4. Ashok A, Huang LC, Neifeld MA. Information optimal compressive sensing: static measurement design. *J Opt Soc Am A*. 2013; 30:831–853.
5. Van Trees, HL. *Detection, Estimation and Modulation Theory. Vol. Part 1*. Wiley; 1968.
6. Anderson, TW. *An Introduction to Multivariate Statistical Analysis*. Wiley; 1971.
7. Metz CE. Basic principles of ROC analysis. *Semin Nucl Med*. 1978; 8:283–298. [PubMed: 112681]
8. Turner DA. An intuitive approach to ROC curve analysis. *J Nucl Med*. 1978; 19:213–220. [PubMed: 627904]
9. Lusted LB. Signal detectability and medical decision making. *Science*. 1971; 171:1217–1219. [PubMed: 5545199]
10. Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. *Invest Radiol*. 1979; 14:109–121. [PubMed: 478799]
11. Swets, JA., Pickett, RM. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Academic; 1982.
12. Swets JA. Measuring the accuracy of diagnostic systems. *Science*. 1988; 240:1285–1293. [PubMed: 3287615]
13. Metz, CE. Fundamental ROC analysis. In: Beutel, J.Kundel, HL., Van Metter, RL., editors. *Handbook of Medical Imaging*. SPIE; 1999. Vol. 1 of Progress in Medical Physics and Psychophysics
14. Barrett HH, Abbey CK, Clarkson E. Objective assessment of image quality III: ROC metrics, ideal observers and likelihood-generating functions. *J Opt Soc Am A*. 1998; 15:1520–1535.
15. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J*. 1948; 27:379–423.
16. Fano, RM. *Transmission of Information: A Statistical Theory of Communication*. Wiley; 1961.
17. Kullback, S. *Information Theory and Statistics*. Wiley; 1959.
18. Fisher RA. *Theory of Statistical Information*. Proc Cambridge Philos Soc. 1925; 22:700–725.
19. Clarkson E, Barrett HH. Approximations to ideal-observer performance on signal detection tasks. *Appl Opt*. 2000; 39:1783–1793. [PubMed: 18345075]
20. Liese F, Vajda I. On divergences and informations in statistics and information theory. *IEEE Trans Inf Theory*. 2006; 52:4394–4412.

21. Guo D, Shamai S, Verdu S. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans Inf Theory*. 2005; 51:1261–1282.
22. Guo D, Shamai S, Verdu S. Mutual information and conditional mean estimation in Poisson channels. *IEEE Trans Inf Theory*. 2008; 54:1837–1849.
23. Clarkson E. Asymptotic ideal observers and surrogate figures of merit for signal detection with list-mode data. *J Opt Soc Am A*. 2012; 29:2204–2216.
24. Shen F, Clarkson E. Using Fisher information to approximate ideal observer performance on detection tasks for lumpy backgrounds. *J Opt Soc Am A*. 2006; 23:2406–2414.
25. Clarkson E, Shen F. Fisher information and surrogate figures of merit for the task-based assessment of image quality. *J Opt Soc Am A*. 2010; 27:2313–2326.

Fig. 1.
ROC curve and SIROC for $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(0, \sigma^2)$.

Fig. 2.
ROC curve and SIROC for $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(0, \sigma^2)$.

Fig. 3.
 A_λ and A_s for $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(0, \sigma^2)$.

Fig. 4.
 A_Λ and A_s for $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(m, 1)$.

(a) $H_0 : x \sim \mathcal{N}(0, 1)$ and $H_1 : x \sim \mathcal{N}(0, \sigma^2)$

(b) $H_0 : x \sim \mathcal{N}(0, 1)$ and $H_1 : x \sim \mathcal{N}(m, 1)$

Fig. 5.
Inequality II verification for Gaussian distributions.

Fig. 6.
 P_e for $\mathcal{N}(0.001, 1.1^2)$.

Fig. 7.
 C_e for $\mathcal{N}(0.001, 1.1^2)$.

Fig. 8.
 C_e for $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(0, \sigma^2)$.

Fig. 9.
 C_e for $H_0: x \sim \mathcal{N}(0, 1)$ and $H_1: x \sim \mathcal{N}(m, 1)$.

Fig. 10.
 C_e and C_a for 1D Gaussian.

Fig. 11.
ROC curve and SIROC for $H_0: \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1: \mathbf{x} \sim \mathcal{N}(\mathbf{0}, K)$.

Fig. 12.
ROC curve and SIROC for $H_0: \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1: \mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{I})$.

Fig. 13.
 A_λ and A_s for $H_0:\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1:\mathbf{x} \sim \mathcal{N}(\mathbf{0}, K)$.

Fig. 14.
 A_λ and A_s for $H_0: \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1: \mathbf{x} \sim \mathcal{N}(\mathbf{m}, \mathbf{I})$.

(a) $H_0 : \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1 : \mathbf{x} \sim \mathcal{N}(\mathbf{0}, K)$

(b) $H_0 : \mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1 : \mathbf{x} \sim \mathcal{N}(m, \mathbf{I})$

Fig. 15.
Verifying Inequality II for multivariate Gaussian.

Fig. 16.
 P_e for $\mathcal{N}(\mathbf{m}, \mathbf{K})$.

Fig. 17.
 C_e for $\mathcal{N}(\mathbf{m}, \mathbf{K})$.

Fig. 18.
 C_e for $H_0: x \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1: x \sim \mathcal{N}(\mathbf{0}, \mathbf{K})$.

Fig. 19.
 C_e for $H_0: x \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $H_1: x \sim \mathcal{N}(\mathbf{m}, \mathbf{K}(\sigma = 1))$.

Fig. 20.
 C_e and C_a for multivariate Gaussian.