Published in final edited form as:

J Opt Soc Am A Opt Image Sci Vis. 2019 July 01; 36(7): 1209–1214. doi:10.1364/JOSAA.36.001209.

Relation between Bayesian Fisher information and Shannon information for detecting a change in a parameter

Eric Clarkson

Department of Medical Imaging and College of Optical Sciences University of Arizona, Tucson, Arizona 85719, USA

Abstract

We derive a connection between the performance of statistical estimators and the performance of the ideal observer on related detection tasks. Specifically, we show how the task-specific Shannon information for the task of detecting a change in a parameter is related to the Fisher information and to the Bayesian Fisher information. We have previously shown that this Shannon information is related via an integral transform to the minimum probability of error on the same task. We then outline a circle of relations starting with this minimum probability of error and ensemble mean squared error for an estimator via the Ziv–Zakai inequality, then the ensemble mean squared error and the Bayesian Fisher information via the van Trees inequality, and finally the Bayesian Fisher information and the Shannon information for a detection task via the work presented here.

1. INTRODUCTION

In imaging, we often use the data generated by the imaging system to perform a task of importance to us. These tasks can be broadly classified as estimation tasks, where we are trying to estimate a set of parameters associated with the object being imaged; classification tasks, where we are trying to decide to which class an object belongs; and tasks that combine estimation and classification. For example, in medical imaging we may be interested in estimating the size of a tumor to determine whether a cancer therapy is having a positive effect. A closely related classification task to this estimation task is to determine whether the parameters defining the size of the tumor have changed or not. In remote sensing we may want to estimate the location of a particular target object. A closely related classification task in this case is to determine whether the target object has moved or not. We have previously shown that the performance of an imaging system on a parameter estimation task, as measured by the Fisher information matrix (FIM), is directly related to the performance on the classification task of detecting a change in the parameters of interest as measured by the area under the receiver operating characteristic (ROC) curve of the ideal observer. In this paper we derive similar relations connecting the FIM and the Bayesian FIM (BFIM) to the task-specific Shannon information (TSI) [1,2] for the task of detecting a change in the parameters.

(clarkson@radiology.arizona.edu).

Disclosures. The author declares that there are no conflicts of interest related to this article.

Fisher information (FI) is normally thought of as related to the performance of an estimator of a parameter in a statistical model for the given data. The ordinary FI is related through the Cramer–Rao bound (CRB) to the variance of an unbiased estimator of the parameter. What we will call the Bayesian FI (BFI) is related by the van Trees inequality to the ensemble mean squared error (EMSE) for an estimator. On the other hand, we have shown previously that the FI is related to the performance of the ideal observer on the task of detecting a small change in the parameter as measured by the area under the ROC curve (AUC) [3–5]. Thus, FI is related to performance on a binary classification task. The Ziv–Zakai inequality relates the EMSE of an estimator to the performance of the ideal observer on the task of distinguishing two different values of the parameter as measured by the minimum probability of error (MPE). Thus, EMSE is also related to performance on a binary classification task. In this work we will derive further connections between the performance of estimators and performance measures for related detection tasks. Specifically, we will show how the TSI for a certain binary classification task is related to the FI and the BFI. We have previously shown that this TSI is related via an integral transform to the MPE on the same task [6,7]. We then have a circle of relations starting with MPE and EMSE via Ziv–Zakai, then EMSE and BFI via van Trees, BFI and TSI via the work presented here, and finally TSI and MPE via the aforementioned integral transform.

2. FISHER INFORMATION

We will consider a scalar parameter θ throughout most of this paper. This parameter takes values on the real line. There are also vector parameter versions of all results. The data is a random vector $\mathbf{g}$ generated by a noisy imaging system and has a conditional probability density function (PDF) denoted by $\text{pr}(\mathbf{g}|\theta)$. The FI is then given by [8,9]

$$F(\theta) = \left\langle \left[\frac{d}{d\theta} \ln \text{pr}(\mathbf{g}|\theta) \right]^2 \right\rangle_{\mathbf{g}|\theta}. \quad (1)$$

Suppose that $\hat{\theta}(\mathbf{g})$ is an estimator that uses the data vector to produce an unbiased estimate of θ . The CRB states that

$$\text{var}(\hat{\theta}) \geq [F(\theta)]^{-1} \quad (2)$$

when

$$\langle \hat{\theta}(\mathbf{g}) \rangle_{\mathbf{g}|\theta} = \theta. \quad (3)$$

In other words, the variance of an unbiased estimator is bounded below by the reciprocal of the FI. For the Bayesian FI there is the van Trees inequality, which is similar to the Cramer–Rao lower bound. This will be discussed further below.

3. FI AND AUC

We consider a classification task to classify $\mathbf{g}$ as a sample drawn from $\text{pr}(\mathbf{g}|\theta)$ or $\text{pr}(\mathbf{g}|\tilde{\theta})$ corresponding to hypotheses H_0 and H_1 , respectively. The ideal-observer AUC for this task computes the likelihood ratio [10]

$$\Lambda(\mathbf{g}) = \frac{\text{pr}(\mathbf{g}|\tilde{\theta})}{\text{pr}(\mathbf{g}|\theta)} \quad (4)$$

and compares it to a threshold Λ_0 in order to make the decision. If the likelihood ratio is greater than the threshold, then decision D_1 is made that $\mathbf{g}$ was drawn from $\text{pr}(\mathbf{g}|\tilde{\theta})$. If the likelihood ratio is less than or equal to the threshold, then decision D_0 is made that $\mathbf{g}$ was drawn from $\text{pr}(\mathbf{g}|\theta)$. The true positive fraction (TPF) is $\text{Pr}(D_1|H_1)$, the probability that decision D_1 is made when H_1 is true. The false positive fraction (FPF) is $\text{Pr}(D_1|H_0)$, the probability that decision D_1 is made when H_0 is true. The receiver operating characteristic (ROC) curve for this task plots TPF versus FPF as the threshold is varied from 0 to ∞ . This is a concave curve that starts at point (1, 1) and ends at (0, 0). The area under the ideal-observer ROC curve, AUC, is a commonly used figure of merit (FOM) for an imaging system on a classification task. In this case we will denote this quantity by $\text{AUC}(\theta, \tilde{\theta})$. For the ideal observer, the AUC is always between 0.5 and 1. This area is related to the detectability $d(\theta, \tilde{\theta})$ by the equation

$$\text{AUC}(\theta, \tilde{\theta}) = \frac{1}{2} + \frac{1}{2} \text{erf} \left[\frac{1}{2} d(\theta, \tilde{\theta}) \right]. \quad (5)$$

The detectability is an equivalent FOM to the AUC and varies from 0 to ∞ . In previous publications we showed that the Taylor series expansion for the ideal-observer detectability in this situation is given by [3–5]

$$d^2(\theta, \theta + \Delta\theta) = (\Delta\theta)^2 F(\theta) + \dots \quad (6)$$

Our goal in this work is to derive a similar relation between the Bayesian FI and the average Shannon information for the classification task that we have defined here.

4. BAYESIAN FI AND EMSE

For the rest of this paper we will assume that we have a prior probability $\text{pr}(\theta)$ on the parameter of interest. The Bayesian FI is given by [11,12]

$$F = \langle F(\theta) \rangle_{\theta} + \left\langle \left[\frac{d}{d\theta} \ln \text{pr}(\theta) \right]^2 \right\rangle_{\theta}. \quad (7)$$

For an alternate expression of the BFI, we define the posterior PDF $\text{pr}(\theta|\mathbf{g})$ on θ by

$$\text{pr}(\theta|\mathbf{g}) = \frac{\text{pr}(\mathbf{g}|\theta)\text{pr}(\theta)}{\text{pr}(\mathbf{g})}, \quad (8)$$

where

$$\text{pr}(\mathbf{g}) = \int_{-\infty}^{\infty} \text{pr}(\mathbf{g}|\theta)\text{pr}(\theta)d\theta. \quad (9)$$

In terms of the posterior PDF, we can show that

$$F = \left\langle \left\langle \left[\frac{d}{d\theta} \ln \text{pr}(\theta|\mathbf{g}) \right]^2 \right\rangle_{\mathbf{g}|\theta} \right\rangle_{\theta}. \quad (10)$$

The EMSE for any estimator of θ is defined by

$$\text{EMSE} = \langle \langle [\hat{\theta}(\mathbf{g}) - \theta]^2 \rangle_{\mathbf{g}|\theta} \rangle_{\theta}. \quad (11)$$

The van Trees inequality states that [11,12]

$$\text{EMSE} \geq F^{-1}. \quad (12)$$

In other words, the EMSE for any estimator of the parameter of interest is bounded below by the reciprocal of the BFI. The van Trees inequality is also called the Bayesian CRB. This is the reason for calling F the Bayesian FI. This inequality, much like the CRB, relates the relevant version of the BFI to performance on an estimation task. On the other hand, the result described in Section 3 relates the FI to a classification task, namely, the detection of a small change in the parameter of interest. We want to connect the BFI to this same task, but before we do this we discuss a relation between the EMSE and the detection task in Section 3.

5. ZIV-ZAKAI INEQUALITY (EMSE AND MPE)

When the prior $\text{pr}(\theta)$ is known, then we can define prior probabilities for the two hypotheses H_0 and H_1 for the detection task in Section 3 via

$$\text{pr}_0(\theta, \tilde{\theta}) = \frac{\text{pr}(\theta)}{\text{pr}(\theta) + \text{pr}(\tilde{\theta})} \quad (13)$$

and

$$\text{pr}_1(\theta, \tilde{\theta}) = \frac{\text{pr}(\tilde{\theta})}{\text{pr}(\theta) + \text{pr}(\tilde{\theta})}. \quad (14)$$

Then the minimum probability of error $P_e(\theta, \tilde{\theta})$ for this detection task is the probability of error for the ideal observer when the threshold used for the likelihood ratio is given by

$$A_0 = y(\theta, \tilde{\theta}) = \frac{\text{pr}_0(\theta, \tilde{\theta})}{\text{pr}_1(\theta, \tilde{\theta})} = \frac{\text{pr}(\theta)}{\text{pr}(\tilde{\theta})}. \quad (15)$$

The minimum probability of error is

$$P_e(\theta, \tilde{\theta}) = \text{FPF}(\theta, \tilde{\theta}) \text{pr}_0(\theta, \tilde{\theta}) + [1 - \text{TPF}(\theta, \tilde{\theta})] \text{pr}_1(\theta, \tilde{\theta}) \quad (16)$$

since $1 - \text{TPF}$ is the false negative fraction (FNF) $\text{Pr}(D_0|H_1)$. Of course, in this expression, TPF and FPF also depend on the pair $(\theta, \tilde{\theta})$. The Ziv–Zakai inequality, in its standard form, states that [13,14]

$$\text{EMSE} \geq \frac{1}{2} \int_0^\infty \int_{-\infty}^\infty [\text{pr}(\theta) + \text{pr}(\theta + x)] P_e(\theta, \theta + x) d\theta dx. \quad (17)$$

We will show that this inequality can be written in an alternate form that connects it to the classification task in Section 3. We start with the substitution $\tilde{\theta} = \theta + x$ and a change in the order of integration so that we have

$$\text{EMSE} \geq \frac{1}{2} \int_{-\infty}^\infty \int_\theta^\infty [\text{pr}(\theta) + \text{pr}(\tilde{\theta})] P_e(\theta, \tilde{\theta}) (\tilde{\theta} - \theta) d\tilde{\theta} d\theta. \quad (18)$$

Due to the range of integration for $\tilde{\theta}$, we may also write this as

$$\text{EMSE} \geq \frac{1}{2} \int_{-\infty}^\infty \int_\theta^\infty [\text{pr}(\theta) + \text{pr}(\tilde{\theta})] P_e(\theta, \tilde{\theta}) |\tilde{\theta} - \theta| d\tilde{\theta} d\theta. \quad (19)$$

Interchanging the order of integration, we have

$$\text{EMSE} \geq \frac{1}{2} \int_{-\infty}^\infty \int_{-\infty}^{\tilde{\theta}} [\text{pr}(\theta) + \text{pr}(\tilde{\theta})] P_e(\theta, \tilde{\theta}) |\tilde{\theta} - \theta| d\theta d\tilde{\theta}. \quad (20)$$

Now we note that $P_e(\theta, \tilde{\theta}) = P_e(\tilde{\theta}, \theta)$ since the two tasks described by the pairs $(\theta, \tilde{\theta})$ and $(\tilde{\theta}, \theta)$ are the same task. Therefore, we may interchange the variables θ and $\tilde{\theta}$ and use the symmetry of all of the functions in the integrand to arrive at

$$\text{EMSE} \geq \frac{1}{2} \int_{-\infty}^\infty \int_{-\infty}^\theta [\text{pr}(\theta) + \text{pr}(\tilde{\theta})] P_e(\theta, \tilde{\theta}) |\tilde{\theta} - \theta| d\tilde{\theta} d\theta. \quad (21)$$

Now, combining the second and fourth inequalities in this chain, we have

$$\text{EMSE} \geq \frac{1}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [\text{pr}(\theta) + \text{pr}(\tilde{\theta})] P_e(\theta, \tilde{\theta}) |\tilde{\theta} - \theta| d\tilde{\theta} d\theta. \quad (22)$$

We separate the right-hand side into two integrals and again use the symmetry in θ and $\tilde{\theta}$ to find out that the two integrals are the same. Therefore, we have the final result

$$\text{EMSE} \geq \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \text{pr}(\theta) P_e(\theta, \tilde{\theta}) |\tilde{\theta} - \theta| d\tilde{\theta} d\theta. \quad (23)$$

Finally, the integral on the right can be written as an expectation as follows:

$$\text{EMSE} \geq \frac{1}{2} \left\langle \int_{-\infty}^{\infty} P_e(\theta, \tilde{\theta}) |\tilde{\theta} - \theta| d\tilde{\theta} \right\rangle_{\theta}. \quad (24)$$

This formulation of the Ziv–Zakai inequality makes it clear that the worst-case scenario, from the EMSE perspective, is when the minimum probability of error $P_e(\theta, \tilde{\theta})$ on the classification task in Section 3 is significant for relatively large values of $|\tilde{\theta} - \theta|$. In the next section we will complete the circle by relating the BFI to performance on the classification task in Section 2 as measured by the average TSI.

6. TSI AND THE BAYESIAN FIM

For the purposes of this section, we will use the notation

$$\Lambda(\mathbf{g}|\theta, \tilde{\theta}) = \frac{\text{pr}(\mathbf{g}|\tilde{\theta})}{\text{pr}(\mathbf{g}|\theta)} \quad (25)$$

for the likelihood ratio for the classification task in Section 3. Keeping in mind that $\Lambda = \Lambda(\mathbf{g}|\theta, \tilde{\theta})$ and $y = y(\theta, \tilde{\theta})$ from Eq. (15), we can write the SI for this classification task as [15]

$$I(y) = \frac{1}{1+y} \left\langle \ln \left[\frac{\Lambda(1+y)}{\Lambda+y} \right] \right\rangle_{\mathbf{g}|\tilde{\theta}} + \frac{y}{1+y} \left\langle \ln \left[\frac{1+y}{\Lambda+y} \right] \right\rangle_{\mathbf{g}|\theta}. \quad (26)$$

If we introduce a binary random variable X such that $X=1$ with probability $\text{pr}_1(\theta, \tilde{\theta})$, and $X=0$ with probability $\text{pr}_0(\theta, \tilde{\theta})$, then the entropy of X depends only on $y = y(\theta, \tilde{\theta})$ and is given by

$$H(y) = -\frac{1}{1+y} \ln \left(\frac{1}{1+y} \right) - \frac{y}{1+y} \ln \left(\frac{y}{1+y} \right). \quad (27)$$

The usual notation for this entropy is $H(X)$, but we are using the nonstandard notation $H(y)$ in this equation to emphasize the functional dependence on y . The conditional entropy for X

given $\mathbf{g}$ is then defined by $C_e(y) = H(y) - I(y)$. Again, the standard notation for this conditional entropy would be $H(X|\mathbf{g})$. The non-standard notation $C_e(y)$ for conditional entropy used here was introduced in an earlier publication [6] that related conditional entropy to MPE via an integral transform in the variable y . After some algebra, we have the expression

$$C_e(y) = \frac{1}{1+y} \left\langle \ln \left(\frac{\Lambda + y}{\Lambda} \right) \right\rangle_{\mathbf{g}|\tilde{\theta}} + \frac{y}{1+y} \left\langle \ln \left(\frac{\Lambda + y}{y} \right) \right\rangle_{\mathbf{g}|\theta}. \quad (28)$$

We are interested in the limiting value of the conditional entropy as $\tilde{\theta}$ approaches θ . After some more simplification involving the definition of the likelihood ratio and the identity

$$\left\langle \ln \left(\frac{\Lambda + y}{\Lambda} \right) \right\rangle_{\mathbf{g}|\tilde{\theta}} = \left\langle \Lambda \ln \left(\frac{\Lambda + y}{y} \right) \right\rangle_{\mathbf{g}|\theta}, \quad (29)$$

we can write the conditional entropy for the task in Section 3 in the form

$$C_e(\theta, \tilde{\theta}) = C_e(y) = \frac{1}{1+y} [\langle (\Lambda + y) \ln(\Lambda + y) \rangle_{\mathbf{g}|\theta} - \langle \Lambda \ln \Lambda \rangle_{\mathbf{g}|\theta} - y \ln y]. \quad (30)$$

Note that only expectations under hypothesis H_0 are needed to compute the conditional entropy and hence the TSI. We will now show that, when $\tilde{\theta} = \theta + \Delta\theta$ and we average over θ , then, to the lowest order, we have the expansion

$$\langle C_e(\theta, \theta + \Delta\theta) \rangle_{\theta} = \ln 2 - \frac{(\Delta\theta)^2}{4} F + \dots \quad (31)$$

Thus, the average value of the conditional entropy for the detection task in Section 3 is closely related to the BFIM for the relevant parameter. The following derivation is straightforward but somewhat involved, and it can be skipped without loss of understanding of the results of this paper. A vector version of this result and some implications for TSI are discussed in the sections following the derivation.

7. DERIVATION OF THE TSI AND BFIM RELATION

In order to keep the notational complications to a minimum, we introduce the function

$$C'_e(\theta) = \frac{d}{d\tilde{\theta}} C_e(\theta, \tilde{\theta}) \Big|_{\tilde{\theta} = \theta}. \quad (32)$$

We also will use the following notation:

$$\Lambda' = \Lambda'(\mathbf{g}|\theta) = \frac{d}{d\tilde{\theta}} \Lambda(\mathbf{g}|\theta, \tilde{\theta}) \Big|_{\tilde{\theta} = \theta} = \frac{\text{pr}'(\mathbf{g}|\theta)}{\text{pr}(\mathbf{g}|\theta)}, \quad (33)$$

where the prime on the far right indicates a derivative with respect to θ . Similarly, we will write

$$y' = y'(\theta) = \left. \frac{d}{d\tilde{\theta}} y(\theta, \tilde{\theta}) \right|_{\tilde{\theta} = \theta} = \frac{\text{pr}'(\theta)}{\text{pr}(\theta)}. \quad (34)$$

Now, using the fact that, when $\tilde{\theta} = \theta$, we have $y = 1$ and $\Lambda = 1$

$$\begin{aligned} C'_e(\theta) = & -\frac{y'}{4} 2 \ln 2 \\ & + \frac{1}{2} [(\Lambda' + y') \ln 2 + (\Lambda' + y') - \Lambda']_{\mathbf{g}|\theta} - y']. \end{aligned} \quad (35)$$

Now we use the fact that $\langle \Lambda' \rangle_{\mathbf{g}|\theta} = 0$ to give us

$$C'_e(\theta) = -\frac{y'}{4} 2 \ln 2 + \frac{1}{2} [y' \ln 2] = 0. \quad (36)$$

This result is to be expected since, as a function of $\tilde{\theta}$, the conditional entropy $C_e(\theta, \tilde{\theta})$ is maximized when $\tilde{\theta} = \theta$.

Now we will look at second derivative terms. We define

$$C''_e(\theta) = \left. \frac{d^2}{d\tilde{\theta}^2} C_e(\theta, \tilde{\theta}) \right|_{\tilde{\theta} = \theta}. \quad (37)$$

We also will use the following notation:

$$\Lambda'' = \Lambda''(\mathbf{g}|\theta) = \left. \frac{d^2}{d\tilde{\theta}^2} \Lambda(\mathbf{g}|\theta, \tilde{\theta}) \right|_{\tilde{\theta} = \theta} = \frac{\text{pr}''(\mathbf{g}|\theta)}{\text{pr}(\mathbf{g}|\theta)} \quad (38)$$

and

$$y'' = y''(\theta) = \left. \frac{d^2}{d\tilde{\theta}^2} y(\theta, \tilde{\theta}) \right|_{\tilde{\theta} = \theta} = \frac{\text{pr}''(\theta)}{\text{pr}(\theta)}. \quad (39)$$

In order to ease the mental computations for the reader, we note that, for any variable z that depends on θ , we have

$$\frac{d}{d\theta} z \ln z = z'(1 + \ln z) \quad (40)$$

and

$$\frac{d^2}{d\theta^2} z \ln z = z''(1 + \ln z) + \frac{(z')^2}{z}. \quad (41)$$

Also note that we use

$$\frac{d}{d\theta} \frac{1}{1+y} = -\frac{y'}{(1+y)^2} \quad (42)$$

and

$$\frac{d^2}{d\theta^2} \frac{1}{1+y} = \frac{2(y')^2}{(1+y)^3} - \frac{y''}{(1+y)^2}. \quad (43)$$

Now, using the Leibniz rule for second derivatives, we have the expansion $C_e''(\theta) = A + 2B + C$, where the terms on the right are given by

$$A = \left\langle \left[\frac{(y')^2 - y''}{4} \right] (2 \ln 2) \right\rangle_{\mathbf{g}|\theta}, \quad (44)$$

$$2B = \left\langle -\frac{y'}{2} [(1 + \ln 2)(\Lambda' + y') - \Lambda' - y'] \right\rangle_{\mathbf{g}|\theta}, \quad (45)$$

and

$$C = \left\langle \frac{1}{2} [(\Lambda'' + y'')(1 + \ln 2) + \frac{1}{2} (\Lambda' + y')^2 - \Lambda'' - (\Lambda')^2 - y'' - (y')^2] \right\rangle_{\mathbf{g}|\theta}. \quad (46)$$

These three terms simplify algebraically to

$$A = \left\langle [(y')^2 - y''] \frac{\ln 2}{2} \right\rangle_{\mathbf{g}|\theta}, \quad (47)$$

$$2B = \left\langle -\frac{\ln 2}{2} [\Lambda' y' + (y')^2] \right\rangle_{\mathbf{g}|\theta}, \quad (48)$$

and

$$C = \left\langle (\Lambda'' + y'') \frac{\ln 2}{2} + \frac{1}{2} \Lambda' y' - \frac{1}{4} (\Lambda')^2 - \frac{1}{4} (y')^2 \right\rangle_{\mathbf{g}|\theta}. \quad (49)$$

Now we use the fact that $\langle \Lambda'' \rangle_{\mathbf{g}|\theta} = 0$ and $\langle \Lambda' y' \rangle_{\mathbf{g}|\theta} = y' \langle \Lambda' \rangle_{\mathbf{g}|\theta} = 0$ to arrive at

$$C_e''(\theta) = \left\langle -\frac{1}{4}(\Lambda')^2 - \frac{1}{4}(y')^2 \right\rangle_{\mathbf{g}|\theta} = -\frac{1}{4}[\langle (\Lambda')^2 \rangle_{\mathbf{g}|\theta} + \langle (y')^2 \rangle]. \quad (50)$$

If we use the prior to average over θ , we then have

$$\langle C_e''(\theta) \rangle_{\theta} = -\frac{1}{4} \langle \langle (\Lambda')^2 \rangle_{\mathbf{g}|\theta} + \langle (y')^2 \rangle \rangle_{\theta} = -\frac{1}{4}F. \quad (51)$$

Since $\langle C_e(\theta, \theta) \rangle_{\theta} = \langle \ln 2 \rangle_{\theta} = \ln 2$, we have the following Taylor expansion of the average conditional entropy for the task in Section 2:

$$\langle C_e(\theta, \theta + \Delta\theta) \rangle_{\theta} = \ln 2 - \frac{(\Delta\theta)^2}{4}F + \dots. \quad (52)$$

Thus, the BFI gives the lowest-order approximation for the average conditional entropy when the task is detecting a small change in the parameter of interest.

The Ziv–Zakai inequality provides a relation between the MPE for the task of using the data vector $\mathbf{g}$ to detect a change in a parameter θ in the conditional PDF $\text{pr}(\mathbf{g}|\theta)$ and the EMSE for the task of using $\mathbf{g}$ to estimate a parameter θ in the conditional PDF $\text{pr}(\mathbf{g}|\theta)$. The van Trees inequality relates this EMSE to the BFI. We have now found a relationship between the BFI and the average conditional entropy for the detection task. This completes the circle since there is an integral relation between the conditional entropy for any classification task and the MPE for the same task [6,7]. This reinforces the idea that there is a strong mathematical relationship between the performance of an imaging system on estimation tasks and the performance of the same system on the detection of a small change in the parameter of interest.

8. VECTOR VERSION

For the vector version of this last result we assume a vector parameter $\boldsymbol{\theta}$ and a perturbation in this parameter of the form $t\mathbf{u}$, where $\mathbf{u}$ is a unit vector in parameter space. We then have a Taylor series expansion in the variable t given by

$$\langle C_e(\boldsymbol{\theta}, \boldsymbol{\theta} + t\mathbf{u}) \rangle_{\boldsymbol{\theta}} = \ln 2 - \frac{t^2}{4} \mathbf{u}^\dagger \mathbf{F} \mathbf{u} + \dots. \quad (53)$$

In this equation the Bayesian Fisher information matrix (BFIM) is given by

$$\mathbf{F} = \langle \mathbf{F}(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}} + \langle [\nabla_{\boldsymbol{\theta}} \ln \text{pr}(\boldsymbol{\theta})][\nabla_{\boldsymbol{\theta}} \ln \text{pr}(\boldsymbol{\theta})]^\dagger \rangle_{\boldsymbol{\theta}}, \quad (54)$$

where the ordinary FIM is defined by

$$\mathbf{F}(\boldsymbol{\theta}) = \langle [\nabla_{\boldsymbol{\theta}} \ln \text{pr}(\mathbf{g}|\boldsymbol{\theta})][\nabla_{\boldsymbol{\theta}} \ln \text{pr}(\mathbf{g}|\boldsymbol{\theta})]^\dagger \rangle_{\mathbf{g}|\boldsymbol{\theta}}. \quad (55)$$

There are vector versions of the van Trees inequality and the Ziv–Zakai inequality, so the circle of relations between EMSE, MPE, and Bayesian FIM also exists for vector parameters.

9. TAYLOR SERIES EXPANSION FOR DETECTION TSI

To get a Taylor series expansion for the TSI similar to the one we have derived for conditional entropy, we must find the expansion for $H(y)$. This function can be written in the form

$$H(y) = \frac{1}{1+y}[(1+y)\ln(1+y) - y\ln(y)]. \quad (56)$$

Using notation similar to that in Section 5, we have

$$H'(\theta) = -\frac{y'}{4}(2\ln 2) + \frac{1}{2}(y'\ln 2 + y' - y') = 0. \quad (57)$$

For the second derivative, we can write $H''(\theta) = a + 2b + c$ with

$$a = \left[\frac{(y')^2}{4} - \frac{y''}{4} \right] (2\ln 2), \quad (58)$$

$$2b = -2\left(\frac{y'}{4}\right)(y'\ln 2 + y' - y'), \quad (59)$$

and

$$c = \left(\frac{1}{2}\right) \left\{ y''[1 + \ln 2] + \frac{(y')^2}{2} - y'' - (y')^2 \right\}. \quad (60)$$

Combining terms, we have

$$H''(\theta) = -\frac{(y')^2}{4}. \quad (61)$$

Now we may combine these computations with the expansion in Section 5 and find that

$$I''(\theta) = \frac{1}{4} \langle (\Lambda')^2 \rangle_{g|\theta} = \frac{1}{4} F(\theta). \quad (62)$$

This gives us another relation between the FI and the detection task in Section 3. The Taylor series expansion for $\mathcal{I}(\theta, \theta + \theta)$ now has the form

$$I(\theta, \theta + \Delta\theta) = \frac{(\Delta\theta)^2}{4} F(\theta) + \dots . \quad (63)$$

Thus, the TSI and the ideal-observer detectability are essentially the same FOM when we are trying to detect a small change in a parameter.

After using the prior on θ to compute an expectation, we have

$$\langle I(\theta, \theta + \Delta\theta) \rangle_\theta = \frac{(\Delta\theta)^2}{4} \langle F(\theta) \rangle_\theta + \dots . \quad (64)$$

Thus, the quantity $\langle F(\theta) \rangle_\theta$ also has an interpretation terms of the average TSI for the detection task in section 3.

10. CONCLUSION

FI is usually thought of in terms of the CRB. BFI is usually thought of in terms of the van Trees inequality. We have shown previously that FI is related to the performance of the ideal observer on the task of detecting a small change in the parameter as measured by the area under the ROC curve. Thus, FI is related to the ideal performance on a binary classification task. The Ziv–Zakai inequality relates the EMSE of an estimator to the performance of the ideal observer on the task of distinguishing two different values of the parameter as measured by the MPE. In this work, we showed how the task-specific Shannon information for the task of detecting a small change in a parameter is related to the BFI. Since we have previously shown that this Shannon information is related via an integral transform to the MPE on the same task, we then have a circle of relations relating MPE on a detection task, EMSE of an estimator, BFI, and Shannon information for the detection task.

One of the useful results of this work is that the TSI for the detection of a small change in a parameter and the ideal-observer AUC for the same task are essentially equivalent figures of merit for a given imaging system. Optimizing a system for one of these quantities will also optimize it for the other. Another point of interest is that, when the parameter is a vector, the interpretation of the BFIM in terms of Shannon information does not require inversion of the matrix. This is similar to the connection between the FIM and the ideal-observer AUC derived previously. Finally, both of these relations are the first terms of a Taylor series expansion and are not inequalities like the CRB and the van Trees inequality. They thus have the potential to provide good approximations to the relevant TSI or AUC values.

Funding.

National Institute of Biomedical Imaging and Bioengineering (NIBIB) (P41-EB002035, R01-EB000803).

REFERENCES

1. Neifeld MA, Ashok A, and Baheti PK, “Task-specific information for imaging system analysis,” J. Opt. Soc. Am. A 24, B25–B41 (2007).

2. Ashok A, Baheti PK, and Neifeld MA, "Compressive imaging system design using task-specific information," *Appl. Opt* 47, 4457–4471 (2008). [PubMed: 18758516]
3. Clarkson E and Shen F, "Fisher information and surrogate figures of merit for the task-based assessment of image quality," *J. Opt. Soc. Am. A* 27, 2313–2326 (2010).
4. Shen F and Clarkson E, "Using Fisher information to approximate ideal observer performance on detection tasks for lumpy-background images," *J. Opt. Soc. Am. A* 23, 2406–2414 (2006).
5. Clarkson E, "Asymptotic ideal observers and surrogate figures of merit for signal detection with list-mode data," *J. Opt. Soc. Am. A* 29, 2204–2216 (2012).
6. Clarkson E and Cushing J, "Shannon information and ROC analysis in imaging," *J. Opt. Soc. Am. A* 32, 1288–1301 (2015).
7. Clarkson E and Cushing J, "Shannon information and receiver operating characteristic analysis for multi-class classification in imaging," *J. Opt. Soc. Am. A* 33, 930–937 (2016).
8. Shao J, *Mathematical Statistics* (Springer, 1999).
9. Barrett HH and Myers KJ, *Foundations of Image Science* (Wiley, 2004).
10. Barrett H, Abbey C, and Clarkson E, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," *J. Opt. Soc. Am. A* 15, 1520–1535 (1998).
11. Van Trees HL, *Detection, Estimation and Modulation Theory* (Wiley, 1968), Part 1.
12. Gill RD and Levit BY, "Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound," *Bernoulli* 1, 59–79 (1995).
13. Ziv J and Zakai M, "Some lower bounds on signal parameter," *IEEE Trans. Inf. Theory* 15, 386–391 (1969).
14. Bell K, Steinberg Y, Ephraim Y, and Van Trees H, "Extended Ziv-Zakai lower bound for vector parameter estimation," *IEEE Trans. Inf. Theory* 43, 624–637 (1997).
15. Cover TM and Thomas JA, *Elements of Information Theory* (Wiley, 2006).