Inequalities and Approximations for Fisher Information in the Presence of Nuisance Parameters

Eric Clarkson

February 14, 2019

Abstract

Many imaging systems are used to estimate a vector of parameters associated with the object being imaged. In many cases there are other parameters in the model for the imaging data that are not of interest for the task at hand. We refer to these as nuisance parameters and use them to form the components of the nuisance parameter vector. If we have a prior probability distribution function (PDF) for the nuisance parameter vector, then we may marginalize over the nuisance parameters to produce a conditional PDF for the data that only depends on the parameters of interest. We will examine this approach to develop inequalities and approximations for the FIM when the data is affected by nuisance parameters.

1 Introduction

The Fisher Information Matrix (FIM) is important for computing figures of merit for imaging systems on specific tasks. For estimation tasks the Cramer Rao Bound (CRB), which is derived from the inverse of the FIM, provides a lower bound for the covariance matrix of an unbiased estimator of the parameter vector of interest. The inverse of the FIM is also used to asymptotically approximate the covariance matrix of the maximum likelihood estimator. The FIM is also related directly to the performance of the ideal observer on the task of detecting a small change in the parameter vector of interest, as measured by the area under the Receiver Operating Characteristic (ROC) curve, also known as the AUC. The Bayesian version of the FIM is related by the van Trees inequality to the Ensemble Mean Squared Error (EMSE) for any estimator of the parameter vector of interest. The lower bound given by the van Trees inequality is also known as the Bayesian CRB, and for this reason we will refer to the Bayesian FIM for the relevant matrix. Elsewhere we have shown that the Bayesian FIM is also directly related to the average Shannon Information (SI) for the task of detecting a small change in the parameter vector of interest.
In many cases there are other parameters in the model for the imaging data that are not of interest for the task at hand. We refer to these as nuisance parameters and use them to form the components of the nuisance parameter vector. One way to deal with nuisance parameters for estimation tasks is to estimate them along with the parameters of interest and then ignore these estimates. If we have a prior probability distribution function (PDF) for the nuisance parameter vector, then we may marginalize over the nuisance parameters to produce a conditional PDF for the data that only depends on the parameters of interest. This is the approach that we will examine in this paper to develop inequalities and approximations for the FIM when the data is affected by nuisance parameters.

In Section 2 we develop and inequality for the FIM when the prior PDF on the nuisance parameter vector is independent of the parameters of interest. In Section 3 we generalize this inequality to the case where the PDF for the nuisance parameters depends on the parameters of interest. Section 4 contains description of the implications of inequalities derived in Section 3 for the Bayesian FIM. Finally in Section 4 we present an approximation to the FIM that could be useful if the prior PDF for the nuisance parameters is narrowly distributed around a nominal value that is known.

2 Signal independent nuisance parameters

We will be considering list-mode data in this paper, but everything is still valid for binned data. For list-mode data the \( n^{th} \) detected photon generates an attribute column vector \( a_n \), and these vectors are then aggregated into a matrix \( A = [a_1 \cdots a_N] \). An attribute vector may include position of the photon detection, direction that the photon was travelling when detected, frequency and/or polarization state. The number \( N \) of photons collected may be fixed or random. For simplicity we will consider \( N \) to be fixed, but all of the results apply equally well when \( N \) is random. We are interested in either estimating or detecting a small change in a parameter vector \( \theta \) using the data \( A \). The vector \( \phi \) contains nuisance parameters, i.e., parameters that affect the data but that we are not interested in estimating. We assume that we have a prior probability distribution function (PDF) \( pr(\phi) \) for the nuisance parameters. This PDF is defined on the \( p \)-dimensional nuisance parameter space \( \Phi \). The model for the data PDF is then given by

\[
pr(A | \theta) = \int_{\Phi} pr(A | \phi, \theta) pr(\phi) d^p \phi
\]

If the attribute space is partitioned into bins and the number of photons in each bin is counted then this will produce an integer data vector \( g \) whose dimension is the number of bins. In this case the data model is given by

\[
Pr(g | \theta) = \int_{\Phi} Pr(g | \phi, \theta) pr(\phi) d^p \phi
\]
where $Pr (g|\theta)$ and $Pr (g|\phi, \theta)$ are probabilities rather than PDFs. We will be using the list-mode model throughout, but all of the results will hold with minor notational changes for binned data.

The score vector $s (A|\theta)$ is defined by $s (A|\theta) = \nabla_{\theta} \ln Pr (A|\theta)$. We can write the score vector as

$$s (A|\theta) = \int_{\phi} Pr (A|\phi, \theta) Pr (\phi) \nabla_{\theta} \ln Pr (A|\phi, \theta) d\phi Pr (A|\theta)$$

Using the posterior PDF

$$Pr (\phi|A, \theta) = \frac{Pr (A|\phi, \theta) Pr (\phi)}{Pr (A|\theta)}$$

and the local score vector

$$s (A|\phi, \theta) = \nabla_{\theta} \ln Pr (A|\phi, \theta)$$

the (global) score vector is given by

$$s (A|\theta) = \int_{\phi} Pr (\phi|A, \theta) s (A|\phi, \theta) d\phi$$

The Fisher Information Matrix (FIM) is defined by the expectation [1,2]

$$F (\theta) = \langle s (A|\theta) s^\dagger (A|\theta) \rangle_{A|\theta}$$

This matrix then leads to the well known Cramer-Rao lower bound (CRB) on the variance of any unbiased estimator of any component of $\theta$. However, there is another aspect of the FIM which is not as well known. We have shown previously that for small $\Delta \theta$, the quantity $\Delta \theta^\dagger F (\theta) \Delta \theta$ is directly related to the performance of the ideal observer [3] (as measured by the area under the ROC curve) on the task of classifying whether a data matrix $A$ was generated by the pdf $Pr (A|\theta)$ or the nearby pdf $Pr (A|\theta + \Delta \theta)$. In other words, this scalar quantity measures our ability to detect small changes in the parameter vector $\theta$ [4,5,6]. It is useful to consider this scalar to develop inequalities and then use them to derive inequalities for the FIM. We now have

$$\Delta \theta^\dagger F (\theta) \Delta \theta = \left\langle \left[ \int_{\phi} Pr (\phi|A, \theta) \Delta \theta^\dagger s (A|\phi, \theta) d\phi \right]^2 \right\rangle_{A|\theta}$$

This expectation gives us the detectability of a small change in the parameter of interest. On the other hand, with $s (A|\phi, \theta) = \nabla_{\theta} \ln Pr (A|\phi, \theta)$, the average FIM component for fixed nuisance parameters is

$$\left\langle \Delta \theta^\dagger F (\phi, \theta) \Delta \theta \right\rangle_{\phi} = \int_{\phi} Pr (\phi) \left\langle \left[ \Delta \theta^\dagger s (A|\phi, \theta) \right]^2 \right\rangle_{A|\phi, \theta} d\phi
We can write this as
\[
\langle \triangle \theta ^\dagger F (\phi, \theta) \triangle \theta \rangle_\phi = \left\langle \int_\phi \text{pr} (\phi | A, \theta) \left[ \triangle \theta ^\dagger s (A | \phi, \theta) \right]^2 d^p \phi \right\rangle_{A|\theta}
\]
This would apply to the case where the nuisance parameters vary randomly, but we have some way to measure them in each case. This expectation gives us the average detectability of a small change in the parameter of interest, where the average is over the random nuisance parameters. The difference between these two quantities is positive and given by
\[
\langle \triangle \theta ^\dagger F (\phi, \theta) \triangle \theta \rangle_\phi - \langle \triangle \theta ^\dagger F (\theta) \triangle \theta \rangle_\phi = \left\langle \text{var} \left[ \triangle \theta ^\dagger s (A | \phi, \theta) | A, \theta \right] \right\rangle_{A|\theta}
\]
This gives us the bound
\[
\triangle \theta ^\dagger F (\theta) \triangle \theta \leq \langle \triangle \theta ^\dagger F (\phi, \theta) \triangle \theta \rangle_\phi
\]
This is not an unexpected result. It says that, on average, it is better to know the values of the nuisance parameters when we are trying to detect a change in the parameters of interest. In terms of the FIMs themselves we have the relation
\[
\langle F (\phi, \theta) \rangle_\phi - F (\theta) = \langle \text{cov} \left[ s (A | \phi, \theta) | A, \theta \right] \rangle_{A|\theta}
\]
and therefore the matrix inequality
\[
F (\theta) \leq \langle F (\phi, \theta) \rangle_\phi
\]
For the Cramer-Rao bound we then have
\[
[F (\theta)]^{-1} \geq [\langle F (\phi, \theta) \rangle_\phi]^{-1}
\]
It is not as easy to provide an interpretation for this inequality since, on the right, we are averaging over the nuisance parameters before we take the inverse. We would like to say that the average CRB when we know the nuisance parameters is less than the CRB when we do not, but this statement would require averaging over the nuisance parameters after we take the inverse of the matrices F (\phi, \theta).

3 Signal dependent nuisance parameters

We may be faced with a situation where the PDF for the nuisance parameters depends on the parameters of interest. In this case we have
\[
\text{pr} (A | \theta) = \int_\phi \text{pr} (A | \phi, \theta) \text{pr} (\phi | \theta) d^p \phi
\]
The score vector is given by
\[
s (A | \theta) = \frac{\int_\phi \text{pr} (A | \phi, \theta) \text{pr} (\phi | \theta) \nabla_\theta \ln \text{pr} (A | \phi, \theta) d^p \phi}{\text{pr} (A | \theta)} + \int_\phi \text{pr} (A | \phi, \theta) \text{pr} (\phi | \theta) \nabla_\theta \ln \text{pr} (\phi | \theta) d^p \phi
\]
We write this as

\[ s(A|\theta) = \int_\phi p_r(\phi|A, \theta) s(A|\phi, \theta) \, d\phi + \int_\phi p_r(\phi|A, \theta) s(\phi|\theta) \, d\phi \]

with

\[ s(A|\phi, \theta) = \nabla_\theta \ln p_r(A|\phi, \theta) \]

and

\[ s(\phi|\theta) = \nabla_\theta \ln p_r(\phi|\theta) \]

We now have

\[ \triangle^\dagger F(\theta) \triangle \theta = \left\langle \left( \triangle^\dagger s(A|\phi, \theta) + \triangle^\dagger s(\phi|\theta) \right)^2 \right\rangle_{\phi|A, \theta} \]

Since the variance of any random variable is always nonegative we can write

\[ \triangle^\dagger F(\theta) \triangle \theta \leq \left\langle \left( \left[ \triangle^\dagger s(A|\phi, \theta) + \triangle^\dagger s(\phi|\theta) \right] \right)^2 \right\rangle_{\phi|A, \theta} \]

Now we reverse the order of the expectations and have

\[ \triangle^\dagger F(\theta) \triangle \theta \leq \left\langle \left[ \triangle^\dagger s(A|\phi, \theta) + \triangle^\dagger s(\phi|\theta) \right] \right\rangle_{\phi|A, \theta} \]

Since the mean of the score vector is zero, the cross term vanishes when we expand the square of the term in square brackets. Thus we have

\[ \triangle^\dagger F(\theta) \triangle \theta \leq \left\langle \left( \triangle^\dagger s(A|\phi, \theta) \right)^2 \right\rangle_{A|\phi, \theta} + \left\langle \left( \triangle^\dagger s(\phi|\theta) \right)^2 \right\rangle_{\phi|\theta} \]

On the right side we have two contributions. One would give us the average over the nuisance parameters of the square of the detectability of a small change in the parameter of interest if we knew the random nuisance parameters via some other measurement. The second term gives us the square of the detectability of a small change in the parameters of interest from the measurement of the nuisance parameters themselves. From the first inequality the difference between the right-hand side and left-hand side in the second inequality is given by the average variance:

\[ \left\langle \text{var}_\phi \left[ \triangle^\dagger s(A|\phi, \theta) + \triangle^\dagger s(\phi|\theta) |A, \theta \right] \right\rangle_{A, \theta} \]

The subscript on the variance function indicates the random vector whose PDF is used in the expectations to compute that variance. The vectors to the right of the vertical bar in the variance function indicate what vectors are being held fixed when computing the expectations. In terms of the relevant matrices we now can write
The last term on the right is also a Fisher information matrix $F_{\phi}(\theta)$. This matrix measures the information contained in the nuisance parameters about the parameters of interest. Now we can write this inequality as

$$F(\theta) \leq \langle F(\phi, \theta) \rangle_{\phi, \theta} + \left\langle |\nabla_{\theta} \ln pr(\phi|\theta)| |\nabla_{\theta} \ln pr(\phi|\theta)|^{\dagger} \right\rangle_{\phi, \theta} \left\langle |\nabla_{\theta} \ln pr(\phi|\theta)| \right\rangle_{\phi, \theta} \left\langle |\nabla_{\theta} \ln pr(\phi|\theta)|^{\dagger} \right\rangle_{\phi, \theta}$$

$$F(\theta) \leq \langle F(\phi, \theta) \rangle_{\phi, \theta} + F_{\phi}(\theta)$$

It is difficult to interpret this inequality in terms of the CRB since the inverse of the sum of two matrices is not easily relatable to the sum of their inverses. The difference between the matrix on the right and the one on the left in this inequality is a covariance matrix

$$\langle \text{cov}_{\phi} \left[ s(A|\phi, \theta) + s(\phi|\theta) | A, \theta \right] \rangle_{A|\theta}$$

### 4 Relation to Bayesian FIM for joint model

The averages of FIMs that appear in the previous two sections are reminiscent of averages over parameters that appear in the van Trees inequality, also known as the Bayesian CRB [7,8]. For this reason we will call the version of the FIM that appears in the Bayesian CRB the Bayesian FIM. To compute the Bayesian FIM we need a prior distribution $pr(\theta)$ on the parameters of interest. The van Trees inequality then uses the Bayesian FIM to provide a lower bound for the EMSE when we are estimating these parameters. This inequality, like the CRB, requires inverting the Bayesian FIM. We can also show that the Bayesian FIM, without inversion, is directly related to the average Shannon information for the task of detecting a small change in the parameters of interest.

For the Bayesian FIM we define posterior score vectors via the posterior distribution as $s_\theta = \nabla_{\theta} \ln pr(\theta, \phi|A)$ and $s_\phi = \nabla_{\phi} \ln pr(\theta, \phi|A)$. The Bayesian FIM for the pair $(\theta, \phi)$ is then given by the matrix

$$\mathbf{F}_J = \left[ \begin{array}{cc} s_\theta s_\theta^\dagger & s_\theta s_\phi^\dagger \\ s_\phi s_\theta^\dagger & s_\phi s_\phi^\dagger \end{array} \right]_{A|\phi, \theta} \left[ \begin{array}{c} s_\phi^\dagger \\ s_\phi \end{array} \right]_{\phi, \theta} = \left[ \begin{array}{cc} F_{\theta\theta} & F_{\theta\phi} \\ F_{\phi\theta} & F_{\phi\phi} \end{array} \right]$$

The subscript $J$ here refers to the fact that the corresponding estimation problem in the van Trees inequality would be estimating the pair $(\theta, \phi)$ jointly. This corresponds to one approach to estimating $\theta$ in the presence of nuisance parameters contained in $\phi$, which is to estimate both vectors and then ignore the estimate of the nuisance vector.

Using the definition of the posterior distribution

$$pr(\theta, \phi|A) = \frac{pr(A|\phi, \theta) pr(\phi|\theta) pr(\theta)}{pr(A)}$$
we find that the posterior score vectors are given by
\[ s_\theta = \frac{\nabla_{\theta \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} + \frac{\nabla_{\theta \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} + \frac{\nabla_{\theta \mathsf{pr}} (\theta)}{pr (\theta)} \]
and
\[ s_\phi = \frac{\nabla_{\phi \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} + \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \]
The matrix \( F_{\theta\theta} \) is a sum of three components: \( F_{\theta\theta} = \langle \langle F_{11} (\theta, \phi) \rangle \rangle_{\phi|\theta} + \langle F_{11} (\theta) \rangle_{\theta} + F_{11} \). The three matrices appearing on the right in this equation are
\[ F_{11} (\theta, \phi) = \langle \left[ \frac{\nabla_{\theta \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} \right] \left[ \frac{\nabla_{\theta \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} \right] \rangle_{A|\phi, \theta} \]
\[ F_{11} (\theta) = \langle \left[ \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \right] \left[ \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \right] \rangle_{\phi|\theta} \]
and
\[ F_{11} = \langle \left[ \frac{\nabla_{\theta \mathsf{pr}} (\theta)}{pr (\theta)} \right] \left[ \frac{\nabla_{\theta \mathsf{pr}} (\theta)}{pr (\theta)} \right] \rangle_{\theta} \]
This is the principal matrix of interest, but for completeness we also provide:
\( F_{\phi\phi} = \langle \langle F_{22} (\theta, \phi) \rangle \rangle_{\phi|\theta} + \langle F_{22} (\theta) \rangle_{\theta} \) with
\[ F_{22} (\theta, \phi) = \langle \left[ \frac{\nabla_{\phi \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} \right] \left[ \frac{\nabla_{\phi \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} \right] \rangle_{A|\phi, \theta} \]
and
\[ F_{22} (\theta) = \langle \left[ \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \right] \left[ \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \right] \rangle_{\phi|\theta} \]
We also have \( F_{\theta\phi} = \langle \langle F_{12} (\theta, \phi) \rangle \rangle_{\phi|\theta} + \langle F_{12} (\theta) \rangle_{\theta} \) with
\[ F_{12} (\theta, \phi) = \langle \left[ \frac{\nabla_{\phi \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} \right] \left[ \frac{\nabla_{\phi \mathsf{pr}} (A|\phi, \theta)}{pr (A|\phi, \theta)} \right] \rangle_{A|\phi, \theta} \]
and
\[ F_{12} (\theta) = \langle \left[ \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \right] \left[ \frac{\nabla_{\phi \mathsf{pr}} (\phi|\theta)}{pr (\phi|\theta)} \right] \rangle_{\phi|\theta} \]
The component \( F_{\phi\phi} \) is the transpose of \( F_{\theta\phi} \).
From the results in Section 3 we have, in the notation in this section,
\[ F (\theta) \leq \langle \langle F_{11} (\theta, \phi) \rangle \rangle_{\phi|\theta} + F_{11} (\theta) \].
The Bayesian FIM for the model in Section 3, where we are marginalizing over \( \phi \), is defined by
\[
F_M = \left\langle \left( \left[ \nabla_\theta \ln p_r(\theta|A) \right] \left[ \nabla_\theta \ln p_r(\theta|A) \right]^\dagger \right) \right\rangle_{\theta|A}. 
\]
The subscript \( M \) refers to the marginalization of the nuisance parameters in the model. We now have \( F_M = (F(\theta))_\theta + F_{11} \), which implies that \( F_M \leq F_{\theta\theta} \). If we are trying to detect a small change \( \Delta \theta \) in the parameter vector of interest, then we have shown that \( \Delta \theta^\dagger F_M \Delta \theta \) is a useful figure of merit related to the average SI [9] for this detection task [10]. This SI can in turn be related to the ideal observer AUC via an integral transform [11,12]. We then have
\[
\Delta \theta^\dagger F_M \Delta \theta \leq \Delta \theta^\dagger F_{\theta\theta} \Delta \theta = \left[ \begin{array}{c} \Delta \theta \\ 0 \end{array} \right]^\dagger F_J \left[ \begin{array}{c} \Delta \theta \\ 0 \end{array} \right] 
\]
On the right in this inequality is the same figure of merit when \( \phi \) is not marginalized out and \( \Delta \phi = 0 \). We have therefore the not too surprising conclusion that our ability to detect a change in the parameter vector of interest is increased, on average, if we know the value of the random nuisance parameter vector.

5 Approximate change in the FIM due to nuisance parameter uncertainty

There are situations where we have nominal values for the nuisance parameters but there is still some uncertainty in their actual values. If the nominal values are the components of the vector \( \phi \), then the probability of the data conditional on the parameters of interest is given by
\[
pr(A|\theta) = \int_\Omega pr(A|\phi, \theta) pr(\Delta \phi) d^q \Delta \phi 
\]
We will assume that the mean of the error vector \( \Delta \phi \) is zero. We want to find an approximation to the FIM that will be useful if the PDF \( pr(\Delta \phi) \) is concentrated around the origin in the nuisance parameter space.

We start with Taylor series expansion for the PDF \( pr(A|\theta) \):
\[
pr(A|\theta) = \int_\Omega \left[ pr(A|\phi, \theta) + \Delta \phi^\dagger \nabla_\phi pr(A|\phi, \theta) + \frac{1}{2} \Delta \phi^\dagger \nabla_\phi \nabla_\phi^\dagger pr(A|\phi, \theta) \Delta \phi + \ldots \right] pr(\Delta \phi) d^q \Delta \phi 
\]

Using the fact that the mean error vector is zero we find the lowest order terms;
\[
pr(A|\theta) = pr(A|\phi, \theta) + \frac{1}{2} \text{tr} \left[ K_\phi \nabla_\phi \nabla_\phi^\dagger pr(A|\phi, \theta) \right] + \ldots 
\]
In this equation the matrix \( K_\phi \) is the covariance matrix for \( \Delta \phi \). We may now derive an approximate expression for the FIM which makes use of the constant coefficient second order differential operator
\[
\mathcal{L}_\phi = \frac{1}{2} \nabla_\phi^\dagger K_\phi \nabla_\phi. 
\]
We can write to lowest order in $K_\phi$,

$$
pr(A|\theta) = pr(A|\phi, \theta) + \mathcal{L}_\phi pr(A|\phi, \theta) + \ldots
$$

Now we can use the series expansion for the logarithm to write

$$
\ln pr(A|\theta) = \ln pr(A|\phi, \theta) + \frac{\mathcal{L}_\phi pr(A|\phi, \theta)}{pr(A|\phi, \theta)} + \ldots
$$

We will use these two expansions to derive a series expansion for the FIM $F(\theta)$.

By definition the FIM in question is given by

$$
F(\theta) = \int_D [\nabla_\theta \ln pr(A|\theta)] [\nabla_\theta \ln pr(A|\theta)]^\dagger pr(A|\theta) d^M g
$$

To lowest order there are three correction terms

$$
F(\theta) = F(\phi, \theta) + F_1(\phi, \theta) + F_2(\phi, \theta) - F_3(\phi, \theta) + \ldots
$$

The correction terms in this expansion are

$$
F_1(\phi, \theta) = \int_D [\nabla_\theta \ln pr(A|\phi, \theta)] [\nabla_\theta \mathcal{L}_\phi pr(A|\phi, \theta)]^\dagger d^M g,
$$

$$
F_2(\phi, \theta) = \int_D [\nabla_\theta \mathcal{L}_\phi pr(A|\phi, \theta)] [\nabla_\theta \ln pr(A|\phi, \theta)]^\dagger d^M g
$$

and

$$
F_3(\phi, \theta) = \int_D [\nabla_\theta \ln pr(A|\phi, \theta)] [\nabla_\theta \ln pr(A|\phi, \theta)]^\dagger \mathcal{L}_\phi pr(A|\phi, \theta) d^M g.
$$

These correction terms can be computed numerically using the same Monte Carlo methods commonly used to compute the FIM $F(\phi, \theta)$.

The expansion derived here and the inequalities derived above leave open the possibility that $F(\theta) > F(\phi, \theta)$ for some particular value $\phi$ of the nuisance parameter. Does this inequality make sense? To see that it can be a valid inequality consider a particular estimation task. Suppose that we have an imaging system at one end of an L-shaped hallway and that there is a small light source around the corner of that hallway. We want to know the location of that light source, so this is the parameter vector of interest $\theta$. Assume that there is a swinging door in the leg of the hallway occupied by the source and that this door, if closed, completely blocks any light from the source from reaching our imaging system. The nuisance parameter $\phi$ will be the angle that the door makes with the wall it is attached to, so that $\phi = 0$ or $\phi = \pi$ when the door is wide open, and $\phi = \pi/2$ when it is closed. If the nominal value for the nuisance parameter is $\phi = \pi/2$, then $F(\phi, \theta) = 0$. In other words, if we are certain that the door is closed, then the FIM for $\theta$ is the zero matrix, since no light from the source is reaching our detector. On the other hand if there is some uncertainty in $\phi$, i.e. if the door might be open a little by some random angle $\Delta \phi$, then the FIM $F(\theta)$ is not the zero matrix, since the PDF $pr(A|\theta)$ includes contributions from configurations where the door is open a little. This does not violate the inequality $F(\theta) \leq \langle F(\phi, \theta) \rangle_{\phi}$ in Section 2 since the matrix on the right also includes contributions from configurations where the door is open a little.
6 Conclusion

In Section 2 we developed and inequality for the FIM when the prior PDF on the nuisance parameter vector is independent of the parameters of interest. This inequality states that the FIM for estimating $\theta$, the vector parameter of interest is smaller than the average FIM for the pair $(\theta, \phi)$, where $(\phi)$ is the vector of nuisance parameters and the average is over these nuisance parameters. Section 3 we generalized this inequality to the case where the PDF for the nuisance parameters depends on the parameters of interest. In this case the inequality involves an extra term which is the FIM for the conditional PDF $p_{\theta}(\phi|\theta)$. In section 4 we described the implications of inequalities derived in Section 3 for the Bayesian FIM. We found, not surprisingly, that in terms of the Bayesian FIM it is, on average, better if we know the values of the nuisance parameters than if we do not. This is, however, an average result which can be violated for individual instances of the nuisance parameter vector. Finally, in Section 4 we presented an approximation to the FIM that could be useful if the prior PDF for the nuisance parameters is narrowly distributed around a nominal value that is known. This approximation only requires knowledge of the covariance matrix of the distribution of the nuisance parameter vector around its mean, which is assumed to be the nominal value.

The inequalities provide upper bounds for the FIM and Bayesian FIM when nuisance parameters are present. In general, these upper bounds are easier to compute than the FIM in question. When the nuisance parameters are known to within some error, then the approximation could be useful and is again easier to compute than the original FIM.

References

[1] J. Shao, *Mathematical Statistics*, Springer, New York (1999).

[2] H. H. Barrett, K. J. Myers, *Foundations of Image Science*, John Wiley & Sons, Hoboken, NJ (2004).

[3] H. Barrett, C. Abbey, E. Clarkson, "Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions," J. Opt. Soc. Am. A 15, 1520-1535 (1998).

[4] E. Clarkson, F. Shen, “Fisher information and surrogate figures of merit for the task-based assessment of image quality,” JOSA A 27, 2313-2326 (2010).

F. Shen, E. Clarkson, “Using Fisher information to approximate ideal observer performance on detection tasks for lumpy-background images,” JOSA A 23, 2406-2414 (2006).
[6] E. Clarkson, "Asymptotic ideal observers and surrogate figures of merit for signal detection with list-mode data," J. Opt. Soc. Am. A 29, 2204-2216 (2012).

[7] H. L. van Trees, Detection, Estimation and Modulation Theory, Part 1, New York, Wiley, (1968).

[8] R.D. Gill, B.Y. Levit, “Applications of the van Trees inequality: a Bayesian Cramér–Rao bound,” Bernoulli 1, 59–79 (1995).

[9] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley & Sons, Hoboken, NJ (2006).

[10] E. Clarkson, “The relation between Bayesian Fisher information and Shannon information for detecting a change in a parameter,” IEEE Trans. Information Theory, (Submitted)

[11] E. Clarkson, J. Cushing, "Shannon information and ROC analysis in imaging," J. Opt. Soc. Am. A 32, 1288-1301 (2015).

[12] E. Clarkson, J. Cushing, "Shannon information and receiver operating characteristic analysis for multiclass classification in imaging," J. Opt. Soc. Am. A 33, 930-937 (2016).