Estimation of Channelized Hotelling Observer Performance With Known Class Means or Known Difference of Class Means

Adam Wunderlich, *Graduate Student Member, IEEE,* and Frédéric Noo*, *Member, IEEE*

Abstract—This paper concerns task-based image quality assessment for the task of discriminating between two classes of images. We address the problem of estimating two widely-used detection performance measures, SNR and AUC, from a finite number of images, assuming that the class discrimination is performed with a channelized Hotelling observer. In particular, we investigate the advantage that can be gained when either 1) the means of the signal-absent and signal-present classes are both known, or 2) when the difference of class means is known. For these two scenarios, we propose uniformly minimum variance unbiased estimators of SNR^2, derive the corresponding sampling distributions and provide variance expressions. In addition, we demonstrate how the bias and variance for the related AUC estimators may be calculated numerically by using the sampling distributions for the SNR^2 estimators. We find that for both SNR^2 and AUC, the new estimators have significantly lower bias and mean-square error than the traditional estimator, which assumes that the class means, and their difference, are unknown.

Index Terms—AUC, class discrimination, estimation, image quality, receiver operating characteristic (ROC), signal-to-noise ratio (SNR).

I. INTRODUCTION

T ASK-BASED assessment of image quality provides a pragmatic paradigm for objective evaluation and comparison of image reconstruction algorithms and imaging systems [1]. One task that is frequently of interest is the detection of a specified object in an image. For example, an important task in radiology is the discrimination of two classes of images: images with a lesion present and with no lesion present.

Lesion detectability may be appraised with both human observers and mathematical model observers [1]. However, human observers come with prominent difficulties. They are highly variable, slow, costly, and subject to fatigue, to the extent that they become impractical, especially for imaging system optimization, as noted by Park *et al.* [2]. By contrast, mathematical model observers [1] can be implemented efficaciously using computers. Thus, they offer an attractive alternative to human observers for task-based assessment of image quality. One type of model observer which has been given significant attention in medical imaging is the channelized Hotelling observer (CHO) [1]. Specifically, CHO methodology for image quality assessment has been applied in nuclear medicine, e.g., [3]–[8] and in X-ray computed tomography (CT) [9]–[12].

To assess the capability of a given model observer to detect lesions, summary measures of observer performance need to be estimated. Two popular measures of observer performance are the observer signal-to-noise ratio (SNR) and the area under the receiver operating characteristic curve, denoted as AUC [1] (see Section II-C). In most situations, it is not possible to calculate SNR and AUC exactly, and they must be estimated from a finite collection of images. Therefore, it is important to understand the small-sample statistical properties of the estimators which are used. We note that two main approaches may be considered for the estimation of SNR and AUC: 1) an SNR-based approach, in which the SNR is first estimated, and then the AUC is obtained from the SNR estimate [1]; 2) an AUC-based approach, in which the AUC is first estimated, and then the SNR is obtained from the AUC estimate [1]. Under conditions of normality, both approaches are valid because SNR and AUC are functions of each other [1]. Here, we focus on the SNR-based approach. Some researchers have investigated direct estimation of the AUC, e.g., [13]–[15]; these works rely on the Mann–Whitney U statistic.

Note that two main estimation strategies may be selected for the two approaches described above. In the first strategy, all available images are used together to estimate the desired figure of merit [1, p. 972]. In the second strategy, the images are divided into two groups, with one group being used to define (aka train) the observer whereas the second group is used to estimate the performance of the trained observer (aka testing the observer) [1, p. 973]. The second strategy has the advantage of yielding a negative (conservative) bias in the estimated performance [16], [17]. In this paper, we adopt the first strategy, and we will see later that negative biases may also be achieved in this case, provided that the class means, or their difference, are known.

In general, deriving the small-sample properties of estimators is difficult. However, analytical results are known for some cases. For example, if SNR^2 for a CHO is estimated using a scalar multiple of the two-sample Hotelling T^2 statistic, then the sampling distribution is known for normally distributed measurements [18, p. 216]; see Section IV-A. Also, results pertaining to direct AUC estimation via the nonparametric

A. Wunderlich is with the Utah Center for Advanced Imaging Research, Department of Radiology, University of Utah, Salt Lake City, UT 84108 USA (e-mail: awunder@ucair.med.utah.edu).

*F. Noo is with the Utah Center for Advanced Imaging Research, Department of Radiology, University of Utah, Salt Lake City, UT 84108 USA (e-mail: noo@ucair.med.utah.edu).

Mann–Whitney U statistic may be found in [15]. When exact analytical results are not readily available, Taylor series approximations may be beneficial [19], [20], or one may resort to analysis based on Monte Carlo simulation and resampling techniques [16], [17], [21]–[27].

As mentioned by Barrett and Myers [1, p. 972], there is potential to improve the estimation accuracy of SNR^2 for a CHO if prior knowledge of the mean images with and without a lesion present is incorporated into the estimation procedure. However, to our knowledge, and according to Barrett and Myers [1, p. 973], the statistical properties of such estimators have not been investigated. We have examined this potential and report on our findings here.

Specifically, we propose uniformly minimum variance unbiased (UMVU) estimators of SNR^2 for a CHO in the situation when the mean images for the lesion present and lesion absent cases are known, and also when only the difference of the mean images is known. We present analytical results for the sampling distributions of these estimators and provide expressions for their variance. Using the sampling distributions of the SNR^2 estimators, we illustrate how the bias and variance for the related AUC estimators may be numerically evaluated, and we compare the results with a traditional estimator which assumes that the class means, and their difference, are unknown. For expositional clarity, all mathematical proofs are deferred to the appendixes.

II. BACKGROUND

In this section, we set our notation and review several necessary concepts. First, some probability distributions that will be used later in the paper are introduced. Afterwards, the classification task and the channelized Hotelling observers which we consider in this paper are covered. Next, two common summary measures for model observer performance, SNR and AUC, are reviewed. The remainder of the section discusses calculation of moments for AUC estimators. Throughout the paper, we write vectors using boldface and denote the transpose of vectors and matrices with the superscript T, e.g., $\mathbf{x}^T$ is the transpose of the vector $\mathbf{x}$.

A. Some Probability Distributions

We assume that the reader is familiar with both the multivariate normal distribution and the univariate noncentral χ^2 distribution, e.g., see [18] and [28]. A $p \times 1$ random vector $\mathbf{X} \in \mathbb{R}^p$ following a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ is denoted $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$. A random variable $X \in \mathbb{R}$ following a noncentral χ^2 distribution with ν degrees of freedom and noncentrality parameter, λ, is denoted $X \sim \chi_\nu^2(\lambda)$. Recall that when $\lambda = 0$, the noncentral χ^2 distribution becomes the central χ^2 distribution.

1) Inverted Gamma: The inverted gamma distribution arises as the distribution of the reciprocal of a gamma variate, and has two positive parameters, α and β. A random variable X is said to have an inverted gamma distribution if its probability density function (pdf) is of the form [29]

$$f_X(x) = \frac{\beta^\alpha e^{-\beta/x}}{\Gamma(\alpha)x^{\alpha+1}} \tag{1}$$

when $x > 0$, and $f_X(x) = 0$ otherwise. Just as the central χ^2 distribution is a special case of the gamma distribution, the inverted central χ^2 distribution is a special case of the inverted gamma distribution. In particular, the reciprocal of a central χ^2 random variable with ν degrees of freedom is an inverted gamma random variable with $\alpha = \nu/2$ and $\beta = 1/2$. An inverted gamma random variable X will be denoted as $X \sim \text{IG}(\alpha, \beta)$. The mean and variance of an inverted gamma deviate are [29]

$$\text{E}[X] = \frac{\beta}{\alpha - 1}, \quad \text{for} \quad \alpha > 1 \tag{2}$$

and

$$\text{Var}[X] = \frac{\beta^2}{(\alpha - 1)^2(\alpha - 2)}, \quad \text{for} \quad \alpha > 2. \tag{3}$$

2) Noncentral F: The noncentral F distribution [18], [28]–[30] appears as the distribution of the ratio of a noncentral χ^2 random variable to an independent, central χ^2 random variable. Specifically, if $X \sim \chi_{\nu_1}^2(\lambda)$, $Y \sim \chi_{\nu_2}^2(0)$ and X and Y are independent, then the random variable $F = (X/\nu_1)/(Y/\nu_2)$ has a noncentral F distribution with ν_1 and ν_2 degrees of freedom and noncentrality parameter, λ. In this case, we write $F \sim F'_{\nu_1, \nu_2}(\lambda)$. An expression for the pdf of the noncentral F may be found in [28].

3) Wishart: The Wishart distribution [18], [28], [31] is a matrix variate generalization of the χ^2 distribution and it emerges naturally as the distribution of the sample covariance matrix for multivariate normal measurements [18, p. 82], [31, p. 92–93]. Let $\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n$ each be independently distributed according to $N_p(\mathbf{0}, \Sigma)$ and consider the $p \times p$ random matrix $W = \sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T$. The matrix W is said to have a Wishart distribution, written as $W \sim W_p(n, \Sigma)$, with n degrees of freedom and $p \times p$ positive definite scale matrix Σ. When $n \geq p$, W is invertible and expressions for the pdf of W may be found in [18], [28], [31]. When $n < p$, W is singular, and the pdf does not exist in the conventional sense, but the distribution is nonetheless defined [18, p. 85].

4) Inverted Wishart: The inverted Wishart distribution originates as the distribution of the inverse of a Wishart distributed random matrix, and it is the matrix variate generalization of the inverted gamma distribution [31, p. 111]. A $p \times p$ random matrix V which is distributed as an inverted Wishart with m degrees of freedom and a $p \times p$ positive definite parameter matrix, Ψ, will be written as $V \sim \text{IW}_p(m, \Psi)$. The inverted Wishart distribution is defined if $m > 2p$ and is undefined otherwise. An expression for the pdf is given in [18], [31]. Lemma 4 in Appendix A clarifies the relationship between Wishart and Inverted Wishart matrices. According to this lemma, when $n \geq p$, $S \sim W_p(n, \Sigma)$ if and only if $S^{-1} \sim \text{IW}_p(n + p + 1, \Sigma^{-1})$.

B. Channelized Hotelling Observers

We consider a binary classification task in which the goal is to discriminate between two classes of noisy images. The two classes will be denoted as class 1 and class 2, respectively. When the classification task corresponds to lesion detection, we assume that class 1 consists of images with no lesion present, and that class 2 consists of images with a lesion present.

For each noisy image realization, a model observer generates a statistic, t, which is compared to a threshold, t_c, in order to classify the image as belonging to either class 1 or class 2. Specifically, if $t > t_c$, the observer indicates that the image belongs to class 2, otherwise, the observer indicates that the image belongs to class 1 [1].

Channel filters are often added to model observers in order to reduce the dimensionality of the image data and to model human performance. Suppose that the noisy image realization is represented as a $m \times 1$ column vector, $\mathbf{g}$. We denote the number of channels by p, where usually, p is much smaller than m. The weights that make up each channel are organized into a column of a $m \times p$ channel matrix, designated as U. A channelized model observer applies the channel matrix to the image to get a $p \times 1$ output vector, $\mathbf{v}$, according to $\mathbf{v} := U^T \mathbf{g}$. Note that a model observer without channels corresponds to the case when the channel matrix is the identity, i.e., $U = I$ and $p = m$.

We denote the means of the channel outputs over class 1 and class 2 by $\boldsymbol{\mu}_i$, for $i = 1, 2$ respectively, and write the difference of these means as $\Delta\boldsymbol{\mu} := \boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$. Also, we write the covariance matrices for $\mathbf{v}$ over the two classes as Σ_i, for $i = 1, 2$, respectively, and designate their average as $\bar{\Sigma} := 0.5(\Sigma_1 + \Sigma_2)$.

One type of model observer is the channelized Hotelling observer (CHO) [1], which is the focus of this paper. This observer computes the observer statistic as the inner product of $\mathbf{v}$ with a $p \times 1$ template, defined as $\mathbf{w}_{\text{CHO}} := \bar{\Sigma}^{-1}\Delta\boldsymbol{\mu}$, i.e., $t := \mathbf{w}_{\text{CHO}}^T \mathbf{v}$.

C. Observer Performance Measures

One measure of an observer's performance is its signal-to-noise ratio (SNR), defined as the difference of the class means for t divided by the pooled standard deviation [1, p. 819]. The SNR is a good measure of class separability when t is normally distributed for each class, higher values of the observer SNR indicating greater class separation [1, p. 819]. The square of the SNR for a CHO may be computed as [1, p. 967]

$$\text{SNR}^2 = \Delta\boldsymbol{\mu}^T \bar{\Sigma}^{-1} \Delta\boldsymbol{\mu}. \tag{4}$$

The performance of an observer may also be characterized by its receiver operating characteristic (ROC) curve, which plots the true positive fraction (TPF) versus the false positive fraction (FPF) for different values of the discrimination threshold [1, p. 814–815]. A useful summary measure for the overall performance of an observer is the area under the ROC curve, denoted as the AUC. The AUC ranges from 0.5 to 1, where higher values of the AUC indicate better class discrimination. The AUC may be interpreted as the average TPF over all FPF values. Another useful interpretation of the AUC is as the probability that the observer makes a correct classification when a randomly selected pair of images from class 1 and class 2 are compared [32, p. 77–78], [1, p. 823]. If the observer statistic, t, is normally distributed for each class, then the AUC may be computed from the SNR as [1, p. 819]

$$\text{AUC} = (1/2)[1 + \text{erf}(\text{SNR}/2)] \tag{5}$$

where $\text{erf}(z)$ is the conventional error function. When t is computed as a linear combination of channel outputs, the observer

statistic is typically well-approximated by normal distributions for each class [1, p. 824].

D. Calculation of Moments for AUC Estimators

Given a generic estimator of SNR^2, denoted as $\widehat{\text{SNR}^2}$, we may define an estimator for the AUC by substituting the square root of this generic estimator into (5). We denote the resulting AUC estimator as $\widehat{\text{AUC}}$.

When the sampling distribution for $\widehat{\text{SNR}^2}$ is known, we may use it to calculate the kth raw moment of $\widehat{\text{AUC}}$ by numerically evaluating the integral

$$\text{E}[\widehat{\text{AUC}}^k] = \int_0^\infty f_X(x) \left(\frac{1}{2}[1 + \text{erf}(\sqrt{x}/2)] \right)^k dx \tag{6}$$

where $f_X(x)$ is the pdf for $X = \widehat{\text{SNR}^2}$. Therefore, we can compute the mean and variance of $\widehat{\text{AUC}}$ by evaluating (6) for $k = 1, 2$.

III. ESTIMATION OF SNR^2 WITH KNOWN CLASS MEANS OR KNOWN DIFFERENCE OF CLASS MEANS

In this section, estimation of SNR^2 is considered for two scenarios. In the first scenario, the means for classes 1 and 2 are assumed to be known. In the second scenario, only the difference of the class means is assumed to be known. We introduce one estimator for scenario 1 and two different estimators for scenario 2. We assume that we have n independent realizations of channel outputs from class 1, $\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_n^{(1)}$, and n independent realizations of channel outputs from class 2, $\mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \ldots, \mathbf{v}_n^{(2)}$. Also, we assume that the channel outputs are normally distributed with equal class covariance matrices, i.e., $\mathbf{v}_i^{(1)} \sim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \sim N_p(\boldsymbol{\mu}_2, \Sigma)$ for $i = 1, 2, \ldots n$. In this case, $\bar{\Sigma} = \Sigma$. The normality is a reasonable assumption in many CHO applications thanks to the multivariate central limit theorem, and the equal covariance assumption is easily met for low-contrast signals. After defining each estimator, we provide an analytical expression for the estimator's sampling distribution and discuss its properties.

A. Scenario 1: Known Class Means

First, we consider the situation when the class means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are known, and the class covariance matrix, Σ, is unknown.

1) Estimator Definition: In this scenario, unbiased estimates for the covariance matrices of class 1 and class 2 are [18, p. 17]

$$\hat{S}_k := \frac{1}{n} \sum_{i=1}^n \left(\mathbf{v}_i^{(k)} - \boldsymbol{\mu}_k \right) \left(\mathbf{v}_i^{(k)} - \boldsymbol{\mu}_k \right)^T \tag{7}$$

for $k = 1, 2$, respectively. Let $\hat{S} := 0.5(\hat{S}_1 + \hat{S}_2)$. From Lemmas 2 and 8(i) in Appendix A, it follows that $((2n - p - 1)/(2n))\hat{S}^{-1}$ is an unbiased estimator for Σ^{-1} if $n > (p+1)/2$. Substituting this estimator for Σ^{-1} into (4) yields an estimator for SNR^2, defined as

$$\hat{\theta}_1 := \left(\frac{2n - p - 1}{2n} \right) \Delta\boldsymbol{\mu}^T \hat{S}^{-1} \Delta\boldsymbol{\mu} \tag{8}$$

for $n > (p+1)/2$.

2) *Sampling Distribution:* For normally distributed samples, the sampling distribution for $\hat{\theta}_1$ may be expressed as an inverted gamma distribution. We state this result in the following theorem, which is proved in Appendix A.

Theorem 1: Suppose that $\hat{\theta}_1$ is computed from independent samples $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$. Then

$$\hat{\theta}_1 \backsim \text{IG}\left(\frac{2n - p + 1}{2}, \left(\frac{2n - p - 1}{2}\right)\text{SNR}^2\right)$$

for $n > (p + 1)/2$.

3) *Estimator Properties:* Since the sampling distribution for $\hat{\theta}_1$ is an inverted gamma distribution, one may easily find expressions for the mean and variance of $\hat{\theta}_1$ by using Theorem 1 together with (2) and (3). We write these expressions in the following corollary.

Corollary 1: If $\hat{\theta}_1$ is computed from independent, normally distributed samples as stated in Theorem 1, then $\hat{\theta}_1$ is an unbiased estimator of SNR^2, i.e., $\text{E}[\hat{\theta}_1] = \text{SNR}^2$ for $n > (p+1)/2$, and its variance is

$$\text{Var}[\hat{\theta}_1] = \left(\frac{2}{2n - p - 3}\right)\text{SNR}^4, \quad \text{for} \quad n > \frac{p + 3}{2}.$$

Also, we have the pleasing result that $\hat{\theta}_1$ is the uniformly minimum variance unbiased (UMVU) estimator for SNR^2 when the class means are known. For reference, we state this result as a theorem; see Appendix B for a proof.

Theorem 2: Suppose that $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are known, and that Σ is unknown. If $\hat{\theta}_1$ is computed from independent, normally distributed samples as stated in Theorem 1 with $n > (p + 1)/2$, then $\hat{\theta}_1$ is the unique UMVU estimator of SNR^2.

B. Scenario 2: Known Difference of Class Means

Next, we examine the case for which the difference of the class means, $\Delta\boldsymbol{\mu}$, is known but $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and Σ are unknown. We consider two estimators of SNR^2 for this scenario.

1) *Estimator Definitions:* The first estimator which we consider for scenario 2 incorporates knowledge of $\Delta\boldsymbol{\mu}$ into the estimation of Σ. The conventional unbiased estimators for the means of classes 1 and 2 are [28, p. 77]

$$\bar{\mathbf{v}}_k := \frac{1}{n}\sum_{i=1}^{n}\mathbf{v}_i^{(k)} \tag{9}$$

for $k = 1, 2$, respectively. Using (9) together with our knowledge of $\Delta\boldsymbol{\mu}$, we define unbiased estimators for the means of classes 1 and 2 as

$$\tilde{\boldsymbol{\mu}}_1 := (1/2)(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 - \Delta\boldsymbol{\mu}) \tag{10}$$

and

$$\tilde{\boldsymbol{\mu}}_2 := (1/2)(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 + \Delta\boldsymbol{\mu}). \tag{11}$$

Using these estimators for the class means, we define estimators for the covariance matrices of classes 1 and 2 as

$$\tilde{S}_k := \frac{1}{n - 1/2}\sum_{i=1}^{n}\left(\mathbf{v}_i^{(k)} - \tilde{\boldsymbol{\mu}}_k\right)\left(\mathbf{v}_i^{(k)} - \tilde{\boldsymbol{\mu}}_k\right)^T \tag{12}$$

for $k = 1, 2$, respectively. Next, we define a pooled estimator of Σ as $\tilde{S} := 0.5(\tilde{S}_1 + \tilde{S}_2)$. From Lemmas 2 and 8(ii) in Appendix A, it follows that $((2n - p - 2)/(2n - 1))\tilde{S}^{-1}$ is an unbiased estimator for Σ^{-1} if $n > (p + 2)/2$. This suggests the following estimator for SNR^2:

$$\hat{\theta}_2 := \left(\frac{2n - p - 2}{2n - 1}\right)\Delta\boldsymbol{\mu}^T\tilde{S}^{-1}\Delta\boldsymbol{\mu} \tag{13}$$

which is defined for $n > (p + 2)/2$.

The second estimator which we consider in this scenario does not use knowledge of $\Delta\boldsymbol{\mu}$ for estimation of Σ. The conventional unbiased estimators for the covariance matrices of classes 1 and 2 have the form [28, p. 77]

$$S_k := \frac{1}{n - 1}\sum_{i=1}^{n}\left(\mathbf{v}_i^{(k)} - \bar{\mathbf{v}}_k\right)\left(\mathbf{v}_i^{(k)} - \bar{\mathbf{v}}_k\right)^T \tag{14}$$

for $k = 1, 2$, respectively. Define a pooled estimator of Σ as $\bar{S} := 0.5(S_1 + S_2)$. From Lemmas 2 and 8(iii) in Appendix A, it follows that $((2n - p - 3)/(2n - 2)))\bar{S}^{-1}$ is an unbiased estimator for Σ^{-1} if $n > (p + 3)/2$. This result motivates an estimator for SNR^2, defined as

$$\hat{\theta}_3 := \left(\frac{2n - p - 3}{2n - 2}\right)\Delta\boldsymbol{\mu}^T\bar{S}^{-1}\Delta\boldsymbol{\mu} \tag{15}$$

for $n > (p + 3)/2$.

2) *Sampling Distributions:* For normally distributed samples, the sampling distributions for $\hat{\theta}_2$ and $\hat{\theta}_3$ may be expressed as inverted gamma distributions. We state these results below. Proofs are provided in Appendix A.

Theorem 3: Suppose that $\hat{\theta}_2$ is computed from independent samples $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$. Then

$$\hat{\theta}_2 \backsim \text{IG}\left(\frac{2n - p}{2}, \left(\frac{2n - p - 2}{2}\right)\text{SNR}^2\right)$$

for $n > (p + 2)/2$.

Theorem 4: Suppose that $\hat{\theta}_3$ is computed from independent samples $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$. Then

$$\hat{\theta}_3 \backsim \text{IG}\left(\frac{2n - p - 1}{2}, \left(\frac{2n - p - 3}{2}\right)\text{SNR}^2\right)$$

for $n > (p + 3)/2$.

3) *Estimator Properties:* Similar to Section III-A3, expressions for the mean and variance of $\hat{\theta}_2$ and $\hat{\theta}_3$ may be easily found by using (2) and (3) together with Theorems 3 and 4, respectively. These results are collected below.

Corollary 2: If $\hat{\theta}_2$ is computed from independent, normally distributed samples as stated in Theorem 3, then $\hat{\theta}_2$ is an unbiased estimator of SNR^2, i.e., $\text{E}[\hat{\theta}_2] = \text{SNR}^2$ for $n > (p+2)/2$, and its variance is

$$\text{Var}[\hat{\theta}_2] = \left(\frac{2}{2n - p - 4}\right)\text{SNR}^4, \quad \text{for} \quad n > \frac{p + 4}{2}.$$

Corollary 3: If $\hat{\theta}_3$ is computed from independent, normally distributed samples as stated in Theorem 4, then $\hat{\theta}_3$ is an unbi-

ased estimator of SNR^2, i.e., $E[\hat{\theta}_3] = SNR^2$ for $n > (p+3)/2$, and its variance is

$$Var[\hat{\theta}_3] = \left(\frac{2}{2n-p-5}\right) SNR^4, \quad \text{for } n > \frac{p+5}{2}.$$

As in Section III-A3, we have a satisfying optimality property. Namely, if only the difference of class means is known, then $\hat{\theta}_2$ is the UMVU estimator for SNR^2. This property is stated precisely as the following theorem. A proof may be found in Appendix B.

Theorem 5: Suppose that $\Delta\boldsymbol{\mu}$ is known, and that $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and Σ are unknown. If $\hat{\theta}_2$ is computed from independent, normally distributed samples as stated in Theorem 3 with $n > (p+2)/2$, then $\hat{\theta}_2$ is the unique UMVU estimator of SNR^2.

From the above results, it is clear that $\hat{\theta}_3$ is not the UMVU estimator for SNR^2 in scenario 2. However, $\hat{\theta}_3$ is the UMVU estimator within the family of unbiased estimators of SNR^2 that disregard the knowledge that $\Delta\boldsymbol{\mu}$ is equal to $\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1$. This statement is clarified by the following theorem. We omit the proof, which is similar to that of Theorem 5 in Appendix B.

Theorem 6: Suppose that $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ are independent samples from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$ with $n > (p+3)/2$, and where $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and Σ are unknown. Let $\boldsymbol{\delta}$ be an arbitrary, known $p \times 1$ column vector, and define the estimator

$$\hat{\phi} := \left(\frac{2n-p-3}{2n-2}\right) \boldsymbol{\delta}^T \bar{S}^{-1} \boldsymbol{\delta}.$$

Then $\hat{\phi}$ is the unique UMVU estimator of $\boldsymbol{\delta}^T \Sigma^{-1} \boldsymbol{\delta}$.

IV. COMPUTATIONAL EVALUATIONS

Now, we further explore the theoretical properties of the estimators introduced in the last section. We will compare these estimators to a conventional estimator for SNR^2, which is reviewed first. As before, we assume that there are $\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_n^{(1)}$ independent realizations of channel outputs from class 1 and $\mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \ldots, \mathbf{v}_n^{(2)}$ independent realizations of channel outputs from class 2. Also, we will be assuming that $p = 40$ channels are used. There is nothing special about $p = 40$, as opposed to another value, such as $p = 4$, which is typical for isotropic channels. Our choice of $p = 40$ is motivated by our current interest in evaluating the impact of anisotropic noise on lesion detectability [11].

A. An Estimator for SNR^2 When the Class Means and Their Difference Are Unknown

Suppose that the class means, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, and their difference, $\Delta\boldsymbol{\mu}$, are unknown. In this situation, the conventional method to construct an estimator for SNR^2 is to substitute the sample estimates $\Delta\bar{\mathbf{v}} := \bar{\mathbf{v}}_2 - \bar{\mathbf{v}}_1$ and $\bar{S}$ for $\Delta\boldsymbol{\mu}$ and Σ in (4). This approach is used in [1, p. 972], and yields an estimator defined as

$$\hat{\theta}_4 := \Delta\bar{\mathbf{v}}^T \bar{S}^{-1} \Delta\bar{\mathbf{v}}, \quad \text{for } n > \frac{p+1}{2}. \tag{16}$$

The sampling distribution for $\hat{\theta}_4$ is known in the case when the samples are normally distributed with equal covariance matrices, i.e., $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ for $i = 1, 2, \ldots n$. Specifically, [18, p. 216]

$$\left(\frac{n(2n-p-1)}{4p(n-1)}\right) \hat{\theta}_4 \backsim F'_{\nu_1, \nu_2}(\lambda), \quad \text{for} \quad n > \frac{p+1}{2} \tag{17}$$

where $F'_{\nu_1, \nu_2}(\lambda)$ is the noncentral F distribution with $\nu_1 = p$ and $\nu_2 = 2n - p - 1$ degrees of freedom and noncentrality parameter, $\lambda = (n/2)SNR^2$. Expressions for the mean and variance of the noncentral F distribution [18], [30] may be used to derive formulae for the bias and variance of $\hat{\theta}_4$. These formulae are [33]

$$bias[\hat{\theta}_4] = E[\hat{\theta}_4] - SNR^2 = \frac{n(p+1)SNR^2 + 4(n-1)p}{2n^2 - np - 3n} \tag{18}$$

for $n > (p+3)/2$, and

$$Var[\hat{\theta}_4] = \frac{32(n-1)^2 \left[\frac{n^2}{4}SNR^4 + (2n-3)(nSNR^2 + p)\right]}{n^2(2n-p-3)^2(2n-p-5)} \tag{19}$$

for $n > (p+5)/2$.

B. Comparison of the SNR^2 Estimators

In the first set of comparisons, the relative error for each of the four SNR^2 estimators is plotted as a function of the total number of images, $2n$; see Fig. 1. The plots assume SNR^2 values of 0.12837, 0.90987, and 3.2847, corresponding to AUC values of 0.6, 0.75, and 0.9, respectively. Here the relative error (in %) is calculated as the square root of the mean-square error (MSE) divided by the SNR^2 value times 100. Using the fact that the MSE is equal to the bias squared plus the variance [34], the MSE for $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ and $\hat{\theta}_4$ was calculated using the expressions from Corollaries 1, 2, and 3 and (18) and (19).

Comparing $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$ in the plots, we observe that their relative errors are very similar. This is not surprising, since they are all unbiased and their variance expressions are comparable. In addition, we note that for fixed n and p, the relative errors for $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_3$ are invariant over the different SNR^2 values. Although this invariance is only shown in the plots for particular values of n and p, it actually holds for any choice of n and p. This follows from the fact that the standard deviation of each estimator divided by SNR^2 only depends on n and p; this quantity is independent of SNR^2 and also of the correlation matrix and the class means.

On the other hand, the relative error for $\hat{\theta}_4$ decreases as the true SNR^2 value increases. Furthermore, we see that the relative error for $\hat{\theta}_4$ is significantly larger than that for $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$. For example, in the case of $SNR^2 = 0.90987$ with $2n = 250$ total images, the relative error for $\hat{\theta}_1$ is 9.8%, compared to a relative error of 112.3% for $\hat{\theta}_4$.
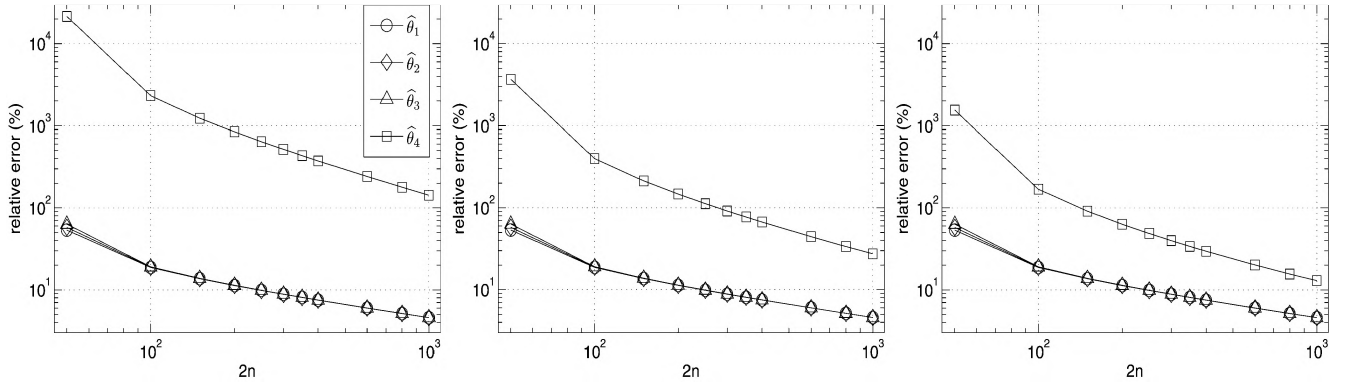
Fig. 1. Relative error of SNR^2 estimators as a function of the total number of images, $2n$. The plots assume $p = 40$ channels with $\mathrm{SNR}^2 = 0.12837$ (left), 0.90987 (middle), and 3.2847 (right), corresponding to AUC values of 0.6, 0.75, and 0.9, respectively. The estimators are denoted as follows: $\hat{\theta}_1$ (circles), $\hat{\theta}_2$ (diamonds), $\hat{\theta}_3$ (triangles), and $\hat{\theta}_4$ (squares). The relative error (in %) is calculated as $(\sqrt{\mathrm{MSE}}/(\mathrm{SNR}^2)) \times 100$, where MSE denotes the mean-square error.
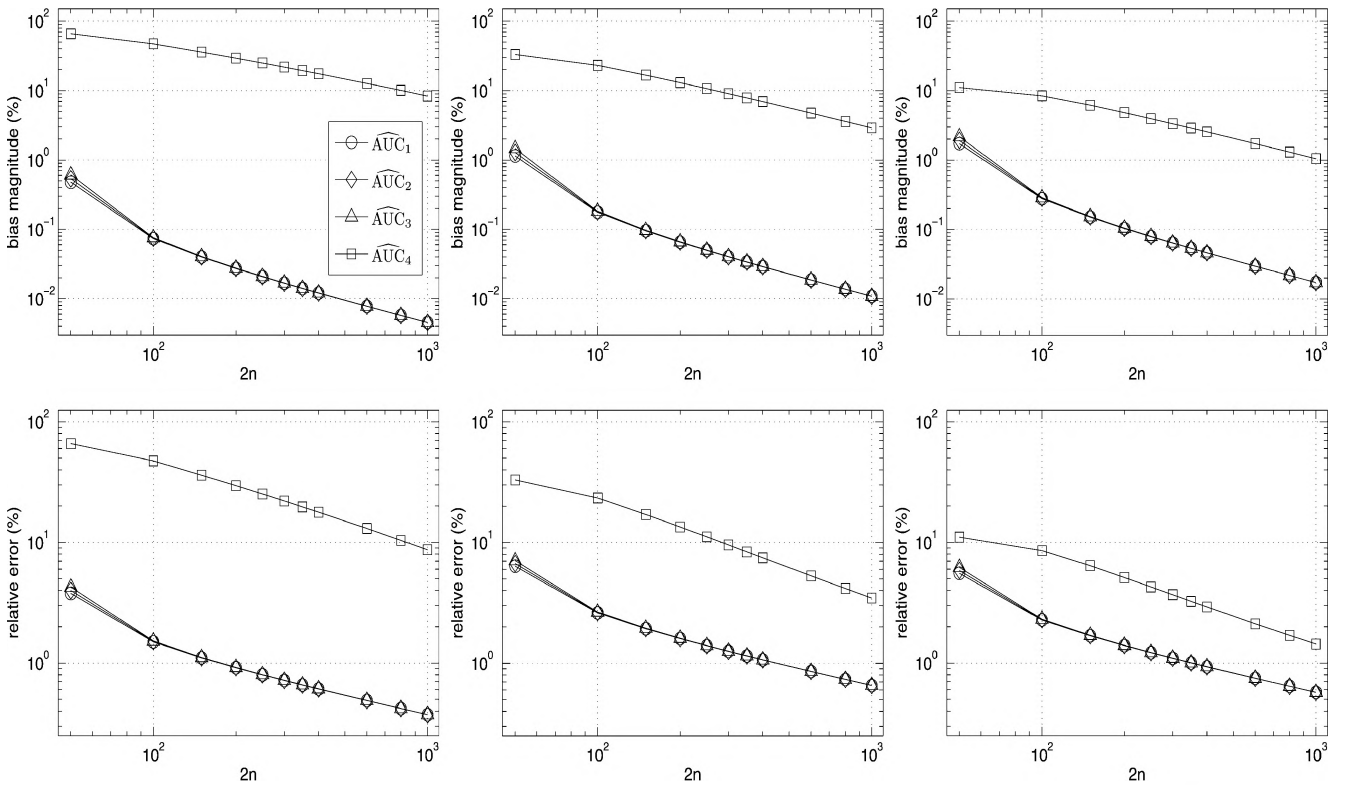

Fig. 2. Bias magnitude and relative error of AUC estimators as a function of the total number of images, $2n$. The plots assume $p = 40$ channels with AUC = 0.6 (left), 0.75 (middle), and 0.9 (right). Bias magnitude is on the top and relative error is on the bottom. The estimators are $\widehat{\mathrm{AUC}}_1$ (circles), $\widehat{\mathrm{AUC}}_2$ (diamonds), $\widehat{\mathrm{AUC}}_3$ (triangles), $\widehat{\mathrm{AUC}}_4$ (squares). For each $i \in \{1, 2, 3, 4\}$, $\widehat{\mathrm{AUC}}_i$ is calculated by substituting the square root of $\hat{\theta}_i$ for SNR in (5). The bias magnitude (in %) for an AUC estimator, $\widehat{\mathrm{AUC}}_i$, is calculated as $(|\mathrm{E}[\widehat{\mathrm{AUC}}_i] - \mathrm{AUC}|/\mathrm{AUC}) \times 100$. The relative error (in %) is calculated as $(\sqrt{\mathrm{MSE}}/\mathrm{AUC}) \times 100$, where MSE denotes the mean-square error.

C. Comparison of the AUC Estimators

For the next set of comparisons, we consider the AUC estimators which are found by substituting the square root of $\hat{\theta}_i$ for SNR in (5). For each $i \in \{1, 2, 3, 4\}$, the resulting AUC estimator will be denoted as $\widehat{\mathrm{AUC}}_i$.

In Fig. 2, the bias magnitude and relative error for each of the AUC estimators are plotted as functions of the total number of images, $2n$. The plots compare the estimator behavior for AUC values of 0.6, 0.75, and 0.9. The bias magnitude (in %) is calculated by dividing the absolute value of the bias by the AUC value and multiplying by 100. The relative error (in %) is calculated as the square root of the MSE divided by the AUC value

times 100. The bias and MSE for each AUC estimator were computed from the estimator's mean and variance, which were calculated numerically using the pdf of the corresponding $\hat{\theta}_i$ and the method described in Section II-D. In all cases, the numerical error for the mean and variance results was constrained to be less than 10^{-6}.

Examining the bias magnitude plots in Fig. 2 (top row), we see that the bias magnitudes for $\widehat{\mathrm{AUC}}_1$, $\widehat{\mathrm{AUC}}_2$, and $\widehat{\mathrm{AUC}}_3$ are very similar, and increase at fixed n with increasing AUC values, but are generally very small. Specifically, for $2n \geq 100$, the bias magnitudes for these three estimators are all less than 0.3%. In contrast, the bias magnitude for $\widehat{\mathrm{AUC}}_4$ decreases with in-

creasing AUC values, but it is always much larger than for the other three estimators. For instance, in the case of $2n = 100$, the bias magnitude for $\widehat{\text{AUC}}_4$ has values of 47.1%, 23.1%, and 8.4%, corresponding to AUC values of 0.6, 0.75, and 0.9, respectively. Although not reflected in the plots, the biases for $\widehat{\text{AUC}}_1$, $\widehat{\text{AUC}}_2$, and $\widehat{\text{AUC}}_3$ were negative in all cases and the bias for $\widehat{\text{AUC}}_4$ was positive in all cases.

Directing our attention to the relative error plots in Fig. 2 (bottom row), the trends of $\widehat{\text{AUC}}_1$, $\widehat{\text{AUC}}_2$, and $\widehat{\text{AUC}}_3$ are again seen to be very similar. For $2n \geq 100$, the relative errors for these three estimators are less than 2.7%. On the other hand, the relative error for $\widehat{\text{AUC}}_4$ is significantly larger. For example, when $2n = 100$, the relative error for $\widehat{\text{AUC}}_4$ has values of 47.4%, 23.4%, and 8.55% corresponding to AUC values of 0.6, 0.75, and 0.9, respectively. Comparing these relative errors to the bias magnitude values mentioned above, we see that most of the relative error for $\widehat{\text{AUC}}_4$ is due to bias.

V. DISCUSSION AND CONCLUSION

In this work, we were interested in the estimation of classical performance metrics for discrimination between two classes of images. The discrimination was assumed to be performed by a channelized Hotelling observer. Also, we assumed that the estimation is to be carried out using a small number of image realizations, as is typically the case in medical imaging.

Unfortunately, access to a limited number of image realizations implies that the estimated performance values will have some statistical errors, quantified, for example, in terms of estimator bias and variance. Barrett and Myers [1, p. 972] suggested that these statistical errors could potentially be reduced if the class means are assumed to be known. We have investigated this issue thoroughly. Assuming that the channel outputs corresponding to the two classes of images have multivariate normal distributions with equal covariance matrices, we have completely characterized the sampling distributions for three estimators of SNR^2 that assume that the class means, or their difference, are known. Using these distributions, we have demonstrated that a significant decrease in estimation error can be achieved when the class means, or their difference, are known.

A useful property of our three SNR^2 estimators is the fact that the ratio of each estimator's standard deviation to its mean is a function of only n and p. This property allows for easy preselection of the number of images needed to achieve a desired level of accuracy.

Another interesting property of our three SNR^2 estimators is that they yield AUC estimators that have negative bias, just as estimators based on a training/testing paradigm [16], [17]. Thus, we demonstrated that the training/testing paradigm is not the only estimation strategy that can result in negative biases. Moreover, recall from Fig. 2 that the biases achieved by our new estimators are fairly small.

Of course, it is natural to ask in which situations the image class means, or their difference, may be found. One situation where this is possible is for image reconstruction from simulated data with the data mean known. For example, in the case of linear reconstruction algorithms (e.g., filtered backprojection (FBP) type), which are still the methods of choice in CT,

the image class means may be found by simply reconstructing the means of the data for each class. Similarly, for many types of (nonlinear) iterative reconstruction methods, the image class means may be found to first order by reconstructing the means of the data for each class [35]. This approximation has been found to be fairly accurate for the cases of expectation maximization (EM) [36], [37] and penalized-likelihood reconstruction [35].

When significant uncertainty exists for the class mean estimates, our estimators may not be satisfactory. This could be the case for iterative reconstruction methods if first-order estimates of the means are not accurate enough. This could also be the case for image quality assessment from real data, because the class means of real data are rarely known exactly. For these situations, it would be useful to understand how our estimators are affected by errors in the mean estimates. Moreover, as refined estimates of the class means may possibly be obtained at low cost, we note that knowledge of the distributions for our estimators may pave the way for new attractive estimators that incorporate prior information pertaining to the class means. These topics define interesting areas of further research.

In order to safely apply our results to a particular imaging application, we emphasize that it is necessary to verify that the channel outputs corresponding to the two classes of images are well-approximated by multivariate normal distributions that have the same covariance matrix. We are currently assessing the validity of this assumption for X-ray CT. This assessment will include a comparison of our approach to nonparametric AUC estimation using the Mann–Whitney U statistic.

Last, note that our estimators allow for a significant improvement in efficiency for the computation of detectability maps. These maps plot SNR or AUC as a function of lesion location [1, p. 858] [38]. This is one way to investigate the dependence of lesion detectability on background variation [1, p. 858–859]. In the future, we will evaluate our estimators for the purpose of building detectability maps for commercially-available FBP reconstruction methods in X-ray CT.

APPENDIX A

In this Appendix, we prove Theorems 1, 3, and 4 which characterize the sampling distributions of $\hat{\theta}_1$, $\hat{\theta}_2$, and $\hat{\theta}_3$. For this task, we first review several useful properties of the Wishart, inverted Wishart, and inverted gamma distributions.

Lemma 1: If $A_1 \backsim W_p(m, \Sigma)$ and $A_2 \backsim W_p(n, \Sigma)$ are independent, $p \times p$ random matrices, then $A_1 + A_2 \backsim W_p(m+n, \Sigma)$.

Proof: See [18, Theorem 3.2.4, p. 91] or [31, Theorem 3.3.8, p. 94].

Lemma 2: Let $n > p + 1$. If $S \backsim W_p(n, \Sigma)$, then

$$\text{E}[S^{-1}] = \frac{1}{n - p - 1}\Sigma^{-1}.$$

Proof: See [18, p. 97] or [28, p. 273].

Lemma 3: Let $n \geq p$. If $S \backsim W_p(n, \Sigma)$ and A is a $q \times p$ matrix of rank $q \leq p$, then

$$(AS^{-1}A^T)^{-1} \backsim W_q(n - p + q, (A\Sigma^{-1}A^T)^{-1}).$$

Proof: Proof: See [18, Theorem 3.2.11, p. 95] or [31, Theorem 3.3.13, p. 97].

Lemma 4: Let $n \geq p$. Then $S \backsim W_p(n, \Sigma)$ if and only if $S^{-1} \backsim \mathrm{IW}_p(n + p + 1, \Sigma^{-1})$.

Proof: See [31, Theorem 3.4.1, p. 111].

Lemma 5: Let $n \geq p$. If $S \backsim W_p(n, \Sigma)$ and A is a $q \times p$ matrix of rank $q \leq p$, then

$$AS^{-1}A^T \backsim \mathrm{IW}_q(n - p + 2q + 1, A\Sigma^{-1}A^T).$$

Proof: This follows immediately from Lemmas 3 and 4.

Lemma 6: If $X \backsim \mathrm{IW}_1(m, \Psi)$ and $m > 2$, then $X \backsim \mathrm{IG}((m - 2)/2, \Psi/2)$.

Proof: Suppose that $X \backsim \mathrm{IW}_1(m, \Psi)$ and $m > 2$. First, note that since X is an inverted Wishart matrix with $p = 1$, X and Ψ are scalars. For the $p = 1$ case, the pdf of the inverted Wishart distribution [31, p. 111] is

$$f_X(x) = \frac{(\Psi/2)^{\frac{1}{2}(m-2)} e^{-\frac{\Psi}{2x}}}{\Gamma\left(\frac{1}{2}(m - 2)\right) x^{\frac{1}{2}m}}. \tag{20}$$

Comparing with (1), we recognize the above expression as the pdf for an inverted gamma random variable with $\alpha = (m-2)/2$ and $\beta = \Psi/2$.

Lemma 7: Let $c > 0$ be an arbitrary constant. If $X \backsim \mathrm{IG}(\alpha, \beta)$, then $Y = cX \backsim \mathrm{IG}(\alpha, c\beta)$.

Proof: Suppose that $X \backsim \mathrm{IG}(\alpha, \beta)$ and let $Y = cX$. The pdf for Y may be expressed in terms of the pdf for X by applying the rule for a monotonic transformation of a random variable [39, p. 51], i.e., $f_Y(y) = (1/c)f_X(y/c)$. Using this rule together with (1), we may express the pdf for Y as

$$f_Y(y) = \frac{1}{c}\left(\frac{\beta^\alpha e^{-c\beta/y}}{\Gamma(\alpha)}\right)\left(\frac{c}{y}\right)^{\alpha+1} = \frac{(c\beta)^\alpha e^{-c\beta/y}}{\Gamma(\alpha)y^{\alpha+1}}. \tag{21}$$

Hence, $Y \backsim \mathrm{IG}(\alpha, c\beta)$.

Next, we collect some results concerning the distributions of $\hat{S}$, $\tilde{S}$, and $\bar{S}$.

Lemma 8: Suppose that $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ are independent samples from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$. If $\hat{S}$, $\tilde{S}$, and $\bar{S}$ are computed from these samples according to (7), (12), and (14), respectively, then (i) $2n\hat{S} \backsim W_p(2n, \Sigma)$, (ii) $(2n - 1)\tilde{S} \backsim W_p(2n - 1, \Sigma)$, and (iii) $(2n - 2)\bar{S} \backsim W_p(2n - 2, \Sigma)$.

Proof:

i) First, observe that $\hat{S}_1$ may be expressed as $n\widehat{S_1} = \sum_{i=1}^{n} \mathbf{Z}_i\mathbf{Z}_i^T$, where $\mathbf{Z}_i = \mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1$. Since $\mathbf{Z}_1, \mathbf{Z}_2, \ldots \mathbf{Z}_n$ are independently distributed according to $N_p(\mathbf{0}, \Sigma)$, it follows from the definition of the Wishart distribution [18, p. 82] that $n\hat{S}_1 \backsim W_p(n, \Sigma)$. By a similar argument, $n\hat{S}_2 \backsim W_p(n, \Sigma)$. Because $n\hat{S}_1$ and $n\hat{S}_2$ are independent, Lemma 1 implies that $2n\hat{S} = n\hat{S}_1 + n\hat{S}_2 \backsim W_p(2n, \Sigma)$.

ii) From the definition of $\tilde{S}$, we have

$$(2n - 1)\tilde{S} = \sum_{i=1}^{n} \left(\mathbf{v}_i^{(1)} - \tilde{\boldsymbol{\mu}}_1\right)\left(\mathbf{v}_i^{(1)} - \tilde{\boldsymbol{\mu}}_1\right)^T$$
$$+ \sum_{i=1}^{n} \left(\mathbf{v}_i^{(2)} - \tilde{\boldsymbol{\mu}}_2\right)\left(\mathbf{v}_i^{(2)} - \tilde{\boldsymbol{\mu}}_2\right)^T. \tag{22}$$

Substituting $(\mathbf{v}_i^{(1)} - \bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_1 - \tilde{\boldsymbol{\mu}}_1)$ for $(\mathbf{v}_i^{(1)} - \tilde{\boldsymbol{\mu}}_1)$ and $(\mathbf{v}_i^{(2)} - \bar{\mathbf{v}}_2 + \bar{\mathbf{v}}_2 - \tilde{\boldsymbol{\mu}}_2)$ for $(\mathbf{v}_i^{(2)} - \tilde{\boldsymbol{\mu}}_2)$ in (22) and rearranging, we find

$$(2n - 1)\tilde{S} = \sum_{i=1}^{n} \left(\mathbf{v}_i^{(1)} - \bar{\mathbf{v}}_1\right)\left(\mathbf{v}_i^{(1)} - \bar{\mathbf{v}}_1\right)^T$$
$$+ \sum_{i=1}^{n} \left(\mathbf{v}_i^{(2)} - \bar{\mathbf{v}}_2\right)\left(\mathbf{v}_i^{(2)} - \bar{\mathbf{v}}_2\right)^T$$
$$+ n\left(\bar{\mathbf{v}}_1 - \tilde{\boldsymbol{\mu}}_1\right)\left(\bar{\mathbf{v}}_1 - \tilde{\boldsymbol{\mu}}_1\right)^T$$
$$+ n\left(\bar{\mathbf{v}}_2 - \tilde{\boldsymbol{\mu}}_2\right)\left(\bar{\mathbf{v}}_2 - \tilde{\boldsymbol{\mu}}_2\right)^T. \tag{23}$$

Inserting the definitions of $\tilde{\boldsymbol{\mu}}_1$ and $\tilde{\boldsymbol{\mu}}_2$ given by (10) and (11) into the last equation, and simplifying, yields

$$(2n - 1)\tilde{S} = \sum_{i=1}^{n} \left(\mathbf{v}_i^{(1)} - \bar{\mathbf{v}}_1\right)\left(\mathbf{v}_i^{(1)} - \bar{\mathbf{v}}_1\right)^T$$
$$+ \sum_{i=1}^{n} \left(\mathbf{v}_i^{(2)} - \bar{\mathbf{v}}_2\right)\left(\mathbf{v}_i^{(2)} - \bar{\mathbf{v}}_2\right)^T$$
$$+ \left(\frac{n}{2}\right)(\Delta\bar{\mathbf{v}} - \Delta\boldsymbol{\mu})(\Delta\bar{\mathbf{v}} - \Delta\boldsymbol{\mu})^T. \tag{24}$$

By a theorem concerning the distribution of the sample covariance matrix [31, Theorem 3.3.6, p. 92], the first and second terms in (24) are distributed as $W_p(n - 1, \Sigma)$. Also, using standard results for the sample mean and multivariate normal distribution [18], it follows that $\sqrt{n/2}(\Delta\bar{\mathbf{v}} - \Delta\mu) \backsim N_p(\mathbf{0}, \Sigma)$. Therefore, by the definition of the Wishart distribution [18, p. 82], we see that the third term in (24) has a $W_p(1, \Sigma)$ distribution. The first and second terms in (24) are clearly independent because they are computed from independent samples. Moreover, for the same reason, we note that the sample means, $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$, are independent of the second and first terms, respectively. Also, by a standard theorem [31, Theorem 3.3.6, p. 92], $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ are independent of the first and second terms, respectively. Hence, the third term is independent of the first and second terms. Because the three terms in (24) are independent, we may apply Lemma 1 to conclude that $(2n - 1)\tilde{S} \backsim W_p(2n - 1, \Sigma)$.

iii) By a theorem for the distribution of the sample covariance matrix [31 Theorem 3.3.6, p. 92], we have $(n - 1)S_1 \backsim W_p(n-1, \Sigma)$ and $(n-1)S_2 \backsim W_p(n-1, \Sigma)$. In addition, since S_1 and S_2 are computed from independent samples, they are independent. Thus, Lemma 1 implies that $(2n - 2)\bar{S} = (n - 1)S_1 + (n - 1)S_2 \backsim W_p(2n - 2, \Sigma)$.

Now, we are ready to prove Theorems 1, 2, and 3. Recall that each of these estimators is computed from independent samples $\mathbf{v}_i^{(1)} \backsim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \backsim N_p(\boldsymbol{\mu}_2, \Sigma)$ from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$.

Proof. of. Theorem 1: Assume $n > (p + 1)/2$. The definition of $\hat{\theta}_2$ in (8) may be rewritten as $(1/(2n - p - 1))\hat{\theta}_1 = \Delta\boldsymbol{\mu}^T(2n\hat{S})^{-1}\Delta\boldsymbol{\mu}$. From Lemma 8(i), $2n\hat{S} \backsim W_p(2n, \Sigma)$. Hence, we may apply Lemma 5 and (4) to find that $(1/(2n - p - 1))\hat{\theta}_1 \backsim \mathrm{IW}_1(2n - p + 3, \mathrm{SNR}^2)$. Lemma 6 then yields $(1/(2n - p - 1))\hat{\theta}_1 \backsim \mathrm{IG}((2n - p + 1)/2, (1/2)\mathrm{SNR}^2)$. Finally, application of Lemma 7 gives the desired result.

Proof of Theorem 3: Assume $n > (p+2)/2$. The definition of $\hat{\theta}_2$ in (13) may be rewritten as $(1/(2n - p - 2))\hat{\theta}_2 = \Delta\boldsymbol{\mu}^T[(2n-1)\tilde{S}]^{-1}\Delta\boldsymbol{\mu}$. From Lemma 8(ii), $(2n-1)\tilde{S} \sim W_p(2n-1, \Sigma)$. Hence, we may apply Lemma 5 and (4) to find that $(1/(2n-p-2))\hat{\theta}_2 \sim \mathrm{IW}_1(2n-p+2, \mathrm{SNR}^2)$. Lemma 6 then yields $(1/(2n-p-2))\hat{\theta}_2 \sim \mathrm{IG}((2n-p)/2, (1/2)\mathrm{SNR}^2)$. Finally, application of Lemma 7 gives the desired result.

Proof of Theorem 4: Assume $n > (p+3)/2$. The definition of $\hat{\theta}_3$ in (15) may be rewritten as $(1/(2n - p - 3))\hat{\theta}_3 = \Delta\boldsymbol{\mu}^T[(2n-2)\bar{S}]^{-1}\Delta\boldsymbol{\mu}$. From Lemma 8(iii), $(2n-2)\bar{S} \sim W_p(2n-2, \Sigma)$. Hence, we may apply Lemma 5 and (4) to find that $(1/(2n-p-3))\hat{\theta}_3 \sim \mathrm{IW}_1(2n-p+1, \mathrm{SNR}^2)$. Lemma 6 then yields $(1/(2n-p-3))\hat{\theta}_3 \sim \mathrm{IG}((2n-p-1)/2, (1/2)\mathrm{SNR}^2)$. Finally, application of Lemma 7 gives the desired result.

APPENDIX B

In this Appendix, we prove Theorems 2 and 5, which state that $\hat{\theta}_1$ and $\hat{\theta}_2$ are UMVU estimators. Again, we assume that each of these estimators is computed from independent samples $\mathbf{v}_i^{(1)} \sim N_p(\boldsymbol{\mu}_1, \Sigma)$ and $\mathbf{v}_i^{(2)} \sim N_p(\boldsymbol{\mu}_2, \Sigma)$ from classes 1 and 2, respectively, where $i = 1, 2, \ldots n$.

Proof of Theorem 2: Assume $n > (p+1)/2$. Also, suppose that $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are known, and that Σ is unknown. First, we show that $\hat{S}$ is a sufficient statistic for the joint pdf of the sample. The joint pdf of $\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_n^{(1)}$ and $\mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \ldots, \mathbf{v}_n^{(2)}$ is

$$f\left(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_n^{(1)}, \mathbf{v}_1^{(2)}, \ldots, \mathbf{v}_n^{(2)}\right) = (2\pi)^{-np}|\Sigma|^{-n}e^{\eta} \tag{25}$$

where

$$\eta = -\frac{1}{2}\sum_i^n \left[\left(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1\right)^T \Sigma^{-1}\left(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1\right) \right. \\ \left. + \left(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2\right)^T \Sigma^{-1}\left(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2\right)\right]. \tag{26}$$

Using the additive and cyclic properties of the trace, denoted tr, we may rewrite η as

$$\eta = -\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\left(\sum_i^n\left[\left(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1\right)\left(\mathbf{v}_i^{(1)} - \boldsymbol{\mu}_1\right)^T \right.\right.\right. \\ \left.\left.\left. + \left(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2\right)\left(\mathbf{v}_i^{(2)} - \boldsymbol{\mu}_2\right)^T\right]\right)\right) \tag{27}$$

i.e.,

$$\eta = (-n)\mathrm{tr}\left(\Sigma^{-1}\hat{S}\right). \tag{28}$$

Hence, the joint pdf has the form

$$f\left(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_n^{(1)}, \mathbf{v}_1^{(2)}, \ldots, \mathbf{v}_n^{(2)}\right) \\ = (2\pi)^{-np}|\Sigma|^{-n}\exp\left[(-n)\mathrm{tr}\left(\Sigma^{-1}\hat{S}\right)\right]. \tag{29}$$

By the Fisher–Neyman factorization theorem [40, Theorem 6.5, p. 35] [41, Prop. IV.C.1, p. 159], $\hat{S}$ is a sufficient statistic. Moreover, because the expression in (29) has the form of a full rank exponential family [40, p. 23–24], $\hat{S}$ is a complete

statistic [40, Theorem 6.22, p. 42]. Since (i) $\hat{S}$ is a complete sufficient statistic, (ii) $\hat{\theta}_1$ is an unbiased estimator of SNR^2, and (iii) $\hat{\theta}_1 = \mathrm{E}[\hat{\theta}_1|\hat{S}]$, i.e., $\hat{\theta}_1$ is a function of $\hat{S}$ only, the Lehmann–Scheffé Theorem [40, Theorem 1.11, p. 88] [41, p. 164] implies that $\hat{\theta}_1$ is the unique UMVU estimator of SNR^2 in scenario 1.

Proof of Theorem 5: Assume $n > (p+2)/2$. Also, suppose that $\Delta\boldsymbol{\mu}$ is known, and that $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$, and Σ are unknown. The joint pdf of $\mathbf{v}_1^{(1)}, \mathbf{v}_2^{(1)}, \ldots, \mathbf{v}_n^{(1)}$ and $\mathbf{v}_1^{(2)}, \mathbf{v}_2^{(2)}, \ldots, \mathbf{v}_n^{(2)}$ is given by (25) and (26). After lengthy algebra, one may rewrite the joint pdf in the form

$$f\left(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_n^{(1)}, \mathbf{v}_1^{(2)}, \ldots, \mathbf{v}_n^{(2)}\right) \\ = (2\pi)^{-np}|\Sigma|^{-n}\exp\left[(-n)\mathrm{tr}\left(\Sigma^{-1}\boldsymbol{\mu}_1\boldsymbol{\mu}_2^T\right)\right] \\ \times \exp\left[-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}((2n-1)\tilde{S}\right.\right. \\ \left.\left. -\frac{n}{2}(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 - \Delta\boldsymbol{\mu})(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 - \Delta\boldsymbol{\mu})^T)\right)\right] \\ \times \exp\left[n\,\mathrm{tr}\left(\Sigma^{-1}\boldsymbol{\mu}_1(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2)^T\right)\right]. \tag{30}$$

By the Fisher–Neyman factorization theorem [40, Theorem 6.5, p. 35] [41, Prop. IV.C.1, p. 159], the statistic

$$T := \begin{bmatrix} \bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 \\ \tau \end{bmatrix} \tag{31}$$

where $\tau = (2n-1)\tilde{S} - (n/2)(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 - \Delta\boldsymbol{\mu})(\bar{\mathbf{v}}_1 + \bar{\mathbf{v}}_2 - \Delta\boldsymbol{\mu})^T$, is sufficient. Moreover, because the expression in (30) has the form of a full rank exponential family [40, p. 23–24], T is a complete statistic [40, Theorem 6.22, p. 42]. Since (i) T is a complete sufficient statistic, (ii) $\hat{\theta}_2$ is an unbiased estimator of SNR^2, and (iii) $\hat{\theta}_2 = \mathrm{E}[\hat{\theta}_2|T]$, i.e., $\hat{\theta}_2$ is a function of T only, the Lehmann–Scheffé Theorem [40, Theorem. 1.11, p. 88] [41, p. 164] implies that $\hat{\theta}_2$ is the unique UMVU estimator of SNR^2 in scenario 2.

REFERENCES

[1] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. New York: Wiley, 2004.

[2] S. Park, H. H. Barrett, E. Clarkson, M. A. Kupinski, and K. J. Myers, "Channelized-ideal observer using Laguerre-Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal," *J. Opt. Soc. Am. A*, vol. 24, no. 12, pp. B136–B150, Dec. 2007.

[3] S. Wollenweber, B. Tsui, D. Lalush, E. Frey, K. LaCroix, and G. Gullberg, "Comparison of Hotelling observer models and human observers in defect detection from myocardial SPECT imaging," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6, pp. 2098–2103, Dec. 1999.

[4] P. Bonetto, J. Qi, and R. M. Leahy, "Covariance approximation for fast and accurate computation of channelized Hotelling observer statistics," *IEEE Trans. Nucl. Sci.*, vol. 47, no. 4, pp. 1567–1572, Aug. 2000.

[5] H. C. Gifford, M. A. King, D. J. de Vries, and E. J. Soares, "Channelized Hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.*, vol. 41, no. 3, pp. 514–521, Mar. 2000.

[6] J.-S. Kim, R. S. Miyaoka, R. L. Harrison, P. E. Kinahan, and T. K. Lewellen, "Detectability comparisons of image reconstruction algorithms using the channelized Hotelling observer with bootstrap resampled data," in *IEEE Nucl. Sci. Symp. Conf. Rec.*, 2001, pp. 1444–1448.

[7] G. L. Zeng and G. T. Gullberg, "A channelized-Hotelling-trace collimator design method based on reconstruction rather than projections," *IEEE Trans. Nucl. Sci.*, vol. 49, no. 5, pp. 2155–2158, Oct. 2002.

[8] J.-S. Kim, P. E. Kinahan, C. Lartizien, C. Comtat, and T. K. Lewellen, "A comparison of planar versus volumetric numerical observers for detection task performance in whole-body PET imaging," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 1, pp. 34–40, Feb. 2004.

[9] S. J. LaRoque. E. Y. Sidky. D. C. Edwards. and X. Pan. "Evaluation of the channelized Hotelling observer for signal detection in 2D tomographic imaging." in *Proc. SPIE Med. Imag.*. 2007. vol. 6515. p. 651514.

[10] E. Y. Sidky. S. J. LaRoque. and X. Pan. "Accurate computation of the Hotelling observer for the evaluation of image reconstruction algorithms in helical. cone-beam CT." in *IEEE Nucl. Sci. Symp. Conf Rec.*. 2007. pp. 3233–3236.

[11] A. Wunderlich and F. Noo. "Evaluation of the impact of tube current modulation on lesion detectability using model observers." in *Proc. 30th Annu. Int. Conf IEEE Eng. Med. Biol. Soc.*. Aug. 2008. pp. 2705–2708.

[12] A. Wunderlich and F. Noo. "Image covariance and lesion detectability in direct fan-beam X-ray computed tomography." *Phys. Med. Biol.*. vol. 53. no. 10. pp. 2471–2493. May 2008.

[13] D. Bamber. "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph." *J. Math. Psych.*. vol. 12. pp. 387–415. 1975.

[14] J. A. Hanley and B. J. McNeil. "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology*. vol. 143. no. 1. pp. 29–36. Apr. 1982.

[15] W. A. Yousef. R. F. Wagner. and M. H. Loew. "Assessing classifiers from two independent data sets using ROC analysis: A nonparametric approach." *IEEE Trans. Pattern Anal. Mach. Intell.*. vol. 28. no. 11. p. 2006. Nov. 2006.

[16] H.-P. Chan and B. Sahiner. "Classifier design for computer-aided diagnosis: Effects of finite sample size on the mean performance of classical and neural network classifiers." *Med. Phys.*. vol. 26. no. 12. pp. 2654–2668. Dec. 1999.

[17] B. Sahiner. H.-P. Chan. and N. Petrick. "Feature selection and classifier performance in computer-aided diagnosis: The effect of finite sample size." *Med. Phys.*. vol. 27. no. 7. pp. 1509–1522. Jul. 2000.

[18] R. J. Muirhead, *Aspects of Multivariate Statistical Theory.*. New York: Wiley. 1982.

[19] K. Fukunaga and R. R. Hayes. "Effects of sample size in classifier design." *IEEE Trans. Pattern Anal. Mach. Intell.*. vol. 11. no. 8. pp. 873–885. Aug. 1989.

[20] C. K. Abbey. H. H. Barrett. and M. P. Eckstein. "Practical issues and methodology in assessment of image quality using model observers." in *Proc. SPIE Med. Imag.*. 1997. vol. 3032. pp. 182–194.

[21] R. Wagner. D. Brown. G. J.-P. . K. Myers. and K. Wear. . H. Barrett and A. Gmitro. Eds.. "Multivariate gaussian pattern classification: Effects of finite sample size and the addition of correlated or noisy features on summary measures of goodness." in *Information Processing in Medical Imaging*. ser. Lecture Notes in Computer Science # 687. New York: Springer-Verlag. 1993. pp. 507–524.

[22] R. Wagner. H.-P. Chan. B. Sahiner. N. Petrick. and J. Mossoba. "Finite-sample effects and resampling plans: Applications to linear classifiers in computer-aided diagnosis." in *Proc. SPIE Med. Imag.*. 1997. vol. 3034. pp. 467–477.

[23] S. V. Beiden. M. A. Maloof. and R. F. Wagner. "A general model for finite-sample effects in training and testing of competing classifiers." *IEEE Trans. Pattern Anal. Mach. Intell.*. vol. 25. no. 12. pp. 1561–1569. Dec. 2003.

[24] B. D. Gallas. "Variance of the channelized-Hotelling observer from a finite number of trainers and testers." in *Proc. SPIE Med. Imag.*. 2003. vol. 5034. pp. 100–111.

[25] W. A. Yousef. R. F. Wagner. and M. H. Loew. "Comparison of nonparametric methods for assessing classifier performance in terms of ROC parameters." in *Proc. IEEE Comp. Soc. 33rd Appl. Imagery Pattern Recognit. Workshop*. 2004. pp. 190–195.

[26] R. M. Gagne. B. D. Gallas. and K. J. Myers. "Toward objective and quantitative evaluation of imaging systems using images of phantoms." *Med. Phys.*. vol. 33. no. 1. pp. 83–95. Jan. 2006.

[27] B. Sahiner. H.-P. Chan. and L. Hadjiiski. "Classifier performance prediction for computer-aided diagnosis using a limited dataset." *Med. Phys.*. vol. 35. no. 4. pp. 1559–1570. Apr. 2008.

[28] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley. 2003.

[29] M. Evans. N. Hastings. and B. Peacock, *Statistical Distributions*. 2nd ed. New York: Wiley. 1993.

[30] N. L. Johnson. S. Kotz. and Balakrishnan, *Continuous Univariate Distributions*. 2nd ed. New York: Wiley. 1995. vol. 2.

[31] A. Gupta and D. Nagar, *Matrix Variate Distributions*. London. U.K.: Chapman & Hall/CRC. 2000.

[32] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford Univ. Press. 2003.

[33] A. Wunderlich and F. Noo. "Accuracy of channelized Hotelling observer performance measures estimated from repeated CT scans." in *IEEE Nucl. Sci. Symp. Conf Rec.*. Oct. 2008. pp. 4444–4447.

[34] S. M. Kay, *Fundamentals of Statistical Signal Processing*. New York: Prentice Hall. 1993. vol. 1.

[35] J. A. Fessler. "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography." *IEEE Trans. Image Process.*. vol. 5. no. 3. pp. 493–506. Mar. 1996.

[36] H. H. Barrett. D. W. Wilson. and B. M. W. Tsui. "Noise properties of the EM algorithm: I. Theory." *Phys. Med. Biol.*. vol. 39. pp. 833–846. 1994.

[37] D. W. Wilson. B. M. W. Tsui. and H. H. Barrett. "Noise properties of the EM algorithm: II. Monte Carlo simulations." *Phys. Med. Biol.*. vol. 39. pp. 847–871. 1994.

[38] A. R. Pineda. M. Schweiger. S. R. Arridge. and H. H. Barrett. "Information content of data types in time-domain optical tomography." *J. Opt. Soc. Am. A*. vol. 23. no. 12. pp. 2989–2996. Dec. 2006.

[39] G. Casella and R. L. Berger, *Statistical Inference*. 2nd ed. : Duxbury Press. 2001.

[40] E. Lehmann and G. Casella, *Theory of Point Estimation*. 2nd ed. New York: Springer. 1998.

[41] H. V. Poor, *An Introduction to Signal Detection and Estimation*. 2nd ed. New York: Springer. 1994.