

A Probabilistic Development of the MRMC Method

Eric Clarkson [Corresponding Author], Matthew A. Kupinski,
and Harrison H. Barrett

College of Optical Sciences, The University of Arizona, 1630 East

University Blvd., Tucson, Arizona 85721;

(TEL) 520-626-7280; (FAX) 520-626-2892;

(email) clarkson@radiology.arizona.edu

Abstract

Rationale and Objectives: Current approaches to ROC analysis use the MRMC (multiple-reader, multiple-case) paradigm in which several readers read each case and their ratings (or scores) are used to construct an estimate of the area under the ROC curve or some other ROC-related parameter. Standard practice is to decompose the parameter of interest according to a linear model into terms that depend in various ways on the readers, cases and modalities. Though the methodological aspects of MRMC analysis have been studied in detail, the literature on the probabilistic basis of the individual terms is sparse. In particular, few papers state what probability law applies to each term and what underlying assumptions are needed for the assumed independence. When probability distributions are specified for these terms, these distributions are assumed to be Gaussians.

Materials and Methods: This paper approaches the MRMC problem from a mechanistic perspective. For a single modality, three sources of randomness are included: the images, the reader skill and the reader uncertainty. The probability law on the reader scores is written in terms of three nested conditional probabilities, and random variables associated with this probability are referred to as triply stochastic. **Results and Discussion:** In this paper, we present the probabilistic MRMC model and apply this model to the Wilcoxon statistic. The result is a seven-term expansion for the variance of the figure of merit. We relate the terms in this expansion to those in the standard, linear MRMC model. Finally, we use the probabilistic model to derive constraints on the coefficients in the seven-term expansion.

Keywords: ROC analysis, multiple reader multiple case, Wilcoxon statistic

1. INTRODUCTION

The multiple-reader, multiple-case paradigm is often used to assess the performance of a new medical-imaging system or to compare the performances of two or more such systems. In this paradigm, we first select a random sample of abnormal and normal cases. Each of these cases is individually read by each member in a sample of readers. Each reader produces a test statistic for each image which measures his or her confidence that an abnormality is present. This array of test statistics is used to generate a figure of merit. An important issue is the variance of this figure of merit as a function of the number of readers and cases. This is the issue addressed by standard, linear MRMC models [1–3] and by the probabilistic model presented here.

The linear model presupposes that the figure of merit can be decomposed as a sum of statistically uncorrelated terms. For a single modality there are 5 terms. The first term is the mean value of the figure of merit and is a constant. The remaining 4 terms, the reader term, the case term, the reader-case term, and the internal noise, are random variables. The reader term is a function of the reader sample only. The case term is a function of the case sample only. The reader-case term is a function of both samples. Finally, the internal noise term accounts for all other sources of variability not accounted for in the previous 3 terms.

The conventional assumption for the linear model is that the random terms in the linear decomposition are mutually independent and normally distributed [1]. As with any model-based decomposition, this assumption cannot be verified directly. In particular, a normality assumption cannot be valid if the figure of merit is the area under the ROC curve since this quantity must be between 0 and 1.

In this paper, we present a probabilistic formulation of the MRMC problem. We account for

case variability, reader variability, and reader uncertainty. We then use the methods and concepts of doubly- and triply-stochastic variables to directly derive an exact seven-term decomposition of the variance of the Wilcoxon statistic [4, 5] as a function of the numbers of readers and cases. Our results are an extension of others who have studied the statistical properties of the Wilcoxon or Mann-Whitney statistics [6–10]. This paper expands upon results first presented in [11]. The probabilistic model introduced in that paper has already been used by B. Gallas [12] to develop a “one-shot” estimate of the components of variance for the Wilcoxon statistic with multiple readers and multiple cases. Here, we provide details of the theoretical foundations and subsequent derivations for the components of variance for the Wilcoxon statistic. We also derive constraints on the MRMC coefficients that result from the theoretical model.

In the probabilistic development [11], there is no need to define intermediate and unobservable random variables. The probabilistic assumptions that go into our model are derived from the physics and intuition of the problem as opposed to the independence assumptions used for the conventional linear model to make the problem tractable. The probabilistic approach also allows us to derive constraints on the coefficients in the seven-term expansion of the variance which cannot be derived from a linear model. Indeed the normality assumption used in the conventional linear model is inconsistent with the statistical properties used to derive these constraints.

Nevertheless, we show that we may rigorously define a decomposition of the figure of merit in terms of uncorrelated, but not necessarily independent or normal, random variables that correspond to the terms in the standard linear model. The variances of these random variables can be identified with terms, or combinations of terms, in the seven-term expansion. Finally, we show that the seven-term expansion turns into a ten-term expansion when replication of the entire study is considered.

2. METHODS

MRMC methodology accounts for multiple readers each reading multiple cases. In general, we will assume that a reader analyzes an image (case) and produces a test statistic that signifies the reader's confidence that the image is abnormal. We do not assume that a given reader will produce the same value for the test statistic on multiple readings of the same image. This is due to the internal noise or reader jitter inherent in the diagnostic process. Thus, the fundamental random quantities in the MRMC problem are the case sample, the reader sample, and the resulting array of test statistics.

A. Cases, readers and test statistics

The image matrix \mathbf{G} (the cases) is composed of column vectors each representing an image. We subdivide this matrix into submatrices of signal-absent cases (i.e., normal cases), \mathbf{G}_0 , and signal-present cases (i.e., abnormal cases), \mathbf{G}_1 .

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 \end{bmatrix}. \quad (1)$$

The matrix \mathbf{G}_0 is $M \times N_0$ and \mathbf{G}_1 is $M \times N_1$, where M is the number of pixels in an image, N_0 is the number of signal-absent cases, and N_1 is the number of signal-present images. The full image matrix \mathbf{G} is $M \times N$ with $N = N_0 + N_1$. The submatrices are decomposed into the individual case vectors as follows:

$$\mathbf{G}_0 = \begin{bmatrix} \mathbf{g}_{01} & \mathbf{g}_{02} & \cdots & \mathbf{g}_{0N_0} \end{bmatrix} \quad (2)$$

$$\mathbf{G}_1 = \begin{bmatrix} \mathbf{g}_{11} & \mathbf{g}_{12} & \cdots & \mathbf{g}_{1N_1} \end{bmatrix}. \quad (3)$$

The \mathbf{g}_{0i} and \mathbf{g}_{1j} are column vectors of image data. For digital imaging, these column vectors are finite dimensional, although this assumption is not required for our probabilistic development.

The reader parameters are also formed into column vectors γ_r , one for each of N_R readers, and then collected into the reader matrix $\mathbf{\Gamma}$:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_{N_R} \end{bmatrix}. \quad (4)$$

This is a $K \times N_R$ matrix, where N_R is the number of readers, and K is the dimension of a reader parameter vector. The reader vectors γ_r may be mathematical constructs, such as a template for a linear model observer, or they may simply be a strings of numbers that are used to identify readers in an observer study. In fact, the γ_r do not even have to be numerical vectors; they could, for example, be the names of the radiologists in an observer study.

A reader produces a test statistic for each image. For a given case and reader this test statistic is a random variable due to internal noise. The test statistics for all of the readers and cases are collected into a matrix \mathbf{T} . This matrix is subdivided into submatrices corresponding to signal-absent cases, \mathbf{T}_0 , and signal-present cases, \mathbf{T}_1 :

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_0 & \mathbf{T}_1 \end{bmatrix}. \quad (5)$$

This is an $N_R \times N$ matrix of the noisy values of the reader test statistics. Rows of the two submatrices correspond to individual readers and give a reader's test statistics for all of the signal-absent images and all of the signal-present images, respectively:

$$\mathbf{T}_0 = \begin{bmatrix} t_{01} \\ t_{02} \\ \vdots \\ t_{0N_R} \end{bmatrix} \quad \text{and} \quad \mathbf{T}_1 = \begin{bmatrix} t_{11} \\ t_{12} \\ \vdots \\ t_{1N_R} \end{bmatrix}. \quad (6)$$

We can also concatenate these row vectors to make a vector of all test statistics for a given reader:

$$\mathbf{t}_r = \begin{bmatrix} t_{0r} & t_{1r} \end{bmatrix}. \quad (7)$$

B. Statistical assumptions

We make some statistical assumptions at this point. The cases are assumed to be drawn independently from signal-absent and signal-present distributions. The reader parameter vectors are assumed to be drawn independently from a distribution of such vectors. The readers are also assumed to be independent of the cases. Finally, the joint conditional density for the noisy test statistics is a product of conditional densities for the individual reader test statistics. Furthermore, this latter distribution depends only on the given reader and the cases. These assumptions can be summarized as follows:

$$pr(\mathbf{\Gamma}, \mathbf{G}) = pr_{\mathbf{\Gamma}}(\mathbf{\Gamma}) pr_{\mathbf{G}}(\mathbf{G}) \quad (8)$$

$$pr_{\mathbf{\Gamma}}(\mathbf{\Gamma}) = \prod_{r=1}^{N_R} pr_{\gamma}(\gamma_r) \quad (9)$$

$$pr_{\mathbf{G}}(\mathbf{G}) = \prod_{i=1}^{N_0} pr_0(\mathbf{g}_{0i}) \prod_{j=1}^{N_1} pr_1(\mathbf{g}_{1j}) \quad (10)$$

$$pr(\mathbf{T}|\mathbf{\Gamma}, \mathbf{G}) = \prod_{r=1}^{N_R} pr_t(\mathbf{t}_r|\gamma_r, \mathbf{G}). \quad (11)$$

The fact that the readers are independent from the cases does not imply that there is no reader-case interaction. In fact, the reader-case interaction is embodied in the distribution $pr_t(\mathbf{t}_r|\gamma_r, \mathbf{G})$ which we discuss in more detail below. This independence assumption simply implies that the selection of the reader sample is not dependent on the selection of the case sample.

C. A note on notation

If x is a random variable with conditional PDF $pr_x(x|y, z)$, conditioned on the random variables y and z , then

$$\langle f(x) \rangle_{x,z|y} = \int \int f(x) pr_x(x|y, z) pr_z(z|y) dx dz \quad (12)$$

stands for the conditional expectation of $f(x)$ conditioned on y . In this expression we are averaging over the distribution of x given (z, y) , and then averaging over the distribution of z given y . To perform this operation we need the conditional densities $pr_x(x|y, z)$ and $pr_z(z|y)$. The end result is a function of y . It appears that Eqn. 12 could be reduced to a single integral,

$$\langle f(x) \rangle_{x,z|y} = \int f(x) pr_x(x|y) dx. \quad (13)$$

However, from an operational point of view, we do not know $pr_x(x|y)$ whereas with the probabilistic assumptions above, we can calculate or approximate the integrals in 12. When z and y are independent, as will often be the case, Eqn. 12 reduces to

$$\langle f(x) \rangle_{x,z|y} = \int \int f(x) pr_x(x|y, z) pr_z(z) dx dz. \quad (14)$$

Note that Eqn. 12 includes the case where x is a deterministic function $x(y, z)$ of y and z , in which case

$$\langle f(x) \rangle_{z|y} = \int f(x(y, z)) pr_z(z|y) dz. \quad (15)$$

D. Figure of merit

Initially we will assume that the figure of merit has the following form

$$\hat{A}(\mathbf{T}) = \frac{1}{N_R} \sum_{r=1}^{N_R} \hat{a}(\mathbf{t}_r) \quad (16)$$

where $\hat{a}(\mathbf{t})$ is some figure of merit for an individual reader. Later we will be more specific about this function.

E. Mean and Variance

As an example of the probabilistic method and the notation introduced above, we compute the mean and variance of the figure of merit shown in Eqn. 16. From the independence assumptions

on the readers the mean of the figure of merit can be written as

$$\langle \widehat{A} \rangle = \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t|\gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}}. \quad (17)$$

The inner angle bracket averages over internal noise with the reader and case sample fixed. The outer angle bracket is then the average of this quantity over readers and case samples.

For the expectation of the square of \widehat{A} we have a double sum, which we decompose into a single sum where the indices match, and a double sum where the indices do not match (see Appendix). The end result is

$$\langle \widehat{A}^2 \rangle = \frac{1}{N_R} \langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{t|\gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}} + \left(1 - \frac{1}{N_R}\right) \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t, \gamma | \mathbf{G}}^2 \rangle_{\mathbf{G}}. \quad (18)$$

Putting the results we have so far together we get an expression for the variance of \widehat{A} in terms of moments of $\widehat{a}(\mathbf{t})$:

$$\begin{aligned} \text{Var}(\widehat{A}) &= \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t, \gamma | \mathbf{G}}^2 \rangle_{\mathbf{G}} - \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t|\gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}}^2 \\ &+ \frac{1}{N_R} \left[\langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{t|\gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}} - \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t, \gamma | \mathbf{G}}^2 \rangle_{\mathbf{G}} \right]. \end{aligned} \quad (19)$$

The three moments we need to calculate in order to proceed further are

$$\#1 : \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t|\gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}}^2 \quad (20)$$

$$\#2 : \langle \langle \widehat{a}(\mathbf{t}) \rangle_{t, \gamma | \mathbf{G}}^2 \rangle_{\mathbf{G}} \quad (21)$$

$$\#3 : \langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{t|\gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}}. \quad (22)$$

Equation 19 is an exact expression of the variance of the overall figure of merit in terms of expectations of the single-reader figure of merit. In order to compute these moments we need to specify our single-reader figure of merit $\widehat{a}(\mathbf{t})$. In the next section we will compute these three moments when $\widehat{a}(\mathbf{t})$ is the Wilcoxon statistic.

F. The Wilcoxon statistic

Suppose reader γ produces test statistics \mathbf{t} given by,

$$\mathbf{t} = \begin{bmatrix} \mathbf{t}_0 & \mathbf{t}_1 \end{bmatrix} \quad (23)$$

with

$$\mathbf{t}_0 = \begin{bmatrix} t_{01} & t_{02} & \dots & t_{0N_0} \end{bmatrix} \quad (24)$$

$$\mathbf{t}_1 = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1N_1} \end{bmatrix}. \quad (25)$$

The Wilcoxon statistic $\hat{a}(\mathbf{t})$ as a function of \mathbf{t} is given by

$$\hat{a}(\mathbf{t}) = \frac{1}{N_0 N_1} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} s(t_{1j} - t_{0i}). \quad (26)$$

In this equation $s(t)$ is the step function, although that fact will not play a role in most of the calculations.

We will also use one more statistical assumption

$$pr_t(\mathbf{t}|\gamma, \mathbf{G}) = \prod_{i=1}^{N_0} pr_t(t_{0i}|\gamma, \mathbf{g}_{0i}) \prod_{j=1}^{N_1} pr_t(t_{1j}|\gamma, \mathbf{g}_{1j}) \quad (27)$$

This equation tells us that, conditional on the reader and cases, the components of \mathbf{t} are independent. It also tells us that the conditional distribution for the internal noise on an individual test statistic only depends on the reader parameter vector and the corresponding case. If, for example, the internal noise is Gaussian, then the mean and variance of the test statistic for a given reader will depend only on the case at hand and the reader parameter vector.

3. RESULTS

A. The seven-term expansion

We will show that the statistical assumptions provided above imply that the variance of the Wilcoxon statistic can be expanded as

$$\text{Var} \left[\widehat{A} \right] = \frac{\alpha_1}{N_0} + \frac{\alpha_2}{N_1} + \frac{\alpha_3}{N_0 N_1} + \frac{\alpha_4}{N_R} + \frac{\alpha_5}{N_R N_0} + \frac{\alpha_6}{N_R N_1} + \frac{\alpha_7}{N_R N_0 N_1}. \quad (28)$$

We will call this the seven-term expansion for the variance of \widehat{A} and find explicit expressions for the coefficients α_n . These expressions will, in turn, lead to constraints on these coefficients. For any given set of values for N_R , N_0 and N_1 , these constraints can be used to provide an upper bound for $\text{Var} \left[\widehat{A} \right]$.

The three moments shown in Eqns.20–22 are all that we need to derive Eqn. 28.

1. Moment #1

For the first moment (Eqn. 20) we have

$$\langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} = \frac{1}{N_0 N_1} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \langle s(t_{1j} - t_{0i}) \rangle_{t_{0i}, t_{1j} | \gamma, \mathbf{g}_{0i}, \mathbf{g}_{1j}} \quad (29)$$

$$= \frac{1}{N_0 N_1} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \bar{s}(\gamma, \mathbf{g}_{0i}, \mathbf{g}_{1j}). \quad (30)$$

The last equality introduces $\bar{s}(\gamma, \mathbf{g}_{0i}, \mathbf{g}_{1j})$, the average performance of reader γ on cases \mathbf{g}_{0i} and \mathbf{g}_{1j} . The average is over the internal noise for this reader. Since the cases are independent and identically distributed random vectors, and are independent from the readers, we have the first of the three moments

$$\left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}}^2 = \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1}^2 = \langle \bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1}^2 = \mu^2. \quad (31)$$

The penultimate equality here introduces $\bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1)$, which is $\bar{s}(\boldsymbol{\gamma}, \mathbf{g}_0, \mathbf{g}_1)$ averaged over readers $\boldsymbol{\gamma}$. Finally, it is notationally convenient to define μ as the overall mean of $\hat{a}(\mathbf{t})$. This will facilitate the comparison with the more standard approach to the MRMC problem below.

2. Moment #2

For the second of the three moments we average over cases after squaring. This gives

$$\left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \boldsymbol{\gamma} | \mathbf{G}}^2 \right\rangle_{\mathbf{G}} = \frac{1}{N_0^2 N_1^2} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \sum_{k=1}^{N_0} \sum_{l=1}^{N_1} \left\langle \bar{s}(\mathbf{g}_{0i}, \mathbf{g}_{1j}) \bar{s}(\mathbf{g}_{0k}, \mathbf{g}_{1l}) \right\rangle_{\mathbf{G}}. \quad (32)$$

This sum involves averaging over observers before multiplying and averaging over cases. By separating the sum into the cases where both indices match, one index matches, and no indices match we get four terms

$$\begin{aligned} \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \boldsymbol{\gamma} | \mathbf{G}}^2 \right\rangle_{\mathbf{G}} &= \frac{1}{N_0 N_1} \left\langle \bar{\bar{s}}^2(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1} \\ &+ \frac{1}{N_0} \left(1 - \frac{1}{N_1} \right) \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}_0, \mathbf{g}'_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \\ &+ \frac{1}{N_1} \left(1 - \frac{1}{N_0} \right) \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}'_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} \\ &+ \left(1 - \frac{1}{N_0} - \frac{1}{N_1} + \frac{1}{N_0 N_1} \right) \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1}^2 \end{aligned} \quad (33)$$

We are now in a position to compute the first part of the overall variance (Eqn. 19), which is the variance of the noise-and-reader-averaged figure of merit with respect to the case randomness.

The result is three terms

$$\left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \boldsymbol{\gamma} | \mathbf{G}}^2 \right\rangle_{\mathbf{G}} - \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t} | \boldsymbol{\gamma}, \mathbf{G}} \right\rangle_{\boldsymbol{\gamma}, \mathbf{G}}^2 = \frac{\alpha_1}{N_0} + \frac{\alpha_2}{N_1} + \frac{\alpha_3}{N_0 N_1} \quad (34)$$

with the coefficients given by

$$\alpha_1 = \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}_0, \mathbf{g}'_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} - \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1}^2 \quad (35)$$

$$\alpha_2 = \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}'_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} - \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1}^2 \quad (36)$$

and

$$\begin{aligned} \alpha_3 = & \langle \bar{s}^2(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1} - \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}_0, \mathbf{g}'_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \\ & - \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}'_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} + \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1}^2. \end{aligned} \quad (37)$$

These equations are very similar to those in Hoeffding [6] and Lehmann [10]. By using independence of cases, we may simplify these expressions. The results are,

$$\alpha_1 = \text{Var} \left[\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_1 | \mathbf{g}_0} \right] \quad (38)$$

$$\alpha_2 = \text{Var} \left[\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0 | \mathbf{g}_1} \right] \quad (39)$$

$$\alpha_3 = \text{Var} \left[\bar{s}(\mathbf{g}_0, \mathbf{g}_1) \right] - \alpha_1 - \alpha_2. \quad (40)$$

In the α_1 expression the quantity inside the square brackets is a random variable since \mathbf{g}_1 has been averaged over but \mathbf{g}_0 has not. The coefficient α_1 is then the variance of this random variable.

Similar remarks apply to α_2 .

3. Moment #3

For the third moment in our list we square before doing any averaging. This leads to a fourfold sum

$$\left\langle \left\langle \widehat{a}^2(\mathbf{t}) \right\rangle_{\mathbf{t} | \gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}} = \frac{1}{N_0^2 N_1^2} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \sum_{k=1}^{N_0} \sum_{l=1}^{N_1} \left\langle \left\langle s(t_{1j} - t_{0i}) s(t_{1l} - t_{0k}) \right\rangle_{\mathbf{t} | \gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}}. \quad (41)$$

As before we can break this down into four sums depending on which indices match, and use our independence assumptions to reduce this expectation to four terms:

$$\begin{aligned}
\left\langle \left\langle \widehat{a}^2(\mathbf{t}) \right\rangle_{t|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}} &= \frac{1}{N_0 N_1} \left\langle \left\langle s^2(t_1 - t_0) \right\rangle_{t_0, t_1 | \gamma, \mathbf{g}_0, \mathbf{g}_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1} \\
&+ \frac{1}{N_0} \left(1 - \frac{1}{N_1} \right) \left\langle \left\langle s(t_1 - t_0) s(t'_1 - t_0) \right\rangle_{t_0, t_1, t'_1 | \gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \\
&+ \frac{1}{N_1} \left(1 - \frac{1}{N_0} \right) \left\langle \left\langle s(t_1 - t_0) s(t_1 - t'_0) \right\rangle_{t_0, t'_0, t_1 | \gamma, \mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} \\
&+ \left(1 - \frac{1}{N_0} - \frac{1}{N_1} + \frac{1}{N_0 N_1} \right) \left\langle \left\langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma}^2 \right\rangle_{\gamma}. \tag{42}
\end{aligned}$$

The last term may require some explanation which is provided in the Appendix. If we use the fact that $s^2(t) = s(t)$, then the first term reduces to

$$\left\langle \left\langle s^2(t_1 - t_0) \right\rangle_{t_0, t_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1 | \gamma, \mathbf{g}_0, \mathbf{g}_1} = \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1} = \langle \bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1} = \mu. \tag{43}$$

We are now ready to compute the second part of the overall variance (Eqn. 19). Combining the expressions we just derived with earlier ones we have

$$\frac{1}{N_R} \left[\left\langle \left\langle \widehat{a}^2(\mathbf{t}) \right\rangle_{t|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}} - \left\langle \left\langle \widehat{a}(\mathbf{t}) \right\rangle_{t|\gamma, \mathbf{G}}^2 \right\rangle_{\mathbf{G}} \right] = \frac{\alpha_4}{N_R} + \frac{\alpha_5}{N_R N_0} + \frac{\alpha_6}{N_R N_1} + \frac{\alpha_7}{N_R N_0 N_1}, \tag{44}$$

with

$$\alpha_4 = \left\langle \left\langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma}^2 \right\rangle_{\gamma} - \langle \bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1}^2 \tag{45}$$

$$\begin{aligned}
\alpha_5 &= \left\langle \left\langle s(t_1 - t_0) s(t'_1 - t_0) \right\rangle_{t_0, t_1, t'_1 | \gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \\
&\quad - \langle \bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1) \bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}'_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} - \alpha_4 \tag{46}
\end{aligned}$$

$$\begin{aligned}
\alpha_6 &= \left\langle \left\langle s(t_1 - t_0) s(t_1 - t'_0) \right\rangle_{t_0, t'_0, t_1 | \gamma, \mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} \\
&\quad - \langle \bar{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1) \bar{\bar{s}}(\mathbf{g}'_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} - \alpha_4 \tag{47}
\end{aligned}$$

and

$$\alpha_7 = \langle \bar{s}(\boldsymbol{\gamma}, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\boldsymbol{\gamma}, \mathbf{g}_0, \mathbf{g}_1} - \left\langle \frac{\bar{s}^2}{\bar{s}}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1} - \alpha_4 - \alpha_5 - \alpha_6. \quad (48)$$

The first two terms in the expression for α_5 are the average of a conditional variance of a random variable. A similar simplification is possible for α_6 and α_7 . The end results are alternate expressions for these coefficients (See Appendix),

$$\alpha_4 = \text{Var} \left[\langle \bar{s}(\boldsymbol{\gamma}, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1 | \boldsymbol{\gamma}} \right] \quad (49)$$

$$\alpha_5 = \left\langle \left\langle \langle s(t_1 - t_0) \rangle_{t_1, \mathbf{g}_1 | t_0, \boldsymbol{\gamma}, \mathbf{g}_0}^2 \right\rangle_{t_0, \boldsymbol{\gamma} | \mathbf{g}_0} - \left\langle \langle s(t_1 - t_0) \rangle_{t_1, \mathbf{g}_1 | t_0, \boldsymbol{\gamma}, \mathbf{g}_0} \right\rangle_{t_0, \boldsymbol{\gamma} | \mathbf{g}_0}^2 \right\rangle_{\mathbf{g}_0} - \alpha_4 \quad (50)$$

$$\alpha_6 = \left\langle \left\langle \langle s(t_1 - t_0) \rangle_{t_0, \mathbf{g}_0 | t_1, \boldsymbol{\gamma}, \mathbf{g}_1}^2 \right\rangle_{t_1, \boldsymbol{\gamma} | \mathbf{g}_1} - \left\langle \langle s(t_1 - t_0) \rangle_{t_0, \mathbf{g}_0 | t_1, \boldsymbol{\gamma}, \mathbf{g}_1} \right\rangle_{t_1, \boldsymbol{\gamma} | \mathbf{g}_1}^2 \right\rangle_{\mathbf{g}_1} - \alpha_4 \quad (51)$$

$$\alpha_7 = \left\langle \left\langle \langle s^2(t_1 - t_0) \rangle_{t_0, t_1, \boldsymbol{\gamma} | \mathbf{g}_0, \mathbf{g}_1} - \langle s(t_1 - t_0) \rangle_{t_0, t_1, \boldsymbol{\gamma} | \mathbf{g}_0, \mathbf{g}_1}^2 \right\rangle_{\mathbf{g}_0, \mathbf{g}_1} \right\rangle_{\mathbf{g}_0, \mathbf{g}_1} - \alpha_4 - \alpha_5 - \alpha_6. \quad (52)$$

The quantity in the outer angle brackets in Eqn. 50 is the variance of the step function averaged over internal noise and cases for the signal-present class. The random variables involved in computing this variance are the internal noise for a signal-absent case and readers. This variance is then averaged over signal-absent cases. A similar description can be applied to the bracketed term in α_6 and α_7 .

B. Bounds

To gain more insight into the significance of α_1 , α_2 , and α_3 , we expand $\bar{s}(\mathbf{g}_0, \mathbf{g}_1)$ as

$$\bar{s}(\mathbf{g}_0, \mathbf{g}_1) = \mu + s_0(\mathbf{g}_0) + s_1(\mathbf{g}_1) + \varepsilon(\mathbf{g}_0, \mathbf{g}_1) \quad (53)$$

where

$$s_0(\mathbf{g}_0) = \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_1 | \mathbf{g}_0} - \mu \quad (54)$$

$$s_1(\mathbf{g}_1) = \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0 | \mathbf{g}_1} - \mu. \quad (55)$$

Thus, $s_0(\mathbf{g}_0)$ is $s(t_1 - t_0)$ averaged over internal noise, readers and signal-present cases when the signal-absent case is \mathbf{g}_0 . Similarly, $s_1(\mathbf{g}_1)$ is $s(t_1 - t_0)$ averaged over internal noise, readers and signal-absent cases when the signal-present case is \mathbf{g}_1 . The random variable $\varepsilon(\mathbf{g}_0, \mathbf{g}_1)$ is defined by Eqn. 53. It is straightforward to verify that the following expectations and conditional expectations vanish

$$\langle s_0(\mathbf{g}_0) \rangle_{\mathbf{g}_0} = 0 \quad (56)$$

$$\langle s_1(\mathbf{g}_1) \rangle_{\mathbf{g}_1} = 0 \quad (57)$$

$$\langle \varepsilon(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0 | \mathbf{g}_1} = 0 \quad (58)$$

$$\langle \varepsilon(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_1 | \mathbf{g}_0} = 0. \quad (59)$$

These equations, combined with the fact that \mathbf{g}_0 and \mathbf{g}_1 are independent, imply that $s_0(\mathbf{g}_0)$, $s_1(\mathbf{g}_1)$ and $\varepsilon(\mathbf{g}_0, \mathbf{g}_1)$ are uncorrelated random variables. This then gives us the expansion

$$\text{Var} [\bar{s}(\mathbf{g}_0, \mathbf{g}_1)] = \text{Var} [s_0(\mathbf{g}_0)] + \text{Var} [s_1(\mathbf{g}_1)] + \text{Var} [\varepsilon(\mathbf{g}_0, \mathbf{g}_1)]. \quad (60)$$

From this expansion, Eqns. 38–40, and the definitions above we can identify the coefficients α_1 , α_2 and α_3 .

$$\alpha_1 = \text{Var} [s_0(\mathbf{g}_0)] \quad (61)$$

$$\alpha_2 = \text{Var} [s_1(\mathbf{g}_1)] \quad (62)$$

$$\alpha_3 = \text{Var} [\varepsilon(\mathbf{g}_0, \mathbf{g}_1)]. \quad (63)$$

A random variable that is constrained to be between 0 and 1 has a maximum variance of $1/4$. This fact and Eqns. 60–63 above lead to the constraints

$$\alpha_1 \geq 0 \quad (64)$$

$$\alpha_2 \geq 0 \quad (65)$$

$$\alpha_3 \geq 0 \quad (66)$$

$$\alpha_1 + \alpha_2 + \alpha_3 \leq \frac{1}{4}. \quad (67)$$

These constraints define a bounded region in the space of points $(\alpha_1, \alpha_2, \alpha_3)$ and thus allow us to compute, for any given values of N_0 and N_1 , the maximum possible contribution to the variance of \hat{A} from the first three terms in the seven-term expansion.

$$\frac{\alpha_1}{N_0} + \frac{\alpha_2}{N_1} + \frac{\alpha_3}{N_0 N_1} \leq \frac{1}{4 \min\{N_0, N_1\}}. \quad (68)$$

This bound represents a worst case scenario. In practice we could expect this sum to be significantly smaller than the upper bound.

Equations 49–52 lead to the following constraints

$$0 \leq \alpha_4 \leq \frac{1}{4} \quad (69)$$

$$0 \leq \alpha_4 + \alpha_5 \leq \frac{1}{4} \quad (70)$$

$$0 \leq \alpha_4 + \alpha_6 \leq \frac{1}{4} \quad (71)$$

$$0 \leq \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 \leq \frac{1}{4}. \quad (72)$$

These constraints define a bounded region in the space of points $(\alpha_4, \alpha_5, \alpha_6, \alpha_7)$. This allows us to compute, for any given N_R , N_0 and N_1 , the maximum contribution of the last four terms in the seven-term expansion to the variance of \hat{A} , i.e.,

$$\frac{\alpha_4}{N_R} + \frac{\alpha_5}{N_R N_0} + \frac{\alpha_6}{N_R N_1} + \frac{\alpha_7}{N_R N_0 N_1} \leq \frac{1}{4 N_R}. \quad (73)$$

Again we could expect this sum to be significantly smaller in practice. However, we can now write an upper bound for the variance of \widehat{A} ,

$$\text{Var} \left[\widehat{A} \right] \leq \frac{1}{4 \min \{N_0, N_1\}} + \frac{1}{4N_R}. \quad (74)$$

This could be useful in simulations where the numbers of cases and readers are easy to change and the computations of the α_n would be tedious.

C. The Moments Needed

To compute the α_n in the full expansion for the variance of \widehat{A} (Eqn. 28), we need four moments at the reader-averaged level

$$\left\langle \overline{\overline{s}}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1} \quad (75)$$

$$\left\langle \overline{\overline{s^2}}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1} \quad (76)$$

$$\left\langle \left\langle \overline{\overline{s}}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_1 | \mathbf{g}_0}^2 \right\rangle_{\mathbf{g}_0} \quad (77)$$

$$\left\langle \left\langle \overline{\overline{s}}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0 | \mathbf{g}_1}^2 \right\rangle_{\mathbf{g}_1}, \quad (78)$$

one at the case-averaged level,

$$\left\langle \left\langle \overline{\overline{s}}(\boldsymbol{\gamma}, \mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1 | \boldsymbol{\gamma}}^2 \right\rangle_{\boldsymbol{\gamma}} \quad (79)$$

and two at the test statistic level

$$\left\langle \left\langle \overline{\overline{s}}(t_1 - t_0) \right\rangle_{t_1, \mathbf{g}_1 | t_0, \boldsymbol{\gamma}, \mathbf{g}_0}^2 \right\rangle_{t_0, \boldsymbol{\gamma}, \mathbf{g}_0} \quad (80)$$

$$\left\langle \left\langle \overline{\overline{s}}(t_1 - t_0) \right\rangle_{t_0, \mathbf{g}_0 | t_1, \boldsymbol{\gamma}, \mathbf{g}_1}^2 \right\rangle_{t_1, \boldsymbol{\gamma}, \mathbf{g}_1}. \quad (81)$$

The α_n are then linear combinations of these moments.

D. Relationship to the Conventional Linear Model

We now wish to see how the expansion given above for the variance of $\widehat{A}(\mathbf{T})$ compares to the more standard approach to MRMC that uses an expansion into uncorrelated components [1,2]. For this purpose we set

$$\widehat{A}(\mathbf{T}) = \mu + r + c + rc + \varepsilon \quad (82)$$

and define each term in this expansion in terms of averages. The first term μ is the overall mean

$$\mu = \langle \widehat{A} \rangle = \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\Gamma, \mathbf{G}} = \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}}. \quad (83)$$

The second term is the reader term

$$r = r(\Gamma) = \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\mathbf{G}} - \mu. \quad (84)$$

This random variable is a function of the reader sample Γ . The third term is the case term

$$c = c(\mathbf{G}) = \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\Gamma} - \mu. \quad (85)$$

This random variable is a function of the case sample \mathbf{G} . Since Γ and \mathbf{G} are statistically independent, the random variables r and c are also statistically independent. The fourth term is the reader/case term

$$rc = rc(\Gamma, \mathbf{G}) = \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} - r - c - \mu. \quad (86)$$

This random variable is a function of Γ and \mathbf{G} . The last term is the only one that depends on the internal noise of the readers via the matrix of test statistics \mathbf{T}

$$\varepsilon = \widehat{A}(\mathbf{T}) - r - c - rc - \mu. \quad (87)$$

We will call this the noise term. It is straightforward to show that

$$\langle \varepsilon \rangle_{\mathbf{T}|\mathbf{\Gamma},\mathbf{G}} = 0 \quad (88)$$

$$\langle rc \rangle_{\mathbf{\Gamma}|\mathbf{G}} = 0 \quad (89)$$

$$\langle rc \rangle_{\mathbf{G}|\mathbf{\Gamma}} = 0 \quad (90)$$

$$\langle r \rangle_{\mathbf{\Gamma}} = 0 \quad (91)$$

$$\langle c \rangle_{\mathbf{G}} = 0. \quad (92)$$

These equations, together with the independence of r and c , can then be used to show that r , c , rc and ε are statistically uncorrelated. This fact gives us the following expansion for the variance of the figure of merit

$$\text{Var} [\widehat{A}] = \text{Var} [r] + \text{Var} [c] + \text{Var} [rc] + \text{Var} [\varepsilon] \quad (93)$$

We will now examine each term in this expansion

1. Variance of the reader term

The reader term may be written as follows

$$r(\mathbf{\Gamma}) = \frac{1}{N_R} \sum_{r=1}^{N_R} \langle \bar{s}(\gamma_r, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma_r} - \mu. \quad (94)$$

For the second moment, which is also the variance, of this random variable we have, via the now familiar manipulations of the square of a sum,

$$\text{Var} [r] = \frac{1}{N_R} \left\langle \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\mathbf{\Gamma},\mathbf{G}} \right\rangle_{\mathbf{G}|\mathbf{\Gamma}}^2 \right\rangle_{\mathbf{\Gamma}} + \left(1 - \frac{1}{N_R}\right) \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\mathbf{\Gamma},\mathbf{G}} \right\rangle_{\mathbf{\Gamma},\mathbf{G}}^2 - \mu^2 \quad (95)$$

$$= \frac{1}{N_R} \left[\left\langle \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma}^2 \right\rangle_{\mathbf{\Gamma}} - \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{\Gamma}, \mathbf{g}_0, \mathbf{g}_1}^2 \right]. \quad (96)$$

The first equality follows from the independence of the readers, the second from the definition of μ , and the third from the definition of $\bar{s}(\boldsymbol{\gamma}, \mathbf{g}_0, \mathbf{g}_1)$. The end result is that

$$\text{Var}[r] = \frac{\alpha_4}{R}. \quad (97)$$

Thus the variance of the reader term can be identified with the fourth term in the seven-term expansion for $\text{Var}[\widehat{A}]$.

2. Variance of the case term

For the case term we can write

$$c(\mathbf{G}) = \frac{1}{N_R} \sum_{r=1}^{N_R} \left\langle \langle \widehat{a}(t_r) \rangle_{t_r | \boldsymbol{\gamma}_r, \mathbf{G}} \right\rangle_{\boldsymbol{\gamma}_r | \mathbf{G}} - \mu \quad (98)$$

$$= \left\langle \langle \widehat{a}(t) \rangle_{t | \boldsymbol{\gamma}, \mathbf{G}} \right\rangle_{\boldsymbol{\gamma} | \mathbf{G}} - \mu \quad (99)$$

$$= \frac{1}{N_0 N_1} \sum_{i=1}^{N_0} \sum_{j=1}^{N_1} \bar{s}(\mathbf{g}_{0i}, \mathbf{g}_{1j}) - \mu. \quad (100)$$

The variance is given by

$$\text{Var}[c] = \left\langle \left\langle \langle \widehat{a}(t) \rangle_{t | \boldsymbol{\gamma}, \mathbf{G}} \right\rangle_{\boldsymbol{\gamma} | \mathbf{G}}^2 \right\rangle_{\mathbf{G}} - \mu^2 \quad (101)$$

$$= \frac{\alpha_1}{N_0} + \frac{\alpha_2}{N_1} + \frac{\alpha_3}{N_0 N_1}. \quad (102)$$

In other words, the first three terms in the seven-term expansion for $\text{Var}[\widehat{A}]$ comprise the variance of the case term.

It should be noted that if $N_1 = \phi N_{\text{total}}$ and $N_0 = (1 - \phi) N_{\text{total}}$, where ϕ is the prevalence, then the variance of the case term is given by,

$$\text{Var}[c] = \frac{\alpha_1/\phi + \alpha_2/(1 - \phi)}{N_{\text{total}}} + \frac{\alpha_3/(\phi(1 - \phi))}{N_{\text{total}}^2}. \quad (103)$$

The first term in Eqn. 103 agrees with standard MRMC models [2]. The second term can contribute substantially when N_{total} is small and will become negligible for N_{total} sufficiently large.

3. Variance of the reader/case term

The reader/case term can be written as

$$rc(\mathbf{\Gamma}, \mathbf{G}) = \frac{1}{N_R} \sum_{r=1}^{N_R} \langle \hat{a}(\mathbf{t}_r) \rangle_{\mathbf{t}_r | \gamma_r, \mathbf{G}} - r(\mathbf{\Gamma}) - c(\mathbf{G}) - \mu \quad (104)$$

For the variance we use the fact that r , c and rc are uncorrelated and have zero mean values to get

$$\text{Var}[rc] = \frac{1}{N_R} \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t} | \gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} + \left(1 - \frac{1}{N_R}\right) \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \gamma | \mathbf{G}}^2 \right\rangle_{\mathbf{G}} - \langle r^2 \rangle - \langle c^2 \rangle - \mu^2. \quad (105)$$

This equation then gives us

$$\text{Var}[rc] = \frac{1}{N_R} \left[\left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t} | \gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} - \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \gamma | \mathbf{G}}^2 \right\rangle_{\mathbf{G}} - \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \mathbf{G} | \gamma}^2 \right\rangle_{\gamma} + \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t} | \gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}}^2 \right]. \quad (106)$$

A new moment appears here that does not appear in the computation of the seven-term expansion for $\text{Var}[\hat{A}]$, i.e., the first term in the square brackets. This moment is discussed further in the Appendix.

4. Variance of the noise term

The noise term is explicitly given by

$$\varepsilon(\mathbf{T}) = \hat{A}(\mathbf{T}) - rc(\mathbf{\Gamma}, \mathbf{G}) - r(\mathbf{\Gamma}) - c(\mathbf{G}) - \mu. \quad (107)$$

By rearranging the variance expansion for \hat{A} we have

$$\langle \varepsilon^2 \rangle = \left\langle \left\langle \hat{A}^2(\mathbf{T}) \right\rangle_{\mathbf{T} | \mathbf{\Gamma}, \mathbf{G}} \right\rangle_{\mathbf{\Gamma}, \mathbf{G}} - \langle rc^2 \rangle - \langle c^2 \rangle - \langle r^2 \rangle - \mu^2. \quad (108)$$

This then gives us

$$\text{Var}[\varepsilon] = \text{Var}[\hat{A}] - \text{Var}[rc] - \text{Var}[c] - \text{Var}[r] \quad (109)$$

$$= \frac{\alpha_5}{N_R N_0} + \frac{\alpha_6}{N_R N_1} + \frac{\alpha_7}{N_R N_0 N_1} - \text{Var}[rc]. \quad (110)$$

Note that it is $rc + \varepsilon$ that accounts for the last three terms in the seven-term expansion. It appears that the separation of $rc + \varepsilon$ into rc and ε is not a very useful concept at this point. Moments appear in the individual variances of rc and ε that cancel out, and therefore do not appear in the expressions for the α_n . It would therefore be somewhat wasteful to compute their variances separately. This situation changes when we consider replication.

E. Replication

Now we replicate the trial K times, with the same cases and readers, and assume that the internal reader noise is independent and identically distributed from one trial to the next (the readers are not learning anything). Then we have an average figure of merit for the K trials

$$\widehat{A}_K(\mathbf{T}) = \frac{1}{K} \sum_{k=1}^K \widehat{A}(\mathbf{T}_k). \quad (111)$$

The mean value of \widehat{A}_K is given by

$$\langle \widehat{A}_K(\mathbf{T}) \rangle = \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\Gamma, \mathbf{G}}. \quad (112)$$

For the variance we need the second moment, which can be expanded as

$$\langle \widehat{A}_K^2(\mathbf{T}) \rangle = \frac{1}{K} \left\langle \left\langle \widehat{A}^2(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\Gamma, \mathbf{G}} + \left(1 - \frac{1}{K}\right) \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}}^2 \right\rangle_{\Gamma, \mathbf{G}}. \quad (113)$$

This expansion follows from the usual independence arguments. We can now write for the variance

$$\begin{aligned} \text{Var} [\widehat{A}_K] &= \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}}^2 \right\rangle_{\Gamma, \mathbf{G}} - \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\Gamma, \mathbf{G}}^2 + \\ &\quad \frac{1}{K} \left[\left\langle \left\langle \widehat{A}^2(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}} \right\rangle_{\Gamma, \mathbf{G}} - \left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}}^2 \right\rangle_{\Gamma, \mathbf{G}} \right]. \end{aligned} \quad (114)$$

The new moment we need is

$$\left\langle \left\langle \widehat{A}(\mathbf{T}) \right\rangle_{\mathbf{T}|\Gamma, \mathbf{G}}^2 \right\rangle_{\Gamma, \mathbf{G}} = \frac{1}{N_R} \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} + \left(1 - \frac{1}{N_R}\right) \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \gamma|\mathbf{G}}^2 \right\rangle_{\mathbf{G}}. \quad (115)$$

This expansion follows from the conditional independence of the internal noise and the independence of the readers. Now we may write

$$\begin{aligned} \text{Var}(\widehat{A}_K) &= \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}}^2 \right\rangle_{\mathbf{G}} - \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}}^2 + \\ &\quad \frac{1}{N_R} \left[\left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} - \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} \right\rangle_{\mathbf{G}}^2 \right] + \\ &\quad \frac{1}{KN_R} \left[\left\langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}} - \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} \right]. \end{aligned} \quad (116)$$

The moments involved here have all been worked out above or in the Appendix. The result is a ten-term expansion which we will describe below.

We may also expand into uncorrelated components as before

$$A_K = \mu_K + c_K + r_K + rc_K + \varepsilon_K \quad (117)$$

$$= \mu + c + r + rc + \varepsilon_K. \quad (118)$$

The second line here follows from the conditional independence between trials. Now we have

$$\text{Var}(A_K) = \text{Var}(c) + \text{Var}(r) + \text{Var}(rc) + \text{Var}(\varepsilon_K) \quad (119)$$

where the first three variances of are given above, and the last variance is given by

$$\text{Var}(\varepsilon_K) = \frac{1}{KN_R} \left[\left\langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}} - \left\langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}|\gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} \right]. \quad (120)$$

The dependencies on numbers of cases, readers and trials are given by

$$\text{Var}(c) = \frac{\alpha_1}{N_0} + \frac{\alpha_2}{N_1} + \frac{\alpha_3}{N_0 N_1} \quad (121)$$

$$\text{Var}(r) = \frac{\alpha_4}{N_R} \quad (122)$$

$$\text{Var}(rc) = \frac{1}{N_R} \left[\frac{\beta_1}{N_0} + \frac{\beta_2}{N_1} + \frac{\beta_3}{N_0 N_1} \right] \quad (123)$$

$$\text{Var}(\varepsilon_K) = \frac{1}{KN_R} \left[\frac{\delta_1}{N_0} + \frac{\delta_2}{N_1} + \frac{\delta_3}{N_0 N_1} \right]. \quad (124)$$

Explicit expressions for the β_n and the δ_n can be written out using equations already provided. The main point is that we can now see that rc and ε_K are now distinguishable in terms of their variances. Of course replicating an MRMC study in this way is probably not practical, except in simulation.

4. CONCLUSIONS

We have developed a probabilistic framework for analyzing MRMC problems. We have applied this framework to the Wilcoxon statistic and derived an exact seven-term expansion for the variance of the figure of merit as a function of the numbers of readers and cases. We have used the probabilistic model to derive constraints on the coefficients in this expansion. These constraints, in turn, provide an upper bound on the variance of the Wilcoxon statistic. We introduced a linear decomposition of the figure of merit into uncorrelated random variables that are defined in terms of conditional expectations over the readers, cases, and test statistics. This linear decomposition has the same structure as the conventional MRMC decomposition. We have shown that the variances of the individual terms in the linear decomposition can be related to the terms in the seven-term expansion. Finally, we have shown that replication of the MRMC experiment results in a ten-term expansion.

In the future, we plan to validate this seven-term expansion of the variance of the Wilcoxon statistic in simulation. We will also apply this methodology to real data. We are especially interested in computing the variance of the Wilcoxon statistic for ideal, Bayesian observers which we calculate using Markov chain Monte Carlo techniques. Finally, we are working on the extension of the probabilistic model to account for multiple modalities as well as multiple readers and multiple cases.

ACKNOWLEDGEMENTS

We thank Drs. Charles Metz, Brandon Gallas and Robert Wagner for their many helpful discussions about this topic. This work was supported by NIH/NCI grant K01 CA87017 and by NIH/NIBIB grants R01 EB002146, R37 EB000803, P41 EB002035.

APPENDIX

Derivation of Equation 18

What follows is a derivation of Eqn. 18.

$$\langle \widehat{A}^2 \rangle \tag{125}$$

$$= \frac{1}{N_R^2} \left[\sum_{r=1}^{N_R} \langle \langle \widehat{a}^2(\mathbf{t}_r) \rangle_{\mathbf{t}_r | \gamma_r, \mathbf{G}} \rangle_{\gamma_r, \mathbf{G}} + \sum_{\substack{r,s=1 \\ r \neq s}}^{N_R} \langle \langle \widehat{a}(\mathbf{t}_r) \widehat{a}(\mathbf{t}_s) \rangle_{\mathbf{t}_r, \mathbf{t}_s | \gamma_r, \gamma_s, \mathbf{G}} \rangle_{\gamma_r, \gamma_s, \mathbf{G}} \right] \tag{126}$$

$$= \frac{1}{N_R^2} \left[\sum_{r=1}^{N_R} \langle \langle \widehat{a}^2(\mathbf{t}_r) \rangle_{\mathbf{t}_r | \gamma_r, \mathbf{G}} \rangle_{\gamma_r, \mathbf{G}} + \sum_{\substack{r,s=1 \\ r \neq s}}^{N_R} \langle \langle \widehat{a}(\mathbf{t}_r) \rangle_{\mathbf{t}_r | \gamma_r, \mathbf{G}} \langle \widehat{a}(\mathbf{t}_s) \rangle_{\mathbf{t}_s | \gamma_s, \mathbf{G}} \rangle_{\gamma_r, \gamma_s, \mathbf{G}} \right] \tag{127}$$

$$= \frac{1}{N_R} \langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{\mathbf{t} | \gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}} + \left(1 - \frac{1}{N_R}\right) \langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \gamma | \mathbf{G}} \langle \widehat{a}(\mathbf{t}') \rangle_{\mathbf{t}', \gamma' | \mathbf{G}} \rangle_{\mathbf{G}} \tag{128}$$

$$= \frac{1}{N_R} \langle \langle \widehat{a}^2(\mathbf{t}) \rangle_{\mathbf{t} | \gamma, \mathbf{G}} \rangle_{\gamma, \mathbf{G}} + \left(1 - \frac{1}{N_R}\right) \langle \langle \widehat{a}(\mathbf{t}) \rangle_{\mathbf{t}, \gamma | \mathbf{G}}^2 \rangle_{\mathbf{G}}. \tag{129}$$

The second equality follows from the independence of the test statistics when the readers and cases are fixed. The third equality follows from the independence of the reader parameters, and the fact that they are identically distributed.

Explanation of the Equation 42

We start with the sum over all four indices with no matched indices in Eqn. 41,

$$\begin{aligned} & \sum_{i \neq k} \sum_{j \neq l} \left\langle \langle s(t_{1j} - t_{0i}) s(t_{1l} - t_{0k}) \rangle_{t|\gamma, \mathbf{G}} \right\rangle_{\gamma, \mathbf{G}} \\ &= \sum_{i \neq k} \sum_{j \neq l} \left\langle \langle s(t_{1j} - t_{0i}) \rangle_{t_{0i}, t_{1j} | \gamma, \mathbf{g}_{0i}, \mathbf{g}_{1j}} \langle s(t_{1l} - t_{0k}) \rangle_{t_{0k}, t_{1l} | \gamma, \mathbf{g}_{0k}, \mathbf{g}_{1l}} \right\rangle_{\gamma, \mathbf{G}} \end{aligned} \quad (130)$$

$$= \sum_{i \neq k} \sum_{j \neq l} \left\langle \bar{s}(\gamma, \mathbf{g}_{0i}, \mathbf{g}_{1j}) \bar{s}(\gamma, \mathbf{g}_{0k}, \mathbf{g}_{1l}) \right\rangle_{\gamma, \mathbf{G}} \quad (131)$$

$$= \sum_{i \neq k} \sum_{j \neq l} \left\langle \langle \bar{s}(\gamma, \mathbf{g}_{0i}, \mathbf{g}_{1j}) \rangle_{\mathbf{g}_{0i}, \mathbf{g}_{1j} | \gamma} \langle \bar{s}(\gamma, \mathbf{g}_{0k}, \mathbf{g}_{1l}) \rangle_{\mathbf{g}_{0k}, \mathbf{g}_{1l} | \gamma} \right\rangle_{\gamma} \quad (132)$$

$$= \sum_{i \neq k} \sum_{j \neq l} \left\langle \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma}^2 \right\rangle_{\gamma} \quad (133)$$

$$= (N_0^2 N_1^2 - N_0 N_1^2 - N_0^2 N_1 + N_0 N_1) \left\langle \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma}^2 \right\rangle_{\gamma}. \quad (134)$$

The first equality follows from independence of the internal noise when readers and cases are fixed.

The fourth equality follows from independence of cases.

Explanation of Equation 50

The first step to derive Eqn. 50 is to rewrite the first term in Eqn. 46 as

$$\begin{aligned} & \left\langle \langle s(t_1 - t_0) s(t'_1 - t_0) \rangle_{t_0, t_1, t'_1 | \gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \right\rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \\ &= \left\langle \langle s(t_1 - t_0) \rangle_{t_1 | t_0, \gamma, \mathbf{g}_0, \mathbf{g}_1} \langle s(t'_1 - t_0) \rangle_{t'_1 | t_0, \gamma, \mathbf{g}_0, \mathbf{g}'_1} \right\rangle_{t_0, \gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \end{aligned} \quad (135)$$

$$= \left\langle \langle s(t_1 - t_0) \rangle_{t_1, \mathbf{g}_1 | t_0, \gamma, \mathbf{g}_0} \langle s(t'_1 - t_0) \rangle_{t'_1, \mathbf{g}'_1 | t_0, \gamma, \mathbf{g}_0} \right\rangle_{t_0, \gamma, \mathbf{g}_0} \quad (136)$$

$$= \left\langle \langle s(t_1 - t_0) \rangle_{t_1, \mathbf{g}_1 | t_0, \gamma, \mathbf{g}_0}^2 \right\rangle_{t_0, \gamma, \mathbf{g}_0} \quad (137)$$

where the first equality follows from conditional independence of the internal noise and the second equality from independence of the cases. The second step is to rewrite the second term in Eqn. 46

as

$$\begin{aligned} & \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \bar{s}(\mathbf{g}_0, \mathbf{g}'_1) \rangle_{\mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} \\ &= \left\langle \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_1 | \mathbf{g}_0} \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}'_1) \right\rangle_{\mathbf{g}'_1 | \mathbf{g}_0} \right\rangle_{\mathbf{g}_0} \end{aligned} \quad (138)$$

$$= \left\langle \left\langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_1 | \mathbf{g}_0}^2 \right\rangle_{\mathbf{g}_0} \quad (139)$$

where again independence of cases is used. Now we use the fact that

$$\left\langle \langle s(t_1 - t_0) \rangle_{t_1, \mathbf{g}_1 | t_0, \gamma, \mathbf{g}_0} \right\rangle_{t_0, \gamma} = \langle \bar{s}(\mathbf{g}_0, \mathbf{g}_1) \rangle_{\mathbf{g}_1 | \mathbf{g}_0} \quad (140)$$

to get the result in Eqn. 50.

The New Moment in the Variance of the Reader/Case Term

The first moment in Eqn. 106 can be expanded as

$$\begin{aligned} \left\langle \langle \hat{a}(\mathbf{t}) \rangle_{\mathbf{t} | \gamma, \mathbf{G}}^2 \right\rangle_{\gamma, \mathbf{G}} &= \frac{1}{N_0 N_1} \langle \bar{s}^2(\gamma, \mathbf{g}_0, \mathbf{g}_1) \rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1} + \\ & \frac{1}{N_0} \left(1 - \frac{1}{N_1}\right) \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}'_1) \rangle_{\gamma, \mathbf{g}_0, \mathbf{g}_1, \mathbf{g}'_1} + \\ & \frac{1}{N_1} \left(1 - \frac{1}{N_0}\right) \langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \bar{s}(\gamma, \mathbf{g}'_0, \mathbf{g}_1) \rangle_{\gamma, \mathbf{g}_0, \mathbf{g}'_0, \mathbf{g}_1} + \\ & \left(1 - \frac{1}{N_0} - \frac{1}{N_1} + \frac{1}{N_0 N_1}\right) \left\langle \left\langle \bar{s}(\gamma, \mathbf{g}_0, \mathbf{g}_1) \right\rangle_{\mathbf{g}_0, \mathbf{g}_1 | \gamma}^2 \right\rangle_{\gamma} \end{aligned} \quad (141)$$

Note that $\text{Var}[rc]$ has no term that varies as N_R^{-1} . This variance will only have terms that vary as $(N_R N_0)^{-1}$, $(N_R N_1)^{-1}$ and $(N_R N_0 N_1)^{-1}$.

References

1. D. D. Dorfman, K. S. Berbaum, and C. E. Metz, "Receiver operating characteristic rating analysis. generalization to the population of readers and patients with the jackknife method," *Investigative Radiology*, vol. 27, pp. 723–731, 1992.
2. S. V. Beiden, R. F. Wagner, and G. Campbell, "Components-of-variance models and multiple-bootstrap experiments: An alternative method for random-effects, receiver operating characteristic analysis," *Academic Radiology*, vol. 7, pp. 342–349, 2000.
3. C. A. Roe and C. E. Metz, "Variance-component modeling in the analysis of receiver operating characteristic index estimates," *Academic Radiology*, vol. 4, no. 8, pp. 587–600, 1997.
4. F. Wilcoxon, "Individual comparison of ranking methods," *Biometrics*, vol. 1, pp. 80–93, 1945.
5. H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Annals of Mathematical Statistics*, vol. 18, pp. 50–60, 1947.
6. W. Hoeffding, "A class of statistics with asymptotically normal distribution," *Annals of Mathematical Statistics*, vol. 19, pp. 293–325, 1948.
7. D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *Journal of Mathematical Psychology*, vol. 12, pp. 387–415, 1975.
8. G. E. Noether, *Elements of Nonparametric Statistics*. New York: Wiley, 1967.

9. E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, pp. 837–845, 1988.
10. E. L. Lehmann, "Consistency and unbiasedness of certain nonparametric tests," *Annals of Mathematical Statistics*, vol. 22, pp. 165–179, 1951.
11. H. H. Barrett, M. A. Kupinski, and E. Clarkson, "Probabilistic foundations of the MRMC method," in *Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment*, pp. 21–31, SPIE, 2005.
12. B. D. Gallas, "One-shot estimate of mrmc variance: Auc," *Academic Radiology*, vol. 13, pp. 353–362, 2006.