Published in final edited form as:

J Appl Stat. 2012 January 1; 39(1): 67–79. doi:10.1080/02664763.2011.578616.

Confidence intervals and bands for the binormal ROC curve revisited

Eugene Demidenko*

Dartmouth Medical School, Hanover, NH 03755, USA

Abstract

Two types of confidence intervals (CIs) and confidence bands (CBs) for the receiver operating characteristic (ROC) curve are studied: pointwise CIs and simultaneous CBs. An optimized version of the pointwise CI with the shortest width is developed. A new ellipse-envelope simultaneous CB for the ROC curve is suggested as an adaptation of the Working–Hotelling-type CB implemented in a paper by Ma and Hall (1993). Statistical simulations show that our ellipse-envelope CB covers the true ROC curve with a probability close to nominal while the coverage probability of the Ma and Hall CB is significantly smaller. Simulations also show that our CI for the area under the ROC curve is close to nominal while the coverage probability of the CI suggested by Hanley and McNail (1982) uniformly overestimates the nominal value. Two examples illustrate our simultaneous ROC bands: radiation dose estimation from time to vomiting and discrimination of breast cancer from benign abnormalities using electrical impedance measurements.

Keywords

area under ROC curve; confidence interval; confidence bands; breast cancer; time to vomiting; ROC curve; Working–Hotelling

1. Introduction

The receiver operating characteristic (ROC) curve has been known for more than 30 years and is widely used in biomedical research to visually represent the trade-off between the sensitivity and specificity of a test. There are many excellent texts and review articles on the ROC curve. We refer the reader to papers [3,16,24,26] and books [21,28,36] for general discussion. The simplest and perhaps the most common ROC curve is when the two discriminating populations have normal distributions, termed the binormal ROC curve [15,24]. The remarkably robust features of the binormal ROC curve are well documented [14,32] and stem from the fact that the ROC curve is invariant to any monotonic transformation applied to the two discriminating populations.

Confidence interval (CI) should be distinguished from confidence band (CB) in the framework of the ROC curve: CI specifies the interval of values of the sensitivity given specificity with probability λ , but CB specifies the set of points in the unit square that the *entire* ROC curve belongs to with probability λ . Two types of CIs have been extensively discussed in the literature. Early authors starting from Bamber [2] developed CI for the area under the curve (AUC), which represents the overall discrimination power of the test. Much

effort has been put into the construction of CB for the ROC curve in a nonparametric setting (see [12,18] among recent references). Several papers consider the situations when observations are subject to measurement error ([6,33] among them).

The topic of CI and CB for the binormal ROC curve is sparsely scattered through the literature. For example, in the major reference on the topic [23], the Working–Hotelling simultaneous CB is expressed in terms of $\Phi(a - bx)$, where Φ is the normal cumulative distribution function (cdf), but an additional and missing step is required to express a and b and their variances through the sample statistics. Likewise, we could not find a closed-form expression for the CI for the area under the binormal ROC curve.

The goal of the present work is to develop optimized (shortest width) CI and CB for the binormal ROC curve: (1) a CI for sensitivity, or pointwise CB, based on the delta method with the shortest width and (2) a simultaneous ellipse-envelope CB for the entire ROC curve as a modification of the Working–Hotelling approach.

The organization of the paper is as follows. In the next section, we introduce notation and give basic definitions in the framework of the binormal ROC curve. The optimized pointwise CI is discussed in Section 3. The CB as the confidence region that covers the entire ROC is discussed in Section 4. Statistical simulations for assessing the coverage probability for the CB and CI for the AUC are reported in Section 5. Section 6 offers two examples, and Section 7 concludes.

2. Basic concepts and definitions

To be specific, consider two populations of people: nondiseased (D_0) and diseased (D_1). We do not know in advance who is healthy and who is sick, but there is a test that aims to discriminate people based on the outcome of the test. For example, a high prostate-specific antigen (PSA) level or another biomarker may warn of prostate cancer. There are two types of errors: false positive (FP), or false alarm, and false negative (FN) when the disease is overlooked. There are two types of correct identification: true positive (TP), called *sensitivity*, and true negative, called *specificity*. The ROC curve represents the trade-off between TP and FP rates when the cut-off value of the test changes. In this paper, it is assumed that the distribution of the test's continuous outcome variable is normal,

$\mathcal{N}(\mu_0, \sigma_0^2)$, and the distribution in the diseased group is $\mathcal{N}(\mu_1, \sigma_1^2)$. Without loss of generality, we assume that on average the nondiseased group has a smaller value of the test that, in terms of the general population, is expressed as $\mu_0 < \mu_1$.

Sensitivity, as the TP probability, is the probability that the test gave a value greater than the cut-off (c) in the diseased population. In the case of the normal distribution, sensitivity can be expressed through the cdf Φ as

$$T(c) = \Phi\left(\frac{\mu_1 - c}{\sigma_1}\right). \quad (1)$$

The FP probability is the probability that the test gave a value greater than the cut-off (c) in the nondiseased population,

$$F(c) = \Phi\left(\frac{\mu_0 - c}{\sigma_0}\right). \quad (2)$$

The binormal ROC curve can be represented on the unit square as the curve $(F(c), T(c))$ parametrized by $c \in (-\infty, \infty)$ or explicitly through the FP probability $p = F(c)$ in a functional form as

$$R(p) = \Phi\left(\frac{\mu_1 - \mu_0 + \sigma_0 \Phi^{-1}(p)}{\sigma_1}\right), \quad p \in (0, 1). \quad (3)$$

It is straightforward to see that the ROC curve starts at $(0, 0)$, is an increasing function of p and ends at $(1, 1)$.

There is a special point on the ROC curve where the sum of the two errors, FP and FN, is minimal. Using the derivation rule of implicit functions, it is easy to show that this point on the ROC curve is where the tangent line has a 45° slope. The optimal cut-off value, c , is found from the quadratic equation

$$\frac{(\mu_0 - c)^2}{\sigma_0^2} - \ln \sigma_1^2 = \frac{(\mu_1 - c)^2}{\sigma_1^2} - \ln \sigma_0^2 \quad (4)$$

with the solution

$$c = \begin{cases} \frac{(\mu_0 \sigma_1^2 - \mu_1 \sigma_0^2) + \sqrt{(\mu_0 \sigma_1^2 - \mu_1 \sigma_0^2)^2 - (\sigma_1^2 - \sigma_0^2)(\sigma_1^2 \mu_0^2 - \sigma_0^2 \mu_1^2 + \sigma_0^2 \sigma_1^2 \ln(\sigma_0^2 / \sigma_1^2))}}{\sigma_1^2 - \sigma_0^2} & \text{if } \sigma_0^2 \neq \sigma_1^2, \\ \frac{1}{2}(\mu_0 + \mu_1) & \text{if } \sigma_0^2 = \sigma_1^2. \end{cases} \quad (5)$$

There is an interesting connection to discriminant analysis: if $\sigma_0^2 \neq \sigma_1^2$, Equation (4) is the solution to the quadratic equation and if $\sigma_0^2 = \sigma_1^2$, Equation (5) is the solution to the standard discriminant analysis that minimizes the total misclassification error [22].

In reality, we do not know the true values of the mean and variance but estimate them from the outcome sample in the nondiseased group $\{x_i^0, i=1, \dots, n_0\}$ and diseased group $\{x_i^1, i=1, \dots, n_1\}$, where n_0 and n_1 are sample sizes in control and diseased groups. It is assumed that observations are independent within the group as well as between groups. Then, we estimate the ROC curve by replacing the true unknown values for (μ_k, σ_k^2) with their unbiased sample estimates $(\bar{x}_k, s_k^2)$, $k=0, 1$.

The AUC serves as an indicator of the discrimination power of the test [2]. In the case of the binormal ROC curve, it admits a closed-form solution

$$\text{AUC} = \Phi\left(\frac{\mu_1 - \mu_0}{\sqrt{\sigma_0^2 + \sigma_1^2}}\right), \quad (6)$$

assuming that $\mu_0 < \mu_1$, which implies that $\text{AUC} > 0.5$. Referring to the PSA example, the AUC can be interpreted as the overall proportion of the prostate cancer patients with a PSA greater than that of the normal controls. Alternatively, one can say that the AUC is the average sensitivity using control-test outcomes as cut-offs.

3. Pointwise CI

Let p be fixed and we want to construct a CI for $R(p)$ specified by Equation (3) using the sample values for the mean and variance. Define the cut-off $c = \Phi^{-1}(p)$ and let

$$X_c = \frac{\bar{x}_1 - \bar{x}_0 + s_0 c}{s_1}, \quad (7)$$

where Φ^{-1} is the inverse cdf. In order to construct a CI for the ROC curve, one needs to approximate $\text{var}(X_c)$. Since the data have a normal distribution, using standard formulas, we obtain

$$\text{var}(\bar{x}_1 - \bar{x}_0) = \frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1}, \quad \text{var}(s_0^2) = \frac{2\sigma_0^4}{n_0 - 1}, \quad \text{var}(s_1^2) = \frac{2\sigma_1^4}{n_1 - 1}. \quad (8)$$

For the normally distributed data, $\bar{x}_1 - \bar{x}_0$, s_0^2 and s_1^2 are mutually independent, so from the delta method [4], similar to what can be found in [35], we obtain

$$\text{var}(X_c) = \frac{1}{\sigma_1^2} \left(\frac{\sigma_0^2}{n_0} + \frac{\sigma_1^2}{n_1} \right) + \frac{c^2 \sigma_0^2}{2\sigma_1^2 (n_0 - 1)} + \frac{(\mu_1 - \mu_0 + \sigma_0 c)^2}{2\sigma_1^2 (n_1 - 1)}.$$

Hence, with fixed c , if α is the significance level, an approximate $\lambda = 1 - \alpha$ CI for $R(p)$ is

$$\Phi(X_c \pm Z_{1-\alpha/2} \sqrt{\widehat{\text{var}}(X_c)}), \quad (9)$$

where $Z_{1-\alpha/2}$ is the $(1 - \alpha/2)$ th quantile (we use the hat over the variance to indicate that the sample estimates are used instead of the true unknown values). We expect that this CI will have the nominal coverage probability for large n_0 and n_1 .

It is curious to note that the minimum $\text{var}(X_c)$, or equivalently the minimum width of the CI, occurs at the cut-off value

$$c_{\min} = -\frac{n_0 - 1}{n_1 + n_2 - 2} \frac{\bar{x}_1 - \bar{x}_0}{s_0}.$$

That is, when the FP rate is less than 0.5.

We note that in our approach, the unbiased moment estimators for variances are used (Equation (8)), which allows us to derive the closed-form expression for the CI and its optimized version in finite samples; Hsieh and Turnbull [19] and Cai and Moskowitz [5] used an asymptotic approach with maximum-likelihood estimation. When the sample size in both samples gets large, the two approaches converge.

3.1 CI with the shortest width

When constructing the CI (9), we usually assume an equal-tail probability, which results in $-Z_{1-\alpha/2}$ and $Z_{1-\alpha/2}$. However, we are at liberty to take any quantiles so that the area under the density between these quantiles is equal to the specified confidence level, $1 - \alpha$. We select those quantiles that minimize the width of the CI for the ROC curve after Φ transformation (on the probability scale), not before as in the standard approach. This idea is realized below.

In an optimized version, when $c = \Phi^{-1}(p)$ is fixed, we represent the estimate of the ROC curve similar to Equation (9) as $\Phi(a+bz)$, where $z \sim \mathcal{N}(0, 1)$, and $a = X_c$ and $b = \sqrt{\text{var}(X_c)}$ are assumed constant. We note that under those assumptions, for any values $z_1 < 0$ and $z_2 > 0$ such that

$$\Phi(z_2) - \Phi(z_1) = \lambda, \quad (10)$$

a CI $(\Phi(a + bz_2), \Phi(a + bz_1))$ will cover the true value of the ROC curve, namely $\Phi(a)$, with probability λ . Therefore, we can choose z_1 and z_2 such that the CI has the shortest width,

$$\Phi(a + bz_2) - \Phi(a + bz_1) = \min.$$

Expressing z_2 through z_1 using Equation (10), we reduce this optimization problem to a one-variable minimization problem ($z = z_1$):

$$H(z) \stackrel{\text{def}}{=} \Phi(a + b\Phi^{-1}(\lambda + \Phi(z))) - \Phi(a + bz) \Rightarrow \min \quad (11)$$

over $-\infty < z < \Phi^{-1}(\alpha)$. Minimization of function $H(z)$ is a trivial task using modern statistical packages such as SAS or R/S-Plus (note that z_1 and z_2 depend on c). It is easy to see that the width of this CI is less than the “standard” width, $H(-Z_{1-\alpha/2}) = \Phi(a + bZ_{1-\alpha/2}) - \Phi(a - bZ_{1-\alpha/2})$.

3.2 The AUC CI

Typically, the λ -CI for the AUC curve given by formula (6) is computed as the transformation of the symmetric λ -CI of the argument with the standard error approximated using the delta method. Here, we apply the above idea of the shortest CI to the AUC with

$$X = \frac{\bar{x}_1 - \bar{x}_0}{\sqrt{s_0^2 + s_1^2}}.$$

Based on the mutual independence of four estimates by the delta method, using formulas (8), we obtain

$$\widehat{\text{var}}(X) \simeq \frac{s_0^2/n_0 + s_1^2/n_1}{s_0^2 + s_1^2} + \frac{(\bar{x}_1 - \bar{x}_0)^2}{2(s_0^2 + s_1^2)^3} \left(\frac{s_0^4}{n_0 - 1} + \frac{s_1^4}{n_1 - 1} \right).$$

Then, the CI for the AUC is the Φ transformation of the interval

$$X \pm Z_{1-\alpha/2} \sqrt{\widehat{\text{var}}(X)}. \quad (12)$$

The shortest CI reduces to a one-variable minimization problem (11). Typically, the CI for the AUC computed on the ROC scale may lead to values outside the interval $[0, 1]$, while the Φ -transformed CI remains in the desired range. We compare the coverage probability for the AUC based on Equation (12) with that suggested earlier by Hanley and McNail [17] via statistical simulations.

4. Simultaneous CBs

The pointwise CI for the ROC curve gives the CI of the sensitivity given the FN rate, p . Instead, one may seek a confidence set on the unit square that covers the *entire* ROC curve with a prespecified probability, λ . This idea leads to the concept of the simultaneous CB as a special case of the confidence region routinely used in multivariate statistics [30]. We argue that CI for the ROC curve (the simultaneous CB) seems more natural than the pointwise CI since the *curve* is of interest. Two types of simultaneous CB are considered in this section: one is based on the original idea in [34] and another is based on the novel idea of the ellipse-envelope approach.

4.1 Working–Hotelling CBs

Ma and Hall [23] adopted the idea of Working and Hotelling [34], originally intended for regression lines, to construct a CB for the ROC curve. Working and Hotelling developed a CB for a linear regression $y = a + b(x - \bar{x})$ with coaxial ellipses:

$$\frac{(a - \alpha)^2}{\sigma_a^2} + \frac{(b - \beta)^2}{\sigma_b^2} = \chi^2(2), \quad (13)$$

where $b = \sum (y_i - \bar{y})(x_i - \bar{x}) / \sum (x_i - \bar{x})^2$ and $a = \bar{y} - b\bar{x}$ are the least squares estimates of the intercept α and slope β . Ma and Hall simply applied this approach to X_c as a linear function of c defined in Equation (7) with intercept $a = (\bar{x}_1 - \bar{x}_0)/s_1$ and slope $b = s_0/s_1$. The attractiveness of this approach is that the same pointwise CB is used but with a different (greater) confidence coefficient. Specifically, instead of $Z_{1-\alpha/2}$ in Equation (9), we use $\sqrt{q_{\lambda,2}}$, where $q_{\lambda,2}$ is the λ th quantile of the chi-square distribution with two degrees of freedom. It is possible to prove that $Z_{1-\alpha/2} < \sqrt{q_{\lambda,2}}$, which implies that these CBs are wider than the pointwise version. For example, for $\lambda = 0.95$, we have $\sqrt{q_{\lambda,2}} = 2.45$, while $Z_{1-\alpha/2} = 1.96$.

4.2 The ellipse-envelope CB

In the Working–Hotelling approach, variances σ_a^2 and σ_b^2 in Equation (13) are assumed fixed and known which is not true even for the linear regression. Now we modify the Working–Hotelling approach to account for the fact that, instead of the true variances, we use $\hat{\sigma}_a^2$ and $\hat{\sigma}_b^2$. This approach is illustrated with the binormal ROC curve.

Recall that the basis of our CB for the ROC curve is the function

$$\gamma(c) = \frac{\mu - c}{\sigma}$$

since the sensitivity (1) and the FN probability (2) are functions of $\gamma(c)$. Letting

$\hat{\gamma}(c) = (\bar{x} - c)/s$ from the delta method [4], we approximate

$$\begin{aligned} \text{var}(\hat{\gamma}(c)) &\simeq \frac{1}{s^2} \text{var}(\hat{\mu}) + \frac{(\bar{\mu} - c)^2}{4s^6} \text{var}(s^2) \\ &= \frac{1}{n} + \frac{(\bar{\mu} - c)^2}{2(n-1)s^2} = \frac{1}{n} + \frac{\hat{\gamma}^2(c)}{2(n-1)}. \end{aligned}$$

As with the Working–Hotelling approach, we define the λ confidence region for the pair $(\gamma_1(c), \gamma_2(c))$ as the interior of an ellipse in terms of x and y ,

$$\frac{(x - \widehat{\gamma}_1(c))^2}{1/n_1 + \widehat{\gamma}_1^2(c)/2(n_1 - 1)} + \frac{(y - \widehat{\gamma}_2(c))^2}{1/n_2 + \widehat{\gamma}_2^2(c)/2(n_2 - 1)} = q_{\lambda, 2}. \quad (14)$$

The envelope to this system of ellipses specifies simultaneous bounds for the ROC curve.

Note that when the terms with $\widehat{\gamma}_1^2(c)$ and $\widehat{\gamma}_2^2(c)$ in the denominator of Equation (14) are ignored, we arrive at the classic Working–Hotelling approach, as implemented by Ma and Hall for the ROC curve.

To find these bounds, we differentiate the above equation with respect to c and equate it to zero. This equation generates a pair of curves that are the lower and the upper bounds for the $(\gamma_1(c), \gamma_2(c))$ curve.

As we show in the appendix, the ellipse envelope is reduced to a polynomial equation of the fourth order. The simultaneous bounds for the ROC curve are obtained after transformations (1) and (2).

5. Statistical simulations

In this section, we describe statistical simulations to assess the coverage probability for: (1) the two methods of simultaneous bands, Ma and Hall (MA) and our ellipse envelope, $\widehat{\lambda}_{MH}$ and $\widehat{\lambda}_{EE}$, respectively, and (2) the CI for the AUC as developed previously in [17] (see also [21, p. 79]) compared with that following from Equation (12), $\widehat{\lambda}_{HM}$ and $\widehat{\lambda}_{(12)}$ respectively. The results of simulations are presented in Table 1; R is used for simulations with 10,000 experiments. Five sample sizes for the two groups (n_0 and n_1) in the range of 50–1000 were used to see whether the coverage probability depends on the sample size. Two significance levels $\alpha = 0.05$ and $\alpha = 0.25$ corresponding to the nominal confidence levels $\lambda_{nom} = 0.95$ and $\lambda_{nom} = 0.75$ were used. Two scenarios for the means were used: $\mu_0 = 1, \mu_1 = 3$ (strong discrimination) and $\mu_0 = 1, \mu_1 = 2$ (weak discrimination). The standard deviations were in the range from 0.5 to 2.5. The chosen parameters yield a wide range of binormal ROC curves with the AUC from 0.634 to 0.998.

In all scenarios, the empirical coverage probability for the simultaneous CB was smaller than the nominal confidence level. However, while the EE method ($\widehat{\lambda}_{EE}$) produced the ROC band that just slightly less frequently covered the true ROC curve compared with the desired confidence level, the coverage probability for the MH method ($\widehat{\lambda}_{MH}$) was significantly less than the nominal one. The latter method had especially poor performance for the nominal coverage $\lambda_{nom} = 0.75$, the coverage probability was half in this case. Remarkably, the coverage probability for both methods remained fairly stable across the sample sizes, means and standard deviations, although for the EE method the coverage probability improved slightly when the sample size got larger due to asymptotics (this is clearly seen for the 95% CB).

In terms of the AUC, the coverage probability of our method ($\widehat{\lambda}_{(12)}$) based on the Φ transformation of the interval (12) was close to nominal, while the Hanley and McNail [17]CI ($\widehat{\lambda}_{HM}$) significantly overestimated the AUC.

6. Examples

We illustrate CIs and CBs with two examples. In the first example, the original data are skewed to right, but after the log transformation makes them look quite normal. In the second example, as follows from a Q–Q plot, the transformation is not needed.

6.1 ROC curve for time to vomiting

Nuclear terrorism is one of today's dangers. In the case of a “dirty bomb” attack, one needs to get a rapid assessment of the radiation dose received by a large number of innocent individuals. It is well documented that victims typically vomit after receiving the radiation – the larger the dose, the sooner vomiting occurs [1]. Time to vomiting is a rapid self-evaluation technique that can be used for triage in the case of mass population involvement.

Several papers investigate the relationship between time to vomiting and the radiation dose received [11,27]. Here, we focus on assessing the sensitivity and specificity of time to vomiting when discriminating individuals who received less than 2 Gy and greater than 2 Gy of radiation, a widely accepted threshold for medical triage in the case of nuclear attack (Gy is a scale of radiation dose, 1 Gy = 100 rad): people who received less than 2 Gy do not require medical attention, but those who received greater than 2 Gy should undergo a thorough evaluation and possibly medical treatment [9].

We use the data on three radiological accidents that happened in the former USSR and the USA in the period 1956–2001 to illustrate the computation of sensitivity and specificity for the vomiting test (Figure 1). The majority of the cases came from the Chernobyl accident. For this example, the 0 group constitutes individuals who received < 2 Gy ($n_0 = 24$), and the 1 group received ≥ 2 Gy ($n_1 = 84$). We want to predict who received radiation > 2 Gy knowing how long after the accident the person vomited. We make a couple of comments: first, since more radiation leads to sooner vomiting, our decision rule is based on “ $<$ ” not the “ $>$ ”, as in the previous sections. The change of the inequality sign does not change the theory, but should be kept in mind at the interpretation stage. Second, before constructing the ROC curve and its confidence bounds, it is advantageous to take $\log_{10}$ of the time to vomiting because the data are skewed to the right. Indeed, as follows from the Q–Q plot in Figure 2, the data in the two groups after the log transformation seem to follow the normal distribution, so that the binormal ROC curve is valid. For the group of individuals who received less than 2 Gy, $\hat{\mu}_0 = 0.633$ and $\hat{\sigma}_0 = 0.305$, and for individuals who received greater than 2 Gy, $\hat{\mu}_1 = 0.115$ and $\hat{\sigma}_1 = 0.455$ on the $\log_{10}$ scale. Apparently, different cut-offs in time to vomiting lead to different FP and FN errors, so the ROC curve becomes a very valuable graphical technique to compare different scenarios of handling a large mass of people.

The ROC curve with its 95% ellipse-envelope CB (shaded area) is presented in Figure 3. The bold curve shows the binormal ROC (3) with sample means and standard deviations. The cut-off time to vomiting is shown on the top axis. For example, if the decision of 2 Gy is made based on the 2 h rule, 60% of victims will be identified correctly at the price of a 10% FP rate. The area under the ROC curve is 0.828, meaning that the probability that a randomly chosen pair of individuals can be truly discriminated based on time to vomiting is 0.828.

The discriminant solution for the cut-off point minimizes the sum of FN and FP errors and corresponds to the point where the 45 degree straight line (dotted) touches the ROC curve. This solution gives $c = 0.34$ which corresponds to $10^{0.34} = 2.2$ h. We show the pointwise 95% CI with the shortest width at this point. Obviously, this interval lies within the simultaneous CB.

The shortest CI for the sensitivity given the specificity requires a univariate minimization defined in Equation (11). Assuming that the cut-off is 2 h after vomiting, which is equivalent to a FP rate of about 10%, the sensitivity lies in the interval (0.35, 0.7) with probability 95%. The standard 95% CI for the AUC is (0.736, 0.896) with $Z = -1.96$, but the 95% CI with the shortest width is (0.743, 0.901) with $z_1 = -1.828$ and $z_2 = 2.138$.

In summary, although time to vomiting is a simple measure for dose reconstruction, it is quite imprecise because it has large CBs. The ROC curve with its CB is an indispensable tool for government agencies when time to vomiting is used for radiation dose reconstruction in the case of nuclear terrorism.

6.2 Electrical impedance tomography breast cancer detection

Traditional breast imaging techniques, such as mammography and ultrasound, tend to overlook breast cancer in young women. Radiologist's interpretations of images is essentially subjective and contribute to FP rates and decreases in sensitivity of malignant abnormality detection. Electrical impedance tomography (EIT) is an alternative physical model-based imaging modality aimed at reducing erroneous diagnosis by estimating electric properties of the breast, such as conductivity and permittivity, at specific suspicious locations of the breast [25,31]. Statistical estimation methods of electrical properties become dominating techniques in breast tissue classification and probability assessment of the malignancy [7]. Fundamental to breast cancer detection are calibration *ex vivo* studies when conductivity and permittivity are measured in excised breast tissues and used for calibration for *in vivo* exams [13]. We use *ex vivo* data presented in [20] to construct a binormal ROC curve to discriminate benign mass from a cancer tumour using 10 current frequencies (Figure 4; as follows from Figure 2, the distribution in the two groups is close to normal). The stepwise line depicts the empirical ROC curve based on 159 breast tissue samples; among them 119 are benign abnormalities classified as fibroadenoma, mastopathy and mammary gland, and 40 are carcinoma tumours. The bold line depicts the binormal ROC curve. Two simultaneous CBs are shown: our ellipse-envelope band is wider than that of Ma and Hall. This can be explained with our simulation results because the coverage probability of this method is consistently less than nominal.

7. Conclusion and future work

The ROC curve is a well-established statistical concept that visually represents the trade-off between the sensitivity and specificity of a test. Although much work has been done on the construction of the CB for the nonparametric ROC curve, surprisingly little research and software is available for the simplest parametric, binormal, ROC curve. For example, we could not find a closed-form expression for the CI of the area under binormal ROC curve in the existing literature.

We have developed optimized (shortest width) CIs and CBs for the binormal ROC curve and demonstrated through simulations that coverage probabilities are much closer to the nominal level than for the existing methods. The R code along with the data on time to vomiting is available online at the author's webpage: www.dartmouth.edu/~eugened.

While the binormal ROC curve has been criticized as being restrictive, we found that in real applications the data can be successfully normalized after an appropriate monotonic transformation, such as the log transformation. The existence of the normalizing transformation for the continuous random variable with density $p(x)$ follows from the theory of ordinary differential equation (ODE). Indeed, if an increasing inverse transformation function is $g(\cdot)$ and ϕ is the density of the normal distribution, we have $\phi(x) = g'(x)p(g(x))$, so that the transformation function obeys the ODE $g'(x) = p(g(x))/\phi(x)$, a nonlinear ODE of the

first order. As follows from Picard's existence theorem [10], a normalizing monotonic transformation exists for any continuous random variable. Recall that nonparametric estimation of the ROC curve is not assumption free; for example, the choice of the width in the kernel density estimation significantly influences the ROC curve.

The performance of the considered CIs and CBs is questionable in small samples because the delta method was used for variance approximation. This suggests a direction for a future work on exact CIs and CBs that would assure the coverage probability equal, or close, to the nominal probability for all n_0 and n_1 . A hint to the construction of an exact and possibly optimal CI is an observation that the CI (C_1, C_2) for the random variable of interest given by Equation (7) can be rewritten as

$$C_1 s_1 < \widehat{\mu}_1 - \widehat{\mu}_0 + s_0 c < C_2 s_1.$$

Since $\widehat{\mu}_1 - \widehat{\mu}_0$, s_0^2 and s_1^2 are mutually independent with known distributions, there is a possibility to find $C_1 < C_2$ such that the probability of the above double inequality is at least λ and $C_2 - C_1 = \min$. The construction of CIs and CBs for the binormal ROC curve is a practically urgent and theoretically challenging task.

Acknowledgments

The author is grateful for the comments of the reviewers that improved the paper. This work was supported by grants CA130880, U19 A1067733 and U54CA151662 from National Cancer Institute. The content is solely the responsibility of the author and does not necessarily represents the official views of the National Institutes of Health.

Appendix. Derivation of the ellipse-envelope CB

Let

$$u = x - \widehat{\gamma}_0(c), \quad v = y - \widehat{\gamma}_1(c) \quad (\text{A1})$$

and

$$A_k = \left(\frac{1}{n_k} + \frac{\widehat{\gamma}_k^2(c)}{2(n_k - 1)} \right)^{-1}, \quad k=0, 1.$$

After differentiating equation (14) with respect to c and setting the derivative to zero, we obtain the equation

$$\frac{2A_0}{\sigma_0} u - \frac{2A_0^2}{\sigma_0} u^2 \widehat{\gamma}_0'(c) + \frac{2A_1}{\sigma_1} v - \frac{2A_1^2}{\sigma_1} v^2 \widehat{\gamma}_1'(c) = 0,$$

which, in combination with Equation (14), defines a couple of curves as functions of c . After some simplification, we come to a system of two equations for u and v :

$$\begin{aligned} A_0 u^2 + A_1 v^2 &= q_{1,2}, \\ B_0 u - D_0 u^2 + B_1 v - D_1 v^2 &= 0, \end{aligned}$$

where

$$B_k = \frac{A_k}{\sigma_k}, \quad D_k = \frac{\widehat{\gamma}_1(c) A_k^2}{\sigma_k (n_k - 1)}.$$

After some algebra, one can find

$$\begin{aligned} v &= \eta, \\ u &= \frac{(A_0 D_1 - D_0 A_1) \eta^2 - A_0 B_1 \eta + D_0 q_{\lambda,2}}{B_0 A_0}, \end{aligned}$$

where η is a root of a fourth-order polynomial equation:

$$p_4 \eta^4 + p_3 \eta^3 + p_2 \eta^2 + p_1 \eta + p_0 = 0$$

with coefficients defined as

$$\begin{aligned} p_4 &= A_1^2 D_0^2 - 2 D_0 D_1 A_0 A_1 + A_0^2 D_1^2, \\ p_3 &= 2 D_0 A_0 A_1 B_1 - 2 A_0^2 D_1 B_1, \\ p_2 &= A_0 A_1 B_2^2 + 2 D_0 A_1 D_1 q_{\lambda,2} + B_1^2 A_0^2 - 2 D_0^2 A_1 q_{\lambda,2}, \\ p_1 &= -2 A_0 D_0 B_1 q_{\lambda,2}, \\ p_0 &= -B_0^2 A_0 q + D_0^2 q_{\lambda,2}^2. \end{aligned}$$

This polynomial has two real roots for each c that, from Equation (A1), determine a couple of curves $x(c)$ and $y(c)$ leading to the simultaneous lower and upper confidence bounds.

References

1. Anno GH, Baum SJ, Withers HR, Young RW. Symptomatology of acute radiation effects in humans after exposure to doses of 0.5–30 Gy. *Health Phys.* 1989; 55:821–838. [PubMed: 2722506]
2. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* 1975; 12:387–415.
3. Begg CB. Advances in statistical methodology for diagnostic medicine in the 1980s. *Stat. Med.* 1991; 10:1887–1895. [PubMed: 1805316]
4. Bickel, PJ.; Doksum, KA. *Mathematical Statistics*. 2nd ed.. Prentice Hall; Upper Saddle River, NJ: 2001.
5. Cai T, Moskowitz CS. Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics.* 2004; 5:573–586. [PubMed: 15475420]
6. Coffin M, Sukhatme S. Receiver operating characteristic studies and measurement errors. *Biometrics.* 1997; 53:823–837. [PubMed: 9333348]
7. Demidenko E, Hartov A, Paulsen K. Statistical estimation of resistance/conductance by electrical impedance tomography measurements. *IEEE Trans. Med. Imaging.* 2004; 23:829–838. [PubMed: 15250635]
8. Demidenko E, Williams BB, Swartz HM. Radiation dose prediction using data on time to emesis in the case of nuclear terrorism. *Radiat. Res.* 2009; 171:310–319. [PubMed: 19267558]
9. Flynn DF, Goans RE. Nuclear terrorism: Triage and medical management of radiation and combined-injury casualties. *Surg. Clin. N. Am.* 2006; 86:601–636. [PubMed: 16781272]
10. Forsyth, AR. *Theory of Differential Equations*. Dover, New York: 1959.
11. Goans RE, Waselenko JK. Medical management of radiological casualties. *Health Phys.* 2005; 89:505–512. [PubMed: 16217194]

12. Hall P, Hyndman RJ, Fan Y. Nonparametric confidence intervals for receiver operating characteristic curves. *Biometrika*. 2004; 91:743–750.
13. Halter RJ, Zhou T, Meaney PM, Hartov A, Barth RJ, Rosenkranz KM, Wells WA, Kogel CA, Borsic A, Rizzo EJ, Paulsen KD. The correlation of in vivo and ex vivo tissue dielectric properties to validate electromagnetic breast imaging: Initial clinical experience. *Physiol. Meas.* 2009; 30:S121–S136. [PubMed: 19491436]
14. Hanley JA. The robustness of the “binormal” assumptions used in fitting ROC curves. *Med. Decis. Making*. 1988; 8:197–203. [PubMed: 3398748]
15. Hanley JA. The use of the “binormal” model for parametric ROC analysis of quantitative diagnostic tests. *Stat. Med.* 1996; 15:1575–1585. [PubMed: 8855483]
16. Hanley, JA. Receiver operating characteristic (ROC) curves. In: Armitage, P.; Colton, T., editors. *Encyclopedia of Biostatistics*. Wiley; New York: 1998.
17. Hanley JA, McNail BJ. The meaning and use of area under an ROC curve. *Radiology*. 1982; 148:839–843. [PubMed: 6878708]
18. Horváth L, Horváth Z, Zhou W. Confidence bands for ROC curves. *J. Statist. Plann. Inference*. 2008; 138:1894–1904.
19. Hsieh F, Turnbull BW. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Ann. Stat.* 1996; 24:25–40.
20. Jossinet J, Schmitt M. A review of parameters for the bioelectrical characterization of breast tissue. *Ann. NY Acad. Sci.* 1999; 873:30–41. [PubMed: 10372147]
21. Krzanowski, WJ.; Hand, DJ. *ROC Curves for Continuous Data*. Chapman and Hall/CRC Press; Boca Raton, FL: 2009.
22. Lachenbruch, PA. *Discriminant Analysis*. Hafner Press; London: 1975.
23. Ma GQ, Hall WJ. Confidence bands for receiver operating characteristic curves. *Med. Decis. Making*. 1993; 13:191–197. [PubMed: 8412547]
24. Metz CE. Basic principles of ROC analysis. *Semin. Nucl. Med.* 1978; 8:283–298. [PubMed: 112681]
25. Ng EYK, Sree SV, Ng KH, Kaw G. The use of tissue electrical characteristics for breast cancer detection: A perspective review. *Technol. Cancer Res. Treat.* 2008; 7:295–308. [PubMed: 18642968]
26. Obuchowski NA. Fundamentals of clinical research for radiologists. *Am. J. Roentgenol.* 2005; 184:364–372. [PubMed: 15671347]
27. Parker DD, Parker JC. Estimating radiation dose from time to emesis and lymphocyte depletion. *Health Phys.* 2007; 93:701–704. [PubMed: 17993851]
28. Pepe, MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; Oxford: 2003.
29. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; Vienna, Austria: ISBN 3-900051-07-0, available at URL <http://www.R-project.org>
30. Rao, CR. *Linear Statistical Inference and Its Applications*. 2nd ed.. Wiley; New York: 1973.
31. Soni NK, Hartov A, Kogel C, Poplack SP, Paulsen KD. Multi-frequency electrical impedance tomography of the breast: New clinical results. *Physiol. Meas.* 2004; 25:301–314. [PubMed: 15005324]
32. Swets JA. Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychol. Bull.* 1986; 99:181–198. [PubMed: 3515382]
33. Tosteson TD, Buonaccorsi JP, Demidenko E, Wells WA. Measurement error and confidence intervals for ROC curves. *Biom. J.* 2005; 47:409–416. [PubMed: 16161800]
34. Working H, Hotelling H. Applications of the theory of errors to the interpretation of trends. *J. Am. Statist. Soc.* 1929; 24:73–85.
35. Xiong C, van Belle G, Miller JP, Morris JC. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Stat. Med.* 2006; 25:1251–1273. [PubMed: 16345029]
36. Zhou, X-H.; Obuchowski, NA.; McClish, DK. *Statistical Methods in Diagnostic Medicine*. Wiley; New York: 2002.

Figure 1.

Time-to-vomiting data as a function of the radiation dose received from three accidents, $n = n_0 + n_1 = 108$. The 2 Gy radiation dose divides people into two groups: those who are OK and those who should go through additional evaluation. The log scale seems appropriate for this data (Figure 2).

Figure 2.
Q-Q plots for the two examples – the distributions are fairly close to normal.

Figure 3. Time-to-vomiting ROC curve as a 2 Gy discriminator with 95% CB and CI with the shortest width at the discriminant solution point (the step line is the empirical ROC curve).

Figure 4.
The binormal ROC curve for breast cancer discrimination based on the EIT measurements.

Table 1

Coverage probabilities from statistical simulations for [23] and the ellipse-envelope CB, and the CI for the AUC using Equation (12) and [17].

Strong discrimination ($\mu_0 = 1, \mu_1 = 3$)								Weak discrimination ($\mu_0 = 1, \mu_1 = 2$)							
Sample size		SD		CB ROC		CI AUC		Sample size		SD		CB ROC		CI AUC	
n_0	n_1	σ_0	σ_1	$\hat{\lambda}_{EE}$	$\hat{\lambda}_{MH}$	$\hat{\lambda}_{(12)}$	$\hat{\lambda}_{HM}$	n_0	n_1	σ_0	σ_1	$\hat{\lambda}_{EE}$	$\hat{\lambda}_{MH}$	$\hat{\lambda}_{(12)}$	$\hat{\lambda}_{HM}$
$\lambda_{nom} = 0.95$															
50	20	0.5	0.5	0.90	0.83	0.97	0.99	50	20	0.5	0.5	0.89	0.81	0.96	0.97
100	50	0.5	0.5	0.92	0.84	0.97	1.00	100	50	0.5	0.5	0.92	0.82	0.96	0.98
300	100	1.0	1.5	0.93	0.82	0.98	0.99	300	100	1.0	1.5	0.93	0.81	0.96	0.98
500	150	1.0	2.0	0.93	0.81	0.98	0.99	500	150	1.0	2.0	0.93	0.81	0.96	0.99
1000	300	1.5	2.5	0.94	0.82	0.97	0.99	1000	300	1.5	2.5	0.94	0.82	0.95	0.98
$\lambda_{nom} = 0.75$															
50	20	0.5	0.5	0.66	0.41	0.81	0.98	50	20	0.5	0.5	0.63	0.34	0.78	0.84
100	50	0.5	0.5	0.68	0.42	0.81	0.99	100	50	0.5	0.5	0.66	0.35	0.79	0.85
300	100	1.0	1.5	0.68	0.31	0.80	0.88	300	100	1.0	1.5	0.68	0.30	0.77	0.84
500	150	1.0	2.0	0.66	0.29	0.81	0.89	500	150	1.0	2.0	0.67	0.28	0.77	0.86
1000	300	1.5	2.5	0.68	0.29	0.78	0.86	1000	300	1.5	2.5	0.68	0.28	0.76	0.84

$\hat{\lambda}_{EE}$: Ellipse-envelope CB coverage probability; $\hat{\lambda}_{MH}$: Ma and Hall [23] CB coverage probability; $\hat{\lambda}_{(12)}$: CI coverage probability of AUC using Equation (12); $\hat{\lambda}_{HM}$: Hanley and McNail [17] CI coverage probability of AUC.