

THE PREDICTIVE DISTRIBUTION AND AREA UNDER THE ROC CURVE

Lyle D. Broemeling

Department of Biostatistics and Applied Mathematics
The University of Texas M D Anderson Cancer Center
1500 Holcombe Boulevard
Houston, TX 77030
lbroemel@mdanderson.org

Abstract

The predictive distribution is used to find the area under the ROC curve when the diagnostic response is discrete. Typically, there are two groups of patients, those who have the condition and those who do not, and on each patient a diagnostic ordinal response is measured. The true status of the patient is determined by a gold standard, and this information together with the diagnostic response allows one to estimate the true positive rate (sensitivity) and the false positive rate (1- specificity), for a given threshold value. Using each possible value of the diagnostic response as a threshold value, the ROC curve is determined, and the area under the curve can be estimated. The area under the curve gives the overall accuracy of the diagnostic procedure. A novel approach using the Bayesian predictive density to estimate the area is proposed. Several examples from diagnostic medicine illustrate the methodology for various multi-reader and/or multi-modality scenarios.

1. INTRODUCTION

The receiver operating characteristic (ROC) curve provides the investigator with the overall accuracy of a diagnostic procedure. See Pepe (2003, Chapter 4) and Zhou, Obuchowski, and McClish (2003, Chapter 2) for a brief history and introduction to the development and current use of the curve. Typically, an ordinal diagnostic response is measured on the two populations of patients: one with the condition, and the other without the condition. In addition, for each patient the true status (gold standard) of the condition is known from an outside source. For example, in breast cancer diagnostic studies with mammography, the response varies from 1 to 5, where 1 indicates a definite benign lesion, and 5 indicates a definite malignant lesion (see Zhou *et al.*, 2002, Chapter 4). Follow-up information from surgery and pathology provides the investigator with the true status of the lesion. Often one of the response values (*e.g.*, the response 4 of “probably malignant”) is used as a threshold, meaning that any rating at least as large as 4 indicates a diagnosis of breast cancer. Using this information along with the gold standard allows one to estimate the true positive rate (sensitivity) and the false-positive rate (1- specificity) of mammography.

If each possible value of the response is considered a threshold, the true positive rate versus the false-positive rate can be plotted and the resulting curve is referred to as the ROC curve. The area under the curve gives a measure of the overall accuracy of the procedure. There are several interpretations of this area, including the following: (1) the average value of sensitivity for all possible values of specificity, (2) the average value of specificity for all values of sensitivity, and (3) the probability that the response of a randomly selected patient with the condition is greater than the response of a randomly selected patient without the condition. It is this interpretation that will be employed for calculating the ROC area.

Figure 1 shows how to interpret the ROC curve and its area (see Pepe, 2003, page 69). The vertical axis gives the true positive rate corresponding to a false-positive rate on the horizontal axis. There are four ROC curves, where one, with an area of .5, is labeled useless, and the one with an area of 1 indicates a perfect diagnostic test. With this test a patient with the disease always has a larger diagnostic value than a patient without the disease.

Curve A has a larger true positive rate for all false-positive rates compared to curve B, and thus has a larger area. If the two curves represent different modalities (or different readers of the results), curve A would be preferred because it has a higher level of diagnostic accuracy.

There are many ways to calculate the ROC curve and its area for both continuous and discrete diagnostic data and with parametric and nonparametric assumptions. (See articles by Pepe [2003] and Zhou *et. al.* [2002] for complete accounts.) For example, assuming a binormal response, that is, where the distribution of the response is normal for both groups of patients, Swets (1986) and Hanley (1988, 1996) provided interesting descriptions. For a nonparametric approach, the Mann Whitney U-statistic estimates the area under the ROC curve. Obuchowski (1997) determined the area and accuracy for the diagnosis of coronary artery disease with this method. Her approach is based on an interpretation using the predictive distribution of future observations.

On the other hand, there are few Bayesian contributions to estimate the ROC area. Assuming the binormal approach, O'Malley, Zhou, Fielding, and Tempany (2001) fitted the ROC curve and estimated its area based on a Markov chain Monte Carlo (MCMC) algorithm to find the posterior distribution of the area. The method is quite general and can be used to incorporate patient covariate information. The investigators analyzed diagnostic tests for PSA levels in prostate cancer and for spiral computed tomography (CT) of ureteral stones. For discrete data, Peng and Hall (1996) also employed an MCMC algorithm along with a model of Tosteson and Begg (1988) for the cumulative probabilities of the ordinal response to calculate the area. The diagnostic accuracy of ultrasound to detect hepatic metastasis is determined by the area of the ROC curve. Using an MCMC approach quite similar to that of Peng and Hall (1996), Hellmich *et. al.* (1998) analyzed the same data by computing the posterior distribution of the area. Both investigations relied on a model where the ordinal responses were viewed as belonging to a partition of the range of the normally distributed latent variable. The posterior mean, median, standard deviation, and 95% credible intervals of the ROC area comprise these Bayesian analyses.

2. THE PREDICTIVE DISTRIBUTION AND ROC AREA

2.1 The Predictive Distribution

The Bayesian approach is used to estimate the ROC area based on the predictive distribution of future observations. If Y and Z are random variables indicating the diagnostic responses of the diseased and nondiseased populations, respectively, Pepe (2003, page 78) shows that the area under the ROC curve is

$$\Pr[Y \geq Z / \text{parameters}]. \quad (1)$$

Suppose the possible values of Y and Z are the consecutive integers a, b, c, d , and e . Let the probability that $Y = a$ be θ_1, \dots , and the probability that $Y = e$ be θ_5 . Also let the probability that $Z = a$ be ϕ_1, \dots , and the probability that $Z = e$ be ϕ_5 .

Then $Y \geq Z$ under the following conditions:

1. $Y = a$ & $Z = a$, with probability $\theta_1\phi_1$,
2. $Y = b$ & $Z = a$ or b with probability $\theta_2(\phi_1 + \phi_2)$, ..
3. $Y = d$ & $Z = a$ or b or c or d with probability $\theta_4(\phi_1 + \phi_2 + \phi_3 + \phi_4)$,
and
4. $Y = e$ & $Z = a$ or b or c or d or e) with probability θ_5 .

Since Y and Z refer to different groups of patients (diseased and nondiseased), Y and Z are assumed to be conditionally independent, given the parameters.

In general, the ROC area is

$$\text{Area}(\theta, \phi) = \sum_{i=1}^{i=5} \sum_{j=1}^{j=i} \theta_i \phi_j. \quad (2)$$

2.2 The Bayesian Analysis

The Bayesian solution is to find the posterior distribution of $\theta = (\theta_1, \theta_2, \dots, \theta_5)$ and $\phi = (\phi_1, \phi_2, \dots, \phi_5)$ and thus of the area.

If one uses a Bayesian approach to this problem, the likelihood function and the prior distribution of θ and ϕ must be specified. Suppose there are m patients with the disease, the likelihood function for Y is

$$L(\theta / y) \propto \prod_{i=1}^{i=5} \theta_i^{y_i}, \quad (3)$$

where y_1 is the frequency of a, \dots , and y_5 the frequency of e , $0 \leq \theta_i \leq 1$, $\sum_{i=1}^{i=5} \theta_i = 1$, and $m = \sum_{i=1}^{i=5} y_i$. In a similar way, the likelihood function for Z is

$$L(\phi / z) \propto \prod_{j=1}^{j=5} \phi_j^{z_j}, \quad (4)$$

and $\sum_{j=1}^{j=5} z_j = n$, where n is the number of the nondiseased patients.

The posterior density of θ and ϕ is the product of (3) and (4) multiplied by the prior density of θ and ϕ . Of course (3) and (4) are recognized as multinomial likelihoods; therefore a possible choice for the prior density is the product of the two Dirichlet densities,

$$g(\theta, \phi) \propto \prod_{i=1}^{i=5} \theta_i^{\alpha_i - 1} \prod_{j=1}^{j=5} \phi_j^{\beta_j - 1}. \quad (5)$$

Multiplying (3), (4), and (5) results in the posterior distribution

$$g(\theta, \phi / y, z) \propto \prod_{i=1}^{i=5} \theta_i^{y_i + \alpha_i - 1} \prod_{j=1}^{j=5} \phi_j^{z_j + \beta_j - 1}, \quad (6)$$

as the product of two independent Dirichlet densities. If little or no prior information is available, let the *alphas* and *betas* of the prior distribution (5) be 1, giving a uniform prior for the parameters. Since the posterior distribution of θ and ϕ are independent Dirichlet distributions, the posterior distribution of the ROC area is easily determined by generating samples from the posterior distribution of θ and ϕ .

The above determination of the joint posterior distribution of the parameters is easily generalized to multi-reader and/or multi-modality scenarios. The details will be given when the relevant examples are discussed.

2.3 Computing the Posterior Distribution of the Area

It is seen from (6), that θ has a posterior distribution that is Dirichlet with parameter $(y_1 + 1, y_2 + 1, \dots, y_5 + 1)$ and is independent of ϕ , which has a Dirichlet distribution with parameter $(z_1 + 1, z_2 + 1, \dots, z_5 + 1)$. The posterior distribution of the ROC area

$$A(\theta, \phi) = \sum_{i=1}^{i=5} \sum_{j=1}^{j=i} \theta_i \phi_j \quad (2)$$

is determined by sampling from the joint posterior distribution of θ and ϕ . Many software packages such as WinBugs®, S-Plus®, and Minitab® have random number generators for the Dirichlet distribution.

Suppose a “large” number M (say 10,000) samples are generated from the posterior distribution of θ and ϕ , then the area is computed via (2) for each of the M samples, providing M samples from the posterior distributions of $A(\theta, \phi)$. Based on those samples, the posterior mean, median, standard deviation, and 95% credible interval (or any other posterior characteristic) are easily computed. This approach will be illustrated with three examples, which follow.

How large is M ? One way to determine M is to choose a value large enough so that the generated posterior mean of θ_1 agrees (to, say, 2 decimal places) with its known true value. Since the posterior distribution of θ_1 is *Beta*, its mean is known, and can be compared to the value computed by resampling.

3. THE ROC AREA IN DIAGNOSTIC MEDICINE

Three examples are given that will illustrate the methodology for estimating the area under the ROC curve. The first example is based on the studies of Peng and Hall (1996) and Hellmich *et. al.* (1998), both of which estimated the area under the ROC curve for the diagnosis of breast cancer metastasis in 23 patients. The diagnosis of metastasis is made with ultrasound on a five point ordinal scale. The second example is from Masaryk *et. al.*(1991) for the diagnosis of arterial atherosclerotic stenosis using magnetic resonance angiography. The third example is taken from a melanoma study of Ekmekcioglu *et. al.* (2004), who investigated a promising new biomarker for predicting morbidity and survival.

3.1 Breast Cancer Metastasis and Ultrasound

Peng and Hall (1996) studied the use of ultrasound to detect breast cancer metastasis in 23 patients. The ultrasound ratings of the 23 patients are given below in Table 1.

Table 1. Rating Data for Detection of Metastasis

Metastasis	Rating					Total N
	1	2	3	4	5	
No	6	5	2	1	0	14
Yes	0	2	0	2	5	9

The ratings vary from 1 to 5, where 1 signifies definitely normal, 2 probably normal, 3 equivocal, 4 probably abnormal, and 5 definitely abnormal. With a uniform prior distribution for the parameters, the posterior distribution of θ is Dirichlet with parameter (1, 3, 1, 3, 6) and independent of ϕ , which is Dirichlet with parameter (7, 6, 3, 2, 1). Using Minitab®, 1000 samples were generated from the joint posterior distribution of the parameters (6) and 1000 values computed, via (2), for the ROC area.

The results were as follows.

Table 2. Posterior Distribution of ROC Area
Breast Cancer Metastasis

Investigator	Mean	Median	SD	95% Credible Interval
Broemeling	.902	.914	.052	.780, .977
Peng & Hall	.987			.927, .999
Hellmich	.903		.073	.72, .99

It should be noted that although Bayesian, the studies of Peng and Hall and Hellmich *et. al.* used different approaches from that taken here. They employed a regression model for the ordinal responses, a normally distributed latent variable, and MCMC techniques for computing the posterior distribution of the area. By contrast, we used a very simple resampling method that generated samples directly from the Dirichlet posterior distribution, and the results were quite similar to those of Hellmich *et. al.*

3.2 Magnetic Resonance and Arterial Stenosis

Masaryk *et. al.*(1991) investigated the diagnostic accuracy of magnetic resonance angiography (MRA) for the diagnosis of arterial atherosclerotic stenosis among 36 patients. Both coronary arteries were imaged, using two readers. The diagnostic marker was continuous and measured the degree of stenosis in the left coronary artery, where larger values provide greater

indication of disease. Obuchowski (1997) examined the effect of clustering (left and right arteries in each patient) within patients on the estimation of the ROC curve. For the purpose of demonstrating the proposed technique, the diagnostic marker will be discretized into two groups. The same data is examined here, but with a discrete diagnostic variable labeled as either 1 or 2, where 1 indicates values from -122 to 60 for the original measure of stenosis, and 2 indicates values from 61 to 100 . The ratings of the two readers are given in the two tables below.

Table 2a. Ratings of Two Readers for Stenosis

12 Patients With Stenosis

Reader 2 Rating

Reader1	1	2	Total
1	0	0	0
2	0	12	12
Total	0	12	12

Table 2b. Ratings of Two Readers for Stenosis

21 Patients Without Stenosis

Reader 2 Rating

Reader1	1	2	Total
1	18	0	18
2	0	3	3
21	18	3	21

In the case of two readers, the results are generalized as follows. For those with disease, let θ_{ij} be the probability that reader 1 gives a rating of i and reader 2 a rating of j to the CT image, and for those without disease, let ϕ_{ij} be the corresponding probability. Also let

$$\theta = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) \text{ and } \phi = (\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}),$$

then $\theta_{i.}$ = the probability that reader 1 gives an i rating for a diseased patient and $\phi_{.j}$ is the probability that reader 2 gives a rating of j to the image from a nondiseased patient. The dot subscript indicates summation of the θ_{ij} over the missing subscript. Since a higher rating indicates stenosis, the area under the ROC curve for reader 1 is

$$A_1(\theta, \phi) = \sum_{i=1}^{i=2} \sum_{j=1}^{j=i} \theta_{i.} \phi_{.j} \quad (7)$$

and for reader 2

$$A_2(\theta, \phi) = \sum_{i=1}^{i=2} \sum_{j=1}^{j=i} \theta_{.i} \phi_{.j} \quad (8)$$

With a uniform prior distribution for the Dirichlet parameters, the posterior distributions of the two reader areas are the nonlinear functions given by (7) and (8). To determine the posterior distribution of the two readers, 10,000 samples were generated from the joint posterior distribution of θ and ϕ . This also provided 10,000 samples from the posterior distribution of the two reader areas and gave the following results.

Table 3. Posterior Distribution of the Two Reader Areas and Difference

Parameter	Mean	Median	SD	95% Interval
Area 1	.974	.980	.020	.921, .997
Area 2	.976	.979	.019	.924, .997
Difference	.0001	.0003	.020	-.040, .041

It is seen that the two readers gave similar readings for accuracy to the MRA images for the diagnosis of stenosis, with virtually identical areas under the ROC curves. This is confirmed by the posterior distribution of the difference in the areas, where the 95% credible interval contains zero. These results also compare favorably with those of Masaryk *et. al.*, where the areas of readers 1 and 2 for the left artery were estimated as .986 and .988, respectively.

3.3 A Melanoma Study

In this study, Ekmekcioglu *et. al.* (2004) investigated the value of iNOS expression (inducible nitric acid synthase) to predict morbidity and survival in 132 patients diagnosed with stage III cutaneous melanoma from 1996 to 1997 at M. D. Anderson Cancer Center in Houston, Texas. The iNOS expression in the tumors of patients with melanoma was examined by immunohistochemistry, where the number of positive cells provided the measure of the expression. The median follow-up time from the stage III diagnosis was 49.5 months. Over the period of observation, 72 patients died and 60 remained alive. The iNOS number (= 0, 1, 2, 3) is given below for the two groups of patients; where iNOS = 0, the number of positive tumor cells is less than 5%.

Table 4. iNOS Number and Survival

Status	INOS Number				Total
	0	1	2	3	
Alive	47	8	4	1	60
Died	22	13	25	12	72
Total	69	21	29	13	132

It appears that the smaller the iNOS number, the better the survival, while larger ratings are an indication of poor survival. Using a uniform prior distribution for θ and ϕ , where the former is the Dirichlet parameter for iNOS among those that died, and has a posterior Dirichlet distribution with parameter (23, 14, 26, 13) and is independent of the latter, which is Dirichlet with parameter (48, 9, 5, 2). Based on resampling with 5,000 samples, the posterior mean area was .876, the posterior median area was .879, the posterior standard deviation was .032, and the 95% credible interval was (.804, .933). Thus the posterior probability that the iNOS number of a person who died from the disease is greater than that for one who survived is on the average .876, indicating that the iNOS number is a good predictor of survival.

4.0 COMMENTS AND CONCLUSIONS

The ROC area is estimated with a Bayesian procedure that is based on the Bayesian predictive distribution of two future observations corresponding to patients selected randomly from two populations. With a diagnostic marker that is ordinal and using a conjugate prior for the parameters of the model, a resampling of the posterior distribution produces the posterior distribution of the area under the ROC curve. Several examples show that the methodology is easy to implement with standard statistical packages.

This approach with the predictive distribution can easily be extended to studies that have several readers and modalities. If additional patient covariates are to be taken into account, regression models for ordinal responses along with the Bayesian predictive distribution will in a similar way determine the ROC area.

Acknowledgements

This effort was supported by the Biostatistics Core of the Melanoma SPORE from NCI award number P50 CA093459 01 A1.

REFERENCES

- Broemeling, L.D. (1985). *Bayesian Analysis of Linear Models*. New York, Marcel-Dekker.
- Ekmekcioglu, S., Ellerhorst, J.A., Prieto, G., Johnson, M.M., Broemeling, L., and Grimm, E.A. (2004). "Tumor Inducible Nitric Oxide Synthase Expression Predicts Survival in Stage III Melanoma," Submitted to New England Journal of Medicine.
- Hanley, J.A. (1966). "The Use of Binormal Model for Parametric ROC Analysis of Quantitative Diagnostic Tests," *Statistics In Medicine*, 15, 1575-85.
- Hanley, J.A. (1988). "The Robustness of the 'Binormal' Assumptions Used in Fitting ROC curve," *Medical Decision Making*, 8, 197-203.
- Hellmich, M., Abrams, K.R., Jones, D.R., Lambert, P.C. (1998). "A Bayesian Approach to a General Regression Model for ROC Curves," *Medical Decision Making*, 18, 436-443.
- Masaryk, A.M., Ross, J.S., DiCello, M.C., Modic, M.T., Paranandi, L., and Masarayk, T.J. (1991). "3DFT MR Angiography of the Carotid Bifurcation: Potential and Limitations as a Screening Examination," *Radiology*, 148, 839-843.
- O'Malley, J.A., Zou, K.H. (2001). "Bayesian Regression Methodology for Estimating a Receiver Operating Characteristic Curve with Two Radiologic Applications, Prostate Biopsy and Spiral CT of Ureteral Stones," *Academic Radiology*, 8, 713-725.
- Obuchowski, N.A. (1997). "Nonparametric Analysis of Clustered ROC Curve Data," *Biometrics*, 53, 567-578.
- Peng, F., and Hall, W.J. (1996). "Bayesian Analysis of ROC Curves Using Markov-Chain Monte Carlo Methods," *Medical Decision Making*, 16, 404-411.

Pepe, M.S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, Oxford University Press.

Swets, J.A. (1986). "Indices of Discrimination or Diagnostic Accuracy: Their ROC's and Implied Models," *Psychological Bulletin*, 99, 100-117.

Tamm, E.P. (2004). Personal Communication.

Tosteson, A.A.N. and Begg, C.B. (1988). "A General Regression Methodology for ROC Curve Estimation," *Medical Decision Making*, 8, 204-215.

Zhou, X.H., McClish, D.K. and Obuchowski, N.A. (2002) *Statistical Methods in Diagnostic Medicine*. New York, Wiley.

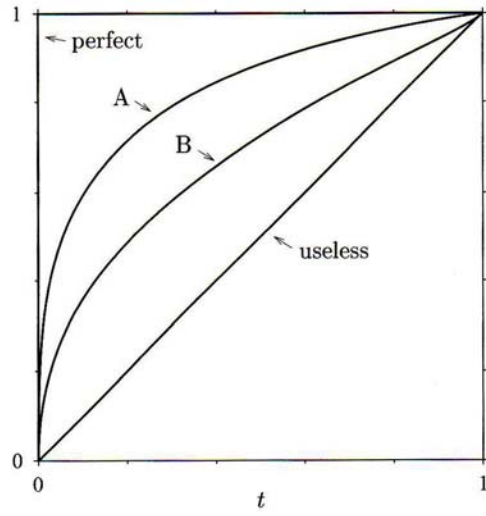


Fig. 1 ROC curves for two tests, A and B, where test A is unequivocally better. Also shown are ROC curves for the useless and perfect tests for comparison