The Effect of Two Priors on Bayesian Estimation of "Proper" Binormal ROC Curves from Common and Degenerate Datasets

Richard M. Zur, BSc, Lorenzo L. Pesce, PhD, Yulei Jiang, PhD

Rationale and Objectives: We showed previously that maximum-likelihood (ML) and Bayesian (with a flat prior on a common parameterization of the model) estimates of "proper" binormal receiver operating characteristic (ROC) curves produce similar results. We propose a new prior that is flat over the area under the ROC curve (AUC) and investigate its effect on the Bayesian estimates.

Materials and Methods: In two simulation studies, we compared Bayesian estimation of the AUC with the two prior probability distributions against ML estimation in terms of root mean squared error (RMSE) and the coverage of 95% confidence (or credible) intervals (both abbreviated CIs). In the first study, we simulated categorical data that tend to be "well-behaved" and produce ROC curve estimates that most would consider reasonable. In the second study, we simulated coarsely discretized categorical data that often included so-called degenerate datasets that cause the ML estimate to be the perfect ROC curve.

Results: For the well-behaved datasets, all three AUC estimates were similar in terms of RMSE and 95% CI coverage. For the coarsely discretized datasets, the RMSE of ML was consistently greater than that of Bayesian estimation and the 95% CI coverage of ML estimation was consistently below nominal, whereas the 95% CI coverage of Bayesian estimation was consistently equal to, or greater than, nominal.

Conclusion: Bayesian estimation with a flat prior on the AUC can provide reasonable inference from datasets with coarsely categorized data that are prone to be degenerate and produce results similar to other estimation methods on well-behaved datasets.

Key Words: ROC curves; Bayesian analysis; degenerate data.

Receiver operating characteristic (ROC) analysis is a fundamental method for the evaluation of diagnostic accuracy (1–3). An ROC curve is a plot of true-positive fraction (TPF, or sensitivity) versus false-positive fraction (FPF, or 1-specificity). The conventional binormal model for ROC analysis provides satisfactory ROC curve fits in a wide variety of practical situations (4–7). However, except for ROC curves that are symmetrical with respect to the negative 45° line in the ROC plot, the conventional binormal model produces ROC curve estimates that contain "hooks" (ie, a change in the ROC curve curvature [eg, from convex to concave], which for the conventional binormal model implies that a portion of the ROC curve falls below the "guessing line" defined by TPF = FPF). These ROC curve estimates are considered unsatisfactory because hooks indicate diagnostic accuracies that are worse than guessing (8–10). ROC models that describe ROC curves guaranteed to have a monotonically decreasing slope are known as "proper" ROC models (3). In this article, we focus on the so-called "proper" binormal model (11,12).

Previously (Zur RM, unpublished data, 2010), we compared maximum-likelihood (ML) and Bayesian estimates of "proper" binormal ROC curves. The Bayesian estimates were based on a prior probability distribution that is flat (ie, constant) over the most common parameterization of the proper binormal model (11). Prior probability distributions are a well-known characteristic of Bayesian estimation and they incorporate information obtained independently from the data at hand (13,14). Because of that, priors are expected either to improve estimations or to bias them. We showed (Zur RM, unpublished data, 2010) that the Bayesian and ML estimates are similar in terms of root mean squared error (RMSE) for the area under the ROC curve (AUC), TPF values at fixed

Acad Radiol 2010; 17:969–979

From the Department of Radiology, The University of Chicago, 5841 South Maryland Avenue, MC2026, Chicago, IL 60637 (R.M.Z., L.L.P., Y.J.). Supported in part by the National Cancer Institute of the National Institutes of Health (NIH) through grants R01 CA092361 and P30 CA14599, by the National Center for Research Resources of NIH through grant S10 RR021039, and by the National Institute of Biomedical Imaging and Bioengineering of NIH through grant R01 EB000863 (Kevin S. Berbaum, PI). Received December 4, 2009; accepted March 26, 2010. **Address correspondence to:** R.M.Z. e-mail: zur@uchicago.edu

doi:10.1016/j.acra.2010.03.020

FPF values and FPF values at fixed TPF values. In this article, we evaluate the effect on Bayesian estimation of proper binormal ROC curves of a new prior that is not flat over the most common parameterization of the proper binormal model but, rather, flat over the AUC values. We refer to both of these flat priors as low information because they do contain information (as will be demonstrated later through reparameterization) and can affect ROC analysis. Our motivation was that, whereas a prior flat on the curve parameters is expected to influence minimally the estimation of the curve parameters, it is probably more desirable to influence minimally the estimation of the AUC—the most commonly reported ROC summary index in radiological research (15). It is not possible to impose a prior that is simultaneously flat on both the curve parameters and the AUC because the curve parameters and the AUC are nonlinear functions of each other (16). Therefore, a tradeoff is unavoidable. Although the effect of prior information or preconceived views is usually not discussed in the context of ROC curve estimation, a review of the radiological literature that reports ROC analysis shows that prior experience and belief do seem to influence our understanding and acceptance of ROC curve estimates (15).

BACKGROUND

The "Proper" Binormal ROC Model

The proper binormal ROC model introduced by Metz et al (11,12) is usually specified by two parameters: d_a and c, which define the signal-present and signal-absent distributions of cases with respect to a latent decision variable, v (for details refer to Metz et al (11)). The minimum of the AUC value (0.5) occurs when $d_a = 0$ and $c = 0$, and the AUC value increases when either d_a increases toward infinity or $|c|$ increases toward 1, or both. The ROC curve is symmetrical around the negative 45° line in the ROC plot when $c = 0$, whereas the ROC curve is skewed toward the left (ie, toward the ordinate) when $c < 0$ and toward the right with respect to the negative 45° line in the ROC plot (ie, toward the line defined by TPF = 1) when $c > 0$.

In this article, we treat all ROC data as ordinal-categorical, because the limited precision of real-world quasi-continuous data implies that they can be mapped to some ordinal-categorical data without loss of information relevant to ROC analysis. We partition the latent decision variable axis into I categories with I-1 cut-points, v_{ci}, $i = 1, 2, \ldots, I$-1, hereafter denoted by $\{v_{ci}\}$ (17–19). If necessary, we reduce the number of categories to $I \leq 20$ using the LABROC5 algorithm (20), which appears to affect the bias and variance of ML estimates negligibly as long as at least 20 categories are used (12,20). As is customary, we consider these cut points as nuisance parameters because we are more interested in estimating the ROC curve than in estimating the cut points. Therefore, the likelihood function—the probability of the data conditional on the model—assumes that the ROC data are obtained through multinomial sampling (20).

ML Estimation of the "Proper" Binormal ROC Model

Let us denote the likelihood function by $L(d_a, c, \{v_{ci}\}; \mathbf{D}) = p(\mathbf{D} \mid d_a, c, \{v_{ci}\})$, where the semicolon indicates that the likelihood function is dependent, but not conditional, on the data, $\mathbf{D}$. The ML estimate is the set of parameter values, d_a, c, and $\{v_{ci}\}$ that maximizes $L(\bullet)$. Uncertainty in d_a, c, and $\{v_{ci}\}$ can be estimated from the inverse of the Fisher information matrix (21). Uncertainty in other quantities of interest (eg, the AUC, TPF, and FPF) can be estimated from the Fisher information matrix using the delta method (16). An extensive discussion on this likelihood function, the ML estimate, and the estimate of the variance-covariance matrix can be found elsewhere (12).

Bayesian Estimation of the "Proper" Binormal ROC Model

From Bayes' rule, the posterior probability distribution of the proper binormal model parameters is given by

$$p(d_a, c, \{v_{ci}\}|\mathbf{D}) = \frac{L(d_a, c, \{v_{ci}\}; \mathbf{D})p(d_a, c, \{v_{ci}\})}{p(\mathbf{D})}, \quad (1)$$

where $p(d_a, c, \{v_{ci}\})$ is the prior probability distribution and $p(\mathbf{D})$ is the marginal likelihood (13). The marginal likelihood can be considered as a normalization constant, which does not affect most estimates of the posterior probability distribution (13). Therefore, Bayesian estimation focuses on the probability of the model given the data (ie, the posterior probability distribution), rather than on the probability of the data given a model (ie, the likelihood function). However, as shown in equation (1), a prior probability distribution is required to estimate the posterior probability distribution. It is not always clear what the best or even a reasonable prior probability distribution is. Furthermore, we are often interested in indices that are different from, or in addition to, the parameters that we estimate directly. In such instances, even after necessary transformations (16), we sometimes find that a prior probability distribution that is reasonable for one set of parameters does not appear to be reasonable for other parameters of interest. Moreover, to estimate a high-dimensional, nonstandard, probability distribution, as is often required with Bayesian estimation, can be computationally challenging. Here, we use Markov chain Monte Carlo (MCMC) algorithms for that purpose (22).

MCMC Estimation of the Bayesian Posterior Probability Distribution

MCMC is a common approach to obtaining random samples from probability distributions, even of large dimensions (eg, >10) (22–24). MCMC algorithms are well-suited for Bayesian estimation of ROC curves because the estimation can involve a large number of dimensions due to potentially large numbers of the cut points, $\{v_{ci}\}$. With MCMC