Journal of Statistical Planning and Inference ▮▮▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Bayesian ROC curve estimation under binormality using a rank likelihood

Jiezhun Gu[a,*], Subhashis Ghosal[b]

[a]Duke Clinical Research Institute, Duke University Medical Center, P.O. Box 17969, Durham, NC 27715, USA
[b]Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

ARTICLE INFO

Article history:
Received 20 June 2007
Received in revised form
18 September 2008
Accepted 19 September 2008

Keywords:
Binormal model
MCMC
Rank-based likelihood
ROC curve
Posterior consistency

ABSTRACT

There are various methods to estimate the parameters in the binormal model for the ROC curve. In this paper, we propose a conceptually simple and computationally feasible Bayesian estimation method using a rank-based likelihood. Posterior consistency is also established. We compare the new method with other estimation methods and conclude that our estimator generally performs better than its competitors.

1. Introduction

Receiver operating characteristic (ROC) analysis was first introduced in the context of medical diagnostic test by Lusted (1960). Since then, it has become a unique tool in measuring the accuracy of diagnostic tests. It has the ability to compare two proportions against each other (true positive and false positive) (Swets, 1973). In the general diagnostic setting with the gold standard for the truth, each patient is known to be from either the non-diseased or the diseased population. Diagnostic result is obtained by comparing the diagnostic test value with the subjective decision threshold value. For example, let X and Y denote two diagnostic test variables from the non-diseased and the diseased populations, respectively. Based on a decision threshold value $c_t \in \mathbb{R}$, a patient from X (or Y) is diagnosed as positive if $X > c_t$ (or $Y > c_t$). The ROC curve can capture the whole decision rule by plotting $t = \Pr(X > c_t)$ (1-specificity) on the x-axis and $R(t) = \Pr(Y > c_t)$ (sensitivity) on the y-axis for varying $c_t \in \mathbb{R}$. This means that ROC can demonstrate all possible decision operating rules in a unit square with the scales of false positive fraction and true positive fraction. The ROC curve is increasing, and is invariant under any monotone increasing transformation on diagnostic variables X and Y. Further, the area under the curve can be interpreted as $\Pr(Y > X)$ (Bamber, 1975). This may be used to quantify the accuracy of different diagnostic tests if the curves do not cross. These inherited features have made ROC analysis flourishing; see the reviews by Swets and Pickett (1982), Hanley (1989), Zhou et al. (2002) and Pepe (2003).

Throughout the literature on estimation of ROC curves for continuous diagnostic variables based on independent observations, the most popular assumption is binormality. Here we assume that the diagnostic test variables of the non-diseased and diseased groups are normally distributed after some monotone increasing transformation H. That is (without loss of generality), $H(X) \sim \text{Normal}(0,1)$ and $H(Y) \sim \text{Normal}(\mu, \sigma^2)$, with the convention that $\mu > 0$. If not, we can consider $-X$ and $-Y$ as measurements which will satisfy this convention. Then the binormal ROC curve is given by $R(t) = \Phi(a + b\Phi^{-1}(t))$, where $a = \mu/\sigma$ and $b = 1/\sigma$. The corresponding AUC has a closed-form expression $\Phi(a/\sqrt{1+b^2})$ (McClish, 1989). Thus, in the binormal model, the distribution of

* Corresponding author. Tel.: +1 919 668 8967; fax: +1 919 668 7049.
E-mail address: jiezhun.gu@duke.edu (J. Gu).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.09.014

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

X and Y can be parameterized by (μ, σ, H). The appropriateness of the binormality assumption were discussed well by Hanley (1988, 1989, 1996), Swets (1986), Metz et al. (1998) and Cai and Moskowitz (2004).

Existing estimation methods for continuous diagnostic variables based on independent observations under binormality assumption are abundant.

Hsieh and Turnbull (1996) considered a generalized least squares (GLS) estimator, which requires choosing a set of grid points. They also studied an alternative estimator—a minimum distance estimator which avoids choosing grid points. However, they did not provide any computational algorithm.

Metz et al. (1998) observed that the "truth-state runs" for the combined ordered data are transformation invariant. Hence the binormal model may be reduced to normality. More specifically, they proposed a method, called LABROC4, by treating continuous diagnostic test data as extension of ordinal data. The rank-ordered data were grouped into some fixed I categories conditional on the "truth-state runs". Both binormal parameters and the nuisance parameters of $I-1$ category boundary points can be estimated by the maximum likelihood (ML) method (Dorfman and Alf, 1969) at the same time. They also proposed LABROC5 algorithm, another version of LABROC4, by combining "truth-state runs" in an ad hoc way, in order to deal with the case when the number of "truth-state runs" is large. Actually, they are not the true ML estimates which maximize $\Pr(X, Y|a, b)$. Pepe (2003) and Cai and Moskowitz (2004) gave more comments on this point.

Later on, several authors focused on the maximum-invariant rank-based likelihood and proposed estimation methods based on this likelihood.

Zou and Hall (2000) considered a method based on a rank likelihood introduced by Dabrowska and Doksum (1988) for linear transformation models. Dabrowska and Doksum's (1988) likelihood can even apply to censored data. The likelihood (in the uncensored case) is defined as the probability of observing the given set of ranks. This was originally used by Kalbfleisch and Prentice (1973) for the special case of Cox proportional hazard model. If this transformation model can be extended to allow heteroscedasticity as well, then the binormal model becomes a special case. Zou and Hall (2000) gave an efficient maxima search algorithm without evaluating the whole likelihood function. However, this method requires costly computations.

Based on the observation that $R(t) = \Pr(\bar{F}(Y) \leqslant t) = \mathrm{E}[1(\bar{F}(Y) \leqslant t)]$, Pepe (2000) and Alonzo and Pepe (2002) considered a generalized linear model (GLM) for $U_{ji} = 1(Y_j > X_i)$ based on $\hat{t}_i = \hat{\bar{F}}_m(X_i)$, $i = 1, \ldots, m$, $j = 1, \ldots, n$ and the probit link; here $\hat{\bar{F}}_m(x)$ denotes the empirical survival function of $X_1, \ldots, X_m$. However, the applicability of the GLM estimation method is not fully justified due to the dependency of U_{ji}'s and the randomness of $\hat{t}_i$'s. Cai and Moskowitz (2004) showed that binormality implies that the ratio of the density function for the diseased group over that for the non-diseased group evaluated at y is equal to the ratio of the normal density with mean μ and standard deviation σ over that of the standard normal evaluated at $-H(y)$. Hence, the full likelihood can be essentially expressed by binormal parameters and nuisance parameters. By profiling this likelihood with respect to the nuisance parameters, Cai and Moskowitz (2004) proposed a maximum profile likelihood method. Their searching method is more efficient than the earlier estimation methods, especially when the sample size is large. In the same paper, they also proposed a pseudo-maximum likelihood estimator (PMLE) based on imputation of the transformation function H depending on the data from two groups of subjects. One nice feature of GLM andPMLE methods is that they can be easily extended to adjust covariate effects. The estimators based on modeling of distribution function do not satisfy the invariance property (under monotone increasing transformations on the diagnostic variables) in general, such as Box–Cox transformation method (Zou and Hall, 2000); see reviews by Zhou et al. (2002), Pepe (2003) and Cai and Moskowitz (2004) for more details.

Here, we propose a Bayesian method using the rank likelihood assuming binormality. This new method treats the transformation as a latent function as in the rank-based likelihood used by Zou and Hall (2000). However, the Bayesian approach avoids function evaluation needed in the method of Zou and Hall (2000). More specifically, through Gibbs sampling, we generate samples from the posterior distribution of (μ, σ) (equivalently, (a, b)) to construct a Bayes estimator. A similar technique was used by Hoff (2007) in the context of Bayesian estimation of semiparametric copula. Unlike Bamber's (1975) approach, where confidence intervals are obtained by large sample approximation, our error estimates and credible intervals are automatically obtained from the posterior samples. We also obtain consistency of the rank-based posterior distribution for all except on a set of true parameter values of Lebesgue measure zero.

The following notations will be used. Two-sided-truncated version of normal distribution with mean μ, variance σ^2 and corresponding truncated region (e_1, e_2) is denoted by $\mathrm{TN}(\mu, \sigma^2, (e_1, e_2))$. Hereafter, ϕ and Φ denote the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of the standard normal distribution, respectively; $\Phi_{\mu,\sigma}$ denotes the c.d.f. of the normal distribution with mean μ and standard deviation σ, that is, $\Phi_{\mu,\sigma}(x) = \Phi((x-\mu)/\sigma)$.

The paper is organized as follows. The method is described in Section 2. Posterior consistency is studied in Section 3. Simulation studies and real data analysis are provided in Sections 4 and 5, respectively.

2. Bayesian estimation using rank-based likelihood

Under the setting of binormality, the joint distribution of (X, Y) depends on (μ, σ, H). The binormal model remains invariant under the action of the group of monotonically increasing transformations. Under binormality, the ROC function does not depend on H. Thus the study of the ROC curve based on a likelihood approach becomes much simpler if likelihood used for that purpose does not involve H at all. Traditionally, this can be achieved by profiling the original likelihood—the joint distribution of (X, Y), with respect to H; see Cai and Moskowitz (2004). However, a much simpler approach is possible based on invariance. This is because the distribution of an invariant statistic depends only on (μ, σ) but not on H. Since the ranks (and the labels) form

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

a maximal invariant statistic, a rank-based likelihood, which is a function of (μ,σ) only, can be used instead of the original likelihood.

A binormal model assumes that $X_1,\ldots,X_m \sim$ independently and identically distributed (i.i.d.) F, $Y_1,\ldots,Y_n \sim$ i.i.d. G, $\mathbf{X}=(X_1,\ldots,X_m)$ and $\mathbf{Y}=(Y_1,\ldots,Y_n)$ are continuous and independent, and there exists a monotone increasing transformation H, such that

$$Z_i = H(X_i) \sim \text{i.i.d. Normal}(0,1), W_j = H(Y_j) \sim \text{i.i.d. Normal}(\mu,\sigma^2). \tag{1}$$

Let $\mathbf{Z}=(Z_1,\ldots,Z_m)$, $\mathbf{W}=(W_1,\ldots,W_m)$; $\mathbf{S}=(\mathbf{X},\mathbf{Y})=(S_1,\ldots,S_N)$, $\mathbf{Q}=(\mathbf{Z},\mathbf{W})=(Q_1,\ldots,Q_N)$, $\widetilde{\mathbf{S}}$ and $\widetilde{\mathbf{Q}}$ be order statistics of $\mathbf{S}$ and $\mathbf{Q}$, respectively. Let the combined ranks be denoted by $\mathbf{R}_N=R(\mathbf{S})=(R_{N1},\ldots,R_{NN})$, $N=m+n$, and labels $\mathbf{L}=L(\mathbf{S})=L(\mathbf{Q})=(L_{N1},\ldots,L_{NN})$, where $L_{Ni}=0$ (or 1) if S_i from $\mathbf{X}$ (or $\mathbf{Y}$), labels $\widetilde{\mathbf{L}}=L(\widetilde{\mathbf{S}})=L(\widetilde{\mathbf{Q}})=(\widetilde{L}_{N1},\ldots,\widetilde{L}_{NN})$, where $\widetilde{L}_{Ni}=0$ (or 1) if $\widetilde{S}_i$ from $\mathbf{X}$ (or $\mathbf{Y}$).

Under the binormality assumption, the ranks and labels of $\mathbf{Q}$ preserve those of $\mathbf{S}$. Therefore, we can define a set $\mathscr{D}_{\text{obs}}$ invariant under H as follows:

$$\begin{aligned}\mathscr{D}_{\text{obs}} &= \{(\mathbf{z},\mathbf{w})\in\mathbb{R}^{m+n} : R(\mathbf{z},\mathbf{w})=R(\mathbf{S}), L(\mathbf{z},\mathbf{w})=L(\mathbf{S})\},\\ &= \{(\mathbf{z},\mathbf{w})\in\mathbb{R}^{m+n} : \underline{z}_k < z_k < \overline{z}_k, \underline{w}_l < w_l < \overline{w}_l, L(\mathbf{z},\mathbf{w})=L(\mathbf{S})\},\end{aligned} \tag{2}$$

where $\mathbf{z}=(z_1,\ldots,z_m)$, $\mathbf{w}=(w_1,\ldots,w_n)$, $\underline{z}_k=\max_i\{z_i : R_{Ni}<R_{Nk}\}\vee\max_j\{w_j : R_{N(m+j)}<R_{Nk}\}$, $\overline{z}_k=\min_i\{z_i : R_{Nk}<R_{Ni}\}\wedge\min_j\{w_j : R_{Nk}<R_{N(m+j)}\}$, $\underline{w}_l=\max_i\{z_i : R_{Ni}<R_{N(m+l)}\}\vee\max_j\{w_j : R_{N(m+j)}<R_{N(m+l)}\}$, $\overline{w}_l=\min_i\{z_i : R_{N(m+l)}<R_{Ni}\}\wedge\min_j\{w_j : R_{N(m+l)}<R_{N(m+j)}\}$, for all $k,i=1,\ldots,m$; $l,j=1,\ldots,n$. Hence, a rank-based likelihood can be defined based on this invariant set (2) as

$$\Pr\{(\mathbf{Z},\mathbf{W})\in\mathscr{D}_{\text{obs}}|\mu,\sigma)\} = \Pr\{(\mathbf{z},\mathbf{w})\in\mathbb{R}^{m+n} : R(\mathbf{z},\mathbf{w})=R(\mathbf{S}), L(\mathbf{z},\mathbf{w})=L(\mathbf{S})|\mu,\sigma\}. \tag{3}$$

Zou and Hall (2000) evaluated the maxima of this likelihood (3) by an efficient search algorithm which does not need evaluation of the whole likelihood.

Instead of evaluating the likelihood function directly, an alternative is to adopt the Bayesian approach and sample from the posterior. The posterior mean of (μ,σ) is considered as Bayes estimator based on rank likelihood (BRL). The posterior median may also be used in place of posterior mean. We shall denote the posterior density of (μ,σ) by $\pi^*(\mu,\sigma)=\pi(\mu,\sigma|R(\mathbf{S}),L(\mathbf{S}))$.

Now, we present a data augmentation technique and implement Gibbs sampling to compute $\pi^*(\mu,\sigma)$. More specifically, we describe a procedure to sample from the conditional distribution of $((\mu,\sigma)|\mathbf{Q},R(\mathbf{S}),L(\mathbf{S}))$ and $(\mathbf{Q}\mid\mu,\sigma,R(\mathbf{S}),L(\mathbf{S}))$, where $\mathbf{Q}=(\mathbf{Z},\mathbf{W})$. We choose the commonly used improper prior $\pi(\mu,\sigma^2)\propto\sigma^{-2}$ for (μ,σ), and order the combined vector $(\mathbf{X},\mathbf{Y})$ denoted as $\widetilde{\mathbf{S}}$ and record the corresponding labels of $\widetilde{\mathbf{S}}$ as $\widetilde{\mathbf{L}}$. Gibbs sampling procedure is shown as follows:

1. Choose an initial value of (μ,σ), and generate the initial values of $\mathbf{Q}$ using (1).
2. Start the iterations:
   (a) Conditional on (μ,σ), update $\widetilde{\mathbf{Q}}$ with constraint $(\mathbf{Z},\mathbf{W})\in\mathscr{D}_{\text{obs}}$, equivalently, labeling the order statistics satisfying $\widetilde{\mathbf{L}}=\mathbf{L}=(\widetilde{L}_{N1},\ldots,\widetilde{L}_{NN})$ through the following sequential truncated normal component-wise simulations $(i=1,\ldots,N)$:

$$\widetilde{Q}_i^{(new)} \sim \begin{cases}\text{TN}(0,1,(\widetilde{Q}_{i-1},\widetilde{Q}_{i+1})) & \text{when } \widetilde{L}_{Ni}=0,\\ \text{TN}(\mu,\sigma^2,(\widetilde{Q}_{i-1},\widetilde{Q}_{i+1})) & \text{when } \widetilde{L}_{Ni}=1,\end{cases}$$

   where $\widetilde{Q}_0=-\infty$, $\widetilde{Q}_{N+1}=\infty$.
   (b) Update $\mathbf{Z}$ and $\mathbf{W}$ values based on $\widetilde{\mathbf{Q}}^{(\text{new})}$ and $\widetilde{\mathbf{L}}$.
   (c) Update μ and σ by their posterior distributions as

$$\begin{aligned}&\sigma^2|\text{rest} \sim \text{inverse gamma}\,((n-1)/2,(n-1)s_w^2/2),\\ &\mu|\text{rest} \sim \text{Normal}(\bar{W}_n,\sigma^2/n),\end{aligned}$$

   where $s_w^2=\sum_{j=1}^n(W_j-\bar{W}_n)^2/(n-1)$, $\bar{W}_n=\sum_{j=1}^n W_j/n$.
3. After burn-in, we obtain the intercept and slope's estimates denoted as $\hat{a}$ and $\hat{b}$ in the binormal model by averaging out the sampled values of μ/σ and $1/\sigma$, respectively. We also calculate $100(1-\alpha)\%$ credible interval for a as $(q_{a,\alpha/2},q_{a,1-\alpha/2})$, where $q_{a,\alpha/2}$ and $q_{a,1-\alpha/2}$ denote $\alpha/2$ and $1-\alpha/2$ quantiles of the sampled values of a. The credible interval of $(q_{b,\alpha/2},q_{b,1-\alpha/2})$ for b is similarly defined.

Remark 1. (1) The prior we used for (μ,σ) is a non-informative prior. Although conjugate prior is a natural alternative, the right choice of the hyperparameters is not trivial. In further simulation not shown in this paper, we found that inappropriate choice of hyperparameters lead to considerable bias, especially for the parameter σ.

(2) We also found through simulation that our method works much better for the parametrization (a,b) rather than (μ,σ), that is, the posterior mean of (a,b) is a much more accurate estimator of (a,b) than the posterior mean of (μ,σ) as an estimator for (μ,σ).

(3) For larger sample sizes, we need to increase the number of MCMC samples to ensure convergence of the Gibbs samples.

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

3. Consistency of the posterior

Let (μ_0, σ_0) be the true value of (μ, σ). It is important to know if the posterior for (μ, σ) is consistent at (μ_0, σ_0), i.e., whether or not the posterior distribution concentrates around (μ_0, σ_0) (Ghosh and Ramamoorthi, 2003). Assume that the disease prevalence is asymptotically stable, i.e., when $N \to \infty$, $n/N \to \lambda$, where $0 < \lambda < 1$. Each observation X_i or Y_j is randomly sampled from the whole population, and with the gold standard for the truth, it can be labeled perfectly as either X or Y sample. Also assume that the joint prior density $\pi(\mu, \sigma)$ is dominated by the Lebesgue measure (denoted as ν).

Theorem 1. *Suppose that $n/N \to \lambda$, $0 < \lambda < 1$, and assume that binormality (1) hold and (μ, σ) has joint prior density $\pi(\mu, \sigma) > 0$ a.e. over $\mathbb{R} \times \mathbb{R}^+$ with respect to the Lebesgue measure ν. Then for (μ_0, σ_0) a.e. $[\nu]$, and for any neighborhood $\mathcal{U}_0$ of (μ_0, σ_0), we have that*

$$\lim_{N\to\infty} \pi((\mu, \sigma) \in \mathcal{U}_0 | R(\mathbf{S}), L(\mathbf{S})) = 1 \quad \text{a.s. } [P^{\infty}_{\mu_0,\sigma_0,H}], \tag{4}$$

where $P^{\infty}_{\mu_0,\sigma_0,H}$ denotes the joint distribution of all $\mathbf{X}$'s and $\mathbf{Y}$'s under the binormal model (1) with (μ_0, σ_0) as the true value of (μ, σ).

The proof is based on an application of Doob's Theorem stated below:

Doob's Theorem. [*cf.* Ghosal and Van der Vaart, 2009]: *Let $X^{(n)}$ be observations whose distribution depends on a parameter θ, both of which take values in Polish spaces. Assume that θ is equivalent to a $\mathfrak{X}^{(\infty)}$-measurable random variable, i.e., there exists a $\mathfrak{X}^{(\infty)}$ measurable function f on $\mathfrak{X}^{(\infty)}$ such that $\theta = f(\omega^{\infty})$ a.e. $[\Pi \times P^{(\infty)}_{\theta}]$. Then the posterior $\Pi(\cdot|X^{(n)})$ is strongly consistent at θ for almost every θ $[\Pi]$.*

Proof of Theorem 1. We can suppose that an observation comes from the case or the control groups randomly with probability λ and $1 - \lambda$, respectively, so the pooled observation $\mathbf{S} = (S_1, \ldots, S_N)$ is unconditionally given by $S_i \sim$ i.i.d. $(1 - \lambda)F + \lambda G$. Thus we have $U_i = ((1 - \lambda)F + \lambda G)(S_i)$ i.i.d. Uniform(0,1). Now, as in Theorem *a* on page 157 of Hájek and Šidák (1967), we have

$$\mathrm{E}\left(U_i - \frac{R_{Ni}}{N+1}\right)^2 = \frac{1}{N}\sum_{j=1}^{N} \mathrm{E}\left[\left(U_i - \frac{j}{N+1}\right)^2 \middle| R_{Ni} = j\right] = \frac{1}{N}\sum_{j=1}^{N} \frac{j(N-j+1)}{(N+1)^2(N+2)} < \frac{1}{N}. \tag{5}$$

Hence, U_i is an in-probability limit of $\mathscr{F}_N$-measurable random variables, where $\mathscr{F}_N$ is the σ-field generated by $\{R_{N1}, \ldots, R_{NN}\}$. Therefore,

$$U_i = \lim_{N'\to\infty} \frac{R_{N'i}}{N'+1} \quad \text{for } i \geqslant 1, \text{ with probability 1 for some subsequence } \{N'\}, \tag{6}$$

and hence U_i is a $\mathscr{F}_\infty$-measurable random variable, where $\mathscr{F}_\infty = \sigma\langle \cup_1^\infty \mathscr{F}_N \rangle$. Note that $U_i = ((1 - \lambda)\Phi + \lambda\Phi_{\mu,\sigma})(N_i)$ are i.i.d. Uniform(0,1), where $N_i = H(S_i)$. Based on the sample size N, we denote the indices of observations conditionally from F as $(i_1, \ldots, i_m)$. Hence, conditionally on $L_{Ni_j} = 0$, $j = 1, \ldots, m$, $N_{i_j} \sim$ Normal(0, 1) and $U_{i_j} = ((1 - \lambda)\Phi + \lambda\Phi_{\mu,\sigma})(N_{i_j}) \sim$ i.i.d. $g_{\mu,\sigma}$ (say), which belongs to a regular parametric family indexed by parameters (μ, σ). More specifically, Cramér type regularity conditions can be verified by applying the inverse function theorem. Thus some consistent estimator for (μ, σ), such as the MLE, a Bayes estimator, the method of moment estimator, or a minimum distance estimator can be easily obtained. Hence, there exists a sequence of functions h_N and h, such that $(\mu, \sigma) = \lim_{N\to\infty} h_N(U_{i_1}, \ldots, U_{i_m}) = h(U_{i_1}, U_{i_2}, \ldots)$.

Therefore, there exists a function h^* of all ranks and labels such that

$$\begin{aligned}(\mu, \sigma) &= h(U_{i_1}, U_{i_2}, \ldots) \\ &= h^*(\{R_{N1}, \ldots, R_{NN}, N \geqslant 1\}) \quad \text{a.s. } [P^{\infty}_{\mu_0,\sigma_0,H}].\end{aligned}$$

By Doob's Theorem, the consistency of posterior (4) at (μ_0, σ_0) holds a.e. $[\pi]$, and hence at (μ_0, σ_0) a.e. $[\nu]$. □

Remark 2. In most infinite dimensional applications of Doob's Theorem, because there is no appropriate analog of the Lebesgue measure, the exceptional set of points where the posterior may be inconsistent is null only when measured with respect to the prior. This severely dilutes the importance of the conclusion since a null set with respect to a given prior may be quite large when measured with respect to another prior. In the present case, existence of the dominating Lebesgue measure removes this arbitrariness.

4. Simulation studies

We compare the BRL estimator of binormal model for ROC curve with other methods, such as maximum profile likelihood estimator (MLE) and PMLE proposed by Cai and Moskowitz (2004), binary regression approach (GLM) by Alonzo and Pepe

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

Table 1
Estimates of binormal ROC's parameter a and b using methods BRL, MLE, PMLE, GLM and LABROC4 (which we abbreviate as LAB inside the table)

m		Bias					Mean square error				
n		BRL	MLE	PML	GLM	LAB	BRL	MLE	PML	GLM	LAB
50	a	0.071	0.079	0.064	0.034	0.032	0.051	0.063	0.064	0.059	0.059
50	b	−0.004	0.043	0.062	−0.009	0.002	0.020	0.022	0.022	0.022	0.022
200	a	−0.007	0.026	0.018	0.013	0.002	0.010	0.013	0.014	0.013	0.013
200	b	−0.006	0.014	0.017	−0.004	0.001	0.004	0.005	0.005	0.005	0.005

Our BRL estimate is based on 100 simulated data sets, with 95 000 Gibbs samples (100 000 MCMC iterations, but first 5000 samples are used to burn-in for each replication). Other estimates are based on 1000 simulated data sets.

Table 2
Sampling properties of BRL and PMLE

m		Sampling SE		Ave($\widehat{SE}$)		95% Coverage	
n		BRL	PMLE	BRL	PMLE	BRL	PMLE
50	a	0.215	0.245	0.234	0.248	0.98	0.974
50	b	0.141	0.135	0.143	0.193	0.97	0.968
200	a	0.098	0.116	0.112	0.113	0.96	0.945
200	b	0.066	0.070	0.064	0.073	0.95	0.949

Our BRL estimate is based on 100 simulated data sets, 95 000 Gibbs samples (100 000 iterations with burn-in at 5000). The PMLE estimate is based on 1000 simulated data sets.

Table 3
Coverage probabilities of AUC and corresponding average lengths of the 90% CI shown beneath in the parentheses obtained by BRL, BN-G and BN-T methods

Data		$m=n=15$			$m=n=50$		
u_x, σ_x	u_y, σ_y	BRL	BN-G	BN-T	BRL	BN-G	BN-T
A		0.97	0.886	0.866	0.89	0.871	0.882
0,1	1, 1	(0.249)	(0.254)	(0.250)	(0.141)	(0.141)	(0.143)
		0.94	0.859	0.801	0.86	0.890	0.873
0,1	3, 3	(0.237)	(0.233)	(0.205)	(0.132)	(0.134)	(0.122)
B		0.93	0.862	0.860	0.91	0.888	0.856
0,1	1, 1	(0.258)	(0.288)	(0.285)	(0.158)	(0.159)	(0.158)
		0.92	0.857	0.910	0.91	0.772	0.925
0,1	3, 3	(0.134)	(0.097)	(0.076)	(0.054)	(0.062)	(0.045)

Our BRL estimate is based on 100 simulated data sets and 95 000 Gibbs samples (100 000 iterations with burn-in at 5000), BN-G and BN-T's estimates are based on 1000 simulated data sets and corresponding 1000 resamples. Simulated data sets are generated by lognormal, location-scale exponential distributions (abbreviated as A, B, respectively) with different combinations of the parameters (A: **X** and **Y** data set are generated from the lognormal with corresponding normal parameters (u_x, σ_x) and (u_y, σ_y), respectively; B: **X** and **Y** are generated from the exponential distribution with rate = 0.5 and the location and scale parameters (u_x, σ_x) and (u_y, σ_y), respectively).

(2002) and LABROC4 by Metz et al. (1998), through the same data generating scheme used in Cai and Moskowitz (2004). We generate $(X_1, \ldots, X_m) \sim$ i.i.d. Normal$(0, 1)$, $(Y_1, \ldots, Y_n) \sim$ i.i.d. Normal(μ, σ^2), where $(\mu, \sigma) = (1.868, 1.5)$. We calculate the bias and the MSE of the estimates using methods BRL, MLE, PMLE, GLM and LABROC4. The true values of parameters are set at $a = 1.245$ and $b = 0.667$, equivalently, at $\mu = 1.868, \sigma = 1.5$. Simulation results in Table 1 also include the information contained in Cai and Moskowitz (2004).

In order to check whether or not the frequentist variability of the estimate can be approximated by the estimated variability in large samples, we examine the sampling standard error (sampling SE), the average of the estimated SE estimator and MCMC SE obtained by PMLE and BRL, respectively, denoted as Ave($\widehat{SE}$) in Table 2. We also compare the coverage probabilities of the 95% confidence interval for the estimates a and b. Simulation results in Table 2 also contain the simulation output of Cai and Moskowitz (2004).

Besides considering the performance of the estimators in the binormal model, we also compare the accuracy of the estimators of AUC functional obtained by our BRL with BN-G (ROC GLM method by Pepe, 2000), BN-T (Box–Cox method by Zou and Hall, 2000). Coverage probabilities of AUC and corresponding average lengths of the 90% CI (shown beneath in the parentheses) are displayed in Table 3.

Based on the limited simulation results shown above, we conclude that the BRL estimator of (a, b) has considerably smaller MSE than the estimators given by MLE, PML, GLM and LABROC4 methods in this simulation setting. When the sample size increases, although BRL estimator tends to have slightly larger bias, its mean square error is much smaller than the others in Table 1. Compared to PMLE, our BRL estimates have smaller sampling variation in most cases and similar coverage probabilities

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

Table 4
Real data analysis CA-125: comparison of estimates (standard errors which are shown beneath in the parentheses) of binormal parameters obtained by our BRL and other semiparametric methods

	BRL $\mu=2,\sigma=2$	BRL $\mu=3,\sigma=2$	MLE	PMLE	Zou and Hall	GLM (2002)	LABROC4
a	0.7636 (0.1836)	0.7651 (0.1847)	0.76 (0.191)	0.719 (0.198)	0.727 (0.190)	0.778 (0.197)	0.720 (0.185)
b	1.097 (0.1328)	1.092 (0.1330)	1.065 (0.140)	1.020 (0.148)	1.007 (0.142)	1.017 (0.167)	1.002 (0.137)

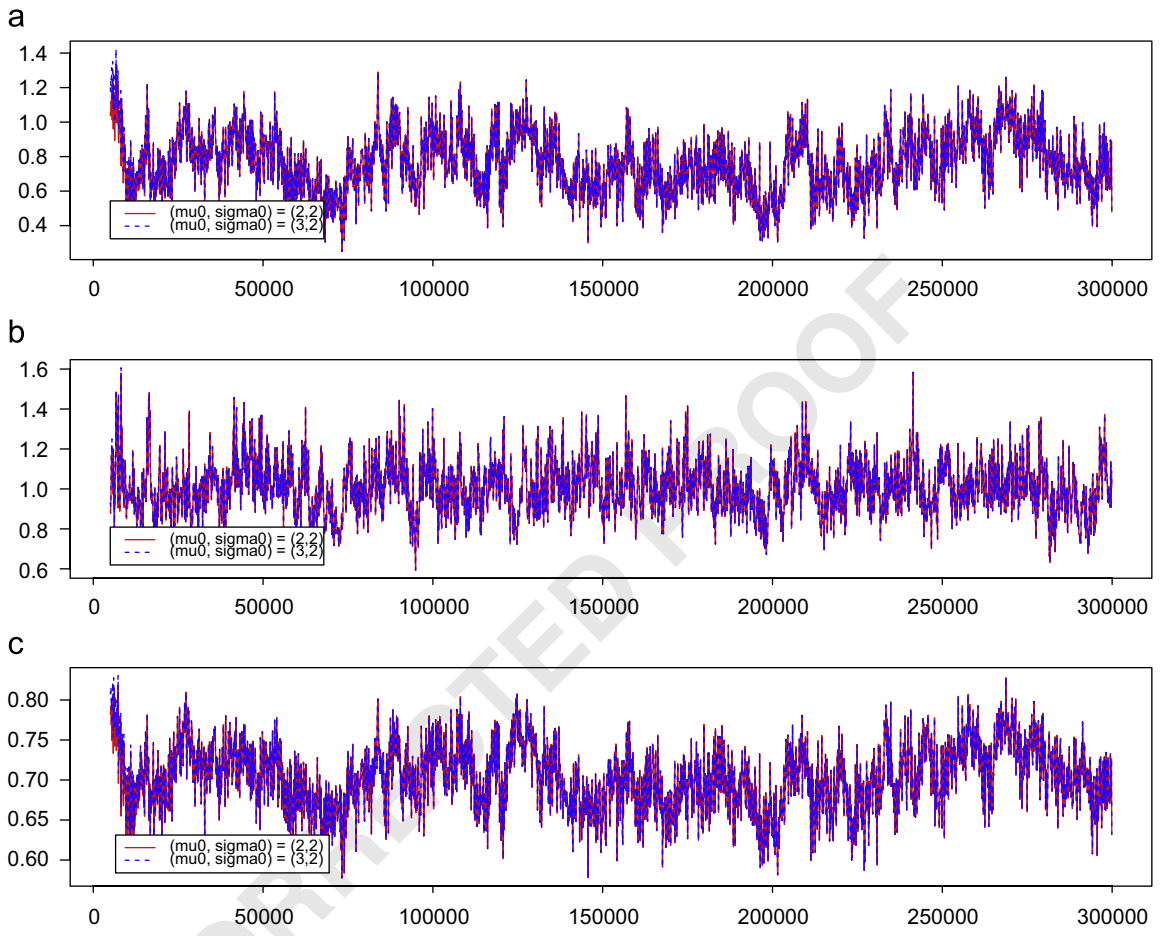

Fig. 1. Trace plots of intercept (a), slope (b) and AUC (c).

in Table 2. Also, the simulation results show that the posterior variability and asymptotic variability using BRL and PMLE method, are not far from the actual sampling SE, and hence the former can be used to estimate the frequentist variability. Moreover, compared to PMLE, our BRL estimator performs better when estimating b based on sample size $m=n=50$. Compared to BN-G and BN-T, our BRL estimator sometimes have larger average length of CI for AUC but with much higher coverage probability when the simulated data sets are from location-scale exponential distributions. When the data are from lognormal, which means the binormality assumption holds exactly, BRL estimator performs slightly better in terms of both accuracy and efficiency when $(u_y,\sigma_y)=(1,1)$.

5. Real data analysis

We will use the data set published by Wieand et al. (1989). This study was based on 51 patients as the control group diagnosed as pancreatitis and 90 patients as the cases group diagnosed as pancreatic cancer by two biomarkers, namely a cancer antigen

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

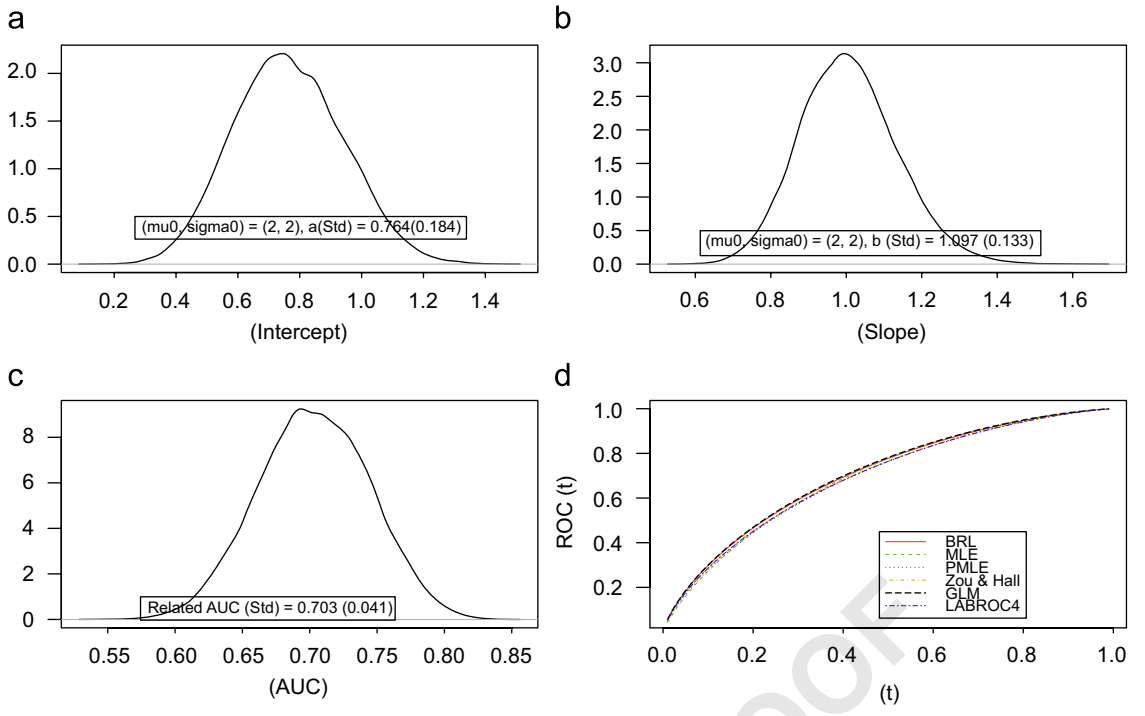

Fig. 2. Density plots of intercept (a), slope (b), AUC (c) and ROC estimates (d).

(CA 125) and a carbohydrate antigen (CA 19-9). The purpose was to decide which marker would better distinguish the case group from the control group. We will compare our BRL estimator with MLE and PMLE by Cai and Moskowitz (2004), Zou and Hall (2000), GLM (Alonzo and Pepe, 2002) and LABROC4 (Metz et al., 1998) using biomarker CA 125 for illustration. Our BRL estimate is based on 295 000 Gibbs samples with 300 000 iterations burn-in at 5000, after which every 10th sample is collected. We use the same prior as described in Section 2. In order to obtain BRL estimate, we choose two sets of initial values as $\mu = 3, \sigma = 2$ and $\mu = 2, \sigma = 2$ for the data set. Table 4 also reflects the information contained in Cai and Moskowitz (2004). Convergence of MCMC is examined by trace plots of estimates of a, b and AUC from MCMC samples by two sets of different initial values in Fig. 1. Posterior consistency is present for different initial values. The posterior density plots of MCMC samples of (a, b), AUC and our BRL estimate of ROC curve are shown in Fig. 2, respectively, when we use the initial values as $\mu = 2, \sigma = 2$. Compared with the other estimates, our BRL estimate tends to have slightly smaller estimated SE.

Acknowledgments

We want to thank editors and reviewers for their correction and valuable comments. Research of the second author is partially supported by NSF Grant no. DMS-0349111.

References

Alonzo, T.A., Pepe, M.S., 2002. Distribution-free ROC analysis using binary regression techniques. Biostatistics 3, 421–432.
Bamber, D., 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J. Math. Psych. 12, 387–415.
Cai, T., Moskowitz, C., 2004. Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. Biostatistics 5, 573–586.
Dabrowska, D.M., Doksum, K.A., 1988. Partial likelihood in transformation models with censored data. Scand. J. Statist. 15, 1–23.
Dorfman, D.D., Alf, E., 1969. Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating-method data. J. Math. Psych. 6, 487–496.
Ghosal, S., Van der Vaart, A.W., 2009. Theory of Nonparametric Bayesian Inference. Cambridge University Press, Cambridge. (to be published). Q1
Ghosh, J.K., Ramamoorthi, R.V., 2003. Bayesian Nonparametrics. Springer, New York.
Hájek, J., Šidák, Z., 1967. Theory of Rank Tests. Academic press, New York.
Hanley, J.A., 1988. The robustness of the "binormal" assumptions used in fitting ROC curves. Medical Decis. Making 8, 197–203.
Hanley, J.A., 1989. Receiver operating characteristic (ROC) methodology: the state of the art. Critical Rev. Diagnostic Imaging 29, 307–335.
Hanley, J.A., 1996. The use of the "binormal" model for parametric ROC analysis of quantitative diagnostic tests. Statist. Medicine 15, 1575–1585.
Hoff, P.D., 2007. Extending the rank likelihood for semiparametric copula estimation. Ann. Appl. Statist. 1, 265–283.
Hsieh, F.S., Turnbull, B.W., 1996. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. Ann. Statist. 24, 25–40.
Kalbfleisch, J.D., Prentice, R.L., 1973. Marginal likelihoods based on Cox's regression and life model. Biometrika 60, 267–278.
Lusted, L.B., 1960. Logical analysis in roentgen diagnosis. Radiology 74, 178–193.
McClish, D.K., 1989. Analyzing a portion of the ROC curve. Medical Decis. Making 9, 190–195.

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014

Metz, C.E., Herman, B.A., Shen, J., 1998. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. Statist. Med. 17, 1033–1053.

Pepe, M.S., 2000. An interpretation for the ROC curve and inference using GLM procedures. Biometrics 56, 352–359.

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

Swets, J.A., 1973. The relative operating characteristic in psychology. Science 182, 990–1000.

Swets, J.A., 1986. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. Psychol. Bull. 99, 100–117.

Swets, J.A., Pickett, R.M., 1982. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Academic Press, New York.

Wieand, S., Gail, M.H., James, B.R., James, K.L., 1989. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. Biometrika 76, 585–592.

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

Zou, K.H., Hall, W.J., 2000. Two transformation models for estimating an ROC curve derived from continuous data. J. Appl. Statist. 27, 621–631.

Please cite this article as: Gu, J., Ghosal, S., Bayesian ROC curve estimation under binormality using a rank likelihood. J. Statist. Plann. Inference (2008), doi: 10.1016/j.jspi.2008.09.014