Statistical inference for the two-sample problem under likelihood ratio ordering, with application to the ROC curve estimation

Dingding Hu | Meng Yuan | Tao Yu | Pengfei Li

1Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ontario, Canada
2Department of Statistics and Data Science, National University of Singapore, Singapore

Correspondence
Tao Yu, Department of Statistics and Data Science, National University of Singapore, Singapore.
Email: yu.tao@nus.edu.sg

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2020-04964; Singapore Ministry of Education Academic Research Tier 1 Fund, Grant/Award Number: A-8000413-00-00

The receiver operating characteristic (ROC) curve is a powerful statistical tool and has been widely applied in medical research. In the ROC curve estimation, a commonly used assumption is that larger the biomarker value, greater severity the disease. In this article, we mathematically interpret “greater severity of the disease” as “larger probability of being diseased.” This in turn is equivalent to assume the likelihood ratio ordering of the biomarker between the diseased and healthy individuals. With this assumption, we first propose a Bernstein polynomial method to model the distributions of both samples; we then estimate the distributions by the maximum empirical likelihood principle. The ROC curve estimate and the associated summary statistics are obtained subsequently. Theoretically, we establish the asymptotic consistency of our estimators. Via extensive numerical studies, we compare the performance of our method with competitive methods. The application of our method is illustrated by a real-data example.

KEYWORDS
area under the ROC curve, Bernstein polynomials, likelihood ratio ordering, ROC curve, Youden index

1 INTRODUCTION

The ROC curve is a powerful statistical tool and has been widely applied in many scientific areas; it evaluates the diagnostic abilities of a binary classifier for varied discrimination thresholds. Consider a medical study, where a biomarker serves as a binary classifier for diagnosing the healthy/diseased status of each patient. The ROC curve plots the “sensitivity” against “one minus the specificity” at all the possible thresholds of the biomarker. Here, sensitivity/specificity refers to the probability of correctly identifying the diseased/healthy individuals. The ROC curve related inferences have been extensively investigated in the literature; we refer to Pepe,7 Zhou et al,2 Qin and Zhang,3 and Chen et al4 for a comprehensive review of the existing developments.

Practically, for an individual, we may assume that the larger biomarker value indicates greater possibility/severity of the disease; therefore, we can classify individuals as diseased if their corresponding biomarkers exceed a given cutoff point \(x\), and vice versa; see Yin and Tian5 for a discussion. Under this assumption, the sensitivity and specificity are \(1 - F_1(x)\) and \(F_0(x)\), where \(F_0(\cdot)\) and \(F_1(\cdot)\) denote the cumulative distribution functions (cdfs) of the biomarkers for individuals that are healthy and diseased, respectively. The ROC curve is then given by

\[
ROC(s) = 1 - F_1 \left( F_0^{-1}(1 - s) \right),
\]

for \(s \in (0, 1)\).
In the literature, there are two common strategies for mathematically accommodating the assumption that larger biomarker associates with the greater severity of the disease. The first interprets it as the biomarker values of diseased individuals are stochastically greater than those of healthy ones; that is $F_1(x) \leq F_0(x)$ for every $x \in \mathbb{R}$. The second assumes that larger biomarker value corresponds to larger probability of getting the disease, with the assumption that “the larger biomarker value indicates greater severity of the disease,” this second strategy is equivalent to interpret “greater severity of the disease” as “larger probability of being diseased.” More specifically, let $X$ and $D$ denote the biomarker and indicator of the membership of an individual: $D = 1$ or $0$ respectively indicates that the individual is diseased or healthy. Using Bayes’ formula, we have

$$P(D = 1|X = x) = \frac{P(D = 1)f_1(x)}{P(D = 1)f_1(x) + P(D = 0)f_0(x)};$$

where $f_1(x)$ and $f_0(x)$ are the probability density functions (pdfs) of the biomarkers in the diseased population and the healthy population, respectively. Hence, the second interpretation equivalently assumes that $f_1(x)/f_0(x)$ is a monotone function of $x$, which is known as the likelihood ratio ordering. We observe that with the likelihood ratio ordering, the derivative of $\text{ROC}(s)$,

$$\frac{d(\text{ROC}(s))}{ds} = \frac{f_0(F_1^{-1}(1-s))}{f_1(F_1^{-1}(1-s))},$$

is a nonincreasing function, which implies that $\text{ROC}(s)$ is concave for $s \in (0, 1)$. We refer to Gneiting and Vogel for the importance of concavity in the interpretation and modeling of ROC curves. Furthermore, we observe that as an assumption, likelihood ratio ordering is also widely adopted in many important applications of the Neyman–Pearson lemma, such as the Karlin–Rubin theorem. In this article, we shall incorporate the likelihood ratio ordering assumption to improve the estimation of ROC curve and its summary statistics.

There are three types of estimation methods for the ROC curves: the parametric, the semiparametric, and the non-parametric methods. We observe that many of these estimation methods did not incorporate the likelihood ratio ordering constraint; ignoring this constraint may result in less accurate ROC estimation. As far as we are aware, they are two estimation methods in the literature that have accommodated this constraint: Dykstra et al. considered the estimation of $F_0$ and $F_1$ with a maximum nonparametric likelihood approach; Yu et al. proposed a smoothed likelihood approach to estimate $f_0$ and $f_1$. The former leads to a non-smooth estimate of $P(D = 1|X = x)$ in (1), resulting in an inefficient estimate of the optimal cutoff point that maximizes the Youden index. For a theoretical comparison of the convergence rates of the optimal cutoff point between the non-smooth and smoothed estimates of $P(D = 1|X = x)$, we refer to Hsieh and Turnbull; the numerical performance of this estimate can be found in Section 1 of the supplementary material. The latter relies on a smoothing parameter the optimal choice of which is not easily obtained.

In this article, we propose a Bernstein polynomial method that accommodates the likelihood ratio ordering assumption, and use the empirical likelihood principle to estimate $F_0$, $F_1$, ROC curve, and two popular summary statistics: the area under the curve (AUC) and the Youden index. With the help of the existing R function glmnet, we establish an algorithm to implement our proposed method; we integrate our algorithm as an R package BPLR available at “https://github.com/Dingding-Hu/BPLR-package.” We have also established the asymptotic consistency of our method. It is noteworthy that the resulting estimate of $P(D = 1|X = x)$ from our method is smooth; through extensive numerical studies, we observe that our ROC curve, Youden index, and the associated optimal cutoff point estimates outperform those from the existing methods in most examples.

The rest of the article is organized as follows. Section 2 proposes the Bernstein polynomial method with the empirical likelihood principle to establish the estimators for $F_0(\cdot)$ and $F_1(\cdot)$, and subsequently result in the estimators of the ROC curve and its summary statistics. The consistency of the proposed method is also studied. Section 3 presents the simulation results. Section 4 compares our method with existing methods in a real-data example, and Section 5 concludes the article with some discussion. Technical details are given in the supplementary document.

## 2 MAIN RESULTS

### 2.1 Bernstein polynomials approach

Denote by $\{X_1, \ldots, X_{n_0}\}$ and $\{Y_1, \ldots, Y_{n_1}\}$ the random samples of biomarkers from the healthy and the diseased populations, respectively. The pdf and cdf of $X_i$’s are $f_0$ and $F_0$; the pdf and cdf of $Y_j$’s are $f_1$ and $F_1$. Let $n = n_0 + n_1$ be the total sample size; denote by $t_1 < \cdots < t_n$ the distinct values of the combined sample.
We assume that \( f_0 \) and \( f_1 \) satisfy the likelihood ratio ordering, that is, \( f_1(x)/f_0(x) \) is an increasing function of \( x \). To incorporate this constraint in the estimation procedure, we propose to model \( \log(f_1(x)/f_0(x)) \) as a linear combination of Bernstein polynomials. The definition of the Bernstein polynomials is given below.

**Definition 1** (Bernstein polynomials). Let \( N \) be a positive integer. For each \( 0 \leq l \leq N \), the Bernstein polynomials are defined to be

\[
B_l(x; N) = \binom{N}{l} x^l (1-x)^{N-l}, \quad l = 0, \ldots, N, \quad x \in [0, 1].
\]

Bernstein polynomials can serve as a set of base functions in the expansion of a nonparametric component. A nice feature of such an expansion is that it is able to incorporate the shape constraints as some condition(s) of the coefficients of polynomials. For a monotonic nonparametric component, we can expand it to be \( \sum_{l=0}^{N} \beta_l B_l(x; N) \), which is an increasing function of \( x \) if

\[
\beta_0 \leq \beta_1 \cdots \leq \beta_N. \tag{2}
\]

We refer to Wang and Ghosh\(^{14} \) for more details of incorporating other types of shape constraints in the Bernstein polynomial expansion.

Next, we apply the Bernstein polynomials to our estimation problem. For presentational convenience, in the development below, we assume that \( N \) is given and all biomarkers are ranged in \([0, 1]\). Remarks 1 and 2 give the solutions of how to choose \( N \) and transform the biomarkers so that they are contained in \([0, 1]\) in practice. Based on the likelihood ratio ordering assumption, \( \log(f_1(x)/f_0(x)) \) is an increasing function of \( x \). We propose to model it by a linear combination of Bernstein polynomials:

\[
\log(f_1(x)/f_0(x)) = \sum_{i=0}^{N} \beta_i B_i(x; N), \tag{3}
\]

with \( (\beta_0, \ldots, \beta_N) \) satisfying (2); and then incorporate the maximum empirical likelihood to estimate \( (\beta_0, \ldots, \beta_N), F_0, \) and \( F_1 \). To this end, let

\[
a_i = \sum_{j=1}^{m} I(X_j = t_i) \quad \text{and} \quad b_i = \sum_{j=1}^{m} I(Y_j = t_i).
\]

Based on the observed two-sample data, the full likelihood function is given by

\[
L = \prod_{i=1}^{m} [f_0(t_i)^{a_i} \cdot (f_1(t_i))^{b_i}].
\]

For \( i = 1, \ldots, m \), let

\[
p_{i0} = f_0(t_i) \quad \text{and} \quad p_{i1} = f_1(t_i).
\]

Then, the empirical likelihood\(^{15} \) is

\[
L = \prod_{i=1}^{m} p_{i0}^{a_i} \cdot p_{i1}^{b_i}.
\]

We assume \( n_1/n \to \lambda \in (0, 1) \) as \( n \to \infty \). For simplicity, hereafter we write \( \lambda = n_1/n \) and assume that it is constant, since it does not affect our technical development. We further define

\[
\phi_i = (1-\lambda)f_0(t_i) + \lambda f_1(t_i) = (1-\lambda)p_{i0} + \lambda p_{i1}.
\]

By model (3), we have

\[
p_{i0} = \frac{\phi_i}{1 - \lambda + \lambda \exp \left\{ \sum_{l=0}^{N} \beta_l B_l(t_i; N) \right\}},
\]

\[
p_{i1} = \frac{\exp \left\{ \sum_{l=0}^{N} \beta_l B_l(t_i; N) \right\} \phi_i}{1 - \lambda + \lambda \exp \left\{ \sum_{l=0}^{N} \beta_l B_l(t_i; N) \right\}}.
\]
That is, \( p_{i0}'s \) and \( p_{i1}'s \) are determined by \( \phi_i's \) and \( (\beta_0, \ldots, \beta_N) \).

With the above reparameterization, the empirical likelihood function of \( (\phi_1, \ldots, \phi_m, \beta_0, \ldots, \beta_N) \) is then given as

\[
L = \prod_{i=1}^{m} p_{i0}^{\alpha_i} p_{i1}^{\beta_i} = (\lambda)^{-n_1} (1 - \lambda)^{-n_0} \cdot L_1(\phi_1, \ldots, \phi_m) \cdot L_2(\beta_0, \ldots, \beta_N),
\]

where

\[
L_1(\phi_1, \ldots, \phi_m) = \prod_{i=1}^{m} \phi_i^{\alpha_i + \beta_i} \quad \text{and} \quad L_2(\beta_0, \ldots, \beta_N) = \prod_{i=1}^{m} \left[ \theta(t_i) \right]^{\beta_i} \{1 - \theta(t_i)\}^{\alpha_i}
\]

with

\[
\theta(x) = \frac{\lambda \exp \left\{ \sum_{l=0}^{N} \beta_l B_l(x; N) \right\}}{1 - \lambda + \lambda \exp \left\{ \sum_{l=0}^{N} \beta_l B_l(x; N) \right\}}.
\]

Note that feasible \( \phi_i's \) satisfy

\[
\phi_i \geq 0, \quad \sum_{i=1}^{m} \phi_i = 1, \quad (4)
\]

and

\[
\sum_{i=1}^{m} \phi_i \theta(t_i) = \lambda, \quad (5)
\]

to ensure that both \( F_0 \) and \( F_1 \) are cdfs. The maximum empirical likelihood estimator (MELE) of \( (\phi_1, \ldots, \phi_m, \beta_0, \ldots, \beta_N) \) is then defined to be

\[
(\hat{\phi}_1, \ldots, \hat{\phi}_m, \hat{\beta}_0, \ldots, \hat{\beta}_N) = \arg \max_{\phi_1, \ldots, \phi_m, \beta_0, \ldots, \beta_N} L,
\]

subject to constraints (4), (5), and the inequality constraints in (2).

To solve the optimization problem above, we consider the following reparameterization of \( \beta_i's \):

\[
\beta_0 = a_0, \quad \beta_1 = a_0 + a_1, \ldots, \beta_N = \sum_{l=0}^{N} a_l.
\]

Then (2) is equivalent to

\[
\alpha_1 \geq 0, \ldots, \alpha_N \geq 0.
\]

The above reparameterization implies that

\[
\sum_{l=0}^{N} \beta_l B_l(x; N) = \sum_{l=0}^{N} a_l B_l^*(x; N),
\]

where \( B_l^*(x; N) = \sum_{k=0}^{N} B_k(x; N) \) for \( l = 1, \ldots, N \) and \( B_0^*(x; N) = 1 \).

With a slight abuse of notation, we write

\[
L_2(a_0, \ldots, a_N) = \prod_{l=1}^{m} \left[ \theta(t_l) \right]^{\beta_l} \{1 - \theta(t_l)\}^{\alpha_l}
\]

with

\[
\theta(x) = \frac{\lambda \exp \left\{ a_0 + \sum_{l=1}^{N} a_l B_l^*(x; N) \right\}}{1 - \lambda + \lambda \exp \left\{ a_0 + \sum_{l=1}^{N} a_l B_l^*(x; N) \right\}}.
\]
The following proposition summarizes the results for calculating the MELEs of \( \phi_i \)'s and \((a_0, a_1, \ldots, a_N)\). The proof is given in the supplementary material.

**Proposition 1.** Let

\[
(\hat{\phi}_1, \ldots, \hat{\phi}_m) = \arg \max_{\phi_1, \ldots, \phi_m} L_1(\phi_1, \ldots, \phi_m) \text{ subject to (4)},
\]

and

\[
(\hat{a}_0, \ldots, \hat{a}_N) = \arg \max_{a_0, \ldots, a_N} L_2(a_0, \ldots, a_N),
\]

subject to \(a_l \geq 0\) for \(l = 1, \ldots, N\). Then

(a) \( \hat{\phi}_i = (a_i + b_i)/n \) for \(i = 1, \ldots, m; \)

(b) \( \sum_{i=1}^{m} \hat{\phi}_i \hat{\theta}(t_i) = \lambda \), where

\[
\hat{\theta}(x) = \frac{\lambda \exp \left\{ \hat{a}_0 + \sum_{i=1}^{N} \hat{a}_i B^*_i(x; N) \right\}}{1 - \lambda + \lambda \exp \left\{ \hat{a}_0 + \sum_{i=1}^{N} \hat{a}_i B^*_i(x; N) \right\}}.
\] (6)

Proposition 1 implies that we can maximize \(L_1\) and \(L_2\) separately to obtain the MELEs of \( \phi_i \)'s and \((a_0, a_1, \ldots, a_N)\). The MELEs of \( \phi_i \)'s have the closed form in Proposition 1 (a). Note that \(L_2(a_0, \ldots, a_N)\) can be viewed as the likelihood for the standard logistic regression with the intercept being \(a_0 + \log(\lambda/(1 - \lambda))\) and covariates being \(B^*_i(x; N), \ldots, B^*_N(x; N)\). Then \((\hat{a}_0, \ldots, \hat{a}_N)\) can be readily calculated by using the existing R function \texttt{glmnet}. Once \( \hat{\phi}_i \)'s and \((\hat{a}_0, \ldots, \hat{a}_N)\) are available, the estimates of \(p_{0i}'s\) and \(p_{1i}'s\) are given by

\[
\hat{p}_{10} = \frac{\hat{\phi}_i}{1 - \lambda + \lambda \exp \left\{ \hat{a}_0 + \sum_{i=1}^{N} \hat{a}_i B^*_i(t_i; N) \right\}},
\]

\[
\hat{p}_{11} = \frac{\exp \left\{ \hat{a}_0 + \sum_{i=1}^{N} \hat{a}_i B^*_i(t_i; N) \right\} \phi_i}{1 - \lambda + \lambda \exp \left\{ \hat{a}_0 + \sum_{i=1}^{N} \hat{a}_i B^*_i(t_i; N) \right\}},
\]

which lead to the estimates for \(F_0(\cdot)\) and \(F_1(\cdot)\):

\[
\hat{F}_0(x) = \sum_{i=1}^{m} \hat{p}_{10} I(t_i \leq x) \quad \text{and} \quad \hat{F}_1(x) = \sum_{i=1}^{m} \hat{p}_{11} I(t_i \leq x).
\] (7)

We make some remarks for the proposed method above.

**Remark 1.** In the development above, we have assumed that \(N\) is known. In practice, we can choose it based on the Bayesian information criterion (BIC). Specifically, since we maximize \(L_2(a_0, \ldots, a_N)\) to obtain the estimators of \((a_0, \ldots, a_N)\), we can use it to establish the BIC criterion. For any \(N\), let

\[
(\bar{a}_0, \bar{a}_1, \ldots, \bar{a}_N) = \arg \max_{a_0, \ldots, a_N} L_2(a_0, \ldots, a_N),
\]

and

\[
\bar{\theta}(x) = \frac{\lambda \exp \left\{ \bar{a}_0 + \sum_{i=1}^{N} \bar{a}_i B^*_i(x; N) \right\}}{1 - \lambda + \lambda \exp \left\{ \bar{a}_0 + \sum_{i=1}^{N} \bar{a}_i B^*_i(x; N) \right\}}.
\]

That is, \(\bar{\theta}(x)\) is the MELE of \(\theta(x)\) without the monotonicity assumption. Denote by \(df_N\) the number of unknown parameters in \(\theta(x)\). We define

\[
\text{BIC}(N) = -2 \log \left[ \prod_{i=1}^{m} (\bar{\theta}_N(t_i))^{b_i} (1 - \bar{\theta}_N(t_i))^{a_i} \right] + (\log n) \cdot df_N.
\]
Consequently, \( N \) is set to be the minimizer of \( \text{BIC}(N) \). We use \( \tilde{\theta}(x) \) instead of \( \hat{\theta}(x) \) to construct the BIC criterion because without the monotonicity assumption, the number of unknown parameters in \( \theta(x) \) can be clearly counted.

**Remark 2.** Practically, a biomarker, \( x \) say, may not be in the range \([0, 1]\). We can consider the transformation

\[
x^* = \frac{x - t_{(1)}}{t_{(m)} - t_{(1)}},
\]

with \( t_{(1)} \) and \( t_{(m)} \) being the minimum and the maximum values of \( t_i \)'s. Clearly \( x^* \in [0, 1] \); we can then apply our method to the transformed biomarkers.

**Remark 3.** We observe that biomarkers may exhibit high variability in practice; applying a log transformation on them may improve the performance. Furthermore, we may include both the original and transformed biomarkers in the model. Specifically, we may consider

\[
\log [f_1(x)/f_0(x)] = a_0 + \sum_{l=1}^{N} a_l B_l^*(x^*; N) + \sum_{l=1}^{N} a_{N+l} B_l^*(z^*; N),
\]

where

\[
z^* = \frac{\log x - \log t_{(1)}}{\log t_{(m)} - \log t_{(1)}},
\]

and \( a_l \geq 0 \) for \( l = 1, \ldots, (2N) \). All our developments above can be similarly applied to (8). This modeling strategy is applied in all the simulation and real data examples.

**Remark 4.** Our method is established on the likelihood ratio ordering assumption. To check the rationale of this assumption in practice, we can plot \((\hat{F}_0, \hat{F}_1)\) in (7) and \((\tilde{F}_0, \tilde{F}_1)\), where \( \hat{F}_0 \) and \( \hat{F}_1 \) are respectively the empirical cdfs of \( \{X_1, \ldots, X_{n_0}\} \) and \( \{Y_1, \ldots, Y_{n_1}\} \). If the plots of \( \hat{F}_i \) are reasonably close to those of \( \tilde{F}_i \) for \( i = 0, 1 \), we may regard the likelihood ratio ordering as a reasonable assumption. For a more rigorous approach, we can utilize the goodness-of-fit test statistics:

\[
\Delta_n = \sup_x |\hat{F}_0(x) - \tilde{F}_0(x)| \quad \text{and} \quad \Delta_n^* = \sup_x |\hat{F}_1(x) - \tilde{F}_1(x)|,
\]

and apply the Bootstrap method to perform the test. It can be demonstrated that \( \Delta_n = \lambda \Delta_n^*/(1 - \lambda) \), and thus the testing outcomes based on either \( \Delta_n \) or \( \Delta_n^* \) are identical. Therefore, we only need to consider one, for example \( \Delta_n \), in practice.

### 2.2 Estimation of the ROC curve and its summary statistics

With the estimate \( \hat{F}_0(x) \) and \( \hat{F}_1(x) \), the ROC curve can be estimated by

\[
\hat{\text{ROC}}(s) = 1 - \hat{F}_1 \left( \hat{F}_0^{-1}(1 - s) \right).
\]

In this article, we further consider two summary statistics based on the ROC curve: AUC and Youden index. The AUC is the total area under the ROC curve, that is,

\[
\text{AUC} = \int_0^1 \text{ROC}(s) \, ds.
\]

Note that for a given threshold or cutoff point of the biomarker, it is desirable to have the corresponding value on y-axis (the sensitivity) to be as large as possible. Hence, the larger AUC value indicates that the binary classifier has stronger classification ability. With \( \hat{\text{ROC}}(s) \), AUC can be estimated by

\[
\hat{\text{AUC}} = \int_0^1 \hat{\text{ROC}}(s) \, ds.
\]
The Youden index \(J \) is defined as the maximum value of the sensitivity plus the specificity minus 1, that is,
\[
J = \max_x \{1 - F_1(x) + F_0(x) - 1\} = \max_x \{F_0(x) - F_1(x)\}.
\]
One advantage of Youden index is that it results in a criterion to choose the “optimal” cutoff point, which is the “\(x\)” where Youden index is achieved. Specifically
\[
C = \arg \max_x \{F_0(x) - F_1(x)\}.
\]
(11)
We refer to Yuan et al\(^{16}\) and the references therein for recent developments in the Youden index and the optimal cutoff point estimation. Note that (11) implies \(f_0(C) = f_1(C)\), and hence \(\theta(C) = \lambda\). Therefore we can solve the equation
\[
\hat{\theta}(\hat{C}) = \lambda,
\]
(12)
to obtain \(\hat{C}\) and as a consequence,
\[
\hat{J} = \hat{F}_0(\hat{C}) - \hat{F}_1(\hat{C}).
\]
(13)

### 2.3  Asymptotic properties

In this section, we establish the consistency of our estimators. We need the following notation. Let \(G(x) = \lambda F_1(x) + (1 - \lambda)F_0(x)\) and denote
\[
\theta_0(x) = \frac{\lambda f_1(x)}{(1 - \lambda f_0(x) + \lambda f_1(x))}.
\]
(14)
We first show the \(L_2\) convergence of \(\hat{\theta}(\cdot)\) to \(\theta_0(\cdot)\), where \(\hat{\theta}(\cdot)\) is defined by (6). For presentational continuity, we give the conditions in the Appendix, and relegate the technical details to the supplementary document.

**Theorem 1.** Assume Conditions A1–A3 in the Appendix. We have
\[
\int_0^1 \left\{ \hat{\theta}(x) - \theta_0(x) \right\}^2 dG(x) = o_p(1).
\]

With Theorem 1, we are able to establish the asymptotic consistency of \((\hat{F}_0, \hat{F}_1)\) in (7), \(\hat{ROC}(s)\) in (9), \(\hat{AUC}\) in (10), \(\hat{C}\) in (12), and \(\hat{J}\) in (13); the results are given in Theorem 2 below.

**Theorem 2.** Assume Conditions A1–A3 in the Appendix. We have
(a) \(\sup_{x \in [0,1]} |\hat{F}_0(x) - F_0(x)| = o_p(1)\) and \(\sup_{x \in [0,1]} |\hat{F}_1(x) - F_1(x)| = o_p(1)\),
(b) \(\sup_{s \in [0,1]} |\hat{ROC}(s) - ROC(s)| = o_p(1)\),
(c) \(\hat{AUC} = AUC + o_p(1)\).

Furthermore, if Condition A4 in the Appendix is also satisfied, then
(d) \(\hat{C} = C + o_p(1)\) and \(\hat{J} = J + o_p(1)\).

### 3  SIMULATION STUDY

### 3.1  Simulation setup

In this section, with simulation examples, we compare the performance of our proposed method (denoted as “BP”) with existing methods in the estimation of the ROC curves and its summary statistics. We consider the following seven competitive methods:
the Box-Cox method in Bantis et al.\textsuperscript{17} denoted as “Box-Cox”; 
- the method in Zhou and Lin\textsuperscript{10} under the binormal model, denoted as “ZL”; 
- the method in Lin et al.\textsuperscript{18} denoted as “LZL”; 
- the ECDF-based method, denoted as “ECDF,” which is also known as the nonparametric estimator in Pepe\textsuperscript{1} and Zhou et al.;\textsuperscript{2} 
- the maximum nonparametric likelihood method under the likelihood ratio ordering,\textsuperscript{8} denoted as “MNLE”; 
- the kernel-based method in Bantis et al.\textsuperscript{17} denoted as “Kernel”; 
- the maximum smoothed likelihood method under the likelihood ratio ordering in Yu et al.\textsuperscript{7} denoted as “MSLE.”

In the Kernel and MSLE methods, we follow Bantis et al.\textsuperscript{17} and Yu et al.\textsuperscript{7} We use the Gaussian kernel with bandwidths $h_0$ and $h_1$ for healthy and diseased groups respectively, where

\begin{align*}
h_0 &= 0.9 \min \left\{ s_0, \frac{q_0}{1.34} \right\} n_0^{-0.2} \\
h_1 &= 0.9 \min \left\{ s_1, \frac{q_1}{1.34} \right\} n_1^{-0.2}.
\end{align*}

Here, $s_0$ and $q_0$ are sample SD and sample interquartile range for the sample from the healthy population, and $s_1$ and $q_1$ are sample SD and sample interquartile range for the sample from the diseased population.

Throughout our numerical studies, we observe that the MNLE and MSLE methods have similar performance as that of the ECDF and kernel methods, respectively. Furthermore, the LZL method is designed to accommodate covariates; without covariates, the performance of this method is very similar to that of the ECDF method. For space limitation, we present the results for MNLE, MSLE, and LZL methods in the supplementary document.

We consider two distributional settings:

1. $f_0 \sim N(\mu_0, \sigma^2_0)$ and $f_1 \sim N(\mu_1, \sigma^2_1)$;
2. $f_0 \sim \text{Gamma}(a_0, b_0)$ and $f_1 \sim \text{Gamma}(a_1, b_1)$.

Here, $\text{Gamma}(a, b)$ denotes the gamma distribution with shape parameter $a$ and rate parameter $b$. Note that Setting (1) corresponds to the case that the model assumption for the Box-Cox method is satisfied, whereas Setting (2) corresponds to the case that the model assumption for the Box-Cox method is violated. In both cases, $f_0$ is fixed. The parameters for $f_1$ are varied such that the corresponding Youden indices are 0.3, 0.5, and 0.7, respectively. The details of these settings are given in Table 1. We consider three different combinations of sample sizes: $(n_0, n_1) = (50, 50), (100, 100)$, and $(150, 50)$. Therefore, for each distributional setting of $f_0$ and $f_1$, we have nine combinations of parameters and sample sizes. For each combination, we repeat the simulation 2000 times.

We observe that the proposed method depends on the choice of $N$. In the supplementary document, we give the frequency of the selected $N$ values based on the BIC criterion given in Remark 1 from 2000 repetitions. Furthermore, we have considered the distributional setting that $f_0$ and $f_1$ follow the Beta distributions; the details and results are also given in the supplementary document.

### 3.2 Comparison of the ROC curve estimation

In this section, we compare our proposed method with the competitive methods in the ROC curve estimation. The criteria for comparison are the $L_1$- and $L_2$-distances between the estimated and the true ROC curves. For a

| Distribution | $J$ | $AUC$ | $\mu_0$ | $\sigma^2_0$ | $\mu_1$ | $\sigma^2_1$ |
|--------------|-----|-------|---------|-------------|---------|-------------|
| Normal       | 0.3 | 0.707 | 10      | 1           | 10.771  | 1           |
| Normal       | 0.5 | 0.830 | 10      | 1           | 11.349  | 1           |
| Normal       | 0.7 | 0.929 | 10      | 1           | 12.073  | 1           |
| Distribution | $J$ | $AUC$ | $a_0$   | $b_0$       | $a_1$   | $b_1$       |
| Gamma        | 0.3 | 0.708 | 2       | 1           | 3       | 0.937       |
| Gamma        | 0.5 | 0.830 | 2       | 1           | 4       | 0.944       |
| Gamma        | 0.7 | 0.929 | 2       | 1           | 5       | 0.827       |
TABLE 2  Averages of $L_1$-distances and $L_2$-distances of five methods for estimating the ROC curve.

| Distribution (J) | Method | (50, 50) | (100, 100) | (150, 50) |
|------------------|--------|----------|------------|-----------|
|                  |        | $L_1$    | $L_2$      | $L_1$     | $L_2$     |
| Normal (0.3)     | BP     | 0.040    | 0.046      | 0.029     | 0.033     | 0.033     | 0.038     |
|                  | Box-Cox| 0.044    | 0.053      | 0.032     | 0.038     | 0.037     | 0.043     |
|                  | ZL     | 0.045    | 0.054      | 0.032     | 0.038     | 0.037     | 0.044     |
|                  | ECDF   | 0.056    | 0.071      | 0.040     | 0.051     | 0.046     | 0.058     |
|                  | Kernel | 0.047    | 0.057      | 0.035     | 0.042     | 0.040     | 0.048     |
| Normal (0.5)     | BP     | 0.032    | 0.041      | 0.023     | 0.029     | 0.026     | 0.034     |
|                  | Box-Cox| 0.036    | 0.047      | 0.025     | 0.034     | 0.029     | 0.038     |
|                  | ZL     | 0.036    | 0.049      | 0.026     | 0.034     | 0.030     | 0.039     |
|                  | ECDF   | 0.045    | 0.064      | 0.032     | 0.046     | 0.037     | 0.051     |
|                  | Kernel | 0.040    | 0.053      | 0.029     | 0.040     | 0.033     | 0.043     |
| Normal (0.7)     | BP     | 0.020    | 0.033      | 0.014     | 0.023     | 0.016     | 0.026     |
|                  | Box-Cox| 0.022    | 0.036      | 0.015     | 0.026     | 0.017     | 0.028     |
|                  | ZL     | 0.023    | 0.039      | 0.016     | 0.027     | 0.019     | 0.030     |
|                  | ECDF   | 0.029    | 0.053      | 0.020     | 0.038     | 0.023     | 0.040     |
|                  | Kernel | 0.027    | 0.044      | 0.020     | 0.033     | 0.022     | 0.035     |
| Gamma (0.3)      | BP     | 0.042    | 0.050      | 0.030     | 0.035     | 0.035     | 0.041     |
|                  | Box-Cox| 0.045    | 0.054      | 0.032     | 0.038     | 0.037     | 0.044     |
|                  | ZL     | 0.047    | 0.056      | 0.032     | 0.039     | 0.038     | 0.045     |
|                  | ECDF   | 0.057    | 0.072      | 0.040     | 0.052     | 0.046     | 0.058     |
|                  | Kernel | 0.051    | 0.062      | 0.038     | 0.046     | 0.044     | 0.051     |
| Gamma (0.5)      | BP     | 0.034    | 0.046      | 0.024     | 0.032     | 0.027     | 0.035     |
|                  | Box-Cox| 0.036    | 0.049      | 0.026     | 0.035     | 0.028     | 0.037     |
|                  | ZL     | 0.037    | 0.050      | 0.026     | 0.035     | 0.028     | 0.038     |
|                  | ECDF   | 0.045    | 0.067      | 0.032     | 0.047     | 0.036     | 0.050     |
|                  | Kernel | 0.046    | 0.060      | 0.035     | 0.044     | 0.038     | 0.046     |
| Gamma (0.7)      | BP     | 0.021    | 0.036      | 0.016     | 0.026     | 0.016     | 0.027     |
|                  | Box-Cox| 0.022    | 0.038      | 0.016     | 0.027     | 0.016     | 0.028     |
|                  | ZL     | 0.023    | 0.040      | 0.016     | 0.028     | 0.018     | 0.031     |
|                  | ECDF   | 0.029    | 0.055      | 0.021     | 0.040     | 0.022     | 0.040     |
|                  | Kernel | 0.033    | 0.050      | 0.026     | 0.038     | 0.028     | 0.038     |

generic ROC curve estimate $\overline{ROC}(s)$, the $L_1$- and $L_2$-distances between $\overline{ROC}(s)$ and the true ROC curve $ROC(s)$ are defined to be:

$$L_1(\overline{ROC}, ROC) = \int_0^1 \left| \overline{ROC}(s) - ROC(s) \right| \, ds,$$

and

$$L_2(\overline{ROC}, ROC) = \left[ \int_0^1 \left\{ \overline{ROC}(s) - ROC(s) \right\}^2 \, ds \right]^{1/2}.$$
From this table and the results given in the supplementary material, we observe that our BP method results in the smallest average $L_1$- and $L_2$-distances for all the examples; the improvement is significant. For example, the ratio of the $L_2$-distances between our BP method and the ECDF method is ranged between 0.61 and 0.71 with an average of 0.66; the ratio of the $L_2$-distances between our BP method and the ZL method is between 0.84 and 0.93 with an average of 0.88.

### 3.3 Comparison of the AUC estimation

In this section, we compare our proposed method with the competitive methods in the AUC estimation. The criteria for comparison are the relative bias (RB) and mean square error (MSE). Suppose we have $B$ point estimates of the AUC $\hat{a}^{(i)}$ for $i = 1, \ldots, B$. The RB in percentage and the MSE are respectively defined to be:

$$RB(\%) = \frac{1}{B} \sum_{i=1}^{B} \frac{\hat{a}^{(i)} - a_0}{a_0} \times 100 \quad \text{and} \quad \text{MSE} = \frac{1}{B} \sum_{i=1}^{B} (\hat{a}^{(i)} - a_0)^2,$$

where $a_0$ is the true value of the AUC. The results are summarized in Table 3.

| Distribution ($J$) | $(a_0, n_0)$ | Method | RB   | MSE  | RB   | MSE  | RB   | MSE  |
|-------------------|--------------|--------|------|------|------|------|------|------|
| Normal (0.3)      | (50, 50)     | BP     | 0.29 | 2.39 | 0.00 | 1.21 | 0.20 | 1.65 |
|                   |              | Box-Cox| 0.44 | 2.44 | 0.06 | 1.23 | 0.20 | 1.68 |
|                   |              | ZL     | 0.44 | 2.47 | 0.09 | 1.24 | 0.24 | 1.69 |
|                   |              | ECDF   | 0.09 | 2.51 | -0.10| 1.27 | -0.01| 1.74 |
|                   |              | Kernel | -1.67| 2.38 | -1.54| 1.28 | -1.53| 1.71 |
|                   | (100, 100)   |        |      |      |      |      |      |      |
|                   |              | BP     | 0.16 | 1.50 | -0.01| 0.75 | 0.16 | 1.01 |
|                   |              | Box-Cox| 0.35 | 1.50 | 0.08 | 0.75 | 0.17 | 1.02 |
|                   |              | ZL     | 0.36 | 1.53 | 0.10 | 0.76 | 0.19 | 1.03 |
|                   |              | ECDF   | 0.03 | 1.56 | -0.07| 0.77 | -0.01| 1.06 |
|                   |              | Kernel | -1.95| 1.77 | -1.68| 0.95 | -1.71| 1.24 |
|                   | (150, 50)    |        |      |      |      |      |      |      |
|                   |              | BP     | 0.05 | 0.58 | -0.01| 0.28 | 0.08 | 0.37 |
|                   |              | Box-Cox| 0.10 | 0.54 | 0.01 | 0.27 | 0.05 | 0.36 |
|                   |              | ZL     | 0.14 | 0.58 | 0.04 | 0.28 | 0.10 | 0.40 |
|                   |              | ECDF   | -0.02| 0.61 | -0.04| 0.29 | -0.01| 0.39 |
|                   |              | Kernel | -1.56| 0.88 | -1.28| 0.46 | -1.33| 0.59 |
| Normal (0.5)      | (50, 50)     | BP     | 0.69 | 2.49 | 0.29 | 1.26 | 0.11 | 1.70 |
|                   |              | Box-Cox| 0.57 | 2.58 | 0.26 | 1.30 | 0.23 | 1.76 |
|                   |              | ZL     | 0.55 | 2.60 | 0.24 | 1.30 | 0.20 | 1.75 |
|                   |              | ECDF   | 0.04 | 2.66 | -0.08| 1.34 | -0.10| 1.78 |
|                   |              | Kernel | -2.36| 2.63 | -2.08| 1.44 | -2.28| 1.86 |
|                   | (100, 100)   |        |      |      |      |      |      |      |
|                   |              | BP     | 0.33 | 1.56 | 0.06 | 0.80 | 0.17 | 0.94 |
|                   |              | Box-Cox| 0.40 | 1.55 | 0.13 | 0.80 | 0.39 | 0.95 |
|                   |              | ZL     | 0.39 | 1.58 | 0.10 | 0.81 | 0.36 | 0.97 |
|                   |              | ECDF   | -0.01| 1.65 | -0.14| 0.84 | 0.13 | 0.99 |
|                   |              | Kernel | -2.65| 2.12 | -2.31| 1.20 | -2.33| 1.37 |
|                   | (150, 50)    |        |      |      |      |      |      |      |
|                   |              | BP     | 0.10 | 0.59 | -0.01| 0.30 | 0.00 | 0.35 |
|                   |              | Box-Cox| 0.08 | 0.55 | -0.01| 0.29 | 0.08 | 0.32 |
|                   |              | ZL     | 0.13 | 0.59 | -0.01| 0.30 | 0.14 | 0.39 |
|                   |              | ECDF   | -0.05| 0.62 | -0.10| 0.31 | 0.00 | 0.35 |
|                   |              | Kernel | -2.19| 1.14 | -1.82| 0.65 | -2.03| 0.80 |
| Normal (0.7)      | (50, 50)     | BP     | 0.33 | 1.56 | 0.06 | 0.80 | 0.17 | 0.94 |
|                   |              | Box-Cox| 0.40 | 1.55 | 0.13 | 0.80 | 0.39 | 0.95 |
|                   |              | ZL     | 0.39 | 1.58 | 0.10 | 0.81 | 0.36 | 0.97 |
|                   |              | ECDF   | -0.01| 1.65 | -0.14| 0.84 | 0.13 | 0.99 |
|                   |              | Kernel | -2.65| 2.12 | -2.31| 1.20 | -2.33| 1.37 |
|                   | (100, 100)   |        |      |      |      |      |      |      |
|                   |              | BP     | 0.33 | 1.56 | 0.06 | 0.80 | 0.17 | 0.94 |
|                   |              | Box-Cox| 0.40 | 1.55 | 0.13 | 0.80 | 0.39 | 0.95 |
|                   |              | ZL     | 0.39 | 1.58 | 0.10 | 0.81 | 0.36 | 0.97 |
|                   |              | ECDF   | -0.01| 1.65 | -0.14| 0.84 | 0.13 | 0.99 |
|                   |              | Kernel | -2.65| 2.12 | -2.31| 1.20 | -2.33| 1.37 |
|                   | (150, 50)    |        |      |      |      |      |      |      |
|                   |              | BP     | 0.10 | 0.59 | -0.01| 0.30 | 0.00 | 0.35 |
|                   |              | Box-Cox| 0.08 | 0.55 | -0.01| 0.29 | 0.08 | 0.32 |
|                   |              | ZL     | 0.13 | 0.59 | -0.01| 0.30 | 0.14 | 0.39 |
|                   |              | ECDF   | -0.05| 0.62 | -0.10| 0.31 | 0.00 | 0.35 |
|                   |              | Kernel | -2.19| 1.14 | -1.82| 0.65 | -2.03| 0.80 |
Comparing the RB values in this table, we observe that our method is small in RB values for all examples, though in most cases, the RB values from the ECDF method are the smallest; the kernel method always results in negative and largest RBs in absolute values. Comparing the MSE values, our BP method has comparable or better performance than competitive methods.

### 3.4 Comparison of the Youden index and optimal cutoff point estimation

In this section, we compare our method with the competitive methods in the estimation of the Youden index and the optimal cutoff point. The criteria for comparison are the RB and MSE, which are similarly defined as those in Section 3.3. The simulation results for estimating the Youden index and the optimal cutoff point are summarized in Tables 4 and 5, respectively.

**TABLE 4** RB (%) and MSE (×1000) of five methods for estimating the Youden index.

| Distribution (J) | (n₀, n₁) | (50,50) | (100,100) | (150,50) |
|------------------|----------|---------|-----------|----------|
|                  | Method   | RB      | MSE       | RB       | MSE       | RB       | MSE       |
| Normal (0.3)     | BP       | 1.90    | 5.69      | 0.44     | 2.87      | 1.25     | 3.93      |
|                  | Box-Cox  | 4.25    | 5.86      | 1.57     | 2.89      | 2.50     | 3.97      |
|                  | ZL       | 4.66    | 5.94      | 1.80     | 2.92      | 2.84     | 4.01      |
|                  | ECDF     | 24.05   | 11.64     | 11.57    | 5.57      | 19.05    | 7.75      |
|                  | Kernel   | −0.88   | 5.61      | −2.53    | 3.08      | −1.83    | 4.05      |
| Normal (0.5)     | BP       | 1.32    | 4.91      | 0.40     | 2.43      | 0.99     | 3.29      |
|                  | Box-Cox  | 2.64    | 5.04      | 1.05     | 2.42      | 1.52     | 3.33      |
|                  | ZL       | 2.89    | 5.24      | 1.15     | 2.47      | 1.70     | 3.40      |
|                  | ECDF     | 11.67   | 9.15      | 7.40     | 4.33      | 9.28     | 6.13      |
|                  | Kernel   | −3.53   | 5.45      | −3.73    | 3.01      | −3.54    | 3.90      |
| Normal (0.7)     | BP       | 1.01    | 3.51      | 0.37     | 1.68      | 0.74     | 2.25      |
|                  | Box-Cox  | 1.64    | 3.38      | 0.70     | 1.61      | 0.95     | 2.20      |
|                  | ZL       | 2.09    | 3.84      | 0.89     | 1.77      | 1.40     | 2.87      |
|                  | ECDF     | 6.29    | 5.71      | 4.06     | 2.79      | 5.01     | 3.80      |
|                  | Kernel   | −3.74   | 4.43      | −3.43    | 2.43      | −3.43    | 3.10      |
| Gamma (0.3)      | BP       | 3.85    | 6.03      | 1.74     | 2.98      | 1.32     | 4.03      |
|                  | Box-Cox  | 5.27    | 6.25      | 2.90     | 3.08      | 3.18     | 4.27      |
|                  | ZL       | 5.60    | 6.32      | 2.92     | 3.08      | 3.19     | 4.23      |
|                  | ECDF     | 23.77   | 11.66     | 15.77    | 5.83      | 19.08    | 7.85      |
|                  | Kernel   | −0.67   | 6.10      | −1.91    | 3.35      | −0.95    | 4.17      |
| Gamma (0.5)      | BP       | 1.92    | 5.27      | 0.57     | 2.56      | 1.08     | 3.09      |
|                  | Box-Cox  | 3.26    | 5.33      | 1.66     | 2.58      | 2.62     | 3.33      |
|                  | ZL       | 3.30    | 5.47      | 1.54     | 2.60      | 2.43     | 3.35      |
|                  | ECDF     | 11.20   | 9.02      | 7.25     | 4.45      | 9.38     | 5.94      |
|                  | Kernel   | −3.42   | 5.91      | −3.64    | 3.21      | −2.96    | 3.55      |
| Gamma (0.7)      | BP       | 1.34    | 3.66      | 0.51     | 1.83      | 0.71     | 2.09      |
|                  | Box-Cox  | 1.88    | 3.46      | 0.95     | 1.75      | 1.37     | 2.08      |
|                  | ZL       | 2.28    | 4.03      | 1.01     | 1.89      | 1.81     | 3.26      |
|                  | ECDF     | 6.17    | 5.87      | 3.85     | 2.86      | 4.95     | 3.62      |
|                  | Kernel   | −3.71   | 4.56      | −3.50    | 2.62      | −3.62    | 2.88      |
From Table 4, we observe that in the estimation of the Youden index, out of 18 simulation settings, our BP method leads to the smallest RB values in 16 settings, and obtains the second smallest RB values in two settings, where the kernel method produces the smallest RB values. Comparing MSE values, the performance of our BP method and the Box-Cox method is comparable, and is better than the other methods in most of the settings.

From Table 5, for estimating the cutoff point, our BP method and the Box-Cox method result in significantly smaller MSE values than other methods. The performance of the BP and the Box-Cox method is mixed. When \( f_0 \) and \( f_1 \) are simulated from the Gamma distribution and the Youden index is 0.7, Box-Cox method produces smaller MSE values, whereas our BP method leads to smaller MSE values in all other settings.

### 4 | REAL DATA APPLICATION

In this section, we apply our method to a dataset on Duchenne Muscular Dystrophy (DMD). DMD is a type of muscular dystrophy that is genetically transmitted from a mother to her children. Muscle loss occurs at an early age for offspring...
with the disease. Female offspring with the disease do not suffer from significant symptoms compared to male offspring who die at a young age. Female carriers do not show sign of disease and therefore detection of potential female carrier is of main interest.

Percy et al\textsuperscript{19} stated that DMD carriers are more likely to have higher measurement of specific biomarkers. Four biomarkers including creatine kinase (CK), hemopexin (H), lactate dehydrogenase (LD), and pyruvate kinase (PK) are measured from the blood serum samples of a healthy group ($n_0 = 127$) and a group of carriers ($n_1 = 67$). The complete dataset was collected by Andrews and Herzberg.\textsuperscript{20} Yuan et al\textsuperscript{16} pointed out that the biomarker CK has the best performance among these four biomarkers since it corresponds to the largest estimate of the Youden index. Therefore, we apply our method and other existing methods on the biomarker CK to compare their performance.

Following Remark 1, we construct the $\text{BIC}(N)$ vs $N$ plot in Panel (a) of Figure 1; it suggests to use $N = 1$ in our BP method. In this figure, it can be seen that $\text{BIC}(N)$ exhibits an almost linear relationship with $N$. This is because BIC consists of two terms: the log-likelihood and a penalty term $(\log n) \cdot df_N$. We computed the maximized log-likelihood values and corresponding BIC values for $N = 1, \ldots, 10$, and found that the maximized log-likelihood values remain almost constant as $N$ ranges from 1 through 10, indicating that the model with $N = 1$ is sufficient to explain the data; numerical details are omitted for brevity. As a result, the penalty term dominates as $N$ increases, resulting in a linearly increasing trend in BIC as $N$ increases. Based on Remark 4, to check whether the likelihood ratio ordering assumption is reasonable, we plot our BP estimates $(\hat{F}_0, \hat{F}_1)$ and the empirical cdf estimates $(\tilde{F}_0, \tilde{F}_1)$ in Panel (b) of Figure 1; this plot suggests that the likelihood ratio ordering assumption might be reasonable. We further perform the goodness-of-fit test suggested in Remark 4; the $P$-value based on 1000 bootstrap sample is 0.975. This reinforces the validity of the likelihood ratio ordering assumption.

In Figure 2, we plot the ROC curve estimates from our BP and competitive methods. The performance of the MNLE and MSLE methods is similar to that of the ECDF and kernel method, respectively; we thus do not include them in this figure. From the figure, we observe that the ROC curve estimate from our BP method is similar to those of Box-Cox, ZL, ECDF, and LZL methods, but different from the kernel method. The difference lies in the fact that the kernel method leads to much smaller sensitivity estimate compared to other methods when 1-specificity is large, greater than 0.5 say. This is mainly because that the CK values for the diseased individuals are very skewed, which inflates the selected bandwidth when the kernel method is used to estimate $F_1$; this in turn makes $F_1$ overestimated and hence causes the sensitivity underestimated when 1-specificity is large or the value of CK is small.

We evaluate the point estimates (PEs) and bootstrap percentile confidence intervals (BPCIs) of the AUC, Youden index, and optimal cutoff point based on different methods, and display them in Table 6. The AUC estimates from all methods are similar, except for the kernel method, which leads to a smaller AUC estimate. This is consistent with the results in Figure 2, and complies with our observation in Table 3: the kernel method has the negative
FIGURE 2  Estimated ROC curves based on different methods.

TABLE 6  PEs and BPCIs of the AUC, Youden index, and optimal cutoff point by using CK as the biomarker.

| Statistics method | AUC       |         | Youden index |         | Optimal cutoff point |         |
|-------------------|-----------|---------|--------------|---------|----------------------|---------|
|                   | PE        | BPCI    | PE           | BPCI    | PE                   | BPCI    |
| BP                | 0.865     | [0.804, 0.914] | 0.588       | [0.480, 0.688] | 58.998               | [51.976, 67.311] |
| Box-Cox           | 0.874     | [0.816, 0.921] | 0.616       | [0.510, 0.710] | 58.017               | [51.869, 65.055] |
| ZL                | 0.868     | [0.805, 0.917] | 0.600       | [0.498, 0.697] | 57.000               | [50.975, 69.000] |
| LZL               | 0.863     | [0.800, 0.914] | 0.612       | [0.502, 0.727] | 56.000               | [43.000, 97.500] |
| ECDF              | 0.863     | [0.800, 0.914] | 0.612       | [0.502, 0.727] | 56.000               | [43.000, 97.500] |
| Kernel            | 0.790     | [0.748, 0.844] | 0.591       | [0.502, 0.669] | 73.356               | [54.689, 79.413] |

relative biases, and all other methods have small RBs and comparable MSEs. The Youden index estimates are similar for all methods: the maximum difference between any two estimates is less than 0.03. For the optimal cutoff point estimates, the kernel method leads to a different result from other methods; other methods give similar results. But the ZL, LZL, and ECDF methods have resulted in wider BPCIs than our method; we conjecture that this may indicate that our method has given smaller SE than other methods in the optimal cutoff point estimation. However, unlike with the simulation studies, in this real data example, we are unable to compare the estimates with the true values.

5  | CONCLUSION

We have proposed a Bernstein polynomial method to estimate the cdfs of the biomarkers in the two-sample problem under the likelihood ratio ordering assumption, and subsequently considered the the estimation of the ROC curve and its summary statistics. With simulation studies, we have compared our method with existing methods; we observe that our method performs well in the estimation of both the ROC curve and the summary statistics. With the empirical process theory, we established the consistency of our estimators. We implemented the numerical algorithm for our method in an R package named BPLR, which is ready to be applied in practice.

We observe that within the framework of this article, there are many interesting topics that can be explored in the future. First, we have established the consistency of our proposed estimators, but not the convergence rates and the
asymptotic distributions; if we are able to establish the asymptotic distributions of them in the future, we can construct the confidence bands/intervals for \( F_0(\cdot) \), \( F_1(\cdot) \), \( \text{ROC}(\cdot) \), and other summary statistics. Second, we have proposed to employ the Youden index as the criterion, and maximize it to determine the optimal cutoff point; however, there may exist other criteria; for example, the closest to (0,1) criterion and equal sensitivity and specificity criterion.\(^{21}\) It could be interesting to explore how to extend our method to accommodate these criteria. Third, in Remark 3, we have suggested to include the log transformed biomarker in our model and have applied this strategy in both the simulation and real data analysis. We observe that there may exist other reasonable transformations that can be incorporated in our method, for example, Yang et al.\(^{22}\) proposed a transformation method based on the likelihood ratio and discussed the scenarios under which that transformation is applicable. We leave this for future research.

**ACKNOWLEDGEMENTS**

The authors thank the editor, the associate editor, and two referees for constructive comments and suggestions that lead to a significant improvement over the article. Dr. Pengfei Li’s work is supported in part by the Natural Sciences and Engineering Research Council of Canada: RGPIN-2020-04964. Dr. Tao Yu’s work is supported in part by Singapore Ministry of Education Academic Research Tier 1 Fund: A-8000413-00-00.

**DATA AVAILABILITY STATEMENT**

The data analyzed in Section 4 are given as a CSV file in the supplementary material.

**ORCID**

Tao Yu \( \text{https://orcid.org/0000-0002-7921-1403} \)

**REFERENCES**

1. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York: Oxford University Press; 2003.
2. Zhou X, McClish DK, Obuchowski NA. *Statistical Methods in Diagnostic Medicine*. 2nd ed. New York: Wiley; 2011.
3. Qin J, Zhang B. Using logistic regression procedures for estimating receiver operating characteristic curves. *Biometrika*. 2003;90:585-596.
4. Chen B, Li P, Qin J, Yu T. Using a monotonic density ratio model to find the asymptotically optimal combination of multiple diagnostic tests. *J Am Stat Assoc*. 2016;111:861-874.
5. Yin J, Tian L. Joint inference about sensitivity and specificity at the optimal cut-off point associated with Youden index. *Comput Stat Data Anal*. 2014;77:1-13.
6. Wang D, Tian L, Zhao Y. Smoothed empirical likelihood for the Youden index. *Comput Stat Data Anal*. 2017;115:1-10.
7. Yu T, Li P, Qin J. Density estimation in the two-sample problem with likelihood ratio ordering. *Biometrika*. 2017;104:141-152.
8. Dykstra R, Kochar S, Robertson T. Inference for likelihood ratio ordering in the two-sample problem. *J Am Stat Assoc*. 1995;90:1034-1040.
9. Gneiting T, Vogel P. Receiver operating characteristic (ROC) curves. arXiv preprint arXiv:1809.04808, 2018.
10. Zhou XH, Lin H. Semi-parametric maximum likelihood estimates for ROC curves of continuous-scale tests. *Stat Med*. 2008;27:5271-5290.
11. Hsieh F, Turnbull B. Nonparametric methods for evaluating diagnostic tests. *Stat Sin*. 1996;6:47-62.
12. Youden WJ. Index for rating diagnostic tests. *Cancer*. 1950;3:32-35.
13. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33:1-22.
14. Wang J, Ghosh S. Shape restricted nonparametric regression with Bernstein polynomials. *Comput Stat Data Anal*. 2012;56:2729-2741.
15. Owen AB. *Empirical Likelihood*. New York: Chapman and Hall/CRC; 2001.
16. Yuan M, Li P, Wu C. Semiparametric inference of the Youden index and the optimal cut-off point under density ratio models. *Can J Stat*. 2021;49:965-986.
17. Bantis LE, Nakas CT, Reiser B. Construction of confidence intervals for the maximum of the Youden index and the corresponding cutoff point of a continuous biomarker. *Biom J*. 2019;61:138-156.
18. Lin H, Zhou XH, Li G. A direct semiparametric receiver operating characteristic curve regression with unknown link and baseline functions. *Stat Sin*. 2012;22:1427-1456.
19. Percy ME, Andrews DF, Thompson MW. Duchenne muscular dystrophy carrier detection using logistic discrimination: serum creatine kinase, hemopexin, pyruvate kinase, and lactate dehydrogenase in combination. *Am J Med Genet*. 1982;13:27-38.
20. Andrews DF, Herzberg AM. *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer; 2012.
APPENDIX. REGULARITY CONDITIONS

We need the following conditions in our technical developments for Theorems 1 and 2. They are not necessarily the weakest possible.

A1. The total sample size \( n = n_0 + n_1 \to \infty \) and \( n_1/n \to \lambda \) for \( \lambda \in (0, 1) \) being a constant.

A2. The order \( N \to \infty \) as \( n \to \infty \) and \( \lim_{n \to \infty} N/n = 0 \).

A3. The likelihood ratio ordering assumption is satisfied, or equivalently, \( \theta_0(x) \) in (14) is a nondecreasing function of \( x \). Furthermore, \( \theta_0(x) \) is a continuous function of \( x \) and there exists a \( \delta > 0 \) such that \( \delta \leq \theta_0(x) \leq 1 - \delta \) for all \( x \in [0, 1] \).

A4. There exists an \( \epsilon_0 > 0 \), such that \( \theta_0(x) \) is strictly increasing for \( x \in [C - \epsilon_0, C + \epsilon_0] \), where \( C \) is defined by (11).