Published in final edited form as:

Stat Biopharm Res. 2009 November 1; 1(4): 407–414. doi:10.1198/sbr.2009.0044.

Empirical Likelihood-Based Confidence Interval of ROC Curves

Haiyan Su [Ph.D. student],

Department of Biostatistics and Computational Biology, University of Rochester Medical Center,
Rochester, NY 14642.

Yongsong Qin [Professor], and

School of Mathematics, Guangxi Normal University, Guilin, Guangxi 541004, China.

Hua Liang [Professor]

Department of Biostatistics and Computational Biology, University of Rochester Medical Center,
Rochester, NY 14642.

Abstract

In this article we propose an empirical likelihood-based confidence interval for receiver operating characteristic curves which are based on a continuous-scale test. The approach is easily understood, simply implemented, and computationally efficient. The results from our simulation studies indicate that the finite-sample numerical performance slightly outperforms the most promising methods published recently. Two real datasets are analyzed by using the proposed method and the existing bootstrap-based method.

Keywords

Bahadur's representation; Bootstrap; Quantile; Sensitivity; Specificity

1. Introduction

Assessing the accuracy of a diagnostic test in discriminating diseased patients from nondiseased ones plays an important role in diagnostic medicine. When the response of the test is binary, the accuracy of the test is usually measured by its sensitivity and specificity. When the response of the test is continuous, its accuracy is generally measured by the receiver operating characteristic (ROC) curve, which is a plot of sensitivity versus 1-specificity, as one varies the definition of a positive test result. Many laboratory tests yield continuous measurements, and ROC curves are the most commonly used measures for the accuracy of these tests. For instance, in the first example in Section 4, we are interested in the accuracy of cerebrospinal fluid CK-BB isoenzyme measured within 24 hours of injury as a means of predicting the outcome of severe head trauma. The response of the CK-BB isoenzyme is on a continuous-scale, and we can use the ROC curve to represent the accuracy of the CK-BB isoenzyme in distinguishing a poor outcome from a good outcome.

Let X and Y be the responses of the diseased and nondiseased subjects, and $F(\cdot)$ and $G(\cdot)$ be the distribution functions of X and Y , respectively. The ROC curve is a plot of $1 - F(t)$ versus $1 - G(t)$, for $-\infty < t < \infty$, and can be written as

$$R(p) = 1 - F\{G^{-1}(1 - p)\}, \quad 0 < p < 1, \quad (1)$$

where p is the false positive rate in $(0, 1)$ as the corresponding cutoff point ranges from $-\infty$ to $+\infty$.

Estimation of ROC curves has been studied for decades. A variety of parametric and semiparametric methods have been developed to estimate the ROC curve $R(p)$. See, for example, Zweig and Campbell (1993) for an extensive review of parametric methods. To avoid misspecification problem, nonparametric techniques are used to estimate the ROC curve. The most commonly used nonparametric estimator for $R(p)$ is the empirical ROC curve. Hsieh and Turnbull (1996) showed that the empirical ROC curve estimator shares the same good asymptotic properties with standard ML estimators. See Zhou et al. (2002) for a comprehensive survey of the estimation methods of ROC curves. Consequently, development of confidence intervals of an ROC curve $R(p)$ has received attention because it is more important than point estimates and more useful for practitioners in making diagnostic decisions.

Linnet (1987) developed parametric and nonparametric methods to construct confidence intervals for the sensitivity of a test at a fixed value of specificity. Platt et al. (2000) pointed out several shortcomings in Linnet's methods and then suggested the use of Efron's bias-corrected acceleration (BCa) bootstrap interval. Zhou and Qin (2005) further showed that the BCa bootstrap interval can still have poor coverage accuracy in many circumstances, and proposed two new intervals for the sensitivity of a test at a fixed value of specificity. The proposed intervals may be regarded as extensions of Agresti and Coull's (1998) interval for a binomial proportion. Hall, Hyndman, and Fan (2004) proposed nonparametric confidence intervals for ROC curves.

The goal of this article is to develop an approach for construction of confidence intervals using the empirical likelihood ratio principle, which has been systematically studied and developed by Owen (1988, 1990), Qin (1994, 1999), and Qin and Lawless (1994). The empirical likelihood ratio has a limiting chi-squared distribution and can be used to obtain tests and CIs for a variety of problems, including linear models, generalized linear models, and estimating equations. The empirical likelihood method has many advantages over its counterparts like the normal-approximation-based method and the boot-strap method (see Hall and La Scala 1990). The most appealing features of the empirical likelihood method include improvement of the confidence region, an increase of accuracy of coverage because of using auxiliary information (Owen 2001), and easy implementation. The method has been applied in a variety of topics, for example, linear models (Owen 1991; Chen 1993, 1994), generalized linear models (Kolaczyk 1994), and the general estimating equation (Qin and Lawless 1994). We define an adjusted empirical function of the ROC function $R(p)$, derive the asymptotic distribution of the resulting empirical likelihood function, and explain how to establish the corresponding CI. In the next section we briefly introduce Zhou and Qin's (2005) bootstrap-based CIs. Section 3 presents our empirical likelihood-based CI and gives an asymptotic distribution of the empirical likelihood function. In Section 4 we report the results of the simulation studies. In Section 5 the approaches are used to establish CIs of ROC curves for two real datasets. Discussions are given in Section 6, and a technical proof on the asymptotic distribution is presented in Appendix A.

2. Bootstrap-Based Confidence Intervals

Let $X_1, \dots, X_n$ be the test responses of the diseased subjects and $Y_1, \dots, Y_m$ be the test responses of the nondiseased subjects. Let $\hat{F}_n(\cdot)$ and $\hat{G}_m(\cdot)$ be the empirical functions of $F(\cdot)$ and $G(\cdot)$, respectively. Then

$$\hat{F}_n(t) = \# \{X_i \leq t\} / n \text{ and } \hat{G}_m(t) = \# \{Y_i \leq t\} / m.$$

The inverse function of $\hat{G}_m(t)$ is defined as $\hat{G}_m^{-1}(p) = \inf \{t : \hat{G}_m(t) \geq p\}$. The empirical ROC curve is defined by

$$\hat{R}_n(p) = 1 - \hat{F}_n\{\hat{G}_m^{-1}(1 - p)\}.$$

The implementation of the confidence intervals developed by Zhou and Qin (2005) can be summarized as follows:

1. Draw a resample of size, m Y_i^* 's, with replacement from the diseased sample Y_i 's and a separate resample of size n , X_j^* 's, with replacement from the nondiseased sample X_j 's.
2. Calculate the bootstrap quantity

$$\tilde{R}^*(p) = \frac{\sum_{i=1}^n I_{\{X_i^* \geq \hat{G}_m^{-1*}(1-p)\}} + z_{1-\alpha/2}^2/2}{m + z_{1-\alpha/2}^2},$$

where $\hat{G}_m^{-1*}(\zeta)$ is the ζ th sample quantile based on the bootstrap resample Y_j^* 's.

3. Repeat the first two steps B times to obtain the set of bootstrap replications $\{\tilde{R}^{*b}(p); b = 1, \dots, B\}$, and then have the bootstrap variance $V^*(p)$ of $R^*(t)$.

Denote z_α the α -quantile of the standard normal distribution. The first $(1 - \alpha)$ 100% level interval, called the BTI interval by Zhou and Qin (2005), of $R(p)$ is defined by

$$\left(\tilde{R}(p) - z_{1-\alpha/2} \sqrt{V^*(p)}, \tilde{R}(p) + z_{1-\alpha/2} \sqrt{V^*(p)} \right),$$

where

$$\tilde{R}(p) = \frac{\sum_{i=1}^n I_{\{X_i \geq \hat{G}_m^{-1}(1-p)\}} + z_{1-\alpha/2}^2/2}{m + z_{1-\alpha/2}^2}.$$

The second $(1 - \alpha)$ 100% level interval, called the BTII interval by Zhou and Qin (2005), of $R(p)$ is defined by

$$\left(\tilde{R}^*(p) - z_{1-\alpha/2} \sqrt{V^*(p)}, \tilde{R}^*(p) + z_{1-\alpha/2} \sqrt{V^*(p)} \right),$$

where $\tilde{R}^*(p)$ is the mean of $\{\tilde{R}^{*b}(p); b = 1, \dots, B\}$.

3. Empirical Likelihood-Based Confidence Interval

In this section, we use the empirical likelihood principle for a development of confidence intervals (CI) of $R(p)$ to avoid estimating the covariance matrix and using normal approximation. The procedure in our context is different from that for usual situations. We propose an adjusted empirical likelihood function, then prove that the resulting log-likelihood ratio statistic is still asymptotically chi-squared. The empirical log-likelihood function is therefore able to be used to construct a confidence interval of $R(p)$.

Denote $\xi_i = I_{\{X_i \geq G^{-1}(1-p)\}} - R(p)$ for a given $p(0 < p < 1)$. Let F_ξ be the distribution function which assigns probability p_i at points ξ_i . Then $\sum_{i=1}^n p_i = 1$ and $p_i \geq 0$ for each i . Note that

$$E[I_{\{X_i \geq G^{-1}(1-p)\}} - R(p)] = 0.$$

If $G(\cdot)$ is known, the empirical likelihood ratio function for $R(p)$ is defined as

$\prod_{i=1}^n np_i$ subject to $\sum_{i=1}^n p_i \xi_i = 0$, $p_i \geq 0$, and $\sum_{i=1}^n p_i = 1$. Because $G(\cdot)$ is unknown, we replace it by its empirical distribution and the resulting adjusted empirical log-likelihood ratio statistic is defined as

$$\begin{aligned} \text{LLR}(p) &= \max \left\{ \sum_{i=1}^n \log(np_i) \mid \sum_{i=1}^n p_i \hat{\xi}_i \right. \\ &\quad \left. = 0, p_i \geq 0, \sum_{i=1}^n p_i = 1 \right\}, \end{aligned} \quad (2)$$

where $\hat{\xi}_i = I_{\{X_i \geq G_m^{-1}(1-p)\}} - R(p)$. Without the restriction conditions, the adjusted empirical likelihood has its maximum value at $p_i = 1/n$, that is, the empirical distribution.

Theorem 1

Suppose that $n/m \rightarrow k(0 < k < \infty)$, and that F and G have probability density functions f and g , and g is continuous at ψ with $f(\psi)g(\psi) > 0$. Then, as $m, n \rightarrow \infty$,

$$-2\text{LLR}(p) \xrightarrow{d} C_{\mathcal{L}\chi_1^2},$$

where $C_{\mathcal{L}} = 1 + \frac{k f^2(\psi) p(1-p)}{g^2(\psi) \theta(1-\theta)}$ with $\psi = G^{-1}(1-p)$ and $\theta = R(p)$.

The proof of Theorem 1 is given in Appendix A. Using the traditional density estimation (Wand and Jones 1995) by appropriately selecting bandwidth, we may obtain consistent estimators, $\hat{f}$ and $\hat{g}$, of f and g . By a substitution of θ, ψ, f , and g by their estimators, we obtain consistent estimators $\hat{C}_{\mathcal{L}}$ of $C_{\mathcal{L}}$. Under the conditions of Theorem 1, we still have

$-2\hat{C}_{\mathcal{L}}^{-1}\text{LLR}(p) \xrightarrow{d} \chi_1^2$. Based on these statements, a confidence interval of $R(p)$ can be derived by $\{R(p): -2\hat{C}_{\mathcal{L}}^{-1}\text{LLR}(p) \leq c_\alpha\}$, where c_α denotes the α quantile of the chi-squared distribution with one degree of freedom. See Owen (2001) for a detailed explanation for how to obtain the confidence interval.

In terms of bandwidth used in the estimation procedure, because the statistic converges to a standard chi-squared random times a constant, which depends on the density functions f and g , we need only a consistent estimator of the constant, and therefore use a traditional density estimation process in our simulation experiments and real data analysis; that is, we use the “optimal” band-width selection (Wand and Jones 1995) for estimating the functions f and g . Note that the biases of these estimators are of an order h^2 , where h is the bandwidth used for estimating density functions f and g . The estimation of $C_{\mathcal{X}}$ will not affect the limiting distribution in a large sample sense. In our simulation studies and real examples, we find that the numerical performances are stable against the different values of the bandwidths, which we shift around the selected ones.

4. Simulation Study

In this section, we conduct extensive simulation studies for an illustration of finite-sample performance of the proposed method and for a comparison with the bootstrap-based alternatives.

Assume that the diseased distribution $F(\cdot)$ and nondiseased distribution $G(\cdot)$ follow one of three pairs: (a) $F \sim W_{(2,3)}$, $G(\cdot) \sim F_{(5,4)}$; (b) $F \sim F_{(7,5)}$, $G(\cdot) \sim F_{(5,4)}$; and (c) $F \sim \text{Normal}(1, 1)$, $G(\cdot) \sim \text{Normal}(0, 1)$. Here $W_{(\alpha,\beta)}$ denotes a Weibull distribution with a cumulative distribution function $1 - \exp(-\alpha x^\beta)$ ($x \geq 0$); $F_{(\alpha,\beta)}$ denotes an F-distribution with α and β degrees of freedom.

For each pair, we consider a combination of $n = 30, 40, 50, 60$ and $m = 30, 40, 50, 60, 70$. For each combination, we generate 200 datasets of size (n, m) and calculate the empirical likelihood-based confidence intervals (EL) and the bootstrap I/II-based confidence intervals (BTI/BTII) of the level 95% at $p = 0.5, 0.7$, and 0.9 for each random sample. Five hundred bootstrap samples are generated from each of the 200 datasets for BTI and BTII. A total of 180 results are obtained. Table 1 gives the results for pair (a) with $p = 0.7$, Table 2 gives the results for pair (b) with $p = 0.5$, and Table 3 gives the results for pair (c) with $p = 0.3$. We summarize our findings as follows. The coverage probabilities of the empirical likelihood-based confidence interval are almost closer to the nominal level than those of the two corresponding bootstrap confidence intervals. When their coverage probabilities are identical, the lengths of the empirical likelihood-based confidence interval are almost shorter than those of the two corresponding bootstrap-based confidence intervals.

5. Real Data Example

In this section we apply the method presented in Section 3 to analyze two real examples. One is on assessing the accuracy of cerebrospinal fluid CK-BB for predicting poor outcomes among patients with severe head trauma (Zhou et al. 2002), and the other one is on the accuracy of a carbohydrate antigenic determinant (CA19-9) in detecting pancreatic cancer (Wieand et al. 1989). Let $\hat{F}_n$ and $\hat{G}_m$ be the estimates of the cumulative distribution function (CDF) of a nondiseased subject and that of a diseased subject, respectively. The ROC curve is estimated by $\hat{R}_n(p) = 1 - \hat{F}_n\{\hat{G}_m^{-1}(1 - p)\}$. For the bootstrap procedure we generate 1,000 bootstrapping samples. The results from the two examples give us shorter empirical likelihood-based CIs than the associated bootstrap-based CIs, and provide a more accurate measurement to distinguish the poor and good outcomes.

5.1 Severe Head Trauma Example

In the CK-BB study, Hans et al. (1985) conducted a study on assessing the accuracy of the cerebrospinal fluid CK-BB isoenzyme measured within 24 hours of injury as a means of predicting the outcome of severe head trauma. Investigators were interested in determining whether patients have a poor outcome (death, vegetative state, or severe disability) after

suffering a severe head trauma. A sample of 60 subjects admitted to a hospital with severe head trauma were considered, with 19 eventually having moderate to full recovery, and the remaining 41 poor or no recovery (Hans et al. 1985). Three confidence intervals of the ROC curve for this study are shown in Figure 1, in which the dashed, dash-dot, and dotted lines represent the empirical likelihood-based, BTII, and BTI CIs. The estimated values $\hat{R}_n(p)$ (solid line) are also presented there. It is readily seen that the empirical likelihood-based CI is superior to BTII and BTI CIs. The length of our CI is shorter than those of the alternatives at almost all points.

5.2 Pancreatic Cancer Example

Wieand et al. (1989) reported a study on the accuracy of a carbohydrate antigenic determinant (CA19-9) in detecting pancreatic cancer. In this study, we are interested in estimating the accuracy of a carbohydrate antigenic determinant (CA19-9) in detecting pancreatic cancer. Concentrations of CA19-9 in sera (mL) from 51 control patients with pancreatitis and 90 cases of pancreatic cancer were collected. As in the previous example, we report three confidence intervals of the ROC curve for this study in Figure 2. Although the difference between our CI and BTI/BTII CI is not so big as in the previous example, the superiority of our CI is still visible.

6. Discussion

In this article, we developed an empirical likelihood method to establish a confidence interval for the ROC curve. The procedure is easy to understand and simple to implement. The principal advantage of the empirical likelihood method is that it avoids estimating a variance and using normal approximation which one encounters when using other techniques, but the normal approximation may not at times be true in practice. The performance of the proposed approach is slightly superior to that of the bootstrap-based confidence intervals. This increased level of performance and availability of easy implementation were illustrated by simulation experiments.

Claeskens et al. (2003) developed an advanced approach of calculating confidence intervals for ROC curves by proposing a smoothed empirical likelihood method. Two concerns with their approach are that band-widths have to be selected and the computation of their method may be very expensive, which may not be feasible in practice unless an effective computation package is developed.

For comparing ROC curves, various nonparametric approaches have been proposed (Qin and Zhou 2006; Li and Zhou 2008; Tang and Zhou 2009). The most commonly used approach is the area under the ROC curve (AUC), for which Qin and Zhou (2006) developed an empirical likelihood inference and derived the asymptotic distribution of their empirical likelihood ratio, a sum of the independent chi-squared distributions with unknown weights. Extension of our approach to the comparison of ROC curves is by no means straightforward and needs further investigation.

A. Appendix

To prove the main result, we need the following Lemma.

Lemma A.1

Under the conditions of Theorem 1, as $m, n \rightarrow \infty$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\xi}_i \xrightarrow{d} N \{0, \theta(1-\theta) + kf^2(\psi)p(1-p)/g^2(\psi)\}, \quad (\text{A.1})$$

$$\frac{1}{n} \sum_{i=1}^n \widehat{\xi}_i^2 = \theta(1-\theta) + o_p(1), \quad (\text{A.2})$$

$$\max_{1 \leq i \leq n} |\widehat{\xi}_i| = O_p(1). \quad (\text{A.3})$$

Proof

Denote $\widehat{\psi} = \widehat{G}_m^{-1}(1-p)$ and

$$U_n(\psi) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \geq \psi\}}.$$

From Bahadur's representation of sample quantile, we know that

$$\begin{aligned} \sqrt{m}(\widehat{\psi} - \psi) &= -\{1/g(\psi)\} m^{-1/2} \sum_{j=1}^m \\ &\quad \times \{I_{\{Y_j < \psi\}} - (1-p)\} + o_p(1). \end{aligned}$$

Thus,

$$n^{1/2}\{U_n(\psi) - \theta, \widehat{\psi} - \psi\} \xrightarrow{d} N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \theta(1-\theta) & 0 \\ 0 & kp(1-p)/g^2(\psi) \end{pmatrix}.$$

Thus, from Theorem 2.13 of Randles (1982), we have (A.1). The proofs of (A.2) and (A.3) are trivial and omitted.

Proof of Theorem 1

A standard simplification as in Owen (2001) yields that

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda \widehat{\xi}_i} \text{ for } i=1, \dots, n, \quad (\text{A.4})$$

where λ is the solution of the equation

$$\sum_{i=1}^n \frac{\widehat{\xi}_i}{1+\lambda\widehat{\xi}_i} = 0. \quad (\text{A.5})$$

Mimicking the proof of Theorem 3.2 from Owen (2001), we have

$$|\lambda| = O_p(n^{-1/2}). \quad (\text{A.6})$$

It follows from $\sum_{i=1}^n p_i = 1$ and (A.5) that

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{\widehat{\xi}_i}{1+\lambda\widehat{\xi}_i} &= \frac{1}{n} \sum_{i=1}^n \widehat{\xi}_i (1 - \lambda\widehat{\xi}_i) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{(\lambda\widehat{\xi}_i)^2 \widehat{\xi}_i}{1+\lambda\widehat{\xi}_i} + o_p(n^{-1/2}). \end{aligned}$$

The second term is $o_p(n^{-1/2})$ by (A.6) and (A.3). It means that

$$\lambda = \left(\sum_{i=1}^n \widehat{\xi}_i^2 \right)^{-1} \sum_{i=1}^n \widehat{\xi}_i + o_p(n^{-1/2}). \quad (\text{A.7})$$

A similar argument yields that

$$\begin{aligned} 0 &= \frac{1}{n} \sum_{i=1}^n \frac{\lambda\widehat{\xi}_i}{1+\lambda\widehat{\xi}_i} \\ &= \frac{1}{n} \sum_{i=1}^n \lambda\widehat{\xi}_i - \frac{1}{n} \sum_{i=1}^n (\lambda\widehat{\xi}_i)^2 + o_p(1/n). \end{aligned}$$

It follows that

$$\sum_{i=1}^n \lambda\widehat{\xi}_i = \sum_{i=1}^n (\lambda\widehat{\xi}_i)^2 + o_p(1). \quad (\text{A.8})$$

Consider LLR (p). Using Taylor expansion, we have

$$\begin{aligned} -\text{LLR}(p) &= \sum_{i=1}^n \log(1+\lambda\widehat{\xi}_i) \\ &= \sum_{i=1}^n \left\{ \lambda\widehat{\xi}_i - \frac{1}{2}(\lambda\widehat{\xi}_i)^2 \right\} + Q_n. \end{aligned}$$

The remainder term Q_n is bounded by

$$|n|^3 \max_{1 \leq i \leq n} |\widehat{\xi}_i| \sum_{i=1}^n |\widehat{\xi}_i|^2 = o_p(1).$$

Combining this with (A.8) and (A.7) implies that

$$-2\text{LLR}(p) = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{\xi}_i \right)^2 \left(\frac{1}{n} \sum_{i=1}^n \widehat{\xi}_i^2 \right)^{-1} + o_p(1).$$

The proof of Theorem follows from (A.1) and (A.2).

Acknowledgments

The authors thank one AE and two referees for their constructive comments and useful suggestions. This research was supported by NI-AID/NIH grants AI62247 and AI059773 and NSF grant DMS806097. Su's research was also supported by a Merck Quantitative Sciences Fellowship grant provided by the Merck Company Foundation.

REFERENCES

- Agresti A, Coull BA. Approximate is Better than Exact for Interval Estimation of Binomial Proportions. *The American Statistician* 1998;52:119–126.
- Chen SX. On the Accuracy of Empirical Likelihood Confidence Regions for Linear Regression Model. *Annals of the Institute of Statistical Mathematics* 1993;45:621–637.
- Chen SX. Empirical Likelihood Confidence Intervals for Linear Regression Coefficients. *Journal of Multivariate Analysis* 1994;49:24–40.
- Claeskens G, Jing BY, Peng L, Zhou W. An Empirical Likelihood Confidence Interval for an ROC Curve. *The Canadian Journal of Statistics* 2003;31:173–190.
- Hall P, La Scala B. Methodology and Algorithms of Empirical Likelihood. *International Statistical Review* 1990;58:109–127.
- Hall P, Hyndman RJ, Fan Y. Nonparametric Confidence Intervals for Receiver Operating Characteristic Curves. *Biometrika* 2004;91:743–750.
- Hans P, Albert A, Born JD, Chapelle JP. Derivation of a Bioclinical Index in Severe Head Trauma. *Intensive Care Medicine* 1985;11:186–191. [PubMed: 4044994]
- Hsieh F, Turnbull BW. Nonparametric and Semiparametric Estimation of the Receiver Operating Characteristic Curve. *Annals of Statistics* 1996;24:25–40.
- Kolaczyk ED. Empirical Likelihood for Generalized Linear Models. *Statistica Sinica* 1994;4:199–218.
- Li G, Zhou KF. A Unified Approach to Nonparametric Comparison of Receiver Operating Characteristic Curves for Longitudinal and Clustered Data. *Journal of the American Statistical Association* 2008;103:705–713. [PubMed: 20209021]
- Linnet K. Comparison of Quantitative Diagnostic Tests: Type I Error, Power, and Sample Size. *Statistics in Medicine* 1987;6:147–158. [PubMed: 3589244]
- Owen AB. Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika* 1988;75:237–249.
- Owen AB. Empirical Likelihood Confidence Regions. *Annals of Statistics* 1990;18:90–120.
- Owen AB. Empirical Likelihood for Linear Models. *Annals of Statistics* 1991;19:1725–1747.
- Owen, AB. *Empirical Likelihood*. London: Chapman and Hall/CRC; 2001.
- Platt RW, Hanley JA, Yang H. Bootstrap Confidence Intervals for the Sensitivity of a Quantitative Diagnostic Test. *Statistics in Medicine* 2000;19:313–322. [PubMed: 10649298]
- Qin GS, Zhou XH. Empirical Likelihood Inference for the Area Under the ROC Curve. *Biometrics* 2006;62:613–622. [PubMed: 16918927]

- Qin J. Semi-Empirical Likelihood Ratio Confidence Intervals for the Difference of Two Sample Means. *Annals of Statistics* 1994;46:117–126.
- Qin J. Empirical Likelihood Ratio Based Confidence Intervals for Mixture Proportions. *Annals of Statistics* 1999;27:1368–1384.
- Qin J, Lawless J. Empirical Likelihood and General Estimating Equations. *Annals of Statistics* 1994;22:300–325.
- Qin Y, Zhao L. Empirical Likelihood Ratio Intervals for the Quantile Differences of Two Populations. *Chinese Annals of Mathematics* 1997;18A:687–694.
- Randles RH. On the Asymptotic Normality of Statistics with Estimated Parameters. *Annals of Statistics* 1982;10:462–474.
- Tang LS, Zhou XH. Semiparametric Inferential Procedures for Comparing Multivariate ROC Curves with Interaction Terms. *Statistica Sinica*. 2009 in press.
- Shapiro DE. The Interpretation of Diagnostic Tests. *Statistical Methods in Medical Research* 1998;8:113–134. [PubMed: 10501649]
- Wand, MP.; Jones, MC. Kernel Smoothing. London: Chapman and Hall; 1995.
- Wieand S, Gail MH, James BR, James KL. A Family of Nonparametric Statistics for Comparing Diagnostic Markers with Paired and Unpaired Data. *Biometrika* 1989;76:585–592.
- Zweig MH, Campbell G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry* 1993;39:561–577. [PubMed: 8472349]
- Zhou, XH.; Obuchowski, NA.; McClish, DK. Statistical Methods in Diagnostic Medicine. New York: Wiley; 2002.
- Zhou XH, Qin GS. Improved Confidence Intervals for the Sensitivity at a Fixed Level of Specificity of a Continuous-Scale Diagnostic Test. *Statistics in Medicine* 2005;24:465–477. [PubMed: 15635678]

The CIs of the ROC Curve for Head-trauma Data Set

Figure 1.

The confidence intervals for Trauma data. Solid line represents the estimates of the ROC curve. The dashed, dash-dot, and dotted lines are the empirical likelihood-based, BTII, and BTI CIs.

The CIs of the ROC Curve for Mayo Clinic Cancer Dataset

Figure 2.

The confidence intervals for Mayo clinic data. Solid line represents the estimates of the ROC curve. The dashed, dash-dot, and dotted lines are the empirical likelihood-based, BTII, and BTI CIs.

Table 1

The length and coverage probabilities (cov) of the confidence intervals based on the empirical likelihood (EL) and bootstrap I (BTI) and II (BTII) methods for pair (a) and $p = 0.7$.

		EL		BTI		BTII	
<i>n</i>	<i>m</i>	length	cov	length	cov	length	cov
30	30	0.439	0.945	0.432	0.915	0.432	0.925
30	40	0.396	0.960	0.392	0.940	0.392	0.945
30	50	0.384	0.965	0.392	0.955	0.392	0.940
30	60	0.367	0.945	0.370	0.940	0.370	0.935
30	70	0.356	0.955	0.366	0.950	0.366	0.970
40	30	0.419	0.955	0.428	0.950	0.428	0.975
40	40	0.397	0.925	0.396	0.910	0.396	0.920
40	50	0.355	0.965	0.373	0.955	0.373	0.965
40	60	0.346	0.955	0.368	0.950	0.368	0.965
40	70	0.327	0.945	0.348	0.955	0.348	0.975
50	30	0.421	0.970	0.421	0.925	0.421	0.945
50	40	0.374	0.910	0.374	0.910	0.374	0.925
50	50	0.344	0.940	0.357	0.935	0.357	0.935
50	60	0.320	0.950	0.330	0.960	0.330	0.960
50	70	0.312	0.965	0.320	0.965	0.320	0.985
60	30	0.415	0.965	0.416	0.930	0.416	0.955
60	40	0.383	0.945	0.377	0.945	0.377	0.955
60	50	0.341	0.965	0.352	0.945	0.352	0.965
60	60	0.319	0.910	0.334	0.915	0.334	0.945
60	70	0.308	0.945	0.320	0.940	0.320	0.950

Table 2

The length and coverage probabilities (cov) of the confidence intervals based on the empirical likelihood (EL) and bootstrap I (BTI) and II (BTII) methods for pair (b) and $p = 0.5$.

EL			BTI		BTII		
<i>n</i>	<i>m</i>	length	cov	length	cov	length	cov
30	30	0.289	0.973	0.318	0.979	0.318	0.984
30	40	0.254	0.979	0.269	0.979	0.269	0.984
30	50	0.253	0.979	0.270	0.964	0.270	0.979
30	60	0.242	0.957	0.248	0.952	0.248	0.957
30	70	0.237	0.989	0.239	0.963	0.239	0.968
40	30	0.250	0.953	0.283	0.990	0.283	0.979
40	40	0.233	0.944	0.258	0.959	0.258	0.975
40	50	0.223	0.959	0.247	0.959	0.247	0.969
40	60	0.221	0.969	0.239	0.969	0.239	0.969
40	70	0.214	0.960	0.226	0.970	0.226	0.975
50	30	0.227	0.970	0.271	0.980	0.271	0.985
50	40	0.205	0.955	0.230	0.970	0.230	0.980
50	50	0.199	0.939	0.228	0.955	0.228	0.955
50	60	0.193	0.939	0.218	0.949	0.218	0.959
50	70	0.202	0.975	0.221	0.975	0.221	0.980
60	30	0.221	0.929	0.265	0.955	0.265	0.965
60	40	0.205	0.923	0.241	0.974	0.241	0.979
60	50	0.192	0.965	0.225	0.985	0.225	0.980
60	60	0.187	0.940	0.216	0.970	0.216	0.990
60	70	0.178	0.940	0.201	0.940	0.201	0.960

Table 3

The length and coverage probabilities (cov) of the confidence intervals based on the empirical likelihood (EL) and bootstrap I (BTI) and II (BTII) methods for pair (c) and $p = 0.3$.

EL			BTI		BTII		
<i>n</i>	<i>m</i>	length	cov	length	cov	length	cov
30	30	0.413	0.970	0.414	0.935	0.414	0.970
30	40	0.396	0.950	0.398	0.950	0.398	0.955
30	50	0.388	0.940	0.387	0.905	0.387	0.925
30	60	0.385	0.960	0.389	0.940	0.389	0.950
30	70	0.368	0.955	0.368	0.940	0.368	0.960
40	30	0.392	0.935	0.395	0.925	0.395	0.950
40	40	0.383	0.945	0.373	0.940	0.373	0.945
40	50	0.370	0.950	0.369	0.935	0.369	0.950
40	60	0.347	0.955	0.345	0.925	0.345	0.950
40	70	0.343	0.952	0.346	0.940	0.346	0.970
50	30	0.373	0.960	0.378	0.950	0.378	0.965
50	40	0.352	0.945	0.353	0.940	0.353	0.955
50	50	0.337	0.945	0.339	0.940	0.339	0.965
50	60	0.317	0.945	0.319	0.930	0.319	0.940
50	70	0.321	0.945	0.327	0.950	0.327	0.965
60	30	0.385	0.950	0.364	0.945	0.364	0.950
60	40	0.344	0.945	0.343	0.925	0.343	0.955
60	50	0.338	0.950	0.330	0.940	0.330	0.940
60	60	0.321	0.945	0.318	0.920	0.318	0.935
60	70	0.320	0.950	0.311	0.920	0.311	0.930