Confidence Intervals for AUC and pAUC by Empirical Likelihood

Yumin Zhao[*], Xue Ding[†], Mai Zhou[‡]

Abstract

The Receiver Operating Characteristic (ROC) Curve is often used to evaluate the performance of medical diagnostic tests. The Area Under the ROC Curve (AUC) is a one-number summary of the diagnostic performance. Sometimes the clinical interest is focused primarily on part of the ROC Curve, and in this case the partial AUC (pAUC) has been proposed by many authors to measure the performance of the diagnostic test. Nonparametric estimators of AUC and pAUC are available when samples of test results from diseased and healthy subjects are collected. We propose and illustrate in this paper a novel empirical likelihood approach to test hypothesis and construct confidence intervals for AUC and pAUC. The empirical likelihood ratio test in our setup yields an asymptotic chi square distribution under null hypothesis. Thus, there is no need to estimate the complicated scale factor or the variance of the nonparametric AUC/pAUC estimators like most other competing methods do. Computation of the proposed empirical likelihood is also studied. Simulations show our method is very competitive. In fact our method tops competitors in every situations we simulated. Real data example (with R code) is presented illustrating the confidence intervals for AUC and pAUC.

Keywords and Phrases: Chi square distribution; Two sample empirical likelihood ratio; Partial AUC; ROC curve; Wilks confidence intervals, Nuisance parameter.

1 Introduction

The purpose of diagnostic tests is to confirm the presence of disease in subjects with disease and to deny the possibility of the disease in healthy subjects. Ideally such tests should correctly identify all patients with the disease (True Positive), and similarly correctly identify all patients who are disease free (True Negative). However, a perfect test is hardly found in reality.

[*]Eli Lilly and Company, Indianapolis, IN 46225, U.S.A. Email: zyyztwo@gmail.com

[†]Dr. Bing Zhang Department of Statistics, University of Kentucky, Lexington, KY 40536, U.S.A.

[‡]Dr. Bing Zhang Department of Statistics, University of Kentucky, Lexington, KY 40536, U.S.A. Email: maizhou@gmail.com

For diagnostic tests which report numeric values on a continuous scale, one may chose a threshold value c such that the values above c will be classified as positive, otherwise negative. The sensitivity and specificity, which are defined as the probabilities of the test correctly identifying the diseased and non-diseased subjects, respectively, can be computed across all possible threshold values c for a test.

The plot of the sensitivity or the true positive rate (TPR) versus 1−specificity or the false positive rate (FPR) as the threshold value for the test vary (from $-\infty$ to ∞) is the receiver operating characteristic (ROC) curve. The sample version of the ROC curve plot would replace those positive rates by their corresponding sample fractions.

The ROC curve was first developed during World War II for detecting enemy objects and soon found other uses in psychology, medicine, radiology, biometrics, and is increasingly applied in machine learning, data mining and artificial intelligence researches. Numerous papers are written on various aspects of this methodology. Several recent books are published with extensive discussions on topics related to the ROC/AUC analysis and more: see [17], [30], [32], [12]. Computation and plots related to ROC analysis are available in the R packages `ROCR` and `pROC`.

1.1 Notation, Definition and Estimations

Let X and Y, with respective distribution functions F and G, be the results of a continuous-scale test for a non-diseased and a diseased subject, respectively.

For a given value c, without loss of generality, we assume that a test value greater than c is indicative of a positive test result. *Sensitivity* or true positive rate (TPR) is defined as $TPR = Pr(Y > c) = 1 - G(c)$. One minus *Specificity* or false positive rate (FPR) is defined as $FPR = Pr(X > c) = 1 - F(c)$. The ROC curve is an $x\,y$ plot of $\{1 - F(c), 1 - G(c)\}$ for all possible c.

The area under an ROC curve (AUC) represents the overall accuracy of a diagnostic test, which can be interpreted as the probability in a randomly selected pair of diseased and non-diseased subjects, the test value of the diseased subject is higher than that of the non-diseased subject (Hanley and McNeil 1982) [7]. A perfect test has an AUC equal to 1.0. A test that is just a random guess has an AUC value of 0.5. A test with an AUC value approaching 1.0 indicates a high sensitivity and specificity.

The AUC of the test results X and Y of the above non-diseased and diseased subjects can be represented by (Hanley and McNeil 1982) [7]:

$$AUC = \int_{-\infty}^{\infty} (1 - G(s))dF(s) = Pr(Y > X)\ . \qquad (1)$$

Given a random sample $X_1, \cdots, X_m$ of test results from non-disease population and independently another random sample $Y_1, \cdots, Y_n$ of test results from the disease population, a non-parametric estimate of AUC is

$$\widehat{AUC} = \frac{1}{nm}\sum_{i=1}^{m}\sum_{j=1}^{n} I[Y_j > X_i] + 0.5 I[Y_j = X_i]\ . \tag{2}$$

See for example: [6] or any of the above mentioned books.

However, the AUC of an ROC is a measure of the overall performance of a diagnostic test. It may not be informative, or even misleading in some cases. For example, two diagnostic tests may have the same AUC but not identical ROCs when the two ROC curves cross each other. One test may be better than the other in the high FPR range; while the other test may be better in the low FPR range. Thus to evaluate two diagnostic tests on a portion of ROC curves, the partial AUC (which is defined as the area under a ROC curve between two FPRs (McClish 1989) [13]), is desirable.

The true value of a partial AUC when FPR $\in (p_1, p_2)$, with $0 < p_1 < p_2 < 1$, can be written as

$$pAUC(p_1, p_2) = \int_{t_1}^{t_2} (1 - G(s)) dF(s) \tag{3}$$

where $t_2 = F^{-1}(1 - p_1)$ and $t_1 = F^{-1}(1 - p_2)$. For the special case of pAUC$(0, p)$ this is

$$pAUC(0, p) = \int_{\tau}^{\infty} (1 - G(s)) dF(s) \tag{4}$$

where $\tau = F^{-1}(1 - p)$. While our proposed empirical likelihood approach will work for both cases specified above, for the ease of presentation we shall focus on the analysis for pAUC of the type pAUC$(0, p)$ in the rest of this paper, except in the Example 1 below where we gave an empirical likelihood confidence interval for pAUC(0.2, 0.7).

The partial AUC of ROC curves has also been studied by many researchers. McClish (1989) [13] first proposed the pAUC assuming binormal data. We shall focus on the non-parametric method in this paper. Dodd and Pepe (2003) [6] proposed the non-parametric estimator for the pAUC$(0, p)$:

$$\widehat{pAUC}(0, p) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} (I[Y_j > X_i] + 0.5 I[Y_j = X_i])\, I[X_i > \tau] \tag{5}$$

where τ is the $(1 - p)th$ quantile of X, or equivalently, $\tau = F^{-1}(1 - p)$.

If the quantile τ in (5) is not known, as is usually the case, Dodd and Pepe (2003) [6] suggested that an empirical quantile estimate be substituted: $\hat{\tau} = \hat{F}^{-1}(1-p) = \inf\{s : \hat{F}(s) \geq 1 - p\}$; where $\hat{F}$ is an empirical distribution based on $X_1, \cdots, X_m$. Dodd and Pepe (2003) [6] also used some linear interpolation technique to improve the empirical quantile estimate. This is equivalent to smoothing, which we shall also discuss in section 2.

Due to the vast amount of research literature on the topic, a complete review of existing statistical inference methods for AUC/pAUC is unpractical here. But they roughly falls into 3 categories.

- 1. Procedures based on the asymptotic normality of the AUC/pAUC estimators like (2) and (5). We shall refer to them as the Wald method. A typical procedure will first obtain an estimate, say $\hat{\sigma}^2$, of the variance of $\widehat{AUC}$, then the p-value of the test $H_0 : AUC = \theta$ vs. $H_A : AUC \neq \theta$, for example, can be computed as $2Pr(Z > |\widehat{AUC} - \theta|/\hat{\sigma})$ where Z denotes a $N(0, 1)$ random variable. A Wald type confidence interval for AUC can be computed by $\widehat{AUC} \pm 1.96\hat{\sigma}$. Wald type statistical analysis for the pAUC can be similarly obtained, except the variance estimator of $\widehat{pAUC}$ is more complicated. Example in this category include Wieand et. al (1989) [24], DeLong [5], Bandos et. al [2], He and Escobar [9], Yang (2017) [26] and those in the R package `pROC`. In this approach, the variance estimator is the main difficult. Conventional plug-in method as well as Jackknife method of estimating the variance were proposed and investigated.

- 2. Procedures based on re-sampling techniques. We shall refer to them as Bootstrap method. Some of those are available in the package `pROC`. The jackknife method of estimating variance mentioned above, however are classified here as Wald.

- 3. Adhoc pseudo empirical likelihood method or hybrid method, since they often need to combine with some re-sampling methods to help estimate scale/variance/pseudo values. We shall do a slightly more detailed review since this is closer to our proposed method.

  Qin, Jin, and Zhou (2011) [18] and Qin and Zhou (2006) [19] applied a pseudo empirical likelihood ratio method to the above AUC and pAUC estimator. Instead of using the two samples of observations to construct an empirical likelihood (both diseased and non-diseased samples), they constructed the empirical likelihood for AUC/pAUC using only the non-diseased sample. The diseased sample only enters the calculation as a plug-in estimator for $1-G(\cdot)$. In addition, plug-in sample quantiles of non-diseased population in the estimator (5) were used. This is called plug-in empirical likelihood method by Hjort, McKeague and Keilegom (2009) [10]. They concluded that the limiting distribution of the test statistic was a scaled chi-square under the null hypothesis.

  Qin, Jin and Zhou (2011) [18] proposed a very complex bootstrap procedure to estimate the scale constant. Adimari and Chiogna (2012) [1] combined jackknife with empirical likelihood to test the pAUC. Yu, et al. (2016) [27] formulated a generalized empirical likelihood test for AUC utilizing both samples, where they incorporated the variance of AUC estimate to the empirical likelihood test statistic. Yang et al. (2017) [26] propose to first use Jackknife to create some pseudo values, then construct a (one-sample) empirical likelihood based on the pseudo values for statistical inference of pAUC.

  One common theme in these approaches is that they all need some type of variance estimator: either in the estimation of the scale factor or pre-process the data with a jackknife. (an exception is the bootstrap percentile method). In contrast, the original empirical likelihood method of Owen (1988) [15] does not involve a variance estimator at all.

In this paper we propose and study a method based on a truly two sample empirical likelihood. The limiting distribution of our empirical likelihood ratio, under null hypothesis, is chi-square without any scale adjustment, thus no variance estimation is needed (nor for the scale factor). Our empirical likelihood is based on the two sample original observations, not jackknife pseudo values.

Furthermore, in the inference for pAUC, instead of plug in an estimator for unknown $(1-p)th$ quantile of X in (5), we used a nuisance parameter/profile trick to avoid explicitly estimating this nuisance parameter. More specifically, we first artificially introduce a nuisance parameter in addition to pAUC. This allows us to easily formulate empirical likelihood for the two parameters jointly and then later apply the profiling technique to arrive the so called profile empirical likelihood for pAUC alone. And a clean limiting chi-square distribution under null hypothesis of $H_0 : pAUC = \theta$ is again ensured.

Indeed, our empirical likelihood is in essence the empirical likelihood for the two sample (generalized) U-statistic. And the chi square limit is ensured as long as $\min(m, n) \to \infty$ and no need to require $m/(m+n) \to \lambda \in (0, 1)$. In addition, the confidence interval of pAUC(p_1, p_2) can be similarly obtained as for pAUC$(0, p)$, by perform one more profiling/minimization operation.

The advantage of Wilks confidence intervals over the Wald is a well known phenomena discussed by many authors, see for example Meeker and Escobar, (1995) [14]. We gave a list in the Discussion section later in this paper. Advantage of empirical likelihood confidence region over pseudo empirical likelihood was discussed in Kim and Zhou (2019) [11].

2 Smoothing

In the previous section, the estimators of AUC, (2), and estimator of pAUC, (5), were defined using the indicator function, and we treated the $[Y_j > X_i]$ and $[Y_j = X_i]$ case separately. If we replace the indicator function by a smoothed version, we can handle these $[Y_j > X_i]$ and $[Y_j = X_i]$ cases with one function. Smoothing the ROC curve is also proposed in Zou, Hall, and Shapiro (1997) [33] among many others.

We hereby specify a typical smoothed indicator function. Since this function is going to replace the indicator function $I[y > x]$, we shall call it $I_\epsilon(y, x)$ where the bandwidth parameter $\epsilon > 0$ controls the degree of smoothing. When $\epsilon \to 0$, the function $I_\epsilon(y, x)$ becomes the original indicator function (except when $x = y$).

$$I_\epsilon(y,x) = \begin{cases} 1\,, & \text{if } (x-y) < -\epsilon\ ; \\ 0.5 - \frac{3(x-y)}{4\epsilon} + \frac{(x-y)^3}{4\epsilon^3}\,, & \text{if } -\epsilon \le (x-y) \le \epsilon\ ; \\ 0\,; & \text{if } (x-y) > \epsilon\ . \end{cases} \tag{6}$$

Other smoothing functions are also possible and should lead to similar results. We pick this function because it is second order smooth and also fast to compute. Notice when $y = x$ we have $I_\epsilon(y, x) = 0.5$ always; and when $x < y - \epsilon$, we have $I_\epsilon(y, x) = 1$ etc.

In addition, the estimator of pAUC, (5), involves the quantile $\tau = F^{-1}(1-p)$. As mentioned earlier, Dodd and Pepe (2003) [6] have used some smoothing when estimate the quantile in the course of estimate the pAUC. Also, using empirical likelihood for testing hypothesis involve quantiles was investigated by Chen and Hall (1993) [4]. One take away message from Chen and Hall paper is that the sample quantile function needs to be smoothed. The smoothing makes the empirical likelihood ratio converge faster to the limiting chi square distribution (thus the empirical likelihood ratio test is more accurate). Indeed smoothing is almost a must whenever sample quantile is involved in any statistical analysis.

After a transformation, the defining equation for quantile τ becomes $F(\tau) = 1 - p$ which can be written as

$$\mathbf{E} I[X \leq \tau] = 1 - p \,. \tag{7}$$

We recall $F(\cdot)$ is the unknown distribution function for X_i.

Chen and Hall (1993) [4] have carefully analyzed the error rates. The recommended choice of smoothing bandwidth ξ should be $> (1/m)\log m$ and $O(m^{-1/2})$ for best convergence rate and $O(m^{-3/4})$ if we also require Bartlett corrections, with m the sample size. We shall use a bandwidth $m^{-3/4}$ in our example and simulation. For details please see their paper.

We therefore shall also smooth the indicator function in (7) by $I_\xi(\cdot,\cdot)$ and replace (7) by

$$\mathbf{E} I_\xi(\tau, X) = 1 - p \,. \tag{8}$$

Here the bandwidth ξ shall be chosen following Chen and Hall's recommendation. The bandwidth ϵ in (9) and (10) may follow other guidelines.

To summarize: we shall use the smoothed estimator of AUC

$$\widehat{AUC_\epsilon} = \frac{1}{nm} \sum_{i=1}^{m} \sum_{j=1}^{n} I_\epsilon(Y_j, X_i) \,. \tag{9}$$

A smoothed estimator of pAUC can be similarly defined to (5):

$$\widehat{pAUC_{\epsilon,\xi}}(0, p) = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I_\epsilon(Y_j, X_i)\, I_\xi(X_i, \tau) \tag{10}$$

where the quantile τ is now defined by (8).

Finally, smoothing also brings a computational convenience for our profile calculation in the next section, since a smoother function is easier to optimize.

3 Two Sample Empirical Likelihood

Owen (1988) [15] was first to coin the term "empirical likelihood" and made many contributions to the theory and practice of empirical likelihood method. The first book on this fascinating topic is also by Owen in 2001 [16]. One of the empirical likelihood theorems contained in this book is the *two-sample empirical likelihood theorem* (his section 11.4). We formulate below a version specific for the inference of AUC (h function in (i)) and the joint inference of pAUC and τ ($\boldsymbol{h}$ function in (ii)) below.

Theorem 1 (Two-Sample Empirical Likelihood Theorem) Suppose $X_1, \cdots, X_m$ are iid random variables with distribution $F(t)$. We further suppose, independent of the X's, that $Y_1, \cdots, Y_n$ are iid random variables with distribution $G(t)$. Let the true parameter $\boldsymbol{\theta} \in R^r$ be defined by the equation $\mathbf{E}\boldsymbol{h}(X, Y, \boldsymbol{\theta}) = \mathbf{0}$, where $\boldsymbol{h}$ is a function with values in R^r specified either in (i) or (ii) below.

(i) *For the inference of AUC*: here $r = 1$ and $\theta = AUC$; and $h(X, Y, \theta) = \{I[Y > X] + 0.5I[Y = X]\} - \theta$. We further assume $Pr(Y \geq X) \neq 1$.

(ii) *For the joint inference of pAUC(0, p) and the $(1-p)$th quantile, τ, of X*: here $r = 2$ and $\boldsymbol{\theta} = (\theta_1, \theta_2) = (pAUC(0, p), \tau)$; and $\boldsymbol{h} = (h_1, h_2)$ where

$$\begin{aligned} h_1(X, Y, \boldsymbol{\theta}) &= \{I[Y > X] + 0.5I[Y = X]\}I[X > \theta_2] - \theta_1 \\ h_2(X, Y, \boldsymbol{\theta}) &= h_2(X, \theta_2) = I[X \leq \theta_2] - (1-p)\ . \end{aligned} \tag{11}$$

We further assume $0 < p < 1$; $F'(\tau) > 0$ and $Pr(Y \geq X | X > \tau) \neq 1$.

Define the two-sample empirical likelihood ratio

$$\begin{aligned} R(\boldsymbol{\theta}) = \sup_{u_i, v_j} \Big\{ \prod_{i=1}^m m u_i \prod_{j=1}^n n v_j\ ;\ \text{s.t. } u_i > 0;\ v_j > 0;\ \sum_{i=1}^m u_i = 1;\ \sum_{j=1}^n v_j = 1; \\ \sum_{i=1}^m \sum_{j=1}^n \boldsymbol{h}(X_i, Y_j, \boldsymbol{\theta}) u_i v_j = \mathbf{0} \Big\}\ . \end{aligned} \tag{12}$$

As sample size $\min(m, n)$ go to infinite, if $\boldsymbol{\theta}$ is the true parameter value, we have

$$-2 \log R(\boldsymbol{\theta}) \xrightarrow{\mathcal{D}} \chi^2_{(r)}$$

where $\chi^2_{(r)}$ denotes a chi squared distribution with r degrees of freedom.

Proof: The empirical likelihood book of Owen (2001) section 11.4 contains a proof of this theorem for the case $r = 1$. The conditions Owen imposed on the h functions are easy to check with our h in (i) or (ii) above. For the case $r = 2$ the proof is similar. ∎

Remark: We have stated the conditions (i) and (ii) in Theorem 1 without smoothing due to its clear connections to AUC/pAUC. If we apply the smoothing as detailed in previous section, the h function in (i) or (ii) of Theorem 1 needs to be modified as follows: all the indicator functions there shall be replaced by either $I_\epsilon(\cdot, \cdot)$ or $I_\xi(\cdot, \cdot)$. Theorem 1 is still valid after this smoothing modification: the chi square limit still hold for the -2 log empirical likelihood ratio. However, the convergence to chi square will be faster with smoothing. When profiling the empirical likelihood (next theorem), smoothing is not only a good idea but a must.

The empirical likelihood theorem above with an h specified in (i) immediately gives us a test of AUC when sample sizes are reasonably large: for testing $H_0 : AUC = \theta^*$ vs. $H_A : AUC \neq \theta^*$ the p-value can be computed as $Pr(\chi^2_{(1)} > -2\log R(\theta^*))$, where $\chi^2_{(1)}$ denotes a chi-square random variable with degree of freedom 1. The 95% confidence interval for AUC is

$$\{\theta^* | \text{ s.t. } -2\log R(\theta^*) < 3.84146 = \chi^2_{(1)}(0.95)\} .$$

For the test of the pAUC$(0, p)$, however, some more work is needed. The above empirical likelihood theorem, with $\boldsymbol{h}$ specified in (ii), only gives us a test of pAUC and τ jointly. We, however, are most likely interested only in testing pAUC alone. This calls for a profiling of the empirical likelihood ratio.

Profile empirical likelihood ratio is studied by Qin and Lawless (1994) [20]. They demonstrated that under reasonable smoothness conditions (so that certain derivatives exist) the profiling of empirical likelihood just behave the same as in the (well known) parametric likelihood case. Owen (2001) [16] Chapter 3 also discussed this topic.

Theorem 2 (Profile Empirical Likelihood) Assume the same conditions specified in Theorem 1 also hold here. We take the $\boldsymbol{h}$ function as specified in (ii) there (for pAUC and τ) but with smoothed indicator functions as discussed in the Remark following Theorem 1. Recall in this case $r = 2$ and $\boldsymbol{\theta} = (\theta_1, \theta_2) = (pAUC, \tau)$.

Define the profile log empirical likelihood ratio

$$W(\theta_1) = \inf_{\theta_2} -2\log R(\boldsymbol{\theta}) , \tag{13}$$

where $R(\boldsymbol{\theta})$ is given in Theorem 1.

As sample size $\min(m, n) \to \infty$, we have $W(\theta_1) \xrightarrow{\mathcal{D}} \chi^2_{(1)}$ if θ_1 is the true pAUC value.

Proof: See a proof in Qin and Lawless (1994) [20] Corollary 5 or Owen (2001) [16] Chapter 3. The proof was based on a two-term Taylor expansion of the log likelihood ratio. The required smoothness conditions can be easily checked since we used smoothed indicator functions. ■

Remark For parametric likelihood ratio tests, profile is a well known methodology. Also, empirical likelihood ratio test itself can, in fact, be thought of as a profile likelihood ratio test: where infinite dimensional nuisance parameters are profiled out, leaving only p parameters of interest.

One consequence of Theorem 2 is that we can use W defined in (13) to test hypothesis and construct confidence intervals for pAUC similar to those procedures we discussed after Theorem 1, using $-2\log R(\theta^*)$ for AUC.

We shall discuss the computational methods for the empirical likelihood ratio defined in Theorems 1 and 2 in next section. But here we end this section with an illustrative example:

Example 1: As a real data example we analyze the performance of the biomarker s100b in the blood of patients at hospital admission after aneurysmal subarachnoid haemorrhage (aSAH) as a predictor of their 6-month outcome. The data is from Robin et al. (2011) [21] and more information can be found in Turck et al. (2010) [22]. It contains 113 patients, among which 41 are classified as poor outcome (diseased) after 6-month. The data values are recorded with precision 0.01, we used a smoothing $\epsilon = 0.005$. The quantile smoothing bandwidth ξ is taken to be $m^{-0.75}$ when analyzing pAUC.

The estimated AUC of the ROC curve for s100b is 0.73137; and the 95% confidence interval using the empirical likelihood method is [0.62303, 0.82149]. We point out that our confidence interval is non-symmetric, i.e. not centered at the estimator: $0.73137 \neq (0.62303 + 0.82149)/2$ which is a great feature of the Wilks type confidence intervals. The detailed computation algorithm used is discussed in section 4, with actual R code in the Appendix. As a comparison, the confidence interval for AUC obtained by `pROC` package `ci.auc` function is [0.63012, 0.83262] using the 'DeLong' method, and using the bootstrap method we got [0.6265, 0.8276]. When doing the bootstrap confidence interval (here and two more below) we `set.seed(123)` and used 250,000 bootstrap repetitions.

Patients with poor post-aSAH outcome require specific health care management, therefore the clinical test must be highly specific. A pAUC with specificity in the 80% to 100% range maybe of interest. The estimator of pAUC(0, 0.2) here for biomarker s100b is 0.08061165. The 95% empirical likelihood confidence interval for the pAUC$(0, 0.2)$ is [0.049708, 0.114085]. As a comparison, the bootstrap 95% confidence interval is [0.05068, 0.1158] using `pROC` package.

Finally, as an illustration, the estimate of pAUC$(0.2, 0.7)$ is 0.3726. The 95% empirical likelihood confidence interval for pAUC$(0.2, 0.7)$ is [0.303962, 0.426242]. Using bootstrap method provided by `pROC` package we got 95% confidence interval [0.2983, 0.4275].

4 Computation

The computation of the $\log R(\theta)$ specified in (12) is not provided by the current R packages like `emplik`, nor discussed in either books [16] and [29]. Therefore, it deserve a more detailed investigation here.

To make the following easier to read, we shall spell out the details only for the h function defined in (i) of Theorem 1, i.e. the case for only one parameter of AUC. The two parameters case, pAUC and τ, is similar but formulas are longer to spell out.

First of all, the computation of $\log R(\theta)$ in Theorem 1 is a typical optimization problem over variables $(u_1, \cdots, u_m, v_1, \cdots, v_n) = \mathbf{x}$ (say) . The objective function (to be maximized) is

$$\log R(\theta) = \left\{ \sum_i \log m u_i + \sum_j \log n v_j \right\} . \tag{14}$$

The constraints imposed on $\mathbf{x} = (u_1, \cdots, u_m, v_1, \cdots, v_n)$ are

$$u_i > 0; \quad v_j > 0; \tag{15}$$

$$\sum u_i - 1 = 0; \quad \sum v_j - 1 = 0; \tag{16}$$

$$\sum_i \sum_j u_i v_j h(X_i, Y_j, \theta) = 0 \ . \tag{17}$$

The objective function, $\sum \log mu_i + \sum \log nv_j$, is nonlinear. However, it is concave and can be well approximated by quadratic functions locally. Thus a sequential quadratic programming method should work well. Indeed, Chen and Zhou (2007) [3] have successfully used the sequential quadratic programming in computing the empirical likelihood for one sample right censored data.

The difficult, however, lies in the last constraint (17). It is quadratic in terms of $\mathbf{x}$. Thus even after we replace the objective with a quadratic function, this is a Quadratic Equality Constrained Quadratic Programming problem. One possibility is to try also linearize the constraint. For work on this direction, see Wood, Do and Broom (1996) [25]. Another possibility is that we may hold u_i fixed and solve the problem for v_j, then hold v_j fixed and solve for u_i and alternating between them. We notice when u_i (or v_j) are held fixed, the problem is a linear equality constrained quadratic programming problem and easier to solve.

On the other hand, we know the similar optimizing problem in one sample empirical likelihood with a linear constraint can be solved quite satisfactorily via Lagrange multiplier method, see Owen (1988, 2001) [15] [16]. In the said case, it reduces the optimizing over n variables to solving r equations for r variables. Here r is the number of parameters and is fixed. Typically r is much smaller than sample size n.

We shall explore this idea next.

4.1 Computation For Inference of AUC

Using Lagrange multiplier method for the above optimization problem (14) – (17), the Lagrangian is

$$L(u_i, v_j, \gamma, \eta, \lambda) = \sum_{i=1}^{m} \log mu_i + \sum_{j=1}^{n} \log nv_j - \gamma\left(\sum u_i - 1\right) - \eta\left(\sum v_j - 1\right) - \lambda \sum_i \sum_j u_i v_j h(X_i, Y_j, \theta)\,. \tag{18}$$

Taking derivatives and set them to zero leads to

$$u_i(\lambda, v_1, \cdots, v_n) = \frac{1}{m + \lambda \sum_j h(X_i, Y_j, \theta) v_j}\,, \qquad i = 1, \cdots, m \tag{19}$$

$$v_j(\lambda, u_1, \cdots, u_m) = \frac{1}{n + \lambda \sum_i h(X_i, Y_j, \theta) u_i}\,, \qquad j = 1, \cdots, n. \tag{20}$$

The above two sets of equations plus the following (constraint requirement)

$$\sum_i \sum_j \frac{h(X_i, Y_j, \theta)}{[m + \lambda \sum_k h(X_i, Y_k, \theta) v_k]\,[n + \lambda \sum_k h(X_k, Y_j, \theta) u_k]} = 0 \tag{21}$$

are the system of equations we need to solve to obtain maximized $\log R(\theta)$. Notice there are $m + n + 1$ equations for $m + n + 1$ variables (u_i, v_j, λ). Compared to the situation of Owen

(1988) [15], we have $m+n$ more equations here: namely (19) and (20). A direct solution seems elusive. But the following iterative method works in our investigations.

Initialize $u_i^{(0)} = 1/m$ and $v_j^{(0)} = 1/n$,

1. Plug $u_i^{(s)}$ and $v_j^{(s)}$ into equation (21), solve for λ, call the solution $\lambda^{(s+1)}$.
2. Using $u_i^{(s)}$ and $v_j^{(s)}$ and $\lambda^{(s+1)}$ obtained above, plug into the right hand side of equations (19) and (20). This yields $u_i^{(s+1)}, v_j^{(s+1)}$.
3. With $u_i^{(s+1)}$ and $v_j^{(s+1)}$, repeat steps 1–2 to obtain $\lambda^{(s+2)}$ and $u_i^{(s+2)}, v_j^{(s+2)}$.
4. Iterate until λ converges.

This is the algorithm we used when computing examples and carry out simulations. We notice the equation (21) is monotone in λ, at least for those λ's that make (19) and (20) a probability.

An implementation of above is coming soon in our new package `emplikAUC`.

4.2 Computation For Inference of pAUC

The first step here is to compute the log empirical likelihood ratio $\log R(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2)$. Here $\boldsymbol{\theta}$ and $\boldsymbol{h}$ are as defined in Theorem 1 equation ()11) but with smoothing. This is similar to the calculation detailed in the above subsection for AUC except one constrain equation, (21), becomes two constrain equations, (22) – (23) now, and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$, $\boldsymbol{h}$ and $\boldsymbol{\theta}$ are now vectors of length two.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} \frac{h_1(X_i, Y_j, \boldsymbol{\theta})}{[m + \boldsymbol{\lambda}^\top \sum_k \boldsymbol{h}(X_i, Y_k, \boldsymbol{\theta}) v_k]\,[n + \boldsymbol{\lambda}^\top \sum_k \boldsymbol{h}(X_k, Y_j, \boldsymbol{\theta}) u_k]} = 0 \tag{22}$$

$$\sum_{i=1}^{m} \frac{h_2(X_i, \theta_2)}{m + \boldsymbol{\lambda}^\top \sum_k \boldsymbol{h}(X_i, Y_k, \boldsymbol{\theta}) v_k} = 0 \tag{23}$$

The above two constrain equations plus (19) – (20), (with the obvious modification of $\lambda \to \boldsymbol{\lambda}$, $h \to \boldsymbol{h}$ and $\theta \to \boldsymbol{\theta}$), are the system of $n+m+2$ equations for $n+m+2$ unknowns $(u_i, v_j, \lambda_1, \lambda_2)$ we need to solve. The iterative algorithm detailed in the previous subsection also works well here.

Now assume we have obtained $\log R(\boldsymbol{\theta})$. The second step is the minimization (profiling) of $\log R(\boldsymbol{\theta})$ over θ_2. This is an unconstrained minimization problem over one variable and an obvious starting point for θ_2 is the sample quantile $\hat{\tau}$. In the simulation and examples, we have used R function `optimize` to accomplish this.

The profile will yield $\log W(\theta_1)$ which can be used to test and construct confidence intervals for pAUC.

5 Simulations

We first compare the confidence intervals for AUC. The empirical likelihood method specified in section 3 and 4, and those available in the R package `pROC`, namely DeLong and Bootstrap, were used. The bootstrap repetitions were set to 5000. Results are listed in Table 1.

We see from Table 1 that

- When compared to both the DeLong and Bootstrap confidence intervals, the empirical likelihood confidence intervals have shorter average length AND higher coverage probabilities (only in one case it has equal coverage probability with Bootstrap, but with shorter length). The differences are more profound for (smaller) sample sizes (60, 40), less so for (100, 100).

- The errors for the empirical likelihood confidence intervals are more balanced. The DeLong confidence intervals missed the true value too much on the "high" side (i.e. lower confidence limits too high). The Bootstrap intervals are better than the DeLong, but still worse than Empirical Likelihood.

- All the coverage probabilities are pretty close to the nominal 95% value for sample sizes (100, 100) but have bigger gaps for (60, 40).

Next we will compare the confidence intervals for pAUC, the results are summarized in table 2. We have used the bootstrap percentile method (BootPT) provided by the package `pROC`, with bootstrap repetition set at 10,000. We also included two methods provided by the `tpAUC` package, namely the method of `"MW"` (Mann-Whitney) and the method `"expect"`. Both methods from package `tpAUC` are using the same jackknife variance estimate to construct Wald confidence intervals. Thus, the length of confidence intervals of these two methods are exactly the same. The only difference is the center of the interval, i.e. the estimator.

Finally the entry "2EL" there is our empirical likelihood method described in sections 3 and 4.

We see from table 2 that

- The "2EL" confidence intervals have shortest average length . Yet at the same time it has highest coverage probability, except in one case it tied with bootstrap as the two highest coverage probabilities (case 150/150).

- Compared with other methods, the errors of the 2EL confidence intervals are more balanced, i.e. missing the true pAUC value on the above/below with similar probability. The worst one is the Mann-Whitney method – the confidence intervals missing the true value mostly because it is too high.

- Among the two Wald methods with jackknife variance estimator, the method `"expect"` performs better than `"MW"`. According to [26] page 361, `"expect"` method is based on

Table 1: Coverage Probability and Average Length of Nominal 95% Confidence Intervals for AUC. I. $X \sim exp(1); Y \sim exp(0.4)$. True $AUC = 0.714286$. II. $X \sim N(0,1); Y \sim N(\mu = 2, sd = 2)$. True $AUC = 0.814453$. Based on 1000 runs. The middle column are the probabilities of the intervals missing the true value on the above/below.

(m, n)	Method	Coverage Probability	Average Length
(100, 100)	DeLong	$0.953 = 1 - (0.034 + 0.013)$	0.1421457
exp(1), exp(0.4)	Boot	$0.956 = 1 - (0.030 + 0.014)$	0.1416802
	2EL	$0.956 = 1 - (0.024 + 0.020)$	0.1406806
(80, 50)	DeLong	$0.939 = 1 - (0.039 + 0.022)$	0.1875612
exp(1), exp(0.4)	Boot	$0.938 = 1 - (0.035 + 0.027)$	0.1860194
	2EL	$0.943 = 1 - (0.025 + 0.032)$	0.1843396
(60, 40)	DeLong	$0.943 = 1 - (0.041 + 0.016)$	0.2120362
exp(1), exp(0.4)	Boot	$0.944 = 1 - (0.036 + 0.020)$	0.210127
	2EL	$0.951 = 1 - (0.028 + 0.021)$	0.2074215
(100, 100)	DeLong	$0.949 = 1 - (0.036 + 0.015)$	0.1234693
N(0,1), N(2,2)	Boot	$0.950 = 1 - (0.031 + 0.019)$	0.1228783
	2EL	$0.951 = 1 - (0.025 + 0.024)$	0.1225727
(80, 50)	DeLong	$0.924 = 1 - (0.063 + 0.013)$	0.168157
N(0,1), N(2,2)	Boot	$0.934 = 1 - (0.047 + 0.019)$	0.1665892
	2EL	$0.941 = 1 - (0.031 + 0.028)$	0.1549231
(60, 40)	DeLong	$0.927 = 1 - (0.062 + 0.011)$	0.1885157
N(0,1), N(2,2)	Boot	$0.937 = 1 - (0.049 + 0.014)$	0.1862026
	2EL	$0.954 = 1 - (0.026 + 0.020)$	0.1856122

an alternative estimator of the pAUC, namely

$$\widetilde{pAUC}(0, p_0) = p_0 - \frac{1}{n} \sum_{j=1}^{n} \min\{1 - \hat{F}_n(Y_j),\, p_0\} \ .$$

However, upon closer inspection, this estimator is exactly the same as the Mann-Whitney estimator (5), except for tie breaker (when $X = Y$). (proof available upon request). In the simulation our random numbers never have ties, so the JackEX entry in table 2 is a bona fide Mann-Whitney based confidence interval. A side note is that the Mann-Whitney option `"MW"` from the package `tpAUC` likely have some bugs. Its performance is also the worst among the four.

On a desktop computer with Intel i7-4790 processor and R version 4.1.1, it takes about 0.8 second to compute a 90% empirical likelihood confidence interval for pAUC$(0, 0.3)$, for sample size 80/120. For sample size 150/150, it takes about 4.2 second to compute such an empirical likelihood confidence interval.

Table 2: Coverage Probability and Average Length of Nominal 90% Confidence Intervals for pAUC(0, 0.3). $X \sim N(0,1); Y \sim N(1,1)$. Based on 1000 simulation runs.

(m, n)	Method	Coverage Probability	Average Length
(50, 50)	BootPT	$0.905 = \quad 1-(0.029+0.066)$	0.08475382
	JackMW	$0.826 = \quad 1-(0.016+0.158)$	0.08660019
	JackEX	$0.909 = \quad 1-(0.037+0.054)$	0.08660019
	2EL	$0.918 = \quad 1-(0.035+0.047)$	0.08379997
(100, 100)	BootPT	$0.889 = \quad 1-(0.037+0.074)$	0.0600507
	JackMW	$0.846 = \quad 1-(0.025+0.129)$	0.06070723
	JackEX	$0.887 = \quad 1-(0.045+0.068)$	0.06070723
	2EL	$0.894 = \quad 1-(0.044+0.062)$	0.05977157
(150, 150)	BootPT	$0.901 = \quad 1-(0.038+0.061)$	0.04917168
	JackMW	$0.863 = \quad 1-(0.025+0.112)$	0.04950833
	JackEX	$0.898 = \quad 1-(0.045+0.057)$	0.04950833
	2EL	$0.901 = \quad 1-(0.045+0.054)$	0.04897593
(120, 80)	BootPT	$0.894 = \quad 1-(0.038+0.068)$	0.06031164
	JackMW	$0.864 = \quad 1-(0.028+0.108)$	0.06084894
	JackEX	$0.890 = \quad 1-(0.047+0.063)$	0.06084892
	2EL	$0.900 = \quad 1-(0.044+0.056)$	0.05989653
(80, 120)	BootPT	$0.893 = \quad 1-(0.033+0.074)$	0.06208052
	JackMW	$0.836 = \quad 1-(0.018+0.146)$	0.06299564
	JackEX	$0.891 = \quad 1-(0.043+0.066)$	0.06299564
	2EL	$0.897 = \quad 1-(0.045+0.058)$	0.06193831

Testing a given hypothesis for pAUC$(0, 0.3)$ by empirical likelihood is much faster. For example at sample size 150/150, it only takes about 0.05 to 0.08 second to do an empirical likelihood test of hypothesis (i.e. obtain p-value). This is a unique feature of the empirical likelihood method: that a testing hypothesis is much faster to compute than a confidence interval.

The Bootstrap confidence interval for sample size 150/150 also takes about 4.2 seconds per interval. We have set the bootstrap repetition to 10000.

The two Jackknife/Wald based method is much faster: for sample size 150/150 at about 0.15 second per confidence interval. Testing a hypothesis will be about the same speed.

6 Discussion and Concluding Remarks

To deal with the inference of pAUC, we view the estimation in the larger frame work of a two-parameter problem, by explicitly include the nuisance parameter τ. This naturally lead to the profile likelihood technique when inference for only one of the two parameters is needed.

When sample sizes goes to infinite, the Wilks and the Wald type confidence intervals (assume both available) are equivalent. However, for smaller samples, it is a generally accepted fact that the Wilks confidence intervals have several advantages over the Wald confidence intervals, see section 3 of Meeker and Escobar [14] and additional references there. The disadvantage they mentioned for the Wilks method is the computational difficulty. But with ever faster computers and publicly available software like R, this is much less of a problem nowadays.

We list here briefly the advantages of Wilks confidence interval:

1). The Wilks confidence intervals are not necessary symmetric about the MLE, rather, it tries to reflect the skewness in the given data.

2). The Wilks confidence intervals are always within the parameter space, while a Wald confidence interval of a probability can include negative values, for example.

3). Once we obtained the Wilks confidence interval for a parameter θ, $[a, b]$ (say), the Wilks confidence interval for $g(\theta)$ is just $[g(a), g(b)]$ (assuming g is increasing).

4). When using Wilks, there is no need to figuring out the variance of the MLE and estimate it.

5). The actual error rates for Wilks intervals are often closer to the nominal than the Wald.

6). Bootstrap re-sampling based procedures rely on the random number generator and the bootstrap repetitions used. A different seed or different number of repetitions will lead to slightly different confidence intervals. Our empirical likelihood confidence interval do not have this problem.

7). The Wilks confidence interval is based on likelihoods and there is well developed theory to handle nuisance parameter in the likelihood analysis. We use this feature to handle nuisance parameter in the inference of pAUCs.

Generalization of the empirical likelihood method proposed in this paper are possible. We only mention one. For sample data that are collected via a biased sampling scheme, (for example, test-result-dependent sampling, see [**?**]), the likelihood function can often be adapted to reflect this new sampling scheme. Once this is done, the empirical likelihood method we proposed in this paper can, in principle, be adapted to give a valid confidence interval, (mind the computational complication).

References

[1] Adimari, G. and Chiogna, M. (2012). Jackknife empirical likelihood based confidence intervals for partial areas under ROC curves. *Statistica Sinica* **22,** 1457-1477.

[2] Bandos, AI. Guo, B. and Gur, D. (2017). Jackknife variance of the partial area under the empirical receiver operating characteristic curve. *Stat Methods Med Res* **26(2)**: 528-541. doi: 10.1177/0962280214551190.

[3] Chen, K. and Zhou, M. (2007). Computation of the empirical likelihood ratio from censored data. *Journal of Statistical Computation and Simulation*, Vol. 77, 1033 - 1042.

[4] Chen, SX. and Hall, P. (1993). Smoothed Empirical Likelihood Confidence Intervals for Quantiles. *The Annals of Statistics* **21,** 1166-1181.

[5] DeLong, ER. DeLong, DM. Clarke-Pearson, DL. (1988). Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44: 837–845.

[6] Dodd, LE. and Pepe, MS. (2003), Partial AUC Estimation and Regression. *Biometrics* **59,** 614-623.

[7] Hanley, JA. and McNeil, BJ. (1982). The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* **143,** 29-36.

[8] Hanley, JA. and McNeil, BJ. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**, (3): 839–843.

[9] He, Y. and Escobar, M. (2008). Nonparametric statistical inference method for partial areas under receiver operating characteristic curves, with application to genomic studies. *Statist. Medicine* **27,** 5291-5308.

[10] Hjort, NL. McKeague, I. and Keilegom, IV. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37,** 1079-1111.

[11] Kim, MO. and Zhou, M. (2019). Comparison of extended empirical likelihood methods: size and shape of test-based confidence regions. *Statistica Sinica* **29**, 371-386 doi:https://doi.org/10.5705/ss.202015.0410

[12] Krzanowski, W. and Hand, D. (2009). *ROC Curves for Continuous Data,* Chapman & Hall/CRC, Boca Raton.

[13] McClish, DK. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making* **9(3),** 190-195.

[14] Meeker, W. and Escobar, L. (1995). Teaching about Approximate Confidence Regions Based on Maximum Likelihood Estimation. *The American Statistician*, Vol. 49, No. 1, pp. 48–53.

[15] Owen, AB. (1988). Empirical Likelihood Ratio Confidence Intervals for a Single Functional. *Biometrika* **75,** 237 - 249.

[16] Owen, AB. (2001). *Empirical Likelihood*. Chapman & Hall/CRC: Boca Raton, London, New York, Washington, D.C.

[17] Pepe, MS. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction,* Oxford University Press Inc., New York.

[18] Qin, G. Jin, X. Zhou, XH. (2011). Non-parametric interval estimation for the partial area under the ROC curve. *The Canadian Journal of Statistics* **39(1),** 17-33.

[19] Qin, G. Zhou, XH. (2006). Empirical likelihood inference for the Area under the ROC curve. *Biometrics* **62,** 613-622.

[20] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22,** 300-325.

[21] Robin, X. Turck, N. Hainard, A. Tiberti, N. Lisacek, F. Sanchez, JC. Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, **12**, 77. DOI: 10.1186/1471-2105-12-77.

[22] Turck, N. Vutskits, L. Sanchez-Pena, P. Robin, X. Hainard, A. Gex-Fabry, M. Fouda, C. Bassem, H. Mueller, M. Lisacek, F. Puybasset, L. and Sanchez, J. (2010). A multiparameter panel method for outcome prediction following aneurysmal subarachnoid hemorrhage. *Intensive Care Med* **36** 107–115.

[23] Venkatraman, ES. and Begg, CB. (1996). A Distribution-Free Procedure for Comparing Receiver Operating Characteristic Curves from a Paired Experiment *Biometrika* 83(4) 835-848

[24] Wieand, HS. Gail, M. and James, B. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* 76: 585–592.

[25] Wood, ATA. Do, KA. and Broom, BM. (1996), Sequential Linearization of Empirical Likelihood Constraints with Application to U-Statistics. *Journal of Computational and Graphical Statistics.* Vol. 5, No. 4 (Dec., 1996), pp. 365-385.

[26] Yang, H. Lu, K. and Zhao, Y. (2017). A nonparametric approach for partial areas under roc curves and ordinal dominance curves. *Statistica Sinica* **27**, 357-371.

[27] Yu, J. Yang, L. Vexler, A. and Hutson, AD. (2016). Easy and accurate variance estimation of the nonparametric estimator of the partial area under the ROC curve and its application. *Statist. Medicine* **35**, Issue 13, 2251-2282.

[28] Zhao, YM. (2016). Statistical Inference on Trimmed Means, Lorenz Curves, and Partial Area Under ROC Curves by Empirical Likelihood Method. University of Kentucky Theses and Dissertations–Statistics. 24. https://uknowledge.uky.edu/statistics_etds/24

[29] Zhou, M. (2016). *Empirical Likelihood Method in Survival Analysis.* CRC Press: Boca Raton.

[30] Zhou, XH. Obuchowski, N. and McClish, D. (2011). *Statistical Methods in Diagnostic Medicine, 2nd Ed.* John Wiley & Sons Inc., Hoboken, New Jersey.

[31] Zhang, DD. Zhou, XH. Daniel, H. Freeman, J. Freeman, JL. (2002). A non-parametric method for the comparison of partial areas under roc curves and its application to large health care data sets. *Statist. Medicine* **21,** 701-715.

[32] Zou, KH. Liu, A. Bandos, AI. Ohno-Machado, L. and Rockette, H. (2012). *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis,* CRC Press, Boca Raton.

[33] Zou, K. Hall, WJ. Shapiro, D. (1997). Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statist. Medicine* **16**, 2143-2156.

7 Appendix

R code for example 1.

The data set we used, `aSAH`, is from R package `pROC`. We also assume the packages `pROC`, `downloader`, `rootSolve` and `emplik2` are installed. These R packages can all be downloaded at https://cran.r-project.org/web/packages/.

```
###### Get data set aSAH for marker s100b ######
library(pROC)
data(aSAH)
Xis <- aSAH$s100b[aSAH$outcome == "Good"]
Yis <- aSAH$s100b[aSAH$outcome == "Poor"]

library(downloader)
source_url("http://www.ms.uky.edu/~mai/Rcode/emplikAUC.R")

############ Get an estimator of AUC ############
rep(1/72, 72)%*%smooth3(x=Xis, y=Yis)%*%rep(1/41, 41)
##           [,1]
## [1,] 0.7313686      ##### estimate of AUC #####
```

Now we compute the 95% confidence interval for AUC.

```
findULNEW(step=0.03, fun=ThetafunAUCone, MLE=0.73, x=Xis, y=Yis)
## $Low
## [1] 0.6230165
##
## $Up
## [1] 0.821502
##
## $FstepL
## [1] 7.450581e-09
##
## $FstepU
```

```
## [1] 7.450581e-09
##
## $Lvalue
## [1] 3.84146
##
## $Uvalue
## [1] 3.84146
###### The 95% (default) confidence interval is seen to be [0.6230165, 0.821502] ###
###### If we just need to test a specific hypothesis, we may do the following #####
eltest4aucONE(x=Xis, y=Yis, true=0.821502, ind=smooth3, tol.u=1e-6, tol.v=1e-6,
              tol.H0=1e-6)$"-2LLR"
## [1] 3.841464    ### Output is the -2log ELR value for testing AUC=true=0.821502.

eltest4aucONE(x=Xis, y=Yis, true=0.623016, ind=smooth3, tol.u=1e-6, tol.v=1e-6,
              tol.H0=1e-6)$"-2LLR"
## [1] 3.841491     ### this chi square df1 value gives a p-value of 5%.
```

Now for testing and confidence interval of pAUC$(0, 0.2)$. We first get an estimator of τ and pAUC$(0, 0.2)$. The quantile estimator τ uses our smoothing, defined in section 3.

```
##### Use our smoothing to get estimators of 80th quantile and pAUC(0,0.2) #####
myEstPaucT(x=Xis, y=Yis, partial=0.2, eps=0.005, epsT=(72)^(-0.75))
## $`tau(1-partial)`
## [1] 0.2083099
##
## $`Pauc(0,partial)`
## [1] 0.08060994
#### The output value can change a little for other smoothing values of eps/epsT.

###### Jointly testing F^{-1}(0.8)=tau=0.19, pAUC(0, 0.2)=0.08, with null
###### distribution of chisq with 2DF  ####

Profun(tau=0.19, pauc=0.08, partial=0.2, x=Xis, y=Yis)      #### output omitted

###### Next, testing pAUC(0, 0.2)=0.08 alone, with null dist. = chisq 1DF #####

optimize(f=Profun,interval=c(0.154, 0.420),pauc=0.08,partial=0.2,x=Xis,y=Yis)

#### Output omitted.  The "interval" is where we search for nuisance parameter.  ##
#### The $objective in the output is the Wilks statistic: -2LLR(pAUC(0,0.2)=0.08) #

#### Verify the 95% confidence interval for pAUC(0,0.2) is [0.049810, 0.114224] ###
#### since testing the upper/lower bound, we get W(.) = chisq(0.95, df=1) ####

optimize(f=Profun,interval=c(0.154,0.420),pauc=0.114224,partial=0.2,x=Xis,y=Yis)
## $minimum
## [1] 0.1777186
##
## $objective
## [1] 3.841459

optimize(f=Profun,interval=c(0.154,0.420),pauc=0.049810,partial=0.2,x=Xis,y=Yis)
## $minimum
## [1] 0.3048885
##
## $objective
## [1] 3.841665

##### We may also use findULNEW( ) to find the confidence interval #####

findULNEW(step=0.03, fun=ThetafunPAUC, MLE=0.08061,x=Xis,y=Yis) #### output omitted
```

Same calculation, using faster code (but may be less robust).

```
#### Jointly testing F^{-1}(0.8)=tau=0.19 and pAUC(0, 0.2) = 0.08
ProfunONE(tau=0.19, pauc=0.08, partial=0.2, x=Xis, y=Yis)
## [1] 0.3047554      ### This is -2 log ELR, null dist. is chi-square 2DF.###

#### Getting ready to search for confidence interval, ONE SIDE at a time.

#### When search for Upper confidence bound, nuiup = MLE of nuisance parameter tau.
findUnew(step=0.01, fun=ThetafunPaucONE, MLE=0.08061, x=Xis, y=Yis, nuilow=0.15,
          nuiup=0.2083)
## $Up
## [1] 0.1142243
##
## $FstepU
## [1] 7.450581e-09
##
## $Uvalue
## [1] 3.84146

#### When search for Lower confidence bound, nuilow=MLE of nuisance parameter tau.
findLnew(step=0.01, fun=ThetafunPaucONE, MLE=0.08061, x=Xis, y=Yis, nuilow=0.2083,
          nuiup=0.42)
## $Low
## [1] 0.04981065
##
## $FstepL
## [1] 7.450581e-09
##
## $Lvalue
## [1] 3.84146
```

Finally, we shall test and find confidence interval for pAUC(0.2, 0.7). From above we know by using our smoothing, the estimator of $F^{-1}(0.8)$ is 0.2083099. Similarly the estimator of $F^{-1}(0.3)$ is 0.08346956 as seen below.

```
quantONE(x=Xis, prob=0.8)
## [1] 0.2083099
quantONE(x=Xis, prob=0.3)
## [1] 0.08346956

tauhat2 <- 0.2083099       ##### estimator of F^{-1}(0.8)=F^{-1}(1-partial1)
tauhat1 <- 0.08346956      ##### estimator of F^{-1}(0.3)=F^{-1}(1-partial2)
myeps <- (length(Xis))^(-0.75)

H11 <- smooth3(x=Xis, y=Yis)
H12 <- as.matrix(  smooth3(x=Xis, y=rep(tauhat2, 41), eps=myeps) )
H13 <- as.matrix(1- smooth3(x=Xis, y=rep(tauhat1, 41), eps=myeps) )
H1 <- H11*H12*H13
rep(1/72,72)%*% H1 %*%rep(1/41,41)
##        [,1]
## [1,] 0.3726                      ###### estimator of pAUC(0.2, 0.7)####

##### Jointly test H_0: pAUC(0.2, 0.7)=0.37; F^{-1}(0.8)=0.208; F^{-1}(0.3)=0.084.

Profun2(tauVec=c(0.084, 0.208), pauc=0.37, partial1=0.2, partial2=0.7, x=Xis,y=Yis)
## [1] 0.01404179    #### so the p-value is 1-pchisq(0.01404179, df=3) = 0.9995593

###### To testing pAUC alone   H_0: pAUC(0.2, 0.7) = 0.33.
###### We just need to minimize over tau1 and tau2 of the Profun2() output.

optim(par=c(0.083,0.208),fn=Profun2,pauc=0.33,partial1=0.2,partial2=0.7,x=Xis,y=Yis)
## $par
## [1] 0.08644508 0.25721451
##
## $value
## [1] 1.587815
##
## $counts
## function gradient
## 61       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

```
##### Or, using another minimization method, with box bounded search. #####

optim(par=c(0.083,0.208),fn=Profun2,pauc=0.33,partial1=0.2,partial2=0.7,
           x=Xis,y=Yis,method="L-BFGS-B",lower=c(0.08,0.19),upper=c(0.09,0.27))
## $par
## [1] 0.08645122 0.25725144
##
## $value
## [1] 1.587816
##
## $counts
## function gradient
## 11       11
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"

#### Either way we got -2LLR=1.5878.  This gives a p-value of
#### 1-pchisq(1.5878, df=1) = 0.2076407. ####

#####  For 95% confidence interval, we check:
#####             testing H_0: pAUC(0,2, 0.7)=0.303962;
#####             testing H_0: pAUC(0.2, 0.7)=0.426242
##### In both tests we got p-value of 0.05. Therefore the 95%
##### Confidence interval for pAUC(0.2, 0.7) is [0.303962, 0.426242].
```