Georgia State University
ScholarWorks @ Georgia State University

Mathematics Theses | Department of Mathematics and Statistics

8-2-2007

Empirical Likelihood-Based NonParametric Inference for the Difference between Two Partial AUCS

Yan Yuan

Follow this and additional works at: http://scholarworks.gsu.edu/math_theses

Recommended Citation
Yuan, Yan, "Empirical Likelihood-Based NonParametric Inference for the Difference between Two Partial AUCS." Thesis, Georgia State University, 2007.
http://scholarworks.gsu.edu/math_theses/32

This Thesis is brought to you for free and open access by the Department of Mathematics and Statistics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Mathematics Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

EMPIRICAL LIKELIHOOD-BASED NONPARAMETRIC INFERENCE FOR THE DIFFERENCE BETWEEN TWO PARTIAL AUCS

by

Yan Yuan

Under the direction of Gengsheng Qin

ABSTRACT

Compare the accuracy of two continuous-scale tests is increasing important when a new test is developed. The traditional approach that compares the entire areas under two Receiver Operating Characteristic (ROC) curves is not sensitive when two ROC curves cross each other. A better approach to compare the accuracy of two diagnostic tests is to compare the areas under two ROC curves (AUCs) in the interested specificity interval. In this thesis, we have proposed bootstrap and empirical likelihood (EL) approach for inference of the difference between two partial AUCs. The empirical likelihood ratio for the difference between two partial AUCs is defined and its limiting distribution is shown to be a scaled chi-square distribution. The EL based confidence intervals for the difference between two partial AUCs are obtained. Additionally we have conducted simulation studies to compare four proposed EL and bootstrap based intervals.

INDEX WORDS: ROC curve, AUC, PAUC, Partial AUC, Empirical Likelihood, Bootstrap, Confidence Interval.

EMPIRICAL LIKELIHOOD-BASED NONPARAMETRIC INFERENCE FOR THE DIFFERENCE BETWEEN TWO PARTIAL AUCS

by

Yan Yuan

A Thesis submitted in partial Fulfillment of the Requirements for the Degree of

Master of Science

In the College of Arts and Science

Georgia State University

2007

Copyright by
Yan Yuan
2007

EMPIRICAL LIKELIHOOD-BASED NONPARAMETRIC INFERENCE FOR THE DIFFERENCE BETWEEN TWO PARTIAL AUCS

by

Yan Yuan

Major Professor: Dr. Gengsheng Qin

Committee: Dr. Yu-Sheng Hsu

Dr. Xu Zhang

Electronic Version Approval:

Office of Graduate Studies

College of Arts and Science

Georgia State University

August 2007

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr.Gengsheng Qin who directed and guided me through the whole process of this thesis. This thesis cannot be finished without him.

I would also like to thank the other members of my committee, Dr. Yu-Sheng Hsu and Dr. Xu Zhang who took the time to participate as proofreaders and committee members. I thank them very much for their comments and suggestions on the improvement of my thesis.

Finally I want to thank my family and friends for their encouragement, love and support.

TABLE OF CONTENTS

LIST OF ABBREVIATION

AUC: Area under the ROC Curve

BS: Bootstrap Method

EL: Empirical Likelihood Method

FPR: False Positive Rate

PAUC: Partial Area under the ROC Curve

ROC: Receiver Operating Characteristic

TPR: True Positive Rate

$\Delta_{p_0 p_1}$: The difference between two partial AUCs

$\delta_{p_0 p_1}$: PAUC over the interval (p_0, p_1)

CHAPTER I

INTRODUCTION

The accuracy of a binary diagnostic test can be measured by its specificity and sensitivity. The sensitivity or true positive rate (TPR) of the test is the proportion of diseased patients who test positive. The specificity or true negative rate (TNR) of the test is the proportion of non-diseased patients who test negative.

When the outcome of a diagnostic test is continuous, a cut-off point for the positive of disease needs to be chosen to compute specificity and sensitivity of the test. Let Y and X be the results of a continuous-scale test for a diseased and a non-diseased subject with cumulative distribution function G and F, respectively. For a given cut-off point c, the sensitivity and specificity of the test are defined as

$$Se = P(Y \geq c) = 1 - G(c); \qquad Sp = P(X \leq c) = F(c)$$

respectively. When specificity is 1-p, the corresponding sensitivity of the test is $R(p) = 1 - G(F^{-1}(1-p))$, where F^{-1} is the inverse function of F .

The receiver operating characteristic (ROC) curve, denoted by $R(p)$, is the plot of sensitivity against the false positive rate (FPR or 1- specificity) as the cut-off point runs through the whole range of possible test values. In fact, the non-diseased population is unknown, and the optimal cut-off point is unknown too. For a continuous-scale diagnostic test, the area under the ROC curve (AUC), defined as $\delta = \int_0^1 R(p)dp$, is commonly used to summarize the accuracy of the diagnostic test across all the possible

cut-off points. The larger is the AUC, the better the diagnostic test will be. Now, the AUC is a very popular tool in diagnostic medicine.

However, the AUC has several limitations that may make it less useful for continuous diagnostic tests (Hilden, 1991). When two ROC curves cross, the two diagnostic tests can have similar AUC even though one test has higher sensitivity for certain specificities while the other test has better sensitivity for other specificities. On the other hand, in diagnostic testing, it is critical to maintain a high sensitivity in order not to miss detecting subjects with "disease" and the interest would be in the region of ROC curve corresponding only to acceptable high sensitivities. For cancer screening, only the lower tail of the ROC curve is of interest because the FPR must be very small to be acceptable (Lilienfeld, 1974). For these reasons, the partial AUC (pAUC) has been proposed as an alternative measure to the full AUC. When using the pAUC, one considers only those regions of the ROC space where data have been observed, or which correspond to clinical relevant values of sensitivity or specificity. The pAUC over the interval (p_0, p_1) of false positive rates, denoted by $\delta_{p_0 p_1}$, is

$$\delta_{p_0 p_1} = \int_{p_0}^{p_1} R(p)dp \qquad \text{for } 0 \le p_0 < p_1 \le 1.$$

It can be described as the cumulative value of sensitivity for all possible values of the false positive rates in the interval (p_0, p_1) .

Let $X_1, X_2, \ldots, X_m$ be the test results from a random sample of non-diseased population with distribution function *F*; let $Y_1, Y_2, \ldots, Y_n$ be the test results from a random sample of diseased population with distribution function *G*. Dodd and Pepe

(2003) proposed the following nonparametric estimator for the pAUC. When the quantiles $q_i = F^{-1}(1-p_i)$ (i=0, 1) are known, the pAUC can be estimated by

$$\tilde{\delta}_{p_0 p_1} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_j \geq X_i) I\left(X_i \in (q_1, q_0)\right).$$

When the quantile q_i's are unknown, the pAUC can be estimated by

$$\hat{\delta}_{p_0 p_1} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_j \geq X_i) I\left(X_i \in (\hat{q}_1, \hat{q}_0)\right)$$

where $\hat{q}_i = \hat{F}^{-1}(1-p_i)$ (i=0,1) and $\hat{F}$ is the empirical distribution of F.

Many approaches have been proposed for constructing a confidence interval for the full or partial AUC. McClish (1989), Thompson and Zucchini (1989), and Jiang, Metz, and Nishikawa (1996) proposed parametric methods for the interval estimation of the pAUC using the bi-normal model. But Walsh (1997) found that the inferences for the pAUC are sensitive to the parametric model assumption. Wieand et al (1989) proposed a generalized nonparametric method for the inference of both the full and the partial AUC. However, their method is involved in density and distribution function estimations and mathematically too complicated to be well applied in practice. Qin and Zhou (2006) proposed an Empirical Likelihood (EL) based approach for the inference on the full AUC and recommended the use of an EL-based approach when the underlying distributions for diseased and non-diseased populations are unknown. Qin, Jin and Zhou (2006) developed bootstrap and EL-based inference for pAUC and did extensive simulation studies to compare three nonparametric confidence intervals (Normal Approximation, Bootstrap, and Empirical Likelihood) for the pAUC. They also recommended the use of EL-based

approach for pAUC when the underlying distributions for diseased and non-diseased populations are unknown.

Comparing two continuous-scale diagnostic tests is increasingly important when a new test is developed and marketed (Delong 1988). How can we know which diagnostic test is better? Investigators often compare the validity of two tests based on the estimated areas under the respective ROC curves. However, the traditional way of comparing entire areas under two ROC curves is not sensitive when two ROC curves cross each other (Zhang et al., 2002). In this thesis, we propose methods to compare the partial area under the curve within a specific range of specificity for two ROC curves, non-parametric methods based on EL and bootstrap have been developed.

This thesis is organized as follows: In Chapter II, we propose two bootstrap confidence intervals for the difference between two partial AUCs. In Chapter III, we propose the EL-based intervals for the difference between two partial AUCs. In Chapter IV, we conduct simulation study to evaluate the performances of these intervals. In Chapter V, we analyze Dermatoscope Example to illustrate the proposed intervals. Finally, the conclusions are discussed in Chapter VI.

CHAPTER II

Bootstrap Confidence Interval for the Difference between Two partial AUCS

Consider two diagnostic tests T_k (k=1, 2). Both tests yield continuous measurements and are performed on the same *m* non-diseased and *n* diseased cases. Let $X_{k1}, X_{k2} \ldots X_{km}$ be i.i.d bivariate test results from a non-diseased population with joint distribution function $F(x_1, x_2)$, and let $Y_{k1}, Y_{k2} \ldots\ldots, Y_{kn}$ i.i.d bivariate test results from a diseased population with joint distribution function $G(y_1, y_2)$. Denote the marginal distribution functions of X_{ki} and Y_{kj} by F_k and G_k, respectively. The pAUC of test T_k (k=1, 2) over the interval (p_0, p_1) of false positive rates, denoted by $\delta_{p_0 p_1}^{(k)}$, is

$$\delta_{p_0 p_1}^{(k)} = \int_{p_0}^{p_1} R_k(p)dp \qquad \text{for } 0 \le p_0 < p_1 \le 1,$$

where $R_k(p) = 1 - G_k(F_k^{-1}(1-p))$ is the ROC curve of test T_k (k=1, 2) . The difference between two pAUCS is $\Delta_{p_0 p_1} = \delta_{p_0 p_1}^{(2)} - \delta_{p_0 p_1}^{(1)}$. Our goal is to construct confidence interval for $\Delta_{p_0 p_1}$ based on test results X_{ki}'s and Y_{kj}'s.

2.1 Normal Approximation Method

For one diagnostic test, Let $\{Y_1, Y_2, \ldots, Y_n\}$ and $\{X_1, X_2, \ldots, X_m\}$ be the results of a continuous-scale test for a diseased and a non-diseased subject with cumulative distribution function *F* and *G*. Dodd and Pepe (2003) defined the restricted placement value of *X* as $V(X) = (1 - G(x))I(X \in (q_1, q_0))$ when assume the quantiles q_1 and q_0 are known. Let $\hat{G}(y) = \frac{1}{n}\sum_{j=1}^{n} I(Y_j \le y)$ be the empirical distribution of *G*, and

$\tilde{V}_i = (1-\hat{G}(X_i))I(X_i \in (q_1, q_0))$, i=1, 2, …, m. Then, $\tilde{\delta}_{p_0 p_1} = \frac{1}{m}\sum_{i=1}^{m}\tilde{V}_i$ is the mean of *m* 'sample' restricted placement value $\tilde{V}_i$'*s*. Noticing that $\tilde{\delta}_{p_0 p_1}$ is a two-sample U-statistic, it follows from the asymptotic normality for U-statistic (Lehmann, 1998) that

$$\frac{1}{m\sigma_{mn}}\sum_{i=1}^{m}(\tilde{V}_i - \delta_{p_0 p_1}) \xrightarrow{L} \mathrm{N}(0, 1),$$

Where

$$\sigma_{mn}^2 = \frac{1}{m}\sigma_1^2 + \frac{1}{n}\sigma_0^2, \quad \sigma_1^2 = Var[V(X)], \quad \sigma_0^2 = Var\,[F\,(min(Y, q_0))].$$

Since both q_1, q_0 are unknown, $\tilde{V}_i$ is still unknown. The above normal approximation cannot be directly used to produce a confidence interval for the pAUC. Therefore Qin, Jin and Zhou (2007) introduced a bootstrap method to produce a confidence interval for pAUC.

For two diagnostic tests T_k (k=1, 2), we can use $\hat{\Delta}_{p_0 p_1} = \hat{\delta}^{(2)}_{p_0 p_1} - \hat{\delta}^{(1)}_{p_0 p_1}$ to estimate $\Delta_{p_0 p_1}$, where $\hat{\delta}^{(k)}_{p_0 p_1} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} I(Y_{kj} \geq X_{ki})I\left(X_{ki} \in (\hat{q}_{k1}, \hat{q}_{k0})\right)$, $\hat{q}_{kl} = \hat{F}_k^{-1}(1-p_l)$ (l = 0, 1), and $\hat{F}_k$ is the empirical distribution of F_k. It can be proved that

$$\sqrt{m+n}\left(\hat{\Delta}_{p_0 p_1} - \Delta_{p_0 p_1}\right) \xrightarrow{L} \mathrm{N}(0, \sigma^2_{p_0 p_1}),$$

where $\sigma^2_{p_0 p_1}$ is the asymptotic variance of $\hat{\Delta}_{p_0 p_1}$. Since $\sigma^2_{p_0 p_1}$ is an unknown function of F_k, G_k, F_k^{-1} and G_k^{-1}, the estimation of $\sigma^2_{p_0 p_1}$ involves in complex density and quantile estimation. This normal approximation cannot be directly used to produce a confidence interval for the $\Delta_{p_0 p_1}$. In next subsection, we will extend the method used in Qin, Jin and

Zhou (2006) to construct confidence intervals for the difference between two partial AUCS.

2.2 Bootstrap Method

Bootstrap method is a popular non-parametric method for constructing confidence intervals of unknown parameter; it can be applied to very complex problems. In this chapter we will propose use bootstrap method to construct confidence interval for the difference between two partial AUCS.

We draw a bootstrap resample $\{ X_{k1}^*, X_{k2}^*, \ldots, X_{km}^* \}$ of size *m* with replacement from$\{ X_{k1}, X_{k2}, \ldots, X_{km} \}$ and a separate bootstrap resample $\{ Y_{k1}^*, Y_{k2}^*, \ldots, Y_{kn}^* \}$ of size *n* with replacement from $\{ Y_{k1}, Y_{k2}, \ldots, Y_{kn} \}$. The partial AUC can be estimated by

$$\hat{\delta}_{p_0 p_1}^{(k)*} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} I(Y_{kj}^* \geq X_{ki}^*) I\left(X_{ki}^* \in (\hat{q}_{k1}^*, \hat{q}_{k0}^*)\right), \quad \text{k=1, 2,}$$

where $\hat{q}_{kl}^* = \hat{F}_k^{-1*}(1 - p_l)$ (l = 0, 1) is the $(1 - p_l)$-th sample quantile based on bootstrap resample $\{ X_{k1}^*, X_{k2}^*, \ldots, X_{km}^* \}$. Then the bootstrap estimate for the difference of two partial AUCs can be calculated as

$$\hat{\Delta}_{p_0 p_1}^* = \hat{\delta}_{p_0 p_1}^{(2)*} - \hat{\delta}_{p_0 p_1}^{(1)*}.$$

After *B* repetitions of above process, *B* bootstrap copies of $\hat{\Delta}_{p_0 p_1}^*$ are obtained

$$\{ \hat{\Delta}_{p_0 p_1 (b)}^* : \text{b=1, 2, …, B} \}.$$

The bootstrap estimator for the variance of $\hat{\Delta}_{p_0 p_1}$ is given by

$$V^* = \frac{1}{B-1} \sum_{b=1}^{B} (\hat{\Delta}_{p_0 p_1 (b)}^* - \overline{\Delta}_{p_0 p_1}^*)^2,$$

where $\overline{\Delta}^*_{p_0 p_1} = \frac{1}{B}\sum_{b=1}^{B} \hat{\Delta}^*_{p_0 p_1(b)}$.

Two bootstrap (1-α)100% confidence intervals for $\Delta_{p_0 p_1}$ can be proposed based on the bootstrap variance estimator V^*.

First one, called BS interval is defined as follows:

$$(\overline{\Delta}^*_{p_0 p_1} - z_{1-\alpha/2}\sqrt{V^*},\ \overline{\Delta}^*_{p_0 p_1} + z_{1-\alpha/2}\sqrt{V^*}).$$

Second one, called BT interval is given by

$$(\hat{\Delta}_{p_0 p_1} - z_{1-\alpha/2}\sqrt{V^*},\ \hat{\Delta}_{p_0 p_1} + z_{1-\alpha/2}\sqrt{V^*}).$$

CHAPTER III

Empirical Likelihood Based Confidence Interval for The difference between two partial AUCs

In this chapter, we will use empirical likelihood method to construct the confidence interval for the difference between two partial AUCs.

Empirical likelihood (EL) (Owen, 1990, 2001) also is a popular non-parametric method traditionally used for providing confidence intervals. The EL method has many advantages over other non-parametric methods. For example, it has better small sample performance than approaches based on normal approximation, it studentizes internally, thereby eliminating the need for a pivot. But the applications of EL method to the ROC study are relatively few. The main challenge of developing the EL-based theory for the difference between two partial AUCs is the standard EL method can't be applied directly when the underlying distributions are unknown (Qin and Zhou 2006) and the empirical log-likelihood ratio for the partial AUC is a sum of non-independent random variables (Qin, Jin and Zhou 2006). Hence, the standard EL theory cannot be directly applied in the partial AUC setting.

For test value X_k from a "non-diseased" subject, Dodd and Pepe (2003) defined the restricted placement value of X_k as

$$V_k(X_k) = (1 - G_k(X_k))I(X_k \in (q_{k1}, q_{k2})), \text{ k=1, 2,}$$

where $q_{kl} = F_k^{-1}(1 - p_l)$, $l = 1, 2$.

When the quantiles are unknown, we can use

$$\hat{V}_k(X_k) = (1 - \hat{G}_k(X_k))I(X_k \in (\hat{q}_{k1}, \hat{q}_{k2})), \text{ k=1, 2,}$$

where $\hat{q}_{kl} = \hat{F}_k^{-1}(1-p_l)$, l = 1, 2.

V_k can be interpreted as the restricted placement value of a given "non-diseased" test value X_k, in the survival function of the results of "diseased". It is evident that

$$E(V_k(X_k)) = p\{Y_k > X_k, X_k \in (q_{k1}, q_{k2})\} = pAUC_k(p_{0,} p_1) = \delta_{p_0 p_1}^{(k)}$$

Therefore,

$$\Delta_{p_0 p_1} = \delta_{p_0 p_1}^{(2)} - \delta_{p_0 p_1}^{(1)} = E(V_2(X_2) - V_1(X_1)).$$

Based on this relationship between the difference between two partial AUCs and the restricted placement values $V_1(X_1)$ and $V_2(X_2)$, the profile empirical likelihood for $\Delta_{p_0 p_1}$ can be defined as

$$L(\Delta_{p_0 p_1}) = \sup\{\prod_{k=1,2}\prod_{i=1}^{m} p_{kj} : \sum_{j=1}^{n} p_{ki} = 1,\ \sum_{i=1}^{m} p_{ki}(\hat{V}_{ki} - \delta_{p_0 p_1}^{(k)}),\ \sum_{i=1}^{m} p_{2i}\hat{V}_{2i} - \sum_{i=1}^{m} p_{1i}\hat{V}_{1i} = \Delta_{p_0 p_1}\},$$

where $\hat{V}_{ki} = \hat{V}_k(X_{ki})$, i=1, 2, …, m, k=1, 2.

Then the corresponding empirical log-likelihood ratio (ELR) for $\Delta_{p_0 p_1}$ is

$$l(\Delta_{p_0 p_1}) = 2[\sum_{i=1}^{m}\log(1 - 2\lambda(\hat{V}_{1i} - \delta_{p_0 p_1}^{(1)})) + \sum_{i=1}^{m}\log(1 + 2\lambda(\hat{V}_{2i} - \delta_{p_0 p_1}^{(2)}))],$$

where λ and $\delta_{p_0 p_1}^{(k)}$ (k=1, 2) are the solutions of the following equations:

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{1i} - \delta_{p_0 p_1}^{(1)}}{1 - 2\lambda(\hat{V}_{1i} - \delta_{p_0 p_1}^{(1)})} = 0 \qquad (1)$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{2i} - \delta_{p_0 p_1}^{(2)}}{1 + 2\lambda(\hat{V}_{2i} - \delta_{p_0 p_1}^{(2)})} = 0 \qquad (2)$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{2i}}{1 + 2\lambda(\hat{V}_{2i} - \delta_{p_0 p_1}^{(2)})} - \frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{1i}}{1 - 2\lambda(\hat{V}_{1i} - \delta_{p_0 p_1}^{(1)})} = \Delta_{p_0 p_1} \qquad (3)$$

Theorem 3.1: If $\Delta_{p_0 p_1}$ is the true value of the difference between two partial AUCs, and $\lim_{m,n\to\infty} \frac{m}{n} = \rho$ is a constant, then the limiting distribution of $l(\Delta_{p_0 p_1})$ is a scaled chi-square distribution with one degree of freedom.

$$C(\Delta_{p_0 p_1})\, l(\Delta_{p_0 p_1}) \xrightarrow{L} \chi_1^2,$$

where $C(\Delta_{p_0 p_1}) = \dfrac{(\sigma_{p_0 p_1}^{(1)\,2} + \sigma_{p_0 p_1}^{(2)\,2})/m}{\sigma_{p_0 p_1}^2/(m+n)}$, $\sigma_{p_0 p_1}^{(k)\,2} = Var[V_k(X_k)]$, k = 1, 2.

Using Theorem 3.1, two empirical and bootstrap based intervals for the difference between two partial AUCs can be constructed as follows:

The first hybrid empirical and bootstrap interval (**EL**) is defined as

$$R_\alpha(\Delta_{p_0 p_1}) = \{\Delta_{p_0 p_1} :\ \hat{C}(\Delta_{p_0 p_1})\, l(\Delta_{p_0 p_1}) \le \chi_1^2(1-\alpha)\},$$

where $\chi_1^2(1-\alpha)$ is the (1-α)-th quantile of the chi-square distribution χ_1^2, $\hat{C}(\Delta_{p_0 p_1})$ is an estimate for $C(\Delta_{p_0 p_1})$:

$$\hat{C}(\Delta_{p_0 p_1}) = \frac{(\hat{\sigma}_{p_0 p_1}^{(1)\,2} + \hat{\sigma}_{p_0 p_1}^{(2)\,2})/m}{V^*}, \qquad \hat{\sigma}_{p_0 p_1}^{(k)\,2} = \frac{1}{m-1}\sum_{i=1}^{m}(\hat{V}_{ki} - \frac{1}{m}\sum_{i=1}^{m}\hat{V}_{ki})^2, \text{ k=1, 2,}$$

and V^* is the bootstrap variance estimate defined in chapter II.

$R_\alpha(\Delta_{p_0 p_1})$ is an approximate confidence intervals for the difference between two partial AUCs with asymptotically correct coverage probability 1-α, i.e.,

$$P(\Delta_{p_0 p_1} \in R_\alpha(\Delta_{p_0 p_1})) = 1\text{-}\alpha + o(1).$$

We can solve the following equations to get the lower and upper bounds of the confidence interval for the difference between the two partial AUCs:

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{1i}-\delta_{p_0p_1}^{(1)}}{1-2\lambda(\hat{V}_{1i}-\delta_{p_0p_1}^{(1)})}=0 \tag{1}$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{2i}-\delta_{p_0p_1}^{(2)}}{1+2\lambda(\hat{V}_{2i}-\delta_{p_0p_1}^{(2)})}=0 \tag{2}$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{2i}}{1+2\lambda(\hat{V}_{2i}-\delta_{p_0p_1}^{(2)})}-\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{1i}}{1-2\lambda(\hat{V}_{1i}-\delta_{p_0p_1}^{(1)})}=\Delta_{p_0p_1} \tag{3}$$

$$\hat{C}(\Delta_{p_0p_1})\, l(\Delta_{p_0p_1}) = \chi_1^2(1-\alpha) \tag{4}$$

In these four equations, λ and $\delta_{p_0p_1}^{(k)}$ (k=1, 2) and $\Delta_{p_0p_1}$ are unknown and can be solved. The $\Delta_{p_0p_1}$ will have two solutions. The smaller one is the lower bound of the **EL** interval and larger one is the upper bound of the **EL** interval.

The second hybrid empirical and bootstrap interval (**HBEL**) is given by

$$R_\alpha^*(\Delta_{p_0p_1})=\{\Delta_{p_0p_1}:\ \ \hat{C}^*(\Delta_{p_0p_1})\, l(\Delta_{p_0p_1})\le\chi_1^2(1-\alpha)\},$$

where $\hat{C}^*(\Delta_{p_0p_1})=\dfrac{(\bar{\hat{\sigma}}_{p_0p_1}^{*(1)\ 2}+\bar{\hat{\sigma}}_{p_0p_1}^{*(2)\ 2})/m}{V^*}$, $\bar{\hat{\sigma}}_{p_0p_1}^{*(k)\ 2}$ is the mean of B bootstrap copies of $\hat{\sigma}_{p_0p_1}^{(k)\ 2}$ (k=1,2), and V^* is the bootstrap variance estimate defined in chapter II.

Similarly, $R_\alpha^*(\Delta_{p_0p_1})$ is an approximate confidence intervals for the difference between two partial AUCs with asymptotically correct coverage probability 1-α, i.e.,

$$P(\Delta_{p_0p_1}\in\ R_\alpha^*(\Delta_{p_0p_1}))=1-\alpha+o(1).$$

The lower and upper bound of **HBEL** interval can be obtained by solving the following equations:

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{1i}-\delta^{(1)}_{p_0p_1}}{1-2\lambda(\hat{V}_{1i}-\delta^{(1)}_{p_0p_1})}=0 \tag{5}$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{2i}-\delta^{(2)}_{p_0p_1}}{1+2\lambda(\hat{V}_{2i}-\delta^{(2)}_{p_0p_1})}=0 \tag{6}$$

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{2i}}{1+2\lambda(\hat{V}_{2i}-\delta^{(2)}_{p_0p_1})}-\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{V}_{1i}}{1-2\lambda(\hat{V}_{1i}-\delta^{(1)}_{p_0p_1})}=\Delta_{p_0p_1} \tag{7}$$

$$\hat{C}^{*}(\Delta_{p_0p_1})\, l(\Delta_{p_0p_1}) \;=\chi^2_1(1-\alpha) \tag{8}$$

CHAPTER IV

Simulation Study

In this chapter, we conduct a simulation study to evaluate coverage accuracy and interval length of the newly proposed four intervals for the difference $\Delta_{p_0p_1}$ of two pAUCs. In the study, the difference $\Delta_{p_0p_1}$ between two pAUCs is taken to be 0 and 0.2. We generate 1000 random samples of size *n* from $G(y_1, y_2)$ for test responses of diseased patients, and another set of independent random samples of size *m* from $F(x_1, x_2)$ for test responses of non-diseased patients.

The distribution $F(x_1, x_2)$ is chosen to be a bivariate normal distribution having means $E(X_1) = 0$, $E(X_2) = 0$ with a common standard deviation 1 and correlation ρ. The distribution $G(y_1, y_2)$ is chosen to be a bivariate normal distribution having means $E(Y_1) = \mu_1$, $E(Y_2) = \mu_2$ with a common standard deviation 2 and correlation ρ. μ_1 and μ_2 are calculated by solving the following equations

$$\delta_{p_0p_1}^{(k)} = \int_{p_0}^{p_1} R_k(p)dp \quad \text{with } R_k(p) = 1 - G_k(F_k^{-1}(1-p)), \text{ k = 1, 2.}$$

When $\Delta_{p_0p_1}$ = 0, we choose three groups of $(\hat{\delta}_{p_0p_1}^{(1)}, \hat{\delta}_{p_0p_1}^{(2)})$ to calculate three groups of (μ_1, μ_2) and generate random samples from the $G(y_1, y_2)$:

(i) $\delta_{(0,0.4)}^{(2)} = \delta_{(0,0.4)}^{(1)} = 0.2$ with (p_0, p_1) = (0, 0.4),

(ii) $\delta_{(0,0.7)}^{(2)} = \delta_{(0,0.7)}^{(1)} = 0.45$ with (p_0, p_1) = (0, 0.7),

(iii) $\delta_{(0.05,0.50)}^{(2)} = \delta_{(0.05,0.50)}^{(1)} = 0.26$ with (p_0, p_1) = (0.05, 0.50).

When $\Delta_{p_0 p_1}$ = 0.2, we also choose three groups of $(\hat{\delta}^{(1)}_{p_0 p_1}, \hat{\delta}^{(2)}_{p_0 p_1})$ to calculate three groups of (μ_1, μ_2) and generate random samples from the $G(y_1, y_2)$:

(i) $\delta^{(2)}_{(0,0.4)} = 0.37, \delta^{(1)}_{(0,0.4)} = 0.17$ with (p_0, p_1) = (0, 0.4),

(ii) $\delta^{(2)}_{(0,0.7)} = 0.61, \delta^{(1)}_{(0,0.7)} = 0.41$ with (p_0, p_1) = (0, 0.7),

(iii) $\delta^{(2)}_{(0.05,0.50)} = 0.39, \delta^{(1)}_{(0.05,0.50)} = 0.19$ with (p_0, p_1) = (0.05, 0.50).

In the bootstrap step, we draw B = 150 bootstrap re-samples from the original samples. We construct both 90% and 95% confidence intervals for $\Delta_{p_0 p_1}$. The results of the simulation study are shown in Table I to Table VIII. From these tables, the following observations were made.

(1) When the correlation ρ = 0 and $\Delta_{p_0 p_1}$ = 0, the four proposed intervals have similar coverage probabilities but the hybrid empirical likelihood and bootstrap intervals (EL and HBEL) have slightly shorter interval length.

(2) When $\Delta_{p_0 p_1}$ > 0, all the intervals over-cover the true difference between two pAUCs when sample sizes are small. As the sample sizes increase, the coverage probabilities of all the intervals approach to the nominal level. Although in most time all the interval have similar coverage probabilities, the EL and HBEL intervals have much shorter interval length than bootstrap (BS and BT) intervals.

(3) When the correlation is positive (ρ = 0.3), bigger sample sizes ($m, n \geq 150$) are needed to get better coverage accuracy for all the intervals.

In summary, the simulation study indicates that the hybrid empirical likelihood and bootstrap based intervals perform better than the bootstrap intervals when two partial AUCs are different. When there is no difference between two partial AUCs, the four

proposed intervals have similar performance. Therefore, we recommend the use of hybrid empirical likelihood and bootstrap method for construction of confidence interval of difference between two pAUCs when the underlying distributions for diseased and non-diseased populations are unknown.

CHAPTER V

Dermatoscope Example

Malignant melanoma (MM) is one of the most deadly kinds of skin disease. Melanomas of less than 1mm are not likely to have spread to the lymph nodes or to other parts of the body, called early stage; they have a very good chance of cure. The thinner the melanoma, the better chance of a complete cure. Early diagnosis of malignant melanoma is essential for cure.

Dermatoscopy is a hand- held instrument for skin surface microscopy at 10 times magnification and is a noninvasive diagnostic technique for the early diagnosis of melanoma and the evaluation of other pigmented and non-pigmented lesions on the skin that are not as well seen with the unaided eye [www.medterms.com]. Stolz et al (1994) studied the accuracy of clinical evaluations with or without the aid of Dermatoscopy in detecting MM by using the ABCD rule (Asymmetry, irregular border, different colors, and Diameter larger than 6mm). In this study, two tests were applied for detecting MM; the first test is the clinical assessment without the aid of dermatoscopy, and the second test is the clinical assessment with the aid of dermatoscopy. The data set we used here includes 21 patients with MM and 51 patients with benign melanocytic lesions; all patients have both tests results. The objective of this study is to find out whether the aid of dermatoscopy can improve for detecting MM. We estimate the difference of two pAUCs of the two tests and construct confidence intervals for the difference by using the proposed methods. The 90% and 95% confidence intervals for the difference between two pAUCs over three intervals of FPR are shown in Table IX and Table X.

The estimates of the differences between two pAUCs over the three intervals (0, 0.4), (0, 0.7) and (0.05, 0.50) of FPR for the two tests are all close to 0. Also, all the confidence intervals for the differences contain zero. Therefore, we conclude that there is no significant advantage in adopting the clinical assessment with the aid of dermatoscopy in detecting MM. The same conclusion has been obtained in Qin, Hsu and Zhou (2006) where they compared those two tests by using the sensitivities at a fixed level of specificity.

CHAPTER VI

Discussion and Conclusion

Comparing the accuracy of two continuous-scale tests is increasingly important when a new test is developed and marketed. There are many ways to do such a comparison. For example, we can compare the sensitivities at a fixed common specificity or we can compare the areas under the ROC curves. But traditional ways of comparing entire areas under two ROC curves are not sensitive when two ROC curves cross each other. Comparing the areas under two ROC curves on the interested FPR interval is a more appropriate way to compare the accuracy of two diagnostic tests. In this thesis, we have proposed two bootstrap based confidence intervals (BS and BT) and two hybrid empirical likelihood and bootstrap confidence intervals (EL and HBEL). The simulation study indicates that two hybrid empirical likelihood and bootstrap intervals performed better than the bootstrap intervals in most cases, especially when there is a difference between two pAUCs. The proposed hybrid empirical likelihood and bootstrap based method combines the power of both bootstrap and empirical likelihood methods. The unknown scale constant in the empirical likelihood theorem can be conveniently and accurately estimated by using bootstrap method. The confidence intervals can be constructed by using the empirical likelihood theorem. Based on this study, we recommend the use of the proposed hybrid empirical likelihood and bootstrap confidence intervals for the difference between two partial AUCs when the underlying distributions for diseased and non-diseased populations are unknown.

REFERENCES

Bamber DC (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic curve graph. *Journal of Mathematical Psychology*, **12**:387-415.

Cai T and Dodd L (2006). Regression analysis for the partial area under the ROC curve. Harvard University Biostatistics Working Paper Series.

DeLong ER, DeLong DM, Clarke-Pearson DL (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* **44**:837-45.

Dodd L and Pepe MS (2003). Partial AUC Estimation and Regression. *Biometrics.* **59**: *614-623.*

Hanley JA, McNeil BJ (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* **148**:839-43.

Hilden J (1991). The area under the ROC curve and its competitors. *Medical Decision making* **11**:95-101.

Jiang Y, Metz CE and Nishikawa RM (1996). A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology* **143**:29-36.

Lilienfeld AM (1974). Some limitations and problems of screening for cancer. *Cancer* **36**:1720-1724

McClish DK (1989). Analyzing a portion of the ROC curve. *Medical Decision Making* **9**: 190–195.

Molodianovitch K, Faraggi D, Reiser B (2006). Comparing the areas under two correlated ROC curves: parametric and non-parametric approaches. *Biom J.* **48**:745-57

Owen A. (2001). Empirical likelihood. Chapman & Hall/CRC, New York.

Pepe MS and Cai T (2002). The analysis of placement values for evaluating discriminatory measures. *Biometrics* **60**:528-535.

Pepe MS (2003). The statistical evaluation of medical tests for classification and biomarkers. Oxford University Press, Cary, NC.

Pepe MS (2003). The statistical evaluation of diagnostic tests and prediction. Oxford University Press, Cary, NC.

Qin, GS and Zhou XH (2006). Empirical Likelihood Inference for the Area Under the ROC Curve. *Biometrics* **62**: 613-622.

Qin, GS, Hsu, YS, and Zhou, XH (2006). New confidence intervals for the difference between two sensitivities at a fixed level of specificity. *Statistics in Medicine* **25**: 3487-3502.

Qin, GS, Jin, XP and Zhou, XH (2006). Bootstrap and Empirical Likelihood-based inference for the Partial Area Under the ROC Curve. *Manuscript.*

Shapiro DE (1999). The Interpretation of Diagnostic Tests. *Statistical Methods in Medical Research,* **8**:113-134

Thompson ML, and Zucchini W (1989). On the statistical analysis of ROC curvers. *Statistics in Medicine*. **8**:1277-1290.

Wieand S (1989). A Family of Non-Parametric Statistic for Comparing Diagnostic Markers with Paired or Unpaired Data. *Biometrika,* **76**:585-592

Walsh SJ(1997). Limitations to the robustness of binormal ROC curves: effects of model misspecification and location of the decision thresholds on bias, precision, size and power. *Statistics in Medicine*. **16**: 669-679

Zhang DD, Zhou XH, Freeman DH and 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. *Statistics in Medicine.* **21**:701-15

www.medterms.com

APPENDIX

APPENDIX A: Simulation Tables

Table I: Level of 95 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0	(20, 20)	HBEL	0.965	0.2810	0.963	0.4118	0.964	0.3203
		EL	0.977	0.2959	0.975	0.4099	0.970	0.3297
		BT	0.981	0.2996	0.971	0.4343	0.974	0.3372
		BS	0.976	0.2996	0.963	0.4343	0.965	0.3372
	(50, 50)	HBEL	0.954	0.1651	0.951	0.2551	0.962	0.1843
		EL	0.959	0.1693	0.953	0.2586	0.964	0.1869
		BT	0.963	0.1699	0.950	0.2647	0.968	0.1879
		BS	0.958	0.1699	0.946	0.2647	0.959	0.1879
	(80, 80)	HBEL	0.944	0.1274	0.942	0.1970	0.950	0.1448
		EL	0.948	0.1294	0.943	0.1975	0.951	0.1461
		BT	0.950	0.1297	0.940	0.1981	0.954	0.1466
		BS	0.950	0.1297	0.939	0.1981	0.942	0.1466
	(50, 20)	HBEL	0.922	0.2280	0.947	0.3486	0.948	0.2575
		EL	0.929	0.2370	0.951	0.3529	0.952	0.2642
		BT	0.942	0.2390	0.945	0.3658	0.955	0.2673
		BS	0.938	0.2390	0.943	0.3658	0.950	0.2673
	(80, 50)	HBEL	0.936	0.1510	0.945	0.2364	0.957	0.1725
		EL	0.943	0.1540	0.947	0.2376	0.961	0.1744
		BT	0.947	0.1545	0.945	0.2389	0.963	0.1752
		BS	0.937	0.1545	0.947	0.2389	0.963	0.1752

Table II: Level of 95 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0.2	(20, 20)	HBEL	0.987	0.1277	0.984	0.2342	0.985	0.1528
		EL	0.987	0.1308	0.976	0.2089	0.988	0.1555
		BT	0.988	0.2693	0.974	0.3906	0.990	0.3084
		BS	0.981	0.2693	0.973	0.3906	0.976	0.3084
	(50, 50)	HBEL	0.968	0.0728	0.959	0.1165	0.963	0.0917
		EL	0.972	0.0736	0.959	0.1162	0.964	0.0925
		BT	0.971	0.1424	0.958	0.2214	0.964	0.1682
		BS	0.960	0.1424	0.956	0.2214	0.959	0.1682
	(80, 80)	HBEL	0.955	0.0513	0.956	0.0822	0.965	0.0602
		EL	0.956	0.0517	0.956	0.0821	0.967	0.0605
		BT	0.957	0.1062	0.955	0.1708	0.969	0.1281
		BS	0.957	0.1062	0.959	0.1708	0.963	0.1281
	(50, 20)	HBEL	0.951	0.0942	0.944	0.1603	0.958	0.1243
		EL	0.953	0.0957	0.945	0.1613	0.959	0.1262
		BT	0.957	0.1936	0.941	0.3209	0.961	0.2332
		BS	0.953	0.1936	0.945	0.3209	0.951	0.2332
	(80, 50)	HBEL	0.949	0.0595	0.947	0.1022	0.953	0.0740
		EL	0.950	0.0600	0.948	0.1025	0.954	0.0745
		BT	0.953	0.1252	0.945	0.2042	0.956	0.1528
		BS	0.948	0.1253	0.944	0.2042	0.953	0.1528

Table III: Level of 95 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0.3$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0	(50, 50)	HBEL	0.976	0.1650	0.981	0.2533	0.982	0.1846
		EL	0.978	0.1693	0.981	0.2543	0.982	0.1872
		BT	0.981	0.1699	0.979	0.2558	0.987	0.1882
		BS	0.982	0.1699	0.977	0.2558	0.979	0.1882
	(80, 80)	HBEL	0.974	0.1278	0.972	0.1974	0.974	0.1447
		EL	0.976	0.1299	0.972	0.1979	0.975	0.1460
		BT	0.978	0.1302	0.972	0.1985	0.975	0.1465
		BS	0.977	0.1302	0.969	0.1985	0.972	0.1465
	(150, 150)	HBEL	0.971	0.0917	0.980	0.1416	0.977	0.1025
		EL	0.973	0.0925	0.979	0.1418	0.978	0.1030
		BT	0.974	0.0926	0.979	0.1420	0.978	0.1032
		BS	0.973	0.0926	0.974	0.1420	0.977	0.1032
	(80, 50)	HBEL	0.970	0.1516	0.975	0.2355	0.965	0.1730
		EL	0.974	0.1546	0.975	0.2368	0.969	0.1750
		BT	0.974	0.1551	0.975	0.2378	0.972	0.1757
		BS	0.979	0.1551	0.975	0.2378	0.967	0.1757
	(150, 80)	HBEL	0.974	0.1171	0.976	0.1829	0.965	0.1320
		EL	0.977	0.1185	0.976	0.1836	0.964	0.1330
		BT	0.979	0.1188	0.975	0.1840	0.965	0.1333
		BS	0.978	0.1188	0.972	0.1840	0.963	0.1333

Table IV: Level of 95 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0.3$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0.2	(50, 50)	HBEL	0.982	0.0747	0.981	0.1158	0.983	0.0899
		EL	0.982	0.0754	0.981	0.1157	0.983	0.0906
		BT	0.983	0.1425	0.981	0.2261	0.984	0.1683
		BS	0.983	0.1425	0.977	0.2261	0.981	0.1683
	(80, 80)	HBEL	0.983	0.0521	0.970	0.0878	0.976	0.0644
		EL	0.984	0.0525	0.970	0.0880	0.977	0.0648
		BT	0.984	0.1062	0.970	0.1741	0.978	0.1285
		BS	0.982	0.1062	0.967	0.1741	0.971	0.1285
	(150, 150)	HBEL	0.963	0.0348	0.972	0.0632	0.968	0.0469
		EL	0.963	0.0349	0.972	0.0632	0.968	0.0471
		BT	0.965	0.0745	0.969	0.1241	0.968	0.0894
		BS	0.961	0.0745	0.967	0.1241	0.967	0.0894
	(80, 50)	HBEL	0.971	0.0595	0.972	0.1088	0.971	0.0742
		EL	0.972	0.0601	0.973	0.1092	0.973	0.0747
		BT	0.971	0.1250	0.968	0.2186	0.974	0.1520
		BS	0.970	0.1250	0.969	0.2186	0.972	0.1520
	(150, 80)	HBEL	0.958	0.0475	0.964	0.0828	0.962	0.0590
		EL	0.961	0.0478	0.965	0.0829	0.963	0.0592
		BT	0.961	0.0941	0.960	0.1616	0.966	0.1143
		BS	0.957	0.0941	0.961	0.1616	0.963	0.1143

Table V: Level of 90 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0	(20, 20)	HBEL	0.919	0.2374	0.939	0.3522	0.935	0.2704
		EL	0.930	0.2506	0.939	0.3529	0.942	0.2792
		BT	0.942	0.2522	0.932	0.3656	0.944	0.2832
		BS	0.934	0.2522	0.925	0.3656	0.933	0.2832
	(50, 50)	HBEL	0.922	0.1390	0.912	0.2130	0.925	0.1556
		EL	0.927	0.1424	0.915	0.2137	0.927	0.1578
		BT	0.927	0.1428	0.908	0.2148	0.929	0.1582
		BS	0.925	0.1428	0.902	0.2148	0.919	0.1582
	(80, 80)	HBEL	0.896	0.1078	0.922	0.1659	0.900	0.1214
		EL	0.902	0.1095	0.922	0.1663	0.906	0.1225
		BT	0.906	0.1097	0.922	0.1667	0.907	0.1228
		BS	0.902	0.1097	0.919	0.1667	0.896	0.1228
	(50, 20)	HBEL	0.880	0.1921	0.910	0.2959	0.907	0.2168
		EL	0.891	0.1997	0.915	0.3002	0.912	0.2223
		BT	0.902	0.2010	0.906	0.3077	0.915	0.2240
		BS	0.894	0.2010	0.905	0.3077	0.908	0.2240
	(80, 50)	HBEL	0.889	0.1276	0.889	0.1987	0.900	0.1445
		EL	0.894	0.1302	0.889	0.1998	0.903	0.1463
		BT	0.898	0.1304	0.887	0.2006	0.908	0.1468
		BS	0.891	0.1304	0.885	0.2006	0.902	0.1468

Table VI: Level of 90 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0.2	(20, 20)	HBEL	0.965	0.1197	0.953	0.1871	0.953	0.1261
		EL	0.966	0.1228	0.953	0.1851	0.957	0.1283
		BT	0.969	0.2250	0.946	0.3305	0.960	0.2577
		BS	0.962	0.2250	0.944	0.3305	0.936	0.2577
	(50, 50)	HBEL	0.938	0.0638	0.927	0.0985	0.942	0.0744
		EL	0.939	0.0645	0.926	0.0986	0.943	0.0750
		BT	0.940	0.1201	0.920	0.1892	0.944	0.1408
		BS	0.938	0.1201	0.920	0.1892	0.932	0.1408
	(80, 80)	HBEL	0.923	0.0446	0.915	0.0715	0.927	0.0521
		EL	0.924	0.0449	0.915	0.0715	0.928	0.0523
		BT	0.924	0.0893	0.911	0.1463	0.928	0.1079
		BS	0.923	0.0893	0.909	0.1463	0.915	0.1079
	(50, 20)	HBEL	0.912	0.0869	0.911	0.1363	0.893	0.1045
		EL	0.916	0.0882	0.912	0.1370	0.896	0.1061
		BT	0.919	0.1630	0.907	0.2700	0.901	0.1963
		BS	0.913	0.1630	0.909	0.2700	0.879	0.1963
	(80, 50)	HBEL	0.921	0.0539	0.892	0.0879	0.915	0.0665
		EL	0.922	0.0544	0.892	0.0880	0.916	0.0670
		BT	0.929	0.1054	0.891	0.1749	0.916	0.1276
		BS	0.919	0.1054	0.887	0.1749	0.900	0.1276

Table VII: Level of 90 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0.3$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0	(50, 50)	HBEL	0.954	0.1387	0.966	0.2138	0.954	0.1550
		EL	0.961	0.1422	0.966	0.2146	0.956	0.1572
		BT	0.965	0.1426	0.964	0.2154	0.957	0.1578
		BS	0.959	0.1426	0.962	0.2154	0.953	0.1578
	(80, 80)	HBEL	0.935	0.1072	0.950	0.1654	0.949	0.1219
		EL	0.939	0.1090	0.950	0.1659	0.951	0.1230
		BT	0.939	0.1091	0.948	0.1662	0.952	0.1233
		BS	0.936	0.1091	0.945	0.1662	0.949	0.1233
	(150,150)	HBEL	0.952	0.0772	0.953	0.1193	0.944	0.0861
		EL	0.953	0.0779	0.954	0.1195	0.945	0.0866
		BT	0.955	0.0780	0.951	0.1197	0.947	0.0867
		BS	0.950	0.0780	0.947	0.1197	0.946	0.0867
	(80, 50)	HBEL	0.935	0.1279	0.941	0.1994	0.946	0.1450
		EL	0.939	0.1305	0.941	0.2004	0.949	0.1467
		BT	0.941	0.1308	0.941	0.2007	0.951	0.1472
		BS	0.934	0.1308	0.936	0.2007	0.945	0.1472
	(150, 80)	HBEL	0.940	0.0984	0.940	0.1541	0.935	0.1109
		EL	0.942	0.0995	0.941	0.1547	0.936	0.1117
		BT	0.947	0.0997	0.939	0.1549	0.938	0.1119
		BS	0.938	0.0997	0.938	0.1549	0.934	0.1119

Table VIII: Level of 90 per cent confidence interval for $\Delta_{p_0 p_1}$. Bivariate normal distribution with $\rho = 0.3$.

True $\Delta_{p_0 p_1}$	Sample size (m,n)	Method	(p_0, p_1)=(0-0.4)		(p_0, p_1)=(0-0.7)		(p_0, p_1)=(0.05-0.5)	
			Coverage probability	Length	Coverage probability	Length	Coverage probability	Length
0.2	(50, 50)	HBEL	0.956	0.0602	0.946	0.0914	0.965	0.0775
		EL	0.958	0.0608	0.946	0.0912	0.965	0.0781
		BT	0.958	0.1197	0.948	0.1902	0.965	0.1411
		BS	0.954	0.1197	0.946	0.1902	0.963	0.1411
	(80, 80)	HBEL	0.940	0.0438	0.935	0.0766	0.948	0.0531
		EL	0.941	0.0441	0.935	0.0765	0.948	0.0534
		BT	0.942	0.0889	0.935	0.1457	0.949	0.1075
		BS	0.938	0.0889	0.934	0.1457	0.935	0.1075
	(150, 150)	HBEL	0.925	0.0310	0.944	0.0510	0.954	0.0380
		EL	0.925	0.0312	0.945	0.0510	0.955	0.0381
		BT	0.925	0.0622	0.941	0.1040	0.954	0.0748
		BS	0.925	0.0622	0.938	0.1040	0.944	0.0748
	(80, 50)	HBEL	0.931	0.0542	0.946	0.0868	0.931	0.0648
		EL	0.934	0.0546	0.947	0.0871	0.936	0.0653
		BT	0.933	0.1053	0.942	0.1752	0.936	0.1279
		BS	0.929	0.1053	0.940	0.1752	0.930	0.1279
	(150, 80)	HBEL	0.909	0.0413	0.936	0.0651	0.931	0.0492
		EL	0.913	0.0415	0.936	0.0652	0.933	0.0494
		BT	0.912	0.0792	0.935	0.1350	0.933	0.0957
		BS	0.909	0.0792	0.932	0.1350	0.925	0.0957

APPENDIX B: Real Data Analysis Tables

Table IX: Dermatoscope Example
Level of 95 per cent confidence interval for $\Delta_{p_0 p_1}$

Method	*CI & Length*	(p_0, p_1)= (0-0.4)	(p_0, p_1)= (0-0.7)	(p_0, p_1)= (0.05-0.5)
HBEL	Lower-Limit	-0.10722	-0.11170	-0.09468
	Upper-Limit	0.10722	0.11170	0.09468
	CI_Length	0.21444	0.22340	0.18936
EL	Lower-Limit	-0.10913	-0.11129	-0.09512
	Upper-Limit	0.10913	0.11129	0.09512
	CI_Length	0.21826	0.22258	0.19024
BT	Lower-Limit	-0.10931	-0.11155	-0.09533
	Upper-Limit	0.10931	0.11155	0.09533
	CI_Length	0.21862	0.22310	0.19066
BS	Lower-Limit	-0.10945	-0.11714	-0.10025
	Upper-Limit	0.10917	0.10596	0.09041
	CI_Length	0.21862	0.22310	0.19066
Estimate of $\Delta_{p_0 p_1}$		-0.00014	-0.00559	-0.00492

Table X: Dermatoscope Example
Level of 90 per cent confidence interval for $\Delta_{p_0 p_1}$

Method	*CI & Length*	(p_0, p_1)= (0-0.4)	(p_0, p_1)= (0-0.7)	(p_0, p_1)= (0.05-0.5)
HBEL	Lower-Limit	-0.08697	-0.08704	-0.08132
	Upper-Limit	0.08697	0.08704	0.08132
	CI_Length	0.17394	0.17408	0.16263
EL	Lower-Limit	-0.08837	-0.08674	-0.08163
	Upper-Limit	0.08837	0.08674	0.08163
	CI_Length	0.17674	0.17349	0.16326
BT	Lower-Limit	-0.08846	-0.08687	-0.08176
	Upper-Limit	0.08846	0.08687	0.08176
	CI_Length	0.17693	0.17374	0.16353
BS	Lower-Limit	-0.08675	-0.08216	-0.08214
	Upper-Limit	0.09018	0.09158	0.08139
	CI_Length	0.17693	0.17374	0.16353
Estimate of $\Delta_{p_0 p_1}$		0.00171	0.00471	-0.00037

APPENDIX C: The Splus code for simulation studies

```
#######################part 1: Functions#####################

## Function R(p)##
  Rp<-function(p, muy, stdd) 1-pnorm(qnorm(1-p),muy, stdd)


## solveNonlinear##
##nlmin can be used to solve a system of nonlinear equations:
  solveNonlinear <- function(f, y0, x, ...){
  # solve f(x) = y0
  # x is vector of initial guesses, same length as y0
  # ... are additional arguments to nlmin (not to f)
  g <- function(x, y0, f) sum((f(x) - y0)^2)
  g$y0 <- y0   # set g's default value for y0
  g$f <- f     # set g's default value for f
  nlmin(g, x, max.fcal = 10000, max.iter = 10000, ...)
  }

##calculate x[1]=y1.mean x[2]=y2.mean##
mu <- function(x){
c( integrate(Rp, muy=x[1], stdd=y1.sd, lower=p0, upper = p1)$integral,
  integrate(Rp, muy=x[2], stdd=y2.sd, lower=p0, upper = p1)$integral )
}

##function for sigma##
my.mean <- function(vv) mean((vv-mean(vv))^2) ;


##solve  x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda
f <- function(x) c( mean((V.hat[,1]-x[1])/(1-2*x[3]*(V.hat[,1]-x[1]))),
                    mean((V.hat[,2]-x[2])/(1+2*x[3]*(V.hat[,2]-x[2]))),
  mean(V.hat[,2]/(1+2*x[3]*(V.hat[,2]-x[2])))-mean(V.hat[,1]/(1-
    2*x[3]*(V.hat[,1]-x[1]))))



##x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda x[4]: delta using C.deltap0p1.hat
g2 <- function(x) c( mean((V.hat[,1]-x[1])/(1-2*x[3]*(V.hat[,1]-x[1]))),
                     mean((V.hat[,2]-x[2])/(1+2*x[3]*(V.hat[,2]-x[2]))),
       mean(V.hat[,2]/(1+2*x[3]*(V.hat[,2]-x[2])))-mean(V.hat[,1]/(1-
    2*x[3]*(V.hat[,1]-x[1])))-x[4],
      C.deltap0p1.hat*(2*(sum( log(abs(1-2*x[3]*(V.hat[,1]-x[1]))))+sum(
    log(abs(1+2*x[3]*(V.hat[,2]-x[2])))))-CritVal)


##x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda x[4]: delta using C.deltap0p1
g1 <- function(x) c( mean((V.hat[,1]-x[1])/(1-2*x[3]*(V.hat[,1]-x[1]))),
                     mean((V.hat[,2]-x[2])/(1+2*x[3]*(V.hat[,2]-x[2]))),
       mean(V.hat[,2]/(1+2*x[3]*(V.hat[,2]-x[2])))-mean(V.hat[,1]/(1-
    2*x[3]*(V.hat[,1]-x[1])))-x[4],
       C.deltap0p1*(2*(sum( log(abs(1-2*x[3]*(V.hat[,1]-x[1]))))+sum(
    log(abs(1+2*x[3]*(V.hat[,2]-x[2])))))-CritVal)


##function for deltapAUC.hat##
deltapAUC <- function(X1X2, Y1Y2, p0, p1, m){

  # Caculate X Quantile of 1-pi (i=0,1) for q.hat
  q0.1.hat<-quantile(X1X2[,1],1-p0);
  q0.2.hat<-quantile(X1X2[,2],1-p0);
```

```
  q1.1.hat<-quantile(X1X2[,1],1-p1);
  q1.2.hat<-quantile(X1X2[,2],1-p1);

   # Caculate V(ki).hat & C(deltap0p1).hat
   V.hat<-matrix(, m, 2)

  for (i in 1 : m){
     V.hat[i,1]<-(1-mean(Y1Y2[,1]<= X1X2[i,1]))*(q1.1.hat <=
    X1X2[i,1])*(X1X2[i,1]<=q0.1.hat)
     V.hat[i,2]<-(1-mean(Y1Y2[,2]<= X1X2[i,2]))*(q1.2.hat <=
    X1X2[i,2])*(X1X2[i,2]<=q0.2.hat)
     }
  delta.pAUC.hat<-mean(V.hat[,2])-mean(V.hat[,1])
  C.deltap0p1.hat<-(my.mean(V.hat[,1])+my.mean(V.hat[,2]))/(m*Vstar)

  list(delta.pAUC.hat, C.deltap0p1.hat, V.hat)
 }


##bootstrap function##
booth.trap <- function(B, X1X2, Y1Y2, m, n, p0, p1){
   delta.pAUC<-0;
  sigma <- matrix(,B, 2)
  for (b in 1:B) {
        X1B <- sample(X1X2[,1], m, replace = T)
        X2B <- sample(X1X2[,2], m, replace = T)
        Y1B <- sample(Y1Y2[,1], n, replace = T)
        Y2B <- sample(Y1Y2[,2], n, replace = T)

        q0B.1.hat<-quantile(X1B, c(1-p0)) # hatq0, hatq1: sample quantiles of F
        q0B.2.hat<-quantile(X2B, c(1-p0)) # hatq0, hatq1: sample quantiles of F
        q1B.1.hat<-quantile(X1B, c(1-p1))
        q1B.2.hat<-quantile(X2B, c(1-p1))

        VB    <- matrix(,m, 2)

for (i in 1:m)
    {
VB[i,1]<- (1-mean(Y1B <= X1B[i])) *(q1B.1.hat <= X1B[i])*(X1B[i] <= q0B.1.hat)
VB[i,2]<- (1-mean(Y2B <= X2B[i])) *(q1B.2.hat <= X2B[i])*(X2B[i] <= q0B.2.hat)
    }

       sigma[b,1]<-my.mean(VB[,1])
       sigma[b,2]<-my.mean(VB[,2])

       delta.pAUC[b]<-mean(VB[,2])-mean(VB[,1])
       }
  list(delta.pAUC, sigma)
  }

############################ End function part ###################

############################ Part2: initial value#################
iter<-1000
B=150
rho=0
#rho=0.3
m<-50; n<-20;
y1.sd<-2; y2.sd<-2;

levelc<-0.95
#levelc<-0.90
```

```
CritVal<-qchisq(levelc,1)
Z<-qnorm(1-(1-levelc)/2)

y1.mean<-y2.mean<-0
p0<-0 ; p1<-0.4

 pAUC1 <- 0.2
 pAUC2 <- 0.2
 deltapAUC.true<- pAUC2-pAUC1

 S<-solveNonlinear(mu, c(  pAUC1, pAUC2), c(0.1, 0.1))
  y1.mean<-S$x[1]
  y2.mean<-S$x[2]

####################  End part2  ######################



#################### Part3:  Loop    ###################
CovCount<-c(0,0,0,0)
CIL<-c(0,0,0)
#LP<-c(0,0,0,0)
#UP<-c(0,0,0,0)


for ( i12 in c(1:iter)){

# generate non-diseased population F(X1, X2)
# the sample from 2-dimensinal multinormal distribution with mean 0 and std=1
   X1X2<-rmvnorm(m, mean=c(0,0), cov=matrix(c(1,rho,rho,1),2))

# generate  diseased population G(Y1,Y2)
# the sample from 2-dimensinal multinormal distribution with mean
#(y1.mean,y2.mean) and std=(y1.sd,y2.sd)
  Y1Y2<-rmvnorm(n, mean=c(y1.mean,y2.mean),
   cov=matrix(c(y1.sd^2,rho*y1.sd*y2.sd, rho*y1.sd*y2.sd, y2.sd^2),2))

  ##### 1. bootstrap ######
  boot.list<- booth.trap(B, X1X2, Y1Y2, m, n, p0, p1)

  delta.pAUC <- boot.list[[1]]
  sigma <- boot.list[[2]]

  delta.pAUCbar.B<-mean(delta.pAUC); delta.pAUCbar.B  # Estimate mean
   difference of two pAUCs by bootstrap

  Vstar<-var(delta.pAUC);        #Variance of delta.pAUC by bootstrap

  C.deltap0p1<-(mean(sigma[,1])+mean(sigma[,2]))/(m*Vstar);

  #####END OF BOOTSTRAP#######


  ###### 2. Caculate delta.pAUC.hat######

  delta.pAUC.hat.list <- deltapAUC(X1X2, Y1Y2, p0, p1, m)

  delta.pAUC.hat <- delta.pAUC.hat.list[[1]]
  C.deltap0p1.hat <- delta.pAUC.hat.list[[2]]
  V.hat <- delta.pAUC.hat.list[[3]]

  ########END OF 2.  #########
```

```
  ###### 3. caculate L.deltap0p1######
  # EL Method #

  #x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda

  x<-solveNonlinear(f, c( 0,0, deltapAUC.true), c(0.1, 0.2, 0))

  p0p1.1<-x$x[1];
  p0p1.2<-x$x[2];
  lamda<-x$x[3];

  l.delta.p0p1<-2*(sum( log(1-2*lamda*(V.hat[,1]-p0p1.1)))+sum(
   log(1+2*lamda*(V.hat[,2]-p0p1.2))))


  Vel<-C.deltap0p1*l.delta.p0p1;
  Vel.hat<-C.deltap0p1.hat*l.delta.p0p1;
####END OF 3. ######

###### 4. Caculate C.I and coverage######

## compute the HBEL interval(Vel from bootstrap)##
      if (Vel <= CritVal)
      CovCount[1]<-CovCount[1]+1

     #x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda x[4]: delta

       bd<-solveNonlinear(g1, c( 0,0,0,0), c(0.3, 0.1, 0.001, -0.9))
      #low.HBEL<-bd$x[4]  # lower limit of the CI

      b<-solveNonlinear(g1, c( 0,0,0,0), c(0.1, 0.3, 0.001, 0.999))
      #up.HBEL<-b$x[4]    # upper limit of the CI

     #low and up band HBEL
     #LP[1]<- LP[1]+bd$x[4]
      #UP[1]<- UP[1]+ b$x[4]

      # The length of HBEL CI
      CIL[1]<- CIL[1]+(b$x[4]- bd$x[4])


## compute the EL interval(Vel.hat)##
      if (Vel.hat <= CritVal)
      CovCount[2]<-CovCount[2]+1

      #x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda x[4]: delta

     lw<-solveNonlinear(g2, c( 0,0,0,0), c(0.3, 0.1, 0.001, -0.999))
     #low.EL<-lw$x[4]      # lower limit of the CI

     upb<-solveNonlinear(g2, c( 0,0,0,0), c(0.1, 0.3, 0.001, 0.999))
     #up.EL<-upb$x[4]        # upper limit of the CI

     #low and up band of El
      #LP[2]<- LP[2]+lw$x[4]
      #UP[2]<- UP[2]+ upb$x[4]

     # The length of EL CI
      CIL[2]<- CIL[2]+(upb$x[4]- lw$x[4])


## compute the BTI interval.
   hwidth<-Z*sqrt(Vstar)
```

```
     # tlow<- delta.pAUC.hat-hwidth           # lower limit of the CI
      #tup<- delta.pAUC.hat+hwidth            # upper limit of the CI

     if (((delta.pAUC.hat-hwidth)<= deltapAUC.true) & ((delta.pAUC.hat+hwidth)
   >= deltapAUC.true)) CovCount[3]<-CovCount[3]+1

      #low and up band
       #LP[3]<-LP[3]+(delta.pAUC.hat-hwidth)
      #UP[3]<-UP[3]+(delta.pAUC.hat+hwidth)
       CIL[3]<- CIL[3]+2*hwidth               # The length of BT and BS CI



## compute the bootstrap(BS) interval
      #bslow<- delta.pAUCbar.B-hwidth         # lower limit of the CI
      #bsup<- delta.pAUCbar.B+hwidth           # upper limit of the CI

       if (((delta.pAUCbar.B-hwidth) <= deltapAUC.true) &
   ((delta.pAUCbar.B+hwidth)>= deltapAUC.true)) CovCount[4]<-CovCount[4]+1
      #low and up band
      #LP[4]<-LP[4]+(delta.pAUCbar.B-hwidth)
      #UP[4]<-UP[4]+(delta.pAUCbar.B+hwidth)


}

 cov<-CovCount/iter; cov
  #bound.L<-LP/iter
  #bound.U<-UP/iter
  wid<-CIL/iter;wid

#Result Output
sink("C:\\Temp\\new5020.txt",append = T)

cat("iter=", iter,"At level=", levelc, "m=", m, "n=",
   n,"rho=",rho,"Delta=",deltapAUC.true, "p0=", p0, "p1=", p1, "\n")
cat("mean1=",y1.mean,"mean2=", y2.mean,"y1std=", y1.sd, "y2std=", y2.sd, "B=",
   B, "\n")

cat("Coverage of the (HBEL, EL, BT, BS) CI's for delta :", cov, "\n")
cat("Average length of (HBEL,EL,BTI&BS):", wid, "\n")

cat("-------------------------------------------------------------------------
   ------------------------------","\n")


sink();
```

APPENDIX D: The Splus code for real data analysis

```
coln<-c("ID", "MTH1", "MTH2","GP")
realdata<-read.table("C:\\TEMP\\Thesis\\exam3.dat",col.names=coln, header=F)
realdata
X1<-realdata$MTH1[realdata$GP==0]; X1
X2<-realdata$MTH2[realdata$GP==0]; X2
Y1<-realdata$MTH1[realdata$GP==1]; Y1
Y2<-realdata$MTH2[realdata$GP==1]; Y2
m=length(X1)
n=length(Y1)

levelc<-0.90;
CritVal<-qchisq(levelc,1)
#Z<-qnorm(levelc);
Z<-qnorm(1-(1-levelc)/2)
p0<-0.05;  p1<-0.5;


###### part 1: Bootstrap  ######
  #### Bootstrap start ####
   B=500;
   sigma=pAUC=matrix(,B, 2)


   for (b in 1:B)
    {
    X1B <- sample(X1, m, replace = T)
    X2B <- sample(X2, m, replace = T)
    Y1B <- sample(Y1, n, replace = T)
    Y2B <- sample(Y2, n, replace = T)

    q0B.1.hat<-quantile(X1B, c(1-p0)) # hatq0, hatq1: sample quantiles of F
    q0B.2.hat<-quantile(X2B, c(1-p0)) # hatq0, hatq1: sample quantiles of F
    q1B.1.hat<-quantile(X1B, c(1-p1))
    q1B.2.hat<-quantile(X2B, c(1-p1))


    VB <- matrix(,m,2)
    for (i in 1:m)
       {
        VB[i,1]<- (1-mean(Y1B <= X1B[i])) *(q1B.1.hat <= X1B[i])*(X1B[i] <=
   q0B.1.hat)
        VB[i,2]<- (1-mean(Y2B <= X2B[i]))*(q1B.2.hat <= X2B[i])*(X2B[i] <=
   q0B.2.hat)
       }

       sigma[b,1]<-mean((VB[,1]-mean(VB[,1]))^2)    ##my.mean(VB[,1]) if using
   function
       sigma[b,2]<-mean((VB[,2]-mean(VB[,2]))^2)


       pAUC[b,1]<-mean(VB[,1])
       pAUC[b,2]<-mean(VB[,2])

   }

  delta.pAUCbar.B<-mean(pAUC[,2]-pAUC[,1])     # Estimate mean difference of
   two pAUCs by bootstrap

   ##Variance of delta.pAUC by bootstrap
```

```
  #Vstar1<-var(pAUC[,1])+var(pAUC[,2]); #Vstar1

#  V12<- sum((pAUC[,1]-mean(pAUC[,1]))*(pAUC[,2]-mean(pAUC[,2])))/(B-1)
 # Vstar2<-var(pAUC[,1])+var(pAUC[,2])-2*V12;   #Vstar2

   Vstar <- var(pAUC[,2]-pAUC[,1]);  Vstar

   C.deltap0p1<-(mean(sigma[,1])+mean(sigma[,2]))/(m*Vstar); C.deltap0p1;
    #bootstrap C.deltap0p1 to caculate HBEL



  ########################Bootstrap end ##################


  ######Part 2:  Caculate delta.pAUC.hat######

 ## Caculate X Quantile of 1-pi (i=0,1) for q.hat
  q0.1.hat<-quantile(X1,1-p0);q0.1.hat
  q0.2.hat<-quantile(X2,1-p0);q0.2.hat
  q1.1.hat<-quantile(X1,1-p1); q1.1.hat
  q1.2.hat<-quantile(X2,1-p1); q1.2.hat

  ## Caculate V(ki).hat & C(deltap0p1).hat
  V1.hat<-V2.hat<-0
  for (i in 1 : m){
     V1.hat[i]<-(1-mean(Y1<= X1[i]))*(q1.1.hat <= X1[i])*(X1[i]<=q0.1.hat)
     V2.hat[i]<-(1-mean(Y2<= X2[i]))*(q1.2.hat <= X2[i])*(X2[i]<=q0.2.hat)
     }

  V1.hat
  V2.hat
  delta.pAUC.hat<-mean(V2.hat)-mean(V1.hat); delta.pAUC.hat

  #V1.hat; #V2.hat
  sigmap0p1.1.hat<-mean((V1.hat-mean(V1.hat))^2)
  sigmap0p1.2.hat<-mean((V2.hat-mean(V2.hat))^2)
  C.deltap0p1.hat<-(sigmap0p1.1.hat+sigmap0p1.2.hat)/(m*Vstar);
   C.deltap0p1.hat



  ######Part 3: Caculate C.I and coverage######

   ## compute the HBEL1 interval(Vel from bootstrap)##

     #x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda x[4]: delta

      g1 <- function(x) c( mean((V1.hat-x[1])/(1-2*x[3]*(V1.hat-x[1]))),
                     mean((V2.hat-x[2])/(1+2*x[3]*(V2.hat-x[2]))),
      mean(V2.hat/(1+2*x[3]*(V2.hat-x[2])))-mean(V1.hat/(1-2*x[3]*(V1.hat-
    x[1])))-x[4],
      C.deltap0p1*(2*(sum( log(abs(1-2*x[3]*(V1.hat-x[1]))))+sum(
    log(abs(1+2*x[3]*(V2.hat-x[2])))))-CritVal)

      bd<-solveNonlinear(g1, c( 0,0,0,0), c(0.3, 0.1, 0.001, -0.9))
      low.HBEL<-bd$x[4]  # lower limit of the CI

      b<-solveNonlinear(g1, c( 0,0,0,0), c(0.1, 0.3, 0.001, 0.999))
      up.HBEL<-b$x[4]    # upper limit of the CI
```

```
      # The length of HBEL CI
      CIL.HBEL=up.HBEL- low.HBEL


    ## compute the EL interval(Vel.hat)##

      #x[1]: p0p1.1  x[2]: p0p1.2  x[3]: lamda x[4]: delta
      g2 <- function(x) c( mean((V1.hat-x[1])/(1-2*x[3]*(V1.hat-x[1]))),
                    mean((V2.hat-x[2])/(1+2*x[3]*(V2.hat-x[2]))),
       mean(V2.hat/(1+2*x[3]*(V2.hat-x[2])))-mean(V1.hat/(1-2*x[3]*(V1.hat-
    x[1])))-x[4],
      C.deltap0p1.hat*(2*(sum( log(abs(1-2*x[3]*(V1.hat-x[1]))))+sum(
    log(abs(1+2*x[3]*(V2.hat-x[2])))))-CritVal)

     lw<-solveNonlinear(g2, c( 0,0,0,0), c(0.3, 0.1, 0.001, -0.999))
     low.EL<-lw$x[4];      # lower limit of the CI

     up<-solveNonlinear(g2, c( 0,0,0,0), c(0.1, 0.3, 0.001, 0.999))
     up.EL<-up$x[4];        # upper limit of the CI

     # The length of EL CI
      CIL.EL=up.EL- low.EL



    ## compute the BT interval.

     hwidth<-Z*sqrt(Vstar); hwidth
      tlow<- delta.pAUC.hat-hwidth            # lower limit of the CI
      tup<- delta.pAUC.hat+hwidth          # upper limit of the CI

       CIL.BT<-2*hwidth             # The length of BT and BS CI


    ## compute the bootstrap(BS) interval.
      bslow<- delta.pAUCbar.B-hwidth       # lower limit of the CI
      bsup<- delta.pAUCbar.B+hwidth            # upper limit of the CI


 up=c(up.HBEL, up.EL, tup, bsup);up
 low=c(low.HBEL, low.EL, tlow, bslow);low
 wid=c(CIL.HBEL, CIL.EL, CIL.BT, CIL.BT);wid


#Result Output;

sink("C:\\temp\\real.txt", append = T)

cat("Real data At level=", levelc, "p0=", p0, "P1=", p1,
 "m=", m, "n=", n, "B=",  B, "\n")
cat("delta.pAUCbar.B=", delta.pAUCbar.B, "delta.pAUC.hat=", delta.pAUC.hat,
   "\n" )
cat("upbound of the (HBEL, EL, BT, BS) CI's for real data are:",
 up , "\n")
cat("lowbound of the (HBEL, EL, BT, BS) CI's for real data are:",
low , "\n")
cat("length of (HBEL,EL,BTI&BS):",
wid, "\n")

cat("----------------------------------------------------------------","\n")


sink();
```