

REU PROJECT

Confidence bands for ROC Curves

Zsuzsanna HORVÁTH*

Department of Mathematics, University of Utah

Faculty Mentor: Davar Khoshnevisan

We develop two methods to construct confidence bands for the receiver operating characteristic (ROC) curve. The first method is based on the smoothed bootstrap while the second method uses the Bonferroni inequality. As an illustration, we provide confidence bands for the ROC curves using data on Duchenne Muscular Dystrophy.

Key words and Phrases: quantile and empirical processes, weak convergence, smoothed bootstrap, Duchenne Muscular Dystrophy, confidence bands

AMS 2000 subject classification: Primary 62G07; Secondary 62G20, 62G09, 62G15

*Supported in part by NSF and NSF-VIGRE grants

1 Introduction

The receiver operating characteristic (ROC) curve is an important tool to summarize the performance of a medical diagnostic test for detecting whether or not a patient has a disease. In a medical test resulting in a continuous measurement T , the disease is diagnosed if $T > t$, for a given a threshold t . Suppose the distribution function of T is F conditional on non-disease and G conditional on disease. The sensitivity of the test is defined as the probability of correctly classifying a diseased individual, $SE(t) = 1 - G(t)$ when threshold t is used. Similarly, the specificity of the test, $SP(t) = F(t)$, is the probability of correctly classifying a healthy patient.

The ROC curve is defined as the graph $(1 - F(t), 1 - G(t))$ for various values of the threshold t , or in other words, sensitivity (true positive fraction) versus 1 - specificity (false positive fraction), power versus size for a test with critical region $\{T > t\}$. This enables one to summarize a test's performance or to compare two diagnostic tests. For more information about ROC curves and their use, we refer to Swets and Pickett (1982), Zhou, Obuchowski and McClish (2002) and references therein.

An alternative definition is

$$(1.1) \quad R(t) = 1 - G\{F^{-1}(1 - t)\} \quad \text{for } 0 \leq t \leq 1,$$

where F^{-1} denotes the generalized inverse function of F defined by

$$F^{-1}(t) = \inf\{x : F(x) \geq t\}.$$

This is closely related to the comparison distribution function $G\{F^{-1}(t)\}$ for the two-sample problem (Parzen, 1998).

Yet another way of understanding the ROC problem is by rewriting the definition of the ROC value $R(t)$ in an equivalent way as $R(t) = P(1 - F(X_1) \leq t)$. Indeed, as already noted by Lloyd (1998), the ROC curve is simply the distribution function of $F_0 = 1 - F(X_1)$. Distribution F_0 can not be estimated directly since we do not observe data from F_0 , but rather from F and G separately.

To estimate this ROC curve, there are several approaches. One way is to model F and G parametrically (see Goddard and Hinberg, 1990). Hsieh and Turnbull (1996) estimate this curve empirically. Let

$$G_n(y) = n^{-1} \sum_{j=1}^n I(Y_j \leq y)$$

denote the empirical distribution function of a sample of size n from the disease population. Similarly we define

$$F_m(x) = m^{-1} \sum_{i=1}^m I(X_i \leq x)$$

and let

$$F_m^{-1}(t) = \inf\{x : F_m(x) \geq t\}$$

be the empirical quantile function of a size m sample from the non-disease population. The empirical or nonparametric estimator for $R(t)$ is defined as

$$\hat{R}(t) = 1 - G_n\{F_m^{-1}(1 - t)\}.$$

For a semi-parametric approach we refer to Li, Tiwari and Wells (1999).

In the literature, the theoretical focus is mainly on obtaining consistency and asymptotic normality of various estimators of $R(p)$, therefore offering the necessary tools to construct pointwise confidence intervals for $R(p)$. This, however, is not sufficient to compare two diagnostic test via their ROC curves. In order to have an idea about the whole curve, one should construct confidence bands for ROC curves. To our knowledge, there is no corresponding literature on confidence bands for ROC curves. In this paper, we will use smoothed bootstrap to construct a confidence band for $\{R(t) : a \leq t \leq b\}$ in the nonparametric case. This means that we define random functions $\hat{R}_\ell(t)$ and $\hat{R}_u(t)$ based on the original data and the smoothed bootstrap data such that

$$\lim_{\min(m,n) \rightarrow \infty} P\{\hat{R}_\ell(t) \leq R(t) \leq \hat{R}_u(t) \text{ for all } t \in [a, b]\} = 1 - \alpha,$$

where $0 < \alpha < 1$ is a given number.

Although the idea of resampling from a smoothed empirical distribution is already mentioned in Efron's (1979) pioneering work, the smoothed bootstrap is not as popular as the usual non-smoothed bootstrap. The obvious reason may be that the smoothed bootstrap is more complicated (as it requires the choice of, at least, a bandwidth) and more extensive in computational terms. Its relative advantages over the ordinary bootstrap have not clearly been stated. There are many problems in which the smoothed bootstrap seems to be a good alternative. For example: the inference on the number of modes (Silverman, 1981), quantiles estimation (Hall and Rieck 2001) and confidence regions of regression curves (Claeskens and van Keilegom, 2003).

The structure of this paper is as follows: Section 2 contains the methodologies and main theorems; our results are then applied to a data set on elevated creatine levels in carriers of Duchenne Muscular Dystrophy compared to non-carriers.

2 Methodology and Main Theorems

Suppose that a data set X_1, X_2, \dots, X_m of measurements from the healthy population with distribution $F(x)$ is available as is a set Y_1, Y_2, \dots, Y_n from the diseased population

with distribution $G(y)$. We denote the empirical distributions of the X and Y data by

$$F_m(x) = m^{-1} \sum_{i=1}^m I(X_i \leq x)$$

and

$$G_n(y) = n^{-1} \sum_{j=1}^n I(Y_j \leq y),$$

respectively. Also we define the empirical quantile function as

$$F_m^{-1}(t) = \inf\{x : F_m(x) \geq t\}.$$

Then the empirical ROC curve is

$$(2.1) \quad \hat{R}(t) = 1 - G_n\{F_m^{-1}(1-t)\}.$$

which for convenience we will also write as $\hat{R}(t) = 1 - G_n F_m^{-1}(1-t)$. Under the conditions that $F(x)$ and $G(y)$ have continuous densities, $f(x)$ and $g(y)$, $g(F^{-1}(t))/f(F^{-1}(t))$ is bounded on any subinterval (a, b) of $(0, 1)$, and $n/m \rightarrow \lambda$ as $\min(n, m) \rightarrow \infty$, Hsieh and Turnbull (1996) proved the existence of a probability space on which one can define two independent sequences of Brownian bridges $\{B_1^{(m)}(t), 0 \leq t \leq 1\}$ and $\{B_2^{(n)}(t), 0 \leq t \leq 1\}$ such that

$$(2.2) \quad \begin{aligned} & \sqrt{n}(\hat{R}(t) - R(t)) \\ &= \sqrt{n} (1 - G_n\{F_m^{-1}(1-t)\} - (1 - G\{F^{-1}(1-t)\})) \end{aligned}$$

$$(2.3) \quad = \lambda^{1/2} \frac{g(F^{-1}(1-t))}{f(F^{-1}(1-t))} B_1^{(m)}(1-t) + B_2^{(n)}(G\{F^{-1}(1-t)\}) + o(1) \quad \text{a.s.}$$

i.e.

$$\begin{aligned} & \sup_{a \leq t \leq b} \left| \sqrt{n}(\hat{R}(t) - R(t)) \right. \\ & \left. - \lambda^{1/2} \frac{g(F^{-1}(1-t))}{f(F^{-1}(1-t))} B_1^{(m)}(1-t) + B_2^{(n)}(G\{F^{-1}(1-t)\}) \right| \rightarrow 0 \quad \text{a.s.} \end{aligned}$$

uniformly on $[a, b]$. A process $B(t), 0 \leq t \leq 1$ is called Brownian bridge if $EB(t) = 0$ for all t and for any $0 \leq t_1, t_2, \dots, t_k \leq 1$ the k -dimensional vector $(B(t_1), B(t_2), \dots, B(t_k))$ is k -variate normal with 0 mean and covariance $EB(t_i)B(t_j) = \min(t_i, t_j) - t_i t_j$. Since the distribution of B_1 and B_2 do not depend on n and m , this implies immediately that for all points of continuity of H we have

$$\lim_{\min(m, n) \rightarrow \infty} P \left\{ \sup_{a \leq t \leq b} |\sqrt{n}(\hat{R}(t) - R(t))| \leq x \right\} = H(x),$$

where

$$H(x) = P \left\{ \sup_{a \leq t \leq b} \left| \lambda^{1/2} \frac{g(F^{-1}(1-t))}{f(F^{-1}(1-t))} B_1^{(m)}(1-t) + B_2^{(n)}(G\{F^{-1}(1-t)\}) \right| \leq x \right\}.$$

Since there are unknown parameters, such as f and g , in (2.3), we can not directly rely on (2.3) to get confidence bands for $R(t)$.

In this paper, we use smoothed bootstrap to bypass those unknown parameters.

To proceed further, let us introduce some notation. Suppose $K_1(x)$ and $K_2(y)$ are distribution functions with densities $k_1(x)$ and $k_2(y)$, respectively. Introduce

$$\begin{aligned} \hat{F}_m(x) &= \frac{1}{m} \sum_{i=1}^m K_1 \left(\frac{x - X_i}{h_1} \right), \quad \hat{G}_n(y) = \frac{1}{n} \sum_{j=1}^n K_2 \left(\frac{y - Y_j}{h_2} \right), \\ \hat{f}_m(x) &= \frac{1}{mh_1} \sum_{i=1}^m k_1 \left(\frac{x - X_i}{h_1} \right), \quad \hat{g}_n(y) = \frac{1}{nh_2} \sum_{j=1}^n k_2 \left(\frac{y - Y_j}{h_2} \right), \end{aligned}$$

the smoothed empirical distribution functions and densities for the X data and the Y data. Let $\{X_1^*, X_2^*, \dots, X_m^*\}$ and $\{Y_1^*, Y_2^*, \dots, Y_n^*\}$ be two independent random samples from $\hat{F}_m(x)$ and $\hat{G}_n(y)$ respectively. The empirical distributions of the X^* and Y^* data are denoted by $F_m^*(x) = m^{-1} \sum_{i=1}^m I(X_i^* \leq x)$ and $G_n^*(y) = n^{-1} \sum_{j=1}^n I(Y_j^* \leq y)$ respectively. For the sake of notational simplicity, the bootstrap sample sizes and the original sample sizes are the same. We would like to point out that all of our results remain true when the bootstrap sample sizes differ from m and n as long as they also go to infinity. The conditional probability given $\{X_1, X_2, \dots, \}$ and $\{Y_1, Y_2, \dots, \}$ will be denoted by P^* .

In the following we list the regularity conditions needed to get the asymptotic distribution of the bootstrapped ROC process. As usual, we need that both k_1 and k_2 are smooth:

$$(2.4) \quad \begin{aligned} &k_1 \text{ and } k_2 \text{ are bounded and have bounded variation on } R, \text{ and the sets} \\ &\text{of discontinuities have Lebesgue measure 0.} \end{aligned}$$

Condition (2.4) means that that the functions k_1 and k_2 are continuously differentiable and have bounded derivatives except a very small set, like countably many points. We also need that $h_1(n), h_2(m)$ both go to zero but not too fast:

$$(2.5) \quad h_1(n) \rightarrow 0 \text{ and } h_2(m) \rightarrow 0;$$

and

$$(2.6) \quad nh_1(n)/\log n \rightarrow \infty \text{ and } mh_2(m)/\log m \rightarrow \infty.$$

Since we are working with smoothed estimators, it is natural to assume that both F and G are smooth functions. Let $t_F = \inf\{x : F(x) > 0\}$ and $t^F = \sup\{x : F(x) < 1\}$ and

(t_F, t^F) be the (open) support of F . We define (t_G, t^G) , the (open) support of G similarly. We assume that

$$(2.7) \quad f(x) > 0 \text{ if } x \in (t_F, t^F),$$

$$(2.8) \quad g(y) > 0 \text{ if } y \in (t_G, t^G),$$

$$(2.9) \quad f \text{ and } g \text{ are uniformly continuous on } R.$$

THEOREM 2.1 *We assume that (2.4)–(2.9) hold, $0 < a < b < 1$. Then we can define two independent sequences of Brownian bridges $\{\hat{B}_1^{(m)}(t), 0 \leq t \leq 1\}$ and $\{\hat{B}_2^{(n)}(t), 0 \leq t \leq 1\}$, independent of $\{X_i, Y_i, 1 \leq i < \infty\}$ such that*

$$(2.10) \quad \lim_{\min(m,n) \rightarrow \infty} P^* \left(\sup_{a \leq t \leq b} \sqrt{n} \left| 1 - G_n^* \{F_m^{*-1}(1-t)\} - (1 - \hat{G}_n \{\hat{F}_m^{-1}(1-t)\}) \right. \right. \\ \left. \left. - \left(\lambda^{1/2} \frac{g(F^{-1}(1-t))}{f(F^{-1}(1-t))} \hat{B}_1^{(m)}(1-t) + \hat{B}_2^{(n)}(G\{F^{-1}(1-t)\}) \right) \right| \geq \delta \right) = 0 \text{ a.s.}$$

for any $\delta > 0$.

Combining (2.3) and (2.10), we get immediately the following result which provides the justification for the applicability of the smoothed bootstrap.

COROLLARY 2.1 *Under the conditions of Theorem 2.1, we have*

$$(2.11) \quad \sup_x \left| P \left(\sup_{a \leq t \leq b} \sqrt{n} \left| 1 - G_n \{F_m^{-1}(1-t)\} - (1 - G\{F^{-1}(1-t)\}) \right| \leq x \right) \right. \\ \left. - P^* \left(\sup_{a \leq t \leq b} \sqrt{n} \left| 1 - G_n^* \{F_m^{*-1}(1-t)\} - (1 - \hat{G}_n \{\hat{F}_m^{-1}(1-t)\}) \right| \leq x \right) \right| \rightarrow 0 \text{ a.s.}$$

as $\min(m, n) \rightarrow \infty$.

In Theorem 2.1 and therefore Corollary 2.1, we can not replace a with 0 and/or with 1. However, a very simple Bonferroni type confidence band can be derived on the interval $[\epsilon_m, 1 - \epsilon_m]$, where $\epsilon_m \rightarrow 0$. For any $\alpha_i \in (0, 1)$ we define c_i as the solution of the equation

$$(2.12) \quad P \left(\sup_{0 \leq t \leq 1} |B(t)| \leq c_i \right) = 1 - \alpha_i, \quad i = 1, 2,$$

where $\{B(t), 0 \leq t \leq 1\}$ denoted a Brownian bridge.

THEOREM 2.2 *If F is continuous, $\epsilon_m \rightarrow 0$ and $m^{1/2}\epsilon_m \rightarrow \infty$ as $m \rightarrow \infty$, then*

$$\liminf_{\min(m,n) \rightarrow \infty} P\left(\hat{R}(t - c_1 m^{-1/2}) - c_2 n^{-1/2} \leq R(t) \leq \hat{R}(t + c_1 m^{-1/2}) + c_2 n^{-1/2}\right. \\ \left. \text{for all } \epsilon_m \leq t \leq 1 - \epsilon_m\right) \geq (1 - \alpha_1)(1 - \alpha_2),$$

where $c_1 = c_1(\alpha_1)$ and $c_2 = c_2(\alpha_2)$ are defined by (2.12).

We conclude this section with two remarks.

REMARK 2.1 *In Theorem 2.1 and Corollary 2.1 we assumed that the bootstrap sample sizes and the original sample sizes are the same. This restriction can be removed, the only assumption needed is the proportionality of the original sample and the bootstrap sample sizes.*

REMARK 2.2 *It follows from the proof that Theorem 2.2 remains true if the X and Y samples are not independent. As long as the X 's are independent, identically distributed and the y 's are independent, identically distributed we have that*

$$\liminf_{\min(m,n) \rightarrow \infty} P\left(\hat{R}(t - c_1 m^{-1/2}) - c_2 n^{-1/2} \leq R(t) \leq \hat{R}(t + c_1 m^{-1/2}) + c_2 n^{-1/2}\right. \\ \left. \text{for all } \epsilon_m \leq t \leq 1 - \epsilon_m\right) \geq 1 - (\alpha_1 + \alpha_2),$$

where $c_1 = c_1(\alpha_1)$ and $c_2 = c_2(\alpha_2)$ are defined by (2.12). Hence we can construct ROC bands for paired observations.

3 An Application

Duchenne Muscular Dystrophy is a genetically transmitted disease, passed from a mother to her children. Carriers of Duchenne Muscular Dystrophy usually have no physical symptoms, but they tend to exhibit elevated levels of certain serum enzymes or protein including, for example, creatine kinase, hemopexin, lactate dehydrogenase, and pyruvate kinase enzymes. Data set 38 in Andrews and Herzberg (1985) contains the measurement of creatine kinase in $m = 134$ non-carriers (X sample) and $n = 75$ carriers. First we used the smoothed bootstrap method to find confidence bands for the ROC curve with 90% coverage on the interval $[.01, .99]$. The bootstrap procedure was repeated 5,000 times. Figure 3.1 contains the confidence band. The shape of the confidence band is very similar to the ROC curve of exponential variables. Hence we simulated $m = 134$ exponential random variables with mean 39 (X sample) and $n = 75$ exponential random variables with mean 187 (Y sample) and constructed 90% confidence band on $[.01, .99]$ using the bootstrap method. The result is on Figure 3.2. Comparing Figures 3.1 and

3.2 we can see the resemblance between the two confidence bands. We checked the accuracy of the bootstrap method for small samples using exponential observations and found that the asymptotic result in Corollary 2.1 can be used in case of small sample sizes like 100. We used the same procedure for the other enzymes. Figure 3.5 contains the bootstrapped confidence band for the hemopexin data. The Bonferroni confidence interval for the hemopexin data can be found in Figure 3.6. Hemopexin levels were measured for $m=134$ non-carriers (X sample) and $n=75$ carriers (Y-sample). The bootstrap procedure was repeated 5,000 times in this case. To obtain a 90% confidence band, we simulate $m=134$ exponential random variables with mean 83 (X-sample) and $n=75$ exponential random variables with mean 87. Data on pyruvate kinase enzyme levels was available on $m=134$ non-carriers and $n=67$ carriers. The simulation for the X-sample and Y-sample was done using exponential random variables with means 12 and 24, respectively. Figures 3.7 and 3.8 contain the corresponding confidence intervals. Similarly, for the lactate dehydrogenase $m=127$, $n=75$. The simulated exponential random variables for the X-sample and Y-sample are generated with means 165 and 256, respectively. Figure 3.7 and Figure 3.8 contain the confidence bands for the lactate dehydrogenase enzyme.

Simulations showed that the Bonferroni inequality in Theorem 2.2 is very conservative and the probability of coverage is much better than the lower bound $(1 - \alpha_1)(1 - \alpha_2)$ if c_1 and c_2 are from equations (2.12). This was expected since we replace two discrete processes with continuous ones, so the Gaussian approximation provides a crude bound for sample sizes less than 200. We used simulations based on exponential data to find c_1 and c_2 providing 90% confidence interval when $m = 134$ and $n = 75$. We tried several exponential samples with different rates and the choice of $c_1 = c_2 = .75$ was a good choice. The confidence interval for the ROC curve in case of the Duchenne Muscular Dystrophy is in Figure 3.3. In Figure 3.4 we also computed a confidence band for simulated exponential data with means 39 and 187. The four confidence intervals are nearly the same suggesting that the ROC curve is t^2 for this data set.

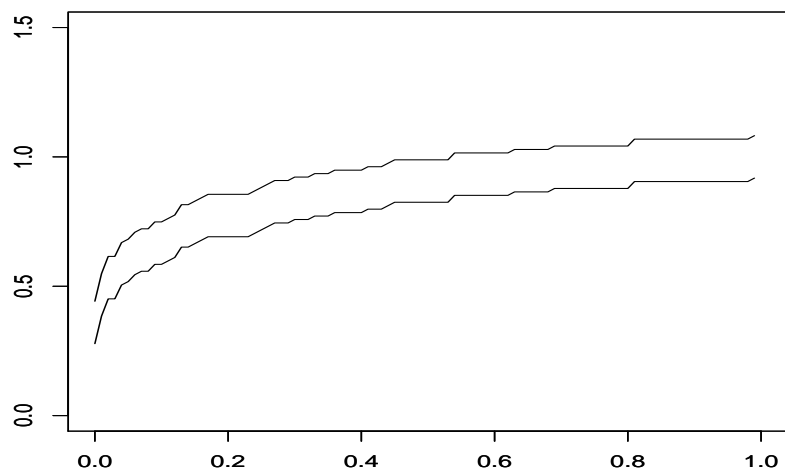


FIGURE 3.1 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing creatine kinase levels using smoothed bootstrap

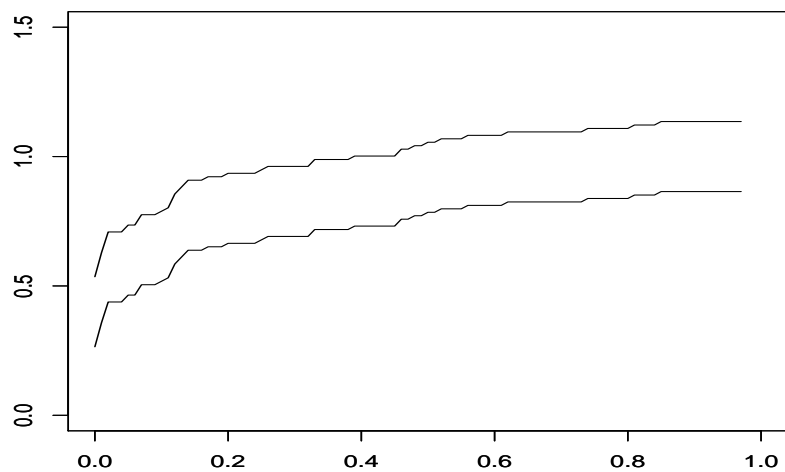


FIGURE 3.2 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on exponential data with $EX = 39$ and $EY = 187$ using smoothed bootstrap

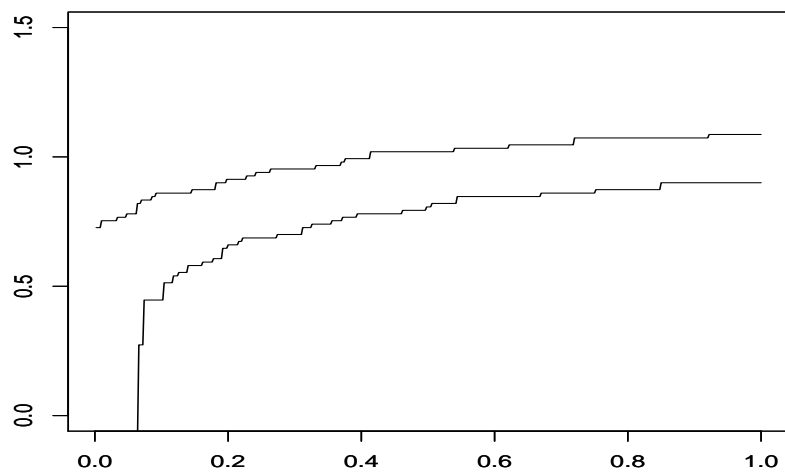


FIGURE 3.3 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing creatine kinase levels using the Bonferroni inequality

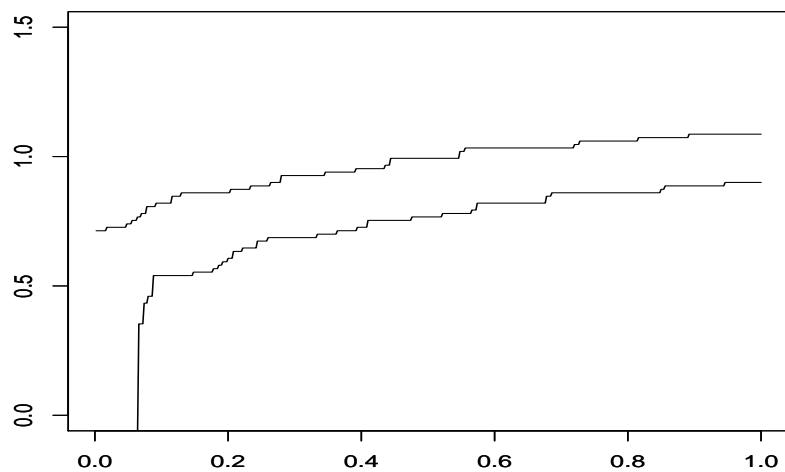


FIGURE 3.4 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on exponential data with $EX = 39$ and $EY = 187$ using the Bonferroni inequality

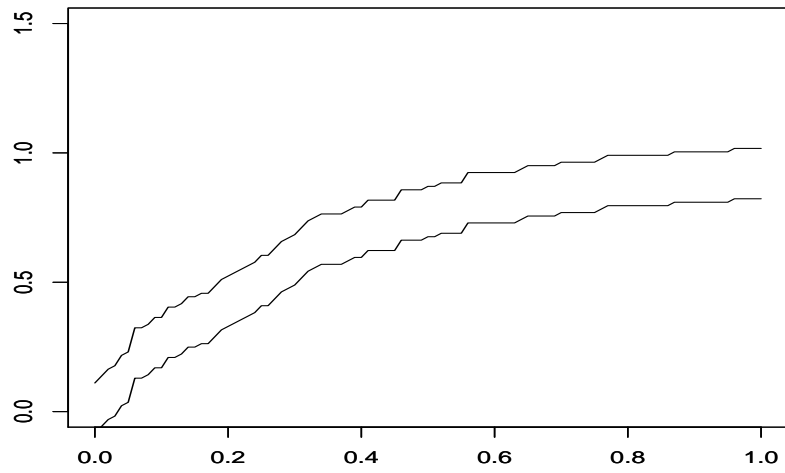


FIGURE 3.5 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing hemopexin levels using smoothed bootstrap

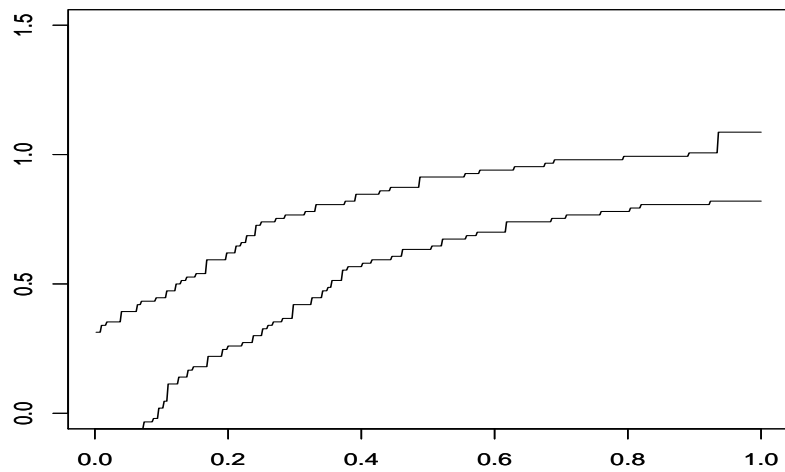


FIGURE 3.6 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing hemopexin levels using the Bonferroni inequality

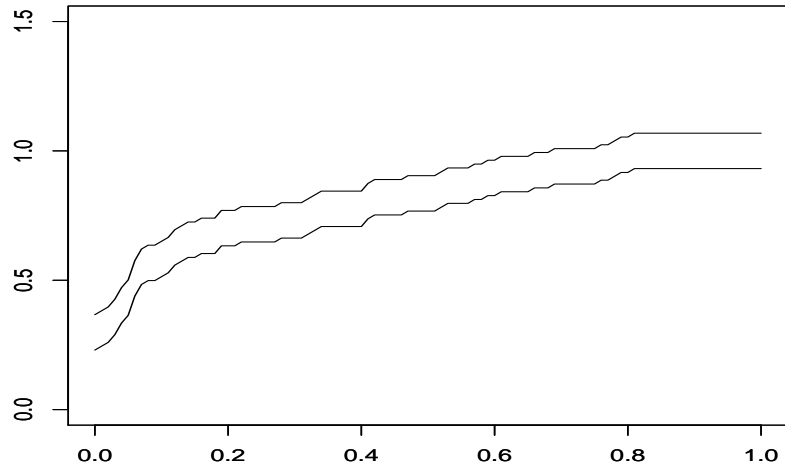


FIGURE 3.7 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing pyruvate kinase enzyme levels using smoothed bootstrap

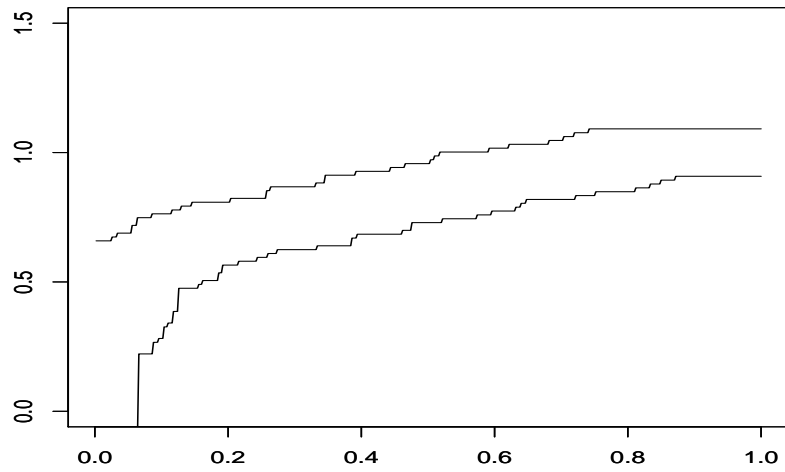


FIGURE 3.8 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing pyruvate kinase enzyme levels using the Bonferroni inequality

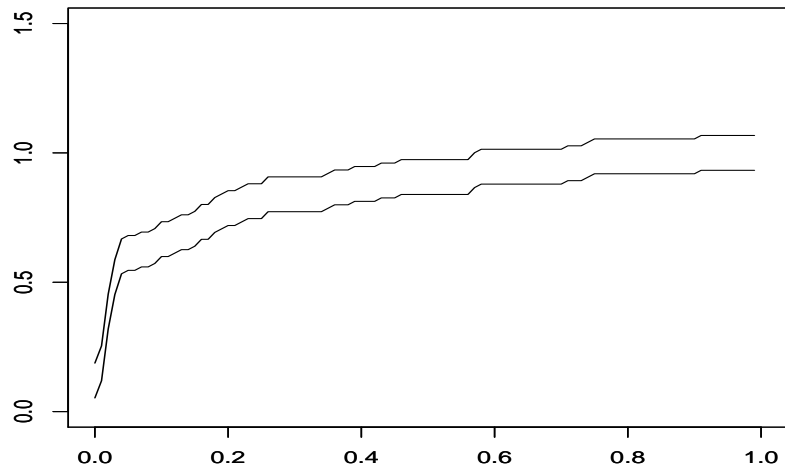


FIGURE 3.9 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing lactate dehydroginase levels using smoothed bootstrap

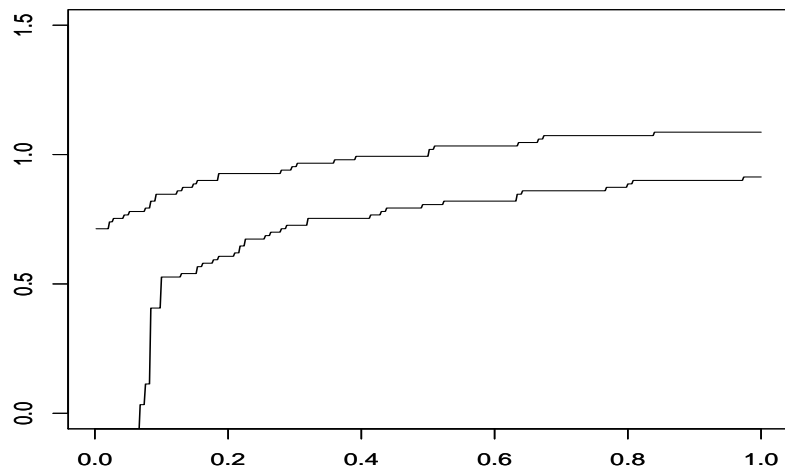


FIGURE 3.10 90% confidence interval for $R(t)$, $0.01 \leq t \leq .99$ based on the Muscular Dystrophy data comparing lactate dehydroginase levels using the Bonferroni inequality

4 Appendix

The first result is the Glivenko–Cantelli (Shorack (2000, p. 223)) lemma for the smoothed empirical distribution function.

LEMMA 4.1 *Suppose $F(x)$ is continuous on R . If (2.5) holds, then*

$$(4.1) \quad \sup_x |\hat{F}_m(x) - F(x)| \rightarrow 0 \quad \text{a.s.}$$

as $m \rightarrow \infty$.

Proof. Since the distribution function F is continuous on R , it is uniformly continuous on R . Let $F_m(x) = m^{-1} \sum_{i=1}^m I(X_i \leq x)$. Then

$$\begin{aligned} \sup_x |\hat{F}_m(x) - F(x)| &\leq \sup_x \left| \int K_1 \left(\frac{x-t}{h_1} \right) d(F_m(t) - F(t)) \right| \\ &\quad + \sup_x \left| \int K_1 \left(\frac{x-t}{h_1} \right) dF(t) - F(x) \right| \\ &= C_{1,m} + C_{2,m}. \end{aligned}$$

Let us estimate $C_{2,m}$ first. Change of variables gives

$$\begin{aligned} (4.2) \quad C_{2,m} &= \sup_x \left| \frac{1}{h_1} \int F(t) k_1 \left(\frac{x-t}{h_1} \right) dt - F(x) \right| \\ &= \sup_x \left| \int (F(x - uh_1) - F(x)) k_1(u) du \right| \\ &\leq \sup_x \int_{-\infty}^{-A} |F(x - uh_1) - F(x)| k_1(u) du \\ &\quad + \sup_x \int_A^{\infty} |F(x - uh_1) - F(x)| k_1(u) du \\ &\quad + \sup_x \int_{-A}^{-A} |F(x - uh_1) - F(x)| k_1(u) du \\ &\leq K_1(-A) + 1 - K_1(A) + \sup_x |F(x + Ah_1) - F(x - Ah_1)| \end{aligned}$$

If A is large enough, the first two terms in (4.2) will be small. The uniform continuity of $F(x)$ implies, for any A , $\sup_x |F(x + Ah_1) - F(x - Ah_1)| \rightarrow 0$ as $h_1 \rightarrow 0$. Hence $C_{2,m}$ can be arbitrarily small as $m \rightarrow \infty$.

Now, let us analyze $C_{1,m}$. As $m \rightarrow \infty$, we have

$$\begin{aligned} C_{1,m} &= \sup_x \left| \frac{1}{h_1} \int (F_m(t) - F(t)) k_1 \left(\frac{x-t}{h_1} \right) dt \right| \\ &= \sup_x \left| \int (F_m(x - uh_1) - F(x - uh_1)) k_1(u) du \right| \\ &\leq \sup_x |F_m(x) - F(x)| \rightarrow 0 \quad \text{a.s.} \end{aligned}$$

by Glivenko–Cantelli theorem (cf. Shorack (2000, p.223)), completing the proof of Lemma 4.1.

The next result is again a version of the Glivenko–Cantelli lemma but now for the kernel estimate of a density.

LEMMA 4.2 *If (2.4)–(2.6) and (2.9) hold, then*

$$\sup_x |\hat{f}_m(x) - f(x)| \rightarrow 0 \quad a.s.$$

as $m \rightarrow \infty$.

Proof. Lemma 4.2 was obtained by Bertrand–Retali (1978) (cf. also Silverman (1986, p72)).

The third lemma is an analogue of lemma 4.1 for the smoothed empirical quantile function. Define for $0 < t < 1$,

$$\hat{F}_m^{-1}(t) = \inf\{x : \hat{F}_m(x) \geq t\}, \quad \hat{G}_n^{-1}(t) = \inf\{y : \hat{G}_n(y) \geq t\}$$

LEMMA 4.3 *If (2.4)–(2.9) hold, then for any $\epsilon > 0$, we have*

$$\sup_{\epsilon \leq t \leq 1-\epsilon} |\hat{F}_m^{-1}(t) - F^{-1}(t)| \rightarrow 0 \quad a.s.$$

as $m \rightarrow \infty$.

Proof. By Lemma 4.2 and (2.7) we can assume that

$$\inf_{A \leq x \leq B} \hat{f}_m(x) > 0 \quad \text{if } m \geq m_0(\omega)$$

with some r.v. $m_0(\omega)$, where $[A, B] \subset (t_F, t^F)$. Hence $\hat{F}_m(x)$ is strictly increasing on $[A, B]$, and therefore

$$\hat{F}_m^{-1}(\hat{F}_m(t)) = t \quad \text{if } A \leq t \leq B$$

and

$$\hat{F}_m(\hat{F}_m^{-1}(t)) = t \quad \text{if } A = \hat{F}_m^{-1}(a) \leq t \leq \hat{F}_m^{-1}(b) = B.$$

By Lemma 4.1 we can assume that $\epsilon > a$ and $1 - \epsilon < b$, if $m \geq m_1(\omega)$ for some r.v. $m_1(\omega)$. Thus we have $\hat{F}_m(\hat{F}_m^{-1}(t)) = t$, if $\epsilon \leq t \leq 1 - \epsilon$, and therefore

$$0 = \hat{F}_m(\hat{F}_m^{-1}(t)) - \hat{F}_m(F^{-1}(t)) + \hat{F}_m(F^{-1}(t)) - F(F^{-1}(t)).$$

Since $\sup_{0 < t < 1} |\hat{F}_m(F^{-1}(t)) - F(F^{-1}(t))| \rightarrow 0$ a.s. by Lemma 4.1, we also must have

$$\sup_{\epsilon \leq t \leq 1-\epsilon} |\hat{F}_m(\hat{F}_m^{-1}(t)) - \hat{F}_m(F^{-1}(t))| \rightarrow 0 \quad a.s..$$

By the mean value theorem, there is a r.v. $\xi = \xi(t)$ such that

$$\sup_{\epsilon \leq t \leq 1-\epsilon} \hat{f}_m(\xi(t)) |\hat{F}_m^{-1}(t) - F^{-1}(t)| \geq \inf_{\epsilon \leq t \leq 1-\epsilon} \hat{f}_m(\xi(t)) \sup_{\epsilon \leq t \leq 1-\epsilon} |\hat{F}_m^{-1}(t) - F^{-1}(t)|.$$

We showed that if $m \geq m_1(\omega)$,

$$\inf_{\epsilon \leq t \leq 1-\epsilon} \hat{f}_m(\xi(t)) \geq \inf_{A \leq x \leq B} \hat{f}_m(x).$$

Using Lemma 4.2, we get

$$\inf_{A \leq x \leq B} \hat{f}_m(x) \rightarrow \inf_{A \leq x \leq B} f(x) > 0$$

Hence we must have

$$\sup_{\epsilon \leq t \leq 1-\epsilon} |\hat{F}_m^{-1}(t) - F^{-1}(t)| \rightarrow 0 \text{ a.s.}$$

Let U_1, U_2, \dots, U_m be uniform r.v., independent of X_1, X_2, \dots, X_m . Let

$$\alpha_m(x) = \sqrt{m}(F_m^*(x) - \hat{F}_m(x)),$$

where $F_m^*(x) = \frac{1}{m} \sum_{j=1}^m I(X_j^* \leq x)$, and $X_i^* = \hat{F}_m^{-1}(U_i)$, $1 \leq i \leq m$. Clearly $X_1^*, X_2^*, \dots, X_m^*$ is a random sample from $\hat{F}_m(x)$, and $F_m^*(x)$ is an empirical distribution of $X^{*'}s$.

LEMMA 4.4 *Under the conditions of Lemma 4.1, we can define a sequence of Brownian bridges $B_m(t)$ independent of $\{X_i, 1 \leq i < \infty\}$, such that*

$$\lim_{m \rightarrow \infty} P^* \left(\sup_x |\alpha_m(x) - B_m(F(x))| \geq \delta \right) = 0 \text{ a.s.}$$

for all $\delta > 0$.

Proof. Let $u_m(t) = \sqrt{m}(U_m(t) - t)$ be the uniform empirical process, where $U_m(t) = \frac{1}{m} \sum_{i=1}^m I(U_i \leq t)$. We can define Brownian bridges $B_m(t), 0 \leq t \leq 1$, which are independent of $\{X_i\}$, such that

$$\sup_{0 \leq t \leq 1} |u_m(t) - B_m(t)| \rightarrow 0 \text{ a.s.}$$

as $m \rightarrow \infty$. (cf. Csörgő and Horváth (1993)). Hence

$$\lim_{m \rightarrow \infty} P^* \left(\sup_x |\alpha_m(x) - B_m(\hat{F}_m(x))| \geq \delta \right) = 0 \text{ a.s.,}$$

so it is enough to prove that

$$\lim_{m \rightarrow \infty} P^* \left(\sup_x |B_m(\hat{F}_m(x)) - B_m(F(x))| \geq \delta \right) = 0 \text{ a.s..}$$

Since the distribution of $B_m(t)$ does not depend on m , it is enough to prove

$$\lim_{n \rightarrow \infty} P^* \left(\sup_x \left| B(\hat{F}_m(x)) - B(F(x)) \right| \geq \delta \right) = 0 \quad \text{a.s.},$$

where $B(t)$ is a Brownian bridge, independent of $\{X_i\}$. By Lemma 4.1, $\sup_x |\hat{F}_m(x) - F(x)| \rightarrow 0$ a.s., so the result follows from the uniform continuity of $B(t)$, (cf. Csörgő and Révész, 1981, p.42).

LEMMA 4.5 *Under the conditions of Lemma 4.3, for any $\epsilon > 0$ and $\delta > 0$, we have*

$$\lim_{m \rightarrow \infty} P^* \left(\sup_{\epsilon \leq t \leq 1-\epsilon} \left| \sqrt{m}(\hat{F}_m^{-1}(t) - F_m^{*-1}(t)) - \frac{1}{f(F^{-1}(t))} B_m(t) \right| > \delta \right) = 0 \quad \text{a.s.}$$

where $B_m(t)$ are the Brownian bridges of Lemma 4.4.

Proof. Since $F_m^{*-1}(t)$ is the generalized inverse of $F_m^*(x) = U_m(\hat{F}_m(x))$, we get

$$F_m^{*-1}(t) = \hat{F}_m^{-1}(U_m^{-1}(t)).$$

It is known (e.g. Csörgő and Horváth, 1993, Section 3.3) that

$$\sup_{0 < t < 1} \left| \sqrt{m} (t - U_m^{-1}(t)) - B_m(t) \right| \rightarrow 0 \quad \text{a.s.}$$

as $m \rightarrow \infty$. Now by the mean value theorem we have

$$\begin{aligned} \sqrt{m} \left(\hat{F}_m^{-1}(t) - F_m^{*-1}(t) \right) &= \sqrt{m} \left(\hat{F}_m^{-1}(t) - \hat{F}_m^{-1}(U_m^{-1}(t)) \right) \\ &= \sqrt{m} (t - U_m^{-1}(t)) / \hat{f}_m(\hat{F}_m^{-1}(\xi)), \end{aligned}$$

where $|t - \xi| \leq |t - U_m^{-1}(t)|$. Hence by Lemmas 4.2 and 4.3, we get

$$\lim_{m \rightarrow \infty} P^* \left(\sup_{\epsilon \leq t \leq 1-\epsilon} \left| \frac{1}{\hat{f}_m(\hat{F}_m^{-1}(\xi))} - \frac{1}{f(F^{-1}(t))} \right| > \delta \right) = 0 \quad \text{a.s.}$$

for any $\delta > 0$, so the lemma is proved.

Proof of Theorem 2.1. Since the two bootstrap samples are independent, we can define two independent sequences of Brownian bridges $\{\bar{B}_1^{(m)}(t), 0 \leq t \leq 1\}$ and $\{\bar{B}_2^{(n)}(t), 0 \leq t \leq 1\}$ such that

$$(4.3) \quad \lim_{m \rightarrow \infty} P^* \left(\sup_{a \leq t \leq b} \left| \sqrt{m}(\hat{F}_m^{-1}(t) - F_m^{*-1}(t)) - \frac{1}{f(F^{-1}(t))} \bar{B}_1^{(m)}(t) \right| > \delta \right) = 0 \quad \text{a.s.}$$

(cf. Lemma 4.5) and

$$(4.4) \quad \lim_{n \rightarrow \infty} P^* \left(\sup_x \left| \sqrt{n} \left(\hat{G}_n(t) - G_n^*(t) \right) - \bar{B}_2^{(n)}(t) \right| > \delta \right) = 0 \text{ a.s.}$$

(cf. Lemma 4.4). Next, we write

$$\begin{aligned} & \sqrt{n} \left(G_n^* \left(F_m^{*-1}(1-t) \right) - \hat{G}_n \left(\hat{F}_m^{-1}(1-t) \right) \right) \\ &= \sqrt{n} \left(G_n^* \left(F_m^{*-1}(1-t) \right) - \hat{G}_n \left(F_m^{*-1}(1-t) \right) \right) \\ & \quad + \sqrt{\frac{n}{m}} \sqrt{m} \left(\hat{G}_n \left(F_m^{*-1}(1-t) \right) - \hat{G}_n \left(\hat{F}_m^{-1}(1-t) \right) \right). \end{aligned}$$

Applying (4.4) we get

$$\begin{aligned} \lim_{m,n \rightarrow \infty} P^* \left(\sup_{0 < t < 1} \left| \sqrt{n} \left(G_n^* \left(F_m^{*-1}(1-t) \right) - \hat{G}_n \left(F_m^{*-1}(1-t) \right) \right) \right. \right. \\ \left. \left. - \bar{B}_2^{(n)} \left(F_m^{*-1}(1-t) \right) \right| > \delta \right) = 0 \quad \text{a.s.} \end{aligned}$$

Putting together Lemma 4.3 and (4.3) we get

$$(4.5) \quad \lim_{m \rightarrow \infty} P^* \left(\sup_{\epsilon \leq t \leq 1-\epsilon} \left| F_m^{*-1}(t) - F^{-1}(t) \right| \geq \delta \right) = 0 \quad \text{a.s.}$$

for any $\epsilon > 0$ and $\delta > 0$.

Using the uniform continuity of the Brownian bridge (cf. Csörgő and Révész (1981)), we get for any n

$$\lim_{m \rightarrow \infty} P^* \left(\sup_{\epsilon \leq t \leq 1-\epsilon} \left| \bar{B}_2^{(n)} \left(F_m^{*-1}(1-t) \right) - \bar{B}_2^{(n)} \left(F^{-1}(1-t) \right) \right| \geq \delta \right) = 0 \quad \text{a.s.}$$

for any $\epsilon > 0$ and $\delta > 0$. (We note that the distribution of $\bar{B}_2^{(n)}$ does not depend on n .)

Thus we have for any $\delta > 0$,

$$(4.6) \quad \lim_{m,n \rightarrow \infty} P^* \left(\sup_{0 < t < 1} \left| \sqrt{n} \left(G_n^* \left(F_m^{*-1}(1-t) \right) - \hat{G}_n \left(F_m^{*-1}(1-t) \right) \right) \right. \right. \\ \left. \left. - \bar{B}_2^{(n)} \left(F^{-1}(1-t) \right) \right| > \delta \right) = 0 \quad \text{a.s.}$$

The mean value theorem gives

$$\sqrt{m} \left(\hat{G}_n \left(F_m^{*-1}(1-t) \right) - \hat{G}_n \left(\hat{F}_m^{-1}(1-t) \right) \right) = \hat{g}_n(\xi) \sqrt{m} \left(F_m^{*-1}(1-t) - \hat{F}_m^{-1}(1-t) \right),$$

where $\xi = \xi_m(t)$ is between $F_m^{*-1}(1-t)$ and $\hat{F}_m^{-1}(1-t)$. So by Lemma 4.3 and (4.5) we have for any $\delta > 0$,

$$\lim_{m \rightarrow \infty} P^* \left(\sup_{a \leq t \leq b} \left| \xi_m(t) - F^{-1}(1-t) \right| \geq \delta \right) = 0 \quad \text{a.s.,}$$

and therefore Lemma 4.2 yields

$$\lim_{m,n \rightarrow \infty} P^* \left(\sup_{a \leq t \leq b} \left| \hat{g}_n(\xi_m(t)) - g(F^{-1}(1-t)) \right| \geq \delta \right) = 0 \quad \text{a.s..}$$

Using (4.3) we conclude

$$(4.7) \quad \lim_{m,n \rightarrow \infty} P^* \left(\sup_{a \leq t \leq b} \left| \sqrt{m} \left(\hat{G}_n(F_m^{*-1}(1-t)) - \hat{G}_n(\hat{F}_m^{-1}(1-t)) \right) - \frac{g(F^{-1}(1-t))}{f(F^{-1}(1-t))} \bar{B}_1^{(m)}(1-t) \right| > \delta \right) = 0 \quad \text{a.s.}$$

Now Theorem 2.1 follows from (4.6) and (4.7) with $\hat{B}_1^{(m)}(t) = -\bar{B}_1^{(m)}(t)$ and $\hat{B}_2^{(n)}(t) = -\bar{B}_2^{(n)}(t)$.

Proof of Theorem 2.2. The weak convergence of the empirical process gives

$$(4.8) \quad \begin{aligned} P \left(G_n(x) - c_2 n^{-1/2} \leq G(x) \leq G_n(x) + c_2 n^{-1/2} \text{ for all } x \right) \\ = P \left(\sup_{-\infty < x < \infty} |B(G(x))| \leq c_2 \right) \\ \geq P \left(\sup_{0 \leq t \leq 1} |B(t)| \leq c_2 \right). \end{aligned}$$

Csörgő and Révész (1984) (cf. also Csörgő and Horváth (1985)) showed that

$$(4.9) \quad \begin{aligned} \lim_{m \rightarrow \infty} P \left(F_m^{-1} \left(1 - (t + c_1 m^{-1/2}) \right) \leq F^{-1}(1-t) \leq F_m^{-1} \left(1 - (t - c_1 m^{-1/2}) \right) \right. \\ \left. \text{for all } \epsilon_m \leq t \leq 1 - \epsilon_m \right) \\ = P \left(\sup_{0 \leq t \leq 1} |B(t)| \leq c_1 \right). \end{aligned}$$

Since the samples are independent, Theorem 2.2 follows immediately from (4.8) and (4.9).

REFERENCES

- Andrews, D.F. and Herzberg, A.M. (1985). *Data*. Springer, New York.
 Bertrand-Rétali, M. (1978). Convergence uniforme d'un estimateur de la densité par la méthode du noyau. (French) *Rev. Roumaine Math. Pures Appl.*, **23**, 361-385.

- Claeskens, G. and van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, **31**, 1852-1884.
- Csörgő, M. and Horváth, L. (1985). Strong approximations of the quantile process of the product-limit estimator. *J. Multivariate Anal.*, **16**, 185-210.
- Csörgő, M. and Horváth, L. (1989). On confidence bands for the quantile function of a continuous distribution function. In *Coll. Math. Soc. J. Bolyai 57 Limit Theorems of Prob. Theory and Statist.*, Pécs, Hungary, 95-106. North-Holland, Amsterdam.
- Csörgő, M. and Horváth, L. (1993). *Weighted approximations in Probability and Statistics*. Wiley, Chichester.
- Csörgő, M. and Révész, P. (1981). *Strong approximations in probability and statistics*. Academic Press, Inc, New York-London.
- Csörgő, M. and Révész, P. (1984). Two approaches to constructing simultaneous confidence bounds for quantiles. *Prob. and Math. Statist.*, **4**, 221-236.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Stat.*, **7**, 1-26.
- Goddard, M. J. and Hinberg, I. (1990). Receiver operating characteristics (ROC) curves and non normal data: an empirical study. *Statist. Med.*, **9**, 325-337.
- Hall, P. and Rieck, A. (2001). Improving coverage accuracy of nonparametric prediction intervals. *J. Roy. Statist. Soc. Ser. B*, **63**, 717-725.
- Hsieh, F. and Turnbull, B.W. (1996). Non-parametric and semi-parametric estimation of the receiver operating characteristic curve. *Ann. Statist.*, **24**, 25-40.
- Li, G., Tiwari, R.C. and Wells, M.T. (1999). Semiparametric inference for a quantile comparison function with applications to receiver operating characteristic curves. *Biometrika*, **86**, 487-502.
- Shorack, G. R. (2000), *Probability for Statisticians*. Springer, New York.
- Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser. B*, **43**, 97-99.
- Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.
- Swets, I.A. and Pickett, R.M. (1982). Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. *New York: Academic Press*.
- Zhou, X.H. and Obuchowski, NA, and McClish, D.K. (2002). Statistical Methods in Diagnostic Medicine. *Wiley: New York*