tes in
Statistics

109

Helmut Rieder (Editor)

Robust Statistics, Data Analysis, and Computer Intensive Methods

In Honor of Peter Huber's 60th Birthday

Springer

The m out of n Bootstrap and Goodness of Fit Tests with Double Censored Data

Peter J. Bickel
University of California, Berkeley

Jian-Jian Ren
University of Nebraska, Lincoln

Abstract

This paper considers the use of the m out of n bootstrap (Bickel, Götze, and van Zwet, 1994) in setting critical values for Cramér–von Mises goodness of fit tests with doubly censored data. We show that, as might be expected, the usual n out of n nonparametric bootstrap fails to estimate the null distribution of the test statistic. We show that if the m out of n bootstrap with $m \to \infty$, $m = o(n)$ is used to set the critical value of the test, the proposed testing procedure is asymptotically level α, has the correct asymptotic power function for $\sqrt{n}$ alternatives and is asymptotically consistent.

Key words and phrases: Bootstrap, Cramér–von Mises statistics, doubly censored data, goodness of fit tests.
AMS 1991 subject classifications: Primary 62G10; secondary 62E20.

1 Introduction

It is logically clear but not always evident or appreciated that the usual nonparametric bootstrap (the n out of n bootstrap) should fail when one tries to estimate the distribution of test statistics under a semiparametric (restricted nonparametric) hypothesis and ignores the restrictions imposed by the hypothesis. For example, Freedman (1981) points out that in setting confidence intervals on the usual slope estimate for regression through the origin, one must resample not the residuals but the residuals centered at their mean. If one considers setting confidence bands as the dual of hypothesis testing, a moment's thought will show that not centering the residuals is tantamount to not imposing the model requirement that the expectation of the error is zero.

For more recent examples, see Härdle and Mammen (1993), Mammen (1992), Bickel, Götze and van Zwet (1994). The same phenomenon occurs when one considers hypothesis testing problems using a censored or incomplete sample. Imposing hypothesis restrictions for censored data can be very complicated given the unknown distributions of the censoring variables. In the situation we consider here, it is impossible to estimate the censoring distribution! Specifically, we consider Cramér–von Mises goodness of fit tests with doubly censored data. The problem is introduced in Section 2, where we show that the usual n out of n nonparametric bootstrap fails to estimate the null distribution of the test statistic. In Section 3, we propose that one uses the m out of n bootstrap to set the critical value of the test and show that the proposed testing procedure is asymptotically consistent and has correct power against $\sqrt{n}$ alternatives with the proofs deferred to Section 5. For the exponential distribution family, Section 4 presents a simulation result which compares the power functions using the true critical value and the one by the m out of n bootstrap.

One should note that the testing procedure proposed here includes the goodness of fit tests for proportional hazards model with right censored data (Wei (1984)) as a special case, and that the proposed procedure easily applies to many other models with censored data or with uncensored data. An alternative to the m out of n bootstrap method in the model of this paper, the Fredholm Integral Equation (FIE) method, has been constructed by Ren (1993). The m out of n bootstrap method, although less powerful to second order, has the advantage of easy implementation and generalization to essentially any semiparametric test.

In collaboration with Götze, van Zwet and others we are in the process of studying this broad applicability and the critical issue of choice of m. In the example we study in Section 4, $m = \sqrt{n}$ is chosen arbitrarily and evidently performs reasonably well.

We close the basic heuristics of this method which are evident in the classic problem of testing $H : \mu = 0$ vs $K : \mu > 0$ when $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, 1)$. The correct test is to reject if $\sqrt{n}\,\bar{X}_n \geq z(1-\alpha)$, where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $z(1-\alpha)$ is the $1-\alpha$ quantile of $\mathcal{N}(0,1)$. Suppose that we instead use the bootstrap critical value of $\sqrt{n}\,\bar{X}_n$. That is call $\hat{c}_{m,n}$ the $1-\alpha$ quantile of the distribution of $\sqrt{m}\,\bar{X}^*_m$, where $X^*_1, \ldots, X^*_m$ are a sample of size m with replacement from $X_1, \ldots, X_n$ and reject if $\sqrt{n}\,\bar{X}_n \geq \hat{c}_{m,n}$. It is easy to see that if $m \to \infty$ with $m = o(n)$, then $\hat{c}_{m,n} = z_{1-\alpha} + o_p(1)$ not only for $\mu = 0$ but also for $\mu = \frac{\Delta}{\sqrt{n}}$, $\Delta > 0$. Thus, the local power function of our test is the same as that of the z test. Furthermore, if $\mu > 0$ is fixed, then $\hat{c}_{m,n} = O_p(\sqrt{m})$ and $P_\mu\{\sqrt{n}\,\bar{X}_n \geq \hat{c}_{m,n}\} \to 1$ since $m = o(n)$. This is the kind of behavior we exhibit in our more complicated context.

2 Goodness of Fit Tests with Doubly Censored Data

In medical follow-up studies and in reliability studies, the data available to be analyzed are often incomplete due to various reasons. Recently, some more complicated types of censored data than right censored data, such as doubly censored data, interval censored data, truncated data, etc., have started to catch the attention of statisticians, as these data occur in important clinical trials.

For instance, doubly censored data are presented in a recent study of the age-dependent growth rate of primary breast cancer (Peer et al. (1993)). Other examples of doubly censored data encountered in practice have been given by Gehan (1965), Mantel (1967), Peto (1973), Turnbull (1974), and others. In statistical analysis, assumptions on the underlying lifetime distributions are often made for various reasons. To check if these assumptions are correct, one would naturally wish to conduct a goodness of fit test. In this study, we are interested in the goodness of fit test with doubly censored data.

Let $X_1, X_2, \ldots, X_n$ be independent observations on a nonnegative r.v. X with a continuous d.f. F. If it is wished to test the null hypothesis

$$H_0 : F = F_0, \tag{2.1}$$

where F_0 is a specified d.f., then the Cramér–von Mises test statistics are given by

$$T_n = n \int_0^\infty [F_n(x) - F_0(x)]^2 dF_0(x), \tag{2.2}$$

where F_n is the empirical d.f. based on $X_1, X_2, \ldots, X_n$. In this study, one observes not $\{X_i\}$ but a doubly censored sample:

$$W_i = \begin{cases} X_i & \text{if } Z_i \le X_i \le Y_i, \quad \delta_i = 1 \\ Y_i & \text{if } X_i > Y_i, \quad \delta_i = 2 \\ Z_i & \text{if } X_i < Z_i, \quad \delta_i = 3 \end{cases} \tag{2.3}$$

where (X_i, Y_i, Z_i), $i = 1, 2, \ldots, n$, are independent observations on (X, Y, Z) for nonnegative variables X, Y and Z, and the r.v.'s Y_i and Z_i are called *right* and *left censoring variables*, respectively. This means that X_i is observable whenever X_i lies in the interval $[Z_i, Y_i]$, and we know whether $X_i < Z_i$ or $X_i > Y_i$ and observe the value of Z_i or Y_i accordingly. The usual assumptions on this model are that X_i and (Y_i, Z_i) are independent and $P\{Y > Z\} = 1$ (see Chang and Yang (1987), Chang (1990), Gu and Zhang (1993)). Right censored data, that is, $Z \equiv -\infty$ in (2.3), is a special case of the doubly censored data (2.3). The problem considered here is to test the goodness of fit of F based on (W_i, δ_i).

Let (W_i, δ_i) be distributed as (W, δ), and denote

$$Q_j(t) = P\{W \le t, \delta = j\}, \quad j = 1, 2, 3,$$

$$S_X(t) = P\{X > t\}, \quad S_Y(t) = P\{Y > t\}, \quad S_Z(t) = P\{Z > t\}.$$

By Turnbull (1974), Tsai and Crowley (1985), the *Nonparametric Maximum Likelihood Estimator* (NPMLE) $\hat{S}_n$ of S_X is one of the solutions of the following equation:

$$\hat{S}_n(t) = 1 - Q^{(n)}(t) + \int_{u \le t} \frac{\hat{S}_n(t)}{\hat{S}_n(u)} dQ_2^{(n)}(u) - \int_{u > t} \frac{1 - \hat{S}_n(t)}{1 - \hat{S}_n(u)} dQ_3^{(n)}(u), \tag{2.4}$$

where

$$Q_j^{(n)}(t) = \frac{1}{n}\sum_{i=1}^{n} I\{W_i \le t, \delta_i = j\}, \qquad j = 1, 2, 3,$$

$$Q^{(n)}(t) = \sum_{j=1}^{3} Q_j^{(n)}(t),$$

with $\int_{u\le t} = 0$ if $S_X^{(n)}(t) = 0$ and $\int_{u>t} = 0$ if $S_X^{(n)}(t) = 1$ (The solutions of (2.4) are not unique, see Gu and Zhang (1993) for examples, but all solutions are asymptotically equivalent). This equation (2.4) is obtained by generalizing the idea of the self-consistent estimator for right censored data originally given by Efron (1967, (7.4) on page 840). One may see Efron (1967) for an intuitive explanation for such an equation. The NPMLE $\hat{S}_n$ may be found numerically by the method suggested by Mykland and Ren (1994), which is based on the EM algorithm.

Let U denote the uniform distribution function on $[0, 1]$, and let

$$\hat{F}_n = 1 - \hat{S}_n \quad \text{and} \quad \hat{U}_n = \hat{F}_n \circ F_0^{-1}. \tag{2.5}$$

$\hat{F}_n$ is the NPMLE of F. For the test (2.1), one would naturally want to replace F_n in (2.2) by $\hat{F}_n$; that is, to use

$$T_n = n\int_0^\infty [\hat{F}_n(x) - F_0(x)]^2 dF_0(x) = n\int_0^1 [\hat{U}_n - U]^2 dU \tag{2.6}$$

as the test statistics. For the rest of the paper, we will always refer T_n to (2.6). If $U_0 = F\circ F_0^{-1}$, under some regularity conditions Gu and Zhang (1993) showed that

$$\sqrt{n}\,[\hat{U}_n - U_0] \text{ weakly converges to } G_F, \qquad \text{as } n \to \infty \tag{2.7}$$

where G_F is a Gaussian process with mean 0 and a covariance function γ. The covariance function γ is determined by S_X, S_Y and S_Z and its specific formulation is given by Ren (1993). Hence, it is easy to see that the null distribution of T_n is given by

$$T_n \xrightarrow{\mathcal{D}} \int_0^1 G_{F_0}^2 dU = \sum_{j=1}^{\infty} \lambda_j Z_j^2, \qquad n \to \infty \tag{2.8}$$

where Z_j are independent normal random variables with mean 0 and variance 1, and λ_j are the eigenvalues for the following eigenvalue problem:

$$\int_0^1 \gamma(s,t)\phi(t)dt = \lambda\phi(s), \qquad \phi \in L^2[0, \beta]. \tag{2.9}$$

One should note that even under H_0, the limiting covariance function γ of the NPMLE $\hat{F}_n$ depends on unknown survival functions S_Y and S_Z. To set the critical value of the test, one needs to estimate the unknown null distribution given by (2.8). As expected, the usual nonparametric n out of n bootstrap fails in this case.

To see the reason for this failure and justify our other arguments, we need to extend the central limit theorem for the bootstrapped empirical process of Giné and Zinn (1990) to the doubly censored case considered in this study. We state the results as below with a proof sketched in Section 5 under the assumption (A1)–(A6) outlined in Chang (1990). Under these conditions (A1)–(A6), Chang (1990) established (2.7) on $[0, T]$ for any $T > 0$.

Let m be the bootstrap sample size satisfying $m \to \infty$ as $n \to \infty$, and let $\hat{F}_m^*$ be the NPMLE based on the bootstrap sample. Then

Proposition 2.1 *Suppose that $\sqrt{n}\,[\hat{F}_n - F]$ weakly converges to a Gaussian process Z as $n \to \infty$. Then, with probability 1, $\sqrt{m}\,[\hat{F}_m^* - \hat{F}_n]$ weakly converges to Z as $n \to \infty$.*

Immediately, we have that with probability 1,

$$\sqrt{m}\,[\hat{U}_m^* - \hat{U}_n] \text{ weakly converges to } G_F, \tag{2.10}$$

where $\hat{U}_m^* = \hat{F}_m^* \circ F_0^{-1}$. Let T_m^* be given by (2.6) based on the bootstrap sample. Hence, for $m = n$, i.e., when the n out of n bootstrap is used, we have

$$\sqrt{n}\,[\hat{U}_n^* - U] = \sqrt{n}\,[\hat{U}_n^* - \hat{U}_n] + \sqrt{n}\,[\hat{U}_n - U] = G_n^* + G_n, \tag{2.11}$$

and thus

$$T_n^* = \int_0^1 [G_n^*]^2 dU + 2\int_0^1 G_n^* G_n dU + \int_0^1 [G_n]^2 dU. \tag{2.12}$$

Let $\boldsymbol{O}_n = \{(W_i, \delta_i);\ 1 \le i \le n\}$, then from (2.7) and (2.10), we know that under H_0,

$$P\{T_n^* \le t \mid \boldsymbol{O}_n\} \longrightarrow P\Big\{\int_0^1 [G_{F_0}^* + G_{F_0}]^2 dU \le t \mid G_{F_0}\Big\}, \quad \text{as } n \to \infty,$$

where $G_{F_0}^*$ and G_{F_0} are two independent centered Gaussian processes with the same covariance function. As stated in the introduction, we need to "bootstrap" taking the restrictions of H_0 into account. For the simpler right censored case, one may draw a sample $X_1^*, \ldots, X_n^*$ from F_0 and a sample $Y_1^*, \ldots, Y_n^*$ from the estimator of S_Y (see Chang and Yang (1987) for the formulation of such an estimator) to obtain the bootstrap sample. It is well known that the n out of n bootstrap method estimates the null distribution consistently. However, this method does not easily generalize to the doubly censored case. The reason is that in the doubly censored case, one would need to obtain a sample of $(Y_1^*, Z_1^*), \ldots, (Y_n^*, Z_n^*)$ from an estimator which estimates the joint distribution of (Y, Z), because Y and Z are usually not independent random variables.

From (2.11) and (2.12), we see that in a general way, the problem can be solved by resampling fewer observations, i.e., using $m = o(n)$ with $m \to \infty$ as the bootstrap sample size. More specifically, with $m = o(n)$, by (2.7), we have that (2.11) becomes

$$\begin{aligned}\sqrt{m}\,|\hat{U}_m^* - U\,] &= \sqrt{m}\,[\hat{U}_m^* - \hat{U}_n] + \sqrt{m}\,[\hat{U}_n - U\,] \\ &= \sqrt{m}\,[\hat{U}_m^* - \hat{U}_n] + o_p(1) = G_m^* + o_p(1).\end{aligned} \tag{2.13}$$

Hence, from (2.10), we have that (2.12) becomes

$$T_m^* = \int_0^1 [G_m^*]^2 dU + o_p(1), \tag{2.14}$$

which along with (2.10) implies that under H_0,

$$P\{T_m^* \le t|\,O_n\} \longrightarrow P\Big\{\int_0^1 [G_{F_0}]^2 dU \le t\Big\}, \qquad \text{as } n \to \infty.$$

Therefore, the m out of n bootstrap is consistent under H_0. Based on this, we propose the m out of n bootstrap tests for the test (2.1) in the next section.

3 The m out of n Bootstrap

For the test (2.1) with a doubly censored sample (2.3), we propose that one uses $m = o(n)$ as the bootstrap sample size, and use C_α^* as the critical value of the test, where for $0 < \alpha < 1$ and $O_n = \{(W_i, \delta_i)\,;\ 1 \le i \le n\}$, C_α^* is given by

$$P\{T_m^* \ge C_\alpha^*|\,O_n\} = \alpha. \tag{3.1}$$

This is called the m *out of* n *bootstrap* method. In the following, we investigate the asymptotic properties of this test procedure

Let the limiting null distribution of T_n be distributed as

$$T = \int_0^1 G_{F_0}^2 dU \tag{3.2}$$

and let C_α^0 be the true critical value given by

$$P\{T \ge C_\alpha^0\} = \alpha. \tag{3.3}$$

Proposition 3.1 *Let* $\|\cdot\|$ *denote the uniform norm and let* $m = o(n)$ *with* $m \to \infty$ *as* $n \to \infty$. *Then, when Proposition 2.1 holds,*

(i) if $\|F - F_0\| = O(1/\sqrt{n})$, *we have*

$$C_\alpha^* \xrightarrow{P} C_\alpha^0, \quad \text{as } n \to \infty; \tag{3.4}$$

(ii) if $F = F_1 \neq F_0$, we have

$$C^*_\alpha \xrightarrow{P} \infty, \quad \text{as } n \to \infty; \tag{3.5}$$

with

$$C^*_\alpha / T_n \xrightarrow{P} 0, \quad \text{as } n \to \infty. \tag{3.6}$$

The proof of Proposition 3.1 is given in Section 5.
From (3.4), (2.7) and (3.3), we have

$$P\{T_n \geq C^*_\alpha | H_0\} \longrightarrow \alpha, \quad \text{as } n \to \infty. \tag{3.7}$$

Consider the contiguous alternatives

$$H_n : F = F_n \neq F_0, \tag{3.8}$$

where for a bounded function Δ, $\sqrt{n}[F_n - F_0] = \Delta_n \to \Delta$, as $n \to \infty$. Denoting $U_n = F_n \circ F_0^{-1}$, we have

$$\sqrt{n}[\hat{U}_n - U] = \sqrt{n}[\hat{U}_n - U_n] + \Delta_n \circ F_0^{-1}.$$

Note that under the condition of Corollary 1 of Gu and Zhang (1993) and under H_0, $\sqrt{n}[\hat{U}_n - U]$ is, in probability, equivalent to a continuous operator of the empirical processes $\sqrt{n}[Q_j^{(n)} - Q_j]$, $j = 1, 2, 3$ (see the proof of Theorem 2 by Gu and Zhang, 1993). Hence, by contiguity, we have that for $\Delta_0 = \Delta \circ F_0^{-1}$,

$$T_n \xrightarrow{\mathcal{D}} T_{\Delta_0} = \int_0^1 [G_{F_0} + \Delta_0]^2 dU = \sum_{j=1}^{\infty} \lambda_j Z^2_{\Delta j}, \quad \text{as } n \to \infty \tag{3.9}$$

where λ_j's are given by (2.8) and $Z_{\Delta j}$ are independent normal r.v.'s with non-zero mean and variance 1. Hence, from (3.4) we have

$$P\{T_n \geq C^*_\alpha | H_n\} \longrightarrow P\{T_{\Delta_0} \geq C^0_\alpha\}, \quad \text{as } n \to \infty \tag{3.10}$$

where $P\{T_{\Delta_0} \geq C^0_\alpha\}$ is the true asymptotic power of the test (2.1) under the contiguous alternatives (3.8). Moreover, under the fixed alternative

$$H_1 : F = F_1 \ (\neq F_0) \tag{3.11}$$

we know that (3.6) implies

$$P\{T_n \geq C^*_\alpha | H_1\} \longrightarrow 1, \quad \text{as } n \to \infty. \tag{3.12}$$

To summarize, (3.7), (3.10) and (3.12) give the following theorem on the asymptotic properties of the proposed m out of n bootstrap test procedure.

Theorem 3.2 *Let $m = o(n)$ with $m \to \infty$ as $n \to \infty$. In the test (2.1) with data (2.3), if T_n given by (2.6) is used as the test statistic and if C^*_α given by (3.1) is used as the critical value of the test, then, when Proposition 2.1 holds,*

(i) *for* $0 < \alpha < 1$, $\lim_{n\to\infty} P\{T_n \geq C_\alpha^* | H_0\} = \alpha$;

(ii) *for contiguous alternatives* H_n *in (3.8)*, $\lim_{n\to\infty} P\{T_n \geq C_\alpha^* | H_n\} = P\{T_{\Delta_0} \geq C_\alpha^0\}$;

(iii) *for the fixed alternative* H_1 *in (3.11)*, $\lim_{n\to\infty} P\{T_n \geq C_\alpha^* | H_1\} = 1$.

4 A Small Simulation Study

In this section, we consider the goodness of fit test (2.1) with $F_0 = \text{Exp}(1)$ against the exponential distributions $\text{Exp}(\mu)$, where $\text{Exp}(\mu)$ denotes the exponential distribution with mean μ. We denote the power functions of the test with the true critical value C_α^0 given by (3.3) and the m out of n bootstrap critical value C_α^* given by (3.1) as

$$P_0(\mu) = P\{T_n \geq C_\alpha^0 | \text{Exp}(\mu)\} \tag{4.1}$$

and

$$P_b(\mu) = P\{T_n \geq C_\alpha^* | \text{Exp}(\mu)\}, \tag{4.2}$$

respectively, where T_n is given by (2.6). For the Fredholm Integral Equation (FIE) method proposed by Ren (1993), we denote the power function as

$$P_f(\mu) = P\{T_n \geq C_\alpha^f | \text{Exp}(\mu)\}, \tag{4.3}$$

where C_α^f is the critical value estimated by FIE such that $P\{T \geq C_\alpha^f\} = \alpha$. The details on the implementation of the FIE method may be found in Ren and Ledder (1995).

In Figure 1, we compare the curves of P_0, P_b and P_f, which are obtained by the simulation with sample size $n = 200$, $m = \sqrt{n}$ and $\alpha = 0.05$. Because the use of the EM algorithm in (2.4) is very time consuming to conduct the simulation study for doubly censored data, in Figure 1 we consider the right censored case with the right censoring variable Y from $\text{Exp}(3)$. One may note that for the right censored case, we have $S_Z \equiv 0$ and the solution $\hat{S}_n$ of (2.4) is the Kaplan–Meier estimator (Chang and Yang (1987)). In our study, the true critical value C_α^0 is obtained from the Monte Carlo method. All simulation results are based on 300 runs, and for each run, the percentiles of P_0, P_b, P_f, C_α^0, C_α^* and C_α^f are obtained from 400 replications of the procedures.

From Figure 1, we easily see that the simulation results are consistent with our asymptotic results in Section 3. As μ gets further away from 1, the power of the test using the m out of n bootstrap critical value C_α^*, although smaller than the true one by C_α^0, goes to 1. As μ gets closer to 1, the power of the test using C_α^* goes to α. We also note that the curve P_f by FIE is closer to P_0 than P_b by bootstrap tests. However, the FIE method method is constructed through solving the Fredholm integral equation of the second kind, and the programming of FIE is no trivial matter. Also, FIE does not easily extend to other hypothesis testing problems.

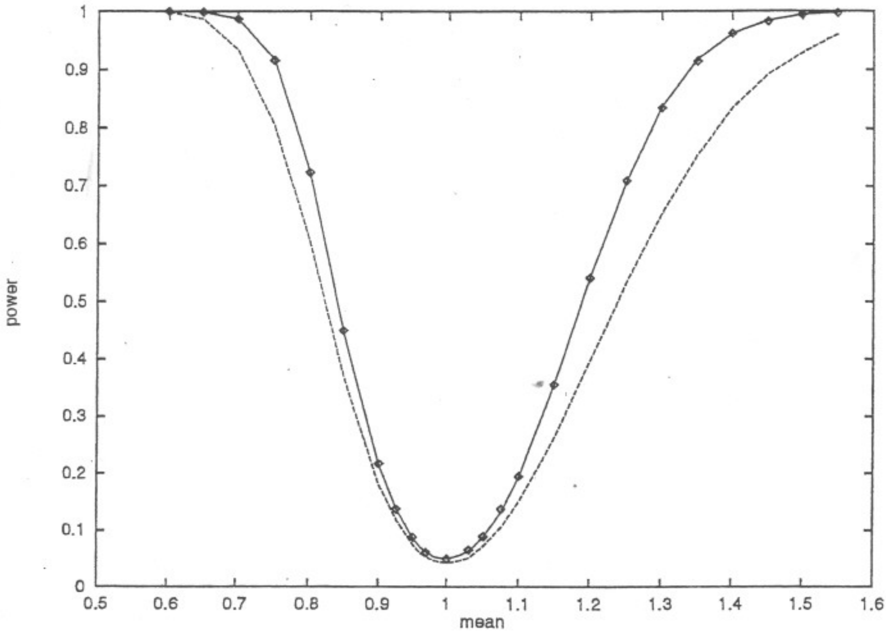

Figure 1: P_0 : — ; P_b : - - - ; P_f : ◇◇◇ ; $X \sim \text{Exp}(\mu)$, $Y \sim \text{Exp}(3)$

A better choice of m and extrapolation as discussed in Bickel, Götze and van Zwet (1994) should also improve the power behavior of the m out of n bootstrap.

5 Proofs

PROOF OF PROPOSITION 2.1 The proposition may be shown under weaker conditions. For simplicity, in this paper we will sketch the proof under the assumptions A1 through A6 outlined in Chang (1990).

Under these assumptions, Chang (1990) showed that for any $T > 0$,

$$\sqrt{n}\,[\boldsymbol{S}_n - \boldsymbol{S}] \text{ weakly converges to } \boldsymbol{Z}, \quad \text{as } n \to \infty \tag{5.1}$$

where $\boldsymbol{Z}$ is a Gaussian process on $[0, T]$,

$$\boldsymbol{S}_n = (S_X^{(n)}, S_Y^{(n)}, S_Z^{(n)}), \quad \boldsymbol{S} = (S_X, S_Y, S_Z), \quad S_X^{(n)} = \hat{S}_n, \tag{5.2}$$

and $S_Y^{(n)}$, $S_Z^{(n)}$ are the estimators of S_Y, S_Z, respectively, given by Chang and Yang (1987). Let

$$\boldsymbol{Q}_n = (Q_1^{(n)}, Q_2^{(n)}, Q_3^{(n)}) \quad \text{and} \quad \boldsymbol{Q} = (Q_1, Q_2, Q_3). \tag{5.3}$$

Chang (1990, (14) of page 395) showed that there exist a continuous operator L_1 and a linear operator L_2 such that

$$\sqrt{n}\,[\boldsymbol{S}_n - \boldsymbol{S}] = L_1 \circ L_2(\sqrt{n}\,[\boldsymbol{Q}_n - \boldsymbol{Q}]) + L_1(\boldsymbol{\theta}_n), \tag{5.4}$$

where $\boldsymbol{\theta}_n$ is given by (9) of Chang (1990, page 394). Noting that $\boldsymbol{Q}_n$ is the empirical process and that $\sqrt{n}\,[\boldsymbol{Q}_n - \boldsymbol{Q}]$ weakly converges to a Gaussian process $\boldsymbol{Z}_Q$, Chang (1990) established (5.1) by showing that

$$\boldsymbol{\theta}_n \xrightarrow{\text{a.s.}} 0, \qquad \text{as } n \to \infty \tag{5.5}$$

uniformly on $[0, T]$.

Let $\boldsymbol{S}^*_m$, $\boldsymbol{Q}^*_m$ and $\boldsymbol{\theta}^*_m$ be $\boldsymbol{S}_n$, $\boldsymbol{Q}_n$ and $\boldsymbol{\theta}_n$ based on the bootstrap sample from (W_i, δ_i), $i = 1, \ldots, n$, respectively. Since with probability 1,

$$\boldsymbol{S}_n \longrightarrow \boldsymbol{S}, \quad \text{as } n \to \infty, \tag{5.6}$$

uniformly on $[0, T]$ (Chang and Yang (1987)), the conditions required for (5.4) hold with probability 1, as $n \to \infty$. Hence, we have that with probability 1,

$$\sqrt{m}\,[\boldsymbol{S}^*_m - \boldsymbol{S}_n] = L_1 \circ L_2(\sqrt{m}\,[\boldsymbol{Q}^*_m - \boldsymbol{Q}_n]) + L_1(\boldsymbol{\theta}^*_m). \tag{5.7}$$

From Giné and Zinn (1990), we know that with probability 1,

$$\sqrt{m}\,[\boldsymbol{Q}^*_m - \boldsymbol{Q}_n] \text{ weakly converges to } \boldsymbol{Z}_Q, \qquad \text{as } n \to \infty. \tag{5.8}$$

Since L_1 is a continuous operator and L_2 is a linear operator, by Theorem 5.1 of Billingsley (1968), (5.8) implies that, as $n \to \infty$, with probability 1, $L_1 \circ L_2(\sqrt{m}\,[\boldsymbol{Q}^*_m - \boldsymbol{Q}_n])$ weakly converges to the Gaussian process $\boldsymbol{Z} = L_1 \circ L_2(\boldsymbol{Z}_Q)$. Therefore, by (5.7), it suffices to show that,

$$\boldsymbol{\theta}^*_m \xrightarrow{\text{a.s.}} 0, \qquad \text{as } n \to \infty \tag{5.9}$$

uniformly on $[0, T]$.

The key to (5.9) is that one needs to show that with probability 1,

$$\|\,\boldsymbol{S}^*_m - \boldsymbol{S}_n\| \xrightarrow{\text{a.s.}} 0, \qquad \text{as } n \to \infty. \tag{5.10}$$

The rest of the proof just follows line by line of Lemma 3.1 through Lemma 3.3 in Chang (1990).

Denoting the equation (2.7) of Chang and Yang (1987) as $\boldsymbol{E}(\boldsymbol{Q}_n, \boldsymbol{S}_n) = \boldsymbol{0}$, $\boldsymbol{S}^*_m$ is the solution of $\boldsymbol{E}(\boldsymbol{Q}^*_m, \boldsymbol{S}^*_m) = \boldsymbol{0}$. Since $\boldsymbol{Q}_n$ is the empirical process, from (5.6) we know that with probability 1, $\boldsymbol{E}(\boldsymbol{Q}_n, \boldsymbol{S}_n)$ has limit $\boldsymbol{E}(\boldsymbol{Q}, \boldsymbol{S})$ as $n \to \infty$. Since each component of $\boldsymbol{S}^*_m$ is uniformly bounded and nonincreasing, by Helly's Theorem, for each fixed t, $\boldsymbol{S}^*_m$ has a convergent subsequence, say with limit $\boldsymbol{S}^*_0$. This limit $\boldsymbol{S}^*_0$ should satisfy the equation $\boldsymbol{E}(\boldsymbol{Q}, \boldsymbol{S}) = \boldsymbol{0}$. From the identifiability for the solution of $\boldsymbol{E}(\boldsymbol{Q}, \boldsymbol{S}) = \boldsymbol{0}$ studied by Chang and Yang (1987), we know that $\boldsymbol{S}^*_0$ must be the solution of $\boldsymbol{E}(\boldsymbol{Q}, \boldsymbol{S}) = \boldsymbol{0}$, i.e., $S_0^* = S$. Since each component of S_m^* is uniformly bounded and non-increasing and since each component of S is continuous, we have that with probability 1,

$$\| S_m^* - S \| \xrightarrow{\text{a.s.}} 0, \qquad \text{as } n \to \infty. \tag{5.11}$$

Hence, (5.10) follows from (5.6) and (5.11). ////

PROOF OF PROPOSITION 3.1 (i) Note that

$$\sqrt{m}\,[\hat{U}_m^* - U\,] = \sqrt{m}\,[\hat{U}_m^* - \hat{U}_n] + \sqrt{m}\,[\hat{U}_n - U_0] + \sqrt{m}\,[U_0 - U\,]. \tag{5.12}$$

Hence, by (2.7) and the assumptions: $\|F - F_0\| = \mathrm{O}(1/\sqrt{n}\,)$ and $m = \mathrm{o}(n)$ with $m \to \infty$ as $n \to \infty$, we have

$$\sqrt{m}\,[\hat{U}_m^* - U\,] = \sqrt{m}\,[\hat{U}_m^* - \hat{U}_n] + \mathrm{o}_p(1), \quad \text{as } n \to \infty \tag{5.13}$$

where $\mathrm{o}_p(1)$ uniformly converges to 0 in probability as $n \to \infty$. From (2.10) and (3.2), we know that with probability 1,

$$\sup_{-\infty<x<\infty} |P\{\hat{T}_m^* \le x|\,\boldsymbol{O}_n\} - P\{T \le x\}| \longrightarrow 0, \quad \text{as } n \to \infty \tag{5.14}$$

where

$$\hat{T}_m^* = m \int_0^1 [\hat{U}_m^* - \hat{U}_n]^2 dU. \tag{5.15}$$

Hence, from (6.13) and (6.14), we have that

$$\sup_{-\infty<x<\infty} |P\{T_m^* \le x|\,\boldsymbol{O}_n\} - P\{T \le x\}| \xrightarrow{P} 0, \quad \text{as } n \to \infty. \tag{5.16}$$

For C_α^* given by (3.1) and C_α^0 by (3.3), we know that (5.16) implies

$$|P\{T_m^* \ge C_\alpha^*|\,\boldsymbol{O}_n\} - P\{T \ge C_\alpha^*\}| = |\alpha - P\{T \ge C_\alpha^*\}|$$

$$= |P\{T \ge C_\alpha^0\} - P\{T \ge C_\alpha^*\}| \;\xrightarrow{P}\; 0, \quad \text{as } n \to \infty.$$

Therefore, (3.4) follows from the continuity of the d.f. of T.
(ii) When $F = F_1 \ne F_0$, from (5.12), (2.7), (2.10) and the assumption that $m = \mathrm{o}(n)$, we have

$$\begin{aligned} m[\hat{U}_m^* - U\,]^2 \;&=\; m[\hat{U}_m^* - \hat{U}_n]^2 + m[U_0 - U\,]^2 + \\ &\quad + 2m[\hat{U}_m^* - \hat{U}_n][\,U_0 - U\,] + \mathrm{o}_p(1) \end{aligned} \tag{5.17}$$

and thus by (5.15) and (2.10),

$$T_m^* = \hat{T}_m^* + m \int_0^1 [\,U_0 - U\,]^2 dU + 2m \int_0^1 [\hat{U}_m^* - \hat{U}_n][\,U_0 - U\,] dU + \mathrm{o}_p(1) \tag{5.18}$$

$$= O_p(1) + m \int_0^1 [U_0 - U]^2 dU + O_p(\sqrt{m})$$

Noting that (3.1) implies

$$P\{T_m^* < C_\alpha^* | O_n\} = 1 - \alpha, \tag{5.19}$$

hence, (3.5) follows from (5.18) and

$$m \int_0^1 [U_0 - U]^2 dU = m \int_0^\infty [F_1 - F_0]^2 dF_0 \longrightarrow \infty, \quad \text{as } n \to \infty. \tag{5.20}$$

To show (3.6), one just needs to notice that by (2.7),

$$T_n = O_p(1) + O_p(\sqrt{n}) + n \int_0^1 [U_0 - U]^2 dU,$$

where $O_p(1)$ is bounded in probability, and that by (3.1),

$$\alpha = P\{T_m^* \geq C_\alpha^* | O_n\} = P\{T_m^*/T_n \geq C_\alpha^*/T_n | O_n\}.$$

Therefore, (3.6) follows from (5.18), (5.20) and $T_m^*/T_n \to 0$ as $n \to \infty$. ////

References

[1] Bickel, P.J., Götze, F. and Van Zwet, W.R. (1994): *Resampling Fewer than n Observations: Gains, Losses and Remedies for Losses*. Preprint.

[2] Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley and Sons, New York.

[3] Chang, M.N. and Yang, G.L. (1987): Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *Ann. Statist.* **15** 1536–1547.

[4] Chang, M.N. (1990): Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *Ann. Statist.* **18** 391–404.

[5] Efron, B. (1967): The two sample problem with censored data. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **4** 831–853.

[6] Freedman, D.F. (1981): Bootstrapping regression models. *Ann. Statist.* **12** 1218–1228.

[7] Gehan, E.A. (1965): A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* **52** 650–653.

[8] Giné, E. and Zinn, J. (1990): Bootstrapping general empirical measures. *Ann. Prob.* **18** 852–869.

[9] Gu, M.G. and Zhang, C.H. (1993): Asymptotic properties of self-consistent estimators based on doubly censored data. *Ann. Statist.* **21**, 611–624.

[10] Härdle, W. and Mammen, E. (1993): Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.

[11] Mammen, E. (1992): *When Does Bootstrap Work?* Springer Verlag, New York.

[12] Mykland, P.A. and Ren, J. (1994): Self-Consistent and Maximum Likelihood Estimation for Doubly Censored Data. Submitted to *Ann. Statist.* (under revision).

[13] Peer, P.G., Van Dijck, J.A., Hendriks, J.H., Holland, R. and Verbeek, A.L. (1993): Age-dependent growth rate of primary breast cancer. *Cancer* **71**(11), 3547–3551.

[14] Peto, R. (1973): Experimental survival curves for interval-censored data. *Appl. Statist.* **22** 86–91.

[15] Ren, J. (1993): Generalized Cramér-von Mises tests of goodness of fit with censored data. *Ann. Inst. Statist. Math.*, to appear.

[16] Ren, J. and Ledder, G. (1995): On Cramér-von Mises Tests of Goodness of Fit with Censored Data. *Communications in Statistics*, to appear.

[17] Tsai, W.Y. and Crowley, J. (1985): A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency. *Ann. Statist.* **13** 1317–1334.

[18] Turnbull, B.W. (1974): Nonparametric estimation of a survivorship function with doubly censored data. *J. Amer. Statist. Assoc.* **69** 169–173.

[19] Wei, L.J. (1984): Testing goodness of fit for the proportional hazards model with censored observations. *J. Amer. Statist. Assoc.* **79**, no. 387, 649–652.