

# Model Based Bootstrap Methods for Interval Censored Data

Bodhisattva Sen and Gongjun Xu

Columbia University and University of Minnesota

December 31, 2013

## **Abstract**

We investigate the performance of model based bootstrap methods for constructing point-wise confidence intervals around the survival function with interval censored data. We show that bootstrapping from the nonparametric maximum likelihood estimator of the survival function is inconsistent for both the current status and case 2 interval censoring models. A model based smoothed bootstrap procedure is proposed and shown to be consistent. In addition, simulation studies are conducted to illustrate the (in)-consistency of the bootstrap methods. Our conclusions in the interval censoring model would extend more generally to estimators in regression models that exhibit non-standard rates of convergence.

## **1 Introduction**

In recent years there has been considerable research on the analysis of interval censored data. Such data arise extensively in epidemiological studies and clinical trials,

especially in large-scale panel studies where the event of interest, which is typically an infection with a disease or some other failure (like organ failure), is not observed exactly but is only known to happen between two consecutive examination times. In particular, large-scale HIV/AIDS studies typically yield various types of interval censored data where interest centers on the distribution of time to HIV infection, but the exact time of infection is only known to lie between two consecutive followups at the clinic.

For general interval censored data, often called *mixed case* interval censoring, an individual is checked at several time points and the status of the individual is ascertained: 1 if the infection/failure has occurred by the time he/she is checked and 0 otherwise. Let  $X$  be the unobserved time of onset of some disease, having distribution function  $F$ , and let  $T_1 \leq T_2 \leq \dots \leq T_K$  be the  $K$  observation times. Here  $X$  and  $(T_1, \dots, T_K)$  are assumed to be independent and  $K$  can be random. We observe  $(T_1, \dots, T_K, \Delta_1, \dots, \Delta_K)$  where  $\Delta_k = \mathbf{1}_{T_{k-1} < X \leq T_k}$ ,  $k = 1, \dots, K$ , with  $T_0 = 0$  and  $\mathbf{1}$  denoting the indicator function. Our data consist of  $n$  independent and identically distributed copies of  $(T_1, \dots, T_K, \Delta_1, \dots, \Delta_K)$ . We are interested in making inference about the value of  $F$  at a pre-specified location  $t_0 > 0$ , assumed to be in the interior of the support of  $F$ .

When the observation number  $K \equiv 1$ , we say that we have *case 1* interval censoring or *current status data*. In this case, our observations are  $(T_i, \Delta_i)$  with  $\Delta_i = \mathbf{1}_{X_i \leq T_i}$ ,  $i = 1, \dots, n$ . The nonparametric maximum likelihood estimator (NPMLE)  $\tilde{F}_n$  of  $F$  maximizes the log-likelihood function

$$\mathbb{F} \mapsto \sum_{i=1}^n \{\Delta_i \log \mathbb{F}(T_i) + (1 - \Delta_i) \log(1 - \mathbb{F}(T_i))\} \quad (1)$$

over all distribution functions  $\mathbb{F}$  and it can be characterized as the left derivative of the greatest convex minorant of the cumulative sum diagram of the data; see page

41 of Groeneboom and Wellner (1992). Let  $G$  be the distribution function of  $T$  and assume that  $F$  and  $G$  are continuously differentiable at  $t_0$  with derivatives  $f(t_0) > 0$  and  $g(t_0) > 0$ , respectively. Under these assumptions, it is well known that

$$\gamma_n := n^{1/3}\{\tilde{F}_n(t_0) - F(t_0)\} \rightarrow \kappa\mathbb{C}, \quad (2)$$

in distribution, where  $\kappa = [4F(t_0)\{1 - F(t_0)\}f(t_0)/g(t_0)]^{1/3}$ ,  $\mathbb{C} = \arg \min_{h \in \mathbb{R}} \{\mathbb{Z}(h) + h^2\}$ , and  $\mathbb{Z}$  is a standard two-sided Brownian motion process, originating from 0.

In the general mixed case interval censoring model, the limiting distribution of the NPMLE is unknown. In fact, in the literature only very limited theoretical results are available on the NPMLE. Groeneboom and Wellner (1992) discussed the asymptotics of the behavior of the NPMLE in a particular version of the case 2 censoring model ( $K = 2$ ); Wellner (1995) studied the consistency of the NPMLE where each subject gets exactly  $k$  examination times; van der Vaart and Wellner (2000a) proved the consistency of the NPMLE of the mixed case interval censoring in the Hellinger distance; see also Schick and Yu (2000) and Song (2004).

We are interested in constructing a pointwise confidence interval for  $F$  at  $t_0$  in the general mixed case censoring model. In the literature, very few results exist that address the construction of pointwise confidence intervals (Song, 2004; Sen and Banerjee, 2007). Even in the current status model, where we know the limiting distribution of the NPMLE, to construct a confidence interval for  $F(t_0)$  we need to estimate the nuisance parameter  $\kappa$ , which is indeed quite difficult – it involves estimation of the derivative of  $F$  and that of the distribution of  $T$ . For the current status model, there exist a few methods that can be used for constructing confidence intervals for  $F(t_0)$ : The  $m$ -out-of- $n$  bootstrap method and subsampling are known to be consistent in this setting (Politis et al., 1999; Lee and Pun, 2006). However, both methods require the choice of a *block* size. In practice, the choice of this tuning parameter

is quite tricky and the confidence intervals vary drastically with different choices of the block size. The estimation of the nuisance parameter  $\kappa$  can be avoided by using the likelihood-ratio test of Banerjee and Wellner (2001). Recently, Groeneboom et al. (2010) proposed estimates of  $F(t_0)$  based on smoothed likelihood function and smoothed NPMLE. However, the limiting distributions depend on the derivative of the density function.

In this paper we consider bootstrap methods for constructing confidence intervals for  $F(t_0)$  and investigate the (in)-consistency and performance of two model-based bootstrap procedures that are based on the NPMLE of  $F$ , in the general framework of mixed case interval censoring. Bootstrap intervals avoid the problem of estimating nuisance parameters and are generally reliable in problems with  $\sqrt{n}$  convergence rates. See Bickel and Freedman (1981), Singh (1981), and Shao and Tu (1995) and references therein.

In regression models, there are two main bootstrapping strategies: “bootstrapping pairs” and “bootstrapping residuals” (see e.g., page 113 of Efron and Tibshirani, 1993). Abrevaya and Huang (2005) considered “bootstrapping pairs”, i.e., bootstrapping from the empirical distribution function of the data, and showed that the procedure is *inconsistent* for the current status model and also other cube-root convergent estimators. In “bootstrapping residuals” one fixes (conditions on) the predictor values and generates the response according to the estimated regression model using bootstrapped residuals. In a binary regression problem, as in the current status model, this corresponds to generating the responses as independent Bernoulli random variables with success probability obtained from the fitted regression model.

In this paper we focus on the “bootstrapping residuals” procedure. In particular, for the mixed-case interval censoring model, conditional on an individual’s observation times  $T_1, \dots, T_K$ , we generate bootstrap sample  $(\Delta_1^*, \dots, \Delta_K^*, 1 - \sum_{k=1}^K \Delta_k^*)$  following

a multinomial distribution with  $n = 1$  and  $p_k = \hat{F}_n(T_k) - \hat{F}_n(T_{k-1})$ ,  $k = 1, \dots, K + 1$ , i.e.,

$$(\Delta_1^*, \dots, \Delta_K^*, 1 - \sum_{k=1}^K \Delta_k^*) \sim \text{Multinomial}(1, \{\hat{F}_n(T_k) - \hat{F}_n(T_{k-1})\}_{k=1}^{K+1}),$$

where  $\hat{F}_n$  is an estimator of  $F$  and  $\hat{F}_n(T_0) = 0, \hat{F}_n(T_{K+1}) = 1$ . We call this a model-based bootstrap scheme, as it uses the inherent features of the model. We study the behavior of the bootstrap method when  $\hat{F}_n = \tilde{F}_n$ , the NPMLE of  $F$ , and  $\hat{F}_n$  is a smooth estimator of  $F$ . Specifically, in Section 2.1 we state a general bootstrap convergence result for the current status model which provides sufficient conditions for any bootstrap scheme to be consistent. In Section 2.2 we illustrate, both theoretically and through simulation, the inconsistency of the NPMLE bootstrap method. The failure of the NPMLE bootstrap is mostly due to the non-differentiability of  $\tilde{F}_n$ . On the other hand, the smoothed NPMLE is differentiable and successfully mimics the local behavior of the true distribution function  $F$  at the location of interest, i.e.,  $t_0$ . As a result, the method yields asymptotically valid confidence intervals; see Section 2.3 where we prove the consistency of the smoothed bootstrap procedure, again in the current status model. The smoothed bootstrap procedure requires the choice of a smoothing bandwidth and we discuss this problem of bandwidth selection in Section 2.4.

Next, in Section 3, we study the case 2 interval censoring model, i.e., when  $K \equiv 2$ . Even in this case, the distribution of the NPMLE is not completely known, although conjectures and partial results exist. Groeneboom (1991) studied a one-step estimate  $F_n^{(1)}$ , obtained at the first step of the iterative convex minorant algorithm (see Groeneboom and Wellner, 1992) and conjectured that  $F_n^{(1)}$  is asymptotically equivalent to the NPMLE. This conjecture is called *the working hypothesis* in this paper and is still unproved. We assume that this conjecture holds and focus on bootstrapping the

distribution of the one-step estimator. We show the inconsistency of bootstrapping from the NPMLE and the consistency of the smoothed bootstrap method.

In general mixed case interval censoring, Sen and Banerjee (2007) introduced a pseudolikelihood method for estimating  $F(t_0)$ . However, the pseudolikelihood does not use the full information in the data and may not be as efficient as the NPMLE (Song, 2004). This is illustrated by a simulation study in Section 4. We compare the finite sample performance of different bootstrap methods under the general setup of mixed interval censoring model. These comparisons illustrate the superior performance of the smoothed bootstrap procedure.

Our results also shed light on the behavior of bootstrap methods in similar non-standard convergence problems, such as the monotone regression estimator (Brunk, 1970), Rousseeuw's least median of squares estimator (Rousseeuw, 1984), and the estimator of the shorth (Andrews et al., 1972; Shorack and Wellner, 1986); see also Groeneboom and Wellner (2001) for statistical problems in which the distribution  $\mathbb{C}$  arises.

## 2 Current status model

### 2.1 A sufficient condition for the consistency of the bootstrap

Under the current status model, our data are  $\mathbf{Z}_n = \{(T_i, \Delta_i)\}_{i=1}^n$ , where  $\Delta_i = \mathbf{1}_{X_i \leq T_i}$ . Each  $X_i$  can be interpreted as the unobserved time of onset of a disease and  $T_i$  is the check-up time at which the  $i$ th patient is observed. We assume that  $X_i \sim F$  and  $T_i \sim G$  are independent and that  $F$  and  $G$  are continuously differentiable at  $t_0 > 0$  ( $t_0$  being a point in the interior of the support of  $F$ ) with derivatives  $f(t_0) > 0$  and  $g(t_0) > 0$ .

We want to approximate the distribution function  $H_n$  of

$$\gamma_n = n^{1/3}\{\tilde{F}_n(t_0) - F(t_0)\}$$

by using bootstrap methods. In our model based bootstrap approach we choose an estimator, say  $F_n$ , of  $F$  (which could be NPMLE  $\tilde{F}_n$  or a smoothed version of it) and generate the bootstrapped response values as  $\Delta_i^* \sim \text{Bernoulli}(F_n(T_i))$ , fixing the values of  $T_i$ . This is the analogue of “bootstrapping residuals” in our setup. Let  $\tilde{F}_n^*$  be the NPMLE of the bootstrap sample.

In the following we establish conditions on  $F_n$  such that the bootstrap procedure is consistent, i.e.,

$$\gamma_n^* = n^{1/3}\{\tilde{F}_n^*(t_0) - F_n(t_0)\}$$

converges weakly to  $\kappa\mathbb{C}$ , as defined in (2), given the data.

We first start by formalizing the notion of *consistency* of the bootstrap. Let  $H_n^*$  be the conditional distribution function of  $\gamma_n^*$ , the bootstrap counterpart of  $\gamma_n$ , given the data. Let  $d$  denote the Levy metric or any other metric metrizing weak convergence of distribution functions. We say that  $H_n^*$  is *weakly consistent* if  $d(H_n, H_n^*) \rightarrow 0$  in probability. If the convergence holds with probability 1, then we say that the bootstrap is *strongly consistent*. If  $H_n$  has a weak limit  $H$ , then consistency requires  $H_n^*$  to converge weakly to  $H$ , in probability; and if  $H$  is continuous, consistency requires  $\sup_{x \in \mathbb{R}} |H_n^*(x) - H(x)| \rightarrow 0$  in probability as  $n \rightarrow \infty$ .

Let  $F_n$  be a sequence of distribution functions that converge weakly to  $F$  and suppose that

$$\lim_{n \rightarrow \infty} \|F_n - F\| = 0, \tag{3}$$

almost surely, where for any bounded function  $h : I \rightarrow \mathbb{R}$ ,  $\|h\| = \sup_{x \in I} |h(x)|$ . As shown in Groeneboom and Wellner (1992), the NPMLE obtained from the bootstrap

sample  $\tilde{F}_n^*$ , defined as the maximizer of (1) over all distribution functions is a step function with possible jumps only at the predictor values  $T_1, \dots, T_n$ . We have the following result on the consistency of bootstrap methods in the current status model.

**Theorem 2.1** *If (3) and*

$$\lim_{n \rightarrow \infty} n^{1/3} |F_n(t_0 + n^{-1/3}t) - F_n(t_0) - f(t_0)n^{-1/3}t| = 0 \quad (4)$$

*hold almost surely, then conditional on the data, the bootstrap estimator  $\gamma_n^*$  converges in distribution to  $\kappa\mathbb{C}$ , as defined in (2), almost surely.*

## 2.2 Inconsistency of bootstrapping from $\tilde{F}_n$

Consider the case when we bootstrap from the NPMLE  $\tilde{F}_n$ . Conditional on the predictor  $T_i$ , we generate the bootstrap response  $\Delta_i^* \sim \text{Bernoulli}(\tilde{F}_n(T_i))$ . Thus we take  $F_n = \tilde{F}_n$  and approximate the sampling distribution of  $\gamma_n$  by the conditional distribution of  $\gamma_n^* = n^{1/3}\{\tilde{F}_n^*(t_0) - \tilde{F}_n(t_0)\}$ , given the data. For this bootstrap procedure to be consistent, the conditional distribution of  $\gamma_n^*$  must converge to that of  $\kappa\mathbb{C}$ , in probability.

**Theorem 2.2** *Unconditionally  $\gamma_n^*$  does not converge in distribution to  $\kappa\mathbb{C}$ , and thus the bootstrap method is inconsistent.*

In fact it can be argued, as in Sen et al. (2010), that conditionally,  $\gamma_n^*$  does not have any weak limit in probability. The inconsistency of bootstrapping from the NPMLE results from the lack of smoothness of  $\tilde{F}_n$ . At a more technical level, the lack of the smoothness manifests itself through the failure of equation (18) in the Appendix.

We illustrate through a simulation study the inconsistency of the NPMLE bootstrap method. The upper panel of Table 1 gives the estimated coverage probabilities

of nominal 90% confidence intervals for  $F(1)$ , where the true distribution of  $F$  is assumed to be  $\text{Exp}(1)$ , or the folded normal distribution,  $|N(0, 1)|$ , and  $G$  is taken as the uniform distribution on  $[0, 2]$ . We use 500 bootstrap samples to compute each confidence interval and construct 500 such intervals. Throughout, we adopt this bootstrap setup unless otherwise specified. Table 1 shows that the coverage probabilities are much smaller than the nominal 90% value and there is no significant improvement as the sample size increases.

Table 1: *Estimated coverage probabilities of nominal 90% CIs for  $F(1)$  for two distributions:  $\text{Exp}(1)$  and  $|Z|$  with  $Z \sim N(0, 1)$ .*

		$n$	100	200	500
NPMLE	$\text{Exp}(1)$		0.73	0.72	0.74
	$ N(0, 1) $		0.69	0.70	0.73
SMLE	$\text{Exp}(1)$		0.89	0.88	0.90
	$ N(0, 1) $		0.88	0.91	0.89

Furthermore, we compare the exact and bootstrapped distributions. Due to limitations of space, we only present results for  $F$  being  $\text{Exp}(1)$ . Figure 1(a) shows the distribution of  $\gamma_n$ , obtained from 10000 random samples of sample size 500, and its bootstrap estimate (that of  $\gamma_n^*$ ) from a single sample based on 10000 bootstrap replicates. We see that the bootstrap distribution is different from that of  $\gamma_n$ .

To illustrate the behavior of the conditional distribution of  $\gamma_n^*$  we show in Figure 2(a) the estimated 0.95 quantiles of the bootstrap distributions for two independent data sequences as the sample size increases from 500 to 5000. The 0.95 quantile of the limiting distribution of  $\gamma_n$  is indicated by the solid line in each panel of Figure 2. We can see that the bootstrap 0.95 quantiles fluctuate enormously as the sample size increases from 500 to 5000 and do not converge to the 0.95 quantile of  $\kappa\mathbb{C}$ . This gives strong empirical evidence that the bootstrapped 0.95 quantiles do not converge.

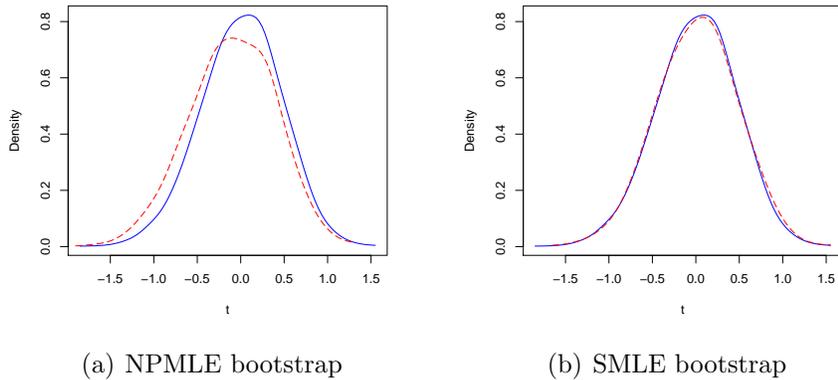


Figure 1: Estimated density functions of  $\gamma_n$  from 10000 Monte Carlo simulation (solid curve) and the bootstrap distribution of  $\gamma_n^*$  when bootstrap samples are drawn from NPMLE  $\tilde{F}_n$  (dashed, left panel) and SMLE  $\check{F}_{n,h}$  with  $h = 0.3$  (dashed, right panel).  $F$  is taken as  $\text{Exp}(1)$  and  $n = 500$ .

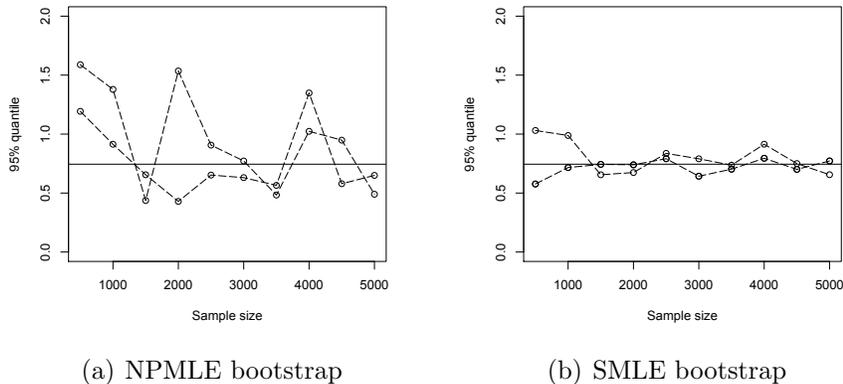


Figure 2: Estimated 0.95 quantiles of the bootstrap (dashed) and the limiting (solid) distribution.

### 2.3 Consistent bootstrap methods

We show that generating bootstrap samples from a suitably smoothed version of  $\tilde{F}_n$  leads to a consistent bootstrap procedure. We propose the following smoothed estimator  $\check{F}_n$  of  $\tilde{F}_n$ ; see Groeneboom et al. (2010). Let  $K$  be a differentiable symmetric kernel density with compact support (say  $[-1, 1]$ ) and let  $\bar{K}(t) = \int_{-\infty}^t K(s) ds$  be the corresponding distribution function. Let  $h$  be the smoothing parameter. Note that

$h$  may depend on the sample size  $n$  but, for notational convenience, in the following we write  $h$  instead of  $h_n$ . Let  $K_h(t) = K(t/h)/h$  and  $\bar{K}_h(t) = \bar{K}(t/h)$ . Then the smoothed maximum likelihood estimator (SMLE) of  $F$  is defined as

$$\check{F}_n(t) \equiv \check{F}_{n,h}(t) = \int \bar{K}_h(t-s) d\tilde{F}_n(s). \quad (5)$$

It can be easily seen that  $\check{F}_{n,h}$  is a non-decreasing function, as for  $t_2 > t_1$ ,  $\bar{K}_h(t_2-s) \geq \bar{K}_h(t_1-s)$  for all  $s$ . Throughout this paper, without further specification, we use the following kernel function to illustrate the performance of the SMLE bootstrap:

$$K(t) \propto (1-t^2)^2 \mathbf{1}_{[-1,1]}(t). \quad (6)$$

$\check{F}_{n,h}$  is a smoothed version of the step function  $\tilde{F}_n$ . As discussed in the previous section, the lack of smoothness of  $\tilde{F}_n$  leads to the inconsistency of the NPMLE bootstrap method. On the other hand, the SMLE successfully mimics the local behavior of  $F$  at  $t_0$ , and consequently gives the desired consistency as shown in Theorem 2.3 below. Recall that when bootstrapping from the SMLE  $\check{F}_{n,h}$  our bootstrap sample is  $\{(\Delta_i^*, T_i)\}_{i=1}^n$  where  $\Delta_i^* \sim \text{Bernoulli}(\check{F}_{n,h}(T_i))$ .

Following Groeneboom and Wellner (1992), we assume that the point of interest  $t_0$  is in the interior of the support of  $F$ ,  $\mathcal{S} = [0, M_0]$  with  $M_0 < \infty$ , on which  $F$  and  $G$  have bounded densities  $f$  and  $g$  staying away from zero, respectively. Furthermore, density  $g$  has a bounded derivative on  $\mathcal{S}$ .

**Theorem 2.3** *Suppose  $F$  and  $G$  satisfy the conditions listed above. Given that  $h \rightarrow 0$  and  $n^{1/3}(\log n)^{-1}h \rightarrow \infty$ , the conditional distribution of  $n^{1/3}\{\tilde{F}_n^*(t_0) - \check{F}_{n,h}(t_0)\}$ , given the data, converges to that of  $\kappa\mathbb{C}$ , in probability. Thus, bootstrapping from  $\check{F}_{n,h}$  is weakly consistent.*

We use simulation to illustrate the consistency of the SMLE bootstrap procedure. The lower panel of Table 1 gives the estimated coverage probabilities of nominal 90% confidence intervals for  $F(1)$  (when  $F$  is assumed to be  $\text{Exp}(1)$  or  $|N(0, 1)|$  and  $G$  is taken as the uniform distribution on  $[0, 2]$ ). Here we take bandwidth  $h = 0.3$ . We see that the coverage probabilities are consistent with the nominal 90% level. Figure 1(b) compares the distributions of  $\gamma_n$ , obtained from 10000 random samples of size 500, and the SMLE bootstrap estimator from a single sample, when  $F$  is  $\text{Exp}(1)$ . In addition, Figure 2(b) shows the estimated 0.95 quantiles of the bootstrap distributions for two independent data sequences. We see that for the SMLE bootstrap, the estimated 0.95 quantile is converging to the appropriate limiting value. This validates our theoretical result.

## 2.4 Choice of the tuning parameter in practice

We propose a bootstrap-based method of choosing the smoothing bandwidth  $h$ , required for computing SMLE  $\check{F}_{n,h}$ . A commonly used criterion for judging the efficacy of bandwidth selection techniques is the mean squared error (MSE) for bandwidth  $h$ ,

$$\text{MSE}(h) = E[\{\check{F}_{n,h}(t_0) - F(t_0)\}^2]. \quad (7)$$

However, the above quantity is not directly computable since  $F$  is unknown in applications. To overcome this difficulty, different procedures have been explored in the literature. Among them, bootstrap is again one of the most widely used methods and we use it to estimate the MSE in our paper. This bootstrap approach works as follows. The idea is to approximate (7) by

$$\text{BMSE}(h) = \frac{1}{B} \sum_{i=1}^B \{\check{F}_{n,h}^*(t_0) - F_n(t_0)\}^2, \quad (8)$$

where  $\check{F}_{n,h}^*(t_0)$  is constructed as in (5) (with bandwidth  $h$ ) from data  $\{(T_i, \Delta_i^*)\}_{i=1}^n$  with  $\Delta_i^* \sim \text{Bernoulli}(F_n(T_i))$ , for  $i = 1, \dots, n$ , and  $B$  is a large number. Throughout, we take  $B = 500$ . In the following we study two choices of  $F_n$ , the NPMLE and the SMLE, and show that the NPMLE does not give consistent estimates of  $\text{MSE}(h)$  while the SMLE performs well.

A natural choice of  $F_n$  is  $\check{F}_n$ , the NPMLE based on the data. However, as shown in Figure 3(a), the estimated MSE curves for different data sets are not consistent with the true curve, simulated from 500 independent samples. Here we use the kernel function in (6). For each MSE curve, we approximate it using  $\text{BMSE}(h_i)$  with  $h_i = i/20$ , for  $i = 1, \dots, 20$ .

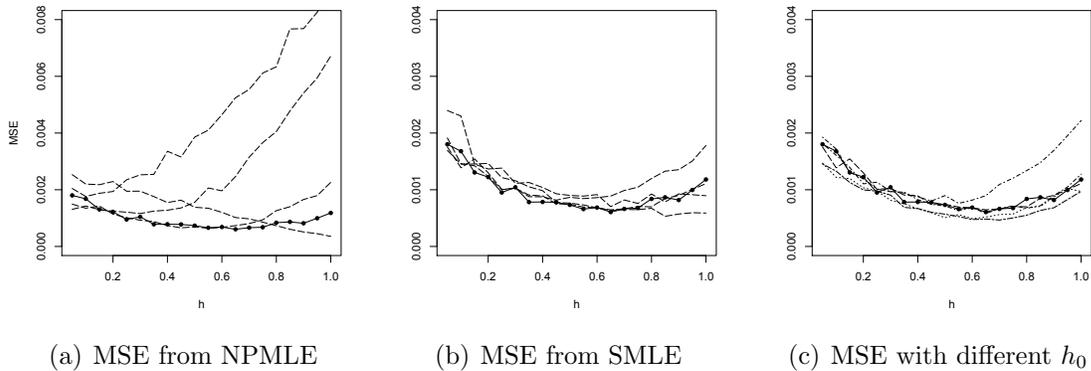


Figure 3: (a) Estimated MSE curves from the NPMLE bootstrap (dashed) and the true MSE based on 500 random samples (solid); (b) estimated MSE from the SMLE with pre-chosen bandwidth  $h_0 = 0.5$  (dashed); (c) estimated MSE from the SMLE with  $h_0 = 0.3, 0.4, 0.5, 0.6$  and  $0.7$  (dashed).  $F$  is taken as  $\text{Exp}(1)$  and  $n = 1000$ .

Another choice of  $F_n$  is  $\check{F}_{n,h_0}$ , the SMLE with a pre-chosen bandwidth  $h_0$ . This strategy is commonly used to select the bandwidth in density estimation; see, e.g., Hazelton (1996) and González-Manteiga et al. (1996). We choose  $h_0$  as the initial bandwidth and sample  $\Delta_i^* \sim \text{Bernoulli}(\check{F}_{n,h_0}(T_i))$ . Then the MSE is estimated by

$$\frac{1}{B} \sum_{i=1}^B \{\check{F}_{n,h}^*(t_0) - \check{F}_{n,h_0}(t_0)\}^2. \quad (9)$$

Under the same simulation setup as for the NPMLE, we show in Figure 3(b) that the estimated MSE curves from the SMLE are consistent with the true curve based on 500 random samples.

A related issue in practice is how to choose the optimal initial smoothing bandwidth  $h_0$ . As in density estimation, where it has been shown that different initial values of bandwidths yield consistent estimation results, we also illustrate that different  $h_0$  values give similar estimated MSE curves and therefore do not affect our final estimation much. We illustrate this through a simulation study. We choose 5 initial values of  $h_0$ ,  $h_0 = 0.3, 0.4, 0.5, 0.6$  and  $0.7$ , and show in Figure 3(c) that the estimated MSE curves with different  $h_0$  values have similar shapes and are consistent with the true MSE curve. The minimum values of the estimated MSE curves are also close to the true minimum.

### 3 Interval censoring, case 2

In case 2 censoring, an individual is checked exactly at two time points  $(T_1, T_2)$ . Suppose that we have  $n$  independent and identically distributed random vectors  $\{(X_i, T_{i,1}, T_{i,2})\}_{i=1}^n$ , where for each pair  $(X_i, T_{i,1}, T_{i,2})$ ,  $X_i \sim F$  and  $(T_{i,1}, T_{i,2})$  are independent and  $T_{i,1} < T_{i,2}$ . For the  $i$ th individual, we observe  $(T_{i,1}, T_{i,2}, \Delta_{i,1}, \Delta_{i,2})$  where  $\Delta_{i,1} = \mathbf{1}_{X_i \leq T_{i,1}}$  and  $\Delta_{i,2} = \mathbf{1}_{T_{i,1} < X_i \leq T_{i,2}}$ . Our goal is to estimate  $F$  at time  $t_0$ , i.e.,  $F(t_0)$ . The NPMLE  $\tilde{F}_n$  for  $F$  maximizes the log-likelihood function

$$\sum_{i=1}^n \{ \Delta_{i,1} \log \mathbb{F}(T_{i,1}) + \Delta_{i,2} \log(\mathbb{F}(T_{i,2}) - \mathbb{F}(T_{i,1})) + (1 - \Delta_{i,1} - \Delta_{i,2}) \log(1 - \mathbb{F}(T_{i,2})) \}$$

over all distribution functions  $\mathbb{F}$ . Deriving the limiting distribution of  $\tilde{F}_n$  in this case is quite difficult and is still an open problem. Groeneboom (1991) instead studied a one-step estimate  $F_n^{(1)}$ , obtained at the first step of the iterative convex minorant

algorithm, starting the iterations from the underlying true distribution function  $F$ , and conjectured that  $F_n^{(1)}$  is asymptotically equivalent to the NPMLE. This conjecture is called *the working hypothesis* and is still unproved. Note that we cannot, of course, compute  $F_n^{(1)}$  in practice. In the following, we assume that the working hypothesis holds and focus on bootstrapping the distribution of the one-step estimator.

Let  $H$  be the distribution function of observation times  $(T_1, T_2)$  and assume that  $F$  and  $H$  are both differentiable at  $t_0$  and  $(t_0, t_0)$ , respectively, with positive derivatives  $f(t_0)$  and  $h(t_0, t_0)$ . From Groeneboom (1991) and Groeneboom and Wellner (1992), we have that

$$(n \log n)^{1/3} \{F_n^{(1)}(t_0) - F(t_0)\} \rightarrow \kappa_1 \mathbb{C}, \quad (10)$$

in distribution, where  $\kappa_1 = \{\frac{3}{4}f(t_0)^2/h(t_0, t_0)\}^{1/3}$  and  $\mathbb{C}$  is as defined in (2).

Under the working hypothesis, the NPMLE also has the above limiting distribution in (10). Again, due to the nuisance parameters present in the limiting distribution, the above result cannot be directly applied to construct a confidence interval for  $F(t_0)$ . In the following we investigate the (in)-consistency of bootstrap methods and show that the smoothed model based bootstrap gives consistent result while bootstrapping from the NPMLE does not.

Let  $F_n$  be a sequence of distribution functions that converge weakly to  $F$ . We condition on  $(T_{i,1}, T_{i,2})$ , and generate the bootstrap response  $(\Delta_{i,1}^*, \Delta_{i,2}^*)$  by sampling  $(\Delta_{i,1}^*, \Delta_{i,2}^*, 1 - \Delta_{i,1}^* - \Delta_{i,2}^*) \sim \text{Multinomial}(1; F_n(T_{i,1}), F_n(T_{i,2}) - F_n(T_{i,1}), 1 - F_n(T_{i,2}))$ . Then based on the bootstrap sample  $\{(T_{i,1}, T_{i,2}, \Delta_{i,1}^*, \Delta_{i,2}^*)\}_{i=1}^n$ , we construct the one step NPMLE estimator  $\tilde{F}_n^{*,(1)}(t)$ , starting the iterations from  $F_n$ .

**Theorem 3.1** *For a sequence of distribution functions  $F_n$  that converge weakly to  $F$ , if the following convergence holds, almost surely, uniformly on compacts (in  $t$ )*

$$\lim_{n \rightarrow \infty} (n \log n)^{1/3} |F_n(t_0 + (n \log n)^{-1/3}t) - F_n(t_0) - f(t_0)(n \log n)^{-1/3}t| = 0, \quad (11)$$

then, the conditional distribution of  $(n \log n)^{1/3} \{F_n^{*,(1)}(t_0) - F_n(t_0)\}$ , given the data, converges to  $\kappa_1 \mathbb{C}$ , almost surely.

The above theorem gives a sufficient condition for the bootstrap procedure to be consistent. In particular, for the SMLE bootstrap, we have  $F_n = \check{F}_{n,h}$  and condition (11) holds. Let  $\tilde{F}_{n,h}^{*,(1)}(t_0)$  be the corresponding one-step bootstrap estimator starting from  $\check{F}_{n,h}$  and let  $\tilde{F}_n^*(t_0)$  be the NPMLE of the bootstrap sample. A similar argument as in the proof of Theorem 2.3 gives that with properly chosen bandwidth  $h$ , conditionally,  $(n \log n)^{1/3} \{\tilde{F}_{n,h}^{*,(1)}(t_0) - \check{F}_{n,h}(t_0)\}$  converges in distribution to  $\kappa_1 \mathbb{C}$ , in probability. Then under the working hypothesis, conditionally,  $(n \log n)^{1/3} \{\tilde{F}_n^*(t_0) - \check{F}_{n,h}(t_0)\}$  converges in distribution to  $\kappa_1 \mathbb{C}$ , in probability, and bootstrapping from  $\check{F}_{n,h}$  is weakly consistent.

Table 2: *Estimated coverage probabilities of nominal 90% CIs for case 2 censoring.*

$n$	100	200	500
NPMLE	0.74	0.76	0.75
SMLE	0.88	0.89	0.91

On the other hand, the NPMLE is a step function and does not satisfy (11), and therefore the above theorem is not applicable. In fact for bootstrapping from the NPMLE, the three convergence results in Lemma 6.2 in the Appendix may not hold. These theoretical arguments are supported by numerical results. Table 2 shows that the NPMLE method has a low coverage rate for nominal 90% confidence intervals while the smoothed bootstrap gives consistent results. Here  $F$  is taken as  $\text{Exp}(1)$  and  $T_1, T_2$  are the order statistics of two uniformly distributed random variables on  $[0, 2]$ . For the SMLE bootstrap, the smoothing bandwidth  $h$  is chosen as  $n^{-1/5}$ .

## 4 Mixed case interval censoring

### 4.1 NPMLE and SMLE bootstrap

Under mixed case interval censoring, for an individual, we have  $K$  observation times  $T_1 \leq \dots \leq T_K$ , where  $K$  is an integer-valued random variable. Let  $\Delta_k = \mathbf{1}_{T_{k-1} < X \leq T_k}$ , for  $k = 1, \dots, K$ , with  $T_0 = 0$ . We observe  $n$  independent and identically distributed copies of  $(T_k, \Delta_k : k = 1, \dots, K)$ , i.e.,  $\{(T_{i,k}, \Delta_{i,k} : k = 1, \dots, K_i)\}_{i=1}^n$ . Again we are interested in estimating  $F$  at  $t_0$ . The NPMLE  $\tilde{F}_n$  for  $F$  maximizes the log-likelihood function

$$\sum_{i=1}^n \left[ \sum_{k=1}^{K_i} \Delta_{i,k} \log(\mathbb{F}(T_{i,k}) - \mathbb{F}(T_{i,k-1})) + \left(1 - \sum_{k=1}^{K_i} \Delta_{i,k}\right) \log(1 - \mathbb{F}(T_{i,K_i})) \right] \quad (12)$$

over all distribution functions  $\mathbb{F}$ . The limiting distribution of the NPMLE is unknown, and this complicates the problem of constructing confidence intervals for  $F(t_0)$ .

In the following, we focus on the NPMLE  $\tilde{F}_n$  and compare empirically the performance of different bootstrap methods in estimating the distribution of  $n^{-1/3}(\tilde{F}_n(t_0) - F(t_0))$ . We illustrate the inconsistency of bootstrapping from the NPMLE and the consistency of bootstrapping from a suitably smoothed NPMLE. As for each subject, only one  $\Delta_{i,k}$ ,  $k = 1, \dots, K_i$ , is 1, the computation of the NPMLE of mixed case interval censoring can be reduced to the case 2 interval censoring, as noted in Huang and Wellner (1997) and Song (2004). This can be done efficiently by the iterative convex minorant algorithm; see Groeneboom and Wellner (1992), Wellner and Zhan (1997), and Jongbloed (1998). In this paper we use R package ‘‘Icens’’ to estimate the NPMLE.

Table 3 shows the estimated coverage probabilities of nominal 90% confidence intervals when bootstrapping from the NPMLE and the SMLE. We can see that the

Table 3: *Estimated coverage probabilities of nominal 90% CIs for mixed case interval censoring.*

	$n$	100	200	500
NPMLE	coverage	0.76	0.77	0.76
	length	0.27	0.20	0.14
SMLE	coverage	0.88	0.87	0.89
	length	0.25	0.20	0.14

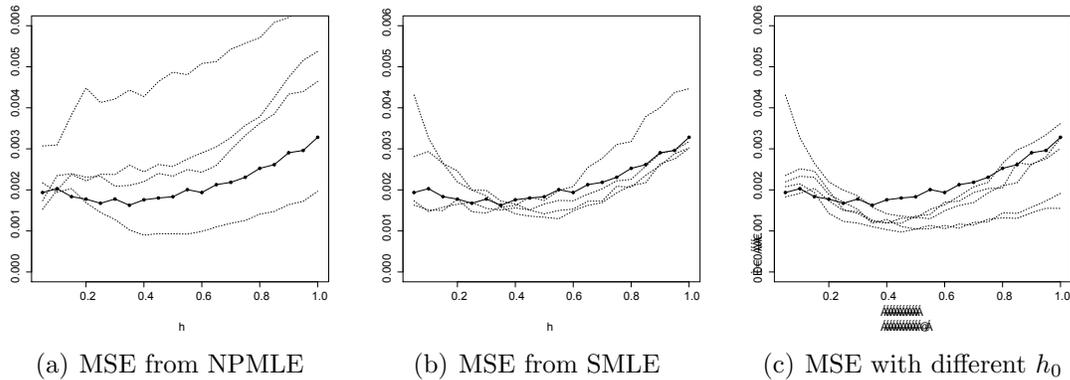


Figure 4: (a) The true MSE based on 500 random samples (solid) and estimated MSE from the NPMLE bootstrap (dashed); (b) MSE from the SMLE with  $h_0 = 0.5$  (dashed); (c) MSE from the SMLE with  $h_0 = 0.3, 0.4, 0.5, 0.6$  and  $0.7$  (dashed).

performance of the SMLE is much better than that of the NPMLE in both coverage and interval lengths. Here  $F$  is taken as  $\text{Exp}(1)$  and  $K$  is chosen from the uniform distribution on set  $\{1, 2, 3\}$ . Given  $K$ , the observation times  $(T_1, \dots, T_K)$  are chosen as the  $K$  order statistics from the uniform distribution on  $[0, 2]$ . The smoothing bandwidth for the SMLE is chosen as  $n^{-1/5}$ .

We minimize the MSE criterion in (7) to estimate the optimal bandwidth  $h$  in the smoothed bootstrap method. Two choices of  $F_n$ , the NPMLE and a SMLE with an initial bandwidth  $h_0$ , are studied. The corresponding estimated MSE curves are displayed in Figure 4. It can be clearly seen that curves based on the NPMLE are not consistent while those from the SMLE consistently estimate the true MSE curve. The effect of choosing different initial values of  $h_0$  to obtain the optimal smoothing

bandwidth is shown in Figure 4(c). As in the current status model, we can see that the estimated curves are robust to the initial choice of  $h_0$ .

## 4.2 Comparison with the existing methods

To further illustrate the superiority of the proposed smoothed bootstrap method, we compare its finite sample performance with the pseudolikelihood method of Sen and Banerjee (2007) and the  $m$ -out-of- $n$  bootstrap method of Lee and Pun (2006).

We present simulations from a mixed case censoring model under the same setup as in Section 3.1 of Sen and Banerjee (2007). We take  $F$  to be  $\text{Exp}(1)$  and  $K$  is chosen uniformly from  $\{1, 2, 3, 4\}$ . Given  $K$ ,  $(T_1, \dots, T_K)$  are the  $K$  order statistics from the uniform distribution on  $[0, 3]$ . We generate 1000 samples for each sample size shown in Table 4 and the corresponding 95% confidence intervals for  $F(\log 2) = 0.5$  are constructed. For the SMLE bootstrap, we choose bandwidths  $h = 0.5$  for  $n = 50, 100, 200, 500$  and  $h = 0.3$  for  $n = 1000, 1500, 2000$ . Table 4 shows that the SMLE bootstrap gives more consistent results than the pseudolikelihood and  $m$ -out-of- $n$  methods in general. The pseudolikelihood method is more anti-conservative while the  $m$ -out-of- $n$  method is in general conservative and has wider intervals.

Table 4: *Estimated coverage probabilities of nominal 95% CIs for mixed case interval censoring. The results for the pseudolikelihood (PL) and  $m$ -out-of- $n$  methods are taken from Table 1 in Sen and Banerjee (2007).*

	$n$	50	100	200	500	1000	1500	2000
SMLE	coverage	0.92	0.94	0.95	0.95	0.94	0.95	0.95
	length	0.51	0.38	0.30	0.20	0.15	0.13	0.12
PL	coverage	0.90	0.92	0.92	0.95	0.94	0.94	0.94
	length	0.41	0.33	0.26	0.20	0.16	0.14	0.12
$m$ -out-of- $n$	coverage	0.97	0.97	0.96	0.96	0.95	0.96	0.97
	length	0.54	0.47	0.31	0.24	0.17	0.16	0.14

## 5 Real data analysis

Finkelstein and Wolfe (1985) considered an interval censored data set of a study of early breast cancer patients. Between 1976 and 1980, 94 patients had been treated at the Joint Center for Radiation Therapy in Boston. They were assigned into two groups: one group treated with primary radiation therapy and adjuvant chemotherapy (48 patients) and the other group with radiotherapy alone (46 patients). Times of cosmetic deterioration, determined by the appearance of breast retraction, were compared between the two treatment groups to determine whether chemotherapy has an impact on the rate of deterioration of the cosmetic state. The patients were checked at clinic every 4 to 6 month. For each patient, the only information available is a time interval when the retraction was present. See Section 5.3 in Finkelstein and Wolfe (1985) and their Table 4 for more details, where all patients' time intervals are provided.

We are interested in estimating the distribution of the retraction time ( $F$ ). We model the data set as case 2 censoring and estimate the distribution functions separately for the two treatment groups, i.e., radiotherapy group ( $T = 0$ ) and radiotherapy and chemotherapy group ( $T = 1$ ). Figure 5(a) shows the NPMLE of  $F$  computed for the two treatment groups. The distribution function of the group with  $T = 1$  dominates that of  $T = 0$  in general, which indicates that patients receiving radiotherapy and chemotherapy have an earlier deterioration time, as measured by the appearance of breast retraction. From Figure 5(b) we see that the MSE curves from different  $h_0$  are consistent with each other, and their values decrease as  $h$  increases from 1 to about 10 and then stay quite close. Based on these observations, we choose our bandwidth  $h = 10$ .

Table 5 presents 90% and 95% confidence intervals of  $F$  at  $t_0 = 20$  and 30 obtained using the NPMLE and SMLE bootstrap methods. The left extremities of the

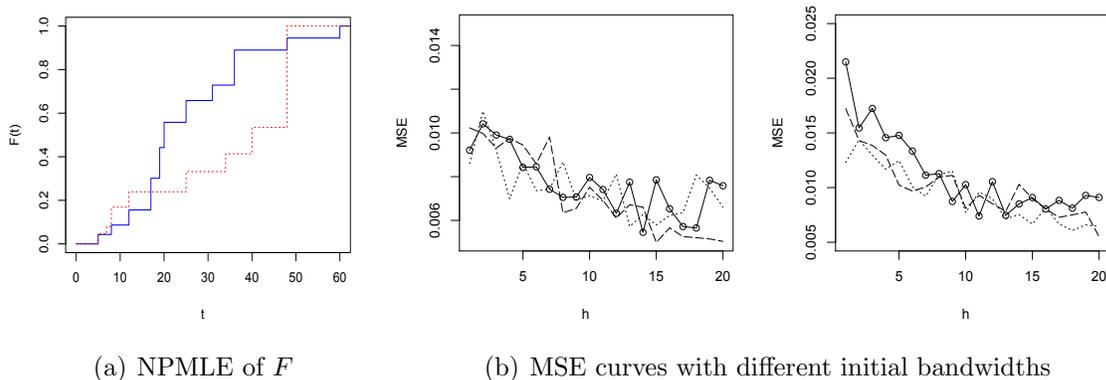


Figure 5: (a) NPMLE of  $F$  for groups  $T = 0$  (dashed) and  $T = 1$  (solid); (b) estimated MSE curves at  $t = 30$  from the SMLE bootstrap with initial bandwidths  $h_0 = 5, 10, 15$ . The left plot is for  $T = 0$  and the right for  $T = 1$ .

Table 5: *CIs of the distribution of retraction time in groups  $T = 0$  and  $T = 1$  at  $t_0 = 20$  and  $30$ .*

		$\tilde{F}_n(20)$	90% CI	95% CI	$\tilde{F}_n(30)$	90% CI	95% CI
$T = 1$	SMLE	0.56	[0.35, 0.79]	[0.31, 0.82]	0.66	[0.48, 0.91]	[0.45, 0.94]
	NPMLE	0.56	[0.41, 0.88]	[0.39, 0.92]	0.66	[0.44, 0.83]	[0.41, 0.86]
$T = 0$	SMLE	0.24	[0.11, 0.38]	[0.08, 0.39]	0.33	[0.17, 0.50]	[0.14, 0.53]
	NPMLE	0.24	[0.12, 0.35]	[0.09, 0.38]	0.33	[0.16, 0.49]	[0.13, 0.51]

confidence intervals for the distribution function in group  $T = 1$  are shifted to the right of those for the corresponding time points in group  $T = 0$ , which indicates the treatment effect. We also see that the confidence intervals from SMLE and NPMLE bootstrap methods are quite different, which may be due to the inconsistency of the NPMLE method.

## References

- Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica*, 73:1175–1204.
- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey,

- J. W. (1972). *Robust Estimates of Location*. Princeton Univ. Press.
- Banerjee, M. and Wellner, J. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.*, 29:1699–1731.
- Bickel, P. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9:1196–1217.
- Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons Inc., New York.
- Billingsley, P. (1995). *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York.
- Brunk, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969)*, pages 177–197. Cambridge Univ. Press, London.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.
- Finkelstein, D. M. and Wolfe, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics*, 41(4):933–945.
- González-Manteiga, W., Cao, R., and Marron, J. S. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *J. Amer. Statist. Assoc.*, 91(435):1130–1140.
- Groeneboom, P. (1989). Brownian motion with a parabolic drift and Airy functions. *Probab. Theory Related Fields*, 81(1):79–109.

- Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. *Technical report, available at <http://statistics.stanford.edu/~ckirby/techreports/NSF/EFS%20NSF%20378.pdf>*.
- Groeneboom, P., Jongbloed, G., and Witte, B. I. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.*, 38(1):352–387.
- Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel.
- Groeneboom, P. and Wellner, J. A. (2001). Computing chernoff’s distribution. *J. Comp. Graph. Statist.*, 10:388–400.
- Hazelton, M. (1996). Bandwidth selection for local density estimators. *Scand. J. Statist.*, 23(2):221–232.
- Huang, J. and Wellner, J. A. (1997). Interval censored survival data: a review of recent progress. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 123–169. Springer.
- Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Statist.*, 7(3):310–321.
- Lee, S. M. S. and Pun, M. C. (2006). On  $m$  out of  $n$  bootstrapping for nonstandard  $m$ -estimation with nuisance parameters. *J. Amer. Statis. Assoc.*, 101:1185–1197.
- Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling*. Springer-Verlag, New York.

- Pollard, D. (1984). *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York.
- Rousseeuw, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.*, 79:871–880.
- Schick, A. and Yu, Q. (2000). Consistency of the GMLE with mixed case interval-censored data. *Scand. J. Statist.*, 27(1):45–55.
- Sen, B. and Banerjee, M. (2007). A pseudolikelihood method for analyzing interval censored data. *Biometrika*, 94(1):71–86.
- Sen, B., Banerjee, M., and Woodroffe, M. (2010). Inconsistency of bootstrap: The Grenander estimator. *Ann. Statist.*, 38(4):1953–1977.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.
- Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- Singh, K. (1981). On the asymptotic accuracy of Efron’s bootstrap. *Ann. Statist.*, 9(6):1187–1195.
- Song, S. (2004). Estimation with univariate “mixed case” interval censored data. *Statist. Sinica*, 14(1):269–282.
- van der Vaart, A. and Wellner, J. A. (2000a). Preservation theorems for Glivenko-Cantelli and uniform Glivenko-Cantelli classes. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pages 115–133. Birkhäuser Boston, Boston, MA.
- van der Vaart, A. W. and Wellner, J. A. (2000b). *Weak Convergence and Empirical Processes With Applications to Statistics*. Springer, New York.

Wellner, J. A. (1995). Interval censoring, case 2: alternative hypotheses. In *Analysis of censored data (Pune, 1994/1995)*, volume 27 of *IMS Lecture Notes Monogr. Ser.*, pages 271–291. Inst. Math. Statist., Hayward, CA.

Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Amer. Statist. Assoc.*, 92(439):945–959.

## 6 Appendix: Proof of Theorems

In this paper stochastic processes are regarded as random elements in  $D(\mathbb{R})$ , the space of right continuous functions on  $\mathbb{R}$  with left limits, equipped with the projection  $\sigma$ -field and the topology of uniform convergence on compacta; see Pollard (1984), Chapters IV and V for background.

For random elements  $(V_n)_{n=1}^\infty$  and  $V$  taking values in a metric space  $(\mathfrak{X}, \rho)$  we say that  $V_n$  converges conditionally (given the data) in probability to  $V$ , almost surely (in probability), if for any given  $\epsilon > 0$ ,  $P(\rho(V_n, V) > \epsilon \mid \mathbf{Z}_n) \rightarrow 0$  almost surely (in probability), where  $\mathbf{Z}_n$  denotes our observed data.

### 6.1 Proof of Theorem 2.1

We denote our bootstrap sample by  $(T_1, \Delta_1^*), \dots, (T_n, \Delta_n^*)$ . Let  $\mathbb{P}_n^*$  denote the induced measure of the bootstrap sample and write

$$\mathbb{P}_n^* f(\Delta, T) = \frac{1}{n} \sum_{i=1}^n f(\Delta_i^*, T_i).$$

Letting  $A = \{(x, t) : x \leq t\}$ , we define the following stochastic processes:

$$V_n^*(t) = \mathbb{P}_n^* \mathbf{1}_A \mathbf{1}_{\mathbb{R} \times [0, t]} = \frac{1}{n} \sum_{i=1}^n \Delta_i^* \mathbf{1}_{T_i \leq t}, \quad G_n^*(t) = \mathbb{P}_n^* \mathbf{1}_{\mathbb{R} \times [0, t]} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{T_i \leq t}.$$

Let  $\mathbb{P}_{T,n}$  be the empirical probability measure of  $\{T_i\}_{i=1}^n$ . Let  $P_n$  be the probability measure induced by  $F_n$  and  $\mathbb{P}_{T,n}$ . Note that under the conditional bootstrap procedure,  $\mathbb{P}_n^* f(T) = P_n f(T) = \mathbb{P}_{T,n} f(T) = \sum_{i=1}^n f(T_i)/n$ . We use  $E_n$  to denote the expectation with respect to  $P_n$ .

Appealing to the characterization of  $\tilde{F}_n^*$  (see pp. 298-299 of van der Vaart and

Wellner, 2000b), we know that

$$\tilde{F}_n^*(t) \leq a \text{ iff } \arg \min_s \{V_n^*(s) - aG_n^*(s)\} \geq T_{(t)} \quad (13)$$

where  $T_{(t)}$  is the largest observation time that does not exceed  $t$ . By (13), the event that  $n^{1/3}\{\tilde{F}_n^*(t_0) - F_n(t_0)\} \leq x$  is equivalent to

$$\arg \min_s \{V_n^*(s) - [xn^{-1/3} + F_n(t_0)]G_n^*(s)\} \geq T_{(t_0)}.$$

This is the same as

$$n^{1/3} \left[ \arg \min_s \{V_n^*(s) - [xn^{-1/3} + F_n(t_0)]G_n^*(s)\} - t_0 \right] \geq n^{1/3}(T_{(t_0)} - t_0).$$

Changing  $s \mapsto t_0 + tn^{-1/3}$  and using the fact that  $n^{1/3}(T_{(t_0)} - t_0) = o(1)$ , the above inequality can be re-expressed as

$$\arg \min_t [V_n^*(t_0 + tn^{-1/3}) - \{xn^{-1/3} + F_n(t_0)\}G_n^*(t_0 + tn^{-1/3})] \geq o(1).$$

The left hand side of the above inequality can be written as

$$\begin{aligned} & \arg \min_t [\mathbb{P}_n^* \mathbf{1}_A \mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - F_n(t_0)G_n^*(t_0 + tn^{-1/3}) - xn^{-1/3}G_n^*(t_0 + tn^{-1/3})] \\ = & \arg \min_t \left[ n^{2/3} \mathbb{P}_n^* \{ \mathbf{1}_A - F_n(t_0) \} (\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right. \\ & \left. - xn^{1/3} [G_n^*(t_0 + tn^{-1/3}) - G_n^*(t_0)] \right]. \\ = & \arg \min_t \left[ n^{2/3} \mathbb{P}_n^* (\mathbf{1}_A - F_n(T)) (\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right. \\ & \left. + n^{2/3} \mathbb{P}_n^* (F_n(T) - F_n(t_0)) (\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right. \\ & \left. - xn^{1/3} [G_n^*(t_0 + tn^{-1/3}) - G_n^*(t_0)] \right]. \end{aligned} \quad (14)$$

To study the distribution of  $\gamma_n^*$ , we start with the distributions of the three terms in (14). This is given in the following Lemma.

**Lemma 6.1** *We have the following convergence results:*

(i) *We have that*

$$xn^{1/3}\{G_n^*(t_0 + tn^{-1/3}) - G_n^*(t_0)\} \rightarrow xg(t_0)t, \quad (15)$$

*uniformly on compacta, almost surely.*

(ii) *Let  $\mathbb{Z}$  be a standard two-sided Brownian motion on  $\mathbb{R}$  such that  $\mathbb{Z}(0) = 0$ . If*

*(3) holds, then, conditional on the data, the process*

$$n^{2/3}\mathbb{P}_n^* \left\{ (\mathbf{1}_A - F_n(T))(\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right\} \xrightarrow{d} \sqrt{F(t_0)[1 - F(t_0)]g(t_0)}\mathbb{Z}(t) \quad (16)$$

*almost surely in the space  $D(\mathbb{R})$ .*

(iii) *If we have the following convergence uniformly on compacta in  $t$*

$$\lim_{n \rightarrow \infty} n^{1/3}|F_n(t_0 + n^{-1/3}t) - F_n(t_0) - f(t_0)n^{-1/3}t| = 0, \quad (17)$$

*then conditionally*

$$n^{2/3}\mathbb{P}_n^* \left\{ (F_n(T) - F_n(t_0))(\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right\} \rightarrow \frac{1}{2}f(t_0)g(t_0)t^2 \quad (18)$$

*uniformly on compacta, almost surely.*

**Proof of Lemma 6.1.** (i) To show the first convergence result, observe that

$$\begin{aligned} & xn^{1/3}\{G_n^*(t_0 + tn^{-1/3}) - G_n^*(t_0)\} \\ = & xn^{1/3}(\mathbb{P}_{T,n} - P)(\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) + xn^{1/3}P(\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}). \end{aligned}$$

By the law of iterated logarithm, this equals

$$o(1) + xn^{1/3}\{G(t_0 + tn^{-1/3}) - G(t_0)\} \rightarrow xg(t_0)t, \quad a.s.,$$

uniformly on compacta.

(ii) To show (16), let  $Z_{n,i}(t) = n^{-1/3}(\Delta_i^* - F_n(T_i))W_{n,i}(t)$  where  $W_{n,i}(t) = \mathbf{1}_{T_i \leq t_0 + tn^{-1/3}} - \mathbf{1}_{T_i \leq t_0}$ . The left-hand side of (16) then can be expressed as  $\sum_{i=1}^n Z_{n,i}(t)$ . Note that  $Z_{n,i}(t)$  has mean 0 and variance  $\sigma_{n,i}^2(t) = n^{-2/3}W_{n,i}^2(t)F_n(T_i)[1 - F_n(T_i)]$ .

Therefore, for  $h > 0$ ,  $s_n^2(t) := \sum_{i=1}^n \sigma_{n,i}^2(t)$  can be simplified as

$$n^{1/3}\mathbb{P}_{T,n}[F_n(T)(1 - F_n(T))(\mathbf{1}_{T \leq t_0 + tn^{-1/3}} - \mathbf{1}_{T \leq t_0})^2].$$

The preceding display is equal to

$$\begin{aligned} & o(1) + n^{1/3}P[F_n(T)(1 - F_n(T))(\mathbf{1}_{T \leq t_0 + tn^{-1/3}} - \mathbf{1}_{T \leq t_0})^2] \\ = & o(1) + n^{1/3} \int_{t_0}^{t_0 + tn^{-1/3}} F_n(u)(1 - F_n(u))g(u)du \\ = & o(1) + \int_0^t F_n(t_0 + sn^{-1/3})[1 - F_n(t_0 + sn^{-1/3})]g(t_0 + sn^{-1/3})ds \\ \xrightarrow{a.s.} & F(t_0)[1 - F(t_0)]g(t_0)t. \end{aligned}$$

By the Lindeberg-Feller CLT (see pp. 359 of Billingsley, 1995) we have

$$\sum_{i=1}^n Z_{n,i}(t) \xrightarrow{d} N(0, F(t_0)[1 - F(t_0)]g(t_0)t)$$

for every  $t \in \mathbb{R}$ . Similarly we have the convergence of the finite dimensional joint distribution. We only need to show the tightness of  $\sum_{i=1}^n Z_{n,i}(t)$ . By Theorem 15.6 in Billingsley (1968), it is sufficient to show that there exists a nondecreasing, continuous function  $H$  such that for any  $t_1 < t < t_2$ ,  $\gamma > 0$  and  $\alpha > 1/2$

$$E_n \left[ \left| \sum_{i=1}^n Z_{n,i}(t_1) - \sum_{i=1}^n Z_{n,i}(t) \right|^\gamma \left| \sum_{i=1}^n Z_{n,i}(t_2) - \sum_{i=1}^n Z_{n,i}(t) \right|^\gamma \right] \leq (H(t_2) - H(t_1))^{2\alpha}. \quad (19)$$

Take  $\gamma = 2$  and  $\alpha = 1$ . Note that the following inequality holds almost surely

$$\begin{aligned} & E_n \left[ \left| \sum_i Z_{n,i}(t_1) - \sum_i Z_{n,i}(t) \right|^2 \left| \sum_j Z_{n,j}(t_2) - \sum_j Z_{n,j}(t) \right|^2 \right] \\ = & n^{-4/3} E_n \left[ \sum_i (\Delta_i^* - F_n(T_i))^2 \mathbf{1}_{t_0 + t_1 n^{-1/3} < T_i \leq t_0 + t n^{-1/3}} \right. \\ & \left. \times \sum_j (\Delta_j^* - F_n(T_j))^2 \mathbf{1}_{t_0 + t n^{-1/3} < T_j \leq t_0 + t_2 n^{-1/3}} \right] \\ \leq & n^{-4/3} \sum_i \mathbf{1}_{t_0 + t_1 n^{-1/3} < T_i \leq t_0 + t n^{-1/3}} \sum_j \mathbf{1}_{t_0 + t n^{-1/3} < T_j \leq t_0 + t_2 n^{-1/3}} \\ \rightarrow & g^2(t_0)(t_2 - t_1)^2. \end{aligned}$$

Therefore we have the desired weak convergence result.

(iii) The left-hand side of (18) can be decomposed as

$$\begin{aligned} & n^{2/3} (\mathbb{P}_n^* - P) \left[ (F_n(T) - F_n(t)) (\mathbf{1}_{\mathbb{R} \times [0, t_0 + t n^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right] \\ + & n^{2/3} P \left[ (F_n(T) - F_n(t)) (\mathbf{1}_{\mathbb{R} \times [0, t_0 + t n^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right]. \end{aligned} \quad (20)$$

The first term in (20) can be shown to converge to 0 uniformly for  $t \in [-K, K]$ , in probability, for any  $K > 0$ , as follows:

$$\begin{aligned}
& n^{2/3} \left| (\mathbb{P}_n^* - P) \left[ (F_n(T) - F_n(t_0)) (\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right] \right| \\
&= n^{2/3} \left| \int_{t_0}^{t_0 + tn^{-1/3}} (F_n(u) - F_n(t_0)) d(\mathbb{P}_{T,n} - P)(u) \right| \\
&= n^{2/3} \left| \left[ (\mathbb{P}_{T,n} - P)(u) (F_n(u) - F_n(t_0)) \right]_{t_0}^{t_0 + tn^{-1/3}} - \int_{t_0}^{t_0 + tn^{-1/3}} (\mathbb{P}_{T,n} - P)(u) dF_n(u) \right| \\
&\leq (\sqrt{n} \|\mathbb{P}_{T,n} - P\|) 2n^{1/6} [F_n(t_0 + Kn^{-1/3}) - F_n(t_0 - Kn^{-1/3})] = o(1), \quad a.s.
\end{aligned}$$

Therefore first term in (20) is ignorable given that (17) holds. The second term in (20) can be simplified as:

$$\begin{aligned}
& n^{2/3} P \left[ (F_n(T) - F_n(t_0)) (\mathbf{1}_{\mathbb{R} \times [0, t_0 + tn^{-1/3}]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \right] \\
&= n^{1/3} \int_0^t [F_n(t_0 + sn^{-1/3}) - F(t_0)] g(t_0 + sn^{-1/3}) ds \\
&= (1 + o(1)) n^{1/3} \int_0^t sn^{-1/3} f(t_0) g(t_0 + sn^{-1/3}) ds \\
&\rightarrow f(t_0) g(t_0) \frac{1}{2} t^2,
\end{aligned}$$

where the last step follows from the assumption that  $g$  is continuous at  $t_0$ . This gives the desired conclusion. ■

We proceed to prove Theorem 2.1.

**Proof of Theorem 2.1.** This follows from a similar argument as in Section 3.2.15 in van der Vaart and Wellner (2000b). We only need to show the uniform tightness of the minimum, i.e., for any  $\epsilon$  and  $B_0 > 0$ , there exists a constant  $B$  such that

$$P_n \left( \max_{x \in [-B_0, B_0]} n^{1/3} \left| \arg \min_s \{ V_n^*(s) - [xn^{-1/3} + F_n(t_0)] G_n^*(s) \} - t_0 \right| > B \right) < \epsilon, \quad a.s.$$

Here recall that  $P_n$  is the probability measure induced by  $F_n$  and the empirical probability measure of  $\{T_i\}_{i=1}^n$ . The above tightness result follows from Theorem 3.4.1 in van der Vaart and Wellner (2000b). In particular, following their notation, we take

$$\begin{aligned} \mathbb{M}_n^*(h) &= \mathbb{P}_n^*(\mathbf{1}_A - F_n(T))(\mathbf{1}_{\mathbb{R} \times [0, t_0+h]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \\ &\quad + \mathbb{P}_n^*(F_n(T) - F_n(t_0))(\mathbf{1}_{\mathbb{R} \times [0, t_0+h]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}) \\ &\quad - xn^{-1/3}[G_n^*(t_0+h) - G_n^*(t_0)], \end{aligned}$$

and

$$\mathbb{M}_n(h) = P_n(\mathbf{1}_A - F_n(T))(\mathbf{1}_{\mathbb{R} \times [0, t_0+h]} - \mathbf{1}_{\mathbb{R} \times [0, t_0]}).$$

Then the conditions of their Theorem 3.4.1 are satisfied with  $\phi_n(\delta) = \sqrt{\delta} + x\delta n^{1/6}$ . Together with the fact that  $\mathbb{M}_n(h)$  converges to 0 in probability, which follows from the results in Lemma 6.1, the tightness result holds and we have the weak convergence of  $\gamma_n^*$ .

The above tightness result and Lemma 6.1 imply that conditional on the data  $n^{1/3} \arg \min_s \{V_n^*(s) - [xn^{-1/3} + F_n(t_0)]G_n^*(s)\} - t_0$  converges weakly to the process

$$T(x) := \operatorname{argmin}_t \left\{ \sqrt{F(t_0)[1 - F(t_0)]g(t_0)}Z(t) + \frac{1}{2}f(t_0)g(t_0)t^2 - xg(t_0)t \right\}$$

almost surely. Therefore conditionally we have the following convergence

$$\begin{aligned} P_n(n^{1/3}(\tilde{F}_n^*(t_0) - F_n(t_0)) \leq x) &\xrightarrow{a.s.} P(T(x) \geq 0) \\ &= P(T(x) - f(t_0)^{-1}x \geq -f(t_0)^{-1}x). \end{aligned}$$

By the stationarity of process  $T(x) - f(t_0)^{-1}x$  as shown in Groeneboom (1989), we have that  $T(x) - f(t_0)^{-1}x$  and  $T(0)$  have the same distribution function. Therefore,

$P_n(n^{1/3}(\tilde{F}_n^*(t_0) - F_n(t_0)) \leq x)$  converges almost surely to

$$\begin{aligned}
& P(T(0) \geq -f(t_0)^{-1}x) \\
&= P\left(\operatorname{argmin}_t \left\{ \sqrt{F(t_0)[1 - F(t_0)]g(t_0)}\mathbb{Z}(t) + \frac{1}{2}f(t_0)g(t_0)t^2 \right\} \geq -f(t_0)^{-1}x\right) \\
&= P\left(\operatorname{argmin}_t \left\{ \mathbb{Z}\left(\frac{f(t_0)^{2/3}g(t_0)^{1/3}}{[4F(t_0)(1 - F(t_0))]^{1/3}}t\right) + \left(\frac{f(t_0)^{2/3}g(t_0)^{1/3}}{[4F(t_0)(1 - F(t_0))]^{1/3}}t\right)^2 \right\} \geq -\kappa^{-1}x\right).
\end{aligned}$$

Here recall that  $\kappa = \{4F(t_0)(1 - F(t_0))f(t_0)/g(t_0)\}^{1/3}$  and the last equation is due to the Brownian scaling. Then, the above display is equal to

$$P(\operatorname{argmin}_t \{\mathbb{Z}(t) + t^2\} \geq -\kappa^{-1}x) = P(\operatorname{argmin}_t \{\mathbb{Z}(t) + t^2\} \leq \kappa^{-1}x),$$

which gives the desired conclusion. ■

## 6.2 Proofs of Theorems 2.2 and 2.3

**Proof of Theorem 2.2.** Proof of Theorem 2.2 follows a similar argument as in the proof of Theorem 3.1 in Sen et al. (2010) and we only show the key steps. The bootstrap NPMLE  $\tilde{F}_n^*$  is the left derivative of the greatest convex minorant of the cumulative sum diagram consisting of points  $(G_n^*(t), V_n^*(t))$ . Let  $\mathbb{F}_n^*$  be the corresponding cumulative sum diagram function, i.e., for  $u \in [G_n^*(T_{(i)}), G_n^*(T_{(i+1)})]$  function  $\mathbb{F}_n^*(u) = V_n^*(T_{(i)})$ , where  $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$  are the order statistics of  $T_1, \dots, T_n$ . Then  $\gamma_n^*$  equals the left derivative at  $t = 0$  of the greatest convex minorant of process

$$\mathbb{Z}_n^*(t) := n^{2/3} \{ \mathbb{F}_n^*(G_n^*(t_0) + n^{-1/3}t) - \mathbb{F}_n^*(G_n^*(t_0)) - \tilde{F}_n^*(t_0)n^{-1/3}t \}.$$

We further write  $\mathbb{Z}_n^*(t)$  as  $\mathbb{Z}_{n,1}^*(t) + \mathbb{Z}_{n,2}^*(t)$ , where

$$\begin{aligned}\mathbb{Z}_{n,1}^*(t) &:= n^{2/3}\{(\mathbb{F}_n^* - \tilde{\mathbb{F}}_n)(G_n^*(t_0) + n^{-1/3}t) - (\mathbb{F}_n^* - \tilde{\mathbb{F}}_n)(G_n^*(t_0))\}, \\ \mathbb{Z}_{n,2}^*(t) &:= n^{2/3}\{\tilde{\mathbb{F}}_n(G_n^*(t_0) + n^{-1/3}t) - \tilde{\mathbb{F}}_n(G_n^*(t_0)) - \tilde{F}_n(t_0)n^{-1/3}t\}.\end{aligned}$$

Here  $\tilde{\mathbb{F}}_n$  is the greatest convex minorant of the cumulative sum diagram function based on the observed data  $(T_i, \Delta_i), i = 1, \dots, n$ . These  $\mathbb{Z}^*$  processes take analogous forms as  $\mathbb{Z}_n^*(h), \mathbb{Z}_{n,1}^*(h), \mathbb{Z}_{n,2}^*(h)$  in Section 3.2 in Sen et al. (2010). For  $f$ , let  $L_{\mathbb{R}}f$  be its greatest convex minorant on  $\mathbb{R}$ . Following the proofs of Theorem 3.1 in Sen et al. (2010) and Lemma 6.1, we have unconditionally

$$\begin{aligned}\mathbb{Z}_{n,1}^*(t) &\xrightarrow{d} \mathbb{U}_1(t) := \sqrt{F(t_0)[1 - F(t_0)]}\mathbb{Z}_1(t), \\ \mathbb{Z}_{n,2}^*(t) &\xrightarrow{d} \mathbb{U}_2(t) := L_{\mathbb{R}}\mathbb{Z}_2^0(t) - L_{\mathbb{R}}\mathbb{Z}_2^0(0) - (L_{\mathbb{R}}\mathbb{Z}_2^0)'(0)t,\end{aligned}$$

where

$$\mathbb{Z}_2^0(t) = \sqrt{F(t_0)[1 - F(t_0)]}\mathbb{Z}_2(t) + \frac{1}{2}f(t_0)g^{-1}(t_0)t^2,$$

and  $\mathbb{Z}_1(t)$  and  $\mathbb{Z}_2(t)$  are two independent two-sided standard Brownian motions. Furthermore, we have that the unconditional distribution of  $\gamma_n^*$  converges to that of  $L_{\mathbb{R}}(\mathbb{U}_1 + \mathbb{U}_2)'(0)$ , which is different from that of  $\kappa\mathbb{C}$ . This gives the inconsistency result. ■

**Proof of Theorem 2.3.** We will apply Theorem 2.1. By Lemma 5.9 in Groeneboom and Wellner (1992), the NPMLE  $\tilde{F}_n$  satisfies  $\|\tilde{F}_n - F\| = O_p(n^{-1/3} \log n)$  under our assumption of  $F$  and  $G$ . Since  $h \rightarrow 0$  and  $n^{1/3}(\log n)^{-1} h \rightarrow \infty$ , we have the following

holds uniformly in  $t$ :

$$\begin{aligned}
|\check{F}_{n,h}(t) - F(t)| &\leq \left| \int \bar{K}_h(t-s) d(\tilde{F}_n(s) - F(s)) \right| + \left| \int \bar{K}_h(t-s) dF(s) - F(t) \right| \\
&= \left| \int (\tilde{F}_n(s) - F(s)) dK_h(t-s) \right| + \left| \int \bar{K}_h(t-s) dF(s) - F(t) \right| + o_p(1) \\
&= O(1) \|\tilde{F}_n - F\| h^{-1} + o_p(1) = o_p(1).
\end{aligned}$$

Thus (3) holds in probability.

Next we show that (17) holds in probability for SMLE  $\check{F}_{n,h}$ , i.e.,

$$\lim_{n \rightarrow \infty} n^{1/3} |\check{F}_{n,h}(t_0 + n^{-1/3}t) - \check{F}_{n,h}(t_0) - f(t_0)n^{-1/3}t| = 0.$$

Note that we have

$$\begin{aligned}
n^{1/3}(\check{F}_{n,h}(t_0 + n^{-1/3}t) - \check{F}_{n,h}(t_0)) &= n^{1/3} \int \{\bar{K}_h(t_0 + n^{-1/3}t - s) - \bar{K}_h(t_0 - s)\} d\tilde{F}_n(s) \\
&= o_p(1) + t \int K_h(t_0 - s) d\tilde{F}_n(s).
\end{aligned}$$

Integrating by parts yields

$$\begin{aligned}
\int K_h(t_0 - s) d\tilde{F}_n(s) &= \int K_h(t_0 - s) d(\tilde{F}_n(s) - F(s)) + \int K_h(t_0 - s) dF(s) \\
&= - \int (\tilde{F}_n(s) - F(s)) dK_h(t_0 - s) + \int K_h(t_0 - s) dF(s) + o_p(1) \\
&= f(t_0) + o_p(1),
\end{aligned}$$

given that  $\|\tilde{F}_n - F\| = O_p(n^{-1/3} \log n)$ ,  $h \rightarrow 0$  and  $n^{1/3}(\log n)^{-1}h \rightarrow \infty$ . ■

### 6.3 Proof of Theorem 3.1

Let  $\mathbb{P}_n^*$  denote the induced measure of the bootstrap sample. For any  $t > 0$ , define

$$W_n^*(t) = W_{n,1}^*(t) + \int_0^t \{F_n(s) - F_n(t_0)\} dW_{n,2}^*(s),$$

where for  $k = 1$  and  $2$ ,

$$\begin{aligned} W_{n,k}^*(t) &= \int_{t_1 \in [0,t], x \leq t_1} \frac{d\mathbb{P}_n^*(x, t_1, t_2)}{\{F_n(t_1)\}^k} + \int_{t_1 \in [0,t], t_1 < x \leq t_2} \frac{d\mathbb{P}_n^*(x, t_1, t_2)}{\{F_n(t_1) - F_n(t_2)\}^k} \\ &+ \int_{t_2 \in [0,t], t_1 < x \leq t_2} \frac{d\mathbb{P}_n^*(x, t_1, t_2)}{\{F_n(t_2) - F_n(t_1)\}^k} + \int_{t_2 \in [0,t], x > t_2} \frac{d\mathbb{P}_n^*(x, t_1, t_2)}{\{F_n(t_2) - 1\}^k}. \end{aligned}$$

Thanks to the characterization of  $\tilde{F}_n^{*,(1)}$  (see, e.g., Groeneboom, 1991), we know that

$$P_n \left[ (n \log n)^{1/3} \{ \tilde{F}_n^{*,(1)}(t_0) - F_n(t_0) \} > x \right] = P_n [ T_n^{*,(0)}(F_n(t_0) + (n \log n)^{-1/3}x) < t_0 ],$$

where  $P_n$  is the probability measure induced by  $F_n$  and the empirical probability measure of  $\{T_{i,1}, T_{i,2}\}_{i=1}^n$ , and

$$T_n^{*,(0)}(x) := \operatorname{sargmin}_t [ W_n^*(t) - \{x - F_n(t_0)\} W_{n,2}^*(t) ].$$

For a function  $w(t)$ ,  $\operatorname{sargmin}_t w(t)$  means the maximum value of the minimizers of function  $w(t)$ . If there is a unique minimizer, then  $\operatorname{sargmin}_t w(t) = \operatorname{argmin}_t w(t)$ . By

the definition of  $T_n^{*,(0)}(x)$ , we can write

$$\begin{aligned}
& (n \log n)^{1/3} \left( T_n^{*,(0)}(F_n(t_0) + (n \log n)^{1/3}x) - t_0 \right) \\
&= \operatorname{sargmin}_t \left\{ n^{2/3} (\log n)^{-1/3} (W_n^*(t_0 + (n \log n)^{1/3}t) - W_n^*(t_0)) \right. \\
&\quad \left. - x (n \log n)^{1/3} (W_{n,2}^*(t_0 + (n \log n)^{-1/3}t) - W_{n,2}^*(t_0)) \right\} \\
&= \operatorname{sargmin}_t \left\{ n^{2/3} (\log n)^{-1/3} (W_{n,1}^*(t_0 + (n \log n)^{1/3}t) - W_{n,1}^*(t_0)) \right. \\
&\quad + n^{2/3} (\log n)^{-1/3} \int_{t_0}^{t_0 + (n \log n)^{1/3}t} (F_n(s) - F_n(t_0)) dW_{n,2}^*(s) \\
&\quad \left. - x n^{1/3} (\log n)^{-2/3} (W_{n,2}^*(t_0 + (n \log n)^{-1/3}t) - W_{n,2}^*(t_0)) \right\}.
\end{aligned}$$

As in the proof of Theorem 6.1, we start with the distributions of three terms in the above display. This is given by the following lemma.

**Lemma 6.2** *For a sequence of distribution functions  $F_n$  that converge weakly to  $F$ , if the following convergence holds uniformly on compacts (in  $t$ )*

$$\lim_{n \rightarrow \infty} (n \log n)^{1/3} |F_n(t_0 + (n \log n)^{-1/3}t) - F_n(t_0) - f(t_0)(n \log n)^{-1/3}t| = 0, \quad (21)$$

*then, conditionally on the data, we have the the following convergence almost surely uniformly on compacta*

$$\begin{aligned}
n^{1/3} (\log n)^{-2/3} (W_{n,2}^*(t_0 + (n \log n)^{-1/3}t) - W_{n,2}^*(t_0)) &\xrightarrow{p} \frac{2}{3f(t_0)} h(t_0, t_0)t, \\
n^{2/3} (\log n)^{-1/3} \int_{t_0}^{t_0 + (n \log n)^{1/3}t} (F_n(s) - F_n(t_0)) dW_{n,2}^*(s) &\xrightarrow{p} \frac{1}{3} h(t_0, t_0)t^2, \\
n^{2/3} (\log n)^{-1/3} (W_{n,1}^*(t_0 + (n \log n)^{1/3}t) - W_{n,1}^*(t_0)) &\xrightarrow{d} \sqrt{\frac{2}{3} h(t_0, t_0)/f(t_0)} \mathbb{Z}(t).
\end{aligned}$$

**Proof of Lemma 6.2.** Proof of Lemma 6.2 follows from a similar argument as in the proof of Theorem 5.3 in Groeneboom (1991). Note that for  $B_0 > 0$ , if (21) holds,

we have that

$$\begin{aligned}
& P_n\left(T_1 < X < T_2, T_1, T_2 \in [t_0, t_0 + B_0(n \log n)^{-1/3}]\right) \\
&= (1 + o(1)) \int_{t_0 \leq t_1 < t_2 \leq t_0 + B_0(n \log n)^{-1/3}} f(t_0)(t_2 - t_1)h(t_0, t_0)dt_2dt_1 \\
&= (1 + o(1))\frac{1}{6}f(t_0)h(t_0, t_0)B_0^3(n \log n)^{-1}, \quad a.s.,
\end{aligned}$$

where  $h$  is the density function of observation times  $T_1$  and  $T_2$ . This mean that the event  $T_1 < X < T_2$  and both  $T_1$  and  $T_2$  are in interval  $[t_0, t_0 + B_0(n \log n)^{-1/3}]$  has probability going to 0 at rate  $n \log n$ . Then, for  $n$  observations  $\{T_{i,1}, T_{i,2}, X_i\}_{i=1}^n$ , we have

$$\begin{aligned}
& P_n\left(\exists i \in \{1, \dots, n\} : T_{i,1} < X_i < T_{i,2}, T_{i,1}, T_{i,2} \in [t_0, t_0 + B_0(n \log n)^{-1/3}]\right) \\
&\leq n \times (1 + o(1))\frac{1}{6}f(t_0)h(t_0, t_0)B_0^3(n \log n)^{-1} = o(1), \quad a.s.
\end{aligned}$$

Therefore, conditionally on  $T$ 's, the following equation holds with probability 1, almost surely,

$$\begin{aligned}
& n^{1/3}(\log n)^{-2/3}(W_{n,2}^*(t_0 + (n \log n)^{-1/3}t) - W_{n,2}^*(t_0)) \\
&= n^{1/3}(\log n)^{-2/3} \\
&\times \left[ \int_{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t]} \frac{\mathbf{1}_{x \leq t_1}}{F_n(t_1)^2} + \frac{\mathbf{1}_{t_1 < x \leq t_2, t_2 > t_1 + B_0(n \log n)^{-1/3}}}{(F_n(t_2) - F_n(t_1))^2} d\mathbb{P}_n^*(x, t_1, t_2) \right. \\
&\left. + \int_{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t]} \frac{\mathbf{1}_{t_1 < x \leq t_2, t_2 > t_1 + B_0(n \log n)^{-1/3}}}{(F_n(t_2) - F_n(t_1))^2} + \frac{\mathbf{1}_{x > t_2}}{(1 - F_n(t_2))^2} d\mathbb{P}_n^*(x, t_1, t_2) \right].
\end{aligned}$$

The above display is equal to

$$\begin{aligned}
& (1 + o(1))n^{1/3}(\log n)^{-2/3} \\
& \times \left[ \int_{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t]} \frac{\mathbf{1}_{x \leq t_1}}{F_n(t_1)} + \frac{\mathbf{1}_{t_1 < x \leq t_2, t_2 > t_1 + B_0(n \log n)^{-1/3}}}{F_n(t_2) - F_n(t_1)} dH(t_1, t_2) \right. \\
& \left. + \int_{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t]} \frac{\mathbf{1}_{t_1 < x \leq t_2, t_2 > t_1 + B_0(n \log n)^{-1/3}}}{F_n(t_2) - F_n(t_1)} + \frac{\mathbf{1}_{x > t_2}}{1 - F_n(t_2)} dH(t_1, t_2) \right] \\
& = (1 + o(1)) \int_{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t]} \frac{\mathbf{1}_{t_1 < x \leq t_2, t_2 > t_1 + B_0(n \log n)^{-1/3}}}{f(t_0)(t_2 - t_1)} h(t_0, t_0) dt_1 dt_2 \\
& \quad + (1 + o(1)) \int_{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t]} \frac{\mathbf{1}_{t_1 < x \leq t_2, t_2 > t_1 + B_0(n \log n)^{-1/3}}}{f(t_0)(t_2 - t_1)} h(t_0, t_0) dt_1 dt_2 + o(1) \\
& = (1 + o(1)) \frac{2h(t_0, t_0)t}{3f(t_0)}.
\end{aligned}$$

In the above display, recall that  $H$  is the distribution function of  $T_1$  and  $T_2$ . This gives the first convergence result.

The second convergence follows from a similar argument. Conditionally on  $T$ 's, the following equation holds with probability 1, almost surely,

$$\begin{aligned}
& n^{2/3}(\log n)^{-1/3} \int_{t_0}^{t_0 + (n \log n)^{1/3}t} (F_n(s) - F_n(t_0)) dW_{n,2}^*(s) \\
& = n^{2/3}(\log n)^{-1/3} \left[ \int_{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t], x \leq t_1} \frac{F_n(t_1) - F_n(t_0)}{F_n(t_1)^2} d\mathbb{P}_n^*(x, t_1, t_2) \right. \\
& \quad + \int_{\substack{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t], t_1 < x \leq t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{F_n(t_1) - F_n(t_0)}{(F_n(t_2) - F_n(t_1))^2} d\mathbb{P}_n^*(x, t_1, t_2) \\
& \quad + \int_{\substack{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t], t_1 < x \leq t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{F_n(t_2) - F_n(t_0)}{(F_n(t_2) - F_n(t_1))^2} d\mathbb{P}_n^*(x, t_1, t_2) \\
& \quad \left. + \int_{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t], x > t_2} \frac{F_n(t_2) - F_n(t_0)}{(1 - F_n(t_2))^2} d\mathbb{P}_n^*(x, t_1, t_2) \right].
\end{aligned}$$

The preceding display is equal to

$$\begin{aligned}
& (1 + o(1)) \int_{\substack{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t], t_1 < x \leq t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{t_1 - t_0}{t_2 - t_1} h(t_0, t_0) dt_1 dt_2 \\
& + (1 + o(1)) \int_{\substack{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t], t_1 < x \leq t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{t_1 - t_0}{t_2 - t_1} h(t_0, t_0) dt_1 dt_2 + o(1) \\
& = (1 + o(1)) \frac{1}{3} h(t_0, t_0) t^2.
\end{aligned}$$

The third convergence result follows from a similar argument as in the proof of Lemma 5.5 in Groeneboom (1991). For  $t \in [0, B_0]$ , let

$$\begin{aligned}
\bar{W}_n^*(t) &= n^{2/3} (\log n)^{-1/3} \left[ \int_{\substack{t_1 \in [t_0, t_0 + (n \log n)^{1/3}t], x \leq t_1, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{1}{F_n(t_1)} d\mathbb{P}_n^*(x, t_1, t_2) \right. \\
&\quad - \int_{\substack{t_1 \in [t_0, t_0 + (n \log n)^{1/3}t], t_1 < x \leq t_2 \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{1}{F_n(t_2) - F_n(t_1)} d\mathbb{P}_n^*(x, t_1, t_2) \\
&\quad + \int_{\substack{t_2 \in [t_0, t_0 + (n \log n)^{1/3}t], t_1 < x \leq t_2 \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{1}{F_n(t_2) - F_n(t_1)} d\mathbb{P}_n^*(x, t_1, t_2) \\
&\quad \left. - \int_{\substack{t_2 \in [t_0, t_0 + (n \log n)^{1/3}t], x > t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{1}{1 - F_n(t_2)} d\mathbb{P}_n^*(x, t_1, t_2) \right].
\end{aligned}$$

We can see that conditional on  $T$ 's,  $\bar{W}_n^*(t)$  is a martingale and its variance is given

by

$$\begin{aligned}
& (1 + o(1))n^{1/3}(\log n)^{-2/3} \left[ \int_{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t], x \leq t_1} \frac{1}{F_n(t_1)} dH(t_1, t_2) \right. \\
& + \int_{\substack{t_1 \in [t_0, t_0 + (n \log n)^{-1/3}t], t_1 < x \leq t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{1}{F_n(t_2) - F_n(t_1)} dH(t_1, t_2) \\
& + \int_{\substack{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t], t_1 < x \leq t_2, \\ t_2 > t_1 + B_0(n \log n)^{-1/3}}} \frac{1}{F_n(t_2) - F_n(t_1)} dH(t_1, t_2) \\
& \left. + \int_{t_2 \in [t_0, t_0 + (n \log n)^{-1/3}t], x > t_2} \frac{1}{1 - F_n(t_2)} dH(t_1, t_2) \right] \\
& = (1 + o(1)) \frac{2h(t_0, t_0)t}{3f(t_0)}, \quad a.s.
\end{aligned}$$

Therefore, by the martingale central limit theorem, we have the conditional weak convergence of  $\bar{W}_n^*(t)$  to  $\sqrt{\frac{2}{3}h(t_0, t_0)/f(t_0)}\mathbb{Z}(t)$ . Next we show that the difference between  $\bar{W}_n^*$  and  $n^{2/3}(\log n)^{-1/3}(W_{n,1}^*(t_0 + (n \log n)^{1/3}t) - W_{n,1}^*(t_0))$  is ignorable. Let

$$\hat{W}_n^* = n^{2/3}(\log n)^{-1/3}(W_{n,1}^*(t_0 + (n \log n)^{1/3}t) - W_{n,1}^*(t_0)) - \bar{W}_n^*.$$

By Markov's inequality and a similar argument as in the proof of the second convergence result, we obtain that

$$P_n \left( \max_{t \in [0, B_0]} |\hat{W}_n^*(t)| > \epsilon \right) \leq O(1) \frac{\epsilon^{-1}n^{2/3}}{(\log n)^{1/3}} \int_{t_1, t_2 \in [t_0, t_0 + (n \log n)^{-1/3}B_0]} dH(t_1, t_2) = o(1) \text{ a.s.}$$

Therefore, we have the third convergence result. ■

**Proof of Theorem 3.1.** We need the following tightness result, whose proof follows from a similar argument as in the proof of Lemma 5.6 in Groeneboom (1991) and is omitted from this paper.

**Lemma 6.3** *If (21) holds, then for any  $\epsilon > 0$  and  $B_0 > 0$ , there exists a constant  $B$*

such that the following result holds almost surely

$$P_n \left( \max_{x \in [-B_0, B_0]} (n \log n)^{1/3} |T_n^{*,(0)}(F_n(t_0) + (n \log n)^{-1/3}x) - t_0| > B \right) < \epsilon.$$

The above lemma and Lemma 6.2 imply that  $(T_n^{*,(0)}(F_n(t_0) + (n \log n)^{-1/3}x) - t_0)$  converges to the process

$$T(x) := \operatorname{sargmin}_t \left\{ \sqrt{\frac{2h(t_0, t_0)}{3f(t_0)}} \mathbb{Z}(t) + \frac{1}{3}h(t_0, t_0)t^2 - x \frac{2h(t_0, t_0)t}{3f(t_0)} \right\} \quad a.s.$$

Since  $P_n((n \log n)^{1/3}(F_n^{*,(1)}(t_0) - F_n(t_0)) \leq x) = P_n(T_n^{*,(0)}(F_n(t_0) + (n \log n)^{-1/3}x) \geq t_0)$ , we have

$$P_n \left( (n \log n)^{1/3}(F_n^{*,(1)}(t_0) - F_n(t_0)) \leq x \right) \xrightarrow{a.s.} P(T(x) - f^{-1}(t_0)x \geq -f^{-1}(t_0)x).$$

By the stationarity of process  $T(x) - f^{-1}(t_0)x$  as given in Groeneboom (1989), we have that  $P_n((n \log n)^{1/3}(F_n^{*,(1)}(t_0) - F_n(t_0)) \leq x)$  converges to  $P(T(0) \geq -f^{-1}(t_0)x)$ . Then by a Brownian scaling argument as in the proof of Theorem 2.1, we obtain the desired conclusion. ■