TU Delft

Delft University of Technology

The nonparametric bootstrap for the current status model

Groeneboom, Piet; Hendrickx, Kim

DOI
10.1214/17-EJS1345

Publication date
2017

Document Version
Final published version

Published in
Electronic Journal of Statistics

Citation (APA)
Groeneboom, P., & Hendrickx, K. (2017). The nonparametric bootstrap for the current status model. *Electronic Journal of Statistics*, *11*(2), 3446-3484. https://doi.org/10.1214/17-EJS1345

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Electronic Journal of Statistics
Vol. 11 (2017) 3446–3484
ISSN: 1935-7524
DOI: 10.1214/17-EJS1345

The nonparametric bootstrap for the current status model

Piet Groeneboom

Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands
e-mail: P.Groeneboom@tudelft.nl

and

Kim Hendrickx

Hasselt University, I-BioStat, Agoralaan, B3590 Diepenbeek, Belgium
e-mail: kim.hendrickx@uhasselt.be

Abstract: It has been proved that direct bootstrapping of the nonparametric maximum likelihood estimator (MLE) of the distribution function in the current status model leads to inconsistent confidence intervals. We show that bootstrapping of functionals of the MLE can however be used to produce valid intervals. To this end, we prove that the bootstrapped MLE converges at the right rate in the L_p-distance. We also discuss applications of this result to the current status regression model.

MSC 2010 subject classifications: Primary 62G09, 62N01.
Keywords and phrases: Bootstrap, current status, MLE, smooth functionals.

Received January 2017.

1. Introduction

In the current status model, the variable of interest is a survival variable X with distribution function F_0. However, instead of observing the exact survival time X, a censoring variable $T \sim G$ is observed together with the indicator $\Delta = 1_{X \leq T}$. Such data arise naturally in clinical trials when a patient can only be checked at one measurement due to destructive testing. A lot of research has been published on the behavior of the maximum likelihood estimator (MLE) F_n of the distribution function F_0. The limiting distribution of $n^{1/3}(F_n(t) - F_0(t))$ is after scaling by the constant $\kappa = \{4F_0(t)(1 - F_0(t))f_0(t)/g(t)\}^{1/3}$ given by

$$\mathbb{C} = \arg\max_t \left\{ W(t) - t^2 \right\},$$

where W is a two-sided Brownian motion with $W(0) = 0$ (see [19]). Other estimators with similar asymptotic properties are Chernoff's estimator of the mode ([6]), the Grenander estimator ([10]) of a nonincreasing density, Manski's maximum score estimator ([27]) and Rouseeuw's least median of squares estimator ([29]). A general framework for cube-root n asymptotics is given in [25].

In this paper we investigate the behavior of Efron's nonparametric bootstrap method ([9]) for constructing confidence intervals for smooth functionals of the MLE. It is known that the nonparametric bootstrap is inconsistent for generating the limit distribution of the MLE. The authors of [2] prove that (conditional on the data),

$$n^{1/3}\{4F_0(t)(1-F_0(t))f_0(t)/g(t)\}^{-1/3}\{\hat{F}_n(t)-F_n(t)\} \xrightarrow{\mathcal{D}} \arg\max_t(W(t)+\hat{W}(t)-t^2)-\arg\max_t(W(t)-t^2),$$

where $\hat{F}_n$ is the bootstrap MLE and W and $\hat{W}$ are two independent two-sided Brownian motions originating at zero. A similar result is obtained in [26] and in [31] for the Grenander estimator. The maximum score estimator of [27] is another example of a cube-root n statistic with asymptotic distribution derived in [25], where inconsistency of the nonparametric bootstrap for this estimator is shown in [2].

Constructing asymptotic confidence intervals for the distribution function in the current status model based on Chernoff's distribution and the normalizing constant κ is complicated by the need to compute the critical values of $\mathbb{C}$ and to estimate the density f_0 consistently. Since this turns out to be a rather difficult task several alternative bootstrap methods have been proposed based on resampling from a smooth estimate. [32] consider a smooth kernel estimate $\tilde{F}$ of F_0 and resample the Δ_i from a Bernoulli distribution with probability $\tilde{F}(T_i)$, while keeping the censoring variables T_i fixed and center the values of the bootstrap samples by subtracting the smooth estimate of the distribution function. [26] and [31] propose similar smooth respampling schemes for the Grenander estimator and a model-based smoothed bootstrap procedure for making inference on the maximum score estimator is developed in [28]. All methods result in consistent estimation of the (suitably standardized) distribution $\mathbb{C}$ conditional on the original data.

A drawback of this approach is that smoothness conditions of F_0 are used which allow faster than cube-root n estimation of F_0. This raises the question if one should really use confidence intervals based on the MLE instead of on a faster converging estimate.

This latter procedure is followed in [14], where the authors consider constructing confidence intervals around the smoothed maximum likelihood estimator (SMLE) of F_0 in the current status model. The SMLE is a kernel estimate based on the MLE with an asymptotic normal distribution, instead of Chernoff's limiting distribution ([16]). The bootstrap method proposed in [14] is however still based on the smooth bootstrap procedure described in [32] and not on Efron's nonparametric bootstrap. We show in this paper that the construction of confidence intervals around the SMLE based on the nonparametric bootstrap can also be proved to be valid, where one does not resample from a smooth estimate of F_0, but just resamples with replacement from the pairs (T_i, Δ_i) in the original sample. This method already has been used without proof in [17] and also in [18] and the present manuscript intends to fill the gap

of the missing proofs here. An important difference with the smooth bootstrap in [14] is that for the centering of the estimates in the nonparametric bootstrap samples the SMLE of the original sample is used, whereas this will not work for the resampling as proposed in [14]; in the latter case one needs to center the estimates in the bootstrap samples by a kernel convolution of the SMLE in the original sample. It is not clear which method is better, and the most striking fact is the similarity of the results of the two methods in our simulations. An advantage of the purely nonparametric bootstrap, discussed in the present paper, might be its conceptual simplicity and the absence of the need to center with a convolution of the SMLE in the centering of the bootstrap samples instead of the SMLE itself. An advantage of the smooth bootstrap, discussed in [14] might be the fact that only the indicators Δ_i are being resampled, and that in this sense one stays closest to the sample distribution of the observation times T_i, which stay fixed in this procedure.

Although it is argued in [8] that the naive bootstrap will not work for their goodness-of-fit test for monotone functions, based on the Grenander estimator, no theoretical justification for this conjecture is given. Other examples where a smooth bootstrap procedure is used, are the likelihood ratio type two-sample test for current status data proposed by [11] and the test for equality of functions under monotonicity constraints proposed by [7]. Both tests establish asymptotic normality for the test statistic considered.

The paper is organized as follows: In Section 2 we introduce the current status model and review some interesting properties of the MLE. The validity of the nonparametric bootstrap is discussed in Section 3. In Section 4 we provide two examples to illustrate the applicability of our result. In the first example we construct pointwise confidence intervals based on the smoothed MLE in the current status model. The second example deals with doing inferences for a finite dimensional regression parameter in the current status linear regression model. For both examples, the theoretical and finite sample behavior of the nonparametric bootstrap is discussed. Section 5 presents some concluding remarks. The proofs of our results are given in Section 6.

2. The current status model and the MLE

Let $Z_1 = (T_1, \Delta_1), \ldots, Z_n = (T_n, \Delta_n)$ be an i.i.d. sample from the probability space $([0, R] \times \{0, 1\}, \mathcal{A}, P)$, where $\Delta_i = 1_{X_i \leq T_i}$ and $R > 0$. The X_i are interpreted as (nonnegative) survival times with distribution function F_0. Instead of observing X, a censoring variable $T \sim G$ is observed (with density g) independent of X. One could say that in the current status model, each observation Z_i represents the current status of the item i at time T_i. The density of Z_i with respect to the product of Lebesgue measure and counting measure on $[0, R] \times \{0, 1\}$ is given by

$$p_{F_0}(t, \delta) = [\delta F_0(t) + (1 - \delta)\{1 - F_0(t)\}]\, g(t).$$

The maximum likelihood estimator F_n is defined as the maximizer of the log likelihood given by (up to a constant not depending on F),

$$\ell_n(F) = n^{-1} \sum_{i=1}^{n} \left[\Delta_i \log F(T_i) + (1 - \Delta_i) \log\{1 - F(T_i)\} \right], \tag{2.1}$$

over all distribution functions $F : [0, \infty] \mapsto [0, 1]$. [19] show that the MLE can be characterized as the left-continuous slope of the greatest convex minorant of a cumulative sum diagram consisting of the points (0,0) and

$$\left(i, \sum_{j \le i} \Delta_{(j)} \right),$$

where we let $T_{(j)}$ denote the jth order statistic of the T_i and $\Delta_{(j)}$ be the Δ_i corresponding to it (assuming no ties are present in the data). An important property of the MLE is the so-called *switch relation*, see [17] p. 69. Let $\mathbb{G}_n$ be the empirical distribution function of $T_1, \ldots, T_n$ and define the process V_n by

$$V_n(t) = n^{-1} \sum_{i=1}^{n} \Delta_i 1_{\{T_i \le t\}}, \tag{2.2}$$

and the process (in a) U_n by

$$U_n(a) = \operatorname{argmin}\{t \in \mathbb{R} : V_n(t) - a\mathbb{G}_n(t)\}. \tag{2.3}$$

Then, taking $a = F_0(t)$, we get the *switch relation*:

$$F_n(t) \ge a \iff U_n(a) \le t,$$

see also Figure 1.

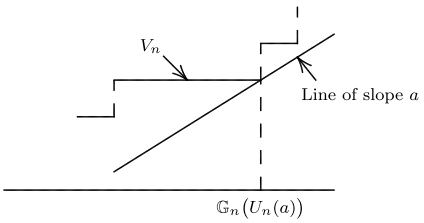


FIG 1. *The switch relation.*

3. Bootstrapping the MLE

In this section we establish properties of the bootstrap MLE $\hat{F}_n$ based on the nonparametric bootstrap proposed by [9]. Our main concern is to show that conditional on the data $Z_1, \ldots, Z_n$, we have

$$E\left\{ \left\| n^{1/3} \left\{ \hat{F}_n - F_0 \right\} \right\|_p \;\middle|\; Z_1, \ldots, Z_n \right\} = O_p(1), \tag{3.1}$$

and

$$\sup_{t\in[0,R]} E\left\{ n^{1/3}\left|\hat{F}_n(t) - F_0(t)\right| \,\middle|\, Z_1,\ldots,Z_n \right\} = O_p(1). \tag{3.2}$$

Denote the empirical probability measure of $Z_1,\ldots,Z_n$ by $\mathbb{P}_n$. The bootstrap empirical measure is

$$\hat{\mathbb{P}}_n = n^{-1}\sum_{i=1}^{n} M_{ni} 1_{Z_i},$$

where 1_{Z_i} denotes the points mass at $Z_i = (T_i, \Delta_i)$ and

$$M_n = (M_{n1},\ldots,M_{nn}) \sim \text{multinomial}(n, n^{-1},\ldots,n^{-1}),$$

is a vector of multinomial weights, independent of $Z_1,\ldots,Z_n$. The bootstrap MLE $\hat{F}_n$ is computed using the weighted cumulative sum diagram formed by the point $(0,0)$ and

$$\left(\sum_{j=1}^{i} M_{n(j)}, \sum_{j=1}^{i} M_{n(j)}\Delta_{(j)}\right),$$

where $M_{n(j)}$ corresponds to the multinomial weight corresponding to $T_{(j)}$. The bootstrap MLE $\hat{F}_n$ is then calculated from the left-continuous slope of the convex minorant of this cusum diagram.

To complete notation, we suppose that the vectors $((Z_1,\ldots,Z_n), M_n), n = 1,2,\ldots$ are defined on the product space $(([0,R]\times\{0,1\})^\infty \times \mathbb{Z}_+^\infty, \mathcal{B}, P_{ZM})$, where $\mathbb{Z}_+$ is the set of nonnegative integers and $\mathcal{B}$ is the collection of Borel sets, generated by the finite dimensional projections. We say that a real-valued function Γ_n defined on the joint probability space is of order $o_{P_M}(1)$ in probability if for all $\epsilon,\eta > 0$:

$$P^*\left(P_{M|Z}\left\{|\Gamma_n| > \epsilon\right\} > \eta\right) \to 0 \text{ as } n\to\infty,$$

where P^* denotes outer probability and $P_{M|Z}$ is the conditional probability measure w.r.t. the weights, given the sample $Z_1,\ldots,Z_n$.

To establish (3.1), we need the following result, which is a bootstrap version of Lemma 11.5 in [17].

Lemma 3.1. *Suppose F_0 has a continuous density f_0 with support [0,R] that satisfies,*

$$0 < \inf_{t\in[0,R]} f_0(t) < \sup_{t\in[0,R]} f_0(t) < \infty.$$

Also suppose that the observation distribution G has a continuous derivative g that stays away form zero and infinity on $[0,R]$. Let

$$U(a) = F_0^{-1}(a) \qquad 0 < a < 1,$$

and define the process

$$\hat{U}_n(a) = \text{argmin}\{t \in [0,R] : \hat{V}_n(t) - a\hat{\mathbb{G}}_n(t)\} \qquad 0 < a < 1,$$

with processes $\hat{V}_n$ *and* $\hat{\mathbb{G}}_n$ *defined by*

$$\hat{V}_n(t) = \int_{u\in[0,t]} \delta \, d\hat{\mathbb{P}}_n(u,\delta) \quad \text{and} \quad \hat{\mathbb{G}}_n(t) = \int_{u\in[0,t]} d\hat{\mathbb{P}}_n(u,\delta) \quad t \in [0,R]. \tag{3.3}$$

Then there are positive constants K_1 *and* K_2*, such that, for all* $a \in (0,1)$ *and for all large* n*:*

$$\left\{\exists x \in [0,R] : P_{M|Z}\left\{n^{1/3}\left|\hat{U}_n(a) - U(a)\right| \geq x\right\} > K_1 e^{-K_2 x^{3/2}}\right\} = o_p(1),$$

where $\{A\}$ *denotes the indicator* 1_A *of the event* A*.*

Lemma 3.1 implies that the probability that for all $x \in [0,R]$, and $a = F_0(t)$,

$$P_{M|Z}\left\{n^{1/3}\left|\hat{U}_n(a) - U(a)\right| \geq x\right\} \leq K_1 e^{-K_2 x^{3/2}},$$

tends to 1 as $n \to \infty$. The proof of Lemma 3.1 is given in Section 6. The proof uses empirical process theory and results on tail probabilities for $\|\sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n)\|_{\mathcal{F}}$ for classes $\mathcal{F}$ with finite entropy integrals. Similar results are proved using martingale theory in Section 11.2 of [17] for the original sample and in [14] for a smooth bootstrap empirical process. Since

$$E_{M|Z}\left[n^{1/3}\{\hat{F}_n(t) - F_0(t)\}_+\right]^p = \int_0^\infty P_{M|Z}\left\{n^{1/3}\{\hat{F}_n(t) - F_0(t)\} \geq x\right\} px^{p-1}dx,$$

where $\{\hat{F}_n(t) - F_0(t)\}_+$ denotes the positive part of $\{\hat{F}_n(t) - F_0(t)\}$ and since,

$$\begin{aligned} &P_{M|Z}\left\{\hat{U}_n\left(a + n^{-1/3}x\right) \leq t\right\} \\ &\quad = P_{M|Z}\Bigg[n^{1/3}\left\{\hat{U}_n\left(a + n^{-1/3}x\right) - U\left(a + n^{-1/3}x\right)\right\} \\ &\qquad\qquad \leq n^{1/3}\left\{t - U\left(a + n^{-1/3}x\right)\right\}\Bigg], \end{aligned}$$

it follows from Lemma 3.1 and the *bootstrapped switch relation* given by

$$P_{M|Z}\left\{n^{1/3}\{\hat{F}_n(t) - F_0(t)\} \geq x\right\} = P_{M|Z}\left\{\hat{U}_n\left(a + n^{-1/3}x\right) \leq t\right\},$$

that there exists a positive constant $K > 0$ such that,

$$\left\{\exists t \in [0,R] : E_{M|Z}\left|\hat{F}_n(t) - F_0(t)\right|^p > Kn^{-p/3}\right\} = o_p(1).$$

In particular, there exists a $K_1 > 0$ such that:

$$P\left\{\sup_{t\in[0,R]} E_{M|Z}\left|\hat{F}_n(t) - F_0(t)\right| > K_1 n^{-1/3}\right\} \longrightarrow 0, \qquad n \to \infty,$$

and likewise there exists a $K_2 > 0$ such that:

$$P\left\{E_{M|Z}\left\|\hat{F}_n - F_0\right\|_2 > K_2 n^{-1/3}\right\} \longrightarrow 0, \qquad n \to \infty.$$

In the next section we show how (3.1) can be used to justify the bootstrap validity for drawing inferences in models which can be estimated using smooth functionals of the MLE. The proofs for deriving the asymptotic behavior of these functionals are in general based on applications of the Cauchy-Schwarz inequality and on showing asymptotic equicontinuity. Both steps involve calculating the L_2-distance which can often be reduced to the L_2-distance between the MLE and the true underlying distribution function. Our main result given in (3.1) is therefore important to show that the asymptotic properties of the estimates obtained in the original sample are still valid in the bootstrap sample conditionally on the data. The asymptotic behavior of the functionals does not depend on the distribution function of the MLE, which is, as shown in Theorem 5 of [2], not the same in the original sample and the bootstrap sample (conditionally on the data). We note that the variances of the corresponding asymptotic distributions however still have the same order $n^{-2/3}$, just like our squared L_p-distances in (3.1).

4. Applications

In this Section we illustrate the applicability of our bootstrap results. In our first example we consider the current status model described in Section 2 and estimate F_0 by the SMLE. In the second example we consider estimating a finite dimensional regression parameter for the current status model, where in addition to observing the vector (T, Δ), also a covariate vector X is observed.

4.1. The Smoothed Maximum Likelihood Estimator (SMLE)

We estimate F_0 by the SMLE $\tilde{F}_{nh}$ obtained by first estimating the MLE F_n and then smoothing this using a smoothing kernel, i.e.,

$$\tilde{F}_{nh}(t) = \int \mathbb{K}\left((t-x)/h\right)\, dF_n(x), \tag{4.1}$$

where $\mathbb{K}$ is an integrated kernel,

$$\mathbb{K}(u) = \int_{-\infty}^{u} K(x)\, dx,$$

and where h is a chosen bandwidth. Here dF_n represents the jumps of the discrete distribution function F_n and K is one of the usual symmetric twice differentiable kernels with compact support, used in density estimation. In our computer experiments, we used the triweight kernel

$$K(u) = \frac{35}{32}\left(1-u^2\right)^3 1_{[-1,1]}(u).$$

For a constant $c > 0$ and $h = cn^{-1/5}$, the SMLE has been proved to converge at rate $n^{-2/5}$ with asymptotic limit distribution,

$$n^{2/5}\left\{\tilde{F}_{nh}(t) - F_0(t)\right\} \xrightarrow{\mathcal{D}} N(\beta(t), \sigma^2(t)),$$

where

$$\beta(t) = \frac{c^2 f_0'(t)}{2}\int u^2 K(u)\,du \quad \text{and} \quad \sigma^2(t) = \frac{F_0(t)\{1-F_0(t)\}}{cg(t)}\int K(u)^2\,du. \tag{4.2}$$

(see [16]). The SMLE is often used in the smooth bootstrap procedures described in Section 1 (see also the numerical example below). Let $\tilde{F}^*_{nh}(t)$ be the bootstrapped SMLE based on replacing F_n in (4.1) by the bootstrapped MLE $\hat{F}_n$, then we have the following result,

$$n^{2/5}\left\{\tilde{F}^*_{nh}(t) - \tilde{F}_{nh}(t)\right\} \xrightarrow{\mathcal{D}} N(0, \sigma^2(t)), \tag{4.3}$$

given the data $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, in probability. Note that, in contrast to the smooth bootstrap method described in [14], we do not need to estimate the convolution SMLE (see (4.7) below).

To prove the asymptotic normality result for the nonparametric bootstrap, given in (4.3), we prove (in Section 6) the following Lemma:

Lemma 4.1. *Assume that the conditions of Lemma 3.1 are satisfied and that g has a bounded derivative g' on $[0, R]$. Let t be an interior point of $[0, R]$ such that f_0 has a continuous derivative f_0' at t. If $h \sim cn^{-1/5}$ then,*

$$\tilde{F}^*_{nh}(t) = \tilde{F}^{(toy)*}_{nh}(t) + o_{P_M}(n^{-2/5}),$$

in probability, where

$$\tilde{F}^{(toy)*}_{nh}(t) = \int \mathbb{K}((t-u)/h)\,dF_0(u) + \int \frac{K((t-u)/h)\ \{\delta - F_0(u)\}}{hg(u)}\,d\hat{\mathbb{P}}_n(u,\delta). \tag{4.4}$$

Since

$$\tilde{F}_{nh}(t) = \tilde{F}^{(toy)}_{nh}(t) + o_p(n^{-2/5}), \tag{4.5}$$

where $\tilde{F}_{nh}^{(toy)}(t)$ is defined by (4.4) with $\hat{\mathbb{P}}_n$ replaced by $\mathbb{P}_n$, we have by Lemma 4.1 that,

$$n^{2/5}\left\{\tilde{F}^*_{nh}(t)-\tilde{F}_{nh}(t)\right\}=n^{2/5}\int\frac{K((t-u)/h)\,\{\delta-F_0(u)\}}{hg(u)}\,d(\hat{\mathbb{P}}_n-\mathbb{P}_n)(u,\delta)+o_{P_M}(1),$$

in probability, which converges, conditional on the data $(T_1,\Delta_1),\ldots(T_n,\Delta_n)$ to the same asymptotic limit as

$$n^{2/5}\int\frac{K((t-u)/h)\,\{\delta-F_0(u)\}}{hg(u)}\,d(\mathbb{P}_n-P)(u,\delta),$$

in probability (see e.g. [21] for more details about the use of the bootstrap for kernel estimators). Finally, applying the central limit theorem on the expression above proves the asymptotic normality result for the bootstrapped SMLE given in (4.3). The proof of Lemma 4.1 is a generalization of the proof for the representation of the SMLE $\tilde{F}_{nh}(t)$ as the "toy-estimator" defined in (4.5). The proof is outlined in Section 11.3 of [17] and uses the result of Theorem 11.3 given in Section 11.2 which is the analogue of our Lemma 3.1 in the original sample.

Remark 4.1. In practice, one should use a boundary correction to ensure consistent estimation of F_0 near the boundaries of the support $[0,R]$. In our experiments we used the method of [30], see also p. 328 in [17]. It is straightforward to show that the nonparametric bootstrap method remains valid under this boundary correction. Moreover, one should also take into account the bias defined in (4.2) when constructing confidence intervals around the SMLE. The bias issue is discussed in more details via a simulation study in Section 4.1.1.

In the remainder part of this Section, we show the applicability of this bootstrap result (4.3) by constructing pointwise confidence intervals (CIs) around the SMLE. We consider two different simulation models and a real data example to illustrate the performance of these CIs.

In the first simulation study we compare our nonparametric bootstrap CIs with (a) the smooth bootstrap CIs proposed in [14], (b) the likelihood ratio intervals around the MLE F_n proposed in [4], (c) the smooth bootstrap MLE-based intervals proposed in [32] and (d) Wald-type CIs, derived from the asymptotic normality of the SMLE.

In a second simulation study, we discuss the difficulties with the construction of pointwise CIs around the SMLE that are not necessarily specific to the bootstrap procedure but that have to be taken into account in order to obtain good CIs around the SMLE under current status data. We first describe a bandwidth selection procedure for choosing the bandwidth of the SMLE and we next discuss the effect of the bias on the performance of the CIs. The algorithms to produce the proposed CIs around the SMLE can be found in the R package `curstatCI`.

4.1.1. Simulation study 1: comparing CIs for the distribution function under current status data

To illustrate the performance of the nonparametric bootstrap procedure for constructing pointwise CIs of the distribution function, we consider a first simulation study based on $N = 5,000$ simulation runs from a model where both X and T have a Uniform(0,2) distribution. In this model the bias $\beta(t)$ defined in (4.2) is zero for all $t \in [0,2]$. The $1-\alpha$ bootstrap interval is given by

$$\left[\tilde{F}_{nh}(t) - Q^*_{1-\alpha/2}(t)\sqrt{S_{nh}(t)}, \tilde{F}_{nh}(t) - Q^*_{\alpha/2}(t)\sqrt{S_{nh}(t)}\right], \tag{4.6}$$

where $Q^*_\alpha(t)$ is the αth quantile of B values of $W^*_{nh}(t)$ defined by

$$W^*_{nh}(t) = \left\{\tilde{F}^*_{nh}(t) - \tilde{F}_{nh}(t)\right\} / \sqrt{S^*_{nh}(t)},$$

where $S_{nh}(t)$ resp. $S^*_{nh}(t)$ are estimates of the variance $\sigma^2(t)$ defined in (4.2) (apart from the factor $cg(t)$ which drops out in the Studentized bootstrap procedure) given by

$$S_{nh}(t) = \frac{1}{n^2}\sum_{i=1}^{n} K_h(t-T_i)^2\left(\Delta_i - F_n(T_i)\right)^2,$$
$$S^*_{nh}(t) = \frac{1}{n^2}\sum_{i=1}^{n} M_{ni}K_h(t-T_i)^2\left(\Delta_i - \hat{F}_n(T_i)\right)^2.$$

In Figure 2(a) we compare the proportion of times that $F_0(t)$ is not in the 95% bootstrap CIs for $t = 0.02, 0.04, \ldots, 2$ with the corresponding proportions obtained with (a) the smooth bootstrap procedure proposed in [14], (b) the likelihood ratio intervals around the MLE F_n proposed in [4] and (c) the smooth

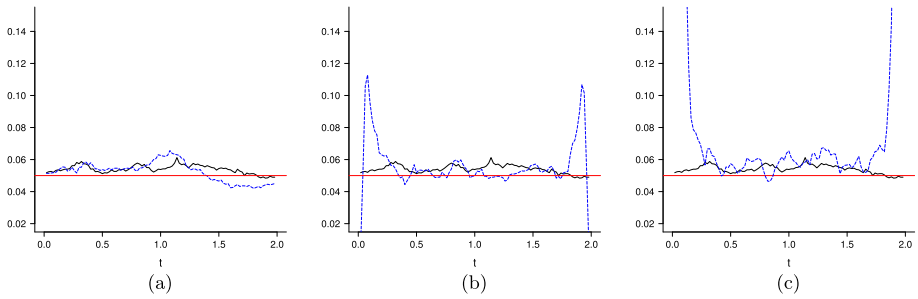


FIG 2. *Uniform samples: Proportion of times that $F_0(t)$, $t = 0.02, 0.04, \ldots$ is not in the 95% CIs for the classical bootstrap CIs defined in (4.6) (black, solid) and (a) the smooth bootstrap (blue, dashed) procedure in constructing CIs around the SMLE of [14], (b) the likelihood ratio CIs of [4] (blue, dashed) and (c) the smooth MLE-based CIs of [32] (blue, dashed). $n = 1,000$, $N = 5,000$, $B = 1,000$ and $h = 2n^{-1/5}$.*

bootstrap MLE-based intervals proposed in [32]. For samples of size $n = 1,000$, $B = 1,000$ bootstrap samples were generated for both methods and the triweight kernel is used for calculation of the SMLE with $h = 2n^{-1/5}$, where the constant $c = 2$ corresponds to the length of the support of the observation variable T. For the smooth bootstrap procedures (a) and (c), first a bootstrap sample $(T_1, \Delta_1^*), \ldots, (T_n, \Delta_n^*)$ is obtained by keeping the T_i in the original sample fixed and by resampling the Δ_i^* from a Bernoulli distribution with probability $\tilde{F}_{nh}(T_i)$, then the bootstrap MLE $\hat{F}_n$ and SMLE $\tilde{F}_{nh}^*$ are estimated based on the $(T_i, \Delta_i^*), i = 1, \ldots, n$. The smooth bootstrap $1 - \alpha$ intervals around the SMLE proposed in [14] are then constructed via (4.6), except that the SMLE $\tilde{F}_{nh}(t)$ in the definition of $W_{nh}^*(t)$ is replaced by the convolution SMLE given by

$$\int \mathbb{K}_h(t-u)\, d\tilde{F}_{nh}(u), \tag{4.7}$$

and that the variance estimate in the bootstrap sample is given by

$$\frac{1}{n^2} \sum_{i=1}^{n} K_h(t - T_i)^2 \left(\Delta_i^* - \hat{F}_n(T_i)\right)^2.$$

The convolution SMLE corresponds to the extra level of smoothing introduced by the smooth bootstrap procedure and is hence not required for the nonparametric bootstrap. The smooth bootstrap CIs of [32] around the MLE are given by

$$\left[F_n(t) - Z_{1-\alpha/2}^*(t), F_n(t) - Z_{\alpha/2}^*(t)\right],$$

where $Z_\alpha^*(t)$ is the αth quantile of B values of $\hat{F}_n(t) - \tilde{F}_{nh}(t)$, where again the extra level of smoothing is introduced (since one subtracts $\tilde{F}_{nh}$ and not F_n) to justify the smooth bootstrap procedure.

The performance of the SMLE-based CIs is comparable. The bootstrap intervals based on the classical bootstrap procedure avoid however calculation of the convolution SMLE defined in (4.7). The CIs in (b) and (c) have similar coverage proportions in the middle of the interval $[0, 2]$ but have a worse behavior near the boundaries of the interval compared to the SMLE-based intervals.

Figure 3(a) shows the average length of both bootstrap intervals around the SMLE in comparison with the average length of the likelihood ratio CIs of [4] and the smooth MLE-based CIs of [32]. The latter intervals are constructed around the MLE F_n instead of the SMLE $\tilde{F}_{nh}$. The length of the MLE-based intervals is larger than the length of the SMLE-based intervals due to the fact that the MLE converges at the slower rate $n^{1/3}$.

Instead of constructing the Studentized bootstrap intervals where the quantiles of the limiting distribution of the SMLE are derived from the bootstrap distribution, one can alternatively consider Wald-type confidence intervals using the quantiles of the normal distribution and an estimate of the asymptotic variance.

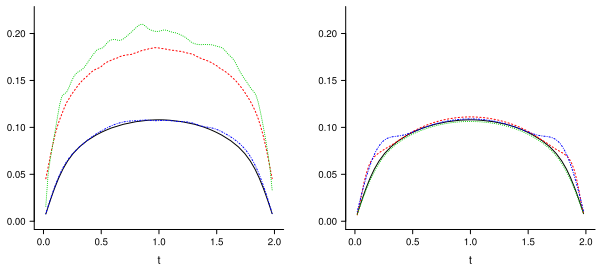


FIG 3. *Uniform samples: average length of the bootstrap CIs defined in (4.6) (black, solid) and (a) the smooth bootstrap CIs (blue, dashed-dotted) of [14], the likelihood ratio CIs of [4] (red, dashed) and the smooth MLE-based CIs of [32] (green,dotted); and (b) Wald-type CIs using the first estimate* $\hat{\sigma}_{1,nh}$ *(red,dashed), the second estimate* $\hat{\sigma}_{2,nh}$ *(blue,dashed-dotted) and the third estimate* $\hat{\sigma}_{3,nh}$ *(green,dotted).* $n = 1,000$, $N = 5,000$, $B = 1,000$ *and* $h = 2n^{-1/5}$.

We compare three different estimates $\hat{\sigma}_{nh}(t)$ for $\sigma(t)$ defined in (4.2) and construct CIs given by

$$[\tilde{F}_{nh}(t) - z_{1-\alpha/2}(n^{-2/5}\hat{\sigma}_{nh}(t)) - \beta(t)n^{-2/5}; \quad \tilde{F}_{nh}(t) - z_{\alpha/2}(n^{-2/5}\hat{\sigma}_{nh}(t)) - \beta(t)n^{-2/5}], \tag{4.8}$$

where z_α is the αth quantile of the standard normal distribution. In this simulation study $\beta(t)$ defined in (4.2) is zero. The effect of $\beta(t)$ on the behavior of the intervals will be discussed in the second simulation study below. A first estimate for $\hat{\sigma}_{nh}(t)$ is given by

$$\hat{\sigma}^2_{1,nh}(t) = \frac{F_n(t)\{1 - F_n(t)\}}{cg_{nh}(t)} \int K(u)^2 \, du, \tag{4.9}$$

where g_{nh} is a classical kernel estimate for the density g of the observation time $T \sim U(0, 2)$, using again the triweight kernel with bandwidth $h = 2n^{-1/5}$. A second estimate for $\sigma(t)$ is inspired by the fact that the SMLE is asymptotically equivalent to the toy-estimator defined in (4.5), which has a sample variance

$$s^2_{nh}(t) = \frac{1}{n^2} \sum_{i=1}^{n} \frac{K_h(t - T_i)^2 \left(\Delta_i - F_0(T_i)\right)^2}{g(T_i)^2}. \tag{4.10}$$

This suggests taking the second estimate $n^{-2/5}\hat{\sigma}_{2,nh}(t)$ equal to the root of (4.10) where F_0 is replaced by the MLE F_n and g is replaced by the kernel density estimate g_{nh}.

Contrary to the bootstrap procedure for constructing CIs defined in (4.6), both estimates $\hat{\sigma}_{1,nh}(t)$ and $\hat{\sigma}_{2,nh}(t)$ require estimating the density g. A bootstrap based estimate for the variance, avoiding estimating g, is finally given

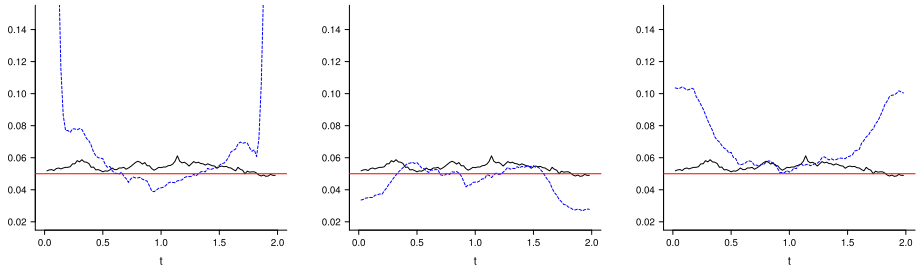


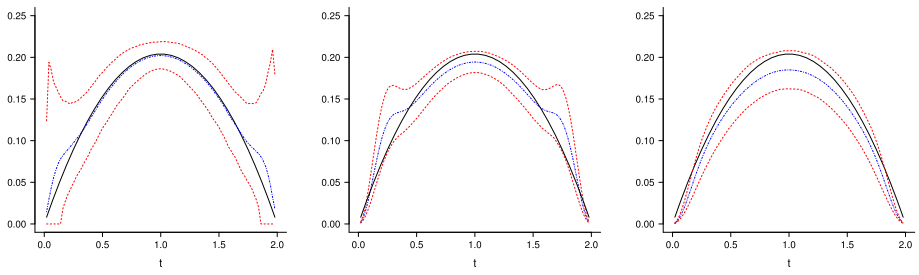
FIG 4. *Uniform samples: Proportion of times that* $F_0(t)$, $t = 0.02, 0.04, \ldots$ *is not in the* 95% *CIs for the bootstrap CIs defined in (4.6) (black, solid) and Wald-type CIs defined in (4.8) using (a) the first estimate* $\hat{\sigma}^2_{1,nh}$ *(blue, dashed), (b) the second estimate* $\hat{\sigma}^2_{2,nh}$ *(blue, dashed) and (c) the third estimate* $\hat{\sigma}^2_{3,nh}$ *(blue,dotted).* $n = 1,000$, $N = 5,000$, $B = 1,000$ *and* $h = 2n^{-1/5}$.

by

$$\hat{\sigma}^2_{3,nh}(t) = \frac{1}{B}\sum_{b=1}^{B}\left(\tilde{F}^{*,b}_{nh}(t) - \tilde{F}_{nh}(t)\right)^2, \tag{4.11}$$

where $\tilde{F}^{*,b}_{nh}(t)$ is the SMLE in the bth bootstrap run. Figure 4 compares the coverage proportions between the bootstrap CIs in (4.6) with the Wald-type CIs in (4.8) using the three different variance estimates described above. Pointwise confidence bands for the variance estimates are illustrated in Figure 5. The curves show the average variance estimate and the 5% and 95% empirical quantiles of the variance estimates at points $t = 0.02, 0.04, \ldots, 2$. The best results for the Wald-type CIs are obtained with the second variance estimate $\hat{\sigma}^2_{2,nh}(t)$ but the coverage proportions and average lengths (shown in Figure 3(b)) are inferior to the results obtained with the bootstrap CIs in (4.6). Estimating the density g in $\hat{\sigma}_{1,nh}(t)$ and $\hat{\sigma}_{2,nh}(t)$ requires an additional bandwidth selection, whereas the estimate $\hat{\sigma}_{3,nh}(t)$ is straightforward to obtain and does not depend on an estimate of g. The variance of the first estimate $\hat{\sigma}^2_{1,nh}(t)$ is larger than the variance of the second and third variance estimates $\hat{\sigma}^2_{2,nh}(t)$ and $\hat{\sigma}^2_{3,nh}(t)$, especially near the boundaries of the support.

Although we have proven validity of the nonparametric bootstrap for constructing pointwise CIs around the SMLE, the performance of the CIs is often influenced by several other aspects that are not specifically due to the nonparametric bootstrap algorithm. In what follows we describe some of these issues further and analyze the problems that can arise in the construction of the CIs. In a second simulation study we investigate the bias effect. Estimation of the bias defined in (4.2) is known to be a rather difficult task since it requires estimating the derivative f_0' of the density f_0 under current status data. Sufficiently accurate estimates of the bias are hard to obtain by direct estimation of f_0'. Besides estimating the derivative directly we therefore also explore the effect of the bandwidth choice on the performance of the pointwise CIs. We first describe a



(a) (b) (c)

FIG 5. *Uniform samples: True variance σ^2 (black, solid), mean estimate (blue, dashed-dotted) and the 5% and 95% empirical quantiles of the estimates (red, dashed) for (a) the first estimate $\hat{\sigma}^2_{1,nh}$, (b) the second estimate $\hat{\sigma}^2_{2,nh}$ and (c) the third estimate $\hat{\sigma}^2_{3,nh}$. $n = 1,000$, $N = 5,000$, $B = 1,000$ and $h = 2n^{-1/5}$.*

procedure for selecting the bandwidth and next examine the quality of (a) a bootstrap based estimate of the bias, (b) a direct estimate of the bias using an estimate of f_0' and (c) undersmoothing the bandwidth on the reduction of the bias effect present in the pointwise CIs.

4.1.2. Bandwidth selection

In the previous simulation study, we considered taking the bandwidth equal to $h = 2n^{-1/5}$, where the factor 2 is based on the size of the support $[0, 2]$ of the density f_0. This choice gave satisfactory results on the performance of the CIs discussed above. A bad choice of the bandwidth can however seriously affect the performance of the SMLE. It is therefore advisable to use an approach that selects the bandwidth with respect to some optimization criteria. We apply the method proposed in [20] to select the bandwidth which uses bootstrap subsamples of smaller size from the original sample to estimate the pointwise mean squared error (MSE) of the SMLE. The method works as follows: to obtain an approximation to the optimal bandwidth minimizing the pointwise MSE, we generate B bootstrap subsamples of size $m = o(n)$ from the original sample using the subsampling principle and take $c_{t,opt}$ as the minimizer of

$$\widehat{MSE}(c) = B^{-1} \sum_{b=1}^{B} \left\{ \tilde{F}^{*,b}_{m,cm^{-1/5}}(t) - \tilde{F}_{n,c_0 n^{-1/5}}(t) \right\}^2, \tag{4.12}$$

where $\tilde{F}_{n,c_0 n^{-1/5}}$ is the SMLE in the original sample of size n using an initial bandwidth $c_0 n^{-1/5}$ for some constant c_0. The bandwidth used for estimating the SMLE is next given by $h = c_{t,opt} n^{-1/5}$ where $c_{t,opt}$ minimizes $\widehat{MSE}(c)$ as a function of c. In the simulation study below we show the results for $m = 50$ when generating subsamples from a sample of size $n = 1,000$. Other subsample sizes $m = 30, 100$ were considered as well which resulted in similar optimal bandwidth choices. We used subsamples $m = 100$ resp. $m = 250$ when we generated data sets of size $n = 5,000$ resp. $n = 10,000$ from the model.

4.1.3. Simulation study 2: correcting the asymptotic bias

To investigate the effect of the bias on the construction of the pointwise CIs in (4.6), we consider a second simulation study where the event times are generated from a truncated exponential distribution on $[0, 2]$ and the censoring times are uniformly distributed on $[0, 2]$. The density of the event times is given by $f_0(t) = \exp(-t)/(1 - \exp(-2))1_{[0,2]}(t)$ and therefore the bias $\beta(t)$ defined in (4.2) will influence the performance of the CIs.

Figure 6 compares the proportion of times that $F_0(t)$ is not in the 95% bootstrap CIs for $t = 0.02, 0.04, \ldots, 2$ with the corresponding proportions in the bias corrected CIs given by

$$[\tilde{F}_{nh}(t) - Q^*_{1-\alpha/2}(t)\sqrt{S_{nh}(t)} - \beta(t)n^{-2/5}, \\ \tilde{F}_{nh}(t) - Q^*_{\alpha/2}(t)\sqrt{S_{nh}(t)} - \beta(t)n^{-2/5}], \quad (4.13)$$

where $Q^*_{1-\alpha/2}(t)$ and $S_{nh}(t)$ are defined above and where $\beta(t)$ is the true bias of the SMLE at timepoint t defined in (4.2). The bandwidth of the SMLE is selected by the procedure described in Section 4.1.2. The coverage proportions of the uncorrected CIs are clearly smaller than the nominal 95%-level at the left endpoint of the interval $[0, 2]$ in correspondence to the region where $\beta(t)$ is largest and correcting for the bias effect is needed to obtain good CIs. Figure 6 suggests that the coverage proportions of the intervals will be satisfying if the bias can be estimated sufficiently accurately. Estimation of the bias requires estimating the density f_0, which is a rather difficult task with current status

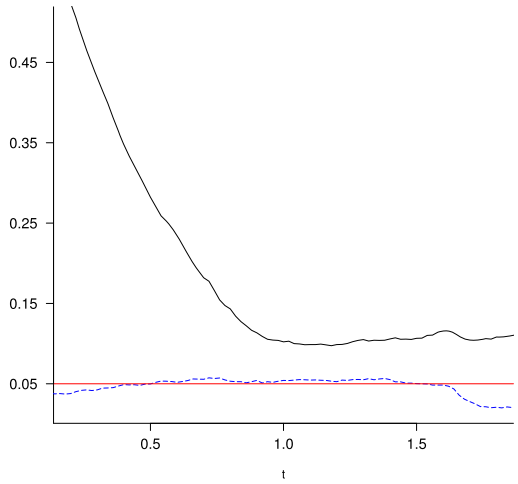


FIG 6. *Exponential samples: Proportion of times that $F_0(t)$, $t = 0.02, 0.04, \ldots$ is not in the 95% CIs for the bootstrap CIs defined in (4.6) (blue, dashed) and the bias corrected CIs defined in (4.13) (black, solid). $n = 1,000$, $N = 5,000$, $B = 1,000, m = 50$ and $h = c_{t,opt}n^{-1/5}$.*

data. A kernel based estimate of f_0' using the MLE F_n is given by

$$\tilde{f}'_{n\bar{h}}(t) = \bar{h}^{-2} \int K'\left((t-x)/\bar{h}\right)\, dF_n(x), \tag{4.14}$$

where the bandwidth $\bar{h} \sim n^{-1/9}$. In our experiments, we take the bandwidth of the estimate $\tilde{f}'_{n\bar{h}}(t)$ equal to $\bar{h} = \bar{c}_{t,opt} n^{-1/9}$ where $\bar{c}_{t,opt}$ is selected by the same bootstrap-MSE approach discussed in Section 4.1.2, but with the SMLE replaced by this derivative estimate. To obtain good estimates of f_0' near the boundaries of the support, we consider the boundary correction method explained in Section 9.2 of [17]. A direct estimator of the actual bias is then obtained by first replacing $f_0'(t)$ in (4.2) by the estimate $\tilde{f}'_{nh}(t)$ and next multiplying with $n^{-2/5}$, i.e. the order of the actual bias that has to be taken into account when constructing the CIs.

Similarly to the estimate of the pointwise MSE defined in (4.12), we can also construct a bootstrap method for estimating the bias by using the subsampling principle described in [20]. Our estimate $\widehat{Bias}(t)$ of the actual bias $\beta(t)n^{-2/5}$, is given by

$$\widehat{Bias}(t) = \left\{ B^{-1} \sum_{b=1}^{B} \left\{ \tilde{F}^{*,b}_{m,c_{t,opt}m^{-1/5}}(t) - \tilde{F}_{nc_0 n^{-1/5}}(t) \right\} \right\} \left(\frac{m}{n}\right)^{2/5}.$$

Figure 7 compares the average true bias effect $\beta(t)n^{-2/5}$ and the average bias estimates obtained by either the direct estimation approach or the bootstrap based bias estimate for sample sizes $n = 1000, 5000$ and $n = 10,000$. Note that, since the bandwidth constant $c_{t,opt}$ used for estimating the SMLE is different in each simulation run, the true bias (depending on $c_{t,opt}$, see (4.2)) in each run is also different and therefore the average true bias is shown in Figure 7. The actual size of the bias decreases with increasing sample size.

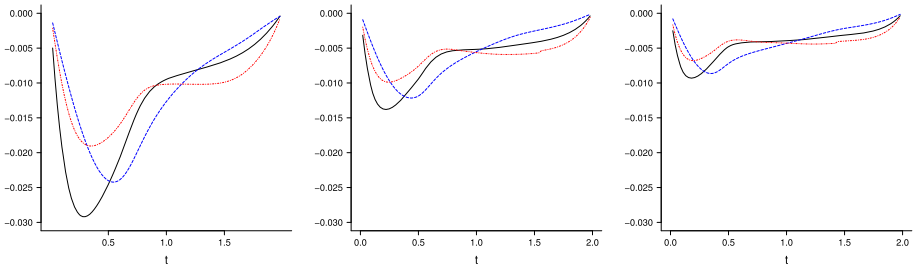


FIG 7. *Exponential samples: Average true bias (black solid) and average estimated bias for the bootstrap based estimate (blue, dashed) and the direct estimate (red, dashed-dotted) for samples (and subsamples) of size (a)* $n = 1,000, m = 50$*, (b)* $n = 5,000, m = 100$ *and (c)* $n = 10,000, m = 250$*.* $N = 5,000$*,* $B = 1,000$ *and* $h = c_{t,opt} n^{-1/5}$*.*

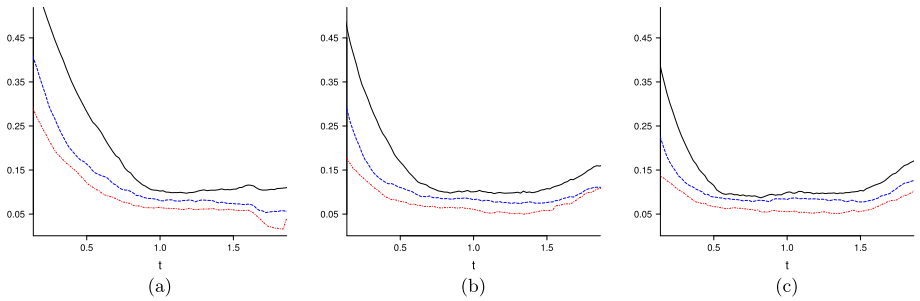


FIG 8. *Exponential samples without undersmoothing: Proportion of times that* $F_0(t)$, $t = 0.02, 0.04, \ldots$ *is not in the* 95% *CIs defined in (4.6) (black, solid) and the bias corrected CIs defined in (4.13) with bootstrap based bias estimate (blue, dashed) and direct bias estimate (red, dashed-dotted) for samples (and subsamples) of size (a)* $n = 1,000, m = 50$, *(b)* $n = 5,000, m = 100$ *and (c)* $n = 10,000, m = 250$. $N = 5,000$, $B = 1,000$ *and* $h = c_{t,opt} n^{-1/5}$.

The proportion of times that $F_0(t)$ is not in the 95% bootstrap CIs, shown in Figure 8, decreases if one corrects for the bias by one of the discussed bias estimates. The results for the direct bias estimate using the estimate $\tilde{f}'_{n\bar{h}}$ are slightly better than the results for the bootstrap estimate of $\beta(t)n^{-2/5}$. The coverage proportions are however still anti-conservative for points at the left end of the support. We also considered constructing the bias corrected CIs in the uniform model used in Section 4.1.1 where the actual bias is zero (results not shown). The results of the uncorrected CIs in (4.6) were slightly better and estimating the bias in this model has a somewhat negative effect on the coverage proportions of the pointwise CIs around the SMLE.

Similarly to the methods proposed in [14] we next investigate how the choice of the bandwidth can affect the coverage proportions and average length of our CIs. To this end, we consider the concept of undersmoothing proposed by [22] and take $c_{t,opt} n^{-1/4}$ as the bandwidth used in constructing the CIs defined in (4.6). The coverage proportions of the CIs for the exponential model, shown in Figure 9, illustrate that the performance of the CIs around the SMLE improve by undersmoothing. We also observed that if we considered a smaller bandwidth choice $h = (1/3)c_{t,opt} n^{-1/5}$, the coverage proportions even improve further and give satisfactory results in the left end point of the support. This illustrates that a smaller bandwidth choice can indeed correct for the bias in the CIs.

The results of the CIs in (4.6) in the uniform model with a bandwidth $h = c_{t,opt} n^{-1/4}$ or $h = (1/3)c_{t,opt} n^{-1/5}$ are in line with the results obtained with a bandwidth $h = c_{t,opt} n^{-1/5}$ and similar to the results shown in Figure 4. This shows that undersmoothing in a model without bias has no negative effect on the coverage proportions of our CIs.

By undersmoothing, the length of our SMLE-based CIs increases but the average length of the CIs remains remarkably smaller than the average length of the CIs around the MLE proposed by [4] and [32] (see Table 1).

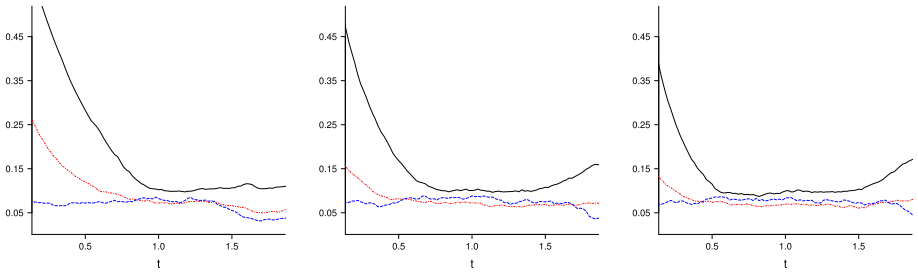


FIG 9. *Proportion of times that* $F_0(t)$, $t = 0.02, 0.04, \ldots$ *is not in the* 95% *CIs defined in (4.6) with* $h = c_{t,opt}n^{-1/5}$ *(black, solid),* $h = c_{t,opt}n^{-1/4}$ *(red, dashed-dotted) and* $h = (1/3)c_{t,opt}n^{-1/5}$ *(blue, dashed) for samples (and subsamples) of size (a)* $n = 1,000, m = 50$*, (b)* $n = 5,000, m = 100$ *and (c)* $n = 10,000, m = 250$*.* $N = 5,000$ *and* $B = 1,000$*.*

TABLE 1

Average length of the SMLE-based CIs defined in (4.6) for different bandwidth choices ($h \sim n^{-1/5}$ *and* $h \sim n^{-1/4}$) *and average length of the MLE-based CIs proposed by [4] and [32] at timepoints* $t = 0.5, 1, 1.5$*.*

	Uniform			Exponential		
Method	$t = 0.5$	$t = 1$	$t = 1.5$	$t = 0.5$	$t = 1$	$t = 1.5$
SMLE ($h \sim n^{-1/5}$)	0.064819	0.077020	0.064976	0.085540	0.087565	0.057716
SMLE ($h \sim n^{-1/4}$)	0.079671	0.092096	0.079757	0.085540	0.087565	0.057716
MLE ([4])	0.164767	0.184590	0.165699	0.204079	0.161122	0.104002
MLE ([32])	0.183982	0.202430	0.186452	0.225882	0.176159	0.118541

4.1.4. Rubella data

We also applied the bootstrap procedures to the Rubella data set described by [24]. The data set contains 230 observations on the prevalence of rubella in Austrian males. For the smooth bootstrap, CIs were calculated in [14] using the bandwidth $h = c_{t,opt}n^{-1/4}$. Figure 10 shows the CIs obtained with the nonparametric bootstrap and illustrates the applicability of our method in a real data example. For comparison, we also show the CIs obtained by the methods of [4] and [32]. The latter CIs were obtained by the Rcpp scripts in [13]. The nonparametric bootstrap SMLE-based CIs, including the data-driven bandwidth procedure, can be generated with the R package `curstatCI`.

4.2. The current status linear regression model

In the current status linear regression model we are interested in the estimation of the regression parameter β_0 based on observations $(T_1, X_1, \Delta_1 = 1_{\{Y_1 \le T_1\}}), \ldots (T_n, X_n, \Delta_n = 1_{\{Y_n \le T_n\}})$ from (T, X, Δ) where we assume that

$$Y_i = \beta_0' X_i + \varepsilon_i \qquad i = 1, 2, \ldots$$

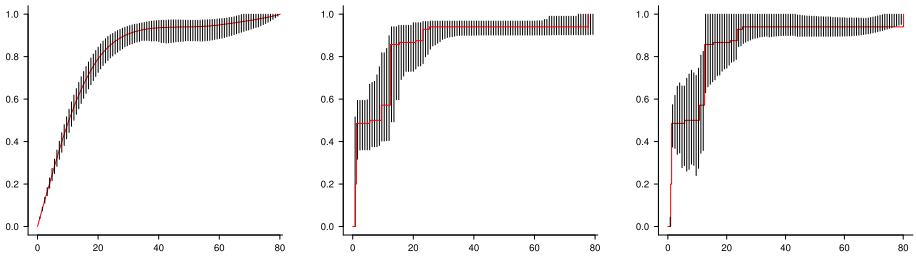


FIG 10. *Rubella data: (a) SMLE (red, solid) and CI defined in (4.6) based on* $n = 230$ *observations using* $B = 1,000$ *bootstrap samples and bandwidth* $h = c_{t,opt} n^{-1/4}$ *(*$c_{t,opt}$ *obtained by subsampling with* $B = 1,000$ *bootstrap samples of smaller size* $m = 50$*). (b) MLE (red, solid) and CI obtained by the method of Banerjee and Wellner [4], (c) MLE (red, solid) and CI obtained by the method of Sen and Xu [32] with* $B = 1,000$ *'smooth' bootstrap samples from the SMLE with bandwidth* $h = 80n^{-1/5}$.

with i.i.d. random error terms ε_i, independent of (T_i, X_i) with unknown distribution function F_0.

In [15] a simple score estimator β_n was introduced depending on the MLE $F_{n,\beta}$ for fixed β, defined as,

$$F_{n,\beta} \stackrel{def}{=} \arg\max_{F \in \mathcal{F}} \sum_{i=1}^{n} \left[\Delta_i \log F(T_i - \beta' X_i) + (1 - \Delta_i) \log\{1 - F(T_i - \beta' X_i)\}\right], \tag{4.15}$$

where $\mathcal{F} = \{F : \Re \mapsto [0,1] : \text{ F is a distribution function}\}$. The estimator β_n for β_0 is next defined as a zero-crossing (see Definition 4.1 in [15]) of

$$\sum_{F_{n,\beta}(T_i - \beta' X_i) \in [\epsilon, 1-\epsilon]} X_i \big\{\Delta_i - F_{n,\beta}(T_i - \beta' X_i)\big\}, \tag{4.16}$$

for some fixed truncation parameter $\epsilon \in (0, 1/2)$. It is proved in [15] that $\sqrt{n}\big\{\beta_n - \beta_0\big\}$ is asymptotically normal with mean zero and variance $V^{-1} W V^{-1}$ where

$$V = E_\epsilon\Big[f_0(T - \beta_0' X)\ \{X - E(X|T - \beta_0' X)\}\{X - E(X|T - \beta_0' X)\}'\Big],$$
$$W = E_\epsilon\Big[F_0(T - \beta_0' X)\{1 - F_0(T - \beta_0' X)\}\{X - E(X|T - \beta_0' X)\} \times \{X - E(X|T - \beta_0' X)\}'\Big],$$

where $E_\epsilon(w(T, X, \Delta)) = \int_{F_0(t - \beta_0' x) \in [\epsilon, 1-\epsilon]} w(t, x, \delta)\, dP(t, x, \delta)$ is the truncated expectation of $w(T, X, \Delta)$ for some deterministic function w and where P denotes the probability measure of (T, X, Δ).

A bootstrap version $\hat{\beta}_n$ based on a bootstrap sample from $\mathbb{P}_n$ is then defined as the zero-crossing of

$$\sum_{\hat{F}_{n,\beta}(T_i-\beta' X_i)\in[\epsilon,1-\epsilon]} M_{ni} X_i \{\Delta_i - \hat{F}_{n,\beta}(T_i - \beta' X_i)\} = 0, \tag{4.17}$$

where $\hat{F}_{n,\beta}$ is the MLE in the bootstrap sample. A straightforward extension of the results given in Section 3 shows that, as n tends to infinity,

$$E_{M|Z} \left| n^{-1/3} \left\{ \hat{F}_{n,\beta}(t - \beta' x) - F_\beta(t - \beta' x) \right\} \right|^p,$$

stays bounded in probability for all $(t, x) \in \{(t, x) : F_\beta(t - \beta' x) \in [\epsilon, 1 - \epsilon]\}$ and for all β in a neighborhood of β_0 where F_β is defined by

$$F_\beta(u) = P\{\Delta_i = 1 \mid T_i - \beta' X_i = u\} = \int F_0(u + (\beta - \beta_0)' x) f_{X|T-\beta' X}(x|u)\, dx. \tag{4.18}$$

The validity of the bootstrap method follows from the fact that, in probability, we have conditionally on the data $(T_1, X_1, \Delta_1), \ldots, (T_n, X_n, \Delta_n)$ that,

$$\begin{aligned} -\sqrt{n} V(\hat{\beta}_n - \beta_n) = \sqrt{n} \int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} & \{x - E(X|T - \beta_0' X = t - \beta_0' x)\} \\ & \cdot \{\delta - F_0(t - \beta_0' x)\}\, d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t, x, \delta) \\ & + o_{P_M}(1 + \sqrt{n}(\hat{\beta}_n - \beta_n)), \end{aligned} \tag{4.19}$$

where the dominant term in the right-hand side of the display above is normally distributed with mean zero and variance W conditional on $(T_1, X_1, \Delta_1), \ldots, (T_n, X_n, \Delta_n)$.

Remark 4.2. The nonparametric bootstrap is also valid for the second estimator of β_0 proposed in [15] based on a different score function involving the MLE $\hat{F}_{n,\beta}$ and the derivative of the SMLE $\tilde{F}_{nh,\beta}$ (constructed by the procedure described in Section 4.1).

To provide more insight into the finite sample behavior of the classical bootstrap estimators, we show in Tables 2 and 3 the results of two simulation studies for a one-dimensional regression model $Y = \beta_0 X + \varepsilon$. In the first simulation setting we take $\beta_0 = 0.5$ and consider Uniform(0,2) distributions for the variables T and X; for the distribution of the random error ε we take $f_0(e) = 384(e - 3/8)(5/8 - e)1_{[3/8,5/8]}(e)$. A picture of the density and distribution function of the random error in model 1 is shown in Figure 11. The first model is also analyzed in [15]. In the second simulation model T, X and ε are independently sampled from a standard normal distribution and $\beta_0 = 1$. A similar model was considered in [1].

With these simulations we want to point out that it is not necessary to use smoothing techniques for doing inferences in the current status linear regression model. We compare the simple score estimator (SSE) described above with

TABLE 2

Simulation model 1: mean, n times the variance and n times MSE. CP: coverage proportion of 95% CIs (Wald-type intervals based on a kernel variance estimate and classical bootstrap intervals) that contain the true parameter value $\beta_0 = 0.5$, AL: Average length of the CIs, for different samples sizes n based on $N = 1,000$ simulation runs and $B = 1,000$ bootstrap samples. $\epsilon = 0.001$. SSE = simple score estimator, MRCE = maximum rank correlation estimator and ESE = efficient score estimator.

Estimate	n	mean	$n\times$var	$n\times$MSE	Wald-type CI		Bootstrap CI	
					CP	AL	CP	AL
SSE	100	0.498943	0.310723	0.310968	0.978	0.265883	0.824	0.204163
	500	0.499717	0.220885	0.220925	0.982	0.097457	0.897	0.080317
	1000	0.500720	0.217415	0.217933	0.977	0.065837	0.924	0.055648
	5000	0.499993	0.195111	0.195112	0.977	0.027159	0.945	0.024423
MRCE	100	0.497996	0.308180	0.308582	0.979	0.268731	0.821	0.205522
	500	0.499761	0.251232	0.251260	0.978	0.098028	0.862	0.089143
	1000	0.500553	0.246388	0.246693	0.973	0.063990	0.911	0.053129
	5000	0.499876	0.208386	0.208462	0.965	0.027197	0.922	0.026987
ESE	100	0.500145	0.337755	0.337757	0.964	0.252687	0.824	0.223849
	500	0.499671	0.217428	0.217482	0.978	0.094390	0.896	0.080003
	1000	0.500742	0.207401	0.207953	0.973	0.063990	0.911	0.053129
	5000	0.500228	0.185614	0.185874	0.972	0.026396	0.904	0.022285

TABLE 3

Simulation model 2: mean, n times the variance and n times MSE. CP: coverage proportion of 95% CIs (Wald-type intervals based on a kernel variance estimate and classical bootstrap intervals) that contain the true parameter value $\beta_0 = 1$, AL: Average length of the CIs, for different samples sizes n based on $N = 1,000$ simulation runs and $B = 1,000$ bootstrap samples. $\epsilon = 0.001$. SSE = simple score estimator, MRCE = maximum rank correlation estimator and ESE = efficient score estimator.

Estimate	n	mean	$n\times$var	$n\times$MSE	Wald-type CI		Bootstrap CI	
					CP	AL	CP	AL
SSE	100	0.935732	4.525330	4.938096	0.922	1.000283	0.855	0.79952
	500	0.966217	4.676249	5.246881	0.926	0.399728	0.902	0.364210
	1000	0.977799	5.032432	5.525339	0.933	0.279928	0.914	0.262449
	5000	0.989466	4.580756	5.135616	0.945	0.124375	0.948	0.121388
MRCE	100	1.038510	8.500588	8.648890	0.925	1.125225	0.889	1.364034
	500	1.006050	6.443404	6.461690	0.932	0.429007	0.912	0.473787
	1000	1.002680	6.294143	6.301326	0.939	0.296537	0.903	0.320908
	5000	0.998502	5.160694	5.171915	0.962	0.129512	0.954	0.136487
ESE	100	0.974199	5.722576	5.789144	0.768	0.604649	0.827	0.910229
	500	0.998806	5.984291	5.985003	0.823	0.290297	0.902	0.430819
	1000	1.005545	6.032743	6.063495	0.841	0.214280	0.928	0.302124
	5000	1.002462	5.244373	5.274692	0.892	0.104281	0.951	0.131427

Han's maximum rank correlation estimator ([23], MRCE) and with the efficient score estimator (ESE) proposed in [15]. The asymptotic behavior of the MRCE for the current status model, also obtained without any smoothing techniques, is established in [1] where the author also proposes consistent kernel-based estimates of the asymptotic variance of the MRCE. We use these variance estimates to construct estimates for V, W and the almost (determined by the truncation parameter ϵ) efficient variance of the SSE. For more details about the variance estimation we refer to [1].

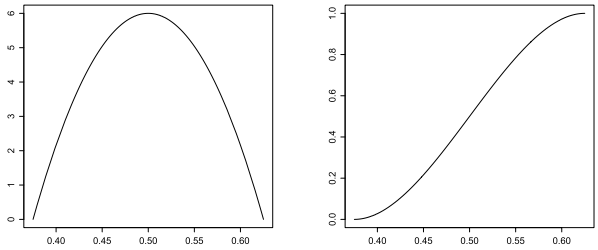


FIG 11. *The density f_0 (left panel) and distribution function F_0 (right panel) of the random error ε in simulation model 1.*

A summary of $N = 1,000$ simulation runs from models 1 and 2 for different sample sizes n is given in Tables 2 and 3. For each estimator, the mean, n times the variance and n times the mean squared-error (MSE) is given in columns 3-5. The asymptotic variance of the estimators equals 0.193612 for the SSE, 0.158699 for the ESE and 0.192857 for the MRCE in model 1 using truncation parameter $\epsilon = 0.001$. The corresponding asymptotic variances in model 2 equal 5.046413, 4.994988 and 5.35448 respectively. The asymptotic variance of the SSE without truncation (i.e. $\epsilon = 0$) equals the asymptotic variance of the MRCE in model 1. The efficient variances are 0.151706 in model 1 and 4.994987 in model 2. Note that the differences between the limiting variances for the different estimation methods are tiny and that the effect of the truncation parameter ϵ on the asymptotic behavior of the score estimators is small. Tables 2 and 3 show that n times the variance tends to converge to the asymptotic variance for all estimators. The ESE performs worse for small sample sizes and the results suggest to use the SSE for point estimation of the regression parameter β_0.

We constructed Wald-type CIs, similar to the intervals proposed in [1], using the asymptotic normal limiting distribution of the estimators and compared the coverage proportion and average length of these intervals with bootstrap CIs based on the nonparametric bootstrap described in this paper using $B = 1,000$ samples from the original data. For the MRCE, the validity of the classical bootstrap is proved in [33]. The Wald-type CIs remain anti-conservative for the ESE in model 2.

We observed (result not shown) that, in both models, the bias in estimating the efficient variance of the ESE remains larger than the bias of the asymptotic variance estimates for the SSE and the MRCE. Tables 2 and 3 show that the coverage proportion of the classical bootstrap CIs converges to the nominal 95%−level and the average length of the CIs obtained by resampling from the original data is smaller than the corresponding length of the Wald-type CIs. We also investigated the behavior of Studentized bootstrap CIs (results not shown) based on the variance estimate used in the construction of the Wald-type CIs, but no improvement was observed for the behavior of the bootstrap intervals.

Our results do not indicate better performances corresponding to smoothing techniques and therefore suggest that smoothing should not be the primary concern in inferences for the current status linear regression model. Note that the

Wald-type CIs are constructed using smoothing kernel estimation for the variance estimate and that the only results obtained without any smoothing are the bootstrap CIs for the SSE and the MRCE. It is noteworthy that the SSE tends to perform better than the MRCE, which is not based on a nuisance parameter that is not estimable at $\sqrt{n}-$rate. Based on these results, we recommend the use of the SSE in combination with the nonparametric bootstrap procedure for doing inference in the current status linear regression model.

5. Discussion

In this paper we studied the behavior of the nonparametric bootstrap in current status models. Asymptotic results show that, given the data, the L_2-distance between the bootstrap MLE $\hat{F}_n$ and the underlying distribution function F_0 is of order $n^{-1/3}$. This result is noteworthy given the fact that the nonparametric bootstrap is inconsistent for generating the distribution of the MLE. Despite this negative result, we show that it is still possible to use the MLE while doing inferences for certain functionals in the current status model. We illustrated the effectiveness of this result by constructing pointwise confidence intervals around the SMLE and proved the validity of interval estimation in the current status linear regression model.

The result is applicable to several other nonparametric estimators depending on a cube-root n convergence class. Because of its connection with the MLE, applications of the nonparametric bootstrap involving the Grenander estimator, such as the smoothed Grenander estimator used in [7] or the goodness-of-fit tests described in [8], are worthy of study in further research.

Extensions to semiparametric models, where one considers bootstrapping a finite dimensional parameter, are also possible such as the score estimator for the semiparametric monotone single index model proposed by [3], which is similar to the current status linear regression estimator discussed in Section 4.2. A general bootstrap consistency result for semiparametric M-estimators is derived in [5]. However, if computations are in first instance based on nonparametric maximum likelihood estimators or least squares estimators of the infinite dimensional parameter, fixing temporarily the finite-dimensional parameter, the use of local smooth functional theory is needed, where the remainder terms involving the cube-root-n M-estimator of the nuisance parameter are shown to be negligible by an application of a result of the type (3.1). The treatment of the remainder terms in this local smooth functional theory is a highly non-trivial matter. On the other hand, in [5], this negligibility is assumed to hold by their condition SB3.

Furthermore, the results in [5] hold for a class of exchangeable bootstrap weights of which the multinomial weights considered in this paper are a special case. Although we did not investigate this in the present paper, extensions of our nonparametric bootstrap results to the more general bootstrap resampling schemes seem possible as well.

Another interesting extension of this research is the construction of confidence bands for the distribution instead of the currently proposed pointwise confidence

intervals. Note that our main result (3.2) does not imply:

$$E\left\{\sup_{t\in[0,R]} n^{1/3}\left|\hat{F}_n(t)-F_0(t)\right|\,\middle|\, Z_1,\ldots,Z_n\right\}=O_p(1). \tag{5.1}$$

A bound on $\sup_{t\in[0,R]} n^{1/3}\left|\hat{F}_n(t)-F_0(t)\right|$ which no doubt would contain logarithmic factors, would be needed for confidence bands instead of our pointwise confidence intervals. The idea is that the process $t\mapsto n^{1/3}\{\hat{F}_n(t)-F_0(t)\}$ will fall apart into asymptotically independent pieces, and that we therefore expect Gumbel-type distributions to enter, via the maximum of independent random variables. The theory for this still has to be developed, however. What struck us in the present simulation studies is how comparatively well the global behavior of our pointwise confidence intervals still was, indicating that the extra logarithmic factors do not have such a very large impact.

Probably results similar to those presented in the current paper will follow for the more challenging interval censoring, type II models where the development of the local limit theory for the MLE has not yet been settled. It is reasonable to believe that the nonparametric bootstrap also allows for inferences with the maximum smoothed likelihood estimator studied in [12].

6. Appendix

6.1. Proof of Lemma 3.1

Before proving Lemma 3.1 we provide two technical lemmas.

Lemma 6.1. *Let $\alpha>0$. There exist constants $K_1,K_2>0$ such that, for each $j\geq 1$, $j\in\mathbb{N}$,*

$$\begin{aligned}&P_{M|Z}\Bigg\{\exists y\in\left[(j-1)n^{-1/3},jn^{-1/3}\right):\\&\qquad\left|\int_{u\in(U(a),U(a)+y]}\{\delta-F_0(u)\}\,d\big(\hat{\mathbb{P}}_n-\mathbb{P}_n\big)(u,\delta)\right|\geq\alpha(j-1)^2n^{-2/3}\Bigg\}\\&\leq K_1\exp\left\{-K_2(j-1)^{3/2}\right\},\end{aligned} \tag{6.1}$$

in probability.

Likewise, there exist constants $K_1,K_2>0$ such that, for each $j\geq 1$, $j\in\mathbb{N}$,

$$\begin{aligned}&P_{M|Z}\Bigg\{\exists y\in\left[-jn^{-1/3},-(j-1)n^{-1/3}\right):\\&\qquad\left|\int_{u\in(U(a)+y,U(a)]}\{\delta-F_0(u)\}\,d\big(\hat{\mathbb{P}}_n-\mathbb{P}_n\big)(u,\delta)\right|\geq\alpha(j-1)^2n^{-2/3}\Bigg\}\end{aligned}$$

$$\leq K_1 \exp\left\{-K_2(j-1)^{3/2}\right\}, \tag{6.2}$$

in probability.

Proof. We only prove (6.1), since the proof of (6.2) is similar. Let $\mathcal{F}_t$ be the (Vapnik-Cervonenkis) class of functions

$$\mathcal{F}_t = \left\{(\delta - F_0(v))1_{(U(a),U(a)+u]}(v) : u \in [0,t],\, \delta \in \{0,1\}\right\},$$

with envelope

$$F_t(v,\delta) = 1_{(U(a),U(a)+t]}(v), \qquad v \in [0,t].$$

To prove (6.1), we use that an exponential tail bound can be derived from a bounded Orlicz norm $\|\cdot\|_{P,\psi}$, i.e., when taking $\psi_1(x) = \exp(x) - 1$, for $x \geq 0$, we get, for $x > 0$ the inequality

$$P(|X| > x) \leq 2\exp\left\{-x/\|X\|_{P,\psi_1}\right\}, \tag{6.3}$$

where

$$\|X\|_{P,\psi_1} = \inf\left\{C > 0 : E\psi_1\left(\frac{|X|}{C}\right) \leq 1\right\}.$$

Using the second statement of Theorem 2.14.5 in [34], with $p = 1$, we get, the following inequality:

$$\begin{aligned}
&\left\|\left\|\sqrt{n}\left(\hat{\mathbb{P}}_n - \mathbb{P}_n\right)\right\|^*_{\mathcal{F}_t}\right\|_{\mathbb{P}_n,\psi_1} \\
&\quad\lesssim \left\|\left\|\sqrt{n}\left(\hat{\mathbb{P}}_n - \mathbb{P}_n\right)\right\|^*_{\mathcal{F}_t}\right\|_{\mathbb{P}_n,1} + n^{-1/2}\{1+\log n\}\|F_t\|_{\mathbb{P}_n,\psi_1},
\end{aligned} \tag{6.4}$$

where $\|\cdot\|^*_{\mathcal{F}_t}$ denotes the so-called measurable majorant of $\|\cdot\|_{\mathcal{F}_t}$ (see [34]). (Note that we use temporarily the "*" notation which is used for bootstrap variables in the rest of the paper.)

Furthermore, we have by the rightmost inequality of Theorem 2.14.1 of [34] that

$$\left\|\left\|\sqrt{n}\left(\hat{\mathbb{P}}_n - \mathbb{P}_n\right)\right\|^*_{\mathcal{F}_t}\right\|_{\mathbb{P}_n,1} \lesssim J\left(1,\mathcal{F}_t\right)\|F_t\|_{\mathbb{P}_n,2},$$

where $J(\delta,\mathcal{F}_t)$ is defined by

$$J(\delta,\mathcal{F}_t) = \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon\|F\|_{Q,2},\mathcal{F}_t,L_2(Q)\right)}\,d\varepsilon,$$

and where the supremum is over all discrete probability measure Q with $\|F_t\|_{Q,2} > 0$. Since $\mathcal{F}_t \subset \mathcal{F}_{R-U(a)}$ for all $t \in [0, R-U(a)]$, and since $\mathcal{F}_{R-U(a)}$ is a Vapnik-Cervonenkis class, $J(\delta,\mathcal{F}_t)$ is bounded by a fixed constant for all $t \in [0, R-U(a)]$, and we get:

$$\left\|\left\|\sqrt{n}\left(\hat{\mathbb{P}}_n - \mathbb{P}_n\right)\right\|^*_{\mathcal{F}_t}\right\|_{\mathbb{P}_n,1} \lesssim \|F_t\|_{\mathbb{P}_n,2},$$

uniformly for all $t \in [0, R - U(a)]$. Note that

$$\|F_t\|^2_{\mathbb{P}_n,2} = \int_{u\in U(a),U(a)+t]} d\mathbb{P}_n(u,\delta) = \int_{u\in U(a),U(a)+t]} d\mathbb{G}_n(u), \tag{6.5}$$

$t \in [U(a), R - U(a)]$. We next evaluate the second term on the right-hand side of (6.4). We have:

$$\int \psi_1\left(\frac{F_t(u,\delta)}{c}\right) d\mathbb{P}_n(u,\delta) = \left\{e^{1/c} - 1\right\} \int 1_{(U(a),U(a)+t]}(u)\, d\mathbb{G}_n(u),$$

and

$$\begin{aligned} &\left\{e^{1/c} - 1\right\} \int 1_{(U(a),U(a)+t]}(u)\, d\mathbb{G}_n(u) \le 1 \\ &\quad\iff c \ge \frac{1}{\log\left\{1 + 1/\int_{u\in U(a),U(a)+t]} d\mathbb{G}_n(u)\right\}}. \end{aligned}$$

Thus (6.4) becomes, using (6.5),

$$\begin{aligned} &\left\| \left\| \sqrt{n}\left(\hat{\mathbb{P}}_n - \mathbb{P}_n\right) \right\|^*_{\mathcal{F}_t} \right\|_{\mathbb{P}_n,\psi_1} \\ &\le c_1 \left\{ \int_{u\in U(a),U(a)+t]} d\mathbb{G}_n(u) \right\}^{1/2} + \frac{1 + \log n}{n^{1/2} \log\left\{1 + 1/\int_{u\in U(a),U(a)+t]} d\mathbb{G}_n(u)\right\}}, \end{aligned} \tag{6.6}$$

for a constant $c_1 > 0$. If $t \ge Kn^{-1/3}$ we get for the second term in probability,

$$\frac{1 + \log n}{n^{1/2} \log\left\{1 + 1/\int_{u\in U(a),U(a)+t]} d\mathbb{G}_n(u)\right\}} \ll c_1 \left\{ \int_{u\in U(a),U(a)+t]} d\mathbb{G}_n(u) \right\}^{1/2}.$$

We have:

$$\begin{aligned} &\int_{u\in[U(a),U(a)+t]} d\mathbb{G}_n(u) \\ &= \int_{u\in[U(a),U(a)+t]} dG(u) + \int_{u\in[U(a),U(a)+t]} d\big(\mathbb{G}_n - G\big)(u) \\ &= \int_{u\in[U(a),U(a)+t]} dG(u) + O_p\left(n^{-1/2}\right) = O(t) + O_p\left(n^{-1/2}\right) \\ &= O(t) + O_{P_M}\left(n^{-1/2}\right), \end{aligned}$$

in probability (since a term defined only on the probability space $(\mathcal{X}, \mathcal{A}, P)$ of order $O_p(1)$ is also of order $O_{P_M}(1)$ in probability). So we obtain, for $j \ge K$ in

probability, conditioning on $(T_1, \Delta_1), (T_2, \Delta_2), \ldots$ using the inequality on Orlicz norms on p. 96 or 239 of [34]:

$$
\begin{aligned}
&P_{M|Z}\Bigg\{\exists y \in \big[(j-1)n^{-1/3}, jn^{-1/3}\big) : \\
&\qquad \left|\int_{u\in(U(a),U(a)+y]} \{\delta - F_0(u)\}\, d\big(\hat{\mathbb{P}}_n - \mathbb{P}_n\big)(u,\delta)\right| \geq \alpha(j-1)^2 n^{-2/3}\Bigg\} \\
&= P_{M|Z}\Bigg\{\exists y \in \big[(j-1)n^{-1/3}, jn^{-1/3}\big) : \\
&\qquad \sqrt{n}\left|\int_{u\in(U(a),U(a)+y]} \{\delta - F_0(u)\}\, d\big(\hat{\mathbb{P}}_n - \mathbb{P}_n\big)(u,\delta)\right| \geq \alpha(j-1)^2 n^{-1/6}\Bigg\} \\
&\leq 2\exp\left\{-m(j-1)^2 n^{-1/6} \Big/ \left\| \left\| \sqrt{n}\left(\hat{\mathbb{P}}_n - \mathbb{P}_n\right)\right\|^*_{\mathcal{F}_{jn^{-1/3}}} \right\|_{\mathbb{P}_n,\psi_1}\right\} \\
&\leq 2\exp\left\{-c_2 m(j-1)^{3/2}\right\},
\end{aligned}
$$

for some $c_2 > 0$. This proves the statement. □

Lemma 6.2. *For each $\varepsilon > 0$ and $x \in [0, R - U(a)]$,*

$$
\left|\int_{u\in(U(a),U(a)+x]} \{\delta - F_0(u)\}\, d\big(\mathbb{P}_n - P\big)(u,\delta)\right| \leq \varepsilon x^2 + O_p\left(n^{-2/3}\right).
$$

Proof. As in the proof of Lemma 6.1, we consider the Vapnik-Cervonenkis collection of functions:

$$
\mathcal{F}_t = \left\{(\delta - F_0(v)) 1_{(U(a),U(a)+u]}(v) : u \in [0,t],\, \delta \in \{0,1\}\right\},
$$

with envelope

$$
F_t(v,\delta) = 1_{(U(a),U(a)+t]}(v), \qquad v \in [0,t].
$$

We have, using Theorem 2.14.1 of [34]:

$$
E_X\left\{\sup_{f\in\mathcal{F}_t} |\mathbb{P}_n - P|\,(f)\right\}^2 \leq K n^{-1} \|F_t\|^2_{P,2}, \tag{6.7}
$$

for some $K > 0$. Since,

$$
\|F_t\|^2_{P,2} = \int_{u\in U(a),U(a)+t]} dP(u,\delta) = \int_{u\in U(a),U(a)+t]} dG(u) = O(t),
$$

for $t \in [U(a), R - U(a)]$, we get, by Markov's inequality,

$$P\left\{ n^{2/3} \left| \int_{u \in (U(a), U(a)+jn^{-1/3}]} \{\delta - F_0(u)\}\, d\big(\mathbb{P}_n - P\big)(u,\delta) \right| > A + \varepsilon(j-1)^2 \right\}$$
$$\leq Kj / \left\{ A + \varepsilon(j-1)^2 \right\}^2 .$$

The result now easily follows, see, e.g., [25]. p. 201. □

As a consequence of Lemma 6.1 and Lemma 6.2 we get the following result.

Lemma 6.3. *Let $\hat{V}_n$ and $\hat{\tilde{V}}_n$ be defined by*

$$\hat{V}_n(t) = \int_{u \in [0,t]} \delta\, d\hat{\mathbb{P}}_n(u,\delta), \qquad \hat{\tilde{V}}_n(t) = \int_{u \in [0,t]} F_0(u)\, d\hat{\mathbb{G}}_n(u), \qquad t \in [0, R]. \tag{6.8}$$

where the process $\hat{\mathbb{G}}_n$ is defined in (3.3), and let $\hat{D}_n = \hat{V}_n - \hat{\tilde{V}}_n$. Then there exist constants $K_1, K_2 > 0$ such that, for each $j \geq 1$, $j \in \mathbb{N}$,

$$P_{M|Z}\left\{ \exists y \in \left[(j-1)n^{-1/3}, jn^{-1/3}\right) : \hat{D}_n(U(a)+y) - \hat{D}_n(U(a)) \right.$$
$$\left. \leq -\int_{U(a)}^{U(a)+y} \left\{F_0(u) - F_0(U(a))\right\} d\hat{\mathbb{G}}_n(u) \right\}$$
$$\leq K_1 \exp\left\{ -K_2 (j-1)^{3/2} \right\}, \tag{6.9}$$

in probability. Likewise, there exist constants $K_1, K_2 > 0$ such that, for each $j \geq 1$, $j \in \mathbb{N}$,

$$P_{M|Z}\left\{ \exists y \in \left[-jn^{-1/3}, -(j-1)n^{-1/3}\right) : \hat{D}_n(U(a)+y) - \hat{D}_n(U(a)) \right.$$
$$\left. \leq -\int_{U(a)+y}^{U(a)} \left\{F_0(u) - F_0(U(a))\right\} d\hat{\mathbb{G}}_n(u) \right\}$$
$$\leq K_1 \exp\left\{ -K_2 (j-1)^{3/2} \right\}, \tag{6.10}$$

in probability.

Proof. We again only prove (6.1), since the proof of (6.2) is similar. First note:

$$P_{M|Z}\left\{ \exists y \in \left[(j-1)n^{-1/3}, jn^{-1/3}\right) : \hat{D}_n(U(a)+y) - \hat{D}_n(U(a)) \right.$$
$$\left. \leq -\int_{U(a)}^{U(a)+y} \left\{F_0(u) - F_0(U(a))\right\} d\hat{\mathbb{G}}_n(u) \right\}$$
$$\leq P_{M|Z}\left\{ \exists y \in \left[(j-1)n^{-1/3}, jn^{-1/3}\right) : \left| \hat{D}_n(U(a)+y) - \hat{D}_n(U(a)) \right| \right.$$
$$\left. \geq \int_{U(a)}^{U(a)+y} \left\{F_0(u) - F_0(U(a))\right\} d\hat{\mathbb{G}}_n(u) \right\}.$$

Furthermore:

$$\begin{aligned}
&\int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, d\hat{\mathbb{G}}_n(u) \\
&\qquad = \int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, d\mathbb{G}_n(u) \\
&\qquad\quad + \int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, d\bigl(\hat{\mathbb{G}}_n - \mathbb{G}_n\bigr)(u) \\
&\qquad = \int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, dG(u) \\
&\qquad\quad + \int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, d\bigl(\mathbb{G}_n - G\bigr)(u) \\
&\qquad\quad + \int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, d\bigl(\hat{\mathbb{G}}_n - \mathbb{G}_n\bigr)(u), \qquad (6.11)
\end{aligned}$$

and for the dominant term on the right-hand side we get:

$$\begin{aligned}
&\int_{U(a)}^{U(a)+y} \bigl\{F_0(u) - F_0(U(a))\bigr\}\, dG(u) \geq m_0 \int_{U(a)}^{U(a)+y} \{u - U(a)\}\, dG(u) \\
&\geq m_0 m_1 \int_{U(a)}^{U(a)+y} \{u - U(a)\}\, du = \tfrac{1}{2} m_0 m_1 \{y - U(a)\}^2,
\end{aligned}$$

where $m_0 = \inf_{u\in[U(a),R]} f_0(u)$ and $m_1 = \inf_{u\in[U(a),R]} g(u)$. We therefore consider the probability:

$$P_{M|Z}\Bigl\{\exists y \in \bigl[(j-1)n^{-1/3}, jn^{-1/3}\bigr) : \Bigl|\hat{D}_n(U(a)+y) - \hat{D}_n(U(a))\Bigr| \geq m(j-1)^2 n^{-2/3}\Bigr\}. \qquad (6.12)$$

where

$$m = \tfrac{1}{2} \min\left\{ \inf_{u\in[t_0,R]} f_0(u), \inf_{u\in[t_0,R]} g(u) \right\}.$$

We also have:

$$\begin{aligned}
&\hat{D}_n(U(a)+y) - \hat{D}_n(U(a)) = \int_{u\in(U(a),U(a)+y]} \{\delta - F_0(u)\}\, d\hat{\mathbb{P}}_n(u,\delta) \\
&= \int_{u\in(U(a),U(a)+y]} \{\delta - F_0(u)\}\, d\bigl(\hat{\mathbb{P}}_n - P\bigr)(u,\delta) \\
&= \int_{u\in(U(a),U(a)+y]} \{\delta - F_0(u)\}\, d\bigl(\hat{\mathbb{P}}_n - \mathbb{P}_n\bigr)(u,\delta) \\
&\quad + \int_{u\in(U(a),U(a)+y]} \{\delta - F_0(u)\}\, d\bigl(\mathbb{P}_n - P\bigr)(u,\delta).
\end{aligned}$$

By Lemma 6.2, we may assume that for $x \in [0, R - U(a)]$,

$$\left| \int_{u \in (U(a), U(a)+x]} \{\delta - F_0(u)\} \, d\big(\mathbb{P}_n - P\big)(u, \delta) \right| \leq \varepsilon x^2 + K n^{-2/3}, \tag{6.13}$$

for some $K > 0$ and $0 < \varepsilon < m/2$. Considering sequences $X = (T_1, \Delta_1)$, $(T_2, \Delta_2) \ldots$, satisfying (6.13), we get:

$$\begin{aligned}
&P_{M|Z}\Big\{\exists y \in \big[(j-1)n^{-1/3}, jn^{-1/3}\big) : \Big|\hat{D}_n(U(a)+y) - \hat{D}_n(U(a))\Big| \\
&\qquad\qquad \geq m(j-1)^2 n^{-2/3}\Big\} \\
&\leq P_{M|Z}\Bigg\{\exists y \in \big[(j-1)n^{-1/3}, jn^{-1/3}\big) : \\
&\qquad \left| \int_{u \in (U(a), U(a)+y]} \{\delta - F_0(u)\} \, d\big(\hat{\mathbb{P}}_n - \mathbb{P}_n\big)(u, \delta) \right| \geq \tfrac{1}{2} m(j-1)^2 n^{-2/3} \Bigg\} \\
&\leq K_1 \exp\Big\{-K_2 (j-1)^{3/2}\Big\},
\end{aligned}$$

with probability tending to one, using Lemma 6.1. □

We now prove Lemma 3.1.

Proof of Lemma 3.1. Suppose that $n^{1/3}|\hat{U}_n(a) - U(a)| > x$ for some $x > 0$, then there exists a y such that, $n^{1/3}\,|y - U(a)| > x$ and $\hat{V}_n(y) - a\hat{\mathbb{G}}_n(y) \leq \hat{V}_n(U(a)) - a\hat{\mathbb{G}}_n(U(a))$. Hence,

$$\begin{aligned}
&P_{M|Z}\left\{n^{1/3}\left|\hat{U}_n(a) - U(a)\right| \geq x\right\} \\
&\leq P_{M|Z}\Bigg(\inf_{y - U(a) \geq n^{-1/3}x} \hat{D}_n(y) - \hat{D}_n(U(a)) \\
&\qquad\qquad \leq -\int_{U(a)}^{y} \big\{F_0(u) - F_0(U(a))\big\} \, d\hat{\mathbb{G}}_n(u)\Bigg) \\
&\leq \sum_{j=i}^{\infty} P_{M|Z}\Bigg(\exists y \in \big[(j-1)n^{-1/3}, jn^{-1/3}\big) : \hat{D}_n(U(a)+y) - \hat{D}_n(U(a)) \\
&\qquad\qquad \leq -\int_{U(a)}^{U(a)+y} \big\{F_0(u) - F_0(U(a))\big\} \, d\hat{\mathbb{G}}_n(u)\Bigg),
\end{aligned}$$

where $x \in [(i-1)n^{-1/3}, in^{-1/3}]$. By Lemma 6.3, this is bounded above by

$$\sum_{j=i}^{\infty} K_1 \exp\Big\{K_2 (j-1)^{3/2}\Big\}$$

$$= K_1 \exp\left\{-K_2(i-1)^{3/2}\right\} \sum_{j=i}^{\infty} \exp\left\{-K_2[(j-1)^{3/2} - (i-1)^{3/2}]\right\}$$
$$\leq K_1' \exp\left\{K_2'(i-1)^{3/2}\right\},$$

for constants $K_1, K_1', K_2, K_2' > 0$. □

6.2. *Proof of Lemma 4.1*

We introduce notations K_h and $\mathbb{K}_h$ to denote the scaled versions of K and $\mathbb{K}$ respectively:

$$K_h(u) = h^{-1}K(u/h) \qquad \text{and} \qquad \mathbb{K}_h(u) = \mathbb{K}(u/h).$$

Proof. Define the function

$$\psi_{t,h}(u) = \frac{K_h(t-u)}{g(u)}.$$

Denote the points of jump of the MLE $\hat{F}_n$ by $\hat{\tau}_1, \ldots, \hat{\tau}_m$ and define the piecewise constant function $\bar{\psi}_{t,h}$ with only jumps at $\hat{\tau}_1, \ldots, \hat{\tau}_m$ by

$$\bar{\psi}_{t,h}(u) = \begin{cases} \psi_{t,h}(\hat{\tau}_i), & \text{if } F_0(u) > \hat{F}_n(\hat{\tau}_i),\ u \in [\hat{\tau}_i, \hat{\tau}_{i+1}), \\ \psi_{t,h}(s), & \text{if } F_0(u) = \hat{F}_n(s), \text{ for some } s \in [\hat{\tau}_i, \hat{\tau}_{i+1}), \\ \psi_{t,h}(\hat{\tau}_{i+1}), & \text{if } \tilde{F}_0(u) < \hat{F}_n(\tau_i),\ u \in [\hat{\tau}_i, \hat{\tau}_{i+1}). \end{cases}$$

By the convex minorant interpretation of $\hat{F}_n$, we have

$$\int \bar{\psi}_{t,h}(u)(\delta - \hat{F}_n(u))d\hat{\mathbb{P}}_n(u,\delta) = 0,$$

(see the discussion of the SMLE in [17], p. 332).

We can write

$$\tilde{F}^*_{nh}(t) = \int \mathbb{K}_h(t-u)\, d\hat{F}_n(u)$$
$$= \int \mathbb{K}_h(t-u)\, d(\hat{F}_n - F_0)(u) + \int \mathbb{K}_h(t-u)\, dF_0(u)$$
$$= \int \psi_{t,h}(u) \left\{\hat{F}_n(u) - F_0(u)\right\} dG(u) + \int \mathbb{K}_h(t-u)\, dF_0(u)$$
$$= \int \psi_{t,h}(u) \left\{F_0(u) - \hat{F}_n(u)\right\} d(\hat{\mathbb{G}}_n - G)(u) + \int \psi_{t,h}(u) \left\{\delta - F_0(u)\right\} d\hat{\mathbb{P}}_n(u,\delta)$$
$$+ \int \left\{\psi_{t,h}(u) - \bar{\psi}_{t,h}(u)\right\} \left\{\hat{F}_n(u) - \delta\right\} d\hat{\mathbb{P}}_n(u,\delta) + \int \mathbb{K}_h(t-u)\, dF_0(u)$$

$$= \tilde{F}_{nh}^{(toy)*}(t) + \int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\hat{\mathbb{G}}_n - G)(u, \delta)$$
$$+ \int \left\{ \psi_{t,h}(u) - \bar{\psi}_{t,h}(u) \right\} \left\{ \hat{F}_n(u) - \delta \right\} d\hat{\mathbb{P}}_n(u, \delta)$$
$$= \tilde{F}_{nh}^{(toy)*}(t) + A_I + A_{II}.$$

We first evaluate A_I and show that this term is $o_{P_M}(n^{-2/5})$ in probability, we have:

$$A_I = \int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\hat{\mathbb{G}}_n - G)(u, \delta)$$
$$= \int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\hat{\mathbb{G}}_n - \mathbb{G}_n)(u, \delta)$$
$$+ \int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\mathbb{G}_n - G)(u, \delta)$$

An argument similar to that of Lemma A.7 in [16] shows that

$$\int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\mathbb{G}_n - G)(u, \delta) = o_p(n^{-2/5}),$$

and hence,

$$\int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\mathbb{G}_n - G)(u, \delta) = o_{P_M}(n^{-2/5}),$$

in probability. Similarly to the proof of Lemma A.7 in [16], we can also show that

$$\int \psi_{t,h}(u) \left\{ F_0(u) - \hat{F}_n(u) \right\} d(\hat{\mathbb{G}}_n - \mathbb{G}_n)(u, \delta) = o_{P_M}(n^{-2/5}), \tag{6.14}$$

in probability, such that,

$$A_I = o_{P_M}(n^{-2/5}) \qquad \text{in probability.}$$

We now study the term A_{II}. Using the same inequality for $\psi_{t,h} - \bar{\psi}_{t,h}$ as used in the second display after (11.49) on p. 333 of [17], we get for some constant $C > 0$ that:

$$\left| \bar{\psi}_{t,h}(u) - \psi_{t,h}(u) \right| \leq C h^{-2} \left| \hat{F}_n(u) - F_0(u) \right|, \tag{6.15}$$

for all u such that f_0 is positive and continuous in a neighborhood around u. We decompose the term A_{II} as follows,

$$A_{II} = \int \left\{ \bar{\psi}_{t,h}(u) - \psi_{t,h}(u) \right\} \left\{ \hat{F}_n(u) - F_0(u) \right\} d\hat{\mathbb{P}}_n(u, \delta)$$
$$+ \int \left\{ \bar{\psi}_{t,h}(u) - \psi_{t,h}(u) \right\} \left\{ F_0(u) - \delta \right\} d\hat{\mathbb{P}}_n(u, \delta). \tag{6.16}$$

For the first term on the right-hand side of the above display we write,

$$\begin{aligned}
&\int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \left\{\hat{F}_n(u) - F_0(u)\right\} d\hat{\mathbb{P}}_n(u,\delta) \\
&\quad = \int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \left\{\hat{F}_n(u) - F_0(u)\right\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(u,\delta) \\
&\qquad + \int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \left\{\hat{F}_n(u) - F_0(u)\right\} d\mathbb{P}_n(u,\delta) \\
&\quad \leq \int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \left\{\hat{F}_n(u) - F_0(u)\right\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(u,\delta) \\
&\qquad + Ch^{-2} \int_{t-h}^{t+h} \left\{\hat{F}_n(u) - F_0(u)\right\}^2 d\mathbb{P}_n(u,\delta), \qquad (6.17)
\end{aligned}$$

where we use (6.15) in the last inequality. The first term in the display above is $o_{P_M}(n^{-2/5})$ in probability by (6.14) and (6.15). Since

$$E_{M|Z} \left\{\hat{F}_n(t) - F_0(t)\right\}^2 < Kn^{-2/3} \qquad \forall t \in (0,R),$$

in probability, we have by Markov's inequality and Fubini's theorem that,

$$\int_{t-h}^{t+h} \left\{\hat{F}_n(u) - F_0(u)\right\}^2 d\mathbb{P}_n(u,\delta) = O_{P_M}\left(hn^{-2/3}\right) \text{ in probability.} \qquad (6.18)$$

Hence, for $h \asymp n^{-1/5}$, we get for the second term in (6.17):

$$\begin{aligned}
&Ch^{-2} \int_{t-h}^{t+h} \left\{\hat{F}_n(u) - F_0(u)\right\}^2 d\mathbb{P}_n(u,\delta) \\
&\quad = O_{P_M}\left(h^{-1}n^{-2/3}\right) = O_{P_M}\left(n^{-7/15}\right) = o_{P_M}\left(n^{-2/5}\right),
\end{aligned}$$

in probability. For the second term of (6.16) we have

$$\begin{aligned}
&\int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \{F_0(u) - \delta\} d\hat{\mathbb{P}}_n(u,\delta) \\
&\quad = \int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \{F_0(u) - \delta\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(u,\delta) \\
&\qquad + \int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \{F_0(u) - \delta\} d\big(\mathbb{P}_n - P\big)(u,\delta).
\end{aligned}$$

Similar to the arguments used in the treatment of term A_I above, we get by using again arguments similar to that of Lemma A.7 in [16] that:

$$\int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \{F_0(u) - \delta\} d\big(\hat{\mathbb{P}}_n - \mathbb{P}_n\big)(u,\delta) = o_{P_M}(n^{-2/5}),$$

and

$$\int \{\bar{\psi}_{t,h}(u) - \psi_{t,h}(u)\} \{F_0(u) - \delta\} d\big(\mathbb{P}_n - P\big)(u,\delta) = o_{P_M}(n^{-2/5}),$$

in probability. □

6.3. The current status linear regression model: bootstrap validity

In this section we give a road map for the proof of the bootstrap validity in the current status linear regression model. We assume that the assumptions stated in Theorem 4.1 of [15] hold. Since the proof is very similar to the proof of Theorem 4.1 in [15], we leave the details to the interested reader. Consider the bootstrap score function

$$\hat{\psi}_n^{(\epsilon)}(\beta) = \int_{\hat{F}_{n,\beta}(t-\beta' x)\in[\epsilon,1-\epsilon]} x\{\delta - \hat{F}_{n,\beta}(t-\beta' x)\}\, d\hat{\mathbb{P}}_n(t,x,\delta), \tag{6.19}$$

for some fixed truncation parameter $\epsilon \in (0, 1/2)$.

The main idea is to show that

$$\begin{aligned}\hat{\psi}_n^{(\epsilon)}(\hat{\beta}_n) = V(\hat{\beta}_n - \beta_0) &+ \int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} \{x - E(X|T-\beta_0' X = t-\beta_0' x)\} \\ &\qquad \cdot \{\delta - F_0(t-\beta_0' x)\}\, d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t,x,\delta) \\ &+ \int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} \{x - E(X|T-\beta_0' X = t-\beta_0' x)\} \\ &\qquad \cdot \{\delta - F_0(t-\beta_0' x)\}\, d(\mathbb{P}_n - P)(t,x,\delta) \\ &+ o_{P_M}(n^{-1/2} + (\hat{\beta}_n - \beta_0)),\end{aligned} \tag{6.20}$$

in probability, where E denotes the unconditional expectation. As in [15] we can work with the definition

$$\hat{\psi}_n^{(\epsilon)}(\hat{\beta}_n) = 0,$$

for the score estimator $\hat{\beta}_n$. Since by the proof of Theorem 4.1 in [15],

$$\begin{aligned}&-\sqrt{n}V(\beta_n - \beta_0) \\ &\quad = \sqrt{n}\int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} \{x - E(X|T-\beta_0' X = t-\beta_0' x)\} \\ &\qquad\qquad \cdot \{\delta - F_0(t-\beta_0' x)\}\, d(\mathbb{P}_n - P)(t,x,\delta) \\ &\qquad + o_p(1 + \sqrt{n}(\beta_n - \beta_0)),\end{aligned}$$

we get that,

$$\begin{aligned}&-\sqrt{n}V(\hat{\beta}_n - \beta_n) \\ &\quad = \sqrt{n}\int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} \{x - E(X|T-\beta_0' X = t-\beta_0' x)\} \\ &\qquad\qquad \cdot \{\delta - F_0(t-\beta_0' x)\}\, d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t,x,\delta) \\ &\qquad + o_{P_M}(1 + \sqrt{n}(\hat{\beta}_n - \beta_0)).\end{aligned}$$

The validity of the bootstrap then follows by the arguments given in Section 4.2. Very important in the proof of (6.20) is the conditional bootstrapped L_2-result,

$$\sup_\beta \int \left\{\hat{F}_{n,\beta}(t-\beta' x) - F_\beta(t-\beta' x)\right\}^2 d\mathbb{P}_n(t,x,\delta) = O_{P_M}\left(n^{-2/3}\right), \tag{6.21}$$

in probability, where F_β is defined in (4.18).

Let $\bar{\phi}_{\hat{\beta}_n,\hat{F}_{n,\hat{\beta}_n}}$ be a (random) piecewise constant version of $\phi_{\hat{\beta}_n}$, where

$$\phi_\beta \stackrel{def}{=} E\left\{X | T-\beta' X = u\right\},$$

and where, for a piecewise constant distribution function F with finitely many jumps at $\tau_1 < \tau_2 < \ldots$, the function $\bar{\phi}_{\beta,F}$ is defined in the following way.

$$\bar{\phi}_{\beta,F}(u) = \begin{cases} \phi_\beta(\tau_i), & \text{if } F_\beta(u) > F(\tau_i),\ u \in [\tau_i,\tau_{i+1}), \\ \phi_\beta(s), & \text{if } F_\beta(u) = F(s),\ \text{for some } s \in [\tau_i,\tau_{i+1}), \\ \phi_\beta(\tau_{i+1}), & \text{if } F_\beta(u) < F(\tau_i),\ u \in [\tau_i,\tau_{i+1}). \end{cases} \tag{6.22}$$

Similar to the proof of Theorem 4.1 in [15], we get that,

$$\|\phi_{\hat{\beta}_n}(u) - \bar{\phi}_{\hat{\beta}_n,\hat{F}_{n,\hat{\beta}_n}}(u)\| \leq K|\hat{F}_{n,\hat{\beta}_n}(u) - F_{\hat{\beta}_n}(u)|, \tag{6.23}$$

for some constant $K > 0$ not depending on β. By the definition of the MLE $\hat{F}_{n,\hat{\beta}_n}$ as the slope of the greatest convex minorant of the corresponding cusum diagram, we can write:

$$\begin{aligned}
&\hat{\psi}_n^{(\epsilon)}(\hat{\beta}_n) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \left\{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\right\}\left\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\right\} d\hat{\mathbb{P}}_n(t,x,\delta) \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \left\{\phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x) - \bar{\phi}_{\hat{\beta}_n,\hat{F}_{n,\hat{\beta}_n}}(t-\hat{\beta}_n' x)\right\} \\
&\qquad\qquad \cdot \left\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\right\} d\hat{\mathbb{P}}_n(t,x,\delta) \\
&= I + II,
\end{aligned}$$

For the second term, we have:

$$\begin{aligned}
II &= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \left\{\phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x) - \bar{\phi}_{\hat{\beta}_n,\hat{F}_{n,\hat{\beta}_n}}(t-\hat{\beta}_n' x)\right\} \\
&\qquad\qquad \cdot \left\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\right\} d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t,x,\delta) \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \left\{\phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x) - \bar{\phi}_{\hat{\beta}_n,\hat{F}_{n,\hat{\beta}_n}}(t-\hat{\beta}_n' x)\right\} \\
&\qquad\qquad \cdot \left\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\right\} d\mathbb{P}_n(t,x,\delta) \\
&= II_a + II_b
\end{aligned}$$

It is shown in the proof of Theorem 4.1 in [15] that

$$II_b = o_p(n^{-1/2} + (\hat{\beta}_n - \beta_0)),$$

and therefore

$$II_b = o_{P_M}(n^{-1/2} + (\hat{\beta}_n - \beta_0)) \text{ in probability.}$$

Using similar arguments as in in the proof of Theorem 4.1 in [15] we can also show that

$$II_a = o_{P_M}(n^{-1/2}) \text{ in probability.}$$

Hence, we get:

$$\begin{aligned}
&\hat{\psi}_n^{(\epsilon)}(\hat{\beta}_n) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\, d\hat{\mathbb{P}}_n(t,x,\delta) \\
&\quad + o_{P_M}(n^{-1/2} + (\hat{\beta}_n - \beta_0)),
\end{aligned}$$

in probability. We now write,

$$\begin{aligned}
&\int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\, d\hat{\mathbb{P}}_n(t,x,\delta) \\
&= \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\} \\
&\qquad\qquad \cdot \{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\, d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t,x,\delta) \\
&\quad + \int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\, d\mathbb{P}_n(t,x,\delta)
\end{aligned}$$

It follows from the proof of Theorem 4.1 in [15] that there exists a random variable R_n of order $o_p(n^{-1/2}+\hat{\beta}_n-\beta_0)$ (and hence of order $o_{P_M}(n^{-1/2}+\hat{\beta}_n-\beta_0)$ in probability) such that,

$$\begin{aligned}
&\int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\, d\mathbb{P}_n(t,x,\delta) \\
&= \int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} \Big\{x - \phi_0(t-\beta_0' x)\Big\}\Big\{\delta - F_0(t-\beta_0' x)\Big\}\, d(\mathbb{P}_n - P)(t,x,\delta) \\
&\quad + \psi_{1,\epsilon}'(\beta_0)(\hat{\beta}_n - \beta_0) + R_n. \qquad (6.24)
\end{aligned}$$

where $\phi_0 \equiv \phi_{\beta_0}$. Therefore, (6.20) follows if we can show that,

$$\begin{aligned}
&\int_{\hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\in[\epsilon,1-\epsilon]} \{x - \phi_{\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\{\delta - \hat{F}_{n,\hat{\beta}_n}(t-\hat{\beta}_n' x)\}\, d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t,x,\delta) \\
&= \int_{F_0(t-\beta_0' x)\in[\epsilon,1-\epsilon]} \Big\{x - \phi_0(t-\beta_0' x)\Big\}\Big\{\delta - F_0(t-\beta_0' x)\Big\}\, d(\hat{\mathbb{P}}_n - \mathbb{P}_n)(t,x,\delta) \\
&\quad + o_{P_M}(n^{-1/2} + (\hat{\beta}_n - \beta_0)). \qquad (6.25)
\end{aligned}$$

Equality (6.25) follows by similar arguments used in the proof of (6.24) based on asymptotic equicontinuity using the closeness of $\hat{F}_{n,\beta}$ to F_β and using entropy results for the functions $u \mapsto \hat{F}_{n,\beta}(u)$ and the simpler parametric functions $u \mapsto F_\beta(u)$ and $u \mapsto \phi_\beta(u)$, parametrized by the finite dimensional parameter β.

Acknowledgements

The research of the second author was supported by the Research Foundation Flanders (FWO) [grant number 11W7315N]. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged. For the simulations we used the infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government - department EWI.

References

[1] Abrevaya, J. (1999). Rank regression for current-status data: asymptotic normality. *Statist. Probab. Lett.* **43** 275–287. MR1708095

[2] Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245

[3] Balabdaoui, F., Groeneboom, P. and Hendrickx, K. (2017). Score estimation in the monotone single index model. working paper.

[4] Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32** 405–424. MR2204627

[5] Cheng, G., Huang, J. Z. et al. (2010). Bootstrap consistency for general semiparametric M-estimation. *Ann. Statist.* **38** 2884–2915. MR2722459

[6] Chernoff, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* **16** 31–41. MR0172382 (30 ##2601)

[7] Durot, C., Groeneboom, P. and Lopuhaä, H. P. (2013). Testing equality of functions under monotonicity constraints. *J. Nonparametr. Stat.* **25** 939–970. MR3174305

[8] Durot, C. and Reboul, L. (2010). Goodness-of-Fit Test for Monotone Functions. *Scandinavian Journal of Statistics* **37** 422–441. MR2724504

[9] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* **7** 1-26. MR0515681

[10] Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39** 125–153 (1957). MR0093415 (19,1243c)

[11] Groeneboom, P. (2012). Likelihood Ratio Type Two-Sample Tests for Current Status Data. *Scandinavian Journal of Statistics* **39** 645–662. MR3000840

[12] Groeneboom, P. (2014). Maximum smoothed likelihood estimators for the interval censoring model. *Ann. Statist.* **42** 2092–2137. MR3262478

[13] Groeneboom, P. (2015). Rcpp scripts. `https://github.com/pietg/book/tree/master/Rcpp_scripts`.

[14] Groeneboom, P. and Hendrickx, K. Confidence Intervals for the Current Status Model. *Scand. J. Statist.* 10.1111/sjos.12294.
[15] Groeneboom, P. and Hendrickx, K. (2017). Current status linear regression. Accepted for publication in Ann. Statist., available at https://arxiv.org/abs/1601.00202.
[16] 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** 352–387. MR2589325
[17] Groeneboom, P. and Jongbloed, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge. MR3445293
[18] Groeneboom, P. and Jongbloed, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43** 2019–2054. MR3375875
[19] Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation. DMV Seminar* **19**. Birkhäuser Verlag, Basel. MR1180321 (94k:62056)
[20] Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32** 177–203. MR1046764 (91i:62060)
[21] Hall, P. (1992a). *The bootstrap and Edgeworth expansion. Springer Series in Statistics*. Springer. MR1145237
[22] Hall, P. (1992b). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587 (93e:62131)
[23] Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics* **35** 303–316. MR0903188
[24] Keiding, N., Begtrup, K., Scheike, T. H. and Hasibeder, G. (1996). Estimation from Current Status Data in Continuous Time. *Lifetime Data Anal.* **2** 119–129.
[25] Kim, J. K. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219. MR1041391 (91f:62059)
[26] Kosorok, M. R. (2008). Bootstrapping the Grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen. Inst. Math. Stat. Collect.* **1** 282–292. Inst. Math. Statist., Beachwood, OH. MR2462212
[27] Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics* **3** 205–228. MR0436905
[28] Patra, R. K., Seijo, E. and Sen, B. (2011). A consistent bootstrap procedure for the maximum score estimator. *arXiv preprint arXiv:1105.1976*.
[29] Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association* **79** 871–880. MR0770281
[30] Schuster, E. F. (1985). Incorporating support constraints into nonparametric estimators of densities. *Comm. Statist. A—Theory Methods* **14** 1123–1136. MR797636 (86m:62078)

[31] Sen, B., Banerjee, M. and Woodroofe, M. B. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880 (2011f:62046)

[32] Sen, B. and Xu, G. (2015). Model based bootstrap methods for interval censored data. *Comput. Statist. Data Anal.* **81** 121–129. MR3257405

[33] Subbotin, V. (2007). Asymptotic and bootstrap properties of rank regressions. Available at SSRN: https://ssrn.com/abstract=1028548.

[34] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics*. Springer-Verlag, New York. With applications to statistics. MR1385671 (97g:60035)