Confidence Intervals for the Current Status Model

Groeneboom, Piet; Hendrickx, Kim

DOI
10.1111/sjos.12294
Publication date
2018
Document Version
Final published version
Published in
Scandinavian Journal of Statistics: theory and applications

Citation (APA)
Groeneboom, P., & Hendrickx, K. (2018). Confidence Intervals for the Current Status Model. *Scandinavian Journal of Statistics: theory and applications*, *45*(1), 135-163. https://doi.org/10.1111/sjos.12294

Scandinavian Journal of Statistics Theory and Applications

Scandinavian Journal of Statistics, Vol. 45: 135–163, 2018
doi: 10.1111/sjos.12294

Confidence Intervals for the Current Status Model

PIET GROENEBOOM
Delft Institute of Applied Mathematics, Delft University of Technology

KIM HENDRICKX
I-BioStat, Hasselt University

ABSTRACT. We discuss a new way of constructing pointwise confidence intervals for the distribution function in the current status model. The confidence intervals are based on the smoothed maximum likelihood estimator, using local smooth functional theory and normal limit distributions. Bootstrap methods for constructing these intervals are considered. Other methods to construct confidence intervals, using the non-standard limit distribution of the (restricted) maximum likelihood estimator, are compared with our approach via simulations and real data applications.

Key words: bootstrap, confidence intervals, current status, maximum likelihood estimator, smoothed maximum likelihood estimator

1. Introduction

Survival models are commonly used to characterize the distribution of a variable X that is not observed directly. Depending on what information is obtained on X, different censoring schemes arise. In this paper, we consider the situation that a variable of interest is only known to lie before or after some random censoring variable T. Each observed sample consists of a set of n inspection times T_i (independent of the other T_j and all $X_j's$, $j = 1, \ldots, n$) and n censoring indicators $\Delta_i = 1_{\{X_i \le T_i\}}$. This type of censored data is known as current status data and arises naturally in reliability and survival studies when the status of an observational unit is only checked at one measurement point, which happens especially when testing is destructive. One could say that the ith observation indicates the *current status* of component i at time T_i. Estimation of the distribution function of the response variable in the current status model is harder than in right-censored models because of the lack of observing an actual event of interest. Groeneboom & Wellner (1992) show that the (non-parametric) maximum likelihood estimator (MLE) $\hat{F}_n$, maximizing the likelihood of the data given by

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

over all possible distribution functions F without making any additional constraints, converges pointwise at cube root n rate to the true distribution function F_0 of X. The Kaplan–Meier estimator (Kaplan & Meier, 1958), which is the MLE for right-censored data, converges on the contrary at a faster square root n rate because of the fact that one has actual observations in addition to the censored ones. In the current status model, all observations are censored.

In this paper, we introduce new methods for constructing pointwise confidence intervals (CIs) for F_0 at time t and compare our techniques with existing methods for interval estimation in current status models. We assume that both X and T have continuously differentiable distribution functions F_0 and G, respectively, with positive derivatives f_0 and g at t. From Groeneboom & Wellner (1992), it is known that

$$n^{1/3}\left\{\hat{F}_n(t) - F_0(t)\right\} \xrightarrow{\mathcal{D}} [4F_0(t)\{1 - F_0(t)\}f_0(t)/g(t)]^{1/3}\mathbb{C}, \tag{2}$$

where $\mathbb{C} = \arg\min_{t\in\mathbb{R}}\{\mathbb{Z}(t) + t^2\}$ and $\mathbb{Z}(t)$ is a standard two-sided Brownian motion process, originating from zero. To construct a CI using the result given in (2), we therefore need estimates of f_0 and g. If one is willing to make assumptions on the underlying distribution functions of X and T, parametric methods can be used. This was performed in, for example, Keiding *et al.* (1996) using Weibull models for both f_0 and g. Non-parametric estimates obtained by kernel smoothing were considered in Banerjee & Wellner (2005). The choice of the tuning parameter is however crucial for a good performance of the CIs.

Banerjee & Wellner (2005) proposed a likelihood-ratio (LR)-based method for constructing pointwise CIs for the distribution function in current status models. Starting from the LR statistic

$$\mathrm{LR}(\theta_0) = 2\left(\log \ell_n(\hat{F}_n) - \log \ell_n(\hat{F}_n^{\theta_0})\right),$$

for testing the null hypothesis $F_0(t) = \theta_0$, which has an asymptotic distribution $\mathbb{D}$ characterized in Banerjee & Wellner (2001), the authors estimate the interval by

$$\{\theta \in (0,1) : \mathrm{LR}(\theta) \le d_{1-\alpha}\},$$

where $d_{1-\alpha}$ is the $(1-\alpha)$th percentile of $\mathbb{D}$. Here, $\hat{F}_n$ denotes the unconstrained MLE maximizing (1), and $\hat{F}_n^{\theta_0}$ denotes the MLE of F_0 under the constrained that $F_0(t) = \theta_0$. The LR-based method avoids estimation of f_0 and g because, under the null hypothesis, the limiting distribution is free of the underlying parameters. In contrast to the situation for the MLE itself and for the smoothed maximum likelihood estimator (SMLE), no analytical information is available for the distribution $\mathbb{D}$, and the distribution $\mathbb{D}$ is estimated via simulations. A short proof of the characterization of the 'Chernoffian' limit distribution of the MLE itself (without being restricted) in terms of Airy functions has recently been given in Groeneboom *et al.* (2015), and the asymptotic distribution of the SMLE is just normal. So, in these cases, tables of the critical values are available (for the MLE, they are given in Groeneboom & Wellner, 2001). Tables to determine the asymptotic critical values for the LR test are available in Banerjee & Wellner (2001).

More recently, bootstrap methods for constructing CIs in the current status model have been considered. It is however proved in Abrevaya & Huang (2005) that the naive bootstrap procedure, which simply resamples the original data, will not work for pointwise CIs for the distribution function F_0 if it is estimated by the MLE $\hat{F}_n$. A consistent model-based bootstrap procedure was introduced in Sen & Xu (2015). Instead of resampling the (T_i, Δ_i), the authors proposed resampling the Δ_i from a Bernoulli distribution with success probability given by $\tilde{F}(T_i)$, where $\tilde{F}$ is an estimator of F_0 satisfying some smoothness conditions (that are not fulfilled by the ordinary MLE $\hat{F}_n$). The obtained bootstrap sample $(T_1, \Delta_1^*), \ldots, (T_n, \Delta_n^*)$ can next be used for interval estimation. In this case, one computes the MLE F_n^* in the bootstrap samples and subtracts the smooth distribution $\tilde{F}$, generating the Δ_i^*. The CIs are then formed by taking

$$\left[\hat{F}_n(t) - V^*_{1-\alpha/2}(t), \hat{F}_n(t) - V^*_{\alpha/2}(t)\right],$$

where V^*_α us the αth quantile of B values of

$$F_n^*(t) - \tilde{F}(t),$$

where B is the number of bootstrap samples taken.

For current status and related models, some research has been reported recommending the use of this smooth bootstrap procedure. A smooth bootstrap calibration was used in Durot & Reboul (2010) for a goodness-of-fit test for monotone functions and in Groeneboom (2012) for an LR-type two-sample test for current status data. Durot *et al.* (2013) used a similar approach to determine the critical value for testing equality of functions under monotonicity constraints. The main motivation for recommending the smooth bootstrap is the negative results by Abrevaya & Huang (2005) and Kosorok (2008) proving the inconsistency of the naive bootstrap for generating the limiting distribution of the MLE.

Recently, it was however proved in Groeneboom & Hendrickx (2017b) that the naive bootstrap of resampling with replacement from the data does work in case the underlying distribution function is estimated by the SMLE or in case interest is in other functionals than the values of the distribution function. The validity of the naive bootstrap for constructing pointwise CIs around the SMLE and for doing inferences in the current status linear regression model (Groeneboom & Hendrickx, 2017a) is illustrated in Groeneboom & Hendrickx (2017b). Although Durot & Reboul (2010) conjecture that the naive bootstrap fails in their setting, the result of Groeneboom & Hendrickx (2017b) suggests that this conjecture might be incorrect and that applications of the naive bootstrap involving the Grenander estimator are worthy of study in further research.

Besides considering the naive or smooth bootstrap, one could moreover consider resampling the Δ_i from the MLE itself. Simulation studies in Durot *et al.* (2013) even suggest that the smooth bootstrap does not necessarily perform better than bootstrapping from the Grenander estimator in their setting. So far, the theoretical properties of the latter bootstrap procedure remain an open problem. As a consequence of the positive result by Groeneboom & Hendrickx (2017b), we conjecture that bootstrapping from the MLE might very well work for pointwise CIs in the current status model, if one uses the right functional of the model as a basis for the intervals.

The outline of this paper is as follows. In Section 2, we introduce the current status model, describe the construction of the SMLE for the distribution function and explain how the smooth bootstrap procedure can be used to construct pointwise CIs for the distribution function. The asymptotic behaviour of the CIs is also given in Section 2 together with some details on how to improve the performance of our intervals. Simulation studies are reported in Section 3 to demonstrate the finite sample behaviour of our CIs and to compare our method with existing methods proposed by Banerjee & Wellner (2005) and Sen & Xu (2015). In Section 4, we illustrate our methods on the hepatitis A dataset and the rubella dataset. Some concluding remarks are pointed out in Section 5. The Appendix contains the proofs of our main results.

The proofs of our results are rather non-trivial and use techniques totally different from the techniques used in Banerjee & Wellner (2005) and Sen & Xu (2015). The latter fact is not unexpected, because the intervals are based on recently developed smooth functional theory (see, e.g. Groeneboom & Jongbloed, 2015) and deal with asymptotically normal limits instead of the non-standard limits for the (restricted) MLE. We hope that the present paper serves the purpose of making these techniques more widely known. Rcpp scripts for all methods, discussed here (also the methods of Banerjee & Wellner, 2005, and Sen & Xu, 2015), are available in Groeneboom (2015).

2. Pointwise confidence intervals in the current status model

Consider an independent identically distributed sample $X_1, \ldots, X_n$ with distribution function F_0, where the distribution corresponding to F_0 has support $[0, M]$, and let F_0 have a density

f_0 staying away from zero on $[0, M]$. The observations in the current status model are $(T_1, \Delta_1 = 1_{\{X_1 \le T_1\}}), \ldots, (T_n, \Delta_n = 1_{\{X_n \le T_n\}})$ where the T_i are independent of all $X_j^{\prime}s$ and have a distribution G with Lebesgue density g with a support that contains $[0, M]$. We assume that g stays away from zero on $[0, M]$ and has a bounded derivative g'. In this section, we develop a method for CI estimation for $F_0(t)$ when t is an interior point of $[0, M]$ and f_0 has a continuous derivative at t. We estimate $F_0(t)$ by the SMLE obtained by first estimating the MLE $\hat{F}_n$ and then smoothing this using a smoothing kernel, that is,

$$\tilde{F}_{nh}(t) = \int \mathbb{K}\left(\frac{t-x}{h}\right) d\hat{F}_n(x), \tag{3}$$

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

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

and where h is a chosen bandwidth. Here, $d\hat{F}_n$ represents the jumps (masses) of the discrete distribution function $\hat{F}_n$, and K is one of the usual kernels used in density estimation (i.e. K is a probability density with support $[-1, 1]$ that is symmetric and twice continuously differentiable on $\mathbb{R}$). We use the notations K_h and $\mathbb{K}_h$ to denote the scaled versions of K and $\mathbb{K}$, respectively, given by

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

It is well known that the MLE $\hat{F}_n$ can be characterized as the left continuous slope of the convex minorant of a cumulative sum diagram formed by the point $(0, 0)$ and

$$\left(\sum_{j=1}^{i} w_j, \sum_{j=1}^{i} f_{1j}\right), \; i = 1, \ldots, m,$$

where the w_j are weights, given by the number of observations at point $T_{(j)}$, assuming that $T_{(1)} < \cdots < T_{(m)}$ (m being the number of different observations in the sample) are the order statistics of the sample $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$ and where f_{1j} is the number of Δ_k equal to one at the jth-order statistic of the sample. When no ties are present in the data (as is indeed the case in our simulations due to continuity assumptions of g, but is often not satisfied in real data examples), $w_j = 1, m = n$ and $f_{1j} = \Delta_{(j)}$, where $\Delta_{(j)}$ corresponds to $T_{(j)}$.

From Groeneboom *et al.* (2010) (theorem 4.2, p. 365), it follows that

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

where

$$\beta = \frac{c^2 f_0^{\prime}(t)}{2} \int u^2 K(u)\, du \quad \text{and} \quad \sigma^2 = \frac{F_0(t)\{1 - F_0(t)\}}{cg(t)} \int K(u)^2\, du. \tag{4}$$

In the remainder of this section, we first introduce a procedure for interval estimation based on a smooth bootstrap resampling scheme and next elucidate some adjustments to improve the performance of the bootstrap CIs.

2.1. The smooth bootstrap

We obtain a bootstrap sample $(T_1, \Delta_1^*), \ldots, (T_n, \Delta_n^*)$ by keeping the T_i in the original sample fixed and by resampling the Δ_i^* from a Bernoulli distribution with probability $\tilde{F}_{nh}(T_i)$. The following bootstrap $1-\alpha$ interval is suggested:

$$\left[\tilde{F}_{nh}(t) - U^*_{1-\alpha/2}(t), \tilde{F}_{nh}(t) - U^*_{\alpha/2}(t)\right], \tag{5}$$

where $U^*_\alpha(t)$ is the αth quantile of B values of

$$Z_{nh}(t) = \tilde{F}^*_{nh}(t) - \int \mathbb{K}_h(t-u)\, d\tilde{F}_{nh}(u).$$

Here, $\tilde{F}^*_{nh}(t)$ is the SMLE in the bootstrap sample defined in the same way as in (3) but with $\hat{F}_n$ replaced by $\hat{F}^*_n$, that is, the MLE in the bootstrap sample.

Under the model assumptions stated at the beginning of this section, we have the following main result showing that $n^{2/5}Z_{nh}(t)$ converges to a normal distribution with the same asymptotic variance as the SMLE. The proof of this result can be found in the Appendix. Some theoretical aspects of the bootstrap MLE $\hat{F}^*_n$, important for proving our main result, are given in Section 2.2.

Theorem 1. *Let $h = h_n \sim cn^{-1/5}$, and let σ^2 be given by (4), then*

$$n^{2/5}\left\{\tilde{F}^*_{nh}(t) - \int \mathbb{K}_h(t-u)\, d\tilde{F}_{nh}(u)\right\} \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

given the data $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$, almost surely (a.s.) along sequences $(T_1, \Delta_1), (T_2, \Delta_2), \ldots$.

Note that we can write

$$\begin{aligned}\int \mathbb{K}_h(t-u)\, d\tilde{F}_{nh}(u) &= \int \mathbb{K}_h(t-u)\left\{\int K_h(u-v) d\hat{F}_n(v)\right\} du\\ &= \int\int \mathbb{K}((t-v)/h - w)K(w)\, dw d\hat{F}_n(v).\end{aligned}$$

In practice, we therefore have to compute the convolution kernel $\tilde{\mathbb{K}}$, defined by

$$\tilde{\mathbb{K}}(x) = \int \mathbb{K}(x-w)K(w)\, dw. \tag{6}$$

A picture of the functions K, $\mathbb{K}$ and $\tilde{\mathbb{K}}$ is given in Fig. 1 using the triweight kernel defined by

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

Remark 1. Note that we subtract the integrated SMLE using the original data instead of the SMLE itself in the definition of $Z_{nh}(t)$ because of the bias of the SMLE $\tilde{F}_{nh}(t)$. This is in line with the method proposed by Sen & Xu (2015) where the authors subtract the SMLE instead of the MLE of the original data for constructing CIs around the MLE. One needs to introduce an additional level of smoothing in order to construct valid intervals using the smooth bootstrap procedure.

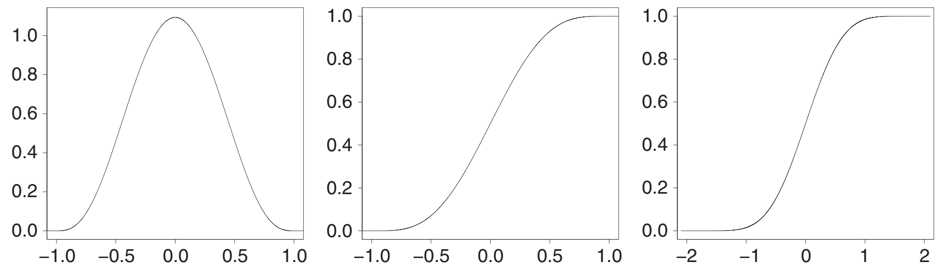

Fig. 1. (a) Triweight kernel $K : x \mapsto \frac{35}{32}(1 - x^2)^3 1_{[-1,1]}(x)$, (b) integrated triweight kernel $\mathbb{K} : x \mapsto \int_{-\infty}^{x} K(w)\,dw$ and (c) the convolution $\tilde{\mathbb{K}} : x \mapsto \int \mathbb{K}(x-w)K(w)\,dw$.

2.2. *Asymptotic properties of the smooth bootstrap*

It is well known that the L_2-distance between the MLE $\hat{F}_n$ in the original sample and the true distribution function F_0 is of order $n^{-1/3}$ (see, e.g. van de Geer, 2000, example 7.4.3). In the proof of theorem 1, we need the following result:

$$\int_{t-h}^{t+h} \left\{\hat{F}_n^*(x) - \tilde{F}_{nh}(x)\right\}^2 dx = O_p^*\left(hn^{-2/3}\right), \tag{7}$$

where $O_p^*\left(hn^{-2/3}\right)$ means that for all $\varepsilon > 0$ and almost all sequences $(T_1, \Delta_1), (T_2, \Delta_2), \ldots,$ there exists an $M > 0$ such that

$$P_n^*\left\{\int_{t-h}^{t+h} \left\{\hat{F}_n^*(x) - \tilde{F}_{nh}(x)\right\}^2 dx \geq Mhn^{-2/3}\right\} < \varepsilon,$$

for all large n. Here, P_n^* denotes the conditional probability measure given (T_1, Δ_1), … (T_n, Δ_n). Note that (7) does not follow from a conditional global bound on the L_2-distance between the MLE $\hat{F}_n^*$ in the bootstrap sample and the SMLE $\tilde{F}_{nh}$ in the original sample of order $n^{-1/3}$ and that this is a refinement of the usual Hellinger distance calculations.

By using the so-called 'switch relation' that reduces the study of the MLE $\hat{F}_n^*$ to the study of an inverse process (see, e.g. Groeneboom & Jongbloed, 2014, p. 320), we show in the Appendix that

$$E_n^*\left\{\hat{F}_n^*(t) - \tilde{F}_{nh}(t)\right\}^2 \leq Kn^{-2/3} \quad \forall t \in [0, M], \tag{8}$$

where E_n^* denotes the conditional expectation given (T_1, Δ_1), … (T_n, Δ_n). From this result, it follows that (7) holds.

In the remainder of this section, we describe techniques to improve the CIs defined in (5) by (i) considering estimation of the variance, (ii) taking into account the boundary effects of kernel estimates, and (iii) estimating the asymptotic bias β defined in (4).

2.3. *Studentized confidence intervals*

Usually, the performance of the bootstrap CIs works best if one uses a pivot, obtained by Studentizing. In each bootstrap sample, we therefore estimate the variance σ^2 defined in (4), apart from the factor $cg(t)$, which drops out in the Studentized bootstrap procedure, by

$$S_{nh}^*(t) = n^{-2}\sum_{i=1}^{n} K_h(t - T_i)^2\left(\Delta_i^* - \hat{F}_n^*(T_i)\right)^2. \tag{9}$$

The variance estimate defined in (9) is inspired by the fact that the SMLE $\tilde{F}_{nh}$ is asymptotically equivalent to the toy estimator

$$\tilde{F}_{nh}^{\text{toy}}(t) = \int \mathbb{K}_h(t-x)\,dF_0(x) + \frac{1}{n}\sum_{i=1}^{n} \frac{K_h(t-T_i)\{\Delta_i - F_0(T_i)\}^2}{g(T_i)},$$

which has sample variance

$$S_n(t) = \frac{1}{n^2}\sum_{i=1}^{n} \frac{K_h(t-T_i)^2\,\{\Delta_i - F_0(T_i)\}^2}{g(T_i)^2}.$$

We next compute

$$W_{nh}^*(t) = \frac{\tilde{F}_{nh}^*(t) - \int \mathbb{K}_h(t-u)\,d\tilde{F}_{nh}(u)}{\sqrt{S_{nh}^*(t)}}.$$

Let $Q_\alpha^*(t)$ be the αth quantile of B values of $W_{nh}^*(t)$, where B is the number of bootstrap samples. Then the following bootstrap $1-\alpha$ interval is suggested:

$$\left[\tilde{F}_{nh}(t) - Q_{1-\alpha/2}^*(t)\sqrt{S_{nh}(t)},\, \tilde{F}_{nh}(t) - Q_{\alpha/2}^*(t)\sqrt{S_{nh}(t)}\right], \tag{10}$$

where $S_{nh}(t)$ is the variance estimate in the original sample obtained by replacing $\Delta_i^* - \hat{F}_n^*(T_i)$ in (9) by $\Delta_i - \hat{F}_n(T_i)$. Note that we do not need an estimate of the density g in each of the observations T_i as a consequence of the fact that $g(u)$ is close to $g(t)$ for $u \in [t-h, t+h]$. If, on the contrary, one wants to consider Wald-type CIs for the distribution function based on the asymptotic normality results of the SMLE, estimation of g is inevitable.

2.4. Boundary correction

It is well known that kernel density and distribution estimators without boundary correction are generally inconsistent at the boundary of the support $[0, M]$. We therefore use the boundary correction method proposed in Groeneboom & Jongbloed (2014) and define the SMLE as

$$\tilde{F}_{nh}^{(\text{bc})}(t) = \int \left\{\mathbb{K}\left(\frac{t-x}{h}\right) + \mathbb{K}\left(\frac{t+x}{h}\right) - \mathbb{K}\left(\frac{2M-t-x}{h}\right)\right\} d\hat{F}_n(x). \tag{11}$$

The boundary corrected version of $Z_{nh}^*(t)$ is defined by

$$Z_{nh}^{(\text{bc})*}(t) = \tilde{F}_{nh}^{(\text{bc})*}(t) - \int \{\mathbb{K}_h(t-x) + \mathbb{K}_h(t+x) - \mathbb{K}_h(2M-t-x)\}\, d\tilde{F}_{nh}^{(\text{bc})}(x).$$

The result of theorem 1 remains valid under this boundary correction. We also have the following lemma.

Lemma 1. *Let the boundary corrected estimate* $\tilde{F}_{nh}^{(\text{bc})}$ *be defined by (11), and let* $\tilde{\mathbb{K}}_h$ *be defined by*

$$\tilde{\mathbb{K}}_h(u) = \tilde{\mathbb{K}}(u/h), \qquad u \in \mathbb{R},$$

where the convolution kernel $\tilde{\mathbb{K}}$ *is defined by (6). Moreover, let* $0 < h \le M/3$. *Then*

$$\begin{aligned} &\int \{\mathbb{K}_h(t-x) + \mathbb{K}_h(t+x) - \mathbb{K}_h(2M-t-x)\}\, d\tilde{F}_{nh}^{(\text{bc})}(x) \\ &= \int \left\{\tilde{\mathbb{K}}_h(t-x) + \tilde{\mathbb{K}}_h(t+x) - \tilde{\mathbb{K}}_h(2M-t-x)\right\} d\hat{F}_n(x). \end{aligned} \tag{12}$$

From lemma 1, it follows that we can write

$$\begin{aligned} Z_{nh}^{(\mathrm{bc})*}(t) &= \int \{\mathbb{K}_h(t-x) + \mathbb{K}_h(t+x) - \mathbb{K}_h(2M-t-x)\}\, d\left(\hat{F}_n^* - \tilde{F}_{nh}^{(\mathrm{bc})}\right)(x) \\ &= \int \{\mathbb{K}_h(t-x) + \mathbb{K}_h(t+x) - \mathbb{K}_h(2M-t-x)\}\, d\hat{F}_n^*(x) \\ &\quad - \int \left\{\tilde{\mathbb{K}}_h(t-x) + \tilde{\mathbb{K}}_h(t+x) - \tilde{\mathbb{K}}_h(2M-t-x)\right\} d\hat{F}_n(x). \end{aligned}$$

The proof of lemma 1 is given in the Appendix. A picture of the MLE, together with the SMLE, both corrected and uncorrected for boundary effects, is shown in Fig. 2A for a sample from the truncated exponential distribution on [0,2] (see Section 3 for a detailed description of the model). Figure 2B presents the boundary corrected and uncorrected integrated SMLE and clearly shows the improvement of the boundary correction.

2.5. Bias estimation

When constructing CIs around the SMLE, one should take into account the bias of the SMLE. Note that this matter does not occur for CIs around the MLE, as in Sen & Xu (2015), because the asymptotic distribution of the MLE is symmetric around zero. Direct estimation of the asymptotic bias β defined in (4) requires a consistent estimate of the second derivative f_0' of the distribution function F_0. Although it is possible to estimate f_0' consistently (see, e.g. Groeneboom & Jongbloed, 2015, p. 243), our computer experiments demonstrated that it is very difficult to estimate the bias term sufficiently accurately. We therefore propose to use an adaptive bandwidth $h = h(t)$ in order to improve the performance of the SMLE-based CIs. When h is monotone increasing and continuous in t, then the SMLE $\tilde{F}_{nh}(t)$ is also monotone increasing and continuous in t. The effect of the adaptive bandwidth will be elucidated further in Section 3.

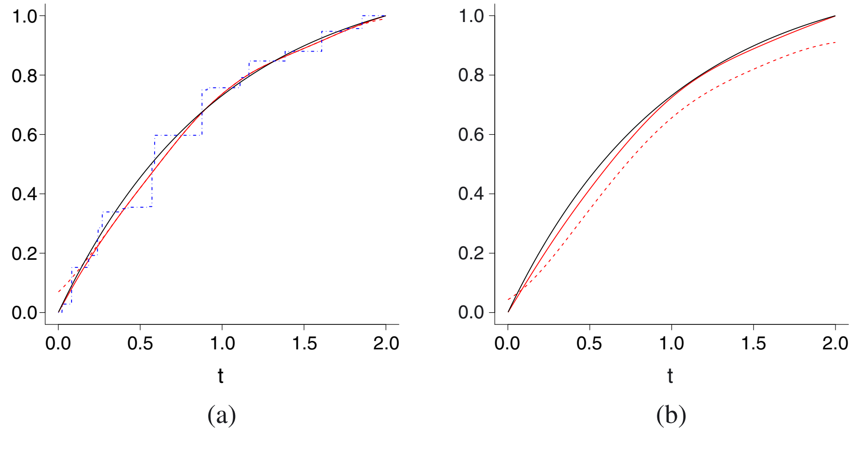

Fig. 2. Truncated exponential samples: (a) F_0 (black, solid), the maximum likelihood estimator (blue, dashed-dotted), the smoothed maximum likelihood estimator (SMLE) with boundary correction (red, solid) and the SMLE without boundary correction (red, dashed) and (b) F_0 (black, solid), the integrated SMLE with boundary correction (red, solid) and the integrated SMLE without boundary correction (red, dashed); $n = 1000$ and $h = 2n^{-1/5}$. [Colour figure can be viewed at wileyonlinelibrary.com]

3. Simulations

In this section, we illustrate the finite sample behaviour of the SMLE-based CIs introduced in Section 2. We also compare our CIs with the LR-based CIs proposed by Banerjee & Wellner (2005) and the CIs introduced in Sen & Xu (2015). The latter CIs are both constructed around the MLE. We use two simulation examples to analyze the effect of Studentizing and the choice of the kernel K on the behaviour of our SMLE-based CIs. We propose a criterion for bandwidth selection and illustrate how undersmoothing the bandwidth can improve the behaviour of the CIs.

In the first simulation setting, both the event times and censoring times are sampled from a Uniform(0,2)-distribution. Because the derivative of the uniform density equals zero, the SMLE is an unbiased estimate of the uniform distribution function, and no bias correction is needed. In the next simulation setup, we consider the model where we generate the event times from a truncated exponential distribution on [0,2] and take Uniform(0,2)-censoring times.

For sample sizes $n = 100, 500, 1000$ and 2000, we generated 5000 datasets from both models. The boundary correction described in Section 2.4 is used each time the SMLE is considered. The number of bootstrap samples within each simulation run equals $B = 1000$.

Table 1 shows the coverage percentage, that is, the number of times (out of the 5000 simulation runs) that $F_0(t)$ is not in the 95% CIs, and the average length of the 95% CIs around $F_0(t)$ for the uniform model and $t = 1$. We use the bandwidth $h = cn^{-1/5}$, where the constant $c = 2.0$ corresponds to the length of the interval $[0, 2]$. We consider two different choices for the kernel, the triweight kernel and the Epanechnikov kernel and compare the results of our SMLE-based CIs (10) with the results for the MLE-based methods of Banerjee & Wellner (2005) and Sen & Xu (2015).

For each point $t_i = 0.02, 0.04, \ldots, 2$, Fig. 3A presents the coverage proportions for the Studentized SMLE-based CIs (10) using the Epanechnikov kernel and the triweight kernel and illustrates that the choice of the kernel has only a small effect on the coverage proportions. The average length of the CIs, shown in Fig. 3B, is smaller for the intervals constructed with the Epanechnikov kernel.

A picture of the proportion of times that $F_0(t_i), t_i = 0.02, 0.04, \ldots, 2$ is not in the 95% CIs for $n = 1000$ is shown in Fig. 4A–C for the uniform model. The average length of our CIs based on the SMLE (both classical CIs (5) (result not shown) and Studentized CIs (10)) remains smaller than the average lengths of the Banerjee–Wellner and Sen–Xu CIs based on the MLE for all points t, as is shown in Fig. 4D. For the uniform samples, our SMLE-based method does not suffer from bias effects; the coverage of the different intervals is comparable for time points in the middle of the interval [0,2], but becomes rather bad at the boundary of the interval for the Banerjee–Wellner and Sen–Xu intervals. Figure 4 is obtained with the results for the Epanechnikov kernel. Similar comparisons were obtained when the triweight kernel was used. Figure 4A also shows that the classical SMLE-based CIs (5) are slightly anti-conservative near

Table 1. *Uniform samples*

	Studentized SMLE-based CI (*10*)						Sen–Xu			
	Triweight		Epanechnikov		Banerjee–Wellner		Triweight		Epanechnikov	
n	CP	L	CP	L	CP	L	CP	L	CP	L
100	0.0326	0.2799	0.0358	0.2376	0.0486	0.3897	0.0568	0.4620	0.0470	0.4625
500	0.0472	0.1473	0.0454	0.1276	0.0504	0.2311	0.0636	0.2532	0.0580	0.2536
1000	0.0626	0.1072	0.0600	0.0928	0.0498	0.1846	0.0654	0.2024	0.0596	0.2028
2000	0.0494	0.0827	0.0502	0.0710	0.0414	0.1466	0.0516	0.1598	0.0482	0.1599

CI, confidence interval; CP, coverage proportion; L, average length ($\alpha = 0.05$); SMLE, smoothed maximum likelihood estimator.

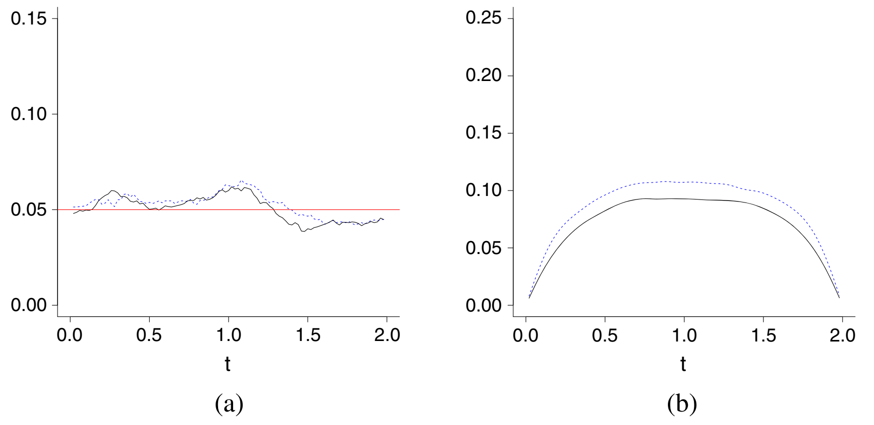

Fig. 3. Uniform samples: (a) proportion of times that $F_0(t_i)$, $t_i = 0.02, 0.04, \ldots$ is not in the 95% confidence intervals (CIs) and (b) average length of the CIs in 5000 samples using 1000 bootstrap samples with the Epanechnikov kernel (black, solid) and the triweight kernel (blue, dashed) for the Studentized smoothed maximum likelihood estimator-based CIs (10). $h = 2n^{-1/5}$. [Colour figure can be viewed at wileyonlinelibrary.com]

the left boundary of the interval and have a coverage that is less good than the Studentized CIs (10). Similar conclusions are also drawn for the exponential samples.

In contrast to the MLE-based intervals, the SMLE-based intervals in the exponential setting are subjected to bias effects. A picture of the asymptotic bias β defined in (4) is shown in Fig. 5. The function $\beta = \beta(t)$ is scaled by a factor $1000^{-2/5}$, and therefore, its magnitude corresponds to the quantity that should be subtracted from the estimated SMLE-based CIs in order to construct unbiased CIs based on $n = 1000$ observations. Accurate procedures to handle the bias are hard to obtain and still need more investigation in further research. We propose to use a combination of a local bandwidth, minimizing an estimate of the mean squared error (MSE) together with undersmoothing in order to reduce the bias effects when constructing CIs around the SMLE. Undersmoothing can be used to correct for bias when the bootstrap is used to construct CIs. As argued by Hall (1992), undersmoothing has the advantage that direct estimation of the bias is no longer necessary and can improve coverage accuracy of the CIs as well as result in narrower intervals. An improvement of the performance of bootstrap-based CIs around the SMLE as a consequence of undersmoothing is also observed in Groeneboom & Jongbloed (2014), section 9.5. (see, e.g. figure 9.19 on p. 272).

3.1. Bandwidth selection

We use a bootstrap procedure to select the optimal local bandwidth at time point t. The selection criterion is based on minimizing the MSE

$$\mathrm{MSE}(h) = E\{\tilde{F}_{nh}(t) - F_0(t)\}^2. \tag{13}$$

Because F_0 is unknown, in practice, we select, for each time point t, the constant $\hat{c}_{t,\mathrm{opt}}$ that minimizes

$$\widehat{\mathrm{MSE}}(c) = B^{-1}\sum_{b=1}^{B}\{\tilde{F}^{b}_{m,cm^{-1/5}}(t) - \tilde{F}_{n\tilde{h}_0}(t)\}^2, \tag{14}$$

where $\tilde{F}^{b}_{m,cm^{-1/5}}$ is the SMLE in a bootstrap sample $(T_1^*, \Delta_1^*), \ldots, (T_m^*, \Delta_m^*)$ of size $m < n$, where the T_i^* are sampled from a kernel estimator for the distribution function G of the

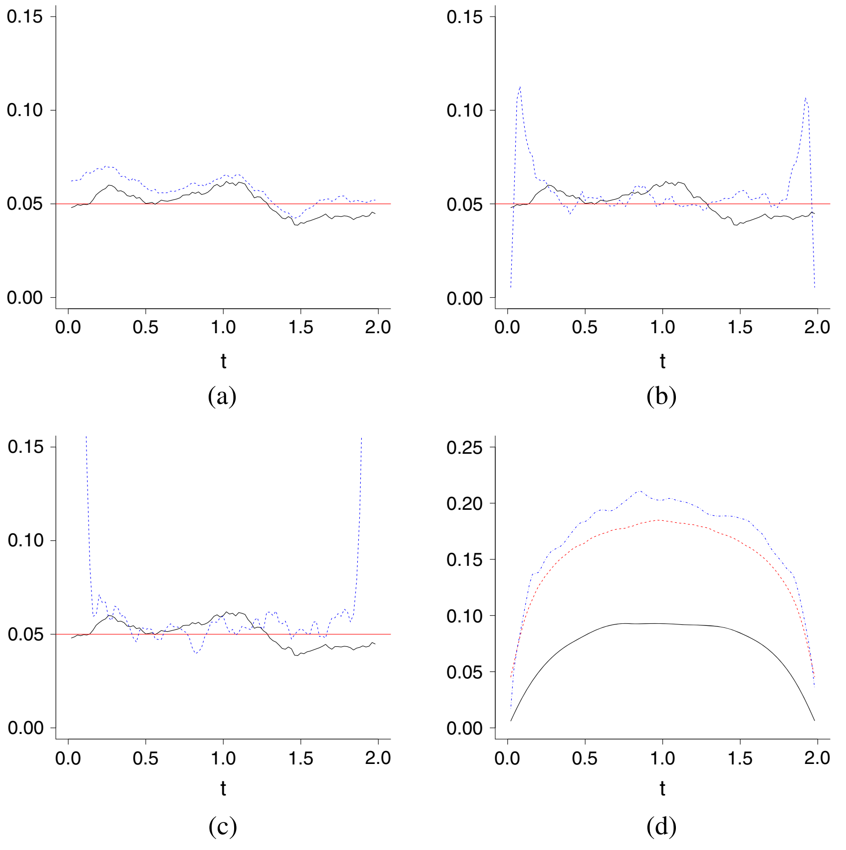

Fig. 4. Uniform samples: proportion of times that $F_0(t_i)$, $t_i = 0.02, 0.04, \ldots$ is not in the 95% confidence intervals (CIs) in 5000 samples using the Epanechnikov kernel and 1000 bootstrap samples for (a) classical smoothed maximum likelihood estimator (SMLE)-based CIs (5) (blue, dashed) and Studentized SMLE-based CIs (10) (black, solid), (b) Banerjee–Wellner CIs (blue, dashed) and Studentized SMLE-based CIs (10) (black, solid) and (c) Sen–Xu CIs (blue, dashed) and Studentized SMLE-based CIs (10) (black, solid). (d) The average length for the SMLE-based CIs (10) (black, solid), Banerjee–Wellner CIs (red, dashed) and Sen–Xu CIs (blue, dashed-dotted). $n = 1000$ and $h = 2n^{-1/5}$. [Colour figure can be viewed at wileyonlinelibrary.com]

censoring variable T and where the Δ_i^* are sampled from a Bernoulli distribution with probability $\tilde{F}_{nh_0}(T_i^*)$. Here, $\tilde{F}_{nh_0}$ denotes the SMLE in the original sample (of size n) using the bandwidth $h_0 = c_0 n^{-1/5}$ for some constant c_0, and B equals the number of bootstrap samples. A similar procedure to select the constant c when interest is in point estimation of $F_0(t)$ is proposed in Groeneboom *et al.* (2010). For each time point t, we next choose the bandwidth

$$\hat{h}_{t,\mathrm{opt}} = \hat{c}_{t,\mathrm{opt}} n^{-1/4},$$

where we use undersmoothing to reduce the bias effect in constructing CIs for $F_0(t)$.

An important point is the fact that we have to use subsampling, that is, bootstrapping with a smaller sample size, for estimating the right bandwidth in a reasonable fashion, as argued convincingly in Hall (1990). In the present case, we took $m = 100$. If one does not use subsampling, the bias/variance comparison is not performed in the right way, whereas our present scheme, taking $m = 100$ versus the original sample size $n = 1000$, seemed to give a reasonable estimate of the MSE, as was borne out by a comparison with the real MSE. We estimated MSE(c) on

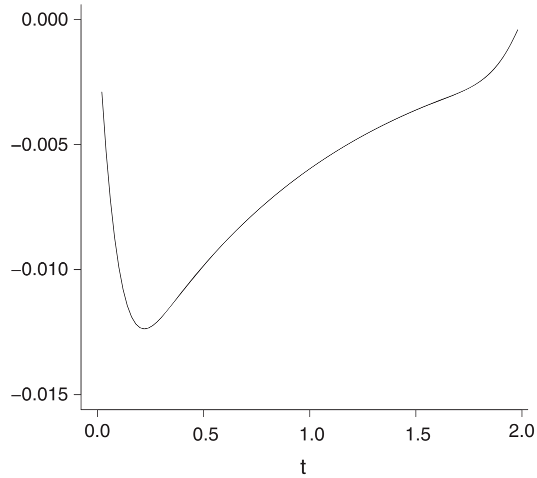

Fig. 5. Truncated exponential samples: bias for a sample of size $\boldsymbol{n = 1000}$, $\boldsymbol{h = 2n^{-1/5}}$.

a grid $c = 0.05, 0.10, \ldots, 5$, for a sample of size $n = 1000$ by a Monte Carlo experiment with $N = 1000$ simulation runs by

$$\widetilde{\mathrm{MSE}}(c) = N^{-1} \sum_{j=1}^{N} \{\tilde{F}^{j}_{n,cn^{-1/5}}(t) - F_0(t)\}^2, \tag{15}$$

where $\tilde{F}^{j}_{n,cn^{-1/5}}(t)$ is the estimate of $F_0(t)$ in the jth simulation run, $j = 1, \ldots, N$. Figure 6A compares the values of c minimizing the Monte Carlo estimate of MSE (15) with the values of c minimizing the bootstrap MSE (14) as a function of t and illustrates that the bootstrap MSE is a good estimate of (13).

Figure 6B compares the proportion of times that $F_0(t_i)$, $t_i = 0.02, 0.04, \ldots$ is not in the 95% Studentized SMLE-based CIs (10) for the truncated exponential model when a fixed bandwidth $h = 2n^{-1/5}$ is used with the proportion obtained when a local bandwidth is used. We use the bandwidth $h(t) = 0.3412{+}0.1280t$ that corresponds to the least squares regression line through the points $(t, cn^{-1/4})$ where c is the value minimizing (15) at time point t. An improvement in the coverage probabilities of the CIs is seen at the left end (i.e. the region where the bias is most prominent), indicating that it is indeed possible to obtain good CIs if undersmoothing in combination with a local optimal bandwidth is considered. The coverage proportions for the MLE-based methods of Banerjee & Wellner (2005) and Sen & Xu (2015) (results not shown) are similar to the proportions obtained for the uniform samples. Under our regularity conditions, our SMLE-based CIs have a better behaviour than the MLE-based intervals near the boundary of the intervals in terms of coverage proportions and in the middle of the interval in terms of the length of the intervals.

The CIs for one sample of size $n = 1000$ are shown in Figs 7 and 8. Note that the Sen–Xu CIs do not have monotone bounds. One may wonder if one really wants to use the MLE for estimating the distribution function and if one resamples from the SMLE as in Sen & Xu (2015) because one uses smoothness conditions that allow to estimate the distribution function at a faster rate than the convergence rate of the MLE. The pointwise CIs around the SMLE change smoothly over the interval, whereas MLE-based intervals change in discrete steps.

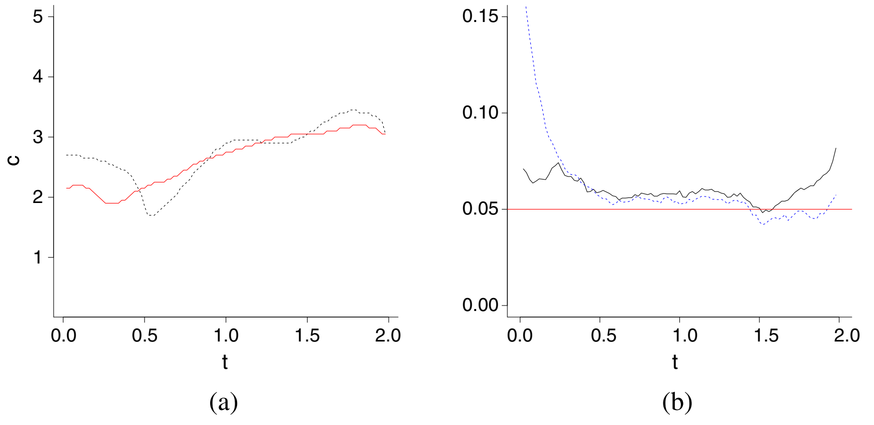

Fig. 6. Truncated exponential samples: (a) optimal bandwidth h at time point $t_0 = 0.02, 0.04, \ldots, 2$ obtained by the minimizer of $\widehat{\text{MSE}}$ (red, solid) using $N = 1000$ Monte Carlo runs and $\widehat{\text{MSE}}$ (black, dashed) using $B = 1000$ bootstrap runs of size $m = 100$ and $h_0 = 2n^{-1/5}$. (b) Proportion of times that $F_0(t_i)$, $t_i = 0.02, 0.04, \ldots$ is not in the 95% Studentized smoothed maximum likelihood estimator-based confidence intervals (10) in 5000 samples using 1000 bootstrap samples with bandwidth $h = 2n^{-1/5}$ (blue, dashed) and $h(t) = 0.3412 + 0.1280t$ (black, solid); $n = 1000$. [Colour figure can be viewed at wileyonlinelibrary.com]

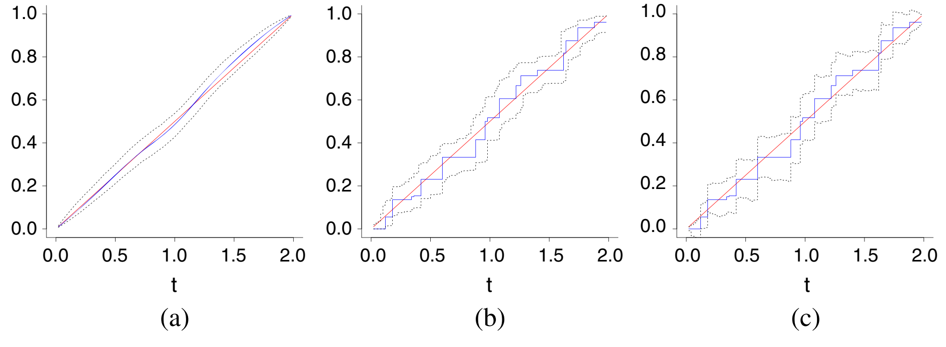

Fig. 7. Uniform samples: F_0 (red, solid). (a) Studentized smoothed maximum likelihood estimator-based confidence interval (CI) (10), (b) Banerjee–Wellner CI and (c) Sen–Xu CI based on one sample of size $n = 1000$ using 1000 bootstrap samples. In (a), the smoothed maximum likelihood estimator (blue, solid) is given, and in (b and c), the maximum likelihood estimator (blue, step function) is given; $h = 2n^{-1/5}$. [Colour figure can be viewed at wileyonlinelibrary.com]

4. Real data analysis

4.1. Hepatitis A

Keiding (1991) considered a cross-sectional study on the hepatitis A virus from Bulgaria. In 1964, samples were collected from schoolchildren and blood donors on the presence or absence of hepatitis A immunity. In total, $n = 850$ individuals ranging from 1 to 86 years old were tested for immunization. It is assumed that, once infected with hepatitis A, lifelong immunity is achieved. We are interested in estimating the seroprevalence for hepatitis A in Bulgaria. We constructed CIs at time points $t_1 = M/100, t_2 = 2M/100, \ldots, M$ where $M = 86$ is the largest observed age using the Studentized SMLE-based CIs (10) described in Section 2 using

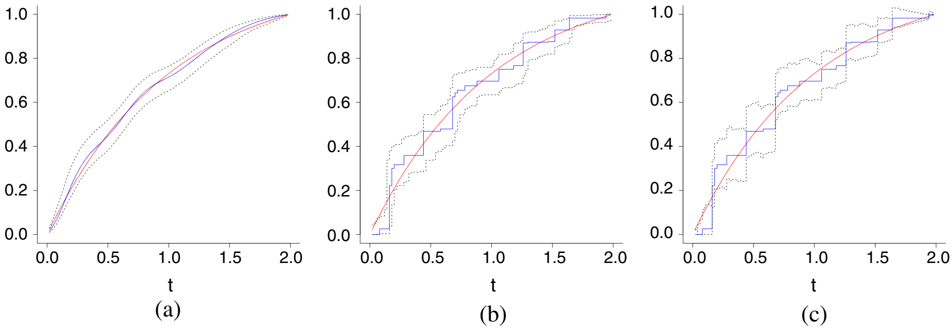

Fig. 8. Truncated exponential samples: F_0 (red, solid). (a) Studentized smoothed maximum likelihood estimator (SMLE)-based confidence interval (CI) (10), (b) Banerjee–Wellner CI and (c) Sen–Xu CI based on one sample of size $n = 1000$ using 1000 bootstrap samples. In (a), the SMLE (blue, solid) is given, and in (b and c), the maximum likelihood estimator (blue, step function) is given. $h(t) = 0.3412+0.1280t$ for SMLE-based CI and $h = 2n^{-1/5}$ for Sen–Xu CI. [Colour figure can be viewed at wileyonlinelibrary.com]

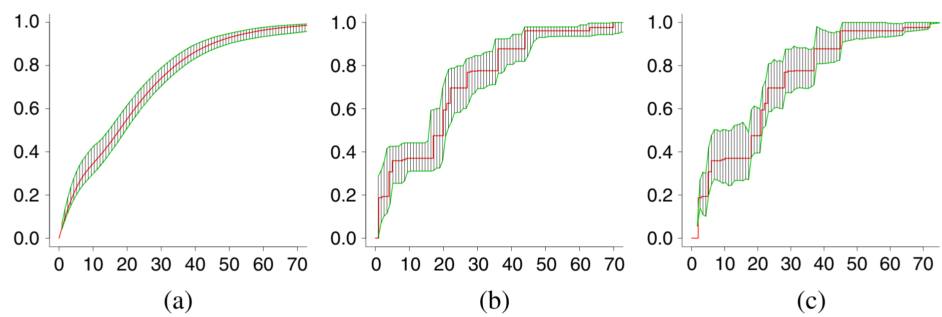

Fig. 9. Hepatitis A data: (a) Studentized smoothed maximum likelihood estimator (SMLE)-based confidence interval (CI) (10), (b) Banerjee–Wellner CI and (c) Sen–Xu CI based on $n = 850$ observations using 1000 bootstrap samples. In (a), the SMLE (red, solid) is given, and in (b and c), the maximum likelihood estimator (red, step function) is given. $h(t) = (43 + 1.5t)n^{-1/5}$ for SMLE-based CI and $h = 86n^{-1/5}$ for Sen–Xu CI. [Colour figure can be viewed at wileyonlinelibrary.com]

a local bandwidth $h(t_i) = (0.5M + 1.5t_i)n^{-1/5}$. A picture of the CIs together with the LR-based CIs of Banerjee & Wellner (2005) and the CIs of Sen & Xu (2015) is given in Fig. 9. The estimated prevalence of hepatitis A at the age of 18 years is 0.51, about half of the infections in Bulgaria happen during childhood. The length of the CIs is smallest for our SMLE-based CIs and largest for the Sen–Xu CIs. The latter CIs have left and right endpoints that are not monotone increasing in age, a property that is not shared by the other two CIs that have monotone increasing bounds. In contrast to the Banerjee–Wellner CIs, the bounds of our SMLE-based CIs are not increasing by construction.

4.2. *Rubella*

Keiding *et al.* (1996) considered a current status data set on the prevalence of rubella in 230 Austrian men older than 3 months. Rubella is a highly contagious childhood disease spread by airborne and droplet transmission. The symptoms (such as rash, sore throat, mild fever and swollen glands) are less severe in children than in adults. Because the Austrian vaccination policy against rubella only vaccinated girls, the male individuals included in the data set

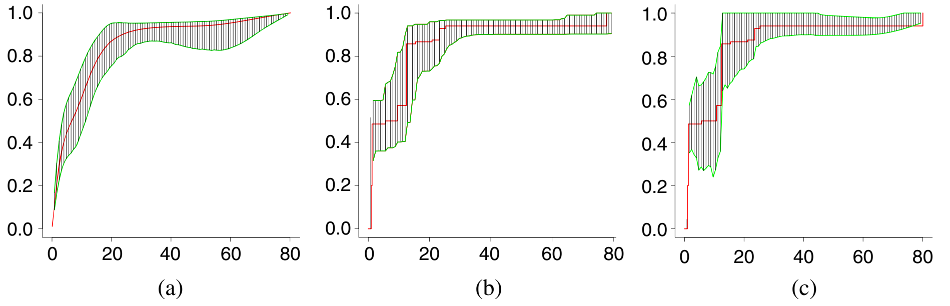

Fig. 10. Rubella data: (a) Studentized smoothed maximum likelihood estimator (SMLE)-based confidence interval (CI) (10), (b) Banerjee–Wellner CI and (c) Sen–Xu CI based on $n = 230$ observations using 1000 bootstrap samples. In (a), the SMLE (red, solid) is given, and in (b and c), the maximum likelihood estimator (red, step function) is given. $h(t) = (0.25M + t)n^{-1/5}$ if $t \leq 20$ and $h(t) = h(20) + 2(t - 20)n^{-1/5}$ else for SMLE-based CI and $h = 80.12n^{-1/5}$ for Sen–Xu CI. [Colour figure can be viewed at wileyonlinelibrary.com]

represent an unvaccinated population, and (lifelong) immunity could only be acquired if the individual obtained the disease. We are interested in estimating the time to immunization (i.e. the time to infection) against rubella using the SMLE. We constructed CIs at time points $t_1 = M/100, t_2 = 2M/100, \ldots, M$ where $M = 80.1178$ is the largest observed age, using CIs defined in (10) with the boundary correction described in Section 2 and a local bandwidth $h(t_i) = (0.25M + t_i)n^{-1/5}$ if $t_i \leq 20$ and $h(t_i) = h(20) + 2(t_i - 20)n^{-1/5}$ else. The bandwidth choice is based on the fact that most infections occurred before the age of 20 years, and a larger bandwidth is needed in the range $[20, M]$ to obtain plausible estimates. The SMLE increases steeply in the ages before adulthood that is in line with the fact that rubella is considered as a childhood disease. As can be seen from Fig. 10, our CIs and the Banerjee–Wellner CIs are favoured over the Sen–Xu intervals because of their remarkable non-increasing behaviour and their large width in the region up to 20 years. A further discussion of statistical aspects of this data set can be found in Banerjee & Wellner (2005) and Groeneboom & Jongbloed (2014).

5. Concluding remarks

In this paper, we presented a method for CI estimation for the distribution function of a random variable that cannot be observed completely because of current status censoring. The CIs are based on a smooth bootstrap procedure. Unfortunately, a rather negative feeling on the usefulness of bootstrap methods in this context is created by the results in Abrevaya & Huang (2005) and Kosorok (2008), showing that the classical bootstrap cannot be used in reproducing the 'Chernoffian' limit distribution of the MLE in current status models and of the Grenander estimator in monotone density estimation. The result in Sen *et al.* (2010) showing that even resampling from the Grenander estimator itself will not result in a consistent bootstrap has further contributed to this negative image of the bootstrap.

A positive bootstrap result, on the other hand, was derived in Sen & Xu (2015) showing that one can in fact reproduce the Chernoffian limit distribution if one resamples from a smooth estimate of the distribution function, such as the SMLE. But we meet a familiar paradox in the field here: if one introduces smoothness conditions (which is also performed in the conditions of limit theorems for the MLE), then one can usually achieve better convergence rates than the MLE achieves. For example, under the smoothness conditions of Groeneboom *et al.* (2010), the SMLE achieves rate $n^{2/5}$ (familiar from density estimation), whereas the MLE only achieves rate $n^{1/3}$ (familiar from histogram estimation). The authors of Sen & Xu (2015) however use bootstrapping by resampling the indicators Δ_i^* from the SMLE, while keeping the observation times T_i fixed in combination with intervals around the MLE instead of the smooth estimate from which the resampling is performed. It seems more natural to construct CIs on the basis of the SMLE instead of the MLE, and this is indeed what we propose in the current paper. We have shown that the procedure, based on the SMLE, gives a consistent bootstrap and has considerably smaller intervals than the intervals in Banerjee & Wellner (2005), who used LR tests, based on the (restricted) MLE, or Sen & Xu (2015) who used intervals, based on the MLE rather than the SMLE.

We showed in this article that the intervals based on the SMLE can be constructed in such a way that one obtains a better boundary behaviour, provided the necessary smoothness conditions are satisfied. The simulations also showed, not unexpectedly, that the Studentized CIs were better than the non-Studentized bootstrap CIs. In contrast to the unbiased MLE, the squared bias and variance for the SMLE are of the same order. We therefore found in our simulations that the performance of our CIs increases considerably if we subtracted the (unobserved) bias in the construction of the CIs. In practice, it is of course not possible to subtract the real bias. However, our simulations showed a remarkable improvement of the behaviour of the CIs if one uses a local bandwidth in combination with undersmoothing instead of one global bandwidth of order $n^{-2/5}$. We propose a bandwidth selection criteria based on the smooth bootstrap procedure developed in this paper and apply the concept of undersmoothing to reduce the bias effect when constructing CIs around the SMLE. Further research related to the development of criteria to decide on how to adapt the bandwidth in order to handle the bias is worth studying.

Rcpp scripts for producing the pictures of this paper and doing simulations can be found in Groeneboom (2015).

Acknowledgements

We are very grateful to Cécile Durot for communicating her approach to the proof of lemma 2 to us. 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

Abrevaya, J. & Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73**, 1175–1204.

Banerjee, M., Durot, C. & Sen, B. (2016). Divide and conquer in non-standard problems and the super-efficiency phenomenon. https://arxiv.org/abs/1605.04446 [Accessed on 17 November 2016].

Banerjee, M. & Wellner, J. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29**, 1699–1731.

Banerjee, M. & Wellner, J. A. (2005). Confidence intervals for current status data. *Scand. J. Statist.* **32**, 405–424.

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

Durot, C. & Reboul, L. (2010). Goodness-of-fit test for monotone functions. *Scand. J. Statist.* **37**, 422–441.

Groeneboom, P. (2012). Likelihood ratio type two-sample tests for current status data. *Scand. J. Statist.* **39**, 645–662.

Groeneboom, P. (2015). Rcpp scripts. https://github.com/pietg/book/tree/master/Rcpp_scripts [Accessed on 1 March 2017].

Groeneboom, P. & Hendrickx, K. (2017a). Current status linear regression. https://arxiv.org/abs/1601.00202 [Accessed on 1 April 2017].

Groeneboom, P. & Hendrickx, K. (2017b). The nonparametric bootstrap for the current status model. https://arxiv.org/abs/1701.07359 [Accessed on 25 January 2017].

Groeneboom, P. & Jongbloed, G. (2014). *Nonparametric estimation under shape constraints*, Cambridge University Press, Cambridge.

Groeneboom, P. & Jongbloed, G. (2015). Nonparametric confidence intervals for monotone functions. *Ann. Statist.* **43**, 2019–2054.

Groeneboom, P., Jongbloed, G. & Witte, B. (2010). Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model. *Ann. Statist.* **38**, 352–387.

Groeneboom, P., Lalley, S. & Temme, N. (2015). Chernoff's distribution and differential equations of parabolic and Airy type. *J. Math. Anal. Appl.* **423**, 1804–1824.

Groeneboom, P. & Wellner, J. (1992). *Information bounds and nonparametric maximum likelihood estimation*, Basel, Birkhäuser Verlag.

Groeneboom, P. & Wellner, J. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Statist.* **10**, 388–400.

Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.* **32**, 177–203.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Stat.* **20**, 675–694.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457–481.

Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective. *J. Roy. Statist. Soc. Ser. A* **154**, 371–412.

Keiding, N., Begtrup, K., Scheike, T. & Hasibeder, G. (1996). Estimation from current status data in continuous time. *Lifetime Data Anal.* **2**, 119–129.

Kosorok, M. (2008). Bootstrapping the grenander estimator. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen, Inst. Math. Stat. Collect.*, 282–292. Beachwood, OH.

Pollard, D. (1984). *Convergence of stochastic processes*, Springer-Verlag, New York.

Sen, B., Banerjee, M. & Woodroofe, M. (2010). Inconsistency of bootstrap: The grenander estimator. *Ann. Statist.* **38**, 1953–1977.

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

van de Geer, S. (2000). *Applications of empirical process theory*, Cambridge University Press, Cambridge.

Received November 2016, in final form July 2017

Kim Hendrickx, I-BioStat, Hasselt University, Agoralaan, B-B3590 Diepenbeek, Belgium.
E-mail: kim.hendrickx@uhasselt.be

Appendix

A1. Proof of theorem 1

We denote the bootstrap sample by $(T_1, \Delta_1^*), \dots, (T_n, \Delta_n^*)$. Note that the sample is produced by keeping the T_i fixed and drawing the Δ_i^* from a Bernoulli distribution with probability $\tilde{F}_{nh}(T_i)$ at each ith draw. Let $\mathbb{G}_n$ be the empirical measure of $T_1, \dots, T_n$, and let $\mathbb{P}_n^*$ denote the empirical measure of $(T_1, \Delta_1^*), \dots, (T_n, \Delta_n^*)$. We write

$$n^{-1} \sum_{i=1}^{n} f(T_i, \Delta_i^*) = \int f(u, \delta^*) \, d\mathbb{P}_n^*(u, \delta^*)$$

for some bounded function $f : [0, M] \times \{0, 1\} \to \mathbb{R}$. Note that for any bounded function $h : [0, M] \to \mathbb{R}$,

$$n^{-1}\sum_{i=1}^{n} h(T_i) = \int h(u)\, d\mathbb{P}_n^*(u,\delta^*) = \int h(u)\, d\mathbb{G}_n(u).$$

Finally, let P_n^* denote the conditional probability measure, given $(T_1,\Delta_1),\dots,(T_n,\Delta_n)$, and note that

$$P_n^*\left(\Delta_i^* = 1\right) = \tilde{F}_{nh}(T_i) \quad i = 1,\dots,n. \tag{16}$$

For the proof of theorem 1, we use the so-called 'switch relation', which reduces the study of $\hat{F}_n^*$ to the study of an inverse process. To this end, we define the process W_n^* by

$$W_n^*(t) = n^{-1}\sum_{i=1}^{n} \Delta_i^* 1_{\{T_i\le t\}}$$

and the process (in a) U_n^* by

$$U_n^*(a) = \text{argmin}\{t\in\mathbb{R} : W_n^*(t) - a\mathbb{G}_n(t)\}. \tag{17}$$

Then, taking $a_n = \tilde{F}_{nh}(t)$, we obtain the *switch relation*

$$\begin{aligned} P_n^*\left\{n^{1/3}\{\hat{F}_n^*(t) - \tilde{F}_{nh}(t)\}\ge x\right\} &= P_n^*\left\{\hat{F}_n^*(t)\ge a_n + n^{-1/3}x\right\}\\ &= P_n^*\left\{U_n^*(a_n + n^{-1/3}x)\le t\right\}. \end{aligned} \tag{18}$$

Now, let U_n be defined by

$$U_n(a) = \inf\{x\in\mathbb{R} : \tilde{F}_{nh}(x)\ge a\}, \qquad a\in(0,1). \tag{19}$$

We have the following result.

Lemma 2. *There are positive constants C_1 and C_2, such that, a.s., for all $x > 0$ and all large n,*

$$P_n^*\left\{n^{1/3}\left|U_n^*(a) - U_n(a)\right|\ge x\right\}\le C_1 e^{-C_2x^3}.$$

Note that, in the unconditional setting, there is theorem 11.3 in Groeneboom & Jongbloed (2014). Let E_n^* denote the conditional expectation, given $(T_1,\Delta_1),\dots,(T_n,\Delta_n)$; it follows from lemma 2 and the switch relation that

$$E_n^*\left\{\hat{F}_n^*(t) - \tilde{F}_{nh}(t)\right\}^2\le Kn^{-2/3} \quad \forall t\in[0,M], \tag{20}$$

which moreover implies that

$$\|\hat{F}_n^* - \tilde{F}_{nh}\|_2 = O_p^*\left(n^{-1/3}\right), \tag{21}$$

where $O_p^*\left(n^{-1/3}\right)$ means that for all $\varepsilon > 0$ and almost all sequences $(T_1,\Delta_1),(T_2,\Delta_2),\dots,$ there exists an $M > 0$ such that

$$P_n^*\left\{n^{1/3}\|\hat{F}_n^* - \tilde{F}_{nh}\|_2\ge M\right\} < \varepsilon,$$

for all large n. We also have, similarly,

$$\int_{t-h}^{t+h}\left\{\hat{F}_n^*(x) - \tilde{F}_{nh}(x)\right\}^2 dx = O_p^*\left(hn^{-2/3}\right), \tag{22}$$

conditionally on $(T_1, \Delta_1), (T_2, \Delta_2), \ldots$. See p. 320 of Groeneboom & Jongbloed (2014) for the relation of lemma 2 to these last statements. We now give the proof of theorem 1, using the result of lemma 2. The proof of lemma 2 is given at the end of this section.

Proof of theorem 1. Define the functions

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

and

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

where the τ_i are the points of jump of $\hat{F}^*_n$. By the convex minorant interpretation of $\hat{F}^*_n$, we have

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

This implies that

$$\begin{aligned} 0 &= \int \bar{\psi}^*_{t,h}(u) \left\{\delta^* - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*) \\ &= \int \psi_{t,h}(u)\left\{\delta^* - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*) + \int \left\{\bar{\psi}^*_{t,h}(u) - \psi_{t,h}(u)\right\}\left\{\delta^* - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*) \\ &= \int \psi_{t,h}(u) \left\{\delta^* - \tilde{F}^*_{nh}(u)\right\} d(\mathbb{P}^*_n - P^*_n)(u, \delta^*) \\ &\quad + \int \psi_{t,h}(u) \left\{\tilde{F}^*_{nh}(u) - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*) \\ &\quad + \int \left\{\bar{\psi}^*_{t,h}(u) - \psi_{t,h}(u)\right\} \left\{\delta^* - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*), \end{aligned}$$

where we write $d(\mathbb{P}^*_n - P^*_n)$ instead of $d\mathbb{P}^*_n$ in the last equality as a result of (16). Using integrating by parts, we have

$$\tilde{F}^*_{nh}(t) - \int \mathbb{K}_h(t-u)\, d\tilde{F}_{nh}(u) = \int \psi_{t,h}(u) \left\{\hat{F}^*_n(u) - \tilde{F}_{nh}(u)\right\} dG(u).$$

So we find

$$\begin{aligned} &\tilde{F}^*_{nh}(t) - \int \mathbb{K}_h(t-u)\, d\tilde{F}_{nh}(u) \\ &= \int \bar{\psi}^*_{t,h}(u) \left\{\delta^* - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*) - \int \psi_{t,h}(u) \left\{\tilde{F}_{nh}(u) - \hat{F}^*_n(u)\right\} dG(u) \\ &= \int \psi_{t,h}(u) \left\{\delta^* - \tilde{F}_{nh}(u)\right\} d(\mathbb{P}^*_n - P^*_n)(u, \delta^*) \\ &\quad + \int \psi_{t,h}(u) \left\{\tilde{F}_{nh}(u) - \hat{F}^*_n(u)\right\} d(\mathbb{G}_n - G)(u, \delta^*) \\ &\quad + \int \left\{\bar{\psi}^*_{t,h}(u) - \psi_{t,h}(u)\right\} \left\{\delta^* - \hat{F}^*_n(u)\right\} d\mathbb{P}^*_n(u, \delta^*) \\ &= A_{\mathrm{I}} + A_{\mathrm{II}} + A_{\mathrm{III}}. \end{aligned}$$

To study the asymptotic distribution of

$$n^{2/5}\left\{\tilde F^*_{nh}(t)-\int \mathbb{K}_h(t-u)\,d\tilde F_{nh}(u)\right\},$$

we therefore have to analyze the three terms $A_{\rm I}$, $A_{\rm II}$ and $A_{\rm III}$. We start with $A_{\rm I}$ and prove

$$n^{2/5}\int \psi_{t,h}(u)\left\{\delta^*-\tilde F_{nh}(u)\right\}d(\mathbb{P}^*_n-P^*_n)(u,\delta^*)\xrightarrow{\mathcal{D}} N\left(0,\sigma^2\right), \tag{23}$$

where σ^2 is defined in (4). Define

$$Z_{nh,i}=n^{-3/5}\psi_{t,h}(T_i)\left\{\Delta^*_i-\tilde F_{nh}(T_i)\right\}.$$

The left-hand side of (23) can then be expressed as $\sum_{i=1}^n Z_{nh,i}$. Conditionally on (T_1,X_1), …, (T_n,X_n), $Z_{nh,i}$ has mean zero and variance

$$\sigma^2_{nh,i}=n^{-6/5}\psi^2_{t,h}(T_i)\tilde F_{nh}(T_i)\left\{1-\tilde F_{nh}(T_i)\right\}.$$

Therefore, along almost all sequences $(T_1,\Delta_1),(T_2,\Delta_2)\ldots$,

$$\begin{aligned}\sum_{i=1}^n\sigma^2_{nh,i}&=n^{-1/5}\int\psi^2_{t,h}(u)\tilde F_{nh}(u)\left\{1-\tilde F_{nh}(u)\right\}d\mathbb{G}_n(u)\\&=n^{-1/5}\int\psi^2_{t,h}(u)\tilde F_{nh}(u)\left\{1-\tilde F_{nh}(u)\right\}dG(u)+o(1)\\&=\int_{-1}^1K^2(u)\tilde F_{nh}(t+hu)\left\{1-\tilde F_{nh}(t+hu)\right\}g(t+hu)du+o(1)\\&\to\frac{F_0(t)\{1-F_0(t)\}}{cg(t)}\int K^2(u)du=\sigma^2,\end{aligned}$$

where we use the a.s. convergence of $F_{nh}(t)\to F_0(t)$ in the last line. By the Lindeberg–Feller central limit theorem, we have

$$\sum_{i=1}^n Z_{nh,i}\xrightarrow{\mathcal{D}}N\left(0,\sigma^2\right).$$

This proves (23).

We next consider $A_{\rm II}$. From the fact that the integrand is the product of h^{-1} times the fixed bounded continuous function $u\mapsto K((t-u)/h)/g(u)$ and the class of functions of bounded variation $\hat F^*_n-\tilde F_{nh}$ that have entropy with bracketing of order ε^{-1} for the L_2-distance and are of order $O^*_p(n^{-1/3})$ for the L_2-distance, again conditionally on $\omega=(T_1,\Delta_1),(T_2,\Delta_2),\ldots$, it follows that $A_{\rm II}$ is of order $O^*_p(h^{-1}n^{-2/3})$. As a consequence, we have for $h\asymp n^{-1/5}$,

$$A_{\rm II}=\int\psi_{t,h}(u)\left\{\tilde F_{nh}(u)-\hat F^*_n(u)\right\}d(\mathbb{G}_n-G)(u)=o^*_p(n^{-2/5}). \tag{24}$$

We finally study the term $A_{\rm III}$. Using similar arguments as in the proof of lemma A.4 in Groeneboom *et al.* (2010), there exists a positive constant C such that

$$\left|\bar\psi^*_{t,h}(u)-\psi_{t,h}(u)\right|\le Ch^{-2}\left|\hat F^*_n(u)-\tilde F_{nh}(u)\right|, \tag{25}$$

for all u such that $\tilde f_{nh} = \tilde F'_{nh}$ is positive and continuous in a neighbourhood around u. By (16), we can write

$$A_{\mathrm{III}} = \int \left\{\bar\psi^*_{t,h}(u) - \psi_{t,h}(u)\right\}\left\{\delta^* - \tilde F_{nh}(u)\right\} d(\mathbb{P}^*_n - P^*_n)(u,\delta^*) \\ + \int \left\{\bar\psi^*_{t,h}(u) - \psi_{t,h}(u)\right\}\left\{\tilde F_{nh}(u) - \hat F^*_n(u)\right\} d\mathbb{G}_n(u). \tag{26}$$

It is clear that

$$\int \left\{\bar\psi^*_{t,h}(u) - \psi_{t,h}(u)\right\}\left\{\delta^* - \tilde F_{nh}(u)\right\} d(\mathbb{P}^*_n - P^*_n)(u,\delta^*) \\ = o^*_p\left(\int \psi_{t,h}(u)\left\{\delta^* - \tilde F_{nh}(u)\right\} d(\mathbb{P}^*_n - P^*_n)(u,\delta^*)\right),$$

which is $o^*_p(n^{-2/5})$ by (23). For the second term on the right-hand side of (26), we obtain by (25) and (22):

$$\left|\int \left\{\bar\psi^*_{t,h}(u) - \psi_{t,h}(u)\right\}\left\{\tilde F_{nh}(u) - \hat F^*_n(u)\right\} d\mathbb{G}_n(u)\right| \\ \le Ch^{-2}\int_{t-h}^{t+h}\left\{\tilde F_{nh}(u) - \hat F^*_n(u)\right\}^2 d\mathbb{G}_n(u) = O^*_p\left(h^{-1}n^{-2/3}\right) = O^*_p\left(n^{-7/15}\right). \tag{27}$$

The proof of theorem 1 now follows by (23), (24) and (27). □

In the next section, we give the proof of lemma 1 about the boundary corrected version of the SMLE.

A2. Proof of lemma 1

Proof. We have

$$\int \{\mathbb{K}_h(t-u) + \mathbb{K}_h(t+u) - \mathbb{K}_h(2M-t-u)\}\, d\tilde F^{(\mathrm{bc})}_{nh}(u) \\ = \int_{u=0}^{M} \{\mathbb{K}_h(t-u) + \mathbb{K}_h(t+u) - \mathbb{K}_h(2M-t-u)\}\, \tilde f^{(\mathrm{bc})}_{nh}(u)\, du.$$

If $t \in [h, M-h]$, we obtain, noting that $\mathbb{K}_h(t+u) = \mathbb{K}_h(2M-t-u) = 1$, if $t \in [h, M-h]$,

$$\int_{u=0}^{M} \{\mathbb{K}_h(t-u) + \mathbb{K}_h(t+u) - \mathbb{K}_h(2M-t-u)\}\, \tilde f^{(\mathrm{bc})}_{nh}(u)\, du \\ = \int_{u=0}^{M} \mathbb{K}_h(t-u)\, \tilde f^{(\mathrm{bc})}_{nh}(u)\, du \\ = \int_{u=0}^{M} \mathbb{K}_h(t-u)\int \{K_h(u-v) + K_h(u+v) + K_h(2M-u-v)\}\, d\hat F_n(v)\, du \\ = \int \left\{\int_{u=0}^{M} \mathbb{K}_h(t-u)\{K_h(u-v) + K_h(u+v) + K_h(2M-u-v)\}\, du\right\} d\hat F_n(v) \\ = \int \left\{\tilde{\mathbb{K}}_h(t-v) + \tilde{\mathbb{K}}_h(t+v) - \tilde{\mathbb{K}}_h(2M-t-v)\right\} d\hat F_n(v).$$

The last transition follows from integration by parts and the symmetry of the kernel K:

$$\begin{aligned}
&\int_{u=0}^{M} \mathbb{K}_h(t-u)\left\{K_h(u-v)+K_h(u+v)+K_h(2M-u-v)\right\} du\\
&=\left[\mathbb{K}_h(t-u)\left\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\right\}\right]_{u=0}^{M}\\
&\qquad +\int K_h(t-u)\left\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\right\} du\\
&=\int K_h(t-u)\left\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\right\} du\\
&=\int \left\{\mathbb{K}_h(t-v-hw)+\mathbb{K}_h(t+v-hw)-\mathbb{K}_h(2M-t-v-hw)\right\} K(w)\, dw\\
&=\tilde{\mathbb{K}}_h(t-v)+\tilde{\mathbb{K}}_h(t+v)-\tilde{\mathbb{K}}_h(2M-t-v),
\end{aligned}$$

if $t \in [h, M-h]$.

We likewise obtain, if $t \in [0,h]$,

$$\begin{aligned}
&\int_{u=0}^{M} \left\{\mathbb{K}_h(t-u)+\mathbb{K}_h(t+u)-\mathbb{K}_h(2M-t-u)\right\} \tilde{f}_{nh}^{(\mathrm{bc})}(u)\, du\\
&=\int_{u=0}^{M} \left\{\mathbb{K}_h(t-u)+\mathbb{K}_h(t+u)-1\right\} \tilde{f}_{nh}^{(\mathrm{bc})}(u)\, du\\
&=\int_{u=0}^{M} \left\{\mathbb{K}_h(t-u)+\mathbb{K}_h(t+u)-1\right\}\\
&\qquad \cdot \int \left\{K_h(u-v)+K_h(u+v)+K_h(2M-u-v)\right\} d\hat{F}_n(v)\, du\\
&=\int \left\{\tilde{\mathbb{K}}_h(t-v)+\tilde{\mathbb{K}}_h(t+v)-\tilde{\mathbb{K}}_h(2M-t-v)\right\} d\hat{F}_n(v).
\end{aligned}$$

In the last transition, we use integration by parts again:

$$\begin{aligned}
&\int_{u=0}^{M} \left\{\mathbb{K}_h(t-u)+\mathbb{K}_h(t+u)-1\right\}\\
&\qquad \cdot \left\{K_h(u-v)+K_h(u+v)+K_h(2M-u-v)\right\} du\\
&=\left[\left\{\mathbb{K}_h(t-u)+\mathbb{K}_h(t+u)-1\right\}\left\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\right\}\right]_{u=0}^{M}\\
&\qquad +\int_{u=0}^{M} \left\{K_h(t-u)-K_h(t+u)\right\}\left\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\right\} du\\
&=\int_{u=0}^{M} \left\{K_h(t-u)-K_h(t+u)\right\}\left\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\right\} du,
\end{aligned}$$

where we use $\mathbb{K}_h(-v)+\mathbb{K}_h(v)=1$ in the last equality (which follows from the symmetry of K). Furthermore,

$$
\begin{aligned}
&\int_{u=0}^{M} \{K_h(t-u)-K_h(t+u)\}\{\mathbb{K}_h(u-v)+\mathbb{K}_h(u+v)-\mathbb{K}_h(2M-u-v)\}\,du\\
&=\int_{w=-1}^{t/h} K(w)\{\mathbb{K}_h(t-v-hw)+\mathbb{K}_h(t+v-hw)-1\}\,dw\\
&\qquad-\int_{w=t/h}^{1} K(w)\{\mathbb{K}_h(-t-v+hw)+\mathbb{K}_h(-t+v+hw)-1\}\,dw\\
&=\int_{w=-1}^{t/h} K(w)\{\mathbb{K}_h(t-v-hw)+\mathbb{K}_h(t+v-hw)-1\}\,dw\\
&\qquad+\int_{w=t/h}^{1} K(w)\{\mathbb{K}_h(t+v-hw)+\mathbb{K}_h(t-v-hw)-1\}\,dw\\
&=\int_{w=-1}^{1} K(w)\{\mathbb{K}_h(t-v-hw)+\mathbb{K}_h(t+v-hw)-1\}\,dw\\
&=\int K(w)\{\mathbb{K}_h(t-v-hw)+\mathbb{K}_h(t+v-hw)-K_h(2M-t-v-hw)\}\,dw\\
&=\tilde{\mathbb{K}}_h(t-v)+\tilde{\mathbb{K}}_h(t+v)-\tilde{\mathbb{K}}_h(2M-t-v),
\end{aligned}
$$

again using the relation $\mathbb{K}_h(x)+\mathbb{K}_h(-x)=1$.

The case $t\in[M-h,M]$ is treated similarly. □

In the remaining subsection, we prove lemma 2 needed in the proof of theorem 1.

A3. *Proof of lemma 2*

In the proof of lemma 2, we use the following (Dvoretsky–Kiefer–Wolfowitz type) inequality from Banerjee *et al.* (2016).

Lemma 3 (lemma 8.1 of Banerjee *et al.* (2016))**.** *Let F be a distribution function on $\mathbb{R}$ with a density f supported on $[0,1]$ and bounded away from zero on $[0,1]$. Let $\mathbb{F}_n$ be the empirical distribution function associated with a sample of n observations from F and let $\mathbb{F}_n^{-1}$ be the corresponding empirical quantile function. With c a lower bound for f, we then have*

$$
\mathbb{P}\left(\sup_{t\in[0,1]}|\mathbb{F}_n^{-1}(t)-F^{-1}(t)|>x\right)\le 4\exp(-2nc^2x^2),
$$

for all n and $x>0$.

Proof of lemma 2. We follow notation, introduced in section 4.1 of Banerjee *et al.* (2016), but now applied to a bootstrap sample $(T_1,\Delta_1^*)\dots,(T_n,\Delta_n^*)$. Just as in the proof of the corresponding theorem 11.3 in Groeneboom & Jongbloed (2014), Doob's inequality and exponential centring play an important role in the proof.

Moreover, we prove the equivalent statement

$$
P_n^*\left\{\left|U_n^*(a)-U_n(a)\right|>x\right\}\le c_1\exp\left\{-c_2nx^3\right\}, \tag{28}
$$

a.s., for all large n, and constants $c_1,c_2>0$ and all $x\in(n^{-1/3},M]$. To see that this is equivalent, first note that

$$
P_n^*\left\{n^{1/3}\left|U_n^*(a)-U_n(a)\right|>x\right\}=P_n^*\left\{\left|U_n^*(a)-U_n(a)\right|>n^{-1/3}x\right\},
$$

so, if (28) holds, we obtain

$$P_n^* \left\{ n^{1/3} \left| U_n^*(a) - U_n(a) \right| > x \right\} \le c_1 \exp\left\{ -c_2 x^3 \right\},$$

for all $x > 0$. Next, note that for $x \in [0, n^{-1/3}]$,

$$c_1 \exp\left\{ -c_2 n x^3 \right\} \ge c_1 \exp\{-c_2\} \ge 1,$$

if $c_1 \ge e^{c_2}$. So we can always adapt the constants in such a way that the inequality is satisfied for $x \in [0, n^{-1/3}]$.

Furthermore, for $x \in [1, M]$, we can write

$$c_1 \exp\left\{ -c_2 n x^2 \right\} \le c_1 \exp\left\{ -(c_2/M) n x^3 \right\}.$$

So for $x \in [1, M]$, we only need an inequality with x^2 in the exponent on the right-hand side and can use lemma 3 to our advantage (see below). Finally, for $x > M$, the probability on the left-hand side of (28) is zero.

Let $\Lambda_n^* : [0,1] \to [0,1]$ be defined by $\Lambda_n^*(0) = 0$, and

$$\Lambda_n^*(i/n) = n^{-1} \sum_{j \le i} \Delta_j^*, \qquad i = 1, \ldots, n,$$

and by linear interpolation at other points of [0, 1]. Furthermore, let λ_n^* be the left-continuous slope of the greatest convex minorant of Λ_n^*. Then

$$\hat{F}_n^*(T_i) = \lambda_n^*(i/n) = \lambda_n^*(\mathbb{G}_n(T_i)),$$

where $\mathbb{G}_n$ is the empirical distribution function of the observations $T_1, \ldots, T_n$ and $\hat{F}_n^*$ is the MLE in the bootstrap sample.

We define analogously $\tilde{\lambda}_n = \tilde{F}_{nh} \circ G^{-1}$, and

$$\tilde{\Lambda}_n(t) = \int_0^t \tilde{\lambda}_n(u)\, du = \int_0^t \tilde{F}_{nh}\left(G^{-1}(u) \right) du, \qquad t \in [0,1].$$

Moreover, we define

$$V_n = \tilde{\lambda}_n^{-1}. \tag{29}$$

With these definitions, we have

$$U_n = G^{-1} \circ \tilde{\lambda}_n^{-1} = G^{-1} \circ V_n, \tag{30}$$

where U_n is defined by (19). By the model assumptions at the beginning of Section 2 for F_0 and G, and the almost sure convergence of $\tilde{F}_{nh}$ and its derivative to F_0 and f_0, respectively, uniformly on $[0, M]$ (using the suggested boundary correction near 0 and M), we may assume that there is a constant $c > 0$ such that $\tilde{\lambda}_n'(t) \ge c$ for all $t \in [0,1]$ and all large n and that therefore, using a Taylor expansion, we obtain

$$\tilde{\Lambda}_n(t) - \tilde{\Lambda}_n(V_n(a)) \ge \left(t - V_n(a)\right)a + \frac{1}{2} c \left(t - V_n(a)\right)^2, \tag{31}$$

for all $t, a \in [0,1]$.

We similarly define

$$V_n^*(a) = \text{argmin}_{u\in[0,1]}\{\Lambda_n^*(u) - au\},$$

where argmin denotes the smallest location of the minimum. Note that, analogously to (30), we have for U_n^* as defined by (17):

$$U_n^* = \mathbb{G}_n^{-1} \circ V_n^*. \tag{32}$$

By the transition of U_n and U_n^* to V_n and V_n^*, respectively, the range of U_n and U_n^* is changed from $[0, M]$ to $[0, 1]$. We now prove

$$P_n^* \left\{\left|V_n^*(a) - V_n(a)\right| > x\right\} \le c_1 \exp\left\{-c_2 n x^3\right\}, \tag{33}$$

a.s., for all large n, and constants $c_1, c_2 > 0$ and all $x \in (n^{-1/3}, 1]$. Note that the probability on the left-hand side of (33) is zero if $x > 1$.

Define

$$\varepsilon_i^* = \Delta_i^* - \tilde F_{nh}(T_i), \qquad i = 1, \ldots, n.$$

Then

$$\begin{aligned}\Lambda_n^*(i/n) &= n^{-1}\sum_{j\le i}\varepsilon_j^* + n^{-1}\sum_{j\le i}\tilde F_{nh}\left(\mathbb{G}_n^{-1}(j/n)\right)\\ &= n^{-1}\sum_{j\le i}\varepsilon_j^* + \int_0^{i/n}\tilde F_{nh}\left(\mathbb{G}_n^{-1}(u)\right)du, \qquad i = 1,\ldots,n,\end{aligned}$$

using the piecewise constancy of $\mathbb{G}_n^{-1}$.

This gives

$$\begin{aligned}&P_n^*\left\{\left|V_n^*(a) - V_n(a)\right| > x\right\}\\ &\le P_n^*\left\{\min_{i:\,|V_n(a)-i/n|>x}\{\Lambda_n^*(i/n) - a\,i/n\} \le \Lambda_n^*(V_n(a)) - aV_n(a)\right\}\\ &\le P_n^*\left\{\min_{i:\,|V_n(a)-i/n|>x}\left\{D_n^*(i/n) - D_n^*(V_n(a)) + \tfrac12 c\left(in^{-1} - V_n(a)\right)^2\right\} \le 0\right\},\end{aligned}$$

where D_n^* is defined by $D_n^* = \Lambda_n^* - \tilde\Lambda_n$ and where we use (31) in the last step. Define

$$B_n^*(t) = D_n^*(t) - \int_0^t\left\{\tilde F_{nh}\left(\mathbb{G}_n^{-1}(u)\right) - \tilde F_{nh}\left(G^{-1}(u)\right)\right\}du.$$

Then

$$B_n^*(i/n) = n^{-1}\sum_{j\le i}\varepsilon_i^*.$$

Moreover, the event $\{\left|V_n^*(a) - V_n(a)\right| > x\}$ is contained in the union of the events

$$E_{n1} = \left\{\sup_{u:\,|V_n(a)-u|>x}\left\{\int_{V_n(a)}^u\left\{\tilde F_{nh}\left(G^{-1}(t)\right) - \tilde F_{nh}\left(\mathbb{G}_n^{-1}(t)\right)\right\}dt - \frac{c}{4}(u - V_n(a))^2\right\} \ge 0\right\}$$

and

$$E_{n2} = \left\{\sup_{i:\,|V_n(a)-i/n|>x}\{B_n^*(V_n(a)) - B_n^*(i/n) - \frac{c}{4}(in^{-1} - V_n(a))^2\} \ge 0\right\}.$$

We have, by the mean value theorem and the bounded differentiability of $\tilde{F}_{nh}$,

$$\left|\int_{V_n(a)}^{u}\left\{\tilde{F}_{nh}\left(G^{-1}(t)\right)-\tilde{F}_{nh}\left(\mathbb{G}_n^{-1}(t)\right)\right\}dt\right|\le c'\left|u-V_n(a)\right|\sup_{t\in[0,1]}\left|\mathbb{G}_n^{-1}(t)-G^{-1}(t)\right|.$$

for a constant $c' > 0$. Hence, we obtain from lemma 3 in the original space:

$$\begin{aligned}P_n(E_{n1})&\le P_n\left\{\sup_{t\in[0,1]}\left|\mathbb{G}_n^{-1}(t)-G^{-1}(t)\right|\ge\frac{cx}{4c'}\right\}\le 4\exp\left\{-Knc^2x^2\right\}\\&\le 4\exp\left\{-Kc^2n^{1/3}\right\},\end{aligned}\tag{34}$$

for some $K > 0$ and $x \in (n^{-1/3}, M]$. This means that we may assume that, a.s., the complement of E_{n1} is satisfied for all large n and all $x \in (n^{-1/3}, 1]$. So we now turn to $P_n^*(E_{n2})$.

We have

$$\begin{aligned}P_n^*(E_{n2})&\le\sum_{k\ge1}P_n^*\left(\sup_{i:|V_n(a)-i/n|\in(kx,(k+1)x]}\left\{B_n^*(V_n(a))-B_n^*(i/n)-\frac{c}{4}(i/n-V_n(a))^2\right\}\ge0\right)\\&\le\sum_{k\ge1}P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}\{B_n^*(V_n(a))-B_n^*(i/n)\}\ge\frac{c}{4}k^2x^2\right).\end{aligned}$$

Using the piecewise linearity of B_n^*, we obtain

$$B_n^*(V_n(a))=B_n^*\left(\frac{\lfloor nV_n(a)\rfloor}{n}\right)+\left(V_n(a)-\frac{\lfloor nV_n(a)\rfloor}{n}\right)\varepsilon^*_{\lfloor nV_n(a)\rfloor+1},$$

where $\lfloor nV_n(a)\rfloor$ denotes the integer part (floor) of $nV_n(a)$. Hence,

$$\begin{aligned}P_n^*(E_{n2})&\le\sum_{k\ge1}P_n^*\left(\left(V_n(a)-\frac{\lfloor nV_n(a)\rfloor}{n}\right)\varepsilon^*_{\lfloor nV_n(a)\rfloor+1}\ge\frac{c}{8}k^2x^2\right)\\&\quad+\sum_{k\ge1}P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}\left\{\sum_{j\le nV_n(a)}\varepsilon_j^*-\sum_{j\le i}\varepsilon_j^*\right\}\ge\frac{nc}{8}k^2x^2\right).\end{aligned}\tag{35}$$

The Markov inequality implies that for all $\theta > 0$, $k \ge 1$, $a \in [0, 1]$ and $x \in (n^{-1/3}, 1]$,

$$\begin{aligned}&P_n^*\left\{\left(V_n(a)-\frac{\lfloor nV_n(a)\rfloor}{n}\right)\varepsilon^*_{\lfloor nV_n(a)\rfloor+1}\ge\frac{c}{8}k^2x^2\right\}\\&\le\exp\left\{-\frac{\theta c}{8}k^2x^2\right\}E_n^*\exp\left\{\theta\left(V_n(a)-\frac{\lfloor nV_n(a)\rfloor}{n}\right)\varepsilon^*_{\lfloor nV_n(a)\rfloor+1}\right\},\end{aligned}$$

where E_n^* denotes the expectation under P_n^*. Because $\varepsilon_i^* \in [-1, 1]$ for all i, we have $\exp(\alpha\varepsilon_i^*) \le K\exp(\alpha^2)$ for all $\alpha \in \mathbb{R}$ and $K \ge \exp(1)$, and therefore, with $\theta = ck^2x^2n^2/16$, we obtain

$$P_n^*\left(\left(V_n(a)-\frac{\lfloor nV_n(a)\rfloor}{n}\right)\varepsilon^*_{\lfloor nV_n(a)\rfloor+1}\geq\frac{c}{8}k^2x^2\right)\leq K\exp\left(-\frac{\theta c}{8}k^2x^2+\frac{\theta^2}{n^2}\right)$$
$$\leq K\exp\left(-\frac{c^2k^4x^4n^2}{256}\right).$$

Using that $k^4\geq k$ for all $k\geq 1$ and $nx\geq 1$ for all $x\in(n^{-1/3},1)$, we conclude that for all $a\in[0,1]$ and $x\in(n^{-1/3},1)$,

$$\begin{aligned}&\sum_{k\geq 1}P_n^*\left(\left(V_n(a)-\frac{\lfloor nV_n(a)\rfloor}{n}\right)\varepsilon^*_{\lfloor nV_n(a)\rfloor+1}\geq\frac{c}{8}k^2x^2\right)\\&\leq K\sum_{k\geq 1}\exp\left(-\frac{c^2kx^3n}{256}\right)\leq K\exp\left(-\frac{c^2x^3n}{256}\right)\sum_{k\geq 0}\exp\left(-\frac{c^2kx^3n}{256}\right)\\&\leq K'\exp(-K_2nx^3),\end{aligned}\tag{36}$$

with any finite K' that satisfies $K'\geq K\sum_{k\geq 0}\exp\left(-c^2k/256\right)$ and $K_2\leq c^2/256$. This takes care of the first term on the right of (35).

We now consider the second term on the right of (35). Just as in the proof of theorem 11.3 in Groeneboom & Jongbloed (2014), we use Doob's submartingale inequality, this time conditionally on $(T_1,\Delta_1),\ldots,(T_n,\Delta_n)$. This gives

$$\begin{aligned}&P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\leq(k+1)x}\left\{\sum_{j\leq nV_n(a)}\varepsilon_j^*-\sum_{j\leq i}\varepsilon_j^*\right\}\geq\frac{nc}{8}k^2x^2\right)\\&\leq 2\exp\left(-\frac{\theta nc}{8}k^2x^2\right)\sup_{i:\,|V_n(a)-i/n|\leq(k+1)x}E_n^*\left[\exp\left(\theta\left(\sum_{j\leq nV_n(a)}\varepsilon_j^*-\sum_{j\leq i}\varepsilon_j^*\right)\right)\right].\end{aligned}$$

Suppose $i/n<V_n(a)$. Then we obtain

$$\begin{aligned}&\log E_n^*\left[\exp\left(\theta\left(\sum_{j\leq nV_n(a)}\varepsilon_j^*-\sum_{j\leq i}\varepsilon_j^*\right)\right)\right]=\log E_n^*\left[\exp\left(\theta\left(\sum_{i<j\leq nV_n(a)}\varepsilon_j^*\right)\right)\right]\\&=\sum_{i<j\leq nV_n(a)}\log\left\{\exp\left\{\theta\{1-\tilde F_{nh}(T_j)\}\right\}\tilde F_{nh}(T_j)+\exp\left\{-\theta\tilde F_{nh}(T_j)\right\}\{1-\tilde F_{nh}(T_j)\}\right\}\\&=n\int_{i/n}^{V_n(a)}\log\left\{\exp\left\{\theta\{1-\tilde F_{nh}(\mathbb{G}_n^{-1}(t))\}\right\}\tilde F_{nh}(\mathbb{G}_n^{-1}(t))\right.\\&\qquad\left.+\exp\left\{-\theta\tilde F_{nh}(\mathbb{G}_n^{-1}(t))\right\}\left\{1-\tilde F_{nh}(\mathbb{G}_n^{-1}(t))\right\}\right\}dt.\end{aligned}$$

Because $\log(1+x)\leq x$, this is bounded earlier by

$$n\int_{i/n}^{V_n(a)}\left\{\exp\left\{\theta\{1-\tilde F_{nh}(\mathbb{G}_n^{-1}(t))\}\right\}\tilde F_{nh}(\mathbb{G}_n^{-1}(t))\right.$$

$$
\begin{aligned}
&+\exp\left\{-\theta\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\}\left\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\}-1\Bigg\}\,dt\\
&\le n\int_{V_n(a)-(k+1)x}^{V_n(a)}\Bigg\{\exp\left\{\theta\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\}\right\}\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\\
&\quad+\exp\left\{-\theta\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\}\left\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\}-1\Bigg\}\,dt,\\
&=n\int_{V_n(a)-(k+1)x}^{V_n(a)}\Bigg\{\sum_{i=2}^{\infty}\frac{\theta^i}{i!}\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\}^i\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\\
&\quad+\sum_{i=2}^{\infty}\frac{\theta^i}{i!}(-1)^i\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))^i\left\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\}\Bigg\}\,dt\\
&=n\sum_{i=2}^{\infty}\frac{\theta^i}{i!}\int_{V_n(a)-(k+1)x}^{V_n(a)}\Bigg\{\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\}^i\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\\
&\quad+(-1)^i\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))^i\left\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\}\Bigg\}\,dt,
\end{aligned}
$$

if $i/n < V_n(a)$ and $|V_n(a)-i/n| \le (k+1)x$. Because $V_n(a) \in [0,1]$, the integrand,

$$
\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\}^i\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))+(-1)^i\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))^i\left\{1-\tilde{F}_{nh}(\mathbb{G}_n^{-1}(t))\right\},
$$

is bounded by 1/2, we obtain

$$
\begin{aligned}
&P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}\left\{\sum_{j\le nV_n(a)}\varepsilon_j^*-\sum_{j\le i}\varepsilon_j^*\right\}\ge\frac{nc}{8}k^2x^2\right)\\
&\le 2\exp\left(-\frac{\theta nc}{8}k^2x^2\right)\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}E_n^*\left[\exp\left(\theta\left(\sum_{j\le nV_n(a)}\varepsilon_j^*-\sum_{j\le i}\varepsilon_j^*\right)\right)\right]\\
&\le 2\exp\left(-\frac{\theta nck^2x^2}{8}+\frac{n(k+1)x}{2}\sum_{i=2}^{\infty}\frac{\theta^i}{i!}\right),
\end{aligned}
$$

for all $x \in (n^{-1/3},1), k \ge 1, a \in [0,1]$ and $\theta > 0$. Therefore, with $\theta = \log(1+\frac{ck^2x}{4(k+1)})$, we arrive at

$$
\begin{aligned}
&P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}\left\{\sum_{j\le nV_n(a)}\varepsilon_j^*-\sum_{j\le i}\varepsilon_j^*\right\}\ge\frac{nc}{8}k^2x^2\right)\\
&\le 2\exp\left(\frac{n(k+1)x}{2}\left\{\frac{ck^2x}{4(k+1)}-\left(1+\frac{ck^2x}{4(k+1)}\right)\log\left(1+\frac{ck^2x}{4(k+1)}\right)\right\}\right).
\end{aligned}
$$

Following Pollard (1984), in his discussion of Bennett's inequality on p. 192, we introduce the function B, defined by $B(0) = 1/2$ and

$$B(u) = u^{-2}\{(1+u)\log(1+u) - u\}.$$

Making the change of variables $u_k = ck^2x/(4(k+1))$, we can write

$$P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}\left\{\sum_{j\le nV_n(a)}\varepsilon_j^* - \sum_{j\le i}\varepsilon_j^*\right\} \ge \frac{nc}{8}k^2x^2\right)$$
$$\le 2\exp\left(-\frac{nc^2k^4x^3}{32(k+1)}B(u_k)\right).$$

Because u_k varies over a finite interval $[0, M']$ and therefore $B(u_k)$ stays away from zero on $[0, M']$, we find that

$$\sum_{k\ge1} P_n^*\left(\sup_{i:\,|V_n(a)-i/n|\le(k+1)x}\left\{\sum_{j\le nV_n(a)}\varepsilon_j^* - \sum_{j\le i}\varepsilon_j^*\right\} \ge \frac{nc}{8}k^2x^2\right)$$
$$\le K_1\sum_{k\ge1}\exp\left(-\frac{nc^2k^4x^3}{32(k+1)}\right) \le K_1\exp\left(-\frac{nc^2x^3}{64}\right)\sum_{k\ge0}\exp\left(-\frac{c^2k^4}{32(k+1)}\right)$$
$$\le K_2\exp\left(-K_3nx^3\right),$$

for appropriate K_1, K_2 and K_3. Combining this with (34) and (36), it follows that

$$P_n^*\left\{\left|V_n^*(a) - V_n(a)\right| > x\right\} \le c_1\exp\{-nc_2x^3)$$

for all large n, a.s. along $(T_1, \Delta_1), \ldots$ for constants $c_1, c_2 > 0$ and $x \in (n^{-1/3}, 1]$.

We now prove that (28) also follows by considering the transition of V_n and V_n^* to U_n and U_n^*. By (30) and (32), we obtain

$$U_n^*(a) - U_n(a) = \mathbb{G}_n^{-1}\circ V_n^*(a) - G^{-1}\circ V_n(a),$$

and hence,

$$\left|U_n^*(a) - U_n(a)\right| \le \sup_{t\in[0,1]}\left|\mathbb{G}_n^{-1}(t) - G^{-1}(t)\right| + k_1\left|V_n^*(a) - V_n(a)\right|,$$

where

$$k_1 = 1/\inf_{x\in[0,M]} g(x).$$

From lemma 3, we obtain in the original space

$$P_n\left\{\sup_{t\in[0,1]}\left|\mathbb{G}_n^{-1}(t) - G^{-1}(t)\right| \ge x/2\right\} \le 4\exp\left\{-Kn^{1/3}\right\},$$

for some $K > 0$ and $x \in (n^{-1/3}, M]$. So we may assume that, a.s., $\left|\mathbb{G}_n^{-1}(t) - G^{-1}(t)\right| < x/2$, for all large n and all $x \in (n^{-1/3}, 1]$. By the foregoing proof, we also have

$$P_n^*\left\{k_1\left|V_n^*(a) - V_n(a)\right| \ge x/2\right\} \le c_1\exp\left\{-c_2nx^3/\left(8k_1^3\right)\right\}.$$

This proves the result. □