FUNCTIONALS OF NONPARAMETRIC MAXIMUM LIKELIHOOD
ESTIMATORS

BY PIET GROENEBOOM

Delft University of Technology

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

Nonparametric maximum likelihood estimators (MLEs) in inverse problems often have non-normal limit distributions, like Chernoff’s distribution. However, if one considers smooth functionals of the model, with corresponding functionals of the MLE, one gets normal limit distributions and faster rates of convergence. We demonstrate this for interval censoring models and a model for the incubation time of Covid-19. The usual approach in the latter models is to use parametric distributions, like Weibull and gamma distributions, which leads to inconsistent estimators. Smoothed bootstrap methods are discussed for choosing a bandwidth and constructing confidence intervals. The classical bootstrap, based on the nonparametric MLE itself, has been proved to be inconsistent in this situation.

1. Introduction. In the original treatment of classical statistical inverse problems such as the current status model it was assumed that the nonparametric maximum likelihood estimator (MLE) would converge as a process at \( \sqrt{n} \) rate and in particular would be “tight”. It was also conjectured that the pointwise limit distribution would be normal ([26], [30], [21]). But it was proved in [6] that, in contrast to the situation for right-censored data, the process is not tight, does not pointwise converge at \( \sqrt{n} \) rate, and that the actual pointwise limit distribution is also not normal, but in fact given by Chernoff’s distribution (see [5] and [20]). This fact was for example noticed in [33], who refer for the result to [19], where it is also given.

On the other hand, if we consider differentiable functionals of the MLE, we get normal limit distributions again. Theorem 3.1 on p. 183 of [32] gives necessary and sufficient conditions for a functional to be differentiable in a very general setting. This fundamental difference between the nonparametric MLE and smooth functionals of the MLE plays a key role in the present paper.

The current status model is the simplest of the models we are going to consider and about which we know the most. In the current status model we have independent nonnegative random variables \( X_1, X_2, \ldots \), with distribution function \( F_0 \) and independent random variables \( T_1, T_2, \ldots \) with distribution function \( G \). Moreover, the \( T_i \)'s and \( X_j \)'s are also independent. For a sample size \( n \), the data available to us are given by

\[
(T_1, \Delta_1), \ldots, (T_n, \Delta_n),
\]

where \( \Delta_i = 1_{\{X_i \leq T_i\}} \). Following usual customs in probability theory, we shall denote indicators like \( 1_{\{X_i \leq T_i\}} \) also by \( \{X_i \leq T_i\} \). A nonparametric MLE of the distribution function \( F_0 \) of the (hidden) random variables \( X_1, \ldots, X_n \) maximizes the log likelihood

\[
\sum_{i=1}^{n} \Delta_i \log F(T_i) + (1 - \Delta_i) \log \{1 - F(T_i)\}
\]

AMS 2000 subject classifications: Primary 62G05, 62N01; secondary 62-04.

Keywords and phrases: nonparametric MLE, nonparametric SMLE, Chernoff’s distribution, kernel estimates, smoothed bootstrap, incubation time, COVID-19.
over all distribution functions $F$ on $\mathbb{R}_+$. Note that the complete log likelihood would also involve the density of the observation times $T_i$, but this part of the likelihood factors out because of the independence condition.

The following result is proved in [6], but also in [19] and [18]. The latter reference will be our standard reference for results of this type.

**Theorem 1.** Let $t_0$ be such that $0 < F_0(t_0), G(t_0) < 1$, and let $F_0$ and $G$ be differentiable at $t_0$, with strictly positive derivatives $f_0(t_0)$ and $g(t_0)$, respectively. Furthermore, let $\hat F_n$ be the MLE of $F_0$. Then we have, as $n \to \infty$,

$$n^{1/3} \left\{ \hat F_n(t_0) - F_0(t_0) \right\} / \left\{ 4 F_0(t_0)(1 - F_0(t_0)) f_0(t_0)/g(t_0) \right\}^{1/3} \overset{D}{\to} Z,$$

where $\overset{D}{\to}$ denotes convergence in distribution, and where $Z$ is the last time where standard two-sided Brownian motion plus the parabola $y(t) = t^2$ reaches its minimum.

In this result we see the cube root rate of convergence and Chernoff’s limit distribution, alluded to above. The MLE can be computed in one step as the left-continuous slope of the greatest convex minorant of the so-called cusum diagram consisting of the point $(0, 0)$ and the points

$$\left( i/n, V_n(T_{(i)}) \right), \quad i = 1, \ldots, n,$$

where $T_{(i)}$ is the $i$th order statistic of observation times $T_1, \ldots, T_n$, and

$$V_n(t) = n^{-1} \sum_{i=1}^n \{ T_i \leq t \}, \quad t \geq 0,$$

see p. 68 of [18]. Here we assume that there are no ties in the observations $T_1, \ldots, T_n$, which will be true if the variables $T_i$ come from a continuous distribution.

If we assume smoothness (as we in fact do in Theorem 1), it seems more natural to use a continuous estimator of the distribution function, which is (for example) given by the smoothed maximum likelihood estimator (SMLE). This estimator is defined by first computing the ordinary nonparametric MLE $\hat F_n$ and then smoothing the MLE by a smoothing kernel. More specifically, define the SMLE $\hat F_{nh}(t)$ at points $t$ away from the boundary by

$$\hat F_{nh}(t) = \int IK_h(t - x) d\hat F_n(x), \quad IK_h(x) = IK(x/h),$$

where $IK$ is an integrated kernel, defined by

$$IK(x) = \int_{-\infty}^{x} K(y) dy,$$

and $K$ is a symmetric kernel of the usual kind, used in density estimation. Near the boundary we use the Schuster-type boundary correction, also used in Section 11.3 of [18] in the definition of the SMLE. We will assume that $K$ has support $[-1, 1]$. For this estimator we have the following result (Theorem 4.2 p. 365 in [17] and Theorem 11.10 in [18]):

**Theorem 2.** Let the distribution corresponding to $F_0$ have support $[0, M]$ and let $F_0$ have a continuous density $f_0$ staying away from zero on $[0, M]$. Furthermore, let $G$ have a density $g$ with a support that contains $[0, M]$ and let $g$ stay away from zero on $[0, M]$, with a bounded derivative $g'$. Finally, let $t$ be an interior point of $[0, M]$ such that $f_0$ has a
continuous derivative \( f'_0(t) \neq 0 \) at \( t \). Then, if \( h \sim cn^{-1/5} \) and the SMLE \( \tilde{F}_{n,h} \) is defined by (1.3),

\[
\frac{n^{2/5}}{2} \left\{ \tilde{F}_{n,h}(t) - F_0(t) \right\} \xrightarrow{D} N(\mu, \sigma^2),
\]

where

\[
\mu = \frac{1}{2} c^2 f'_0(t) \int u^2 K(u) \, du
\]

and

\[
\sigma^2 = \frac{F_0(t) \{1 - F_0(t)\}}{cg(t)} \int K(u)^2 \, du.
\]

Note the faster convergence in Theorem 2 compared to the convergence in Theorem 1 and also the asymptotic normality of the SMLE \( \tilde{F}_{n,h} \). A crucial difference between Theorem 1 and Theorem 2 is that a bandwidth is introduced in Theorem 2. So it is very important to have a procedure which gives a method for producing a good bandwidth.

Methods for choosing the bandwidth using subsampling are discussed in [17], [15], [13] and [14]. They were inspired by a method proposed in [22]. A method which does not use subsampling is suggested in [28] to estimate the locally optimal \( \tilde{h}_n(t) = \tilde{c}(t)n^{-1/5} \) by performing a smoothed bootstrap experiment, where we minimize over \( c > 0 \):

\[
B^{-1} \sum_{i=1}^{B} \left\{ \tilde{F}_{n,cn^{-1/5}}^*(t) - \tilde{F}_{n,cn^{-1/5}}(t) \right\}^2,
\]

over \( B \) bootstrap experiments, drawing samples of size \( n \) from the SMLE \( \tilde{F}_{n,cn^{-1/5}} \) for a “pilot bandwidth” \( c_0n^{-1/5} \). However, as noticed in an analogous procedure for ordinary density estimation in [23], this does not seem to lead to the choice of a bandwidth that is asymptotically equivalent to the optimal bandwidth.

We discuss this matter in Section 2, where we suggest to take a pilot bandwidth of order \( n^{-1/9} \). We then discuss the construction of pointwise confidence intervals, based on the SMLE. These intervals are considerably narrower than the intervals based on the MLE, as proposed in [2] and [28], and also behave better near the boundary (see, in particular, Figure 3).

We then apply the theory to a model for the distribution of the incubation time of Covid-19, which was the original motivation of this paper. We show that this model is equivalent to a special case of the so-called mixed case interval censoring model and prove the convergence in distribution of the nonparametric MLE to Chernoff’s distribution. This result is of the same type as a result proved for the MLE in the interval censoring, case 2, model in [8]. It is somewhat unfortunate that no other proofs of the latter result have been given since 1996 when this proof was published, since the proof is rather complicated.

Anticipating proofs of this type for the more general cases of interval censoring, a proof using the characterization of the MLE as a stationary point of the iterative convex minorant algorithm with random “self-induced” weights was given for the current status model in [19], section 5.1 of Part 2 (and previously in the Stanford lectures [7] which in fact coincide with Part 2 of [19]). An easier straightforward proof with non-random weights for the current status model which cannot be used to this end was sketched in the exercises of the same chapter in [19].
The process that characterizes the nonparametric MLE for the incubation time model is the nonnegative process $W_{n,F_n}$ given by

$$W_{n,F_n}(t) = \int_{\min(S_i) \leq s \leq \max(S_i - E_i)} \frac{1}{\tilde{F}_n(s) - \hat{F}_n(s - e)} d\mathbb{Q}_n(e, s)$$

$$\mathbb{Q}_n$$ is the empirical distribution of the incubation time data $(E_1, S_1), \ldots, (E_n, S_n)$, and where the $E_i$ and $S_i$ are the exposure times and the times of becoming symptomatic, respectively, see Lemma 2 in Section 5.

The key to the proof of the limit result for the nonparametric MLE is showing that for the asymptotics $\tilde{F}_n(s - e)$ can be replaced by $F_0(s - e)$ in the first term and $\hat{F}_n(s)$ can be replaced by $F_0(s)$ in the second term of (1.8), where $F_0$ is the distribution function of the incubation time. This proof heavily relies on smooth functional theory and properties of integral equations, and showing this is the hard part of the proof and something that is not needed at all in the proof of the limit result for the MLE in the current status model. In this way the proof becomes intimately connected with the other results on smooth functionals in the present paper. Because of the technical nature of these methods, we put the proofs in the supplementary material.

We then compare the behavior of smooth estimates, based on the nonparametric MLE, with the parametric estimates which are usually used in the epidemiological literature and which are inherently inconsistent, since there is no reason to assume the distribution of the incubation time to be of a particular parametric type.

2. Bandwidth selection. We consider observations of the form (1.1) under the conditions of Theorem 2. It is proved in [17] that, under the conditions of Theorem 4.2, the locally optimal bandwidth for the smoothed maximum likelihood estimator (SMLE) $\tilde{F}_{nh}(t)$ is given by

$$\tilde{h}_n = \tilde{c}(t)n^{-1/5},$$

where

$$\tilde{c}(t) = \left\{ \frac{F_0(t)}{g(t)} \int K(u)^2 du \right\}^{1/5} \left\{ \int \frac{f'_0(t)^2}{u^2} \left( \int u^2 K(u) du \right) \right\}^{-1/5}.$$

As discussed in Section 1, it is suggested in [28] to estimate the optimal $\tilde{h}_n(t) = \tilde{c}(t)n^{-1/5}$ by performing a smoothed bootstrap experiment, where we minimize (1.7) over $c > 0$, drawing samples of size $n$ from the SMLE $\tilde{F}_{nh,c_0n^{-1/5}}$ for a “pilot bandwidth” $h_0 = c_0n^{-1/5}$. We investigate the conditional variance + squared bias for the bootstrap samples:

$$E_n^* \left\{ \tilde{F}_{nh,c_0n^{-1/5}}(t) - \int IK_h(t - y) d\tilde{F}_{nh,c_0n^{-1/5}}(y) \right\}^2 + \left\{ \int IK_h(t - y) d\tilde{F}_{nh,c_0n^{-1/5}}(y) - \tilde{F}_{n,h_0}(t) \right\}^2,$$

where $E_n^*$ denotes the conditional expectation of the bootstrap sample, given $(T_1, \Delta_1), \ldots, (T_n, \Delta_n)$. Investigating the bias, we get:

$$\int IK_h(t - y) d\tilde{F}_{nh,c_0n^{-1/5}}(y) - \tilde{F}_{n,h_0}(t)$$

$$= \int \{IK_h(t - x) K_{h_0}(x - y) dx - IK_{h_0}(t - y) \} d\tilde{F}_n(y)$$
\[ \int \left\{ \int IK_h(t-x)K_{h_0}'(x-y)\,dx - K_{h_0}(t-y) \right\} \hat{F}_n(y)\,dy \]
\[ = \int \left\{ \int K_h(t-x)K_{h_0}(x-y)\,dx - K_{h_0}(t-y) \right\} \hat{F}_n(y)\,dy \]
\[ = \int K_h(t-x)\hat{F}_{n,h_0}(x)\,dx - \hat{F}_{n,h_0}(t), \]

where
\[ \hat{F}_{n,h_0}(x) = (K_{h_0} \ast \hat{F}_n)(x) = \int K_{h_0}(x-y)\hat{F}_n(y)\,dy. \]

and
\[ (2.1) \]
\[ K_h(u) = h^{-1}K(u/h). \]

Note:
\[ \int K_h(t-x)\hat{F}_{n,h_0}(x)\,dx - \hat{F}_{n,h_0}(t) = \int K_h(t-x) \left\{ \hat{F}_{n,h_0}(x) - \hat{F}_{n,h_0}(t) \right\} \,dx \]
\[ = \int K(u) \left\{ \hat{F}_{n,h_0}(t-hu) - \hat{F}_{n,h_0}(t) \right\} \,du \]
\[ = h\hat{F}_{n,h_0}'(t) \int uK(u)\,du + \frac{1}{2}h^2\hat{F}_{n,h_0}''(t) \int u^2K(u)\,du + o(h^2) \]
\[ = \frac{1}{2}h^2\hat{F}_{n,h_0}''(t) \int u^2K(u)\,du + o(h^2). \]

We have the following result.

**Lemma 1.** Let the conditions of Theorem 2 be satisfied for an interior point \( t \) of \([0, M]\). Moreover, let \( h_{n,0} \sim c_0n^{-1/9} \), as \( n \to \infty \). Then
\[ \hat{F}_{n,h_{n,0}}(t) \overset{p}{\to} f'_0(t), \quad n \to \infty. \]

The proof of Lemma is given in Section 12. The following corollary is immediate.

**Corollary 1.** Let the conditions of Theorem 2 be satisfied for an interior point \( t \) of \([0, M]\). Moreover, let \( h_n \sim cn^{-1/5} \) and \( h_{n,0} \sim c_0n^{-1/9} \), as \( n \to \infty \). Let \( \hat{F}_n \) be the nonparametric MLE. Then
\[ \int \left\{ IK_{h_n}(t-x)K_{h_{n,0}}(x-y)\,dx - IK_{h_{n,0}}(t-y) \right\} d\hat{F}_n(y) \sim \frac{1}{2}h_n^2f'_0(t) \int u^2K(u)\,du, \]

in probability.

So we lose in this way (asymptotically) the dependence on the pilot bandwidth \( h_{n,0} \). This suggests, as in [23], to take the pilot bandwidth \( h_{n,0} = c_0n^{-1/9} \) for some \( c_0 > 0 \) (taking the optimal order for a bandwidth for estimating the derivative \( f'_0 \)) instead of \( h_{n,0} = c_0n^{-1/5} \), and to minimize over \( h > 0 \)

\[ (2.2) \]
\[ B^{-1} \sum_{i=1}^{B} \left\{ \hat{F}^*_n(t) - \hat{F}_{n,h_{n,0}}(t) \right\}^2. \]
The minimizer \( \tilde{h}_n \) would, conditionally on sequences

\[
(T_i, \{X_i \leq T_i\}), \quad i = 1, 2, \ldots,
\]

asymptotically be given by

\[
\tilde{h}_n \sim \frac{g(t)^{-1} F_0(t) \{1 - F_0(t)\} \int K(u)^2 du}{\left\{ f_0(t)^2 \left( \int u^2 K(u) du \right)^2 \right\}^{1/5}} n^{-1/5}.
\]

A pilot bandwidth of the type \( h_{n,0} = c_0 n^{-1/5} \), as chosen in [28], will not give this result, since in that case the convergence of Lemma 1 does not hold. See also [29] for information on the order of the bandwidth needed in estimating the derivatives of densities, but note that, in view of the fact that we deal with inverse problems, the estimation of the distribution function is in our case comparable with the estimation of the density in ordinary density estimation, and so on; the estimation of the first derivative of the density is therefore comparable to the estimation of the second derivative in ordinary density estimation (which explains the bandwidth order \( n^{-1/9} \)).

Fig 1: The local bandwidth, selected by minimizing (2.2) (blue, solid), compared with the bandwidth, obtained by minimizing (2.3 (red, dashed), where the \( X_i \) are drawn from a truncated standard exponential distribution on \([0, 2]\) and the observations \( T_i \) from a Uniform distribution on \([0, 2]\). Sample size is \( n = 1000 \). The local bandwidths were computed on a grid of equidistant points 0.02, 0.04, . . .

An example is given in Figure 1, where the choice of the bandwidth based on (2.2), using 1000 bootstrap samples, is compared to the choice, based on minimizing

\[
N^{-1} \sum_{i=1}^{N} \left\{ \tilde{F}_{n,h}(t) - F_0(t) \right\}^2,
\]

computed for \( N = 1000 \) simulated samples of size \( n = 1000 \).
It is clear that we can compute in a similar way a globally asymptotically optimal bandwidth by a smoothed bootstrap experiment, in which case one replaces (2.2) by

\begin{equation}
B^{-1} \sum_{i=1}^{m} \left\{ \hat{F}_{n,h}(t_i) - \tilde{F}_{n,h}(t_i) \right\}^2,
\end{equation}

over a grid of points \( t_1, t_2, \ldots, t_m \). As noted in [15], if we select the bandwidth locally, we do not necessarily get a distribution function, the monotonicity can fail. For this reason we prefer a globally selected bandwidth, where we do not have this difficulty.

3. Confidence intervals for the current status model. For the current status model many different methods for constructing confidence intervals have been discussed in the literature. It has been proved that the ordinary bootstrap, resampling with replacement from \((T_1, \Delta_1), \ldots, (T_n, \Delta_n)\) and computing the MLE from the bootstrap samples \((T_1^*, \Delta_1^*), \ldots, (T_n^*, \Delta_n^*)\) is inconsistent. This was proved in [28]. Related results for the Grenander estimator were proved in [25].

It is suggested in [28] to base confidence intervals on the MLE \( \hat{F}_n \) by performing a smoothed bootstrap from the SMLE \( \bar{F}_{nh} \), and by basing the confidence intervals around the MLE \( \hat{F}_n \) on the fluctuation of \( \hat{F}_n(t) - \bar{F}_{nh}(t) \) in the bootstrap samples. But since we need the SMLE in this case, one could also consider bootstrap confidence intervals around the SMLE \( \tilde{F}_{nh} \). This is done in [13] and [15].

In [13] a classical bootstrap is considered, where we base the confidence intervals on the fluctuations around the SMLE \( \bar{F}_{nh}(t) \) of \( \hat{F}_{nh}(t) - \bar{F}_{nh}(t) \) in the bootstrap samples and resample with replacement from the sample \((T_1, \Delta_1), \ldots, (T_n, \Delta_n)\). In [15] a smoothed bootstrap is considered, where one bases the confidence intervals on the fluctuations of \( \hat{F}_{nh}(t) - \int IK_h(t-y) d\bar{F}_{nh}(t) \) around the SMLE \( \bar{F}_{nh}(t) \) in the (smoothed) bootstrap samples. In the latter case one keeps the observation times \( T_i \) fixed in the bootstrap samples and only varies the indicators \( \Delta_i \) in the bootstrap samples, which take values 1 or 0 with probabilities given by the SMLE \( \bar{F}_{nh} \).

Usually the performance of the bootstrap confidence intervals is best if one uses a pivot, obtained by “Studentization”, by dividing by the square root of an estimate of the variance, as is done in [15] and also in the \( \mathbb{R} \) package [14]. In each bootstrap sample we estimate the variance \( \sigma^2 \) defined by (1.6), apart from the factor \( cg(t) \), which drops out in the Studentized bootstrap procedure, by:

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

where \( \hat{F}_n^*(T_i) \) is the (non-smoothed) MLE in the bootstrap sample, and where \( \Delta_i^* \) is equal to 1 or 0 with probability \( \hat{F}_{nh,0}(T_i) \) and \( 1 - \hat{F}_{nh,0}(T_i) \), respectively. The variance estimate defined in (3.1) is inspired by the fact that the SMLE \( \bar{F}_{nh} \) is asymptotically equivalent to the toy estimator (“toy” because it is not a statistic we can actually compute on the basis of a sample alone),

\[ \hat{F}_{nh}^{toy}(t) = \int IK_h(t-x) dF_{0,x} + \frac{1}{n} \sum_{i=1}^{n} K_h(t-T_i) \{ \Delta_i - F_{0}(T_i) \} \],

which has sample variance

\[ S_{nh}(t) = \frac{1}{n^2} \sum_{i=1}^{n} K_h(t-T_i) \left( \Delta_i - F_{0}(T_i) \right)^2. \]
For the (smoothed) bootstrap confidence intervals we compute

\[
W_{nh}^*(t) = \frac{\tilde{F}_{nh}(t) - \int IK_h(t-u) d\tilde{F}_{n,h_0}(u)}{\sqrt{S_{nh}(t)}}.
\]

Let \( Q_{\alpha}^*(t) \) be the \( \alpha \)th 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:

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

where \( S_{nh}(t) \) is the estimate in the original sample obtained by replacing \( \Delta_i^* - \hat{F}_n(T_i) \) in (3.1) 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 \( u \in [t-h, t+h] \).

Fig 2: (a): Confidence intervals obtained the method using (3.2) The red piecewise constant function is the nonparametric MLE, the black solid curve the SMLE, and the dashed blue curve is the real truncated exponential distribution function \( F_0 \). (b): confidence intervals, also for sample size \( n = 500 \) in the same model, using the method of [28], with a band-width \( 2n^{-1/5} \) for the SMLE from which the smoothed bootstrap samples are generated. The curves have the same meaning as in (a). (c): confidence intervals in the same model, using the method of [2], based on likelihood ratio tests. The red curve is the MLE and the dashed curve is \( F_0 \). Sample size is again \( n = 500 \) and the confidence intervals were evaluated on an equidistant grid of 100 points.

The confidence intervals obtained in this way, with \( h = 1.5 n^{-1/5} \) (chosen by a procedure based on (2.4)) and \( h_0 = 2n^{-1/9} \), are shown in Figure 2 for one such sample of size \( n = 500 \). We compare them with the intervals of the method in [28]. We also compare the results with the confidence intervals in [2], based on likelihood ratio tests and the asymptotic distribution of the likelihood ratio statistic, which is a modified Chernoff-type distribution. The jumps in the confidence intervals together with the jumps of the MLE in these methods are somewhat unpleasant in the case of a continuous distribution function, which is nevertheless an assumption of the methods of [2] and [28].

A comparison of the coverage percentages of the three methods is given in Figure 3. As can be seen, the methods based on the MLE itself from [2] and [28] do not perform very well near the boundary. R scripts for all these methods are given in [9].
4. A model for the incubation time of a disease. We consider the following model, used for estimating the distribution of the incubation time of a disease. In this model there is an infection time $U$, uniformly distributed on an interval $[0, E]$, where $E$ ("exposure time") has an absolutely continuous distribution function $F_E$ on an interval $[0, M_2]$, and where $U$ is uniform on $[0, E]$, conditionally on $E$. Moreover there is an incubation time $V$ with an absolutely continuous distribution $F_0$ on an interval $[0, M_1]$ and a time for getting symptomatic $S$, where $S = U + V$. We assume that $U$ and $V$ are independent, conditionally on $E$, and that $E$ and $V$ have continuous distributions. Our observations consist of the pairs $(E_i, S_i)$, $i = 1, \ldots, n$.

The model is for example considered in [27], [4], [1] and [11].

We define the (convolution) density $q_F$ by

$$q_F(e, s) = e^{-1} \{ F(s) - F(s - e) \} = e^{-1} \int_{u=(s-e)+}^{s} dF(u), \quad e > 0, \quad s \in [0, M_1 + M_2],$$

w.r.t. $\mu$, which is the product of the measure $dF_E$ of the exposure time $E$ and Lebesgue measure on $[0, M_1 + M_2]$, where $M_1 > 0$ is the upper bound for the incubation time and $M_2 > 0$ is the upper bound for the exposure time.

For estimating the distribution function $F_0$ of the incubation time, usually parametric distributions are used, like the Weibull, log-normal or gamma distribution. However, in [11] the nonparametric maximum likelihood estimator is used. The maximum likelihood estimator $\hat{F}_n$ maximizes the function

$$\ell(F) = n^{-1} \sum_{i=1}^{n} \log \left\{ F(S_i) - F(S_i - E_i) \right\} = \int \log \left\{ F(s) - F(s - e) \right\} dQ_n(e, s)$$

over all distribution functions $F$ on $\mathbb{R}$ which satisfy $F(x) = 0, \quad x \leq 0$, see [11]. Here $Q_n$ is the empirical distribution function of the pairs $(E_i, S_i), \quad i = 1, \ldots, n$.

One can compare the model with a model where the exposure time is degenerate at a fixed point. The simplest model of this type is the model for uniform deconvolution, where $E_i \equiv 1$, the distribution of the infection time is uniform on $[0, 1]$ and the distribution of the

Fig 3: The proportion of times the real values $F_0(t)$ $t = 0.02, 0.04, \ldots$ did not belong to the 95% confidence intervals in 1000 samples of size 500 in same model as above for (a) the method using (3.2), (b) the method of [28], where the bandwidth of the SMLE is $h = 2n^{-1/5}$, (c) the method of [2]. The proportions were evaluated on an equidistant grid of 100 points.
incubation time is an unknown distribution, also concentrated on $[0, 1]$. In this case the model is equivalent to the current status model, discussed in the preceding sections.

This equivalence is outlined in Exercise 2 of Section 2.3 of Part 2 of [19], p. 61, and also in Exercise 10.4 in [18]. The matter is further discussed in Section 10.1 of [18]. Introducing the indicators $\Delta_i$ by defining $\Delta_i = \{S_i \leq E_i\} = \{S_i \leq 1\}$, the log likelihood (divided by $n$) (4.2) can be written:

$$\ell(F) = n^{-1} \sum_{i=1}^{n} \{\Delta_i \log F(S_i) + (1 - \Delta_i) \log \{1 - F(S_i - 1)\}\}. \quad (4.3)$$

Following the path of this exercise, we have that the random variables $T_i$, defined by

$$T_i = \begin{cases} S_i, & \text{if } \Delta_i = 1, \\ S_i - 1, & \text{if } \Delta_i = 0, \end{cases} \quad (4.4)$$

are uniformly distributed on $[0, 1]$. This leads to the following result.

**Theorem 3.** Let the observations consist of the pairs $(E_i, S_i)$, where $E_i \equiv 1$ and $S_i = U_i + V_i$, where $U_i$ is uniform on $[0, 1]$ and the $V_i$ have a distribution function $F_0$, concentrated on $[0, 1]$. Moreover, let the random variables $T_i$ be defined by (4.4). Defining (conversely)

$$\Delta_{i1} = \{S_i = T_i\}, \quad \Delta_{i2} = \{S_i = T_i + 1\}$$

the model is equivalent to the current status model, where the $T_i$ are uniformly distributed on $[0, 1]$, and the log likelihood for the distribution function $F_0$ is given by

$$\sum_{i=1}^{n} \{\Delta_{i1} \log F(T_i) + \Delta_{i2} \log \{1 - F(T_i)\}\},$$

where $\Delta_{i1}$ is a Bernoulli random variable, with success probability $F_0(T_i)$, conditionally on $T_i$, and $\Delta_{i2} = 1 - \Delta_{i1}$.

**Proof.** It is easily checked that the random variables $T_i$ have a uniform distribution on $[0, 1]$. We have, for $t \in (0, 1)$:

$$\mathbb{P}\{T_i \in dt, \Delta_i = 1\} = F_0(t) \, dt,$$

and similarly

$$\mathbb{P}\{T_i \in dt, \Delta_i = 0\} = \{1 - F_0(t)\} \, dt$$

So the density of $(T_i, \Delta_i)$ w.r.t. Lebesgue measure times counting measure on $\{0, 1\}$ factors out into the uniform density of $T_i$ (which equals 1 on $(0, 1)$) and the probability that $\Delta_i$ equals 1 ($= F_0(t)$) or 0 ($= 1 - F_0(t)$).

However, even in this simple case where the exposure time $E$ has a degenerate distribution at the point 1, the situation becomes more complicated if the support of the distribution of $V$ (in our model the incubation time) is no longer contained in $[0, 1]$. This situation was discussed in [16] (see also Section 11.2.3.e of [31]), but we shall now go in a different direction and discuss the relation with interval censoring models. Suppose again that $E$ is degenerate at a point $e > 0$, and that the support of the incubation time distribution is $[0, M]$, where $M > e$. We now get:
Theorem 4. Let the observations consist of the pairs \((E_i, S_i)\), where \(E_i \equiv e \geq 0\) and let \(S_i = U_i + V_i\), where \(U_i\) is uniform on \([0, e]\) and the \(V_i\), independently of \(U_i\), have an absolutely continuous distribution function \(F_0\), concentrated on \([0, M]\), where \(M > e\). Moreover, let the random variables \(T_i\) be defined by
\[
T_i = S_i - je, \quad \text{if } S_i \in (je, (j + 1)e], \quad j = 0, 1, \ldots.
\]
Defining
\[
\Delta_{ij} = \{S_i = T_i + je\}, \quad j = 0, \ldots, m,
\]
where \(m\) is the smallest integer \(j\) such that \(F_0(T_i + je) = 1\), the model is equivalent to the interval censoring, case \(m\), model, where the log likelihood for the distribution function \(F_0\) is proportional to
\[
\sum_{i=1}^{n} \sum_{j=0}^{m} \Delta_{ij} \log \{F(T_i + je) - F(T_i + (j - 1)e)\},
\]
and where, conditionally on the \(T_i\), the indicators \(\Delta_{ij}\) have a multinomial \(M(1; p_{i1}, \ldots, p_{im})\) distribution, where
\[
p_{ij} = F_0(T_i + je) - F_0(T_i + (j - 1)e).
\]

Theorem 4 follows from the facts that we have:
\[
P\{S_i \in je + dt\} = e^{-1} \{F_0(t + je) - F_0(t + (j - 1)e)\} \, dt,
\]
and
\[
P\{T_i \in dt\} = P\{S_i - \lfloor S_i/E_i \rfloor E_i \in dt\}
= e^{-1} \sum_{j=0}^{m} \{F_0(t + je) - F_0(t + (j - 1)e)\} \, dt = e^{-1} \, dt, \quad t \in (0, e).
\]

For the general case, where we condition on values \(E_i\), which have a common absolutely continuous distribution function \(F_E\), we now get immediately:

Theorem 5. Let the observations consist of the pairs \((E_i, S_i)\), where \(E_i\) has an absolutely continuous distribution on an interval \([0, M_2]\). Let \(S_i = U_i + V_i\), where \(U_i\) and \(V_i\) are independent, conditionally on \(E_i\), where \(U_i\) is uniform on \([0, E_i]\) and \(V_i\) has an absolutely continuous distribution function \(F_0\), concentrated on an interval \([0, M_1]\). Moreover, let the random variables \(T_i\) be defined by
\[
T_i = S_i - jE_i, \quad \text{if } S_i \in (jE_i, (j + 1)E_i], \quad j = 0, 1, \ldots.
\]
Furthermore, let conversely the indicators \(\Delta_{ij}\) be defined by
\[
\Delta_{ij} = \{S_i = T_i + jE_i\}, \quad j = 0, 1, \ldots, m_i,
\]
where \(m_i\) is the smallest integer \(j\) such that \(F_0(T_i + jE_i) = 1\). Then, conditionally on the \(E_i\), the \(T_i = S_i - \lfloor S_i/E_i \rfloor E_i\) are uniformly distributed on \([0, E_i]\), and the model is equivalent to a mixed case interval censoring model, where the log likelihood for the distribution function \(F_0\) is proportional to
\[
\sum_{i=1}^{n} \sum_{j=0}^{m_i} \Delta_{ij} \log \{F(T_i + jE_i) - F(T_i + (j - 1)E_i)\},
\]
and where the \(\Delta_{ij}\) have a multinomial \(M(1; p_{i1}, \ldots, p_{im_i})\) distribution, conditionally on the \((T_i, E_i)\), where
\[
p_{ij} = F_0(T_i + jE_i) - F_0(T_i + (j - 1)E_i),
\]
5. Characterization of the nonparametric maximum likelihood estimator (MLE) for the incubation time distribution. Theorem 5 in Section 4 showed that the model for the incubation time distribution, with the exposure times $E_i$ and the times of becoming symptomatic $S_i$, is equivalent to the so-called mixed case interval censoring model, where the observation intervals for the $i$th variable are

$$(0, T_i), (T_i, T_i + E_i), (T_i + E_i, T_i + 2E_i), \ldots,$$

with $T_i = S_i - \lfloor S_i/E_i \rfloor E_i$, uniformly distributed on $[0, E_i)$, conditionally on $E_i$ and where the indicators $\Delta_{ij}$ have multinomial distributions, conditionally on $T_i$ and $E_i$.

Let $Q_n$ be the empirical distribution function of the pairs $(E_i, S_i)$. Then, for a distribution function $F$ on $\mathbb{R}$, which is zero on $(-\infty, 0]$, we define the process

$$W_{n,F}(t) = \int_{\min(S_i) \leq s \leq t} \frac{\{s : F(s) - F(s - e) > 0\}}{F(s) - F(s - e)} dQ_n(e,s) - \int_{\min(S_i) \leq s - e \leq t} \frac{\{s : F(s) - F(s - e) > 0\}}{F(s) - F(s - e)} dQ_n(e,s),$$

(5.1)

with $0/0 = 0$, where $Q_n$ is the empirical distribution of $(E_1, S_1), \ldots, (E_n, S_n)$.

Note that the MLE is only determined at the points $(S_i - E_i)_{+}$ and $S_i$. We therefore restrict the set of candidates for the MLE to discrete distribution functions which only have mass at these points. But because the likelihood for a distribution $F$ can be enlarged if $F$ has a point of mass at a point $S_i - E_i$ not coinciding with a point $S_j$, for $j \neq i$ by removing this mass to a point $S_j$, we can restrict the set of distribution functions to the set of discrete distribution functions, having mass at points $S_i$ only. We denote this set of distribution functions by $\mathcal{F}_n$.

The following lemma characterizes the MLE.

**Lemma 2.** Let $\mathcal{F}_n$ be the set of discrete distribution functions $F$, which only have mass at the points $S_i$. Then $\hat{F}_n \in \mathcal{F}_n$ maximizes

$$\ell(F) = \int \log \{F(s) - F(s - e)\} dQ_n(e,s)$$

(5.2)

over $F \in \mathcal{F}_n$ if and only if

(i) \[ \int_{u \in [t, \infty)} dW_{\hat{F}_n}(u) \leq 0, \quad t \geq 0, \]

(ii) \[ \int W_{n,F}(t-) \, d\hat{F}_n(t) = 0. \]

Moreover, $\hat{F}_n$ is uniquely determined at the points $S_i$ by (5.2) and (5.3).

The proof of Lemma 2 follows the steps of the proof of Proposition 1.3 of Part 2 of [19] and is therefore omitted. Note that condition (5.3) is condition (1.27) of Proposition 1.3 and obtained by integration by parts from that conditions.

A picture of the point process $\{(T_i, W_{\hat{F}_n}(T_i)), i = 1, 2, \ldots\}$ in a simulation of the incubation time distribution, for $T_i$ running through points $S_i$ and $S_i - E_i$ between the minimum of the $S_i$ and the maximum of the $S_i - E_i$, is given in Figure 4 for sample size $n = 100$.

The process $W_{n,\hat{F}_n}$ touches zero at points just to the left of points of mass of $\hat{F}_n$. This shift of one point is due to the fact that we need the left-continuous slope of the greatest convex minorant of the cusum diagram, but use right-continuous distribution functions.
Fig 4: The point process \( \{(T_i, W_n \hat{F}(T_i)), i = 1, 2, \ldots \} \) for points \( T_i \), running through points \( S_i \) and \( S_i - E_i \) between the minimum of the \( S_i \) and the maximum of the \( S_i - E_i \). Sample size \( n = 100 \). The data correspond to a truncated Weibull distribution for the incubation time distribution, used in simulations of the incubation time distribution. The points are connected by line segments.

As explained in [11], one can compute the MLE by the iterative convex minorant algorithm, where one computes iteratively the greatest convex minorant of the cusum diagram with points \((0, 0)\) and points

\[
\left( G_{n,F}(t), \int_{u \in [0,t]} F(u) \, dG_{n,F}(u) + W_{n,F}(t) \right),
\]

where the “weight process” \( G_{n,F} \) is defined by

\[
G_{n,F}(t) = \int_{\min(S_i) \leq s \leq \max(S_i - E_i)} \left\{ \frac{\{F(s) - F(s - e) > 0\}}{\{F(s) - F(s - e)\}^2} \right\} \, dQ_n(e,s) \\
+ \int_{\min(S_i) \leq s - e \leq \max(S_i - E_i)} \left\{ \frac{\{F(s) - F(s - e) > 0\}}{\{F(s) - F(s - e)\}^2} \right\} \, dQ_n(e,s),
\]

where \( F \) is the temporary estimate of the distribution function at an iteration. The MLE \( \hat{F}_n \) corresponds to a stationary point of this algorithm and is given by the left-continuous slope of the greatest convex minorant of the cusum diagram, see Figure 5. See [11] for further remarks on this algorithm.

We define the process \( V_n \) by

\[
V_n(t) = \int_{u \in [0,t]} \hat{F}_n(u) \, dG_n(u) + W_{n,\hat{F}_n}(t),
\]
Fig 5: The cusum diagram \( \{(G_n(T_i), V_n(T_i)), i = 1, 2, \ldots \} \), where \( T_i \) runs through the ordered points \((S_i - E_i)_i\) and \( S_i \) and \( V_n \) is defined by (5.6), together with its greatest convex minorant (red curve). Sample size \( n = 100 \).

where \( G_n = G_n,F_n \) is defined by (5.5) for \( F = F_n \). Thus \( F_n \) is obtained by taking the left-continuous slope of the “self-induced” cusum diagram, defined by \((0, 0)\) and points

\[
(G_n,F(t), V_n(t)), \quad t \geq 0.
\]

(5.7)

6. Consistency of the MLE. We have the following result.

THEOREM 6. Let \( F_0 \) have a strictly positive density \( f_0 \) on \((0, M_1)\), for some \( M_1 > 0 \). Furthermore, let \( F_E \) be zero on an interval \([0, \varepsilon]\), where \( 0 < \varepsilon < M_1 \) and have a strictly positive continuous density \( f_E \) on the interval \((\varepsilon, M_2)\). Let \( F_n \in \mathcal{F}_n \) be the MLE, where \( \mathcal{F}_n \) is the set of distribution functions with mass at the set of points \( S_i \). Then the MLE \( F_n \) converges almost surely to \( F_0 \) on \([0, M_1]\).

There are a lot of different ways to prove consistency, but we feel a preference for the elegant method in [24], which is used in the proof below.

PROOF. We start by observing that, by the fact that \( F_n \) is the MLE, we must have:

\[
\sum_{0}^{1} \int \frac{F_0(s) - F_0(s - e)}{F_n(s) - F_n(s - e)} dQ_n(e, s) \leq 1.
\]

The empirical probability measure \( Q_n \) converges weakly to the underlying measure \( Q_0 \) on a set of elements \( \omega \) which has probability one. Fixing an \( \omega \) we get by the Helly compactness theorem a subsequence \((F_{n_k})(\cdot; \omega)\) converging vaguely to a subdistribution function \( F_{\omega} \), for which we get the inequality:

\[
\sum_{0}^{1} \int e^{-1} \left\{ \int \frac{(F_0(s) - F_0(s - e))^2}{F(s) - F(s - e)} ds \right\} dF_{\omega}(e) \leq 1.
\]

(6.1)

The minimum of

\[
\int e^{-1} \left\{ \int \frac{(F_0(s) - F_0(s - e))^2}{G(s) - G(s - e)} ds \right\} dF_{\omega}(e)
\]

(6.2)
over subdistribution functions $G$ is attained by a nondegenerate distribution function $G$, since otherwise (6.2) could be made smaller by multiplying $G$ by a constant bigger than 1. This means that we may assume that the minimizer $G$ of (6.2) satisfies

\begin{equation}
(6.3) \quad \int e^{-1} \left\{ \int \left\{ G(s) - G(s - e) \right\} ds \right\} dF_E(e) = 1,
\end{equation}

Minimizing (6.2) under the condition (6.3) is the same as minimizing

\begin{equation}
\int e^{-1} \int \left\{ \frac{(F_0(s) - F_0(s - e))^2}{G(s) - G(s - e)} + G(s) - G(s - e) \right\} ds dF_E(e),
\end{equation}

without this condition, using a Lagrange multiplier argument (with Lagrange multiplier $\lambda = 1$).

For $s \in (0, \varepsilon]$ we have $F_0(s - e) = G(s - e) = 0$ for $e \geq \varepsilon$ and the minimum of

\[
\frac{F_0(s)^2}{x} + x, \quad x > 0,
\]

is attained by taking $x = F_0(s)$. If $F_0(s) > F_0(s - e) > 0$, we find that

\[
\frac{(F_0(s) - F_0(s - e))^2}{y - x} + y - x, \quad x \geq 0, y - x > 0,
\]

is minimized by taking $y - x = F_0(s) - F_0(x)$, but since the minimizing values on the interval $u \in (0, \varepsilon]$ are equal $F_0(u)$, we must have $G(s) = F_0(s)$ and $G(s - e) = F_0(x - e)$ for the minimizing function $G$.

So the minimum of (6.2) is equal to 1 and attained for $G = F_0$. This means that the limit $F$ the subsequence $(\hat{F}_{n_k})$ must be equal to $F_0$, since otherwise the left-hand side of (6.1) would be strictly bigger than 1. Since this holds for all subsequences $(\hat{F}_{n_k})$, the result now follows.

7. Asymptotic distribution of the MLE in the model for the incubation time. We have the following result for the non-smoothed MLE in the model for the incubation time, corresponding to Theorem 1 in the current status model.

**THEOREM 7.** Let $F_0$ have a continuous density $f_0$, staying away from zero on its support $[0, M_1]$, $M_1 > 0$, and let the exposure time $E$ have a continuous density $f_E \geq c > 0$ on its support $[\varepsilon, M_2]$, for some $c > 0$, with a bounded derivative on the interval $(\varepsilon, M_2)$. Let $\hat{F}_n \in F_n$ be the MLE, where the set of distribution functions $F_n$ has the same meaning as in Theorem 6. Then we have at a point $t_0 \in (0, M_1)$:

\begin{equation}
(7.1) \quad n^{1/3} \left\{ \hat{F}_n(t_0) - F_0(t_0) \right\} / (4f_0(t_0)/c_E)^{1/3} \xrightarrow{d} \arg\min \left\{ W(t) + t^2 \right\},
\end{equation}

where $W$ is two-sided Brownian motion on $\mathbb{R}$, originating from zero and where the constant $c_E$ is given by:

\begin{equation}
(7.2) \quad c_E = \int e^{-1} \left[ \frac{1}{F_0(t_0) - F_0(t_0 - e)} + \frac{1}{F_0(t_0 + e) - F_0(t_0)} \right] dF_E(e),
\end{equation}

The result shows that the limit distribution is again given by Chernoff’s distribution. The jump in difficulty of the proof in going from the corresponding result for the current status model to more general cases of interval censoring models (to which the model for the incubation time also can be reduced, as we showed in the preceding sections) is considerable.

The proof of Theorem 7 is given in the supplementary material, section B.
8. Confidence intervals in the incubation time model. We now construct confidence intervals for the distribution function in the same way as we did for the current status model, based on the SMLE. But for justifying these methods, we need a result, corresponding to Theorem 2 for the current status model.

The SMLE $\tilde{F}_{nh}$ is again defined by:

\[ \tilde{F}_{nh}(t) = \int IK((t - y)/h) d\hat{F}_n(y), \quad (8.1) \]

where $\hat{F}_n$ is the MLE and $IK$ is defined by (1.4).

We have the following result, analogous to Theorem 2 above for the current status model.

**Theorem 8.** Let the conditions of Theorem 7 be satisfied and let $h_n$ be a bandwidth such that $h_n \sim kn^{-1/5}$, as $n \to \infty$, for some $k > 0$. Moreover, let the density $f_0$ be differentiable at $t \in (0, M_1)$, where $f'_0(t) \neq 0$ and let $K_h$ be defined as in (2.1), for the symmetric kernel $K$ which is the derivative of $IK$. Finally, let the SMLE $\tilde{F}_{nh}$ be defined by (8.1). Then

\[ n^{2/5} \{ \tilde{F}_{n,h_n}(t) - F_0(t) \} \xrightarrow{d} N(\mu, \sigma^2), \quad (8.2) \]

where

\[ \mu = \frac{1}{2} k^2 f'_0(t) \int u^2 K(u) du, \quad (8.3) \]

and

\[ \sigma^2 = \lim_{n \to \infty} n^{-1/5} \int \phi_n(y) K_{h_n}(t - y) dy \]

where the function $\phi_n$ solves the equation

\[ \int_{e > 0} e^{-1} \left[ \frac{\phi(v + e) - \phi(v)}{F_0(v + e) - F_0(v)} - \frac{\phi(v) - \phi(v - e)}{F_0(v) - F_0(v - e)} \right] dF_E(e) = -K_{h_n}(t - v), \quad v \in (0, M_1). \quad (8.4) \]

The proof of this result is given in the supplementary material, Section C, but we here give the relation to the proof for the current status model. Let $Q_0$ be the probability measure of the $(T_i, \Delta_i)$ in the current status model. In [17] the key step is to write

\[ \int IK_h(t - u) d(\hat{F}_n - F_0)(u) = - \int \phi_n(u) \{ \delta - \hat{F}_n(u) \} dQ_0(u, \delta), \]

where

\[ \phi_n(u) = \frac{K_{h_n}(t - u)}{g(u)}, \]

and $g$ is the density of the observation times, see (4.2), p. 364, of [17]. We next show that

\[ - \int \phi_n(u) \{ \delta - \hat{F}_n(u) \} dQ_0(u, \delta) = \int \phi_n(u) \{ \delta - F_0(u) \} d(Q_n - Q_0)(u, \delta) + o_p \left( n^{-2/5} \right) \]

\[ = \int \phi_n(u) \{ \delta - F_0(u) \} d(Q_n - Q_0)(u, \delta) + o_p \left( n^{-2/5} \right), \]

to get the asymptotic normality result.
In Theorem 8 the function $\phi_n$ plays a similar role, but this time we do not have an explicit representation for it, we can only say that it is the solution of the integral equation (8.4). So we have to show that this solution exists and derive the result from its properties.

Assuming Theorem 8, we can construct confidence intervals, using the smoothed bootstrap and the interpretation as mixed interval censoring model. Analogously to the method of Section 3 in [28], we keep the $E_i$ and $T_i = S_i - \lfloor S_i/E_i \rfloor E_i$ fixed and sample the multinomial random vectors $(\Delta_i^*, \ldots, \Delta_i^*_{m_i})$ according to the probabilities given in Theorem 4, where we replace the distribution function $F_0$ by its estimate the SMLE $\tilde{F}_{nh}$. Note that the authors in [28] mention in their section 3.1 that the limiting distribution of the MLE is unknown, but that we give the limit distribution in our case in Theorem 7, under the conditions given there.

An example is given in Figure 6.

The construction of the 95% bootstrap confidence intervals figure 6 proceeded in the following way.

We took sample size $n = 500$. The (original) sample is $(E_1, S_1), \ldots, (E_n, S_n)$, where $E_i$ is uniform on $[0, 30]$ and $S_i = U_i + V_i$, where $U_i$ is uniform on $[0, E_i]$ and (independently, given $E_i$) $V_i$ has a Weibull distribution, truncated on $[0, 20]$, with distribution function

$$ x \mapsto F_{\alpha, \beta}(x) = \begin{cases} 0 & , x < 0, \\ \{1 - \exp(-\beta x^{-\alpha})\} / \{1 - \exp(-20 x^{-\alpha})\} & , x \in [0, 20], \\ 1 & , x > 20, \end{cases}$$

where $\alpha = 3.03514$ and $\beta = 0.002619$. These values were found in a study of 88 travelers from Wuhan in [11], if one tries to estimate the distribution of the incubation time by maximum likelihood, assuming that it has a Weibull distribution.

Defining $T_i = S_i - \lfloor S_i/E_i \rfloor E_i$, the multinomials $(\Delta_i^*, \ldots, \Delta_i^*_{m_i})$ were sampled according to the probabilities given in Theorem 4, where $F_0$ is replaced by $\tilde{F}_{nh}$. The bandwidth $h$ was
determined to be $h = 3.2$, using (2.4), over a grid of 100 equidistant points, using a pilot
bandwidth $h_0 = 5 \approx 10 n^{-1/9}$.

Our bootstrap sample consists of:

$$(E_1, S^*_1), \ldots, (E_n, S^*_n),$$

where

$$S^*_i = \sum_{j=0}^{m_i} \Delta^*_ij (T_i + j E_i),$$

and where $m_i$ is the smallest $j$ such that $\bar{F}_{nh}(T_i + j E_i) = 1$. We take here

$$\mathbb{P}^* \{ \Delta^*_ij = 1 \} = \bar{F}_{n,h_0}(T_i + j E_i) - \bar{F}_{n,h_0}(T_i + (j - 1) E_i), \quad j = 0, \ldots, m_i,$$

where $\mathbb{P}^*$ denotes the probabilities for the bootstrap sample, conditionally on $(T_1, E_1), \ldots, (T_n, E_n)$. Note that we keep the $E_i$ fixed, relieving us from the duty of estimating its distribu-
tion. For each bootstrap sample the MLE $\hat{F}^*_n$ and the SMLE $\bar{F}^*_nh$ were computed.

For Figure 6 (a) all 1000 values $\bar{F}_{nh}(t) - \int IK_h(t - y)d\bar{F}_{n,h_0}(y)$, and the per-
centiles $\bar{P}_{0.025}(t)$ and $\bar{P}_{0.975}(t)$ were determined. Note that we do not subtract
$\bar{F}_{nh}(t)$, but instead subtract the convolution of the kernel $IK_h$ with $d\bar{F}_{nh_0}$. This gives the bootstrap intervals:

$$\left[ \bar{F}_{nh}(t) - \bar{P}_{0.975}(t), \bar{F}_{nh}(t) - \bar{P}_{0.025}(t) \right].$$

For Figure 6 (b) all 1000 values $\bar{F}_{n}(t) - \bar{F}_{nh}(t)$, and the percentiles $P_{0.025}(t)$ and $P_{0.975}(t)$
were determined. This gives the bootstrap intervals:

$$\left[ \bar{F}_{n}(t) - P_{0.975}(t), \bar{F}_{n}(t) - P_{0.025}(t) \right].$$

9. Estimation of quantiles and comparison with parametric methods. We now illustrate the difference between the nonparametric approach and the approach using distributions
like the Weibul, log-normal, etc. for the incubation time distribution. This is shown for the problem of estimating the 95th percentile of the distribution. To this end we generated 1000
samples of size $n = 500$ and also size $n = 1000$, using the same Weibull distribution to gen-
orate the incubation time distribution as we used in Section 8 for constructing confidence intervals. This example is also given (for sample size $n = 500$) in [12].

In the Weibull approach to the problem, we maximize for $\alpha, \beta > 0$:

$$\sum_{i=1}^{n} \log \left\{ F_{\alpha,\beta}(S_i) - F_{\alpha,\beta}(S_i - E_i) \right\}, \quad (9.1)$$

where $F_{\alpha,\beta}$ is defined by (8.5). This gives a maximum likelihood estimate $F_{\hat{\alpha},\hat{\beta}}$ of the distribution function, where $(\hat{\alpha}, \hat{\beta})$ maximizes (9.1) over $(\alpha, \beta).$ The estimate of the 95th percentile is then defined by $F^{-1}_{\hat{\alpha},\hat{\beta}}(0.95)$, where $F^{-1}_{\hat{\alpha},\hat{\beta}}$ denotes the inverse function.

In the log-normal approach to the problem, we maximize for $\alpha \in \mathbb{R}$ and $\beta > 0$:

$$\sum_{i=1}^{n} \log \left\{ G_{\alpha,\beta}(S_i) - G_{\alpha,\beta}(S_i - E_i) \right\}, \quad (9.2)$$

where $G_{\alpha,\beta}$ is defined by

$$G_{\alpha,\beta}(x) = \Phi \left( (\log x - \alpha) / \beta \right), \quad (9.3)$$
for \( x > 0 \) (zero otherwise), where \( \beta > 0 \) and \( \Phi \) is the standard normal distribution function. The estimate of the percentile is then given by \( G^{-1}_{\hat{\alpha},\hat{\beta}}(0.95) \), where \((\hat{\alpha}, \hat{\beta})\) maximizes (9.2) over \((\alpha, \beta)\).

In the nonparametric maximum likelihood approach we simply maximize

\[
\sum_{i=1}^{n} \log \{ F(S_i) - F(S_i - E_i) \},
\]

over all distribution functions \( F \). This give the nonparametric MLE \( \hat{F}_n \), from which we compute the SMLE \( \hat{F}_{n,h} \) and the estimate of the 95th percentile \( \hat{F}_{n,h}^{-1}(0.95) \). The bandwidth \( h \) was chosen to be \( h_n = 6n^{-1/5} \) here. Note that, using the delta method, we find:

\[
n^2/5 \left\{ \hat{F}_{n,h}^{-1}(0.95) - F_{\alpha,\beta}^{-1}(0.95) \right\} = -n^{2/5} \left\{ \hat{F}_{n,h}(F_{\alpha,\beta}^{-1}(0.95)) - 0.95 \right\} \big/ f_{\alpha,\beta}(F_{0}^{-1}(0.95)) + o_p(1),
\]

where \( f_{\alpha,\beta} \) is the (truncated) Weibull density, corresponding to the distribution function \( F_{\alpha,\beta} \), defined by (8.5).

The results of this simulation for 1000 samples of size \( n = 500 \) are shown in the box plot Figure 7 and the corresponding picture for sample size \( n = 1000 \) in Figure 8.

Fig 7: Box plot of 95th percentile estimates for the nonparametric, Weibull and log-normal maximum likelihood estimators for 1000 samples of size \( n = 500 \). The incubation time data are generated from a Weibull distribution. The red line denotes the value of the true percentile.

The black line segments in the boxes are at the position of the median. Finally, the red line denotes the value of \( F_{\alpha,\beta}^{-1}(0.95) \approx 10.17716 \), where \((\alpha, \beta)\) are the parameters of the Weibull distribution. R scripts for all methods are given in the directory “simulations” of [10].

It can be seen that, since the incubation time data were generated from a Weibull distribution, the estimates of the quantiles assuming this distribution have indeed the smallest
variation. But the nonparametric estimates, not making the assumption that the distribution is of the Weibull type, are also pretty good, whereas the estimates, assuming a log-normal distribution are completely off (in fact, these estimate are inconsistent). The SMLE adapts to the underlying distribution and provides consistent estimates, using the consistency of the MLE itself, derived in Section 6 and the consistency of the SMLE, which can be deduced from this.

One sees that the uncertainty about what interest us, is not much larger when one only uses the nonparametric SMLE than when one assumes that the incubation time distribution is Weibull (which is the correct distribution in this simulation setting), but much larger when one assumes log-normal. While there is absolutely no scientific (medical) reason to “believe” Weibull, or to “believe” log-normal. They lead to completely different statistical inferences, hence could lead to completely different policy recommendations.

10. Other smooth functionals. The relation between the model for the incubation time distribution and the mixed case interval censoring model was discussed in Section 4. Smooth functionals for the simplified situation where the exposure time is degenerate at a point in situations where we do not have to deal with an extra bandwidth parameter were for example discussed in Section 10.1 of [18] and in Section 11.2.3.e of [31].

The first moment is the prototype of such a smooth functional, The asymptotic normality and $\sqrt{n}$ convergence of the estimate

$$\int x \, d\hat{F}_n(x)$$

where $\hat{F}_n$ is the nonparametric MLE was given for the current status model in [19], Theorem 5.5 of Part 2. The asymptotic variance is given by

$$\sigma^2 = \int \frac{F_0(t)(1 - F_0(t))}{g(t)} \, dt,$$
where \( g \) is the density of the observation times and \( F_0 \) the distribution function of the hidden estimate.

Similar results for more general cases of interval censoring are given in Chapter 10 of [18], but in those cases the expression for the asymptotic variance is coming from the solution of an integral equation and no longer explicit as in the case of the current status model. A similar situation holds for the model for the incubation time distribution.

We have the following asymptotic normality result for the estimate of the first moment, based on the nonparametric MLE \( \hat{F}_n \) for the incubation time model, if the support of the incubation time distribution is \([0, M_1]\):

\[
\sqrt{n} \left\{ \int x \, d\hat{F}_n(x) - \int x \, dF_0(x) \right\} \xrightarrow{D} N(0, \sigma^2),
\]

where \( N(0, \sigma^2) \) is a normal distribution with mean zero and variance

\[
\sigma^2 = -\int_0^{M_1} \phi_{F_0}(x) \, dx.
\]

and \( \phi_{F_0} \) is also the solution of the following equation in \( \phi \):

\[
\int_{e > 0} e^{-1} \left[ \frac{\phi(v + e) - \phi(v)}{F(v + e) - F(v)} - \frac{\phi(v) - \phi(v - e)}{F(v) - F(v - e)} \right] \, dF_E(e) = 1, \quad v \in [0, M_1].
\]

The distribution function \( F_E \) of the exposure time was again chosen to be the uniform distribution function on \([1, 30]\).

---

Fig 9: Box plot of estimation of the first moment of the incubation distribution for the nonparametric, Weibull and log-normal maximum likelihood estimators for 1000 samples of size \( n = 5000 \). The incubation time data are generated from a Weibull distribution. The red line denotes the value of the actual real first moment.

The derivation of this result is given in the supplementary material, Section D. The inconsistency of the estimate based on the log normal model is again clearly seen from the boxplot.
Figure 9. However, if we would have generated the incubation time distribution from a log normal distribution, the estimate based on the Weibull distribution would be inconsistent, so Figure 9 cannot be interpreted as showing the superiority of the Weibull distribution.

Examples of the behavior of the density estimate

$$\hat{f}_{nh}(t) = \int K_h(t - y) d\hat{F}_n(y),$$

which converges at rate $n^{2/7}$, where $h \sim cn^{-1/7}$, $c > 0$, are given in [11] and [12].

11. Conclusion. We discussed the convergence to Chernoff’s distribution for the non-parametric maximum likelihood estimator (MLE) in the current status model and the convergence to a normal distribution of the smoothed MLE (the SMLE), as proved in [17]. The latter result needs the choice of a bandwidth and we discussed the choice of a bandwidth using the smoothed bootstrap, using oversmoothing for generating the bootstrap samples in Section 2. The choice of a bandwidth, converging at a lower rate than the optimal bandwidth for generating the bootstrap samples, seems crucial in this procedure.

Next we discussed confidence intervals, and compared the confidence intervals based on the SMLE with confidence intervals, based on the MLE (but using the SMLE to generate smoothed bootstrap samples) in [28] and confidence intervals, proposed in [2], also based directly on the (restricted) MLE. Figures 2 and 3 give the typical behavior of the three methods. It has been proved that the ordinary bootstrap, just resampling the observations, is inconsistent for this model (see [28]).

We proved that the nonparametric MLE in an often used model for the incubation time distribution converges in distribution, after standardization, to Chernoff’s distribution. We also showed that the model is equivalent to the mixed case interval censoring model. The rate of convergence is cube root $n$, if $n$ is the sample size, under a separation condition for the exposure time. We also discussed (locally) differentiable functionals of the model, estimated by corresponding functionals of the nonparametric MLE, which converge after standardization to a normal distribution at faster rates, where the constants are given by the solution of an integral equation.

This provides an alternative for the parametric models that are usually applied in this context, estimating the incubation time distribution by, e.g., Weibull, gamma or log-normal distributions. If the parametric model is not right (there is in fact no scientific or medical reason to choose for Weibull, gamma, Erland, log-normal, etc., and one sees for this reason usually these distribution applied at the same time) the estimates are inconsistent if the chosen model does not hold, as we demonstrate in Sections 9 and 10.

As shown in Section 10, for parameters like the first moment, we do not have to choose a bandwidth parameter, while the behavior of the estimate based on the nonparametric MLE is competitive to the parametric estimates in this case, even if the model for the parametric estimate is right.

\(\mathbb{R}\) scripts for computing the estimates are given in [10].

12. Appendix.

PROOF OF LEMMA 1. We have:

$$\hat{F}_{n,h,0}''(t) - h_{n,0}^{-3} \int K_n'' \left( \frac{(t - y)}{h_{n,0}} \right) F_0(y) \, dy$$

$$= h_{n,0}^{-3} \int K_n'' \left( \frac{(t - y)}{h_{n,0}} \right) \left\{ \hat{F}_n(y) - F_0(y) \right\} \, dy,$$
and
\[
\int K'' \left( (t - y)/h_{n,0} \right) \left\{ \hat{F}_n(y) - F_0(y) \right\} dy \\
= \int \frac{K'' \left( (t - y)/h_{n,0} \right)}{g(y)} \left\{ \hat{F}_n(y) - F_0(y) \right\} dG(y) \\
= \int \frac{K'' \left( (t - y)/h_{n,0} \right)}{g(y)} \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta),
\]
where \( G \) is the distribution function of the \( T_i \) and \( P_0 \) the probability measure of the pairs \((T_i, \Delta_i)\). We now define the function \( \psi_{t,h} \) by
\[
\psi_{t,h}(y) = \frac{K'' \left( (t - y)/h \right)}{h^3 g(y)}.
\]
Analogously to the development in Section 11.3 of [18], p. 332, we introduce a piecewise constant version \( \tilde{\psi}_{t,h} \) of \( \psi_{t,h} \) such that
\[
\tilde{\psi}_{t,h}(y) = \begin{cases}
\psi_{t,h}(\tau_i), & \text{if } F_0(y) > \hat{F}_n(\tau_i), \ u \in [\tau_i, \tau_{i+1}), \\
\psi_{t,h}(s), & \text{if } F_0(y) = \hat{F}_n(s), \ \text{for some } s \in [\tau_i, \tau_{i+1}), \\
\psi_{t,h}(\tau_{i+1}), & \text{if } F_0(y) < \hat{F}_n(\tau_i), \ u \in [\tau_i, \tau_{i+1}),
\end{cases}
\]
where the \( \tau_i \) are successive points of jump of \( \hat{F}_n \). So we can write:
\[
\int K'' \left( (t - y)/h_{n,0} \right) \left\{ \hat{F}_n(y) - F_0(y) \right\} dy = \int \frac{K'' \left( (t - y)/h_{n,0} \right)}{g(y)} \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta)
\]
\[
= \int \psi_{n,h_{n,0}}(y) \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta)
\]
\[
= \int \{ \psi_{n,h_{n,0}}(y) - \tilde{\psi}_{n,h_{n,0}}(y) \} \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta) + \int \tilde{\psi}_{n,h_{n,0}}(y) \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta)
\]
\[
= \int \{ \psi_{n,h_{n,0}}(y) - \tilde{\psi}_{n,h_{n,0}}(y) \} \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta)
\]
\[
+ \int \tilde{\psi}_{n,h_{n,0}}(y) \left\{ \hat{F}_n(y) - \delta \right\} d(P_0 - P_n) (y, \delta),
\]
using the characterization of the MLE \( \hat{F}_n \) in the last step (see Section 11.3 of [18]). We now get:
\[
\int \{ \psi_{n,h_{n,0}}(y) - \tilde{\psi}_{n,h_{n,0}}(y) \} \left\{ \hat{F}_n(y) - \delta \right\} dP_0(y, \delta)
\]
\[
= O_p \left( h_{n,0}^{-4} n^{-2/3} \right) = O_p \left( n^{-2/3} \right),
\]
and
\[
\int \tilde{\psi}_{n,h_{n,0}}(y) \left\{ \hat{F}_n(y) - \delta \right\} d(P_0 - P_n) (y, \delta) = O_p \left( h_{n,0}^{-3/2} n^{-1/2} \right) = O_p \left( n^{-1/3} \right).
\]
So the conclusion is:
\[
\hat{F}_{n,h_{n,0}}(t) - h_{n,0}^{-3} \int K'' \left( (t - y)/h_{n,0} \right) F_0(y) dy = O_p \left( n^{-2/3} \right).
\]
But under the conditions of Theorem 2 we have:

\[ h_{n,0}^{-3} \int K''((t-y)/h_{n,0}) F_0(y) \, dy = f_0'(t) + o(1), \quad n \to \infty. \]

SUPPLEMENTARY MATERIAL

Section A
Score operators and adjoint equation.

Section B
Proof of Theorem 7.

Section C
Proof of Theorem 8.

Section D
Proof of (10.1)

REFERENCES

[1] Backer, J. A., Klindexberg, D. and Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-nCov) infections among travellers from Wuhan, China, 20-28 January 2020. Euro Surveill. 25
[2] Banerjee, M. and Wellner, J. A. (2005). Confidence intervals for current status data. Scand. J. Statist. 32 405–424. MR2204627
[3] Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). Efficient and adaptive estimation for semiparametric models. Springer-Verlag, New York. Reprint of the 1993 original. MR1623559 (99c:62076)
[4] Britton, T. and Scalia Tomba, G. (2019). Estimation in emerging epidemics: bases and remedies. J. R. Soc. Interface 16.
[5] Chernoff, H. (1964). Estimation of the mode. Ann. Inst. Statist. Math. 16 31–41. MR0172382 (30 #2601)
[6] Groeneboom, P. (1987). Asymptotics for incomplete censored observations. Report 87-18, Mathematical Institute, University of Amsterdam.
[7] Groeneboom, P. (1991). Nonparametric maximum likelihood estimators for interval censoring and deconvolution. Technical Report 378, Department of Statistics, Stanford University.
[8] Groeneboom, P. (1996). Lectures on inverse problems. In Lectures on probability theory and statistics (Saint-Flour, 1994). Lecture Notes in Math. 1648 67–164. Springer, Berlin. MR1600884 (99c:62029)
[9] Groeneboom, P. (2015). Repp scripts. https://github.com/pietg/book/tree/master/Rcpp_scripts.
[10] Groeneboom, P. (2020). Incubation Time. https://github.com/pietg/incubationtime.
[11] Groeneboom, P. (2021a). Estimation of the incubation time distribution for COVID-19. Stat. Neerl. 75 161–179. MR4245907
[12] Groeneboom, P. (2021b). Chernoff’s distribution and the bootstrap. Nieuw Archief voor Wiskunde 22 241-244.
[13] Groeneboom, P. and Hendrickx, K. (2017a). The nonparametric bootstrap for the current status model. Electron. J. Stat. 11 3446-3484.
[14] Groeneboom, P. and Hendrickx, K. (2017b). curstatCI R package. Version 0.1.1.
[15] Groeneboom, P. and Hendrickx, K. (2018). Confidence Intervals for the Current Status Model. Scand. J. Stat. 45 135-163. doi: 10.1111/sjos.12294.
[16] Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. Statist. Neerlandica 57 136–157. MR2035863 (2004k:62121)
[17] 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.
[18] Groeneboom, P. and Jongbloed, G. (2014). Nonparametric Estimation under Shape Constraints. Cambridge Univ. Press, Cambridge.
[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] Groeneboom, P. and Wellner, J. A. (2001). Computing Chernoff’s distribution. J. Comput. Graph. Statist. 10 388–400. MR1939706
[21] Grüber, J. (1986). Nichtparametrische Analyse sporadisch beobachtbarer Krankheitsverlaufsdaten, Ph.D. dissertation, Technische Universität, Dortmund, Germany.
[22] 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)
[23] Hazelton, M. (1996). Bandwidth selection for local density estimators. Scand. J. Statist. 23 221–232. MR1394655
[24] Jewell, N. P. (1982). Mixtures of exponential distributions. The Annals of Statistics 479–484.
[25] 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.
[26] Peto, R. (1973). Experimental survival curves for interval-censored data. J.R. Statist. Soc. Series C 22 86–91.
[27] Reich, N. G., Lessler, J., Cummings, D. A. T. and Brookmeyer, R. (2009). Estimating incubation period distributions with coarse data. Stat. Med. 28 2769–2784. MR2750164
[28] Sen, B. and Xu, G. (2015). Model based bootstrap methods for interval censored data. Comput. Statist. Data Anal. 81 121–129. MR3257405
[29] Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. Ann. Statist. 6 177–184. MR0471166 (57 #10904)
[30] Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. J. Roy. Statist. Soc. Ser. B 38 290–295. MR652727
[31] Van de Geer, S. A. (2000). Applications of empirical process theory. Cambridge Series in Statistical and Probabilistic Mathematics 6. Cambridge University Press, Cambridge. MR1739079 (2001h:62002)
[32] van der Vaart, A. W. (1991). On differentiable functionals. Ann. Statist. 19 178–204. MR1091845 (92i:62100)
[33] Yu, Q., Schick, A., Li, L. and Wong, G. Y. C. (1998). Asymptotic properties of the GMLE in the case 1 interval-censorship model with discrete inspection times. Canad. J. Statist. 26 619–627. MR1671976