

Nonparametric confidence intervals for monotone functions

Piet Groeneboom and Geurt Jongbloed

Abstract: We study nonparametric isotonic confidence intervals for monotone functions. In [1] pointwise confidence intervals, based on likelihood ratio tests using the restricted and unrestricted MLE in the current status model, are introduced. We extend the method to the treatment of other models with monotone functions, and demonstrate our method by a new proof of the results in [1] and also by constructing confidence intervals for monotone densities, for which still theory had to be developed. For the latter model we prove that the limit distribution of the LR test under the null hypothesis is the same as in the current status model. We compare the confidence intervals, so obtained, with confidence intervals using the smoothed maximum likelihood estimator (SMLE), using bootstrap methods. The “Lagrange-modified” cusum diagrams, developed here, are an essential tool both for the computation of the restricted MLEs and for the development of the theory for the confidence intervals, based on the LR tests.

AMS 2000 subject classifications: Primary 62G05, 62N01; secondary 62G20.

Keywords and phrases: LR test, MLE, confidence intervals, isotonic estimate, smoothed MLE, bootstrap.

1. Introduction

In many situations one would like to estimate functions under the condition that they are monotone. Apart from giving algorithms for computing such estimates and from deriving their (usually asymptotic) distribution theory, it is also important to construct confidence intervals. These intervals can be uniform (in which case they are usually called confidence bands) as well as pointwise.

In this paper we consider two methods to obtain pointwise confidence intervals for distribution functions and monotone densities, based on nonparametric estimators. One approach, that of a (nonparametric) likelihood ratio (LR) test, based on the maximum likelihood estimator (MLE) in the model, is related to the one taken in [1] and [2]. The other approach, using a smoothed maximum likelihood estimator (SMLE) is based on an estimator introduced in [5] and further analyzed in [8]. Our methods can also be applied to monotone nonparametric least squares estimates of monotone regression functions. The examples and simulations we discuss suggest that the distinction between the uniform bounds and the pointwise bounds might be less dramatic than one would perhaps be inclined to assume.

There are some important differences between the approaches, based on the MLE and SMLE, respectively. How appropriate it is to use the MLE will largely depend on whether one expects (or allows) that the underlying monotone function will have jumps. Secondly, the bias of the MLE does not play a role in the construction of the confidence intervals based on the MLE. But if one constructs confidence intervals, using the SMLE with an optimal bandwidth, the bias will not be negligible in the limiting distribution. There is an extensive literature on how to deal with the bias in nonparametric function estimation, some approaches use undersmoothing, other approaches oversmoothing. A recent paper, discussing this literature and giving a solution for confidence bands, is [10]. We will use undersmoothing, as suggested in [9].

The method of constructing confidence intervals, based on the likelihood ratio test for the MLE, and the method using the SMLE are both asymptotically pivotal. For the method, based on the likelihood ratio test for the MLE, this arises from the universality properties of likelihood ratio tests. For the intervals, based on the SMLE, this is based on using bootstrap intervals for a “studentized” statistic, together with the undersmoothing. We now first describe two models that will be studied thoroughly in this paper.

Example 1.1. (*Monotone density functions*) The classical example of a monotone estimate of a monotone function is the so-called *Grenander estimator*. Let X_1, \dots, X_n be a sample of random variables, generated by a decreasing density f_0 on $[0, \infty)$. The MLE \hat{f}_n of f_0 is the Grenander estimator, which is by definition the left derivative of the least concave majorant of the empirical distribution function \mathbb{F}_n of X_1, \dots, X_n , as proved in [3] (see also Lemma 2.2 in [8]). This is also the first example in [1], where there is the (implicit) conjecture

that pointwise confidence intervals, based on the Grenander estimate, will have similar properties as the confidence intervals for the current status model (see the next example), based on a likelihood ratio test for the MLE. The difficulty in proving this result for the monotone density model resides in the constraint that the density integrates to 1, a condition which does not play a role in constructing LR tests for the current status model. We shall prove that the conjecture in [1] is correct and that one can use the same critical values as in the current status model in the construction of the asymptotic confidence intervals. We also compare the confidence intervals, obtained in this way, with confidence intervals, based on the SMLE, using bootstrap methods and asymptotic normality of the SMLE.

Example 1.2. (*The current status model*) Consider a sample X_1, X_2, \dots, X_n , drawn from a distribution with distribution function F_0 . Instead of observing the X_i 's (which can be thought of as an event time, such as 'getting infected'), one only observes for each i whether or not $X_i \leq T_i$ for some random (inspection time) T_i (independently of the other T_j 's and all X_j 's). More formally, instead of observing X_i 's, one observes

$$(T_i, \Delta_i) = (T_i, 1_{[X_i \leq T_i]}), \quad 1 \leq i \leq n. \quad (1.1)$$

One could say that the i -th observation represents the *current status* of item i at time T_i .

The problem is to estimate the unknown distribution function F based on the data given in (1.1). Denote the ordered realized T_i 's by $t_1 < t_2 < \dots < t_n$ and the associated realized values of the Δ_i 's by $\delta_1, \dots, \delta_n$. For this problem the log likelihood function in F (conditional on the T_i 's) is given by

$$\ell(F) = \sum_{i=1}^n \{\delta_i \log F(t_i) + (1 - \delta_i) \log(1 - F(t_i))\}. \quad (1.2)$$

The MLE maximizes ℓ over the class of *all* distribution functions. Since distribution functions are by definition nondecreasing, the problem belongs to the class of problems we want to study. As can be seen from the structure of (1.2), the value of ℓ only depends on the values that F takes at the observed time points t_i ; the values of F in between are not relevant as long as F is nondecreasing. Hence one can choose to consider only distribution functions that are constant between successive observed time points t_i . Lemma 2.1 below shows that this estimator can be characterized in terms of a greatest convex minorant of a certain diagram of points.

The main result of [1] is that confidence intervals, based on an LR test for the MLE, can be constructed, and that this is a pivotal way of constructing asymptotic confidence intervals, since the limit distribution does not depend on the parameters (under certain conditions). We will give a new proof, which is in line with our proof for the monotone density model.

There are numerous other models where our approach can be adopted. Examples include the model where one has a monotone hazard rate and right censored observations (see Sections 2.6 and 11.6 in [8]), the competing risk model with current status observations (see [6]) and monotone regression.

The methods based on the LR tests for the MLEs in the context of Example 1.1 and 1.2 follow the same line of argument, where, in both cases, an essential role is played by the penalization parameter $\hat{\mu}_n$, which is of order $O_p(n^{-2/3})$. Our methods rely on cumulative sum (cusum) diagrams which could be called 'Lagrange-modified' cusum diagrams, since they incorporate the Lagrange multipliers for the penalties. Asymptotic distribution theory is derived from the asymptotic properties of the Lagrange multipliers, used to construct these cusum diagrams. Once this has been done, the theory for the confidence intervals follows.

2. Confidence intervals for the current status model

The following lemma characterizes the unrestricted MLE in the current status model. This is Example 1.2 in Section 1, and we use the notation introduced there.

Lemma 2.1 (Lemma 2.7 in [8]). *Consider the cumulative sum diagram consisting of the points $P_0 = (0, 0)$ and*

$$P_i = \left(i, \sum_{j=1}^i \delta_j \right), \quad 1 \leq i \leq n, \quad (2.1)$$

recalling that the δ_i 's correspond to the t_i 's, which are sorted. Then the unrestricted MLE \hat{F}_n is given at the point t_i by the left derivative of the greatest convex minorant of this diagram of points, evaluated at the point i . This maximizer is unique among all subdistribution functions with mass concentrated on the inspection times t_1, \dots, t_n .

Remark 2.1. The left derivative of the convex minorant at P_i determines the value of \hat{F}_n at t_i and hence (by right continuity of the step function) on $[t_i, t_{i+1})$, a region to the right of t_i .

The characterization via Lemma 2.1 is well-known and a proof can e.g. be found in [14] and [8].

For the confidence intervals based on likelihood ratio tests for the MLE, we also have to compute the MLE under the restriction that its value is equal to a prescribed value a at a point t_0 . There are different ways to do this. It is suggested in [1] to compute the restricted MLE in two steps. The restricted MLE $\hat{F}_n^{(0)}$ is computed for values at points t to the left of t_0 under the restriction that $\hat{F}_n^{(0)}(t) \leq a$ and for values at points t to the right of t_0 under the restriction that $\hat{F}_n^{(0)}(t) \geq a$. To this end two cusum diagrams of type (2.1) are formed. Let m be such that $t_m \leq t_0 \leq t_{m+1}$. Then, a diagram of type (2.1) is formed, with n replaced by m , for the values to the left of t_0 . Next the minimum of a and the left derivative of the greatest convex minorant of this diagram of points is taken as the solution to the left of t_0 . For the points on the right side of t_0 the cusum diagram consisting of the points

$$P_0 = (0, 0) \text{ and } P_i = \left(i, \sum_{j=1}^i (1 - \delta_{n-j+1}) \right), \quad 1 \leq i \leq n - m, \quad (2.2)$$

is considered and the maximum of a and 1 minus the left derivatives of the greatest convex minorant of this diagram of points, with the obvious renumbering, is taken as the solution $\hat{F}_n^{(0)}(t_i)$ to the right of t_0 . Note that in this approach there is not necessarily a point t_i where $\hat{F}_n^{(0)}(t_i) = a$ is actually achieved; we only have inequalities. Of course, in view of the log likelihood, allowing an extra jump of the distribution function at t_0 , the value of $\hat{F}_n^{(0)}(t_0)$ can be taken equal to a if this is required.

In view of our general approach, where we also will prove the result for monotone densities, we will follow a different path, where we make the connection with the penalization methods, studied in, e.g., [4] and earlier in [19]. We have the following result.

Lemma 2.2. Let $0 < a < 1$ and $1 < i_0 < n$ be such that $\delta_i = 1$ for some $i \leq i_0$ and $\delta_i = 0$ for some $i > i_0$. Moreover, let $t_0 \in (t_{i_0}, t_{i_0+1})$. Denote by $(\hat{F}_1, \dots, \hat{F}(t_0), \dots, \hat{F}_n)$ the vector of values of a piecewise constant nondecreasing function \hat{F} at the observation points and at the point t_0 , where $\hat{F}_{i_0} \leq \hat{F}(t_0) \leq \hat{F}_{i_0+1}$. Then

- (i) If \hat{F}_i is given by the left-hand slope of the greatest convex minorant of the cusum diagram with points $(0, 0)$ and

$$\left(i, \sum_{j=1}^i \delta_j \right), \quad i = 1, \dots, n, \quad (2.3)$$

and if $\hat{F}_{i_0} \leq a \leq \hat{F}_{i_0+1}$, we put $\hat{F}(t_0) = a$, and $\hat{F}^{(0)} = (\hat{F}_1, \dots, \hat{F}(t_0), \dots, \hat{F}_n)$ is the maximizer of $\sum_{i=1}^n \{\delta_i \log F_i + (1 - \delta_i) \log(1 - F_i)\}$, under the side condition $F(t_0) = a$.

- (ii) If $(\hat{F}_1, \dots, \hat{F}_n)$ is defined as in (i), but $\hat{F}_{i_0} > a$ or $\hat{F}_{i_0+1} < a$, we define $\hat{\mu} \in \mathbb{R}$ to be the solution (in μ) of the equation

$$\max_{k \leq i_0} \min_{i \geq i_0} \frac{\sum_{j=k}^i \delta_j + n\mu a(1-a)}{i - k + 1} = a. \quad (2.4)$$

and define $\hat{F}_i^{(0)}$ by the left-hand slope of the greatest convex minorant of the cusum diagram with points $(0, 0)$ and

$$\left(i, \sum_{j=1}^i \{\delta_j + n\hat{\mu} a(1-a) 1_{\{j=i_0\}}\} \right), \quad i = 1, \dots, n. \quad (2.5)$$

We put $\hat{F}^{(0)}(t_0) = a$. Then $\hat{F}^{(0)} = (\hat{F}_1^{(0)}, \dots, \hat{F}^{(0)}(t_0), \dots, \hat{F}_n^{(0)})$ is the maximizer of $\sum_{i=1}^n \{\delta_i \log F_i + (1 - \delta_i) \log(1 - F_i)\}$, under the side condition $F(t_0) = a$.

Remark 2.2. The condition $\delta_i = 0$ for some $i > i_0$ is to avoid trivialities for the case that $\delta_i = 1$ for all $i \geq i_0$, in which case the only reasonable value of F_i is 1 for $i \geq i_0$. A similar remark holds for the condition that $\delta_i = 0$ for some $i \leq i_0$. If this were not the case, we would put F_i equal to 0 for $i \leq i_0$. For the asymptotic confidence intervals we concentrate on *interior* points of the support of the distribution F_0 .

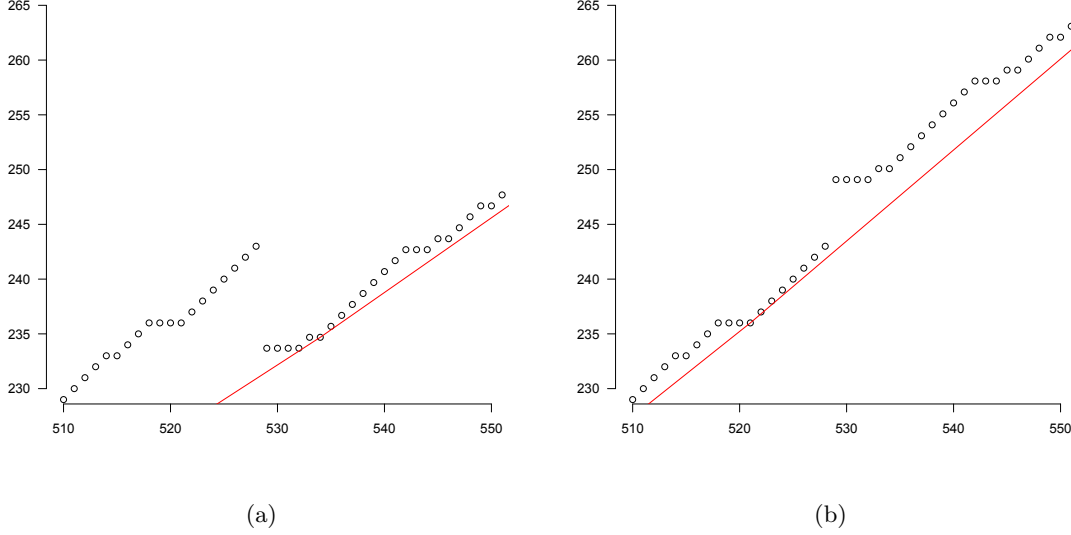


Fig 1: Pieces of two “Lagrange modified” cusum diagrams for the current status model, for sample size 1000 from the truncated exponential distribution function F_0 on $[0, 2]$; the observation distribution is uniform on $[0, 2]$, and $t_0 = 1$, $F_0(t_0) = 0.731058$. The unrestricted MLE \hat{F} has value 0.722892 at t_0 . In the example we have: $t_{529} < t_0 = 1 < t_{530}$.

Here (a) gives the local cusum diagram for $\hat{F}^{(0)}(t_0) = a = F_0(t_0) - 0.1 = 0.631058$, $\hat{\mu} = -0.039998$, and (b) gives the cusum diagram for $\hat{F}^{(0)}(t_0) = a = F_0(t_0) + 0.1 = 0.831058$, $\hat{\mu} = 0.0433549$.

Proof. (i) If \hat{F}_i is given by the left-hand slope of the greatest convex minorant of the cusum diagram (2.3), then $\hat{F} = (\hat{F}_1, \dots, \hat{F}_n)$ maximizes $\sum_{i=1}^n \{\delta_i \log F_i + (1 - \delta_i) \log(1 - F_i)\}$ without the side condition $\hat{F}(t_{i_0}) \leq a \leq \hat{F}(t_{i_0+1})$. Since the side condition is also satisfied under (i), \hat{F} is also the maximizer under this side condition in this case.

(ii) We can reduce the proof to the situation where $\delta_n = 0$. For if $\delta_j = 1$ for $j \geq i$, we put $F_j = 1$ for $j \geq i$. For similar reasons we can assume $\delta_1 = 1$. A similar reduction of the maximization problem was used in Proposition 1.3, p. 46 of [7]. An advantage of this reduction is that maximizing ℓ over all vectors $F = (F_1, \dots, F_n)$ with $0 \leq F_1 \leq \dots \leq F_n \leq 1$ is equivalent to maximizing ℓ over the cone $C = \{F = (F_1, \dots, F_n) : 0 \leq F_1 \leq \dots \leq F_n\}$.

Now first suppose $\hat{F}(t_{i_0}) > a$ for the unrestricted solution in (i). Then we have to make $\hat{F}(t_{i_0})$ smaller to allow $\hat{F}(t_0) = a$. We do this by changing the object function to be maximized over C into:

$$\phi_\mu(F_1, \dots, F_n) = \ell(F) + n\mu(F_{i_0} - a) = \sum_{i=1}^n \{\delta_i \log F_i + (1 - \delta_i) \log(1 - F_i)\} + n\mu(F_{i_0} - a), \quad (2.6)$$

where $\mu < 0$ is a suitable Lagrange multiplier.

The elements of the cone C can be uniquely expressed as positive linear combinations of its so-called generators

$$g_1 = (0, 0, \dots, 0, 0, 1), g_2 = (0, 0, \dots, 0, 1, 1), \dots, g_n = (1, 1, \dots, 1, 1, 1).$$

The necessary and sufficient Fenchel conditions for maximizing a concave function over a convex cone, (7.35) of [8], applied to these generators, lead to the following inequalities:

$$\langle \nabla \phi_\mu(F), g_j \rangle = \sum_{j=i}^n \left\{ \frac{\delta_j - F_j}{F_j \{1 - F_j\}} + n\mu 1_{\{j=i_0\}} \right\} \leq 0, \quad i = 1, \dots, n, \quad (2.7)$$

where $\nabla \phi_\mu(F)$ is the nabla vector $(\frac{\partial}{\partial F_1} \phi_\mu, \dots, \frac{\partial}{\partial F_n} \phi_\mu)$ at F and μ of the function (2.6). These inequalities can be rewritten as

$$\sum_{j=i}^n \left\{ \frac{\delta_j - F_j + n\mu 1_{\{j=i_0\}} a(1-a)}{F_j \{1 - F_j\}} \right\} \leq 0, \quad j = 1, \dots, n.$$

We also have the equality part of the Fenchel conditions,

$$\langle \nabla \phi_\mu(F), F \rangle = \sum_{j=1}^n \frac{\delta_j - F_j}{1 - F_j} + n\mu a = 0. \quad (2.8)$$

Multiplying this relation on blocks of constancy of F by $1 - F_j$ (see the proof of Lemma 2.3 in [8]), we find:

$$\sum_{j=1}^n (\delta_j - F_j) + n\mu a(1-a) = 0. \quad (2.9)$$

The Fenchel conditions (2.7) and (2.8) or (2.9) are necessary and sufficient conditions for the MLE, restricted to be equal to a at t_{i_0} .

It now follows that $\hat{F}^{(0)}$ is given by the left derivatives of the greatest convex minorant of the cusum diagram (2.5), where $\hat{\mu}$ is the solution of the equation (2.4). For the left derivative of the greatest convex minorant of the cusum diagram at i_0 is given by the left side of (2.4), by a well-known maxmin characterization, see, e.g., Theorem 1.4.4 in [14], and if (2.4) holds, we also have (2.9), since the greatest convex minorant will be equal to the second coordinate $\sum_{j=1}^n \{\delta_j + \hat{\mu} a(1-a) 1_{\{j=i_0\}}\}$ of the cusum diagram at n . Since $\hat{F}_{i_0}^{(0)} = a$, we can also let $\hat{F}^{(0)}(t_0) = a$.

If $\hat{F}(t_{i_0+1}) < a$, for \hat{F} as in (i), we also have $\hat{F}(t_{i_0}) < a$, and we reason in a similar way, this time for a Lagrange multiplier $\hat{\mu} > 0$. This will again give $\hat{F}_{i_0}^{(0)} = a$, and we can define $\hat{F}^{(0)}(t_0) = a$ again. \square

Remark 2.3. Cusum diagrams, incorporating the penalty, are shown in Figure 1. We have $\hat{\mu} > 0$ if $\hat{F}_{i_0+1} < a$ for the unrestricted solution of the maximization problem, and the cusum diagram for the restricted maximization problem is moved upward at i_0 . If $\hat{F}_{i_0} > a$, it is the other way around. The penalties give a local deviation of the restricted MLE $\hat{F}^{(0)}$ from the unrestricted MLE, but outside a local neighborhood of the point of restriction, $\hat{F}^{(0)}$ and \hat{F} will coincide again, where $\hat{F}^{(0)}$ picks up the same points of jump as \hat{F} .

Note, however, that we cannot say $\hat{F}^{(0)}(t) = a$ for the values t where $\hat{F}^{(0)}(t) \neq \hat{F}(t)$. A typical picture is shown in Figure 2, where, on the region where $\hat{F}^{(0)}$ and \hat{F} are different, the points of jump of $\hat{F}^{(0)}$ and \hat{F} are at different locations. There is also not a “contained in” relation in either direction for the sets of points of jump.

The proof of Theorem 2.1 below will use the following lemma, which is of a similar nature as results in [4]. To indicate the dependence on the sample size n , we now will denote the unrestricted and restricted MLE by \hat{F}_n and $\hat{F}_n^{(0)}$, respectively.

Lemma 2.3. *Under the conditions of Theorem 2.1 we have, if $a = F_0(t_0)$, as $n \rightarrow \infty$:*

$$\hat{\mu}_n = O_p(n^{-2/3}).$$

Proof. Consider the function

$$\phi(\mu) = \max_{k \leq i_0} \min_{i \geq i_0} \frac{\sum_{j=k}^i \delta_j + n\mu a(1-a)}{i - k + 1}, \quad a = F_0(t_0).$$

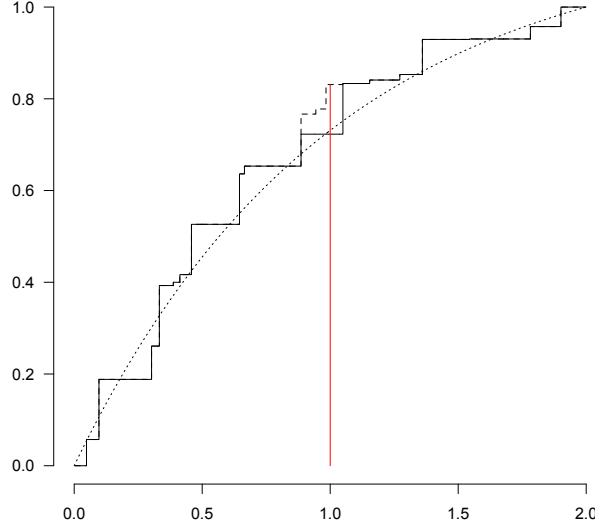


Fig 2: The unrestricted MLE and restricted MLE for the same data as in Figure 1, where F_0 (dotted) is the truncated exponential on $[0, 2]$ and the observation distribution is uniform on $[0, 2]$. Moreover, $\hat{F}^{(0)}(1) = F_0(1) + 0.1$. The deviation of the restricted MLE $\hat{F}^{(0)}$ from the unrestricted MLE \hat{F} is dashed. The jumps of the restricted and unrestricted MLE do not coincide on the interval of deviation.

By the conditions of Theorem 2.1 we may assume that there exist points t_{i_0} and t_{i_0+1} , as in Lemma 2.2, such that $t_0 \in (t_{i_0}, t_{i_0+1})$. By the maxmin characterization of the unrestricted MLE \hat{F}_n , we have

$$\phi(0) = \max_{k \leq i_0} \min_{i \geq i_0} \frac{\sum_{j=k}^i \delta_j}{i - k + 1} = \hat{F}_n(t_{i_0}).$$

Let $k_1 \leq i_0$ and $i_1 \geq i_0$ be the indices, satisfying

$$\hat{F}_n(t_{i_0}) = \frac{\sum_{j=k_1}^{i_1} \delta_j}{i_1 - k_1 + 1} = \max_{k \leq i_0} \min_{i \geq i_0} \frac{\sum_{j=k}^i \delta_j}{i - k + 1}.$$

This means that between the points P_{k_1-1} and P_{i_1} of the cusum diagram, with x -coordinates $k_1 - 1$ and i_1 , the least concave majorant of the cumulative sum diagram equals the line segment connecting P_{k_1-1} and P_{i_1} . Suppose $a > \hat{F}_n(t_{i_0})$ and let, for $\mu > 0$, $i_\mu \geq i_0$ be the index such that

$$\frac{\sum_{j=k_1}^{i_\mu} \delta_j + n\mu a(1-a)}{i_\mu - k_1 + 1} = \min_{i \geq i_0} \frac{\sum_{j=k_1}^i \delta_j + n\mu a(1-a)}{i - k_1 + 1}.$$

Then, since the function

$$\mu \mapsto \min_{i \geq i_0} \frac{\sum_{j=k_1}^i \delta_j + n\mu a(1-a)}{i - k_1 + 1}$$

is continuous and increasing in μ and tends to ∞ , as $\mu \rightarrow \infty$, there exists a $\mu > 0$ such that

$$\frac{\sum_{j=k_1}^{i_\mu} \delta_j + n\mu a(1-a)}{i_\mu - k_1 + 1} = \min_{i \geq i_0} \frac{\sum_{j=k_1}^i \delta_j + n\mu a(1-a)}{i - k_1 + 1} = a.$$

Using $a = F_0(t_0)$ and denoting the empirical measure of $\{(T_j, \Delta_j) : 1 \leq j \leq n\}$ by \mathbb{P}_n , this means that

$$\mu F_0(t_0)(1 - F_0(t_0)) = \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t_0) - \delta\} d\mathbb{P}_n(t, \delta). \quad (2.10)$$

where $\tau_- = t_{k_1}$ is the last jump point of \hat{F}_n before t_{i_0} . By a well-known fact on the jump points of the MLE in the current status model (see, e.g., Lemma 5.4 and its proof on p. 95 of [7]), we have that $t_0 - \tau_- = O_p(n^{-1/3})$. By the same type of argument, we can choose for each $\varepsilon > 0$ an $M > 0$ such that

$$\mathbb{P} \left\{ \int_{u \in [\tau_-, t]} \{F_0(t_0) - \delta\} d\mathbb{P}_n(u, \delta) < 0 \right\} > 1 - \varepsilon,$$

if $t > t_0 + Mn^{-1/3}$. Denote the distribution function of the observations T_i by G , with corresponding empirical distribution function \mathbb{G}_n . Then, since we must have

$$\begin{aligned} 0 &< \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t_0) - \delta\} d\mathbb{P}_n(t, \delta) \\ &= \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t_0) - F_0(t)\} d\mathbb{G}_n(t) + \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t) - \delta\} d\mathbb{P}_n(t, \delta) \\ &= \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t_0) - F_0(t)\} dG(t) + \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t_0) - F_0(t)\} d(\mathbb{G}_n - G)(t) \\ &\quad + \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t) - \delta\} d\mathbb{P}_n(t, \delta) \end{aligned} \quad (2.11)$$

by the positivity of μ , relation (2.10) and the conditions of Theorem 2.1, it now follows that $t_{i_\mu} - t_0 = O_p(n^{-1/3})$, and therefore

$$\mu F_0(t_0)(1 - F_0(t_0)) = \int_{t \in [\tau_-, t_{i_\mu}]} \{F_0(t_0) - \delta\} d\mathbb{P}_n(t, \delta) = O_p(n^{-2/3}),$$

since $t_0 - \tau_- = O_p(n^{-1/3})$, $t_{i_\mu} - t_0 = O_p(n^{-1/3})$ and therefore all three expressions on the right-hand side of (2.11) are $O_p(n^{-2/3})$.

Hence $\mu = O_p(n^{-2/3})$ and

$$\phi(\mu) = \max_{k \leq i_0} \min_{i \geq i_0} \frac{\sum_{j=k}^i \delta_j + n\mu a(1-a)}{i-k+1} \geq \min_{i \geq i_0} \frac{\sum_{j=k_1}^i \delta_j + n\mu a(1-a)}{i-k_1+1} = a.$$

By the monotonicity and continuity of the function ϕ we can now conclude

$$0 \leq \hat{\mu}_n \leq \mu = O_p(n^{-2/3}).$$

The case $a < \hat{F}_n(t_0)$ can be treated in a similar way. \square

Remark 2.4. The crux of the matter in proving a result like $t_{i_\mu} - t_0 = O_p(n^{-1/3})$ in the proof of Lemma 2.3 (see the discussion below (2.11)), is that, outside a neighborhood of order $n^{-1/3}$, the last two terms of the three terms on the right-hand side of (2.11) cannot cope with the negative parabolic drift of the first term. Arguments of this type are familiar by now, and were for example also used in the proofs of Lemma 3.5 in [8] and Lemma 5.4 on p. 95 of [7]. Arguments of this type can also be found in [12].

The preceding lemmas enable us to prove the following result, which corresponds to Theorem 2.5 in [1]. The proof is given in Section 5.

Theorem 2.1. *Let F_0 and G be distribution functions with continuous densities f_0 and g in a neighborhood of the point t_0 such that $0 < F_0(t_0) < 1$ and $f_0(t_0)$ and $g(t_0)$ are strictly positive. Let \hat{F}_n be the unrestricted*

MLE and let $\hat{F}_n^{(0)}$ be the MLE under the restriction that $\hat{F}_n^{(0)}(t_0) = F_0(t_0)$. Moreover, let the log likelihood ratio statistic $2 \log \ell_n$ be defined by

$$2 \log \ell_n = 2 \sum_{i=1}^n \left\{ \Delta_i \log \frac{\hat{F}_n(T_i)}{\hat{F}_n^{(0)}(T_i)} + (1 - \Delta_i) \log \frac{1 - \hat{F}_n(T_i)}{1 - \hat{F}_n^{(0)}(T_i)} \right\}.$$

Then

$$2 \log \ell_n \xrightarrow{\mathcal{D}} \mathbb{D},$$

where \mathbb{D} is the universal limit distribution as given in [1].

Construction of SMLE based confidence intervals for the distribution function

Let F_0 be defined on an interval $[a, b]$ with $a < b$ satisfying $F_0(a) = 0$ and $F_0(b) = 1$. Then we can estimate F_0 by the SMLE, using a boundary correction:

$$\tilde{F}_{nh}(t) = \int \left\{ \mathbb{K} \left(\frac{t-x}{h} \right) + \mathbb{K} \left(\frac{t+x-2a}{h} \right) - \mathbb{K} \left(\frac{2b-t-x}{h} \right) \right\} d\hat{F}_n(x), \quad (2.12)$$

where \hat{F}_n is the MLE, $\mathbb{K}(x) = \int_{-\infty}^x K(u) du$, and K is a symmetric kernel density, like the triweight kernel. If $t \in [a+h, b-h]$ the SMLE coincides with the familiar

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

the other two terms in (2.12) are only there for correction at the left and right boundary. For simplicity we take $a = 0$ in the following (the usual case), and the interval, containing the support of F_0 will now be denoted by $[0, b]$.

For the construction of the $1-\alpha$ confidence interval we take a number of bootstrap samples $(T_1^*, \Delta_1^*), \dots, (T_n^*, \Delta_n^*)$ with replacement from $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$. For each such sample we compute the SMLE \tilde{F}_{nh}^* , using the same bandwidth h as used for the SMLE \tilde{F}_{nh} in the original sample, and the same type of boundary correction. Next we compute at the points t :

$$\begin{aligned} Z_{n,h}^*(t) &= \frac{\tilde{F}_{nh}^*(t) - \tilde{F}_{nh}(t)}{\sqrt{n^{-2} \sum_{i=1}^n \{K_h(t - T_i^*) - K_h(t + T_i^*) - K_h(2b - t - T_i^*)\}^2 (\Delta_i - \hat{F}_n(T_i^*))^2}}, \end{aligned} \quad (2.13)$$

where \hat{F}_n^* is the ordinary MLE (not the SMLE!) of the bootstrap sample $(T_1^*, \Delta_1^*), \dots, (T_n^*, \Delta_n^*)$.

Let $U_{\alpha}^*(t)$ be the α -th percentile of the B bootstrap values $Z_{n,h}^*(t)$. Then, disregarding the bias for the moment, the following bootstrap $1 - \alpha$ interval is suggested:

$$\left[\tilde{F}_{nh}(t) - U_{1-\alpha/2}^*(t) S_{nh}(t), \tilde{F}_{nh}(t) + U_{\alpha/2}^*(t) S_{nh}(t) \right], \quad (2.14)$$

where

$$S_{nh}(t)^2 = n^{-2} \sum_{i=1}^n \{K_h(t - T_i) - K_h(t + T_i) - K_h(2b - t - T_i)\}^2 (\Delta_i - \hat{F}_n(T_i))^2.$$

The bootstrap confidence interval is inspired by the fact that the SMLE is asymptotically equivalent to the toy estimator

$$\begin{aligned} F_{nh}^{toy}(t) &= \int \{ \mathbb{K}_h(t-u) + \mathbb{K}_h(t+u) - \mathbb{K}_h(2b-t-u) \} dF_0(u) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \frac{\{K_h(t - T_i) - K_h(t + T_i) - K_h(2b - t - T_i)\} \{\Delta_i - F_0(T_i)\}}{g(T_i)}, \end{aligned}$$

the variance of which can be estimated by

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

and also by Theorem 4.2, p. 365 in [5], which states that, if $h \sim cn^{-1/5}$, under the conditions of that theorem, for each $t \in (0, b)$,

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

where

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

and

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

We now first study the behavior of intervals of type (2.14) for a situation where the asymptotic bias plays no role (the uniform distribution) and compare the behavior of the intervals with the confidence intervals, based on LR tests for the MLE.

Simulation for uniform distributions

We generated 1000 samples $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ by generating T_1, \dots, T_n , $n = 1000$, from the uniform distribution on $[0, 2]$ and, independently, a sample X_1, \dots, X_n , also from the uniform distribution on $[0, 2]$. If $X_i \leq T_i$ we get a value $\Delta_i = 1$, otherwise $\Delta_i = 0$. For each such sample $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ we generated 1000 bootstrap samples, and computed the 25th and 975th percentile of the values (2.13) at the points $t_j = 0.02, 0.04, \dots, 1.98$. On the basis of these percentiles we constructed the confidence intervals (2.14) for all of the (99) t_j 's and checked whether $F_0(t_j)$ belonged to it. The percentages of simulation runs that $F_0(t_j)$ did not belong to the interval are shown in Figure 3. We likewise computed the confidence interval, based on the LR test for the MLE for each t_j , and also counted the percentages of times that $F_0(t_j)$ did not belong to the interval. The corresponding confidence intervals for one sample are shown in Figure 4.

Simulation for truncated exponential distributions

To investigate the role of the asymptotic bias of the SMLE, we also generated 1000 samples $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ by generating T_1, \dots, T_n , $n = 1000$, from the uniform distribution on $[0, 2]$ and, independently, X_1, \dots, X_n , from the truncated exponential distribution on $[0, 2]$, with density

$$f_0(x) = \frac{e^{-x}}{1 - e^{-2}}, \quad x \in [0, 2].$$

If $X_i \leq T_i$ we get $\Delta_i = 1$, otherwise $\Delta_i = 0$. For each such sample $(T_1, \Delta_1), \dots, (T_n, \Delta_n)$ we generated $B = 1000$ bootstrap samples, and computed the confidence intervals in the same way as for the uniform samples, discussed above, where the interval is of the form (2.14) and bias is neglected. This is compared in Figure 5 with the results for confidence intervals of the form

$$\left[\tilde{F}_{nh}(t) - \beta(t) - U_{1-\alpha/2}^*(t) S_n(t), \tilde{F}_{nh}(t) - \beta(t) - U_{\alpha/2}^*(t) S_n(t) \right], \quad (2.15)$$

where $U_{\alpha/2}^*$, $U_{1-\alpha/2}^*$ and $S_n(t)$ are as in (2.14), and where $\beta(t)$ is the actual asymptotic bias, which is, for $t \in [h, 2 - h]$, given by

$$\frac{1}{2} f'_0(t) h^2 \int u^2 K(u) du = - \frac{h^2 e^{-t} \int u^2 K(u) du}{2\{1 - e^{-2}\}}.$$

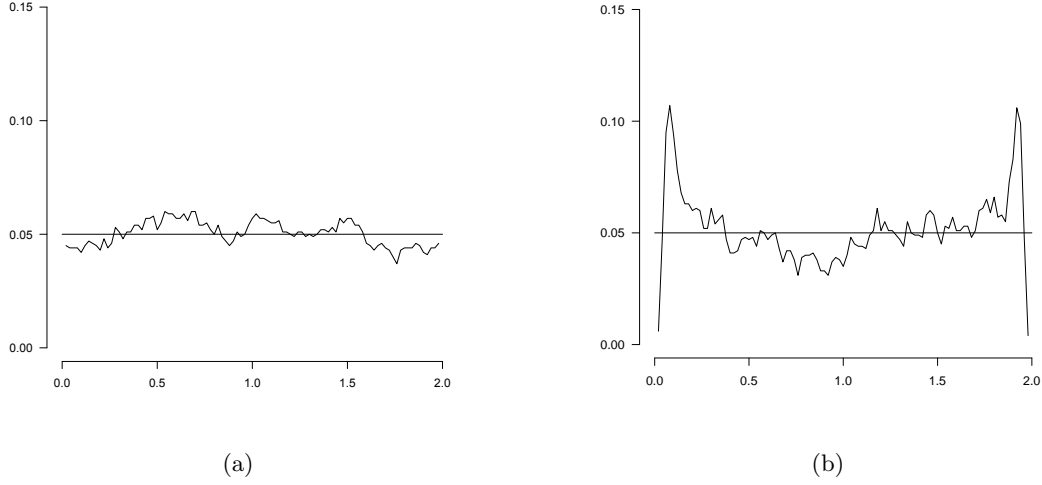


Fig 3: Uniform samples. Proportion of times that $F_0(t_i)$, $t_i = 0.02, 0.02, \dots, 1.98$ is not in the 95% CI's in 1000 samples $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$ using the SMLE and 1000 bootstrap samples from the sample $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$. In (a), the SMLE is used with CI's given in (2.14). In (b) CI's are based on the LR test. The observations are based on two independent samples of T_i 's and i 's, uniformly distributed on $[0, 2]$.

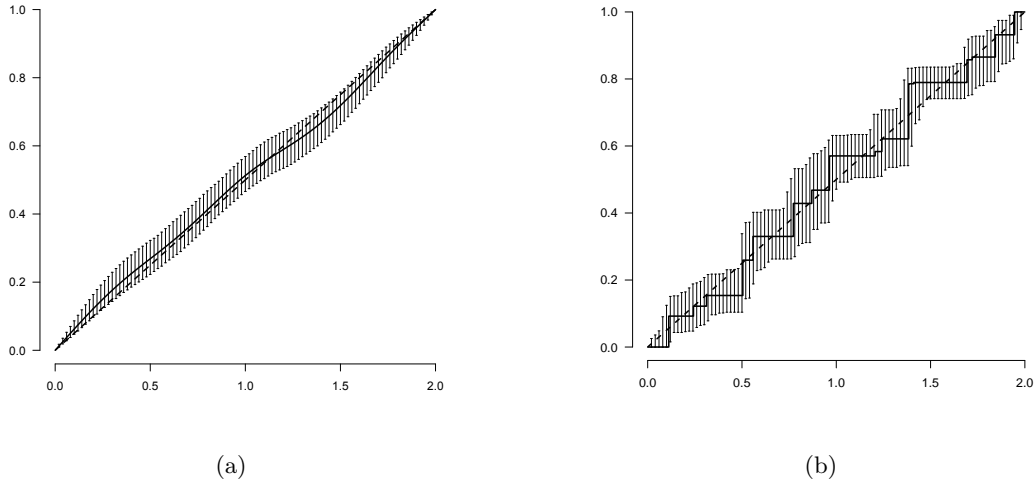


Fig 4: Uniform samples: 95% confidence intervals for $F_0(t_i)$, $t_i = 0.02, 0.02, \dots, 1.98$ for one sample $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$. For (a) the SMLE and 1000 bootstrap samples are used; F_0 is dashed and the SMLE solid. For (b) the LR test is used; F_0 is dashed and the MLE solid.

For $t \notin [h, 2 - h]$ this expression is of the form

$$-\frac{h^2 e^{-t} \left\{ \int u^2 K(u) du - 2 \int_v^1 (u-v)^2 K(u) du \right\}}{2\{1 - e^{-2}\}},$$

where $v = t/h$, if $t \in [0, h)$ and $v = (2 - t)/h$ if $t \in (2 - h, 2]$.

It is seen in Figure 5 that if we use the bandwidth $2n^{-1/5}$ and do not use bias correction for the SMLE, the 95% coverage is off at the left end (where the bias is largest), but that the intervals are ‘on target’ if we add the asymptotic bias to the intervals, as in (2.15). However, we cannot use the method of Figure 5b in practice, since the actual bias will usually not be available. We are faced here with a familiar problem in nonparametric confidence intervals, and we can take several approaches. Two possible solutions are *estimation of the bias* and *undersmoothing*.

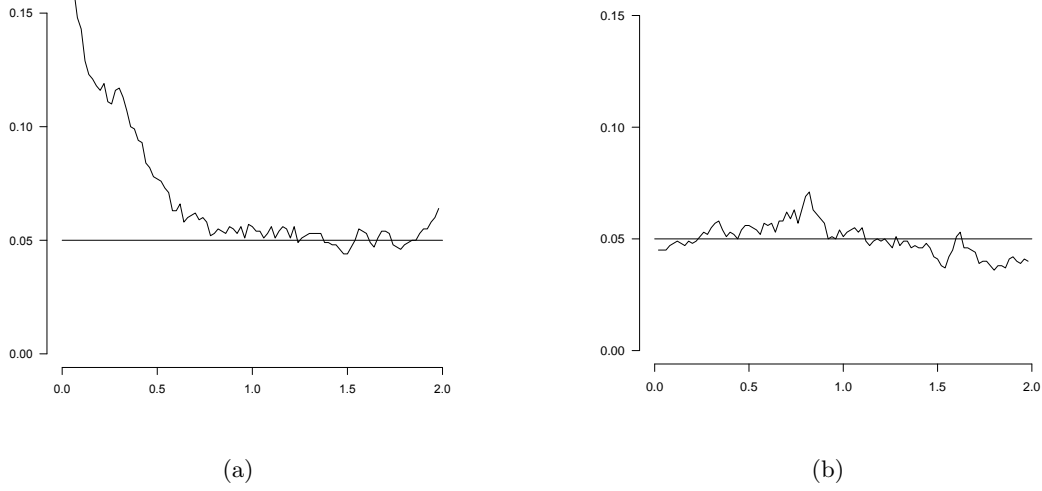


Fig 5: Coverage for the truncated exponential distribution function F_0 . Proportion of times that $F_0(t_i)$, $t_i = 0.01, 0.02, \dots$ is not in the 95% CI's in 1000 samples $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$. In (a) the confidence intervals (2.14) are used, in (b) the bias corrected confidence intervals (2.15). The bandwidth is $h = 2n^{-1/5}$

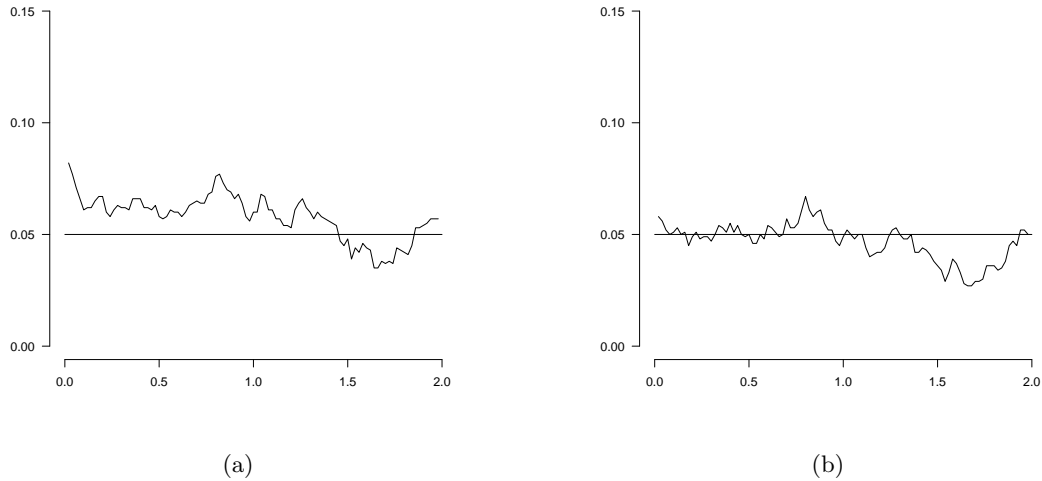


Fig 6: Coverage for the truncated exponential distribution function F_0 . Proportion of times that $F_0(t_i)$, $t_i = 0.01, 0.02, \dots$ is not in the CI's in 1000 samples $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$. In (a) the SMLE and (2.14) are used for $\alpha = 0.025$ with undersmoothing. In (b), (2.14) is used with $\alpha = 0.02$ instead of $\alpha = 0.025$ and the same undersmoothing as in (a). The bandwidth is $h = 2n^{-1/4}$

In the present case it turns out to be very difficult to estimate the bias term sufficiently accurately. Moreover, [9] argues that undersmoothing has several advantages; one of these is that estimation of the bias term is no longer necessary. For the present model, we changed the bandwidth of the SMLE from $2n^{-1/5}$ to $2n^{-1/4}$ (with $n = 1000$) and computed the confidence intervals again by the bootstrap procedure, given above. This gave a remarkable improvement of the coverage at the left end, as is shown in Figure 6. Nevertheless, the undersmoothing has the tendency to make the confidence interval slightly liberal (anti-conservative), as can be seen from Figure 6a, so one might prefer to take for example the 20th and 980th percentile if one wants to have a coverage $\geq 95\%$. The effect of this method is shown in Figure 6b and the coverage of this method is compared to the coverage of the method, using the LR test, as in [2], in Figure 7. Undersmoothing, together with the method of Figure 6b, will generally of course still produce narrower confidence intervals than the method, based on the LR test (which is based on cube root n asymptotics), under the appropriate smoothness conditions, as can be seen in Figure 8.

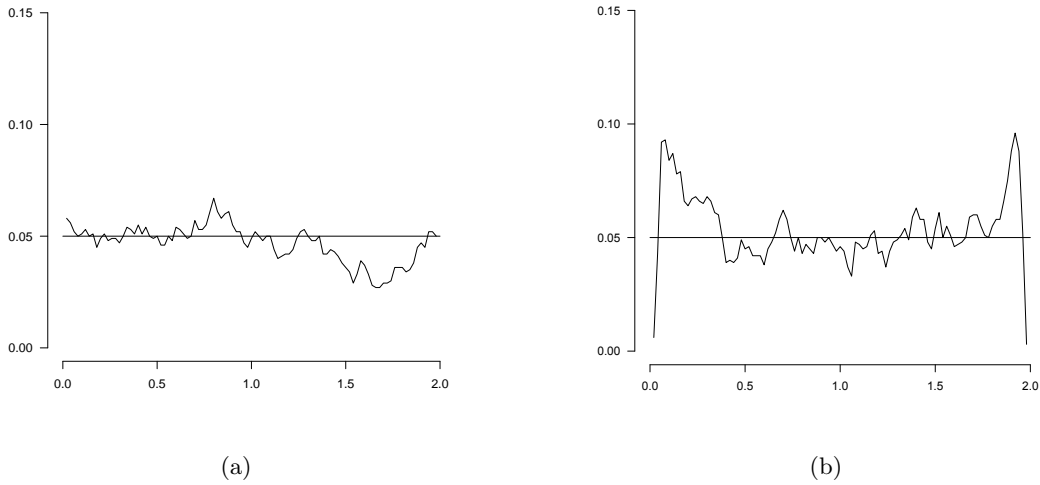


Fig 7: Truncated exponentials for F_0 . Proportion of times that $F_0(t_i)$, $t_i = 0.01, 0.02, \dots$ is not in the CI's in 1000 samples $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$. Figure (a) uses the SMLE with the method of Figure 6b. In (b) the LR test for the MLE is used.

Another way of bias correction is to use a higher order kernel in the definition of the SMLE, for example a 4-th order kernel, but still use a bandwidth of order $n^{-1/5}$. Since a 4-th order kernel has necessarily negative parts, and since the estimate of F_0 will be close to zero or 1 at the boundary of the interval, this gives difficulties at the end of the interval. We therefore stick to the method described above.

3. Confidence intervals for the monotone density case

In this section we construct confidence intervals for a decreasing density, in the setting of Example 1.1. We start by considering the confidence intervals based on the LR tests. To this end, we first give a characterization of the restricted MLE. In view of Example 3.1 below, in which the observations are on a discrete scale and therefore have ties in the observations, we denote the number of observations at the ordered points t_i by w_i . The number of strictly different observation times is denoted by m , and the total number of observations is again denoted by n , so $n = \sum_{j=1}^m w_j$.

We recall the definition of the unrestricted MLE in this case.

Lemma 3.1. *Let $\hat{f} = (\hat{f}_1, \dots, \hat{f}_m)$ be the vector of left-continuous slopes of the least concave majorant of*

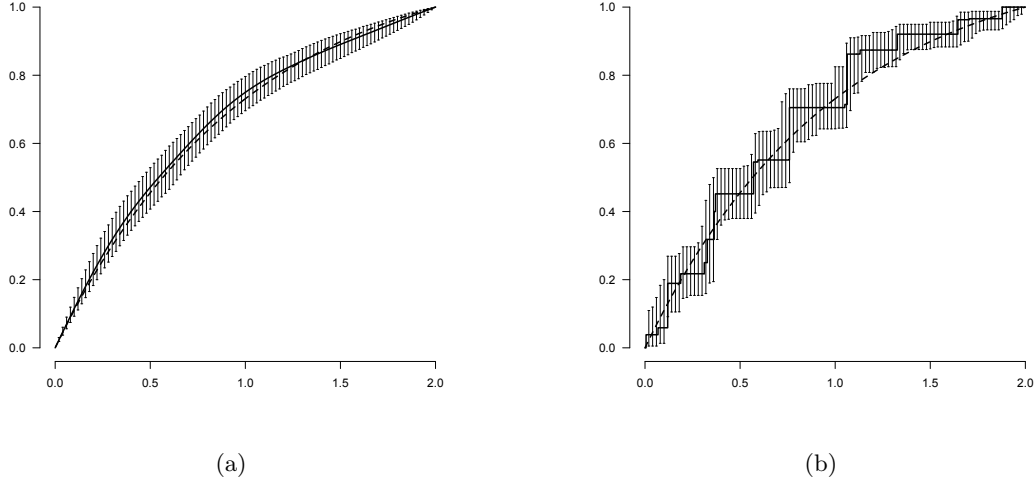


Fig 8: Truncated exponentials for F_0 : 95% confidence intervals for $F_0(t_i)$, $t_i = 0.01, 0.02, \dots$ for one sample $(T_1, \Delta_1) \dots, (T_n, \Delta_n)$. In (a) the SMLE is used with undersmoothing and the method of Figure 6b. Dashed: real F_0 , solid: SMLE. In (b) the LR test for the MLE is used. Dashed: real F_0 , solid: MLE.

the cusum diagram with points $(0, 0)$ and

$$\left(t_j, \frac{1}{n} \sum_{i=1}^j w_i \right), \quad j = 1, \dots, m, \quad (3.1)$$

Then \hat{f} maximizes $\sum_{i=1}^m w_i \log f_i$, under the condition that f is nonincreasing and the side condition $\sum_{i=1}^m f_i (t_i - t_{i-1}) = 1$.

For convenience, we provide the proof below.

Proof. Introducing the Lagrange multiplier λ , we get the maximization problem of maximizing

$$\phi_{\lambda, \mu}(f_1, \dots, f_m) = \frac{1}{n} \sum_{i=1}^m w_i \log f_i - \lambda \left\{ \sum_{i=1}^m f_i (t_i - t_{i-1}) - 1 \right\}, \quad (3.2)$$

over the convex cone $C_m = \{(f_1, \dots, f_m) : f_1 \geq f_2 \geq \dots \geq f_m \geq 0\}$, where we look for $\hat{\lambda} \in \mathbb{R}_+$ such that the maximizer $\hat{f} = (\hat{f}_1, \dots, \hat{f}_m)$ satisfies

$$\sum_{i=1}^m \hat{f}_i (t_i - t_{i-1}) = 1.$$

Using the equality part of the Fenchel conditions for this maximization problem, the solution has to satisfy

$$\langle \nabla \phi_{\hat{\lambda}}(\hat{f}), \hat{f} \rangle = \frac{1}{n} \sum_{i=1}^m w_i - \hat{\lambda} \sum_{i=1}^m (t_i - t_{i-1}) \hat{f}_i = 1 - \hat{\lambda} \sum_{i=1}^m (t_i - t_{i-1}) \hat{f}_i = 1 - \hat{\lambda} = 0. \quad (3.3)$$

So $\hat{\lambda} = 1$.

The generators of the cone C_m are of the form

$$g_1 = (1, 0, 0, \dots, 0, 0), g_2 = (1, 1, 0, \dots, 0, 0), \dots, g_m = (1, 1, 1, \dots, 1, 1).$$

The inequality part of the Fenchel conditions can therefore be written as

$$\langle \nabla \phi_{\hat{\lambda}, \hat{\mu}}(\hat{f}), g_j \rangle = \sum_{i=1}^j \left\{ \frac{w_i}{n f_i} - (t_i - t_{i-1}) \right\} \leq 0, \quad j = 1, \dots, m.$$

Using that $\hat{f}_m > 0$, these conditions are equivalent to:

$$\sum_{i=1}^j \left\{ \frac{w_i}{n} - (t_i - t_{i-1}) \hat{f}_i \right\} \leq 0, \quad j = 1, \dots, m,$$

Since, by (3.3), we also have:

$$\sum_{i=1}^m \left\{ \frac{w_i}{n} - (t_i - t_{i-1}) \hat{f}_i \right\} = 0,$$

this proves our claim. \square

We now add the condition $f(t_0) = a$ and proceed in a similar way as in the preceding section to characterize the solution under this restriction.

Lemma 3.2. *Let $t_0 \in (t_{i_0}, t_{i_0+1})$. Denote by $(\hat{f}_1, \dots, \hat{f}(t_0), \dots, \hat{f}_m)$ the vector of values of a piecewise constant nonincreasing function \hat{f} at the observation points t_i and at the point t_0 , where $\hat{f}_{i_0} \geq \hat{f}(t_0) \geq \hat{f}_{i_0+1}$. Then*

- (i) *If \hat{f}_i is given by the left-hand slope of the least concave majorant of the cusum diagram with points $(0, 0)$ and*

$$\left(t_j, \frac{1}{n} \sum_{i=1}^j w_i \right), \quad j = 1, \dots, m, \quad (3.4)$$

and if $\hat{f}_{i_0} \geq a \geq \hat{f}_{i_0+1}$, we put $\hat{f}(t_0) = a$, and $\hat{f}^{(0)} = (\hat{f}_1, \dots, \hat{f}(t_0), \dots, \hat{f}_n)$ is the maximizer of $\sum_{i=1}^n \{\delta_i \log F_i + (1 - \delta_i) \log(1 - F_i)\}$, under the side condition $F(t_0) = a$.

- (ii) *If $(\hat{f}_1, \dots, \hat{f}_n)$ is defined as in (i), but $\hat{f}_{i_0} < a$ or $\hat{f}_{i_0+1} > a$, we define $\hat{\mu} \in \mathbb{R}$ to be the solution (in μ) of the equation*

$$\min_{1 \leq i \leq i_0} \max_{i_0 \leq j \leq m} \frac{\sum_{k=i}^j w_k/n + \mu a}{(t_j - t_{i-1})} = a\{1 + \mu a\}. \quad (3.5)$$

and define $\hat{f}_i^{(0)}$ by the left-hand slope of the least concave majorant of the cusum diagram with points $(0, 0)$ and cusum diagram with points $(0, 0)$ and

$$\left((1 + \hat{\mu}a) t_j, \sum_{i=1}^j \left\{ \frac{w_i}{n} + \hat{\mu}a 1_{\{i=i_0\}} \right\} \right), \quad j = 1, \dots, m, \quad (3.6)$$

Then \hat{f} maximizes $\sum_{i=1}^m w_i \log f_i$, under the condition that f is nonincreasing and the side conditions $\sum_{i=1}^m f_i (t_i - t_{i-1}) = 1$ and $f(t_0) = a$.

Remark 3.1. Although \hat{f}_i and $\hat{f}_i^{(0)}$ are defined by left-continuous slopes of a concave majorant, we define the functions \hat{f} and $\hat{f}^{(0)}$ as a piecewise right-continuous functions, having the values \hat{f}_i and $\hat{f}_i^{(0)}$ at t_i . In this sense the definition is analogous to the definition of the distribution functions \hat{F} and $\hat{F}^{(0)}$ in the preceding section.

Proof. (i) This follows from Lemma 3.1.

(ii) Introducing the Lagrange multipliers λ and μ , we get the maximization problem of maximizing

$$\phi_{\lambda, \mu}(f_1, \dots, f_m) = \frac{1}{n} \sum_{i=1}^m w_i \log f_i - \lambda \left\{ \sum_{i=1}^m f_i (t_i - t_{i-1}) - 1 \right\} + \mu (f_{i_0} - a), \quad (3.7)$$

over the convex cone $C_m = \{(f_1, \dots, f_m) : f_1 \geq f_2 \geq \dots \geq f_m \geq 0\}$, where we look for $(\hat{\lambda}, \hat{\mu}) \in \mathbb{R}_+ \times \mathbb{R}$ such that the maximizer $\hat{f} = (\hat{f}_1, \dots, \hat{f}_n)$ satisfies

$$\sum_{i=1}^m \hat{f}_i (t_i - t_{i-1}) = 1 \quad \text{and} \quad \hat{f}_{i_0} = a.$$

Using the equality part of the Fenchel conditions for this maximization problem, the solution has to satisfy

$$\begin{aligned} \langle \nabla \phi_{\hat{\lambda}, \hat{\mu}}(\hat{f}), \hat{f} \rangle &= \frac{1}{n} \sum_{i=1}^m w_i - \hat{\lambda} \sum_{i=1}^m (t_i - t_{i-1}) \hat{f}_i + \hat{\mu} \hat{f}_{i_0} \\ &= 1 - \hat{\lambda} \sum_{i=1}^m (t_i - t_{i-1}) \hat{f}_i + \hat{\mu} \hat{f}_{i_0} = 1 - \hat{\lambda} + \hat{\mu} a = 0. \end{aligned}$$

This yields the following relation between the two Lagrange multipliers:

$$\hat{\mu} = \frac{\hat{\lambda} - 1}{a}. \quad (3.8)$$

The generators of the cone C_m are of the form

$$g_1 = (1, 0, 0, \dots, 0, 0), g_2 = (1, 1, 0, \dots, 0, 0), \dots, g_m = (1, 1, 1, \dots, 1, 1).$$

The inequality part of the Fenchel conditions can therefore be written as

$$\begin{aligned} \langle \nabla \phi_{\hat{\lambda}, \hat{\mu}}(\hat{f}), g_j \rangle &= \sum_{i=1}^j \left\{ \frac{w_i}{n f_i} - \hat{\lambda} (t_i - t_{i-1}) \right\} + \hat{\mu} 1_{\{j \geq i_0\}} \\ &= \sum_{i=1}^j \left\{ \frac{w_i}{n f_i} - \hat{\lambda} (t_i - t_{i-1}) + \hat{\mu} 1_{\{i=i_0\}} \right\} \leq 0, \quad j = 1, \dots, m. \end{aligned}$$

Using that $\hat{f}_m > 0$, these conditions are equivalent to:

$$\sum_{i=1}^j \left\{ \frac{w_i}{n} - \hat{\lambda} (t_i - t_{i-1}) \hat{f}_i + \hat{\mu} 1_{\{i=i_0\}} a \right\} \leq 0, \quad j = 1, \dots, m,$$

which we obtain by multiplying the i -th component of the inner product with \hat{f}_i .

We now consider the equation:

$$g(\lambda, \mu, a) = a \quad (3.9)$$

where

$$g(\lambda, \mu, a) = \min_{1 \leq i \leq i_0} \max_{i_0 \leq j \leq m} \frac{\sum_{k=i}^j w_k/n + \mu a}{\lambda (t_j - t_{i-1})} = \frac{1}{\lambda} \min_{1 \leq i \leq i_0} \max_{i_0 \leq j \leq m} \frac{\sum_{k=i}^j w_k/n + \mu a}{(t_j - t_{i-1})}.$$

Note that $g(\lambda, \mu, a)$ is the left hand slope of the least concave majorant of the cusum diagram with points $(0, 0)$ and

$$\left(\lambda t_j, \sum_{i=1}^j \left\{ \frac{w_i}{n} + \mu a 1_{\{i=i_0\}} \right\} \right), \quad j = 1, \dots, m,$$

evaluated at λt_{i_0} . So $g(\lambda, \mu, a)$ should be equal to the value of the restricted MLE at t_0 and hence should be equal to a .

On the other hand, using the identity $\lambda = 1 + a\mu$, (3.9) turns into

$$\mu = \frac{1}{a^2} \min_{1 \leq i \leq i_0} \max_{i_0 \leq j \leq m} \frac{\sum_{k=i}^j w_k/n + \mu a}{(t_j - t_{i-1})} - \frac{1}{a}.$$

Multiplying by a^2 yields (3.5). □

The cusum diagram for the restricted MLE is shown in Figure 9 for a sample of size $n = 1000$ from a truncated exponential distribution on $[0, 2]$, where we subtract the the line connecting $(0, 0)$ and $(\hat{\lambda} t_m, 1 + a\hat{\mu})$ for clearer visibility of the difference between the least concave majorant and the values of the cusum diagram. We took $i_0 = 700$, which gave $t_{i_0} = 0.909047$ and a value $\hat{f}_n(t_{i_0}) = 0.519022$ for the unrestricted MLE at t_{i_0} . The restricted MLE was specified to have the value $0.519022 + 0.1 = 0.619022$ at t_{i_0} . The computation

of the restricted MLE gave $\hat{\mu} = 0.064020$ and t_{i_0} was transformed into the point 0.945073 on the axis of the cumulative weights by multiplying by $1 + a\hat{\mu}$.

The lifting of the cusum diagram at $(1 + a\hat{\mu})t_{i_0}$ is clearly visible in part (a) of Figure 9. Part (b) of this figure shows that the unrestricted MLE is globally changed over the whole interval instead of the only local change of the MLE in the current status model. Nevertheless, the (universal) limit distribution of the log likelihood ratio statistic is the same as in the current status model, as we show below.

Remark 3.2. Note that it is clear from the geometric construction that the penalty in the cusum diagram will only locally lead to different points of jump of the restricted MLE on an interval D_n with respect to the unrestricted MLE. Outside D_n the points of jump will be the same. This correspondence also follows from the minmax characterization of the MLEs. The correspondence of the points of jump outside D_n is also clearly visible in part (b) of Figure 9, where the restricted and unrestricted MLE are plotted in the same scale.

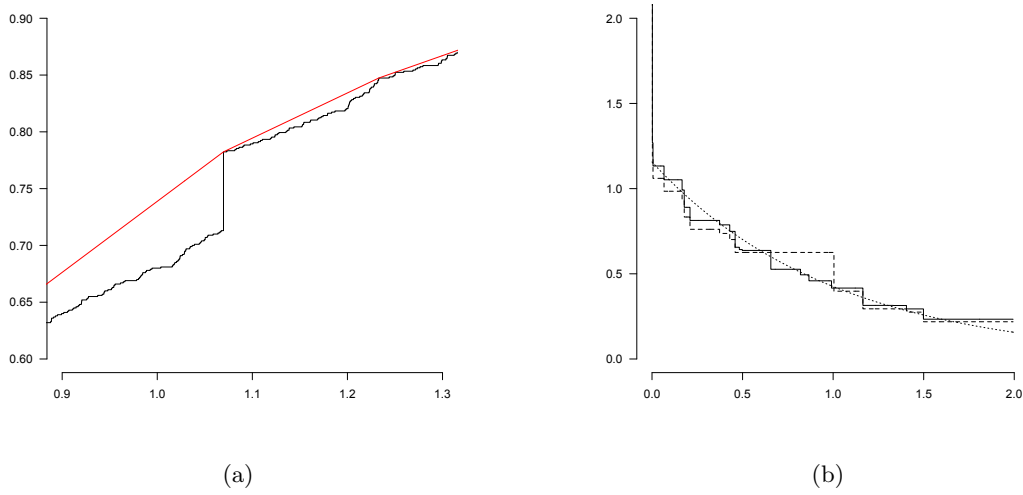


Fig 9: Cusum diagram and MLEs for a sample of size $n = 1000$ from a truncated exponential distribution on $[0, 2]$. (a): cusum diagram with added penalty for the restricted MLE, where we subtract the line connecting $(0, 0)$ and $(\hat{\lambda}t_m, 1 + a\hat{\mu})$ to get a clearer picture. The penalty is added at the location 0.945073 on the x -axis. (b): the restricted MLE (dashed) and the unrestricted MLE (solid).

The proof of Theorem 3.1 below will use the following lemma, which is similar to Lemma 2.3.

Lemma 3.3. *Under the conditions of Theorem 3.1 we have, if $a = f_0(t_0)$,*

$$\hat{\mu}_n = O_p\left(n^{-2/3}\right).$$

Proof. Suppose $t_0 \in (t_{i_0}, t_{i_0+1})$. Consider the function

$$\phi : \mu \mapsto \min_{k \leq i_0} \max_{i \geq i_0} \frac{\sum_{j=k}^i w_j / n + \mu a}{(1 + \mu a)(t_i - t_{k-1})}, \quad a = f_0(t_0).$$

By the least concave majorant characterization of the unrestricted MLE \hat{f}_n , we have

$$\phi(0) = \min_{k \leq i_0} \max_{i \geq i_0} \frac{\sum_{j=k}^i w_j}{n(t_i - t_{k-1})} = \hat{f}_n(t_0).$$

Let $k_1 \leq i_0$ and $i_1 \geq i_0$ be the indices, satisfying

$$\hat{f}_n(t_{i_0}) = \frac{\sum_{j=k_1}^{i_1} w_j}{n(t_{i_1} - t_{k_1-1})} = \min_{k \leq i_0} \max_{i \geq i_0} \frac{\sum_{j=k}^i w_j}{n(t_i - t_{k-1})}.$$

Note that, by the definition of \hat{f}_n , t_{i_1+1} is the first point of jump of \hat{f}_n to the right of t_{i_0} and t_{k_1} the last point of jump before t_{i_0} .

Suppose $a > \hat{f}_n(t_{i_0})$ and let, for $\mu > 0$, $k_\mu \leq i_0$ be the index such that

$$\frac{\sum_{j=k_\mu}^{i_1} w_j/n + \mu a}{t_{i_1} - t_{k_\mu-1}} = \min_{k \leq i_0} \frac{\sum_{j=k}^{i_1} w_j/n + \mu a}{t_{i_1} - t_{k-1}}$$

Then, if $a(t_{i_1} - t_{k_\mu-1}) \neq 1$, there exists a $\mu > 0$ such that

$$\frac{\sum_{j=k_\mu}^{i_1} w_j + n\mu a}{n(t_{i_1} - t_{k_\mu-1})} = \min_{k \leq i_0} \frac{\sum_{j=k}^{i_1} w_j + n\mu a}{n(t_{i_1} - t_{k-1})} = a(1 + \mu a),$$

and this μ is given by:

$$\mu = \frac{a(t_{i_1} - t_{k_\mu-1}) - \sum_{j=k_\mu}^{i_1} w_j/n}{a\{1 - a(t_{i_1} - t_{k_\mu-1})\}}.$$

Using $a = f_0(t_0)$, this can be written in the form:

$$0 < \mu f_0(t_0) = \frac{\int_{t \in (t_{k_\mu-1}, t_{i_1}]} f_0(t_0) dt - \int_{t \in (t_{k_\mu-1}, \tau_{i_1}]} d\mathbb{F}_n(t)}{1 - \int_{t \in (t_{k_\mu-1}, \tau_{i_1}]} f_0(t_0) dt}, \quad (3.10)$$

where \mathbb{F}_n is defined by

$$\mathbb{F}_n(t) = n^{-1} \sum_{i: t_i \leq t} w_i.$$

As noted above, t_{i_1+1} is the first point of jump of \hat{f}_n to the right of t_{i_0} . Let $\tau_+ = t_{i_1}$. As in the proof of Lemma 2.3, we have: $\tau_+ - t_{i_0} = O_p(n^{-1/3})$. To see this, note that, by (3.10), we must have:

$$\int_{t \in (t_{k_\mu-1}, \tau_+]} f_0(t_0) dt - \int_{t \in (t_{k_\mu-1}, \tau_+]} d\mathbb{F}_n(t) > 0,$$

and

$$\begin{aligned} & \int_{t \in (t_{k_\mu-1}, \tau_+]} f_0(t_0) dt - \int_{t \in (t_{k_\mu-1}, \tau_+]} d\mathbb{F}_n(t) \\ &= \int_{t \in (t_{k_\mu-1}, \tau_+]} \{f_0(t_0) - f_0(t)\} dt - \int_{t \in (t_{k_\mu-1}, \tau_+]} d(\mathbb{F}_n - F_0)(t), \end{aligned}$$

where the first term on the right gives a negative parabolic drift which cannot be compensated by the second random term outside a neighborhood of order $O_p(n^{-1/3})$ of t_0 .

By the same type of argument, we can choose for each $\varepsilon > 0$ an $M > 0$ such that

$$\mathbb{P} \left\{ \frac{\int_{u \in (t, \tau_+]} f_0(t_0) du - \int_{u \in (t, \tau_+]} d\mathbb{F}_n(u)}{1 - \int_{u \in (t, \tau_+]} f_0(t_0) du} < 0 \right\} > 1 - \varepsilon,$$

if $t < t_{i_0} - Mn^{-1/3}$. But since we must have

$$\int_{t \in (t_{k_\mu-1}, \tau_+]} f_0(t_{i_0}) dt - \int_{t \in (t_{k_\mu-1}, \tau_+]} d\mathbb{F}_n(t) > 0,$$

by the positivity of μ and relation (3.10), it now follows that $t_{i_0} - t_{k_\mu-1} = O_p(n^{-1/3})$ and therefore

$$\mu f_0(t_{i_0}) = \frac{\int_{t \in (t_{k_\mu-1}, \tau_+]} f_0(t_{i_0}) dt - \int_{t \in (t_{k_\mu-1}, \tau_+]} d\mathbb{F}_n(t)}{1 - \int_{t \in (t_{k_\mu-1}, \tau_+]} f_0(t_{i_0}) dt} = O_p(n^{-2/3}).$$

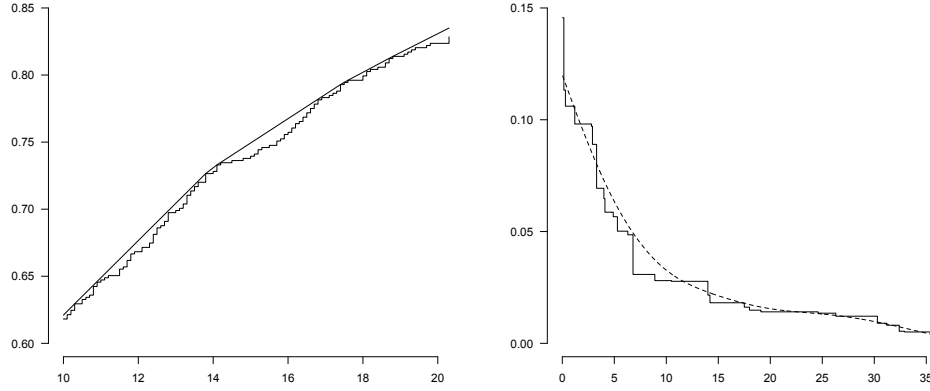


Fig 10: The left panel shows the empirical distribution function and its least concave majorant for the values between 10 and 20 months of the 618 current durations ≤ 36 months. The resulting Grenander estimate (the MLE) of the observation density on the interval $[0, 36]$ is shown in the right panel, together with its smoothed version (dashed, the SMLE)

Hence $\mu = O_p(n^{-2/3})$ and

$$\phi(\mu) = \min_{k \leq i_0} \max_{i \geq i_0} \frac{\sum_{j=k}^i w_j/n + a\mu}{t_i - t_{k-1}} \geq \min_{k \leq i_0} \frac{\sum_{j=k}^{i_1} w_j/n + a\mu}{t_i - t_{k-1}} = a(1 + a\mu).$$

As in the proof of Lemma 2.3 we can now conclude

$$0 \leq \hat{\mu}_n \leq \mu = O_p(n^{-2/3}).$$

The case $a < \hat{f}_n(t_{i_0})$ can be treated in a similar way. □

We can now prove the following result. The proof is given in Section 6.

Theorem 3.1. *Let f_0 be a decreasing density, which is continuous and has a continuous strictly negative derivative f'_0 in a neighborhood of t_0 . Let \hat{f}_n be the unrestricted MLE and let $\hat{f}_n^{(0)}$ be the MLE under the restriction that $\hat{f}_n^{(0)}(t_0) = f_0(t_0)$. Moreover, let the log likelihood ratio statistic $2 \log \ell_n$ be defined by*

$$2 \log \ell_n = 2 \sum_{i=1}^n \log \frac{\hat{f}_n(T_i)}{\hat{f}_n^{(0)}(T_i)}.$$

Then

$$2 \log \ell_n \xrightarrow{\mathcal{D}} \mathbb{D},$$

where \mathbb{D} is the universal limit distribution as given in [1].

Remark 3.3. The condition that f_0 has a continuous strictly negative derivative f'_0 in a neighborhood of t_0 corresponds to “condition A” in [1] for the current status model, which is the condition that the derivative f'_0 of F_0 is strictly positive at t_0 and continuous in a neighborhood of t_0 . A condition of this type is necessary for getting Brownian motion with parabolic drift in the limit distribution of the MLEs. This fails if we take f_0 uniform, in which case we get a different type of asymptotics. See section 3.10 in [8].

Example 3.1. Suppose we have a sample Z_1, \dots, Z_n from the length biased distribution, associated with an unknown distribution function F of interest. This means that the distribution function of Z_i is given by

$$\bar{F}(z) = P(Z_i \leq z) = \frac{1}{m_F} \int_0^z x dF(x) \quad (3.11)$$

where $m_F = \int_0^\infty x dF(x)$ is assumed to be nonzero and finite. However, instead of observing the values of Z_i directly, we only observe the data X_1, \dots, X_n where X_i is a uniform random fraction of Z_i . More specifically, we observe

$$X_i = U_i Z_i,$$

where U_1, \dots, U_n is a random sample from the uniform distribution on $[0, 1]$, independent of the Z_i 's. Now the density of X_i can be seen to be

$$g(x) = \frac{1}{m_F} (1 - F(x)), \quad x \geq 0, \quad (3.12)$$

see (2.5) in Section 2.2 and Exercise 2.4 in [8]. This means that the survival function $1 - F(x)$ is given by $g(x)/g(0)$.

Hence, by monotonicity of the initial distribution function F and the fact that $0 < m_F < \infty$, it follows that sampling density g is bounded and decreasing on $[0, \infty)$. Moreover, if no additional assumptions are imposed on F , any density of this type can be represented by (3.12). The density g can be estimated by the Grenander estimator of a decreasing density. See [18] and [17] for applications of this model.

In [16] a data set of current durations of pregnancy in France is studied. The aim is to estimate the distribution of the time it takes for a woman to become pregnant after having started unprotected sexual intercourse. For 867 women the current duration of unprotected intercourse, measured in months, was recorded and this is the basis of part of the research, reported in [16].

Given that the woman in the study is currently trying to become pregnant, the actual recorded data (current duration) can be viewed as uniform random fraction of the true, total duration. In that sense, the model as given in (3.12) is not unreasonable. The left panel of Figure 10 shows a part of the empirical distribution function of 618 recorded current durations, kindly provided to us by Niels Keiding, where the data are truncated at 36 months and are of a similar nature as the data in [16]. Based on the least concave majorant, the right panel of Figure 10 is computed, showing the resulting MLE of the decreasing density of the observations together with its smoothed version, the smoothed maximum likelihood estimator (SMLE), defined by

$$\tilde{g}_{nh}(t) = - \int \mathbb{K}((t-x)/h) d\hat{g}_n(x), \quad \mathbb{K}(x) = \int_x^\infty K(u) du, \quad (3.13)$$

where \hat{g}_n is the Grenander estimator (the MLE) and K is a symmetric kernel, for which we took the triweight kernel

$$K(u) = \frac{35}{32} (1 - u^2)^3 1_{[-1,1]}(u), \quad u \in \mathbb{R}.$$

The bandwidth h was chosen to be

$$h = 36n^{-1/5} \approx 9.95645,$$

where $n = 618$. Near the boundary points 0 and 36 the same boundary correction as in section 2 was used. For $t \in [h, b-h]$, where $b = 36$, the SMLE is asymptotically equivalent to the ordinary kernel density estimator

$$\int K_h(t-x) d\mathbb{F}_n(x), \quad K_h(u) = h^{-1} K(u/h), \quad (3.14)$$

which, however, will in general not be monotone, so not belong to the allowed class.

The 95% confidence intervals for the density (3.12), based on the SMLE and the LR test for the MLE, respectively, are shown in Figure 11. The survival function for the time until pregnancy or end of the period of unprotected intercourse is given by $g(x)/g(0)$, where g is the density of the observations. The 95% confidence intervals for the survival function at the 99 equidistant points 0.36, 0.72, \dots , 35.64, are constructed from 1000 bootstrap samples T_1^*, \dots, T_n^* , also of size n , drawn from the original sample, and in these samples we computed

$$\tilde{g}_{nh}^*(t)/\tilde{g}_{nh}^*(0) - \tilde{g}_{nh}(t)/\tilde{g}_{nh}(0), \quad (3.15)$$

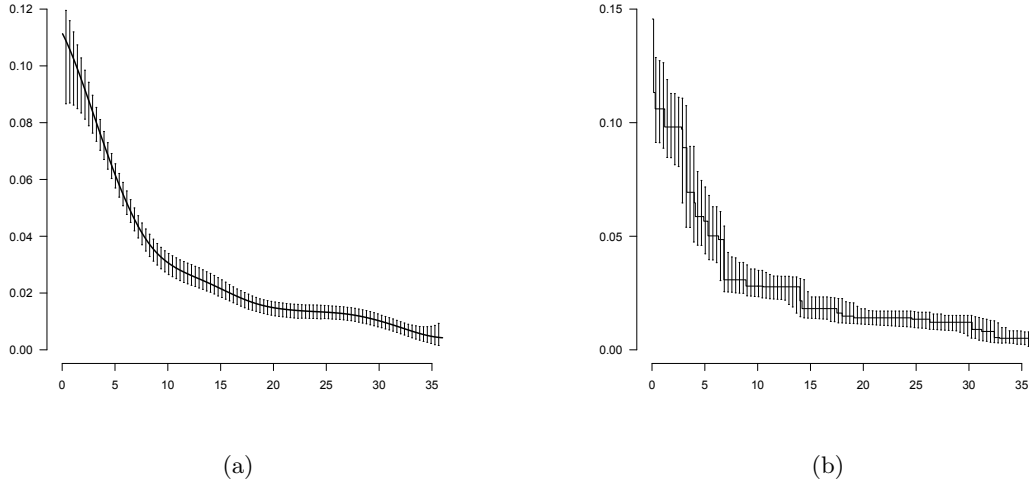


Fig 11: 95% confidence intervals, based on the SMLE (part (a)) and MLE (part (b)), respectively, for the data in [16] at the points $0.36, 0.72, \dots, 35.64$. The chosen bandwidth for the SMLE was $36n^{-1/4} \approx 7.2203$. The time is measured in months.

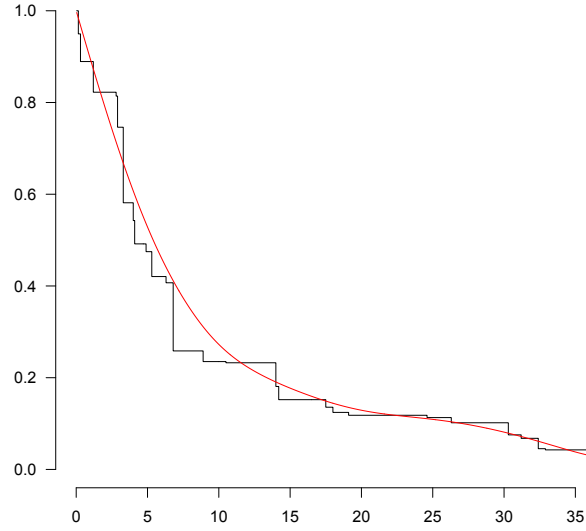


Fig 12: Estimates of the survival function, based on the MLE (step function) and SMLE (smooth function), where the MLE is restricted to have the same value at zero as the (consistent) SMLE.

where \tilde{g}_{nh} and \tilde{g}_{nh}^* are the SMLEs in the original sample and the bootstrap sample, respectively. The chosen bandwidth was $36n^{-1/4} \approx 7.2203$, so (according to the method of undersmoothing, see section 2), smaller than the bandwidth used in Figure 10, which uses a bandwidth for which the squared bias and variance are

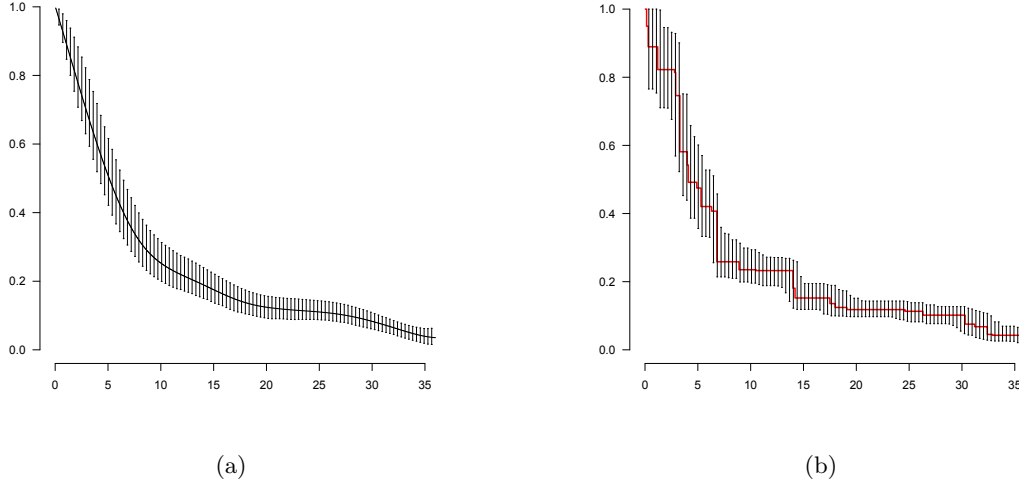


Fig 13: 95% confidence intervals, based on the SMLE (part (a)) and MLE (part (b)), respectively, for the survival functions in [16] at the points $0.36, 0.72, \dots, 35.64$. The chosen bandwidth for the SMLE was $36n^{-1/4} \approx 7.2203$ and the MLE was restricted to have the same value as the (consistent) SMLE at zero.

approximately in equilibrium. The 95% asymptotic confidence intervals are given by:

$$[\tilde{g}_{nh}(t)/\tilde{g}_{nh}(0) - U_{0.975}^*, \tilde{g}_{nh}(t)/\tilde{g}_{nh}(0) - U_{0.025}^*],$$

where $U_{0.025}^*$ and $U_{0.975}^*$ are the 2.5% and 97.5% percentiles of the bootstrap values (3.15). The result is shown in Figure 13a and should be compared with the confidence intervals in part A of Figure 2, p. 1495 of [16], based on a parametric (generalized gamma) model.

We have here the easiest, but also somewhat unusual, situation that the isotonic estimator is asymptotically equivalent to an ordinary non-isotonic estimator. The more usual situation is that we only can find a so-called “toy estimator”, which is asymptotically equivalent to the MLE or SMLE, but still contains parameters that have to be estimated. This is the case in the current status model as seen in section 2.

In [16] and [11] also parametric models are considered for analyzing these data. We compute the MLE as the slope of the smallest concave majorant of the data ≤ 36 months, where the x -values are only the strictly different values, and where we use the number of values at a tie as the increase of the second coordinate of the cusum diagram. In this way we get 618 values ≤ 36 , but only 248 strictly different ones. It is clear that the SMLE has a somewhat intermediate position w.r.t. the parametric models and the fully nonparametric MLE, considered in [16] and [11].

In the model considered here, the nonparametric MLE is inconsistent at zero and can therefore not be used as an estimate of $g(0)$ and therefore also not as an estimate of the survival function $g(x)/g(0)$, unless we also use penalization at zero. This is in contrast with the SMLE, which is consistent at zero due to the boundary correction. This difficulty with the inconsistency of the MLE at zero for the present model is discussed in [11]. We solve this difficulty by adding a penalty at zero, as in [19], and maximize the function

$$\phi_{\alpha, \lambda, \mu}(f_1, \dots, f_m) = \frac{1}{n} \sum_{i=1}^m w_i \log f_i - \lambda \left\{ \sum_{i=1}^m f_i (t_i - t_{i-1}) - 1 \right\} + \mu (f_{i_0} - a) - \alpha (f_1 - b), \quad (3.16)$$

where b is the value of a consistent estimator at zero (for example, the value of the SMLE); we switch to the notation $f = (f_1, \dots, f_m)$ again (instead of using g) to be in line with the presentation in the preceding

section. The solution has to satisfy

$$\begin{aligned} \langle \nabla \phi_{\hat{\alpha}, \hat{\lambda}, \hat{\mu}}(\hat{f}), \hat{f} \rangle &= \frac{1}{n} \sum_{i=1}^m w_i - \hat{\lambda} \sum_{i=1}^m (t_i - t_{i-1}) \hat{f}_i + \hat{\mu} \hat{f}_{i_0} - \hat{\alpha} \hat{f}_1 \\ &= 1 - \hat{\lambda} \sum_{i=1}^m (t_i - t_{i-1}) \hat{f}_i + \hat{\mu} \hat{f}_{i_0} - \alpha \hat{f}_1 = 1 - \hat{\lambda} + \hat{\mu} a - \hat{\alpha} b = 0, \end{aligned}$$

and hence

$$\hat{\mu} = \frac{\hat{\lambda} - 1 + \hat{\alpha} b}{a}. \quad (3.17)$$

Analogously to Lemma 3.2, we now get the following lemma.

Lemma 3.4. *Let $\hat{f} = (\hat{f}_1, \dots, \hat{f}_m)$ be the vector of slopes of the least concave majorant of the cusum diagram with points $(0, 0)$ and*

$$\left(\hat{\alpha} + \hat{\lambda} t_j, \sum_{i=1}^j \left\{ \frac{w_i}{n} + (\hat{\lambda} - 1 + \hat{\alpha} b) \{i = i_0\} \right\} \right), \quad j = 1, \dots, m, \quad (3.18)$$

where $(\hat{\alpha}, \hat{\lambda})$ is the solution of the equations (in (α, λ))

$$\min_{1 \leq i \leq i_0} \max_{i_0 \leq j \leq m} \frac{\sum_{k=i}^j w_k/n + \lambda - 1 + \alpha b}{\lambda(t_j - t_{i-1}) + 1_{\{i=1\}} \alpha} = a, \quad \max_{i \geq 1} \frac{\sum_{j=1}^i w_j/n}{\alpha + \lambda t_i} = b. \quad (3.19)$$

Then \hat{f} maximizes $\sum_{i=1}^m w_i \log f_i$, under the condition that f is nonincreasing and the boundary conditions

$$\sum_{i=1}^m f_i (t_i - t_{i-1}) = 1, \quad f_1 = b \quad \text{and} \quad f_{i_0} = a.$$

We now restrict the MLE of the density to have a value at zero, given by a consistent estimator at zero. There are several possible choices; we took the value of the SMLE at zero for illustrative purposes. The resulting estimate of the survival function, based on the MLE restricted at zero to have the same value as the SMLE, is shown in Figure 12. It is also possible to take histogram-type estimates at zero if one wants to impose more lenient conditions. Next we can compute the 95% confidence intervals again by the likelihood ratio method, where one restricts the MLE to have a value at zero, prescribed by the consistent estimate. Using Lemma 3.4 we can then compute the LR tests again for the values of f_{i_0} . The result is shown in part (b) of Figure 13, where we used the same asymptotic critical values as before.

4. Computational aspects and concluding remarks

There are several ways of computing the restricted MLE's. One way of computing the restricted MLE for the current status model was given in [1], see the discussion following Remark 2.1 in Section 2. We computed the restricted MLE by first solving equations (2.4), (3.5) or (3.19) for the Lagrange multiplier $\hat{\mu}$ or $\hat{\alpha}$ and $\hat{\lambda}$, and next computing in one step the left derivative of the greatest convex minorant, resp. the smallest concave majorant, of the cusum diagrams which were constructed using the Lagrange multipliers. So the iterative part of the algorithm is in determining the solution $\hat{\mu}$ or $\hat{\alpha}$ and $\hat{\lambda}$. For the monotone density case it is not clear that a completely non-iterative method for computing the restricted MLE exists (as in the current status model, if one adapts the definition in terms of inequalities in [1]). For solving the non-linear equations for $\hat{\mu}$ or $\hat{\alpha}$ and $\hat{\lambda}$ in Lemma 3.4 we wrote C programs, which seems to work fine.

In practice we would recommend to use the methods based on the MLE or SMLE in conjunction; the intervals based on the LR test for the MLE seem pretty much on target, except perhaps for values close to the boundary, and use less assumptions. On the other hand, the intervals, based on the SMLE are narrower and based on asymptotically normal limit distributions, which enables the use of bootstrap methods in constructing the confidence intervals. Direct bootstrap methods have been shown to fail for the MLE, see [13] and [15].

5. Appendix A

Proof of Theorem 2.1. Let D_n be the smallest interval $[a_n, b_n)$ such that \hat{F}_n and $\hat{F}_n^{(0)}$ coincide on D_n^c and such that the boundary points of D_n are points of jump of \hat{F}_n and $\hat{F}_n^{(0)}$; we assume \hat{F}_n and $\hat{F}_n^{(0)}$ to be right-continuous. Then $\hat{\mu}_n = O_p(n^{-2/3})$ and, as argued in the proof of Lemma 2.3, the nearest points of jump to t_0 of $\hat{F}_n^{(0)}$ and \hat{F}_n are at distance $O_p(n^{-1/3})$ of t_0 .

Suppose $t_\ell > t_{i_1}$, where $(t_{k_1-1}, t_{i_1}]$ is the interval around t_0 where $\hat{F}_n^{(0)}$ is constant. The maxmin characterization of $\hat{F}_n^{(0)}$ then gives

$$\hat{F}_n^{(0)}(t_\ell) = \max_{i_0 < k \leq \ell} \min_{i \geq \ell} \frac{\sum_{j=k}^i \delta_j}{i - k + 1}.$$

Note that the term $\hat{\mu} a 1_{\{i=i_0\}}$ does no longer occur in the minmax characterization, since the relevant intervals do not contain i_0 . Likewise, if $t_\ell > t_{i'_1}$, where $(t_{k'_1-1}, t_{i'_1}]$ is the interval around t_0 where \hat{F}_n is constant, the maxmin characterization of \hat{F}_n gives

$$\hat{F}_n(t_\ell) = \max_{i_0 < k \leq \ell} \min_{i \geq \ell} \frac{\sum_{j=k}^i \delta_j}{i - k + 1}.$$

Since we have $t_{i_1} - t_0 = O_p(n^{-1/3})$ and $t_{i'_1} - t_0 = O_p(n^{-1/3})$, we get therefore that the functions \hat{F}_n and $\hat{F}_n^{(0)}$ coincide with high probability for values $t \geq t_0 + Mn^{-1/3}$, is $M > 0$ is sufficiently large. The same argument holds on intervals to the left of t_0 . In other words: the length of the interval $D_n = [a_n, b_n)$ is of order $O_p(n^{-1/3})$. By the monotonicity of the functions \hat{F}_n and $\hat{F}_n^{(0)}$ and the properties of the unrestricted \hat{F}_n , this also implies:

$$\sup_{t \in D_n} |\hat{F}_n(t) - F_0(t_0)| = O_p(n^{-1/3}) \quad \text{and} \quad \sup_{t \in D_n} |\hat{F}_n^{(0)}(t) - F_0(t_0)| = O_p(n^{-1/3}). \quad (5.1)$$

We now have, by (5.1), and the fact that the length of D_n is of order $O_p(n^{-1/3})$,

$$\begin{aligned} & 2n \int_{D_n} \left\{ \delta \log \frac{\hat{F}_n(t)}{\hat{F}_n^{(0)}(t)} + (1 - \delta) \log \frac{1 - \hat{F}_n(t)}{1 - \hat{F}_n^{(0)}(t)} \right\} d\mathbb{P}_n(t, \delta) \\ &= -2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{\hat{F}_n(t)} - (1 - \delta) \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{1 - \hat{F}_n(t)} \right\} d\mathbb{P}_n(t, \delta) \\ & \quad + n \int_{D_n} \left\{ \delta \frac{\{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}^2}{\hat{F}_n(t)^2} + (1 - \delta) \frac{\{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}^2}{\{1 - \hat{F}_n(t)\}^2} \right\} d\mathbb{P}_n(t, \delta) + o_p(1). \end{aligned} \quad (5.2)$$

Denote the empirical distribution function of the T_i 's by \mathbb{G}_n and define the probability measure $P_{\hat{F}_n}$ by

$$\int \psi(t, \delta) dP_{\hat{F}_n}(t, \delta) = \int \psi(t, \hat{F}_n(t)) d\mathbb{G}_n(t),$$

for bounded Borel measurable functions on $\psi : \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}$. We similarly define $P_{\hat{F}_n^{(0)}}$ by:

$$\int \psi(t, \delta) dP_{\hat{F}_n^{(0)}}(t, \delta) = \int \psi(t, \hat{F}_n^{(0)}(t)) d\mathbb{G}_n(t).$$

Then

$$\begin{aligned} & 2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{\hat{F}_n(t)} - (1 - \delta) \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{1 - \hat{F}_n(t)} \right\} dP_{\hat{F}_n}(t, \delta) \\ &= 2n \int_{D_n} \frac{\{\delta - \hat{F}_n(t)\} \{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}}{\hat{F}_n(t) \{1 - \hat{F}_n(t)\}} dP_{\hat{F}_n}(t, \delta) \\ &= 2n \int_{D_n} \frac{\{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\} \{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}}{\hat{F}_n(t) \{1 - \hat{F}_n(t)\}} d\mathbb{G}_n(t) = 0, \end{aligned}$$

and hence:

$$\begin{aligned}
& -2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{\hat{F}_n(t)} - (1-\delta) \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{1 - \hat{F}_n(t)} \right\} d\mathbb{P}_n(t, \delta) \\
& = -2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{\hat{F}_n(t)} - (1-\delta) \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{1 - \hat{F}_n(t)} \right\} d(\mathbb{P}_n - P_{\hat{F}_n})(t, \delta) \\
& = -2n \int_{D_n} \frac{\{\delta - \hat{F}_n(t)\} \{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}}{\hat{F}_n(t) \{1 - \hat{F}_n(t)\}} d(\mathbb{P}_n - P_{\hat{F}_n})(t, \delta) \\
& = -2n \int_{D_n} \frac{\{\delta - \hat{F}_n(t)\} \{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}}{\hat{F}_n(t) \{1 - \hat{F}_n(t)\}} d(\mathbb{P}_n - P_0)(t, \delta) \\
& \quad - 2n \int_{D_n} \frac{\{\delta - \hat{F}_n(t)\} \{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)\}}{\hat{F}_n(t) \{1 - \hat{F}_n(t)\}} d(P_0 - P_{\hat{F}_n})(t, \delta)
\end{aligned}$$

We can write:

$$\begin{aligned}
& -2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{1 - F_0(t_0)} \right\} d(\mathbb{P}_n - P_{\hat{F}_n})(t, \delta) \\
& = -2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - F_0(t_0)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d(\mathbb{P}_n - P_{\hat{F}_n})(t, \delta) \\
& \quad + 2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d(\mathbb{P}_n - P_{\hat{F}_n})(t, \delta) \\
& = 2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d(P_{\hat{F}_n^{(0)}} - P_{\hat{F}_n})(t, \delta) + o_p(1).
\end{aligned}$$

The first term of the next to last expression is zero by the characterization of \hat{F}_n , since

$$\begin{aligned}
& \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - F_0(t_0)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d(\mathbb{P}_n - P_{\hat{F}_n})(t, \delta) \\
& = \frac{1}{F_0(t_0) \{1 - F_0(t_0)\}} \int_{D_n} \{\delta - \hat{F}_n(t)\} \{\hat{F}_n(t) - F_0(t_0)\} d\mathbb{P}_n(t, \delta)
\end{aligned}$$

The characterization of $\hat{F}_n^{(0)}$ implies:

$$\begin{aligned}
& \int_{D_n} \left\{ \delta \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d\mathbb{P}_n(t, \delta) \\
& = \frac{1}{F_0(t_0) \{1 - F_0(t_0)\}} \int_{D_n} \{\delta - \hat{F}_n^{(0)}(t)\} \{\hat{F}_n^{(0)}(t) - F_0(t_0)\} d\mathbb{P}_n(t, \delta) \\
& = \frac{1}{F_0(t_0) \{1 - F_0(t_0)\}} \int_{D_n} \{\delta - \hat{F}_n^{(0)}(t)\} \{\hat{F}_n^{(0)}(t) - a\} d\mathbb{P}_n(t, \delta) = 0.
\end{aligned}$$

Note that the last equality holds because $\hat{F}_n^{(0)}(t) - a = 0$ at $t = t_{i_0}$, at which point there is the added jump of $\hat{\mu}_n a(1-a)$ in the cusum diagram. By what we just derived, we also have:

$$\begin{aligned}
& \int_{D_n} \left\{ \delta \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - (1-\delta) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d(\mathbb{P}_n - P_{\hat{F}_n^{(0)}})(t, \delta) \\
& = \frac{1}{F_0(t_0) \{1 - F_0(t_0)\}} \int_{D_n} \{\delta - \hat{F}_n^{(0)}(t)\} \{\hat{F}_n^{(0)}(t) - F_0(t_0)\} d\mathbb{P}_n(t, \delta) = 0,
\end{aligned}$$

$$\begin{aligned} & \int_{D_n} \left\{ \delta \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - (1 - \delta) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} dP_{\hat{F}_n^{(0)}}(t, \delta) \\ & \int_{D_n} \left\{ \hat{F}_n^{(0)}(t) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - \{1 - \hat{F}_n^{(0)}(t)\} \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d\mathbb{G}_n(t) \end{aligned}$$

So we can conclude:

$$\begin{aligned} & -2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{\hat{F}_n(t)} - (1 - \delta) \frac{\hat{F}_n(t) - \hat{F}_n^{(0)}(t)}{1 - \hat{F}_n(t)} \right\} d\mathbb{P}_n(t, \delta) \\ & = 2n \int_{D_n} \left\{ \delta \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{F_0(t_0)} - (1 - \delta) \frac{\hat{F}_n^{(0)}(t) - F_0(t_0)}{1 - F_0(t_0)} \right\} d(P_{\hat{F}_n^{(0)}} - P_{\hat{F}_n})(t, \delta) + o_p(1) \\ & = 2n \int_{D_n} \frac{\{\hat{F}_n^{(0)}(t) - F_0(t_0)\} \{\hat{F}_n^{(0)}(t) - \hat{F}_n(t)\}}{F_0(t_0) \{1 - F_0(t_0)\}} d\mathbb{G}_n(t) + o_p(1). \end{aligned}$$

So we have reduced the problem of finding the asymptotic distribution of:

$$2n \int_{D_n} \frac{\{\hat{F}_n^{(0)}(t) - F_0(t_0)\} \{\hat{F}_n^{(0)}(t) - \hat{F}_n(t)\}}{F_0(t_0) \{1 - F_0(t_0)\}} d\mathbb{G}_n(t) \quad (5.3)$$

Combining this with (5.2), we find:

$$\begin{aligned} & 2n \int_{D_n} \left\{ \delta \log \frac{\hat{F}_n(T_i)}{\hat{F}_n^{(0)}(T_i)} + (1 - \delta) \log \frac{1 - \hat{F}_n(T_i)}{1 - \hat{F}_n^{(0)}(T_i)} \right\} d\mathbb{P}_n(t, \delta) \\ & = n \int_{D_n} \frac{\{\hat{F}_n(t) - F_0(t_0)\}^2 - \{\hat{F}_n^{(0)}(t) - \hat{F}_n(t_0)\}^2}{F_0(t_0) \{1 - F_0(t_0)\}} d\mathbb{G}_n(t) + o_p(1) \\ & = n^{2/3} g(t_0) \int_{t_0 + n^{1/3}(a_n - t_0)}^{t_0 + n^{1/3}(b_n - t_0)} \frac{\{\hat{F}_n(t_0 + n^{-1/3}t) - F_0(t_0)\}^2 - \{\hat{F}_n^{(0)}(t_0 + n^{-1/3}t) - F_0(t_0)\}^2}{F_0(t_0) \{1 - F_0(t_0)\}} dt + o_p(1). \end{aligned}$$

The result now follows, using the joint convergence of the processes

$$\frac{n^{1/3} \{\hat{F}_n(t_0 + n^{-1/3}t) - F_0(t_0)\}}{\sqrt{F_0(t_0) \{1 - F_0(t_0)\}}} \quad \text{and} \quad \frac{n^{1/3} \{\hat{F}_n^{(0)}(t_0 + n^{-1/3}t) - F_0(t_0)\}}{\sqrt{F_0(t_0) \{1 - F_0(t_0)\}}}$$

on bounded intervals to the slope of the least convex minorant of Brownian motion plus the function t^2 at zero without restriction and with the restriction that the slope is zero, respectively. \square

6. Appendix B

Proof of Theorem 3.1. We extend the values \hat{f}_{ni} and $\hat{f}_{ni}^{(0)}$ of the solution \hat{f}_n and $\hat{f}_n^{(0)}$ as vectors to right-continuous functions \hat{f}_n and $\hat{f}_n^{(0)}$ on $[0, \infty)$. Note that the fact that the vector solutions are left-continuous slopes of the least concave majorants of the respective cusum diagrams does not force us to make their extensions to function on \mathbb{R}_+ also left-continuous.

Let D_n be the smallest interval $[a_n, b_n)$ such that \hat{f}_n and $\hat{f}_n^{(0)}$ have the same points of jump on D_n^c and such that the boundary points of D_n are points of jump of \hat{f}_n and $\hat{f}_n^{(0)}$ (see Remark 3.2). Since $\hat{\mu}_n = O_p(n^{-2/3})$, we get:

$$2n \int_{D_n^c} \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) = 2n \log\{1 + a\hat{\mu}_n\} \int_{D_n^c} d\mathbb{F}_n(t) = 2na\hat{\mu}_n \int_{D_n^c} d\mathbb{F}_n(t) + O_p(n^{-1/3}).$$

The function $\hat{f}_n^{(0)}$ must satisfy

$$\int \hat{f}_n^{(0)}(x) dx = 1.$$

So

$$\int_{D_n} \hat{f}_n^{(0)}(t) dt + \int_{D_n^c} \hat{f}_n^{(0)}(t) dt = 1,$$

and hence, using $\hat{\mu}_n = O_p(n^{-2/3})$,

$$\begin{aligned} 2n \int_{D_n} \hat{f}_n^{(0)}(t) dt &= 2n \left\{ 1 - \int_{D_n^c} \hat{f}_n^{(0)}(t) dt \right\} = 2n \left\{ 1 - \{1 + \hat{\mu}_n a\}^{-1} \int_{D_n^c} d\mathbb{F}_n(t) \right\} \\ &= 2n \left\{ \int_{D_n} d\mathbb{F}_n(t) + \hat{\mu}_n a \int_{D_n^c} d\mathbb{F}_n(t) \right\} + O_p(n^{-1/3}). \end{aligned}$$

So we obtain

$$\begin{aligned} &2n \int \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) \\ &= 2n \int_{D_n} \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) + 2n \int_{D_n^c} \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) \\ &= 2n \int_{D_n} \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) + 2na\hat{\mu}_n \int_{D_n^c} d\mathbb{F}_n(t) + O_p(n^{-1/3}) \\ &= 2n \int_{D_n} \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) - 2n \int_{D_n} d(\mathbb{F}_n - \hat{F}_n^{(0)})(t) + O_p(n^{-1/3}) \end{aligned}$$

Expanding in \hat{f}_n , we find:

$$\begin{aligned} &2n \int_{D_n} \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) - 2n \int_{D_n} (\hat{f}_n(t) - \hat{f}_n^{(0)}(t)) dt \\ &= -2n \int_{D_n} \frac{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)}{\hat{f}_n(t)} d\mathbb{F}_n(t) + n \int_{D_n} \frac{\{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)\}^2}{\hat{f}_n(t)^2} d\mathbb{F}_n(t) \\ &\quad - 2n \int_{D_n} (\hat{f}_n(t) - \hat{f}_n^{(0)}(t)) dt + o_p(1) \\ &= -2n \int_{D_n} \frac{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)}{\hat{f}_n(t)} d(\mathbb{F}_n - \hat{F}_n)(t) + n \int_{D_n} \frac{\{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)\}^2}{\hat{f}_n(t)^2} d\mathbb{F}_n(t) + o_p(1) \\ &= -2n \int_{D_n} \frac{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)}{f_0(t_{i_0})} d(\mathbb{F}_n - \hat{F}_n)(t) + n \int_{D_n} \frac{\{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)\}^2}{f_0(t_{i_0})^2} d\mathbb{F}_n(t) + o_p(1). \end{aligned}$$

We now get:

$$\begin{aligned} &\int_{D_n} \{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)\} d(\mathbb{F}_n - \hat{F}_n)(t) \\ &= \int_{D_n} \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\} d(\mathbb{F}_n - \hat{F}_n)(t) - \int_{D_n} \{\hat{f}_n(t) - f_0(t_{i_0})\} d(\mathbb{F}_n - \hat{F}_n)(t) \\ &= \int_{D_n} \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\} d(\mathbb{F}_n - \hat{F}_n)(t) \\ &= \int_{D_n} \{\hat{f}_n^{(0)}(t) - f_0(t_0)\} d(\mathbb{F}_n - \hat{F}_n^{(0)})(t) - \int_{D_n} \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\} d(\hat{F}_n^{(0)} - \hat{F}_n)(t) \\ &= - \int_{D_n} \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\} \{\hat{f}_n^{(0)}(t) - \hat{f}_n(t)\} dt. \end{aligned}$$

In the last equality we use the fact that $\hat{f}_n^{(0)}(t_{i_0}) - a = \hat{f}_n^{(0)}(t_{i_0}) - f_0(t_{i_0}) = 0$, which implies that the first term in the next to last expression is zero, using that the increments w.r.t. $d\mathbb{F}_n$ and $d\hat{F}_n^{(0)}$ are equal on an interval between jumps of $\hat{f}_n^{(0)}$, if the interval does not contain t_{i_0} . So, gathering terms, we obtain:

$$\begin{aligned}
& 2n \int \log \frac{\hat{f}_n(t)}{\hat{f}_n^{(0)}(t)} d\mathbb{F}_n(t) \\
&= n \int_{D_n} \frac{\{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)\}^2}{f_0(t_{i_0})^2} d\mathbb{F}_n(t) - \frac{2n}{f_0(t_{i_0})} \int_{D_n} \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\} \{\hat{f}_n^{(0)}(t) - \hat{f}_n(t)\} dt + o_p(1) \\
&= n \int_{D_n} \frac{\{\hat{f}_n(t) - f_0(t_{i_0})\}^2 - \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\}^2}{f_0(t_{i_0})} dt + n \int_{D_n} \frac{\{\hat{f}_n(t) - \hat{f}_n^{(0)}(t)\}^2}{f_0(t_{i_0})^2} d(\mathbb{F}_n - F_0)(t) \\
&\quad + o_p(1) \\
&= n \int_{D_n} \frac{\{\hat{f}_n(t) - f_0(t_{i_0})\}^2 - \{\hat{f}_n^{(0)}(t) - f_0(t_{i_0})\}^2}{f_0(t_{i_0})} dt + o_p(1).
\end{aligned}$$

The expression on the last line converges under the null hypothesis in distribution to the limit distribution of the likelihood ratio test for the current status model, as described in Theorem 2.5, p. 1710 of [1]. \square

References

- [1] M. Banerjee and J.A. Wellner, *Likelihood ratio tests for monotone functions*, Ann. Statist. **29** (2001), 1699–1731. [MR1891743 \(2003c:62072\)](#)
- [2] ———, *Confidence intervals for current status data*, Scand. J. Statist. **32** (2005), 405–424. [MR2204627](#)
- [3] U. Grenander, *On the theory of mortality measurement. II*, Skand. Aktuarietidskr. **39** (1956), 125–153 (1957). [MR0093415 \(19,1243c\)](#)
- [4] P. Groeneboom and G. Jongbloed, *Smooth and non-smooth estimates of a monotone hazard*, From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner, Institute of Mathematical Statistics, 2013, pp. 174–196.
- [5] P. Groeneboom, G. Jongbloed, and B.I. Witte, *Maximum smoothed likelihood estimation and smoothed maximum likelihood estimation in the current status model*, Ann. Statist. **38** (2010), 352–387.
- [6] P. Groeneboom, M.H. Maathuis, and J.A. Wellner, *Current status data with competing risks: consistency and rates of convergence of the MLE*, Ann. Statist. **36** (2008), 1031–1063. [MR2418648 \(2009h:62039\)](#)
- [7] P. Groeneboom and J.A. Wellner, *Information bounds and nonparametric maximum likelihood estimation*, DMV Seminar, vol. 19, Birkhäuser Verlag, Basel, 1992. [MR1180321 \(94k:62056\)](#)
- [8] Piet Groeneboom and Geurt Jongbloed, *Nonparametric estimation under shape constraints*, Cambridge Univ. Press, Cambridge, 2014.
- [9] P. Hall, *Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density*, Ann. Statist. **20** (1992), 675–694. [MR1165587 \(93e:62131\)](#)
- [10] P. Hall and J.L. Horowitz, *A simple bootstrap method for constructing nonparametric confidence bands for functions*, Ann. Statist. **41** (2013), 1892–1921. [MR3127852](#)
- [11] N. Keiding, O.K. Højbjerg Hansen, D.N. Sørensen, and R. Slama, *The current duration approach to estimating time to pregnancy*, Scand. J. Stat. **39** (2012), 185–204. [MR2927018](#)
- [12] J.K. Kim and D. Pollard, *Cube root asymptotics*, Ann. Statist. **18** (1990), 191–219. [MR1041391 \(91f:62059\)](#)
- [13] M.R. Kosorok, *Bootstrapping the Grenander estimator*, Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen, Inst. Math. Stat. Collect., vol. 1, Inst. Math. Statist., Beachwood, OH, 2008, pp. 282–292.
- [14] T. Robertson, F.T. Wright, and R.L. Dykstra, *Order restricted statistical inference*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester, 1988. [MR961262 \(90b:62001\)](#)
- [15] B. Sen, M. Banerjee, and M.B. Woodroffe, *Inconsistency of bootstrap: the Grenander estimator*, Ann. Statist. **38** (2010), 1953–1977. [MR2676880 \(2011f:62046\)](#)
- [16] R. Slama, O.K. Højbjerg Hansen, B. Ducot, A. Bohet, D. Sørensen, L. Allemand, M.J. Eijkemans, L. Rosetta, J.C. Thalabard, N. Keiding, et al., *Estimation of the frequency of involuntary infertility on a nation-wide basis*, Human reproduction **27** (2012), 1489–1498.
- [17] Y. Vardi, *Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation*, Biometrika **76** (1989), 751–761. [MR1041420 \(91i:62057\)](#)
- [18] G.S. Watson, *Estimating functionals of particle size distributions*, Biometrika **58** (1971), 483–490. [MR0312639 \(47 #1195\)](#)
- [19] M.B. Woodroffe and J. Sun, *A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing*, Statist. Sinica **3** (1993), 501–515. [MR1243398 \(94i:62044\)](#)