NONPARAMETRIC LEAST SQUARES ESTIMATORS FOR INTERVAL CENSORING

By Piet Groeneboom[a]

Delft Institute of Applied Mathematics, Mekelweg 4, 2628 CD Delft, The Netherlands.
[a]*P.Groeneboom@tudelft.nl*

The limit distribution of the nonparametric maximum likelihood estimator for interval censored data with more than one observation time per unobservable observation, is still unknown in general. For the so-called separated case, where one has observation times which are at a distance larger than a fixed $\epsilon > 0$, the limit distribution was derived in [4]. For the non-separated case there is a conjectured limit distribution, given in [9], Section 5.2 of Part 2. But the findings of the present paper suggest that this conjecture may not hold.

We prove consistency of a related nonparametric isotonic least squares estimator and give a sketch of the proof for a result on its limit distribution. We also provide simulation results to show how the nonparametric MLE and least squares estimator behave in comparison. Moreover, we discuss a simpler least squares estimator that can be computed in one step, but is inferior to the other least squares estimator, since it does not use all information.

For the simplest model of interval censoring, the current status model, the nonparametric maximum likelihood and least squares estimators are the same. This equivalence breaks down if there are more observation times per unobservable observation. The computations for the simulation of the more complicated interval censoring model were performed by using the iterative convex minorant algorithm. They are provided in the GitHub repository [6].

1. Introduction. The simplest and most studied interval censoring model is the so-called interval censoring, case 1, or current status model. This model can be defined in the following way (see, e.g., Section 2.3 of [7]).

Consider a sample $X_1, X_2, \ldots, X_n$, drawn from a distribution with distribution function F_0. Instead of observing the X_i's, one only observes for each i whether or not $X_i \le T_i$ for some random T_i (independent of the other T_j's and all X_j's). More formally, instead of observing the X_i's, one observes

$$(T_i, \Delta_i) = (T_i, 1_{\{X_i \le T_i\}}). \tag{1.1}$$

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

We want to estimate the unknown distribution function F_0 based on the data given in (1.1). For this problem the log likelihood function for distribution functions F (conditional on the T_i's) is given by

$$\ell(F) = \sum_{i=1}^{n} \left(\Delta_i \log F(T_i) + (1 - \Delta_i) \log(1 - F(T_i))\right). \tag{1.2}$$

The (nonparametric) maximum likelihood estimator $\hat{F}_n$ maximizes ℓ over the class of *all* distribution functions.

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

Keywords and phrases: interval censoring, nonparametric least squares estimators, nonparametric maximum likelihood, asymptotic distribution theory, integral equations, smooth functional theory, iterative convex minorant algorithm.

Since distribution functions are by definition nondecreasing, computing the maximum likelihood estimator poses a shape constrained optimization problem. As can be seen from (1.2), the value of ℓ only depends on the values that F takes at the observed time points T_i. Hence one can choose to consider only distribution functions that are constant between successive observed time points T_i. The lemma below shows that this estimator can be characterized in terms of a greatest convex minorant of a certain diagram of points. The following result is Lemma 2.7 in [7] (and Proposition 1.2 in [9]).

LEMMA 1. *[Characterization of the nonparametric ML estimator in the current status model] Consider the cumulative sum diagram consisting of the points* $P_0 = (0,0)$ *and*

$$P_i = \left(i, \sum_{j=1}^{i} \Delta_j\right), \;\; 1 \le i \le n,$$

where the Δ_i*'s correspond to the* T_i*'s, which are supposed to be ordered* $0 < T_1 \cdots < T_n$ *(one can also allow ties, but we disregard this further complication here). Then the nonparametric MLE* $\hat F_n(T_i)$ *is given by the left derivative of the convex minorant of this diagram of points, evaluated at the point* i*. This maximizer is unique.*

REMARK 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 distribution function $\hat F_n$) on $[T_i, T_{i+1})$, a region to the *right* of T_i.

Lemma 1 shows that $\hat F_n$ is in fact the *isotonic regression* on the indicators $\Delta_i = 1_{\{X_i T_i\}}$, that is, it minimizes

$$\sum_{i=1}^{n} \{F(T_i) - \Delta_i\}^2 \tag{1.3}$$

over all monotone nondecreasing (not necessarily bounded by 0 and 1) functions F, see p. 43 in section 1.1 of part 2 of [9]. It also follows from Theorem 1.2.1 on p. 7 of [10], where the connection brtween the derivative of the greatest convex minorant and the solution of the isotonic regression problem is given.

So we have the two characterizations of the nonparametric maximum likelihood estimator, the characterization as a maximizer of (1.2) and the characterization as a minimizer of (1.3) (see, e.g., [9]). Note that the weights are equal and constant in the least sauares problem.

If we turn to the (common) situation where there are more observation times per unobservable X_i, the situation is considerably more complicated, and the limit distribution of the nonparametric MLE is still unknown generally. We consider here the simplest extension, where one has two observation times per unobservable X_i (*interval censoring, case 2*). Instead of observing the X_i's, one observes

$$(U_i, V_i, \Delta_{i0}, \Delta_{i1}) \stackrel{\text{def}}{=} (U_i, V_i, 1_{\{X_i \le U_i\}}, 1_{\{U_i < X_i \le V_i\}}), \tag{1.4}$$

so instead of only observing that X_i is larger than or smaller than an observation T_i, we now have an observation interval (U_i, V_i), $U_i < V_i$, and we know whether our unobservable X_i is inside the interval or to the left or right of it. Just as in the current status model, we assume that X_i is distributed independently of (U_i, V_i). The log likelihood (1.2) changes into

$$\ell(F) = \sum_{i=1}^{n} \left\{\Delta_{i0} \log F(U_i) + \Delta_{i1} \log(F(V_i) - F(U_i)) + \Delta_{i2} \log(1 - F(V_i))\right\}, \tag{1.5}$$

where $\Delta_{i2} = 1 - \Delta_{i0} - \Delta_{i1}$. It is not clear that there is an equivalent to the minimization of (1.3) in this situation. We cannot use Theorem 1.5.1 in [10] because we have the difference of $F(V_i)$ and $F(U_i)$ in the terms $\log(F(V_i) - F(U_i))$ instead of just $F(U_i)$ or $F(V_i)$ by itself only.

The perhaps most natural least squares approach is to consider minimization of

$$\sum_{i=1}^{n} \left\{ \{F(U_i) - \Delta_{i0}\}^2 + \{F(V_i) - F(U_i) - \Delta_{i1}\}^2 + \{1 - F(V_i) - \Delta_{i2}\}^2 \right\} \tag{1.6}$$

over all distribution functions F, so an isotonic regression on the indicators $\Delta_{i0}, \Delta_{i1}, \Delta_{i2}$, $i = 1, \ldots, n$.

As an example, we analyze the behavior of the nonparametric MLE, maximizing (1.5) and the least squares estimator minimizing (1.6) in Example 1.

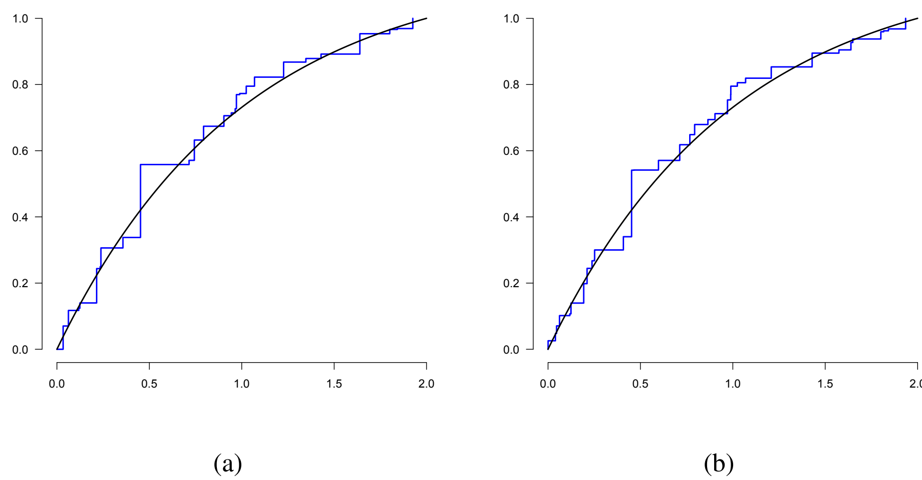


Fig 1: (a) Nonparametric MLE (blue) of F_0 for a sample of size $n = 1000$, (b) Nonparametric least squares estimate (blue) minimizing (1.6) of F_0 for the same sample. The solid black curve shows F_0.

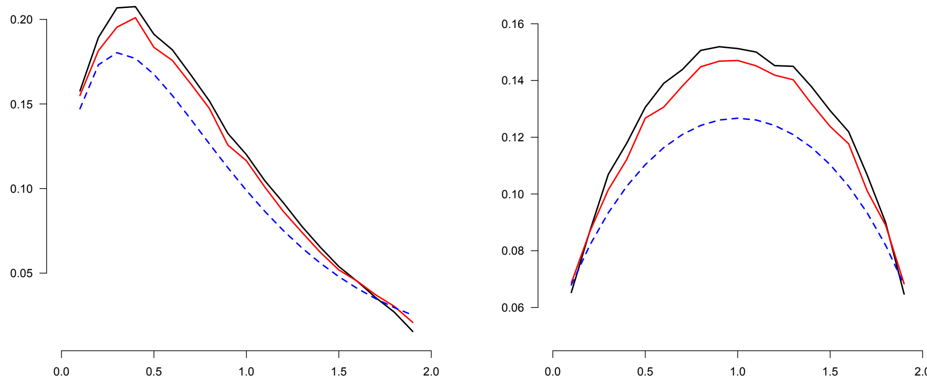
EXAMPLE 1. Suppose $X_1, \ldots, X_n$ is either a sample from the truncated exponential distribution on $[0, 2]$, with density

$$f_0(x) = \frac{\exp\{-x\}}{1 - \exp\{-2\}} 1_{[0,2]}(x), \qquad x \in \mathbb{R},$$

or a sample from the Uniform distribution on $[0, 2]$, and let the (U_i, V_i), $V_i > U_i$, be the order statistics of the Uniform distribution on $[0, 2]^2$ for a sample of size 2. Note that this is a prototype of the non-separated case, where we can have arbitrarily small observation intervals $[U_i, V_i]$

Figure 1 shows the nonparametric MLE and the nonparametric least squares estimate for a simulated sample of size $n = 1000$ for these models. In both cases the estimate has to be computed iteratively, we do not have a one step algorithm as in the current status model.

We computed the variances of the estimates times $n^{2/3}$ for $10,000$ samples of size $n = 1000$ for $t_i = i \cdot 0.1$, $i = 1, \ldots, 19$. A comparison of the simulated variances is shown



(a) (b)

Fig 2: (a) Simulated variances, times $n^{2/3}$, of the nonparametric MLE (black solid curve) and the least squares estimate (red) minimizing (1.6), for $t_i = 0.1, 0.2, \dots, 1.9$, linearly interpolated between values at the t_i for the model of Example 1. The blue dashed curve is the theoretical limit curve one obtains from Theorem 1 in Section 2 for the LS estimator. The simulated variances are based on $10,000$ simulations of samples of size $n = 1000$ for the truncated exponential distribution function F_0 on $[0, 2]$ and the order statistics of the uniform distribution on $[0, 2]^2$ as observation times. (b) The same comparison, but now for F_0 uniform on $[0, 2]$.

in Figure 2. It suggests that overall the least squares estimator is slightly better than the nonparametric MLE for this sample size in this model.

Inspired by the conjectured faster rate of convergence of the MLE in the non-separated case (i.e., observation intervals can be arbitrarily small) in Section 5.2 of Part 2 of [9], Lucien Birgé constructed in [2] a histogram estimator which actually achieves rate $(n \log n)^{1/3}$ locally (but suffers severely from bias, in contrast with the MLE). The asymptotic (normal) distribution of the histogram estimator was derived in [8]. At the time that [8] was written, it was still believed that the nonparametric MLE would also achieve this faster rate. The matter has still not been settled, but the present paper casts doubts on this conjecture.

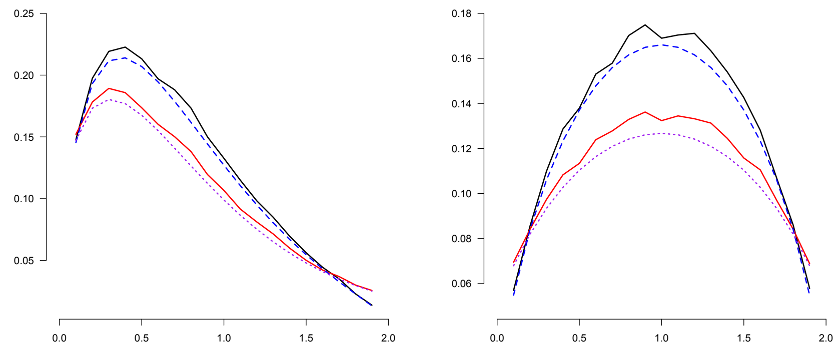
It is also possible to define a least squares estimator which can be computed in one step, using the convex minorant (or "pool adjacent violators") algorithm. An estimator of this type was proposed for the uniform deconvolution problem in [11]. This estimator minimizes

$$\sum_{i=1}^{n} \left\{ \{F(U_i) - \Delta_{i0}\}^2 + \{F(V_i) - \Delta_{i0} - \Delta_{i1}\}^2 \right\} \tag{1.7}$$

over all distribution functions F, so is also an isotonic regression on the indicators $\Delta_{i0}, \Delta_{i1}, \Delta_{i2}$, $i = 1, \dots, n$. Criterion (1.7) can also be written

$$\sum_{i=1}^{n} \left\{ \{F(U_i) - \Delta_{i0}\}^2 + \{1 - F(V_i) - \Delta_{i2}\}^2 \right\}. \tag{1.8}$$

We compare the two least squares estimators for sample size $n = 10,000$ in Figure 3, from which the superiority of the estimator, based on minimizing (1.6), seems clear.



(a) (b)

Fig 3: (a) Simulated variances, times $n^{2/3}$, of the simple nonparametric LS estimator, minimizing (1.7) (black solid curve) and the least squares estimator, minimizing (1.6) (red), for $t_i = 0.1, 0.2, \ldots, 1.9$, linearly interpolated between values at the t_i for the model of Example 1. The blue dashed curve and purple dotted curves are the theoretical limit curves discussed in Section 2 for the LS estimators, minimizing (1.7) and (1.6), respectively The simulated variances are based on $10,000$ simulations of samples of size $n = 10,000$ for the truncated exponential distribution function F_0 on $[0,2]$ and the order statistics of the uniform distribution on $[0,2]^2$ as observation times. (b) The same comparison, but now for F_0 uniform on $[0,2]$.

Our proof of the limit result Theorem 1 is not complete. We still have to show that the "off-diagonal terms" (5.8) and (5.9) are $o_p(n^{-2/3})$, see Section 5.2. In fact, terms of this type were shown to be $O_p(n^{-5/6})$ in a similar problem, analyzed in [5]. In Section 3 we discuss the theory, needed for the completion of the proof.

2. The least squares estimators. In the characterization of the nonparametric MLE one can use the fact that the logarithms provide a natural logarithmic boundary, preventing values of the solution to leave the interval $[0,1]$. This is no longer true for the least squares estimate minimizing (1.6) and for this reason we use Lagrange multipliers in its characterization.

Let, for a distribution function F, the process $W_{n,F}$ be defined by

$$\begin{aligned} &W_{n,F}(t) \\ &= \int_{u\le t} \{\delta_0 - F(u)\}\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1) - \int_{u\le t} \{\delta_1 - \{F(v)-F(u)\}\}\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1) \\ &\quad + \int_{v\le t} \{\delta_1 - \{F(v)-F(u)\}\}\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1) + \int_{v\le t} \{\delta_0+\delta_1 - F(v)\}\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1), \end{aligned} \tag{2.1}$$

where $\mathbb{Q}_n$ is the empirical probability measure of the $(U_i, V_i, \Delta_{i0}, \Delta_{i1})$.

The least squares estimate of F_0 has the following characterization.

LEMMA 2. *Let the process $W_{n,F}$ be defined by (2.1) and let the Lagrange multipliers $\lambda_{1,F}$ and $\lambda_{2,F}$ be defined by*

$$\lambda_{1,F} = -\int_{(u,v):F(u)=0 \text{ or } F(v)=0} dW_{n,F}(u,v,\delta_0,\delta_1), \tag{2.2}$$

and

$$\lambda_{2,F} = \int_{(u,v):F(u)=1 \text{ or } F(v)=1} dW_{n,F}(u,v,\delta_0,\delta_1). \tag{2.3}$$

Then the distribution function $\hat F_n$ minimizes (1.6) over all distribution functions F on $\mathbb{R}_+$ if and only if the following conditions are satisfied

(i)

$$\lambda_{1,\hat F_n} + W_{n,\hat F_n}(t) \ge 0, \qquad t \ge 0,$$

(ii)

$$\int \hat F_n(t)\, dW_{n,\hat F_n}(t) - \lambda_{2,\hat F_n} = 0.$$

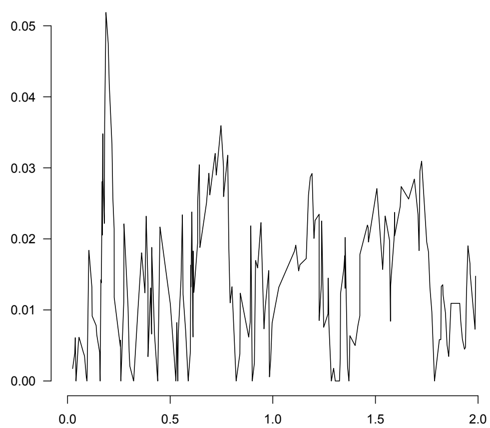


Fig 4: The process $W_{n,\hat F_n}$ as a function of the $2n$ ordered observations U_i and V_i for Example 1 and $n = 100$. For this example $\lambda_{1,\hat F_n} = 0.003148$ and $\lambda_{2,\hat F_n} = 0.014758$.

The proof is based on the so-called Fenchel duality condition, as also used in [7] in the characterization of the nonparametric MLE for interval censoring (apart from the Lagrange multipliers in the present case) and is therefore omitted. A picture of the process $W_{n,\hat F_n}$ for $n = 100$ is given in Figure 4 for Example 1 above. $W_{n,\hat F_n}$ touches zero at points just to the left of points where the least squares estimator $\hat F_n$ has a jump. The iterative convex minorant algorithm for computing $\hat F_n$ is based on this characterization.

The algorithm proceeds in the following way. At each step we solve the isotonic regression problem without the Lagrange multipliers, under the restriction $y_1 \le \cdots \le y_{2n}$, where $y_i = F(T_i)$ represents the value of the distribution function at the ith order statistic of set of U_i's and V_i's (ties can also be handled, but we disregard this further complication here). Typically, this will give values $y_i < 0$ and $y_i > 1$. If $y_i < 0$ we put $y_i = 0$ and if $y_i > 1$ we put $y_i = 1$. For this new value of the y_i we compute the (preliminary) Lagrange multipliers λ_1 and λ_2, using (2.2) and (2.3). We repeat this procedure until the conditions (i) and (ii) of Lemma 2 are satisfied up to an accuracy of say 10^{-8}. Convergence of this algorithm is very fast.

One can also compute the estimate by the interior point method, using a logarithmic barrier function. This algorithm is a kind of opposite of the iterative convex minnorant algorihm, since it converges to the solution from the interior of the parameter space, whereas the iterative convex minorant algorithm immediately hits the boundary in the first iteration step. Both algorithms were programmed for the present problem and give exactly the same result, though.

The least squares estimator minimizing (1.6) is consistent, as the following lemma shows.

LEMMA 3. *Let $\hat F_n$ be the isotonic least squares estimator, minimizing (1.6) over distribution functions F under the conditions of Theorem 1 below. Then $\hat F_n$ converges almost surely to F_0 in the supremum metric.*

PROOF. Let the function ψ be defined by

$$\psi(F) = \tfrac{1}{2}\int \big\{\{F(u)-\delta_0\}^2+\{F(v)-F(u)-\delta_1\}^2+\{F(v)-\delta_0-\delta_1\}^2\big\}\,d\mathbb{Q}_n(u,v,\delta_0,\delta_1). \tag{2.4}$$

Since $\hat F_n$ minimizes (2.4) over F, we must have

$$\lim_{\epsilon\downarrow 0}\epsilon^{-1}\Big[\psi\big((1-\epsilon)\hat F_n+\epsilon F_0\big)-\psi(\hat F_n)\Big]\ge 0.$$

Note that the limit exists by the convexity of the function ψ. This means

$$\begin{aligned}&\int\{F_0(u)-\hat F_n(u)\}\{\hat F_n(u)-\delta_0\}\,d\mathbb{Q}_n(u,v,\delta_0,\delta_1)\\&\quad+\int\{F_0(v)-F_0(u)-\hat F_n(v)+\hat F_n(u)\}\{\hat F_n(v)-\hat F_n(u)-\delta_1\}\,d\mathbb{Q}_n(u,v,\delta_0,\delta_1)\\&\quad+\int\{F_0(v)-\hat F_n(u)\}\{\hat F_n(v)-\delta_0-\delta_1\}\,d\mathbb{Q}_n(u,v,\delta_0,\delta_1)\\&\ge 0.\end{aligned}$$

Proceeding as in Section 4 of Part 2 of [9], using the Helly compactness theorem, we get from this, for a limit point F of a subsequence of $\hat F_n$:

$$\begin{aligned}&\int\{F_0(u)-F(u)\}\{F(u)-F_0(u)\}\,h(u,v)\,du\,dv\\&\quad+\int\{F_0(v)-F_0(u)-F_(v)-F(v)+F(u)\}\\&\qquad\qquad\cdot\{F(v)-F(u)-F_0(v)+F_0(u)\}\,h(u,v)\,du\,dv\\&\quad+\int\{F_0(v)-F(v)\}\{F(v)-F_0(v)\}\,h(u,v)\,du\,dv\\&=-\int\{F(u)-F_0(u)\}^2\,h(u,v)\,du\,dv-\int\{F(v)-F(u)-F_0(v)+F_0(u)\}^2\,h(u,v)\,du\,dv\\&\quad-\int\{F(v)-F_0(v)\}^2\,h(u,v)\,du\,dv\\&\ge 0.\end{aligned}$$

By the assumptions on F_0 and h this means that $F = F_0$. This implies that $\hat{F}_n$ converges uniformly to F_0. □

In order to state the limit result of the least squares estimator, minimizing (1.6), we need the following notation. Let, for $t_0 \in (0, M)$, a_{t_0} be defined as the positive square root of

$$
\begin{aligned}
a_{t_0}^2 &= F_0(t_0)\{1 - F_0(t_0)\}\{h_1(t_0) + h_2(t_0)\} \\
&\quad + \int_{v=t_0}^{M} \{F_0(v) - F_0(t_0)\}\left[1 - \{F_0(v) - F_0(t_0)\}\right] h(t_0, v)\, dv \\
&\quad + \int_{u=0}^{t_0} \{F_0(t_0) - F_0(u)\}\left[1 - \{F_0(t_0) - F_0(u)\}\right] h(u, t_0)\, du \\
&\quad + 2F_0(t_0) \int_{v=t_0}^{M} \{F_0(v) - F_0(t_0)\}\, h(t_0, v)\, dv \\
&\quad + 2\{1 - F_0(t_0)\} \int_{u=0}^{t_0} \{F_0(t_0) - F_0(u)\}\, h(u, t_0)\, du, \qquad (2.5)
\end{aligned}
$$

and let b_{t_0} be defined by:

$$b_{t_0} = h_1(t_0) + h_2(t_0), \qquad (2.6)$$

where h_1 and h_2 are the marginals of h. Then we have the following result of which the proof is sketched in the Appendix.

THEOREM 1. *Let F_0 have a continuous positive density f_0 on $[0, M]$ and let (U, V) be the order statistics of an absolutely continuous distribution. We assume that (U, V) has a a positive continuous density h on*

$$S = \{(u, v) : 0 \le u < v \le M\},$$

with first and second marginals h_1 and h_2, respectively, with bounded second partial derivatives.We assume that X_i is independent of (U_i, V_i).

Let a_{t_0} and b_{t_0} be defined by (2.5) and (2.6), respectively. Moreover let, for $t_0 \in (0, M)$, σ_{t_0} be defined by

$$\sigma_{t_0} = \left(a_{t_0} f_0(t_0)/b_{t_0}\right)^{1/3}, \qquad t_0 \in (0, M). \qquad (2.7)$$

Then we get at a fixed point $t_0 \in (0, M)$ for the LS estimate $\hat{F}_n$, minimizing

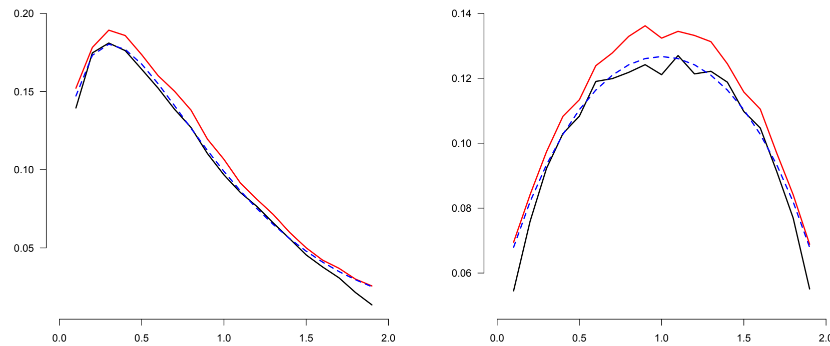
$$\sum_{i=1}^{n} \left\{ \{F(U_i) - \Delta_{i0}\}^2 + \{F(V_i) - F(U_i) - \Delta_{i1}\}^2 + \{F(V_i) - \Delta_{i0} - \Delta_{i1})\}^2 \right\} \qquad (2.8)$$

over all distribution functions F:

$$n^{1/3} \left\{ \hat{F}_n(t_0) - F_0(t_0) \right\} / \sigma_{t_0} \xrightarrow{\mathcal{D}} Z,$$

where Z is the argmin of $t \mapsto W(t) + t^2$, and W is standard two-sided Brownian motion.

REMARK 2. The conditions on the observation density h are similar to the conditions in [3] and used in the smooth functional theory needed in treating so-called “off-diagonal terms”, see Section 5.2.



(a) (b)

Fig 5: (a) Simulated variances, times $n^{2/3}$, of the nonparametric MLE (black solid curve) and the least squares estimate (red), minimizing (1.6), for $t_i = 0.1, 0.2, \ldots, 1.9$, linearly interpolated between values at the t_i for the model of Example 1. The blue dashed curve is the theoretical limit curve one obtains from Theorem 1 below. The simulated variances are based on $10,000$ simulations of samples of size $n = 10,000$ for the truncated exponential distribution function F_0 on $[0,2]$ and the order statistics of the uniform distribution on $[0,2]^2$ as observation times. (b) The same comparison, but now for F_0 uniform on $[0,2]$.

As an example, the asymptotic variance of the LS estimator is shown in Figure 5 as the blue dashed curve for in Example 1. Remarkably, the curve of the variances of the MLE is even closer to the theoretical asymptotic curve than the curve for the variances of the least squares estimator for this sampe size of $n = 10,000$. This should be compared with Figure 2, where the sample size was $n = 1000$ and the variances for the LS estimator were actually smaller that those of the nonparametric MLE. The faster decrease of the variances of the MLE than those of the LS when we increase the sample size from 1000 to $10,000$ might give a hint of the possibility of a rate of $(n \log n)^{1/3}$ again, but the form of the curve is certainly not in accordance with the conjecture in [9].

We here make some additional remarks about the proof of Theorem 1, discussed in the Appendix. Results of this type were proved for the nonparametric MLE in [4] and [5]. One has to show that the leading asymptotic behavior is provided by replacing the estimator by the underlying distribution function F_0 in the characterization, given in Lemma 2. To this end, one has to show that certain terms, called the "off-diagonal terms", are of lower order. Doing this is rather hard in our experience ([5] needed 53 pages), but presently there seems to be no other way. We give a sketch of these calculations in Section 3 and in the Appendix. For similar calculations, see, e.g., Chapter 10 of [7] and [5]. The first result of the type of Theorem 1 was for the current status model in Theorem 5.5 of Part 2 of [9].

For the least squares estimator, minimizing (1.7) we can define the following process.

$$W_{n,F}^{(2)}(t)$$

(2.9)

$$= \int_{u \le t} \{\delta_0 - F(u)\}\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1) + \int_{v \le t} \{\delta_0 + \delta_1 - F(v)\}\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1),\ t \ge 0,$$

where $\mathbb{Q}_n$ is the empirical probability measre of the $(U_i, V_i, \Delta_{i0}, \Delta_{i1})$ and F a distribution function. We now get the following characterization the least squares estimator, minimizing (1.7).

LEMMA 4. *Let the process $W_{n,F}^{(2)}$ be defined by (2.9). Then the distribution function $\hat{F}_n$ minimizes (1.7) over all distribution functions F on $\mathbb{R}_+$ if and only if the following conditions are satisfied*

(i)

$$W_{n,\hat{F}_n}^{(2)}(t) \geq 0, \qquad t \geq 0,$$

(ii)

$$\int \hat{F}_n(t)\, dW_{n,\hat{F}_n}^{(2)}(t) = 0.$$

The proof follows familiar lines and is therefore omitted. Alternatively, one can say that the esitimator is the left-continuous slope of the greatest convex minorant of the cusum diagram starting at $(0,0)$ and running through the $(2n)$ points

$$\left(\sum 1_{\{U_i \leq t\}} + \sum 1_{\{V_i \leq t\}}, \sum \Delta_{i0} 1_{\{U_i \leq t\}} + \sum (\Delta_{i0} + \Delta_{i1}) 1_{\{V_i \leq t\}}\right), \qquad t \geq 0.$$

Since we work with observations from absolutely continuous distributions, we get for the left coordinates just $1, 2, \ldots$.

So, in constructing the cusum diagram, we run through all $2n$ observation points; if we meet a point U_i, we add δ_{i0} to the cusum diagram, if we meet a point V_i we add $\delta_{i0} + \delta_{i1}$ to the diagram. In the latter case we only record whether the unobservable variable lies to the left of V_i (then $\delta_{i0} + \delta_{i1} = 1$) or to the right of V_i (then $\delta_{i0} + \delta_{i1} = 0$).

For the isotonic least squares estimator, minimizing (1.7) we have the following result.

THEOREM 2. *Let the assumptions of Theorem 1 be satisfied and let a'_{t_0} and b'_{t_0} be defined by*

$$(a'_{t_0})^2 = F_0(t_0)\{1 - F_0(t_0)\}\{h_1(t_0) + h_2(t_0)\} \tag{2.10}$$

and

$$b'_{t_0} = \{h_1(t_0) + h_2(t_0)\}/4, \tag{2.11}$$

where h_1 and h_2 are the marginals of h. Moreover let, for $t_0 \in (0, M)$, σ_{t_0} be defined by

$$\sigma'_{t_0} = \left(a'_{t_0} f_0(t_0)/b'_{t_0}\right)^{1/3}, \qquad t_0 \in (0, M). \tag{2.12}$$

Then we get at a fixed point $t_0 \in (0, M)$ for the LS estimate $\hat{F}_n$, minimizing

$$\sum_{i=1}^{n} \left\{\{F(U_i) - \Delta_{i0}\}^2 + \{F(V_i) - \Delta_{i0} - \Delta_{i1})\}^2\right\} \tag{2.13}$$

over all distribution functions F:

$$n^{1/3}\left\{\hat{F}_n(t_0) - F_0(t_0)\right\}/\sigma'_{t_0} \xrightarrow{\mathcal{D}} Z,$$

where Z is the argmin of $t \mapsto W(t) + t^2$, and W is standard two-sided Brownian motion.

Both the computation and the proof of Theorem 2 are easier than for the estimator minimizing (2.8), but the performance of the estimator, minimizing (1.7) is clearly inferior to that of the estimator, minimizing (1.6), see Figure 3.

3. Smooth functional theory. The smooth functional theory for the nonparametric MLE in the separated case is discussed in [7] and in [4] and for the non-separated case of interval censoring in [3]. We give the treatment here for the LS estimator, minimizing (1.6), but the methods could also be used to simplify the theory given in [7] and [3] for the MLE.

We define, for the LS estimate $\hat F_n$,

$$\theta_{\hat F_n}(u,v,\delta_0,\delta_1) = \{\delta_0 - \hat F_n(u)\}\phi_{\hat F_n}(u) + \{\delta_1 - \hat F_n(v) + \hat F_n(u)\}\{\phi_{\hat F_n}(v) - \phi_{\hat F_n}(u)\} - \{\delta_2 - (1 - \hat F_n(v)\}\phi_{\hat F_n}(v). \tag{3.1}$$

where $\phi_{\hat F_n}$ is either an integrated score function

$$\phi_{\hat F_n}(x) = \int_{y\in[0,x]} a(y)\, d\hat F_n(y), \tag{3.2}$$

or a function belonging to the closure (in an L_2-sense) of the space of such functions (the latter will actually be the case for the functions $\phi_{\hat F_n}$, found below for the LS estimator)

Note that, if $\phi_{\hat F_n}$ is a right-continuous piecewise constant function, constant on the same intervals as $\hat F_n$, and satisfying $\phi_{\hat F_n}(t) = 0$ if $\hat F_n(t) = 0$ or $\hat F_n(t) = 1$, we have:

$$\begin{aligned}
&\int \theta_{\hat F_n}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n u,v,\delta_0,\delta_1)\\
&= \int \Big[\delta_0 - \hat F_n(u) - \{\delta_1 - \hat F_n(v) + \hat F_n(u)\}\Big]\, \phi_{\hat F_n}(u)\, d\mathbb{Q}_n(u,v,\delta_0,\delta_1)\\
&\qquad + \int \Big[\delta_1 - \hat F_n(v) + \hat F_n(u) - \delta_2 + \{1 - \hat F_n(v)\}\Big]\, \phi_{\hat F_n}(v)\, d\mathbb{Q}_n u,v,\delta_0,\delta_1)\\
&= 0.
\end{aligned} \tag{3.3}$$

We have, using Fubini's theorem,

$$\begin{aligned}
&\int \theta_{\hat F_n}\, dQ_0\\
&= \int \{F_0(u) - \hat F_n(u)\}\phi_{\hat F_n}(u)\, h_1(u)\, du + \int \{F_0(v) - \hat F_n(v)\}\phi_{\hat F_n}(v)\, h_2(v)\, dv\\
&\qquad + \int \{F_0(v) - F_0(u) - \hat F_n(v) + \hat F_n(u)\}\{\phi_{\hat F_n}(v) - \phi_{\hat F_n}(u)\}\, h(u,v)\, du\, dv\\
&= \int_{x\in[0,M]} \Big[\int_{x\le u} \phi_{\hat F_n}(u)\, \{h_1(u) + h_2(u)\}\, du\\
&\qquad + \int_{u<x\le v} \{\phi_{\hat F_n}(v) - \phi_{\hat F_n}(u)\} h(u,v)\, du\, dv\Big]\, d(\hat F_0 - \hat F_n)(x).
\end{aligned}$$

If we want to estimate a smooth functional of the type $\int \kappa_{F_0}(x)\, dF_0(x)$, we approximately want to have for the inner part between brackets of this integral:

$$\int_{u\ge x} \phi_{\hat F_n}(u)\, \{h_1(u) + h_2(u)\}\, du + \int_{u<x\le v} \{\phi_{\hat F_n}(v) - \phi_{\hat F_n}(u)\}\, h(u,v)\, du\, dv = \kappa_{F_0}(x), \tag{3.4}$$

because we then get

$$\int \kappa_{F_0}\, d\big(\hat F_n - F_0\big) \approx \int \theta_{\hat F_n}\, dQ_0 = \int \theta_{\hat F_n}\, d\big(Q_0 - \mathbb{Q}_n\big),$$

that is, we can represent the functional in the hidden space by an integral in the observation space which is $\sqrt{n}$ convergent and asymptotically normal.

Differentiating equation (3.4) w.r.t. x, we get

$$
\begin{aligned}
&-\phi_{\hat F_n}(x)\{h_1(x)+h_2(x)\}-\int_{u\le x}\bigl\{\phi_{\hat F_n}(x)-\phi_{\hat F_n}(u)\bigr\}h(u,x)\,du\\
&\qquad\qquad\qquad\qquad\qquad\qquad+\int_{v\ge x}\bigl\{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(x)\bigr\}h(x,v)\,dv\\
&=\kappa'_{F_0}(x).
\end{aligned}
$$

Defining, as in [3], $\phi_{\hat F_n}$ by

$$
\phi_{\hat F_n}(x)=\int_{y\in(x,M]}a(y)\,d\hat F_n(y)
$$

instead of $\phi_{\hat F_n}(x)=\int_{y\in[0,x]}a(y)\,d\hat F_n(y)$, and also allowing functions $\phi_{\hat F_n}$ on the boundary of the space, we get the equation

$$
\begin{aligned}
&\phi_{\hat F_n}(x)\{h_1(x)+h_2(x)\}+\int_{u\le x}\bigl\{\phi_{\hat F_n}(x)-\phi_{\hat F_n}(u)\bigr\}h(u,x)\,du\\
&\qquad\qquad\qquad\qquad\qquad\qquad-\int_{v\ge x}\bigl\{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(x)\bigr\}h(x,v)\,dv\\
&=\kappa'_{F_0}(x).
\end{aligned}
$$

This leads to the matrix equation

$$
\begin{aligned}
&y_i\left\{\Delta_i(h_1)+\Delta_i(h_2)+\sum_{j<i}\Delta_{ji}(h)+\sum_{j>i}\Delta_{ij}(h)\right\}\\
&=\kappa_{F_0}(t_{i+1})-\kappa_{F_0}(t_i)+\sum_{j<i}\Delta_{ji}(h)y_j+\sum_{j>i}\Delta_{ij}(h)y_j,\qquad i=1,\dots,m-1,
\end{aligned}
\tag{3.5}
$$

where the t_i are the points of jump of $\hat F_n$ and $y_i=\phi_{\hat F_n}(t_i)$; $\phi_{\hat F_n}(t)=0$ is zero if $\hat F_n(t)=0$ or $\hat F_n(t)=1$. This is analogous to (iii) of Theorem 3.1 on p. 646 of [3]. Here $\Delta_i(h_j)$ and $\Delta_{ij}(h)$ are defined by

$$
\Delta_i(h_j)=\int_{t_i}^{t_{i+1}}h_j(t)\,dt, i=1,\dots,m-1,\qquad j=1,2,
\tag{3.6}
$$

and

$$
\Delta_{ij}(h)=\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}}h(u,v)\,du\,dv,\qquad 1\le i<j\le m-1.
\tag{3.7}
$$

A prototype of a smooth functional is the functional for the mean

$$
T_F: x\mapsto x-\int y\,dF(y).
$$

For this case we get the equation

$$
-\phi_{\hat F_n}(x)\{h_1(x)+h_2(x)\}-\int_{u\le x}\bigl\{\phi_{\hat F_n}(x)-\phi_{\hat F_n}(u)\bigr\}h(u,x)\,du
$$

$$+\int_{v\ge x}\bigl\{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(x)\bigr\}h(x,v)\,dv$$
$$=1.$$

As an example, for the situation that F_0 is the uniform distribution function on $[0,1]$ and (U,V) is uniformly distributed on the upper triangle of the unit square with vertices $(0,0)$, $(0,1)$ and $(1,1)$ and, moreover, κ_{F_0} is the mean functional, this gives the solution $\phi_{\hat F_n}\equiv 1/2$. Note that this $\phi_{\hat F_n}$ does not have an interpretation as an integrated score function.

This yields as asymptotic variance for $m(\hat F_n)=\int x\,d\hat F_n(x)$, if $\hat F_n$ is the LS estimator, minimizing (1.6):

$$n\,\mathrm{var}\Bigl(m(\hat F_n)\Bigr)\longrightarrow\int\theta_{F_0}(u,v,\delta_0,\delta_1)^2\,d\mathbb{Q}_0(u,v,\delta_0,\delta_1),\qquad n\to\infty,$$

where

$$\begin{aligned}
&\int\theta_{F_0}(u,v,\delta_0,\delta_1)^2\,d\mathbb{Q}_0(u,v,\delta_0,\delta_1)\\
&=\tfrac14\int\left[\delta_0-F_0(u)-\delta_2+(1-F_0(v))\right]^2\,dQ_0(u,v,\delta_0,\delta_1)\\
&=\tfrac14\int F_0(u)(1-F_0(u)\}h_1(u)\,du+\tfrac14\int_0^1F_0(v)(1-F_0(v)\}h_2(v)\,dv\\
&\qquad\qquad+\tfrac12\int F_0(u)\{1-F_0(v)\}h(u,v)\,du\,dv\\
&=\tfrac12\int_0^1u(1-u)(1-u)\,du+\tfrac12\int_0^1v(1-v)v\,dv+\int_{0<u<v<1}u(1-v)\,du\,dv\\
&=\frac18=0.125.
\end{aligned}$$

So in this case we get the following result for the LS estimator:

$$\sqrt{n}\Bigl\{m(\hat F_n)-m(F_0)\Bigr\}\stackrel{\mathcal{D}}{\longrightarrow}N(0,\sigma^2),$$

where

$$\sigma^2=\frac18.$$

Other smooth functionals can be treated in a similar way, also giving the $\sqrt{n}$ behavior and asymptotic normality of these functionals of the LS estimator, which means that we can follow the analysis in [5] to show that the "off-diagonal" terms in Section 5.2

$$\int_{u\in[t_0,t_0+n^{-1/3}t],\,v\ge u}\Bigl\{F_0(v)-\hat F_n(v)\Bigr\}\,dH(u,v).$$

and

$$\int_{v\in[t_0,t_0+n^{-1/3}t],\,u\le v}\Bigl\{F_0(u)-\hat F_n(u)\Bigr\}\,dH(u,v).$$

are of order $O_p(n^{-5/6})$.

On the other hand, the analysis of the behavior of the smooth functionals for the nonparametric MLE starts by defining

$$\tilde\theta_{\hat F_n}(u,v,\delta_0,\delta_1)=\delta_0\frac{\phi_{\hat F_n}(u)}{\hat F_n(u)}+\delta_1\frac{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(u)}{\hat F_n(v)-\hat F_n(u)}-\delta_2\frac{\phi_{\hat F_n}(v)}{(1-\hat F_n(v))}\,,\tag{3.8}$$

where $\phi_{\hat F_n}$ is defined by (3.2), but now for the nonparametric MLE $\hat F_n$. By Proposition 3.1 on p. 644 of [3] we have:

$$\int \tilde\theta_{\hat F_n}\,d\mathbb{Q}_n=0, \tag{3.9}$$

(compare with (3.3)). Moreover, using Fubini's theorem again, we get

$$\begin{aligned}
&\int \tilde\theta_{\hat F_n}\,dQ_0\\
&=\int F_0(u)\frac{\phi_{\hat F_n}(u)}{\hat F_n(u)}\,h_1(u)\,du-\int\{1-F_0(v)\}\frac{\phi_{\hat F_n}(v)}{1-\hat F_n(v)}\,h_2(v)\,dv\\
&\qquad\qquad+\int\{F_0(v)-F_0(u)\}\frac{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(u)}{\hat F_n(v)-\hat F_n(u)}\,dH(u,v)\\
&=\int\{F_0(u)-\hat F_n(u)\}\frac{\phi_{\hat F_n}(u)}{\hat F_n(u)}\,h_1(u)\,du+\int\{F_0(v)-\hat F_n(v)\}\frac{\phi_{\hat F_n}(v)}{1-\hat F_n(v)}\,h_2(v)\,dv\\
&\quad+\int\{F_0(v)-F_0(u)-\hat F_n(v)+\hat F_n(u)\}\frac{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(u)}{\hat F_n(v)-\hat F_n(u)}\,h(u,v)\,du\,dv\\
&=\int\left[\int_{x\le u}\frac{\phi_{\hat F_n}(u)}{\hat F_n(u)}\,h_1(u)\,du+\int_{x\le v}\frac{\phi_{\hat F_n}(v)}{1-\hat F_n(v)}\,h_2(v)\,dv\right.\\
&\qquad\qquad\left.+\int_{u<x\le v}\frac{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(u)}{\hat F_n(v)-\hat F_n(u)}\,h(u,v)\,du\,dv\right]\,d\big(F_0-\hat F_n\big)(x)
\end{aligned}$$

So this time we want to solve (approximately) the equation

$$\begin{aligned}
&\int_{x\le u}\left\{\frac{\phi_{\hat F_n}(u)}{\hat F_n(u)}\,h_1(u)+\frac{\phi_{\hat F_n}(u)}{1-\hat F_n(u)}\,h_2(u)\right\}\,du+\int_{u<x\le v}\frac{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(u)}{\hat F_n(v)-\hat F_n(u)}\,h(u,v)\,du\,dv\\
&=\kappa_{F_0}(x).
\end{aligned}$$

Differentiating w.r.t. x, we get the equation

$$\begin{aligned}
&-\frac{\phi_{\hat F_n}(x)}{\hat F_n(x)}h_1(x)-\frac{\phi_{\hat F_n}(x)}{1-\hat F_n(x)}h_2(x)-\int\frac{\phi_{\hat F_n}(x)-\phi_{\hat F_n}(u)}{\hat F_n(x)-\hat F_n(u)}\,h(u,x)\,du\\
&\qquad\qquad+\int\frac{\phi_{\hat F_n}(v)-\phi_{\hat F_n}(x)}{\hat F_n(v)-\hat F_n(x)}\,h(x,v)\,dv\\
&=\kappa'_{F_0}(x).
\end{aligned}$$

As before, we can switch again to the representation $\phi(x)=\int_{y\in(x,M]}a(y)\,d\hat F_n(y)$ of integrated score functions a and use $\int_{y\in(x,M]}a(y)\,d\hat F_n(y)=-\int_{y\in[0,x]}a(y)\,d\hat F_n(y)$. This yields the preceding equation with minus and plus signs reversed on the left side. Now we get instead of the equations (3.5) the equations:

$$\begin{aligned}
&y_i\left\{\frac{\Delta_i(h_1)}{\hat F_n(t_i)}+\frac{\Delta_i(h_2)}{(1-\hat F_n(t_i))}+\sum_{j<i}\frac{\Delta_{ji}(h)}{\hat F_(t_i)-\hat F_n(t_j)}+\sum_{j>i}\frac{\Delta_{ij}(h)}{\hat F_n(t_j)-\hat F_n(t_i)}\right\}\\
&=\kappa_{F_0}(t_{i+1})-\kappa_{F_0}(t_i)+\sum_{j<i}\frac{\Delta_{ji}(h)\,y_j}{\hat F_(t_i)-\hat F_n(t_j)}+\frac{\Delta_{ij}(h)\,y_j}{\hat F_n(t_j)-\hat F_n(t_i)},\qquad i=1,\dots,m-1,
\end{aligned}\tag{3.10}$$

As noted in [3], p. 669, this is an equation of type $A\boldsymbol{y} = \boldsymbol{b}$, where A is a symmetrix, strictly diagonally dominant M-matrix (also called a Stieltjes matrix), which means that the equation has a unique solution, see [1]. This implies that the right-continuous piecewise constant function $\phi_{\hat{F}_n}$, constant on the same intervals as $\hat{F}_n$, is uniquely determined. Of course the same considerations hold for the equations (3.5).

In contrast with the situation for the LS estimator, the function $\phi_{\hat{F}_n}$ now has an interpretation as an integrated score function, but we do not have an explicit expression for it (see below). For the example, discussed above for the LS estimator, we computed $\phi_{\hat{F}_n}$ for a sample of size $n = 1000$, using the matrix equation (3.10), where this time $\hat{F}_n$ is the nonparametric MLE.

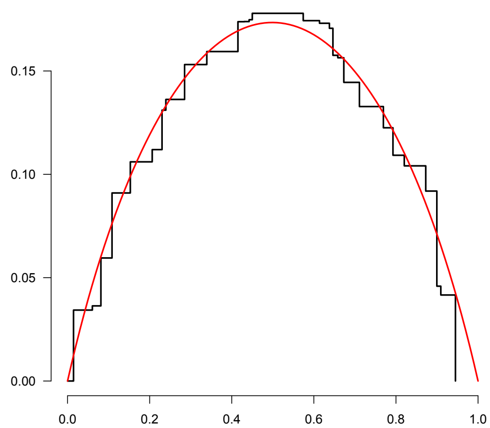


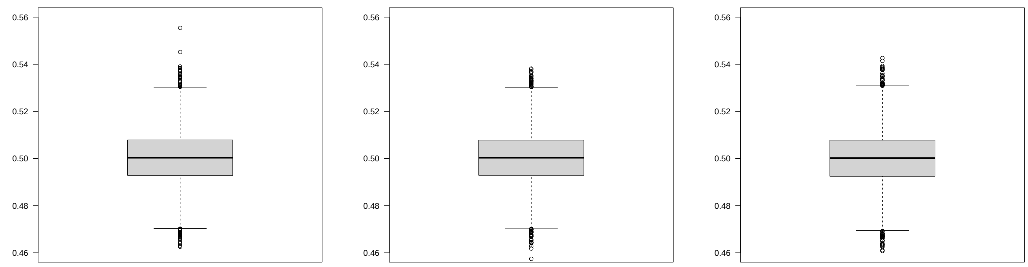
Fig 6: The function $\phi_{\hat{F}_n}$, solving equation (3.10) for the nonparametric MLE, for a sample of size $n = 1000$, where F_0 is the uniform distribution function on $[0, 1]$ and (U, V) is uniformly distributed on the upper triangle of the unit square. The red curve gives the solution of the corresponding integral equation for the underlying model.

We simulated the mean for the nonparametric MLE and the least squares estimators, minimizing (1.6) and (1.7), respectively, using the representation

$$m(F) = \int_0^1 (1 - F(x))\, dx = \sum_{i=0}^{m} (1 - F(t_i))\big(t_{i+1} - t_i\big),$$

where $t_1 < \cdots < t_m$ are the points of jump of F, and $t_0 = 0$ and $t_{m+1} = 1$. For this situation the information lower bound of the asymptotic variance was numerically computed to be approximately 0.1198987, see Section 4.2 of [3]. There is no explicit expression for it since it depends on the solution of an integral equation, for which there is no explicit form either. Based on $10,000$ simulations for sample size 1000, we got as estimates of the asymptotic values of $n\text{var}(m(\hat{F}_n))$ the value 0.1237086 for the nonparametric MLE, and the values 0.1240233 and 0.1271471 for the least squares estimators, minimizing (1.6) and (1.7), respectively.

All estimates give values reasonably close to the information lower bound and the least squares estimates also give $\sqrt{n}$-consistent estimates of the smooth functionals. The boxplots for the three estimates are shown in Figure 7, based on $10,000$ samples of size $n = 1000$.



(a) (b) (c)

Fig 7: (a) Mean estimates of MLE (b) Mean estimates of LS estimates, based on (1.6) (c) Mean estimates of LS estimates, based on (1.7). All boxplots are based on $10,000$ samples of size $n = 1000$.

It is clear that the differences of the three estimators are rather small in estimating the first moment. It should be noted, though, that the finer distinctions in the pointwise behavior of the estimators are "washed away" in the estimation of the smooth functionals.

4. Conclusion. We stated a limit result for the least squares estimator minimizing (1.6) for interval censoring, case 2, in the "non-separated case" ("observation intervals can be arbitrariy small"), in which situation the limit behavior of the nonparametric MLE is still unknown. This is Theorem 1 in section 2. For the separated case the limit distribution of the nonparametric MLE was derived in [4]. One could say (as has been done) that the separated case is the more important case ("there will always be a positive time interval between examination times of the doctor"), but the situation where one does not have the limit for the non-separated case is rather unsatisfactory. After all, allowing arbitrarily small intervals is mostly a matter of type of scaling one allows.

There is a conjecture that the nonparametric MLE has a faster rate of convergence in section 5.2 of part 2 of [9] and there also has been the construction of a histogram-type estimator with a faster local rate in [2], but the result for the least squares estimator and simulations suggest that this conjecture may not hold for the MLE. See, in particular, Figure 5 for sample size $n = 10,000$ in the present paper. But it is conceivable that the $(n \log n)^{1/3}$-rate starts showing up for astronomical sample sizes. It is true that for sample size $n = 10,000$ its variance has become smaller than that of the LS estimate, reversing the relation between these variances for $n = 1000$.

For the least squares estimator minimizing (1.6) one can extend the result to cover the whole range of situations from separated to non-separated. The least squares estimator minimizing (1.6) seems generally to have a smaller (pointwise) variance than the nonparametric MLE for moderate sample sizes (like $n = 1000$).

We also discussed a simpler least squares estimatior, minimizing (1.7) which does not have to be calculated by an iterative procedure and also does not need Lagrange multipliers to keep the values of the estimator inside the interval $[0, 1]$. But this estimator is clearly inferior to the least squares estimator minimizing (1.6) in the pointwise estimation problem. In estimating smooth functionals, like the first moment, the estimators seem to have a rather similar behavior, as shown in Section 3

Algorithms for computing the least squares estimator were discussed and in Section 3 we provide smooth functional theory needed to show that the so-called "off-diagonal terms" in the derivation of the asymptotic distribution are negligible. For the well-known case of current status data (the simplest case of interval censoring) the least squares approach coincides with the aproach based on analyzing the likelihood, since in that case the least square estimate and the nonparametric MLE are the same.

C++ routines with makefiles to check the algorithms and simulations are given in [6]. They use the iterative convex minorant algorithm.

5. Appendix.

5.1. *Variance and drift of the relevant limit process for Theorem 1.* Let, for a distribution function F, the functions $\psi_{1,F}$ and $\psi_{2,F}$ defined by

$$\psi_{1,F}(u,v,\delta_0,\delta_1) = \delta_0 - F(u) - \{\delta_1 - F(v) + F(u)\} \tag{5.1}$$

and

$$\psi_{2,F}(u,v,\delta_0,\delta_1) = \delta_1 - \{F(v) - F(u)\} + \{\delta_0 + \delta_1 - F(v)\}. \tag{5.2}$$

Then we have, for fixed $t_0 \in (0,M)$, and $W_{m,\hat{F}_n}$ is defined by (2.1)

$$\begin{aligned}
&W_{n,\hat{F}_n}(t_0 + n^{-1/3}t) - W_{n,\hat{F}_n}(t_0) = \int_{s\in[t_0,t_0+n^{-1/3}t]} dW_{n,\hat{F}_n}(s)\\
&= \int_{u\in[t_0,t_0+n^{-1/3}t]} \psi_{1,\hat{F}_n}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n + \int_{v\in[t_0,t_0+n^{-1/3}t]} \psi_{2,\hat{F}_n}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\int_{u\in[t_0,t_0+n^{-1/3}t]} \psi_{1,\hat{F}_n}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n + \int_{v\in[t_0,t_0+n^{-1/3}t]} \psi_{2,\hat{F}_n}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n\\
&= \int_{u\in[t_0,t_0+n^{-1/3}t]} \left\{\psi_{1,\hat{F}_n}(u,v,\delta_0,\delta_1) - \phi_{1,F_0}(u,v,\delta_0,\delta_1)\right\} d\mathbb{Q}_n\\
&\qquad + \int_{u\in[t_0,t_0+n^{-1/3}t]} \left\{\psi_{2,\hat{F}_n}(u,v,\delta_0,\delta_1) - \psi_{2,F_0}(u,v,\delta_0,\delta_1)\right\} d\mathbb{Q}_n\\
&\qquad + X_n(t),
\end{aligned} \tag{5.3}$$

where

$$\begin{aligned}
&X_n(t)\\
&= \int_{u\in[t_0,t_0+n^{-1/3}t]} \psi_{1,F_0}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n + \int_{v\in[t_0,t_0+n^{-1/3}t]} \psi_{2,F_0}(u,v,\delta_0,\delta_1)\, d\mathbb{Q}_n\\
&= \int_{u\in[t_0,t_0+n^{-1/3}t]} \psi_{1,F_0}(u,v,\delta_0,\delta_1)\, d\big(\mathbb{Q}_n - Q_0\big)\\
&\qquad\qquad + \int_{v\in[t_0,t_0+n^{-1/3}t]} \psi_{2,F_0}(u,v,\delta_0,\delta_1)\, d\big(\mathbb{Q}_n - Q_0\big),
\end{aligned}$$

where Q_0 is the underlying probability measure for the $(U_i, V_i, \Delta_{i0}, \Delta_{i1})$. We also define $X_n(t)$ for $t<0$ by changing the interval $[t_0, t_0+n^{-1/3}t]$ to $[t_0+n^{-1/3}t, t_0]$. For this process, we have the following lemma.

LEMMA 5. *Let the conditions of Theorem 1 be satisfied and let $\hat F_n$ be the least squares estimate, minimizing (2.8). Then $n^{2/3}X_n$ converges in distribution, in the Skorohod topology, to the process*

$$t \mapsto \sigma_{t_0} W(t), \qquad t \in \mathbb{R},$$

where W is standard two-sided Brownian motion and σ_{t_0} is defined by (2.7).

REMARK 3. Note the similarity with Lemma 9.4 in [5], but note that we now take F_0 instead of $\hat F_n$ in the definition of X_n.

PROOF. Note that we can also write:

$$\begin{aligned} X_n(t) = n^{-1} &\sum_{j: t_0 \le U_j \le t_0 + n^{-1/3}t} \{\Delta_{j0} - F_0(U_j)\} \\ &- n^{-1} \sum_{j: t_0 \le U_j \le t_0 + n^{-1/3}t} \{\Delta_{j1} - \{F_0(V_j) - F_0(U_j)\}\} \\ &+ n^{-1} \sum_{j: t_0 \le V_j \le t_0 + n^{-1/3}t} \{\Delta_{j1} - \{F_0(V_j) - F_0(U_j)\}\} \\ &+ n^{-1} \sum_{j: t_0 \le V_j \le t_0 + n^{-1/3}t} \{\Delta_{j0} + \Delta_{j1} - F_0(V_j)\}\} . \end{aligned} \tag{5.4}$$

We have:

$$\begin{aligned} &n^{4/3}\text{var}\big(X_n(t)\big) \\ &\sim tF_0(t_0)\{1-F_0(t_0)\}\big\{h_1(t_0)+h_2(t_0)\big\} \\ &\quad + t\int_{v=t_0}^{M} \{F_0(v)-F_0(t_0)\}\left[1-\{F_0(v)-F_0(t_0)\}\right] h(t_0,v)\,dv \\ &\quad + t\int_{u=0}^{t_0} \{F_0(t_0)-F_0(u)\}\left[1-\{F_0(t_0)-F_0(u)\}\right] h(u,t_0)\,du \\ &\quad + 2tF_0(t_0)\int_{v=t_0}^{M} \{F_0(v)-F_0(t_0)\}\, h(t_0,v)\,dv \\ &\quad + 2t\{1-F_0(t_0)\}\int_{u=0}^{t_0} \{F_0(t_0)-F_0(u)\}\, h(u,t_0)\,du, \end{aligned} \tag{5.5}$$

where h_1 and h_2 are the marginal densities of h.

This is seen as follows. Conditionally on (U_j, V_j), the triple $(\Delta_{j0}, \Delta_{j1}, \Delta_{j2})$ has a trinomial distribution, with parameters

$$p = F_0(U_j), \qquad q = F(V_j) - F_0(U_j), \qquad 1-p-q = 1-F(V_j).$$

This yields the first three terms on the right of (5.5), since the conditional variances are of the form $p(1-p)$, $q(1-q)$ and $(p+q)(1-p-q)$. The last two terms on the right of (5.5) are due to the covariance of Δ_{j0} and $-\Delta_{j1}$ in the first two terms and Δ_{j1} and $-\Delta_{j2}$ in the last two term on the right-hand side of (5.4), respectively. Other covariances are of lower order, since these involve values of U_j and V_j in the shrinking interval $[t_0, t_0 + n^{-1/3}t]$. So we get

$$n^{2/3}X_n \xrightarrow{\mathcal{D}} \sigma_{t_0} W, \qquad n \to \infty.$$

from tightness and the central limit theorem. □

We can now write

$$W_{n,\hat F_n}(t_0+n^{-1/3}t)-W_{n,\hat F_n}(t_0)=X_n(t)+Y_n(t),$$

where Y_n is defined by

$$\begin{aligned}Y_n(t)=&\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(u)-\hat F_n(u)\right\}\,d\mathbb{H}_n(u,v)\\&-\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-F_0(u)-\hat F_n(v)+\hat F_n(u)\right\}\,d\mathbb{H}_n(u,v)\\&+\int_{v\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-F_0(u)-\hat F_n(v)+\hat F_n(u)\right\}\,d\mathbb{H}_n(u,v)\\&+\int_{v\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}\,d\mathbb{H}_n(u,v),\end{aligned}\tag{5.6}$$

and where $\mathbb{H}_n$ is the empirical distribution function of the (U_i,V_i).

Let the process G_n be defined by

$$G_n(t)=\int_{u\in[0,t]}d\mathbb{H}_n(u,v)+\int_{v\in[0,t]}d\mathbb{H}_n(u,v),\qquad t\ge 0.$$

Then $\hat F_n$ is the left-continuous slope of the (“self-induced”) cusum diagram, consisting of the points $(0,0)$ and

$$\left(G_n(t),\int_{w\in[0,t]}\hat F_n(w)\,dG_n(w)+W_{n,\hat F_n}(t)\right),\qquad t\ge 0,$$

truncated at 0 and 1 (if the slope at t is <0, we put $\hat F_n(t)=0$ and similarly, if the slope at t is >1, we put $\hat F_n(t)=1$).

We have:

$$\begin{aligned}&\int_{u\in[t_0,t_0+n^{-1/3}t]}\hat F_n(w)\,dG_n(w)+Y_n(t)\\&=2\int_{u\in[t_0,t_0+n^{-1/3}t]}F_0(u)\,d\mathbb{H}_n(u,v)+2\int_{v\in[t_0,t_0+n^{-1/3}t]}F_0(u)\,d\mathbb{H}_n(u,v)\\&\quad-\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}\,d\mathbb{H}_n(u,v)\\&\quad-\int_{v\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(u)-\hat F_n(u)\right\}\,d\mathbb{H}_n(u,v).\end{aligned}\tag{5.7}$$

The last two terms are smooth functionals of the model and can be shown to be of order $O_p(n^{-5/6})$. We shall discuss this matter below in sections 5.2 and Section 3.

The conclusion is that

$$\begin{aligned}&\int_{u\in[t_0,t_0+n^{-1/3}t]}\hat F_n(w)\,dG_n(w)+W_{n,\hat F_n}(t_0+n^{-1/3}t)-W_{n,\hat F_n}(t_0)\\&=2\int_{u\in[t_0,t_0+n^{-1/3}t]}F_0(u)\,d\mathbb{H}_n(u,v)+2\int_{v\in[t_0,t_0+n^{-1/3}t]}F_0(v)\,d\mathbb{H}_n(u,v)\\&\quad+X_n(t)+O_p(n^{-5/6}).\end{aligned}$$

This means that $n^{1/3}\{\hat F_n(t_0)-F_0(t_0)\}$ behaves asymptotically as the slope of the greatest convex minorant of the process

$$2n^{2/3}\int_{u\in[t_0,t_0+n^{-1/3}t]}\{F_0(u)-F_0(t_0)\}\,d\mathbb{H}_n(u,v)$$
$$+\,2n^{2/3}\int_{v\in[t_0,t_0+n^{-1/3}t]}\{F_0(u)-F_0(t_0)\}\,d\mathbb{H}_n(u,v)+n^{2/3}X_n(t)+O_p(n^{-1/6}).$$

in the time scale

$$n^{1/3}\{G_n(t_0+n^{-1/3})-G_n(t_0)\}\sim t\{h_1(t)+h_2(t)\}.$$

Since

$$\begin{aligned}&2n^{2/3}\int_{u\in[t_0,t_0+n^{-1/3}t]}\{F_0(u)-F_0(t_0)\}\,d\mathbb{H}_n(u,v)\\&\qquad+2n^{2/3}\int_{v\in[t_0,t_0+n^{-1/3}t]}\{F_0(u)-F_0(t_0)\}\,d\mathbb{H}_n(u,v)\\&\sim f_0(t_0)\{h_1(t_0)+h_2(t_0)\}t^2,\end{aligned}$$

the result now follows from Brownian scaling. using a similar computation on the interval $[t_0+n^{-1/3}t,t_0]$ for $t<0$.

5.2. *Negligibility of the "off-diagonal terms".* In (5.7) we met two terms which we call "off-diagonal terms", because, for example in the interior point method for solving the minimization problem, they correspond to non-zero off-diagonal terms of the Hessian. This is a difficulty we do not have to deal with in the current status model. They are the terms

$$\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}d\mathbb{H}_n(u,v) \tag{5.8}$$

and

$$\int_{v\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(u)-\hat F_n(u)\right\}d\mathbb{H}_n(u,v). \tag{5.9}$$

In (5.8) the variable v roughly varies between t_0 and M, since we have $v\ge u$. Likewise, in in (5.9) the variable u roughly varies between 0 and t_0.

We now take a closer look at (5.8). First of all, we can write

$$\begin{aligned}&\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}d\mathbb{H}_n(u,v)\\&=\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}d\big(\mathbb{H}_n-H\big)(u,v)\\&\qquad+\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}dH(u,v).\end{aligned}$$

For the first term on the right we have:

$$\begin{aligned}&\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}d\big(\mathbb{H}_n-H\big)(u,v)\\&=\int_{u\in[t_0,t_0+n^{-1/3}t]}F_0(v)\,d\big(\mathbb{H}_n-H\big)(u,v)-\int_{u\in[t_0,t_0+n^{-1/3}t]}\hat F_n(v)\,d\big(\mathbb{H}_n-H\big)(u,v),\end{aligned}$$

where both terms are of order $O_p(n^{-5/6})$. This is obvious for the first term, and for the second term if follows from the entropy with bracketing for bounded monotone functions.

So we turn to the term

$$\int_{u\in[t_0,t_0+n^{-1/3}t]}\left\{F_0(v)-\hat F_n(v)\right\}dH(u,v).$$

This term is also expected to be of order $O_p(n^{-5/6})$. The key to this is the fact that integrals of the form

$$\int_a^b\left\{F_0(x)-\hat F_n(x)\right\}dx$$

for $b>a$ are of order $O_p(n^{-1/2})$. This follows from smooth functional theory, discussed in Section 3. We cannot work with simple upper bounds for the distance between $\hat F_n$ and F_0, since these will only give us order $O_p(n^{-2/3})$ instead of order $O_p(n^{-5/6})$.

REFERENCES

[1] BERMAN, A. and PLEMMONS, R. J. (1979). *Nonnegative matrices in the mathematical sciences*. *Computer Science and Applied Mathematics*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London. MR544666

[2] BIRGÉ, L. (1999). Interval censoring: a nonasymptotic point of view. *Math. Methods Statist.* **8** 285–298. MR1735467 (2001g:62025)

[3] GESKUS, R. B. and GROENEBOOM, P. (1999). Asymptotically optimal estimation of smooth functionals for interval censoring, case 2. *Ann. Statist.* **27** 627–674. https://doi.org/10.1214/aos/1018031211 MR1714713 (2000j:60044)

[4] 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. https://doi.org/10.1007/BFb0095675 MR1600884 (99c:62092)

[5] GROENEBOOM, P. (2024). Nonparametric estimation of the incubation time distribution. *Electron. J. Stat.* **18** 1917–1969. https://doi.org/10.1214/24-ejs2243 MR4736274

[6] GROENEBOOM, P. (2025). Interval Censoring. https://github.com/pietg/interval-censoring.

[7] GROENEBOOM, P. and JONGBLOED, G. (2014). *Nonparametric Estimation under Shape Constraints*. Cambridge Univ. Press, Cambridge.

[8] GROENEBOOM, P. and KETELAARS, T. (2011). Estimators for the interval censoring problem. *Electron. J. Stat.* **5** 1797–1845. https://doi.org/10.1214/11-EJS657 MR2870151

[9] GROENEBOOM, P. and WELLNER, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. *DMV Seminar* **19**. Birkhäuser Verlag, Basel. MR1180321 (94k:62056)

[10] ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order restricted statistical inference*. *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. John Wiley & Sons Ltd., Chichester. MR961262 (90b:62001)

[11] VAN ES, A. J. and VAN ZUIJLEN, M. C. A. (1996). Convex minorant estimators of distributions in nonparametric deconvolution problems. *Scand. J. Statist.* **23** 85–104. MR1380484 (97c:62071)