NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATORS FOR INTERVAL CENSORING AND DECONVOLUTION

BY

PIET GROENEBOOM

TECHNICAL REPORT NO. 378

JULY 1991

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Nonparametric Maximum Likelihood Estimators

for Interval Censoring and Deconvolution

By

Piet Groeneboom

Technical Report No. 378

July 1991

Prepared Under the Auspices

Of

National Science Foundation Grant DMS89–05874

Department of Statistics

Stanford University

Stanford, California

Nonparametric Maximum Likelihood Estimators
for Interval Censoring and Deconvolution

Piet Groeneboom

Delft University of Technology
Department of Mathematics and Computer Science
Mekelweg 4
2628 CD Delft
The Netherlands

INTRODUCTION

The following notes are lecture notes of a summer course given at Stanford University, Department of Statistics, in the summer of 1990. An attempt is made to develop theory for nonparametric maximum likelihood estimators (NPMLE's) for the distribution function in certain inverse problems.

The first chapter deals with the interval censoring problem. In this case one only has information about an interval to which the variable of interest belongs, a situation quite common in medical research. The classical approach (see, e.g., Turnbull (1974)) is to use the EM-algorithm. This is shown to correspond to the so-called "self-consistency equations" for the NPMLE. These equations yield necessary but not sufficient conditions for the characterization of the NPMLE and have (so far) not been very useful in developing distribution theory. For this reason we turn to another approach, based on isotonic regression theory, which gives necessary and sufficient conditions. Moreover, it yields efficient algorithms and also leads to distribution theory, to be developed in Chapter 5.

The second chapter follows a similar path as the first chapter, but this time for deconvolution problems. Chapter 3 discusses algorithms and in chapter 4 consistency is proved for the NPMLE in all situations discussed in the preceding chapters. Chapter 5 deals with distribution theory and discusses open problems and conjectures.

I want to thank Persi Diaconis, Iain Johnstone and David Siegmund for providing the opportunity of doing the present work in the pleasant surroundings of Stanford University and for their helpful comments during the summer course.

Table of Contents

1. THE INTERVAL CENSORING PROBLEM

1.1. Characterization of the non-parametric maximum likelihood estimators.

We will study the following two cases of interval censoring.

Case 1. Let $(X_1, T_1), \ldots, (X_n, T_n)$ be a sample of random variables in $I\!\!R_+^2$, where X_i and T_i are independent (nonnegative) random variables with continuous distribution functions F_0 and G, respectively. The only observations which are available are T_i ("observation time") and $\delta_i = \{X_i \leq T_i\}$. Here and (often) in the sequel we will denote the indicator of an event A (such as $\{X_i \leq T_i\}$) just by A, instead of 1_A. The log likelihood for F_0 is given by the function

$$(1.1) \qquad F \mapsto \sum_{i=1}^{n} \left\{ \delta_i \log F(T_i) + (1 - \delta_i) \log(1 - F(T_i)) \right\},$$

where F is a right-continuous distribution function.

Case 2. Let $(X_1, T_1, U_1), \ldots, (X_n, T_n, U_n)$ be a sample of random variables in $I\!\!R_+^3$, where X_i is a (nonnegative) random variable with continuous distribution function F_0, and where T_i and U_i are (nonnegative) random variables, independent of X_i, with a joint continuous distribution function H and such that $T_i \leq U_i$ with probability one. The only observations which are available are (T_i, U_i) (the "observation times") and $\delta_i = \{X_i \leq T_i\}$, $\gamma_i = \{X_i \in (T_i, U_i]\}$. In this case the log likelihood for F_0 is given by the function

$$(1.2) \quad F \mapsto \sum_{i=1}^{n} \left\{ \delta_i \log F(T_i) + \gamma_i \log(F(U_i) - F(T_i)) + (1 - \delta_i - \gamma_i) \log(1 - F(U_i)) \right\}.$$

The second case of interval censoring is related to "double censoring", which is studied e.g., in Chang (1990), but the important difference is that with interval censoring the value of X_i is unknown, even if we know that $T_i < X_i \leq U_i$, whereas in the case of double censoring the value of X_i is known if $T_i < X_i \leq U_i$. The remaining parts of the models for double censoring and for interval censoring, case 2, are the same.

We first study the likelihood equations for interval censoring, case 1. The log likelihood, divided by n, can be written in the following way:

$$(1.3) \qquad \psi(F) \stackrel{\text{def}}{=} \int_{I\!\!R^2} \left\{ 1_{\{x \leq t\}} \log F(t) + 1_{\{x > t\}} \log\{1 - F(t)\} \right\} dP_n(x, t),$$

where P_n is the empirical probability measure of the pairs (X_i, T_i), $1 \leq i \leq n$. The *nonparametric maximum likelihood estimator* (NPMLE) $\hat{F}_n$ of F is a (right-continuous) distribution function F, maximizing (1.3).

1

Remark 1.1. Note that only the values of $\hat{F}_n$ at the observation points T_i matter for the maximization problem. To avoid trivialities, we will take as "the" NPMLE a distribution function which is piecewise constant, and only has jumps at the observation points T_i. It may happen that the likelihood function is maximized by a function F such that $F(T_i) < 1$, for each T_i. In this case we do not specify the location of the remaining mass to the right of the biggest observation point T_i. We shall show that, under these conventions, the NPMLE is uniquely determined, both in case 1 and case 2 of the interval censoring problem.

First we give the usual characterization of the NPMLE in terms of the so-called self-consistency equations. This characterization is used by Turnbull (1974) who attributes the concept to Efron (1967). Let, for each $t \in \mathbb{R}$, the function $A_t : \mathbb{R} \to \mathbb{R}$ be defined by

$$(1.4) \qquad A_t(u) = \begin{cases} 0, & \text{if } F(t) = 0,\, u \in \mathbb{R}, \\ F(u \wedge t)/F(t) - F(u), & \text{if } F(t) > 0,\, u \in \mathbb{R}, \end{cases}$$

where $u \wedge t = \min\{u, t\}$. Differentiating the function ψ in the direction A_t yields:

$$(1.5) \qquad \begin{aligned} &\lim_{h \downarrow 0} \frac{\psi(F + hA_t) - \psi(F)}{h} \\ &= \int_{\mathbb{R}^2} \left\{ 1_{\{x \le u\}} \frac{F(u \wedge t)/F(t) - F(u)}{F(u)} - 1_{\{x > u\}} \frac{F(u \wedge t)/F(t) - F(u)}{1 - F(u)} \right\} dP_n(x, u). \end{aligned}$$

If F is the NPMLE, then (1.5) should equal zero, for each t. So we get for the NPMLE $\hat{F}_n$ the equation:

$$\hat{F}_n(t) = \int_{\mathbb{R}^2} \left\{ 1_{\{x \le u\}} \frac{\hat{F}_n(u \wedge t)}{\hat{F}_n(u)} + 1_{\{x > u\}} \frac{\hat{F}_n(t) - \hat{F}_n(u \wedge t)}{1 - \hat{F}_n(u)} \right\} dP_n(x, u).$$

A nicer way of writing this equation is:

$$(1.6) \qquad \hat{F}_n(t) = E_{\hat{F}_n} \left\{ F_n(t) \mid T_1, \ldots, T_n, \delta_1, \ldots, \delta_n \right\},$$

where F_n is the (unobservable) empirical distribution function of the random variables $X_1, \ldots, X_n$. So $\hat{F}_n(t)$ is the conditional expectation of the empirical distribution function F_n at t, given the available information $T_1, \ldots, T_n, \delta_1, \ldots, \delta_n$, under the (self-induced) probability measure $P_{\hat{F}_n}$. We will see in Chapter 3 that the fixed point equation (1.6) immediately yields the iteration steps of the EM algorithm. We note in passing here that the "self-consistency equation" (1.6) does not uniquely determine the NPMLE $\hat{F}_n$, even under the conventions of Remark 1.1.

We now want to give a different characterization of the NPMLE, using concepts from the theory of isotonic regression. For this, some notation is needed. Let $T_{(i)}$ be the ith order statistic of $T_1, \ldots, T_n$, and let $\delta_{(i)}$ be the corresponding indicator, i.e., if $T_j = T_{(i)}$, then

$\delta_{(i)} = 1_{\{X_j \leq T_j\}}$. The NPMLE corresponds to a vector $\tilde{y} = (y_1, \ldots, y_n) \in I\!\!R^n$, maximizing the function

$$(1.7) \qquad \phi(\tilde{x}) = \sum_{i=1}^{n} \left\{ \delta_{(i)} \log x_i + (1 - \delta_{(i)}) \log(1 - x_i) \right\}, \ \tilde{x} \in I\!\!R^n,$$

under the side condition

$$(1.8) \qquad 0 \leq x_1 \leq \cdots \leq x_n \leq 1.$$

Note that if $\delta_{(i)} = 0$, $i = 1, \ldots, k$, the $\tilde{y}$, maximizing (1.7) should satisfy $y_1 = \cdots = y_k = 0$, since this makes the corresponding second term in the sum (1.7) as big as possible and puts no additional constraints on the values of y_i, for $i > k$. Likewise, if $\delta_{(i)} = 1$, $j \leq i \leq n$, the $\tilde{y}$, maximizing (1.7) should satisfy $y_j = \cdots = y_n = 1$.

So in the maximization problem we may assume without loss of generality that $\delta_{(1)} = 1$ and $\delta_{(n)} = 0$. We then can also assume without loss of generality that $y_1 > 0$ and $y_n < 1$, if $\tilde{y}$ maximizes (1.7), since otherwise we would have $\phi(\tilde{y}) = -\infty$.

The following proposition gives necessary and sufficient conditions for $\tilde{y}$ to be a vector, maximizing (1.7), under the constraint (1.8) and the just mentioned restrictions.

PROPOSITION 1.1. *Let* $\delta_{(1)} = 1$ *and* $\delta_{(n)} = 0$, *and let* $\tilde{y} = (y_1, \ldots, y_n)$ *satisfy (1.8), with* x_i *replaced by* y_i. *Then* $\tilde{y}$ *maximizes (1.7) if and only if*

$$(1.9) \qquad \sum_{j \geq i} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} \leq 0, \ i = 1, \ldots, n,$$

and

$$(1.10) \qquad \sum_{i=1}^{n} \left\{ \frac{\delta_{(i)}}{y_i} - \frac{1 - \delta_{(i)}}{1 - y_i} \right\} y_i = 0.$$

Moreover, $\tilde{y}$ *is uniquely determined by (1.9) and (1.10).*

PROOF: First suppose that $\tilde{y}$ satisfies (1.9) and (1.10). Since we may assume $y_1 > 0$ and $y_n < 1$, all terms in (1.9) are finite. The function ϕ is concave, so if $\tilde{x}$ satisfies (1.8), we get

$$(1.11) \qquad \phi(\tilde{x}) - \phi(\tilde{y}) \leq \langle \nabla \phi(\tilde{y}), \tilde{x} - \tilde{y} \rangle,$$

where $\nabla \phi(\tilde{y})$ is the vector of partial derivatives

$$\nabla \phi(\tilde{y}) = \left(\frac{\delta_{(1)}}{y_1} - \frac{1 - \delta_{(1)}}{1 - y_1}, \ldots, \frac{\delta_{(n)}}{y_n} - \frac{1 - \delta_{(n)}}{1 - y_n} \right).$$

3

Furthermore, $\langle \nabla \phi(\tilde{y}), \tilde{y} \rangle = 0$, if $\tilde{y}$ satisfies (1.10).

Now note that each $\tilde{x}$, satisfying (1.8), can be written in the following form

$$(1.12) \qquad \tilde{x} = \sum_{i=1}^{n} \alpha_i \tilde{1}_i,$$

where $\alpha_i = x_{n-i+1} - x_{n-i}$, $\tilde{1}_i$ is a vector which has 1's as its last i components and zeros as its first $n - i$ components, and where $x_0 \stackrel{\text{def}}{=} 0$. Hence, if $\tilde{x}$ satisfies (1.8), we get

$$
\begin{aligned}
\langle \nabla \phi(\tilde{y}), \tilde{x} - \tilde{y} \rangle = \langle \nabla \phi(\tilde{y}), \tilde{x} \rangle &= \sum_{i=1}^{n} \alpha_i \langle \nabla \phi(\tilde{y}), \tilde{1}_i \rangle \\
(1.13) \\
&= \sum_{i=1}^{n} \alpha_i \sum_{j \geq i} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} \leq 0.
\end{aligned}
$$

Thus $\tilde{y}$ maximizes (1.7).

Conversely, suppose that $\tilde{y}$ maximizes (1.7) under the constraint (1.8). If $0 < \epsilon < 1 - y_n$, the vector $\tilde{y} + \epsilon \tilde{1}_i$ satisfies (1.8), for each i, $1 \leq i \leq n$ (here we use $y_n < 1$). Since $\tilde{y}$ maximizes (1.7) we must have:

$$\lim_{\epsilon \downarrow 0} \frac{\phi(\tilde{y} + \epsilon \tilde{1}_i) - \phi(\tilde{y})}{\epsilon} = \sum_{j \geq i} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} \leq 0, \ 1 \leq i \leq n.$$

This yields (1.9).

Again using $y_1 > 0$ and $y_n < 1$, we obtain (1.10) by observing that

$$\lim_{h \to 0} \frac{\phi(\tilde{y} + h\tilde{y}) - \phi(\tilde{y})}{h} = 0,$$

since otherwise we could find an h close to zero such that $\phi((1+h)\tilde{y}) > \phi(\tilde{y})$, while $(1+h)\tilde{y}$ still satisfies (1.8).

Since ϕ is concave and, moreover, the (diagonal) matrix of second derivatives is non-singular at $\tilde{y}$, it follows that the vector $\tilde{y}$, satisfying (1.9) and (1.10), is the *unique* maximizing vector. ∎

Remark 1.2. Proposition 1.1 is actually a special case of Fenchel's duality theorem, see e.g., Rockafellar (1970), Theorem 31.4. Since things simplify somewhat in the present setting, we chose to write down a proof of this special case.

Example 1.1. Let $n = 5$, $\delta_{(1)} = \delta_{(3)} = \delta_{(4)} = 1$, and $\delta_{(2)} = \delta_{(5)} = 0$. Then the vector $\tilde{y}$, maximizing (1.7) is given by

$$y_1 = y_2 = \tfrac{1}{2}, \ y_3 = y_4 = y_5 = \tfrac{2}{3}.$$

4

Note that y_m, $1 \leq m \leq 5$, is given by

$$y_m = \max_{i \leq m} \min_{k \geq m} \frac{\sum_{i \leq j \leq k} \delta_{(j)}}{k - i + 1}.$$

This is the so-called "max-min formula" for the solution of the maximization problem.

The solution can be found graphically by plotting the points $\left(i, \sum_{j \leq i} \delta_{(j)}\right)$ in the plane, and drawing the *(greatest) convex minorant* of these points on the interval $[0, 5]$. The convex minorant is defined as the function $H^* : [0, 5] \to I\!R$ such that

$$H^*(t) = \sup\{H(t) : H(i) \leq \sum_{j \leq i} \delta_{(j)}, \text{ for each } i, \, 0 \leq i \leq 5, \text{ and } H : [0, n] \to I\!R \text{ is convex}\},$$

for $t \in [0, 5]$, where $\sum_{j \leq i} \delta_{(j)} \stackrel{\text{def}}{=} 0$, if $i = 0$. Then y_i is the *left derivative* of H^* at i, for $i = 1, \ldots, 5$.

The next proposition shows that the representation of $\tilde{y}$, given in Example 1.5, holds generally.

PROPOSITION 1.2. *Let H^* be the convex minorant of the points $\left(i, \sum_{j \leq i} \delta_{(j)}\right)$ on $[0, n]$, i.e.,*

$$H^*(t) = \sup\{H(t) : H(i) \leq \sum_{j \leq i} \delta_{(j)}, \text{ for each } i, \, 0 \leq i \leq n, \text{ and } H : [0, n] \to I\!R \text{ is convex}\},$$

for $t \in [0, n]$. Moreover, let y_i be the left derivative of H^* at i. Then $\tilde{y} = (y_1, \ldots, y_n)$ is the unique vector maximizing (1.7) under the constraint (1.8).

Remark 1.3. Note that in Proposition 1.2 (in contrast to Proposition 1.1) no restriction is made on $\delta_{(1)}$ and $\delta_{(n)}$.

PROOF OF PROPOSITION 1.2: We start by first looking at some special (trivial) cases. If $\delta(i) = 0$, for each i, then $H^*(t) \equiv 0$, and hence $y_i = 0$, for each i. If $\delta(i) = 1$, for each i, then $H^*(t) = t$, $t \in [0, n]$, and hence $y_i = 1$, for each i. Finally, if $\delta_{(i)} = 0$, $1 \leq i \leq k$ and $\delta_{(i)} = 1$, $k < i \leq n$, for some $k < n$, then $y_i = 0$, $1 \leq i \leq k$ and $y_i = 1$, $k < i \leq n$. It is clear that in all these cases $\tilde{y}$ maximizes (1.7).

So we suppose that there is at least one $\delta_{(i)} = 1$, followed by a $\delta_{(j)} = 0$, for some $j > i$. Let k_0 be the smallest index i such that $\delta_{(i)} = 1$, and let m_0 be the largest index i such that $\delta_{(i)} = 0$. Then $y_i = 0$, if $i < k_0$, $y_i \in (0, 1)$, if $k_0 \leq i \leq m_0$, and $y_i = 1$, if $i > m_0$. Let k_1 be the largest index $k > k_0$ such that

$$\frac{\sum_{k_0 \leq j < k} \delta_{(j)}}{k - k_0} = \min_{m > k_0} \frac{\sum_{k_0 \leq j < m} \delta_{(j)}}{m - k_0}.$$

5

Then

$$y_i = y_{k_0} = \frac{\sum_{k_0 \leq j < k_1} \delta(j)}{k_1 - k_0}, \; k_0 \leq i < k_1,$$

and

$$\frac{\sum_{k_0 \leq j < k} \delta(j)}{k - k_0} \geq y_{k_0}, \; k > k_0.$$

Hence we get, for each k, $k_0 < k \leq k_1$:

$$\frac{\sum_{k_0 \leq j < k}(1 - \delta(j))}{\sum_{k_0 \leq j < k} \delta(j)} = \frac{k - k_0}{\sum_{k_0 \leq j < k} \delta(j)} - 1 \leq \frac{1}{y_{k_0}} - 1 = \frac{1 - y_{k_0}}{y_{k_0}},$$

implying

$$(1.14) \qquad \sum_{k_0 \leq j < k} \frac{\delta(j)}{y_{k_0}} \bigg/ \sum_{k_0 \leq j < k} \frac{1 - \delta(j)}{1 - y_{k_0}} = \frac{1 - y_{k_0}}{y_{k_0}} \bigg/ \frac{\sum_{k_0 \leq j < k}(1 - \delta(j))}{\sum_{k_0 \leq j < k} \delta(j)} \geq 1.$$

Thus:

$$(1.15) \qquad \sum_{k_0 \leq j < k} \left\{ \frac{\delta(j)}{y_{k_0}} - \frac{1 - \delta(j)}{1 - y_{k_0}} \right\} \geq 0, \; k_0 < k \leq k_1.$$

If $k = k_1$, we obtain equality in (1.14) (and hence in (1.15)), since $\sum_{k_0 \leq j < k_1} \delta(j) \big/ (k_1 - k_0)$ is equal to y_{k_0}.

We can now repeat this argument for indices $i > k_1$: by definition

$$y_{k_1} = \max_{i \leq k_1} \min_{k \geq k_1} \frac{\sum_{i \leq j \leq k} \delta(j)}{k - i + 1}.$$

Suppose k_2 is the largest index k such that

$$\frac{\sum_{k_1 \leq j < k} \delta(j)}{k - k_1} = \max_{i \leq k_1} \min_{m \geq k_1} \frac{\sum_{i \leq j \leq m} \delta(j)}{m - i + 1}.$$

Then we have, for $k_1 < k \leq k_2$:

$$\frac{\sum_{k_1 \leq j < k}(1 - \delta(j))}{\sum_{k_1 \leq j < k} \delta(j)} = \frac{k - k_1}{\sum_{k_1 \leq j < k} \delta(j)} - 1 \leq \frac{1}{y_{k_1}} - 1 = \frac{1 - y_{k_1}}{y_{k_1}},$$

implying

$$(1.16) \qquad \sum_{k_1 \leq j < k} \left\{ \frac{\delta(j)}{y_{k_1}} - \frac{1 - \delta(j)}{1 - y_{k_1}} \right\} \geq 0, k_1 < k \leq k_2,$$

(compare with (1.15)). If $k = k_2$, we obtain equality in (1.16).

Generally we get for a "block" of consecutive indices $k_m, \ldots, k_{m+1} - 1$ on which the function $i \mapsto y_i$ is constant:

$$(1.17) \qquad \sum_{k_m \leq j < k} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} \geq 0, \; k_m < k \leq k_{m+1},$$

with equality if $k = k_{m+1}$. Summing these (in)equalities over the successive blocks yields

$$\sum_{k_0 \leq j < m_0 + 1} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} = 0,$$

and, using (1.17):

$$(1.18) \qquad \sum_{i \leq j \leq m_0} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} \leq 0, \; i = k_0, \ldots, m_0.$$

Moreover, since $i \mapsto y_i$ is constant on blocks of consecutive indices $k_m, \ldots, k_{m+1} - 1$ and

$$(1.19) \qquad \sum_{k_m \leq j < k_{m+1}} \left\{ \frac{\delta_{(j)}}{y_j} - \frac{1 - \delta_{(j)}}{1 - y_j} \right\} = 0,$$

on each block, we obtain

$$(1.20) \qquad \sum_{i=k_0}^{m_0} \left\{ \frac{\delta_{(i)}}{y_i} - \frac{1 - \delta_{(i)}}{1 - y_i} \right\} y_i = 0.$$

But (1.18) and (1.20) correspond to (1.9) and (1.10), with the index 1 replaced by k_0 and the index n replaced by m_0.

The uniqueness of $\tilde{y}$ now easily follows from Proposition 1.1. ∎

Proposition 1.2 shows that $i \mapsto y_i$ is the *isotonic regression* of the function $i \mapsto \delta_{(i)}$ in the class of all isotonic functions $i \mapsto x_i$, with respect to the simple ordering $x_1 \leq \cdots \leq x_n$. This means that the function $i \mapsto y_i$ minimizes

$$\sum_{i=1}^{n} \left\{ \delta_{(i)} - x_i \right\}^2,$$

in the class of such isotonic functions $i \mapsto x_i$, see, e.g., Theorem 1.2.1 on p. 7 of Robertson et al. (1988), where the connection between the derivative of the convex minorant and the solution of the isotonic regression problem is given. So we get that the NPMLE of F_0, maximizing (1.1), is given by

$$\hat{F}_n(T_{(i)}) = y_i,$$

7

where $\tilde{y}$ is the isotonic regression of the function $i \mapsto \delta_{(i)}$.

We now turn to interval censoring, case 2. Again we have two approaches, either based on the self-consistency equations or on the theory of isotonic regression. We start by deriving the self-consistency equations. The argument is similar to the argument, used to derive (1.5). The log likelihood, divided by n, can now be written in the following way:

$$(1.21) \qquad \psi(F) \stackrel{\text{def}}{=} \int_{\mathbb{R}^3} \phi_F(x,t,u) \, dP_n(x,t,u),$$

where

$$\phi_F(x,t,u) \stackrel{\text{def}}{=} 1_{\{x \leq t\}} \log F(t) + 1_{\{t < x \leq u\}} \log\{F(u) - F(t)\} + 1_{\{x > u\}} \log\{1 - F(u)\},$$

and P_n is the empirical probability measure of the triples (X_i, T_i, U_i), $1 \leq i \leq n$.

Defining, for $t_0 \in \mathbb{R}$, the function $A_{t_0} : \mathbb{R} \to \mathbb{R}$ by

$$A_{t_0}(t) = \begin{cases} 0, & \text{if } F(t_0) = 0,\ t \in \mathbb{R}, \\ F(t \wedge t_0)/F(t_0) - F(t), & \text{if } F(t_0) > 0,\ t \in \mathbb{R}, \end{cases}$$

we get:

$$\lim_{h \downarrow 0} \frac{\psi(F + hA_{t_0}) - \psi(F)}{h}$$

$$= \int \left\{ 1_{\{x \leq t\}} \frac{F(t \wedge t_0)/F(t_0) - F(t)}{F(t)} \right.$$

$$+ 1_{\{t < x \leq u\}} \frac{\big(F(u \wedge t_0) - F(t \wedge t_0)\big)/F(t_0) - \big(F(u) - F(t)\big)}{F(u) - F(t)}$$

$$\left. + 1_{\{x > u\}} \frac{\big(1 - F(u \wedge t_0)/F(t_0)\big) - \big(1 - F(u)\big)}{1 - F(u)} \right\} dP_n(x,t,u).$$

If F is the NPMLE, then this limit should be zero, for each t_0. This leads to the following equation for the NPMLE:

$$\hat{F}_n(t_0)$$

$$(1.22) \qquad = \int \left\{ 1_{\{x \leq t\}} \frac{\hat{F}_n(t \wedge t_0)}{\hat{F}_n(t)} + 1_{\{t < x \leq u\}} \frac{\hat{F}_n(u \wedge t_0) - \hat{F}_n(t \wedge t_0)}{\hat{F}_n(u) - \hat{F}_n(t)} \right.$$

$$\left. + 1_{\{x > u\}} \frac{\hat{F}_n(t_0) - \hat{F}_n(u \wedge t_0)}{1 - \hat{F}_n(u)} \right\} dP_n(x,t,u),$$

which can also be written

$$(1.23) \qquad \hat{F}_n(t) = E_{\hat{F}_n} \left\{ F_n(t) \mid T_1, \ldots, T_n, U_1, \ldots, U_n, \gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_n \right\},$$

8

where F_n is the empirical distribution function of the random variables $X_1, \ldots, X_n$ (compare with (1.6)). So, in analogy with (1.6), we get that $\hat{F}_n(t_0)$ is the conditional expectation of the empirical distribution function F_n at t_0, given the available information $T_1, \ldots, T_n, U_1, \ldots, U_n$ and $\gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_n$ under the (self-induced) probability measure $P_{\hat{F}_n}$. Equation (1.23) yields the iteration steps of the EM algorithm, as will be shown in Chapter 3. Again the "self-consistency equation" (1.23) does not uniquely determine the NPMLE $\hat{F}_n$, even under the conventions of Remark 1.1.

We now turn to the isotonic regression approach. Let $\mathcal{F}$ be the class of distribution functions F satisfying

$$
\begin{aligned}
(1.24) \qquad & F(T_i) > 0 && \text{, if } X_i \leq T_i, \\
& F(U_i) - F(T_i) > 0 && \text{, if } T_i < X_i \leq U_i, \\
& 1 - F(U_i) > 0 && \text{, if } X_i > U_i,
\end{aligned}
$$

and having mass concentrated on the set of observation points augmented with an extra point bigger than all observation points (see Remark 1.1). Note that if F maximizes the log likelihood (1.21), then F has to satisfy (1.24), since otherwise

$$ \psi(F) = -\infty. $$

For distribution functions $F \in \mathcal{F}$, we define the process $t \mapsto W_F(t)$ by

$$
\begin{aligned}
(1.25) \qquad W_F(t) = & \int_{t' \in [0,t],\, x \leq t'} F(t')^{-1} dP_n(x, t', u) \\
& - \int_{t' \in [0,t],\, t' < x \leq u} \{F(u) - F(t')\}^{-1} dP_n(x, t', u) \\
& + \int_{u \in [0,t],\, t' < x \leq u} \{F(u) - F(t')\}^{-1} dP_n(x, t', u) \\
& - \int_{u \in [0,t],\, x > u} \{1 - F(u)\}^{-1} dP_n(x, t', u),
\end{aligned}
$$

$$ \text{for } t \geq 0, $$

where P_n is the empirical probability measure of the points (X_i, T_i, U_i), $i = 1, \ldots, n$. The process W_F can only have a jump at an observation point t which is either a T_i or a U_i, and is such that the corresponding X_i either belongs to the interval to the right or to the interval to the left of this observation point. If, for example, $X_i \leq T_i$, then we will meet no terms of the form

$$ \frac{1}{n(F(U_i) - F(T_i))} \qquad \text{or} \qquad \frac{1}{n(1 - F(U_i))} $$

in the process (1.25). This corresponds to the fact that there will be no terms of the form

$$ \log\{F(U_i) - F(T_i)\} \qquad \text{or} \qquad \log\{1 - F(U_i)\} $$

9

in the log likelihood function. Likewise, if $\{X_i > U_i\}$, we will meet no terms of the form

$$\frac{1}{nF(T_i)} \quad \text{or} \quad \frac{1}{n\big(F(U_i) - F(T_i)\big)}$$

in the process (1.25).

So we can "throw away" these irrelevant observation points. For later convenience, we will denote this "thinned" set of observation points by J_n, as is expressed by the following definition.

DEFINITION 1.1. *Let $J_n^{(1)}$ be the set of observation times T_i such that X_i either belongs to $[0, T_i]$ or to $(T_i, U_i]$, and let $J_n^{(2)}$ be the set of observation times U_i such that X_i either belongs to $(T_i, U_i]$ or to (U_i, ∞). Furthermore, let $J_n = J_n^{(1)} \cup J_n^{(2)}$, and let $T_{(j)}$ be the jth order statistic of the set J_n.*

In the maximization problem we may assume, without loss of generality, that $T_{(1)}$ corresponds to an observation point T_i such that $\{X_i \leq T_i\} = 1$, and, similarly, we may assume that the largest order statistic in J_n, say $T_{(m)}$, corresponds to an observation point U_i such that $\{X_i > U_i\} = 1$. The reasons for this are similar to those discussed before Proposition 1.1. If, for example, $T_{(1)}$ would correspond to an observation point T_i such that $\{T_i < X_i \leq U_i\} = 1$, the distribution function F, maximizing (1.21), should satisfy $F(T_{(1)}) = 0$, since this makes the term $n^{-1}\log\{F(U_i) - F(T_i)\}$ as big as possible, without putting additional constraints on F. Similarly, if, for example, the largest order statistic $T_{(m)} \in J_n$ would correspond to an observation time U_i such that $\{T_i < X_i \leq U_i\} = 1$, then the maximizing F should satisfy $F(U_i) = 1$. In this case we can redefine the observation time T_i to be a right endpoint U_i' of an interval such that $\{X_i > U_i'\} = 1$. Since the left endpoint of such an interval would not belong to the set J_n, we get an equivalent maximization problem, with the left endpoint T_i replaced by a right endpoint U_i'. Finally, if the largest order statistic $T_{(m)} \in J_n$ would correspond to an observation time T_i such that $\{X_i \leq T_i\} = 1$, then the maximizing F satisfies $F(T_{(m)}) = 1$, and we can just remove this observation point from the set J_n, without altering the maximization problem.

We now get the following proposition, analogous to Proposition 1.1.

PROPOSITION 1.3. *Let $T_{(1)}$ correspond to an observation point T_i such that $\{X_i \leq T_i\} = 1$, and let the largest order statistic $T_{(m)} \in J_n$ correspond to an observation point U_i such that $\{X_i > U_i\} = 1$. Then $\hat{F}_n$ maximizes (1.21) over all $F \in \mathcal{F}$ if and only if*

$$(1.26) \qquad \int_{[t,\infty)} dW_{\hat{F}_n}(t') \leq 0, \qquad \forall t \geq 0,$$

and

$$(1.27) \qquad \int \hat{F}_n(t) \, dW_{\hat{F}_n}(t) = 0,$$

10

where W_F is defined by (1.24). Moreover, $\hat{F}_n$ is uniquely determined by (1.26) and (1.27).

PROOF: Suppose $\hat{F}_n$ satisfies (1.26) and (1.27). Then, for all $F \in \mathcal{F}$,

$$\psi(F) - \psi(\hat{F}_n) \leq \int (F(t) - \hat{F}_n(t))\, dW_{\hat{F}_n}(t).$$

This is shown in a similar way as (1.11) in the proof of Proposition 1.1. In fact, defining the function ϕ on the set

$$S = \left\{ \tilde{x} \in (0,1)^m : \tilde{x} = \big(F(T_{(1)}), \ldots, F(T_{(m)})\big), \text{ for some } F \in \mathcal{F} \right\}$$

by

$$\phi(\tilde{x}) = \psi(F), \text{ if } \tilde{x} = \big(F(T_{(1)}), \ldots, F(T_{(m)})\big),$$

it is seen that the maximization problem boils down to the problem of maximizing $\phi(\tilde{x})$ over the set S. Moreover, if $\tilde{x} = \big(F(T_{(1)}), \ldots, F(T_{(m)})\big)$, we have

$$\frac{\partial}{\partial x_i} \phi(\tilde{x}) = W_F(T_{(i)}) - W_F(T_{(i-1)}), \; i = 1, \ldots, m,$$

where $T_{(0)} \stackrel{\text{def}}{=} 0$. Hence, if $\tilde{y} = \big(\hat{F}_n(T_{(1)}), \ldots, \hat{F}_n(T_{(m)})\big)$, we get, since ϕ is concave

$$\psi(F) - \psi(\hat{F}_n) = \phi(\tilde{x}) - \phi(\tilde{y}) \leq \langle \nabla \phi(\tilde{y}), \tilde{x} - \tilde{y} \rangle$$
$$= \int (F(t) - \hat{F}_n(t))\, dW_{\hat{F}_n}(t) = \int F(t)\, dW_{\hat{F}_n}(t),$$

using (1.27) in the last equality. Since $\tilde{x}$ can be represented as

$$\tilde{x} = \sum_{i=1}^{m} \alpha_i \tilde{1}_i,$$

where α_i and $\tilde{1}_i$ are defined as in (1.12), we get

$$\int F(t)\, dW_{\hat{F}_n}(t) = \sum_{i=1}^{m} \alpha_i \langle \nabla \phi(\tilde{y}), \tilde{1}_i \rangle = \sum_{i=1}^{m} \alpha_i \int_{[T_{(i)}, \infty)} dW_{\hat{F}_n}(t) \leq 0.$$

Conversely, suppose that $\tilde{y} = \big(\hat{F}_n(T_{(1)}), \ldots, \hat{F}_n(T_{(m)})\big)$ and that $\hat{F}_n$ maximizes (1.21) over $\mathcal{F}$. Then

$$\lim_{\epsilon \downarrow 0} \frac{\phi(\tilde{y} + \epsilon \tilde{1}_i) - \phi(\tilde{y})}{\epsilon} = \int_{[T_{(m-i+1)}, \infty)} dW_{\hat{F}_n}(t) \leq 0 \qquad , 1 \leq i \leq m.$$

11

and

$$\lim_{h \to 0} \frac{\phi(\tilde{y} + h\tilde{y}) - \phi(\tilde{y})}{h} = \int \hat{F}_n(t)\, dW_{\hat{F}_n}(t) = 0.$$

Finally, we get by a Taylor expansion with a Lagrangian remainder term

$$\phi(\tilde{x}) - \phi(\tilde{y}) - \langle \nabla \phi(\tilde{y}), \tilde{x} - \tilde{y} \rangle$$

$$= -\frac{1}{2n} \sum_{i=1}^{n} p_i \Big\{ \big(F(T_i) - \hat{F}_n(T_i)\big)^2 \{X_i \le T_i\}$$

$$+ \big\{ F(U_i) - F(T_i) - (\hat{F}_n(U_i) - \hat{F}_n(T_i)) \big\}^2 \{T_i < X_i \le U_i\}$$

$$+ \big(F(U_i) - \hat{F}_n(T_i)\big)^2 \{X_i > U_i\} \Big\},$$

(1.28)

where $p_i > 0$, $1 \le i \le n$. If, for example $\{X_i \le T_i\} = 1$, then $p_i = 1/z_i^2$, where z_i is a point in the open interval with endpoints x_i and y_i, if $x_i \ne y_i$, and where $z_i = y_i$, if $x_i = y_i$. The values $\hat{F}_n(T_{(i)})$ are "linked" to each other in terms of the form $n^{-1} \log\{\hat{F}_n(U_i) - \hat{F}_n(T_i)\}$ and $n^{-1} \log\{\hat{F}_n(U_j) - \hat{F}_n(T_j)\}$, with $U_i < T_j$ and $\hat{F}_n(U_i) = \hat{F}_n(T_j)$ (see Exercise 7 and Examples 1.2 and 1.3 below). Moreover, by (1.28) and the assumptions $\{X_i \le T_{(1)}\} = 1$ and $\{X_j > T_{(m)}\} = 1$ for some indices i and j, we have $F(T_{(1)}) = \hat{F}_n(T_{(1)})$ and $F(T_{(m)}) = \hat{F}_n(T_{(m)})$. From this it is easily checked that the right-hand side of (1.28) can only be zero if $F(T_{(i)}) = \hat{F}_n(T_{(i)})$, for $i = 1, \ldots, m$. This proves the uniqueness. ∎

Example 1.2. Consider the following maximization problem. Maximize

$$\phi(x_1, \ldots, x_5) = \log x_1 + \log(x_4 - x_2) + \log(1 - x_3) + \log(1 - x_5),$$

under the restriction

$$0 \le x_1 \le \cdots \le x_5 \le 1.$$

This corresponds to an interval censoring problem, case 2, where $n = 4$ and

$$X_1 \le T_1,\ T_2 < X_2 \le U_2,\ X_3 > U_3,\ X_4 > U_4,$$

$$T_{(1)} = T_1,\ T_{(2)} = T_2,\ T_{(3)} = U_3,\ T_{(4)} = U_2,\ \text{and}\ T_{(5)} = U_4.$$

Let

$$\tilde{y} = (y_1, \ldots, y_5) = \big(\hat{F}_n(T_{(1)}), \ldots, \hat{F}_n(T_{(5)})\big).$$

Then (1.27) yields:

$$2 - \frac{y_3}{1 - y_3} - \frac{y_5}{1 - y_5} = 0,$$

and (1.26) yields:

$$-\frac{1}{1 - y_5} \le 0, \quad \frac{1}{y_4 - y_2} - \frac{1}{1 - y_5} \le 0, \quad -\frac{1}{1 - y_3} + \frac{1}{y_4 - y_2} - \frac{1}{1 - y_5} \le 0,$$

$$-\frac{1}{1 - y_3} - \frac{1}{1 - y_5} \le 0, \quad \frac{1}{y_1} - \frac{1}{1 - y_3} - \frac{1}{1 - y_5} \le 0.$$

These equations and inequalities are satisfied for

$$y_1 = y_2 = y_3 = \tfrac{1}{4}, \; y_4 = y_5 = \tfrac{5}{8}.$$

The following example shows that the solution vector may have irrational components.

Example 1.3. Consider maximizing

$$\phi(x_1, \ldots, x_{12}) = \log x_1 + \log(x_4 - x_2) + \log(1 - x_3) + \log x_5 + \log(1 - x_6)$$
$$+ \log x_8 + \log(1 - x_9) + \log x_{10} + \log(x_{11} - x_7) + \log(1 - x_{12}),$$

under the restriction

$$0 \le x_1 \le \cdots \le x_{12} \le 1.$$

This would correspond to an interval censoring problem with $n = 10$ and

$$\delta_i = 1, \; \text{if } i = 1, 4, 7 \text{ and } 9,$$
$$\gamma_i = 1, \; \text{if } i = 2 \text{ and } 6,$$
$$\gamma_i = \delta_i = 0, \; \text{otherwise.}$$

The set of (relevant) observation times is ordered as follows.

$$T_1 < T_2 < U_3 < U_2 < T_4 < U_5 < T_6 < T_7 < U_8 < T_9 < U_6 < U_{10}.$$

In this case the NPMLE $\hat{F}_n$ (with $n = 10$) and the maximizing vector $\tilde{y}$ are given by

$$\hat{F}_n(T_{(i)}) = y_i = \begin{cases} \tfrac{1}{2} - \tfrac{1}{6}\sqrt{3}, & 1 \le i \le 3, \\ \tfrac{1}{2}, & 4 \le i \le 9, \\ \tfrac{1}{2} + \tfrac{1}{6}\sqrt{3}, & 10 \le i \le 12. \end{cases}$$

We will now give a characterization of the NPMLE as the left derivative of the convex minorant of a cumulative sum diagram with "self-induced weights". We start by introducing the processes on which this cumulative sum diagram is based.

Let the processes G_F and V_F be defined by

$$
\begin{aligned}
G_F(t) = \; & \int_{t' \in [0,t], \, x \le t'} F(t')^{-2} dP_n(x, t', u) \\
& + \int_{t' \in [0,t], \, t' < x \le u} \{F(u) - F(t')\}^{-2} dP_n(x, t', u) \\
& + \int_{u \in [0,t], \, t' < x \le u} \{F(u) - F(t')\}^{-2} dP_n(x, t', u), \\
& + \int_{u \in [0,t], \, x > u} \{1 - F(u)\}^{-2} dP_n(x, t', u),
\end{aligned}
$$

(1.29)

$$\text{for } t \ge 0,$$

13

and

$$(1.30) \qquad V_F(t) = W_F(t) + \int_{[0,t]} F(t') \, dG_F(t'), \ t \geq 0.$$

The following proposition characterizes the NPMLE $\hat{F}_n$ as the slope of the convex minorant of a self-induced cumulative sum diagram.

PROPOSITION 1.4. *Let $T_{(1)}$ correspond to an observation point T_i such that $\{X_i \leq T_i\} = 1$, and let the largest order statistic $T_{(m)} \in J_n$ correspond to an observation point U_i such that $\{X_i > U_i\} = 1$. Then $\hat{F}_n$ is the NPMLE of F_0 if and only if $\hat{F}_n$ is the left derivative of the convex minorant of the "cumulative sum diagram", consisting of the points*

$$P_j = \left(G_{\hat{F}_n}(T_{(j)}), V_{\hat{F}_n}(T_{(j)}) \right),$$

where $P_0 = (0,0)$ and $T_{(j)} \in J_n$, $j = 1, 2, \dots$

PROOF: For simplicity of notation, we will write V_n, W_n and G_n instead of $V_{\hat{F}_n}$, $W_{\hat{F}_n}$ and $G_{\hat{F}_n}$, respectively.

By definition, the left derivative h_n of the convex minorant of the cumulative sum diagram is given by

$$h_n(\tau_i) = \frac{V_n(\tau_i) - V_n(\tau_{i-1})}{G_n(\tau_i) - G_n(\tau_{i-1})}$$

at the successive locations τ_i of the vertices of the convex minorant of the cumulative sum diagram. Moreover, defining $T_{(0)} = 0$ and

$$\Delta V_{n,i} = V_n(T_{(i)}) - V_n(T_{(i-1)}), \ \text{and} \ \Delta G_{n,i} = G_n(T_{(i)}) - G_n(T_{(i-1)}),$$

we have that h_n minimizes

$$\sum_{T_{(i)} \in J_n} \left\{ h(T_{(i)}) - \frac{\Delta V_{n,i}}{\Delta G_{n,i}} \right\}^2 \Delta G_{n,i}$$

over all nondecreasing functions h, such that $h(0) = 0$. This means by Theorem 1.5, p. 28 in Barlow et al. (1972) or Theorem 1.3.2 in Robertson et al. (1988) that

$$(1.31) \qquad \sum_i \left\{ \frac{\Delta V_{n,i}}{\Delta G_{n,i}} - h_n(T_{(i)}) \right\} h(T_{(i)}) \Delta G_{n,i} \leq 0,$$

and

$$(1.32) \qquad \sum_i \left\{ \frac{\Delta V_{n,i}}{\Delta G_{n,i}} - h_n(T_{(i)}) \right\} h_n(T_{(i)}) \Delta G_{n,i} = 0,$$

14

for all nondecreasing h, such that $h(0) = 0$. But it is easily verified that (1.31) implies (1.26) and that (1.32) implies (1.27), with $\hat{F}_n$ replaced by h_n. Proposition 1.3 now implies that h_n is the NPMLE.

Conversely, if $\hat{F}_n$ satisfies (1.26) and (1.27), then $\hat{F}_n$ also satisfies (1.31) and (1.32), with h_n replaced by $\hat{F}_n$. This in turn implies that $\hat{F}_n$ is the left derivative of the convex minorant of the cumulative sum diagram. ∎

We will see in Chapter 3 that Proposition 1.4 actually leads to an iterative convex minorant algorithm for computing the NPMLE, which seems to converge much faster than the EM algorithm.

Remark 1.4. We note that the "weight function" G_F could also be chosen in another way, still yielding the same characterization of the NPMLE as the left derivative of the convex minorant of a self-induced cumulative sum diagram. The particular choice of G_F, made here, is based on a second order expansion of the log likelihood function. We will return to this point in Chapter 3.

1.2. Exercises.

1. Show that the value of the left derivative of the function H^* in Proposition 1.2 is given by
$$h^*(m) = \max_{i \le m} \min_{k \ge m} \frac{\sum_{i \le j \le k} \delta_{(j)}}{k - i + 1}.$$

2. Show that the following definitions of h^* are equivalent to the definition, given in Exercise 1:
$$h^*(m) = \min_{k \ge m} \max_{i \le m} \frac{\sum_{i \le j \le k} \delta_{(j)}}{k - i + 1}$$
$$h^*(m) = \max_{i \le m} \min_{k \ge i} \frac{\sum_{i \le j \le k} \delta_{(j)}}{k - i + 1}$$
$$h^*(m) = \min_{k \ge m} \max_{i \le k} \frac{\sum_{i \le j \le k} \delta_{(j)}}{k - i + 1}.$$

3. Show that the function ϕ, defined in Example 1.3, is maximized by taking
$$x_i = \begin{cases} \frac{1}{2} - \frac{1}{6}\sqrt{3}, & 1 \le i \le 3, \\ \frac{1}{2}, & 4 \le i \le 9, \\ \frac{1}{2} + \frac{1}{6}\sqrt{3}, & 10 \le i \le 12. \end{cases}$$

4. Can condition (1.26) in Proposition 1.3 be replaced by the following condition?

(1.26*) $$\int_{[t,\infty)} \hat{F}_n(u)\, dW_{\hat{F}_n}(u) \le 0, \quad \forall t \ge 0.$$

5. Can condition (1.27) in Proposition 1.3 be replaced by the following condition?

$$(1.27^*) \qquad\qquad \int dW_{\hat{F}_n}(t) = 0.$$

(The following exercise gives an alternative proof of Proposition 1.2.)

6. Let the function $H : [0, n] \to \mathbb{R}$ be defined by

$$(E.1) \qquad\qquad H(i) = \sum_{j=1}^{i} \delta_{(j)}, \ 1 \le i \le n, \ H(0) = 0,$$

where $\delta_{(j)}$ is defined as in Proposition 1.2, and where H is defined by linear interpolation at values $x \in [0, n] \backslash \{0, \ldots, n\}$. We want to maximize the expression

$$\int_{[0,n]} \log y(x) \, dH(x) + \int_{[0,n]} \log\{1 - y(x)\} \, d(x - H(x)),$$

over all nondecreasing functions $y : [0, n] \to [0, 1]$, such that

$$(E.2) \qquad\qquad y(x) = y(i), \ x \in (i - 1, i], \ 1 \le i \le n, \ y(0) = 0.$$

Let H^* be the convex minorant of H on $[0, n]$, i.e.,

$$H^*(x) = \max\{f(x) : f : [0, n] \to \mathbb{R} \text{ is convex and } f \le H\}, \ x \in [0, n].$$

(a) Show that

$$(E.3) \qquad\qquad \int_{[0,n]} \log y(x) \, dH(x) \le \int_{[0,n]} \log y(x) \, dH^*(x),$$

and

$$(E.4) \qquad \int_{[0,n]} \log\{1 - y(x)\} \, d(x - H(x)) \le \int_{[0,n]} \log\{1 - y(x)\} \, d(x - H^*(x)),$$

for all nondecreasing functions $y : [0, n] \to [0, 1]$, satisfying (E.2) (defining $0 \cdot (-\infty) = 0$, $a \cdot (-\infty) = -\infty$, if $a > 0$).

(b) Let h^* be the left derivative of H^*. Show that

$$\int_{[0,n]} \log h^*(x) \, dH(x) + \int_{[0,n]} \log\{1 - h^*(x)\} \, d(x - H(x))$$

$$= \int_{[0,n]} \log h^*(x) \, dH^*(x) + \int_{[0,n]} \log\{1 - h^*(x)\} \, d(x - H^*(x)).$$

(c) Show that

$$\int_{[0,n]} \log\{h(x)/h^*(x)\} \, dH^*(x) + \int_{[0,n]} \log\{(1 - h(x))/(1 - h^*(x))\} d(x - H^*(x)) \le 0,$$

for each non-decreasing function $h : [0, n] \to [0, 1]$.

16

(d) Deduce Proposition 1.2 from (a), (b) and (c).

7. Show that the function ϕ, defined by

$$\phi(\tilde{x}) = \psi(F), \ \text{if} \ \tilde{x} = \big(F(T_{(1)}), \ldots, F(T_{(m)})\big),$$

in the proof of Proposition 1.3, is generally not *strictly* concave. Show that one can always make a preliminary reduction, "linking" components of the vector $\tilde{x}$, such that we get a strictly concave function on a (convex subset of a) space of lower dimension. For example, in Example 1.2, we can immediately make the following reduction:

$$x_1 = x_2 = x_3; \quad x_4 = x_5.$$

So in this case the function ϕ reduces to a function $\bar{\phi}$ of two variables:

$$\bar{\phi}(x_1, x_2) = \log x_1 + \log(x_2 - x_1) + \log(1 - x_1) + \log(1 - x_2).$$

The function $\bar{\phi}$ is strictly concave on the set $\{(x_1, x_2) : 0 < x_1 < x_2 < 1\}$.

2. THE DECONVOLUTION PROBLEM

2.1. Decreasing densities and non-negative random variables.

We first consider the deconvolution problem in a model with non-negative random variables and disturbances with a decreasing density. Formally, let $Z_1, \ldots, Z_n$ be a sample from a distribution function H with density

$$(2.1) \qquad h(z) = \int g(z - x) \, dF_0(x), \; z \in \mathbb{R},$$

where g is a decreasing density on $[0, \infty)$, and F_0 an unknown distribution function, concentrated on $[0, \infty)$. For example, g could be the exponential density

$$g(x) = e^{-x} 1_{[0,\infty)}(x), \; x \in \mathbb{R},$$

or the Uniform (0,1) density

$$g(x) = 1_{[0,1]}(x), \; x \in \mathbb{R}.$$

An NPMLE of F_0 is a distribution function, maximizing

$$(2.2) \qquad \psi(F) = \int \log \left\{ \int g(z - x) \, dF(x) \right\} dH_n(z),$$

as a function of F, where H_n is the empirical distribution function of the sample $Z_1, \ldots, Z_n$.

In order to define the self-consistency equations, we proceed as before. Defining the direction A_t as in (1.4), we find

$$(2.3) \qquad \lim_{h \downarrow 0} \frac{\psi(F + hA_t) - \psi(F)}{h} = \int \frac{\int g(z - x) \, dA_t(x)}{\int g(z - x) \, dF(x)} \, dH_n(z).$$

Putting (2.3) equal to zero yields the following equation for the NPMLE $\hat{F}_n$:

$$
\begin{aligned}
(2.4) \qquad \hat{F}_n(t) &= \int \frac{\int_{(-\infty, t]} g(z - x) \, d\hat{F}_n(x)}{\int g(z - x) \, d\hat{F}_n(x)} \, dH_n(z) \\
&= n^{-1} \sum_{i=1}^{n} P_{\hat{F}_n} \{ X_i \leq t \mid Z_1, \ldots, Z_n \}.
\end{aligned}
$$

Letting $Z_i = X_i + Y_i$, where $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are independent samples from distributions with distribution function F_0 and density g, respectively, we can write (2.4) as

$$(2.5) \qquad \hat{F}_n(t) = E_{\hat{F}_n} \{ F_n(t) \mid Z_1, \ldots, Z_n \},$$

where F_n is the empirical distribution function of $X_1, \ldots, X_n$. So we again get a self-consistency equation, similar to (1.6) (again not uniquely determining $\hat{F}_n$).

18

Turning to the isotonic regression approach, we first note that we may assume that a distribution function F, maximizing (2.2), is a discrete distribution function, with mass concentrated at the points $Z_1, \ldots, Z_n$. Let $Z_{(1)}, \ldots, Z_{(n)}$ be the order statistics of the sample $Z_1, \ldots, Z_n$. Then we may also assume that the maximizing F satisfies $F(Z_{(1)}) > 0$ and $F(Z_{(n)}) = 1$ (see Exercise 1 of this chapter). In analogy with (1.25), we introduce a process $t \mapsto W_F(t)$, defined by

$$(2.6) \qquad W_F(t) = \sum_{Z_{(i)} \leq t} \Delta W_F(Z_{(i)}),$$

where

$$(2.7) \qquad \begin{aligned} \Delta W_F(Z_{(i)}) &= \int \frac{g(z - Z_{(i)}) - g(z - Z_{(i+1)})}{\int g(z - x)\, dF(x)}\, dH_n(z), \ 1 \leq i < n, \\ \Delta W_F(Z_{(n)}) &= \int \frac{g(z - Z_{(n)})}{\int g(z - x)\, dF(x)}\, dH_n(z). \end{aligned}$$

Let $\mathcal{F}$ be the class of discrete distribution functions, with masses concentrated at the points $Z_1, \ldots, Z_n$, and satisfying $F(Z_{(1)}) > 0$. We get the following proposition, analogous to Proposition 1.3.

PROPOSITION 2.1. *The distribution function $\hat{F}_n \in \mathcal{F}$ maximizes (2.2) over $\mathcal{F}$ if and only if*

$$(2.8) \qquad \int_{[0,t]} dW_{\hat{F}_n}(t') \geq 0, \qquad \forall t \geq 0,$$

and

$$(2.9) \qquad \int \left\{ \hat{F}_n(t) - 1 \right\} dW_{\hat{F}_n}(t) = 0,$$

where W_F is defined by (2.6). Moreover, $\hat{F}_n \in \mathcal{F}$ is uniquely determined by (2.8) and (2.9).

PROOF: We proceed in a similar way as in the proof of Proposition 1.3. Define the function ϕ on the set

$$S = \left\{ \tilde{x} \in [0,1)^n : \tilde{x} = (1 - F(Z_{(n)}), \ldots, 1 - F(Z_{(1)})), \text{ for some } F \in \mathcal{F} \right\}$$

by

$$\phi(\tilde{x}) = \psi(F), \text{ if } \tilde{x}_i = 1 - F(Z_{(n-i+1)}), \ 1 \leq i \leq n.$$

Then
$$\frac{\partial}{\partial x_i}\phi(\tilde{x}) = -\Delta W_F(Z_{(n-i+1)}), \, i = 1, \ldots, n,$$

where $\Delta W_F(Z_{(i)})$ is defined by (2.7). Note that ϕ is well defined, since, by definition, $F(Z_{(1)}) > 0$, if $F \in \mathcal{F}$.

Hence, if $\tilde{y} = (1 - \hat{F}_n(Z_{(n)}), \ldots, 1 - \hat{F}_n(Z_{(1)}))$, we get, since ϕ is concave

$$\psi(F) - \psi(\hat{F}_n) = \phi(\tilde{x}) - \phi(\tilde{y}) \leq \langle \nabla\phi(\tilde{y}), \tilde{x} - \tilde{y} \rangle$$
$$= \int \{(1 - \hat{F}_n(t)) - (1 - F(t))\} \, dW_{\hat{F}_n}(t) = - \int (1 - F(t)) \, dW_{\hat{F}_n}(t) \leq 0,$$

using (2.9) in the last equality, and (2.8) in the last inequality. The remaining part of the proof is analogous to the proof of Proposition 1.3, except for the proof of the uniqueness.

For the proof of the uniqueness, let A be the matrix of second derivatives of the function ϕ at $\tilde{x}$, where $\tilde{x} = (1 - F(Z_{(n)}), \ldots, 1 - F(Z_{(1)}))$, for some $F \in \mathcal{F}$.

Then we get, if $\tilde{u} \in I\!\!R^n$:

$$(2.10) \qquad \tilde{u}'A\tilde{u} = -n^{-1} \sum_{j=1}^{n} \beta_j^{-2} \Big\{ \sum_{i \leq j} \alpha_{ij} u_{n-i+1} \Big\}^2,$$

where

$$\beta_j = \int g(Z_{(j)} - x) \, dF(x),$$
$$\alpha_{ji} = \alpha_{ij} = g(Z_{(j)} - Z_{(i)}) - g(Z_{(j)} - Z_{(i+1)}), \, i < j \leq n,$$

and

$$\alpha_{ii} = g(0), \, 1 \leq i \leq n.$$

It is seen from (2.10) that $\tilde{u}'A\tilde{u}$ can only be zero if $\tilde{u} = 0$. Hence A is non-singular, implying that ϕ is strictly concave. ∎

Remark 2.1. The reason for considering the vector $\tilde{x} = (1 - F(Z_{(n)}), \ldots, 1 - F(Z_{(1)}))$ in the proof of Proposition 2.1, instead of the vector $(F(Z_{(1)}), \ldots, F(Z_{(n)}))$, is the fact that we then only have to consider a maximization problem over the whole *cone* $\{\tilde{x} : 0 \leq x_1 \leq \cdots \leq x_n\}$ instead of a maximization problem over the more complicated bounded convex region $\{\tilde{x} : 0 < x_1 \leq \cdots \leq x_n \leq 1\}$ (defining the function to be $-\infty$ if $x_{(n)} \geq 1$).
We also note that the NPMLE $\hat{F}_n$ always satisfies $\hat{F}_n(T_{(n)}) = 1$, but that we do not have to build this into the conditions of Proposition 2.1.

Although in some cases the NPMLE can be found by a 1-step procedure, just by computing the slope of the convex minorant of a certain function (see Exercises 2 and 3 for deconvolution with, respectively, the uniform and the exponential densities as convolution kernels), this does not seem to be possible in general. However, we can (as in the case

of interval censoring) give a characterization of the NPMLE as the left derivative of the convex minorant of a cumulative sum diagram with self-induced weights. This leads to an iterative convex minorant algorithm, which (in our experience) converges rather fast and will be discussed in Chapter 3.

Let the process G_F be defined by

$$(2.11) \qquad G_F(t) = \sum_{Z_{(i)} \leq t} \Delta G_F(Z_{(i)}),$$

where

$$(2.12) \qquad \begin{aligned} \Delta G_F(Z_{(i)}) &= \int \frac{\left\{ g(z - Z_{(i)}) - g(z - Z_{(i+1)}) \right\}^2}{\left\{ \int g(z - x) \, dF(x) \right\}^2} \, dH_n(z), \quad 1 \leq i < n, \\ \Delta G_F(Z_{(n)}) &= \int \frac{g(z - Z_{(n)})^2}{\left\{ \int g(z - x) \, dF(x) \right\}^2} \, dH_n(z), \end{aligned}$$

and let the process V_F be defined by

$$(2.13) \qquad V_F(t) = W_F(t) + \int_{[0,t]} F(t') \, dG_F(t'), \quad t \geq 0.$$

The following proposition, analogous to Proposition 1.4, characterizes the NPMLE $\hat{F}_n$ as the slope of the convex minorant of a "self-induced" cumulative sum diagram.

PROPOSITION 2.2. *The distribution function $\hat{F}_n \in \mathcal{F}$ maximizes (2.2) over $\mathcal{F}$ if and only if $\hat{F}_n$ satisfies*

$$\hat{F}_n(T_{(i)}) = c_n(T_{(i)}) \wedge 1, \quad 1 \leq i \leq n,$$

where c_n is the left derivative of the convex minorant of the "cumulative sum diagram", consisting of the points

$$P_j = \left(G_{\hat{F}_n}(Z_{(j)}), V_{\hat{F}_n}(Z_{(j)}) \right), \quad 0 \leq j \leq n,$$

and where $P_0 = (0,0)$.

The proof is similar to the proof of Proposition 1.4 and is therefore omitted. The process G_F contains "second derivatives on the diagonal" of the log likelihood function, but has again a certain arbitrariness in the sense that other weight functions would give the same type of characterization in Proposition 2.2. However, computer experiments show a superiority of this weight function with respect to certain other possibilities.

Remark 2.2. The results of this section can be generalized to the situation where g is non-decreasing on an interval which may be different from $[0, \infty)$ and where F_0 need not be zero on $(-\infty, 0)$, but we will not go into this here.

2.2. Convolution with symmetric densities.

The deconvolution problem with symmetric kernels offers some new features, which even make the computation of the NPMLE more difficult. The most important difference is that, unlike in the situation discussed in section 2.1, we generally cannot assume that the NPMLE corresponds to a probability distribution, with mass concentrated on the observation points.

We will assume that the convolution kernel g satisfies the following conditions:

(2.14) $\qquad$ g is symmetric about the origin, i.e., $g(x) = g(-x)$, $x \in \mathbb{R}$

and

(2.15) $\qquad$ g is continuous and decreases on $\mathbb{R}_+$.

Examples of such densities are: the standard normal, the Cauchy, the Laplace (or double exponential) and the triangular densities. Again we assume that the sample of "observables" is $Z_1, \ldots, Z_n$ where Z_i has density

$$(2.16) \qquad h(z) = \int g(z - x)\, dF_0(x), \ z \in \mathbb{R},$$

There is a voluminous literature on this problem and on methods for estimating the density f_0, corresponding to the distribution function F_0, see, e.g., Carroll and Hall (1988), Fan (1988), Stefanski and Carroll (1987) and Zhang (1990). In these papers the estimation method is invariably based on Fourier inversion. Almost nothing seems to be known about the behavior of the NPMLE (we will, however, establish consistency of the NPMLE in Chapter 4).

The self-consistency equation (2.5) applies without change, and can actually be used in the EM algorithm. However, the convergence of the EM algorithm is in this case so painfully slow that it almost seems useless for practical purposes. Below some information about the general structure of the NPMLE is listed. The line of argument is similar to arguments used by Jewell (1982) in characterizing the NPMLE of the mixing distribution in scale mixtures of exponential distributions.

It will first be shown that an NPMLE (i.e., a distribution function $\hat{F}_n$, maximizing $\psi(F)$, defined by (2.2), with g satisfying (2.14) and (2.15)) always exists.

LEMMA 2.1. *There always exists an NPMLE. Moreover, the vector*

$$\left(\int g(Z_{(1)} - x)\, d\hat{F}_n(x), \ldots, \int g(Z_{(n)} - x)\, d\hat{F}_n(x) \right)$$

has the same value for each NPMLE $\hat{F}_n$.

PROOF: Let the function $\tilde{x} \mapsto k(\tilde{x})$, $\tilde{x} \in [0,1]^n$ be defined by

$$k(\tilde{x}) = \begin{cases} n^{-1} \sum_{i=1}^{n} \log x_i & , \ x_i \in (0,1],\ 1 \le i \le n \\ -\infty & , \ \text{if } x_i = 0,\ \text{for some } i. \end{cases}$$

Moreover, let $\mathcal{F}$ be the set of subdistribution functions on $\mathbb{R}$, and let the function $\chi : \mathcal{F} \to [0,1]^n$ be defined by

$$\chi(F) = \left(\int g(Z_{(1)} - x)\, dF(x), \dots, \int g(Z_{(n)} - x)\, dF(x) \right), \quad F \in \mathcal{F}.$$

The set $M = \{\chi(F) : F \in \mathcal{F}\}$ is convex and compact (in the vague topology) and hence there exists a subdistribution function $\bar{F} \in \mathcal{F}$ such that

$$k\big(\chi(\bar{F})\big) = \sup_{\tilde{x} \in M} k(\tilde{x}).$$

Since the set $\{\chi(F) : F \in \mathcal{F}, \chi(F) \in (0,1]^n\}$ is nonempty, we may assume $\chi(\bar{F}) \in (0,1]^n$, since otherwise $k\big(\chi(\bar{F})\big) = -\infty$. Moreover, since k is strictly concave on $(0,1]^n$, there is a *unique* $\tilde{y} \in (0,1]^n$ such that

$$k(\tilde{y}) = \sup_{F \in \mathcal{F}} k\big(\chi(F)\big) = \sup_{\tilde{x} \in M} k(\tilde{x}).$$

Finally, if $\bar{F}$ maximizes $k\big(\chi(F)\big)$, then $\bar{F}$ has to be a distribution function, since otherwise there would exist an $\epsilon > 0$ such that

$$k\left(\chi\left(\bar{F} + \epsilon 1_{[0,\infty)} \right) \right) > k\big(\chi(\bar{F})\big). \qquad \blacksquare$$

The following proposition yields some information about the set on which the NPMLE is concentrated.

PROPOSITION 2.3. *Let $P_{\hat{F}_n}$ the probability measure on $\mathbb{R}$, corresponding to an NPMLE $\hat{F}_n$, and let the set M be defined by*

$$M = \left\{ y \in \mathbb{R} : \int \left\{ g(z - y) \Big/ \int g(z - x)\, d\hat{F}_n(x) \right\} dH_n(z) = 1 \right\}.$$

Then $P_{\hat{F}_n}(M) = 1$. Moreover,

$$\int \frac{g(z - y)}{\int g(z - x)\, d\hat{F}_n(x)}\, dH_n(z) < 1, \quad \text{if } y \notin M.$$

PROOF: Fix $x \in \mathbb{R}$. Then

$$\lim_{h \downarrow 0} h^{-1} \left\{ \psi\big(\hat{F}_n + h(1_{[x,\infty)} - \hat{F}_n)\big) - \psi(\hat{F}_n) \right\}$$

$$(2.17) \qquad = \int \frac{g(z - x) - \int g(z - y)\, d\hat{F}_n(y)}{\int g(z - y)\, d\hat{F}_n(y)}\, dH_n(z)$$

$$= \int \frac{g(z - x)}{\int g(z - y)\, d\hat{F}_n(y)}\, dH_n(z) - 1 \leq 0,$$

since $\hat{F}_n$ is an NPMLE. By integrating with respect to the measure $P_{\hat{F}_n}$ (and by Fubini's theorem) we get

$$(2.18) \qquad \int_{x \in \mathbb{R}} \left\{ \int_{z \in \mathbb{R}} \frac{g(z-x)}{\int g(z-y)\, d\hat{F}_n(y)} \, dH_n(z) \right\} d\hat{F}_n(x) = 1.$$

So we must have $P_{\hat{F}_n}(M) = 1$, since otherwise, by (2.17), the left side of (2.18) would be strictly smaller than 1. ∎

Proposition 2.3 shows that in many cases the support of $P_{\hat{F}_n}$ is a finite set. For example, if g is a normal density we get the following result.

COROLLARY 2.1. *Let g be a normal density, symmetric about zero. Then $P_{\hat{F}_n}$ is concentrated on a finite set of at most n points.*

PROOF: Without loss of generality we may assume that g is the standard normal density. Proposition 2.3 implies

$$\int \left\{ \exp\{-\tfrac{1}{2}(z-y)^2\} \Big/ \int \exp\{-\tfrac{1}{2}(z-x)^2\} \, d\hat{F}_n(x) \right\} dH_n(z) = 1,$$

if $y \in M$. This means that y satisfies an equation of the form

$$(2.19) \qquad \sum_{i=1}^{n} \alpha_i \exp\{-\tfrac{1}{2}(Z_{(i)} - y)^2\} - 1 = 0,$$

with $\alpha_i > 0$, $1 \le i \le n$. By Karlin and Studden (1966), pp. 9-11, Examples 1 and 5, this equation can have at most n roots. ∎

Remark 2.3. It is clear that Corollary 2.1 also holds for normal densities which are not symmetric about zero.

As a consequence of the following corollary, we get a stronger result in the case of the double exponential density.

COROLLARY 2.2. *Let g be strictly convex on $[0, \infty)$. Then $P_{\hat{F}_n}$ is concentrated on the set of observation points.*

PROOF: Let $\alpha_i = 1 \big/ \int g(Z_{(i)} - x) \, d\hat{F}_n(x)$, $1 \le i \le n$. Then the function

$$x \mapsto n^{-1} \sum_{i=1}^{n} \alpha_i g(x - Z_{(i)}), \ x \in \mathbb{R},$$

24

is strictly convex on each interval $\big(Z_{(i-1)}, Z_{(i)}\big)$, $1 \le i \le n+1$, where $Z_{(0)} = -\infty$ and $Z_{(n+1)} = \infty$. Thus, if the set M is defined as in Proposition 2.3, we get:

$$x \in M \Longrightarrow x = Z_i, \text{ for some } i, 1 \le i \le n,$$

since

$$x \notin \{Z_1, \ldots, Z_n\} \Longrightarrow n^{-1} \sum_{i=1}^{n} \alpha_i g\big(x - Z_{(i)}\big) < 1. \qquad \blacksquare$$

The following Corollary of Proposition 2.3 can also be obtained by direct methods.

COROLLARY 2.3. *Let g satisfy*

$$g(x) < g(0), \text{ for all } x \ne 0.$$

Then $P_{\hat{F}_n}\big[Z_{(1)}, Z_{(n)}\big] = 1$, i.e., $P_{\hat{F}_n}$ is concentrated on the range of the data points.

Example 2.1. (The following curious facts have been communicated to me by Rudolf Grübel.) Let g be the standard normal density. We want to find the support of $\hat{F}_n$ if the sample size is $n = 2$. After a suitable shift we may assume $Z_{(1)} = -a$, $Z_{(2)} = a$. We have:

$$\psi(F) = \tfrac{1}{2}\Big\{\log \int \phi(a - x)\, dF(x) + \log \int \phi(a + x)\, dF(x)\Big\}.$$

If $\tilde{F}$ is defined by $\tilde{F}(x) = 1 - F(-x-)$, where

$$F(-x-) = \lim_{y \uparrow -x,\, y < -x} F(y),$$

we get

$$\psi(\tilde{F}) = \psi(F).$$

Moreover, since ψ is concave, we have

$$\psi\left(\tfrac{1}{2}(F + \tilde{F})\right) \ge \tfrac{1}{2}\{\psi(F) + \psi(\tilde{F})\}.$$

Hence, if $\psi(F)$ is maximal, then $\psi(F^s)$ is also maximal, where F^s is the symmetrized distribution function $F^s = \tfrac{1}{2}\{F + \tilde{F}\}$.

So, in order to find an NPMLE, we may restrict ourselves to symmetric distributions F. For these we have

$$\int_{(-\infty, 0)} \phi(-a - x)\, dF(x) = \int_{(0, \infty)} \phi(-a + x)\, dF(x),$$

and likewise

$$\int_{(-\infty,0)} \phi(a-x)\,dF(x) = \int_{(0,\infty)} \phi(a+x)\,dF(x),$$

implying (by the symmetry of ϕ):

$$\psi(F) = \log\left\{\phi(a)P_F(\{0\}) + \int_{(0,\infty)} \{\phi(a-x) + \phi(a+x)\}\,dF(x)\right\}.$$

Since, by Proposition 2.3, the support of a maximum likelihood estimator consists of at most 2 points, we either get the degenerate distribution at zero, or a (symmetric) distribution concentrated on the two points y and $-y$, where $y > 0$ satisfies

$$\phi(a-y) + \phi(a+y) = \sup_{x>0}\{\phi(a-x) + \phi(a+x)\}.$$

We have:

$$\frac{d}{dy}\{\phi(a-y) + \phi(a+y)\} = 0 \iff \frac{a-y}{a+y} = \frac{\phi(a+x)}{\phi(a-x)}.$$

For $u = y/a$ this yields

(2.20)
$$\frac{1-u}{1+u} = \exp\{-2a^2u\}.$$

As a function of u, the left hand side of (2.20) is strictly decreasing to -1, as $u \to \infty$, and the right hand side of (2.20) tends to 0 in this case. Both are differentiable with strictly increasing derivatives, and take the value 1 at $u = 0$. Hence another point of intersection exists if and only if the derivative of the left hand side is strictly smaller than the corresponding quantity for the right hand side, i.e., if and only if $a > 1$.

Putting all this together, we get that generally the NPMLE will be the degenerate distribution at the mean $\bar{z} = \frac{1}{2}\{z_1 + z_2\}$ of the two observations if $a = \frac{1}{2}\{z_{(2)} - z_{(1)}\} \le 1$ and will "split" into two masses of size $1/2$ each, located at the points $\bar{z} \pm ac(a)$, where $c(a) = u$ is the (unique) positive solution of (2.20). Some values of $c(a)$ are:

$$c(1.01) = 0.24121, \quad c(1.1) = 0.66993, \quad c(2) = 0.99933.$$

A similar situation arises in the Cauchy and analogous cases. Generally, the "bifurcation distance" of the two data points will be $2u$, where u separates the convex and concave part of the symmetric density g.

2.3. Exercises.

1. Show that, under the conditions of section 2.1, there exists an NPMLE $\hat{F}_n$ with mass concentrated on the set of observation points, and satisfying $\hat{F}_n(Z_{(1)}) > 0$ and $\hat{F}_n(Z_{(n)}) = 1$. Consider in particular the case $g = 1_{[0,1]}$ (i.e., g is the Uniform $(0,1)$ density). Is the NPMLE always uniquely determined, if we drop the condition that its mass is concentrated on the set of observation points?

2. Let g be the Uniform $(0,1)$ density, and suppose that we know that the unknown distribution function F_0 is also concentrated on $[0,1]$. Defining $\delta_i = 1_{\{Z_i \leq 1\}}$, the function ψ, defined by (2.2), can be written

$$\psi(F) = n^{-1} \sum_{i=1}^{n} \left\{ \delta_i \log F(Z_i) + (1 - \delta_i) \log(1 - F(Z_i - 1)) \right\}.$$

(a) Let $Y_i, 1 \leq i \leq n$, be defined by

$$Y_i = \begin{cases} Z_i & , \text{ if } \delta_i = 1, \\ Z_i - 1 & , \text{ if } \delta_i = 0. \end{cases}$$

Show that $Y_1, \ldots, Y_n$ is distributed as a sample from a Uniform $(0,1)$ distribution.

(b) Let $Y_{(1)}, \ldots, Y_{(n)}$ be the set of order statistics of the set $Y_1, \ldots, Y_n$, and define

$$\delta_{(i)} = \begin{cases} 1 & , \text{ if the } X_k, \text{ corresponding to } Y_{(i)}, \text{ is } \leq 1, \\ 0 & , \text{ otherwise.} \end{cases}$$

Show that the value at $Y_{(i)}$ of the NPMLE $\hat{F}_n$, maximizing $\psi(F)$ over $\mathcal{F}$, is given by the left continuous derivative at the point i of the convex minorant of the function $K_n : [0, n] \to \mathbb{R}$, defined by

$$K_n(i) = \sum_{j \leq i} \delta_{(j)}, \quad K_n(0) = 0,$$

at the points i, and by linear interpolation at other points of $[0, n]$.

3. (This example is due to R. Grübel. For a different approach, see Vardi (1989).) Let g be the exponential density

$$g(x) = e^{-x} 1_{[0,\infty)}(x), \quad x \in \mathbb{R},$$

and let the function ψ be defined by (2.2). Show that the value at $Z_{(i)}$ of the NPMLE $\hat{F}_n$, maximizing $\psi(F)$ over $\mathcal{F}$ can be found as follows.
Let the points P_i, $0 \leq i \leq n$, be defined by

$$P_i = (x_i, y_i) = \begin{cases} (0,0) & , \text{ if } i = 0, \\ \left(e^{-Z_{(1)}} - e^{-Z_{(i+1)}}, i/n \right) & , \text{ if } 1 \leq i < n, \\ \left(e^{-Z_{(1)}}, 1 \right), & , \text{ if } i = n. \end{cases}$$

Let f_n be the left derivative on $(0, x_n]$ of the convex minorant of the cumulative sum diagram, consisting of the points P_i, and let $f_n(0) = 0$. Then the value at $Z_{(i)}$ of the NPMLE $\hat{F}_n$ is given by

$$\hat{F}_n(Z_{(i)}) = \sum_{j \leq i} e^{-Z_{(j)}} \left\{ f_n(x_j) - f_n(x_{j-1}) \right\}.$$

Hints: Since f_n is the left derivative of the convex minorant of the cumulative sum diagram, consisting of the points P_i, the function $i \mapsto f_n(x_i)$ is the isotonic regression of $u_i \stackrel{\text{def}}{=} 1/\{n(x_i - x_{i-1})\}$, $1 \leq i \leq n$, with weights $\Delta x_i = x_i - x_{i-1}$, i.e., this function minimizes

$$\sum_{i=1}^{n} \{f(i) - u_i\}^2 \Delta x_i,$$

over all functions $i \mapsto f(i)$, which are nondecreasing in i. This implies, by e.g., Robertson et al. (1988), Theorem 1.3.2:

$$\sum_{i=1}^{n} f_n(x_i) \{u_i - f_n(x_i)\} \Delta x_i = 0,$$

and

$$\sum_{j \geq i} \{u_j - f_n(x_j)\} \Delta x_j \leq 0, \ 1 \leq i \leq n.$$

Show that the conditions of Proposition 2.3 follow from these relations.

4. Deduce from Karlin and Studden (1966), pp. 9-11, Examples 1 and 5, that equation (2.19) can have at most n roots.

5. Let g be the triangular density

$$g(x) = (1 - |x|) 1_{[-1,1]}(x), \ x \in I\!R.$$

Can we assume (as in the double exponential case) that the mass of the NPMLE is concentrated on the set of data points?

3. Algorithms

3.1. The EM algorithm.

We illustrate the properties of the EM algorithm by studying its behavior in the interval censoring problem, Case 1. As argued on p. 1 of Chapter 1, we may, in our search for the NPMLE, restrict attention to a class $\mathcal{F}$ of purely discrete distribution functions F with mass concentrated on the set of points $T_{(1)}, \ldots, T_{(n)}, T_{(n+1)}$, where $T_{(n+1)}$ is an arbitrary point $t > T_{(n)}$ (we need a point to put the remaining mass on). Let

$$p_i = P_F\{X = T_{(i)}\}, 1 \le i \le n + 1.$$

The EM algorithm runs as follows. We take a starting distribution $P_{F^{(0)}}$, with positive masses at all points $T_{(i)}$, $1 \le i \le n + 1$. For example, we could take the discrete uniform distribution

$$(3.1) \qquad p_i^{(0)} = P_{F^{(0)}}\{X = T_{(i)}\} = 1/(n + 1), 1 \le i \le n + 1.$$

We then do the "E"-step ("E" for "Expectation"), i.e., we compute the conditional expectation of the log likelihood

$$(3.2) \qquad E^{(0)}\Big\{\sum_{i=1}^{n} \log f(X_i) \mid \delta_1, \ldots, \delta_n, T_1, \ldots, T_n\Big\},$$

where $f(x) = P_F\{X = x\}$ and $E^{(0)}$ is the expectation under the probability measure $P_{F^{(0)}}$. Next we maximize (3.2) over all discrete distribution with probability density f with respect to counting measure on the set $\{T_{(1)}, \ldots, T_{(n+1)}\}$. This yields new probability masses $p_i^{(1)}$, $i = 1, \ldots, n + 1$.

Now we repeat the E and M step, starting with the probability distribution with masses $p_i^{(1)}$ instead of the $p_i^{(0)}$, defined by (3.1), etc.

In order to establish the connection with the self-consistency equations (1.6), we take a closer look at what happens in the E and M steps. We can rewrite (3.2) in the following way:

$$
\begin{aligned}
(3.3) \qquad & E^{(0)}\Big\{\sum_{i=1}^{n} \log f(X_i) \mid \delta_1, \ldots, \delta_n, T_1, \ldots, T_n\Big\} \\
&= \sum_{i=1}^{n} E^{(0)}\big\{\log f(X_i) \mid \delta_1, \ldots, \delta_n, T_1, \ldots, T_n\big\} \\
&= \sum_{i=1}^{n}\Big\{\sum_{k=1}^{n+1} (\log p_k) P^{(0)}\{X_i = T_{(k)} \mid \delta_i, T_i\}\Big\} \\
&= \sum_{k=1}^{n+1} (\log p_k) \sum_{i=1}^{n} P^{(0)}\{X_i = T_{(k)} \mid \delta_i, T_i\},
\end{aligned}
$$

where we write $P^{(0)}$ instead of $P_{F^{(0)}}$ (generally we will write $P^{(m)}$ instead of $P_{F^{(m)}}$). For the M-step, we have to maximize (3.3) over all discrete distributions with masses p_i at $T_{(i)}$. It is easily seen (for example by using Lagrangian multipliers) that (3.3) is maximized by taking

$$p_k = n^{-1} \sum_{i=1}^n P^{(0)}\{X_i = T_{(k)} \mid \delta_i, T_i\}, \, 1 \le k \le n+1.$$

Hence the combined E and M steps yield

$$p_k^{(1)} = n^{-1} \sum_{i=1}^n P^{(0)}\{X_i = T_{(k)} \mid \delta_i, T_i\}, \, 1 \le k \le n+1.$$

Generally, we get

(3.4)
$$p_k^{(m+1)} = n^{-1} \sum_{i=1}^n P^{(m)}\{X_i = T_{(k)} \mid \delta_i, T_i\}, \, 1 \le k \le n+1,$$

which leads to

$$F^{(m+1)}(t) = n^{-1} \sum_{i=1}^n P^{(m)}\{X_i \le t \mid \delta_i, T_i\} = E^{(m)}\{F_n(t) \mid \delta_1, \ldots, \delta_n, T_1, \ldots, T_n\},$$

where F_n is the empirical distribution function of the (unobservable) sample $X_1, \ldots, X_n$. This corresponds to the self-consistency equation (1.6).

One can write (3.4) in the following explicit form:

$$p_k^{(m+1)} = n^{-1} \sum_{i=1}^n \left\{ \frac{\delta_i}{F^{(m)}(T_i)} 1_{\{T_i \ge T_{(k)}\}} + \frac{1 - \delta_i}{1 - F^{(m)}(T_i)} 1_{\{T_i < T_{(k)}\}} \right\} p_k^{(m)}, \, 1 \le k \le n+1,$$

since

$$P^{(m)}\{X_i = T_{(k)} \mid \delta_i, T_i\} = p_k^{(m)} \left\{ \frac{\delta_i}{F^{(m)}(T_i)} 1_{\{T_i \ge T_{(k)}\}} + \frac{1 - \delta_i}{1 - F^{(m)}(T_i)} 1_{\{T_i < T_{(k)}\}} \right\}.$$

It follows that the probability masses $\hat{p}_k$, corresponding to the NPMLE, should satisfy

$$\hat{p}_k = n^{-1} \sum_{i=1}^n \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} 1_{\{T_i \ge T_{(k)}\}} + \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} 1_{\{T_i < T_{(k)}\}} \right\} \hat{p}_k, \, 1 \le k \le n+1.$$

Hence, if $\hat{p}_k > 0$, we get:

(3.5)
$$1 = n^{-1} \sum_{i=1}^n \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} 1_{\{T_i \ge T_{(k)}\}} + \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} 1_{\{T_i < T_{(k)}\}} \right\},$$

and if $\hat{p}_k = 0$, we get

$$(3.6) \qquad 1 \geq n^{-1} \sum_{i=1}^{n} \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} 1_{\{T_i \geq T_{(k)}\}} + \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} 1_{\{T_i < T_{(k)}\}} \right\}.$$

Note that we do not necessarily get (3.5) and (3.6) if we would not start with positive masses at all points $T_{(k)}$! If the starting distribution puts zero masses at some points $T_{(k)}$, the EM algorithm would converge to a solution of the self-consistency equations, but this solution would not necessarily maximize the likelihood. This is caused by the fact that the EM algorithm will, at each iteration step, put mass zero at a point where the starting distribution puts mass zero. So, if the NPMLE puts positive mass at such a point, the EM algorithm will not converge to the NPMLE, if the starting distribution puts zero mass at that point.

There is the general empirical finding that the number of iteration steps, needed by the EM algorithm to get, say, the NPMLE in two accurate decimals, will increase with the sample size. Below we offer some speculations on why this will be true. In order to make things as simple as possible, we consider interval censoring, Case 1, and suppose that F_0 and G are both the uniform distribution function on $[0, 1]$.

The arguments in Wu (1983), which show that the EM algorithm (starting with positive weights at all points $T_{(1)}, \ldots, T_{(n+1)}$ will actually converge to the NPMLE, give no information about the speed of convergence and it is not clear that one has a contraction with a constant strictly smaller than 1. Moreover, it is not clear what norm should be used, if one wants to show that the iteration steps (3.4) really correspond to a contraction. We will now give some arguments which indicate that one indeed has a (local) contraction with respect to a certain L_2-norm (to be specified below), but with a constant which can be arbitrarily close to 1. By "local" we mean that the values of $F^{(m)}$ should not be too far from the values of the solution $\hat{F}_n$. We conjecture that one does not have a (local) contraction with respect to the supremum norm (on the space of distribution functions).

Let E be a linear space of bounded functions on $[0, 1]$, containing differences of (two) distribution functions, and let the mapping $\phi : E \to E$ be defined by

$$(3.7) \qquad [\phi(h)](t) = h(t) - \int_{u \geq t} \frac{t}{u} h(u) \, du - \int_{u < t} \frac{1-t}{1-u} h(u) \, du$$

The mapping ϕ can be considered as an approximate derivative of the mapping, defined by the iteration steps (3.4), since, for n *and* m large, we get

$$F^{(m+1)}(t) - F^{(m)}(t)$$

$$\approx \frac{1}{n} \sum_{T_i \geq t} \frac{\delta_i}{F^{(m)}(T_i)} \left(F^{(m)}(t) - F^{(m-1)}(t) \right) + \frac{1}{n} \sum_{T_i < t} \frac{1 - \delta_i}{1 - F^{(m)}(T_i)} \left(F^{(m)}(t) - F^{(m-1)}(t) \right)$$

$$- \frac{1}{n} \sum_{T_i \geq t} \frac{\delta_i}{F^{(m)}(T_i)^2} F^{(m)}(t) \left(F^{(m)}(T_i) - F^{(m-1)}(T_i) \right)$$

$$- \frac{1}{n} \sum_{T_i < t} \frac{1 - \delta_i}{\left(1 - F^{(m)}(T_i) \right)^2} \left(1 - F^{(m)}(t) \right) \left(F^{(m)}(T_i) - F^{(m-1)}(T_i) \right)$$

$$\approx F^{(m)}(t) - F^{(m-1)}(t) - \int_{u \geq t} \frac{t}{u} \left(F^{(m)}(u) - F^{(m-1)}(u) \right) du$$

$$- \int_{u < t} \frac{1-t}{1-u} \left(F^{(m)}(u) - F^{(m-1)}(u) \right) du,$$

where the approximations are (partly) justified by the (strong or weak) law of large numbers.

Continuing on this path (and leaving lots of problems aside), we obtain the following approximation to the squared L_2-distance between $F^{(m+1)}$ and $F^{(m)}$:

$$\|F^{(m+1)} - F^{(m)}\|^2$$

$$\approx \|F^{(m)} - F^{(m-1)}\|^2 - \tfrac{1}{2} \iint_{[0,1]^2} \left\{ \frac{t \wedge u}{t \vee u} + \frac{1-t \vee u}{1-t \wedge u} \right\}$$

$$\cdot \left(F^{(m)}(t) - F^{(m-1)}(t) \right) \left(F^{(m)}(u) - F^{(m-1)}(u) \right) dt \, du,$$

where the L_2-distance is just the ordinary L_2-distance between (complex- or real-valued) functions on $[0,1]$ with respect to Lebesgue measure. So, in the limit (as the sample size tends to infinity), we have to deal with the mapping $f \mapsto Uf$, defined by

$$[Uf](t) = \tfrac{1}{2} \int_{[0,1]} \left\{ \frac{t \wedge u}{t \vee u} + \frac{1-t \vee u}{1-t \wedge u} \right\} f(u) \, du,$$

which is a compact self-adjoint operator on the pre-Hilbert space of continuous complex-valued functions f, defined on $[0,1]$. Note that the eigenvalues λ of U satisfy $0 \leq \lambda \leq 1$ (as is seen using the Cauchy-Schwarz inequality and the fact that U is self-adjoint). We guess that the operator U has eigenvalues arbitrarily close to zero.

On the other hand, for each n, the operator corresponds to a symmetric matrix with elements

$$\frac{1}{2n} \left\{ \frac{T_i \wedge T_j}{T_i \vee T_j} + \frac{1 - T_i \vee T_j}{1 - T_i \wedge T_j} \right\}.$$

For example, if $n = 2$ and $T_1 = 1/3$, $T_2 = 2/3$ we get the matrix

$$C_1 = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

with eigenvalues $\frac{3}{4}$ and $\frac{1}{4}$, and if $n = 3$ and $T_i = i/4$ we get the matrix

$$C_2 = \begin{pmatrix} \frac{1}{3} & \frac{7}{36} & \frac{1}{9} \\ \frac{7}{36} & \frac{1}{3} & \frac{7}{36} \\ \frac{1}{9} & \frac{7}{36} & \frac{1}{3} \end{pmatrix}$$

The eigenvalues of the matrix $I - C_2$ are: 0.8916, 0.7778 and 0.3306.

This has a rather interesting general structure: the elements of the matrix C_i become smaller if one moves away from the diagonal, corresponding to a decreasing "coverage" of the points T_i and T_j by intervals $[0, T_k]$ and $(T_l, 1]$ in the estimation problem. One gets the impression that the biggest eigenvalue of the matrix $I - C_n$ will tend to 1, as $n \to \infty$, which would mean that the "contraction constant" also tends to 1, as $n \to \infty$. This would explain why the convergence of the EM algorithm becomes slower, if the sample size increases.

Remark 3.1. The following type of example suggests that we cannot expect to have a contraction with respect to the supremum norm. Let $f_\epsilon : [0, 1] \to I\!R$ be defined by

$$f_\epsilon(x) = \begin{cases} -1, & t \in \left[0, \frac{1}{2}\right), \\ 1, & t \in \left[\frac{1}{2}, \frac{1}{2} + \epsilon\right), \\ 0, & t \in \left[\frac{1}{2} + \epsilon, 1\right]. \end{cases}$$

Then:

$$\lim_{\epsilon \downarrow 0} [\phi f_\epsilon] \left(\tfrac{1}{2}\right) = 1 + \int_0^{\frac{1}{2}} \frac{1}{2(1 - u)} = 1 + \tfrac{1}{2} \log 2,$$

where ϕ is defined by (3.7). So, for sufficiently small $\epsilon > 0$, we get $[\phi f_\epsilon](\tfrac{1}{2}) > 1$. This example can easily be changed into a similar example for matrices (showing that the infinite dimensionality of the function space is not essential here).

3.2. The iterative convex minorant algorithm.

We discuss the properties of the iterative convex minorant algorithm for interval censoring, Case 2. The idea is to reduce the maximization problem to a series of weighted isotonic regression problems, where the weights are induced by the values of the solution at the preceding step.

Suppose $F^{(m)}$ is the estimator of F_0, found at the mth iteration step. Then $F^{(m+1)}$ is the distribution function, maximizing the expression

$$(3.8) \qquad \int \left(F(t) - F^{(m)}(t)\right) dW_{F^{(m)}}(t) - \tfrac{1}{2} \int \left(F(t) - F^{(m)}(t)\right)^2 dG_{F^{(m)}}(t),$$

as a function of $F \in \mathcal{F}$, where $W_{F^{(m)}}$ and $G_{F^{(m)}}$ are defined by (1.25) and (1.29), respectively, and $\mathcal{F}$ is the set of purely discrete distribution function with mass concentrated on the set of points $T_{(i)}$ and possibly with mass at an extra point bigger than all observation points (see the beginning of the chapter and Remark 1.1 in Chapter 1). One can consider (3.8) as a second order approximation to $\psi(F) - \psi(F^{(m)})$, where the off-diagonal second order terms are omitted. It can be shown (Exercise 5) that the distribution function $F^{(m+1)}$, maximizing (3.8), is the left derivative of the convex minorant of the cumulative sum diagram, consisting of the points

$$P_j = \left(G_{F^{(m)}}(T_{(j)}), V_{F^{(m)}}(T_{(j)})\right),$$

where $P_0 = (0,0)$ and $T_{(j)} \in J_n$, $j = 1, 2, \ldots$, provided we make a preliminary reduction of the observation points $T_{(i)}$ such that the conditions of Proposition 1.4 are satisfied.

In the iteration steps of the algorithm we use "buffers" to prevent that $F^{(m)}$ would not be a distribution function or that the likelihood at $F^{(m)}$ would become zero. For example, one can always take a small positive number $c > 0$ such that $F^{(m)}(U_i) - F^{(m)}(T_i) \geq c$, for each pair (T_i, U_i) such that $\{T_i < X_i \leq U_i\} = 1$, since for $c > 0$ sufficiently small a term $\log\big(F^{(m)}(U_i) - F^{(m)}(T_i)\big) \leq \log c$ would make the log likelihood smaller than that of the discrete uniform distribution on the set of points (with the extra point bigger than all observation points). Likewise one can deal with values $F^{(m)}(T_i)$ or $1 - F^{(m)}(T_i)$ that would be too small.

In the full second order approximation to $\psi(F) - \psi(F^{(m)})$, we would have to add to (3.8) the term

$$
(3.9) \qquad \int_{t < x \leq u} \frac{\big(F(t) - F^{(m)}(t)\big)\big(F(u) - F^{(m)}(u)\big)}{\big(F^{(m)}(u) - F^{(m)}(t)\big)^2} \, dP_n(x, t, u).
$$

By a simple change of the algorithm, one can actually maximize the full second order approximation to $\psi(F) - \psi(F^{(m)})$. In this modification, the "time scale" $G_{F^{(m)}}$ at step m is the same as in (1.29), but the process W_F is replaced by

$$
\begin{aligned}
W_F^{(m)}(t) &= \int_{t' \in [0,t], \, x \leq t'} F^{(m)}(t')^{-1} dP_n(x, t', u) \\
&\quad + \int_{t' \in [0,t], \, t' < x \leq u} \frac{F(u) - F(t') - 2\{F^{(m)}(u) - F^{(m)}(t')\}}{\{F^{(m)}(u) - F^{(m)}(t')\}^2} \, dP_n(x, t', u) \\
&\quad - \int_{t' \in [0,t], \, t' < x \leq u} \frac{F(u) - F(t') - 2\{F^{(m)}(u) - F^{(m)}(t')\}}{\{F^{(m)}(u) - F^{(m)}(t')\}^2} \, dP_n(x, t', u) \\
&\quad - \int_{u \in [0,t], \, x > u} \{1 - F^{(m)}(u)\}^{-1} dP_n(x, t', u),
\end{aligned}
$$

$$
\text{for } t \geq 0,
$$

and the process V_F is replaced by

$$
V_F^{(m)}(t) = W_F^{(m)}(t) + \int_{[0,t]} F(t') \, dG_{F^{(m)}}(t'), \; t \geq 0.
$$

So, within the mth iteration we have another sequence of iterations, where $F^{(m,i-1)}$ changes into $F^{(m,i)}$ at the ith step, and where $F^{(m,i)}$ is the left derivative of the convex minorant of the "cumulative sum diagram", consisting of the points

$$
P_j = \Big(G_{F^{(m)}}(T_{(j)}), V_{F^{(m,i-1)}}^{(m)}(T_{(j)}) \Big),
$$

34

with $P_0 = (0,0)$ and $T_{(j)} \in J_n$, $j = 1, 2, \ldots$ (see Definition 1.1 and Proposition 1.4).

By this modification, the "outer iterations" of the algorithm get Newton-type properties: it is more important to start close to the solution of the maximization problem, but once one is close, the convergence is very fast. However, the only similarity with a real Newton algorithm is that we maximize a second order approximation; *the crucial difference is that we hit the boundary of the parameter space at each step.* Only in the lower dimensional boundary set partial derivatives of the likelihood function become zero, but finding this lower dimensional boundary set is the main part of the problem!

If we define the L_2-distance $\| \cdot \|$ on $\mathcal{F}$ by

$$\|F_1 - F_2\|^2 = \int \left(F_1(t) - F_2(t)\right)^2 dG_{F^{(m)}}(t),$$

(suppressing the dependence on m in order to simplify notation), we have for the "inner" iteration steps of the modified algorithm:

$$\|F^{(m,i+1)} - F^{(m,i)}\| \leq \|F^{(m,i)} - F^{(m,i-1)}\|.$$

This can be shown as follows. Let the function

$$t \mapsto \frac{dV_F^{(m)}}{dG_{F^{(m)}}}(t)$$

be defined by

$$\frac{dV_F^{(m)}}{dG_{F^{(m)}}}(t) = \begin{cases} \frac{V_F^{(m)}(t) - V_F^{(m)}(t-)}{G_{F^{(m)}}(t) - G_{F^{(m)}}(t-)} & \text{, if } G_{F^{(m)}}(t) > G_{F^{(m)}}(t-), \\ 0 & \text{, otherwise.} \end{cases}$$

Then $F^{(m,i+1)}$ minimizes

$$\left\| F - \frac{dV_{F^{(m,i)}}^{(m)}}{dG_{F^{(m)}}} \right\|$$

as a function of F, where F is a purely discrete distribution function with support contained in the set of points $T_{(i)}$.

Let the mapping $T : \mathcal{F} \to \mathcal{F}$ be defined by

$$\left\| T(F) - \frac{dV_F^{(m)}}{dG_{F^{(m)}}} \right\| = \min_{H \in \mathcal{F}} \left\| H - \frac{dV_F^{(m)}}{dG_{F^{(m)}}} \right\|.$$

By Theorem 8.2.5 in Robertson et al (1988), this mapping is a distance reducing operator (a projection on the closed convex set $\mathcal{F}$. Hence

$$(3.11) \qquad \left\| F^{(m,i+1)} - F^{(m,i)} \right\| \leq \left\| \frac{dV_{F^{(m,i)}}^{(m)}}{dG_{F^{(m)}}} - \frac{dV_{F^{(m,i-1)}}^{(m)}}{dG_{F^{(m)}}} \right\|.$$

35

But the square of the term at the right-hand side of (3.11) can be written

$$
\int_{t'\in[0,t],\, t'<x\leq u} \frac{\left\{F^{(m,i)}(u) - F^{(m,i-1)}(u)\right\}^2}{\left\{F^{(m)}(u) - F^{(m)}(t')\right\}^2}\, dP_n(x,t',u)
$$

(3.12)
$$
+ \int_{u\in[0,t],\, t'<x\leq u} \frac{\left\{F^{(m,i)}(t') - F^{(m,i-1)}(t')\right\}^2}{\left\{F^{(m)}(u) - F^{(m)}(t')\right\}^2}\, dP_n(x,t',u)
$$

$$
\leq \left\|F^{(m,i)} - F^{(m,i-1)}\right\|^2
$$

Moreover, under the conditions of Proposition 1.4, there can only be equality in (3.12), if $F^{(m,i)} = F^{(m,i-1)}$. Hence by a subsequence argument and the fact that the mapping $T : \mathcal{F} \to \mathcal{F}$ has a unique fixed point $F^{(m,\infty)}$, $F^{(m,i)}$ converges to $F^{(m,\infty)}$ at each point $T_{(j)}$, as $i \to \infty$.

Our general experience with the two versions of the iterative convex minorant algorithm is that the simple version converges rather quickly from any starting point and that the modified version is considerable slower for large data set, mainly because the "inner iterations" converge so slowly at the beginning steps of the algorithm. So we have the peculiar situation that there are rather convincing arguments for the (at least local) convergence of the modified algorithm (the outer steps maximize the full second order approximation to the log likelihood and the inner steps can be shown to converge), but that the simple iterative convex minorant algorithm, maximizing only a diagonal approximation to the log likelihood at each step, shows much better convergence properties, although no convergence proof is available at the moment!

As indicated in Chapter 1, we can always make a preliminary reduction, linking values of the solution. A simple rule for doing this runs as follows. Consider the process W_F, defined by (1.25). Under the conditions of Proposition 1.3, the first jump of the process will be upward, and the last jump will be downward. We now form blocks of upward jumps followed by downward jumps. Specifically, if an upward jump is denoted by $+$ and a downward jump by $-$, we consider consecutive blocks of the form

$$
+\cdots-\cdots,
$$

where each block is made as big as possible. In Example 1.2 we would get the two consecutive blocks:

$$
+--
$$
$$
+-
$$

and in Example 1.3 we would get the 4 consecutive blocks:

$$
+--
$$
$$
++--
$$
$$
+-
$$
$$
+--
$$

It is clear that the solution has to be constant for values of $T_{(i)}$, corresponding to such a block. So in Example 1.3 we immediately get the following relations, to be satisfied by the solution:

$$\hat{F}_n(T_{(1)}) = \hat{F}_n(T_{(2)}) = \hat{F}_n(T_{(3)})$$

$$\hat{F}_n(T_{(4)}) = \cdots = \hat{F}_n(T_{(7)})$$

$$\hat{F}_n(T_{(8)}) = \hat{F}_n(T_{(9)})$$

$$\hat{F}_n(T_{(10)}) = \hat{F}_n(T_{(11)}) = \hat{F}_n(T_{(12)})$$

Generally we can always take as our initial estimate a distribution function $F^{(0)}$ which is constant for values of $T_{(i)}$, corresponding blocks which have the structure described above. However, the NPMLE will typically be constant on much bigger blocks. The initial reduction gives the maximum number of blocks on which the NPMLE will be constant.

3.3. Exercises.

1. Is the following statement true or false:
   the EM algorithm converges, in the interval censoring problem, Case 1, to the NPMLE if and only if the starting distribution has positive mass at the points where the NPMLE puts mass.

2. Let, in the interval censoring problem, Case 1, the sample size be $n = 3$, and let $\delta_1 = \delta_2 = 1$, $\delta_3 = 0$. Take as starting distribution for the EM algorithm:

   $$p_i^{(0)} = P^{(0)}\{X = T_{(i)}\} = 1/4,$$

   where $T_{(1)} < T_{(2)} < T_{(3)}$ are the observation times, and $T_{(4)}$ is an arbitrary point, bigger than $T_{(3)}$. Show that the values at the nth iteration step ($n \geq 1$) are given by

   $$p_1^{(n)} = \tfrac{2}{3} - \tfrac{1}{3} \cdot 2^{-n},$$

   $$p_2^{(n)} = \tfrac{1}{3} \cdot 2^{-n},$$

   $$p_3^{(n)} = 0,$$

   $$p_4^{(n)} = \tfrac{1}{3}.$$

3. Let, in the interval censoring problem, Case 1, $\delta_1 = 1$ and $\delta_n = 0$, and let $y_i = \hat{F}_n(T_{(i)})$, $1 \leq i \leq n$. Show that (1.9) and (1.10) follow from (3.5) and (3.6).

4. What is the analogue of (3.5) and (3.6) for interval censoring, Case 2?

5. Let the conditions of Proposition 1.4 be satisfied. Show that the distribution function $F^{(m+1)}$, maximizing (3.8), is the left derivative of the convex minorant of the cumulative sum diagram, consisting of the points

   $$P_j = \left(G_{F^{(m)}}(T_{(j)}), V_{F^{(m)}}(T_{(j)})\right),$$

   where $P_0 = (0,0)$ and $T_{(j)} \in J_n$, $j = 1, 2, \ldots$.

6.(a) Describe an iterative convex minorant algorithm for the deconvolution problem, discussed in Section 2.1.

 (b) What would correspond to, respectively, the E and the M step of the EM algorithm for this problem?

4. CONSISTENCY

Consistency of the NPMLE in the cases of interval censoring and deconvolution can be proved by a general method which has been used by Jewell (1982) in proving consistency of the NPMLE for the mixing distribution in scale mixtures of exponential distributions. We first illustrate the method for interval censoring, case 1.

4.1. Interval censoring, case 1.

Let $(X_1, T_1), \ldots, (X_n, T_n)$ be a sample of random variables in $I\!R_+^2$, where X_i and T_i are independent (nonnegative) random variables with continuous distribution functions F_0 and G, respectively, satisfying $P_{F_0} \ll P_G$ (the probability measure P_{F_0}, induced by F_0, is absolutely continuous with respect to the probability measure P_G, induced by G).

If $\hat{F}_n$ is the NPMLE, we have, for each $\epsilon \in (0, 1)$:

$$\psi\big((1 - \epsilon)\hat{F}_n + \epsilon F_0\big) - \psi(\hat{F}_n) \leq 0,$$

where $\psi(\hat{F}_n)$ is defined by (1.3). Hence we obtain

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1} \left\{ \psi\big((1 - \epsilon)\hat{F}_n + \epsilon F_0\big) - \psi(\hat{F}_n) \right\} \leq 0.$$

Evaluating this limit, we obtain

$$(4.1) \qquad \int \left\{ \frac{F_0(t)}{\hat{F}_n(t)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_n(t)} 1_{\{x > t\}} \right\} \, dP_n(x, t) \leq 1.$$

We now take as our sample space Ω the space of all (infinite) sequences

$$(X_1, T_1), (X_2, T_2), \ldots,$$

and denote a point of this sample space by ω. To indicate the dependence on ω, we will write $\hat{F}_n(t\,;\omega)$ instead of $\hat{F}_n$, and likewise $P_n(x, t\,;\omega)$ instead of $P_n(x, t)$. The set Ω will be endowed with the usual Borel σ-algebra, generated by the product topology on $\prod_1^\infty I\!R^2$, and the product measure P^∞. We shall denote the latter probability measure by $I\!P$.

Fix $\epsilon \in (0, \frac{1}{2})$ and let a and b be chosen such that

$$(4.2) \qquad F_0(a) = \epsilon, \ F_0(b) = 1 - \epsilon.$$

Using the strong law of large numbers (and a separability argument), it is seen that $P_n(\cdot, \cdot\,;\omega)$ converges weakly to P, for all ω in a set B such that $I\!P(B) = 1$. Fix an $\omega \in B$. By the Helly compactness theorem it is seen that the sequence of functions $\hat{F}_n(\cdot\,;\omega)$ has a subsequence $\hat{F}_{n_k}(\cdot\,;\omega)$, converging vaguely to a nondecreasing right continuous function F,

39

taking values in $[0,1]$. Note that we cannot assume (at this point) that F is a distribution function.

We may assume that $1/\hat{F}_n(t;\omega)$ and $1/\big(1-\hat{F}_n(t;\omega)\big)$ are bounded for $t \in [a,b]$ and all n sufficiently large. This follows from (4.2) and the fact that $P_n(\cdot,\cdot;\omega)$ converges weakly to P. Moreover, by the vague convergence of $\hat{F}_{n_k}(\cdot;\omega)$ to F we may also assume that $1/F(t)$ and $1/\big(1-F(t)\big)$ are bounded for $t \in [a,b]$. Hence we assume

$$(4.3) \qquad\qquad 1/F(a) + 1/\big(1-F(b)\big) \le M$$

and

$$(4.4) \qquad\qquad 1/\hat{F}_n(a;\omega) + 1/\big(1-\hat{F}_n(b;\omega)\big) \le M,$$

for a constant $M > 0$ and all n sufficiently large.

We now get the following Lemma.

LEMMA 4.1. *Let the points a and b be chosen in such a way that (4.2) is satisfied. Then we have:*

$$(4.5) \qquad \begin{aligned} &\lim_{k\to\infty} \int_{\mathbb{R}\times[a,b]} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t;\omega)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-\hat{F}_{n_k}(t;\omega)} 1_{\{x>t\}} \right\} dP_{n_k}(x,t;\omega) \\ &= \int_{\mathbb{R}\times[a,b]} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-F(t)} 1_{\{x>t\}} \right\} dP(x,t). \end{aligned}$$

Moreover,

$$(4.6) \qquad \int_{\mathbb{R}\times[a,b]} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-F(t)} 1_{\{x>t\}} \right\} dP(x,t) \le 1.$$

PROOF: Fix $0 < \delta < 1$ and take a grid of points $a = u_0 < u_1 < \cdots < u_m = b$ on $[a,b]$ such that $m = 1 + \big[1/\delta^2\big]$ and

$$G(u_i) - G(u_{i-1}) = \big\{G(b) - G(a)\big\}/m, \quad i = 1,\dots,m.$$

First we suppose, for simplicity, that the points u_i are points of continuity for the function F, which is the (vague) limit of the sequence of functions $\hat{F}_{n_k}(\cdot;\omega)$.

Let K be the (possibly empty) set of indices i, $i = 1,\dots,m$ such that

$$\max\big\{1/F(u_{i-1}) - 1/F(u_i), 1/\big(1-F(u_i)\big) - 1/\big(1-F(u_{i-1})\big)\big\} \ge \delta.$$

By (4.3), the number of indices of this type is not bigger than $1 + [M/\delta]$. Let L be the remaining set of indices i, $i = 1, \ldots, m$. Denoting the interval $[u_0, u_1]$ by J_1 and the intervals $(u_{i-1}, u_i]$ by J_i, $i > 1$, we get

$$\int_{\mathbb{R} \times [a,b]} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP_{n_k}(x, t\,;\,\omega)$$

$$= \sum_{i=1}^{m} \int_{\mathbb{R} \times J_i} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP_{n_k}(x, t\,;\,\omega).$$

Since $\hat{F}_{n_k}(u_i\,;\,\omega)$ converges to $F(u_i)$ for each i, $0 \leq i \leq m$, we get, for sufficiently large k:

$$(4.7) \quad \begin{aligned} 1/\hat{F}_{n_k}(u_{i-1}\,;\,\omega) - 1/\hat{F}_{n_k}(u_i\,;\,\omega) < 2\delta, \ i \in L \\ 1/\left(1 - \hat{F}_{n_k}(u_i\,;\,\omega)\right) - 1/\left(1 - \hat{F}_{n_k}(u_{i-1}\,;\,\omega)\right) < 2\delta, \ i \in L. \end{aligned}$$

Hence

$$(4.8) \quad \begin{aligned} &\int_{\mathbb{R} \times [a,b]} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP_{n_k}(x, t\,;\,\omega) \\ &= \sum_{i \in K} \int_{\mathbb{R} \times J_i} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP_{n_k}(x, t\,;\,\omega) \\ &+ \sum_{i \in L} \int_{\mathbb{R} \times J_i} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP_{n_k}(x, t\,;\,\omega) \\ &= \int_{\mathbb{R} \times [a,b]} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP(x, t) + r_k(\omega), \end{aligned}$$

where $|r_k(\omega)| \leq c \cdot \delta$, for a constant $c > 0$. This can be seen by replacing $\hat{F}_{n_k}(t\,;\,\omega)$ on each interval J_i by its value $\hat{F}_{n_k}(u_i\,;\,\omega)$ at the right endpoint of the interval, and by noting that for large k

$$\left| 1/\hat{F}_{n_k}(t\,;\,\omega) - 1/\hat{F}_{n_k}(u_i\,;\,\omega) \right| < 2\delta,$$

if $i \in L$ (with a similar inequality for $1/\{1 - \hat{F}_{n_k}(t\,;\,\omega)\}$). On the intervals J_i, with $i \in K$, we use (4.4). Note that

$$\sum_{i \in K} P(\mathbb{R} \times J_i) \to 0, \ \text{if} \ \delta \downarrow 0,$$

since $P(\mathbb{R} \times J_i)$ is of order $O(\delta^2)$, while the number of intervals J_i such that $i \in K$ is of order $O(1/\delta)$.

On the other hand we have by dominated convergence:

$$(4.9) \quad \begin{aligned} &\lim_{k \to \infty} \int_{\mathbb{R} \times [a,b]} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x > t\}} \right\} dP(x, t) \\ &= \int_{\mathbb{R} \times [a,b]} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \leq t\}} + \frac{1 - F_0(t)}{1 - F(t)} 1_{\{x > t\}} \right\} dP(x, t). \end{aligned}$$

41

Combining (4.8) and (4.9) we obtain

$$
\begin{aligned}
(4.10) \quad & \int_{\mathbb{R}\times[a,b]} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x>t\}} \right\} \, dP_{n_k}(x,t\,;\,\omega) \\
& = \int_{\mathbb{R}\times[a,b]} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-F(t)} 1_{\{x>t\}} \right\} \, dP(x,t) + r_k'(\omega),
\end{aligned}
$$

where $\left|r_k'(\omega)\right| \le c'\delta$.

If one or more of the points u_i is not a point of continuity of F, we shift the point u_i a bit to the left or right, in order to get continuity points (using the fact that the continuity points of F are dense). So in all cases we get a relation of type (4.10). Since δ can be chosen arbitrarily small, (4.5) now follows. Relation (4.6) immediately follows from (4.5) and (4.1). ∎

By monotone convergence we now obtain from (4.6):

$$
\begin{aligned}
(4.11) \quad & \int_{\mathbb{R}^2} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-F(t)} 1_{\{x>t\}} \right\} \, dP(x,t) \\
& = \lim_{a\downarrow 0, b\to\infty} \int_{\mathbb{R}\times[a,b]} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-F(t)} 1_{\{x>t\}} \right\} \, dP(x,t) \le 1.
\end{aligned}
$$

This, however, can only happen if $F = F_0$. For we can write

$$
\begin{aligned}
& \int_{\mathbb{R}^2} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x\le t\}} + \frac{1-F_0(t)}{1-F(t)} 1_{\{x>t\}} \right\} \, dP(x,t) \\
& = \int_{\mathbb{R}} \left\{ \frac{F_0(t)^2}{F(t)} + \frac{(1-F_0(t))^2}{1-F(t)} \right\} \, dG(t),
\end{aligned}
$$

and the latter expression is strictly bigger than 1, unless $F = F_0$.

This is proved in the following way. If $0 < F_0(t) < 1$ and $y \in (0,1)$, then

$$
(4.12) \quad \frac{F_0(t)^2}{y} + \frac{(1-F_0(t))^2}{1-y} = \begin{cases} 1, & \text{if } y = F_0(t) \\ > 1, & \text{if } y \ne F_0(t). \end{cases}
$$

By the monotonicity of F, the monotonicity and continuity of F_0, and the absolute continuity of P_{F_0} with respect to P_G, we have

$$
F \ne F_0 \implies F(t) \ne F_0(t) \text{ on an interval of increase of } G.
$$

Hence by (4.12)

$$
\int_{\mathbb{R}} \left\{ \frac{F_0(t)^2}{F(t)} + \frac{(1-F_0(t))^2}{1-F(t)} \right\} \, dG(t) > 1,
$$

42

if $F \neq F_0$. This contradicts (4.11).

Thus we have proved that for each ω outside a set of probability zero, each subsequence of the sequence of functions $\hat{F}_n(\cdot\,;\omega)$ has a vaguely convergent sequence, and that all these convergent subsequences have the same limit F_0. This proves that the sequence $\hat{F}_n$ converges weakly to F_0, with probability one. Since F_0 is continuous, this is the same as saying that $\hat{F}_n$ converges with probability one to F_0 in the supremum distance on the set of distribution functions, i.e.

$$IP \left\{ \lim_{n \to \infty} \sup_{t \in I\!R} \left| \hat{F}_n(t) - F_0(t) \right| = 0 \right\} = 1.$$

We next illustrate the method for the deconvolution problem.

4.2. Convolution with a symmetric density.

Let $Z_1, \ldots, Z_n$ be a sample from a distribution with density

$$z \mapsto \int g(z - y)\,dF_0(y), \ z \in I\!R,$$

where g is a symmetric density, satisfying (2.14) and (2.15), and where F_0 is an unknown continuous distribution function.

Proceeding as before, we find

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1} \left\{ \psi((1 - \epsilon)\hat{F}_n + \epsilon F_0) - \psi(\hat{F}_n) \right\} \leq 0,$$

implying

(4.13)
$$\int \frac{\int g(z - y)\,dF_0(y)}{\int g(z - y)\,d\hat{F}_n(y)}\,dH_n(z) \leq 1.$$

For each ω in a set B of probability 1, the sequence $H_n(\cdot\,;\omega)$ converges weakly to the distribution function H_0, with density

$$h_0(z) = \int g(z - y)\,dF_0(y), \ z \in I\!R.$$

Since H_0 is continuous, we have

$$\sup_{z \in I\!R} \left| H_n(z; \omega) - H_0(z) \right| \to 0, \ \text{as } n \to \infty.$$

Fix $\omega \in B$ and consider a nonempty closed set of the form

$$A_\epsilon = \left\{ z \in I\!R : \int g(z - y)\,dF_0(y) \geq \epsilon \right\},$$

where $\epsilon > 0$. The sequence of functions $\hat{F}_n(\cdot\,;\omega)$ has a subsequence $\left(\hat{F}_{n_k}(\cdot\,;\omega)\right)_{k=1}^{\infty}$, converging vaguely to a subdistribution function F, and there exists a $\delta > 0$ such that

$$(4.14) \qquad \inf_{z \in A_\epsilon} \int g(z - y)\, d\hat{F}_{n_k}(y\,;\omega) \geq \delta,$$

for each sufficiently large k, since otherwise (4.13) would be violated. We get the following lemma, analogous to Lemma 4.1.

LEMMA 4.2. *We have:*

$$(4.15) \qquad \begin{aligned} &\lim_{k \to \infty} \int_{A_\epsilon} \frac{\int g(z - y)\, dF_0(y)}{\int g(z - y)\, d\hat{F}_{n_k}(y\,;\omega)}\, dH_{n_k}(z\,;\omega) \\ &= \int_{A_\epsilon} \frac{\int g(z - y)\, dF_0(y)}{\int g(z - y)\, dF(y)}\, dH_0(z). \end{aligned}$$

Moreover,

$$(4.16) \qquad \int_{A_\epsilon} \frac{\int g(z - y)\, dF_0(y)}{\int g(z - y)\, dF(y)}\, dH_0(z) \leq 1.$$

PROOF: Since $\hat{F}_{n_k}(\cdot\,;\omega)$ converges vaguely to F, we have

$$\lim_{k \to \infty} \int g(z - y)\, d\hat{F}_{n_k}(y\,;\omega) = \int g(z - y)\, dF(y),$$

Define the functions h_n and h by

$$h_n(z) = \int g(z - y)\, d\hat{F}_n(y\,;\omega), \quad h(z) = \int g(z - y)\, dF(y), \quad z \in \mathbb{R}.$$

Since g is bounded and uniformly continuous on $\mathbb{R}$, the sequence of continuous functions h_{n_k} converges uniformly to the (continuous) limit function h. Moreover, by (4.14)

$$\inf_{z \in A_\epsilon} h(z) \geq \delta > 0.$$

(note that A_ϵ is a compact set).

It now follows that the functions

$$z \mapsto \int g(z - y)\, dF_0(y) / h_{n_k}(z), \quad z \in A_\epsilon$$

44

are bounded continuous functions, converging uniformly to the bounded continuous function

$$z \mapsto \int g(z - y) \, dF_0(y) / h(z), \ z \in A_\epsilon.$$

Since $H_{n_k}(\cdot \, ; \omega)$ converges weakly to H_0, (4.15) easily follows, and (4.16) is implied by (4.15). ∎

We now proceed as before. The monotone convergence theorem and (4.16) imply

$$(4.17) \qquad \int \frac{\int g(z - y) \, dF_0(y)}{\int g(z - y) \, dF(y)} \, dH_0(z) \le 1.$$

This can only happen if $F = F_0$. For we get from (4.17):

$$\int h_0(z)^2 / h(z) \, dz \le 1,$$

where h_0 is the density of H_0. But $\int h_0(z)^2 / h(z) \, dz$ is minimized by taking $h = h_0$. This is seen by first observing that we may assume that

$$\int h(z) \, dz = 1,$$

since otherwise we could make the integral $\int h_0(z)^2 / h(z) \, dz$ smaller by multiplying h by a constant bigger than 1, and secondly by observing that the integrand of the integral

$$\int \left\{ \frac{h_0(z)^2}{h(z)} + h(z) \right\} \, dz$$

is pointwise minimized by taking $h(z) = h_0(z)$ for each z.

Now $F \ne F_0$ would imply

$$h(z) = \int g(z - y) \, dF(y) \ne h_0(z),$$

for z in an interval of positive length, and hence

$$(4.18) \qquad \int \frac{\int g(z - y) \, dF_0(y)}{\int g(z - y) \, dF(y)} \, dH_0(z) = \int \frac{h_0(z)^2}{h(z)} \, dz > 1,$$

contradicting (4.17).

We can conclude from this that every subsequence of the sequence of functions $\hat{F}_n(\cdot \, ; \omega)$ has a convergent subsequence, and that all these subsequences have the same (weak) limit F_0. This implies the consistency of the NPMLE.

We finally demonstrate the method for the more complicated case of interval censoring.

4.3. Interval censoring, case 2.

Let $(X_1, T_1, U_1), \ldots, (X_n, T_n, U_n)$ be a sample of random variables in $I\!R_+^3$, where X_i is a (nonnegative) random variable with continuous df F_0, and where T_i and U_i are (nonnegative) random variables, independent of X_i, with a joint continuous distribution function H and such that $T_i \leq U_i$ with probability one. Moreover, we asume that H has a density h with respect to Lebesgue measure, satisfying

$$(4.19) \qquad h(t, u) > 0, \text{ if } 0 < F_0(t) < F_0(u) < 1.$$

As in section 4.1, we obtain

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1} \left\{ \psi((1 - \epsilon)\hat{F}_n + \epsilon F_0) - \psi(\hat{F}_n) \right\} \leq 0,$$

implying

$$(4.20) \quad \int \left\{ \frac{F_0(t)}{\hat{F}_n(t)} 1_{\{x \leq t\}} + \frac{F_0(u) - F_0(t)}{\hat{F}_n(u) - \hat{F}_n(t)} 1_{\{t < x \leq u\}} + \frac{1 - F_0(u)}{1 - \hat{F}_n(u)} 1_{\{x > u\}} \right\} dP_n(x, t, u) \leq 1.$$

We now proceed as in section 4.1. Using the strong law of large numbers, it is seen that $P_n(\cdot, \cdot, \cdot; \omega)$ converges weakly to P, for all ω in a set B such that $I\!P(B) = 1$. Fix an $\omega \in B$. By the Helly compactness theorem it is seen that the sequence of functions $\hat{F}_n(\cdot; \omega)$ has a subsequence $\hat{F}_{n_k}(\cdot; \omega)$, converging vaguely to a subdistribution function F.

Fix $\epsilon \in (0, 1/2)$ and define the set A_ϵ by

$$A_\epsilon = \{(t, u) : F_0(t) \geq \epsilon, \; F_0(u) - F_0(t) \geq \epsilon, \; 1 - F_0(u) \geq \epsilon\}.$$

Moreover, let a and b be chosen such that

$$(4.21) \qquad F_0(a) = \epsilon, \; F_0(b) = 1 - \epsilon.$$

By (4.20) we may assume that there exists an $M > 0$ such that

$$(4.22) \qquad 1/\hat{F}_n(t; \omega) + 1/(\hat{F}_n(u; \omega) - \hat{F}_n(t; \omega)) + 1/(1 - \hat{F}_n(u; \omega)) \leq M,$$

for $(t, u) \in A_\epsilon$ and for all sufficiently large n. Note that the left side of (4.22) cannot tend to ∞ at points $(t, u) \in A_\epsilon$ such that $F_0(u) - F_0(t) = \epsilon$, since in that case the left side of (4.22) would become arbitrarily large on a subset of $A_{\epsilon/2}$ with Lebesgue measure bounded away from zero, as $n \to \infty$, which cannot happen because of (4.19), (4.21), the continuity of F_0 and the weak convergence of $P_n(\cdot, \cdot, \cdot; \omega)$ to P.

We get the following lemma, analogous to Lemma 4.1.

46

LEMMA 4.3. *We have:*
(4.23)
$$
\lim_{k \to \infty} \int_{I\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \le t\}} + \frac{F_0(u) - F_0(t)}{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{t < x \le u\}} \right.
$$
$$
\left. + \frac{1 - F_0(u)}{1 - \hat{F}_{n_k}(u\,;\,\omega)} 1_{\{x > u\}} \right\} dP_{n_k}(x, t, u\,;\,\omega)
$$
$$
= \int_{I\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \le t\}} + \frac{F_0(u) - F_0(t)}{F(u) - F(t)} 1_{\{t < x \le u\}} + \frac{1 - F_0(u)}{1 - F(u)} 1_{\{x > u\}} \right\} dP(x, t, u).
$$

Moreover,
(4.24)
$$
\int_{I\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \le t\}} + \frac{F_0(u) - F_0(t)}{F(u) - F(t)} 1_{\{t < x \le u\}} + \frac{1 - F_0(u)}{1 - F(u)} 1_{\{x > u\}} \right\} dP(x, t, u) \le 1.
$$

PROOF: Let the points a and b be defined by (4.21). We fix $0 < \delta < 1$ and consider a grid of points t_i, $t_0 < \cdots < t_m$, such that $m = 1 + [1/\delta^2]$, $t_0 = a$, $t_m = b$, and

$$
\int_{t \in [t_{i-1}, t_i]} h(t, u)\, dt\, du = m^{-1} \int_{t \in [a, b]} h(t, u)\, dt, \quad i = 1, \ldots, m.
$$

Likewise, we define a grid of points u_i, $u_0 < \cdots < u_m$, such that $u_0 = a$, $u_m = b$, and

$$
\int_{u \in [u_{i-1}, u_i]} h(t, u)\, dt\, du = m^{-1} \int_{u \in [a, b]} h(t, u)\, dt, \quad i = 1, \ldots, m.
$$

We denote the intervals $(t_{i-1}, t_i]$ by J_i and the intervals $(u_{i-1}, u_i]$ by J_i'. Furthermore, let K be the (possibly empty) set of indices i, $i = 1, \ldots, m$ such that

$$
\max \left\{ 1/F(t_{i-1}) - 1/F(t_i), 1/(1 - F(t_i)) - 1/(1 - F(t_{i-1})) \right\} \ge \delta,
$$

and let L be the remaining set of indices i of the points t_i. Likewise, let K' be the (possibly empty) set of indices i, $i = 1, \ldots, m$ such that

$$
\max \left\{ 1/F(u_{i-1}) - 1/F(u_i), 1/(1 - F(u_i)) - 1/(1 - F(u_{i-1})) \right\} \ge \delta,
$$

and let L' be the remaining set of indices i of the points u_i.

First suppose that the t_i and u_i are continuity points of F. The term

$$
\int_{I\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \le t\}} + \frac{1 - F_0(u)}{1 - \hat{F}_{n_k}(u\,;\,\omega)} 1_{\{x > u\}} \right\} dP_{n_k}(x, t, u\,;\,\omega)
$$

47

can be treated as before, showing that

$$\int_{I\!\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1-F_0(u)}{1-\hat{F}_{n_k}(u\,;\,\omega)} 1_{\{x > u\}} \right\} dP_{n_k}(x,t,u\,;\,\omega)$$

$$= \int_{I\!\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{1-F_0(u)}{1-\hat{F}_{n_k}(u\,;\,\omega)} 1_{\{x > u\}} \right\} dP(x,t,u) + r_k(\omega),$$

where $|r_k(\omega)| \leq c \cdot \delta$, for a constant $c > 0$. Similarly, for the "middle term"

$$\int_{R \times A_\epsilon} \left\{ \frac{F_0(u) - F_0(t)}{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{t < x \leq u\}} \right\} dP_{n_k}(x,t,u\,;\,\omega)$$

$$= \int_{R \times A_\epsilon} \left\{ \frac{F_0(u) - F_0(t)}{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{t < x \leq u\}} \right\} dP(x,t,u) + r'_k(\omega),$$

where $|r'_k(\omega)| \leq c \cdot \delta$, for a constant $c > 0$. This is seen by replacing $\hat{F}_{n_k}(t\,;\,\omega)$ on each interval J_i by its value $\hat{F}_{n_k}(t_{i-1}\,;\,\omega)$ at the *left* endpoint of the interval, and by replacing $\hat{F}_{n_k}(u\,;\,\omega)$ on each interval J'_i by its value $\hat{F}_{n_k}(u_i\,;\,\omega)$ at the *right* endpoint of the interval, and noting that for large k, and $(t,u) \in A_\epsilon$:

$$\left| 1/\{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t_{i-1}\,;\,\omega)\} - 1/\{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)\} \right| < 2\delta M^2,$$

if $i \in L$ and that

$$\left| 1/\{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)\} - 1/\{\hat{F}_{n_k}(u_i\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)\} \right| < 2\delta M^2,$$

if $i \in L'$.

If $t \in J_i$, $i \in K$ or $u \in J'_i$, $i \in K'$, we use (4.22). Note that

$$\sum_{i \in K} \int_{I\!\!R \times J_i \times I\!\!R} dP(x,t,u) \to 0, \text{ if } \delta \downarrow 0,$$

and likewise

$$\sum_{i \in K'} \int_{I\!\!R^2 \times J'_i} dP(x,t,u) \to 0, \text{ if } \delta \downarrow 0,$$

since $P(I\!\!R \times J_i \times I\!\!R)$ and $P(I\!\!R^2 \times J'_i)$ are of order $O(\delta^2)$, while the number of intervals J_i such that $i \in K$ is of order $O(1/\delta)$, and likewise the number of intervals J'_i such that $i \in K'$ is of order $O(1/\delta)$.

By dominated convergence we get
(4.25)

$$\lim_{k \to \infty} \int_{I\!\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{\hat{F}_{n_k}(t\,;\,\omega)} 1_{\{x \leq t\}} + \frac{F_0(u) - F_0(t)}{\hat{F}_{n_k}(u\,;\,\omega) - \hat{F}_{n_k}(t\,;\,\omega)} 1_{\{t < x \leq u\}} \right.$$

$$\left. + \frac{1-F_0(u)}{1-\hat{F}_{n_k}(u\,;\,\omega)} 1_{\{x > u\}} \right\} dP(x,t,u\,;\,\omega)$$

$$= \int_{I\!\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \leq t\}} + \frac{F_0(u) - F_0(t)}{F(u) - F(t)} 1_{\{t < x \leq u\}} + \frac{1-F_0(u)}{1-F(u)} 1_{\{x > u\}} \right\} dP(x,t,u).$$

48

If one or more of the points of the (two) grids is not a point of continuity of F, we argue as in the proof of Lemma 4.1. This yields (4.23), and (4.24) easily follows. ∎

By monotone convergence we now obtain from (4.24):

(4.26)
$$\int_{I\!\!R^3} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \le t\}} + \frac{F_0(u) - F_0(t)}{F(u) - F(t)} 1_{\{t < x \le u\}} + \frac{1 - F_0(u)}{1 - F(u)} 1_{\{x > u\}} \right\} dP(x, t, u)$$

$$= \lim_{\epsilon \downarrow 0} \int_{I\!\!R \times A_\epsilon} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \le t\}} + \frac{F_0(u) - F_0(t)}{F(u) - F(t)} 1_{\{t < x \le u\}} + \frac{1 - F_0(u)}{1 - F(u)} 1_{\{x > u\}} \right\} dP(x, t, u)$$

$$\le 1.$$

This, however, can again only happen if $F = F_0$. For we can write

$$\int_{I\!\!R^3} \left\{ \frac{F_0(t)}{F(t)} 1_{\{x \le t\}} + \frac{F_0(u) - F_0(t)}{F(u) - F(t)} 1_{\{t < x \le u\}} + \frac{1 - F_0(u)}{1 - F(u)} 1_{\{x > u\}} \right\} dP(x, t, u)$$

$$= \int_{I\!\!R^2} \left\{ \frac{F_0(t)^2}{F(t)} + \frac{(F_0(u) - F_0(t))^2}{F(u) - F(t)} + \frac{(1 - F_0(u))^2}{1 - F(u)} \right\} dH(t, u),$$

and the latter expression is strictly bigger than 1, unless $F = F_0$.

The proof is analogous to the proof for interval censoring, case 1. If $0 < F_0(t) < F_0(u) < 1$ and $0 < x < y < 1$, then

(4.27) $$\frac{F_0(t)^2}{x} + \frac{(F_0(u) - F_0(t))^2}{y - x} + \frac{(1 - F_0(u))^2}{1 - y} = \begin{cases} 1, & \text{if } x = F_0(t),\ y = F_0(u) \\ > 1, & \text{otherwise.} \end{cases}$$

By the monotonicity of F, the monotonicity and continuity of F_0, and (4.19) we now have

$$\int_{I\!\!R^2} \left\{ \frac{F_0(t)^2}{F(t)} + \frac{(F_0(u) - F_0(t))^2}{F(u) - F(t)} + \frac{(1 - F_0(u))^2}{1 - F(u)} \right\} dH(t, u) > 1,$$

if $F \ne F_0$. This contradicts (4.26).

As in the case of interval censoring, case 1, we get

$$P\left\{ \lim_{n \to \infty} \sup_{t \in R} |\hat{F}_n(t) - F_0(t)| \doteq 0 \right\} = 1.$$

4.4. Exercises.

1. Let F_0 be a continuous distribution function of a non-negative random variable, and let g be a decreasing probability density on $[0, \infty)$, which is continuous on $[0, \infty)$. Show by the method of this chapter that the NPMLE is strongly consistent for the deconvolution model, consisered in section 2.1.

The following exercises give an idea of the "entropy approach" in proving consistency. We consider the interval censoring model, case 1, under the continuity assumptions, specified in section 4.1.

2. Let $\mathcal{F}$ be the set of distribution functions on $I\!\!R$, and let $\mathcal{A}$ be the class of functions $\{\phi_F : F \in \mathcal{F}\}$, where

$$\phi_F(x,t) = \{x \le t\}\sqrt{F(t)} + \{x > t\}\sqrt{1 - F(t)} \qquad , x,t \in I\!\!R.$$

Let $N_2(\delta, P_n, \mathcal{A})$ be the minimum number of balls with radius $\delta > 0$, needed to cover the set $\mathcal{A}$, using the L_2-distance

$$\|\phi_F - \phi_G\|_2 \stackrel{\text{def}}{=} \left\{ \int |\phi_F(x,t) - \phi_G(x,t)|^2 \, dP_n(x,t) \right\}^{1/2}, \quad F, G \in \mathcal{F},$$

where P_n is the empirical measure of $(X_1, T_1), \ldots, (X_n, T_n)$. Show that

$$I\!\!P\left\{ \lim_{n \to \infty} n^{-1} \log N_2(\delta, P_n, \mathcal{A}) = 0 \right\} = 1.$$

3. Let $\mathcal{A}_0$ be the class of functions $\psi_F : F \in \mathcal{F}\}$, where

$$\psi_F(x,t) = 1_{\{(x,t):\phi_{F_0}(x,t)>0\}} \phi_F(x,t)/\phi_{F_0}(x,t) \qquad , x,t \in I\!\!R,$$

and let $N_1(\delta, P_n, \mathcal{A})$ be the minimum number of balls with radius $\delta > 0$, needed to cover the set $\mathcal{A}$, using the L_1-distance

$$\|\phi_F - \phi_G\|_1 \stackrel{\text{def}}{=} \int |\phi_F(x,t) - \phi_G(x,t)| \, dP_n(x,t), \quad F, G \in \mathcal{F}.$$

Show

$$I\!\!P\left\{ \lim_{n \to \infty} n^{-1} \log N_1(\delta, P_n, \mathcal{A}) = 0 \right\} = 1.$$

(**Hint:** use Cauchy-Schwarz.)

4. Deduce from Exercise 3:

$$\sup_{F \in \mathcal{F}} \int \psi_F(x,t) \, d(P_n - P)(x,t) \to 0,$$

with probability one.

5. Let $\hat{F}_n$ be the NPMLE of F_0. Show

$$0 \le \int \log(\psi_{\hat{F}_n}^2) \, dP_n \le 2 \int \psi_{\hat{F}_n} \, dP_n - 2$$

$$(4.28)$$

$$= 2 \int \psi_{\hat{F}_n} \, d(P_n - P)(x,t) - 2 \int \{\psi_{\hat{F}_n} - \psi_{F_0}\}^2 \, dP.$$

Deduce from Exercise 4 and (4.28) that, with probability one:

$$(4.29) \qquad\qquad \int \left\{ \psi_{\hat{F}_n} - \psi_{F_0} \right\}^2 dP \to 0, \ \text{as } n \to \infty.$$

6. Derive from (4.29) the strong consistency of $\hat{F}_n$. Can you give an interpretation of the left-hand side of (4.29) in terms of a (squared) Hellinger distance?

Remark. The outline of the "entropy proof" of the consistency of the NPMLE in the case of interval censoring, case 1, is based on a personal communication by Sara van de Geer. For extensions and other approaches, see van de Geer (1990); in the latter report also rates of convergence are derived by using entropy methods.

5. Distribution theory

Parts of the present chapter have a somewhat heuristic character in the sense that we use a working hypothesis which can be formulated as follows.

Working hypothesis. Starting with the real underlying distribution function F_0, the iterative convex minorant algorithm will give at the first iteration step an estimator which is asymptotically equivalent with the maximum likelihood estimator.

By "asymptotically equivalent" we mean the following. Suppose $F_n^{(1)}$ is the estimator of F_0, obtained at the first step of the iterative convex minorant algorithm, with starting distribution $F^{(0)} = F_0$, and suppose

$$\alpha_n\big(F_n^{(1)}(t) - F_0(t)\big) \xrightarrow{\mathcal{D}} Z, \text{ as } n \to \infty,$$

where Z is some nondegenerate random variable, t is a fixed point in the interior of the support of F_0, α_n is a norming constant, and where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Then also

$$\alpha_n\big(\hat{F}_n(t) - F_0(t)\big) \xrightarrow{\mathcal{D}} Z, \text{ as } n \to \infty,$$

where $\hat{F}_n$ is the maximum likelihood estimator of F_0.

The working hypothesis is supported by computer experiments, and certainly holds (under some condition on F_0 and G in the neighborhood of t) for interval censoring, Case 1, but so far there is no proof for the general situation. Roughly speaking, the fact one would have to show is that "off-diagonal elements of the matrix of second derivatives of the log likelihood function can be neglected" (some arguments for why this might be true will be given below).

We will first demonstrate the method for interval censoring, Case 1. In this situation the off-diagonal elements of the matrix of second derivatives of the log likelihood function are zero. So we don't have to deal with the problem mentioned in the preceding paragraph, and still can demonstrate the general method.

5.1. Interval censoring, Case 1.

We will use the same set-up as in Section 4.1. The following result will be proved.

THEOREM 5.1. *Let t_0 be such that $0 < F_0(t_0), G(t_0) < 1$, and let F_0 and G be differentiable at t_0, with strictly positive derivatives $f_0(t_0)$ and $g(t_0)$, respectively. Furthermore, let $\hat{F}_n$ be the NPMLE of F_0. Then we have, as $n \to \infty$,*

$$n^{1/3}\big\{\hat{F}_n(t_0) - F_0(t_0)\big\}\big/\big\{\tfrac{1}{2}F_0(t_0)(1 - F_0(t_0))f_0(t_0)/g(t_0)\big\}^{1/3} \xrightarrow{\mathcal{D}} 2Z,$$

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

We first study the limiting behavior of $F_n^{(1)}$. To this end, we consider the process $W_n^{(0)}$, defined by

$$(5.1) \qquad W_n^{(0)}(t) = \int_{t' \in [0,t]} \left\{ F_0(t')^{-1} 1_{\{x \leq t'\}} - \left(1 - F_0(t')\right)^{-1} 1_{\{x > t'\}} \right\} dP_n(x,t'), \, t \geq 0,$$

and the process $G_n^{(0)}$, defined by

$$(5.2) \qquad G_n^{(0)}(t) = \int_{t' \in [0,t]} \left\{ F_0(t')^{-2} 1_{\{x \leq t'\}} + \left(1 - F_0(t')\right)^{-2} 1_{\{x > t'\}} \right\} dP_n(x,t'), \, t \geq 0,$$

i.e., $W_n^{(0)}$ and $G_n^{(0)}$ are defined in a similar way as the processes W_F and G_F, for interval censoring, Case 2 (see (1.25) and (1.29)), with $F = F_0$.

Furthermore, we define the process $V_n^{(0)}$ by

$$(5.3) \qquad V_n^{(0)}(t) = W_n^{(0)}(t) + \int_{[0,t]} (F_0(t') - F_0(t_0)) \, dG_n^{(0)}(t'), \, t \geq 0.$$

The process $V_n^{(0)}$ has the following property.

LEMMA 5.1. *Let the process $U_n^{(0)}$ be defined by*

$$U_n^{(0)}(t) = n^{2/3} \left\{ V_n^{(0)}(t_0 + n^{-1/3}t) - V_n^{(0)}(t_0) \right\}, \, t \in \mathbb{R},$$

where $U_n^{(0)}(t) = 0$, if $t \leq -t_0 n^{1/3}$. Then $U_n^{(0)}$ converges in distribution, in the topology of uniform convergence on compacta on the space of locally bounded real-valued functions on $\mathbb{R}$, to the process U, defined by

$$(5.4) \qquad U(t) = \sqrt{g(t_0)/(F_0(t_0)(1 - F_0(t_0)))} \, W(t) + \tfrac{1}{2} \frac{f_0(t_0)g(t_0)}{F_0(t_0)(1 - F_0(t_0))} t^2, \, t \in \mathbb{R},$$

where W is (standard) two-sided Brownian motion on $\mathbb{R}$, originating from zero.

PROOF: For $t \geq 0$ we can write:

$$(5.5) \qquad\qquad\qquad U_n^{(0)}(t) = P_n g_n(\cdot, \cdot; t),$$

where

$$
g_n(x, u\,; t) = n^{2/3} 1_{(t_0, t_0 + n^{-1/3}t]}(u) \left\{ \frac{\{x \leq u\}}{F_0(u)} - \frac{\{x > u\}}{1 - F_0(u)} \right.
$$
$$
\left. + \frac{\{x \leq u\}(F_0(u) - F_0(t_0))}{F_0(u)^2} + \frac{\{x > u\}(F_0(u) - F_0(t_0))}{(1 - F_0(u))^2} \right\}.
$$

53

We have:

$$(5.6) \qquad \mathbb{P} g_n(\cdot, \cdot; t) \sim \frac{1}{2} \frac{f_0(t_0) g(t_0)}{F_0(t_0)\left(1 - F_0(t_0)\right)} \, t^2, \text{ as } n \to \infty,$$

uniformly for t in a bounded interval $[0, M]$. Using the representation (5.5), it is easily checked that the process $U_n^{(0)}$ satisfies the following type of stochastic equicontinuity condition: for each $\epsilon > 0$, $\delta > 0$, and $M > 0$, there exists a $\delta > 0$ such that

$$(5.7) \qquad \limsup_{n \to \infty} \mathbb{P} \left\{ \sup_{0 \le t, t' \le M, \, |t - t'| \le \delta} \left| U_n^{(0)}(t) - U_n^{(0)}(t') \right| > \eta \right\} < \epsilon.$$

(a "tightness condition"). Moreover, the variance of $U_n^{(0)}(t)$ converges, for $t > 0$, to

$$\frac{g(t_0)}{F_0(t_0)\left(1 - F_0(t_0)\right)} \, t,$$

uniformly for t in bounded intervals. It now follows from (5.6) and (5.7) that the process $U_n^{(0)}$, restricted to $[0, \infty)$, converges in distribution to the process U, restricted to $[0, \infty)$ (in the topology of uniform convergence on compacta). Since a similar line of argument holds for $U_n^{(0)}$, restricted to $(-\infty, 0]$, the result follows. ∎

Remark 5.1. Lemma 5.1 can also be proved by a martingale argument. Note, for example, that $\left\{ n^{2/3} \left(W_n^{(0)}(t_0 + n^{-1/3} t) - W_n^{(0)}(t_0) \right) : t \ge 0 \right\}$ is a martingale. For negative values of t, we can use a martingale with time running backward.

We now define, for each $a > 0$, the random variable $T_n^{(0)}(a)$ by

$$(5.8) \qquad T_n^{(0)}(a) = \sup\{ t \in [0, T_{(n)}] : V_n^{(0)}(t) - \left(a - F_0(t_0) \right) G_n^{(0)}(t) \text{ is minimal} \}.$$

Then the process

$$\left\{ \left(G_n^{(0)}(T_n^{(0)}(a)), V_n^{(0)}(T_n^{(0)}(a)) + F_0(t_0) G_n^{(0)}(T_n^{(0)}(a)) \right) : a \in (0, 1) \right\}$$

runs through the vertices of the cumulative sum diagram S_n, consisting of the points

$$P_j = \left(G_n^{(0)}(T_{(j)}), V_n^{(0)}(T_{(j)}) + F_0(t_0) G_n^{(0)}(T_{(j)}) \right), 1 \le j \le n,$$

and $P_0 = (0, 0)$.

In a similar way, we define the process $\{ T(a) : a \in \mathbb{R} \}$ by

$$(5.9) \qquad T(a) = \sup \left\{ t \in \mathbb{R} : U(t) - a \cdot \frac{g(t_0) t}{F_0(t_0)\left(1 - F_0(t_0)\right)} \text{ is minimal} \right\},$$

where the process U is defined by (5.5). The processes $T_n^{(0)}$ and T are similar to processes studied in Groeneboom (1988) and (1989). They run through the locations of the vertices of the convex minorant of respectively, the process $V_n^{(0)}$ (in the "time scale" $G_n^{(0)}$) and the process U. We now show that a rescaled version of the process $T_n^{(0)}$ converges in distribution (in the space $D(I\!\!R)$ with the Skorohod topology) to the process T. For completeness, we define a Skorohod metric on $D(I\!\!R)$. This is done in a similar way as it is done in Pollard (1984), p. 123, for the space $D[0, \infty)$.

Definition 5.1. For each finite $M > 0$ and each pair of functions x and y in $D(I\!\!R)$, the distance $d_M(x, y)$ is defined as the set of all those values of δ for which there exist grids

$$s_0 < s_1 < \cdots < s_k, \text{ and } t_0 < t_1 < \cdots < t_k,$$

such that $s_0 \vee t_0 \leq -M$, $s_k \wedge t_k \geq M$, $|s_i - t_i| \leq \delta$, for $i = 0, \ldots, k$, and

$$|x(s) - y(t)| \leq \delta, \text{ if } s_i \leq s < s_{i+1}, \text{ and } t_i \leq t < t_{i+1},$$

for $i = 0, \ldots, k - 1$. The weighted sum

$$d(x, y) = \sum_{k=1}^{\infty} 2^{-k} \min\{1, d_k(x, y)\}$$

defines the Skorohod metric on $D(I\!\!R)$.

LEMMA 5.2. *Let the conditions of Lemma 5.1 be satisfied, and let $a_0 = F_0(t_0)$. Moreover, let $T_n^{(0)}(a)$ be defined by (5.8) for $a \in (0, 1)$, and let $T_n^{(0)}(a) = 0$, $a \leq 0$, and $T_n^{(0)}(a) = T_{(n)}, a \geq 1$. Then the process*

(5.10) $$\left\{ n^{1/3}\{T_n^{(0)}(a_0 + n^{-1/3}a) - t_0\} : a \in I\!\!R \right\}$$

converges distribution, in the space $D(I\!\!R)$ with the Skorohod topology, to the process $\{T(a) : a \in I\!\!R\}$, defined by (5.9).

PROOF: By a simple scaling argument, it is seen that

$$n^{1/3}\{T_n^{(0)}(a_0 + n^{-1/3}a) - t_0\}$$
$$= \sup\left\{ t : U_n^{(0)}(t) - a \cdot n^{1/3}\{G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0)\} \text{ is minimal} \right\}.$$

Since, with probability one,

$$\lim_{n \to \infty} n^{1/3}\left\{ G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0) \right\} = \frac{g(t_0)t}{F_0(t_0)(1 - F_0(t_0))},$$

uniformly for t in bounded intervals, the process

$$(5.11) \qquad \left\{ U_n^{(0)}(t) - a \cdot n^{1/3} \left\{ G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0) \right\} : (t,a) \in I\!\!R^2 \right\}$$

converges in distribution (in the topology of uniform convergence on compacta), to the process

$$(5.12) \qquad \left\{ U(t) - a \cdot \frac{g(t_0)t}{F_0(t_0)(1 - F_0(t_0))} : (t,a) \in I\!\!R^2 \right\}.$$

It is proved below (Lemma 5.3) that for each $\epsilon > 0$ and $M_1 > 0$ an $M_2 > 0$ can be found such that

$$(5.13) \qquad I\!\!P \left\{ \sup_{a \in [-M_1, M_1]} n^{1/3} \left| T_n^{(0)}(a_0 + n^{-1/3}a) - t_0 \right| > M_2 \right\} < \epsilon,$$

for all sufficiently large n. The process $\{T(a) : a \in I\!\!R\}$ is a function of the process (5.12), and it is shown in Groeneboom (1989) that $\{T(a) : a \in I\!\!R\}$ is an increasing Markovian jump process, with almost surely a finite number of jumps in each finite interval and that $\{T(a) - f_0(t_0)^{-1}a : a \in I\!\!R\}$ is a stationary process. Therefore, by (5.13) and an almost sure construction, similar to Theorem 2.7 in Kim and Pollard (1990) (which is a form of the continuous mapping theorem that can be used for "argmax functionals"), it follows that the process (5.10), considered as a function of the process (5.11), converges in distribution (in the Skorohod topology) to the corresponding function of the process (5.12). However, the latter function is just the process $\{T(a) : a \in I\!\!R\}$. ∎

The following lemma establishes (5.13).

LEMMA 5.3. *For each $\epsilon > 0$ and $M_1 > 0$ an $M_2 > 0$ can be found such that*

$$I\!\!P \left\{ \max_{a \in [-M_1, M_1]} n^{1/3} \left\{ T_n^{(0)}(a_0 + n^{-1/3}a) - t_0 \right\} > M_2 \right\} < \epsilon,$$

and

$$I\!\!P \left\{ \min_{a \in [-M_1, M_1]} n^{1/3} \left\{ T_n^{(0)}(a_0 + n^{-1/3}a) - t_0 \right\} < -M_2 \right\} < \epsilon,$$

for all sufficiently large n.

PROOF: We only prove the first inequality, since the second inequality can be proved in a completely analogous way. First note that

$$I\!\!P \left\{ \max_{a \in [-M_1, M_1]} n^{1/3} \left\{ T_n^{(0)}(a_0 + n^{-1/3}a) - t_0 \right\} > M_2 \right\}$$

$$= I\!\!P \left\{ n^{1/3} \left\{ T_n^{(0)}(a_0 + n^{-1/3}M_1) - t_0 \right\} > M_2 \right\}$$

56

since the process $T_n^{(0)}$ is nondecreasing. Furthermore,

$$\mathbb{P}\left\{n^{1/3}\left\{T_n^{(0)}(a_0 + n^{-1/3}M_1) - t_0\right\} > M_2\right\}$$

$$\leq \mathbb{P}\left\{U_n^{(0)}(t) - n^{1/3}M_1\left\{G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0)\right\} < 0, \text{ for some } t \geq M_2\right\}.$$

We have

$$n^{-2/3}U_n^{(0)}(t) - n^{-1/3}M_1 \cdot \left\{G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0)\right\}$$

$$= V_n^{(0)}(t_0 + n^{-1/3}t) - V_n^{(0)}(t_0) - n^{-1/3}M_1\left\{G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0)\right\}$$

$$= W_n^{(0)}(t_0 + n^{-1/3}t) - W_n^{(0)}(t_0)$$

$$+ \int_{[t_0,t_0+n^{-1/3}t]}\left\{F_0(t') - F_0(t_0) - n^{-1/3}M_1\right\}dG_n^{(0)}(t').$$

The process $\left\{W_n^{(0)}(t_0 + u) - W_n^{(0)}(t_0) : u \geq 0\right\}$, is a martingale with respect to the self-induced filtration $\{\mathcal{F}_u : u \geq 0\}$, where

$$\mathcal{F}_u = \sigma\left\{W_n^{(0)}(t_0 + u') - W_n^{(0)}(t_0) : 0 \leq u' \leq u\right\}.$$

Let $Z_n^{(0)}(u) = \left\{W_n^{(0)}(t_0 + u) - W_n^{(0)}(t_0)\right\}^2$, then $\{Z_n^{(0)}(u) : u \geq 0\}$ is a submartingale and, by Doob's inequality:

$$\mathbb{P}\left\{\sup_{0 \leq u \leq u_0} Z_n^{(0)}(u)\right\} \leq 4\mathbb{P}Z_n^{(0)}(u_0) = 4n^{-1}\left\{\frac{g(t_0)}{F_0(t_0)(1 - F_0(t_0))}u_0 + \mathcal{O}(u_0^2)\right\},$$

if u_0 satisfies $F_0(t_0 + u_0) < 1$. So in particular we get, for $\epsilon > 0$ and $A > 0$:

$$\mathbb{P}\left\{\exists u \in [(j-1)n^{-1/3}, jn^{-1/3}) : n^{2/3}\left|W_n^{(0)}(t_0 + u) - W_n^{(0)}(t_0)\right| > \epsilon(j-1)^2 + A\right\}$$

$$4 \leq \mathbb{P}Z_n^{(0)}(jn^{-1/3})/\left\{\epsilon(j-1)^2 + A\right\}^2$$

$$= 4n^{4/3}\left\{n^{-1}c \cdot jn^{-1/3}\right\}/\left\{\epsilon(j-1)^2 + A\right\}^2 = c \cdot j/\left\{\epsilon(j-1)^2 + A\right\}^2,$$

for some constant $c > 0$, if $0 \leq jn^{-1/3} \leq u_0$, where c does not depend on $jn^{-1/3}$, for $jn^{-1/3} \leq u_0$.

By the same arguments as used in Lemma 4.1 in Kim and Pollard (1990), we get from this that for each $\epsilon > 0$ there exist random variables A_n of order $\mathcal{O}_p(1)$ such that

$$(5.14) \qquad \left|W_n^{(0)}(t_0 + u) - W_n^{(0)}(t_0)\right| \leq \epsilon u^2 + n^{-2/3}A_n, \text{ if } 0 \leq u \leq u_0.$$

By the fact that F_0 is differentiable in t_0, with a strictly positive derivative $f_0(t_0)$, there exist, for each $M > 0$ and $\eta > 0$ an $M_2 > 0$ and $\epsilon > 0$ such that
(5.15)

$$\int_{[t_0,t_0+u]}\left\{F_0(t') - F_0(t_0) - n^{-1/3}M_1\right\}dG_n^{(0)}(t') \geq \max\{Mn^{-2/3}, \epsilon u^2\}, \forall u \in [M_2n^{-1/3}, u_0],$$

with probability bigger than $1 - \eta$. Combining (5.14) and (5.15) we get

$$\mathbb{P}\left\{ U_n^{(0)}(t) - n^{1/3}M_1\left\{ G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0)\right\} < 0, \text{ for some } t,\ M_2 \le t \le u_0 n^{1/3}\right\}$$
$$\le \mathbb{P}\{A_n > M\} + \eta$$

and the last expression can be made smaller than 2η, by taking M sufficiently large.

Finally, we have to deal with the behavior of the process for values of $t \ge u_0 n^{1/3}$, where u_0 is chosen in such a way that $F_0(t_0 + u_0)$ is close to 1. But for these values of t the process

$$U_n^{(0)}(t) - n^{1/3}M_1\left\{ G_n^{(0)}(t_0 + n^{-1/3}t) - G_n^{(0)}(t_0)\right\},\ t \ge 0,$$

will be increasing for all sufficiently large n, since negative terms of the form

$$-1\big/\left\{ n\big(1 - F_0(T_{(i)})\big)\right\}$$

in the martingale part of the process will be compensated by terms of the form

$$n^{-1}\left\{ F_0(T_{(i)}) - F_0(t_0) - n^{-1/3}M_1\right\}\big/ \left\{ F_0(T_{(i)})\big(1 - F_0(T_{(i)})\big)\right\}^2,$$

in the "drift" part of the process. ∎

The distribution of the 1-step estimator $F_n^{(1)}$ at t_0 can now be found by using the following relation between $F_n^{(1)}$ and $T_n^{(0)}$:

(5.16) $$\mathbb{P}\{F_n^{(1)}(t_0) - F_0(t_0) > x\} = \mathbb{P}\{T_n^{(0)}(a_0 + x) < t_0\}.$$

Relation (5.16) is easily verified by drawing a picture of the situation. The preceding results yield the limiting distribution of $F_n^{(1)}(t_0)$.

THEOREM 5.2. *For each $x \in \mathbb{R}$ we have*

$$\lim_{n \to \infty} \mathbb{P}\left\{ n^{1/3}\left\{ F_n^{(1)}(t_0) - F_0(t_0)\right\} > x\right\} = \mathbb{P}\{T(0) > c \cdot x\},$$

where $T(0)$ is defined by (5.9), and where

(5.17) $$c = f_0(t_0)^{-1}.$$

PROOF: Using (5.16), we only have to consider

$$\mathbb{P}\left\{ T_n^{(0)}\big(a_0 + x n^{-1/3}\big) < t_0\right\}.$$

58

But we have, by Lemma 5.2, for fixed $x \in \mathbb{R}$,

$$\mathbb{P}\left\{T_n^{(0)}(a_0 + xn^{-1/3}) < t_0\right\}$$
$$= \mathbb{P}\left\{n^{1/3}\left\{T_n^{(0)}(a_0 + xn^{-1/3}) - t_0\right\} < 0\right\} \to \mathbb{P}\{T(x) < 0\},$$

as $n \to \infty$. Using the stationarity of the process $\{T(a) - c \cdot a : a \in \mathbb{R}\}$, proved in Groeneboom (1989), we get

$$\mathbb{P}\{T(x) < 0\} = \mathbb{P}\{T(x) - c \cdot x < -c \cdot x\} = \mathbb{P}\{T(0) < -c \cdot x\}.$$

The statement now follows from the fact that $T(0)$ has a density which is symmetric about zero. ∎

We will now derive the asymptotic distribution of $\hat{F}_n(t_0)$, where $\hat{F}_n$ is the NPMLE. We first show that the distance between $\hat{F}_n(t)$ and $F_0(t_0)$ is $\mathcal{O}_p(n^{-1/3})$, for t in an interval of the form $[t_0 - Mn^{-1/3}, t_0 + Mn^{-1/3}]$.

LEMMA 5.4. For each $M > 0$ we have:

$$\sup_{t \in [-M,M]} \left|\hat{F}_n(t_0 + n^{-1/3}t) - F_0(t_0)\right| = \mathcal{O}_p(n^{-1/3}).$$

PROOF: We first show that the probability that (for large n) $\hat{F}_n$ does not have a jump in an interval of the form $[t_0 - Mn^{-1/3}, t_0 + Mn^{-1/3}]$ can be made arbitrarily small by taking M sufficiently large. Specifically, we will show that for each $\epsilon > 0$, there exists an $M > 0$ such that

(5.18) $$\mathbb{P}\left\{\hat{F}_n(t_0 - Mn^{-1/3}) \geq F_0(t_0)\right\} < \epsilon,$$

for all large n. Since it can be shown in a similar way that for each $\epsilon > 0$, there exists an $M > 0$ such that

(5.19) $$\mathbb{P}\left\{\hat{F}_n(t_0 + Mn^{-1/3}) \leq F_0(t_0)\right\} < \epsilon,$$

for all large n, we get that

$$\hat{F}_n(t_0 - Mn^{-1/3}) < F_0(t_0) < \hat{F}_n(t_0 + Mn^{-1/3}),$$

with probability $\geq 1 - 2\epsilon$ for all large n and a suitably chosen M, implying that $\hat{F}_n$ has a jump in $[t_0 - Mn^{-1/3}, t_0 + Mn^{-1/3}]$.

For the proof of (5.18), we will assume, as in Proposition 1.1, that $\delta_{(1)} = 1$ and $\delta_{(n)} = 0$, since, by the assumptions on F_0 and G, we can always assume that (for large

n) there exist observation times T_i and T_j such that $T_i < T_j$, $\delta_i = 1_{\{X_i \le T_i\}} = 1$ and $\delta_j = 1_{\{X_j > T_j\}} = 0$ (see the discussion in the first paragraph of the proof of Proposition 1.2).

Let τ_n be the last jump time of $\hat{F}_n$ before $t_0 - Mn^{-1/3}$, i.e.,

$$\tau_n = \max\{t \le t_0 - Mn^{-1/3} : \hat{F}_n(t-) \ne \hat{F}_n(t)\}.$$

By the assumption just made, τ_n is, for each large n, the maximum over a non-empty set, since $T_{(1)}$ belongs to that set for all large n. By Proposition 1.1, we must have

$$\sum_{\tau_n \le T_i < T_{(j)}} \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} - \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} \right\} \ge 0,$$

for all $T_{(j)} > \tau_n$. Hence, if $\hat{F}_n(t_0 - Mn^{-1/3}) \ge F_0(t_0)$, we get

$$(5.20) \quad \sum_{\tau_n \le T_i < T_{(j)}} \left\{ \frac{\delta_i}{F_0(t_0)} - \frac{1 - \delta_i}{1 - F_0(t_0)} \right\} \ge \sum_{\tau_n \le T_i < T_{(j)}} \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} - \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} \right\} \ge 0,$$

for all $T_{(j)} > \tau_n$. But we have

$$(5.21) \quad \begin{aligned} &\sum_{\tau_n \le T_i < t_0 - \frac{1}{2}Mn^{-1/3}} \left\{ \frac{\delta_i}{F_0(t_0)} - \frac{1 - \delta_i}{1 - F_0(t_0)} \right\} \\ &= \sum_{\tau_n \le T_i < t_0 - \frac{1}{2}Mn^{-1/3}} \left\{ \frac{F_0(T_i)}{F_0(t_0)} - \frac{1 - F_0(T_i)}{1 - F_0(t_0)} + \frac{\delta_i - F_0(T_i)}{F_0(t_0)(1 - F_0(t_0))} \right\}. \end{aligned}$$

The right-hand side of (5.21) can be written

$$\{F_0(t_0)(1 - F_0(t_0))\}^{-1} \sum_{\tau_n \le T_i < t_0 - \frac{1}{2}Mn^{-1/3}} \left\{ \delta_i - F_0(T_i) - (F_0(t_0) - F_0(T_i)) \right\},$$

and by a similar technique as used to derive (5.14) (using a martingale with time running backward), we get

$$\mathbb{P}\left\{ \sup_{t \le t_0 - Mn^{-1/3}} \sum_{t \le T_i \le t_0 - \frac{1}{2}Mn^{-1/3}} \left\{ \delta_i - F_0(T_i) - (F_0(t_0) - F_0(T_i)) \right\} \ge 0 \right\} < \epsilon,$$

for M and n large. This, however, contradicts (5.20), and (5.18) follows.

The remaining part of the proposition is proved in a similar way. Using the same arguments as above, one can show that, for large $M > 0$, $\hat{F}_n$ will have with high probability a jump in the interval

$$[t_0 - 2Mn^{-1/3}, t_0 - Mn^{-1/3}].$$

60

Let τ_n be such a jump time, and suppose that

$$\hat{F}_n(\tau_n) \leq F_0(t_0 - c \cdot n^{-1/3}),$$

where $c > 2M$. By Proposition 1.1 we have

$$(5.22) \qquad \sum_{T_{(j)} < T_i \leq \tau_n} \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} - \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} \right\} \leq 0,$$

for all $T_{(j)} < \tau_n$. On the other hand, if $\hat{F}_n(\tau_n) \leq F_0(t_0 - c \cdot n^{-1/3})$, we get

$$\sum_{t_0 - c \cdot n^{-1/3} < T_i \leq \tau_n} \left\{ \frac{\delta_i}{\hat{F}_n(T_i)} - \frac{1 - \delta_i}{1 - \hat{F}_n(T_i)} \right\}$$

$$\geq \sum_{t_0 - c \cdot n^{-1/3} < T_i \leq \tau_n} \left\{ \frac{\delta_i}{F_0(t_0 - c \cdot n^{-1/3})} - \frac{1 - \delta_i}{1 - F_0(t_0 - c \cdot n^{-1/3})} \right\},$$

and

$$n^{-1/3} \sum_{t_0 - c \cdot n^{-1/3} < T_i \leq \tau_n} \left\{ \frac{\delta_i}{F_0(t_0 - c \cdot n^{-1/3})} - \frac{1 - \delta_i}{1 - F_0(t_0 - c \cdot n^{-1/3})} \right\}$$

$$(5.23) \qquad = n^{-1/3} \sum_{t_0 - c \cdot n^{-1/3} < T_i \leq \tau_n} \left\{ \frac{F_0(T_i)}{F_0(t_0 - c \cdot n^{-1/3})} - \frac{1 - F_0(T_i)}{1 - F_0(t_0 - c \cdot n^{-1/3})} \right\}$$

$$+ n^{-1/3} \sum_{t_0 - c \cdot n^{-1/3} < T_i \leq \tau_n} \frac{\delta_i - F_0(T_i)}{F_0(t_0 - c \cdot n^{-1/3})(1 - F_0(t_0 - c \cdot n^{-1/3}))}.$$

The last term at the right-hand side of (5.23) has expectation zero and a variance which is asymptotically bounded above by

$$(c - M)g(t_0) / \{ F_0(t_0)(1 - F_0(t_0)) \}.$$

The first term at the right-hand side of (5.23) is (almost surely) asymptotically bounded below by

$$\tfrac{1}{2}(c - 2M)^2 f_0(t_0)g(t_0) / \{ F_0(t_0)(1 - F_0(t_0)) \}.$$

This implies that, with high probability, (5.23) will be strictly positive, for large n and large c (with c only depending on M). This contradicts (5.23). Hence, with high probability

$$(5.24) \qquad \hat{F}_n(t_0 - Mn^{-1/3}) \geq \hat{F}_n(\tau_n) \geq F_0(t_0 - cn^{-1/3}).$$

Combining (5.18) and (5.24) we find that for each $\epsilon > 0$ there exist positive constants c and M such that

$$I\!P\{ F_0(t_0 - cn^{-1/3}) \leq \hat{F}_n(t_0 - Mn^{-1/3}) \leq F_0(t_0) \} > 1 - \epsilon,$$

for all sufficiently large n. Likewise we get that for each $\epsilon > 0$ there exist positive constants c and M such that

$$\mathbb{P}\left\{F_0(t_0 + cn^{-1/3}) \geq \hat{F}_n(t_0 + Mn^{-1/3}) \geq F_0(t_0)\right\} > 1 - \epsilon.$$

The lemma now follows from the monotonicity of $\hat{F}_n$. ∎

We can now give a proof of Theorem 5.1.

PROOF OF THEOREM 5.1: Fix an interval $[t_0 - Mn^{-1/3}, t_0 + Mn^{-1/3}]$, where M is a positive constant, and let τ_n^- and τ_n^+ be, respectively, the last jump time $\leq t_0 - Mn^{-1/3}$ and the first jump time $\geq t_0 + Mn^{-1/3}$ of $\hat{F}_n$. By Lemma 5.4 and the strict monotonicity of F_0 at t_0, we have $\tau_n^- - t_0 = \mathcal{O}_p(n^{-1/3})$ and likewise $\tau_n^+ - t_0 = \mathcal{O}_p(n^{-1/3})$. The convex minorant of the process $V_{\hat{F}_n}$, defined by (1.30) coincides on the interval $(\tau_n^-, \tau_n^+]$ with the convex minorant of the process $V_{\hat{F}_n}$, *restricted to the interval* $(\tau_n^-, \tau_n^+]$. Moreover, also by Lemma 5.4, the process

(5.25)

$$t \mapsto n^{2/3}\left\{W_{\hat{F}_n}(t_0 + n^{-1/3}t) - W_{\hat{F}_n}(t_0)\right\}$$
$$+ n^{2/3}\int_{[0, t_0 + n^{-1/3}t]}(\hat{F}_n(t') - F_0(t_0))\, dG_{\hat{F}_n}(t') - n^{2/3}\int_{[0, t_0]}(\hat{F}_n(t') - F_0(t_0))\, dG_{\hat{F}_n}(t')$$

converges in distribution (in the topology of uniform convergence on compacta) to the process $\{U(t) : t \in \mathbb{R}\}$, defined by (5.4). This means that the process

(5.26)
$$t \mapsto n^{1/3}\left\{\hat{F}_n(t_0 + n^{-1/3}t) - F_0(t_0)\right\}$$

will converge in distribution (in the Skorohod topology) to the left derivative of the convex minorant of the process U, since, for each $M > 0$, jump times τ_n^- and τ_n^+ can be found, defined as above, such that the process (5.26) coincides on the interval $(-n^{1/3}(t_0 - \tau_n^-), n^{1/3}(\tau_n^+ - t_0)]$ with the left derivative of the convex minorant of the process (5.25), restricted to this interval. This shows that the process (5.26) has the same limiting distribution as the process

(5.27)
$$t \mapsto n^{1/3}\left\{F_n^{(1)}(t_0 + n^{-1/3}t) - F_0(t_0)\right\}.$$

Hence we get from Theorem 5.2:

$$\lim_{n \to \infty} \mathbb{P}\left\{n^{1/3}\{\hat{F}_n(t_0) - F_0(t_0)\} > x\right\} = \mathbb{P}\{T(0) > c \cdot x\},$$

where $T(0)$ is defined by (5.9), and $c = f_0(t_0)^{-1}$ (see (5.17)). Thus

(5.28)
$$\mathbb{P}\left\{n^{1/3}\{\hat{F}_n(t_0) - F_0(t_0)\}/\left\{\tfrac{1}{2}F_0(t_0)(1 - F_0(t_0))f_0(t_0)/g(t_0)\right\}^{1/3} > x\right\}$$
$$\to \mathbb{P}\left\{T(0) > x\left\{\tfrac{1}{2}F_0(t_0)(1 - F_0(t_0))/\{g(t_0)f_0(t_0)^2\}\right\}^{1/3}\right\}, \ n \to \infty.$$

Now $T(0)$ is the last time where the process

$$t \mapsto W\Big(t \cdot g(t_0)/(F_0(t_0)(1 - F_0(t_0)))\Big) + \tfrac{1}{2}\frac{f_0(t_0)g(t_0)}{F_0(t_0)(1 - F_0(t_0))}t^2$$

reaches its minimum. By Brownian scaling (i.e., the property that $t \mapsto W(t)$ has the same distribution as $t \mapsto c^{-1/2}W(ct)$, for each constant $c > 0$), this means that

$$(5.29) \qquad \tfrac{1}{2}T(0)\Big/\Big\{\tfrac{1}{2}F_0(t_0)(1 - F_0(t_0))/\{g(t_0)f_0(t_0)^2\}\Big\}^{1/3}$$

is the last time where

$$t \mapsto W(t) + t^2$$

reaches its minimum. Using the symmetry of the distribution of Brownian motion with repect to the time axis, this means that (5.29) has the same distribution as the last time where

$$t \mapsto W(t) - t^2$$

reaches its maximum. The result now follows from (5.28). ∎

5.2. Interval censoring, Case 2.

We shall prove the following result for the 1-step estimator $F_n^{(1)}$.

THEOREM 5.3. *Let F_0 and H be differentiable at t_0 and (t_0, t_0), respectively, with strictly positive derivatives $f_0(t_0)$ and $h(t_0, t_0)$. By continuous differentiability of H at (t_0, t_0) is meant that the density $h(t, u)$ is continuous in (t, u) if $t < u$ and (t, u) is sufficiently close to (t_0, t_0) and that $h(t, t)$, defined by*

$$h(t, t) = \lim_{u \downarrow t} h(t, u),$$

is continuous in t, for t in a neighborhood of t_0.

Let $0 < F_0(t_0), H(t_0, t_0) < 1$, and let $F_n^{(1)}$ be the estimator of F_0, obtained at the first step of the iterative convex minorant algorithm. Then

$$(n \log n)^{1/3}\big\{F_n^{(1)}(t_0) - F_0(t_0)\big\}\Big/\big\{\tfrac{3}{4}f_0(t_0)^2/h(t_0, t_0)\big\}^{1/3} \xrightarrow{\mathcal{D}} 2Z,$$

where Z is the last time where standard two-sided Brownian motion minus the parabola $y(t) = t^2$ reaches its maximum.

According to our working hypothesis, formulated at beginning of this chapter, this leads us to believe that the NPMLE has the same limiting behavior at t_0 as $F_n^{(1)}$. Proceeding as before, we introduce the processes $W_n^{(0)}$ and $G_n^{(0)}$ defined by

$$W_n^{(0)} = W_{F_0}, \quad \text{and} \quad G_n^{(0)} = G_{F_0},$$

where W_{F_0} and G_{F_0} are defined by (1.25) and (1.29), respectively. The process $V_n^{(0)}$ is defined as in (5.3). We have the following result for $V_n^{(0)}$.

LEMMA 5.5. *Let $\delta_n = (n \log n)^{-1/3}$ and let the process $U_n^{(0)}$ be defined by*

$$U_n^{(0)}(t) = \delta_n^{-2}(\log n)^{-1}\left\{V_n^{(0)}(t_0 + \delta_n t) - V_n^{(0)}(t_0)\right\}, \ t \in I\!R,$$

where $U_n^{(0)}(t) = 0$, if $t \leq -t_0\delta_n^{-1}$. Then $U_n^{(0)}$ converges in distribution, in the topology of uniform convergence on compacta on the space of locally bounded real-valued functions on $I\!R$, to the process U, defined by

$$(5.30) \qquad U(t) = \sqrt{\tfrac{2}{3}h(t_0,t_0)/f_0(t_0)}\,W(t) + \tfrac{1}{3}h(t_0,t_0)t^2, \ t \in I\!R,$$

where W is (standard) two-sided Brownian motion on $I\!R$, originating from zero.

PROOF: We first show that the process

$$(5.31) \qquad t \mapsto \delta_n^{-2}(\log n)^{-1}\left\{W_n^{(0)}(t_0 + \delta_n t) - W_n^{(0)}(t_0)\right\}, \ t \geq 0$$

converges, in the topology of uniform convergence on compacta, to the process

$$(5.32) \qquad t \mapsto \sqrt{\tfrac{2}{3}h(t_0,t_0)/f_0(t_0)}\,W(t), \ t \geq 0.$$

It is clear that the process (5.31) is not a martingale, but by removing some terms, it can be made a martingale. Fix $M > 0$ and define the sets of observation times A_n and B_n by

$$(5.33) \qquad \begin{aligned} A_n =&\{T_i \in (t_0, t_0 + M\delta_n] : U_i - T_i > M\delta_n\} \\ &\cup \{U_i \in (t_0, t_0 + M\delta_n] : U_i - T_i > M\delta_n\}, \end{aligned}$$

and

$$(5.34) \qquad B_n = \{T_i \in (t_0, t_0 + M\delta_n] : T_i \notin A_n\} \cup \{U_i \in (t_0, t_0 + M\delta_n] : U_i \notin A_n\}.$$

Furthermore, define the processes $W_{n,1}$ and $W_{n,2}$ on $[0, M]$ by

$$W_{n,1}(t) = \delta_n^{-2}(\log n)^{-1}\int_{t' \in (t_0, t_0 + \delta_n t], t' \in A_n} dW_n^{(0)}(t'),$$

$$W_{n,2}(t) = \delta_n^{-2}(\log n)^{-1}\int_{t' \in (t_0, t_0 + \delta_n t], t' \in B_n} dW_n^{(0)}(t'),$$

for $t \in [0, M]$. Then we get

$$\max_{t \in [0,M]} |W_{n,2}(t)| \leq \delta_n^{-2}(n \log n)^{-1} \sum_{T_{(i)} \in B_n} |\Delta W_{n,2}(T_{(i)})|,$$

where the $\Delta W_{n,2}(T_{(i)})$ are the jumps of the process $W_{n,2}$. Computing expectations and using Markov's inequality, we obtain, for $\epsilon > 0$,

$$\mathbb{P}\left\{\max_{t \in [0,M]} |W_{n,2}(t)| > \epsilon\right\} \leq 2\epsilon^{-1}\delta_n^{-2}(\log n)^{-1}\int_{t,u \in (t_0, t_0 + M\delta_n]} dH(t,u) = \mathcal{O}((\log n)^{-1}).$$

On the other hand, the process $t \mapsto W_{n,1}(t)$ is a martingale, and a straightforward computation shows that the variance of $W_{n,1}(t)$ is given by

$$n\delta_n^2 \int_{t' \in (t_0, t_0 + t\delta_n], \, u - t' > M\delta_n} \left\{\frac{1}{F_0(t')} + \frac{1}{F_0(u) - F_0(t')}\right\} dH(t', u)$$

$$+ n\delta_n^2 \int_{u \in (t_0, t_0 + t\delta_n], \, u - t' > M\delta_n} \left\{\frac{1}{F_0(u) - F_0(t')} + \frac{1}{1 - F_0(u)}\right\} dH(t', u),$$

which is asymptotically equivalent to

$$\tfrac{2}{3}h(t_0, t_0)/f_0(t_0)\, t, \ t > 0.$$

This shows that the process (5.31) converges to the Wiener process (5.32). Since a similar argument holds for negative values of t, we get that the process

$$t \mapsto \delta_n^{-2}(\log n)^{-1}\left\{W_n^{(0)}(t_0 + \delta_n t) - W_n^{(0)}(t_0)\right\},$$

converges, in the topology of uniform convergence on compacta, to the process

$$t \mapsto \sqrt{\tfrac{2}{3}h(t_0, t_0)/f_0(t_0)}\, W(t), \ t \in \mathbb{R}.$$

Since, for each $M > 0$

$$\mathbb{P}\{T_i < X_i \leq U_i, \ T_i, U_i \in [t_0, t_0 + M\delta_n]\} \sim \tfrac{1}{6}f_0(t_0)h(t_0, t_0)M^3\delta_n^3,$$

we get

$$\mathbb{P}\{T_i < X_i \leq U_i, \ T_i, U_i \in [t_0, t_0 + M\delta_n], \text{ for some } i\}$$
$$\leq n\mathbb{P}\{T_1 < X_1 \leq U_1, \ T_1, U_1 \in [t_0, t_0 + M\delta_n]\}$$
$$\sim \tfrac{1}{6}h(t_0, t_0)M^3(\log n)^{-1} \to 0, \text{ as } n \to \infty.$$

Furthermore, with probability one,

$$\delta_n^{-2}(\log n)^{-1}\int_{t' \in [t_0, t_0 + \delta_n t], t' < x \leq u, u - t' > M\delta_n} \frac{F_0(t') - F_0(t_0)}{\big(F_0(u) - F_0(t')\big)^2} dP_n(x, t', u) \to \tfrac{1}{6}h(t_0, t_0)t^2,$$

and likewise

$$\delta_n^{-2}(\log n)^{-1}\int_{u \in [t_0, t_0 + \delta_n t], t' < x \leq u, u - t' > M\delta_n} \frac{F_0(u) - F_0(t_0)}{\big(F_0(u) - F_0(t')\big)^2} dP_n(x, t', u) \to \tfrac{1}{6}h(t_0, t_0)t^2,$$

65

uniformly for t in a bounded interval $[0, M]$, as $n \to \infty$. Since similar relations hold for $t < 0$, the result now follows. ∎

Next we define, in analogy with (5.8), for each $a > 0$, the random variable $T_n^{(0)}(a)$ by

$$T_n^{(0)}(a) = \sup\{t \in J_n : V_n^{(0)}(t) - (a - F_0(t_0))G_n^{(0)}(t) \text{ is minimal}\},$$

where the index set J_n is defined as in Definition 1.1. Likewise, in analogy with (5.9), we define the process $\{T(a) : a \in I\!\!R\}$ by

$$(5.35) \qquad T(a) = \sup\left\{t \in I\!\!R : U(t) - a \cdot \frac{2h(t_0, t_0)t}{3f_0(t_0)} \text{ is minimal}\right\}.$$

Lemma 5.2 now leads us to expect that

$$\left\{\delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n a) - t_0\} : a \in I\!\!R\right\}$$

will converge in distribution to the process $\{T(a) : a \in I\!\!R\}$. This is indeed the case. The proof is similar to the proof of Lemma 5.2, where the (crucial) Lemma 5.3 is replaced by the following lemma.

LEMMA 5.6. *For each $\epsilon > 0$ and $M_1 > 0$ an $M_2 > 0$ can be found such that*

$$I\!\!P\left\{\max_{a \in [-M_1, M_1]} \delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n a) - t_0\} > M_2\right\} < \epsilon,$$

and

$$I\!\!P\left\{\min_{a \in [-M_1, M_1]} \delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n a) - t_0\} < -M_2\right\} < \epsilon,$$

for all sufficiently large n.

PROOF: We (again) only prove the first inequality. First note that

$$I\!\!P\left\{\max_{a \in [-M_1, M_1]} \delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n a) - t_0\} > M_2\right\}$$
$$= I\!\!P\left\{\delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n M_1) - t_0\} > M_2\right\}.$$

Furthermore,

$$I\!\!P\left\{\delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n M_1) - t_0\} > M_2\right\}$$
$$\leq I\!\!P\left\{U_n^{(0)}(t) - (\delta_n \log n)^{-1}M_1\{G_n^{(0)}(t_0 + \delta_n t) - G_n^{(0)}(t_0)\} < 0, \text{ for some } t \geq M_2\right\}.$$

Similarly as in the proof of Lemma 5.3, we have

$$
\begin{aligned}
\delta_n^2(\log n) & U_n^{(0)}(t) - \delta_n(\log n)M_1 \cdot \left\{ G_n^{(0)}(t_0 + \delta_n t) - G_n^{(0)}(t_0) \right\} \\
&= V_n^{(0)}(t_0 + \delta_n t) - V_n^{(0)}(t_0) - \delta_n^{-1} M_1 \left\{ G_n^{(0)}(t_0 + \delta_n t) - G_n^{(0)}(t_0) \right\} \\
&= W_n^{(0)}(t_0 + \delta_n t) - W_n^{(0)}(t_0) \\
&\quad + \int_{[t_0, t_0 + \delta_n t]} \left\{ F_0(t') - F_0(t_0) - \delta_n M_1 \right\} dG_n^{(0)}(t').
\end{aligned}
$$

Let the process X_n be defined by

$$
X_n(t) = \delta_n^{-2}(\log n)^{-1} \left\{ W_n^{(0)}(t_0 + t) - W_n^{(0)}(t_0) \right\}, \ t \geq 0.
$$

We first derive an upper bound for the probabilities

$$
I\!\!P \left\{ |X_n(j\delta_n)| > \epsilon j^2 + m \right\}, j = 1, 2, \ldots,
$$

where ϵ and m are arbitrarily chosen positive numbers.

Let the sets of observation times $A_{n,j}$ and $B_{n,j}$ be defined by

$$
\begin{aligned}
A_{n,j} &= \left\{ T_i \in (t_0, t_0 + j\delta_n] : F_0(U_i) - F_0(T_i) > \delta_n \right\} \\
&\quad \cup \left\{ U_i \in (t_0, t_0 + j\delta_n] : F_0(U_i) - F_0(T_i) > \delta_n \right\}, \\
B_{n,j} &= \left\{ T_i \in (t_0, t_0 + j\delta_n] : T_i \notin A_{n,j} \right\} \cup \left\{ U_i \in (t_0, t_0 + j\delta_n] : U_i \notin A_{n,j} \right\}.
\end{aligned}
$$

Furthermore, let the random variables $X'_{n,j}$ and $X''_{n,j}$ be defined by

$$
X'_{n,j} = \int_{t \in (0, j\delta_n], \, t \in A_{n,j}} dX_n(t),
$$

and

$$
X''_{n,j} = \int_{t \in (0, j\delta_n], \, t' \in B_{n,j}} dX_n(t).
$$

We have:

$$
\begin{aligned}
I\!\!P \left\{ |X'_{n,j}| > \tfrac{1}{2}(\epsilon j^2 + m) \right\} &\leq 4 I\!\!P \left(X'_{n,j}(j\delta_n) \right)^2 / (\epsilon j^2 + m)^2 \\
&\leq c \cdot n^{-1} \delta_n^{-4}(\log n)^{-1} \left\{ j\delta_n \right\} / (\epsilon j^2 + m)^2 \\
&= c \cdot j / (\epsilon j^2 + m)^2,
\end{aligned}
$$

and

$$
\begin{aligned}
I\!\!P \left\{ |X''_{n,j}| > \tfrac{1}{2}(\epsilon j^2 + m) \right\} &\leq 2 I\!\!P |(X''_{n,j}| / (\epsilon j^2 + m) \\
&\leq c \cdot \delta_n^{-2}(\log n)^{-1} \delta_n^2 / (\epsilon j^2 + m) \\
&= c / \left\{ (\epsilon j^2 + m) \log n \right\},
\end{aligned}
$$

67

for some constant $c > 0$, not depending on j, if $j\delta_n$ satisfies $j\delta_n \leq \delta$, and $\delta > 0$ is chosen in such a way that $t \mapsto f_0(t)$ and $t \mapsto h(t,t)$ satisfy the local positivity and continuity conditions of Theorem 5.3 in the neighborhood $[t_0, t_0 + \delta]$. Moreover, we use the fact that, if $t_0 < T_i < U_i \leq t_0 + j\delta_n$, terms of the form $\pm\{T_i < X_i \leq U_i\}/\{F_0(U_i) - F_0(T_i)\}$ will give no contribution to $X'_{n,j}$ or $X''_{n,j}$, since these terms will occur with opposite signs and will cancel in the summation.

We also have:

$$\mathbb{P}\Big\{\sup_{t\in[0,\delta_n]}|W_{n,1}(t)| > \tfrac{1}{2}(\epsilon j^2 + m)\Big\} \leq 4\mathbb{P}\big(W_{n,1}(\delta_n)\big)^2/(\epsilon j^2 + m)^2$$
$$\leq c/(\epsilon j^2 + m)^2,$$

and

$$\mathbb{P}\Big\{\sup_{t\in[0,\delta_n]}|W_{n,2}(t)| > \tfrac{1}{2}(\epsilon j^2 + m)\Big\} \leq c/\big\{(\epsilon j^2 + m)\log n\big\},$$

for some constant $c > 0$, where $W_{n,1}$ and $W_{n,2}$ are defined as in the proof of Lemma 5.5, with $M = 1$. Hence

$$\mathbb{P}\Big\{\sup_{t\in[0,\delta_n]}|X_n(t)| > \tfrac{1}{2}(\epsilon j^2 + m)\Big\} \leq c/(\epsilon j^2 + m)^2 + c/\big\{(\epsilon j^2 + m)\log n\big\}.$$

It is clear that we get the same kind of upper bound for

$$\mathbb{P}\Big\{\sup_{t\in((j-1)\delta_n,j\delta_n]}|X_n(t) - X_n((j-1)\delta_n)| > \tfrac{1}{2}(\epsilon j^2 + m)\Big\},$$

if $j\delta_n \leq \delta$. Combining these results, we obtain that for $\epsilon > 0$ and $m > 0$:

$$\mathbb{P}\big\{\exists t \in ((j-1)\delta_n, j\delta_n] : |X_n(t)| > \epsilon(j-1)^2 + m\big\}$$
$$\leq c \cdot j/\big\{\epsilon(j-1)^2 + m\big\}^2 + c/\big\{(\epsilon(j-1)^2 + m)\log n\big\},$$

if $j\delta_n \leq \delta$. As in the proof of Lemma 5.3 (see (5.14)), this yields

$$\delta_n^2|X_n(t)| \leq \epsilon t^2 + \delta_n^2 R_n, \text{ if } 0 \leq t \leq \delta,$$

where $R_n = \mathcal{O}_p(1)$.

As in (5.15), there exist, for each $M > 0$ and $\eta > 0$ an $M_2 > 0$ and $\epsilon > 0$ such that

$$(\log n)^{-1}\int_{[t_0,t_0+u]}\big\{F_0(t') - F_0(t_0) - \delta_n M_1\big\}dG_n^{(0)}(t') \geq \max\{M\delta_n^2, \epsilon u^2\},$$

$$\forall u \in [M_2\delta_n, \delta],$$

with probability bigger than $1 - \eta$. This is easily checked by taking a constant $k > 0$ such that

$$F_0(k\delta_n) - F_0(t_0) \geq \delta_n M_1,$$

and by splitting the points of increase of $G_n^{(0)}$ on the interval $[t_0, t_0 + k\delta_n]$ into two sets A_n and B_n, as in (5.33) and (5.34), with M replaced by constant k. On the interval $[t_0, t_0 + k\delta_n]$ we replace the integral by an integral over points T_i and U_i, belonging to the set A_n, and use that the probability of having an observation point in B_n, giving a non-zero contribution to the integral, will tend to zero, as $n \to \infty$. On the remaining interval we also remove points T_i and U_i such that $U_i - T_i < \delta_n$, showing that, with a probability tending to one, the integral

$$(\log n)^{-1} \int_{[t_0, t_0+u]} \{F_0(t') - F_0(t_0) - \delta_n M_1\} dG_n^{(0)}(t')$$

is bounded below by a term which is asymptotically equivalent to

$$\frac{2}{3} \int_{[t_0, t_0+u]} \{F_0(t') - F_0(t_0) - \delta_n M_1\} \frac{h(t,t)}{f_0(t)} dt.$$

Combining the preceding results, we obtain

$$\mathbb{P}\Big\{ U_n^{(0)}(t) - \delta_n^{-1} M_1 (\log n)^{-1} \{G_n^{(0)}(t_0 + \delta_n t) - G_n^{(0)}(t_0)\} < 0,$$

$$\text{for some } t, \ M_2 \delta_n \leq t \leq \delta \Big\}$$

$$\leq \mathbb{P}\{A_n > M\} + \eta,$$

and the last expression can be made smaller than 2η, by taking M sufficiently large.

Finally, we have to deal with the interval $[t_0 + \delta, \infty)$. But on this interval we will only get jumps downward, if

$$(5.36) \qquad -\Delta W_n^{(0)}(T_{(i)}) \geq \{F_0(T_{(i)}) - F_0(t_0) - \delta_n M_1\} \Delta G_n^{(0)}(T_{(i)}),$$

where $T_{(i)}$ is an observation point in $[t_0 + \delta, \infty)$, and where $\Delta W_n^{(0)}(T_{(i)})$ and $\Delta G_n^{(0)}(T_{(i)})$ are jumps of the processes $W_n^{(0)}$ and $G_n^{(0)}$, respectively. But this can only happen if

$$n \big| \Delta W_n^{(0)}(T_{(i)}) \big| \leq \{F_0(t_0 + \delta) - F_0(t_0) - \delta_n M_1\}^{-1},$$

implying that the sum of the terms $\Delta W_n^{(0)}(T_{(i)})$, satisfying (5.36), for $T_{(i)} \in [t_0 + \delta, \infty)$, is bounded below by a fixed constant. But since, for some constant c, $0 < c < 1$,

$$\int_{[t_0, t_0+\delta]} \{F_0(t) - F_0(t_0) - \delta_n M_1\} dG_n^{(0)}(t) > c \big| W_n^{(0)}(t_0 + \delta) - W_n^{(0)}(t_0) \big|$$

with a probability tending to one, and since

$$\int_{[t_0, t_0+\delta]} \{F_0(t) - F_0(t_0) - \delta_n M_1\} dG_n^{(0)}(t)$$

will tend to ∞ with probability one, we will get

$$W_n^{(0)}(t_0 + t) - W_n^{(0)}(t_0) + \int_{[t_0, t_0+t]} \{F_0(t') - F_0(t_0) - \delta_n M_1\} dG_n^{(0)}(t') > 0, \forall t \geq t_0 + \delta,$$

with probability tending to one. ∎

Using Lemma 5.6, we can now give a proof of Theorem 5.3. The proof is quite similar to the proof of Theorem 5.2.

PROOF OF THEOREM 5.3: First of all, similarly to (5.16), we have:

$$\mathbb{P}\{F_n^{(1)}(t_0) - F_0(t_0) > \delta_n a\} = \mathbb{P}\{T_n^{(0)}(a_0 + \delta_n a) < t_0\}.$$

Moreover, using Lemmas 5.5 and 5.6, we get that the process

$$\left\{\delta_n^{-1}\{T_n^{(0)}(a_0 + \delta_n a) - t_0\} : a \in \mathbb{R}\right\}$$

converges in the Skorohod topology on $D(\mathbb{R})$ to the process $\{T(a) : a \in \mathbb{R}\}$, defined by (5.35). As before, the process

$$\{T(a) - f_0(t_0)^{-1}a : a \in \mathbb{R}\}$$

is stationary, since $T(a)$ is the last time where

$$t \mapsto \sqrt{\tfrac{2}{3}h(t_0, t_0)/f_0(t_0)}\, W(t) + \tfrac{1}{3}h(t_0, t_0)\big(t - f_0(t_0)^{-1}a\big)^2,\ t \in \mathbb{R},$$

is minimal. Hence,

$$\mathbb{P}\left\{\delta_n^{-1}\{F_n^{(1)}(t_0) - F_0(t_0)\}\big/\{\tfrac{3}{4}f_0(t_0)^2/h(t_0, t_0)\}^{1/3} > x\right\}$$
$$\rightarrow \mathbb{P}\left\{T(0) > x\{\tfrac{4}{3}f_0(t_0)h(t_0, t_0)\}^{-1/3}\right\},\ n \rightarrow \infty.$$

Now, $T(0)$ is the last time where

$$t \mapsto W\big(\tfrac{2}{3}h(t_0, t_0)f_0(t_0)\big) + \tfrac{1}{3}h(t_0, t_0)f_0(t_0)t^2$$

reaches its minimum. By Brownian scaling, this means that

$$\tfrac{1}{2}T(0)\{\tfrac{4}{3}f_0(t_0)h(t_0, t_0)\}^{1/3} = T(0)\{\tfrac{1}{6}f_0(t_0)h(t_0, t_0)\}^{1/3}$$

is the last time where

$$t \mapsto W(t) + t^2,\ t \in \mathbb{R},$$

reaches its minimum. ∎

As a final step in proving that the NPMLE $\hat{F}_n$ has the same limiting behavior as $F_n^{(1)}$, one would have to prove two things:

70

(1) $\hat{F}_n(t) - F_0(t_0) = \mathcal{O}_p(\delta_n)$, for t in an interval of the form $[t_0 - M\delta_n, t_0 + M\delta_n]$ (compare with Lemma 5.4).

(2) The process

$$t \mapsto \delta_n^{-2} \left\{ W_{\hat{F}_n}(t_0 + \delta_n t) - W_{\hat{F}_n} \right\}$$

(5.37)
$$+ \delta_n^{-2} \int_{[0, t_0 + \delta_n t]} (\hat{F}_n(t') - F_0(t_0)) \, dG_{\hat{F}_n}(t')$$

$$- \delta_n^{-2} \int_{[0, t_0]} (\hat{F}_n(t') - F_0(t_0)) \, dG_{\hat{F}_n}(t')$$

has the same asymptotic behavior as the process $U_n^{(0)}$, defined in Lemma (5.5) (note that $U_n^{(0)}$ is (5.37), with $\hat{F}_n$ replaced by $F_n^{(1)}$). Assuming that the δ_n^{-1}-consistency of property 1 holds, we can, for $t \geq 0$, write (5.37) in the form

$$U_n^{(0)}(t) - \delta_n^{-2} \int_{t_0 < t' \leq t_0 + \delta_n t, \, t' < x \leq u} \frac{\hat{F}_n(u) - F_0(u)}{(F_0(u) - F_0(t'))^2} \, dP_n(x, t', u)$$

$$- \delta_n^{-2} \int_{t_0 < u \leq t_0 + \delta_n t, \, t' < x \leq u} \frac{\hat{F}_n(t') - F_0(t')}{(F_0(u) - F_0(t'))^2} \, dP_n(x, t', u) + R_n(t),$$

where the process $t \mapsto R_n(t)$ will vanish in the limit, with a similar expansion for $t < 0$. The reason for believing that the other two processes added to $U_n^{(0)}(t)$ will also vanish in the limit, is that the integrands will have both positive and negative terms at the points t which have positive mass (since, roughly speaking, the expectation of $\hat{F}_n(t) - F_0(t)$ will be zero), and that $\hat{F}_n(T_{(i)}) - F_0(T_{(i)})$ and $\hat{F}_n(T_{(j)}) - F_0(T_{(j)})$ will (most likely) be almost independent for $T_{(i)}$ and $T_{(j)}$ far apart, (say $|T_{(i)} - T_{(j)}| > n^{-1/3} \log n$). The latter property is in contrast with the behavior of a term like

$$\delta_n^{-2} \int_{t_0 < t' \leq t_0 + \delta_n t, \, t' < x \leq u} \frac{\hat{F}_n(t') - F_0(t')}{(F_0(u) - F_0(t'))^2} \, dP_n(x, t', u),$$

where the restriction of t' to the interval $(t_0, t_0 + \delta_n t]$ causes strong dependence of terms of the above type.

The conjectures mentioned here have actually been supported by computer experiments, which also gave support of the conjecture that the asymptotic variance of the NPMLE at t_0 has the form

$$\left\{ \tfrac{3}{4} f_0(t_0)^2 / h(t_0, t_0) \right\}^{2/3} 4(EZ^2) \delta_n^2,$$

given in Theorem 5.3.

5.3. Deconvolution with a decreasing density.

We consider the situation, discussed in Section 2.1. In this set-up we shall prove the following result for the (fictitious) 1-step estimator $F_n^{(1)}$.

THEOREM 5.4. *Let g be a right-continuous decreasing density on $[0, \infty)$, having only a finite number of discontinuity points $a_0 = 0 < a_1 < \cdots < a_m$. Moreover, suppose that g has a derivative $g'(x)$ at points $x \neq a_i$, $i = 0, \ldots, m$, satisfying*

$$(5.38) \qquad \int_{(0,\infty)} \frac{g'(x)^2}{g(x)} \, dx < \infty,$$

where the integrand is defined to be zero at the points a_i and at points x where g is zero, and where g' is bounded and continuous on the intervals (a_{i-1}, a_i), $i = 1, \ldots, m+1$, with $a_{m+1} \overset{\text{def}}{=} \infty$.

Furthermore, assume that there exist positive constants k_1 and k_2 such that the derivative g' of g satisfies the relation

$$|g'(t+u)| \leq k_1 |g'(t)|,$$

for all $t > 0$ and $0 < u < k_2$, such that $a_i < t < t + u < a_{i+1}$ for some i, $0 \leq i \leq m$.

Let the convolution density h be given by

$$h(z) = \int g(z - x) \, dF_0(x), \ z \geq 0,$$

where the distribution function F_0 of the (non-negative) random variables $X_i, 1 \leq i \leq n$, is continuously differentiable at $z_0 > 0$, with derivative $f_0(z_0) > 0$ at z_0. Then

$$n^{1/3} \left\{ F_n^{(1)}(z_0) - F_0(z_0) \right\} f_0(z_0)^{-1/3} \left\{ 2 \sum_{i=0}^{m} \left(g(a_i) - g(a_i-) \right)^2 \Big/ h(z_0 + a_i) \right\}^{1/3} \overset{\mathcal{D}}{\to} 2Z,$$

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

Example 5.1. If g is the uniform density on $[0,1]$, then the two discontinuity points are $a_0 = 0$ and $a_1 = 1$. Since we define g to be right-continuous, we get $g(0) - g(0-) = 1$ and $g(1) - g(1-) = -1$. Let $F_0(1) = 1$, i.e., the probability distribution of the X_i's has support contained in $[0,1]$. Theorem 5.4 now yields:

$$n^{1/3} \left\{ F_n^{(1)}(z_0) - F_0(z_0) \right\} \Big/ \left\{ \tfrac{1}{2} F_0(z_0)(1 - F_0(z_0)) f_0(z_0) \right\}^{1/3} \overset{\mathcal{D}}{\to} 2Z,$$

since

$$\sum_{i=0}^{m} \left(g(a_i) - g(a_i-) \right)^2 \Big/ h(z_0 + a_i) = 1 / \left\{ F_0(z_0)(1 - F_0(z_0)) \right\}$$

in this case.

It can also be deduced from Exercise 2 of Chapter 2 that in fact

$$n^{1/3}\left\{\hat{F}_n(z_0) - F_0(z_0)\right\} \Big/ \left\{\tfrac{1}{2}F_0(z_0)(1 - F_0(z_0)f_0(z_0)\right\}^{1/3} \xrightarrow{\mathcal{D}} 2Z,$$

where $\hat{F}_n$ is the NPMLE of F_0. The latter result is also given in van Es (1988), Theorem 4.5, where it is shown that the variance of the NPMLE corresponds to a lower bound for the minimax risk of estimating F_0 at z_0, apart from a constant not depending on F_0 and the density g.

Example 5.2. If g is the exponential density on $[0, \infty)$, with scale parameter 1, Theorem 5.4 yields:

$$n^{1/3}\left\{F_n^{(1)}(z_0) - F_0(z_0)\right\} \Big/ \left\{\tfrac{1}{2}f_0(z_0)h(z_0)\right\}^{1/3} \xrightarrow{\mathcal{D}} 2Z.$$

According to our working hypothesis, we study the process $V_n^{(0)}$, defined by

$$V_n^{(0)}(t) = W_{F_0}(t) + \int_{[0,t]} \left\{F_0(t') - F_0(t_0)\right\} dG_{F_0}(t'), \; t \geq 0,$$

For convenience of notation, we shall denote W_{F_0} by $W_n^{(0)}$ and W_{G_0} by $G_n^{(0)}$. The following limit result holds for the process $V_n^{(0)}$.

LEMMA 5.7. *Let, under the conditions of Theorem 5.4, the process $U_n^{(0)}$ be defined by*

$$U_n^{(0)}(t) = n^{2/3}\left\{V_n^{(0)}(z_0 + n^{-1/3}t) - V_n^{(0)}(z_0)\right\}, \; t \in \mathbb{R},$$

where $U_n^{(0)}(t) = 0$, if $t \leq -n^{1/3}t_0$. Then $U_n^{(0)}$ converges in distribution, in the topology of uniform convergence on compacta on the space of locally bounded real-valued functions on $\mathbb{R}$, to the process U, defined by

$$U(t) = \left\{\sum_{i=0}^{m}(g(a_i) - g(a_i-))^2 / h(z_0 + a_i)\right\}^{1/2} W(t)$$

$$+ \tfrac{1}{2}\left\{\sum_{i=0}^{m}(g(a_i) - g(a_i-))^2 / h(z_0 + a_i)\right\} f_0(z_0)t^2, \; t \in \mathbb{R},$$

where W is (standard) two-sided Brownian motion on $\mathbb{R}$, originating from zero.

PROOF: We have, for $t > 0$:

(5.39)
$$n^{2/3}\left\{W_n^{(0)}(z_0 + tn^{-1/3}) - W_n^{(0)}(z_0)\right\} = n^{2/3}\sum_{z_0 < Z_{(i)} \leq z_0 + tn^{-1/3}} \Delta W_n^{(0)}(Z_{(i)})$$

$$= n^{2/3}\int \frac{g(z - \tau_n^-) - g(z - \tau_n^+)}{\int g(z - x)\, dF_0(x)}\, dH_n(z),$$

where τ_n^- and τ_n^+ are defined by

$$\tau_n^- = \max\{Z_i : Z_i \leq z_0\}, \text{ and } \tau_n^+ = \max\{Z_i : Z_i \leq z_0 + tn^{-1/3}\}.$$

This follows from the fact that

$$\sum_{z_0 < Z_{(i)} \leq z_0 + tn^{-1/3}} \int \frac{g(z - Z_{(i)}) - g(z - Z_{(i+1)})}{\int g(z - x)\,dF_0(x)}\,dH_n(z)$$

$$= \int \frac{g(z - \tau_n^-) - g(z - \tau_n^+)}{\int g(z - x)\,dF_0(x)}\,dH_n(z),$$

(telescoping sums). It now follows from (5.38) and the representation (5.39) that the variance of

$$(5.40) \qquad\qquad n^{2/3}\big\{W_n^{(0)}(z_0 + tn^{-1/3}) - W_n^{(0)}(z_0)\big\}$$

converges to

$$t \cdot \sum_{i=0}^{m}\big(g(a_i) - g(a_i-)\big)^2 / h(z_0 + a_i),$$

since the remaining part of the variance of (5.40) is bounded above by

$$c \cdot n^{-1/3} \int \frac{g'(x)^2}{g(x)}\,dx.$$

Here we use that, for some $\delta > 0$,

$$\int \frac{\big(g(z - z_0) - g(z - z_0 - t \cdot n^{-1/3})\big)^2}{h(z)}\,dz$$

$$\leq \delta^{-1} \int \frac{\big(g(z - z_0) - g(z - z_0 - t \cdot n^{-1/3})\big)^2}{g(z - z_0 - t \cdot n^{-1/3})}\,dz$$

which follows from the fact that the density f_0 satisfies $f_0(z) \geq \epsilon > 0$, for some $\epsilon > 0$, in a neighborhood $[z_0, z_0 + \eta]$ of z_0, implying

$$h(z) \geq \tfrac{1}{2}g(z - z_0 - t \cdot n^{-1/3})\epsilon\eta,$$

for large n. We also use the boundedness and continuity of g' on intervals between points of jump of g and apply the mean value theorem.

Next we note that, under the conditions of the Lemma,

$$\int \frac{g(z - z_0) - g(z - z_0 - tn^{-1/3})}{\int g(z - x)\,dF_0(x)}\,dH_0(z), = 0$$

74

and by empirical process theory it is easily seen that the process

$$t \mapsto n^{2/3} \int \frac{g(z - z_0) - g(z - z_0 - tn^{-1/3})}{\int g(z - x) \, dF_0(x)} \, d(H_n - H)(z)$$

converges in distribution, in the topology of uniform convergence on compacta on the space of locally bounded real-valued functions on $\mathbb{R}$, to the process

$$t \mapsto \left\{ \sum_{i=0}^{m} (g(a_i) - g(a_i-))^2 / h(z_0 + a_i) \right\}^{1/2} W(t), \ t \geq 0.$$

Since $\tau_n^- - z_0$ and $\tau_n^+ - z_0 - tn^{-1/3}$ are $\mathcal{O}_p(n^{-1} \log n)$, uniformly in t, the process (5.39) has the same limiting behavior.

Similarly, defining $G_n^{(0)} = G_{F_0}$, we get that the process

$$n^{2/3} \int_{(z_0, z_0 + n^{-1/3}t]} \left\{ F_0(t') - F_0(t_0) \right\} \, dG_n^{(0)}(t'), \ t \geq 0,$$

converges on bounded intervals almost surely to the deterministic function

$$t \mapsto \tfrac{1}{2} \left\{ \sum_{i=0}^{m} (g(a_i) - g(a_i-))^2 / h(z_0 + a_i) \right\} f_0(z_0) t^2, \ t \geq 0.$$

To see this, note that, by (2.12), we have to deal with terms of the form

$$\{ F_0(Z_{(i+1)}) - F_0(z_0) \} \Delta G_F(Z_{(i)})$$

$$= \{ F_0(Z_{(i+1)}) - F_0(z_0) \} \int \frac{\left\{ g(z - Z_{(i)}) - g(z - Z_{(i+1)}) \right\}^2}{\left\{ \int g(z - x) \, dF_0(x) \right\}^2} \, dH_n(z),$$

and $\left\{ g(z - Z_{(i)}) - g(z - Z_{(i+1)}) \right\}^2$ behaves as $\left\{ g'(z - Z_{(i)})(Z_{(i+1)} - Z_{(i)}) \right\}^2$ at the interior of the set of points $z - Z_{(i)}$ where g' is differentiable.

Since negative values of t can be dealt with in a completely similar way, the result follows. ∎

In accordance with the methods used in the preceding sections, we introduce, for each $a > 0$, the random variable $T_n^{(0)}(a)$ by

$$T_n^{(0)}(a) = \sup\{ t \geq 0 : V_n^{(0)}(t) - (a - F_0(t_0))G_n^{(0)}(t) \text{ is minimal} \}.$$

Likewise, in analogy with (5.9) and (5.35), we define the process $\{ T(a) : a \in \mathbb{R} \}$ by

$$T(a) = \sup\left\{ t \in \mathbb{R} : U(t) - a \cdot \left\{ \sum_{i=0}^{m} (g(a_i) - g(a_i-))^2 / h(z_0 + a_i) \right\} f_0(z_0) t \text{ is minimal} \right\}.$$

As in Lemma 5.2, we get that the process

$$\left\{ n^{1/3} \left\{ T_n^{(0)}(a_0 + n^{-1/3}a) - t_0 \right\} : a \in \mathbb{R} \right\}$$

converges in distribution to the process $\{ T(a) : a \in \mathbb{R} \}$. This follows from the following lemma.

75

LEMMA 5.8. *For each $\epsilon > 0$ and $M_1 > 0$ an $M_2 > 0$ can be found such that*

$$\mathbb{P}\left\{\max_{a \in [-M_1, M_1]} n^{1/3}\left\{T_n^{(0)}(a_0 + n^{-1/3}a) - t_0\right\} > M_2\right\} < \epsilon,$$

and

$$\mathbb{P}\left\{\min_{a \in [-M_1, M_1]} n^{1/3}\left\{T_n^{(0)}(a_0 + n^{-1/3}a) - t_0\right\} < -M_2\right\} < \epsilon,$$

for all sufficiently large n.

PROOF: The proof proceeds along similar lines as the proof of Lemma 5.3. Let $Z_n^{(0)}(u) = \left\{W_n^{(0)}(z_0 + u) - W_n^{(0)}(z_0)\right\}^2$. Then,

$$\mathbb{P}\left\{\sup_{0 \leq u \leq u_0} Z_n^{(0)}(u)\right\} \leq cn^{-1}\left\{u_0 \cdot \sum_{i=0}^{m}(g(a_i) - g(a_i-))^2/h(z_0 + a_i) + \mathcal{O}(u_0^2)\right\},$$

for a constant $c > 0$, if u_0 satisfies $H(t_0 + u_0) < 1$. This can, e.g., be seen by applying Lemma 4.1 in Kim and Pollard (1989) to the functions ϕ_β, defined by

$$\phi_\beta(u, \beta) = (g(u - z_0) - g(u - z_0 - \beta))/h(u), \ \beta \geq 0, \ u \geq z_0.$$

Under the conditions of Theorem 5.4 the class of functions $\{|\phi_\beta| : 0 \leq \beta \leq R\}$ has an envelope

$$G_R(u) = h(u)^{-1}\left\{\sum_{i=0}^{m}|g(a_i) - g(a_i-)| 1_{[z_0 + a_i, z_0 + a_i + R]}(u) + R \sup_{t \in [0, R], \ t \notin \{a_0, \dots, a_m\}} |g'(t)|\right\},$$

for $u \geq z_0$, and we get

$$PG_R^2 \leq c_1 R + c_2 \cdot R^2,$$

for positive constants c_1, c_2. The remaining part of the argument is similar to the proof of Lemma 5.3. ∎

We can now give the proof of Theorem 5.4.

PROOF OF THEOREM 5.4: First of all, we have:

$$\mathbb{P}\{F_n^{(1)}(t_0) - F_0(t_0) > n^{-1/3}a\} = \mathbb{P}\{T_n^{(0)}(a_0 + n^{-1/3}a) < t_0\},$$

where $a_0 = F_0(t_0)$. Using Lemmas 5.7 and 5.8, we get that the process

$$\left\{n^{1/3}\left\{T_n^{(0)}(a_0 + n^{-1/3}a) - t_0\right\} : a \in \mathbb{R}\right\}$$

converges in the Skorohod topology on $D(I\!\!R)$ to the process $\{T(a) : a \in I\!\!R\}$. As in the preceding sections, the process

$$\{T(a) - f_0(t_0)^{-1}a : a \in I\!\!R\}$$

is stationary. Hence,

$$IP\left\{n^{1/3}\{F_n^{(1)}(t_0) - F_0(t_0)\}\left\{2\sum_{i=0}^{m}(g(a_i) - g(a_i-))^2/(f_0(z_0)h(z_0 + a_i))\right\}^{1/3} > x\right\}$$

$$\to IP\left\{T(0) > x\Big/\left\{2f_0(z_0)^2\sum_{i=0}^{m}(g(a_i) - g(a_i-))^2/h(z_0 + a_i)\right\}^{1/3}\right\}, \; n \to \infty.$$

Since $T(0)$ is the last time where

$$t \mapsto W\left(t\sum_{i=0}^{m}(g(a_i) - g(a_i-))^2/h(z_0 + a_i)\right) + \tfrac{1}{2}f_0(z_0)t^2\sum_{i=0}^{m}\frac{(g(a_i) - g(a_i-))^2}{h(z_0 + a_i)}$$

reaches its minimum, this means that

$$\tfrac{1}{2}T(0)\left\{2f_0(z_0)^2\sum_{i=0}^{m}(g(a_i) - g(a_i-))^2/h(z_0 + a_i)\right\}^{1/3}$$

is the last time where

$$t \mapsto W(t) + t^2, \; t \in I\!\!R,$$

reaches its minimum. ∎

Computer experiments, using simulated samples of convolution densities, again indicate that in many cases the NPMLE has the same asymptotic behavior as the 1-step estimator $F_n^{(1)}$. But for a proof of this, one has to prove two final steps, which are of the same type as the two remaining steps in the proof for the asymptotic behavior of the NPMLE for interval censoring, Case 2.

We will deal with this problem elsewhere. It seems that the limiting behavior of the NPMLE is the same as that of the 1-step estimator, given in Theorem 5.4, at least when the support of F_0 is contained in an interval not bigger than the minimum distance between discontinuities of g. If the latter condition is not fulfilled, it is not clear that the limiting behavior will be the same.

5.4. Estimation of the mean.

As can be expected from the general theory on differentiable functionals (see e.g., van der Vaart (1989)), efficient estimators of smooth functionals like the mean

$$\mu_{F_0} = \int t \, dF_0(t)$$

should have $\sqrt{n}$–behavior. We will discuss a result of this type for interval censoring, Case 1, using the same set-up as in section 4.1. Let $(X_1, T_1), \ldots, (X_n, T_n)$ be a sample of random variables in $\mathbb{R}_+^2$, where X_i and T_i are independent (nonnegative) random variables with continuous distribution functions F_0 and G, respectively.

Furthermore, we assume that the support of P_{F_0} is a bounded interval $I = [0, M]$, and that F_0 and G have densities f_0 and g, respectively, satisfying

$$g(t) \geq \delta > 0, \text{ and } f_0(t) \geq \delta > 0, \qquad \text{if } t \in I,$$

for some $\delta > 0$. Finally we assume that g has a bounded derivative on I. An example of this situation is the case where F_0 and G are both the uniform distribution function on $[0, 1]$. It is certainly possible to prove the following theorem under weaker conditions, but at the cost of an increasing number of technicalities.

THEOREM 5.5. *Let F_0 and G satisfy the conditions, listed above, and let $\hat{F}_n$ be the NPMLE of F_0. Then*

$$\sqrt{n} \int_I (\hat{F}_n(t) - F_0(t)) \, dt \xrightarrow{D} U,$$

where U has a normal distribution with mean zero and variance

$$\sigma_{F_0}^2 = \int \frac{F_0(t)(1 - F_0(t))}{g(t)} \, dt.$$

Before starting the proof of Theorem 5.5, we give an outline of the steps in the argument. First we note that

$$\int \frac{\{x > t\}}{g(t)} \, dP(x, t) = \int_I (1 - F_0(t)) \, dt = \mu_{F_0},$$

where P is the probability measure of a pair (X_i, T_i). Secondly, let $\tau_1 < \tau_2 < \cdots < \tau_m$ be the points of jump of $\hat{F}_n$ on the interval $[0, T_{(n)}]$, and let $\tau_0 = 0$, $\tau_{m+1} = T_{(n)}$. Then, defining the function $\tilde{g}_n$ on I by $\tilde{g}_n(0) = g(0)$, and

$$\tilde{g}_n(t) = g(\tau_i), \quad \tau_{i-1} < t \leq \tau_i, \quad 1 \leq i \leq m + 1,$$

we get

$$\int_I (1 - \hat{F}_n(t))\, dt - \mu_{F_0} = \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{g(t)}\, dP(x,t)$$

$$= \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)}\, dP(x,t)$$

(5.41)
$$+ \int \{1 - \hat{F}_n(t) - \{x > t\}\}\left\{\frac{1}{g(t)} - \frac{1}{\tilde{g}_n(t)}\right\}\, dP(x,t),$$

$$= \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)}\, dP(x,t)$$

$$- \int \{\hat{F}_n(t) - F_0(t)\}\left\{\frac{1}{g(t)} - \frac{1}{\tilde{g}_n(t)}\right\}\, dG(t).$$

By the characterization of the NPMLE, given in Proposition 1.2, we have:

$$\int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)}\, dP_n(x,t) = 0,$$

where P_n is the empirical measure of $(X_1, T_1), \ldots, (X_n, T_n)$. Hence

(5.42)
$$\sqrt{n} \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)}\, dP(x,t)$$

$$= -\sqrt{n} \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)}\, d(P_n - P)(x,t).$$

We shall prove that the right-hand side of (5.42) converges to U, where U is defined as in Theorem 5.5, and that the last term of (5.41) is $o_P(n^{-1/2})$.

The distance between the successive jumps of $\hat{F}_n$ plays an essential role in the proof. Roughly speaking, they are of order $\mathcal{O}_p(n^{-1/3})$, but we need a uniform upper bound for this distance. To this end we introduce, for $a \in (0,1)$, the random variables $T_n(a)$, defined by

(5.43)
$$T_n(a) = \sup\{t \in [0, T_{(n)}] : V_n(t) - aG_n(t) \text{ is minimal}\},$$

where G_n is the empirical distribution function of $T_1, \ldots, T_n$, and V_n is defined by

(5.44)
$$V_n(t) = n^{-1} \sum_{T_i \le t} 1_{\{X_i \le T_i\}} = \int_{u \le t} 1_{\{x \le u\}}\, dP_n(x,u).$$

Note the similarity with the definitions of the random variables $T_n^{(0)}(a)$, discussed in the preceding sections.

The process $\left\{(G_n(T_n(a)), V_n(T_n(a))) : a \in (0,1)\right\}$ runs through the vertices of the cumulative sum diagram, consisting of the point $(0,0)$ and the points

$$(G_n(T_{(i)}), V_n(T_{(i)})),\ 1 \le i \le n.$$

79

The following lemma gives an upper bound for the probability that the distance of successive jumps of $\hat{F}_n$ is bigger that $n^{-1/3}\log n$. The proof is quite similar to the proof of Lemma 2.4 in Groeneboom (1987). It is based on a stopping time argument for a suitably chosen exponential martingale.

LEMMA 5.9. *Let, for each* $a \in (0,1)$, $F_0^{-1}(a)$ *be defined by* $F_0(F_0^{-1}(a)) = a$ *(note that* $F_0^{-1}(a)$ *is well-defined, since* F_0 *has a strictly positive density on I). Then, for each* $a \in (0,1)$,

$$\mathbb{P}\{\left|T_n(a) - F_0^{-1}(a)\right| \geq n^{-1/3}\log n\} \leq c_1 \exp\{-c_2(\log n)^2\},$$

for positive constants c_1 *and* c_2, *not depending on* a *and* n.

PROOF: Let $a \in (0,1)$ and suppose $t_0 = F_0^{-1}(a)$. Let the process U_n be defined by

$$U_n(t) = V_n(t) - V_n(t_0) - a\{G_n(t) - G_n(t_0)\}, \ t \in I.$$

Then we have

$$\mathbb{P}\{\left|T_n(a) - F_0^{-1}(a)\right| \geq n^{-1/3}\log n\}$$
$$\leq \mathbb{P}\{U_n(t) \leq 0, \text{ for some } t \in I\backslash(t_0 - n^{-1/3}\log n, t_0 + n^{-1/3}\log n)\}.$$

We only give an upper bound for the probability

$$\mathbb{P}\{U_n(t) \leq 0, \text{ for some } t \geq t_0 + n^{-1/3}\log n\},$$

since the probability

$$\mathbb{P}\{U_n(t) \leq 0, \text{ for some } t \leq t_0 - n^{-1/3}\log n\},$$

can be treated similarly.

Define the process W_n by

$$W_n(t) = U_n(t) - \int_{[t_0,t]} F_0(t') \, dG_n(t'), \ t \in I, \ t \geq t_0.$$

Moreover, we define $\Delta W_{n,i}$ and $\Delta G_{n,i}$ by

$$\Delta W_{n,i} = W_n(T_{(i)}) - W_n(T_{(i-1)}), \text{ and } \Delta G_{n,i} = G_n(T_{(i)}) - W_n(T_{(i-1)}),$$

where $T_{(0)} \overset{\text{def}}{=} 0$. Then we can write

$$\mathbb{P}\{U_n(t) \leq 0, \text{ for some } t \geq t_0 + n^{-1/3}\log n\}$$
$$= \mathbb{P}\left\{W_n(t) + \int_{[t_0,t]} \{F_0(t') - F_0(t_0)\} \, dG_n(t') \leq 0, \text{ for some } t \geq t_0 + n^{-1/3}\log n\right\}.$$

Conditionally on the observation times T_i, the terms $\Delta W_{n,i}$, with $T_{(i)} \in [t_0, t]$ are independent random variables, with values in $[-n^{-1}, n^{-1}]$, and we have, for $t \geq t_0$,

(5.45)

$$E\Big\{\exp\{n^{2/3} W_n(t)\} \mid T_1, T_2, \dots \Big\}$$

$$= E\left\{\exp\Big\{n^{-1/3} \sum_{T_{(i)} \in [t_0, t]} \Delta W_{n,i}\Big\} \,\Big|\, T_1, T_2, \dots \right\}$$

$$= \prod_{T_{(i)} \in [t_0, t]} \Big\{ F_0(T_{(i)}) \exp\{n^{-1/3}(1 - F_0(T_{(i)}))\} + (1 - F_0(T_{(i)})) \exp\{-n^{-1/3} F_0(T_{(i)})\} \Big\}$$

$$= \exp\Big\{-\tfrac{1}{2}(1 + r_n(t)) n^{-2/3} \sum_{T_{(i)} \in [t_0, t]} F_0(T_{(i)})(1 - F_0(T_{(i)})) \Big\}$$

$$= \exp\Big\{-\tfrac{1}{2}(1 + r_n(t)) n^{1/3} \int_{[t_0, t]} F_0(t')(1 - F_0(t')) \, dG_n(t') \Big\}$$

where $r_n(t)$ tends uniformly to zero for each realization of the observation times $T_{(i)}$, since $n^{-1/3} F_0(T_{(i)})$ and $n^{-1/3}(1 - F_0(T_{(i)}))$ tend uniformly to zero, as $n \to \infty$.

Let the process $\{Z_n(t) : t \in I, t \geq t_0\}$ be defined by

$$Z_n(t) = \exp\{-n^{2/3} W_n(t)\} E\Big\{\exp\{n^{2/3} W_n(t)\} \mid T_1, T_2, \dots \Big\}.$$

Then, conditionally on the observation times $T_1, \dots, T_n$, the process Z_n is a martingale with respect to the filtration $\{\mathcal{F}_{n,t} : t \in I, t \geq t_0\}$, where

$$\mathcal{F}_{n,t} = \sigma\Big\{ 1_{\{X_i \leq T_i\}} : T_i \in [t_0, t] \Big\}.$$

We now define the stopping time τ_n by

$$\tau_n = \inf\{t \in [t_0 + n^{-1/3} \log n, M] : U_n(t) \leq 0\}.$$

If $U_n(t) > 0$, for all $t \in [t_0 + n^{-1/3} \log n, M]$, we define $\tau_n = \infty$. We shall derive an upper bound for the probability

$$I\!\!P\{\tau_n < \infty \mid T_1, \dots, T_n\},$$

and denote the conditional probabilities $I\!\!P\{\cdot \mid T_1, \dots, T_n\}$ and the conditional expectations $E\{\cdot \mid T_1, \dots, T_n\}$ both by $I\!\!P_n$.

Since (in the conditional set-up), Z_n is a martingale, we have

$$I\!\!P_n Z_n(M \wedge \tau_n) = I\!\!P_n Z_n(t_0) = 1.$$

Furthermore, by (5.45) we have:

(5.46)

$$I\!\!P_n Z_n(M \wedge \tau_n) \geq I\!\!P_n Z_n(\tau_n) 1_{\{\tau_n < \infty\}}$$

$$= I\!\!P_n \exp\Big\{-n^{2/3} W_n(\tau_n) - \tfrac{1}{2}(1 + r_n(\tau_n)) n^{1/3} \int_{[t_0, \tau_n]} F_0(t)(1 - F_0(t)) \, dG_n(t) \Big\} 1_{\{\tau_n < \infty\}}.$$

81

But for $t \geq t_0 + n^{-1/3} \log n$ we have, if $U_n(t) \leq 0$,

(5.47)

$$- n^{2/3} W_n(t) - \tfrac{1}{2}(1 + r_n(t)) n^{1/3} \int_{[t_0,t]} F_0(t)(1 - F_0(t)) \, dG_n(t)$$

$$= -n^{2/3} U_n(t) + n^{2/3} \int_{[t_0,t]} \{ F_0(t') - F_0(t_0) \} \, dG_n(t')$$

$$- \tfrac{1}{2}(1 + r_n(t)) n^{1/3} \int_{[t_0,t]} F_0(t)(1 - F_0(t)) \, dG_n(t)$$

$$\geq n^{2/3} \int_{[t_0,t]} \{ F_0(t') - F_0(t_0) \} \, dG_n(t') - \tfrac{1}{2}(1 + r_n(t)) n^{1/3} \int_{[t_0,t]} F_0(t')(1 - F_0(t')) \, dG_n(t')$$

$$\geq \tfrac{1}{2} f_0(t_0) n^{2/3} \int_{[t_0, t_0 + n^{-1/3} \log n]} (t' - t_0) \, dG_n(t')$$

$$+ \tfrac{1}{2} f_0(t_0) n^{1/3} \{ G_n(t) - G_n(t_0 + n^{-1/3} \log n) \} \log n$$

$$- \tfrac{1}{2}(1 + r_n(t)) n^{1/3} \int_{[t_0,t]} F_0(t)(1 - F_0(t)) \, dG_n(t)$$

$$\geq \tfrac{1}{2} f_0(t_0) n^{2/3} \int_{[t_0, t_0 + n^{-1/3} \log n]} \left(t - t_0 - 2 n^{-1/3} F_0(t)(1 - F_0(t)) / f_0(t_0) \right) \, dG_n(t),$$

for $n \geq n_0$, where n_0 only depends on $f_0(t_0)$ (and not on t or G_n). Here we use that, by the conditions on F_0, $F_0(t) - F_0(t_0) \geq \tfrac{1}{2} f_0(t_0)(t - t_0)$, if $0 \leq t - t_0 \leq n^{-1/3} \log n$ and n is sufficiently large. From (5.46) and (5.47) we obtain, for $n \geq n_0$:

(5.48)

$$\mathbb{P}_n \{ \tau_n < \infty \} \leq$$

$$\mathbb{P}_n \exp \left\{ -\tfrac{1}{2} n^{2/3} f_0(t_0) \int_{[t_0, t_0 + n^{-1/3} \log n]} \left(t - t_0 - 2 n^{-1/3} \frac{F_0(t)(1 - F_0(t))}{f_0(t_0)} \right) \, dG_n(t) \right\}.$$

By (5.48), the probability of a large deviation of the process U_n is reduced to the probability of a large deviation of the empirical distribution function G_n. We now note that

$$n^{2/3} \int_{[t_0, t_0 + n^{-1/3} \log n]} \left(t - t_0 - 2 n^{-1/3} F_0(t)(1 - F_0(t)) / f_0(t_0) \right) \, dG_n(t)$$

$$= n^{2/3} \int_{[t_0, t_0 + n^{-1/3} \log n]} \left(t - t_0 - 2 n^{-1/3} F_0(t)(1 - F_0(t)) / f_0(t_0) \right) \, dG(t)$$

$$+ n^{2/3} \int_{[t_0, t_0 + n^{-1/3} \log n]} \left(t - t_0 - 2 n^{-1/3} F_0(t)(1 - F_0(t)) / f_0(t_0) \right) \, d(G_n - G)(t)$$

$$= (1 + o(1)) g_0(t_0) \left\{ \tfrac{1}{2} (\log n)^2 - \frac{F_0(t_0)(1 - F_0(t_0))}{f_0(t_0)} \log n \right\}$$

$$+ n^{2/3} \int_{[t_0, t_0 + n^{-1/3} \log n]} \left(t - t_0 - 2 n^{-1/3} F_0(t)(1 - F_0(t)) / f_0(t_0) \right) \, d(G_n - G)(t),$$

82

and, moreover, that

$$\operatorname{var}\left\{n^{2/3}\int_{[t_0, t_0+n^{-1/3}\log n]}\left(t - t_0 - 2n^{-1/3}F_0(t)(1 - F_0(t))/f_0(t_0)\right)d(G_n - G)(t)\right\}$$

$$\sim n^{1/3}g(t_0)\int_{[t_0, t_0+n^{-1/3}\log n]}\left(t - t_0 - 2n^{-1/3}F_0(t)(1 - F_0(t))/f_0(t_0)\right)^2 dt$$

$$\sim \tfrac{1}{3}n^{-2/3}g(t_0)(\log n)^3, \text{ as } n \to \infty.$$

Hence, by Bernstein's inequality (see, e.g., Pollard (1984), p. 193), we get:

$$\mathbb{P}\left\{n^{2/3}\int_{[t_0, t_0+n^{-1/3}\log n]}\left(t - t_0 - 2n^{-1/3}F_0(t)(1 - F_0(t))/f_0(t_0)\right)d(G_n - G)(t)\right.$$

$$\left.\le -\tfrac{1}{2}g(t_0)\left\{\tfrac{1}{2}(\log n)^2 - \frac{F_0(t_0)(1 - F_0(t_0))}{f_0(t_0)}\log n\right\}\right\}$$

$$\le 2\exp\left\{-c \cdot n^{2/3}/\log n\right\},$$

for a constant $c > 0$. Applying this on (5.48), we obtain

$$\mathbb{P}\{\tau_n < \infty\} \le \exp\left\{-\tfrac{1}{2}g(t_0)\left\{\tfrac{1}{2}(\log n)^2 - \frac{F_0(t_0)(1 - F_0(t_0))}{f_0(t_0)}\log n\right\}\right\}$$

$$+ 2\exp\left\{-c \cdot n^{2/3}/\log n\right\}.$$

This gives the desired bound, since

$$\mathbb{P}\{U_n(t) \le 0, \text{ for some } t \ge t_0 + n^{-1/3}\log n\} = \mathbb{P}\{\tau_n < \infty\},$$

and since $f_0(t)$ and $g(t)$ are uniformly bounded away from zero for $t \in I$. ∎

Now let $a_1, a_2, \ldots$ be such that

$$a_i = i \cdot n^{-1/3}\log n, \ i = 1, 2, \ldots,$$

and let $m_n = [n^{1/3}/\log n]$, where $[a]$ denotes the biggest integer $\le a$. Then we get from Lemma 5.9:

$$\mathbb{P}\left\{\left|T_n(a_i) - F_0^{-1}(a_i)\right| \ge n^{-1/3}\log n\right\} \le c_1\exp\{-c_2(\log n)^2\}, \ 1 \le i \le m_n.$$

This implies, by the monotonicity of the function $a \mapsto T_n(a)$ and the conditions on F_0 that there exists a constant $c_3 > 0$ such that

$$\mathbb{P}\left\{\sup_{a \in (0,1)}\left|T_n(a) - F_0^{-1}(a)\right| \ge c_3 n^{-1/3}\log n\right\} \le m_n c_1\exp\{-c_2(\log n)^2\}$$

$$\le c_1\exp\{-\tfrac{1}{2}c_2(\log n)^2\},$$

83

for sufficiently large n. This, in turn, implies that both the maximum distance between two successive points of jump of $\hat{F}_n$ and the maximum distance between $\hat{F}_n$ and F_0 are of order $n^{-1/3} \log n$. Hence we get, for the last term in (5.41):

$$\int \{\hat{F}_n(t) - F_0(t)\} \left\{ \frac{1}{g(t)} - \frac{1}{\tilde{g}_n(t)} \right\} dG(t) = \mathcal{O}_p\left(n^{-2/3}(\log n)^2\right),$$

where we use that g has a bounded derivative on I. Furthermore, by (5.42):

$$\sqrt{n} \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)} dP(x,t)$$

$$= -\sqrt{n} \int \frac{1 - \hat{F}_n(t) - \{x > t\}}{\tilde{g}_n(t)} d(P_n - P)(x,t)$$

$$= -\sqrt{n} \int \frac{1 - F_0(t) - \{x > t\}}{g(t)} d(P_n - P)(x,t)$$

$$\quad - \sqrt{n} \int \{1 - F_0(t) - \{x > t\}\} \left\{ \frac{1}{\tilde{g}_n(t)} - \frac{1}{g(t)} \right\} d(P_n - P)(x,t)$$

$$\quad - \sqrt{n} \int \frac{\hat{F}_n(t) - F_0(t)}{\tilde{g}_n(t)} d(P_n - P)(x,t)$$

Since the class of functions

$\{(F - F_0)/g_n : F$ is a distribution function, and g_n is a piecewise constant version of g,

with distance $\leq c \cdot n^{-1/3} \log n$ between successive points of jump$\}$

is "manageable" in the sense of Pollard (1989) (see Birman and Solomjak (1967) and van de Geer (1990) for the relevant entropy calculations), it follows that

$$\sqrt{n} \int \frac{\hat{F}_n(t) - F_0(t)}{\tilde{g}(t)} d(P_n - P)(x,t) = o_p(1), \; n \to \infty.$$

Similarly,

$$\sqrt{n} \int \{1 - F_0(t) - \{x > t\}\} \left\{ \frac{1}{\tilde{g}(t)} - \frac{1}{g(t)} \right\} d(P_n - P)(x,t) = o_p(1).$$

Theorem 5.5 now follows from (5.41).

5.5. Exercises.

1. Let $T_n(a)$ be defined by (5.43), and let the conditions of Theorem 5.1 be satisfied. Moreover, let $a_0 = F_0(t_0)$, where t_0 satisfies the conditions of Theorem 5.1. Show, using the same techniques as in the proof of Lemma 5.9, that for each $M_1 > 0$ and $\epsilon > 0$ there exists an $M_2 > 0$ such that, uniformly in $x \in [-M_1, M_1]$,

$$\mathbb{P}\left\{ \left| T_n(a_0 + x n^{-1/3}) - t_0 \right| \geq M_2 \right\} < \epsilon,$$

for all sufficiently large n.

84

2. Let $U_n(t)$ be defined as in the proof of Lemma 5.9. Show that the process

$$t \mapsto \begin{cases} n^{2/3} U_n(t_0 + n^{-1/3}t) & , t_0 + n^{-1/3}t \in I, \\ 0 & , t_0 + n^{-1/3}t \notin I, \end{cases}$$

converges in distribution, in the topology of uniform convergence on compacta, to the process U, defined by

$$U(t) = \sqrt{g(t_0)F_0(t_0)(1 - F_0(t_0))}\, W(t) + \tfrac{1}{2}f_0(t_0)g(t_0)t^2, \ t \in \mathbb{R},$$

where W is standard two-sided Brownian motion on $\mathbb{R}$, originating from zero.

3. Let $T_n(a)$ and $U(t)$ be defined as in Exercises 1 and 2. Show that the process

$$\left\{ n^{1/3}\left\{ T_n(a_0 + n^{-1/3}a) - t_0 \right\} : a \in \mathbb{R} \right\}$$

converges in distribution, in the space $D(\mathbb{R})$ with the Skorohod topology, to the process $\{T(a) : a \in \mathbb{R}\}$, where

$$T(a) = \sup\left\{ t \in \mathbb{R} : U(t) - a \cdot tg(t_0) \text{ is minimal} \right\}.$$

4. Deduce Theorem 5.1 from Exercises 1 to 3.

5. Let, in Theorem 5.4, g be the standard exponential density on $[0, \infty)$, and let F_0 and z_0 be such that the conditions of Theorem 5.4 are satisfied. Deduce from Exercise 3 in Chapter 2:

$$n^{1/3}\left\{ \hat{F}_n(z_0) - F_0(z_0) \right\} / \left\{ \tfrac{1}{2}f_0(z_0)h(z_0) \right\}^{1/3} \xrightarrow{\mathcal{D}} 2U,$$

where U is the last time where standard two-sided Brownian motion minus the parabola $y(t) = t^2$ reaches its maximum, and $\hat{F}_n$ is the NPMLE of F_0.

6. Deduce from Exercise 2 in Chapter 2 that, under the same conditions and with the same notation as in Exercise 5, but with g replaced by the uniform density on $[0, 1]$,

$$n^{1/3}\left\{ \hat{F}_n(z_0) - F_0(z_0) \right\} / \left\{ \tfrac{1}{2}F_0(z_0)(1 - F_0(z_0))f_0(z_0) \right\}^{1/3} \xrightarrow{\mathcal{D}} 2U.$$

7. Let F_0 and G satisfy the same conditions as in Theorem 5.5, and let $\hat{F}_n$ be the NPMLE of F_0. Show that

$$\sqrt{n}\left\{ \int_I \hat{F}_n^2(t)\, dt - \int_I F_0^2(t)\, dt \right\} \xrightarrow{\mathcal{D}} U,$$

where U has a normal distribution with mean zero and variance

$$\sigma_{F_0}^2 = 4 \int \frac{F_0^3(t)(1 - F_0(t))}{g(t)}\, dt.$$

85

REFERENCES

[1]   Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., Silverman, E. (1955), *An empirical distribution function for sampling with incomplete information*, Ann. Math. Statist. **26**, 641–647.

[2]   Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D. (1972), "Statistical Inference under Order Restrictions," Wiley, New York.

[3]   Bennett, G., (1962), *Probability inequalities for sums of independent random variables*, Amer. Statist. Assoc. **57**, 33–45.

[4]   Birgé, L., Massart, P., (1991), *Rates of convergence for minimum contrast estimators*, Rapport Technique Nr. 140, Université Paris VI, Laboratoire de Statistique Théorique et Appliquée.

[5]   Birman, M.Š., Solomjak, M.Z., (1967), *Piecewise-polynomial approximations of functions in the classes W_p^α*, Math. Sbornik. **73**, 295–317.

[6]   Carroll, R.J., Hall, P., (1988), *Optimal rates of convergence for deconvolving a density*, Amer. Statist. Assoc. **83**, 1184–1186.

[7]   Chang, M.N. (1990), *Weak convergence of a self-consistent estimator of the survival function with doubly censored data*, Ann. Statist. **18**, 391–404.

[8]   Efron, B. (1967), *The two-sample problem with censored data*, Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability, 831–853, University of California Press.

[9]   Fan, J. (1988), *On the optimal rate of convergence for nonparametric deconvolution problem*, Technical Report, Department of Statistics, University of California, Berkeley..

[10] Gill, R.D. (1989), *Non- and semi-parametric maximum likelihood estimators and the von Mises method, part I*, Scand. J. Statistics, **16**, 79–128.

[11] Gill, R.D. (1991), *Non- and semi-parametric maximum likelihood estimators and the von Mises method, part II*, Preprint Nr. 664, Department of Mathematics, University Utrecht.

[12] Groeneboom, P. (1987), *Asymptotics for interval censored observations*, Technical Report 87-18, Department of Mathematics, University of Amsterdam,.

[13] Groeneboom, P. (1989), *Brownian motion with a parabolic drift and Airy functions*, Probability theory and related fields, **81**, 79–109.

[14] Jewell, N.P. (1982), *Mixtures of exponential distributions*, Ann. Statist. **10**, 1184–1186.

[15] Karlin, S., Studden, W.J. (1966), "Tchebycheff Systems: with Applications in Analysis and Statistics," Interscience, New York.

[16] Kim, J., Pollard, D. (1990), *Cube root asymptotics*, Ann. Statist. **18**, 191–219.

[17] Pollard, D. (1984), "Convergence of Stochastic Processes," Springer-Verlag, New York.

[18] Pollard, D. (1989), *Asymptotics via empirical processes (with discussion)*, Statist. Sci. **4**, 341–366.

[19] Pollard, D. (1990), "Empirical Processes: Theory and Applications," NSF-CBMS Regional Conference Series in Probability and Statistics, 2.

[20] Robertson,T., Wright, F.T., Dykstra, R.L. (1988), "Order Restricted Statistical Inference," Wiley, New York.

[21] Rockafellar, R.T. (1988), "Convex Analysis," Princeton University Press.

[22] Stefanski, L.A., Carroll, R.J. (1987), *Deconvoluting kernel density estimators*, University of North Carolina Mimeo Series **1623**.

[23] Tsai, W.I., Crowley, J. (1985), *A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency*, Ann. Statist. **13**, 1317–1334.

[24] Tsai, W.I., Crowley, J. (1990), *Correction on: A large sample study of generalized maximum likelihood estimators from incomplete data via self-consistency*, Ann. Statist. (1990) **18**, 470.

[25] Turnbull, B.W. (1974), *Nonparametric estimation of a survivorship function with doubly censored data*, Amer. Statist. Assoc. **69**, 169–173.

[26] Turnbull, B.W. (1976), *The empirical distribution function with arbitrarily grouped censored and truncated data*, J.R. Statist. Soc. B **38**, 290–295.

[27]   van Es, A.J. (1988), "Aspects of Nonparametric Density Estimation," Ph.D. Dissertation, University of Amsterdam (to appear as Mathematical Centre Tract).

[28]   van de Geer, S. (1990), *Hellinger-consistency of certain nonparametric maximum likelihood estimators*, Preprint Nr. 614, Department of Mathematics, University Utrecht.

[29]   van der Vaart, A.W. (1991), *On differentiable functionals*, Ann. Statist. (1990) **18**, 178–205.

[30]   Cun-Hui Zhang (1990), *Fourier methods for estimating mixing densities and distributions*, Ann. Statist. **18**, 816–831.