# Nonparametric estimation with application to renewal processes

A thesis submitted

by

Luc Watelet

to

the School of Graduate Studies

of the University of Ottawa

in partial fulfillment of the requirements

for the degree of

Master of Science

in the subject of

Mathematics

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

Table of Contents

Abstract.

Suppose that G is the distribution function of a nonnegative random variable, with finite expectation $\mu = \int_0^\infty (1-G(s))ds$, and that F is the probability distribution function which is linked to G by the following formula:

$$\text{(i)} \qquad F(x) = \begin{cases} \int_0^x \frac{1-G(s)}{\mu}\, ds & , x \geq 0, \\ 0 & , x < 0 . \end{cases}$$

Given an independent sequence $(X_i)_1^\infty$ with common distribution function F, we want to construct a nonparametric estimator of G in terms of the random variables $(X_i)_1^n$; further, we want to do this so that the estimator has strong consistency properties when $n \to \infty$. This problem arises when the statistician observes n independent renewal processes which started long before the start of observation, and the statistician can only observe the n "residual lives" from the start of observation to the next renewal in each process. If G is the underlying distribution of the renewal processes, the residual lives can be represented by random variables with distribution function F related to G by (i).

We consider the case where G is continuous on $\mathbb{R}$ and the case where, with $\tau = \inf\{x: G(x) = 1\} < \infty$, G is continuous everywhere but at $\tau$, i.e., $G(\tau^-) < 1$.

The distribution F given above has a density f given by

$$f(x) = \begin{cases} \dfrac{1 - G(x)}{\mu} & , x \geq 0 , \\ 0 & , x < 0 , \end{cases}$$

where f is nonincreasing and right-continuous. Since G(0) = 0 by hypothesis, we have $\mu = (f(0))^{-1}$ and therefore

$$G(x) = \begin{cases} 1 - (f(0))^{-1} f(x) & , x \geq 0 , \\ 0 & , x < 0 . \end{cases}$$

Now, if $f_n$ is an estimator of f then

$$\text{(ii)} \qquad G_n(x) = \begin{cases} 1 - (f_n(0))^{-1} f_n(x) & , x \geq 0 , \\ 0 & , x < 0 \end{cases}$$

is an obvious estimator of G.

Since G is nondecreasing and right-continuous it is natural to seek a nondecreasing and right-continuous estimator for G by a nonincreasing and right-continuous estimator for f. In this thesis we develop two monotonization methods for transforming a histogram estimator $\hat{f}_n$ of f into a histogram estimator $\tilde{f}_n$ which is monotone nonincreasing on $[0,\infty)$. We examine those methods in detail and show that

$$\sup_{0 \leq t < \infty} |f(t) - \tilde{f}_n(t)| \leq \sup_{0 \leq t < \infty} |f(t) - \hat{f}_n(t)|$$

when f is any density which vanishes on $(-\infty,0)$ and is nonincreasing on $[0,\infty)$. We use this to obtain strong uniform consistency and rate of convergence results for the monotone histogram estimator $\tilde{f}_n$. These developments rely on an examination of properties of estimators $\tilde{p}_1,\ldots,\tilde{p}_m$ of probabilities $p_1,\ldots,p_m$ subject to the restriction $\tilde{p}_1 \geq \cdots \geq \tilde{p}_m$ ; it is shown that if $p_1 \geq \ldots \geq p_m$ then

$$\max_{1\leq k\leq m} |p_k - \tilde{p}_k| \leq \max_{1\leq k\leq m} |p_k - \hat{p}_k|$$

where $\hat{p}_k$ is the usual "sample proportion" estimator of $p_k$. We finally show that the results for the monotonized histogram apply to the estimator $\tilde{G}_n$ obtained from $\tilde{f}_n$ by (ii).

## 1. Introduction.

### 1.1. Introductory example.

Before stating the general problem which is the central motivation for this thesis, we describe a situation which illustrates the context in which that problem can arise.

Suppose that a statistician is asked to determine an optimal policy for changing the bulbs of the street lamps in a city. To this end, the statistician would surely want to estimate the distribution function of the lifetime of the bulbs used for the street lamps in that city. The information given to the statistician is the following. Each street lamp was installed long ago and, since then, each bulb has been replaced as soon as it failed. No data are available regarding past replacements. Instead, the statistician will know, for each of n street lamps, the residual life of the bulb; i.e., the length of time, from the start of the study, until that bulb fails. The statistician's problem is: how can the n observed residual lives be used to estimate the distribution function of bulb lifetimes? This question will be answered in 1.4 on the basis of a mathematical model for the situation described above.

### 1.2. Mathematical model.

Our model for the situation described in 1.1 will be developed in the framework of renewal theory. Recall that a renewal process is a sequence $(Z_i)_1^\infty$ of nonnegative independent random variables with common distribution function G such that $G(0) < 1$, and that G is said to be arithmetic if and only if there exists a $\delta > 0$ such that G is the distribution function of a discrete random variable whose possible values are 0, $\pm\delta$, $\pm 2\delta$, ... . Suppose that an object is put in use until it fails, at which time it is promptly replaced by another object of the same type, and so on, so that the sequence of lifetimes can be represented by a sequence $(Z_i)_1^\infty$ of independent random variables with common distribution function G. Consider such a process at time t. Let X(t) represent the residual life of the object in use at time t. It is known [see, e.g., Feller (1971, p. 370)] that if G is nonarithmetic and the expectation $\mu = \int_0^\infty (1 - G(s))ds$ is finite then

$$\lim_{t\to\infty} P[X(t) \le x] = \int_0^x \frac{1-G(s)}{\mu}\, ds$$

where

$$\text{(i)} \qquad F(x) = \begin{cases} \int_0^x \frac{1-G(s)}{\mu}\, ds & , x \ge 0 , \\ 0 & , x < 0 , \end{cases}$$

is a probability distribution function. Thus, when t is large, the residual life of the object in use at t can be represented by a random variable with distribution function F.

In our introductory example, let G be the distribution function of a random variable representing the lifetime of a randomly selected bulb. It is reasonable to represent the residual life of a randomly selected bulb in use at the start of the study by a random variable X with distribution function F related to G by (i). For n street lamps, the n residual lives of the bulbs in use at the time of the study can be represented by the sequence $(X_i)_1^n$ of independent random variables with common distribution function F. The statistician's problem is to construct an estimator of G, using the independent random variables $(X_i)_1^n$ with common distribution function F. To study large sample properties of the estimator, the statistician will consider an infinite sequence $(X_i)_1^\infty$ and will study the behaviour of the estimator in the limit, as $n \to \infty$ .

1.3. The problem.

The problem can now be stated in mathematical terms. Suppose that G is the distribution function of a nonnegative random variable, with $\mu = \int_0^\infty (1 - G(s))ds < \infty$ , and that F is the distribution function which is linked to G by formula (i). Given a sequence of independent random variables $(X_i)_1^\infty$ with common distribution function F, we want to construct a nonparametric estimator of G in terms of the random variables $(X_i)_1^n$; further, we want to do this so that the estimator has strong consistency properties when $n \to \infty$. Since G = 0 on $(-\infty,0)$ we will only consider the behaviour of its estimator on $[0,\infty)$.

We consider the case where G is continuous. In particular, G(0) = 0; in physical terms this means that we only consider the case where there are no instantaneous failures. In practical terms, the continuity hypothesis for G is not realistic because it is unlikely that the precise residual lives will be observed. But the estimator of G will be seen to be based on the number of failures in each interval of a preassigned partition of $[0,\infty)$. Therefore it is not necessary to know the precise time of each failure, as long as the number of failures is known in each interval of the preassigned partition of $[0,\infty)$.

Let $\tau = \inf\{x: G(x) = 1\}$. When $\tau$ is finite, we consider two cases: $G(\tau^-) = 1$ and $G(\tau^-) < 1$. In relation to the introductory example, the case $G(\tau^-) < 1$ corresponds to the case where no bulb is left in a street lamp for more than a preassigned time $\tau > 0$. In this case, G is discontinuous at $\tau$ but is still continuous on $(-\infty,\tau)$ and on $(\tau,\infty)$.

## 1.4. Method of solution.

Our solution of the problem described in 1.3 is based on the observation that the distribution function F given by (i), which is the distribution function of the residual lives $X_1, X_2, \ldots,$ has a density f given by

$$\text{(ii)} \qquad f(x) = \begin{cases} \dfrac{1 - G(x)}{\mu} & , x \geq 0 ; \\ 0 & , x < 0 ; \end{cases}$$

f is nonincreasing on $[0,\infty)$ and right-continuous at 0. Since $G(0) = 0$ by hypothesis, we have $\mu = (f(0))^{-1}$ and therefore

$$\text{(iii)} \qquad G(x) = \begin{cases} 1 - (f(0))^{-1} f(x) & , x \geq 0 , \\ 0 & , x < 0 . \end{cases}$$

Now, if $f_n$ is an estimator of f then

$$\text{(iv)} \qquad G_n(x) = \begin{cases} 1 - (f_n(0))^{-1} f_n(x) & , x \geq 0 , \\ 0 & , x < 0 \end{cases}$$

is an obvious estimator of G.

As G is a probability distribution function, it is natural to want to estimate it by an estimator which is right-continuous and nondecreasing; if G is estimated as in (iv) then this amounts to wanting that the density estimator $f_n$ be right-continuous and monotone nonincreasing on $[0,\infty)$. To obtain such a monotone estimator of f, we consider a histogram estimator $\hat{f}_n$ of f and transform it into an estimator $\tilde{f}_n$ of f which is monotone nonincreasing on $[0,\infty)$. Thus we obtain a monotone estimator $\tilde{G}_n = 1 - (\tilde{f}_n(0))^{-1}\tilde{f}_n$ of G on $[0,\infty)$.

In chapter 3, we give two methods for obtaining $\tilde{f}_n$ by monotonizing a histogram estimator of f. These methods are based on suggestions by Ayer et al. (1955) for transforming the usual estimates $(\hat{p}_1,\ldots,\hat{p}_m)$ of the probabilities $(p_1,\ldots,p_m)$ into a monotone sequence of estimates $(\tilde{p}_1,\ldots,\tilde{p}_m)$ such that $\tilde{p}_1 \geq \ldots \geq \tilde{p}_m$. We consider this transformation in chapter 2.

We obtain strong uniform consistency results for $\tilde{f}_n$ in chapters 3 and 4. This yields analogous results for $\tilde{G}_n$ which we consider in chapter 5.

The approach described above, based on (iv) and the use of a "monotonized histogram", was suggested by Cox (1969). He suggests that the estimation of f at 0 be done independently from estimating f elsewhere; we, on the other hand, use the value of our estimator $\tilde{f}_n$ at 0 to estimate f(0). Though Cox proposed this approach to estimating G, he did not consider large-sample properties of $\tilde{G}_n$.

## 1.5 Summary and results.

Suppose that some objects have lifetimes which can be represented by a random variable with distribution function G as in 1.2. Suppose that the residual life of such an object at the time of the study can be represented by a random variable with distribution function F related to G by (i). In this thesis, we obtain an estimator $\tilde{G}_n$ of G which is monotone and right-continuous, and we prove strong uniform consistency results and rate of convergence results for $\tilde{G}_n$. Those results are stated in chapter 5. Our estimator $\tilde{G}_n$ of G is obtained by formula (iv), with $f_n = \tilde{f}_n$, where $\tilde{f}_n$ is an estimator of f which is monotone nonincreasing on $[0,\infty)$. Our other chapters are mainly concerned with constructing a monotone non-increasing estimator of f which has some strong uniform consistency and rate of convergence properties.

In chapter 2, we consider the estimation of a sequence of probabilities $(p_1,\ldots,p_m)$ by a sequence of estimators $(\tilde{p}_1,\ldots,\tilde{p}_m)$ such that $\tilde{p}_1 \geq \ldots \geq \tilde{p}_m$ ; this is appropriate when it can be assumed that $p_1 \geq \ldots \geq p_m$. Our reason for considering this problem is that methods for transforming the usual estimator $(\hat{p}_k)_1^m$ into $(\tilde{p}_k)_1^m$ yield analogous methods for transforming a histogram $\hat{f}_n$ into a monotonized histogram $\tilde{f}_n$. Notation and arguments are much simpler in the case of estimation of a finite sequence of probabilities than in the case of densities, and can easily be generalized to the case of densities. The estimation of a finite monotone sequence of probabilities is considered by Ayer et al.; as discussed in 2.1, our presentation completes theirs.

Two monotonization methods are considered in chapter 2. Each yields an analogous monotonization method for the case of histograms; these methods, considered in chapter 3, transform a histogram estimator $\hat{f}_n$ of f into an estimator $\tilde{f}_n$ monotone on $[0,\infty)$. In chapters 3 and 4, f is not necessarily the density specified in (ii) but is, more generally, any density function which vanishes on $(-\infty,0)$, is nonincreasing on $[0,\infty)$ and is right-continuous. Our analysis of the monotonization methods enables us to compare the distance of $\tilde{f}_n$ to f with the distance of $\hat{f}_n$ to f; this comparison shows that the rate of convergence of $\tilde{f}_n$ is at least as good as the rate of convergence of $\hat{f}_n$. On the basis of this observation, and adapting some results by Révész (1972), we obtain some rate of convergence results for $\tilde{f}_n$.

The arguments in chapter 3 also yield some strong uniform consistency results (without rate of convergence) but similar results - with weaker conditions - can be obtained by different arguments; that is done in chapter 4. Our estimator $\tilde{f}_n$ was considered by Barlow et al. (1972); they obtained strong consistency results for $\tilde{f}_n$ at points where f is continuous. In chapter 4, we complete the work of Barlow et al. by obtaining strong consistency at 0 (where f is not continuous), and at $\tau$ (defined in 1.3) when $\tau < \infty$ and f is not continuous at $\tau$. Our arguments are based in part on a proof by Révész (1968). From the strong pointwise consistency results and monotonicity, we obtain strong uniform consistency results.

In chapter 5, results from chapters 3 and 4 are applied to the specific case where f is the density specified in (ii) to obtain the already mentioned strong uniform consistency results for $\tilde{G}_n$.

## 1.6. Notation.

As already noted in 1.5, f has two meanings in this thesis. In chapters 3 and 4, f is any density function which vanishes on $(-\infty,0)$, is nonincreasing on $[0,\infty)$ and is right-continuous. In chapter 5, f is the particular density function which is described in section 1.4. The usual histogram estimator of f is denoted by $\hat{f}_n$, and the monotonized histogram estimator of f is denoted by $\tilde{f}_n$. We use F for the distribution function with density function f and $F_n$ for the empirical distribution function estimating F (e.d.f.), computed from $(X_i)_1^n$; in this thesis, we work with the left-continuous form of the e.d.f., namely

$F_n(x) = n^{-1} \sum_{i=1}^{n} I_{(-\infty,x)}(X_i)$. In the remainder of this thesis, when the word consistency is used it should be understood to mean strong consistency; i.e., convergence of estimators is established in the "almost sure" mode of convergence. When convenient, we write a.s. instead of "with probability one". Finally, when we say that a function g is continuous on an interval we mean that g is continuous in the relative topology of that interval; e.g., to say that g is continuous on $[a,b]$ means that g is continuous at each x in $(a,b)$, is right-continuous at a, and left-continuous at b.

## 2. Monotonization methods.

### 2.1. Introduction

Throughout this chapter, we consider the situation where, for $k = 1,\ldots,m$, a number $N_k$ of independent trials are made of an event which occurs with probability $p_k$. Let $a_k$ denote the number of successes in the k-th trial, and $\hat{p}_k$ the ratio $a_k/N_k$. For each k, as $N_k \to \infty$, $\hat{p}_k \to p_k$, with probability one, by the strong law of large numbers.

The problem considered in this chapter is the following. Suppose that one has reasons for assuming that the probabilities $p_k$ satisfy the inequalities $p_1 \geq \ldots \geq p_m$. Then it is natural to seek estimates $\tilde{p}_1,\ldots,\tilde{p}_m$ which satisfy $\tilde{p}_1 \geq \ldots \geq \tilde{p}_m$; the usual estimates $\hat{p}_1,\ldots,\hat{p}_m$ may fail to have this property.

We consider two approaches to the problem stated above. In 2.6, we state a procedure which uses the data $\left((a_k,N_k)\right)_{k=1}^m$ to transform the sequence $\hat{P} = (\hat{p}_k)_{k=1}^m$ into $\tilde{P} = (\tilde{p}_k)_{k=1}^m$ so that $\tilde{p}_1 \geq \ldots \geq \tilde{p}_m$. On the other hand, in 2.14, we use the data $\left((a_k,N_k)\right)_{k=1}^m$ to obtain a sequence $0 = k_o < k_1 < \ldots < k_\sigma = m$ and put

$$\text{(i)} \qquad \tilde{p}_k = \frac{\sum_{k_{i-1}+1 \leq j \leq k_i} a_k}{\sum_{k_{i-1}+1 \leq j \leq k_i} N_k} \quad \text{when} \quad k_{i-1}+1 \leq k \leq k_i .$$

The first part of this chapter is devoted to showing that those two approaches yield the same result. The second part uses the first of the above-mentioned approaches to compare the distance between $(\tilde{p}_k)_{k=1}^m$ and

$(p_k)_{k=1}^m$ with the distance between $(\hat{p}_k)_{k=1}^m$ and $(p_k)_{k=1}^m$.

Of our two approaches to the monotonization of $(\hat{p}_k)_{k=1}^m$, the first construction of $(\tilde{p}_k)_{k=1}^m$ is based on a suggestion by Ayer et al. (1955), which we consider in more detail in 2.2 . Looking at $(\tilde{p}_k)_{k=1}^m$ this way will yield the comparison of distances mentioned above; the result, given in theorem 2.17, is that, when $p_1 \geq p_2 \geq \cdots \geq p_m$ ,

(ii) $\max_k |p_k - \tilde{p}_k| \leq \max_k |p_k - \hat{p}_k|$ .

The consistency of $(\tilde{p}_k)_{k=1}^m$, when the number of trials increases indefinitely, follows from (ii) and is stated in theorem 2.19 .

Inequality (ii) is not obtained in any of the papers that we consulted; it complements, in the context of the monotonization of estimates of probabilities, the inequality

$$\sum_{1 \leq k \leq m} (p_k - \tilde{p}_k)^2 N_k \leq \sum_{1 \leq k \leq m} (p_k - \hat{p}_k)^2 N_k \ .$$

mentioned by Ayer et al. . In the spirit of this thesis though, inequality (ii) is the first step before a more general analogous inequality is obtained for histograms in chapter 3 .

The second of the above-mentioned constructions of $(\tilde{p}_k)_{k=1}^m$, also based on a suggestion by Ayer et al., is useful for proving that the first method has a well defined, unique final output; this is done in theorem 2.13 . Ayer et al. propose a min-max formula to compute each $\tilde{p}_k$. An analogous formula is used throughout the book of Barlow et al. (1972)

in cases of monotonization more general than those considered in this chapter. For each k,

$$\text{(iii)} \qquad \tilde{p}_k = \min_{1 \le r \le k} \left\{ \max_{k \le s \le m} \frac{\sum_{r \le j \le s} a_j}{\sum_{r \le j \le s} N_j} \right\}.$$

Those are "explicit formulas for the determination of the $(\tilde{p}_k)_{k=1}^{m}$, but are not recommended for calculation" (Ayer et al., 1955, p. 644). Our formula (i), on the other hand, will be seen to be convenient for calculations.

The monotonization methods presented in this chapter are generalized in section 2.21 . The work done in this chapter and that generalization can be adapted to the analogous problem of monotonizing a histogram, which is of interest when estimating a monotone density. That problem will be considered in chapter 3.

The content of this chapter is related, in part, to some sections of Barlow et al. (1972). In that book, there are extensive discussions of some very general ideas which are gathered into the concept of "isotonic regression". They deal with the estimation of isotonic regression in a very general framework; particular instances of their estimator turn out to be precisely our monotone estimators $\tilde{p} = (\tilde{p}_k)_{k=1}^{m}$ , in the context of the estimation of probabilities, and $\tilde{f}_n$, in the context of the estimation of monotone density functions. However, the results which interest us (e.g., theorem 2.17 and theorem 3.12) cannot be obtained from the general considerations of isotonic regression in Barlow et al..

## 2.2. The monotonization method by Ayer et al.

In this section we present a method of monotonization of estimates $(\hat{p}_k)_{k=1}^m$ of $(p_k)_{k=1}^m$ suggested by Ayer et al.. We then present a discussion of the results obtained by Ayer et al. from this method. Finally we make two remarks to clarify the steps involved in this method. Ayer et al. describe their method in the following terms (with the notation changed in order to match to notation herein):

"The maximum likelihood estimates $\tilde{p}_1,\ldots,\tilde{p}_m$ of the numbers $p_1,\ldots,p_m$ may be found in the following way. If $\hat{p}_1 \geq \ldots \geq \hat{p}_m \geq 0$, then $\tilde{p}_k = \hat{p}_k$, $k = 1,2,\ldots,m$. If $\hat{p}_i \leq \hat{p}_{i+1}$ for some $i$ $(i = 1,2,\ldots,m-1)$, then $\tilde{p}_i = \tilde{p}_{i+1}$; the ratios $\hat{p}_i = a_i/N_i$ and $\hat{p}_{i+1} = a_{i+1}/N_{i+1}$ are then replaced in the sequence $\hat{p}_1,\hat{p}_2,\ldots,\hat{p}_m$ by the single ratio $(a_i+a_{i+1})/(N_i+N_{i+1})$, obtaining an ordered set of m-1 ratios. This procedure is repeated until an ordered set of ratios is obtained which are monotone nonincreasing. Then, for each k, $\tilde{p}_k$ is equal to that one of the final set of ratios to which the original ratio $a_k/N_k$ contributed."

Clearly, the method is not fully specified: there might be several orders in which it can be applied, and it is not obvious that whichever order is chosen the method will necessarily yield the same monotone nonincreasing sequence. As shown in 2.12, this is indeed the case. --Ayer et al. state the following results - without proof for the last three.

(1) The existence of a maximum likelihood estimator $(\tilde{p}_k)_{k=1}^m$ subject to the conditions $\tilde{p}_1 \geq \tilde{p}_2 \geq \dots \geq \tilde{p}_m$ .

(2) The uniqueness of the maximum likelihood estimator $(\tilde{p}_k)_{k=1}^m$ which is, for each k, given by the min-max formula 2.1. (iii).

(3) Those $\tilde{p}_k$ can also be obtained by the method of monotonization quoted above.

(4) Since $\hat{p}_k$ converges with probability one to $p_k$ $(k = 1,2,\dots,m)$, by the strong law of large numbers, $\tilde{p}_k$ converges, with probability one, to $p_k$ $(k = 1,2,\dots,m)$.

In sections 2.5 and 2.6 we propose a new description of this method. Theorem 2.12 proves the uniqueness of the monotone output sequence obtained by the method and shows that its k-th element is indeed given by the min-max formula 2.1. (iii). We then show in theorem 2.16 that the monotonized estimate is not worse than the original and statement (4) follows.

Remarks.

A. When $\hat{p}_k$ and $\hat{p}_{k+1}$ are replaced by the single ratio $\pi_k$ then, at the next iteration, $\pi_k$ is treated as a single term and not as an adjacent pair of equal terms. This is illustrated in Figure 2.A, where $a_1 < a_2 = a_3$ and $N_1 = N_2 = N_3 = N$ .

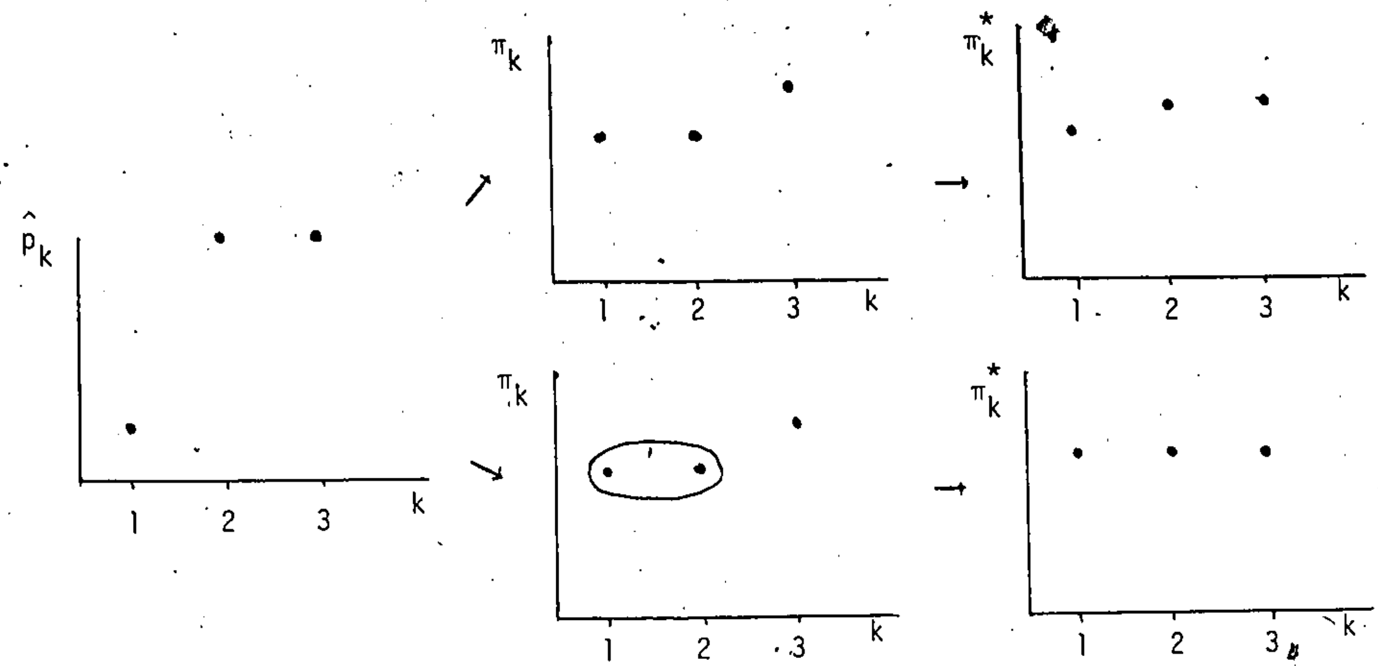



Figure 2.A.

In Figure 2.A, each arrow represents an iteration of the monotonization method. The downward path illustrates how the monotonization should be applied according to Ayer et al.; i.e., the ratios $\hat{p}_1 = a_1/N$ and $\hat{p}_2 = a_2/N$ are replaced in the sequence $(\hat{p}_1,\hat{p}_2,\hat{p}_3)$ by the single ratio $(a_1+a_2)/(2N)$, obtaining an ordered set of 2 ratios. The upward path shows how not to apply the method of monotonization. The first iteration is applied to the pair $\hat{p}_1$ and $\hat{p}_2$ as in the downward path but three distinct ratios are kept. Each further iteration is applied to a pair of ratios and three distinct ratios are kept at each step; this yields a sequence of estimates of $(p_1,p_2,p_3)$ in which either $\pi_1 = \pi_2 < \pi_3$ or $\pi_1 < \pi_2 = \pi_3$ is always the case and the sequence does not converge in finitely many iterations to a nonincreasing estimate.

B. It is not obvious that the method of monotonization always yields a unique monotone nonincreasing estimate. Figure 2.B shows different orders of application of the method; it is not clear at a glance that, e.g., the sequences obtained at state (1a) and at state (2a) can both be transformed to the same monotone nonincreasing estimate. Here again an arrow refers to an iteration of the method of monotonization. Assume that $N_1 = \ldots = N_6 = N$ and that $(a_1,\ldots,a_6) = (3,2,1,3,4,3)$ . Now the starting sequence is $(\hat{p}_1,\hat{p}_2,\ldots,\hat{p}_6) = (3/N,2/N,1/N,3/N,4/N,3/N)$ . We first note that, as $\hat{p}_3 < \hat{p}_4$ and $\hat{p}_4 < \hat{p}_5$, this sequence is not monotone nonincreasing. Therefore we can apply the first iteration either to the pair $\hat{p}_3$ and $\hat{p}_4$ or to the pair $\hat{p}_4$ and $\hat{p}_5$. If we apply it to the pair $\hat{p}_4$ and $\hat{p}_5$ we obtain state (2a), where $\hat{p}_4 = 3/N$ and $\hat{p}_5 = 4/N$ are replaced by the single ratio $(3+4)/(N+N) = 7/2N$ . The new sequence is $(\pi_1,\ldots,\pi_6) = (3/N,2/N,1/N,7/2N,7/2N,3/N)$ . According to Ayer et al., the new sequence should contain only 5 terms because $\hat{p}_4$ and $\hat{p}_5$ should be replaced by only one ratio. It is however more natural to keep sequences of the same length, and we will do so, but it is to be remembered that the new ratio should be considered as replacing <u>both</u> $\hat{p}_4$ and $\hat{p}_5$ in the next iteration; for future iterations, $\hat{p}_4$ and $\hat{p}_5$ are replaced by the single object $(\pi_4,\pi_5)$, where $\pi_4 = \pi_5 = 7/2N$ . This is represented in Figure 2.B by encircling the corresponding points. We apply another iteration to the sequence $(\pi_1,\ldots,\pi_6)$ in state (2a). It can only be applied to the pair $\pi_3$ and $(\pi_4,\pi_5)$. This yields state (2b), where $\pi_3 = 1/N$ and $(\pi_4,\pi_5) = (7/2N,7/2N)$ are replaced by $(8/3N,8/3N,8/3N)$.

Figure 2.B represents all possible paths that the method of monotonization of Ayer et al. could follow if the sequences $(a_1,\ldots,a_6) = (3,2,1,3,4,3)$, $(N_1,\ldots,N_6) = (N,\ldots,N)$ and $(\hat{\beta}_1,\ldots,\hat{\beta}_6) = (3/N,2/N,1/N,3/N,4/N,3/N)$ were given. Note that all paths yield the same final output sequence of state (d).

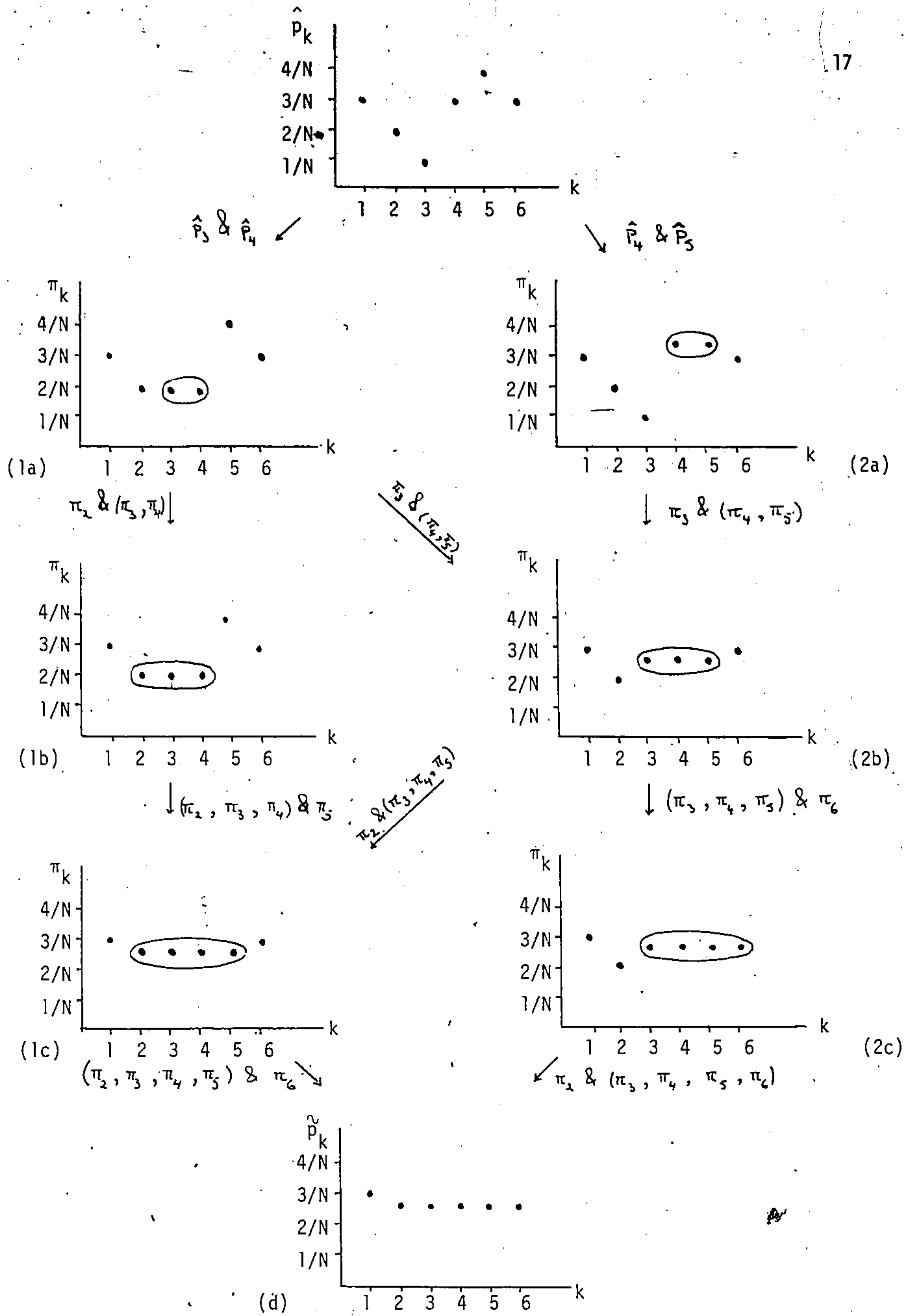
The following sequences are the estimates $(\pi_1,\ldots,\pi_6)$ at each state of Figure 2.B.

(1a) $(\frac{3}{N}, \frac{2}{N}, \frac{4}{2N}, \frac{4}{2N}, \frac{4}{N}, \frac{3}{N})$ (2a) $(\frac{3}{N}, \frac{2}{N}, \frac{1}{N}, \frac{7}{2N}, \frac{7}{2N}, \frac{3}{N})$

(1b) $(\frac{3}{N}, \frac{6}{3N}, \frac{6}{3N}, \frac{6}{3N}, \frac{4}{N}, \frac{3}{N})$ (2b) $(\frac{3}{N}, \frac{2}{N}, \frac{8}{3N}, \frac{8}{3N}, \frac{8}{3N}, \frac{3}{N})$

(1c) $(\frac{3}{N}, \frac{10}{4N}, \frac{10}{4N}, \frac{10}{4N}, \frac{10}{4N}, \frac{3}{N})$ (2c) $(\frac{3}{N}, \frac{2}{N}, \frac{11}{4N}, \frac{11}{4N}, \frac{11}{4N}, \frac{11}{4N})$

(d) $(\frac{3}{N}, \frac{13}{5N}, \frac{13}{5N}, \frac{13}{5N}, \frac{13}{5N}, \frac{13}{5N})$ .



p̂k
4/N
3/N
2/N
1/N
k
1 2 3 4 5 6
p̂3 & p̂4
p̂4 & p̂5
πk
(1a)
(2a)
π2 & (π3, π4)
π3 & (π4, π5)
(1b)
(2b)
(π2, π3, π4) & π5
π2 & (π3, π4, π5)
(π3, π4, π5) & π6
(1c)
(2c)
(π2, π3, π4, π5) & π6
π2 & (π3, π4, π5, π6)
p̃k
(d)


Figure 2.B.

## 2.3. Definitions and remarks.

A. Definition: Let $\Pi = (\pi_1,\ldots,\pi_m)$ be a finite sequence of real numbers. A cluster of $\Pi$ is a subsequence of $\Pi$ consisting of consecutive and equal terms. A maximal cluster is a cluster which cannot be enlarged to yield another cluster; i.e., $(\pi_k: k_1 \leq k \leq k_2)$ is a maximal cluster iff $\pi_{k_1} = \ldots = \pi_{k_2}$ and $\pi_{k_1-1} \neq \pi_{k_1}$ (unless $k_1 = 1$), and $\pi_{k_2+1} \neq \pi_{k_2}$ (unless $k_2 = m$). The cluster $C(r,s)$ of $\Pi$ will denote the subsequence $(\pi_k: r \leq k \leq s)$.

B. Remark: The sets of integers which are indices of distinct maximal clusters form a partition of $\{1,2,\ldots,m\}$. In Figure 2.C the maximal clusters are $(\pi_1)$, $(\pi_2,\pi_3,\pi_4)$ and $(\pi_5)$. The corresponding sets of indices are $\{1\}$, $\{2,3,4\}$ and $\{5\}$; they partition $\{1,2,\ldots,5\}$. Figure 2.C illustrates also that this is not the case for nonmaximal clusters.

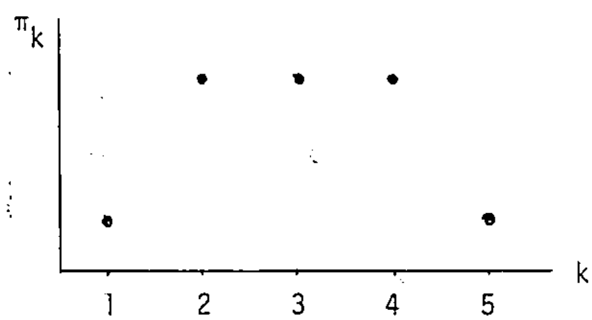



Figure 2.C.

C. Definition: Let $D = \Big((a_1,N_1),\ldots,(a_m,N_m)\Big)$ be the data described in 2.1. For integers r and s such that $1 \leq r \leq s \leq m$, put

$$A(r,s) = \frac{\sum_{r\leq k\leq s} a_k}{\sum_{r\leq k\leq s} N_k}.$$

A cluster $C(r,s)$ of a sequence $\Pi = (\pi_1,\ldots,\pi_m)$ is said to have the average property (with respect to D) if $\pi_k = A(r,s)$ is true for every $\pi_k \in C(r,s)$, i.e., for every $\pi_k$ with $r \le k \le s$.

D. Notation: A sequence will be said to be MNI if it is monotone nonincreasing.

Lemma 2.4. Let $A(\cdot,\cdot)$ be as in definition 2.3.C. For any integers r and s such that $1 \le r \le s \le t \le m$,

(i) $A(r,s) = A(r,t) \Leftrightarrow A(r,t) = A(s+1,t)$,

(ii) $A(r,s) < A(r,t) \Leftrightarrow A(r,t) < A(s+1,t)$,

(iii) $A(r,s) > A(r,t) \Leftrightarrow A(r,t) > A(s+1,t)$,

(iv) $A(r,s) = A(s+1,t) \Rightarrow A(r,s) = A(r,t) = A(s+1,t)$,

(v) $A(r,s) < A(s+1,t) \Rightarrow A(r,s) < A(r,t) < A(s+1,t)$,

(vi) $A(r,s) > A(s+1,t) \Rightarrow A(r,s) > A(r,t) > A(s+1,t)$.

Proof. Let a,b,c, and d be reals, with c and d positive. Now

$$(*)\quad \frac{a}{c} = \frac{a+b}{c+d} \Leftrightarrow a(c+d) = c(a+b) \Leftrightarrow ad = cb$$
$$\Leftrightarrow ad+bd = cb+db \Leftrightarrow \frac{a+b}{c+d} = \frac{b}{d}$$

and

$$(**)\quad \frac{a}{c} = \frac{b}{d} \Rightarrow ad = bc \Rightarrow \begin{cases} ad+ac = bc+ac \Rightarrow \dfrac{a}{c} = \dfrac{a+b}{c+d} \\ ad+bd = bc+bd \Rightarrow \dfrac{a+b}{c+d} = \dfrac{b}{d} \end{cases}.$$

Put

$$a = \sum_{r\le k\le s} a_k, \quad b = \sum_{s+1\le k\le t} a_k,$$

$$c = \sum_{r\le k\le s} N_k, \quad d = \sum_{s+1\le k\le t} N_k.$$

Then (i) follows from (*), (ii) and (iii) follow from (*) with "=" replaced by "<" and ">" respectively, (iv) follows directly from (**), and (v) and (vi) follow from (*) with "=" replaced by "<" and ">" respectively. □

## 2.5. The Standard Monotonization Transformation.

We apply the label Standard Monotonization Transformation (SMT) to a slight modification of the basic step in the procedure proposed by Ayer et al. It is used to transform a non MNI estimate $\Pi = (\pi_1,\ldots,\pi_m)$ of $(p_1,\ldots,p_m)$ into an estimate $\Pi^*$ which is, so to speak, one step closer to being MNI. The SMT can be formalized as follows.

The SMT is applied in relation to the basic data $D = \Big((a_1,N_1),\ldots,(a_m,N_m)\Big)$. As defined in 2.3.C, $A(r,s) = (\sum_{r\le k\le s} a_k)/(\sum_{r\le k\le s} N_k)$ . The input to the SMT is a non MNI sequence $\Pi = (\pi_1,\ldots,\pi_k)$ whose maximal clusters all have the average property (with respect to D), as defined in 2.3.C. The output $\Pi^* = (\pi_1^*,\ldots,\pi_m^*)$ is obtained as follows. Select a pair of contiguous maximal clusters $C(k_1,k_2)$ and $C(k_2+1,k_3)$ of $\Pi$ such that $\pi_{k_2} = A(k_1,k_2) < A(k_2+1,k_3) = \pi_{k_2+1}$; put

$$\pi_k^* = \pi_k \qquad , \text{ if } k \notin \{k_1,\ldots,k_3\},$$

$$\pi_k^* = A(k_1,k_3) = (\sum_{k_1\le k\le k_3} a_k)/(\sum_{k_1\le k\le k_3} N_k) \ , \text{ if } k \in \{k_1,\ldots,k_3\} .$$

Remarks.

A. The maximal clusters of the sequence $(\hat{p}_k)_{k=1}^m = (a_k/N_k)_{k=1}^m$ all have the average property; this will be shown in 2.7.A.

B. Let $\Pi = (\pi_1,\ldots,\pi_m)$ be a non MNI input whose maximal clusters all have the average property. If the SMT is applied to $\Pi$ then the maximal clusters of the output sequence $\Pi^* = (\pi_1^*,\ldots,\pi_m^*)$ all have the average property; this will be shown in 2.7.D.

C. If the input sequence contains $\nu$ maximal clusters then the corresponding output sequence contains at most $\nu-1$ maximal clusters; this will be shown in 2.7.B.

2.6. Monotonization procedure.

This procedure starts with the basic data $D = \left((a_1,N_1),\ldots,(a_m,N_m)\right)$. The initial estimate of $p_k$ is $\hat{p}_k = a_k/N_k$. The procedure yields, as its final output, an MNI estimate $(\tilde{p}_k)_{k=1}^m$.

Procedure.

(1) If $(\hat{p}_k)_{k=1}^m$ is MNI, put $(\tilde{p}_k)_{k=1}^m = (\hat{p}_k)_{k=1}^m$ and stop. Otherwise:

(2) put $\Pi = (\hat{p}_k)_{k=1}^m$ ;

(3) apply the SMT to the input $\Pi$, get a new sequence $\Pi^*$. If $\Pi^*$ is MNI, put $(\tilde{p}_k)_{k=1}^m = \Pi^*$ and stop. Otherwise,

(4) put $\Pi = \Pi^*$ (i.e., replace $\Pi$ by $\Pi^*$) and return to step (3).

Note that all sequences we use as inputs to the SMT are legitimate inputs. By definition of the SMT, the input sequence to the SMT must be such that each of its maximal clusters has the average property. It will be shown in 2.7.A and 2.7.E that, in the procedure described above, the initial input to the SMT, and all subsequent inputs to it, have the required property.

Notes.

A. As the number of maximal clusters decreases at each application of the SMT, the procedure must stop after finitely many iterations.

B. At each iteration, there may be several pairs of maximal clusters which may be selected for application of the SMT. But it will be shown in 2.12 that, no matter how the selection is done, the procedure yields a unique MNI sequence.

2.7. Notes on the monotonization procedure.

A. Fact. The maximal clusters of $(\hat{p}_k)_{k=1}^m$ all have the average property. -- Indeed, given any maximal cluster C(r,s) of $(\hat{p}_k)_{k=1}^m$, we have, by the definition of maximal clusters, $\hat{p}_r = \ldots = \hat{p}_s$; i.e., $a_r/N_r = \ldots = a_s/N_s$ . Then, for $r \leq k \leq s$ , $\hat{p}_k = (a_r+\ldots+a_s)/(N_r+\ldots+N_s) = A(r,s)$ .

B. Fact. If $\Pi = (\pi_1,\ldots,\pi_m)$ is a non MNI sequence with $\nu$ maximal clusters, all with the average property, then the corresponding output sequence $\Pi^*$ is composed of at most $\nu$-1 maximal clusters. -- Indeed, the SMT transforms a pair of maximal clusters into a single cluster and leaves the other clusters unchanged, so the output $\Pi^*$ contains at most $\nu$-1 maximal clusters.

C. Remark. The new cluster formed by the SMT is not necessarily maximal. For instance, in Figure 2.B, the first iteration yielding step (1a) transforms the pair of contiguous maximal clusters C(3,3) and C(4,4) of $(\hat{p}_k)_{k=1}^6$ into the cluster C(3,4) of of $\Pi = (\pi_1,\ldots,\pi_m)$ . Now C(3,4) is not maximal because it is adjacent to a cluster with elements equal to those of C(3,4); C(3,4) is however contained in the maximal cluster C(2,4).

D. Fact. Let $\Pi = (\pi_1,\ldots,\pi_m)$ be a non MNI sequence whose maximal clusters all have the average property, and let $\Pi^*$ be the output sequence obtained by applying the SMT to $\Pi$. Then the maximal clusters of $\Pi^*$ all have the average property.

Proof.

(1) First, we note that if two contiguous clusters C(r,s) and C(s+1,t) of a sequence $\Pi = (\pi_k)_{k=1}^{m}$ have the average property, and are such that $A(r,s) = A(s+1,t)$ , then the concatenated cluster C(r,t) of $\Pi$ has the average property. Indeed, if C(r,s) and C(s+1,t) are contiguous clusters with equal terms then C(r,t) is a cluster. Now it follows from lemma 2.4.(iv) that if $A(r,s) = A(s+1,t)$ then $A(r,s) = A(s+1,t) = A(r,t)$ . But then all the elements of C(r,t) are equal to A(r,t); that is, C(r,t) has the average property.

(2) Now, for some $\sigma$, with $k_0 = 0$ and $k_\sigma = m$, consider the maximal clusters of $\Pi = (\pi_1,\ldots,\pi_m)$:

(i) $C(1,k_1),\ldots,C(k_{i-1}+1,k_i),\ C(k_i+1,k_{i+1}),\ldots,C(k_{\sigma-1}+1,m)$ .

Suppose that they have the average property and that $\Pi$ is not MNI so that the SMT can be applied. Let i be such that $A(k_{i-1}+1,k_i) < A(k_i+1,k_{i+1})$ and such that the pair of maximal clusters $C(k_{i-1}+1,k_i)$ and $C(k_i+1,k_{i+1})$ is selected for concatenation.

The new cluster $C(k_{i-1}+1,k_{i+1})$ has the average property by definition of the SMT. Now, if the sequence (i) of maximal clusters is replaced by a sequence which differs from (i) only in that the pair $C(k_{i-1}+1,k_i)$ and $C(k_i+1,k_{i+1})$ is replaced by $C(k_{i-1}+1,k_{i+1})$, one obtains the sequence of clusters

(ii) $C(1,k_1),\ldots,C(k_{i-1}+1,k_{i+1}),\ldots,C(k_{\sigma-1}+1,m)$ .

All those are clusters of the output sequence $\Pi^*$ and each has the average property. Now the maximal clusters of $\Pi^*$ are either as shown, or concatenations of contiguous clusters with equal elements. There are four possibilities:

(iii) $A(k_{i-2}+1,k_{i-1}) \neq A(k_{i-1}+1,k_{i+1}) \neq A(k_{i+1}+1,k_{i+2})$

(iv) $A(k_{i-2}+1,k_{i-1}) \neq A(k_{i-1}+1,k_{i+1}) = A(k_{i+1}+1,k_{i+2})$

(v) $A(k_{i-2}+1,k_{i-1}) = A(k_{i-1}+1,k_{i+1}) \neq A(k_{i+1}+1,k_{i+2})$

(vi) $A(k_{i-2}+1,k_{i-1}) = A(k_{i-1}+1,k_{i+1}) = A(k_{i+1}+1,k_{i+2})$ .

In case (iii), the input maximal clusters $C(k_{i-1}+1,k_i)$ and $C(k_i+1,k_{i+1})$ are replaced by the maximal cluster $C(k_{i-1}+1,k_{i+1})$, and the other input maximal clusters remain unchanged. As already noted in (1), $C(k_{i-1}+1,k_{i+1})$ has the average property.

In case (iv), the input maximal clusters $C(k_{i-1}+1,k_i)$ and $C(k_i+1,k_{i+1})$ are replaced by the cluster $C(k_{i-1}+1,k_{i+1})$, which is then concatenated with $C(k_{i+1}+1,k_{i+2})$ to arrive at the maximal cluster $C(k_{i-1}+1,k_{i+2})$; the other input maximal clusters remain unchanged. As already noted in (1), $C(k_{i-1}+1,k_{i+2})$ has the average property.

Cases (v) and (vi) are similar to case (iv). □

E. Fact. Each maximal cluster of a sequence $\Pi = (\pi_1,\ldots,\pi_m)$ obtained in the course of procedure 2.6 has the average property.

Proof. If $(\beta_k)_{k=1}^m$ is MNI then $(\tilde{p}_k)_{k=1}^m = (\beta_k)_{k=1}^m$; there is no application of the SMT and this is the only sequence obtained in the course of procedure 2.6. In view of 2.7.A, each maximal cluster of $(\beta_k)_{k=1}^m$ has the average property. Thus the stated Fact is indeed true when $(\beta_k)_{k=1}^m$ is MNI.

If $(\beta_k)_{k=1}^m$ is not MNI then, in view of 2.7.A and 2.7.D, the output of the first application of the SMT is a sequence $\Pi$ each of whose maximal clusters has the average property. In view of 2.7.D, it follows that this remains true for the output of each subsequent applications of the SMT. In particular, this remains true for the final output sequence $(\tilde{p}_k)_{k=1}^m$. □

Proposition 2.8. With $1 \leq \nu \leq m$, let L be the largest integer $\ell \geq \nu$ for which

$$A(\nu,\ell) = \max_{\nu\leq s\leq m} A(\nu,s)$$

is true. Then, for all $\nu \leq k \leq L$,

$$A(\nu,L) = \max_{\nu\leq s\leq m} A(\nu,s) = \min_{\nu\leq r\leq k} \left\{ \max_{k\leq s\leq m} A(r,s) \right\} .$$

Proof. Observe first that, for $\nu \leq k \leq L$,

$$A(\nu,L) = \max_{\nu\leq s\leq m} A(\nu,s) \geq \max_{k\leq s\leq m} A(\nu,s) \geq A(\nu,L) .$$

Thus

(i) $$A(\nu,L) = \max_{k\le s\le m} A(\nu,s) \text{ when } \nu \le k \le L .$$

Consider the case where $\nu < L$. If $\nu < u \le L$ then $\nu \le u-1 < L$, hence

$$A(\nu,u-1) \le \max_{\nu\le s\le m} A(\nu,s) = A(\nu,L) ;$$

now, by lemma 2.4(i) and (ii), with $r = \nu$ , $s = u-1$, and $t = L$,

(ii) $$A(\nu,L) \le A(u,L) .$$

As (ii) is trivially true when $u = \nu$ , it follows that (ii) is true when $\nu \le u \le L$. Thus, without assuming that $\nu < L$,

(iii) $$\max_{k\le s\le m} A(u,s) \ge \max_{L\le s\le m} A(u,s) \ge A(u,L) \ge A(\nu,L)$$

$$\text{when } \nu \le u \le k \le L .$$

From (i) and (iii) we conclude that, for $\nu \le k \le L$,

$$A(\nu,L) = \max_{k\le s\le m} A(\nu,s) \ge \min_{\nu\le r\le k} \left\{ \max_{k\le s\le m} A(r,s) \right\} \ge A(\nu,L) .$$ □

<u>Proposition 2.9</u>. With $1 \le \nu \le m$, let L be the largest integer $\ell \ge \nu$ for which

$$A(\nu,\ell) = \max_{\nu\le s\le m} A(\nu,s)$$

is true. If $L < m$ then, for $L+1 \le k \le m$,

(*) $$\min_{\nu\le r\le k} \left\{ \max_{k\le s\le m} A(r,s) \right\} = \min_{L+1\le r\le k} \left\{ \max_{k\le s\le m} A(r,s) \right\} < A(\nu,L) .$$

<u>Proof</u>. Suppose that $L < m$ and consider $k \in \{L+1,\dots,m\}$.

If $L+1 \le w \le m$ then, from the definition of L, $A(\nu,L) > A(\nu,w)$.

From this and lemma 2.4.(iii), with $r = \nu$, $s = L$, and $t = w$, it follows that $A(\nu,w) > A(L+1,w)$. Combining the two inequalities, we have

(i) $\quad A(L+1,w) < A(\nu,L)$ when $w \geq L+1$ .

As noted in (ii) in the proof of proposition 2.8,

(ii) $\quad A(\nu,L) \leq A(u,L)$ when $\nu \leq u \leq L$ .

By (i) and (ii),

$$A(L+1,w) < A(u,L) \text{ when } \nu \leq u \leq L < L+1 \leq w .$$

From this and lemma 2.4.(vi), with $r = u$, $s = L$, and $t = w$, it follows that

(iii) $\quad A(L+1,w) < A(u,w)$ when $\nu \leq u \leq L < L+1 \leq w$ .

Consider $u$ such that $\nu \leq u \leq L$. Since $k \geq L+1$ , (iii) is true for every $w \geq k$ ; hence

(iv) $\quad \max\limits_{k \leq w \leq m} A(L+1,w) < \max\limits_{k \leq w \leq m} A(u,w)$ when $\nu \leq u \leq L$ .

From (i),

$$A(L+1,w) < A(\nu,L)$$

is true for every $w \geq k$; it follows that

$$\max_{k \leq w \leq m} A(L+1,w) < A(\nu,L) .$$

We conclude that

(v) $$\min_{\nu\le r\le k}\left\{\max_{k\le w\le m} A(r,w)\right\} < A(\nu,L) .$$

Let $\rho \in \{\nu,\ldots,k\}$ be such that

(vi) $$\max_{k\le s\le m} A(\rho,s) = \min_{\nu\le r\le k}\left\{\max_{k\le s\le m} A(r,s)\right\} .$$

Now, since $k \ge L+1$,

$$\min_{\nu\le r\le k}\left\{\max_{k\le s\le m} A(r,s)\right\} \le \max_{k\le s\le m} A(L+1,s) ;$$

from this and (iv) and (vi),

$$\max_{k\le s\le m} A(\rho,s) < \max_{k\le s\le m} A(u,s)$$

is true for all u such that $\nu \le u \le L$ ; it follows that $\rho \notin \{\nu,\ldots,L\}$.
Therefore (vi) holds for some $\rho$ such that $L+1 \le \rho \le k$ ; we conclude from this remark that

$$\min_{L+1\le r\le k}\left\{\max_{k\le s\le m} A(r,s)\right\} = \min_{\nu\le r\le k}\left\{\max_{k\le s\le m} A(r,s)\right\} .$$

Now (*) follows from this and (v). □

Now we claim that the sequence $(\tilde{p}_k)_{k=1}^m$ obtained as a final output of procedure 2.6 is unique and can be computed, for each k, by the formula

$$\tilde{p}_k = \min_{1\le r\le k}\left\{\max_{k\le s\le m} A(r,s)\right\} .$$

But before we prove this claim in theorem 2.12, we obtain some properties of the sequence $\Gamma = \left(\min_{1\le r\le k}\left\{\max_{k\le s\le m} A(r,s)\right\}\right)_{k=1}^m .$

In the rest of this chapter we will use the sequence $0 = k_o < k_1 < \ldots < k_\sigma = m$ whose importance has already been pointed out in 2.1. It is obtained as follows.

Definition 2.10.

Put $k_o = 0$; for $i \geq 1$, if $k_{i-1} < m$ let $k_i$ be the largest integer $k \leq m$ for which

$$A(k_{i-1}+1,k) = \max_{k_{i-1}+1 \leq s \leq m} A(k_{i-1}+1,s) .$$

For some integer $\sigma$, $k_\sigma = m$.

In the next two results, we denote a cluster $C(r,s)$ of a given sequence $\Pi = (\pi_1,\ldots,\pi_m)$ by $C(r,s;\Pi)$.

Proposition 2.11. Put $\Gamma = (\gamma_1,\ldots,\gamma_m)$, with $\gamma_k = \min_{1 \leq r \leq k} \left\{ \max_{k \leq s \leq m} A(r,s) \right\}$ for $1 \leq k \leq m$, and let $(k_o,\ldots,k_\sigma)$ be as in 2.10. Now:

(a) The sequence $\Gamma$ is MNI.

(b) Each maximal cluster of $\Gamma$ has the average property; i.e., if $C(r,s;\Gamma)$ is a maximal cluster of $\Gamma$ then $\gamma_k = A(r,s)$ for all $\gamma_k \in C(r,s;\Gamma)$.

(c) The maximal clusters of $\Gamma$ are $C(k_{i-1}+1,k_i;\Gamma)$, $1 \leq i \leq \sigma$.

Proof. For $1 \leq i \leq \sigma$, let (*) be the following statement:

$C(k_{i-1}+1,k_i;\Gamma)$ is a maximal cluster of $\Gamma$; it has the average property and, if $i < \sigma$, then

$$\text{(i)} \qquad \gamma_k = \min_{1 \leq r \leq k} \left\{ \max_{k \leq s \leq m} A(r,s) \right\} = \min_{k_i+1 \leq r \leq k} \left\{ \max_{k \leq s \leq m} A(r,s) \right\} < A(k_{i-1}+1,k_i)$$

for $k_i+1 \leq k \leq m$.

We show first that (*) is true for $i = 1$. For $k \geq k_1+1$, (i) follows from proposition 2.9, with $\nu = 1$ and $L = k_1$. Further, by proposition 2.8, with $\nu = 1$ and $L = k_1$, $\gamma_k = A(1,k_1)$ for $1 \le k \le k_1$; therefore $C(1,k_1;\Gamma)$ is a cluster and it has the average property. By this remark and (i), $\gamma_{k_1+1} < A(1,k_1) = \gamma_{k_1}$ and it follows that $C(1,k_1;\Gamma)$ is a maximal cluster.

Next, suppose that (*) is true for $i = j < \sigma$. If $j+1 < \sigma$ then, from (i),

$$\text{(ii)} \qquad \gamma_k = \min_{k_j+1 \le r \le k} \left\{ \max_{k \le s \le m} A(r,s) \right\} \text{ when } k_{j+1}+1 \le k \le m;$$

further, from proposition 2.9, with $\nu = k_j+1$ and $L = k_{j+1}$,

$$\text{(iii)} \qquad \min_{k_j+1 \le r \le k} \left\{ \max_{k \le s \le m} A(r,s) \right\} = \min_{k_{j+1}+1 \le r \le k} \left\{ \max_{k \le s \le m} A(r,s) \right\}$$

$$< A(k_j+1,k_{j+1}) \quad \text{when } k_{j+1}+1 \le k \le m .$$

We conclude from (ii) and (iii) that

$$\text{(iv)} \qquad \gamma_k = \min_{k_j+1 \le r \le k} \left\{ \max_{k \le s \le m} A(r,s) \right\} = \min_{k_{j+1}+1 \le r \le k} \left\{ \max_{k \le s \le m} A(r,s) \right\}$$

$$< A(k_j+1,k_{j+1}) \quad \text{when } k_{j+1}+1 \le k \le m .$$

Further, by proposition 2.8, with $\nu = k_j+1$ and $L = k_{j+1}$,
$\min_{k_j+1 \le r \le k} \left\{ \max_{k \le s \le m} A(r,s) \right\} = A(k_j+1,k_{j+1})$ for $k_j+1 \le k \le k_{j+1}$. It now follows from (ii) that $\gamma_k = A(k_j+1,k_{j+1})$ when $k_j+1 \le k \le k_{j+1}$, hence $C(k_j+1,k_{j+1};\Gamma)$ is a cluster with the average property. By this remark and (iv), $\gamma_{k_{j+1}+1} < A(k_j+1,k_{j+1}) = \gamma_{k_{j+1}}$; and, since (i) is assumed true for $i = j$, $\gamma_{k_j+1} < A(k_{j-1}+1,k_j) = \gamma_{k_j}$. Thus $\gamma_{k_j} > \gamma_{k_j+1} = \dots = \gamma_{k_{j+1}} > \gamma_{k_{j+1}+1}$. Therefore the cluster $C(k_j+1,k_{j+1};\Gamma)$

is maximal and has the average property. From this and (iv), (*) is true for $i = j+1$ .

Now it follows by induction that statement (*) is true for $1 \leq i \leq \sigma$ . This proves conclusions (b) and (c). But then, for $1 \leq i \leq \sigma$ and $k_{i-1}+1 \leq k \leq k_i$ ,

$$\gamma_k = A(k_{i-1}+1, k_i) .$$

It follows from this and statement (*) that

$$\begin{array}{lllll}
\gamma_1 & = \cdots = \gamma_{k_1} & = A(1,k_1) & > \gamma_{k_1+1} & = A(k_1+1,k_2) , \\
\gamma_{k_1+1} & = \cdots = \gamma_{k_2} & = A(k_1+1,k_2) & > \gamma_{k_2+1} & = A(k_2+1,k_3) , \\
\vdots & & \vdots & \vdots & \\
\gamma_{k_{\sigma-2}+1} & = \cdots = \gamma_{k_{\sigma-1}} & = A(k_{\sigma-2}+1,k_{\sigma-1}) & > \gamma_{k_{\sigma-1}+1} & = A(k_{\sigma-1}+1,m) .
\end{array}$$

Thus $A(1,k_1) > A(k_1+1,k_2) > \ldots > A(k_{\sigma-1}+1,m)$ and

$$\gamma_1 = \cdots = \gamma_{k_1} > \gamma_{k_1+1} = \cdots = \gamma_{k_2} > \cdots \gamma_{k_{\sigma-1}} > \gamma_{k_{\sigma-1}+1} = \cdots = \gamma_m ,$$

which concludes the proof of (a). □

Proposition 2.12. Let $(k_0,\ldots,k_\sigma)$ be as in definition 2.10. If $i < \sigma$ then

$$A(u,k_i) > A(k_i+1,w) \quad \text{when } u \leq k_i < k_i+1 \leq w .$$

Proof. Let $i$ be such that $1 \leq i < \sigma$ . Consider first the case $i > 1$. Let j be an integer such that $1 \leq j+1 \leq i$ . By proposition 2.11,

$$A(1,k_1) > \ldots > A(k_{\sigma-1}+1,m) ;$$

for integers $\alpha$ and $\beta$ such that $0 \le \alpha < \beta \le \sigma-1$, it follows from this and lemma 2.4. (vi) that

$$A(k_\alpha+1,k_\beta) > A(k_\beta+1,k_{\beta+1}) .$$

Therefore, since $0 \le j < i \le \sigma-1$,

(i) $\quad A(k_j+1,k_i) > A(k_i+1,k_{i+1})$ .

By definition of $k_{i+1}$,

$$A(k_i+1,k_{i+1}) \ge A(k_i+1,w) \quad \text{when} \quad w \ge k_i+1 .$$

Thus from this last inequality and (i), if $0 \le j < i \le \sigma-1$ then

(ii) $\quad A(k_j+1,k_i) > A(k_i+1,w)$ when $w \ge k_i+1$ :

Now, as noted just after (ii) in the proof of proposition 2.9, with $\nu = k_{\ell-1}+1$ and $L = k_\ell$ ,

(iii) $\quad A(u,k_\ell) > A(k_\ell+1,w)$ when $k_{\ell-1}+1 \le u \le k_\ell < k_\ell+1 \le w$ .

With $\ell = 1$, (iii) yields

(iv) $\quad A(u,k_1) > A(k_1+1,w)$ when $u \le k_1 < k_1+1 \le w$ .

If $1 \le j < i \le \sigma-1$ then, with $\ell = j$ and $w = k_i$ , (iii) yields

$$A(u,k_j) > A(k_j+1,k_i) \quad \text{when} \quad k_{j-1}+1 \le u \le k_j .$$

By lemma 2.4. (vi), this yields

$$A(u,k_i) > A(k_j+1,k_i) \quad \text{when} \quad k_{j-1}+1 \leq u \leq k_j .$$

Therefore, from this last inequality and (ii), for each j such that $1 \leq j < i$ ,

(v) $\qquad A(u,k_i) > A(k_i+1,w)$ when $k_{j-1}+1 \leq u \leq k_j$ and $k_i+1 \leq w$ .

When $i > 1$, if $u \leq k_i$ then either $k_{i-1}+1 \leq u \leq k_i$ or $u \leq k_{i-1}$ ; the conclusion of the proposition follows from (iii), with $\ell = i$ , in the first situation - and from (v) in the second situation. Finally, if $i = 1$ then the conclusion is given by (iv). □

Theorem 2.13. The final MNI output $(\tilde{p}_k)_{k=1}^m$ of procedure 2.6 is unique and, for $1 \leq k \leq m$ ,

$$\tilde{p}_k = \min_{1 \leq r \leq k} \left\{ \max_{k \leq s \leq m} A(r,s) \right\} .$$

Proof. As in proposition 2.11, we write $\gamma_k$ for $\min_{1 \leq r \leq k} \left\{ \max_{k \leq s \leq m} A(r,s) \right\}$ and $\Gamma$ for $(\gamma_k)_{k=1}^m$ .

Let $\tilde{\mathcal{P}} = (\tilde{p}_k)_{k=1}^m$ be a final output obtained by procedure 2.6. The maximal clusters of both $\tilde{\mathcal{P}}$ and $\Gamma$ have the average property, by 2.7.E and 2.11 respectively. Therefore, to show that the two sequences $\tilde{\mathcal{P}}$ and $\Gamma$ are equal, it suffices to prove that their maximal clusters are the same. This also shows the uniqueness of the final output.

Let $(k_0,\ldots,k_\sigma)$ be as in definition 2.10 and consider the following statements.

(a) $\tilde{p}_{k_i} > \tilde{p}_{k_i+1}$ when $1 \leq i < \sigma$ .

(b) $C(k_{i-1}+1, k_i; \tilde{P})$ is a maximal cluster of $\tilde{P}$ for every $i \in \{1, \ldots, \sigma\}$ .

Consider i such that $1 \leq i < \sigma$ . From proposition 2.12,

(i) $\quad A(u,k_i) > A(k_i+1,w)$ when $u \leq k_i < k_{i+1} \leq w$ .

With $u = k_i$ and $w = k_i+1$ , (i) yields

(ii) $\quad \hat{p}_{k_i} = A(k_i,k_i) > A(k_i+1,k_i+1) = \hat{p}_{k_i+1}$ .

If $\hat{P} = (\hat{p}_k)_{k=1}^m$ is MNI then procedure 2.6 has only one step - where $\hat{P}$ is the input and $\tilde{P} = \hat{P}$ is the final output. Thus (a) is true in this case.

Next, suppose that $\hat{P}$ is not MNI. We show that, in this case, if the input $\Pi$ to procedure 2.6 is such that $\pi_{k_i} > \pi_{k_i+1}$ then the output $\Pi^*$ is such that $\pi^*_{k_i} > \pi^*_{k_i+1}$ . Suppose that $\pi_{k_i} > \pi_{k_i+1}$ . Note that $\pi_{k_i}$ is the last element of a maximal cluster of $\Pi$ and $\pi_{k_i+1}$ is the first element of an adjacent maximal cluster, say $C(u,k_i)$ and $C(k_i+1,w)$, respectively. But since $\pi_{k_i} > \pi_{k_i+1}$, the SMT will not be applied to this pair of maximal clusters. If neither $C(u,k_i)$ nor $C(k_i+1,w)$ is affected by the SMT then $\pi^*_{k_i} = \pi_{k_i} > \pi_{k_i+1} = \pi^*_{k_i+1}$ . If $C(r,u-1)$ gets combined with $C(u,k_i)$ then $\pi^*_{k_i} = A(r,k_i)$ while $\pi^*_{k_i+1} = \pi_{k_i+1} = A(k_i+1,w)$; and, by (i) $\pi^*_{k_i} > \pi^*_{k_i+1}$ . Similarly if the SMT affects $C(k_i+1,w)$. -- Now, by (ii) and what was just shown, the first output sequence $\Pi^*$ is such that $\pi^*_{k_i} > \pi^*_{k_i+1}$ . This sequence is in turn the input sequence $\Pi$ for the second application of the SMT, if a second application is needed. By what was just shown, the output

$\Pi^*$ of the second application again has the property $\pi^*_{k_i} > \pi^*_{k_i+1}$. Continuing thus until the last application of the SMT, whose output is $\tilde{P}$, we see that (a) is again true.

Now consider i such that $1 \leq i \leq \sigma$. In view of (a), if $C(k_{i-1}+1,k_i;\tilde{P})$ is a cluster of $\tilde{P}$ it must be maximal; indeed, $\tilde{p}_{k_{i-1}} > \tilde{p}_{k_{i-1}+1}$, so that $\tilde{p}_{k_{i-1}+1}$ is the first element of a maximal cluster, and $\tilde{p}_{k_i} > \tilde{p}_{k_i+1}$, so that $\tilde{p}_{k_i}$ is the last element of a maximal cluster.

Suppose that $C(k_{i-1}+1,k_i;\tilde{P})$ __is not__ a cluster of $\tilde{P}$. Then, in view of (a), there exists integers $u_0$ and $u_1$ such that $k_{i-1}+1 \leq u_0 < u_1 \leq k_i$ and such that $C(k_{i-1}+1,u_0;\tilde{P})$ and $C(u_1,k_i;\tilde{P})$ are two maximal clusters of $\tilde{P}$. Those two clusters have the average property by 2.7.E; therefore $\tilde{p}_{k_{i-1}+1} = A(k_{i-1}+1,u_0)$ and $\tilde{p}_{k_i} = A(u_1,k_i)$. Furthermore, $\tilde{p}_{k_{i-1}+1}$ and $\tilde{p}_{k_i}$ are elements of the MNI sequence $\tilde{P}$ and they belong to distinct maximal clusters by hypothesis; therefore

$$A(k_{i-1}+1,u_0) > A(u_1,k_i) .$$

But, as noted in (ii) in the proof of proposition 2.9, with $\nu = k_{i-1}+1$ and $L = k_i$,

$$A(u,k_i) \geq A(k_{i-1}+1,k_i) \quad \text{when} \quad k_{i-1}+1 \leq u \leq k_i .$$

It follows from the last two inequalities that

$$A(k_{i-1}+1,u_0) > A(k_{i-1}+1,k_i) ,$$

which contradicts the definition of $k_i$. We conclude that our supposition is false; i.e., $C(k_{i-1}+1,k_i;\tilde{P})$ __is__ a cluster of $\tilde{P}$. As already noted, this implies that $C(k_{i-1}+1,k_i;P)$ is in fact a maximal cluster of $\tilde{P}$. As i was arbitrary, (b) follows.

From (b) and 2.7.E, the maximal clusters of $\tilde{P}$ are $C(1,k_1;\tilde{P}),\ldots,C(k_{\sigma-1},m;\tilde{P})$ and $A(k_{i-1}+1,k_i)$ is the value of each term in $C(k_{i-1}+1,k_i;\tilde{P})$. From proposition 2.11, the maximal clusters of $\Gamma$ are $C(1,k_1;\Gamma),\ldots,C(k_{\sigma-1},m;\Gamma)$ and $A(k_{i-1}+1,k_i)$ is the value of each term in $C(k_{i-1}+1,k_i;\Gamma)$ .-- We conclude that $\tilde{P}$ and $\Gamma$ are <u>one and the same sequence</u>. The conclusion of the theorem now follows from the definition of $\gamma_k$. □

## <u>2.14. Monotonization algorithm</u>.

Let $D = \left((a_1,N_1),\ldots,(a_m,N_m)\right)$ be the starting sequence of data, with $a_k$ and $N_k$ as in 2.1. Let $\hat{P} = (\hat{p}_k)_{k=1}^m$ , where $\hat{p}_k = a_k/N_k$ , be the starting sequence to be monotonized.

<u>Algorithm</u>.

(1) If $\hat{P}$ is MNI, put $\tilde{P} = \hat{P}$ and stop. Otherwise, put $k_0 = 0$ and $i = 1$ and go to step (2).

(2) Given $k_{i-1} < m$ , let $k_i$ be the largest integer $\ell$ such that

$$A(k_{i-1}+1,\ell) = \max_{k_{i-1}+1\leq s\leq m} A(k_{i-1}+1,s) .$$

For $k_{i-1}+1 \leq k \leq k_i$ , put $\tilde{p}_k = A(k_{i-1}+1,k_i)$ . Go to step (3).

(3) If $k_i = m$ , stop. Otherwise, increase i by 1 and return to step (2).

Note that, since m is finite, there is an interger $\sigma$ such that $k_\sigma = m$ . The sequence $(k_i)_{i=0}^{\sigma}$ obtained here is the same as the one obtained by the procedure in definition 2.10. It follows from proposition 2.11 and theorem 2.13 that the final output sequence of this algorithm is the same as the one obtained in procedure 2.6.

Proposition 2.15. Let a,b and c be nonnegative real numbers. Let $I_1$ and $I_2$ be two nonempty real intervals where $I_1$ is to the left of $I_2$. Now, if h is any real-valued function, nonincreasing on $I_1 \cup I_2$, and if $a < b < c$, then

$$(*) \qquad \sup_{t\varepsilon I_1 \cup I_2} |h(t)-b| < \max\left\{\sup_{t\varepsilon I_1} |h(t)-a|, \sup_{t\varepsilon I_2} |h(t)-c|\right\} .$$

Proof. The proposition follows if inequality (*) is verified, under the stated hypotheses, in the following three cases:

(i) $\inf_{t\varepsilon I_1 \cup I_2} h(t) > b$ ,

(ii) $\inf_{t\varepsilon I_1 \cup I_2} h(t) \le b \le \sup_{t\varepsilon I_1 \cup I_2} h(t)$ ,

(iii) $\sup_{t\varepsilon I_1 \cup I_2} h(t) < b$ .

Since h is nonincreasing, $\sup_{t\varepsilon I_1 \cup I_2} h(t) = \sup_{t\varepsilon I_1} h(t)$ and $\inf_{t\varepsilon I_1 \cup I_2} h(t) = \inf_{t\varepsilon I_2} h(t)$ .

Case (i): Suppose that $\inf_{t\in I_1 \cup I_2} h(t) > b$ . Then

$$\sup_{t\in I_1\cup I_2} |h(t)-b| = \sup_{t\in I_1\cup I_2} h(t)-b = \sup_{t\in I_1} h(t)-b$$
$$< \sup_{t\in I_1} h(t)-a = \sup_{t\in I_1} |h(t)-a|$$
$$\leq \max\left\{\sup_{t\in I_1} |h(t)-a|, \sup_{t\in I_2} |h(t)-c|\right\}.$$

Case (ii): Suppose that $\inf_{t\in I_1\cup I_2} h(t) \leq b \leq \sup_{t\in I_1\cup I_2} h(t)$. Then

$$\sup_{t\in I_1\cup I_2} |h(t)-b| = \max\left\{ \sup_{t\in I_1\cup I_2} h(t)-b , b- \inf_{t\in I_1\cup I_2} h(t)\right\}$$
$$< \max\left\{\sup_{t\in I_1} h(t)-a , c-\inf_{t\in I_2} h(t)\right\}$$
$$\leq \max\left\{\sup_{t\in I_1} |h(t)-a|, \sup_{t\in I_2} |h(t)-c|\right\} .$$

Case (iii): Suppose that $\sup_{t\in I_1\cup I_2} h(t) < b$. Now

$$\sup_{t\in I_1\cup I_2} |h(t)-b| = \sup_{t\in I_1\cup I_2} \{b-h(t)\} = b - \inf_{t\in I_1\cup I_2} h(t) = b - \inf_{t\in I_2} h(t)$$
$$< c - \inf_{t\in I_2} h(t) = \sup_{t\in I_2} |h(t)-c|$$
$$\leq \max\left\{\sup_{t\in I_1} |h(t)-a|, \sup_{t\in I_2} |h(t)-c|\right\} .$$

□

<u>Lemma 2.16</u>. Given $\Pi = (\pi_k)_{k=1}^m$ , a non MNI sequence whose maximal clusters have the average property, let $\Pi^* = (\pi_k^*)_{k=1}^m$ be the output sequence obtained when the SMT is applied to $\Pi$. If $(b_k)_{k=1}^m$ is an MNI sequence of real numbers then

$$\max_{1\le k\le m}|b_k - \pi_k^*| \le \max_{1\le k\le m}|b_k - \pi_k| .$$

Proof. The non MNI sequence $\Pi = (\pi_k)_{k=1}^m$ is composed of maximal clusters, each of which has the average property. Apply the SMT to $\Pi$. Then there is a pair of maximal clusters $C(k_1,k_2)$ and $C(k_2+1,k_3)$ of $\Pi$ such that $A(k_1,k_2) < A(k_2+1,k_3)$ and such that the output $\Pi^*$ is

$$\begin{aligned} \pi_k^* &= \pi_k \quad , &&\text{if } k \notin \{k_1,\dots,k_3\}, \\ \pi_k^* &= A(k_1,k_3) \quad , &&\text{if } k \in \{k_1,\dots,k_3\}. \end{aligned}$$

We then have

(i) $\max_\ell|b_\ell-\pi_\ell^*| = \max_\ell|b_\ell-\pi_\ell|$ when $\ell \notin \{k_1,\dots,k_3\}$ .

Now let $I_1 = [k_1,k_2+1)$ and $I_2 = [k_2+1,k_3+1)$ , and define h as follows:

$$h(t) = \begin{cases} b_k & , \text{ if } t \in [k,k+1),\ k_1 \le k \le k_3 , \\ \text{arbitrary or undefined outside } I_1 \cup I_2 . \end{cases}$$

Since $A(k_1,k_2) < A(k_2+1,k_3)$ , it follows by lemma 2.4.(v) with $r = k_1$ , $s = k_2$ and $t = k_3$ that $A(k_1,k_2) < A(k_1,k_3) < A(k_2+1,k_3)$; hence $\pi_{k_1} = \dots = \pi_{k_2} < \pi_{k_1}^* = \dots = \pi_{k_3}^* < \pi_{k_2+1} = \dots = \pi_{k_3}$ . Then it follows by proposition 2.15 that

$$\begin{aligned} \max_{k_1\le k\le k_3}|b_k - \pi_k^*| &= \sup_{t\in I_1\cup I_2}|h(t) - \pi_{k_1}^*| \\ &< \max\left\{\sup_{t\in I_1}|h(t)-\pi_{k_1}|,\ \sup_{t\in I_2}|h(t)-\pi_{k_2+1}|\right\} \\ &= \max\left\{\max_{k_1\le k\le k_2}|b_k-\pi_k|,\ \max_{k_2+1\le k\le k_3}|b_k-\pi_k|\right\} \\ &= \max_{k_1\le k\le k_3}|b_k-\pi_k| . \end{aligned}$$

From (i) and last inequality, when $\ell \notin \{k_1,\dots,k_3\}$,

$$\begin{aligned}\max_{1\le k\le m} |b_k-\pi_k^*| &= \max\left\{\max_{\ell}|b_\ell-\pi_\ell^*|, \max_{k_1\le k\le k_3} |b_k-\pi_k^*|\right\} \\ &\le \max\left\{\max_{\ell}|b_\ell-\pi_\ell|, \max_{k_1\le k\le k_3} |b_k-\pi_k|\right\} \\ &= \max_{1\le k\le m} |b_k-\pi_k| .\end{aligned}$$

□

Theorem 2.17. If $p_1 \ge p_2 \ge \dots \ge p_m$ and $(\tilde{p}_k)_{k=1}^m$ is obtained by applying procedure 2.6 to $(\hat{p}_k)_{k=1}^m$ then

$$\max_{1\le k\le m} |p_k - \tilde{p}_k| \le \max_{1\le k\le m} |p_k - \hat{p}_k| .$$

Proof. If $(\hat{p}_k)_{k=1}^m$ is MNI then $(\tilde{p}_k)_{k=1}^m = (\hat{p}_k)_{k=1}^m$ and the inequality is trivially satisfied. If $(\hat{p}_k)_{k=1}^m$ is not MNI then, as noted in 2.6.A, $(\hat{p}_k)_{k=1}^m$ is monotonized after finite~~ly~~ many applications of the SMT. After each application of the SMT, use lemma 2.16, with $b_k = p_k$ . Then for any SMT output $\Pi^* = (\pi_k^*)_{k=1}^m$ ,

(i) $$\max_{1\le k\le m} |p_k - \pi_k^*| \le \max_{1\le k\le m} |p_k - \hat{p}_k| .$$

In particular, (i) is also true for the final output $\Pi^* = \tilde{P}$ . □

Example 2.18. Theorem 2.17 is false if the condition $p_1 \ge p_2 \ge \dots \ge p_m$ is omitted. Indeed, suppose that $m = 2$ and $p_1 < p_2$ and that, after the SMT is applied with the input $(\hat{p}_1,\hat{p}_2)$, we get a pair $\tilde{p}_1$, $\tilde{p}_2$ . If $p_1 < \hat{p}_1 < \hat{p}_2 < p_2$ then, as indicated in Figure 2.E,

$$|\tilde{p}_i - p_i| > |\hat{p}_i - p_i| .$$

Therefore

$$\max_{i=1,2} |\tilde{p}_i - p_i| > \max_{i=1,2} |\hat{p}_i - p_i| ,$$

which contradicts the conclusion of theorem 2.17.

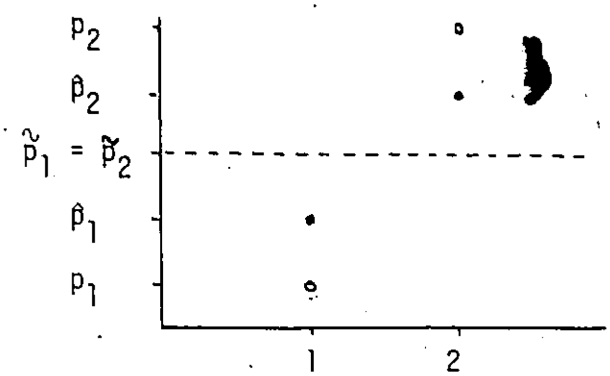



Figure 2.E.

Theorem 2.19. For $1 \leq k \leq m$ , suppose that $N_k \to \infty$. Then $\hat{p}_k$ converges to $p_k$ with probability one. Further, if $p_1 \geq p_2 \geq \ldots \geq p_m$ then $\tilde{p}_k$ also converges to $p_k$ with probability one.

Proof. Given $1 \leq k \leq m$ , $\hat{p}_k$ converges to $p_k$ by the strong law of large numbers. Now, if $p_1 \geq \ldots \geq p_m$ then, by theorem 2.17, $\tilde{p}_k$ also converges to $p_k$ with probability one. ▯

Remark 2.20. In theorem 2.18, as m is fixed, there exists a set $\Omega_o$ such that

(i) $\Omega_o$ has probability one,

(ii) for $\omega \in \Omega_o$, $(\tilde{p}_1(\omega,N_1),\ldots,\tilde{p}_m(\omega,N_m)) \to (p_1,\ldots,p_m)$ in $R^m$ when $\min\{N_1,\ldots,N_m\} \to \infty$ .

## 2.21. Generalization.

This chapter deals with sequences of data $\left((a_k,N_k)\right)_{k=1}^{m}$. All the concepts and results obtained in this chapter for such sequences are based on the properties of the operator $A(\cdot,\cdot)$ which are described in lemma 2.4.

Consider sequences $\left((b_k,d_k)\right)_{k=1}^{m}$ where $b_k$ and $d_k$ are reals, with $d_k$ positive. The operator $Av(\cdot,\cdot)$ associated to such sequences is as follows. For integers r and s such that $1 \leq r \leq s \leq m$, let

$$Av(r,s) = \frac{\sum_{r\leq k\leq s} b_k}{\sum_{r\leq k\leq s} d_k} .$$

Clearly lemma 2.4 remains true when $A(\cdot,\cdot)$ is replaced by $Av(\cdot,\cdot)$. It follows that all the concepts and results obtained in this chapter for sequences $\left((a_k/N_k)\right)_{k=1}^{m}$ can be transferred to sequences $\left((b_k,d_k)\right)_{k=1}^{m}$, with $A(\cdot,\cdot)$ replaced by $Av(\cdot,\cdot)$. E.g., the initial sequence $\hat{\Pi} = (b_k/d_k)_{k=1}^{m}$ can be monotonized by procedure 2.6 and the final result $\tilde{\Pi} = (\tilde{\pi}_k)_{k=1}^{m}$ is such that, for each k,

$$\tilde{\pi}_k = \min_{1\leq r\leq k}\left\{\max_{k\leq s\leq m} Av(r,s)\right\} .$$

# 3. Monotonization methods for histograms.

## 3.1. Introduction.

Throughout this chapter we consider the situation where $(X_i)_{i=1}^{\infty}$ is a sequence of nonnegative independent random variables with common density function f which vanishes on $(-\infty,0)$ and is nonincreasing on $[0,\infty)$; since f is real-valued at 0, it follows that f is bounded. For $n \in \mathbb{N}^+$, $(X_i)_1^n$ is a sample which can be used to estimate f nonparametrically, for instance by means of a histogram. To construct a histogram from the sample, we introduce a partition of $[0,\infty)$, namely intervals $\{[t_k^{(n)}, t_{k+1}^{(n)}): k = 1,2,\ldots\}$ where $0 = t_1^{(n)} < t_2^{(n)} < \ldots$; without risk of confusion, we refer to the sequence $t_1^{(n)}, t_2^{(n)},\ldots$ itself as "a partition of $[0,\infty)$". For each $k \in \mathbb{N}^+$, we denote the interval $[t_k^{(n)},t_{k+1}^{(n)})$ by $\Delta_k^{(n)}$, its length by $\delta_k^{(n)}$, and its indicator function by $I_k^{(n)}$. For the sample $(X_i)_{i=1}^n$, we denote the number of observations in $\Delta_k^{(n)}$ by $a_k^{(n)}$; i.e., $a_k^{(n)} = \sum_{i=1}^{n} I_k^{(n)}(X_i)$. Throughout this chapter, M is is a random integer such that $\max_{1\le i\le n} X_i \in \Delta_M^{(n)}$. The histogram estimator $\hat{f}_n$ of f related to the partition above is defined as follows:

$$\hat{f}_n(t) = \begin{cases} a_k^{(n)}/(n\delta_k^{(n)}) & , t \in \Delta_k^{(n)} , k \in \mathbb{N}^+ , \\ 0 & , t < 0 . \end{cases}$$

Note that $a_k^{(n)} = 0$ when $k > M$, hence $\hat{f}_n(t) = 0$ when $t \notin \cup_{k=1}^M \Delta_k^{(n)}$ .

As $f$ is nonincreasing on $[0,\infty)$, and vanishes on $(-\infty,0)$, it is natural to seek an estimator $\tilde{f}_n$ of $f$ which vanishes on $(-\infty,0)$ and is nonincreasing on $[0,\infty)$. The main result of this chapter is that we can construct such an estimator $\tilde{f}_n$, with rate of convergence at least as good as that of $\hat{f}_n$. This is shown in theorem 3.12. Applications of this theorem are presented in 3.13. One application is the uniform consistency of $\tilde{f}_n$, shown in 3.13.B, but it should be noted that similar results will be obtained in chapter 4 under milder conditions and by different methods. However, the methods used in chapter 4 do not yield rate of convergence results for $\tilde{f}_n$; on the other hand, the main purpose of chapter 2 and this one is to develop methods which do yield rate of convergence results for $\tilde{f}_n$.

The estimator $\tilde{f}_n$ is obtained by applying a monotonization method, described in 3.4 and 3.14, to the initial histogram $\hat{f}_n$. The monotonization procedure presented in 3.4 is useful in the proof of theorem 3.12. In 3.14, we obtain an algorithm which gives an explicit formula for the final histogram $\tilde{f}_n$. In 3.16, we give a geometric interpretation of this algorithm.

In 3.15.A, the histogram estimator $\hat{f}_n$ will be seen to be the slope of the polygonal graph joining the points $\left(t_k, F_n(t_k)\right)$, $k = 1,\ldots,M+1$; and the monotone estimator $\tilde{f}_n$ will be seen to be the slope of the polygonal graph joining the points $\left(t_{k_i+1}, F_n(t_{k_i+1})\right)$, $i = 0,\ldots,\sigma$ - where $(k_i)_0^\sigma$ is a sequence described in definition 3.6, and $F_n$ is the e.d.f. related to the sample $(X_i)_1^n$. From this we already see that the second polygonal graph is smoother than the first one.

In this chapter, a function ($\mathbb{R} \to \mathbb{R}$) is said to be MNI if it vanishes on $(-\infty,0)$ and is nonincreasing on $[0,\infty)$. In order to simplify the notation, the dependence on n will (in general) not be explicitly indicated in the remainder of this chapter: $t_k$, $\Delta_k$, $\delta_k$, $I_k$, and $a_k$ will be written instead of $t_k^{(n)}$, $\Delta_k^{(n)}$, $\delta_k^{(n)}$, $I_k^{(n)}$, and $a_k^{(n)}$, respectively.

We say that a sequence $\Pi = (\pi_k)_{k=1}^m$ and the sequence of adjacent intervals $(J_k)_{k=1}^m$ describe a histogram $\Gamma$ iff

$$\Gamma(t) = \begin{cases} \pi_k & , \text{ if } t \in J_k,\ 1 \le k \le m , \\ 0 & , \text{ elsewhere} . \end{cases}$$

For instance, the sequence of ratios $\left(a_k/(n\delta_k)\right)_{k=1}^M$ and the sequence of intervals $(\Delta_k)_{k=1}^M$ describe $\hat{f}_n$.

Now to monotonize $\hat{f}_n$, it suffices to monotonize its sequence of ratios $\hat{\Pi} = \left(a_k/(n\delta_k)\right)_{k=1}^M$. If this is done, a nonincreasing sequence is obtained which, when considered together with $(\Delta_k)_{k=1}^M$, describes a histogram that is nonincreasing on $[0,\infty)$. The two methods of monotonization we present start with the basic sequence $H = \left((a_k, n\delta_k)\right)_{k=1}^M$, the sequence of ratios $\hat{\Pi} = \left(a_k/(n\delta_k)\right)_{k=1}^M$, and the sequence of intervals $(\Delta_k)_{k=1}^M$. In view of 2.21, procedure 2.6 can be applied to the basic sequence H and the sequence of ratios $\hat{\Pi}$. Our first histogram monotonization method is developed from procedure 2.6 and is presented as a finite number of applications of a standard transformation. After each transformation, a new estimate of f is obtained - whose distance to f can be compared to that of the previous estimate. We do this in lemma 3.9. That permits us to compare, in theorem 3.10, the distance between f and the MNI estimator $\tilde{f}_n$ with the distance between f and the histogram estimator $\hat{f}_n$. That yields our main result, namely theorem 3.12.

3.2. Definition and remarks.

A. Definition. Let $H = \left((a_1, n\delta_1), \ldots, (a_M, n\delta_M)\right)$ be the data described in 3.1. For integers r and s such that $1 \leq r \leq s \leq M$, put

$$Av(r,s) = \frac{\sum_{r \leq k \leq s} a_k}{\sum_{r \leq k \leq s} n\delta_k} .$$

Clusters and maximal clusters of sequences are defined as in 2.3.A. A cluster C(r,s) of a sequence $\Pi = (\pi_1, \ldots, \pi_M)$ is said to have the average property (with respect to H) if $\pi_k = Av(r,s)$ is true for every $\pi_k \in C(r,s)$; i.e., for $\pi_k$ with $r \leq k \leq s$.

Throughout this chapter the average property will always be considered to be with respect to H.

B. Remark. Lemma 2.4 remains valid when $A(\cdot,\cdot)$ is replaced by $Av(\cdot,\cdot)$, with m = M. The above-defined average $Av(\cdot,\cdot)$ plays in this chapter a role analogous to the role played by $A(\cdot,\cdot)$ in chapter 2.

C. Remark. It will be seen in 3.15.A that

$$Av(r,s) = \frac{F_n(t_{s+1}) - F_n(t_r)}{t_{s+1} - t_r} .$$

Therefore Av(r,s) is the slope of the e.d.f. straight-line interpolation between abscissas $t_r$ and $t_{s+1}$.

## 3.3. SMT for histograms.

The standard monotonization transformation for histograms (SMTH) is used to transform a nonincreasing estimate $\Pi = (\pi_1,\ldots,\pi_M)$ of $(f(t_1),\ldots,f(t_M))$ into $\Pi^*$ which is, in the same way as in the SMT in 2.5, one step closer to being monotone. It can be formalized as follows.

The SMTH is applied in relation to the basic data $H = \left((a_1,n\delta_1),\ldots,(a_M,n\delta_M)\right)$. As defined in 3.2.A, $Av(r,s) = (\sum_{r\le k\le s} a_k)/(\sum_{r\le k\le s} n\delta_k)$. The input to the SMTH is a sequence $\Pi = (\pi_k)_{k=1}^{M}$, which is not nonincreasing, whose maximal clusters all have the average property (with respect to H) as defined in 3.2.A. The output $\Pi^* = (\pi_1^*,\ldots,\pi_M^*)$ is obtained as follows. Select a pair of contiguous maximal clusters $C(k_1,k_2)$ and $C(k_2+1,k_3)$ of $\Pi$, such that $\pi_{k_2} = Av(k_1,k_2) < A(k_2+1,k_3) = \pi_{k_2+1}$; put

$$\pi_k^* = \pi_k \qquad , \text{ if } k \notin \{k_1,\ldots,k_3\},$$

$$\pi_k^* = Av(k_1,k_3) = (\sum_{k_1\le k\le k_3} a_k)/(\sum_{k_1\le k\le k_3} n\delta_k) \quad , \text{ if } k \in \{k_1,\ldots,k_3\}.$$

The following remarks are analogous to the remarks in 2.5 and can be obtained by arguments similar to those presented there.

Remarks.

A. $\hat{\Pi} = \left(a_k/(n\delta_k)\right)_{k=1}^{M}$ is a sequence whose maximal clusters all have the average property.

B. Let $\Pi = (\pi_1,\ldots,\pi_M)$ be an input, which is not nonincreasing, whose maximal clusters all have the average property. If the SMTH is applied to $\Pi$ then the maximal clusters of the output sequence $\Pi^* = (\pi_1^*,\ldots,\pi_M^*)$ all have the average property.

C. If the input sequence contains $\nu$ maximal clusters then the corresponding output sequence contains at most $\nu - 1$ maximal clusters.

## 3.4. Monotonization procedure for histograms.

As mentioned in 3.3, this procedure starts with the initial histogram $\hat{f}_n$; more explicitely, it starts with a sequence of data $\left((a_1,n\delta_1),\ldots,(a_M,n\delta_M)\right)$, a sequence of ratios $\left(\hat{f}_n(t_k)\right)_{k=1}^M = \left(a_k/(n\delta_k)\right)_{k=1}^M$, and an associated sequence of intervals $(\Delta_k)_{k=1}^M$. After each application of the SMTH, the procedure yields a histogram. If this histogram is not nonincreasing on $\cup_{k=1}^M \Delta_k$ then the SMTH is applied again. The procedure yields, as its final output, a histogram estimate $\tilde{f}_n$ of f which is MNI in the sense specified in 3.1.

Procedure.

(1) If the sequence $\left(\hat{f}_n(t_k)\right)_{k=1}^M$ is nonincreasing then the histogram $\hat{f}_n$ is MNI. Put $\tilde{f}_n = \hat{f}_n$ and stop. Otherwise:

(2) put $\Pi = \left(\hat{f}_n(t_k)\right)_{k=1}^M$ ;

(3) apply the SMTH to the input $\Pi$, get a new sequence $\Pi^* = (\pi_k^*)_{k=1}^M$.

Put

$$\Gamma^*(t) = \begin{cases} \pi_k^* & , t \in \Delta_k, 1 \le k \le M, \\ 0 & , \text{elsewhere} . \end{cases}$$

If the sequence $\Pi^*$ is nonincreasing then the histogram $\Gamma^*$ is MNI. Put $\tilde{f}_n = \Gamma^*$ and stop. Otherwise,

(4) put $\Pi = \Pi^*$ (i.e., replace $\Pi$ by $\Pi^*$) and return to step (3).

By definition of the SMTH, the input sequence to the SMTH must be such that it is not nonincreasing and each of its maximal clusters has the average property. In the procedure described above, the initial input to the SMTH, and all subsequent inputs to it, have the required property; see comments in 3.5.

Note. As the number of maximal clusters decreases at each application of the SMTH (remark 3.3.C), the procedure must stop after finitely many iterations.

3.5. Notes on the monotonization procedure for histograms.

The following notes are analogous to 2.7.A, 2.7.D and 2.7.E respectively. They follow because $Av(\cdot,\cdot)$ and $A(\cdot,\cdot)$ share the required properties.

A. Fact. The maximal clusters of $\left(\hat{f}_n(t_k)\right)_{k=1}^{M}$ all have the average property with respect to H.

B. Fact. Let $\Pi = (\pi_k)_{k=1}^{M}$ be a sequence, which is not nonincreasing, whose maximal clusters all have the average property, and let $\Pi^*$ be the output sequence obtained by applying the SMTH to $\Pi$. Then the maximal clusters of $\Pi^*$ all have the average property.

C. Fact. Each maximal cluster of a sequence $\Pi = (\pi_k)_{k=1}^{M}$ obtained in the course of procedure 3.4 has the average property. (This follows from A and B above as Fact 2.7.E follows from 2.7.A and 2.7.D).

In particular, this is true for the final output sequence $\left(\tilde{f}_n(t_k)\right)_{k=1}^{M}$.

We use a procedure analogous to that in 2.10 in order to formulate a result analogous to proposition 2.11.

Definition 3.6. Put $k_0 = 0$; for $i \geq 1$, if $k_{i-1} < M$ let $k_i$ be the largest k for which

$$Av(k_{i-1}+1,k) = \max_{k_{i-1}+1 \leq s \leq M} Av(k_{i-1}+1,s) .$$

For some integer $\sigma$, $k_\sigma = M$ .

Proposition 2.11 is obtained from arguments depending only on lemma 2.4. By remark 3.2.B, we then have the following corresponding proposition.

Proposition 3.7. Put $\Gamma = (\gamma_1,\ldots,\gamma_M)$, with $\gamma_k = \min_{1 \leq r \leq k} \left\{ \max_{k \leq s \leq M} Av(r,s) \right\}$ for $1 \leq k \leq M$ , and let $(k_0,\ldots,k_\sigma)$ be as in 3.6. Now:

(a) The sequence $\Gamma$ is nonincreasing.

(b) Each maximal cluster of $\Gamma$ has the average property; i.e., if $C(r,s;\Gamma)$ is a maximal cluster of $\Gamma$ then $\gamma_k = Av(r,s)$ for all $\gamma_k \in C(r,s;\Gamma)$.

(c) The maximal clusters of $\Gamma$ are $C(k_{i-1}+1,k_i;\Gamma)$, $1 \leq i \leq \sigma$ .

The final nonincreasing sequence $\tilde{\Pi} = (\tilde{\pi}_k)_{k=1}^{M}$ obtained by procedure 3.4 can also be obtained by procedure 2.6 in relation to the sequence of data H, with $A(\cdot,\cdot)$ replaced by $Av(\cdot,\cdot)$. Theorem 2.13 is obtained from arguments depending on lemma 2.14 and from the nature of procedure

2.6. By remark 3.2.B, we then have the following corresponding theorem.

Theorem 3.8. The final nonincreasing output sequence $\tilde{\Pi}$ of procedure 3.4 is unique and, for $1 \le k \le M$,

$$\tilde{\pi}_k = \min_{1 \le r \le k} \left( \max_{k \le s \le M} Av(r,s) \right) .$$

It follows that procedure 3.4 yields a unique final MNI histogram $\tilde{f}_n$ given by

$$\tilde{f}_n(t) = \begin{cases} \min_{1 \le r \le k} \left\{ \max_{k \le s \le M} Av(r,s) \right\} & , t \in \Delta_k , 1 \le k \le M , \\ 0 & , \text{elsewhere} . \end{cases}$$

Lemma 3.9. Let $\Pi = (\pi_k)_{k=1}^M$ be an input sequence for the SMTH (see 3.3); $\Pi$ is a sequence which is not nonincreasing and whose maximal clusters all have the average property. Let $\Pi^* = (\pi_k^*)_{k=1}^M$ be the output sequence obtained when the SMTH is applied to $\Pi$. Let $\Gamma$ be the histogram described (see 3.1) by $\Pi$ and $(\Delta_k)_{k=1}^M$ and let $\Gamma^*$ be the histogram described by $\Pi^*$ and $(\Delta_k)_{k=1}^M$. Now:

(a) If h is nonincreasing on $[0,\infty)$ then

$$\sup_{0 \le t < \infty} |h(t) - \Gamma^*(t)| \le \sup_{0 \le t < \infty} |h(t) - \Gamma(t)| .$$

(b) For any $T \in [t_M,\infty)$, if h is nonincreasing on $[0,T]$ then

$$\sup_{0 \le t \le T} |h(t) - \Gamma^*(t)| \le \sup_{0 \le t \le T} |h(t) - \Gamma(t)| .$$

Proof. The steps of this proof are basically the same as those in the proof of lemma 2.16. By assumption,

$$\Gamma(t) = \begin{cases} \pi_k & , t \in \Delta_k, 1 \leq k \leq M , \\ 0 & , \text{elsewhere}, \end{cases}$$

and

$$\Gamma^*(t) = \begin{cases} \pi_k^* & , t \in \Delta_k, 1 \leq k \leq M , \\ 0 & , \text{elsewhere} . \end{cases}$$

Suppose that the maximal clusters $C(k_1,k_2;\Pi)$ and $C(k_2+1,k_3;\Pi)$ are such that $Av(k_1,k_2) < Av(k_2+1,k_3)$ , with $1 \leq k_1 \leq k_2 < k_2+1 \leq k_3 \leq M$, and that these two maximal clusters are chosen for concatenation. Then $\Pi^*$ is obtained from the SMTH by the following transformation:

(i)
$$\pi_k^* = \pi_k \qquad , \text{if } k \notin \{k_1,\ldots,k_3\} ,$$
$$\pi_k^* = Av(k_1,k_3) \qquad , \text{if } k \in \{k_1,\ldots,k_3\} .$$

It follows that for each $t \in \Delta_\ell$, with $\ell \notin \{k_1,\ldots,k_3\}$, $\Gamma^*(t) = \Gamma(t)$ .

Consider $\bigcup_{k_1 \leq k \leq k_3} \Delta_k = J_1 \cup J_2$ , where

$$J_1 = \bigcup_{k_1 \leq k \leq k_2} \Delta_k = [t_{k_1}, t_{k_2+1})$$

and

$$J_2 = \bigcup_{k_2+1 \leq k \leq k_3} \Delta_k = [t_{k_2+1}, t_{k_3+1}) .$$

In view of the above,

(ii)
$$\sup_{t \notin J_1 \cup J_2} |h(t) - \Gamma^*(t)| = \sup_{t \notin J_1 \cup J_2} |h(t) - \Gamma(t)| .$$

Next, consider $t \in J_1 \cup J_2$. By (i), $\Gamma^*(t) = Av(k_1,k_3)$ on $J_1 \cup J_2$ and, by hypothesis,

$$\Gamma(t) = \begin{cases} Av(k_1,k_2) & , \text{ if } t \in J_1 , \\ Av(k_2+1,k_3) & , \text{ if } t \in J_2 . \end{cases}$$

By assumption $Av(k_1,k_2) < Av(k_2+1,k_3)$; hence, by lemma 2.4.(v), with $A(\cdot,\cdot)$ replaced by $Av(\cdot,\cdot)$,

(iii) $\quad Av(k_1,k_2) < Av(k_1,k_3) < Av(k_2+1,k_3)$ .

Apply proposition 2.15 with $I_1 = J_1$, $I_2 = J_2$, $a = Av(k_1,k_2)$, $b = Av(k_1,k_3)$ and $c = Av(k_2+1,k_3)$. Note that $J_1$ and $J_2$ are nonempty intervals and, by (iii), $a < b < c$. This yields

(iv)
$$\begin{aligned} \sup_{t\in J_1\cup J_2} |h(t) - \Gamma^*(t)| &= \sup_{t\in J_1\cup J_2} |h(t) - Av(k_1,k_3)| \\ &< \max\left\{\sup_{t\in J_1} |h(t) - Av(k_1,k_2)|,\ \sup_{t\in J_2} |h(t) - Av(k_2+1,k_3)|\right\} \\ &= \max\left\{\sup_{t\in J_1} |h(t) - \Gamma(t)|,\ \sup_{t\in J_2} |h(t) - \Gamma(t)|\right\} \\ &= \sup_{t\in J_1\cup J_2} |h(t) - \Gamma(t)| . \end{aligned}$$

Now, from (ii) and (iv), taking the supremum over all $t \in [0,\infty)$,

(v)
$$\begin{aligned} \sup_t |h(t) - \Gamma^*(t)| &= \max\left\{\sup_{t\notin J_1\cup J_2} |h(t) - \Gamma^*(t)|,\ \sup_{t\in J_1\cup J_2} |h(t) - \Gamma^*(t)|\right\} \\ &\leq \max\left\{\sup_{t\notin J_1\cup J_2} |h(t) - \Gamma(t)|,\ \sup_{t\in J_1\cup J_2} |h(t) - \Gamma(t)|\right\} \\ &= \sup_t |h(t) - \Gamma(t)| . \end{aligned}$$

This proves statement (a).

Next we consider $T \in [t_M,\infty)$. Note that (ii) remains true when the supremum is taken over all $t \in [0,T] \setminus (J_1 \cup J_2)$. (The set $[0,T] \setminus (J_1 \cup J_2)$ may be empty, but then both terms in (ii) are $-\infty$ ).

To show that (iv) remains true with t restricted to [0,T], we need to show that $I_1 = J_1 \cap [0,T] \neq \phi$ and that $I_2 = J_2 \cap [0,T] \neq \phi$ . If either $I_1$ or $I_2$ is empty, proposition 2.15 cannot be applied. Since $k_2 < k_2+1 \leq k_3 \leq M$ , we see that $t_{k_2} < t_{k_2+1} \leq t_{k_3} \leq t_M \leq T$ . Thus

$$J_1 \cap [0,T] = [t_{k_1},t_{k_2+1}) \cap [0,T] = [t_{k_1},t_{k_2+1}) = J_1$$

and

$$J_2 \cap [0,T] = \begin{cases} J_2 & , \text{ if } T \geq t_{k_3+1} , \\ [t_{k_2+1},t_{k_3}) \cup [t_{k_3},T] & , \text{ if } t_{k_3} \leq T < t_{k_3+1} . \end{cases}$$

Since $t_{k_1} \in J_1 \cap [0,T]$ and $t_{k_3} \in J_2 \cap [0,T]$, we have that $J_1 \cap [0,T]$ and $J_2 \cap [0,T]$ are both nonempty intervals. Therefore (iv) remains true with t restricted to [0,T]. Now, from (ii) and (iv), with t restricted to [0,T], statement (b) follows by (v), with t restricted to [0,T].

□

<u>Theorem 3.10</u>. Let $\hat{f}_n$ be the histogram estimator of f and let $\tilde{f}_n$ be obtained by procedure 3.4. Now:

(a) $\sup_{0 \le t < \infty} |f(t) - \tilde{f}_n(t)| \le \sup_{0 \le t < \infty} |f(t) - \hat{f}_n(t)|$ .

(b) For any $T \in [t_M, \infty)$,

$$\sup_{0 \le t \le T} |f(t) - \tilde{f}_n(t)| \le \sup_{0 \le t \le T} |f(t) - \hat{f}_n(t)| .$$

<u>Proof</u>. If $\hat{f}_n$ is MNI then $\tilde{f}_n = \hat{f}_n$ and the conclusion of the theorem is trivially true. If $\hat{f}_n$ is not MNI, apply procedure 3.4. After the first application of the SMTH an output sequence $\Pi^* = (\pi_k^*)_{k=1}^M$ is obtained which, together with $(\Delta_k)_{k=1}^M$ , describes a histogram estimate $\Gamma^*$ of f. As f is assumed nonincreasing, one can apply lemma 3.9.(a), with h = f, to obtain

(i) $$\sup_{0 \le t < \infty} |f(t) - \Gamma^*(t)| \le \sup_{0 \le t < \infty} |f(t) - \hat{f}_n(t)| .$$

By using lemma 3.9.(a), with h = f, after each application of the SMTH we obtain (i) for any histogram estimate $\Gamma^*$ obtained in the course of procedure 3.4. In particular, (i) is also true for the final output histogram $\tilde{f}_n$. This shows statement (a).

Statement (b) follows by considering $T \in [t_M, \infty)$ and the same argument as above, but using lemma 3.9.(b) instead of 3.9.(a). □

## 3.11. Remarks on theorem 3.10.

A. The conclusion of theorem 3.10 may be false if the condition that $f$ be nonincreasing on $[0,\infty)$ is omitted. This is shown by an example such as 2.18, adapted for histograms.

B. If $T \in [0,t_M)$ then the conclusion of theorem 3.10.(b) may be false. We illustrate this in Figure 3.A where we suppose that all the observations of a sample of size $n$ lie in the two intervals $[t_1,t_2)$ and $[t_2,t_3)$ of the partition; i.e., $M = 2$. Let $f$, $\hat{f}_n$ and $\tilde{f}_n$ be as in Figure 3.A. If $t_1 \leq T < t_2$ then

$$\sup_{0\leq t\leq T} |f(t) - \tilde{f}_n(t)| > \sup_{0\leq t\leq T} |f(t) - \hat{f}_n(t)| ,$$

which contradicts the conclusions of theorem 3.10.(b).

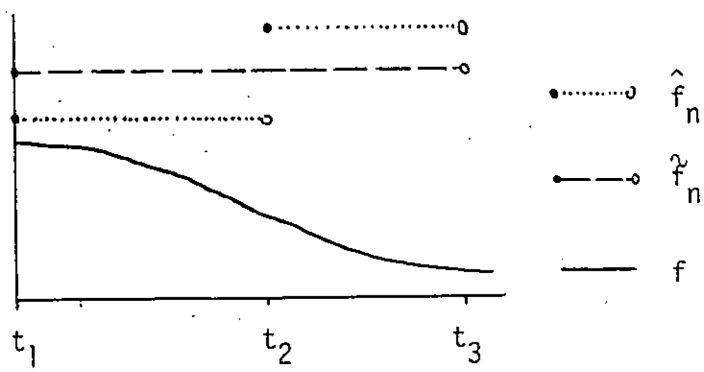



Figure 3.A.

Theorem 3.12. Suppose that $f$ vanishes outside an interval $J$, where $J$ is $[0,\infty)$ or $[0,T]$ with $T < \infty$, and let $||\cdot||$ be the supnorm over $J$. If $f$, the sequence of partitions $((t_k^{(n)})_{k=1}: n = 1,2,\ldots)$ , and

$\chi(n)$ are such that

$$P[\lim_{n\to\infty} \chi(n)||f - \hat{f}_n|| = 0] = 1$$

then

$$P[\lim_{n\to\infty} \chi(n)||f - \tilde{f}_n|| = 0] = 1 .$$

Proof. If $J = [0,\infty)$, the conclusion of the theorem is obtained by theorem 3.10.(a). Consider the case where $J = [0,T]$, with $T < \infty$, and f vanishes outside J. Then, with probability one, $\max_{1\leq i\leq n} X_i \leq T$ for all n. Now, since M is such that $t_M^{(n)} \leq \max_{1\leq i\leq n} X_i$, is follows that, for all n, $T \geq t_M^{(n)}$ a.s.. Thus when $J = [0,T]$, by theorem 3.10.(b), for each n,

$$\chi(n)||f - \tilde{f}_n|| \leq \chi(n)||f - \hat{f}_n|| \quad \text{a.s. .}$$

Therefore the conclusion of the theorem follows also in the case of bounded intervals J.

3.13. Applications of theorem 3.12.

A. The theorems on the rate of convergence of $\hat{f}_n$ that we are aware of require that f be continuous on $\mathbb{R}$. In our case, f is nonincreasing on $[0,\infty)$ and vanishes on $(-\infty,0)$, and is therefore discontinuous at 0. However, a minor modification of a proof given by Révész (1972, p. 101) yields the result stated below. The modification consists simply in using partitions of $[0,\infty)$ instead of partitions of $(-\infty,\infty)$ as in Révész's argument. The proof remains entirely the same except that, in $\sup|f(t) - \hat{f}_n(t)|$, the supremum is taken over $t \in [0,\infty)$ instead of ove

$t \in (-\infty,\infty)$. The modified Révész result is that

$$P\left[\frac{n^{1/3}}{\log n} \sup_{0 \le t < \infty} |f(t) - \hat{f}_n(t)| \to 0\right] = 1$$

when:

(a) $f$ vanishes on $(-\infty,0)$ and $\hat{f}_n$ is defined, as in 3.1, relative to a partition of $[0,\infty)$;

(b) $f$ has a bounded derivative on $(0,\infty)$ and is differentiable from the right at 0;

(c) there exists a $\delta > 0$ such that

$$\lim_{x \to \infty} x^{1+\delta} f(x) = 0 ;$$

(d) for each $n$ and all $k$, $\delta_k^{(n)} = cn^{-1/3}$.

Applying this to our case, where it is (throughout this chapter) assumed that $f$ vanishes on $(-\infty,0)$ and is nonincreasing on $[0,\infty)$, we obtain the following consequence of theorem 3.12.

<u>THEOREM</u>: If conditions (b), (c) and (d) above are satisfied then

$$P\left[\frac{n^{1/3}}{\log n} \sup_{0 \le t < \infty} |f(t) - \tilde{f}_n(t)| \to 0\right] = 1 .$$

B. There are also consistency results for the histogram estimator $\hat{f}_n$ which, by theorem 3.10, give consistency results for the MNI histogram $\tilde{f}_n$ obtained by procedure 3.4. By the same modification as described in A above we can adapt another Révész result (1972, p. 97) to conclude that

$$P\left[\sup_{0\leq t<\infty} |f(t) - \hat{f}_n(t)| \to 0\right] = 1$$

when:

(a) f vanishes on $(-\infty,0)$ and $\hat{f}_n$ is defined, as in 3.1, relative to a partition of $[0,\infty)$;

(b) f is uniformly continuous on $[0,\infty)$;

(c) there exists a $\delta > 0$ such that

$$\int_0^\infty x^\delta f(x)\,dx < \infty ;$$

(d) for some $r > 1$, for each n and all k,

$$\delta_k^{(n)} \geq \frac{(\log n)^r}{n} .$$

(e) $\sup_k \delta_k^{(n)} \to 0$ as $n \to \infty$ .

Applying this to our case, where f vanishes on $(-\infty,0)$ and is nonincreasing on $[0,\infty)$, we obtain the following consequence of theorem 3.12.

<u>THEOREM</u>: If conditions (b), (c) and (d) above are satisfied then

$$P\left[\sup_{0\leq t<\infty} |f(t) - \tilde{f}_n(t)| \to 0\right] = 1 .$$

Note that, in our case, where f is nonincreasing on $[0,\infty)$, condition (b) is equivalent to requiring that f be continuous on $[0,\infty)$.

In chapter 4, we obtain the uniform consistency of $\tilde{f}_n$ with milder conditions. I.e., the condition $\int_0^\infty x^\delta f(x)dx$ will be omitted and the conditions on $\delta_k^{(n)}$ will be restricted to those $\delta_k^{(n)}$ for which $\Delta_k^{(n)}$ is in a neighbourhood of a point of discontinuity of f. In itself the condition that f is nonincreasing on $[0,\infty)$ is a strong property. From this it suffices that its MNI estimator $\tilde{f}_n$ is pointwise consistent on a closed interval for $\tilde{f}_n$ to be uniformly consistent (see 4.5). This yields a direct argument for the consistency of $\tilde{f}_n$, with no consideration of $\hat{f}_n$, and where the only conditions are those which ensure that $\tilde{f}_n$ is pointwise convergent on a closed interval . However, the arguments used in chapter 4 yield no results on rates of convergence of $\tilde{f}_n$.

## 3.14. Monotonization algorithm for histograms.

In this section we give an algorithm for constructing the monotone estimator $\tilde{f}_n$ of f. The algorithm starts with the initial data $H = \left((a_k, n\delta_k)\right)_{k=1}^{M}$ and a sequence of intervals $(\Delta_k)_{k=1}^{M}$. The algorithm monotonizes directly the sequence of ratios $\hat{\Pi} = (\hat{\pi}_k)_{k=1}^{M}$, where $\hat{\pi}_k = a_k/(n\delta_k)$, into a nonincreasing sequence $\tilde{\Pi} = (\tilde{\pi}_k)_{k=1}^{M}$. It yields the MNI estimator $\tilde{f}_n$ as the histogram described by $\tilde{\Pi}$ and $(\Delta_k)_{k=1}^{M}$.

Algorithm.

(1) If $\hat{\Pi}$ is nonincreasing, put $\tilde{f}_n = \hat{f}_n$ and stop. Otherwise, put $k_0 = 0$ and $i = 1$ and go to step (2).

(2) Given $k_{i-1} < M$, let $k_i$ be the largest integer $\ell$ such that

$$Av(k_{i-1}+1,\ell) = \max_{k_{i-1}+1 \le s \le M} Av(k_{i-1}+1,s) .$$

For $k_{i-1}+1 \le k \le k_i$, put $\tilde{\pi}_k = Av(k_{i-1}+1,k_i)$ . Go to step (3).

(3) If $k_i = M$, go to step (4). Otherwise, increase i by 1 and return to step (2).

(4) Put

$$\tilde{f}_n(t) = \begin{cases} \tilde{\pi}_k & , t \in \Delta_k; 1 \le k \le M, \\ 0 & , \text{elsewhere} . \end{cases}$$

Note that since M is finite, there is an integer $\sigma$ such that $k_\sigma = M$. The sequence $(k_i)_{i=0}^{\sigma}$ obtained here is the same as the one obtained by the procedure in definition 3.6. It then follows from proposition 3.7 and theorem 3.8 that the final output sequence $\tilde{\Pi}$ of this algorithm is the same as the one obtained by procedure 3.4. Thus procedure 3.4 and algorithm 3.14 both yield the same final histogram output $\tilde{f}_n$.

3.15. Relation with the e.d.f..

Recall that in this thesis, $F_n(x) = n^{-1} \sum_{i=1}^{n} I_{(-\infty,x)}(X_i)$ is the left-continuous form of the e.d.f..

A. Note that $\sum_{k=r}^{s} (\sum_{i=1}^{n} I_k(X_i))$ is the number of observations, in the sample $(X_i)_{i=1}^{n}$, with values in $\cup_{k=r}^{s} \Delta_k = [t_r, t_{s+1})$. Therefore

$$\text{(i)} \quad Av(r,s) = \frac{\sum_{r \le k \le s} a_k}{\sum_{r \le k \le s} n\delta_k} = \frac{\sum_{r \le k \le s} (\sum_{i=1}^{n} I_k(X_i))}{n} \cdot \frac{1}{\sum_{r \le k \le s} (t_{k+1} - t_k)}$$

$$= \frac{F_n(t_{s+1}) - F_n(t_r)}{t_{s+1} - t_r} .$$

By definition of $Av(\cdot,\cdot)$,

$$\hat{f}_n(t) = \begin{cases} Av(k,k) & , \quad t \in \Delta_k, \ 1 \le k \le M, \\ 0 & , \quad \text{elsewhere} . \end{cases}$$

By (i), this can be rewritten as follows:

$$\text{(ii)} \quad \hat{f}_n(t) = \begin{cases} \dfrac{F_n(t_{k+1}) - F_n(t_k)}{t_{k+1} - t_k} & , \ t \in \Delta_k, \ 1 \le k \le M , \\ 0 & , \ \text{elsewhere} . \end{cases}$$

By (ii), we note that the histogram estimator $\hat{f}_n$ can be seen as the slope of the e.d.f. straight-line interpolation between $t_k$ and $t_{k+1}$ for each k. By (i) we note that Av(r,s) is the slope of the e.d.f. straight-line interpolation between $t_r$ and $t_{s+1}$. Hence, by algorithm 3.14, $\tilde{f}_n$ can be seen as the slope of the e.d.f. straight-line interpolation between $t_{k_{i-1}+1}$ and $t_{k_i+1}$, for each i, where $(k_i)_{i=0}^{\sigma}$ is as in definition 3.6.

B. Recall that M is such that $\max\limits_{1 \le i \le n} X_i \in [t_M, t_{M+1})$ . Then, for $t \ge t_{M+1}$ , $F_n(t) = 1$ . It follows that, for $r \le M$ and $M+1 < s$ ,

$$\frac{F_n(t_s) - F_n(t_r)}{t_s - t_r} = \frac{1 - F_n(t_r)}{t_s - t_r} < \frac{1 - F_n(t_r)}{t_{M+1} - t_r} = \frac{F_n(t_{M+1}) - F_n(t_r)}{t_{M+1} - t_r} .$$

Consequently, for $r \le k \le M$,

$$\text{(i)} \quad \max_{k+1 \le s < \infty} \frac{F_n(t_s) - F_n(t_r)}{t_s - t_r} = \max_{k+1 \le s \le M+1} \frac{F_n(t_s) - F_n(t_r)}{t_s - t_r} .$$

But then, for $r \le k \le M$, from (i) in A above,

(ii) $$\max_{k\le s\le M} Av(r,s) = \max_{k\le s\le M} \frac{F_n(t_{s+1}) - F_n(t_r)}{t_{s+1} - t_r}$$

$$= \max_{k+1\le s\le M+1} \frac{F_n(t_s) - F_n(t_r)}{t_s - t_r} = \max_{s\ge k+1} \frac{F_n(t_s) - F_n(t_r)}{t_s - t_r}$$

Further, recall from proposition 3.7 that the sequence $\tilde{\Pi} = (\tilde{\pi}_k)_{k=1}^{M}$ of 3.14 can be obtained as follows: for each k,

$\tilde{\pi}_k = \min_{1\le r\le k}\left\{\max_{k\le s\le M} Av(r,s)\right\}$. Thus $\tilde{f}_n$ can also be expressed as follows:

$$\tilde{f}_n(t) = \begin{cases} \min_{1\le r\le k}\left\{\max_{k\le s\le M} Av(r,s)\right\} & , t \in \Delta_k,\ 1 \le k \le M \\ 0 & , \text{elsewhere} . \end{cases}$$

Therefore

(iii) $$\tilde{f}_n(t) = \begin{cases} \min_{1\le r\le k}\ \max_{s\ge k+1} \dfrac{F_n(t_s) - F_n(t_r)}{t_s - t_r} & , t \in \Delta_k,\ 1 \le k \le M, \\ 0 & , \text{elsewhere} . \end{cases}$$

Consider $k \ge M+1$; recall that, for such k, $F_n(t_k) = 1$. Then, for any $r \le k$,

$$\max_{s\ge k+1} \frac{F_n(t_s) - F_n(t_r)}{t_s - t_r} = \max_{s\ge k+1} \frac{1 - F_n(t_r)}{t_s - t_r}$$

$$= \frac{1 - F_n(t_r)}{t_{k+1} - t_r} ,$$

hence

$$0 \le \min_{1\le r\le k}\ \max_{s\ge k+1} \frac{F_n(t_s) - F_n(t_r)}{t_s - t_r} = \min_{1\le r\le k} \frac{1 - F_n(t_r)}{t_{k+1} - t_r}$$

$$\le \frac{1 - F_n(t_k)}{t_{k+1} - t_k} = \frac{1 - 1}{t_{k+1} - t_k} = 0 .$$

From this and (iii), we conclude that

(iv) $$\tilde{f}_n(t) = \begin{cases} \min\limits_{1 \le r \le k} \ \max\limits_{s \ge k+1} \dfrac{F_n(t_s) - F_n(t_r)}{t_s - t_r} & , t \in \Delta_k, k \ge 1 \\ 0 & , t < 0 . \end{cases}$$

This form of $\tilde{f}_n$ will be particularly useful in chapter 4.

3.16. Geometric interpretation of algorithm 3.14.

Assume $\hat{f}_n$ is not monotone nonincreasing on $[0,\infty)$. By 3.15.A.(i), $Av(r,s)$ can be replaced by $\dfrac{F_n(t_{s+1}) - F_n(t_r)}{t_{s+1} - t_r}$ and, by 3.15.B.(ii), $\max\limits_{k \le s \le M} Av(r,s)$ can be replaced by $\max\limits_{s \ge k+1} \dfrac{F_n(t_s) - F_n(t_r)}{t_s - t_r}$. If we do this, algorithm 3.14 can be rewritten as follows.

Algorithm

Let $\tilde{f}_n$ be 0 outside $[t_1, t_{M+1})$.

(1) Put $k_o = 0$ and $i = 1$ and go to step (2).

(2) Given $k_{i-1} < M$, let $k_i$ be the largest integer $\ell$ such that

$$\frac{F_n(t_{\ell+1}) - F_n(t_{k_{i-1}+1})}{t_{\ell+1} - t_{k_{i-1}+1}} = \max_{s \ge k_{i-1}+2} \frac{F_n(t_s) - F_n(t_{k_{i-1}+1})}{t_s - t_{k_{i-1}+1}} .$$

Put

$$\tilde{f}_n(t) = \frac{F_n(t_{k_i+1}) - F_n(t_{k_{i-1}+1})}{t_{k_i+1} - t_{k_{i-1}+1}} \text{ when } t \in \bigcup_{k_{i-1}+1 \le k \le k_i} \Delta_k$$

Go to step (3).

(3) If $k_i = M$, stop. Otherwise, increase i by 1 and return to step (2).

This method ~~of obtaining~~ $\tilde{f}_n$ is quite simple geometrically. In words, it consists of the following steps.

Let the graph $\mathcal{G}$ consist of the following points $\left((t_j, F_n(t_j))\right)_{j=1}^{M+1}$, and let $\tilde{f}_n$ be 0 outside $[t_1, t_{M+1})$. Start with $\ell_0 = 1$. With $i \geq 1$ and $\ell_{i-1} \leq M$, consider all possible segments joining $(t_{\ell_{i-1}}, F_n(t_{\ell_{i-1}}))$ to any other point of $\mathcal{G}$ to the right of this one; among these segments, choose the one with the largest slope. That yields $\ell_i$ because, by (2) above [with $\ell_i = k_i+1$], this segment is the one joining the point $(t_{\ell_{i-1}}, F_n(t_{\ell_{i-1}}))$ to the point $(t_{\ell_i}, F_n(t_{\ell_i}))$. In this manner, we obtain $\ell_0, \ldots, \ell_\sigma$ where σ is defined by $t_{\ell_\sigma} = t_{M+1}$. Finally, construct the polygonal graph joining the points $(t_{\ell_i}, F_n(t_{\ell_i}))_{i=0}^{\sigma}$; on each interval $[t_{\ell_i}, t_{\ell_{i+1}})$, from step (2) above, $\tilde{f}_n$ is the slope of the polygonal graph on that interval.

## 3.17. Illustration of algorithm 3.14 and the min-max formula.

We first illustrate algorithm 3.14 by illustrating its geometric interpretation. Then we will illustrate, with the same example, the geometric interpretation of the min-max formula. Illustrative graphs are given in Figures 3.B, C, D and E, at the end of this chapter.

Let $F_n$ be as in Figure 3.B.(i) and let $0 = t_1 < t_2 < \dots$ be a partition of $[0,\infty)$. In this example, $M = 3$. We obtain a graph consisting of the points $\left((t_k, F_n(t_k))\right)_{k=1}^{4}$; this is sketched in Figure 3.B.(ii). By joining consecutive points of the graph in Figure 3.B.(ii) we obtain the polygonal graph with slope $\hat{f}_n$ in each interval $[t_k, t_{k+1})$; this is sketched in Figure 3.B.(iii). The histogram $\hat{f}_n$ is shown in Figure 3.B.(iv).

Figure 3.C shows how to obtain the polygonal graph with slope $\tilde{f}_n$ in each interval $[t_{k_{i-1}+1}, t_{k_i+1})$. This polygonal graph is sketched in Figure 3.C.(iii). In Figure 3.C.(i) we show all possible segments joining $(t_1, F_n(t_1))$ to another point of the graph. The segment with largest slope is the one joining $(t_1, F_n(t_1))$ to $(t_2, F_n(t_2))$. By 3.16,

$\tilde{f}_n(t) = \dfrac{F_n(t_2) - F_n(t_1)}{t_2 - t_1}$ on $[t_1, t_2)$. Similarly, in Figure 3.C.(ii) we show all possible segments joining $(t_2, F_n(t_2))$ to another point of the graph of Figure 3.B.(ii) to the right of $(t_2, F_n(t_2))$. The segment with the largest slope is the one joining $(t_2, F_n(t_2))$ to $(t_4, F_n(t_4))$. By 3.16, $\tilde{f}_n(t) = \dfrac{F_n(t_4) - F_n(t_2)}{t_4 - t_2}$ on $[t_2, t_4)$. The resulting $\tilde{f}_n$ is shown in Figure 3.C.(iv).

Now, for a geometric interpretation of the min-max formula, recall that $Av(\cdot,\cdot)$ is the slope of a segment joining two distinct points in the sequence $\left((t_j, F_n(t_j))\right)_{j=1}^{M+1}$. For each k, the min-max formula yields one such slope. For any fixed k, the min-max formula is

(i) $$\tilde{f}_n(t_k) = \min_{1 \le r \le k} \left\{ \max_{k \le s \le M} Av(r,s) \right\} .$$

It is not possible to know, from a glance at the graph of the points $\left((t_j,F_n(t_j))\right)_{j=1}^{M+1}$, which slope the min-max formula will yield.

Consider the e.d.f. of Figure 3.B.(i), with the partition $0 = t_1 < t_2 < \dots$ and $M = 3$, and the graph of Figure 3.B.(ii) obtained from the points $\left((t_j,F_n(t_j))\right)_{j=1}^{4}$.

Put $k = 1$ and consider the min-max formula (i). It yields

$$\tilde{f}_n(t_1) = \max_{1\le s\le 3} Av(1,s) = \max_{1\le s\le 3} \frac{F_n(t_{s+1}) - F_n(t_1)}{t_{s+1} - t_1} .$$

As in the case of algorithm 3.14, the segment with slope $\tilde{f}_n(t_1)$ is the one joining the point $(t_1,F_n(t_1))$ to the point of the graph of Figure 3.B.(ii) which yields the largest slope. This is represented in Figure 3.D.(i), which is identical to Figure 3.C.(i).

Now we obtain $\tilde{f}_n(t_2)$ from the min-max formula, with $k = 2$:

$$\tilde{f}_n(t_2) = \min_{1\le r\le 2}\left\{\max_{2\le s\le 3} Av(r,s)\right\}$$
$$= \min\left\{\max_{2\le s\le 3} Av(1,s),\ \max_{2\le s\le 3} Av(2,s)\right\} .$$

Figure 3.D.(ii) shows that, in this example,

$$\max_{2\le s\le 3} Av(1,s) = \max_{2\le s\le 3} \frac{F_n(t_{s+1}) - F_n(t_1)}{t_{s+1} - t_1}$$
$$= \frac{F_n(t_3) - F_n(t_1)}{t_3 - t_1} .$$

Figure 3.D.(iii) shows that

$$\max_{2\le s\le 3} Av(2,s) = \max_{2\le s\le 3} \frac{F_n(t_{s+1}) - F_n(t_2)}{t_{s+1} - t_2}$$
$$= \frac{F_n(t_4) - F_n(t_2)}{t_4 - t_2}.$$

Figure 3.D.(iv) shows that

$$\frac{F_n(t_4) - F_n(t_2)}{t_4 - t_2} < \frac{F_n(t_3) - F_n(t_1)}{t_3 - t_1} .$$

It follows that

$$\hat{f}_n(t_2) = \min_{1\le r\le 2}\left\{\max_{2\le s\le 3} Av(r,s)\right\} = \frac{F_n(t_4) - F_n(t_2)}{t_4 - t_2} .$$

Now we obtain $\hat{f}_n(t_3)$ from the min-max formula, with $k = 3$:

$$\hat{f}_n(t_3) = \min_{1\le r\le 3} Av(r,3) = \min_{1\le r\le 3} \frac{F_n(t_4) - F_n(t_r)}{t_4 - t_r} .$$

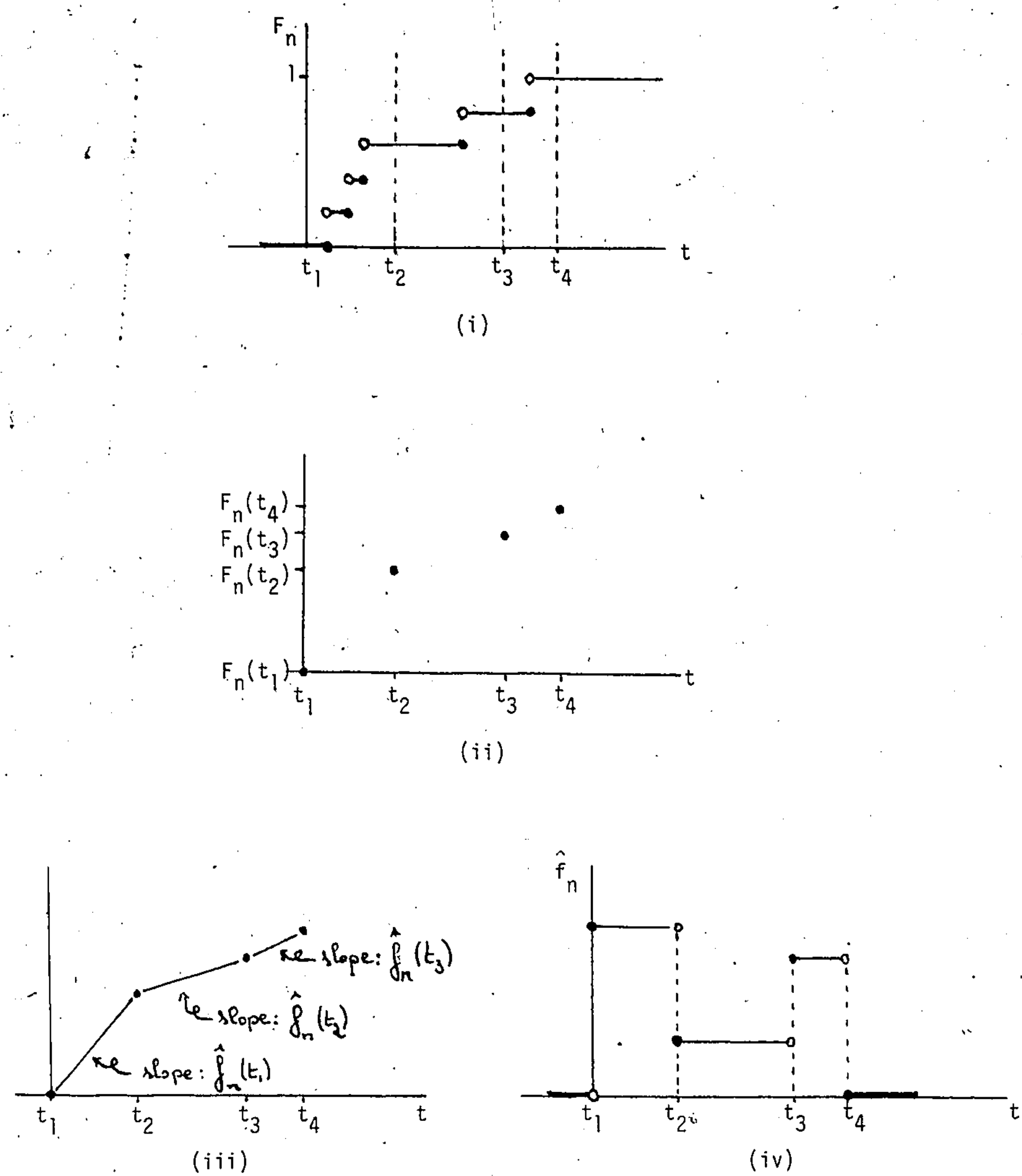
But a glance at Figure 3.D. - for example, 3.D.(iv) - shows that

$$\hat{f}_n(t_3) = \frac{F_n(t_4) - F_n(t_2)}{t_4 - t_2} = \hat{f}_n(t_2) .$$

Finally the polygonal graph with slopes given by the min-max formula is sketched in Figure 3.D.(v) and is indeed identical to the polygonal graph obtained in the illustration of algorithm 3.14 sketched in Figure 3.C.(iii).

It is seen that algorithm 3.14 is geometrically more appealing than the min-max formula.

In conclusion, $\hat{f}_n$ and $\tilde{f}_n$ are shown together, for ease of comparison, in Figure 3.E.



F n
1
t
t 1
t 2
t 3
t 4
(i)
F n(t 4)
F n(t 3)
F n(t 2)
F n(t 1)
t
t 1
t 2
t 3
t 4
(ii)
slope: f̂ n(t 3)
slope: f̂ n(t 2)
slope: f̂ n(t 1)
t 1
t 2
t 3
t 4
t
(iii)
f̂ n
t 1
t 2
t 3
t 4
t
(iv)


Figure 3.B.

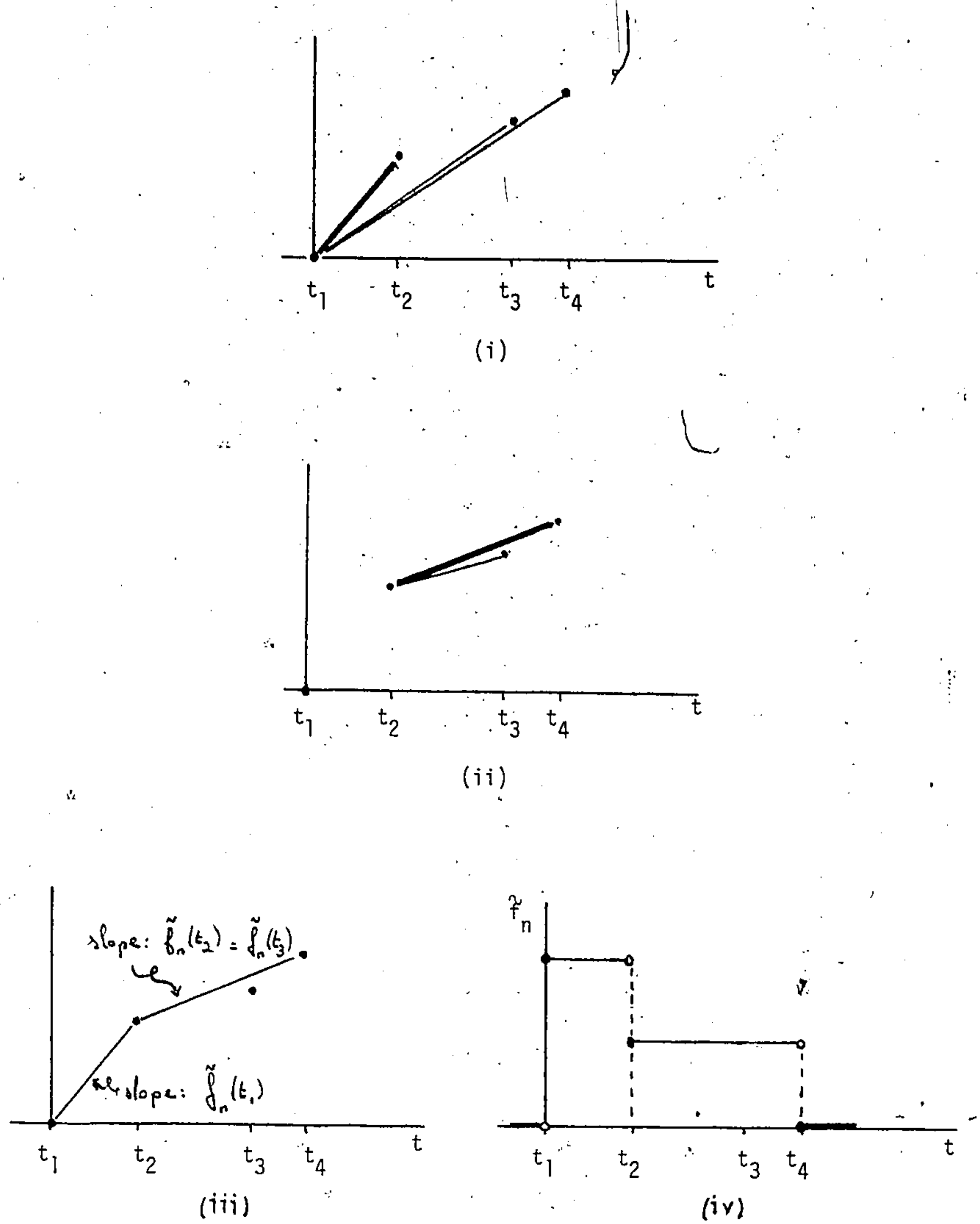



Figure 3.C.

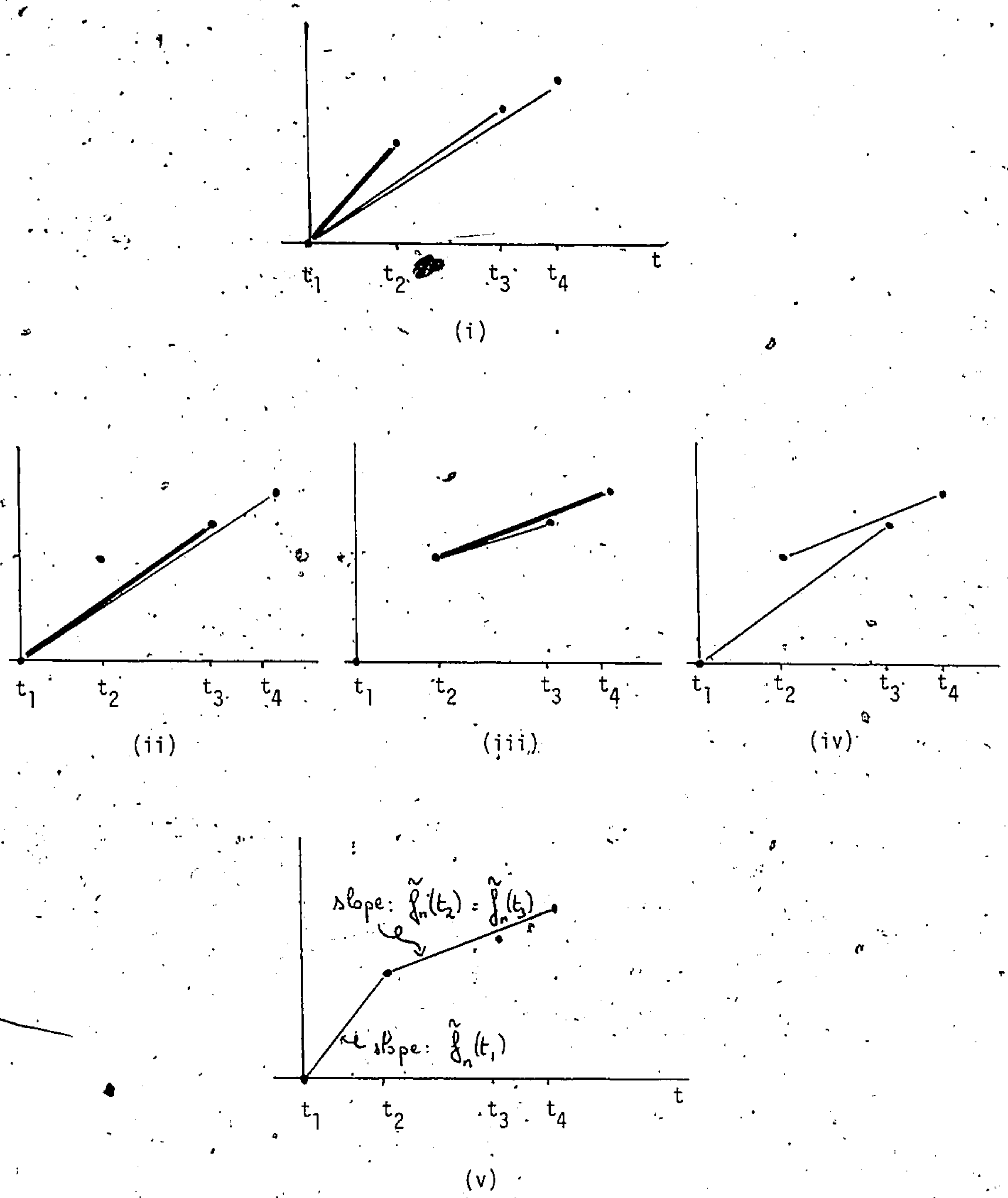



Figure 3.D.

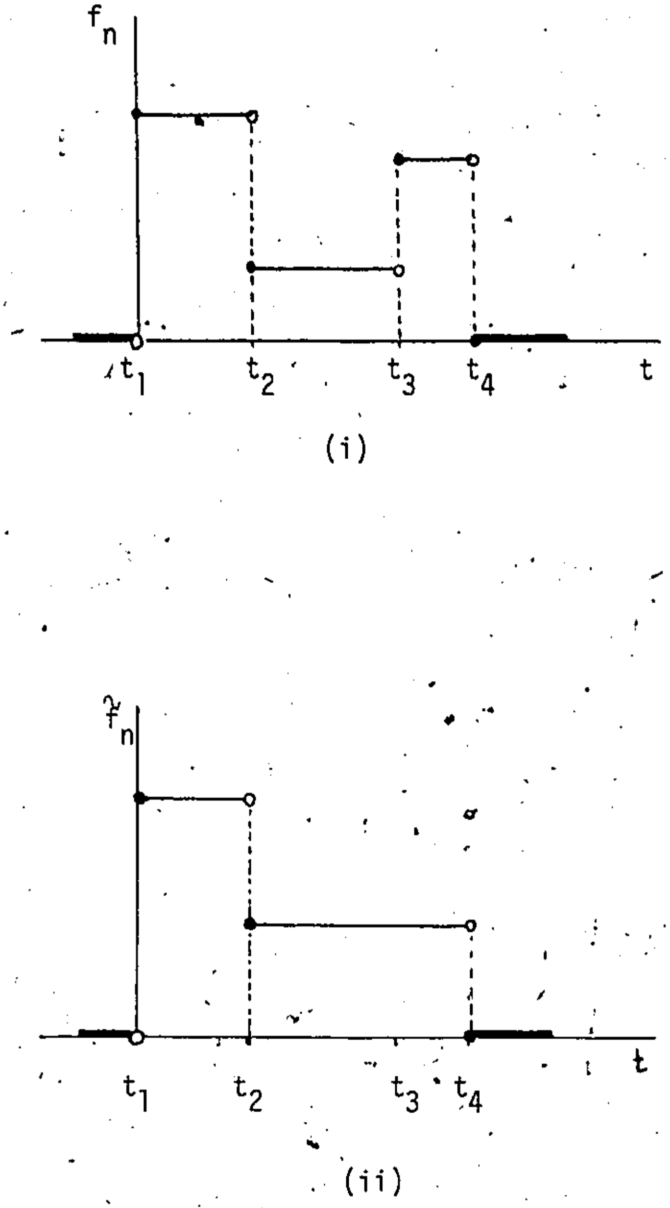



Figure 3.E.

# 4. Uniform consistency of the monotonized estimator.

## 4.1. Introduction.

As in chapter 3, let $(X_i)_{i=1}^{\infty}$ be a sequence of independent nonnegative random variables with common probability density function f, where f is nonincreasing on $[0,\infty)$. Since f is real-valued at 0, it follows that f is bounded. -- The distribution function of the $X_i$'s is

$F(t) = \int_0^t f(x)dx$, and we let $\tau = \inf\{t: F(t) = 1\}$. If $X_1$ is not a.s. bounded, we have that $F(t) < 1$ for all t, hence $\tau = \inf\{t: F(t) = 1\} = \inf \phi = \infty$ ;otherwise $\tau$ is finite. For every $n \in \mathbb{N}^+$, let $0 = t_1^{(n)} < t_2^{(n)} < \ldots$ be a partition (see 3.1) of $[0,\infty)$. As f is monotone on $[0,\infty)$, the limit $f(x^+)$ exists for all $x \in [0,\infty)$ and the limit $f(x^-)$ exists for all $x \in (0,\infty)$. We consider the version f of the density of the $X_i$'s which is right-continuous on $[0,\infty)$. We note that the assumption that f is right-continuous at 0 is, in fact, essential for the application to renewal processes in chapter 5. The symbols $\Delta_k^{(n)}$, $\delta_k^{(n)}$, $I_k^{(n)}$, and $a_k^{(n)}$, and the histogram estimator $\hat{f}_n$ and the monotonized estimator $\tilde{f}_n$, are defined as in chapter 3.

The purpose of this chapter is to obtain results on the uniform consistency of $\tilde{f}_n$. Those results are given in theorems 4.6 and 4.7.1. If $\tau < \infty$ and $f(\tau^-) > 0$ , some interesting results can be obtained when, for each n, the partition can be chosen so that $t_{\ell_n}^{(n)} = \tau$ for some integer $\ell_n$. In practice, $\tau$ is generally unknown; and clearly $\tau$ can be chosen as a partition point only if it is known . Therefore, when $\tau$ is finite we consider two cases: $\tau$ known, and $\tau$ unknown. In theorem

4.6 we consider the case where f is continuous on $[0,\infty)$, and the case where f is not continuous at $\tau$ ($\tau < \infty$) with $\tau$ unknown. In theorem 4.7.1 we consider the case where f is not continuous at $\tau$ ($\tau < \infty$) and $\tau$ is known.

The uniform consistency of $\hat{f}_n$ is obtained by first showing its pointwise consistency on an interval and then extending this result to uniform consistency on that interval by means of analytic methods (see 4.5).

There exist pointwise consistency results for $\hat{f}_n$ on $(0,\tau)$ in Barlow et al. (1972) but we are not aware of any such results for intervals including 0 or $\tau$ ( if $\tau < \infty$ ). To obtain uniform consistency on $[0,\infty)$, when $\tau < \infty$, we establish consistency at 0 and at $\tau$ as well as on $(0,\tau) \cup (\tau,\infty)$. To obtain consistency at 0 we need some conditions on the rate at which $\delta_k^{(n)}$ decreases as $n \to \infty$, for those k for which $\Delta_k^{(n)}$ is in a neighborhood of 0. To obtain consistency at $\tau$, when $\tau$ is known, we let, for each n, $t_{\ell_n}^{(n)} = \tau$ for some $\ell_n$. This has the effect of making $\hat{f}_n = 0 = f$ a.s. on $[\tau,\infty)$. We then consider the consistency of $\hat{f}_n(\tau^-)$. To do this, we need some conditions on the rate at which $\delta_k^{(n)}$ decreases, as $n \to \infty$, for those k for which $\Delta_k^{(n)}$ is in a neighborhood, and at the left, of $\tau$. To obtain the consistency at 0 and at $\tau$, we partly follow some arguments used by Révész (1968, pp. 163-167). We give those results in 4.3.6 and 4.7.

In lemma 4.2, we show the pointwise consistency of $\hat{f}_n$ at points of continuity of f. The result and argument are adapted from Barlow et al. (1972, p. 149). Those authors consider what they call the "generalized failure rate function", a function that includes the usual failure rate function and density as special cases. When the generalized failure rate function becomes the density f, the corresponding estimator given by Barlow et al. becomes precisely the monotone estimator $\tilde{f}_n$ introduced in chapter 3. -- By departing from the generality of the argument in Barlow et al., adapting their argument to the specific case of density estimation, we obtain an argument we believe to be more "transparent" than theirs - and a result which is valid with fewer conditions. More explicitly, the general result in Barlow et al. implies that $\tilde{f}_n$ is consistent at points $x_0$ of continuity of f satisfying $0 < F(x_0) < 1$. We establish consistency for continuity points of f on $(0,\infty)$, and do that without any reliance on some of their conditions (e.g., the way the sequence of partitions should "become dense" and the existence of a finite expectation for $X_1$).

Throughout this chapter we refer to the set

$$\Omega_1 = \{\omega: \lim_{n\to\infty} \sup_t |F(t) - F_n(t;\omega)| = 0\}$$

which, by the Glivenko-Cantelli theorem, has probability one. [As we use the left-continuous form of the e.d.f., the theorem in fact asserts that $\sup_{x\in\mathbb{R}}|F(x^-) - F_n(x;\omega)| \to 0$ a.s.; but $F(x^-) = F(x)$ at continuity points and having a density, F is continuous].

Lemma 4.2. Suppose that $\sup_k \delta_k^{(n)} \to 0$ when $n \to \infty$. For any fixed $x_0 \in (0,\infty)$ and for all $\omega \in \Omega_1$, as $n \to \infty$,

$$f(x_0^+) \le \underline{\lim}\, \tilde{f}_n(x_0;\omega) \le \overline{\lim}\, \tilde{f}_n(x_0;\omega) \le f(x_0^-)$$

and

$$\underline{\lim}\, \tilde{f}_n(0,\omega) \ge f(0).$$

Proof. Consider an arbitrary $\omega \in \Omega_1$.

We first prove that, for an arbitrary $x_o \in (0,\infty)$,

$$\overline{\lim}\, \hat{f}_n(x_o,\omega) \leq f(x_o^-) .$$

Consider an arbitrary $\xi$ such that $0 \leq \xi < x_o$. Select partition points $t_{j_n}^{(n)}$, $n = 1,2,\ldots$, so that $t_{j_n}^{(n)} \geq \xi$ and $t_{j_n}^{(n)} \to \xi$ as $n \to \infty$. There exists an integer N such that, for $n \geq N$, $\xi \leq t_{j_n}^{(n)} < x_o$. Given $n \geq N$, there exists k such that $x_o \in \Delta_k^{(n)} = [t_k^{(n)},t_{k+1}^{(n)})$; for such an n, using (iv) in 3.15.B,

$$\text{(i)} \qquad \hat{f}_n(x_o;\omega) = \min_{1\leq r\leq k}\left\{\max_{s\geq k+1} \frac{F_n(t_s^{(n)};\omega) - F_n(t_r^{(n)};\omega)}{t_s^{(n)} - t_r^{(n)}}\right\}$$

$$= \min_{t_r^{(n)}\leq x_o}\left\{\max_{t_s^{(n)}>x_o} \frac{F_n(t_s^{(n)};\omega) - F_n(t_r^{(n)};\omega)}{t_s^{(n)} - t_r^{(n)}}\right\}$$

$$\leq \max_{t_s^{(n)}>x_o} \frac{F_n(t_s^{(n)};\omega) - F_n(t_{j_n}^{(n)};\omega)}{t_s^{(n)} - t_{j_n}^{(n)}}$$

$$\leq \sup_{x_o\leq x<\infty} \frac{F_n(x;\omega) - F_n(t_{j_n}^{(n)};\omega)}{x - t_{j_n}^{(n)}} .$$

Consider an arbitrary $x \in [x_o,\infty)$. For each $n \geq N$, $x - t_{j_n}^{(n)} \geq x - \xi > 0$.

Given $\varepsilon > 0$, let n be large enough so that, by the Glivenko-Cantelli theorem,

$$\sup_{0 \leq t < \infty} |F_n(t;\omega) - F(t)| < \varepsilon$$

and, by the right-continuity of F at $\xi$,

$$|F(t_{j_n}^{(n)}) - F(\xi)| < \varepsilon .$$

It follows that

$$\text{(ii)} \qquad F_n(x;\omega) - F_n(t_{j_n}^{(n)};\omega) = F_n(x;\omega) - F(x) + F(t_{j_n}^{(n)}) - F_n(t_{j_n}^{(n)};\omega)$$

$$+ F(\xi) - F(t_{j_n}^{(n)}) + F(x) - F(\xi)$$

$$\leq 3\varepsilon + F(x) - F(\xi) .$$

Now

$$\text{(iii)} \qquad \sup_{x_0 \leq x < \infty} \frac{F(x) - F(\xi)}{x - t_{j_n}^{(n)}} = \sup_{x_0 \leq x < \infty} \left\{ \frac{F(x) - F(\xi)}{x - \xi} \; \frac{x - \xi}{x - t_{j_n}^{(n)}} \right\}$$

$$\leq \left\{ \sup_{x_0 \leq x < \infty} \frac{F(x) - F(\xi)}{x - \xi} \right\} \left\{ \sup_{x_0 \leq x < \infty} \frac{x - \xi}{x - t_{j_n}^{(n)}} \right\}$$

$$= \left\{ \sup_{x_0 \leq x < \infty} \frac{F(x) - F(\xi)}{x - \xi} \right\} \frac{x_0 - \xi}{x_0 - t_{j_n}^{(n)}}$$

$$\to \sup_{x_0 \leq x < \infty} \frac{F(x) - F(\xi)}{x - \xi}$$

and

(iv) $$\sup_{x_o \le x < \infty} \frac{3\varepsilon}{x - t_{j_n}^{(n)}} = \frac{3\varepsilon}{x_o - t_{j_n}^{(n)}} \to \frac{3\varepsilon}{x_o - \xi} .$$

By (i) and (ii),

$$\tilde{f}_n(x_o;\omega) \le \sup_{x_o \le x < \infty} \frac{3\varepsilon}{x - t_{j_n}^{(n)}} + \sup_{x_o \le x < \infty} \frac{F(x) - F(\xi)}{x - t_{j_n}^{(n)}}$$

Then, from (iii) and (iv) and the arbitrariness of $\varepsilon$,

$$\overline{\lim}\, \tilde{f}_n(x_o;\omega) \le \sup_{x_o \le x < \infty} \frac{F(x) - F(\xi)}{x - \xi}$$

As f is nonincreasing on $[0,\infty)$ and $\xi < x_o$, if $x_o \le x < \infty$ then

$$\frac{F(x) - F(\xi)}{x - \xi} = \frac{1}{x - \xi} \int_\xi^x f(t)dt \le \frac{1}{x - \xi} \int_\xi^x f(\xi)dt = f(\xi) .$$

We conclude that

$$\overline{\lim}\, \tilde{f}_n(x_o;\omega) \le f(\xi) .$$

As $\xi$ was arbitrary in $[0,x_o)$, $\overline{\lim}\, \tilde{f}_n(x_o;\omega) \le f(x_o^-)$ .

We now prove that, for an arbitrary $x_o$ in $[0,\infty)$, $\underline{\lim}\, \tilde{f}_n(x_o;\omega) \ge f(x_o^+)$ . Consider an arbitrary $\xi$ such that $x_o < \xi < \infty$. Select partition points $t_{j_n}^{(n)}$, n = 1, 2,..., so that $t_{j_n}^{(n)} \le \xi$ and $t_{j_n}^{(n)} \to \xi$ as $n \to \infty$.

There exists an integer N such that, for $n \geq N$, $x_o < t_{j_n}^{(n)} \leq \xi$. Given $n \geq N$, there exists k such that $x_o \in \Delta_k^{(n)} = [t_k^{(n)}, t_{k+1}^{(n)})$; for such an n, by an argument similar to that in (i),

$$\text{(v)} \qquad \mathfrak{F}_n(x_o;\omega) \geq \inf_{0 \leq x \leq x_o} \frac{F_n(t_{j_n}^{(n)};\omega) - F_n(x;\omega)}{t_{j_n}^{(n)} - x} .$$

Consider an arbitrary $x \in [0,x_o]$. Given $\varepsilon > 0$, let n be large enough so that, by the Glivenko-Cantelli theorem, by the left-continuity of F at $\xi$, and by arguments such as those which yield (ii), we obtain

$$\text{(vi)} \qquad F_n(t_{j_n}^{(n)};\omega) - F_n(x;\omega) \geq -3\varepsilon + F(\xi) - F(x) .$$

Now, since $\xi \geq t_{j_n}^{(n)} > x_o$,

$$\text{(vii)} \qquad \inf_{0 \leq x \leq x_o} \frac{F(\xi) - F(x)}{t_{j_n}^{(n)} - x} \geq \inf_{0 \leq x \leq x_o} \frac{F(\xi) - F(x)}{\xi - x}$$

and

$$\text{(viii)} \qquad \inf_{0 \leq x \leq x_o} \frac{-3\varepsilon}{t_{j_n}^{(n)} - x} \geq \frac{-3\varepsilon}{t_{j_n}^{(n)} - x_o} \to \frac{-3\varepsilon}{\xi - x_o} .$$

By (v) and (vi),

$$\mathfrak{F}_n(x_o;\omega) \geq \inf_{0 \leq x \leq x_o} \frac{-3\varepsilon}{t_{j_n}^{(n)} - x} + \inf_{0 \leq x \leq x_o} \frac{F(\xi) - F(x)}{t_{j_n}^{(n)} - x} .$$

Then, from (vii) and (viii) and the arbitrariness of $\varepsilon$,

$$\underline{\lim}\ \hat{f}_n(x_0;\omega) \geq \inf_{0 \leq x \leq x_0} \frac{F(\xi) - F(x)}{\xi - x} .$$

As $f$ is nonincreasing on $[0,\infty)$ and $x_0 < \xi$, if $0 \leq x \leq x_0$ then

$$\frac{F(\xi) - F(x)}{\xi - x} \geq f(\xi) .$$

We conclude that

$$\underline{\lim}\ \hat{f}_n(x_0;\omega) \geq f(\xi) .$$

As $\xi$ was arbitrary in $(x_0,\infty)$, $\underline{\lim}\ \hat{f}_n(x_0;\omega) \geq f(x_0^+)$, and it follows for $x_0 = 0$ that $\underline{\lim}\ \hat{f}_n(0;\omega) \geq f(0^+) = f(0)$ . □

4.3. Consistency at 0.

In this section we follow some arguments which are, in part, based on arguments used by Révész (1968, pp. 163-166), to prove the consistency of $\hat{f}_n$ at 0. We start with a useful inequality.

Proposition 4.3.1. If $x \leq 1$ then $1+x+x^2 \geq e^x$.

Proof. Suppose that $x \leq -1$. As $x \to -\infty$, $1+x+x^2$ increases whereas $e^x$ decreases and, at $x = -1$, $1+x+x^2 \geq e^x$; therefore $1+x+x^2 \geq e^x$ on $(-\infty,-1]$.

On $[-1,1]$, noting that $\sum_{n=0}^{2} 1/n! = 5/2$, we have

$$1+x+x^2 = (1+x+\tfrac{1}{2}x^2) + \tfrac{1}{2}x^2$$

$$\geq (1+x+\tfrac{1}{2}x^2) + (e - \tfrac{5}{2})x^2$$

$$= (1+x+\tfrac{1}{2}x^2) + (\sum_{n\geq 3} \tfrac{1}{n!})x^2$$

$$\geq (1+x+\tfrac{1}{2}x^2) + \sum_{n\geq 3} \tfrac{x^n}{n!} = e^x .$$

□

Proposition 4.3.2. For a real random variable X, and any $x \in \mathbb{R}$,

$$P[X \geq x] \leq e^{-x}\mathbb{E}e^X .$$

Proof. This is obtained for example from inequality 5.26 in Billingsley (1979, p. 65). □

Proposition 4.3.3. If $n \in \mathbb{N}^+$ and $0 \leq \lambda_n < \infty$, and x and y are such that $y - x \geq n^{-1}\lambda_n$, then

$$\mathbb{E} \exp \left\{\lambda_n\left( \frac{F_n(y) - F_n(x)}{y - x} - \frac{F(y) - F(x)}{y - x}\right)\right\} \leq \exp \left(\frac{\lambda_n^2}{n(y-x)} f(x)\right) .$$

Proof. Suppose that x, y and $\lambda_n$ satisfy the stated conditions. Put

$$\alpha_i = I_{[x,y)}(X_i) - \mathbb{E}I_{[x,y)}(X_i)$$

and note that the random variables $\alpha_i$ $(i = 1,2,\ldots)$ are independent, have mean 0, and have variance bounded as follows:

$$\mathbb{E}\, \alpha_i^2 = \mathbb{E}[I_{[x,y)}(X_i) - \mathbb{E}I_{[x,y)}(X_i)]^2$$

$$= \mathbb{E}I^2_{[x,y)}(X_i) - (\mathbb{E}I_{[x,y)}(X_i))^2 ,$$

$$\leq \mathbb{E}I^2_{[x,y)}(X_i) = \mathbb{E}I_{[x,y)}(X_i) .$$

Consequently,

$$\text{(i)} \qquad \frac{1}{y-x}\mathbb{E}\,\alpha_i^2 \le \frac{1}{y-x}\int_x^y f(s)ds$$

$$\le \frac{1}{y-x}\int_x^y f(x)ds = f(x) .$$

Now

$$\text{(ii)} \qquad \frac{F_n(y) - F_n(x)}{y - x} - \frac{F(y) - F(x)}{y - x} = \frac{F_n(y) - F_n(x)}{y - x} - \mathbb{E}\left(\frac{F_n(y) - F_n(x)}{y - x}\right)$$

$$= \frac{1}{y - x}\,\frac{1}{n}\sum_{i=1}^{n}\left[I_{[x,y)}(X_i) - \mathbb{E}(I_{[x,y)}(X_i))\right]$$

$$= \frac{1}{y - x}\,\frac{1}{n}\sum_{i=1}^{n}\alpha_i .$$

As $\alpha_i \le 1$ , and $y-x \ge n^{-1}\lambda_n$ by hypothesis, we have

$$(n(y-x))^{-1}\lambda_n\alpha_i \le (n(y-x))^{-1}\lambda_n \le 1 .$$

By the independence of the random variables $\alpha_i$ $(i = 1,\ldots,n)$, proposition 4.3.1, and (i) and (ii), we obtain a bound:

$$\mathbb{E}\exp\left\{\lambda_n\left(\frac{F_n(y) - F_n(x)}{y - x} - \frac{F(y) - F(x)}{y - x}\right)\right\}$$

$$= \mathbb{E}\exp\left\{\lambda_n \sum_{i=1}^{n} \frac{1}{n(y-x)} \alpha_i\right\}$$

$$= \prod_{i=1}^{n} \mathbb{E}\exp\left\{\frac{\lambda_n}{n(y-x)} \alpha_i\right\}$$

$$\leq \prod_{i=1}^{n} \mathbb{E}\left\{1 + \frac{\lambda_n}{n(y-x)} \alpha_i + \left(\frac{\lambda_n}{n(y-x)} \alpha_i\right)^2\right\}$$

$$= \prod_{i=1}^{n}\left\{1 + \left(\frac{\lambda_n^2}{n^2(y-x)}\right) \frac{1}{y-x} \mathbb{E}\,\alpha_i^2\right\}$$

$$\leq \prod_{i=1}^{n}\left\{1 + \frac{\lambda_n^2}{n^2(y-x)} f(x)\right\}$$

$$\leq \left(\exp\left\{\frac{\lambda_n^2}{n^2(y-x)} f(x)\right\}\right)^n = \exp\left\{\frac{\lambda_n^2}{n(y-x)} f(x)\right\}. \qquad \square$$

<u>Proposition 4.3.4</u>. If $n \in \mathbb{N}^+$ and $1 \leq \lambda_n < \infty$, and x and y are such that $0 \leq x < y < \infty$ and $y - x \geq n^{-1}\lambda_n$ then, for any $\mu_n \in \mathbb{R}$,

$$P\left[\frac{F_n(y) - F_n(x)}{y - x} - \frac{F(y) - F(x)}{y - x} \geq \frac{\mu_n}{\lambda_n}\right] \leq e^{-\mu_n} \exp\left\{\frac{\lambda_n^2}{n(y-x)} f(0)\right\}.$$

<u>Proof</u>. Suppose that x, y and $\lambda_n$ satisfy the stated conditions. Now, by propositions 4.3.3 and 4.3.2,

$$e^{-\mu_n} \exp\left\{\frac{\lambda_n^2}{n(y-x)} f(0)\right\} \geq e^{-\mu_n} \exp\left\{\frac{\lambda_n^2}{n(y-x)} f(x)\right\}$$

$$\geq e^{-\mu_n} \mathbb{E} \exp\left\{\lambda_n\left(\frac{F_n(y) - F_n(x)}{y - x} - \frac{F(y) - F(x)}{y - x}\right)\right\}$$

$$\geq P\left[\lambda_n\left(\frac{F_n(y) - F_n(x)}{y - x} - \frac{F(y) - F(x)}{y - x}\right) \geq \mu_n\right] . \qquad \square$$

Proposition 4.3.5. Suppose that, for some $0 < T < \infty$

$$\rho = \inf \{t_{k+1}^{(n)} - t_k^{(n)} : 0 \leq t_k^{(n)} < T\} \geq n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \geq 3$. Then

$$\overline{\lim}\ \tilde{f}_n(0) \leq f(0) \quad \text{a.s.}$$

Proof. Observe first that, for any real $\lambda_n$ and $\mu_n$,

(i) $$P\left[\sup_{0<t_k^{(n)}\leq T} \left(\frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}}\right) \geq \frac{\mu_n}{\lambda_n}\right]$$

$$= P\left(\bigcup_{0<t_k^{(n)}\leq T} \left[\frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \geq \frac{\mu_n}{\lambda_n}\right]\right)$$

$$\leq \sum_{0<t_k^{(n)}\leq T} P\left[\frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \geq \frac{\mu_n}{\lambda_n}\right] .$$

Suppose that the partition points $t_k^{(n)}$ satisfy the stated conditions and and put $\lambda_n = (n\rho \log n)^{1/2}$. Now $\lambda_n \geq 1$ and

$t_{j+1}^{(n)} - t_j^{(n)} \geq \rho \geq n^{-1/2}\rho^{1/2}(\log n)^{1/2} = n^{-1}\lambda_n$ for $n \geq 3$ and $0 \leq t_j^{(n)} < T$, hence $t_k^{(n)} \geq n^{-1}\lambda_n$ whenever $0 < t_k^{(n)} \leq T$. For $n \geq 3$, apply proposition 4.3.4 to each term in the sum in (i), with $x = 0$ and $y = t_k^{(n)}$, to obtain

$$\text{(ii)} \qquad P\left[\sup_{0<t_k^{(n)}\leq T}\left(\frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}}\right) \geq \frac{\mu_n}{\lambda_n}\right]$$

$$\leq \sum_{0<t_k^{(n)}\leq T} e^{-\mu_n} \exp\left\{\frac{\lambda_n^2}{n\rho} f(0)\right\} .$$

The hypothesis

$$\inf\{t_{k+1}^{(n)} - t_k^{(n)} : 0 \leq t_k^{(n)} < T\} \geq n^{-1}(\log n)^r$$

implies that the sum in (ii) has at most $T/\{n^{-1}(\log n)^r\}$ terms, hence is bounded by

$$\text{(iii)} \qquad \frac{nT}{(\log n)^r} e^{-\mu_n} \exp\left\{\frac{n\rho(\log n)}{n\rho} f(0)\right\} = \frac{T}{(\log n)^r} e^{-\mu_n} n^{(1+f(0))} .$$

Put $\mu_n = C \log n$, with $C > 2+f(0)$. Now, for $n \geq 3$,

$$\frac{\mu_n}{\lambda_n} = \frac{C \log n}{(n\rho \log n)^{1/2}} \leq \frac{C \log n}{(\log n)^{(r+1)/2}} ;$$

thus, due to $r > 1$, $\mu_n/\lambda_n \to 0$. In view of (ii) and (iii),

$$\sum_{n=3}^{\infty} P\left[\sup_{0<t_k^{(n)}\leq T}\left(\frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}}\right) \geq \frac{\mu_n}{\lambda_n}\right]$$

$$\leq \sum_{n=3}^{\infty} \frac{T}{(\log n)^r} n^{-C+1+f(0)} < \infty .$$

By the Borel-Cantelli lemma,

$$P(\limsup_{n\to\infty} A_n) = 0 \quad \text{and} \quad P(\liminf_{n\to\infty} A_n^c) = 1$$

where

$$A_n = \left[ \sup_{0<t_k^{(n)}\le T} \left( \frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) \ge \frac{\mu_n}{\lambda_n} \right]$$

and

$$A_n^c = \left[ \sup_{0<t_k^{(n)}\le T} \left( \frac{F_n(t_k^{(n)})}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) < \frac{\mu_n}{\lambda_n} \right] .$$

Consider an arbitrary $\omega$ in $\liminf_{n\to\infty} A_n^c$. Now there exists an integer $n_o(\omega)$ such that $\omega \in A_n^c$ when $n \ge n_o(\omega)$, hence

$$\overline{\lim_{n\to\infty}} \sup_{0<t_k^{(n)}\le T} \left( \frac{F_n(t_k^{(n)};\omega)}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) \le \overline{\lim} \frac{\mu_n}{\lambda_n} = 0 .$$

Recall that

$$\Omega_1 = \{\omega: \lim_{n\to\infty} \sup_t |F_n(t;\omega) - F(t)| = 0\}$$

and put

$$\Omega_2 = \left\{ \omega: \overline{\lim_{n\to\infty}} \sup_{0<t_k^{(n)}\le T} \left( \frac{F_n(t_k^{(n)};\omega)}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) \le 0 \right\} .$$

Now $P(\Omega_1) = 1$ by the Glivenko-Cantelli theorem and $P(\Omega_2) = 1$ because $\liminf A_n^c \subset \Omega_2$. -- Consider an arbitrary $\varepsilon > 0$ and $\omega \in \Omega_1 \cap \Omega_2$. There exists $n_1(\omega)$ such that, when $n \geq n_1(\omega)$,

$$\sup_t |F_n(t;\omega) - F(t)| < \varepsilon T \ ,$$

which implies

$$\text{(iv)} \qquad \sup_{t_k^{(n)} > T} \left( \frac{F_n(t_k^{(n)};\omega)}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) < \varepsilon \ .$$

And there exists $n_2(\omega)$ such that, when $n \geq n_2(\omega)$,

$$\text{(v)} \qquad \sup_{0 < t_k^{(n)} \leq T} \left( \frac{F_n(t_k^{(n)};\omega)}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) < \varepsilon \ .$$

Recall from 3.16 that, for any $\omega$ and $n$,

$$\mathcal{F}_n(0;\omega) = \frac{F_n(t_{k_1+1}^{(n)};\omega)}{t_{k_1+1}^{(n)}}$$

where $k_1$ is determined by algorithm 3.14. Now by (iv) and (v), for $\omega \in \Omega_1 \cap \Omega_2$ and $n \geq \max\{n_1(\omega), n_2(\omega)\}$,

$$\mathcal{F}_n(0;\omega) - \frac{F(t_{k_1+1}^{(n)})}{t_{k_1+1}^{(n)}} = \frac{F_n(t_{k_1+1}^{(n)};\omega)}{t_{k_1+1}^{(n)}} - \frac{F(t_{k_1+1}^{(n)})}{t_{k_1+1}^{(n)}}$$

$$\leq \sup_{t_k^{(n)} > 0} \left( \frac{F_n(t_k^{(n)};\omega)}{t_k^{(n)}} - \frac{F(t_k^{(n)})}{t_k^{(n)}} \right) < \varepsilon \ .$$

It follows that

$$\hat{f}_n(0;\omega) < \varepsilon + \frac{F(t_{k_1+1}^{(n)})}{t_{k_1+1}^{(n)}} = \varepsilon + \frac{1}{t_{k_1+1}^{(n)}} \int_0^{t_{k_1+1}^{(n)}} f(s)ds$$

$$\leq \varepsilon + f(0) \quad .$$

As $\varepsilon$ was arbitrary, and $\omega$ is an arbitrary element of $\Omega_1 \cap \Omega_2$, this implies

$$\overline{\lim}\, \hat{f}_n(0) \leq f(0) \quad \text{a.s.} \quad .$$

□

Theorem 4.3.6. Suppose that $\sup_k \delta_k^{(n)} \to 0$ and that, for some $T > 0$,

$$\inf\{t_{k+1}^{(n)} - t_k^{(n)}: 0 \leq t_k^{(n)} < T\} \geq n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \geq 3$. Then

$$\hat{f}_n(0) \to f(0) \quad \text{a.s.} \quad .$$

Proof. Suppose that the partition points $t_k^{(n)}$ satisfy the above conditions. By lemma 4.2, the condition that $\sup_k \delta_k^{(n)} \to 0$ as $n \to \infty$ implies

$$\underline{\lim}\, \hat{f}_n(0) \geq f(0) \quad \text{a.s.}$$

and, by proposition 4.3.5, the other condition in the theorem implies

$$\overline{\lim}\, \hat{f}_n(0) \leq f(0) \quad \text{a.s.} \quad .$$

From those two inequalities we conclude that

$$\lim_n \hat{f}_n(0) = f(0) \quad \text{a.s.} \quad .$$

□

We now obtain the theorem on pointwise consistency of $\hat{f}_n$ .

<u>Theorem 4.4</u>. Suppose that $\sup_k \delta_k^{(n)} \to 0$ and that, for some $T > 0$,

$$\inf\{t_{k+1}^{(n)} - t_k^{(n)} : 0 \leq t_k^{(n)} < T\} \geq n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \geq 3$ .

(a) If f is continuous on $[0,\infty)$ - $\tau$ finite or not - then, for each $x_0 \in [0,\infty)$ and all $\omega$ in a set $\Omega_0$ of probability one,

$$\lim_n \hat{f}_n(x_0,\omega) = f(x_0) .$$

(b) Suppose that $\tau < \infty$. If f is continuous on $[0,\tau)$ then, for arbitrary $T_1$ and $T_2$ such that $T_1 < \tau < T_2$, for all $x_0 \in [0,T_1] \cup [T_2,\infty)$, and for all $\omega$ in a set $\Omega_0$ of probability one,

$$\lim_n \hat{f}_n(x_0,\omega) = f(x_0) .$$

<u>Proof</u>.

(a) Suppose that f is continuous on $[0,\infty)$. By lemma 4.2, for each $x_0 \in (0,\infty)$ and all $\omega \in \Omega_1$, $\hat{f}_n(x_0,\omega) \to f(x_0)$;and,by theorem 4.3.6, $\hat{f}_n(0) \to f(0)$ a.s. .

(b) Suppose that $\tau < \infty$ and that f is continuous on $[0,\tau)$. Consider an arbitrary $T_1 < \tau$. By lemma 4.2, for each $x_0 \in (0,T_1)$ and all $\omega \in \Omega_1$, $\hat{f}_n(x_0,\omega) \to f(x_0)$. By theorem 4.3.6, $\hat{f}_n(0) \to f(0)$ a.s. . With probability one, none of the observations $X_i$ has value greater than $\tau$. Fix an arbitrary $T_2 > \tau$. For n sufficiently large, there is a partition point $t_k^{(n)}$ such that $\tau < t_k^{(n)} \leq T_2$. Then, for all such n, $\hat{f}_n = 0 = f$ a.s. on $[T_2,\infty)$. □

To extend this result to the uniform consistency of $\tilde{f}_n$ we use the following lemma.

Lemma 4.5. Suppose that I is a closed interval whose left endpoint is finite; i.e., I = [a,b] or [a,∞), with $-\infty < a < b < \infty$ . Suppose that probability density functions $(h_n)_1^\infty$ converge pointwise to a continuous probability density function h. If h and each $h_n$ are nonincreasing on I then $h_n$ converges uniformly to h on I.

This can be proved by an argument analogous to that used to prove Polya's result - that if probability distribution functions $(H_n)_1^\infty$ converge pointwise to a continuous probability distribution function H then the convergence is in fact uniform.

The following theorem states our main results for cases where τ is unknown.

Theorem 4.6. Suppose that $\sup_k \delta_k^{(n)} \to 0$ and that, for some T > 0,

$$\inf\{t_{k+1}^{(n)} - t_k^{(n)} : 0 \leq t_k^{(n)} < T\} \geq n^{-1}(\log n)^r$$

for some r > 1 and all $n \geq 3$ .

(a) If f is continuous on [0,∞) -τ finite or not- then

$$P\left[\sup_{0\leq t<\infty} |f(t) - \tilde{f}_n(t)| \to 0\right] = 1 .$$

(b) Suppose that τ < ∞. If f is continuous on [0,τ) then, for arbitrary $T_1$ and $T_2$ such that $T_1 < \tau < T_2$ ,

$$P[\sup |f(t) - \tilde{f}_n(t)| \to 0] = 1$$

where the supremum is taken over all $t \in [0,T_1] \cup [T_2,\infty)$.

<u>Proof</u>. Statement (a) follows directly from lemma 4.5 and result (a) of theorem 4.4. Then from this and result (b) of theorem 4.4,

$$P\left[\sup_{0\le t\le T_1} |f(t) - \hat{f}_n(t)| \to 0\right] = 1$$

and

$$P\left[\sup_{T_2\le t<\infty} |f(t) - \hat{f}_n(t)| \to 0\right] = 1 \quad . \qquad \square$$

## 4.7. Consistency when $\tau$ is finite and known.

In this section, we consider the case where $\tau$ is finite and known, and $f$ is continuous on $[0,\tau)$, $f(\tau^-) > 0$ (i.e., $f$ is not continuous at $\tau$). We obtain the uniform consistency of $\hat{f}_n$ on $[0,\infty)$. In view of theorem 4.6.(b), the crucial point is at $\tau$.

Consider the set

$$\Omega_3 = \{\omega: X_i(\omega) < \tau, \quad i \in \mathbb{N}^+\} .$$

By definition of $\tau$, $P(\Omega_3) = 1$. Recall that, given $n$, the random variable $M$ is such that $\max_{1\le i\le n} X_i \in \Delta_M^{(n)}$; see 3.1. For each $n$, let $t_{\ell_n}^{(n)} = \tau$ for some integer $\ell_n$. Thus, if $\omega \in \Omega_3$ then $t_{M(\omega)}^{(n)} \le t_{\ell_n-1}^{(n)}$ . Note that $\hat{f}_n(t_{M(\omega)+1}^{(n)};\omega) = 0$ for each $n$, by algorithm 3.14, so that $\hat{f}_n(\tau;\omega) = 0$ for each $n$. Also, since $\hat{f}_n(\tau^-;\omega) = \hat{f}_n(t_{\ell_n-1}^{(n)};\omega)$,

$$\text{(i)} \qquad \hat{f}_n(\tau^-;\omega) = \begin{cases} 0 & , \text{ if } \tau > t_{M(\omega)+1} , \\ \dfrac{F_n(t^{(n)}_{k_\sigma+1};\omega) - F_n(t^{(n)}_{k_{\sigma-1}+1};\omega)}{t^{(n)}_{k_\sigma+1} - t^{(n)}_{k_{\sigma-1}+1}} & , \text{ if } \tau = t_{M(\omega)+1} , \end{cases}$$

where $k_\sigma = M(\omega)$ and $k_{\sigma-1} \leq k_\sigma - 1$ is determined by algorithm 3.14.

Now assume that, for some $T \in (0,\tau)$,

$$\inf\{t^{(n)}_k - t^{(n)}_{k-1}:\ \tau\text{-}T < t^{(n)}_k \leq \tau\} \geq n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \geq 3$. Consider the probability

$$P\left[\sup_{\tau-T\leq t^{(n)}_k<\tau}\left(\frac{F(\tau) - F(t^{(n)}_k)}{\tau - t^{(n)}_k} - \frac{F_n(\tau) - F_n(t^{(n)}_k)}{\tau - t^{(n)}_k}\right) \geq \frac{\mu_n}{\lambda_n}\right]$$

instead of the one considered in (i) in the proof of proposition 4.3.4. Using arguments analogous to those used in section 4.3, and in particular to those used in the proof of proposition 4.3.4, one obtains $P(\Omega_2) = 1$ where

$$\Omega_2 = \left\{\omega:\ \overline{\lim_{n\to\infty}}\ \sup_{\tau-T\leq t^{(n)}_k<\tau}\left(\frac{F(\tau)-F(t^{(n)}_k)}{\tau-t^{(n)}_k} - \frac{F_n(\tau;\omega)-F_n(t^{(n)}_k;\omega)}{\tau-t^{(n)}_k}\right) \leq 0\right\}$$

Recall that $f(\tau^-) > 0$ by hypothesis. Since

$$F(\tau) - F(t^{(n)}_{\ell_n-1}) = \int_{t^{(n)}_{\ell_n-1}}^{\tau} f(s)ds \geq (\tau - t^{(n)}_{\ell_n-1})f(\tau^-) > 0 ,$$

for $\omega \in \Omega_2 \cap \Omega_3$ there exists $n_3(\omega)$ such that, when $n \geq n_3(\omega)$,

$$\frac{F_n(\tau;\omega) - F_n(t^{(n)}_{\ell_n-1};\omega)}{\tau - t^{(n)}_{\ell_n-1}} > 0 .$$

Thus there are some observations $X_j(\omega)$ in $\Delta^{(n)}_{\ell_n-1}$ for all $n \geq n_3(\omega)$. Furthermore, for each $n \geq n_3(\omega)$, $\max_{1\leq i\leq n} X_i(\omega) < \tau = t^{(n)}_{\ell_n}$ and then $M(\omega) = t^{(n)}_{\ell_n-1}$, i.e., $\tau = t^{(n)}_{M(\omega)+1}$ for all $n \geq n_3(\omega)$. It follows from this and from (i) that, for each $n \geq n_3(\omega)$,

$$\text{(ii)} \qquad \tilde{f}_n(\tau^-;\omega) = \frac{F_n(\tau;\omega) - F_n(t^{(n)}_{k_{\sigma-1}+1};\omega)}{\tau - t^{(n)}_{k_{\sigma-1}+1}} ,$$

where $k_{\sigma-1}$ is determined by algorithm 3.14. Recall that $P(\Omega_1) = 1$, where

$$\Omega_1 = \{\omega: \lim_{n\to\infty} \sup_t |F(t) - F_n(t;\omega)| = 0\} .$$

Consider an arbitrary $\varepsilon > 0$ and $\omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3$. There exists $n_1(\omega)$ such that, when $n \geq n_1(\omega)$,

$$\sup_t |F(t) - F_n(t;\omega)| < \frac{\varepsilon T}{2} ,$$

which implies

$$\text{(iii)} \qquad \sup_{0\leq t_k^{(n)}<\tau-T} \left( \frac{F(\tau) - F(t_k^{(n)})}{\tau - t_k^{(n)}} - \frac{F_n(\tau;\omega) - F_n(t_k^{(n)};\omega)}{\tau - t_k^{(n)}} \right) < \varepsilon .$$

And there exists $n_2(\omega)$ such that, when $n \geq n_2(\omega)$,

(iv) $$\sup_{\tau-T\leq t_k^{(n)}<\tau} \left( \frac{F(\tau) - F(t_k^{(n)})}{\tau - t_k^{(n)}} - \frac{F_n(\tau;\omega) - F_n(t_k^{(n)};\omega)}{\tau - t_k^{(n)}} \right) < \varepsilon .$$

Now, by (ii), (iii) and (iv), for $\omega \in \Omega_1 \cap \Omega_2 \cap \Omega_3$ and $n \geq \max\{n_1(\omega), n_2(\omega), n_3(\omega)\}$ ,

$$\frac{F(\tau) - F(t^{(n)}_{k_{\sigma-1}+1})}{\tau - t^{(n)}_{k_{\sigma-1}+1}} - \tilde{f}_n(\tau^-;\omega)$$

$$= \frac{F(\tau) - F(t^{(n)}_{k_{\sigma-1}+1})}{\tau - t^{(n)}_{k_{\sigma-1}+1}} - \frac{F_n(\tau;\omega) - F_n(t^{(n)}_{k_{\sigma-1}+1};\omega)}{\tau - t^{(n)}_{k_{\sigma-1}+1}}$$

$$\leq \sup_{0\leq t_k^{(n)}<\tau} \left( \frac{F(\tau) - F(t_k^{(n)})}{\tau - t_k^{(n)}} - \frac{F_n(\tau;\omega) - F_n(t_k^{(n)};\omega)}{\tau - t_k^{(n)}} \right) < \varepsilon .$$

It follows that

$$\tilde{f}_n(\tau^-;\omega) > \frac{F(\tau) - F(t^{(n)}_{k_{\sigma-1}+1})}{\tau - t^{(n)}_{k_{\sigma-1}+1}} - \varepsilon$$

$$= \frac{1}{\tau - t^{(n)}_{k_{\sigma-1}+1}} \int_{t^{(n)}_{k_{\sigma-1}+1}}^{\tau} f(s)ds - \varepsilon$$

$$\geq f(\tau^-) - \varepsilon .$$

As $\varepsilon$ was arbitrary, and $\omega$ is an arbitrary element of $\Omega_1 \cap \Omega_2 \cap \Omega_3$, this implies

(v) $\qquad \underline{\lim}\, \tilde{f}_n(\tau^-) \geq f(\tau^-)$ a.s. .

Now, because $\tilde{f}_n$ is nonincreasing on $[0,\infty)$, for $0 < x_o < \tau$ and for any $\omega$,

$$\tilde{f}_n(x_o;\omega) \geq \tilde{f}_n(\tau^-;\omega) .$$

Suppose that $\sup_k \delta_k^{(n)} \to 0$. Then by lemma 4.2 and by the continuity of f, for $\omega \in \Omega_1$,

$$\overline{\lim}\, \tilde{f}_n(\tau^-;\omega) \leq \overline{\lim}\, \tilde{f}_n(x_o;\omega) \leq f(x_o) .$$

As $x_o \in (0,\tau)$ was arbitrary, for any $\omega$,

$$\overline{\lim}\, \tilde{f}_n(\tau^-;\omega) \leq f(\tau^-) .$$

This and (v) yield

(vi) $\qquad \lim_{n\to\infty} \tilde{f}_n(\tau^-) = f(\tau^-)$ a.s. .

Also, for $t \in (0,\tau)$ and $\omega \in \Omega_1$, by lemma 4.2,

(vii) $\qquad \lim_{n\to\infty} \tilde{f}_n(t;\omega) = f(t)$ .

Suppose further that, for some $T > 0$,

$$\inf\{t_{k+1}^{(n)} - t_k^{(n)} : 0 \leq t_k^{(n)} < T\} \geq n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \geq 3$, so that, by theorem 4.3.6,

(viii) $\qquad \lim_{n\to\infty} \tilde{f}_n(0) = f(0)$ a.s.

For each n, put

$$\hat{F}_n^*(t) = \begin{cases} \hat{F}_n(t) & , t \neq \tau, \\ \hat{F}_n(\tau^-) & , t = \tau, \end{cases}$$

and

$$f^*(t) = \begin{cases} f(t) & , t \neq \tau, \\ f(\tau^-) & , t = \tau. \end{cases}$$

In view of (vi), (vii) and (viii), there exists $\Omega_4$ such that $P(\Omega_4) = 1$ and, for each $t \in [0,\tau]$ and any $\omega \in \Omega_4$,

$$\lim_{n\to\infty} \hat{F}_n^*(t;\omega) = f^*(t) .$$

From this and lemma 4.5,

(ix) $$P\left[\sup_{0\leq t\leq\tau} |f^*(t) - \hat{F}_n^*(t)| \to 0\right] = P(\Omega_4) = 1 .$$

Now note that, for any $\omega$,

$$\sup_{0\leq t<\tau} |f(t) - \hat{F}_n(t;\omega)| = \sup_{0\leq t<\tau} |f(t) - \hat{F}_n^*(t;\omega)|$$

$$\leq \sup_{0\leq t\leq\tau} |f(t) - \hat{F}_n^*(t;\omega)| .$$

From this and (ix),

$$\text{(x)} \qquad P\left[\sup_{0\le t<\tau} |f(t) - \hat{f}_n(t)| \to 0\right] = 1 .$$

Now, if $\omega \in \Omega_3$ then, as noted before (i), $\hat{f}_n(\tau;\omega) = 0$. But then $\hat{f}_n(t;\omega) = f(t) = 0$ on $[\tau,\infty)$. We thus finally obtain the following result for the case where $\tau$ is finite and known.

Theorem 4.7.1. Suppose that $\sup_k \delta_k^{(n)} \to 0$. Suppose that, for some $T_1 > 0$,

$$\inf\{t_{k+1}^{(n)} - t_k^{(n)}: 0 \le t_k^{(n)} < T_1\} \ge n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \ge 3$ ; and that, for some $T_2 \in (0,\tau)$,

$$\inf\{t_k^{(n)} - t_{k-1}^{(n)}: \tau-T_2 < t_k^{(n)} \le \tau\} \ge n^{-1}(\log n)^s$$

for some $s > 1$ and all $n \ge 3$ . If $\tau < \infty$ and, for each n, $t_{\ell_n}^{(n)} = \tau$ for some $\ell_n$, and if f is continuous on $[0,\tau)$ then

$$P\left[\sup_{0\le t<\infty} |f(t) - \hat{f}_n(t)| \to 0\right] = 1 .$$

Remark

To obtain this result we assumed also that $f(\tau^-) > 0$ but, in view of theorem 4.6.(a), this result is also valid when $f(\tau^-) = 0$ because then f is continuous on $[0,\infty)$.

# 5. Application to renewal processes.

## 5.1. Introduction.

In this chapter we use the now familiar notations $\Delta_k^{(n)}$, $\delta_k^{(n)}$, $\hat{f}_n$, $\tilde{f}_n$ defined in 3.1. Consider the problem stated in chapter 1. We make observations of the residual life of each of n objects put on test in n independent renewal processes which have started long ago in a manner such that these residual lives can be represented by independent identically distributed random variables $(X_i)_1^n$. We want to estimate the distribution function G of the random variable which represents the lifetime of an object of this type. Recall from chapter 1 that if X is a random variable representing the above-mentioned residual lives then X has a distribution function F with density function given by

$$f(t) = \begin{cases} \dfrac{1 - G(t)}{\mu} & , \ t \geq 0 \, , \\ 0 & , \ t < 0 \, . \end{cases}$$

where $\mu = \int_0^\infty s \, dG(s) < \infty$. As mentioned in chapter 1, we consider here only the case where G is continuous at 0; i.e., f is right-continuous at 0. From this assumption we noted in chapter 1 that $\mu = (f(0))^{-1}$ and that $G(t) = 1 - (f(0))^{-1}f(t)$ on $[0,\infty)$.

In this chapter we obtain two results on the nonparametric estimation of G. The first one, considered in theorem 5.3, is on the consistency of the monotone estimator $1 - (\tilde{f}_n(0))^{-1}\tilde{f}_n(t)$ . The second one, considered in theorem 5.4, is a result on the rate of convergence of the monotone estimator $1 - (\tilde{f}_n(0))^{-1}\tilde{f}_n(t)$ as well as the not necessarily monotone estimator $1 - (\hat{f}_n(0))^{-1}\hat{f}_n(t)$ . Both results are obtained from our discussions and results in chapters 3 and 4. Both $\hat{f}_n$ and $\tilde{f}_n$ are computed, by procedures or formulas given in chapter 3, from the observed residual lives $X_1,\ldots,X_n$.

<u>Lemma 5.2</u>. Suppose that $(h_n)_1^\infty$ is a sequence of real-valued functions such that, for some sequence $(\chi(n): n = 1,2,\ldots)$ and some set J containing 0,

$$\chi(n) \sup_{x\in J} |h(x) - h_n(x)| \to 0 .$$

If $0 < |h(0)| < \infty$ then

$$\chi(n) \sup_{x\in J} \left|\frac{h(x)}{h(0)} - \frac{h_n(x)}{h_n(0)}\right| \to 0 .$$

<u>Proof</u>. Clearly, for n so large that $h_n(0) \neq 0$,

$$\left|\frac{1}{h(0)} - \frac{1}{h_n(0)}\right| = \frac{1}{|h_n(0)h(0)|}|h_n(0) - h(0)| .$$

Since h(0) is finite and nonzero, and since $\chi(n)|h(0) - h_n(0)| \to 0$ by hypothesis, it follows that

(i) $\chi(n)\left|\frac{1}{h(0)} - \frac{1}{h_n(0)}\right| \to 0$ as $n \to \infty$ .

Also, by hypothesis,

(ii) $\chi(n) \sup_{x\varepsilon J}|h(x) - h_n(x)| \to 0$ as $n \to \infty$ .

Now, from (i) and (ii),

$$\chi(n) \sup_{x\varepsilon J}\left|\frac{h(x)}{h(0)} - \frac{h_n(x)}{h_n(0)}\right|$$

$$\leq |h(x)|\ \chi(n) \sup_{x\varepsilon J}\left|\frac{1}{h(0)} - \frac{1}{h_n(0)}\right|$$

$$+ \frac{1}{|h_n(0)|}\ \chi(n) \sup_{x\varepsilon J}|h(x) - h_n(x)|$$

$$\to 0 \text{ as } n \to \infty .$$

Theorem 5.3. Let $\tau = \inf\{x: G(x) = 1\} = \inf\{x: F(x) = 1\}$ .

Suppose that $\sup_k \delta_k^{(n)} \to 0$ and that, for some $T_0 > 0$ ,

$$\inf\left\{t_{k+1}^{(n)} - t_k^{(n)}: 0 \leq t_k^{(n)} < T_0\right\} \geq n^{-1}(\log n)^r$$

for some $r > 1$ and all $n \geq 3$ .

(a) If G is continuous on $[0,\infty)$ - $\tau$ finite or not - then

$$P\left[\sup_{0\leq t<\infty}|G(t) - \{1 - (\hat{f}_n(0))^{-1}\hat{f}_n(t)\}| \to 0\right] = 1.$$

(b) Suppose that $\tau < \infty$. If G is continuous on $[0,\tau)$ then, for arbitrary $T_1$ and $T_2$ such that $T_1 < \tau < T_2$,

$$P[\sup|G(t) - \{1 - (\hat{f}_n(0))^{-1}\hat{f}_n(t)\}| \to 0] = 1 ,$$

where the supremum is taken over all t in $[0,T_1] \cup [T_2,\infty)$.

(c) Suppose that $\tau < \infty$ and that G is continuous on $[0,\tau)$. Suppose that, for some $T_3 \in (0,\tau)$,

$$\inf\left\{t_k^{(n)} - t_{k-1}^{(n)}: \ \tau-T_3 < t_k^{(n)} \leq \tau\right\} \geq n^{-1}(\log n)^s$$

for some $s > 1$ and all $n \geq 3$. If, for each n, $t_{\ell_n}^{(n)} = \tau$ for some $\ell_n$ then

$$P\left[\sup_{0\leq t<\infty} |G(t) - \{1 - (\hat{f}_n(0))^{-1}\hat{f}_n(t)\}| \to 0\right] = 1 .$$

<u>Proof</u>. Put h = f; i.e., $\frac{h(x)}{h(0)} = \frac{f(x)}{f(0)} = 1 - G(x)$. For each $n \in \mathbb{N}^+$, put $h_n = \hat{f}_n$ and $\chi(n) = 1$. The conclusions of statements (a) and (b) follow from the results (a) and (b) of theorem 4.6, respectively, and lemma 5.2, with h, $h_n$ and $\chi(n)$ as above. The conclusion of statement (c) follows from theorem 4.7.1 and lemma 5.2. □

<u>Theorem 5.4</u>. Let $||\cdot||_I$ be the supnorm over an interval I.

(a) If f, the sequence of partitions $\left((t_k^{(n)})_{k=1}^{\infty}: n = 1,2,\ldots\right)$, the interval I, and $\chi(n)$ are such that

$$P[\lim_n \chi(n)||f - \hat{f}_n||_I = 0] = 1$$

then

$$P[\lim_n \chi(n)||G - \{1 - (\hat{f}_n(0))^{-1}\hat{f}_n\}||_I = 0] = 1 .$$

(b) Let $\tau = \inf\{x: G(x) = 1\} = \inf\{x: F(x) = 1\}$ . Let I be $[0,\infty)$ or $[0,\tau]$ with $\tau < \infty$. If f, the sequence of partitions $\left((t_k^{(n)})_{k=1}^{\infty}: n = 1,2,\ldots\right)$, and $\chi(n)$ are such that

$$P[\lim_n \chi(n)||f - \hat{f}_n||_I = 0] = 1$$

then

$$P[\lim_n \chi(n)||G - \{1 - (\tilde{f}_n(0))^{-1}\tilde{f}_n\}||_I = 0] = 1 .$$

<u>Proof</u>. Put h = f; i.e., $\frac{h(x)}{h(0)} = \frac{f(x)}{f(0)} = 1 - G(x)$. The conclusion of statement (a) follows from lemma 5.2, with $h_n = \hat{f}_n$ and h as above. The conclusion of statement (b) follows from theorem 3.12 and lemma 5.2, with $h_n = \tilde{f}_n$ and h as above. □

Bibliography.

Ayer, M., H.D. Brunk, G.M. Ewing, W.T. Reid and E. Silverman (1955). "An empirical distribution function for sampling with incomplete information". Ann. Math. Statist., 26, pp. 641-647.

Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk (1972). "Statistical inference under order restrictions". John Wiley and Sons.

Billingsley, P. (1979). "Probability and measure". John Wiley and Sons.

Cox, D.R. (1969). "Some sampling problems in technology". In N.L. Johnson and H. Smith, Jr. (eds.), pp. 506-527. "New developments in survey sampling". Wiley-Interscience.

Feller, W. (1971). "An introduction to probability theory and its applications", vol. II, 2nd ed. . John Wiley and Sons.

Révész, P. (1968). "The laws of large numbers". Academic Press.

Révész, P. (1972). "On empirical density function". Periodica Mathematica Hungarica, 2, pp. 85-110.

Remerciements.

Mes premiers remerciements vont à mon directeur de thèse, B.B. Winter, pour le sujet de cette thèse, mais surtout pour sa patience "angélique" lors de la première lecture, de la seconde, de la troisième... de mes différents brouillons. La recherche conduisant à ce travail fut facilitée par l'aide financière offerte par le Conseil de recherches en sciences naturelles et en génie et par l'Université d'Ottawa. La frappe rapide et juste de D. Iscoe ne restera sûrement par sans échos et je ne suis sans doute pas le dernier à en témoigner.

Je garde le souvenir du professeur V. Linis qui, par ses problèmes hebdomadaires, m'a décidé à entrer au département de mathématiques. Finalement, je ne saurais oublier Dieu qui ne jouant plus aux dés me les a refilés.