# CONTRIBUTIONS TO STRONG APPROXIMATIONS IN TIME SERIES WITH APPLICATIONS IN NONPARAMETRIC STATISTICS AND FUNCTIONAL LIMIT THEOREMS

Getulio Borges da Silveira Filho

Submitted for the degree of Ph. D. in Statistics

March 1991

London School of Economics and Political Sciences

## ABSTRACT

This thesis is concerned with applications in probability and statistics of approximation theorems for weakly dependent random vectors. The basic approach is to approximate partial sums of weakly dependent random vectors by corresponding partial sums of independent ones. In chapter 2 we apply such a general idea so as to obtain a almost sure invariance principle for partial sums of $\mathbb{R}^d$–valued absolutely regular processes. In chapter 3 we apply the results of chapter 2 to obtain functional limit theorems for non–stationary fractionally differenced processes. Chapter 4 deals with applications of approximation theorems to nonparamatric estimation of density and regression functions under weakly dependent samples. We consider $L_1$–consistency of kernel and histogram density estimates. Universal consistency ofthe partition estimates of the regression function is also studied. Finally in chapter 5 we consider necessary conditions for $L_1$–consistency of kernel density estimates under weakly dependent samples as an application of a Poisson approximation theorem for sums of uniform mixing Bernoulli random variables.

## ACKNOWLEDGEMENTS

I am greatly indebted to my thesis supervisor Prof. P. M. Robinson for several reasons. Some of them are worth mentioning here. First of all, without his encouragement and wise advice this work would take much longer to be completed. Also, a number of talks to him proved to be extremely valuable for improving my understanding of statistics. Finally, I must thank Peter for not complaining too much about my idyosincrasies.

This work has been funded by the Brazilian government through a scholarship provided by the Conselho Nacional de Desenvolvimento Cientifico e Tecnologico (CNPq). It is also partiaaly based on research funded by the Economic and Social Research Council (ESRC) ref: R000231441. The typing of this thesis was carefully performed by Sue Kirkbride to whom I express my gratitude.

Finally, I am very grateful to my wife Lucia Goncalves. Her support and patience throughout our staying in the United Kingdom proved to be most fundamentall for the realization of this work.

## Table of Contents

# CHAPTER 1

## 1.1 Scope of the thesis

Generally speaking this thesis is concerned with some applications of the general idea of directly approximating partial sums of a sequence of dependent random vectors by sums of independent ones. More precisely let $X_1, \ldots, X_n$ be dependent random vectors. The common element in most applications we deal with in this work is that we approximate $X_1+\ldots+X_n$ by $AX_1+\ldots+AX_n$, where the AX's are independent. We exploit the fact that if our approximation is good enough in a probabilistic sense then asymptotics for the original process will follow from corresponding results for the partial sums of independent random vectors.

In a sense we may regard the above described approach as a two step procedure. In the first step we filter out dependence features. In the second one we only have to cope with independent random vectors. The effectiveness of such a procedure depends heavily on the goodness of the approximation performed in the first step.

Now let us consider a more primitive issue. Suppose that X and Y are are random vectors taking values in $\mathbb{R}^c$ and $\mathbb{R}^d$ respectively. Assume that the "degree of dependence" between X and Y in a suitable "scale" is no greater than $\alpha(X,Y)$ (we formalize the concepts of "scale " and "degree of dependence" in section 1.2), where the more dependent two vectors are the higher the corresponding $\alpha$ must be. A number of authors have dealt with the problem of constructing an approximating vector AY such that AY is independent of X and

$$P\{ \mid Y-AY \mid \geqslant \varepsilon \} \leqslant f(\alpha),$$

where f(.) is a nondecreasing real function. Here f(.) depends on the particular "measure of dependence" one is considering. Also in some cases $\epsilon$ depends on some characteristics of the distribution of Y. Approximations of this style will be our main tool for implementing the above mentioned two step procedure.

If on one hand it seems obvious that such a method is far from being optimal (loosely speaking), on the other it leads, in some cases, to much simpler arguments than those needed in a "one step" approach. An example of the non-optimal character of the method we adopt here is shown in Yoshihara ( 1978, page 327 ).

When dealing with stationary stochastic processes an alternative approach might be used. It is based on the so called Gordin decomposition which enables us to write a partial sum of a stochastic process as a sum of stationary ergodic martingale differences plus a random term of small probability order. We refer the reader to Hall and Heyde (1980) for a comprehensive discussion on this issue.

## 1.2 Mixing Processes

Let $\{X_t\}$ be a $\mathbb{R}^d$-valued strictly stationary process, defined on a probability space $(\Omega,\mathcal{A},P)$. Let $\mathcal{B}_t$ be the $\sigma$-field generated by $\{ X_s , s \leqslant t \}$, which we denote by $\mathcal{B}_t = \sigma\{ X_s , s \leqslant t \}$. Also let us put $\mathcal{F}_t = \sigma\{ X_s , s \geqslant t \}$ and $\mathcal{B}_\infty = \sigma\{ \cup_t \mathcal{B}_t \}$. We say that $\{X_t\}$ is a regular process if the $\sigma$-field

$$\mathcal{B}_{-\infty} := \cap \{ \mathcal{B}_t , t \in \mathbb{R} \}$$

is trivial in the sense that it contains only events with probability zero or one.

A simple example of a non-regular stationary process is the following. Let $\{ W, Z_t, t= \dots -1, 0, 1, \dots \}$ be a collection of iid random variables. Consider the $\mathbb{R}^2$-valued stationary process $\{X_t\} := \{ (W,Z_t) \}$. We can easily see that for this case $\mathcal{B}_{-\infty} \supset \sigma\{ W \}$. Thus if W is not a constant then $\{X_t\}$ is not regular.

The above example ( though very artificial ) spots the kind of process one is ruling out when restricting the class of stationary processes to the class of regular ones. Namely, when considering regular processes the present evolution of the process must be approximately independent of its ( remote ) past history. In the above example this is obviously false. An equivalent characterization of regular processes is given by the following

<u>Proposition</u> ( See Thm. 17.1.1 in Ibragimov and Linnik(1970) ) In order that a stationary process be regular, it is necessary and sufficient that, for all $B \in \mathcal{B}_\infty$,

$$\lim \sup\{ |P(AB) - P(A)P(B)| , A \in \mathcal{B}_t \} = 0, \text{ as } t\to-\infty . \tag{1.1}$$

It is not difficult to see that regularity implies ergodicity for the process $\{X_t\}$. In this work we will not consider processes which are merely regular. In fact we will deal with processes for which (1.1) holds in "more uniform" senses. Let us define for any two $\sigma$-fields $\mathcal{B}$ and $\mathcal{F}$

$$\alpha(\mathcal{B},\mathcal{F}) = \sup\{ |P(AB) - P(A)P(B)|, A \in \mathcal{B}, B \in \mathcal{F} \},$$

$$\beta(\mathcal{B},\mathcal{F}) = 1/2 \sup\{ \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_iB_j) - P(A_i)P(B_j)| \}$$

where the supremum above is taken over all possible $\mathcal{B}$-measurable ( $\mathcal{F}$-measurable, resp. ) partitions $\{ A_i , 1\leqslant i\leqslant I \}$ ( $\{ B_j ,1\leqslant j\leqslant J \}$, resp. ) of $\Omega$ ( if $\mathcal{G}$ is a $\sigma$-field of subsets of $\Omega$ a $\mathcal{G}$-measurable partition of $\Omega$ is a partition whose elements belong to $\mathcal{G}$ ). Finally we define

$\varphi(\mathcal{B},\mathcal{F}) = \sup\{ |P(B|A) - P(B)|, B \in \mathcal{F}, A \in \mathcal{B} \}.$

Now we can define the first kind of mixing process we consider in this work

Definition 1.1 Let $\{X_t\}$, $t \in Z$ be a stationary process. We say that $\{X_t\}$ is a strong mixing ( or completely regular or $\alpha$-mixing ) process with mixing weights $\{\alpha_k\}$ if

$\alpha_k := \alpha(\mathcal{B}_i , \mathcal{F}_{i+k}) = o(1)$ □

Strong mixing processes were introduced in Rosenblatt(1956) and have appeared in a huge number of works since then. In this work, strong mixing is the weakest type of regularity we deal with. An important property associated with strong mixing processes is the following:

Proposition ( Davydov's inequality ) Suppose that X and Y are random variables which are $\mathcal{B}$- and $\mathcal{F}$-measurable, respectively, and that $\| X \|_p =1$, $\| Y \|_q =1$, for some $1< p,q \leqslant\infty$, $p^{-1} + q^{-1} \leqslant 1$ ( here and throughout this work we denote $\| X \|_p = \{ E|X|^p \}^{1/p}$ ). Then

$$| E(XY) - (EX)(EY) | \leqslant 8 \{ \alpha(\mathcal{B},\mathcal{F}) \}^{1/s},$$

where $s^{-1} + p^{-1} + q^{-1} = 1$.

Proof: See Davydov(1968).

The next type of mixing process we consider is required to satisfy a stronger condition than the $\alpha$-mixing processes.

Definition 1.2 Let $\{ X_t , t= \ldots, -1, 0, 1, \ldots \}$ be a stationary process. We say that $\{X_t\}$ is a absolutely regular ( or weak Bernoulli or $\beta$-mixing ) process with mixing weights ( or mixing sequence ) $\{\beta_n\}$ if

$$\beta_n := \beta( \mathcal{B}_p, \mathcal{F}_{p+n} ) = o(1).$$

Absolutely regular processes were first studied in Volkonskii and Rozanov(1961). They attribute the definition of absolutely regular processes to Kolmogorov. It is not difficult to see that $\alpha_k \leqslant \beta_k$ and thus any weak Bernoulli process is strong mixing. The following result provides an alternative definition of the $\beta$ weights.

Proposition( Volkonskii and Rozanov (1961) )

$$\beta_k = \sup_p E[ \sup \{ | P(B|\mathcal{B}_p) - P(B) |, B \in \mathcal{F}_{p+k} \} ]. \square$$

Absolutely regular processes are in general a bit easier to be handled than strong mixing ones. In particular when one is dealing with asymptotics for U-statistics, proofs and conditions are much neater under absolute regularity than under strong mixing ( See e.g. Yoshihara(1976) and Denker and Keller(1983) ). Also, when using the two-step procedure described in section 1.1 the existing approximations for absolutely regular processes are better than their strong mixing counterparts. We carry out a more detailed discussion on this matter in chapters 2 and 4.

A further type of stochastic process that will be considered in this work is represented by the class of linear processes. We say that a process $\{X_t\}$ is linear if it can be represented as

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} , \qquad (1.2)$$

where $\Sigma(a_k)^2 < \infty$ and $\{\varepsilon_t\}$ is a sequence of iid random variables with $E\varepsilon_1=0$ and $Var(\varepsilon_1)<\infty$. The assumptions of independence and/or equality of distribution for the sequence $\{\varepsilon_t\}$ are sometimes relaxed. The equality above is in quadratic mean i.e. $X_t$ is the limit in quadratic mean of the partial sums of the series in (1.2).

At this point a question which one could naturally make is whether or not linear processes are strong mixing or absolutely regular.The following result estabilishes sufficient conditions for linear processes being absolutely regular.

<u>Proposition</u> Let $\{X_t\}$ be as in (1.2) above. Suppose that the following conditions hold true.

(i) $\int | g(v-u) - g(u) | dv \leqslant C |u|$, all $u$,

where $g(.)$ is the pdf of $\varepsilon_t$.

(ii) $\Sigma |a_k| < \infty$ and $\Sigma_k a_k z^k \neq 0$, for all $|z|\leqslant 1$.

(iii) $E|\varepsilon_1|^p < \infty$, for some $p>0$.

(iv) $\sum_{k=1}^{\infty} \{ \sum_{j=k}^{\infty} |a_j| \}^{p/(1+p)} < \infty$.

Then $\{ X_t \}$ is weak Bernoulli with mixing weights such that

$$\beta_n = O \left[ \sum_{k=n}^{\infty} \{ \sum_{j=k}^{\infty} |a_j| \}^{p/(1+p)} \right] .$$

<u>Proof</u>: See Phan and Tran(1985).

Similar results hold for strong mixing processes as well (See Gorodetskii(1977b)).

However there do exist linear processes which are not strong mixing. We will deal with such a kind of processes in chapter 3. Let us just give an example of a Gaussian linear process which is not strong mixing.

Take $a_k = k^{-\theta}$ and $\varepsilon_1 = N(0,1)$ in (1.2), where $1/2 < \theta < 1$. It is not difficult to see that $\mathrm{Var}(X_1 + \dots + X_n) = O(n^{3-2\theta})$. Now by theorem 18.1.1 in Ibragimov and Linnick(1970) we know that if $\{X_t\}$ is strong mixing and for a nondegenerate Z we have

$$(b_n)^{-1} \{ X_1 + \dots + X_n \} \Rightarrow Z, \tag{1.3}$$

then the law of Z is necessarily stable ( here and throughout this work the symbol $\Rightarrow$ stands for weak convergence ). Further if the exponent of the law of Z is $\alpha$ ( i.e. the characteristic function of Z is proportional to $\exp(-c|z|^{\alpha})$, for some $c>0$, $0<\alpha\leqslant 2$ ) then $b_n = h(n)n^{1/\alpha}$, where h(.) is a slowly varying function in the sense of Karamata. Now it is obvious that (1.3) holds true with $b_n = \mathrm{Var}(X_1 + \dots X_n)^{1/2}$ and $Z \sim N(0,1)$. Therefore we should have

$$O(n^{3-2\theta}) = \mathrm{Var}(X_1 + \dots + X_n) = n(h(n))^2,$$

which cannot hold since $\theta<1$. Thus $\{X_t\}$ is not strong mixing.

Finally the last class of mixing processes we will consider in this thesis is represented by the uniform mixing processes.

<u>Definition 1.3</u> Let $\{ X_t , t= \dots, -1, 0, 1, \dots \}$ be a stationary process. We say that $\{X_t\}$ is a uniform mixing ( or $\varphi$-mixing ) process with mixing weights ( or mixing sequence ) $\{\varphi_n\}$ if

$\varphi_n := \varphi( \mathcal{B}_p, \mathcal{F}_{p+n} ) = o(1).$

The following result, due to Ibragimov, will be repeatedly used in chapter 5.

<u>Proposition</u> ( Ibragimov's inequality ) Suppose that X and Y are random variables which are $\mathcal{B}$- and $\mathcal{F}$-measurable, respectively, and that $\|X\|_p = 1 = \|Y\|_q$, where $1 \leqslant p,q \leqslant \infty$ and $p^{-1} + q^{-1} = 1$. Then

$$| E(XY) - (EX)(EY) | \leqslant 2\{ \varphi(\mathcal{B},\mathcal{F}) \}^{1/p}.$$

## 1.3 <u>Plan of the thesis</u>

In chapter 2 we prove an almost sure invariance principle ( i.p. )for partial sums of absolutely regular processes. The rate of convergence depends on moment assumptions and on the rate of decay for the mixing sequence. However it is faster than $n^{1/2}$ in the cases we consider.

In chapter 3 we provide an application of the i.p. obtained in chapter 2 in the study of a non-stationary fractional model. Weak convergence of this model is discussed under a variety of non-iid assumptions for the sequence of underlying innovations.

In chapter 4 we deal with the subject of $L_1$-convergence for some nonparametric estimates of the regression and density functions. We consider sufficient conditions for both weak and strong $L_1$-consistency for those estimates under mixing assumptions on the samples.

In chapter 5 we study necessary conditions for $L_1$-consistency of kernel density estimates under uniform mixing samples. Such conditions turn out to be the same as in the iid case.

## 1.4 Some conventions

In this small section we set some notations and conventions for the rest of the thesis.

(1) We denote by $[\![x]\!]$ or [[x]] the integer part of x.

(2) $X_n \Longrightarrow X$ stands for the sequence of processes $\{X_n\}$ weakly converges to X. We obviously use the same notation for convergence of distributions of random vectors.

(3) We adopt in some places Vinogradov's symbol $\ll$ instead of big "Oh". That is $a_n \ll b_n$ has the same meaning of $a_n = O(b_n)$.

(4) Given a sequence of random vectors $\{X_n\}$ $1 \leq n \leq N \leq \infty$ we denote by $\sigma\{X_n, 1 \leq n \leq N\}$ the $\sigma$-field generated by $\{X_n\}$, that is the smallest $\sigma$-field $G$ which makes all the X's $G$-measurable.

(5) The $L_p$ norm of a vector v in $\mathbb{R}^d$ is denoted by $\|v\|_p$ or $|v|_p$. If p is equal to 2 we often omit the subscript. If X is a random vector and $1 \leq p < \infty$ we denote $\|X\|_p = [E|X|^p]^{1/p}$. Also $\|X\|_\infty$ stands for the essential supremum of $|X|$.

The rest of the notations and conventions adopted in this work are standard in the statistical literature.

# CHAPTER 2

## 2.1 Introduction

This chapter is devoted to the study of an almost sure approximation theorem for sums of $\mathbb{R}^d$-valued stationary absolutely regular process. More precisely given $\{X_n\}$ as above we will be interested in finding a sequence $\{Z_n\}$ of i.i.d. Gaussian random vectors such that

$$\left| \sum_{k=1}^{n} X_k - Z_k \right| = o(n^{1/t}) \qquad \text{a.s.,}$$

provided $E|X_1|^t < \infty$. There exist some similar results in the literature (see e.g. Dehling and Philipp (1982), Dehling (1983)) but in general they are mainly concerned with the issue of constructing $\{Z_n\}$ so that

$$\left| \sum_{k=1}^{n} X_k - Z_k \right| = o(n^{\frac{1}{2}-\epsilon}) \qquad \text{a.s.}$$

for some $\epsilon > 0$. Our study, in a sense, goes in the opposite direction: we fix our rate of convergence and then try to find out for which sequences $\{X_k\}$ such rate is achievable.

Notice that the latter approximation theorem yields all the classical invariance principles (law of iterated logarithm and functional central limit theorems). A naive question is then the following: why to specify a (somewhat) "much" smaller rate than $n^{\frac{1}{2}}$? In Chapter Three we present an example where such small rate is indeed necessary.

The rest of the chapter is divided as follows. In section 2 we introduce some concepts on invariance principles and discuss some previous work. In section 3 we state our main result and prove an important corollary to it. As the proof of our main result is a bit long we present some guidelines to it in section 4.

Finally in section 5 we discuss the issue of constructing random vectors in "rich" probability spaces (p-spaces). The proof of our main result is divided into appendices A, B and C.

## 2.2 Invariance Principles

The designation invariance principle was coined by Erdos and Kac (1946) in a weak convergence environment. They wanted to evaluate limit distributions such as

$$\lim_n P\{n^{-\frac{1}{2}} \max_{1 \leq k \leq n} S_k \leq y\},$$

where $\{X_n\}$ is a sequence of i.i.d. r.v. and $S_n = X_1 + \ldots + X_n$. The authors realized that for a particular sequence $\{X_n\}$ such limit distributions could be obtained through available analytical methods. In other words when $X_1$ has a specific distribution F one can evaluate in a (somewhat) straightforward manner the above limit. If we could show that the above limit were <u>invariant</u> for all F in a broad class then the problem of finding out the limit distribution above could be reduced to studying such limits in a particular case.

Strassen (1964) introduced the concept of almost sure invariance principle for partial sums of i.i.d. r.v.'s. We can informally assess the difference between Strassen's and Erdos and Kac's (or its generalizations) concepts of invariance principle (i.p.) by means of the usual central limit theorem (an i.p. in Erdos and Kac sense). Let $\{X_n\}$ be i.i.d. r.v.'s and suppose that $EX_n = 0$, $EX_n^2 = 1$ and $\{Z_n\}$ be i.i.d. r.v.'s with $Z_1 \sim N(0,1)$. The central limit theorem reads

$$\lim_n |P\{X_1 + \ldots + X_n \leq \sqrt{n}\, y\} - P\{Z_1 + \ldots + Z_n \leq \sqrt{n}\, y\}| = 0.$$

Now (as convergence in probability implies convergence in distribution) one could ask whether or not such a result is consequence of a stronger theorem ensuring the existence of such a $\{Z_n\}$ that

$$|X_1 + \ldots + X_n - (Z_1 + \ldots + Z_n)| = o_p(n^{\frac{1}{2}}) \ ,$$

or perhaps

$$|X_1 + \ldots + X_n - (Z_1 + \ldots + Z_n)| = o(n^{\frac{1}{2}}) \quad \text{a.s.}$$

Strassen i.p. provides a partial answer to such questions (actually it is not possible, in general, to obtain such a rate of convergence assuming only the existence of second moments, see Major (1976)). He proved that we can construct a sequence $\{Z_n\}$ as above such that

$$|X_1 + \ldots + X_n - (Z_1 + \ldots + Z_n)| = o((n \log \log n)^{\frac{1}{2}}) \quad \text{a.s.}$$

More precisely

<u>THEOREM (Strassen's invariance principle)</u>. Let $\{X_n\}$ be a sequence of i.i.d. r.v.'s such that $EX_1 = 0$ and $EX_1^2 = 1$. Assume that $\{X_n\}$ is defined on a p-space $(\Omega,F,P)$. Then we can construct a p-space $(\Omega_1,F_1,P_1)$ and two sequences of i.i.d. r.v.'s $\{\tilde{X}_n\}$ and $\{Z_n\}$ defined on it such that

(i) $\{\tilde{X}_n\}$ is a copy (in distribution) of $\{X_n\}$

(ii) $Z_1 \sim N(0,1)$

(iii) $\left|\sum_{k=1}^{n} \tilde{X}_k - Z_k\right| = o((n \log \log n)^{\frac{1}{2}})$ a.s.

We refer the reader to section 2.5 for a discussion on the need of redefining $\{X_n\}$ on a new p-space. From now on, for simplicity, when stating almost sure i.p.'s we will typically say "we can construct a sequence $\{Z_n\}$ ... such that

$$|X_1 + \ldots + X_n - (Z_1 + \ldots + Z_n)| = o(q(n)) \quad \text{a.s.".}$$

The reader should keep in mind that such a construction holds valid in a rich enough p-space where a copy of $\{X_n\}$ is defined on.

Strassen obtained as a by-product of his i.p. a law of iterated logarithm. Now, the rate obtained in (iii) above is not good enough even to imply the usual C.L.T. However we can obtain better rates of convergence provided higher order moments are assumed to exist. Indeed, from the well known articles of Komlos, Major and Tusnady (KMT) we obtain

<u>THEOREM (Komlos, Major, Tusnady (1976), Major (1976)).</u> Let $\{X_n\}$ be a sequence of i.i.d. r.v.'s such that $E|X_1|^p < \infty$ for some $p > 2$, $EX_1 = 0$ and $EX_1^2 = 1$. Then we can construct a sequence $\{Z_n\}$ of i.i.d. r.v.'s with $Z_1 \sim N(0,1)$ such that

$$|X_1 + \ldots + X_n - (Z_1 + \ldots + Z_n)| = o(n^{1/p}) \quad \text{a.s.} \qquad \square$$

Furthermore, the following result, due to Breiman tells us that the rate in KMT theorem is the best possible.

<u>THEOREM (Breiman (1967)).</u> Let $\{X_n\}$ and $\{Z_n\}$ be sequences of i.i.d. r.v.'s such that $EX_1 = 0$, $EX_1^2 = 1$, $Z_1 \sim N(0,1)$. Then

$$\limsup_n \left|\sum_{k=1}^{n} X_k - Z_k\right|/n^{1/p} = +\infty \qquad \text{a.s.,}$$

provided $E|X_1|^p = +\infty$. $\square$

KMT has been recently generalized by Einmahl (1989) to sequences of $\mathbb{R}^d$-valued i.i.d. random vectors. We refer the reader to Csorgo and Revesz (1981) for a textbook exposition on strong invariance problems related to sequences of i.i.d. r.v.'s.

When the assumption of independence is relaxed to weak dependence the rates are not, so far, as good as in KMT theorem.

The first comprehensive study on almost sure i.p.'s for non-i.i.d. situation was done in Philipp and Stout (1975). Previous results on non-i.i.d. processes were in most of the cases related to the law of iterated logarithm (see e.g. Heyde and Scott (1973), Reznik (1968), Strassen (1967)).

Philipp and Stout (1975) exploited the fact that the block sums of weakly dependent random variables behave approximately as martingale differences, to which Skorohod embedding theorem can be applied (see Skorohod (1965), Sawyer (1967), Hall and Heyde (1980)). In this way they proved a number of i.p.'s in a variety of dependence situations.

However as pointed out in Berkes and Philipp (1979) the generalization of such approach to random vectors is far from easy. Berkes and Philipp (1979) proposed a new method which works in any number of dimensions. Their method consists of approximating block sums of weakly dependent random vectors by independent random vectors. Their basic approach was subsequently refined in Dabrowski (1982), and Bradley (1983).

From Bradley (1983) we learn

THEOREM (Bradley's approximation). Suppose X and Y are random vectors taking their values in $R^m$ and $R$, respectively; suppose U is a uniform-[0,1] r.v. independent of (X,Y), and suppose q and $\gamma$ are positive numbers such that $q < \|Y\|_\gamma < \infty$. Then there exists a r.v. $Y^* = f(X,Y,U)$ where f is a measurable function from $R^m \otimes R \otimes [0,1]$ into $R$ such that

(i) $Y^*$ is independent of X

(ii) The probability distributions of $Y^*$ and Y are identical, and

(iii) $P\{|Y^* - Y| > q\} < 18(\|Y\|_\gamma/q)^{\gamma/(2\gamma+1)}\alpha^{2\gamma/(2\gamma+1)}$,

where

$$\alpha \quad = \quad \alpha(\sigma(X),\ \sigma(Y)), \quad \text{and}$$

for any pair of $\sigma$-algebras $(\mathcal{A},\ \mathcal{B})$ we define

$$\alpha(\mathcal{A},\mathcal{B}) \quad = \quad \sup\{|P(A \cap B) - P(A)P(B)|,\ A \in \mathcal{A},\ B \in \mathcal{B}\}. \qquad \square$$

Actually Bradley's theorem is a bit more general. Further, it can be easily generalized for Y taking values in $\mathbb{R}^d$. Let us define

$$\beta(\mathcal{A},\mathcal{B}) \quad = \quad \frac{1}{2} \sup \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)|$$

where the sup above is taken over all pairs of partitions $\{A_1,\ldots,A_I\}$ and $\{B_1,\ldots,B_J\}$ of $\Omega$ such that each $A_i \in \mathcal{A}$ and each $B_j \in \mathcal{B}$.

Bradley's result depends on a relation between $\alpha(\mathcal{A},\mathcal{B})$ and $\beta(\mathcal{A},\mathcal{B})$ and the following

THEOREM (Berbee (1979)). Suppose X and Y are r.v.'s taking their values in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, and suppose U is a uniform-[0,1] r.v. independent of (X,Y). Then there exists a $\mathbb{R}^n$-valued r.v. $Y^* = f(X,Y,U)$, where f is a measurable function from $\mathbb{R}^m \otimes \mathbb{R}^n \otimes \mathbb{R}$ into $\mathbb{R}^n$, such that

(i) $Y^*$ is independent of X

(ii) The probability distributions of Y and $Y^*$ are identical, and

(iii) $P\{Y^* \neq Y\} \le \beta(\sigma(X),\ \sigma(Y))$. $\square$

Berbee's theorem will play an important role in our theorem 2.1. It will allow suitable block sums of weakly dependent random vectors to be approximated by independent ones.

Our final remark in this section is related to the general idea of proving limit theorems for weakly dependent sequences by a direct approximation to suitable independent ones. It is very unlikely that this method cannot, in general, be outperformed by another one which takes into account particularities of the dependence structure. However, the simplicity and flexibility of such a principle makes it extremely appealing.

## 2.3 Main Result

In this section we are going to state a a.s. i.p. for sums of weakly dependent stationary random vectors. We are going to consider only absolutely regular processes with polynomial sequence of mixing weights. The methodology employed in our proof could be adopted when establishing a similar result for strong mixing processes. However, in that case we would need stronger assumptions on the relations between moment conditions and the rate of decay in the (strong) mixing sequence. Indeed a close look at the proof of our theorem 2.1 reveals that if instead of Berbee's theorem one uses Bradley's approximation then no further modification in the structure of the proof will be necessary. The strengthening of the assumptions on the mixing weights and the moment conditions is necessary because Bradley's approximation yields worse errors than Berbee's one (as we could expect since strong mixing is a weaker form of dependence than absolute regularity).

Let us now state our main theorem in this chapter

THEOREM 2.1. Let $\{X_n\}$ be a zero mean stationary absolutely regular process taking values in $\mathbb{R}^d$. Let $\{\beta_k\}$ be the sequence of mixing weights. Assume that

(a) $\quad \beta_k = k^{-\nu}$ ,

where $\nu > 0$. Suppose that for some $\lambda > 1$, $0 < \theta < 1$ we have

(b) $\quad E|X_1|^{2\lambda(1+\theta)} < \infty,$

where $\lambda$, $\theta$ and $\nu$ satisfy

(c) $$\theta + \frac{1}{\nu+1} + \frac{1}{2\lambda} < \frac{1}{2}$$

(d) $$\frac{1+2\theta}{\lambda(1+\theta)} + \frac{2\lambda(1+\theta)}{\nu+2\lambda(1+\theta)} + \frac{1}{(\nu+1)\lambda(1+\theta)} < 1$$

(e) $$\frac{2(1+\theta)}{1-\theta} < \nu$$

(f) $$2(1+\theta)\frac{\lambda}{\lambda-1} < \nu \ .$$

Then enlarging the original p-space, if necessary, we can define a copy of $\{X_n\}$ (which we will also denote by $\{X_n\}$) and a sequence of independent zero mean Gaussian random vectors $\{Z_n\}$ such that

(i) For each $m \geq 1$, $\{Z_k, 2^m < k \leq 2^{m+1}\}$ are i.i.d.

(ii) If

$$\Gamma = EX_1X_1^T + \sum_{k>1} EX_1X_k^T + EX_kX_1^T$$

is a finite positive definite matrix then

$$\|\mathrm{Cov}\,(Z_{2^m}) - \Gamma\| = 0(2^{-m\tau}), \quad \text{where}$$

$\tau > \theta/2(1+\theta)$.

(iii) $$\left|\sum_{k=1}^{n} X_k - Z_k\right| = o(n^{1/2(1+\theta)}) \quad \text{a.s.}$$

<u>Proof</u>. See Appendix 2-C. □

Let us now make a few comments on theorem 2.1. First of all, it is somewhat unusual to impose a moment condition characterized by two parameters ($\lambda$ and $\theta$). A more common statement of a moment condition would be $E|X_1|^{2+\delta} < \infty$. We have chosen the two-fold parameterization basically for three reasons.

Firstly it provides a sort of location ($\theta$) vs. scale ($\lambda$) interpretation for the order of the required moment.

Secondly, conditions (e) and (f) allows one to evaluate the effect, in a somewhat separate manner, of the joint relations between location and degree of dependence, and scale and degree of dependence (note $\theta < 1$).

Finally, adopting a two-parameter characterization provides insights on the effect of weak dependence by a direct comparison to the i.i.d. case. In other words, it is known that if $\{X_n\}$ were i.i.d. then it would be possible to construct $\{Z_n\}$ i.i.d. such that $Z_n \sim N(0,\mathrm{cov}(X_1))$ and

$$\left| \sum_{k=1}^{n} X_k - Z_k \right| = o(n^{1/2(1+\theta)}) \qquad \text{a.s.},$$

provided $E|X_1|^{2(1+\theta)} < \infty$ (see Berger (1982), Einmahl (1987 and 1989)). What our result above hints (note I am not being conclusive here) is that in order that such a rate be achieved we need to impose the existence of higher order moments. It goes without saying that such a "hint" is for absolutely regular processes with polynomial mixing rate only.

A second point worth making is related to conditions (c), (d), (e) and (f) as a whole. A close look at the proof of our results indicates that such conditions are so imposed that a suitable blocking procedure exist. On this respect, it may seem that condition (c) represents a structural limitation of our method of proof as far as the range of $\theta$ is concerned. In other words it may seem that, on using our method of proof we can only obtain errors of order at most $n^{1/3}$ ($\theta = \frac{1}{2}$), no matter how large $\nu$ and $\lambda$ are. This is not quite true. I will not go into details but what (c) actually represents is a feasibility condition for two inequalities. Namely

$$\gamma < \frac{1}{2}\,\frac{\lambda-1}{\lambda(1+\theta)} ,$$

and

$$\gamma > \frac{\nu\theta + (1+\theta)}{(\nu+1)(1+\theta)} ,$$

where $\gamma$ is a parameter in our blocking procedure. Now, the second inequality is structural, that is it must hold valid whichever blocking one chooses. The first one, on the other hand, is not structural. It may be possible to replace the upper bound in the first inequality by a more convenient one. In any case, we will not deal with that issue in this work.

Nonetheless, we would like to point out that there is a structural bound for $\theta$. When $\{X_n\}$ are i.i.d. one can only show (iii) by the method we have adopted (which was first used in Einmahl (1987) for $\theta < 1$.

Finally, we would like to make a brief comment on a methodological aspect of the proof of Theorem 2.1. Accurate probability inequalities play an important role in most works on asymptotics in statistics or probability. In our particular case good rates of convergence in the Marcinkiewicz–Zygmund law of large numbers (see e.g. Chow and Teicher (1988) Theorem 5.2.2) turned out to be essential for Theorem 2.1. Namely good estimates for quantities like

$$a_n = P\{|\sum_{k=1}^{n} X_k| > n^t\} ,$$

where $t > \frac{1}{2}$ and $\{X_n\}$ is weakly dependent, proved to be necessary.

In works where such kind of estimates are needed the commonest way of bounding $a_n$ is based on Markov inequality and a moment inequality. Let us assume

$t = 1$ for simplicity. $a_n$ can be bounded from above by

$$b_n = n^{-u} E \left| \sum_{k=1}^{n} X_k \right|^u .$$

Now, provided that some moment and mixing conditions (which obviously include $E|X_1|^u < \infty$) hold true (see e.g. Roussas (1988), Roussas and Ioannides (1987), Yokoyama (1980)) it can be shown that $b_n$ behaves like its iid counterpart, that is

$$b_n = O(n^{-u/2}).$$

However, it is well known that better rates of convergence for $a_n$ can be obtained under iid assumptions. From Theorem IX-27 in Petrov (1975) we obtain

$$a_n = o(n^{-u+1}),$$

provided $E|X_1|^u < \infty$ and $EX_1 = 0$. Such estimate is better than the one derived through moment inequalities provided $u > 2$.

In the proof of theorem 2.1 we estimate $a_n$ by means of an easy generalization of Petrov's result and Berbee's theorem (see Lemmas 2A1, 2A2, 2A3, 2A4).

Now, theorem 2.1 yields the following.

<u>Corollary 2.1</u>: Let $\{X_n\}$ be as in theorem 2.1. Then the conclusion of theorem 2.1 holds true with $\{Z_n\}$ replaced by $\{\tilde{Z}_n\}$ such that

(i) $\{\tilde{Z}_n\}$ are iid.

(ii) $\{\tilde{Z}_n\} \sim N(0,\Gamma)$.

<u>Proof</u>. It suffices to show that

$$\left| \sum_{k=1}^{n} Z_k - \tilde{Z}_k \right| = o(n^{1/2(1+\theta)}) \quad \text{a.s.},$$

which in turn is a consequence of

$$\sum_{m=1}^{\infty} P\{ \max_{2^m<k<2^{m+1}} | \sum_{k=1+2^m} Z_k - \tilde{Z}_k | > \epsilon\, 2^{m/2(1+\theta)}\} < \infty,$$

all $\epsilon > 0$ (see the proof of theorem 2.1). Assume with no loss in generality that $\Gamma = I$. By virtue of (ii) in theorem 2.1 we can assume w.l.o.g. that $\Gamma_k :=$ $\mathrm{cov}(Z_k)$ is positive definite for all k. Define $\tilde{Z}_k = \Gamma_k^{-\frac{1}{2}} Z_k$. Kolmogorov's inequality (multivariate) yields

$$P\{ \max_{2^m<k<2^{m+1}} | \sum_{j=2^m}^{k} Z_j - \tilde{Z}_j | > \epsilon\, 2^{m/2(1+\theta)}\}$$

$$< \frac{d}{\epsilon^2}\, 2^m\, E|Z_{2^m} - \tilde{Z}_{2^m}|^2/2^{m/(1+\theta)}. \tag{2.1}$$

Now $Z_{2^m} - \tilde{Z}_{2^m} \sim N(0, (\Gamma_{2^m}^{\frac{1}{2}} - I)^2)$, and therefore

$$E|Z_{2^m} - \tilde{Z}_{2^m}|^2 < \|(\Gamma_{2^m}^{\frac{1}{2}} - I)^2\|$$

$$< \|\Gamma_{2^m} - I\|^2 = O(2^{-2m\tau}), \tag{2.2}$$

where $2\tau > \theta/(1+\theta)$. (2.1) and (2.2) imply our result. □

## 2.4 An overview of the proof of the main result

As the proof of theorem 2.1 is rather long it seems worthwhile to present an overview of it. Suppose this result has been shown. Let $\{Z_n\}$ be the corresponding sequence of independent Gaussian random vectors. Let us define

$$S_n = X_1 + \ldots + X_n$$

and

$$T_n = Z_1 + \ldots + Z_n.$$

It is not difficult to show that $|S_n - T_n| = o(n^{1/2(1+\theta)})$ almost surely holds true provided that

$$\Sigma_m P\{\max_{2^{m-1}<k\leq 2^m} |S_k - T_k - S_{2^{m-1}} + T_{2^{m-1}}| > \delta 2^{m/2(1+\theta)}\} < \infty \quad (A)$$

for any $\delta > 0$. Now let us define

$$S_k^{(m)} = S_k - S_{2^{m-1}}, \quad 2^{m-1} < k \leq 2^m$$

$$T_k^{(m)} = T_k - T_{2^{m-1}} .$$

Partition the set $\{2^{m-1} + 1,...,2^m\}$ into alternating small and large blocks. Namely consider

$$2^{m-1} + 1 = a_1^{(1)} < b_1^{(1)} < a_1^{(2)} < b_1^{(2)} < a_2^{(1)} < \ldots < b_u^{(2)} = 2^m,$$

with

$$a_{k+1}^{(2)} = b_k^{(1)} + 1$$

$$a_{k+1}^{(1)} = b_k^{(2)} + 1,$$

and define as the K-th small (large) block the set

$$\{j: a_k^{(1)} \leq j \leq b_k^{(1)}) \ (\{j: a_k^{(2)} \leq j \leq b_k^{(2)}\}, \text{ respectively}).$$

The small blocks, to be of any use, must be constructed in such a way that they can be considered negligible. In our case we will design the small blocks so that

$$\Sigma_m P\{\max_{2^{m-1}<K\leq 2^m} |\Sigma^{(1,k,m)} X_i| > \delta 2^{m/2(1+\theta)}\} < \infty \quad (B)$$

and

$$\Sigma_m P\{\max_{2^{m-1}<K\leq 2^m} |\Sigma^{(1,k,m)} Z_i| > \delta 2^{m/2(1+\theta)}\} < \infty \quad (CB)$$

where $\Sigma^{(1,k,m)}$ ($\Sigma^{(2,k,m)}$, respectively) stands for the summation over those i's which are not greater than k and belong to some small (large, respectively) block. Actually we will use an approximation argument in order to prove (B).

Now we truncate the original random vectors whose indexes belong to some large block. Such a truncation is so performed that

$$\sum_{k=1}^{n} X_k - \tilde{X}_k = o(n^{1/2(1+\theta)}) \quad \text{a.s.},$$

($\{\tilde{X}_k\}$ are the truncated random vectors) which in turn allows us to deal with $\tilde{X}_k$ instead of $X_k$. Such truncation is purely technical.

Next we approximate the sums of the $\tilde{X}_i$'s whose indexes belong to a particular large block by independent random vectors. In other words, let us put

$$W_{k,m} = \sum_{j:\ j \text{ belongs to large block } \#k} \tilde{X}_j .$$

We approximate each $W_{k,m}$ by random vectors $AW_{k,m}$ constructed in such a fashion that $\{AW_{k,m}\}$ is a collection of independent random vectors and $AW_{k,m}$ has the same distribution as $W_{k,m}$. Such approximation is so carried out that

$$\Sigma_m P\{\max_k | \sum_{j=1}^{k} W_{j,m} - AW_{j,m}| > \delta 2^{m/2(1+\theta)}\} < \infty .$$

As a matter of fact we need a bit more than that. Notice that we have performed an approximation of large <u>blocks</u>. We still need to ensure that the error incurred when one "rounds up" a partial sum $\Sigma^{(2,k,m)}\tilde{X}_i$ to the "nearest" partial sum of large blocks is negligible. This can be obtained if

$$\Sigma_m P\{ \max_{1\leq t\leq u} \max_{a_t^{(2)}\leq j\leq b_t^{(2)}} | \sum_{k=a_t^{(2)}}^{j} \tilde{X}_k | > \delta 2^{m/2(1+\theta)}\} < \infty .$$

Our next step is a further grouping procedure. Consider a partition of $\{1,2,\ldots,u(m)\}$ induced by

$$0 = c_0^{(m)} < c_1^{(m)} < \ldots < c_q^{(m)} = u(m) .$$

We approximate each partial sum

$$SW_{j,m} := \sum_{k=c_{j-1}^{(m)}}^{c_j^{(m)}} AW_{k,m}$$

by a Gaussian random vector $SZ_{j,m}$ with the same mean and covariance matrix as $SW_{j,m}$, such that

$$\Sigma_m P\{ \max_{1<j<q} | \sum_{k=1}^{j} SW_{j,m} - SZ_{j,m}| > \delta 2^{m/2(1+\theta)}\} < \infty .$$

Also, the "rounding errors" are shown to be negligible and the random vectors $SZ_{j,m}$ are chosen in such a manner that they are independent.

The final step of our construction is to write each $SX_{j,m}$ as a sum of independent homoskedastic Gaussian random vectors (the Z's which appear in the very beginning of this section). To conclude the proof we show that all the remaining "rounding errors" are negligible.

## 2.5 Constructing Sequences of Random Vectors

In this subsection we will consider some technical subtleties related to the construction of random vectors. Every process of constructing a random variable in a given p–space $(\Omega,F,P)$ depends on the richness of such space. For instance if $F$ is the trivial $\sigma$–field then one cannot construct any random variable in $(\Omega,F,P)$ but the constants.

When dealing with the construction of a single random element taking its values in an Euclidean space and following a particular probability law one has only to assume the existence of a uniformly distributed random variable U defined on the original p–space. Things grow a bit tougher when we need to construct a whole process (see e.g. the proof of Kolmogorov's existence theorem).

When proving almost sure invariance principles for sums of independent random vectors we need a further requirement to be fulfilled. Namely, we must construct a process that obeys a given law and is close to a previously defined process in a probabilistic sense.

By far, the most popular ways of dealing with such additional requirement rely heavily on Strassen–Dudley theorem associated with an estimate of the convergence rate in the central limit theorem. The latter is usually given in terms of the Prohorov distance between the distribution of $S_{a(m+1)} - S_{a(m)}$ and a suitable Gaussian random vector, where $\{S_n\}$ is the partial sum process and $\{a(m)\}$ is a suitable integer sequence. Strassen–Dudley theorem (or a variant of it) allows one constructing a Gaussian random vector which is close to $S_{a(m+1)} - S_{a(m)}$, provided the underlying p–space is rich enough. Under the absence of such richness we must redefine our process in a larger p–space where the construction of above mentioned Gaussian random vectors can be carried out.

The general idea of the proof of our main result does not differ in essence from the procedure just described. There is, however, an intermediate step which aims at approximating the sequence $\{S_{c(m)} - S_{b(m)}\}$ by another sequence of independent random vectors $\{Y_m\}$ (say) so that the distributions of $S_{c(m)} - S_{b(m)}$ and $Y_m$ coincide (here $\{c(m)\}$ and $\{b(m)\}$ are suitable integer sequences). Such an approximation is done by means of Berbee's theorem. Then we approximate the partial sum process for $\{Y_m\}$ by a Gaussian process by means of a variant of a normal approximation theorem due to Einmahl (see Lemma 2–B–2).

When using Berbee's result one must explicitly assume the existence of a random variable U, which is uniformly distributed over [0,1] and is independent of the process of interest. Actually we need a sequence $\{U_n\}$ of iid random variables, such that $U_1$ is distributed as U above and each $U_k$ is independent of the process of interest (we show below that the existence of only one U as above implies the existence of such a sequence).

On the other hand, Einmahl's original statement of his result is not so explicit about our U. Einmahl uses the classical form of Strassen–Dudley theorem which reads: there exists a p–space .... It is not difficult to see that working with the original form of Einmahl's result yields no loss in generality (one can use Lemma A.1 of Berkes and Philipp (1979) at any stage of the construction of the approximating Gaussian process). However it does yield, in our case, notational problems. Besides it hides a technical detail which was first pointed out in Berkes and Philipp (1979). Namely, suppose that one is interested in constructing two independent Gaussian random variables $Z_1$ and $Z_2$ so that $Z_1 + Z_2$ is close, in a probabilistic sense, to $X_1 + X_2$, where $X_1$ and $X_2$ are given independent random vectors. One has, in principle, two possible ways of doing this.

(i) We first construct $Z_1$, close to $X_1$, using some variant of Strassen–Dudley theorem. A typical statement for such construction could be: enlarging the original p–space if necessary, we can construct $Z_1$ so that $X_1$ is "close" to $Z_1$. Typically $Z_1$ will be a measurable function of $X_1$ and a uniform [0,1] random variable which is independent of $X_1$. The "enlargement" of the original p–space is so done as to ensure the existence of such a $U_1$. Actually we need $U_1$ to be independent of $(X_1,X_2)$. If this were not so we could not guarantee $Z_1$ to be independent of $X_2$ and hence the existence of a $Z_2$ independent of $Z_1$ and close to $X_2$ would not be necessarily true. The next step is to construct $Z_2$. Again we could appeal to Strassen–Dudley and enlarging the p–space, construct $Z_2$. Notice that the above described procedure brings about three "nested" p–spaces $(\Omega,F,P)$, $(\Omega_1,F_1,P_2)$ and $(\Omega_2,F_2,P_2)$ (say). If we were to stop in $Z_2$ everything would be fine. This is, however, seldom the case. Usually we are interested in constructing a whole sequence $\{Z_n\}$ of independent Gaussian random variables such that

$$\sum_{k=1}^{n} X_k - Z_k = o(a_n), \quad \text{a.s.}$$

where $\lim a_n = \infty$. Now we must deal with the following question. What does a.s. mean? In other words which is the underlying p-space one is considering above? Such a problem would be solved if we could find a "limit" for the underlying sequence of "nested" p-spaces. However, following the approach proposed by Berkes and Philipp we do not need such unnecessary sophistication. We will describe their approach in the sequel.

(ii) The second way of constructing $Z_1$ and $Z_2$ is the following. Construct Z in such a fashion that Z is close to $X_1 + X_2$. Next "deconvolve" Z into $Z_1$ and $Z_2$. That is, construct $Z_1$ and $Z_2$ so that $Z = Z_1+Z_2$ with $Z_1$ and $Z_2$ independent. Again we would need, in principle, "nested" p-spaces (see the proof of Lemma 2-B-2) so as to ensure the existence of suitable uniformly distributed random variables, which in turn, as in (i) above, would lead to unnecessary sophistications.

In our proof of Theorem 2.1 we face both the deconvolution and "construction" problems described above. Berkes and Philipp's approach, outlined below, provides a neat way of overcoming the possible problems just discussed.

Consider a p-space $(\Omega_0,\mathcal{F}_0,P_0)$ and a sequence of random vectors $\{X_n\}$, defined on it. We first enlarge $(\Omega_0,\mathcal{F}_0,P_0)$ to $(\Omega,\mathcal{F},P)$ so as to obtain one uniformly distributed random variable U and a sequence of random vectors $\{\tilde{X}_n\}$ (all of them defined on the enlarged p-space) such that

(a) the distributions of $\{\tilde{X}_n\}$ and $\{X_n\}$ are identical;

(b) U is independent of $\{\tilde{X}_n\}$.

Such enlargement can be performed as follows. Consider the p-space

$$(\Omega, F, P) = (\Omega_0, F_0, P_0) \otimes ([0,1], B, \lambda) ,$$

where $B$ is the class of borelians in $[0,1]$ and $\lambda$ is the Lebesgue measure restricted
to the unit interval. Let us define

$$\tilde{X}(w,t) = X(w)$$

and

$$U(w,t) = t,$$

where $(w,t) \in \Omega_0 \otimes [0,1]$. It is not difficult to show that both (a) and (b) above hold true. Now we construct an $(\Omega,F,P)$ a sequence $\{U_n\}$ of iid random variables with $U_1$ being uniformly distributed over $[0,1]$ such that each $U_k$ is a measurable function of U (whence (a) and (b) hold valid with U replaced by $U_k$ for any k). This can be accomplished in the following way. Consider any one-to-one function $f: Z^+ \otimes Z^+ \rightarrow Z^+$, such that $f(n,k) < f(n,k+1)$, all n,k (for instance if $\{P_n\}$ is an enumeration of the prime numbers, define $f(n,k) = P_n^k$). Let

$$U(w,t) = \sum_{j=1}^{\infty} 2^{-j} a(j,w,t)$$

be the dyadic representation of U. Define

$$U_n(w,t) = \sum_{k=1}^{\infty} 2^{-k} a(f(n,k), w, t).$$

It is easy to show that $\{U_n\}$ is a sequence of independent random variables uniformly distributed over $[0,1]$.

What the above construction tells us is that to any element of a $Z^+$-partition $\{Q_m, m \geq 1\}$ we can attach a collection of iid-U(0,1) random variables with arbitrary (though denumerable) cardinality. Besides the union of such collections is $\{U_n, n \geq 1\}$ and the intersection of any two distinct collections is empty (in other words the family of such collections partition $\{U_n\}$).

Now let us take a typical step in a construction procedure. Suppose that $\{\tilde{X}_n\}$ (note: $\{\tilde{X}_n\}$ is defined on the enlarged p-space) is a sequence of independent random vectors. Assume that we are dealing with the problem of constructing (with no further enlargements of the p-space) a sequence of independent Gaussian random vectors $\{Z_n\}$ such that (for instance)

$$\Sigma_m P\{|\sum_{k=2^m+1}^{2^{m+1}} \tilde{X}_k - Z_k| > c_m\} < \infty .$$

Let us denote

$$W_m := \sum_{k=2^m+1}^{2^{m+1}} \tilde{X}_k .$$

Assume that the Prohorov distance between the distributions of $W_m$ and a $N(0,\Gamma_m)$ is small. Then (see Lemma 2-B-2 for details) we can construct a zero mean Gaussian random vector $T_m$ with covariance matrix given by $\Gamma_m$ such that $T_m$ and $W_m$ are close in a probabilistic sense and $W_m$ is $\sigma(T_m, V)$-measurable, where V is an element of the collection $R_m$ (say) of uniform random variables associated to $Q_m = \{2^m+1,\ldots,2^{m+1}\}$.

Now we perform the deconvolution operation on $T_m$ (again we refer the reader to Lemma 2-B-2 for details) and write

$$T_m = \sum_{k=2^m+1}^{2^{m+1}} Z_k ,$$

where each $Z_k$ is $\sigma(T_m, V_1,\ldots,V_q)$-measurable, with $V_1,\ldots,V_q$ in $R_m$ and $\{Z_m, 2^m < k \leq 2^{m+1}\}$ being a set of independent Gaussian random vectors.

Finally, as $\{W_m\}$ is a sequence of independent random vectors and $\{R_m\}$ partitions $\{U_n\}$ it easily follows that $\{Z_k\}$ is a sequence of independent Gaussian random vectors defined on $(\Omega,F,P)$ (note: no further enlargements were necessary).

As a final comment we would like to remind that our proof of theorem 2.1 relies heavily on Berbee's theorem which also assumes the existence of a sequence like $\{U_n\}$. Obviously our only-one-enlargement approach works with such "additional" requirement.

Let us just make a few remarks to conclude this section. This section is very much a methodological one. Its main goals are

(A) To spot some possible sources of misunderstanding in the long proof of theorem 2.1.

(B) To make possible the use of a "not so heavy" notation in the proof of theorem 2.1.

(C) To justify why we are not going to use either the classical "enlarging the original p-space ..." or Berkes and Philipp Lemma A.1

The reader should keep this in mind when reading the proof of theorem 2.1.

## APPENDIX 2-A

LEMMA 2-A-1. Consider the following array of random variables

$$
\begin{array}{lll}
X_{11} & & \\
X_{21} & & X_{22} \\
\vdots & & \\
X_{n1} & \cdots & X_{nn} \\
\vdots & &
\end{array}
$$

Let us assume that

(i) For each n, $\{X_{nk}, 1 \leq k \leq n\}$ is a collection of iid random variables.

(ii) For each $\epsilon > 0$, $P\{|X_{n1}| \geq \epsilon n^u\} = o(1/n^{1+t})$, where u and t are positive real numbers.

(iii) $\int_{|x|<n} x dF_n(x) = o(1)$, where $F_n(.)$ is the distribution of $X_{n1}$.

(iv) $S_{nn} = X_{n1} + \ldots + X_{nn} = O_p(n^u)$.

Then

$$P\{|S_{nn}| > \epsilon n^u\} = o(n^{-t}).$$

Proof. Consider the sequence of symmetrized random variables

$\{\tilde{X}_{nk}, 1 \leq k \leq n, n \geq 1\}$, i.e. $\tilde{X}_{nk} = X_{nk} - X_{nk}'$ and

$\{X_{nk}', 1 \leq k \leq n\}$ is an independent copy of $\{X_{nk}, 1 \leq k \leq n\}$. Let us denote the median of a random variable Z by $\mu(Z)$. It is easy to see that for any $a, \epsilon > 0$ we have

$$
\begin{aligned}
\tfrac{1}{2} P\{|X_{n1} - \mu(X_{n1})| > \epsilon\} &\leq P\{|\tilde{X}_{m1}| \geq \epsilon\} \\
&\leq 2P\{|X_{n1} - a| \geq \tfrac{\epsilon}{2}\}, \qquad (1.1)
\end{aligned}
$$

and hence, the symmetrized random variables $\tilde{X}_{nk}$ satisfy (i) to (iv) above. Further, as $S_{nn} = o_p(n^u)$ we have $\mu(|S_{nn}/n^u|) = o(1)$ and thus

$$P\{|S_{nn}| > \epsilon n^u\} = o(n^{-t}), \quad \text{all } \epsilon > 0$$

is equivalent to

$$P\{|S_{nn}/n^u - \mu(S_{nn}/n^u)| > \epsilon\} = o(n^{-t}),$$

all $\epsilon > 0$. Therefore (1.1) enables us to assume with no loss in generality that $\{X_{nk}, n \geq 1, 1 \leq k \leq n\}$ are symmetric random variables. Let us define

$$\bar{X}_{nk} = \begin{cases} X_{nk} & \text{if} \quad |X_{nk}| \leq n^u \\ o & \text{if} \quad |X_{nk}| > n^u \end{cases} .$$

Also let us put

$$\bar{S}_{nn} = \sum_{k=1}^{n} \bar{X}_{nk} .$$

We have

$$n^t P\{|S_{nn}| > n^u \epsilon\} \leq n^{1+t} P\{|X_{n1}| > n^u\} + n^t P\{|\bar{S}_{nn}| > n^u \tfrac{\epsilon}{2}\}$$

$$= o(1) + n^t P\{|\bar{S}_{nn}| > n^u \tfrac{\epsilon}{2}\} . \tag{1.2}$$

Let r be an even integer such that $r > 2t+1$. We have

$$n^t P\{|\bar{S}_{nn}| \geq \epsilon n^u\} \leq n^t \epsilon^{-r} n^{-ur} . E|\bar{S}_{nn}|^r$$

$$\leq n^{t-ur} \epsilon^{-r} \{nE\bar{X}_{n1}^r + n(n-1)E\bar{X}_{n1}^{r-2} E\bar{X}_{n2}^2 + \ldots\}. \tag{1.3}$$

Let $r = 2i_1 + \ldots + 2i_m$ be a representation of r as a sum of positive even integers. The corresponding term in the right hand side of (1.3) is upper bounded by

$$C . n^{t-ru+m} E\bar{X}_{n1}^{2i_1} \ldots E\bar{X}_{n1}^{2i_m} . \tag{1.4}$$

Now notice that

$$E\bar{X}_{n1}^{2j} = \int_0^{n^u} x^{2j-1}P\{|X_{n1}| > x\}dx$$

$$= \int_0^n y^{u(2j-1)}\, P\{|X_{n1}| > y^u\}y^{u-1}dy$$

$$= \begin{cases} o(n^{2ju-t-1}) & \text{if} \quad 2ju > t+1, \\ o(\log n) & \text{if} \quad 2ju = t+1, \\ O(1) & \text{if} \quad 2ju < t+1 \end{cases} \tag{1.5}$$

Now let us split the set $\{i_j,\ j = 1,\ldots,m\}$ into the following three classes

$$L = \{i_j,\ 2ui_j < 1+t\}$$

$$E = \{i_j,\ 2ui_j = 1+t\}$$

$$G = \{i_j,\ 2ui_j > 1+t\}.$$

Also let us denote the cardinalities of L, E and G by $n_L$, $n_E$ and $n_G$ respectively. We have

$$n_L + n_E + n_G = m. \tag{1.6}$$

It is easy to see, taking (1.5) into account, that (1.4) is bounded from above by

$$\begin{cases} o^{(nt-ur+m+\Gamma-n_G t} . (\log n)^{n_E}), & \text{if} \quad n_E+n_G > 0 \\ O(n^{t-ur+m}), \quad \text{if} \quad n_E+n_G = 0, \end{cases}$$

where $\Gamma = \sum_{i_j \in G} (2ui_j - 1)$. Now notice that

$$t-ur+m < t=ur+r/2 = t - (u-\tfrac{1}{2})r < -\tfrac{1}{2}, \tag{1.7}$$

as $u > 1$ and $2t+1 < r$. Thus

$$(I) := n^{t-ru+m}E\bar{X}_{n1}^{2i_1} \ldots E\bar{X}_{n1}^{2i_m} = O(n^{-\frac{1}{2}}) = o(1),$$

if $n_E=n_G = 0$. On the other hand, if $n_E=n_G > 0$ we can write

$$\Gamma = ur - m - \Lambda - tn_E,$$

where

$$\Lambda = \sum_{i_j \in L} (2u_j - 1).$$

Thus, if $n_E+n_G > 0$ we can write (I) as

$$o(n^{t-\Lambda-t(n_E+n_G)}(\log n)^{n_E}) = (II).$$

If $n_E+n_G > 1$ then we have (II) = o(1) (as $\Lambda > 0$). On the other hand if $n_E = 0$ and $n_G = 1$ we have (II) = $o(n^{-\Lambda})$ = o(1). Finally if $n_E = 1$ and $n_G = 0$, then we have $n_L > 0$ (if this were not so we would have $t = ur-1 > u(2t+1) - 1 \geq 2t$), and thus (II) = o(1). Therefore

$$n^t P\{|\bar{S}_{nn}| \geq \epsilon n^u\} = o(1). \tag{1.8}$$

Equations (1.2) and (1.8) imply our result. □

<u>LEMMA 2-A-2</u>. Let $\{X_n\}$, $n \geq 1$ be a stationary absolutely regular process, taking values in $\mathbb{R}$. Suppose that

(i) $EX_1 = 0$

(ii) $E|X_1|^{2\lambda(1+\theta)} < \infty$

(iii) $\beta(k) \leq k^{-\nu}$,

where $\lambda > 1$, $0 < \theta < 1$, $\nu > 0$ and $\{\beta(k)\}$ stands for the sequence of (absolutely regular) mixing weights relative to the process $\{X_n\}$. Assume that

(iv) $\nu > \lambda(3+2\theta)/(\lambda-1)$.

Finally let $\gamma$ and $\omega$ be real numbers such that $0 < \gamma,\omega < 1$ and

(v) $\gamma(\nu-1) > 1$

(vi) $\dfrac{1}{\nu-1} < \gamma\omega < \min\left\{1, \dfrac{\nu-1-\theta}{(1+\nu)(1+\theta)}, \dfrac{\nu(\lambda-1) - 2\lambda(1+\theta)}{\nu + 2\lambda(1+\theta)}\right\}.$

Then for any $\delta > 0$

$$P\{|S_n| > \delta\, n^A\} = o(n^{-(\epsilon+1/(\gamma\omega))}),$$

where $\epsilon$ is a suitable positive real number and

$$A = \frac{1}{2\gamma\omega(1+\theta)} .$$

__Proof.__ Assumption (vi) implies that we can find t so that

$$\nu^{-1}[1 + 1/(\gamma\omega)] < t$$
$$< \min\{1, \frac{1}{\gamma\omega(1+\theta)} - 1, \frac{1}{2\lambda(1+\theta)} [\frac{\lambda-1}{\gamma\omega} - 1]\}. \qquad (2.1)$$

Let us define

$$\ell = [[n^t]]$$
$$m = [[n^{1-t}]] .$$

By virtue of Theorem 1 in Yoshihara (1978) we can write

$$P\{|S_n| > 2\delta n^A\} \leq \ell\, P\{|T_m| > \delta n^A/\ell\} + n\beta(\ell)$$
$$+ \ell\, P\{|X_1| > \delta\, n^A/\ell\} = (I) + (II) + (III),$$

where $T_u = Z_1 + \ldots + Z_u$, with $Z_1, \ldots, Z_m$ iid random variables such that $Z_1$ has the same distribution as $X_1$. Now, as

$$\ell n^{-t} - 1 = o(1),$$
$$mn^{-(1-t)} - 1 = o(1),$$

we can assume with no loss in generality that

$$\ell = n^t$$
$$m = n^{1-t}.$$

We have

$$(II) \leq n\ell^{-\nu} = n^{1-\nu t} \leq n^{-(\epsilon_1 + 1/(\gamma\omega))}, \qquad (2.2)$$

for a small positive real number $\epsilon_1$, as a consequence of the first inequality in (2.1). On the other hand, we can write

$$P\{|T_m| > \delta n^A/\ell\} \approx P\{|T_m| > \delta m^{A/(1-t)} . m^{-t/(1-t)}\}$$

$$= P\{|T_m| > \delta m^u\}, \quad (\text{say}).$$

Now notice that

$$u = (A-t)/(1-t) = \left[\frac{1}{2\gamma\omega(1+\theta)} - t\right]/(1-t) > \tfrac{1}{2}$$

by our choice of t. Thus, the Central Limit Theorem implies that

$$T_m = o_p(m^u). \tag{2.3}$$

On the other hand

$$P\{|X_1| > \delta m^u\} = P\{|Z_1| > \delta m^u\}$$
$$= m^{-2\lambda(1+\theta)u} o(1), \tag{2.4}$$

by Markov inequality and assumption (ii). Now, by Lemma 2-A-1, (2.3) and (2.4), we have

$$P\{|T_m| > \delta m^u\} = o(m^{1-2\lambda(1+\theta)u}),$$

whence

$$(\mathrm{I}) = \ell P\{|T_m| > \delta m^u\} = o(m^v),$$

where

$$v = \frac{t}{1-t} + 1 - \frac{2\lambda(1+\theta)}{1-t}\left[\frac{1}{2\gamma\omega(1+\theta)} - t\right]$$
$$= \frac{1}{1-t}\left[1 + 2\lambda t(1+\theta) - \frac{\lambda}{\gamma\omega}\right]$$
$$< -\frac{1}{1-t}\cdot\frac{1}{\gamma\omega},$$

by our choice of t in (2.1). Therefore we can write

$$(\mathrm{I}) = o(m^{-(\epsilon_2 + 1/(1-t)(\gamma\omega))}),$$

or

$$(\mathrm{II}) = o(n^{-(\epsilon_3 + 1/(\gamma\omega)}), \tag{2.5}$$

where $\epsilon_2$ and $\epsilon_3$ are suitable small positive numbers. Finally (2.4) implies

$$(III) = o(\ell . m^{-2\lambda(1+\theta)u}) = o(n^s),$$

where

$$s = t - 2\lambda(1-\theta)\left[\frac{1}{2\gamma\omega(1+\theta)} - t\right]$$

$$< -\left(\epsilon_3 + \frac{1}{\gamma\omega}\right). \tag{2.6}$$

(2.2), (2.5) and (2.6) imply our result. □

<u>LEMMA 2-A-3</u>. Let $\{X_n\}$, $n \geq 1$ be a zero mean stationary absolutely regular process taking values in $\mathbb{R}$. Suppose that

(i) $E|X_1|^{2\lambda(1+\theta)} < \infty$,

(ii) $\beta(k) \leq k^{-\nu}$,

where $\lambda > 1$, $0 < \theta < 1$, $\nu > 0$ and $\{\beta(k)\}$ stands for the sequence of (absolutely regular) mixing weights relative to the process $\{X_n\}$. Let us also assume that

(iii) $\nu > \lambda(3+2\theta)/(\lambda-1)$.

Let $\gamma$ be any real number satisfying

(iv) $\frac{1}{\nu-1} < \gamma < \min\left\{1, \frac{\nu-1-\theta}{(1+\nu)(1+\theta)}, \frac{\nu(\lambda-1) - 2\lambda(1+\theta)}{\nu + 2\lambda(1+\theta)}\right\}$.

Take $\delta > 0$ and assume that for $n \geq n_0$ (say) we have

(v) $c = c(n,\delta) := \max_{j \leq n} P\{|S_n - S_j| > \delta A\} < 1$,

where $S_j = \sum_{k=1}^{j} X_k$, and $A = A(n) = n^{1/(2\gamma(1+\theta))}$. Take $h > 0$, then for n large enough we have the following Ottaviani-like inequality

$$P\{\max_{1 \leq k \leq n} |S_k| \geq 4\delta A\} \leq 2(1-c)^{-1}\left[P\{|S_n| > \delta A\} + \frac{n^{(\gamma-1)/\gamma}}{(\log n)^{1+h}}\right]$$

**<u>Proof</u>**. Let us define $T = \min\{k;\ |S_k| > 4\delta A,\ 1 \leq k \leq n\}$. Put

$$R = \{k \le n;\ P\{T=k\} > 1/[n^{1/\gamma}.(\log n)^{1+h}]\}, \quad \text{and}$$

$$\bar{R} = \{1,2,\ldots,n\} - R.$$

We have

$$P\{|S_n| > 2\delta A,\ \max_{k\le n} |S_k| > 4\delta A\}$$

$$\ge \sum_{k\le n} P\{|S_n - S_k| \le 2\delta A,\ T=k\}. \tag{3.1}$$

Take $k \in R$. We have

$$P\{|S_n - S_k| \le 2\delta A,\ T=k\}$$

$$= P\{T=k\}[1 - P\{|S_n-S_k| > 2\delta A,\ T=k\}/P\{T=k\}]. \tag{3.2}$$

Now Let $\omega \in (0,1)$ be so chosen that

$$\frac{1}{\nu-1} < \gamma\omega < \min\{1,\ \frac{\nu-1-\theta}{(1+\nu)(1+\theta)},\ \frac{\nu(\lambda-1) - 2\lambda(1+\theta)}{\nu + 2\lambda(1+\theta)}\}.$$

Notice that the existence of such $\omega$ is guaranteed by assumption (v) above. Let us denote

$$p = [[n^{w}]]. \tag{3.3}$$

We have

$$P\{|S_n - S_k| > 2\delta A,\ T=k\}$$

$$\le P\{|S_n - S_{k+p}| > \delta A,\ T=k\} + P\{|S_{k+p} - S_k| > \delta A\}$$

$$= (I) + (II),\quad (\text{say}). \tag{3.4}$$

We have

$$(I) \le P\{|S_n - S_{k+p}| > \delta A\}P\{T=k\} + \beta(p)$$

$$\le P\{T=k\}[c + \beta(p)n^{1/\gamma}(\log n)^{1+h}]$$

$$\le P\{T=k\}[c + n^{-\omega\nu}.n^{1/\gamma}(\log n)^{1+h}]$$

$$= P\{T=k\}[c + o(1)], \tag{3.5}$$

where for the first inequality above we have used the simple fact that $\alpha(A,B) \le$

$\beta(A,B)$ for any pair of $\sigma$-fields $A$, $B$. For the second inequality above we have used the definition of c (assume WLG $n > n_0$) and the fact that $k \in R$. Finally the equality holds by virtue of our choice of w. Now, in order to obtain a bound for (II) we will apply Lemma 2-A-2. Notice that

$$\begin{aligned}(II) &= P\{|S_{k+p}-S_k| > \delta A\} \\ &\approx P\{|S_p| > \delta\, p^B\},\end{aligned}$$

where $B = 1/[2\gamma\omega(1+\theta)]$. Assumptions (i)-(vi) in Lemma 2-A-2 are obviously in force and thus we obtain

$$\begin{aligned}(II) \approx P\{|S_p| > \delta\, p^B\} &= o(p^{-(\epsilon+1/(\gamma\omega))}) \\ &= o(n^{-(\epsilon'+1/\gamma)}),\end{aligned} \tag{3.6}$$

where $\epsilon$ and $\epsilon'$ are suitable positive real numbers. Thus

$$P\{|S_{k+p}-S_k| > \delta A\} = P\{T=k\}.o(1). \tag{3.7}$$

Now (3.2), (3.4), (3.5) and (3.7) imply that

$$\begin{aligned}&\sum_{k\in R} P\{|S_n-S_k| < 2\delta A,\ T=k\} \\ &> \sum_{k\in R} P\{T=k\}[1-c+o(1)] \\ &> \tfrac{1}{2} \sum_{k\in R} P\{T=k\}(1-c),\end{aligned} \tag{3.8}$$

if n is large enough. On the other hand

$$\begin{aligned}&\sum_{K\in\bar{R}} P\{|S_n-S_k| < 2\delta A,\ T=k\} > 0 \\ &> \sum_{K\in\bar{R}} P\{T=k\} - n/(n^{1/\gamma}(\log n)^{1+h}).\end{aligned}$$

Therefore

$$\begin{aligned}&P\{|S_n| > 2\delta A\} > P\{|S_n| > 2\delta A,\ \max_{k<n} |S_k| > 4\delta A\} \\ &> \frac{1-c}{2}\Big[\sum_{k<n} P\{T=k\}\Big] - n^{(\gamma-1)/\gamma}(\log n)^{-(1+h)}\end{aligned}$$

$$= \frac{1-c}{2} P\{\max_{k \leq n} |S_k| > 4\{A\} - n^{(\gamma-1)/\gamma}(\log n)^{-(1+h)}. \qquad \square$$

<u>LEMMA 2-A-4</u>. Let $\{X_n\}$, $n \geq 1$ be a zero mean stationary absolutely regular process taking values in $\mathbb{R}$. Suppose that

(i) $E|X_1|^{2\lambda(1+\theta)} < \infty$

(ii) $\beta(k) \leq k^{-\nu}$,

where $\lambda > 1$, $0 < \theta < 1$, $\nu > 0$ and $\{\beta(k)\}$ stands for the sequence of mixing weights relative to the process $\{X_n\}$. Assume that

(iii) $\nu > 2\lambda(1+\theta)/(\lambda-1)$.

Let $\gamma$ be any real number satisfying

(iv) $\frac{1}{\nu} < \gamma < \min\{1, \frac{(\lambda-1)}{2\lambda(1+\theta)}, \frac{1}{1+\theta} - \frac{1}{\nu}\}$.

Take $\delta > 0$. Then

$$P\{|S_n| > \delta A\} = o(n^{-(1-\gamma)(1+\epsilon)/\gamma}),$$

where $\epsilon$ is a suitably chosen positive real number and $A = A(n) = n^{1/(2\gamma(1+\theta))}$.

<u>Proof</u>: Assumption (iv) enables us to obtain t such that

$$\frac{1}{\gamma\nu} < t < \min\{1, \frac{\lambda-1}{2\gamma\lambda(1+\theta)}, \frac{1}{\gamma(1+\theta)} - 1\}. \tag{4.1}$$

Let us define

$$\ell = [[n^t]]$$

$$m = [[n^{1-t}]].$$

By Theorem 1 in Yoshihara (1978) we can write

$$P\{|S_n| > \delta A\} \leq \ell\, P\{|T_m| > \delta A/\ell\} + n\beta(\ell)$$

$$+ \ell\, P\{|X_1| > \delta A/\ell\} = (I) + (II) + (III),$$

where $T_u = Z_1 + \ldots + Z_u$, with $Z_1, \ldots, Z_m$ iid random variables such that $Z_1$ has the same distribution of $X_1$. We can (and we will) assume WLG that (cf.

the proof of Lemma 1–A–2)

$$\ell = n^t$$

$$m = n^{1-t}.$$

We have by (4.1)

$$(II) = n\beta(\ell) < n\ell^{-\nu} = n^{1-\nu t} < n^{-(1-\gamma)(1+\epsilon_1).\gamma},$$

where $\epsilon_1$ is a suitably chosen small positive real number. On the other hand

$$P\{|T_m| > \delta A/\ell\} \approx P\{|T_m| > \delta m^B\},$$

where $B = \frac{1}{1-t}\left[\frac{1}{2\gamma(1+\theta)} - t\right] > \frac{1}{2}$, because of our choice of t.

Therefore

$$T_m = o_p(m^B), \tag{4.2}$$

as a consequence of the Central Limit Theorem. Now

$$P\{|Z_1| > m^B\} = o(m^{-2\lambda(1+\theta)B}), \tag{4.3}$$

by virtue of Markov inequality and assumption (i). Thus a simple application of Lemma 2–A–1 yields

$$P\{|T_m| > \delta m^B\} = o(m^{1-2\lambda(1+\theta)B}). \tag{4.4}$$

Hence

$$(I) = o(\ell m^{1-2\lambda(1+\theta)B}) = o(m^{t/(1-t)+1-2\lambda(1+\theta)B}).$$

But

$$\frac{t}{1-t} + 1 - 2\lambda(1+\theta)B = \frac{1}{1-t}\left[1 - \frac{\lambda}{\gamma} + 2\lambda(1+\theta)t\right]$$

$$< -\frac{1}{1-t}\,\frac{(1-\gamma)}{\gamma}\,.\,(1+\epsilon_2), \tag{4.5}$$

where $\epsilon_2$ is a small positive real number, by (4.1). Therefore

$$(I) = o(m^{-(1-\gamma)(1+\epsilon_2)/[\gamma(1-t)]})$$

$$= o(n^{-(1-\gamma)(1+\epsilon_2)/\gamma}). \tag{4.6}$$

Now (4.3), Markov inequality and (4.1) entail

$$(\mathrm{III}) \quad = \quad o(n^{-(1-\gamma)(1+\epsilon_3)/\gamma}). \qquad (4.7)$$

In order to finish the proof take $\epsilon = \min\{\epsilon_1, \epsilon_2, \epsilon_3\}$ amd consider the bounds for (I), (II) and (III). □

<u>LEMMA 2-A-5</u>. Let $\{X_n\}$, $n \geq 1$ be a zero mean stationary absolutely regular process taking values in $\mathbb{R}$. Let $\{\beta(k)\}$, $k \geq 1$ be the corresponding sequence of mixing weights. Assume that

(i) $E|X_1|^{2\lambda(1+\theta)} < \infty$

(ii) $\beta(k) = k^{-\nu}$,

where $0 < \theta < 1$, $\lambda > 1$, $\nu > 0$. Let us define

$$W_k = X_k\, I\{|X_k| > N^{\alpha}\},$$

$$Z_k = X_k - W_k,$$

$$\overline{W}_k = W_k - EW_k, \quad \text{and}$$

$$\overline{Z}_k = Z_k - EZ_k,$$

where $N \geq 1$ is an integer and $\alpha = 1/2\gamma\lambda(1+\theta)$, with $0 < \gamma < 1$. If we assume that

(iii) $0 < \sigma^2 = EX_1^2 + 2 \sum_{k>1} EX_1X_k$ .

Then

$$|\sigma^2 - \operatorname{var}(N^{-\frac{1}{2}}(\overline{Z}_1 + \ldots + \overline{Z}_N))| \quad = \quad O(N^{-u}) \ ,$$

where

$$u = \max\{[\lambda(1+\theta) - 1]/2\lambda\gamma(1+\theta) - 1/\gamma\nu, \quad 1 - 1/\gamma(\nu+1)\}$$

provided

(iv) $\gamma\nu > 1$.

**Proof.** We have

$$\mathrm{Var}(N^{-\frac{1}{2}}(\bar{Z}_1 + \dots + \bar{Z}_N))$$

$$= E\bar{Z}_1^2 + 2\sum_{k=1}^{N-1} E\bar{Z}_1\bar{Z}_{k+1} - 2N^{-1}\sum_{k=1}^{N-1} kE\bar{Z}_1\bar{Z}_{k+1}$$

$$= E(X_1-\bar{W}_1)^2 + 2\sum_{k=1}^{N-1} E(X_1-\bar{W}_1)(X_{k+1}-\bar{W}_{k+1}) - 2N^{-1}\sum_{k=1}^{N-1} KE\bar{Z}_1\bar{Z}_{k+1}$$

$$= (\mathrm{I}) + (\mathrm{II}) - (\mathrm{III}). \tag{6.1}$$

It is easy to see that

$$(\mathrm{I}) = EX_1^2 - EW_1^2 - E^2W_1. \tag{6.2}$$

But

$$|EW_1^2 + E^2W_1| \leq 2\, EW_1^2 = 2EX_1^2 I\{|X_1| > N^{\alpha}\}$$

$$\leq 2\, E|X_1|^{2\lambda(1+\theta)}/N^{(\lambda(1+\theta)-1)/\lambda\gamma(1+\theta)} . \tag{6.3}$$

On the other hand

$$(\mathrm{II}) = 2\sum_{k=1}^{N-1} EX_1X_{k+1} - 2\sum_{k=1}^{N-1} E\bar{W}_1X_{k+1} - 2\sum_{k=1}^{N-1} EX_1\bar{W}_{k+1}$$

$$+ 2\sum_{k=1}^{N-1} E\bar{W}_1\bar{W}_{k+1} = 2\sum_{k=1}^{N-1} EX_1X_{k+1} - (\mathrm{IV}) - (\mathrm{V}) + (\mathrm{VI}). \tag{6.4}$$

Let us define

$$M = [[N^{\epsilon}]],$$

where $\epsilon = 1/2\gamma\nu$.

Take $K \leq M$. We have

$$E\bar{W}_1X_{k+1} = EW_1X_{k+1} \leq \|X_{k+1}\|_2\|W_1\|_2 . \tag{6.5}$$

The above inequality and (6.3) imply

$$E\bar{W}_1X_{k+1} \ll N^{-(\lambda(1+\theta)-1)/2\lambda\gamma(1+\theta)} . \tag{6.6}$$

On the other hand

$$|E\bar{W}_1X_{k+1}| \le 12\|\bar{W}_1\|_{2\lambda(1+\theta)}\|X_{k+1}\|_{2\lambda(1+\theta)}[\beta(k)]^{1-1/\lambda(1+\theta)},$$

by a well known inequality [see e.g. Roussas and Ioannides (1987]. Therefore

$$\sum_{k=M+1}^{N} E\bar{W}_1X_{k+1} \le \sum_{k>M} |E\bar{W}_1X_{k+1}|$$

$$<< M^{-\nu(1-1/\lambda(1+\theta))+1} . \qquad (6.7)$$

Now (6.6) and (6.7) yield

$$(IV) = \sum_{k=1}^{N} E\bar{W}_1X_{k+1} << MN^{-(\lambda(1+\theta)-1)/2\lambda\gamma(1+\theta)}$$

$$+ M^{-\nu(1-1/\lambda(1+\theta))+1} << N^{-u}, \qquad (6.8)$$

by our choice of M. Similarly we can show that

$$(V) << (VI) << N^{-u}. \qquad (6.9)$$

Now let $\delta = 1/2\gamma(\nu+1)$ and $L = N^{\delta}$. We have

$$(III) = N^{-1}\sum_{k=1}^{L} kE\bar{Z}_1\bar{Z}_k + N^{-1}\sum_{k=L+1}^{N} kE\bar{Z}_1\bar{Z}_k$$

$$= (VII) + (VIII).$$

Schwartz inequality implies

$$(VII) = O(N^{-s}), \qquad (6.10)$$

where

$$s = 1 - 1/\gamma(\nu+1).$$

On the other hand an argument similar to that used to obtain (6.8) yields

$$(VIII) << L^{-t}N^{-1} << N^{-(1+\delta t)}$$

where

$$t = \nu\left[1 - \frac{1}{\lambda(1+\theta)}\right] - 2.$$

After some arithmetic we obtain

$$1 + \delta t > [\lambda(1+\theta) - 1]/2\lambda\gamma(1+\theta) - 1/\gamma\nu,$$

and therefore

$$(\mathrm{VIII}) = O(N^{-u}). \tag{6.11}$$

Now (6.8), (6.9), (6.10) and (6.11) yield

$$|EX_1^2 + 2\sum_{k=1}^{N} EX_1X_k - \mathrm{Var}[N^{-\frac{1}{2}}(Z_1 + \dots + Z_N)]|$$

$$= O(N^{-u}).$$

Hence, in order to conclude the proof we need only to show that

$$|\sum_{k>N} EX_1X_k| = O(N^{-u})$$

which can be proved using the same arguments as in (6.7). □

<u>LEMMA 2-A-6</u>. Let $\{V_n\}$, $n \geq 1$ be a zero mean stationary absolutely regular process taking values in $\mathbb{R}^d$. Let $\{\beta(k)\}$ be the corresponding sequence of mixing weights. Assume that

(i) $E|V_1|^{2\lambda(1+\theta)} < \infty$

(ii) $\beta(k) = k^{-\nu}$,

where $0 < \theta < 1$, $\lambda > 1$, $\nu > 0$. Let us define for $j = 1,\dots,d$

$$W_k^j = V_k^j\, I\{|V_k^j| > N^{\alpha}\},$$

$$Z_k^j = V_k^j - W_k^j,$$

$$\bar{W}_k^j = W_k^j - EW_k^j, \quad \text{and}$$

$$\bar{Z}_k^j = Z_k^j - EZ_k^j,$$

where $N \geq 1$ is a fixed integer, $\alpha = 1/2\gamma\lambda(1+\theta)$, with $0 < \gamma < 1$ and $\{V_k^j\}$,

$k \geq 1$, $j = 1,\ldots,d$ are the coordinate processes of $\{V_n\}$, $n \geq 1$. Assume that

(iii) $$\Omega = EV_1V_1^T + \sum_{k>1} E[V_1V_k^T + V_kV_1^T]$$

is a positive definite matrix. Then

$$\|\Omega - \mathrm{cov}(N^{-\frac{1}{2}}(\bar{Z}_1 + \ldots + \bar{Z}_N))\| = O(N^{-u}),$$

where we have denoted

$$\bar{Z}_k = (\bar{Z}_k^1, \ldots, \bar{Z}_k^d)^T,$$

and

$$u = \max\{[\lambda(1+\theta) - 1]/2\lambda\gamma(1+\theta) - 1/\gamma\nu,\ 1 - 1/\gamma(\nu+1)\},$$

provided

(iv) $\gamma\nu > 1$.

<u>Proof</u>: The proof consists of showing that each element of the matrix $\Omega - \mathrm{cov}(N^{-\frac{1}{2}}(\bar{Z}_1 + \ldots + \bar{Z}_N))$ is $O(N^{-u})$, which can be shown using the same sort of arguments we used in the proof of Lemma 2-A-5 and Schwartz inequality at the appropriate places.

## APPENDIX 2-B

<u>LEMMA 2-B-1</u>. Let $\{X_n\}$ be a sequence of identically distributed random vectors taking values in $\mathbb{R}^d$. Suppose that $E|X_1|^t < \infty$, $t > 1$. Let us define

$$Z_n = X_n I\{|X_n| < 2^{m/t}\},$$

for $2^{m-1} < n \le 2^m$. Then we have

$$\sum_{k=1}^{N} X_k - [Z_k - EZ_k] = o(N^{1/t}) \quad \text{a.s.},$$

provided that $EX_1 = 0$.

<u>Proof</u>: As $EX_n = 0$ we have

$$X_n - [Z_n - EZ_n] = X_n I\{|X_n| > 2^{m/t}\} - EX_n I\{|X_n| > 2^{m/t}\},$$

for $2^{m-1} < n \le 2^m$. Let us denote

$$H(n) = 2^{m/t} \quad \text{if} \quad 2^{m-1} < n \le 2^m, \quad m \ge 1.$$

We have

$$X_n I\{|X_n| \ge H(n)\} \neq 0 \quad \text{iff} \quad |X_n| \ge H(n).$$

Also

$$\Sigma_n P\{|X_n| \ge H(n)\} \le \Sigma_m 2^m P\{|X_{2m}| \ge 2^{m/t}\}$$

$$= \Sigma_m 2^m P\{|X_1|^t \ge 2^m\}. \tag{1.1}$$

But $E|X_1|^t < \infty$ implies $\Sigma P\{|X_1|^t \ge K\} < \infty$ and thus (as $P\{|X_1|^t > K\}$ is nonincreasing in K)

$$\Sigma 2^m P\{|X_1|^t \ge 2^m\} < \infty. \tag{1.2}$$

(1.1), (1.2) and Borel–Cantelli lemma imply

$$P\{|X_n| \ge H(n) \text{ i.o.}\} = 0$$

and thus (as $H(n) \to \infty$)

$$\sum_{k=1}^{n} X_k I\{|X_k| > H(k)\} = O(1) = o(H(n)) \quad \text{a.s.} \tag{1.3}$$

On the other hand if $2^{m-1} < n \leq 2^m$ then

$$E|X_n| I\{|X_n| > H(n)\} = E|X_n| I\{|X_n| > 2^{m/t}\}$$
$$\leq [E|X_n|^t I\{|X_n| > 2^{m/t}\}]\, 2^{-m(t-1)/t},$$

whence

$$[\sum_{k=1}^{n} E|X_k| I\{|X_n| > H(k)\}]/H(n)$$
$$\leq [\sum_{j=1}^{m} 2^j E|X_1| I\{|X_1| > 2^{j/t}\}]/H(n)$$
$$\leq [\sum_{j=1}^{m} 2^j E|X_1|^t I\{|X_1| > 2^{j/t}\} 2^{-j(t-1)/t}]/H(n)$$
$$= [\sum_{j=1}^{m} 2^j E|X_1|^t I\{|X_1|^t > 2^j\}]/H(n)$$
$$<< [\sum_{j=1}^{m} 2^{j/t} E|X_1|^t I\{|X_1|^t > 2^j\}]/\sum_{k=1}^{m} 2^{k/t}$$
$$= \quad (I)$$

Now as $E|X_1|^t I\{|X_1|^t > 2^j\} = o(1)$, a simple application of Toeplitz lemma shows that

$$(I) = o(1)$$

and thus

$$\sum_{k=1}^{n} E|X_k| I\{|X_k| > H(k)\} = o(H(n))$$

along with (1.3) imply the lemma. □

The following Lemma is not original. It is just a restatement of Theorem 7 in Einmahl (1987) in a more rigorous and convenient fashion.

<u>LEMMA 2-B-2</u>. Let $AW_1,\ldots,AW_n$ be zero mean independent random vectors taking values in $R^d$ and defined on $(\Omega,F,P)$. Assume that there exists a random variable U, uniformly distributed over [0,1] defined on $(\Omega,F,P)$ and independent of $(AW_1,\ldots,AW_n)$. Let us suppose that

$$|AW_k| \le \frac{\epsilon}{14\sqrt{d}} \sqrt{\varphi_n \log(1/\rho_n)} \qquad \text{a.s.,} \quad k = 1,\ldots,n,$$

where $\varphi_n$ ($\psi_n$) is the smallest (largest) eigenvalue of $\Gamma_n := \text{cov}(AW_1 + \ldots + AW_n)$, $\rho_n := \varphi_n^{-3/2}\Sigma_1^n E[|AW_k|^3]$, $\epsilon \in (0,2)$. Then there exists a random vector $T_n$ defined on $(\Omega,G,P)$ taking its values in $R^d$ such that

$$T_n \sim N(0,\Gamma_n)$$
$$E|T_n - (AW_1 + \ldots + AW_n)|^2 \le D_\epsilon \psi_n \rho_n^{2-\epsilon},$$

where $D_\epsilon$ is a positive constant depending on $\epsilon$ and d only, and $G = \sigma(AW_1,\ldots,AW_n, U)$. Further $T_n$ admits of the following representation

$$T_n = Z_1 + \ldots + Z_n,$$

where $\{Z_k\}$, $1 \le k \le n$ is a collection of iid random vectors defined on $(\Omega,G,P)$ and $Z_1$ is Gaussian.

<u>Proof</u>: We can assume without loss of generality that $\rho_n \le \frac{1}{2}$. Let $G_n$ be the distribution of $\Gamma_n^{-\frac{1}{2}}\Sigma_1^n AW_k$. Let us define

$$\lambda(F,G,\delta) = \sup\{F(A) - G(A^\delta)\},$$

where the sup above is taken over all closed sets A in $R^d$ and F and G are distributions in $R^d$. From theorem 6 in Einmahl (1987) we infer that

$$\lambda(G_n,\ N(0,I),\ C_\epsilon \rho_n^{1-\epsilon/2}) \le C'_\epsilon \rho_n^6, \tag{2.1}$$

where $C_\epsilon$ and $C'_\epsilon$ are positive constants depending on $\epsilon$ and d only. Now approximate $S_n = \Gamma_n^{-\frac{1}{2}}\Sigma_1^n AW_k$ by a discrete random vector $S_n^*$ such that

$$P\{|S_n - S_n^*| > (1/3)C_\epsilon \rho_n^{1-\epsilon/2}\} = 0. \tag{2.2}$$

Denote by $G_n^*$ the distribution of $S_n^*$. A triangle inequality argument for $\lambda(.,.,\delta)$ yields

$$\lambda(G_n^*, N(0,I), \tfrac{4}{3} C_\epsilon \rho_n^{1-\epsilon/2}) < C_e' \rho_n^6 . \tag{2.3}$$

Now let $V_1, V_2, \ldots, C_M$ be iid random variables uniformly distributed over [0,1] such that each $V_k$ is a measurable function of U, where $M > dn$. The existence of such random variables can be shown as follows. First we construct two random variables $U_1$ and $U_2$ which are independent, uniformly distributed and measurable functions of U (for instance let the binary expansions of $U_1$ and $U_2$ consist of alternating digits in the binary expansion of U). Carry on in an inductive fashion. Let us denote by $F_d$ the distribution of a standard d-variate normal random vector. By virtue of Strassen-Dudley Theorem [see Dudley (1968), Theorem 2] and (2.3) we obtain a probability measure Q defined on the Borelians of $\mathbb{R}^d \otimes \mathbb{R}^d$ such that Q admits of $G_n^*$ and $F_d$ as its marginals and

$$Q\{(x,y);\ |x-y| > \tfrac{4}{3} C_\epsilon \rho_n^{1-\epsilon/2}) < C_\epsilon' \rho_n^6. \tag{2.4}$$

Now Lemma 2.4 in Berkes and Philipp (1979) and (2.4) imply that we can construct a standard d-variate normal random vector $T_n^*$ on $(\Omega, \mathcal{G}_1, P)$ such that $T_n^*$ and $S_n^*$ have joint distribution Q, where $\mathcal{G}_1 = \sigma(AW_1, \ldots, AW_n, V_{1+nd})$. From (2.2) and (2.4) we conclude

$$P\{|S_n - T_n^*| > 2C_\epsilon \rho_n^{1-\epsilon/2}\} < C_\epsilon' \rho_n^6 . \tag{2.5}$$

Now argue as in the proof of Theorem 7 in Einmahl (1987) to obtain

$$E|AW_1 + \ldots + AW_n - \Gamma_n^{\frac{1}{2}} T_n^*|^2 < D_\epsilon \Psi_n \rho_n^{2-\epsilon}. \tag{2.6}$$

Put $T_n = \Gamma_n^{\frac{1}{2}} T_n^*$. To conclude the proof we must show the existence of $Z_1, \ldots, Z_n$, such that $T_n = Z_1 + \ldots + Z_n$, with $\{Z_k, 1 < k < n\}$ iid. Let $X_1$ and $X_2$ be independent zero mean d-variate Gaussian random vectors with fullrank covariance matrices. It is not difficult to show that there exists a matrix $A_2$ such that

$$EX_1 \mid (X_1 + A_2X_2) \sim N(0, \theta \operatorname{cov}(X_1)) \mid, \tag{2.7}$$

where $0 < \theta < 1$. Now using $V_1,\dots,V_d$ construct via quantile functions a standard d-variate normal random vector $X_2$ (which is necessarily independent of $T_n$). Choose $A_2$ as in (2.5) with $\theta = 1/n$ and $X_1 = T_n$. Write

$$Z_1 = ET_n \mid (T_n + A_2X_2), \tag{2.8}$$

$$T_n = Z_1 + T_n - Z_1 . \tag{2.9}$$

Notice that $Z_1$ is independent of $T_n - Z_1$. Let us denote $T_{n-1} = T_n - Z_1$. Construct, using $V_{d+1},\dots,V_{2d}$ a standard d-variate normal random vector $X_3$. Notice that $X_3$ is independent of $(Z_1, T_{n-1})$. Choose $A_3$ so that

$$\operatorname{cov}(T_{n-1} \mid T_{n-1} + A_3X_3) = \frac{1}{n} \operatorname{cov}(T_n)$$

and denote

$$Z_1 = ET_{n-1} \mid T_{n-1} + A_3X_3$$

$$T_{n-2} = T_{n-1} - Z_2 .$$

It is easy to see that $Z_1$, $Z_2$ and $T_{n-2}$ are independent. Carry on inductively and put $Z_n = T_1$. □

## APPENDIX 2–C

Proof of Theorem 2.1: Let us denote for $m \geq 1$

$$J_m = \{2^m+1, \ldots, 2^{m+1}\}, \quad \text{and}$$

$$J_o = \{1,2\}.$$

Let $\pi$ and $\gamma$ be positive real numbers such that $\pi < \gamma < 1$. $\pi$ and $\gamma$ will be chosen more precisely later on. Consider the partition of $J_m$ into consecutive nonoverlapping blocks

$$P_{1,m},\ G_{1,m},\ \ldots,\ P_{u,m},\ G_{u,m},\ P_{u+1,m},$$

where

$$\# \quad G_{k,m} = [[2^{m\gamma}]]$$

$$\# \quad P_{k,m} = [[2^{m\pi}]], \qquad 1 \leq k \leq u$$

and

$$u = u(m) = [[2^m/([[2^{m\gamma}]] + [[2^{m\pi}]])]]$$

$$\tilde{=} [[2^{m(1-\gamma)}]],$$

where [[x]] stands for the integer part of x. One obviously has

$$\# \quad P_{u+1,m} < [[2^{m\gamma}]] + [[2^{m\pi}]] .$$

We can assume with no loss in generality (see section 2.5.1) the existence of a sequence $\{U_n\}$ of random variables defined on $(\Omega,F,P)$ such that

- The $U_n$'s are iid
- $U_1$ is uniformly distributed over (0,1), (2.1)
- $\{U_n\}$ is independent of $\{X_n\}$.

Let us define

$$\tilde{X}_k = X_k I\{|X_k| \leq 2^{m/2\lambda(1+\theta)}\} - EX_k I\{|X_k| \leq 2^{m/2\lambda(1+\theta)}\},$$

where $2^{m-1} < k \leq 2^m$. Also let us put

$$W_{t,m} = \sum_{k \epsilon G_{t,m}} \tilde{X}_k ,$$

$$V_{t,m} = \sum_{k \epsilon P_{t,m}} X_t , \quad 1 \leq t \leq u \quad \text{and}$$

$$V_{u+1,m} = \sum_{k \epsilon P_{u+1,m}} X_k .$$

Now Lemma 2-B-1 yields

$$|\tilde{S}_n - S_n| = o(n^{1/2(1+\theta)}) \quad \text{a.s.} \tag{2.2}$$

where

$$\tilde{S}_n = \sum_{k=1}^{n} \bar{X}_k ,$$

with

$$\bar{X}_k = \tilde{X}_k \quad \text{if} \quad k \in G_{t,m}, \quad \text{for some } t,m$$

$$= X_k \quad \text{otherwise.}$$

Now by Berbee's Theorem and (2.1) we can find random vectors $AW_{t,m}$, $m \geq 0$, $1 \leq t \leq u(m)$ such that

- $AW_{t,m}$ and $W_{t,m}$ are identically distributed (2.3)
- $AW_{t,m}$ is independent of

$$\{AW_{s,q}, \ s < t \text{ and } q = m \text{ or } q < m \text{ and } 1 \leq s \leq u(q)\} \tag{2.4}$$

- $P\{AW_{t,m} \neq W_{t,m}\} \leq \beta([\,[2^{m\pi}]\,])$. (2.5)

A simple induction argument shows that $\{AW_{t,m}\}$ is a (double) sequence of independent random vectors. From now on we will drop the double bracket symbol from quantities like $[\,[2^{m\pi}]\,]$. In other words we will assume that $[\,[2^{m\pi}]\,] = 2^{m\pi}$. It is easy, though rather tedious, to show that it does not represent any loss of generality so far as our results are concerned. Now (2.5) and Borel–Cantelli imply

$$P\{AW_t \neq W_{t,m} \ \text{ i.o.}\} = 0, \tag{2.6}$$

provided

$$\Sigma\ 2^{m(1-\gamma)}\beta(2^{m\pi}) < \infty,$$

which holds true if

$$1-\gamma < \nu\pi. \tag{2.7}$$

A similar argument shows that there exists a double sequence of random vectors $\{AV_{t,m}\}$, $m \geqslant 0$, $1 \leqslant t \leqslant u(m)$ such that

- $AV_{t,m}$ and $V_{t,m}$ are identically distributed (2.8)
- $\{AV_{s,m},\ m \geqslant 0,\ 1 \leqslant s \leqslant u(m)\}$ is a collection of independent random vectors (2.9)
- $P\{AV_{t,m} \mp V_{t,m}\} \leqslant \beta(2^{m\gamma})$ (2.10)
- $P\{AV_{t,m} \mp V_{t,m}\ \text{i.o.}\} = 0$ if (2.7) holds true (notice $\pi < \gamma$). (2.11)

Now let us denote

$$G_{t,m} = \{a_{t,m},\ a_{t,m}+1, \ldots, b_{t,m}\},$$
$$P_{t,m} = \{c_{t,m},\ c_{t,m}+1, \ldots, d_{t,m}\}.$$

Let n be such that $2^{m-1} < n \leqslant 2^m$. We have

$$\begin{aligned}
|\tilde{S}_n - \Sigma' AV_{s,q} - \Sigma'' AW_{s,q}| &\leqslant \Big|\sum_{q=0}^{m-1}\sum_{s=1}^{u(q)} V_{s,q} - AV_{s,q}\Big| \\
&+ \Big|\sum_{q=0}^{m-1}\sum_{s=1}^{u(q)} W_{s,q} - AW_{s,q}\Big| + \max_{1\leqslant t\leqslant u(q)} \Big|\sum_{s=1}^{t} V_{s,m} - AV_{s,m}\Big| \\
&+ \max_{1\leqslant t\leqslant u(q)} \Big|\sum_{s=1}^{t} W_{s,m} - AW_{s,m}\Big| \\
&+ \max_{1\leqslant t\leqslant u(q)}\ \max_{c_{t,m}\leqslant s\leqslant d_{t,m}} \Big|\sum_{k=c_{t,m}}^{s} X_k\Big| \\
&+ \max_{1\leqslant t\leqslant u(q)}\ \max_{a_{t,m}\leqslant s\leqslant b_{t,m}} \Big|\sum_{k=a_{t,m}}^{s} \tilde{X}_k\Big| \\
&+ \sum_{q=0}^{m-1} |V_{u(q)+1,q}| + \max_{b_{u(m),m}\leqslant s\leqslant 2^m} \Big|\sum_{k=1+b_{u(m),m}}^{2^m} X_k\Big|
\end{aligned}$$

$$= \text{(I)} + \text{(II)} + \ldots + \text{(VIII)},$$

where $\Sigma'$ ($\Sigma''$, respectively) stands for the summation over the pairs (s,q) such that $c_{s,q} \leq n$ ($a_{s,q} \leq n$, respectively). Now if (2.7) holds true then (2.6) and (2.11) imply

$$\text{(I)} + \ldots + \text{(IV)} = O(1) = o(n^{1/2(1+\theta)}) \quad \text{a.s.} \tag{2.12}$$

Now choose $\delta > 0$

$$\max_{j\leq 2^{m\pi}} P\{|X_j + \ldots + X_{2^{m\pi}}| > \delta 2^{m/2(1+\theta)}\}$$

$$\leq \max_{j\leq 2^{m\gamma}} P\{|X_j + \ldots + X_{2^{m\gamma}}| > \delta 2^{m/2(1+\theta)}\}$$

$$\leq \delta^{-2} \max_{j\leq 2^{m\gamma}} E|X_1 + \ldots + X_{2^{m\gamma}}|^2/2^{m/(1+\theta)} .$$

Standard arguments (similar to those employed in the proof of Lemma 2–A–5) can be used to show that

$$E|X_1 + \ldots + X_{2^{m\gamma}}|^2 = O(2^{m\gamma}).$$

Whence if m is large enough and

$$\gamma < \frac{1}{1+\theta} , \tag{2.13}$$

we obtain

$$\max_{j\leq 2^{m\pi}} P\{|X_j + \ldots + X_{2^{m\pi}}| > \delta 2^{m/2(1+\theta)}\} < \tfrac{1}{2} . \tag{2.14}$$

Now Lemmas 2–A–3 and 2–A–4 and (2.14) imply

$$P\{\max_{1\leq k\leq 2^{m\pi}} |X_1 + \ldots + X_k| > \delta 2^{m/2(1+\theta)}\}$$

$$\leq P\{\max_{1\leq k\leq 2^{m\gamma}} |X_1 + \ldots + X_k| > \delta 2^{m/2(1+\theta)}\}$$

$$= O(1/2^{m(1-\gamma)}m^{1+h}) , \tag{2.15}$$

where $h > 0$, provided

$$\frac{1}{\nu-1} < \gamma < \min\{1, \frac{\lambda-1}{2\lambda(1+\theta)}, \frac{1}{1+\theta} - \frac{1}{\nu}, \frac{\nu-1-\theta}{(1+\nu)(1+\theta)}, \frac{\nu(\lambda-1)-2\lambda(1+\theta)}{\nu+2\lambda(1+\theta)}\} \quad (2.16)$$

Now stationarity and (2.15) imply

$$\begin{aligned} &\sum_{m=1}^{\infty} P\{ \max_{1\le t\le u(m)} \max_{c_{t,m}\le s\le d_{t,m}} |\sum_{k=c_{t,m}}^{s} X_k| > \delta 2^{m/2(1+\theta)} \\ &\quad \le \sum_{m=0}^{\infty} 2^{m(1-\gamma)} P\{ \max_{1\le k\le 2^{m\pi}} |X_1 + \ldots + X_k| > \delta 2^{m/2(1+\theta)}\} \\ &\quad < \infty \end{aligned} \quad (2.17)$$

Borel–Cantelli and arbitrariness of $\delta$ imply

$$(IX) = \max_{1\le t\le u(m)} \max_{c_{t,m}\le s\le d_{t,m}} |\sum_{k=c_{t,m}}^{s} X_k| = o(2^{m/2(1+\theta)}),$$

almost surely. Now as $(V) \le (IX) = o(2^{(m-1)/2(1+\theta)})$, almost surely and $n \ge 2^{(m-1)}$ we obtain

$$(V) = o(n^{1/2(1+\theta)}) \quad \text{a.s.} \quad (2.18)$$

Now notice that

$$\begin{aligned} (VI) \le &\max_{1\le t\le u(q)} \max_{a_{t,m}\le s\le b_{t,m}} |\sum_{k=a_{t,m}}^{s} X_k| \\ &+ \sum_{k=1}^{n} |X_k - \tilde{X}_k| . \end{aligned}$$

The first term in the above inequality is $o(n^{1/2(1+\theta)})$ almost surely by the same arguments we used to deal with (V). The second one is $o(n^{1/2(1+\theta)})$ by virtue of Lemma 2–B–1. Thus

$$(VI) = o(n^{1/2(1+\theta)}) \quad \text{a.s.}$$

As far as (VII) is concerned it is not difficult to show that (2.13) yields

$$\sum_{m} P\{|V_{u(m)+1,m}| > \delta 2^{m/2(1+\theta)}\} < \infty,$$

for any $\delta > 0$.

Therefore (Borel–Cantelli)

$$\sum_{k=1}^{m} |V_{u(m)+1,m}| \leq C(\omega) + \sum_{k=1}^{m} \delta 2^{k/2(1+\theta)}$$

$$\leq C(\omega) + 2\delta\, 2^{m/2(1+\theta)}, \qquad \text{a.s.}$$

which yields

$$\sum_{k=1}^{m} |V_{u(k)+1,k}| = o(2^{m/2(1+\theta)}) \qquad \text{a.s.}$$

whence

$$(\mathrm{VII}) = o(n^{1/2(1+\theta)}) \qquad \text{a.s.}$$

Finally, it is not difficult to show that (VIII) = $o(n^{1/2(1+\theta)})$ almost surely using Lemma 2–A–3. Thus we conclude that

$$|\tilde{S}_n - \Sigma' AV_{s,q} - \Sigma'' AW_{s,q}| \leq (\mathrm{I}) + \ldots + (\mathrm{VIII})$$
$$= o(n^{1/2(1+\theta)}) \qquad \text{a.s.} \tag{2.19}$$

Now notice that by Kolmogorov's inequality (multivariate)

$$P\{ \max_{1 \leq s \leq u(m)} |\sum_{k=1}^{s} AV_{k,m}| > \delta\, 2^{m/2(1+\theta)} \}$$

$$\leq d \sum_{s=1}^{u(m)} E|AV_{s,m}|^2 / \delta^2 2^{m/(1+\theta)}$$

$$<< 2^{m(1-\gamma)}\, 2^{m\pi} / 2^{m/(1+\theta)} . \tag{2.20}$$

(2.20), and some standard manipulations imply that

$$\Sigma' AV_{s,q} = o(n^{1/2(1+\theta)}) \qquad \text{a.s.} \tag{2.21}$$

provided

$$1-\gamma + \pi < \frac{1}{1+\theta} \quad . \tag{2.22}$$

Now we will show that we can construct independent random vectors $\{A\tilde{Z}_{s,q}\}$, $q \geq 0$, $1 \leq s \leq u(q)$ such that $A\tilde{Z}_{s,q}$ is normally distributed with zero mean and $\mathrm{cov}(A\tilde{Z}_{s,q}) = \mathrm{cov}(AW_{s,q})$. Such random vectors will be shown to satisfy

$$\sum_{m=1}^{\infty} P\{ \max_{1\leq s\leq u(m)} |\sum_{k=1}^{s} AW_{k,m} - A\tilde{Z}_{k,m}| > \epsilon 2^{m/2(1+\theta)}\} < \infty$$

for any $\epsilon > 0$. Let us assume that

$$\gamma < (\lambda-1)2\lambda(1+\theta) \quad . \tag{2.23}$$

Take $\alpha$ such that

$$\gamma + \frac{1}{\lambda(1+\theta)} < \alpha < 1-\gamma \ ,$$

and (2.24)

$$\alpha + \gamma < \frac{1}{1+\theta} \ .$$

Assume without loss of generality that $2^{m\alpha}$ is integer,

$$\Gamma = EX_1X_1^T + \sum_{k>1} EX_1X_k^T + EX_kX_1^T = I_d \ . \tag{2.25}$$

Now Lemma 2–A–6 implies, in particular, that

$$\|\Gamma - 2^{-m\gamma/2}\mathrm{Cov}(AW_{k,m})\| < 1/2 \tag{2.26}$$

if $m \geq m_0$. Let us denote $N = 2^{m\alpha}$, $L = 2^{m(1-\gamma)}/N$. Also, let us define for $1 \leq j \leq L$

$$\rho_j = \varphi_j^{-3/2} \sum_{k=(j-1)N+1}^{jN} E|2^{-m\gamma/2}AW_{k,m}|^3 \ ,$$

where $\varphi_j$ ($\psi_j$, respectively) is the smallest (largest) eigenvalue of $H_j = \sum_{(j-1)N+1}^{jN} \mathrm{cov}(2^{-m\gamma/2}AW_{k,m})$. Now (2.26) yields

$$\tfrac{1}{2} N \leq \varphi_j \leq \psi_j \leq \tfrac{3}{2} N, \qquad 1 \leq j \leq L. \tag{2.26}$$

On the other hand

$$\begin{aligned} E|2^{-m\gamma/2} AW_{k,m}|^3 &\leq 2^{-m\gamma/2} 2^{m\gamma} 2^{m/2\lambda(1+\theta)} E|2^{-m\gamma/2} AW_{k,m}|^2 \\ &<< 2^{m\gamma/2} 2^{m/2\lambda(1+\theta)}, \end{aligned} \tag{2.27}$$

by Lemma 2–A–6. Now (2.27) and (2.26) imply

$$\begin{aligned} \rho_j &<< N^{-\frac{1}{2}} 2^{m\gamma/2} 2^{m/2\lambda(1+\theta)} \\ &<< [2^{-m\alpha+m\gamma+m/\lambda(1+\theta)}]^{\frac{1}{2}}, \end{aligned} \tag{2.28}$$

for all $1 \leq j \leq L = L(m)$.

(2.28) and (2.24) show in particular that $\rho_j = \rho_j(m) = o(1)$. Now take $\epsilon > 0$ such that

$$1 < (1 - \epsilon/2)(\alpha - \gamma - 1/\lambda(1+\theta)) + 1/(1+\theta). \tag{2.29}$$

Such a choice is possible because of (2.23) and (2.24). As $\rho_j = o(1)$ we obtain

$$2.2^{m/2\lambda(1+\theta)}.2^{m\gamma/2} \leq \frac{\epsilon}{14\sqrt{d}} (\varphi_j \log (1/\rho_j))^{\frac{1}{2}},$$

for all $m \geq m_1$ (say) and $1 \leq j \leq L = L(m)$. Whence

$$|2^{-m\gamma/2} AW_{k,m}| \leq \frac{\epsilon}{14\sqrt{d}} (\varphi_j \log (1/\rho_j))^{\frac{1}{2}}$$

Using Lemma 2–B–2 we obtain independent random vectors $AZ_{k,m}$ such that

$$AZ_{k,m} \sim N(0, \mathrm{cov}(2^{-m\gamma/2} AW_{k,m})) \tag{2.30}$$

and

$$\begin{aligned} E\Big[\Big|2^{-m\gamma/2} \sum_{k=(j-1)N+1}^{jN} AW_{k,m} - \sum_{k=(j-1)N+1}^{jN} AZ_{k,m}\Big|^2\Big] \\ \leq D_\epsilon\, \psi_j\, \rho_j^{2-\epsilon}, \end{aligned} \tag{2.31}$$

for all $1 \leq j \leq L$ and $m \geq \max\{m_0, m_1\}$. Now (2.28) and (2.26) imply that the right hand side of (2.31) can be bounded from above by

$$\tau_m = D_\epsilon' \, 2^{m\alpha}\{2^{-m\alpha+m\gamma+m/\lambda(1+\theta)}\}^{(1-\epsilon/2)} . \tag{2.32}$$

Take $\delta > 0$ and define for $1 \leq t \leq 2^{m(1-\gamma)}$

$$AS_t = \sum_{k=1}^{t} AW_{k,m} ,$$

$$AT_t = \sum_{k=1}^{t} AZ_{k,m} .$$

From Kolmogorov's inequality (multivariate) we obtain

$$P\{\max_{1\leq j\leq l} |2^{-m\gamma/2}AS_{jN} - AT_{jN}| > \delta 2^{m/2(1+\theta)}2^{-m\gamma/2}\}$$

$$\leq \frac{d}{\delta^2} \cdot 2^{-m/(1+\theta)}2^{m\gamma} \sum_{j=1}^{L} E|R_j|^2, \tag{2.33}$$

where we have put

$$R_j = 2^{-m\gamma/2} \sum_{k=(j-1)N+1}^{jN} AW_{k,m} - \sum_{k=(j-1)N+1}^{jN} AZ_{k,m} .$$

By (2.32) we can bound the right hand side of (2.33) by

$$\tau_m' = \frac{d}{\delta^2} D_\epsilon' \, 2^{-m/(1+\theta)}2^m[2^{-m\alpha+m\gamma+m/\lambda(1+\theta)}]^{(1-\epsilon/2)} .$$

From (2.29) we obtain $\Sigma\, \tau_m' < \infty$. Therefore

$$\sum_m P\{\max_{1\leq j\leq L(m)} |AS_{jN} - 2^{m\gamma/2}AT_{jN}| > \delta 2^{m/2(1+\theta)}\} < \infty \tag{2.34}$$

From Kolmogorov inequality we obtain

$$P\{\max_{(j-1)N\leq K\leq jN} |AS_k - AS_{(j-1)N}| > \delta 2^{m/2(1+\theta)}\}$$

$$\leq \frac{d}{\delta^2} \frac{1}{2^{m/(1+\theta)}} N\, E|AW_{1,m}|^2 << N2^{m\gamma}2^{-m/(1+\theta)}$$

$$<< 2^{m(\alpha+\gamma)}2^{-m/(1+\theta)}) , \tag{2.35}$$

where the second inequality holds true by virtue of Lemma 2-A-6. Therefore if m is large enough (2.24) yields

$$\max_{(j-1)N<k\le jN} P\{|AS_{jN} - AS_k| > \tfrac{\delta}{2}\, 2^{m/2(1+\theta)}\}$$

$$\le P\{\max_{(j-1)N<k\le jN} |AS_k - AS_{(j-1)N}| > \tfrac{\delta}{2}\, 2^{m/2(1+\theta)}\}$$

$$< \tfrac{1}{2}.$$

Ottaviani inequality yields (if m is large enough)

$$P\{\max_{(j-1)N<k\le jN} |AS_k - AS_{(j-1)N}| > \delta\, 2^{m/2(1+\theta)}\}$$

$$\le 2\, P\{|AS_{jN} - AS_{(j-1)N}| > \tfrac{\delta}{2}\, 2^{m/2(1+\theta)}\}$$

$$\le 2\, P\{|R_j| > \tfrac{\delta}{4}\, 2^{m/2(1+\theta)} 2^{-m\gamma/2}\}$$

$$+ 2\, P\{|AT_{jN} - AT_{(j-1)N}| > \tfrac{\delta}{4}\, 2^{m/2(1+\theta)} 2^{-m\gamma/2}\}. \qquad (2.36)$$

Hence

$$P\{\max_{1\le j\le L} \max_{(j-1)N<k\le jN} |AS_k - AS_{(j-1)N}| > \delta\, 2^{m/2(1+\theta)}\}$$

$$\le L\, P\{\max_{(j-1)N<k\le jN} |AS_k - AS_{(j-1)N}| > \delta\, 2^{m/2(1+\theta)}\}$$

$$\ll \tau_m' + L\, P\{|AT_{jN} - AT_{(j-1)N}| > \tfrac{\delta}{4}\, 2^{m/2(1+\theta)} 2^{-m\gamma/2}\}$$

$$\ll \tau_m', \qquad (2.37)$$

where the last inequality can be obtained by an exponential inequality for normal random vectors and the fact that

$$E|AT_{jN} - AT_{(j-1)N}|^2 = o(2^{-m/1+\theta} 2^{m\gamma} 2^{-m\epsilon'})$$

for some small $\epsilon' > 0$ (note (2.30) and Lemma 2-A-6). Also, by making use of the facts just mentioned and Ottaviani inequality we can easily show that

$$P\{\max_{1\le j\le L}\ \max_{(j-1)N<k\le jN} |AT_k - AT_{(j-1)N}| > \delta\, 2^{-m\gamma/2} 2^{m/2(1+\theta)}\}$$

$$\ll \tau_m' . \tag{2.38}$$

The next step is to deconvolve each random vector $2^{m\gamma/2}AZ_{k,m}$ into a sum of independent homoscedastic normal random vectors. Namely write

$$2^{m\gamma/2}AZ_{k,m} = \sum_{j=a_{k,m}}^{b_{k,m}} Z_j ,$$

where $\mathrm{cov}(Z_j) = \mathrm{cov}(AZ_{1,m})$, if $j \in J_m$. We can choose our Z's in such a way that

$$\{Z_j;\ a_{k,m} \le j \le b_{k,m} \text{ for some } k,m\}$$

is a collection of independent random vectors (see Section 2.5.1). Adjoin to these collections (just for completeness) a further collection of zero mean independent Gaussian random vectors

$$\{Z_j;\ c_{k,m} \le j \le d_{k,m} \text{ for some } k,m\},$$

such that this second collection is independent of the first one and

$$\mathrm{cov}(Z_j) = \mathrm{cov}(AZ_{1,m}) \text{ if } j \in J_m .$$

Now, provided that we can choose $\gamma$ and $\pi$ satisfying (2.7), (2.13), (2.16) and (2.22) we only need to show

$$\infty > \sum_m P\{\max_{1\le k\le 2^{m(1-\gamma)}}\ \max_{a_{k,m}\le j\le b_{k,m}} |\sum_{i=a_{k,m}}^{j} \tilde{X}_i| > \delta\, 2^{m/2(1+\theta)}\},$$

$$\infty > \sum_m P\{\max_{1\le k\le 2^{m(1-\gamma)}}\ \max_{a_{k,m}\le j\le b_{k,m}} |\sum_{i=a_{k,m}}^{j} Z_i| > \delta\, 2^{m/2(1+\theta)}\},$$

$$\infty > \sum_m P\{\max_{1\le k\le 2^{m(1-\gamma)}} |\sum_{j=1}^{k} \sum_{i=c_{k,m}}^{d_{k,m}} Z_i| > \delta\, 2^{m/2(1+\theta)}\},$$

and

$$\infty > \sum_m P\{\max_{1\le k\le 2^{m(1-\gamma)}}\ \max_{c_{k,m}\le j\le d_{k,m}} |\sum_{i=a_{k,m}}^{j} Z_i| > \delta\, 2^{m/2(1+\theta)}\}.$$

The first relation above is a direct consequence of Lemma 2–A–4. The fourth one is a consequence of the second. The second and third can be shown by combining Ottaviani inequality and exponential bounds for normal random vectors.

Finally it is easy to show that our conditions c, d, e and f imply the existence of $\pi$ and $\gamma$ satisfying (2.7), (2.13), (2.16) and (2.22). Also, relation (ii) is a trivial consequence of Lemma 2–A–6. □

# CHAPTER 3

## 3.1 Introduction

Our main goal in this chapter is to improve and generalize the following result due to Akonom and Gourieroux:

Theorem (See Akonom and Gourieroux(1987)) Let $\{\eta_t\}$, $t \in Z$ be a sequence of iidrv's. Define $\varepsilon_t = \eta_t I\{\ t>0\ \}$. Let $d>1/2$ and assume that $\varepsilon_1$ has absolute moment of order p, where p is strictly greater than $\max\{2,2/(2d-1)\}$. Consider the process defined by

$$X_t = (1-L)^{-d} A^*(L)\varepsilon_t \ ,$$

where L is the lag operator ( $LZ_t = Z_{t-1}$ ) and

$$A^*(L) = \sum_0^{\infty} a_k^* L^k \ ,$$

with $|a_k^*| = O(\rho^{+k})$, for some $0<\rho<1$. Then the process

$$\left[\ X_T^*(r) = T^{(1-2d)/2}\, X_{[\![Tr]\!]},\ r \in [0,1]\ \right]$$

weakly converges to

$$\left[\ \sigma A^*(1)/\Gamma(d) \int_0^r (r-s)^{d-1} dB(s),\ r \in [0,1]\ \right],$$

where B(.) is a standard Brownian motion in [0,1] and $Var(\varepsilon_1) = \sigma^2$. □

The process $\{X_t\}$ is a modification of the so-called fractional ARIMA processes. Such a modification consists on assuming $\varepsilon_t=0$ if $t \leqslant 0$ and was imposed so as to deal with nonstationarity ( which is a consequence of $d>1/2$ ) in a convenient way.

Our results, to be described below, generalize Akonom and Gourieroux's theorem insofar as we allow for a weak dependent sequence of innovations. Also we consider a less restrictive rate of decay for the sequence $\{a_k\}$ defined above. Finally we allow for trending heteroskedasticity in the innovation sequence.

On the other hand we improve the above mentioned result by showing that a natural smoothing of $X_T(.)$ satisfies a functional limit theorem under minimal moment assumptions on the innovation sequence. Functional limit theorems like ours can be applied when deriving tests in regression models under nonstationarity assumptions ( See Gourieroux, Maurel and Monfort(1987) ).

The rest of the chapter is divided as follows. In section 3.2 we make a partial review of the literature concerning fractionally differenced linear processes. In section 3.3 we discuss some intuitive aspects of our results. A proof of our first generalization of Akonom and Gourieroux's result is presented in section 3.4. Finally, some generalizations are discussed in section 3.5.

## 3.2 Fractionally Differenced Linear Processes

In the early seventies the field of time series experienced a big boost in both theoretical and applied terms. As far as the applications are concerned the book of Box and Jenkins(1970) is by far the most important element for the popularization of time series models. In their book they present a comprehensive ( though a bit informal ) study of the so-called ARIMA models. Such a class of models is the subject of study of an enormous number of works.

We define below ARMA and ARIMA models just to set some notation. Let $\{\varepsilon_t\}$,

$t \in Z$ be a sequence of iidrv's such that $E\varepsilon_1=0$ and $Var(\varepsilon_1)= \sigma^2$. We say that $\{X_t\}$ is a ARMA(p,q) process if we can write

$$P(L)X_t = Q(L)\varepsilon_t \ , \ t \in Z,$$

where

$$P(L)= 1 - a_1L - \ldots - a_pL^p \ ,$$

$$Q(L)= 1 + b_1L + \ldots + b_qL^q \ ,$$

and L is the lag operator. It is well known that if the roots of $P(z)=0$ lie outside the unit circle then $\{X_t\}$ as above is stationary. An ARIMA process, in turn, is non-stationary though it keeps the simplicity, so far as the functional describing its evolution through time is concerned, of ARMA processes. We say that $\{W_t\}$ is a ARIMA process ( where d is a non-negative integer ) if $Z_t = (1-L)^dW_t$ is a ARMA(p,q) process. Let us just mention that a ARMA(p,q) process is a linear process.

Once we abandon the class of linear processes, the range of parametric time series models (and their associated processes) grows uncountably. Examples of such non-linear processes are bilinear models (Granger and Andersen (1978)), threshold autoregressive models (Tong and Lim (1980)), amplitude dependent autoregressive models (Hagan and Ozaki (1981)), state dependent models (Priestley (1980)) and many others.

In practice one should either have theoretical justification or to rely on a kind of selection procedure so as to choose a particular non-linear model. It seems sensible that such a procedure should be nonparametric in nature. A partially successful attempt in this direction is provided by the state dependent formulation of some non-linear time series models ( See Haggan et al (1984) for details ).

On the other hand it is debatable whether such a two step procedure should be preferable to a direct nonparametric approach. We refer the reader to Robinson (1983) for a study on nonparametric estimates of regression and density functions under weak dependent samples. For an account on nonparametric methods in specification see Robinson (1986).

Although there exists evidence that a number of popular real time series follow nonlinear models (for instance the sunspot numbers and the Lynx Cannadensis data, see e.g. Haggan et al (1984) and Robinson (1983)), the class of linear time series models is still appropriate for a variety of real data collections.

Now let us consider a sample $\{ X_1,\ldots,X_T \}$, drawn from a linear process. Very loosely, we could describe an estimation problem as a procedure of choosing the particular model which generated the sample based only on the information provided by the sample. If we do not restrict the space of possible choices our problem is nonparametric (again, very informally) since the space of our possible choices is infinite dimensional.

When we reduce the space of possible choices to the class of ARMA(p,q) processes ( p and q fixed ) we very much restrict the generality of the family of linear models inside which the true process is assumed to be. Such a restriction occurs, to begin with, insofar as we are imposing an exponential rate of decrease for the coefficients of the true model which generates our sample. From a <u>purely technical point of view</u> it would be desirable to consider a parametric model that could accomodate linear processes with algebraic rate of decrease for the coefficients. Loosely speaking such a model would be able to exploit the generality of the class of linear models in better terms than ARMA models.

A possible parametric model which allows for algebraic rates for the coefficients of linear processes can be obtained from when we consider fractionally differenced ARMA processes. Let us define the operator $(1-L)^{-d}$ by

$$(1-L)^{-d} = \Gamma(d)^{-1} \sum_{k=0}^{\infty} \{ \Gamma(k+d)/\Gamma(k+1) \} L^{k}, \; d \neq 0, -1, -2, \ldots$$

where $\Gamma(.)$ is Euler's gamma function

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x} dx, \; p>0.$$

When $p<0$ and p is not integer we define $\Gamma(p)$ recursively by

$$\Gamma(p) = p^{-1}\Gamma(p+1).$$

We say that $\{W_t\}$ is a fractional ARIMA(p,d,q) process if

$$(1-L)^{d}P(L)W_t = Q(L)\varepsilon_t \; ,$$

where $\{\varepsilon_t\}$ is a white noise with finite variance and P and Q are polynomials of order p and q respectively, with roots outside the unit circle. If $-1/2<d<1/2$ then $\{W_t\}$ is stationary and can be written as

$$W_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k} \; ,$$

where $a_k = O(k^{d-1})$, as k goes to infinity. If on the other hand $d>1/2$ then $\{W_t\}$ is nonstationary.

In other words we say that $\{W_t\}$ is a fractional ARIMA(p,d,q) if the fractional difference of order d of $\{W_t\}$ follows a ARMA(p,q) process. Fractionally differenced processes were introduced by Mandelbrot and Van Ness(1968). Asymptotic properties of stationary fractional processes have been studied by a number of authors ( See e.g. Adenstedt(1974), Rosenblatt(1976), Granger(1978), Granger and Joyeux(1980), Hosking(1981) )

Models involving fractional processes have been recently the subject of a number of works in the econometric literature. As argued in Akonom and Gourieroux(1988) and Granger (1980) what makes fractional models extremely appealing for economic modelling is the fact that this type of process is naturally introduced when we consider the aggregation of heterogeneous time series. Moreover, nonstationary fractional models (see below) provide a smoother and broader description of the intrinsically nonstationary nature of macroeconomic data than the traditional ARIMA models. Further, fractional models seem to be especially suitable for the study of models with cointegrated time series (see Granger(1986) and Engle and Granger (1987)). We refer the reader to Robinson (1990) for a discussion on rececnt developments on estimation and testing issues on fractional time series models.

## 3.3 An informal look at the functional limit theorem

In this section we present some intuitive background for the functional limit theorem we prove later on. We will consider a variant of the nonstationary process discussed in section 3.2. Let $\{\eta_t\}$, $t \in Z$ be a $\mathbb{R}$-valued stationary process, such that $E\eta_1=0$, $Var(\eta_1)= 1$. We will consider the evolution of a process which is quite similar (though not equivalent) to the nonstationary fractional process ARIMA(p,d,q), where $d>1/2$. Let us define $\{X_t\}$ , $t \in Z$, by

$$(1-L)^d X_t = A(L)\epsilon_t,$$

where $\epsilon_t = \eta_t I\{ t>0 \}$ and $A(L)= \Sigma_k \alpha_k L^k$.

The coefficients of A(.) will be considered in detail later on. For the moment let us just assume that $\{\alpha_k\}$ is absolutely summable. Now let us compare the process $\{X_t\}$ with the corresponding (say) usual fractional process given by

$$(1-L)^d Y_t = A(L)\eta_t.$$

Let us suppose for the sake of simplicity that $A(L)\equiv 1$. Also, let us assume that

$EY_0 = 0$. If d=1 then we can write

$$X_t = \sum_{k=1}^{t} \varepsilon_k, \quad \text{and}$$

$$Y_t = Y_0 + \sum_{k=1}^{t} \varepsilon_k .$$

Clearly, in this case, asymptotics for $\{X_t\}$ and $\{Y_t\}$ will follow the same pattern. For instance, normalizing factors in central limit theorems are the same and so forth.

Statistical models based on processes like $\{X_t\}$ are sometimes much more realistic than those based on $\{Y_t\}$. This is particularly true when dealing with economic data. The reason is simple: the process ( or the data generation mechanism ) simply did not exist prior to time t=0.

Now let us go back to our heuristic discussion on the functional limit theorem. As $\varepsilon_t=0$ if $t\leqslant 0$ we can write

$$X_t = (1-L)^{-d}A(L)\varepsilon_t .$$

Let

$$(1-L)^{-d}A(L) = \sum_{k=0}^{\infty} \pi_k^{(d)} L^k,$$

be the formal representation of $A(L)(1-L)^{-d}$. We can write

$$X_t = \sum_{k=1}^{t} \pi_{t-k}^{(d)} \varepsilon_k .$$

Let us define for each $T\geqslant 1$,

$$X_T^*(r) = T^{1/2-d} X_{[\![Tr]\!]} , \quad r \in [0,1].$$

We aim at proving a functional limit theorem for the sequence of D[0,1]-valued random elements just defined. We will show that

$$X_T^*(.) \Longrightarrow W(.) \ ,$$

where

$$W(r)= c \int_0^r (r-s)^{d-1} \, dB(s) \ , \ r \in [0,1],$$

where

$$\sigma^2 = E\varepsilon_1^2 + 2 \sum_{k=2}^{\infty} E\varepsilon_1\varepsilon_k \ ,$$

which is assumed to be a positive real number, and with $c=A(1)\sigma/\Gamma(d)$. Also, $\{B(t), \ t \in [0,1]\}$ is a standard Brownian motion, and the integral above is defined in the sense of Ito (actually we could use a less powerful definition for the stochastic integral above but this would not make our proofs either easier or shorter). Now let us denote

$$S_t = \sum_{k=1}^{t} \varepsilon_k .$$

We can write

$$X_T^*(r) = T^{1/2-d} \sum_{k=1}^{[\![Tr]\!]} \pi^{(d)}_{[\![Tr]\!]-k} ( S_k - S_{k-1} ).$$

Now we can show that

$$\pi^{(d)}_{t-k} \approx c(t-k)^{d-1}/\sigma.$$

Now using a suitable approximation theorem ( for instance Komlos, Major and Tusnady's or our corollary 2.1 ) we can write

$$X_T^*(r) \approx T^{1/2-d} \sum_{k=1}^{[\![Tr]\!]} c \ ( [\![Tr]\!]-k )^{d-1} \ ( B(k)-B(k-1) ),$$

where $\{ B(t), \ t \geqslant 0 \}$ is a standard Brownian motion. The distribution of the above process equals to that of

$$T^{1-d} \sum_{k=1}^{[\![Tr]\!]} c \ ( [\![Tr]\!]-k )^{d-1} \ ( B( k/T )-B( (k-1)/T ) )$$

$$\approx c \sum_{k=1}^{[\![Tr]\!]} ( r-k/T )^{d-1} ( B( k/T )-B( (k-1)/T ) ).$$

Finally the sum above is an approximation to the stochastic integral

$$c \int_0^r (r-s)^{d-1} dB(s) .$$

The justification of the above equivalences are done in the next section. Let us now consider a further process. Define

$$R_T^*(r) = T^{d+1/2} \left[ \sum_{k=1}^{[\![Tr]\!]} X_k \right] , \ r \in [0,1]$$

It is not difficult to see that

$$R_T^*(r) \approx \int_0^r X_T^*(s)\, ds.$$

Therefore a simple application of the continuous mapping theorem yields

$$R_T^*(r) \Rightarrow c \int_0^r \int_0^u ( u-s )^{d-1} dB(s)\, du.$$

However as we will see, the above mentioned almost sure approximation theorem depends on the existence of moments of order higher than two for $\varepsilon_1$ ( even when $\{\varepsilon_t\}$ is iid ). On the other hand we can write

$$R_T^*(r) = T^{-1/2-d} R_{[\![Tr]\!]} , \ r \in [0,1].$$

where

$$R_t = \sum_{k=1}^{t} X_k = (1-L)^{-1-d} A(L) \varepsilon_k.$$

If we follow the same lines outlined above we can show (at least for the iid case) that the last weak convergence theorem holds valid with the minimal assumption that moments of order two for the innovation process exist ( See section 3.5 ).

## 3.4 A Functional Limit Theorem for $X^*$

Throughout we follow the notation and definitions of section 3.3. Our main goal in this section is to prove the following

Theorem 3.1 Let us assume that $\{\eta_t\}$ is a zero mean stationary, $\mathbb{R}$-valued, absolutely regular process with sequence of mixing weights given by $\{\beta_k\}$. Suppose that there exists $0<\nu$, $1<\lambda$, $0<\theta<1$ such that

(A) $\beta_k = k^{-\nu}$

(B) $E\ |\eta_1|^{2\lambda(1+\theta)} < \infty$.

Assume that all the conditions of theorem 2.1 hold valid and

(C) $\alpha_k = O(\ k^{-4}\ )$ .

Then if $2(1+\theta) > \max\{2,\ 2/(2d-1)\}$ we have

$$X_T(.) \implies \left[\ c \int_0^r (\ r-s\ )^{d-1}\ dB(s)\ ,\quad r \in [0,1]\ \right]. \quad \square$$

Let us just make a few comments before we start proving theorem 3.1. First of all, the stochastic integral above is defined in Ito's sense. We refer the reader to Chung and Williams(1983) for an exposition on the theory of stochastic integrals.

Secondly, we note that all the stochastic processes we consider in this chapter take their values in D[0,1], the space of real functions defined on [0,1] which are right continuous (cad) and have finite left limits (lag). D[0,1] is also called the space of cadlag functions. The paths of $X^*(.)$ are obviously elements of D[0,1]. We will show below that the paths of W(.) are in D[0,1] with probability one.

We adopt the most popular definition of weak convergence. Namely we say that a sequence of D[0,1]-valued random elements $\{\ V_n(.)\ \}$ weakly converges to V(.), whose pathes are D[0,1]-elements with probability one if

$$\lim_n Ef(V_n) = Ef(V),$$

for all f: $D[0,1] \to \mathbb{R}$, bounded and continuous with respect to the Skorohod topology in D[0,1]. We refer the reader to Billingsley(1968) for details ( See also Pollard(1984) for an alternative approach to weak convergence ).

Now in order to show that $V_n(.) \Longrightarrow V(.)$ we must prove

(a) $(V_n(t_1), \ldots, V_n(t_k))$ converges in distribution to

$(V(t_1), \ldots, V(t_k))$, for all $0 \leqslant t_1 < t_2 < \ldots < t_k \leqslant 1$, $r \geqslant 1$.

(b) There exist positive constants $\alpha$, $\beta$, D, independent of n such that

$$E\{ | V_n(r) - V_n(q) |^{\alpha} | V_n(s) - V_n(r) |^{\alpha} \} \leqslant$$

$$\leqslant D | s-q |^{1+\beta}, \quad \text{all } 0 \leqslant q < r < s \leqslant 1.$$

(c) The trajectories of V(.) are D[0,1]-elements with probability one.

Now if $V_n(.)$ and V(.) are Gaussian processes the verification of (a), (b) and (c) above is quite simple. Indeed in order to verify (a) we only have to show that

(a1) $\lim_n EV_n(t) = EV(t)$ , all t.

(a2) $\lim_n E[ V_n(r).V_n(t) ] = E[ V(r) V(t) ]$, all r, t.

On the other hand, Schwartz inequality yields

$$E\{ | V_n(r) - V_n(q) |^{\alpha} | V_n(s) - V_n(r) |^{\alpha} \}$$

$$\leqslant E^{1/2} | V_n(r) - V_n(q) |^{2\alpha} E^{1/2} | V_n(s) - V_n(r) |^{2\alpha}$$

$$= D_1 [ E| V_n(r) - V_n(q) |^2 E| V_n(s) - V_n(r) |^2 ]^{\alpha/2} ,$$

where $D_1 = E|N(0,1)|^{2\alpha}$. Therefore (b) will follow if we can show that there exist positive constants $\gamma$, $D^*$ such that

(b1) $E|V_n(r) - V_n(q)|^2 E|V_n(s) - V_n(r)|^2 \leqslant D^*|s-q|^\gamma$,

for all $n \geqslant 1$ and $0 \leqslant q < r < s \leqslant 1$ ( take $\alpha > 2/\gamma$, $1+\beta = \alpha\gamma/2$ and $D \leqslant D^* D_1$ in (b)).

Finally, in order to show (c) it is enough to show that W(.) is continuous with probability one, which in turn follows from ( See Prakasa Rao(1987), page 85 )

(d) $E|V(r) - V(q)|^\alpha \leqslant D|s-q|^{1+\beta}$,

for some positive constants $\alpha$, $\beta$ and D and all $0 \leqslant r < q \leqslant 1$. Now the same argument used in (b) shows that (d) holds true provided

(c1) $E|V(r) - V(q)|^2 \leqslant D|s-q|^\gamma$ ,

for some positive D and $\gamma$ and all $0 \leqslant r < q \leqslant 1$.

Our last comment is related to the operator A(L). Notice that A(L) could be, for instance the ratio of two polynomials B(L) and C(L) with roots outside the unit circle. In other words, an ARIMA(p,d,q) is a particular case of the class of processes we are considering here. We can now proceed with the proof of theorem 3.1.

<u>Proof of theorem 3.1</u>: By corollary 2.1 there exists a sequence $\{Z_n\}$ of iid random variables such that $Z_1 \cong N(0,\sigma^2)$ and

$$\left| \sum_{k=1}^{n} \varepsilon_k - Z_k \right| = o( n^{1/2(1+\theta)} ) \text{, a.s.} \tag{3.1}$$

We can write

$$X_T^*(r) = T^{1/2-d} \sum_{k=1}^{[\![Tr]\!]} \pi^{(d)}_{[\![Tr]\!]-k}( S_k - S_{k-1} )$$

$$= Q_{1T}(r) + Q_{2T}(r) + Q_{3T}(r) + Q_{4T}(r),$$

where

$$Q_{1T}(r) = c \sum_{k=1}^{b-1} ( r-k/T )^{d-1} ( V(k)-V(k-1) )/T^{1/2},$$

$$Q_{2T}(r) = \sigma\, T^{1/2-d} \sum_{k=1}^{b-1} \pi_{b-k}^{(d)} \left[ \frac{S_k - S_{k-1}}{\sigma} - ( V(k)-V(k-1) ) \right],$$

$$Q_{3T}(r) = \sigma\, T^{1/2-d} \sum_{k=1}^{b-1} \left[ \pi_{b-k}^{(d)} - \frac{A(1)}{\Gamma(d)} ( Tr-k )^{d-1} \right] (V(k)-V(k-1))$$

$$Q_{4T}(r) = T^{1/2-d} ( S_b - S_{b-1} )$$

where $b = b(r,T) = [\![Tr]\!]$, and $\sigma V(k) = \sum_{j=1}^{k} Z_j$. Our result will follow if

(i) $Q_{1T}(.) \Longrightarrow W(.)$

(ii) $Q_{2T}(.) = op(1)$

(iii) $Q_{3T}(.) = op(1)$

(iv) $Q_{4T}(.) = op(1)$

(i) Proof of $Q_{1T}(.) \Longrightarrow W(.)$.

First of all notice that $Q_{1T}(.)$ is a Gaussian D[0,1]-valued random process. Thus by the discussion right after the statement of theorem 3.1 it suffices to prove that (a1), (a2), (b1) and (c1) hold true.

(a1): obvious.

(a2): take $0 \leqslant r \leqslant s \leqslant 1$. Notice that

$$E\, W(r)W(s) = c^2 \int_0^r ( r-x )^{d-1} ( s-x )^{d-1} dx.$$

If T is large enough we have

$$E\, Q_{1T}(r)Q_{1T}(s) = \sum_{k=1}^{b(r,T)-1} c^2\, T^{-1} \left[ r - \frac{k}{T} \right]^{d-1} \left[ s - \frac{k}{T} \right]^{d-1}.$$

Easy arguments yield lim E $Q_{1T}(r)Q_{1T}(s)$ = E $W(r)W(s)$.

(b1): see lemma 3-A-3

(c1): a continuous version ( the arguments are exactly the same ) of the first part of the proof of lemma 3-A-3 yields

$$E[\, W(s)-W(r)\, ]^2 \leqslant D\, (s-r)^{\varepsilon}.$$

Therefore Q $(.) \Longrightarrow W(.)$.

(ii) Proof of $Q_{2T}(.) = op(1)$.

Summation by parts yield

$$Q_{2T}(r) = \sigma\, T^{1/2-d} \sum_{k=1}^{b-1} \pi_{b-k}^{(d-1)} \left[ \frac{S_k}{\sigma} - V(k) \right],$$

whence

$$\sup_{0 \leqslant r \leqslant 1} |\, Q_{2T}(r)\, | \leqslant$$

$$\leqslant \sup_{0 \leqslant r \leqslant 1} \sum_{k=1}^{b-1} \left| \pi_{b-k}^{(d-1)} \right| \sup_{j \leqslant T} \sigma T^{1/2-d} \left| \frac{S_j}{\sigma} - V(j) \right|,$$

$$\leqslant D \sum_{k=1}^{T-1} (\, T-k\, )^{d-2} \sup_{j \leqslant T} \sigma T^{\,1/2-d} \left| \frac{S_j}{\sigma} - V(j) \right|, \qquad (3.2)$$

where the second inequality holds by virtue of lemma 3-A-2 and D is a suitable large constant. Now by (3.1) we can write the right hand side of (3.2) as

$$o_P(1)\; T^{1/2(1+\theta)} \sum_{k=1}^{T-1} k^{\,d-2}\, T^{\,1/2-d}, \qquad (3.3)$$

Now if $d \leqslant 1$, (3.2) and (3.3) yield

$$\sup_{0 \leqslant r \leqslant 1} |\, Q_{2T}(r)\, | \leqslant T^{\alpha} \,[\log T]\, o_P(1),$$

where $\alpha = 1/[2(1+\theta)] - d + 1/2$. Which in turn yields

$$\sup_{0 \leqslant r \leqslant 1} |\, Q_{2T}(r)\, | = o_P(1),$$

since $\alpha<0$. On the other hand if $d>1$, (3.2) and (3.3) yield

$$\sup_{0\leqslant r\leqslant 1} | Q_{2T}(r) | \leqslant T^{\beta} o_P(1),$$

where $\beta = 1/[2(1+\theta)] + d - 1 - [\, d-1/2 \,] < 0$. Thus in both cases (ii) holds.

(i) Proof of $Q_{3T}(.) = o_P(1)$

we can write

$$\sup_{0\leqslant r\leqslant 1} | Q_{3T}(r) | \leqslant$$

$$\leqslant \sigma\, T^{1/2-d} \sup_{0\leqslant r\leqslant 1} \sum_{k=1}^{b-1} \left[ \pi_{b-k}^{(d)} - \frac{A(1)}{\Gamma(d)} (\, b-k \,)^{d-1} \right] \max_{j\leqslant T} |Z(j)|$$

$$+ \sigma\, T^{1/2-d} \sup_{0\leqslant r\leqslant 1} \sum_{k=1}^{b-1} \frac{A(1)}{\Gamma(d)} | (\, b-k \,)^{d-1} - (\, Tr-k \,)^{d-1} | \max_{j\leqslant T} |Z(j)|$$

$$= (I) + (II).$$

We have

$$(I) \leqslant \sigma\, T^{1/2-d} \sum_{k=1}^{T} \left[ \pi_{k}^{(d)} - \frac{A(1)}{\Gamma(d)} k^{d-1} \right] \max_{j\leqslant T} |Z(j)|$$

$$\leqslant D\, T^{1/2-d} \sum_{k=1}^{T} k^{d-2} \max_{j\leqslant T} |Z(j)| ,$$

by virtue of lemma 3-A-2, where D is a sufficiently large constant depending on d only. Now if $d\leqslant 1$ we can bound the right hand side of (3.4) by

$$D\, T^{1/2-d} \log(T) \max_{j\leqslant T} |Z(j)| \leqslant T^{-\delta'} \max_{j\leqslant T} |Z(j)|, \tag{3.5}$$

where $\delta'>0$. On the other hand if $d>1$ we bound (3.4) by

$$D\, T^{1/2-d} \max_{j\leqslant T} |Z(j)|\, T^{d-1} = D\, T^{-1/2} \max_{j\leqslant T} |Z(j)|. \tag{3.6}$$

As far as (II) is concerned we note that ( Mean Value Theorem )

$$| (\, Tr-j \,)^{d-1} - (\, b-j \,)^{d-1} | = | Tr-b | \, | d-1 | \, x^{d-2},$$

where $b-j\leqslant x\leqslant Tr-j\leqslant b+1-j$. Thus

$$| ( Tr-j )^{d-1} - ( b-j )^{d-1} | \leqslant | d-1 | \max( (b-j)^{d-2}, (b-j+1)^{d-2} ).$$

With this bound and performing algebraic manipulations similar to thos e used to deal with (I) we conclude that

$$(II) \leqslant T^{-\delta''} \max_{k\leqslant T} |Z(k)|, \qquad (3.7)$$

where $\delta''>0$. Thus

$$\sup_{0\leqslant r\leqslant 1} |Q_{3T}(r)| \leqslant T^{-\delta} \max_{k\leqslant T} |Z(k)|,$$

for some $\delta>0$. Now for any $x>0$

$$P\{ T^{-\delta} \max_{k\leqslant T} |Z(k)| > x \} \leqslant T\ P\{ |Z(1)| > xT^{\delta} \}$$

$$\leqslant T \exp( -T^{\delta}x/(2\sigma^2) ) = o(1).$$

(iv) Proof of $Q_{4T}(.) = o_p(1)$

We can write

$$P\{ \sup_{0\leqslant r\leqslant 1} |Q_{4T}(r)| > x \} \leqslant P\{ \max_{k\leqslant T} |\varepsilon_k| > xT^{d-1/2} \}$$

$$\leqslant T\ P\{ |\varepsilon_1| \geqslant xT^{d-1/2} \}$$

$$= o( T.T^{2(1+\theta)(1/2-d)} ) = o(1),$$

since $2(1+\theta)(1/2-d)+1 < 0$. □

## 3.5 Some generalizations of Theorem 3.1

Our goal in this section is to provide generalizations of theorem 3.1 in two directions. First we show that smoothed versions of the processes considered in last section weakly converges to corresponding smoothed versions of W(.). In principle, one could say, this is just a trivial application of the continuous mapping theorem (CMT). However our result does not follow from the CMT because we will relax the moment assumptions on the innovation sequence.

The second kind of generalizations is concerned with relaxing the stationarity of the innovation process. We allow the sequence of variances of the innovations to increase polynomially. Throughout we keep the notations and definitions of section 3.3. We focus only on the process defined by

$$R_T^*(r) = T^{-d-1/2} \sum_{k=1}^{b} X_k \quad , \quad 0 \leqslant r \leqslant 1,$$

where $b = b(r,T) = [\![ Tr ]\!]$.

Our first result assumes that the innovation sequence is a white noise .

Theorem 3.2 Assume that $\{ \varepsilon_t \}$ is a sequence of iidrv's such that $E(\varepsilon_1)=0$ and $E(\varepsilon_1)^2=\sigma^2 < \infty$. Then without any further moment assumptions we have

$$R_T^* \Longrightarrow Y(.) \ ,$$

where

$$Y(r) = c^* \int_0^r ( r-s )^d \, dB(s), \quad \text{where}$$

$$c^* = \sigma A(1)/\Gamma(1+d). \quad \square$$

Let us just make a couple of comments regarding this result. First, $R_T^*(.)$ is a kind of smoothed version of $X_T^*(.)$. Such smoothing yields a more robust functional limit theorem in the sense that it holds true ( iid case ) under the minimal assumption that $E(\varepsilon_1)^2 < \infty$. Therefore when using either $R_T^*(.)$ or $X_T^*(.)$ to (for instance) construct a test with asymptotic justification the former will, in principle, be preferable to the latter.

Secondly, for the iid case, it may be possible to show that $X_T^* \Longrightarrow W(.)$ assuming only $E(\varepsilon_1)^2 < \infty$. However a different sort of technique should be emploied. For the proof of theorem 3.2 we will need the following invariance principle which is due to Major(1979).

Proposition 3.1 ( Major (1979) ) Let $\{\varepsilon_n\}$, $n\geqslant 1$ be iidrv's such that $E\varepsilon_1=0$ and $Var(\varepsilon_1)=\sigma^2<\infty$. Then enlarging the original p-space if necessary we can construct a sequence $\{Z_n^*\}$ of independent random variables such that $Z_n^*\cong N(0,\sigma_n^2)$ and

(i) $|\ \varepsilon_1 + \dots + \varepsilon_n - ( Z_1^* + \dots + Z_n^* )\ | = o(n^{1/2})$, a.s.

(ii) $\lim \sigma_n = \sigma$. □

We can now prove our result.

Proof of theorem 3.2 Let $\{Z_k^*\}$ be as above. Notice that we can write for $0\leqslant r\leqslant 1$

$$T^{d+1/2}\, R_T^*(.) = ( 1-L )^{-1} X_b = ( 1-L )^{-(1+d)} A(L)\varepsilon_b.$$

Decompose $R_T^*$ as follows

$$R_T^*(r) = M_{1T}(r) + M_{2T}(r) + M_{3T}(r) + M_{4T}(r) + M_{5T}(r),$$

where

$$M_{1T}(r) = c^* \sum_{k=1}^{b-1} (r - k/T)^d \ (U(k) - U(k-1))/T^{1/2},$$

$$M_{2T}(r) = \sigma\, T^{-1/2-d} \sum_{k=1}^{b-1} \pi_{b-k}^{(1+d)} \left[ \frac{S_k - S_{k-1}}{\sigma} - (U^*(k) - U^*(k-1)) \right]$$

$$M_{3T}(r) = \sigma T^{-1/2-d} \sum_{k=1}^{b-1} \left[ \pi_{b-k}^{(1+d)} - \frac{A(1)}{\Gamma(1+d)} (Tr - k)^d \right] (U(k) - U(k-1)$$

$$M_{4T}(r) = T^{-1/2-d}\, (S_b - S_{b-1})$$

$$M_{5T}(r) = \sigma\, T^{-1/2-d} \sum_{k=1}^{b-1} \pi_{b-k}^{(1+d)} \left[\sigma^{-1} Z_k^* - \sigma_k^{-1} Z_k^*\right] ,$$

where $b = b(r,T) = [\![Tr]\!]$, $\sigma U^*(k) = \sum_{j=1}^{k} Z_j^*$ , and $U(k) = \sum_{j=1}^{k} Z_j^* \sigma_j^{-1}$.

Our result will follow if

(i) $M_{1T}(.) \Longrightarrow Y(.)$

(ii) $M_{2T}(.) = o_P(1)$

(iii) $M_{3T}(.) = o_P(1)$

(iv) $M_{4T}(.) = o_P(1)$

(v) $M_{5T}(.) = o_P(1)$

The proofs of (i), (iii) and (iv) are exactly the same as those of (i), (iii) and (iv) in theorem 3.1, respectively. The proof of (ii) follows the same pattern of its counterpart in theorem 3.1. We can write

$$\sup_{0\leqslant r\leqslant 1} | M_{2T}(r) | \leqslant$$

$$\leqslant \sup_{0\leqslant r\leqslant 1} \sum_{k=1}^{b-1} \left| \pi_{b-k}^{(d)} \right| \sup_{j\leqslant T} \sigma T^{-1/2-d} \left[ \frac{S_j}{\sigma} - U^*(j) \right]$$

$$\leqslant D \sum_{k=1}^{T-1} ( T-k )^{d-1} \sup_{j\leqslant T} \sigma T^{-1/2-d} \left[ \frac{S_j}{\sigma} - U^*(j) \right]$$

$$\leqslant D' \, T^d \, T^{-d-1/2} \sup_{j\leqslant T} \left[ \frac{S_j}{\sigma} - U^*(j) \right] = o_P(1),$$

where the first inequality comes from summation by parts, the second comes from lemma 3-A-2, the third from the fact that 1-d<1 and the equality comes from Major's theorem. Now following the same arguments as above we obtain

$$\sup_{0\leqslant r\leqslant 1} |M_{5T}(r)| \leqslant T^{-1/2} \sup_{j\leqslant T} \left[ \sum_{k=1}^{j} (\sigma_k - \sigma) \frac{Z_k^*}{\sigma_k} \right] = o_P(1),$$

by a simple application of Kolmogorov's inequality and the fact that $\sigma_k=\sigma+o(1)$. (Note: the Z's are normally distributed). □

Actually minor modifications in the proof of theorem 3.2 yield

<u>Theorem 3.3</u> Let $\{\varepsilon_k\}$ be a sequence of identically distributed random

variables. Suppose that there exists a sequence of iidrv's $\{Z_k\}$ such that $Z_1 \cong N(0,\sigma^2)$ and

$$\max_{k \leq T} \left| \sum_{j=1}^{k} \varepsilon_j - Z_j \right| = o_P( T^{1/2}),$$

then $R_T^*(.) \Longrightarrow Y(.)$, where

$$Y(r) = [\sigma A(1)/\Gamma(1+d)] \int_0^r (r - s)^d \, dB(s). \ \square$$

Theorem 3.3 allows us considering $\{\varepsilon_k\}$ strong mixing with a rather slow rate of decreasing for the mixing weights.

Corollary 3.1 Suppose $\{\varepsilon_k\}$ is a strictly stationary strong mixing sequence of random variables with $E\varepsilon_1=0$, $Var(\varepsilon_1)<\infty$ and $Var(\varepsilon_1+\ldots+\varepsilon_T)\to\infty$, as $T\to\infty$. Let $\{\alpha(k)\}$ be the corresponding sequence of mixing weights. Suppose $\delta>0$ and $\lambda>1+3/\delta$ are real numbers, $\alpha(k)=o( (\log(k))^{-\lambda} )$ and

$$\sup_T [ E(\varepsilon_1 +\ldots+ \varepsilon_T)^{2+\delta} ]/[ Var(\varepsilon_1 +\ldots+ \varepsilon_T) ]^{(2+\delta)/2} < \infty.$$

Then there exists $\sigma^2$, $0<\sigma^2<\infty$ such that

$$\lim_T T^{-1} Var(\varepsilon_1 +\ldots+ \varepsilon_T) = \sigma^2, \text{ and}$$

$$R_T^*(.) \Longrightarrow Y(.).$$

Proof: Theorem 4 in Bradley(1983) implies the existence of $\{Z_k\}$ as in theorem 3.3 above. $\square$

Now in order to deal with trending heteroskedasticity in the sequence of innovations we will need the following result due to Einmahl.

Proposition ( See Einmahl(1987) ) Let $\{X_n\}$ be a sequence of independent random variables with zero means and $Var(X_n)=(\sigma_n)^2$. Assume that for some $0<\delta<2$ the following holds true

$\Sigma_n \, (a_n)^{-(2+\delta)} \, E|X_n|^{2+\delta} < \infty,$

where $\lim a_n = \infty$. Then there exists a sequence of independent random variables $\{Z_n\}$ with $Z_n \cong N[0,(\sigma_n)^2]$ such that

$$\sum_{k=1}^{n} X_k - Z_k = o(a_n) \; , \quad \text{a.s.} \quad \square$$

We can now state our last result in this chapter.

Theorem 3.4 Let $\{\epsilon_k\}$ be a sequence of independent random variables such that there exist $0<\delta<2$, $\theta<\delta/(2+\delta)$ and

$\Sigma_n \, n^{-(2+\delta)/2} \, E|\epsilon_n|^{2+\delta} < \infty,$

$E|\epsilon_n|^2 = n^\theta + O(n^\beta),$

where $\beta<\theta$. Then the process defined by

$$R_T^{**}(r) = T^{-(d+\theta+1/2)} \sum_{k=1}^{[\![Tr]\!]} X_k \; , \quad r \in [0,1],$$

weakly converges to

$$Y^*(r) = c^* \int_0^r (r-s)^d s^\theta \, dB(s) \; , \quad r \in [0,1],$$

where $c^* = A(1)/\Gamma(1+d)$.

Proof: Use Einmahl's result with $a_n = n^{1/2}$. The rest of the proof is pratically identical to that of theorem 3.2. $\square$

## APPENDIX 3-A

Lemma 3-A-1 Let d be a real number such that $d>-1/2$. Then

$$\left| \frac{\Gamma(n+d)}{\Gamma(n+1)\Gamma(d)} - \frac{n^{d-1}}{\Gamma(d)} \right| = O(n^{d-2}).$$

Proof: From Abramovitz and Stegun(1970), formula 6.1.47, we obtain

$$x^{b-a}\frac{\Gamma(x+a)}{\Gamma(x+b)} = 1 + \frac{(a-b)(a+b-1)}{2x} + o\left[\frac{1}{x}\right],$$

as $x\to\infty$. Whence

$$\frac{\Gamma(n+d)}{\Gamma(n+1)} = n^{d-1}\left[1 + \frac{(d-1)d}{2n} + o\left[\frac{1}{n}\right]\right]. \ \square$$

Lemma 3-A-2 For any fixed $d>-1/2$ we have

$$\left| \pi_k^{(d)} - \frac{A(1)}{\Gamma(d)} k^{d-1} \right| = O(k^{d-2}).$$

Proof: We have

$$\left| \pi_k^{(d)} - \frac{A(1)}{\Gamma(d)} k^{d-1} \right| =$$

$$\left| \sum_{j=0}^{k} \alpha_{k-j} \frac{\Gamma(j+d)}{\Gamma(j+1)} - k^{d-1} \sum_{j=0}^{\infty} \alpha_j \right| \frac{1}{\Gamma(d)}$$

$$\leqslant \Gamma(d)^{-1} [ (I) + (II) + (III) + (IV) ], \qquad (2.1)$$

where

$$(I) = |\alpha_k| \Gamma(d)$$

$$(II) = \sum_{j=1}^{k} |\alpha_{k-j}| \left| \frac{\Gamma(j+d)}{\Gamma(j+1)} - j^{d-1} \right|$$

$$(III) = \sum_{j=1}^{k} |\alpha_{k-j}| \, |j^{d-1} - k^{d-1}|$$

$$(IV) = \left| \sum_{j=k}^{\infty} \alpha_j \right| k^{d-1} [\Gamma(d)]^{-1}.$$

We have

$$(II) = O\Big( \sum_{j=1}^{k} |\alpha_{k-j}| j^{d-2} \Big)$$

by lemma 3-A-1. We must consider two cases. Namely $d \geqslant 2$ and $d<2$. If $d>2$, summability of $\{|\alpha_n|\}$ yields $(II)=O(k^{d-2})$. If on the other hand $d<2$ we have

$$(II) \ll \sum_{j=1}^{m} |\alpha_{k-j}| \; j^{d-2} + \sum_{j=m+1}^{k} |\alpha_{k-j}| \; j^{d-2}$$

$$= (V) + (VI),$$

where we have denoted $m = [k/2]$. Now

$$(V) \leqslant \sum_{j \geqslant m} |\alpha_j| \ll m^{-3} \ll k^{d-2},$$

because $|\alpha_n| \ll n^{-4}$. Also

$$(VI) \leqslant m^{d-2} \sum_{j=m+1}^{k} |\alpha_{k-j}| \ll k^{d-2},$$

again by summability of $\{|\alpha_n|\}$. Thus

$$(II) = O(k^{d-2}). \tag{2.2}$$

Now the mean value theorem yields ( excude the trivial case $d=1$ )

$$(III) \ll \sum_{j=1}^{k-1} |\alpha_{k-j}| \; (k - j)\beta_j^{d-2},$$

where $j \leqslant \beta \leqslant k$. If $d \geqslant 2$ summability of $\{n|\alpha_n|\}$ implies $(III)=O(k^{d-2})$.

If $-1/2<d<0$ we have

$$(III) \ll \sum_{j=1}^{m} |\alpha_{k-j}| \; |k^{1-d} - j^{1-d}| \; k^{d-1} j^{d-1} +$$

$$+ \sum_{j=m+1}^{m} |\alpha_{k-j}| \; |k - j| \; \beta_j^{d-2}$$

$$= (VII) + (VIII),$$

where $j \leqslant \beta_j \leqslant k$. Now,

$$(VII) \leqslant k^{d-1} \sum_{j=1}^{m} |\alpha_{k-j}| \; |k^{1-d} - j^{1-d}|$$

$$\leqslant k^{d-1} \sum_{j=1}^{m} |\alpha_{k-j}| \; (k - j)\theta_j^{-d}, \qquad (2.3)$$

where $j \leqslant \theta \leqslant k$. The RHS of (2.3) can be bounded from above by

$$k^{d-1}k^{-d} \sum_{j=1}^{m} |\alpha_{k-j}| \; (k - j)$$

$$\leqslant k^{-1} \sum_{j \geqslant m} |\alpha_j| \; j \leqslant k^{-3} \leqslant k^{d-2} \; , \qquad (2.4)$$

since $|\alpha_n| \leqslant n^{-4}$. Also

$$(VIII) \leqslant m^{d-2} \sum_j j|\alpha_j| \leqslant k^{d-2},$$

by summability of $\{j|\alpha_j|\}$. If $0 \leqslant d < 2$ we can treat (III) in a similar manner to obtain

$$(III) = \sum_{j=1}^{m} |\alpha_{k-j}| \; (k - j) \; \beta_j^{d-2} + \sum_{j=m+1}^{k-1} |\alpha_{k-j}| \; (k - j) \; \beta_j^{d-2}$$

$$\leqslant m^{-2} + m^{d-2} \leqslant k^{d-2} \; .$$

Therefore we conclude that

$$(III) = O( k^{d-2} ).$$

Finally

$(I) \leqslant k^{d-2}$ and $(IV) \leqslant k^{d-2}$ are obvious. □

<u>Lemma 3-A-3</u> Let us define for each $T \geqslant 1$

$$W(r) = \sum_{k=1}^{[\![Tr]\!]} \left[ r - \frac{k}{T} \right]^{d-1} |B(\tfrac{k}{T}) - B(\tfrac{k-1}{T})| \; , \quad r \in [0,1]$$

where $B(.)$ is a standard Brownian motion. Then there exists $c>0$, $\varepsilon>0$ such that

$$E[W_T(r) - W_T(q)]^2 \; E[W_T(s) - W_T(r)]^2 \leqslant c(s - q)^{\varepsilon} \; ,$$

for all $T \geq 1$, and $0 \leq q < r < s \leq 1$, provided $d > 1/2$.

Proof: Take q, r, s as above. We can write

$$E[W_T(r) - W_T(q)]^2 =$$

$$T^{-1} \sum_{k=1}^{[\![Tq]\!]} \left[ \left(r - \frac{k}{T}\right)^{d-1} - \left(q - \frac{k}{T}\right)^{d-1} \right]^2 +$$

$$T^{-1} \sum_{k=1+[\![Tq]\!]}^{[\![Tr]\!]} \left(r - \frac{k}{T}\right)^{2d-2} = D_1 + D_2 .$$

where we have adopted the convention $\sum_{k=m}^{m} a_k = 0$ if $m > n$. We first bound $D_1 = D_1(q,r)$. There are three cases to be considered. Suppose first that $d \geq 2$. We have for $0 \leq x \leq q$.

$$| (r - x)^{d-1} - (q - x)^{d-1} |$$

$$\leq \sup_{0 \leq y \leq 1} (d - 1)(r - q) y^{d-2}$$

$$\leq (d - 1)(r - q),$$

by the mean value theorem. Hence

$$D_1 \leq (d - 1)^2 (r - q)^2 . \tag{3.1}$$

Now suppose $1 < d < 2$. Minkovski inequality ( note: $0 \leq d-1 < 1$ ) yields

$$(r - x)^{d-1} - (q - x)^{d-1} \leq (r - q)^{d-1} ,$$

and therefore

$$D_1 \leq (r - q)^{2d-2} . \tag{3.2}$$

Finally assume $1/2 < d < 1$ ( note: $D_1 = 0$ if $d = 1$ ). It is easy to see that $f(x) = (q - x)^{d-1} - (r - x)^{d-1}$ is non-decreasing. Thus

$$D \leq \int_0^q [f(x)]^2 \, dx$$

$$= (r - q)^{2d-2} \int_0^q \left[ \left[1 + \frac{r-x}{r-q}\right]^{d-1} - \left[\frac{r-x}{r-q}\right]^{d-1} \right]^2 dx$$

$$= (r - q)^{2d-1} \int_0^{q/(r-q)} [ (1 + v)^{d-1} - v^{d-1} ]^2 \, dx$$

$$< (r - q)^{2d-1} \int_0^{\infty} [ (1 + v)^{d-1} - v^{d-1} ]^2 \, dx . \qquad (3.3)$$

Now, as $d<3/2$ the integral above is finite and then

$$D_1 \leqslant (r - q)^{2d-1} c_1 . \qquad (3.4)$$

(3.1), (3.2) and (3.4) imply that

$$D_1 \leqslant a (r - q)^{\alpha} , \text{ all } q<r, \text{ all } T, \qquad (3.5)$$

where $a>0$, $\alpha>0$ depend on d only. Now we bound $D_2 = D_2(q,r)$. If $[\![Tr]\!]=[\![Tq]\!]$ then $D_2 = 0$. Otherwise we have two cases to consider. First, if $d\geqslant 1$ we have

$$D_2 \leqslant T^{-1} \{[\![Tr]\!] - [\![Tq]\!]\}$$

$$\leqslant T^{-1} \{ Tr - Tq + 1 \} = r - q + T^{-1} \qquad (3.6)$$

If $1/2<d<1$ we have

$$D_2 \leqslant \int_{1+[\![Tq]\!]/T}^{[\![Tr]\!]/T} (r - x)^{2d-2} dx +$$

$$+ T^{-1} [r - [\![Tr]\!]/T]^{2d-2}$$

$$\leqslant \int_q^r (r - x)^{2d-2} dx + T^{2d-3}$$

$$= (2d - 1)^{-1} (r - q)^{2d-1} + T^{2d-3} . \qquad (3.7)$$

(3.6) and (3.7) imply that

$$D_2 \leqslant b (r - q)^{\beta} + T^{-\delta} , \qquad (3.8)$$

where $b>0$, $\beta>0$, $\delta>0$ depend only on d.

Now if $s-q < 1/T$ then either $D_2(q,r) = 0$ or $D_2(r,s) = 0$. Assume w.l.o.g. $D_2(q,r) = 0$. We have

$$E[W_T(r) - W_T(q)]^2 \; E[W_T(r) - W_T(s)]^2 =$$

$$= D_1(q,r)\,[D_1(r,s) + D_2(r,s)]$$

$$\leqslant a^2\,(r - q)^{\alpha} + ab\,(r - q)^{\alpha} + a\,(r - q)^{\alpha} \,. \tag{3.9}$$

If on the other hand $s\text{-}q \geqslant 1/t$ then

$$E[W_T(r) - W_T(q)]^2 \; E[W_T(r) - W_T(s)]^2 \leqslant$$

$$\leqslant [\;(a + b)^2 + 2a + 2b\;] \max\{\,(r - q)^{\gamma},\,(s - r)^{\gamma}\,\} + T^{-2\delta}$$

$$\leqslant [\;(a + b)^2 + 2a + 2b\;] \max\{\,(r - q)^{\gamma},\,(s - r)^{\gamma}\,\} + (s - q)^{-2\delta}$$

where $\gamma = \min\{\,\alpha,\,\beta\,\}$. (3.9) and (3.10) imply our result. □

# CHAPTER 4

## 4.1 Introduction

In this chapter we are going to deal with the application of approximation theorems to nonparametric statistics. We will be interested in obtaining asymptotic results concerning density and regression function nonparametric estimates. In particular we will show that under mild weak dependence assumptions we can obtain consistency results which parallel those under iid samples.

Some of our conclusions appear to be unexpected from an intuitive point of view. To be slightly more precise, one could expect that the weak dependence which appears in our samples should propagate into our asymptotics. Such a propagation could occur, for instance, via a strengthening of our basic assumptions so as to ensure that a consistency result which holds true under iid samples is still true under weak dependent samples. In some cases to be considered below such a strengthening is seen not to be necessary

As a matter of fact, there exists numerous asymptotic results in the literature concerning nonparametrics under weak dependence which hold true under the same assumptions imposed on their iid counterparts. From a finite sample (or perhaps, practical) point of view, such robustness should not, as remarked by Robinson (1983) and other authors, be taken too seriously.

## 4.2 $L_1$-consistency for Density Estimates

In this section we deal with $L_1$-consistency for two kinds of density estimates under weak dependent samples. The relationship between $L_1$ distance of density estimates to the true density is considered in section 4.2.1. Kernel and Histogram density estimates are discussed in sections 4.2.2 and 4.2.3. We present our main results in section 4.2.4. Some straightforward consequences of our main theorems are treated in section 4.2.5.

### 4.2.1 An Informal Framework for $L_1$-consistency

Let $\{X_n\}$, $n \geq 1$ be a sequence of identically distributed random vectors taking values in $\mathbb{R}^d$, and suppose that $X_1$ admits of a density f (say). Let us consider a density estimate $f_n$ (say) based on $\{X_k\}$, $1 \leq k \leq n$. Define the $L_1$ error between $f_n$ and the true underlying density by

$$J_n = \int |f_n(x) - f(x)| \, dx.$$

We will be interested in studying consistency properties (strong and weak) of $\{f_n\}$ based on $\{J_n\}$, under weak dependence assumptions on $\{X_n\}$, $n \geq 1$. We say that a sequence of density estimates $\{f_n\}$ of f is strongly (weakly, respectively) consistent in $L_1$ if $J_n$ converges to zero almost surely (in probability, respectively).

Implicit in the above definition is the fact that each $f_n$ is integrable. This implies that in any study concerning $L_1$-consistency of density estimates one must, to begin with, rule out, for instance, the class of Nearest Neighbor density estimates (see Loftsgaarden and Quesenberry(1965), Mack and Rosenblatt (1979) and Moore and Yackel(1977a,b)). Going into details on Nearest Neighbour estimates is beyond the scope of this work. Let us just mention that such a class does not provide estimates which are integrable because their tails behave like $\|x\|^{-1}$.

$L_1$-consistency for density estimates has been studied by a number of authors. Devroye (1979, 1983), Abou-Jaoude (1976a,b,c), Gyorfy (1987) and Tran (1989) are some examples. In most of these works a common feature is the fundamental role played by the relation between $J_n$ and multinomial distributions, on obtaining asymptotics for $J_n$. Such a relation can be seen (very informally) as follows. Write $J_n$ as

$$\int_{D^c} |f_n - f| + \int_D |f_n - f| ,$$

where D is a large compact set. Now partition D into $\{E_1,\ldots,E_N\}$. We have

$$\int_D |f_n - f| = \sum_{k=1}^{N} \int_{E_k} |f_n - f| .$$

Let $\mu_n$ be the empirical measure based on $X_1,\ldots,X_n$ and $\mu$ be the underlying probability distribution and $\lambda(.)$ be the Lebesgue measure. Approximate $f_n(x)$ by $\mu_n(E_k)/\lambda(E_k)$ and $f(x)$ by $\mu(E_k)/\lambda(E_k)$ when x is in $E_k$. We have

$$\sum_{k=1}^{N} \int_{E_k} |f_n - f| \approx \sum_{k=1}^{N} \int_{E_k} |\mu_n(E_k) - \mu(E_k)|/\lambda(E_k)$$
$$= \sum_{k=1}^{N} |\mu_n(E_k) - \mu(E_k)| .$$

Now $(n\mu_n(E_1),\ldots,n\mu_n(E_N),\ n\mu_n(D^c))$ is, under the assumption of an independent sample, multinomially distributed with mean $(n\mu(E_1),\ldots,n\mu(E_N),\ n\mu(D^c))$. Also, the other term in our decomposition can be bounded from above by

$$\int_{D^c} |f_n - f| < \int_{D^c} |f_n| + \int_{D^c} |f| \approx \mu_n(D^c) + \mu(D^c)$$
$$< |\mu_n(D^c) - \mu(D^c)| + 2\mu(D^c) .$$

Now $\mu(D^c)$ can be made small by our choice of D and therefore $J_n$ can be written approximately as

$$n^{-1} \cdot \sum_{k=1}^{N+1} |Z_k - EZ_k| \ ,$$

where the random vector $(z_1,\ldots,z_{N+1})$ is multinomially distributed with mean $(n\mu(E_1),\ldots,n\mu(E_N),\ n\mu(D^c))$. Consistency results will then follow if

(a) We can choose N and the D-partition appropriately;

(b) We can obtain good bounds for

$$P\{ \sum_{k=1}^{N+1} |Z_k - EZ_k| > n\epsilon \} \ ;$$

(c) We can provide rigorous justification for our informal approximation.

Nonetheless, when dealing with weakly dependent samples, we must handle a further problem. Namely, the vector $(n\mu_n(E_1),\ldots,n\mu_n(E_N),\ n\mu_n(D^c))$ is no longer multinomially distributed. Our approach to solve such a problem will be to decompose such a vector into a sum of "sub-vectors" in such a way that each of these subvectors can be well approximated by suitable multinominal random vectors.

4.2.2 Kernel Density Estimates

Let $\{X_n\}$ be a sequence of identically distributed $\mathbb{R}^d$-valued random vectors admitting of a common density f. Let $W:\mathbb{R}^d \to \mathbb{R}$ be an integrable function which integrates to 1. A kernel density estimate of f is defined as

$$f_n(x) = (nb^d)^{-1} \sum_{k=1}^{n} W(\ b^{-1}\ (x - X_k)\ ).$$

where $b = b(n)$ is a sequence of positive real numbers ( the sequence of bandwidths or smoothing parameters ).

Kernel density estimates (KDE) were introduced in Rosenblatt(1956) ( See also

Parzen(1962) and Cacoullos(1966) ) and is probably the best known of all density estimates. When W(.) is also a density we can easily see that $f_n$ is the density of the convolution of the empirical distribution function with a distribution which admits $b^{-d}W(x/b)$ as its density. For this reason KDE's are sometimes called empirical density functions ( See Csorgo and Revesz(1981) section 6.2 ).

We will make no attempt to perform a detailed analysis of previous works on KDE's. We refer the reader to the survey papers Bean and Tsokos(1980), Fryer(1977) and Wertz and Schneider(1979) for an account on KDE's under independent samples. See also chapter 3 in Prakasa Rao(1983) for some $L_p$ properties of KDE's. Practical aspects of such estimates are discussed in Devroye(1987), Silverman(1986) and Tapia and Johnson(1978).

For dependent samples, asymptotic results on KDE's are much more scattered and, there is hardly any single volume providing a comprehensive study of the subject. See however the recent book by Gyorfi et al (1989). Pointwise central limit theorems were established in Robinson(1983) for strong mixing samples. Ahmad(1982) studied the integrated mean square error for KDE's under weak dependence. For an account on strong uniform consistency of KDE's under various forms of weak dependent samples see Roussas(1988). See Robinson(1987) and Cheng and Robinson(1990) for central ( and non-central ) limit theorems for KDE's under strongly correlated samples.

In all asymptotic results for KDE's the condition $b + (nb^d)^{-1} = o(1)$ turns out to be minimal ( for instance such a condition is necessary for pointwise consistency for continuous f ). Such a condition imposes upper and lower limits for the rate at which the sequence of smoothing parameters goes to zero. Now in pointwise terms, the smoothing factor b can be thought of as a measure of the relevance we are attributing to observations far from the particular point we are interested in. Thus, as ( very loosely ) a sample of n iid random vectors contains "more information" on the common distribution than a sample of dependent ones ( same size and same underlying density ) we could expect a slower lower limit rate for the bandwidths in dependent samples. However, as we show in section 4.3.2, this is not necessarily the case. As a matter of fact if the dependence structure is weak enough the minimal condition above turns out to be ( as in the iid case ) necessary and sufficient for strong $L_1$ consistency of KDE's ( see chapter 5 ).

4.2.3 Histogram Density Estimates

Let $\{X_n\}$ be a sequence of identically distributed $R^d$–valued random vectors having f as their common density. Let $\{Q_n\}$ be a sequence of $R^d$–partitions, $Q_n=\{ A_{nj}, j\geq 1 \}$. Suppose that $\{Q_n\}$ is rich enough so that

$$\bigcap_{n=1}^{\infty} \sigma\Big( \bigcup_{m=n}^{\infty} Q_m \Big) = \{ \text{Borelians of } R^d \}$$

In that case we define the Histogram Density Estimate of f, based on the sample $\{X_k , 1\leq k\leq n\}$ and subordinated to the sequence of partitions $\{Q_n\}$ by

$$f_n(x) = n^{-1} \sum_{k=1}^{n} I\{X_k \in A_{nj}\} [ \lambda(A_{nj}) ]^{-1} , \text{ for } x \in A_{nj} ,$$

where $\lambda(.)$ is the Lebesgue measure on $R^d$. The above definition is a generalization of the classical histogram estimate where each $A_{nj}$ is a d–fold product of finite intervals. Namely each $A_{nj}$ can be written as

$$\prod_{k=1}^{d} (ba_k i , ba_k(i + 1)]$$

where $a_1$ ,..., $a_d$ are fixed positive real numbers, $b = b(n)$ is a smoothing parameter and $i \in Z$. HDE's as defined above were considered in Abou–Jaoude

(1976a,b).

We refer the reader to the same monographs mentioned in last section for practical and theoretical aspects in the classical case under iid samples. HDE's for weak dependent samples were studied in Gyorfi(1987) and Tran(1990). Studies of HDE's on samples exhibiting strong forms of dependence do not appear in the literature.

### 4.2.4 Main Results

In this section we will present sufficient conditions for weak and strong $L_1$-consistency of KDE's and HDE's under weak dependence assumptions on the sample. As outlined in section 4.2.1 we will handle those issues by approximating the random vector $(n\mu_n(E_1),\ldots,n\mu_n(E_N))$, where $\{E_k,\ 1\leq K\leq N\}$ is a partition of $\mathbb{R}^d$, by a sum of multinomially distributed random vectors. Our main tools for such an approximation will be Berbee's theorem (see Chapter 2) and the following result due to Bradley.

THEOREM 4.2.1 (Bradley, 1983). Suppose X and Y are r.v.'s taking their values in Borel spaces $S_1$ and $S_2$ respectively (a Borel space is a measurable space (S,D) which is bimeasurably isomorphic to a Borel subset of the real line $\mathbb{R}$; for instance, $\mathbb{R}^t$ is a Borel space); and suppose V is uniform $-[0,1]$ r.v. independent of (X,Y). Suppose N is a positive integer and $\mathcal{H} = \{H_1,\ldots,H_N\}$ is a measurable partition of $S_2$. Then there exists a $S_2$-valued r.v. $Y^* = f(X,Y,V)$ where f is a measurable function from $S_1\times S_2\times[0,1]$ into $S_2$ such that

(i) $Y^*$ is independent of X,

(ii) the probability distributions of $Y^*$ and Y on $S_2$ are identical, and

(iii) $P\{Y^*$ and Y are not elements of the same $H \in \mathcal{H}\} \leq (8N)^{\frac{1}{2}}\alpha(\mathcal{B}(X),\mathcal{B}(Y))$.

where $\alpha(\mathcal{B}(X),\mathcal{B}(Y)) = \sup\{|P(A\cap B) - P(A)P(B)|,\ A\in\mathcal{B}(X),\ B\in\mathcal{B}(Y)\}$, and $\mathcal{B}(X) = \{X^{-1}(D),\ D\in\mathcal{D}\}$, with $\mathcal{D}$ being the $\sigma$-algebra that accompanies $S_1$ ($\mathcal{B}(Y)$ is defined analogously). □

Now we can state and prove our approximation result.

<u>THEOREM 4.2.2</u> Let $\{X_n\}$, $n\geq 1$ be a strong mixing (weak Bernoulli, respectively) stationary process such that $X_1$ takes its values in $\mathbb{R}^d$. Let $\mathcal{H} = \{H_j,\ 1\leq j\leq N\}$ be a measurable partition of $\mathbb{R}^d$. Let m be a positive integer and $\epsilon$ be a positive real number satisfying

$$\frac{n}{m} > \frac{4}{\epsilon}\ . \tag{4.2.1}$$

Let $\mu_n(.)$ be the empirical measure based on $X_1,\ldots,X_n$, and $\mu(.)$ be the probability distribution of $X_1$. Then

$$P\{\sum_{k=1}^{N} |n\mu_n(H_k) - n\mu(H_k)| > n\epsilon\}$$

$$\leq m\theta(m) + 2^N.\exp\{-n\epsilon^2/(64m)\},$$

where $\theta(m)$ equals $\sqrt{8N}\ \alpha(m)$ if $\{X_n\}$ is strong mixing and equals $\beta(m)$ if $\{X_n\}$ is weak Bernoulli.

<u>Proof</u>. Let us define $[[n/m]] = M$. We have

$$P\{\sum_{i=1}^{N} |\mu_n(H_i) - \mu(H_i)| > \epsilon\}$$

$$= P\{\sum_{i=1}^{N} |\sum_{k=1}^{n} I(X_k \in H_i) - n\mu(H_i)| > n\epsilon\}$$

$$\leq \sum_{j=1}^{m} P\{\sum_{i=1}^{N} |\sum_{k=j,m}^{mM} I(X_k \in H_i) - M\mu(H_i)| > M\epsilon/2\}$$

$$+ P\{\sum_{i=1}^{N} |\sum_{k=mM+1}^{n} [I(X_k \in H_i) - \mu(H_i)]| > n\epsilon/2\}, \tag{4.2.2}$$

where we have defined

$$\sum_{k=j,m}^{R} a_k = \sum_{k:\ k=j+im,\, i\geq 0,\, k\leq R} a_k .$$

Now the second term in the right hand side of 4.2.2 equals zero by virtue of 4.2.1, whereas the first one can be written as

$$mP\{ \sum_{i=1}^{N} | \sum_{k=1}^{M} I(X_{km} \in H_i) - M\mu(H_i) | > M\epsilon/2\}, \qquad (4.2.3)$$

by stationarity. Now by THEOREM 4.2.1 (Berbee's theorem, respectively) there exist random vectors $Z_m, Z_{2m}, \ldots$, such that

(i) $Z_{mj}$ and $X_{mj}$ are identically distributed,

(ii) $Z_{mj}$ is independent of $\{Z_{ms},\ 1\leq s<j\}$,

(iii) $P\{Z_{mj}$ and $X_{mj}$ are not elements of the same $H_i \in H\} \leq$ $(8N)^{\frac{1}{2}}\alpha(m)$ [$\leq \beta(m)$, respectively].

Now, by (ii) and a simple induction argument we obtain that $\{Z_m, Z_{2m}, \ldots\}$ is a collection of independent random vectors. We can write

$$P\{ \sum_{i=1}^{N} | \sum_{k=1}^{M} I(X_{km} \in H_i) - M\mu(H_i) | > M\epsilon/2\}$$

$$\leq P\{ \sum_{i=1}^{N} | \sum_{k=1}^{M} I(X_{km} \in H_i) - I(Z_{km} \in H_i) | > M\epsilon/4\}$$
$$+ P\{ \sum_{i=1}^{N} | \sum_{k=1}^{M} I(Z_{km} \in H_i) - M\mu(H_i) | > M\epsilon/4\}$$
$$= (I) + (II).$$

We have,

$$(I) \leq (M\epsilon/4)^{-1}) E \sum_{i=1}^{N} | \sum_{k=1}^{M} I(X_{km} \in H_i) - I(Z_{km} \in H_i)|$$
$$\leq (M\epsilon/4)^{-1} \sum_{k=1}^{M} E \sum_{i=1}^{N} |I(X_{km} \in H_i) - I(Z_{km} \in H_i)|$$
$$\leq (4/\epsilon) \max_{k} E \sum_{i=1}^{N} |I(X_{km} \in H_i) - I(Z_{km} \in H_i)|$$

$$< (4/\epsilon) \cdot 2 \cdot \theta(m), \tag{4.2.4}$$

where the last inequality holds valid because of (iii) above. On the other hand

$$(II) < 2^N \max \{P\{ \sum_{k=1}^{M} I(Z_{km} \in G) - M\mu(G) > M\epsilon/8\}, G \in \sigma(H)\}, \tag{4.2.5}$$

where $\sigma(H)$ is the $\sigma$-algebra generated by $H$ (which possesses $2^N$ elements). The inequality above was obtained using the simple fact that if $\mu$ and $\nu$ are probability measures defined on a measurable space $(\Omega,F)$ and $H = \{H_1,\ldots,H_N\}$ is a measurable partition of $\Omega$, then

$$\sum_{i=1}^{N} |\nu(H_i) - \mu(H_i)| = \Sigma \sup \{\nu(G) - \mu(G), G \in \sigma(H)\}.$$

Now, Hoeffding's inequality (see Hoeffding (1963)) and 4.2.5 imply

$$(II) < 2^N \exp(-2\epsilon^2 M/64) < 2^N \exp(-\epsilon^2 n/(64m)). \tag{4.2.6}$$

4.2.3, 4.2.4 and 4.2.6 imply the result. □

It seems worthwhile to make a few comments on the above result. The first one is concerned with similar (though weaker) results that could be obtained by application of a different sort of technique. There is quite a number of bounds for probabilities of large deviations for centered sums of weakly dependent Bernoulli random variables (see e.g. Roussas and Ioannides (1988), Doukhan et.al. (1984), Carbou (1983), Yoshihara (1978)). Such results could be used so as to establish, in a more direct fashion, bounds for

$$P\{ \sum_{i=1}^{N} | \sum_{k=1}^{n} I(X_k \in H_i) - n\mu(H_i)| > n\epsilon\}. \tag{4.2.7}$$

This approach was adopted in Gyorfi (1987) for uniform mixing samples, obtaining rates equivalent to i.i.d. samples. However, when we deal with less restrictive mixing assumptions, the approach in the proof of our theorem 4.2.2 seems to work better. For instance, under the assumption of exponentially strong mixing samples (i.e. the mixing weights decrease exponentially), any of the above references would provide bounds for

$$P\{|\sum_{k=1}^{n} I(X_k \in H_i) - n\mu(H_i)| > n\epsilon\}$$

which are at most like $\exp(-An^{1-\delta})$, where $0 < \delta < 1$ and $A > 0$. Whence the best upper bound for 4.2.7 one can obtain (see 4.2.5) is

$$2^N \exp\{-An^{1-\delta}\} ,$$

which is clearly worse than the bound obtained in THEOREM 4.2.2. The second remark is much more a digression on a conjecture. As will be seen later on, our proofs of strong $L_1$-consistency are in fact deeper than necessary. Namely we will show strong $L_1$-consistency by proving complete $L_1$-consistency (we recall that a sequence of random vectors $\{Z_n\}$ converges completely to a random vector Z iff $\Sigma P\{|Z_n-Z| > \epsilon\} < \infty$, for all $\epsilon > 0$; Borel-Cantelli lemma implies that complete convergence is stronger than almost sure convergence). More specifically let $N = N(n)$ be such that $\lim_n n^{-1}N = 0$ and suppose that we have a sequence of $\mathbb{R}^d$-partition $\{\mathcal{H}_n\}$, where $\mathcal{H}_n$ possesses N(n) elements. Our proofs of strong $L_1$-consistency are invariably made by showing that

$$T_n = \sum_{H \in \mathcal{H}_n} |\mu_n(H) - \mu(H)| \to 0, \quad \text{completely.}$$

It seems natural to ask whether $T_n \to 0$, a.s. under ergodic samples. The answer is negative for general $\mathcal{H}_n$'s. A counter example has been built in Shields (1973). He showed, as an application of Rohlin's

theorem that there exists a sequence of Borel sets $\{A_n\}$ such that

$$P\{|\mu_n(A_n) - \mu(A_n)| \geq \tfrac{1}{2}\} > \tfrac{1}{8} ,$$

for all $n \geq 1$, provided that we only assume that the underlying process $\{X_n\}$ is stationary and ergodic.

Now we can state

THEOREM 4.2.3. Let $\{X_n\}$, $n \geq 1$ be a stationary process taking its values in $R^d$ and assume that $X_1$ admits of a density $f: R^d \to R$. Let $\{f_n\}$ be a sequence of KDE's with corresponding sequence of bandwidths given by $\{b(n)\}$ (obviously we are assuming that $f_n$ is based on $X_1,\ldots,X_n$). Also assume that the underlying kernel $W: R^d \to R$ is integrable and integrates to 1.

(A) If $\{X_n\}$ is weak Bernoulli with mixing sequence $\{\beta_n\}$ and there exists a sequence $m = m(n)$ of positive integers such that

$$m\beta_m + m(nb^d)^{-1} = o(1), \tag{4.2.8}$$

then $f_n$ is weakly $L_1$-consistent. Further if we also have

$$\sum_n m(n)\beta_{m(n)} < \infty, \quad \text{and} \tag{4.2.9}$$

$$\frac{m\log n}{n} = o(1), \tag{4.2.10}$$

then $f_n$ is strongly $L_1$-consistent.

(B) If $\{X_n\}$ is strong mixing with mixing sequence $\{\alpha_n\}$ and there exists a sequence $m = m(n)$ of positive integers such that

$$mb^{-d/2}\alpha_m + m(nb^d)^{-1} = o(1) \tag{4.2.11}$$

then $f_n$ is weakly $L_1$ consistent. Further if we also have

$$\sum_n m(n)[b(n)]^{-d/2}\alpha_{m(n)} < \infty, \quad \text{and} \tag{4.2.12}$$

$$\frac{m\log n}{n} = o(1) \tag{4.2.13}$$

then $f_n$ is strongly $L_1$-consistent.

Proof: See Appendix 4-B.

Now we gather some corollaries to the above result.

Corollary 4.2.4: If $\{X_n\}$ is weak Bernoulli, $\Sigma\beta_n < \infty$ and $b+(nb^d)^{-1} = o(1)$ then $f_n$ is weakly $L_1$-consistent.

Proof: Let $m = m(n) = [\![\sqrt{nb^d}]\!]$. Obviously $m(nb^d)^{-1} = o(1)$. On the other hand as $\Sigma\beta_n < \infty$, and $\lim m(n) = \infty$ we also have $m\beta_m = o(1)$.□

The above result improves Corollary 3.1 in Tran (1989). We relax both Tran's assumption on the sequence of bandwidths and his assumption on the sequence of $\beta$-weights.

Corollary 4.2.5: If $\{X_n\}$ is exponentially strong mixing and $b+(\log n)(nb^d)^{-1} = o(1)$ then $f_n$ is strongly $L_1$-consistent.

Proof: Take A so that $\alpha(A[[\log n]]) \leqslant n^{-2-\epsilon}$ for all n, and define $m(n) = A[[\log n]]$. We have, for n large enough,

$$mn^{-d/2}\alpha(m) < \sqrt{n}\ .\ \sqrt{n}\ .\ \alpha(m) < nn^{-2-\epsilon} = n^{-1-\epsilon};$$

Thus 4.2.12 holds valid. The other condition, 4.2.13 is obviously valid. □

The above result improves Corollary 2.1 in Tran (1989), where $L_1$ strong consistency was shown, provided that $(\log n)^3(nb^d)^{-1} = o(1)$.

Corollary 4.2.6: If $\{X_n\}$ is strong mixing with $\alpha_n = n^{-(1+\theta)}$, where $\theta > 0$ and $(n^{2\theta/(1+2\theta)}b^d)^{-1} = o(1)$ then $f_n$ is weakly $L_1$-consistent.

Proof: Define $m = [[n^{1/(1+2\theta)}]]$. We have

$$m(nb^d)^{-1} = 0(n^{2\theta/(1+2\theta)}b^d)^{-1} = o(1);$$

On the other hand

$$mb^{-d/2}\alpha_m = 0(n^{1/(1+2\theta)}).o(n^{\theta/(1+2\theta)}).0(n^{-(1+\theta)/(1+2\theta)})$$
$$= o(1).$$

Hence 4.2.11 holds valid. □

The above result improves Corollary 2.2 in Tran (1989), where a more restrictive rate of decay for $\{\alpha_n\}$ was imposed. Further Corollary 4.2.6 above allows the sequence of bandwidths decreasing faster than in Tran's result.

<u>Corollary 4.2.7</u>: If $\{X_n\}$ is strong mixing with $\alpha_n = n^{-(2+\delta)}$, where $\delta > 0$ and $(\log\ n)^{(1+\epsilon)/(2+\delta)}(n^{\delta/2+\delta}b^d)^{-1} = o(1)$, for some $\epsilon > 0$ then $f_n$ is strongly $L_1$-consistent.

<u>**Proof**</u>: Take $m = [[n^{2/(2+\delta)}.(\log\ n)^{(1+\epsilon)/(2+\delta)}]]$. We have

$$mb^{-d/2}\alpha_m = m.o((nm^{-1})^{\frac{1}{2}}).\alpha_m = \sqrt{mn}.\alpha_m.o(1)$$
$$\leq n\alpha_m.o(1) = [n.(\log(n))^{1+\epsilon}]^{-1}.o(1)$$

Thus $\Sigma mb^{-d/2}\alpha_m < \infty$. On the other hand 4.2.13 and $m(nb^d)^{-1} = o(1)$ are obviously valid. □

Corollary 4.2.7 above ameliorates Corollary 2.3 in Tran (1989) in the same fashion as Corollary 4.2.6 does over Tran's Corollary 2.2. It also improves Theorem 2 in Gyorfi (1987) provided that $\delta > 1$. For
$0 < \delta < 1$, Gyorfi's rates are better than ours.

<u>Corollary 4.2.8</u>: If $\{X_n\}$ is weak Bernoulli with $\beta_n = n^{-(2+\delta)}$ and $(\log\ n)^{(1+\epsilon)(1+\delta)}(n^{\delta/1+\delta}b^d)^{-1} = o(1)$ for some $\epsilon > 0$ then $f_n$ is strongly $L_1$-consistent.

**Proof**: Take $m = [[(n^2b^d(\log n)^{1+\epsilon})^{1/2+\delta}]]$. Simple algebraic manipulations show that

$$m = o(nb^d),$$

and therefore

$$m\beta_m = o(nb^d)\beta_m = o(nb^d).0(n^{-2}b^{-d}(\log(n))^{-1-\epsilon}),$$

thus $\Sigma m\beta_m < \infty$. On the other hand 4.2.10 is obviously true. □

Corollary 4.2.8 above improves Corollary 3.2 in Tran (1989) in the same way as Corollary 4.2.6 does over Tran's Corollary 2.2.

As far as $L_1$-consistency for HDE's is concerned we have the following:

THEOREM 4.2.9: Let $\{X_n\}$ be a stationary process taking values in $\mathbb{R}^d$ and having common density f. Let $\{Q_n\}$, $n > 1$ be a sequence of $\mathbb{R}^d$-partitions and $f_n$ be the HDE of f subordinated to $Q_n = \{A_{nj}, j>1\}$ (say). Assume that

(HO) $\cap_{n=1}^{\infty}\sigma(\cup_{m=n}^{\infty}Q_m) = \{\text{Borel sets of } \mathbb{R}^d\}$,

(H1) For all Borel set A in $\mathbb{R}^d$ with $0 < \lambda(A) < \infty$ and for all $\epsilon > 0$ there exists an $n_0$ such that we can find $A_n$ in $\sigma(Q_n)$ with $\lambda(A \Delta A_n) < \epsilon$, where $\Delta$ stands for the symmetric difference,

(H2) There exists a sequence $m = m(n)$ of positive integers such that for all $M>0$ and all Borel set C in $\mathbb{R}^d$, with $\lambda(C) < \infty$ we have

$$\lim\sup_n \lambda(U_jA_{nj} \cap C) = 0,$$

where the union above is over the integers j satisfying $\lambda(A_{nj} \cap C) \leq Mm/n$.

Then either (A1) or (B1) below imply that $f_n$ is weakly $L_1$-consistent.

Furthermore either (A1) and (A2) or (B1) and (B2) imply that $f_n$ is strongly $L_1$-consistent.

(A1) $\{X_n\}$ is weak Bernoulli with mixing sequence $\{\beta_n\}$ satisfying

$$m\beta_m = o(1).$$

(A2) $\Sigma m(n)\beta_{m(n)} < \infty$, and

$$\frac{m\log n}{n} = o(1).$$

(B1) $\{X_n\}$ is strong mixing with mixing sequence $\{\alpha_n\}$ satisfying

$$n\alpha_m = o(1).$$

(B2) $\Sigma n\alpha_m < \infty$, and

$$\frac{m\log n}{n} = o(1).$$

<u>Proof</u>: See Appendix 4-B.

Obviously a number of corollaries to theorem 4.2.4 could be obtained in the same fashion as we did for theorem 4.2.3. We will not state such corollaries so as to avoid repetition.

### 4.2.5 <u>Some Consequences</u>

In this section we state some simple consequences of the results shown in section 4.2.4. The first of them is a strengthened variant of Glivenko-Contelli theorem.

<u>Corollary 4.2.10</u>: Let $\{X_n\}$ be a stationary process taking values in $R^d$ and having common density f. Let us also assume that $\{X_n\}$ is exponentially strong mixing. Let $f_n$ be a kernel density estimate based on $X_1,\ldots,X_n$, a kernel W, which we

require to be a density, and a bandwidth $b = b(n)$. Let us define

$$F_n(x_1,x_2,\ldots,x_d) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_d} f_n(\underset{\sim}{x})d\underset{\sim}{x},$$

where $\underset{\sim}{x} = (x_1,\ldots,x_d)$. We will make no distinction between $F_n$ as above defined and the measure in $\mathbb{R}^d$ generated by $F_n$. If $b^d + (\log n)(nb^d)^{-1} = o(1)$ then

$$\sup_{B \in \mathcal{B}^d} |F_n(B) - F(B)| \to 0, \quad \text{a.s. as } n \to \infty,$$

where $F(.)$ is the probability measure in $\mathbb{R}^d$ induced by $X_1$, and $\mathcal{B}^d$ is the class of Borel sets in $\mathbb{R}^d$.

<u>**Proof**</u>: $P\{\sup_{B \in \mathcal{B}^d} |F_n(B) - F(B)| > \epsilon\} = P\{\sup_{B \in \mathcal{B}^d} |\int_B f_n - \int_B f| > \epsilon\}$

Now, the result follows by virtue of Corollary 4.2.5. □

Obviously the result above could have been written in more general terms, taking into account the assumptions on theorem 4.2.3. We have chosen the exponential strong mixing assumption so as to keep the statement short.

Glivenko–Cantelli theorem states that for iid samples we have

$$\sup_{B \in \mathcal{D}} |\mu_n(B) - F(B)| \to 0, \quad \text{a.s.} \quad \text{as } n \to \infty,$$

where $\mathcal{D} = \{(-\infty,\underset{\sim}{x}],\ \underset{\sim}{x} \in \mathbb{R}^d\}$, and $\mu_n$ is the empirical measure based on $X_1,\ldots,X_n$. When $X_1$ admits of a density we have

$$\sup_{B \in \mathcal{B}^d} |\mu_n(B) - F(B)| = 1, \quad \text{a.s.,} \quad \text{for all } n,$$

because for each w the support of $\mu_n(.,w)$ has null F-measure. Hence a result like Corollary 4.2.10 is obviously not true if we replace $F_n$ by $\mu_n$.

An interesting study on the estimate $F_n$ above is presented in Sarda (1991). He presents an alternative characterization of $F_n$ and addresses the issue of optimality choosing the sequence of bandwidths.

A second simple consequence is the following.

<u>Corollary 4.2.11</u>: If $\{X_n\}$ is a scalar process satisfying the assumptions of Corollary 4.2.10 then the d-dimensional joint density of $X_1,\ldots,X_d$ (assumed to exist) can be $L_1$-consistently estimated by considering the process $Y_n$ = $(X_n,\ldots,X_{n+d-1})$.

<u>Proof</u>: Immediate. □

### 4.3 <u>Regression and Conditional Density Estimates</u>

This section is devoted to the study of $L_1$-consistency in a different context. We will deal with a somewhat similar concept of $L_1$-consistency suitably adapted to regression function estimates. Such a concept was originated in Stone (1977) in a framework much more general than that we are going to study below. We will consider a particular kind of nonparametric regression function estimate and prove its strong universal consistency. The concept of universal consistency is discussed in subsection 4.3.1. The definition and some relevant results on the regression estimate with which we will be concerned are treated in subsection 4.3.2. Strong universal consistency for such estimates are proved in subsection 4.3.3. Finally in subsection 4.3.4 we show a strong $L_1$-consistency result for conditional density estimates.

### 4.3.1 Universal Consistency and Regression Estimation

Let (X,Y) be a random vector taking values in $R^d \otimes R$. Suppose that $E|Y| < \infty$. The regression function of Y given X is defined by the conditional expectation of Y given X as follows

$$r(x) = E(Y|X=x), \quad x \in R^d.$$

The issue of estimating (either parametrically or nonparametrically) regression functions is perhaps the most discussed one in the statistical literature. The importance of good nonparametric estimates can be justified on the grounds of at least two factors. Firstly, in a exploratory stage practitioners might gain insights on the choice of particular parametric models. Nonparametric regression estimates may also help a great deal in a confirmatory stage, when certain features of a particular parametric model are to be checked (e.g. the functional form of the assumed parametric model). Secondly, depending on the amount and quality of data, nonparametric regression function estimates might be used, avoiding possible misspecification problems brought about by a particular parametric functional form.

Let us now define, in its full generality, a nonparametric estimate of regression functions. Let $Z_n = (X_n,Y_n)$, $n \geq 1$ be a stationary process taking values in $R^d \otimes R$ and put

$$r(x) = E(Y|X=x).$$

A sequence $\{r_n\}$ of nonparametric estimates of r(x) is defined as

$$r_n(x) = r_n(x,Z_1,\ldots,Z_n), \quad x \in R^d,$$

where $r_n$ is a measurable function of its arguments.

The possible advantages of nonparametrically estimating a regression function over adopting a parametric estimation strategy are obviously connected to the generality of the former. In other words, in an ideal situation, one would like that criteria of "good" performance of estimates (e.g. asymptotic unbiasedness, consistency, efficiency etc.) were valid for $\{r_n\}$ independently of the joint distribution of $Z_1$. Obviously that is a formidable task. Prior to Stone's paper (op.cit.) a number of authors have established several theorems on the asymptotic behaviour of particular regression function estimates. Such results deal with such important asymptotic criteria of performance as uniform consistency, asymptotic normality, and so forth. Most of those results rely, in one way or another, on a sort of smoothness assumption either on r(.) or on the distribution of X. Clearly, even with such assumptions, the generality of nonparametric estimates is still enormous.

Stone (1977) introduced the following concept of universal consistency for nonparametric estimates of the regression function.

<u>Definition</u>: Let $Z_n = (X_n, Y_n)$, $n > 0$ be a stationary process taking values in $\mathbb{R}^d \otimes \mathbb{R}$. We say that $r_n$ is weak $L_p$-$L_q$ universally consistent for $r(.) = E(Y_0 | X_0 = .)$ iff

$$\lim_n E\int |r_n(x) - r(x)|^p \mu(dx) = 0,$$

for all possible distributions of $(X_0, Y_0)$ with $\|Y_0\|_q < \infty$, $q > p > 1$, where $\mu$ denotes the probability distribution of X.

The concept of universal consistency is thoroughly characterized when one deals with a particular (though very general) class of nonparametric regression estimates. Namely consider the following estimate of r(.)

$$r_n(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i,$$

where the sequence of weights $W_{ni}(x) = W_{ni}(x, X_1, \ldots, X_n)$ satisfy

$$\sum_{i=1}^{n} W_{ni}(x) = 1, \quad \text{all } x \text{ in } R^d, \quad \text{and}$$

$$W_{ni}(x) \geq 0, \quad 1 \leq i \leq n, \quad \text{all } x \text{ in } R^d.$$

In Stone (1977) we learn

**THEOREM 4.3.1**: Let $r_n$ be as above and suppose that $\{Z_n\}$ $n \geq 0$, is a i.i.d. sequence of random vectors. The following conditions are necessary and sufficient for the weak $L_p$–$L_p$ universal consistency of $r_n$.

(i) There exists a constant $C \geq 1$ such that for every nonnegative Borel measurable function f on $R^d$,

$$E[\Sigma_i W_{ni}(X_0) f(X_i)] \leq CEf(X_i), \quad n \geq 1,$$

(ii) $\Sigma_i W_{ni}(x_i) I\{\|X_i - X_0\| > \epsilon\} = o_p(1)$, for all $\epsilon > 0$, as $n \to \infty$.

(iii) $\max_i |W_{ni}(X_0)| = O_p(1)$, as $n \to \infty$. □

Stone applied the results above to prove weak $L_p$–$L_p$ universal consistency for nearest neighbour estimates. Weak $L_p$–$L_p$ consistency for kernel estimates was considered in Devroye and Wagner (1980) and Spiegelman and Sacks (1980). Recently, some research has been done on strong universal consistency for regression estimates.

**Definition**: We say that $r_n$ is strongly $L_p$–$L_q$ universal consistent iff

$$\lim_n \int |r_n(x) - r(x)| P\mu(dx) = 0, \quad \text{a.s.},$$

for all possible distributions of (X,Y) with $\|Y\|_q < \infty$, for $q \geq p \geq 1$.

Devroye and Krzyzak (1987) proved that weak and strong $L_1$–$L_\infty$ universal consistency are equivalent for kernel regression estimates under iid sampling. Gyorfi established strong $L_p$-$L_p$ consistency for a modified version of the partitioning estimate.

**Definition**: Let $Q_n = \{A_{n1}, A_{n2}, \ldots\}$ be a sequence of $\mathbb{R}^d$ partitions. We define the partitioning estimate of r(.) subordinated to $\{Q_n\}$ by

$$r_n(x) = \begin{cases} \nu_n(A_{ni})/\mu_n(A_{ni}) & \text{if } x \in A_{ni} \text{ and } \mu_n(A_{ni}) > 0, \\ \frac{1}{n} \sum_{i=1}^{n} Y_i & \text{if } \mu_n(A_{ni}) = 0, \text{ and } x \in A_{ni} \end{cases}$$

where $\mu_n(.)$ is the empirical measure based on $(X_1, \ldots, X_n)$ and

$$\nu_n(A) = n^{-1} \sum_{i=1}^{n} I\{X_i \in A\} Y_i . \qquad \square$$

The motivation for such an estimate can be outlined as follows: Let us write

$$E(Y|X=x) = \int y dF(y|X=x) .$$

Now estimate $F(y|X=x)$ by

$$F_n(Y|X=x) = \begin{cases} \dfrac{\sum_{k=1}^{n} I\{Y_k < y,\ X_k \in A_{ni}\}}{\sum_{k=1}^{n} \{X_k \in A_{ni}\}}, & \text{if } x \in A_{ni} \text{ and } \mu_n(A_{ni}) > 0 \\ n^{-1} \sum_{k=1}^{n} I\{Y_k < y\}, & \text{if } \mu_n(A_{ni}) = 0 \text{ and } x \in A_{ni} \end{cases}$$

With such an estimate we obviously have

$$r_n(x) = \int y dF_n(y|X=x) .$$

Devroye and Gyorfi (1985b) proved strong $L_1$-$L_\infty$ consistency for the partitioning estimate based on iid samples. The assumption of independence was relaxed in

Gyorfy et.al. (1989).

### 4.3.2 Universal $L_1-L_\infty$ consistency for the partitioning estimate under mixing assumptions

Our aim in this subsection is to study sufficient conditions for universal $L_1-L_\infty$ consistency of the partitioning estimate under dependent samples. As in subsection 4.2.4, our main tool will be an approximation theorem.

Let $Z_n = (X_n,Y_n)$ be a stationary sequence of random vectors taking values in $R^d \otimes R$. Let us define

$$\nu_n(A) = n^{-1} \sum_{i=1}^{n} I(X_i \in A)\, Y_i, \quad A \in \mathcal{B}^d, \quad \text{and}$$

$$\nu(A) = \int_A r(x)\mu(dx) = E\nu_n(A), \quad A \in \mathcal{B}^d.$$

<u>THEOREM 4.3.2</u>: Suppose that $\{(X_n,Y_n)\}$, $n \geq 1$ is a strong mixing (weak Bernoulli, respectively) process such that $|Y_1| \leq C$, a.s. Let $\mathcal{H} = \{H_j,\ 1 \leq j \leq N\}$ be a measurable partition of $R^d$. Let m be a positive integer and $\epsilon$ be a positive real number such that

$$4Cm < n\epsilon \tag{4.3.1}$$

then

$$P\{\sum_{k=1}^{N} |n\nu_n(H_k) - n\nu(H_k)| > n\epsilon\}$$

$$\leq m\theta(m) + 5\ 2^N \exp(-n\epsilon^2 / (m[32C]^2))$$

where $\theta(m) = 384\ (C/\epsilon)^{3/2}\ \alpha(m)$ $\quad$ ( $\theta(m) = \beta(m)$, resp. )

**<u>Proof</u>**: Define $M = [[n/m]]$. We have

$$
\begin{aligned}
&P\{\sum_{i=1}^{N} |\nu_n(H_i) - \nu(H_i)| > \epsilon\} \\
&= \; P\{\sum_{i=1}^{N} | \sum_{k=1}^{n} \; I(X_k \in H_i)Y_k - n\nu(H_i)| > n\epsilon\} \\
&\le \; \sum_{j=1}^{m} \; P\{\sum_{i=1}^{N} | \sum_{k=j,m}^{mM} \; I(X_k \in H_i)Y_k - M\nu(H_i)| > M\epsilon/2\} \\
&+ P\{\sum_{i=1}^{N} | \sum_{k=mM+1}^{n} \; \{I(X_k \in H_i)Y_k - \nu(H_i)| > n\epsilon/2\} \\
&= \; mP\{\sum_{i=1}^{N} | \sum_{k=1}^{M} \; I(X_{km} \in H_i)Y_k - M\nu(H_i)| > M\epsilon/2\} \\
&P\{\sum_{i=1}^{N} | \sum_{k=mM+1}^{n} \; \{I(X_k \in H_i)Y_k - \nu(H_i)| > n\epsilon/2\}, \qquad (4.3.2)
\end{aligned}
$$

where as before

$$
\sum_{k=j,m}^{R} a_k \;=\; \sum_{k:k=j+im} a_k \; , \qquad i \ge 0, \; k \le R,
$$

and the last equality in 4.3.2 holds valid by stationarity. Now

$$
\begin{aligned}
&\sum_{i=1}^{N} | \sum_{k=mM+1}^{n} \; [I(X_k \in H_i)Y_k - \nu(H_i)]| \\
&\le \; \sum_{k=mM+1}^{n} \; \sum_{i=1}^{N} \; |I(X_k \in H_i)Y_k| + \sum_{k=mM+1}^{n} \; \sum_{i=1}^{N} |\nu(H_i)| \\
&\le C(n-mM) + \sum_{k=mM+1}^{n} \int |E(Y|X=x)|\mu(dx) \\
&\le \; 2Cm.
\end{aligned}
$$

Thus the second term in 4.3.2 equals zero because of 4.3.1.

Now let $\ell$ be an integer such that

$$
(\ell-1)\,\frac{\epsilon}{9} \; < \; 2C \; \le \; \frac{\ell\epsilon}{9} \; ,
$$

and define

$$
J_i \; = \; [-C + (i-1)\epsilon/9, \; -C + i\epsilon/9), \qquad 1 \le i < \ell, \qquad \text{and}
$$

$$J_\ell = [-C + (\ell-1)\epsilon/9,\ C].$$

Now consider the following partition of $R^d \otimes R$

$$H^* = \{H;\ \text{either } H = H_i \otimes J_j,\ 1 \leq i \leq N,\ 1 \leq j \leq \ell \quad \text{or}$$
$$H = R^d \otimes R - \bigcup_{i=1}^{N} \bigcup_{j=1}^{\ell} H_i \otimes J_j\}$$

Obviously $H^*$ possesses $N\ell+1$ elements. Now by THEOREM 4.2.1 (Berbee's theorem, respectively) we can find (enlarging the original p–space if necessary) random vectors $T_m$, $T_{2m}$, ... such that

(i) $T_{mj} = (U_{mj}, V_{mj})$ and $Z_{mj} = (X_{mj}, Y_{mj})$ are identically distributed,

(ii) $T_{mj}$ is independent of $\{T_{ms}, 1 \leq s < j\}$,

(iii) $P\{T_{mj}$ and $Z_{mj}$ are not elements of the same $H \in H^*\} \leq (8(N\ell+1))^{\frac{1}{2}}\alpha(m)[\leq \beta(m)$, respectively$]$.

It is easy to show that $\{T_m, T_{2m}, ...\}$ is a collection of independent random vectors.

The first term in the right hand side of 4.3.2 can be majorized by

$$P\{\sum_{i=1}^{N} |\sum_{k=1}^{M} I(X_{km} \in H_i)Y_{km} - I(U_{km} \in H_i)V_{km}| > M\epsilon/4\}$$

$$P\{\sum_{i=1}^{N} |\sum_{k=1}^{M} I(U_{km} \in H_i)V_{km} - M\nu(H_i)| > M\epsilon/4\}$$

$$= (I) + (II)\ .$$

Now let us define

$$A_k = \{Z_{km} \text{ and } T_{km} \text{ belong to the same } H \in H^*\}\ .$$

We can write

$$(I) \leq P\{\sum_{k=1}^{M} \sum_{i=1}^{N} |I(X_{km} \in H_i)Y_{km} - I(U_{km} \in H_i)V_{km}| > M\epsilon/4\}$$

$$= P\{\sum_{k=1}^{M}\sum_{i=1}^{N} | I(X_{km}\in H_i)Y_{km} - I(U_{km}\in H_i)V_{km}| I(A_k) > M\epsilon/8\}$$

$$+ P\{\sum_{k=1}^{M}\sum_{i=1}^{N} | I(X_{km}\in H_i)Y_{km} - I(U_{km}\in H_i)V_{km}| I(A_k^c) > M\epsilon/8\}$$

$$= (III) + (IV) .$$

Now, as $|Y_{km}| \leq C$ we have for each $k \leq M$

$$\sum_{i=1}^{N} | I(X_{km}\in H_i)Y_{km} - I(U_{km}\in H_i)V_{km}| I(A_k) \leq \epsilon/9.$$

Whence (III) = 0. On the other hand, by Markov inequality,

$$(IV) \leq (M\epsilon/8)^{-1} \sum_{k=1}^{M} E \sum_{i=1}^{N} | I(X_{km}\in H_i)Y_{km} - I(U_{km}\in H_i)V_{km}| I(A_k^c)$$

$$\leq (8/\epsilon).\max_{k\leq M} E \sum_{i=1}^{N} | I(X_{km}\in H_i)Y_{km} - I(U_{km}\in H_i)V_{km}| I(A_k^c)$$

$$\leq (8/\epsilon).(2C).\max_{k\leq M} P\{A_k^c\}$$

$$\leq (16C/\epsilon).(8(N\ell+1))^{\frac{1}{2}}\alpha(m)[\leq \beta(m), \text{ respectively}]$$

$$\leq 64C/\epsilon.(N\ell)^{\frac{1}{2}}\alpha(m)[\leq \beta(m), \text{ respectively}]$$

$$\leq 64C/\epsilon.(36C/\epsilon)^{\frac{1}{2}}.N^{\frac{1}{2}}.\alpha(m)[\leq \beta(m), \text{ respectively}], \tag{4.3.3}$$

where we have assumed w.l.o.g. that $1 \leq 18C/\epsilon$. Now, as far as (II) is concerned we can write

$$\sum_{i=1}^{N} | \sum_{k=1}^{M} I(U_{km}\in H_i)V_{km} - M\nu(H_i)|$$

$$\leq \sum_{i=1}^{N} | \sum_{k=1}^{M} [ I(U_{km}\in H_i)V_{km} + (EV_{km})I(U_{km}\in H_i) - \mu(H_i)V_{km}$$

$$- EV_{km}I(U_{km}\in H_i) ]|$$

$$+ \sum_{i=1}^{N} | \sum_{k=1}^{M} [ (EV_{km})I(U_{km}\in H_i) - (EV_{km})\mu(H_i) ]|$$

$$+ \sum_{i=1}^{N} | \sum_{k=1}^{M} [ \mu(H_i)V_{km} - (EV_{km})\mu(H_i) ]|$$

$$= (V) + (VI) + (VII).$$

We have

$$(\mathrm{VII}) = \left| \sum_{k=1}^{M} V_{km} - ME(V_{1m}) \right|, \quad \text{and}$$

$$(\mathrm{VI}) \leq |EV_{1m}| \sum_{i=1}^{N} \left| \sum_{k=1}^{M} I(U_{km} \in H_i) - M\mu(H_i) \right|$$

$$\leq C \sum_{i=1}^{N} \left| \sum_{k=1}^{M} I(U_{km} \in H_i) - M\mu(H_i) \right| .$$

Now notice that

$$\sum_{i=1}^{N} \sum_{k=1}^{M} \Big[ I(U_{km} \in H_i)V_{km} + (EV_{km})I(U_{km} \in H_i)$$

$$- \mu(H_i)V_{km} - EV_{km}I(U_{km} \in H_i) \Big]$$

$$= 0.$$

Hence

$$(\mathrm{V}) = \sum_{i=1}^{N} \left| \sum_{k=1}^{M} \Big[ I(U_{km} \in H_i)V_{km} + (EV_{km})I(U_{km} \in H_i) - \mu(H_i)V_{km} \right.$$

$$\left. - EV_{km}I(U_{km} \in H_i) \Big] \right|$$

$$= 2 \sup_{G\in\sigma(H)} \sum_{k=1}^{M} \Big[ I(U_{km} \in G)V_{km} + (EV_{km}I(U_{km} \in G) - \mu(G)V_{km}$$

$$- EV_{km}I(U_{km} \in G) \Big].$$

Thus, taking into account the bounds for (V), (VI) and (VII), we have

$$(\mathrm{II}) \leq P\left\{ \left| \sum_{k=1}^{M} V_{km} - ME(V_{1m}) \right| > M\epsilon/16 \right\}$$

$$+ P\left\{ \sum_{i=1}^{N} \left| \sum_{k=1}^{M} I(U_{km} \in H_i) - M\mu(H_i) \right| > M\epsilon/(16C) \right\}$$

$$+ P\left\{ \sup_{G\in\sigma(H)} \sum_{k=1}^{M} \Big[ I(U_{km} \in G)V_{km} - EV_{km}I(U_{km} \in G) \Big] > M\epsilon/32 \right\}$$

$$+ P\left\{ \sup_{G\in\sigma(H)} \sum_{k=1}^{M} \Big[ (EV_{km})I(U_{km} \in G) - \mu(G)V_{km} \Big] > M\epsilon/32 \right\} .$$

Now by Hoeffding's inequality and the fact that $\sigma(H)$ possesses $2^N$ elements we can easily see that the right hand side of the above inequality can be upper bounded by

$$2 \exp\{-2\ M\epsilon^2/(16C)^2\} + 2^N.3.\exp\{-2\ M\epsilon^2/(32C)^2\}$$
$$< \ 5.2^N.\exp\{-\ n\epsilon^2/[m(32C)^2]\}.$$

Such bound and 4.3.3 imply the result. □

Devroye and Gyorfi (1985b) have shown strong $L_1$–$L_\infty$ universal consistency for the partitioning estimate of regression functions under iid samples. Their smoothness assumptions were the following

(A1) For any sphere S in $R^d$ we have

$$\lim_n n^{-1} \# \{i;\ A_{ni} \cap S \neq \Phi\} = 0,$$

(A2) For any sphere S in $R^d$ we have

$$\lim_n \max_{i:A_{ni}\cap S\neq\Phi} \{\mathrm{diam}(A_{ni})\} = 0,$$

where $\mathrm{diam}(A) = \sup\{|x-y|;\ x,y \in A\}$.

It is easy to show that if $\{Q_n\}$ is a sequence of cubic partitions, that is each $A_{ni}$ is a rectangle of form

$$\prod_{j=1}^{d} [a_j k_{ij} b(n),\ a_j(k_{ij} + 1)b(n)],$$

where $a_1,\ldots,a_d$ are positive real numbers and $\{k_{ij}\}$ are integers then (A1) is equivalent to

$$\lim_n nb^d = \infty,$$

whereas (A2) is equivalent to

$$\lim_n b = 0.$$

The following result is essential for dealing with the nonstochastic part (bias) of

$\int |r_n(x) - r(x)|\mu(dx)$.

THEOREM 4.3.3. (Devroye and Gyorfi (1985b)). Let S be a sphere and put

$$I_{Sn} = \{i: A_{ni} \cap S \neq \Phi\},$$

then under (A2) we get

$$\lim_n \sum_{i \in I_{Sn}} \int_{A_{ni}} \left|\frac{\nu(A_{ni})}{\mu(A_{ni})} - r(x)\right| \mu(dx) = 0. \qquad \square$$

We can now state our main result.

THEOREM 4.3.4: Let $\{Z_n = (X_n, Y_n)\}$, $n \geq 1$ be a stationary process taking values in $\mathbb{R}^d \otimes \mathbb{R}$, and assume that $|Y_1| < C$. Let $\{r_n(.)\}$ be a sequence of partitioning estimates of $r(.) = E(Y_1 | X_1 = .)$, subordinated to the sequence of $\mathbb{R}^d$-partitions $\{Q_n\}$, where $Q_n = \{A_{n1}, A_{n2}, \ldots\}$.

(I) if $\{Z_n\}$ is weak Bernoulli with mixing sequence $\{\beta_n\}$ and there exists a sequence $m = m(n)$ of positive integers such that for every sphere S in $\mathbb{R}^d$

$$m\beta(m) = o(1), \tag{4.3.4}$$

$$m \# \{i; A_{ni} \cap S \neq \Phi\} = o(n), \quad \text{and} \tag{4.3.5}$$

$$\max_{i: A_{ni} \cap S \neq \Phi} \{\text{diam}(A_{ni})\} = o(1), \tag{4.3.6}$$

then $E \int |r_n(x) - r(x)|\mu(dx) = o(1)$. Further if we also have

$$\sum m\beta(m) < \infty, \quad \text{and} \tag{4.3.7}$$

$$\frac{m \log m}{n} = o(1) \tag{4.3.8}$$

then $\int |r_n(x) - r(x)|\mu(dx) = o(1)$ a.s.

(II) If $\{Z_n\}$ is strong mixing with mixing sequence $\{\alpha_n\}$ and there exists a sequence $m = m(n)$ of positive integers such that for every sphere S in $\mathbb{R}^d$

$$m\alpha_m\,[\#\{i:\ A_{ni} \cap S \neq \Phi\}]^{\frac{1}{2}} = o(1) \qquad (4.3.9)$$

$$m\ \#\ \{i:\ A_{ni} \cap S \neq \Phi\} = o(n), \quad \text{and} \qquad (4.3.10)$$

$$\max_{i:A_{ni} \cap S \neq \Phi} \{\text{diam}\ (A_{ni})\} = o(1) \qquad (4.3.11)$$

then $E \int |r_n(x) - r(x)|\mu(dx) = o(1)$. Further if we also have

$$\Sigma\ mA_m[\#\ \{i:\ A_{ni} \cap S \neq \Phi\}]^{\frac{1}{2}} < \infty, \quad \text{and} \qquad (4.3.12)$$

$$m \log m = o(n) \qquad (4.3.13)$$

then $\int |r_n(x) - r(x)|\mu(dx) = o(1)$ a.s.

__Proof__: First of all let us just remark that as $\{\int |r_n(x) - r(x)|\mu(dx)\}$ is a sequence of uniformly bounded random variables, we have that $E\int |r_n(x) - r(x)\mu(dx)| = o(1)$ is equivalent to $\int |r_n(x) - r(x)|$ $\mu(dx) = o_p(1)$. Let S be a sphere in $\mathbb{R}^d$. Using the same notation as in theorem 4.3.3 we have:

$$\int |r_n(x) - r(x)|\mu(dx) \leq \int_{S^c} |r_n(x) - r(x)|\mu(dx)$$

$$+ \sum_{i \in I_{S_n}} \int_{A_{ni}} \left|r_n(x) - \frac{\nu(A_{ni})}{\mu(A_{ni})}\right| \mu(dx)$$

$$+ \sum_{i \in I_{S_n}} \int_{A_{ni}} \left|r(x) - \frac{\nu(A_{ni})}{\mu(A_{ni})}\right| \mu(dx)$$

$$= (I) + (II) + (III).$$

We have

$$(I) \leq 2\ \|Y_1\|_\infty . \mu(S^c) ,$$

which can be made small by our choice of S. Also,

$$(III) = o(1),$$

by virtue of theorem 4.3.3, (4.3.5) and (4.3.10). On the other hand

$$
\begin{aligned}
(I) &= \sum_{i\in I_{S_n}} \int_{A_{ni}} \left| \frac{\nu_n(A_{ni})}{\mu_n(A_{ni})} - \frac{\nu(A_{ni})}{\mu(A_{ni})} \right| \mu(dx) \\
&= \sum_{i\in I_{S_n}} \left| \frac{\mu(A_{ni})\nu_n(A_{ni})}{\mu_n(A_{ni})} - \nu(A_{ni}) \right| \\
&< \sum_{i\in I_{S_n}} \frac{1}{\mu_n(A_{ni})} |\mu(A_{ni})\nu_n(A_{ni}) - \mu(A_{ni})\nu(A_{ni})| \\
&+ \sum_{i\in I_{S_n}} \frac{1}{\mu_n(A_{ni})} |\mu_n(A_{ni})\nu_n(A_{ni}) - \mu_n(A_{ni})\nu(A_{ni})| \\
&= \sum_{i\in I_{S_n}} \frac{\nu_n(A_{ni})}{\mu_n(A_{ni})} |\mu_n(A_{ni}) - \mu(A_{ni})| \\
&\quad + \sum_{i\in I_{S_n}} |\nu_n(A_{ni}) - \nu(A_{ni})| \\
&< \|Y_1\|_\infty \sum_{i\in I_{S_n}} |\mu_n(A_{ni}) - \mu(A_{ni})| \\
&\quad + \sum_{i\in I_{S_n}} |\nu_n(A_{ni}) - \nu(A_{ni})| .
\end{aligned}
$$

The result follows from a simple application of theorems 4.3.2 and 4.2.2. □

As in theorem 4.2.3, a number of corollaries could be drawn from the above result. We will not state them for brevity. Let us just remark that for strong mixing processes with $\alpha_n = O(n^{-(2+\delta)})$, our (not stated) corollary improved Theorem 3.2.1 in Gyorfi et.al. (1989).

### 4.3.3 $L_1$-strong consistency for conditional density estimates

Our goal in this subsection is to establish a simple result which parallels theorem

4.3.4 above, when one is estimating conditional densities, instead of regression functions. Let $\{Z_n\}$, $n \geq 1$ be a stationary process taking its values in $R^{c+d}$, with $c,d > 0$. Put $Z_n = (X_n,Y_n)$, where $X_n$ is c–dimensional and $Y_n$ is d–dimensional. Let us assume that both $Z_1$ and $X_1$ admit of densities given by $f: R^{c+d} \to R$ and $g: R^c \to R$. In this case we have a conditional density for $Z_1$ given $X_1 = x$, given by

$$h(.\,|x) = \begin{cases} f(x,.)/g(x) & \text{if } g(x) > 0 \\ q(.) & \text{if } g(x) = 0, \end{cases}$$

where $q: R^d \to R$ is an arbitrary density in $R^d$. A natural estimate of $h(.\,|.)$ is the following

$$h_n(y|x) = \begin{cases} f_n(x,y)/g_n(x) & \text{if } g_n(x) > 0 \\ q(y) & \text{if } g_n(x) = 0, \end{cases}$$

where $f_n(x,y)$ [$g_n(x)$, respectively] is a nonparametric estimate of $f(x,y)$ [$g(x)$, respectively]. It is possible, obviously, to consider different classes of estimates for $f_n$ and $g_n$ (e.g. to adopt a KDE for $f_n$ and a HDE for $g_n$). However, on doing so one, in general, does not obtain the following property (which holds for $h(.\,|.)$).

(A) for each fixed $x \in R^c$, $h_n(.\,|x)$ is a density in $R^d$.

<u>EXAMPLE 1</u>: Let $V: R^d \to R$ and $W: R^c \to R$ be densities. If we take

$$f_n(x,y) = (nb^{c+d})^{-1} \sum_{k=1}^{n} V[(y-Y_k)/b]W[(x-X_k)/b], \quad \text{and}$$

$$g_n(x) = (nb^c)^{-1} \sum_{k=1}^{n} W[(x-X_k)/b],$$

and define $h_n(y|x)$ as above then (A) holds valid. More generally, take $U: R^{c+d} \to R$ a density in $R^{c+d}$. Take $W: R^d \to R$ as its first marginal, that is

$$W(x) = \int U(x,y)dy,$$

and define

$$f_n(x,y) = (nb^{c+d})^{-1} \sum_{k=1}^{n} U[(x-X_k)/b,\ (y-y_k)/b]$$

$$g_n(x) = (nb^c)^{-1} \sum_{k=1}^{n} W[(x-X_k)/b].$$

Taking $h_n(y|x)$ as above implies that (A) is valid.

EXAMPLE 2: Let $\{Q_n\}$, $n > 1$ ($\{R_n\}$, respectively) be a sequence of $R^c$–partitions ($R^d$–partitions, respectively) such that $Q_n = \{A_{n1}, A_{n2}, \ldots\}$ and $R_n = \{B_{n1}, B_{n2}, \ldots\}$. Define $T_n = \{A \otimes B\colon A \in Q_n \text{ and } B \in R_n0\}$.

Define

$$f_n(x,y) = \sum_{D \in T_n} I\{(x,y) \in D\}\nu_n(D)/\lambda_{c+d}(D), \quad \text{and}$$

$$g_n(x) = \sum_{A \in Q_n} I\{x \in A\}\mu_n(A)/\lambda_c(D)$$

where $\nu_n(.)$ [$\mu_n(.)$, respectively] denotes the empirical measure based on $(X,Y),\ldots,(X_n,Y_n)$ [$X_1,\ldots,X_n$, respectively], and $\lambda_{c+d}(.)$ [$\lambda_c(.)$, respectively] stands for the Lebesgue measure on $R^{c+d}$ [$R^c$, respectively]. Now defining $h_n(.|.)$ as before, it is straightforward to verify (A).

EXAMPLE 3: If one wishes to estimate f and g by

$$f_n(x,y) = (nb^{c+d})^{-1} \sum_{k=1}^{n} U[(x-X_k)/b,(y-Y_k)/b],$$

$$g_n(x) = (nb^c)^{-1} \sum_{k=1}^{n} W[(x-X_k/b],$$

where U and W are densities in $R^{c+d}$ and $R^c$, respectively, such that W is not the first marginal of U, then (A) does not hold.

We can now state our

<u>THEOREM 4.3.5</u>: Let $f_n$ and $g_n$ be strong $L_1$-consistent density estimates of f and g respectively and assume that (A) holds valid. Then

$$\iint |h_n(y|x) - h(y|x)|dy\ \mu(dx) = o(1), \quad a.s.$$

where $\mu$ is the probability distribution of $X_1$, provided that the following assumption holds

$$\text{(B)} \quad f_n(x,y) = 0, \quad \text{for all } y \text{ if } g_n(x) = 0.$$

<u>Proof</u>: By virtue of (B), we can assume without loss of generality $g_n(x) > 0$, for all x and n. Indeed

$$\int_{x:g_n(x)=0}\int |h_n(y|x) - h(y|x)|dy\ \mu(dx)$$

$$= \int_{x:g_n(x)=0<g(x)}\int |q(y).g(x) - f(x,y)|dy\ dx$$

$$< \iint q(y)|g(x) - g_n(x)|dydy + \iint |f_n(x,y) - f(x,y)|dydx$$

$$= \int |g(x) - g_n(x)|dx + \iint |f_n(x,y) - f(x,y)|dydx$$

$$= 0(1), \quad a.s.$$

Besides, we can also assume with no loss of generality that $g(x) > 0$, for all x. Indeed

$$\int_{x:g(x)=0}\int |h_n(y|x) - h(y|x)|dy\ \mu(dx) = 0,$$

because $\mu(x: g(x) = 0) = 0$.

Now, as

$$\int |g_n(x) - g(x)|dx = o(1) \quad a.s.,$$

we have

$$\iint f(x,y)/g(x)|g(x) - g_n(x)|dydx = o(1) \quad a.s.,$$

or

$$\iint f(x,y)|1 - g_n(x)/g(x)|dxdy = o(1) \quad a.s..$$

But

$$\iint |f_n(x,y) - f(x,y)|dxdy = o(1) \quad a.s..$$

The last two equations imply

$$\iint g_n(x)|f_n(x,y)/g_n(x) - f(x,y)/g(x)|dydx$$
$$= \iint |f_n(x,y) - f(x,y)g_n(x)/g(x)|dydx = o(1) \quad a.s..$$

Now, as for each fixed $\underline{x}$ both $f_n/g_n$ and $f/g$ are densities on $\mathbb{R}^d$, we can write

$$\iint g(x)|f_n(x,y)/g_n(x) - f(x,y)/g(x)|dydx$$

$$= \int_{x:g(x)<g_n(x)} \int + \int_{x:g(x)>g_n(x)} \int$$

$$< \iint g_n(x)|f_n(x,y)/g_n(x) - f(x,y)/g(x)|dydx$$

$$+ 2\int_A g(x) \int_B (f(x,y)/g(x) - f_n(x,y)/g_n(x))dydx,$$

with $A = \{x: g(x) > g_n(x)\}$ and $B = \{y; f(x,y)/g(x) > f_n(x,y)/g_n(x)\}$, where we have used Scheffe's theorem: $\int |u-v| = 2\int (u-v)_+$ if u and v are density. Now the second term on the right hand side of last inequality is less than

$$2 \int_A \int_B (f(x,y) - f_n(x,y))dydx \leq 2 \iint |f_n - f|.$$

The last three relations imply the result. □

Let us just point out that condition (B) is valid for the estimates given in example 1 and 2 above. Finally, let us state a corollary to THEOREM 4.3.5.

<u>THEOREM 4.3.6</u>: Let us consider the estimates given in EXAMPLE 1 above, and suppose that $\{Z_n\}$ is weak Bernoulli (let us recall that $Z_n = (X_n,Y_n)$) with mixing sequence $\{\beta_n\}$. let us assume that there exists a sequence $m = m(n)$ of positive integers such that

$$\sum_n m(n)\beta_{m(n)} < \infty,$$

$$m(nb^{c+d})^{-1} = o(1), \quad \text{and}$$

$$\frac{m \log n}{n} = o(1).$$

Then $\iint |h_n(y|x) - h(y|x)|dy\ \mu(dx) = o(1)$ a.s.

<u>Proof</u>: Immediate from THEOREMS 4.3.5 and 4.2.3. □

Obviously, a similar result is valid for strong mixing (instead of weak Bernoulli) processes; it suffices to use the second half of THEOREM 4.2.3 instead of the first. Also using THEOREM 4.2.9 instead of THEOREM 4.2.3 one could obtain strong $L_1$-consistency for the conditional density estimates introduced in EXAMPLE 2 above.

## APPENDIX 4–A

**LEMMA 4–A–1**: Let $f: R^d \to R$ be an integrable function in $R^d$, Suppose that $W: R^d \to R$ satisfies $\int W = 1$. Then

$$\lim_{b \to 0} \int | \int b^{-d} W(y/b) f(x-y) dy - f(x) | dx = 0$$

Proof: See Devroye and Gyorfi (1985a), Theorem 2.1. □

**LEMMA 4–A–2**: Let $\mu$ be a probability defined in $(R^d, B^d)$, where $B^d = \{\text{Borel sets in } R^d\}$. Take $A \in B^d$. Then

$$\int \mu(x+A) dx = \lambda(A),$$

where $\lambda$ is the Lebesgue measure in $R^d$.

Proof: Let us first suppose that A is the d–fold product of finite intervals. Put $A = \prod_{i=1}^{d} [a_i, b_i]$. We have

$$\int \mu(x+a) dx = \int \int_{x+b}^{x+b} \mu(dy) dx$$

$$= \int \int_{y-b}^{y-a} dx \, \mu(dy) = \int \lambda(A) \mu(dy) = \lambda(A),$$

by Tonelli–Fubini's theorem. We used the following convention

$$\int_u^v P(dy) = \int_{\prod_{i=1}^{d} [u_i, v_i]} P(dy),$$

where $v = (v_1, v_2, \ldots, v_d)$ and $u = (u_1, \ldots, u_d)$, with $v_i > u_i$.

Now let us define

$$\mathcal{G}_n = \{A \in \mathcal{B}^d; \int \mu(x + A\cap I_n)dx = \lambda(A\cap I_n)\},$$

where $I_n = \prod_{i=1}^{d} [-n,n)$. Clearly, any finite interval belongs to $\mathcal{G}_n$ (the argument is the same as above). Also

$$\int \mu(x + R^d\cap I_n)dx = \int \mu(x + I_n)dx = \lambda(I_n)$$

$$= \lambda(R^d\cap I_n),$$

thus

(i) $\quad R^d \in \mathcal{G}_n$.

Also if A, B are $\mathcal{G}_n$-elements such that $A \subseteq B$, then

$$\int \mu(x + (B-A)\cap I_n)dx = \int [\mu(x + B\cap I_n) - \mu(x + A\cap I_n)]dx$$

$$= \lambda(B\cap I_n) - \lambda(A\cap I_n) = \lambda(B-A)\cap I_n).$$

Thus

(ii) $\quad B-A \in \mathcal{G}_n$ .

Finally, if $\{A_m\}$ is a sequence of $\mathcal{G}_n$ elements such that $A_m \subseteq A_{m+1}$ then Lebesgue's Monotone convergence theorem implies

(iii) $\quad \cup A_m \in \mathcal{G}_n$ .

Therefore, $\mathcal{G}_n$ is a $\lambda$-system containing the $\pi$-system of the d-fold product of finite intervals. By Dinkyn's $\pi$-$\lambda$ theorem we have

$$\mathcal{B}^d \supseteq \mathcal{G}_n \supseteq \sigma\{\text{product of finite intervals}\} = \mathcal{B}^d.$$

In other words if $A \in \mathcal{B}^d$ then

$$\int \mu(x + A\cap I_n)dx = \lambda(A\cap I_n), \text{ all } n.$$

Take the limit as n goes to infinity and use Lebesgue's monotone convergence theorem to conclude the proof. $\square$

__LEMMA 4-A-3__: Let $f_n$ be the HDE considered in THEOREM 4.2.9 above. Assume that (H0) holds valid. Then

$$\int |Ef_n - f| \to 0$$

if and only if (H1) holds valid.

__Proof__: See Abou-Jaoude (1976, pp.216-219)). □

# APPENDIX 4–B

## 1. Proof of THEOREM 4.2.3

Let us define

$$g_b(x) = Ef_n(x) = \int b^{-d}W_b(x-y)f(y)dy,$$

where we have put $W_b(z) = W(z/b)$. We have

$$\int |f_n - f| < \int |f_n - g_b| + \int |g_b - f|.$$

But

$$\int |g_b - f| = \int |\int b^{-d}W_b(x-y)f(y)dy - f(x)|dx$$

$$= \int |\int b^{-d}W_b(y-f(x-y)dy - f(x)|dx$$

$$= o(1),$$

by Lemma 4–A–1. Thus we need only to consider

$$J_n = \int |f_n - g_b|, \quad \text{(say)}.$$

For a given $\epsilon > 0$, we can choose constructs M, L, N, $Q_1,\ldots,Q_N$, and finite disjoint rectangles $A_1,\ldots,A_N$ such that

$$\max\{|a_k|, 1 < K < N\} < M,$$

$$\bigcup_{i=1}^{N} A_i \subseteq [-L,L]^d, \quad \text{and}$$

$$\int |W(x) - W^*(x)|dx < \epsilon,$$

where we have put $W^* = \sum_{i=1}^{N} a_i I_{A_i}(x)$. Define $g_b^*$ and $f_n^*$ as $g_b$ and $f_n$ with $W^*$ instead of W. We have

$$J_n < \int |f_n - f_n^*| + \int |f_n^* - g_b^*| + \int |g_b^* - g_b|$$

$$< \int b^{-d} \int |W_b(x-y) - W_b(x-y)|\mu_n(dy)dx$$

$$+ \int b^{-d} \int |W_b(x-y) - W_b(x-y)|f(y)dydx$$

$$+ \int |f_n^* - g_b^*| < 2\epsilon + \int |f_n^* - g_b^*| ,$$

by Fubini's theorem. Let us just recall that we are denoting by $\mu_n$ ($\mu$, respectively) the empirical measure based on $X_1,\ldots,X_n$ (the probability distribution of $X_1$, respectively). We have

$$\int |f_n^* - g_b^*| < \sum_{i=1}^{N} |a_i| \int |b^{-d} \int_{x+bA_i} f(y)dy - b^{-d} \int_{x+bA_i} \mu_n(dy)|dx$$

$$< Mb^{-d} \int |\mu(x+bA_i) - \mu_n(x+bA_i)|dx .$$

Thus it suffices to consider

$$b^{-d} \int |\mu(x+bA) - \mu_n(x+bA)|dx,$$

where A is an arbitrary finite rectangle of $\mathbb{R}^d$.

Let A be as above. We can write

$$A = \prod_{i=1}^{d} [x_i, x_i + a_i).$$

Take $\epsilon > 0$ and choose $T > \max\{2/\min_i a_i, 4.(\Sigma_i a_i^{-1})\lambda(A)\}$ where $\lambda(A) = \prod_{i=1}^{d} a_i$, is the Lebesgue measure of A. Now let us consider the $\mathbb{R}^d$-partition whose elements are d-fold products of intervals of the form $[(i-1)b/T, ib/T)$, where i is an integer. Call this partition $\Psi$. Let us define

$$A^* = \prod_{i=1}^{d} [x_i + 1/T, x_i + a_i - 1/T], \quad \text{and}$$

$$C_x = x + bA - \bigcup_{\substack{B\in\Psi \\ B\subseteq x+bA}} B \subseteq x + b(A-A^*) = C_x^* .$$

We have

$$\int |\mu(x+bA) - \mu_n(x+bA)|dx$$

$$< \int \sum_{\substack{B\epsilon\Psi \\ B\subseteq x+bA}} |\mu(B) - \mu_n(B)|dx$$

$$+ \int [\mu(C_x^*) + \mu_n(C_x^*)]dx. \tag{1}$$

Now Lemma 4–A–2 above implies that the last term in the above inequality can be written as

$$2\lambda(b(A-A^*)) = 2b^d\lambda(A-A^*) = 2b^d(\prod_{i=1}^{d} a_i - \prod_{i=1}^{d} (a_i - \frac{2}{T}))$$

$$= 2b^d\lambda(A)(1 - \prod_{i=1}^{d} (1 - \frac{2}{Ta_i}))$$

$$< 4b^d\lambda(A) \sum_{i=1}^{d} \frac{a_i^{-1}}{T} < \epsilon b^d .$$

Now choose R so large that

$$\mu(S_R^c) < \epsilon/2\lambda(A) , \tag{2}$$

where $S_R$ is the sphere centered at the origin with radius R. We can bound the first term in (1) from above by

$$\sum_{\substack{B\epsilon\Psi \\ B\cap S_R\neq\Phi}} |\mu_n(B) - \mu(B)| \int_{B\subseteq x+bA} dx$$

$$+ \sum_{\substack{B\epsilon\Psi \\ B\cap S_R=\Phi}} |\mu_n(B) - \mu(B)| \int_{B\subseteq x+bA}$$

$$< \sum_{\substack{B\epsilon\Psi \\ B\cap S_R\neq\Phi}} |\mu_n(B) - \mu(B)|\lambda(A).b^d$$

$$+ [|\mu_n(S_R^c) - \mu(S_R^c)| + 2\mu(S_R^c)]\lambda(A)b^d, \tag{3}$$

where we have used the fact that

$$\int_{B\subseteq x+bA} dx \leq b^d\lambda(A),$$

which is easily showing as both A and B are d-fold products of finite intervals. Now as our processes are mixing, they are ergodic and thus $|\mu_n(S_{\mathfrak{F}}) - \mu(S_{\mathfrak{F}})| = o(1)$, a.s., thus the second term in (3) can be bounded from above by $b^d o(1) + \epsilon b^d$ a.s. On the other hand the collection of sets $B \in \Psi$ with $B \cap S_R \mp \Phi$ has at most $(2RT/b + 2)^d = O(b^{-d})$ elements. Now apply theorem 4.2.2 to conclude the proof. For instance, in order to show that $f_n$ is strongly $L_1$-consistent if (4.2.8), (4.2.9) and (4.2.10) we need only (by virtue of the majorizations obtained above) prove that

$$\sum_n P\{\sum_{K=1}^{N} |\mu_n(A_{nk}) - \mu(A_{nk})| > \epsilon\} < \infty,$$

for any sequence of $\mathbb{R}^d$-partitions $\{Q_n\}$ such that

$$Q_n = \{A_{n1},\ldots,A_{NN}\},$$
$$N = N(n) = O(b^{-d}),$$

which can be trivially accomplished by theorem 4.2.2. □

## 2. Proof of THEOREM 4.2.9

Let us define $g_n(.) = Ef_n(.)$. By Lemma 4-A-3 we have $\int |g_n-f| = o(1)$. Hence we need only to consider $\int |f_n-g_n|$. We can write

$$\int |f_n - g_n| = \sum_j |\mu_n(A_{nj}) - \mu(A_{nj})|.$$

Let M and C be a positive real constant and a $\mathbb{R}^d$-Borelian (respectively) to be chosen later. Let us denote by $H_n$ the set $\{j: \lambda(A_{nj} \cap C) > Mm/n\}$. We have

$$\sum_j |\mu_n(A_{nj}) - \mu(A_{nj})|$$

$$\leq \sum_{j \in H_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + \sum_{j \in H_n^c} |\mu_n(A_{nj}) - \mu(A_{nj})|$$

$$\leq \sum_{j \in H_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + |\mu_n(B_n) - \mu(B_n)| + 2\mu(B_n), \qquad (4)$$

where $B_n = \cup\{A_{nj};\ j \text{ not in } H_n\}$. Take $\epsilon > 0$. Now notice that $H_n$ has at most $1 + n\lambda(C).(Mm)$ elements (for each j in $H_n$ one has $\lambda(A_{nj} \cap C) > Mm/n$). Thus $\{B_n, A_{nj} : j \in H_n\}$ is a partition of $R^d$ with at most $2 + n\lambda(C)/(Mm)$ elements. Choose C and M so that

$$(2 + n\lambda(C)/Mm) \log 2 < \frac{1}{2} \cdot \frac{n\epsilon^2}{64m}, \qquad \text{and}$$

$$\mu(C^c) < \epsilon/2.$$

Now a simple application of THEOREM 4.2.2 yields

$$P\{\sum_{j \in H_n} |\mu_n(A_{nj}) - \mu(A_{nj})| + |\mu_n(B_n) - \mu(B_n)| > \epsilon\}$$

$$\leq m\theta(m) + 2 \exp\{- n\epsilon^2/(128m)\}, \qquad (5)$$

where $\theta$ is as in the statement of THEOREM 4.2.2. The remaining term in the right hand side of (4) can be written as

$$2\mu(B_n) = 2\mu(B_n \cap C) + 2\mu(B_n \cap C^c)$$

$$\leq o(1) + \epsilon,$$

because of our choice of C and $\lambda(B_n \cap C) = o(1)$ (notice that $\mu$ is absolutely continuous with respect to the Lebesgue measure $\lambda$). To finish the proof simply study each of the assumed relations between m and n separately. □

# CHAPTER 5

## 5.1 Introduction

In Chapter 4 we dealt with the issue of sufficient conditions on the sequence of bandwidths so as to ensure consistency in $L_1$ for kernel and histogram density estimates in weakly dependent samples. Our goal in this chapter is to study necessary conditions to be satisfied by the sequence of smoothing parameters provided that consistency in $L_1$ holds true for kernel density estimates under a particular type of weak dependence assumption. This topic fits in the global theme of the thesis insofar as a Poisson approximation theorem for series of dependent random variables will be our basic tool for proving the main theorem.

## 5.2 Necessary Conditions for $L_1$ Consistency

Under iid samples, the problem of consistency in $L_1$ was thoroughly studied in Devroye (1983) - kernel estimates - and Abou-Jaoude (1977) - histogram estimates. In each of these works, necessary and sufficient conditions for almost sure consistency in $L_1$ were obtained. As previously mentioned, the "sufficiency" part of Devroye's theorem has been generalized by a number of authors so as to accommodate dependent samples. On the other hand, the "necessity" part of his result does not seem to have been studied under dependence.

An important tool in the first half (sufficiency) of Devroye's result was an upper bound for probabilities of large deviations for multinomial random vectors. This bound was obtained via an elegant Poissonization trick. On the other hand, for the second half of his result, Devroye used a simple bound for the probability of a binomial random variable being equal to zero. It turns out that such a bound is much more difficult to obtain in the dependent case. One way of tackling this problem is via a Poisson approximation for sums of dependent Bernoulli random variables.

### 5.3 Poisson Approximations

The well known Poisson theorem states the weak convergence of binomial random variables $X_n \sim B(n,p_n)$ (say) to a Poisson random variable $X \sim P(\lambda)$ (say), provided that $np_n \to \lambda$, as n goes to infinity. Rates of convergence in the Poisson theorem are usually stated in terms of the variation distance between the distribution of $X_n$ and the corresponding Poisson distribution with the same mean (we recall that the variation distance between two probabilities $\mu$ and $\nu$ defined on a measurable space $(\Omega,F)$ is defined to be $d(\mu,\nu) = \sup\{|\mu(A)-\nu(A)|,\ A \in F\}$).

Hodges and LeCam in the early sixties obtained the following rate of convergence in the Poisson theorem.

<u>Theorem 5.1</u> (Hodges and LeCam (1960)) Let $A_1,\dots,A_n$ be independent events in a p-space $(\Omega,F,P)$. Let $N = \Sigma_1^n I_{A_i}$ be the number of such events which occur.

$$a = \sum_1^n P(A_i) \quad \text{and} \quad \epsilon = \sum_1^n P(A_i)^2 .$$

Suppose $\epsilon < 2$. Let $N^*$ be Poisson with parameter a. Define

$$K_1(q) = q^{-1}[-\log(1-q)], \quad \text{and}$$

$$K_2(q) = q^{-2}[-\log(1-q) - q],$$

where $0 < q < 1$. Let

$$\alpha = \frac{1}{2} K_1(\epsilon^{\frac{1}{2}})^2 + K_2(\epsilon^{\frac{1}{2}}) .$$

Then

$$d(N,N^*) \leq \alpha\epsilon \qquad \square$$

Let us just point out that $\alpha = \alpha(\epsilon)$ remains bounded as $\epsilon$ goes to zero.

Theorem 5.1 was subsequently generalized and improved by many authors. Freedman (1974) succeeded in allowing "dependences" to be taken into account on his generalization of Theorem 5.1. He obtained rates of convergence in the Poisson theorem similar to those in Theorem 5.1.

Difference rates of convergence for sums of dependent Bernoulli random variables were obtained in Chen (1975). Chen used the elegant Stein's methodology so as to get a number of useful rates in the Poisson theorem under uniform mixing assumptions. We refer the reader to Chen (1979) for a survey on applications of Stein's methodology in limit theorems for dependent data. See also Stein (1986) for a more informal description of Stein's method and its broad spectrum of applications. As far as the Poisson theorem is concerned, Stein's methodology was also used in Barbour and Hall (1984), Barbour (1987) and Barbour and Holst (1989).

Further approaches to obtaining rates of convergence in Poisson and related theorems can be learnt from Deheuvels and Pfeiffer (1986) and Borovkov (1988).

Our first result in this chapter depends heavily on a Poisson approximation theorem for sums of dependent Bernoulli random variables, due to David Friedman. By ingeniously embedding such a sum in a Poisson process, Friedman obtained.

<u>Theorem 5.2</u> (Friedman (1974))

Let $A_1,A_2,\ldots$ be dependent events in a p-space $(\Omega,F,P)$. Let $F_i$ be the field generated by $A_1,\ldots,A_i$, and let

$$p_i = P\{A_i \mid F_{i-1}\}$$

Let $\zeta$ be a stopping time relative to $\{F_i\}$. Let N be the number of $A_i$ which occur with $i \leq \zeta$:

$$N = \sum_{i=1}^{\zeta} 1_{A_i} .$$

Let $a \leq b$ be nonnegative real numbers. Let $\epsilon$, $\delta$ be nonnegative real numbers less than 1. Suppose

$$P\{a \leq p_1 + \ldots + p_\zeta \leq b \text{ and } p_1^2 + \ldots + p_\zeta^2 \leq \epsilon\} \geq 1-\alpha.$$

Let $N^*$ be Poisson with parameter a. Define $\alpha$ as in Theorem 5.1. Then

$$d(N,N^*) \leq \alpha\epsilon + (b-a) + 2\delta \qquad \square$$

We will not use Theorem 5.2 in its full generality. We will only need to consider the case when $\zeta = n$ and $\delta = 0$.

## 5.4 The Concept of Uniformly Mixing in-both-directions-of-time Sequences of Random Vectors

In Chapter 1 we defined the concept of uniform mixing processes and presented a couple of fundamental related results. Uniform mixing is the strongest form of weak dependence dealt with in this work. Indeed, as can be easily seen, every uniform mixing process is necessarily an absolutely regular or weak Bernoulli one (and thus a strong mixing process).

However, uniform mixing processes do not possess a "desirable" property which is shared by weak Bernoulli and strong mixing processes. Namely, if we reverse the direction of time in a uniform mixing process, then it is not guaranteed that the new process is still uniform mixing. To be more precise, let $\{X_n\}$ $n \in Z$ be a uniform mixing process with associated sequence of mixing weights given by $\{\varphi_k\}$, $k \geq 1$.

Let us define the process $\{\bar{X}_n\}$, $n \in Z$ requiring that $\bar{X}_n = X_{-n}$, for all $n \in Z$. As $\{X_n\}$ is a uniform mixing process, $\{\varphi_k\}$, $k \geq 1$ is a null sequence, where

$$\varphi_k = \sup_{n \in Z} \{|P(B|A) - P(B)|,\ A \in F_{-\infty}^{n},\ B \in F_{n+k}^{\infty}\},$$

with

$$F_a^b = \sigma\{X_k,\ a \leq k \leq b\}.$$

This fact, however, does not imply that the sequence of uniform mixing weights for the process $\{\bar{X}_n\}$ is also a null one. This can be easily seen in the following.

<u>Example</u>

Let $\{Z_n\}$, $n \geq 1$ be a sequence of iid random variables such that

$$P\{Z_1=1\} = \tfrac{1}{2} = P\{Z_1=0\}.$$

Let us define $Z_n = 1$ if $n \leq 0$ and $Q_n = \prod_{k \leq n} Z_k$. We will show that $\{Q_n\}$ is a uniform mixing process and that $\{\bar{Q}_n\} = \{Q_{-n}\}$ is not a uniform mixing process.

Let

$$\varphi_k = \sup \{|P(B|A) - P(B)|,\ A \in \sigma(\ldots, Q_{n-1}, Q_n),$$
$$B \in \geq (Q_{n+k}, Q_{n+k+1}, \ldots)\}$$

In order to show that $\{Q_n\}$ is uniform mixing we must prove that $\varphi_k = o(1)$. We will only consider a particular kind of event A and B as above. Extending our arguments (see below) so as to take A and B in their full generality is straightforward though a bit tedious.

Let $A = \{Q_n=a\}$, $B = \{Q_{n+k}=b\}$, where a,b are either 0 or 1. We have

$$|P\{B|A\} - P\{B\}| = \begin{cases} 1/2^{n+k}, & \text{if } a = b = 0 \\ 1/2^{n+k}, & \text{if } a = 0,\ b = 1 \\ 1/2^k - 1/2^{n+k}, & \text{if } a = 1,\ b = 0 \\ 1/2^k - 1/2^{n+k}, & \text{if } a = 1,\ b = 0 \end{cases}$$

Therefore $\varphi_k < 1/2^k = o(1)$.

Now, to prove that $\{\bar{Q}_n\}$ is not a uniform mixing process we only need to find events $A = A(n+k)$, $B = B(n)$, $k > 1$ and $n > 1$ such that

$$A \in \sigma(\ldots, \bar{Q}_{-(n+k+1)}, \bar{Q}_{-(n+k)})$$
$$B \in \sigma(\bar{Q}_{-n}, \bar{Q}_{-n+1}, \ldots)$$

and

$$\sup_k |P(B|A) - P(B)| > \tfrac{1}{2} .$$

Let us define

$$A = \{\bar{Q}_{-(n+k)} = 1\} = \{Q_{n+k} = 1\}, \quad \text{and}$$
$$B = \{\bar{Q}_{-n} = 1\} = \{Q_n = 1\} .$$

We have

$$|P(B|A) - P(B)| = |1 - \frac{1}{2^n}| ,$$

and thus $\{\bar{Q}_n\}$ is not a uniform mixing process. □

The need of a symmetric–under–time–reversibility weak dependence is especially crucial when one is dealing with U–statistics. We refer the reader to Denker and Keller (1983) for a discussion on such technical subtleties.

In our case, the above mentioned symmetry is essential insofar as we need that the "head" and "tail" $\sigma$–fields be jointly weakly dependent of "middle" $\sigma$–fields. To be more precise, we will need that the $\sigma$–field generated by $F^m_{-\infty}$ and $F^{\infty}_{m+p+q+r}$ be weakly dependent of $F^{m+p+q}_{m+p}$. The degree (or type) to which such a weak dependence holds could be imposed as an assumption. In our particular problem, the assumption of our process under study being uniform mixing in both directions of time will provide us a suitable weak dependence between the above mentioned $\sigma$-algebras.

<u>Definition</u>: Let $\{X_n\}$, $n \in Z$ be a sequence of random vectors defined on a common p-space and taking values in $R^d$. We say that $\{X_n\}$ is uniformly mixing in both directions of time if $\{\theta_k\}$, $k \geq 1$ is a null sequence where

$$\theta_k = \max\{\overset{1}{\varphi_k}, \overset{2}{\varphi_k}\}, \quad \text{and}$$

$$\overset{1}{\varphi_k} = \sup_{n \in Z}\{|P(B|A) - P(B)|,\ A \in F^n_{-\infty},\ B \in F^{\infty}_{n+k}\},$$

$$\overset{2}{\varphi_k} = \sup_{n \in Z}\{|P(A|B) - P(A)|,\ A \in F^n_{-\infty},\ B \in F^{\infty}_{n+k}\}. \qquad \square$$

When dealing with processes for which the time-parameter runs over only the nonnegative integers (i.e. processes like $\{X_n\}$, $n \geq 1$) we implicitly assume that $X_{-n} = 0$, for $n \geq 0$.

## 5.5 <u>Main Result</u>

We are to introduce some notation for the proof of Theorem 5.3 below. We will use the classical partition of the sample into small and large blocks. Namely, we are going to split the set of the first n integers into alternate blocks of length $l_n$ (large blocks) and $s_n$ (small blocks). Let

$$\Lambda_{ni} = \{(i-1)(l_n+s_n) + 1, \ldots, (i-1)(l_n+s_n) + l_n\}, \quad 1 \leq i \leq r_n,$$

where $r_n$ is the largest integer j such that $(j-1)(l_n+s_n) < n$. Also let

$$\psi_{ni} = \{(i-1)(l_n+s_n) + l_n+1, \ldots, i(l_n+s_n)\}, \quad i \leq i \leq r_n, \quad \text{and}$$

$$\psi_{nr_n} = \{(r_n-1)(l_n+s_n) + l_n+1, \ldots, n\}.$$

For $1 \leq k \leq n$ and belonging to some $\Lambda_n$; let $\gamma(k)$ be the order of the large block to which k belongs (that is $\gamma(k) = i$) iff $k \in \Lambda_{ni}$). For $1 \leq k \leq n$

belonging to some $\psi_{ni}$, let $\delta(k) = i$ iff $k \in \psi_{ni}$. We will also need the following convention: for each large block $\Lambda_{ni}$ let $\Gamma_{ni}$ be the innermost portion of $\Lambda_{ni}$ far apart from the boundaries by at least $s_n$ "numbers". Namely let

$$\Gamma_{ni} = \Lambda_{ni} - (\{(i-1)(l_n+s_n) + 1,\ldots,(i-1)(l_n+s_n) + s_n\}$$
$$\cup \{(i-1)(l_n+s_n) + l_n-s_n + 1,\ldots,(i-1)(l_n+s_n) + l_n\}).$$

Finally, let

$$\Lambda = \bigcup_{i=1}^{r_n} \Lambda_{ni}, \quad \Psi = \bigcup_{i=1}^{r_n} \Psi_{ni} \quad \text{and} \quad \Gamma = \bigcup_{i=1}^{r_n} \Gamma_{ni},$$

where we have dropped the subscript n from $\Lambda$, $\Psi$ and $\Gamma$ for notational convenience.

We can now state and prove our

<u>Theorem 5.3</u>

Let $\{X_n\}$ be a stationary process taking values in $R^d$ and let us assume that $X_1$ admits of a density f (say). Let $f_n(.)$ be the kernel density estimate for f based on $\{X_i\}$, $1 \leq i \leq n$, that is

$$f_n(x) = \frac{1}{nb^d} \sum_{k=1}^{n} W\left(\frac{x-X_k}{b}\right),$$

where $W : R^d \to R$ is a density in $R^d$. Let us also assume that $\{X_n\}$ is uniform mixing in both directions of time and that the corresponding sequence of mixing weights satisfy $\Sigma\theta_n^q < \infty$ for some $q > 1$. Then, if $J_n \to 0$, in probability, we necessarily have $b + 1/(nb^d) \to 0$, where

$$J_n = \int |f_n-f|.$$

<u>Proof</u>: The proof consists of two parts. For the first part we must show that $J_n \to 0$, in probability, for some density f implies that $b(n) \to 0$, as $n \to \infty$. The proof of this result can be found in Devroye et al. (1985), and therefore will be

omitted. In the second part, we must show that $J_n \to 0$, in probability, for some density f implies that as $n \to \infty$ we have $n[b(n)]^d \to \infty$.

The proof of the latter part will be carried out by showing that no subsequence of $\{n[b(n)]^d\}$ can converge to a finite limit. Suppose we have such a subsequence having finite limit, say, $\beta$. Since $J_n \leq 2$ and $J_n \to 0$, in probability, we have

$$E(J_n) = E[\int |f_n(x) - f(x)|dx] = o(1).$$

Now let $\{X_n^*\}$, $n \geq 1$, be an independent copy of the process $\{X_n\}$, $n \geq 1$. The superscript "*" on $f_n$ below is to be understood as the estimated density by the former process. We have

$$\int |f_n(x) - f(x)|dx + \int |f_n^*(x) - f(x)|dx \geq \int |f_n(x) - f_n^*(x)|dx,$$

therefore $J_n \to 0$, in probability and $J_n$ being uniformly bounded imply

$$\int |f_n^*(x) - f(x)|dx = o(1),$$

in probability, and thus

$$\int E|f_n(x) - f_n^*(x)|dx = E\int |f_n(x) - f_n^*(x)|dx = o(1).$$

Let R be the positive number such that:

$$\int_{B_R(a)} W(z)dz > 0, \tag{5.1}$$

where $B_r(a)$ stands for the open ball in $R^d$ centered at a and with radius r. Let $A = A(x)$ be the event that

$$X_k - x \notin B_{bR}(a), \quad k \in \Lambda \quad \text{and} \quad X_k^* - x \notin B_{bR}(a), \quad k \in \psi.$$

Where we have chosen $l_n$ and $s_n$ in the following manner: take $\alpha$ so that $(1-\alpha)q < \alpha < 1$, choose $l_n = [[n^\alpha]]$, $s_n = [[n^{(1-\alpha)q}]]$ and, as a consequence, $r_n = 0(n^{1-\alpha})$. We have:

$$E\int |f_n^* - f_n| = \int E|(f_n^* - f_n)| \geq \int E[(f_n^* - f_n)I_A].$$

$$= n^{-1}\Sigma' \int E[(W_b(x-X_k^*) - W_b(x-X_k))I_A]dx$$

$$+ n^{-1}\Sigma'' \int EW_b(x-X_k^*)I_A\, dx$$

$$- n^{-1}\Sigma'' \int EW_b(x-X_k)I_A\, dx$$

$$= (I) + (II) + (III), \tag{5.2}$$

where $\Sigma'$ ($\Sigma''$, respectively) stands for the summation over all k in $\Gamma$ ($\{1,2,\ldots,n\} - \Gamma$, respectively) and $W_b(z) = (1/b^d)W(z/b)$. We will need some further notation. let

$$D(x) = \{X_j \notin B_{bR}(x),\ j \in \Lambda\}, \tag{5.3}$$

$$D_{-k}(x) = \{X_j \notin B_{bR}(x),\ j \in \Lambda-\Lambda_{n\gamma(k)}\},\ k \in \Lambda, \tag{5.4}$$

$$D^*(x) = \{X_j^* \notin B_{bR}(x),\ j \in \Psi\}, \tag{5.5}$$

$$D_{-k}^*(x) = \{X_j^* \notin B_{bR}(x),\ j \in \Psi-\Psi_{n\delta(k)}\}. \tag{5.6}$$

For k $\in \Gamma$ we have, by the use of Ibragimov's inequality and by Lemma 5.A.2 (see Appendix 5.A):

$$E\ W_b(x-X_k^*)I_A = E[W_b(x-X_k^*)I_{D^*}]P(D)$$

$$\geq E[W_b(x-X_k^*)]P(D)P(D^*) - 6\theta_{s_n}EW_b(x-X_k^*).P(D)\,; \tag{5.7}$$

Also,

$$E\ W_b(x-X_k)I_A = E[W_b(x-X_k)I_D]P(D^*)$$

$$\leq E[W_b(x-X_k)I\{X_k \notin B_{bR}(x)\}I_{D_{-k}}]P(D^*)$$

$$\leq E[W_b(x-X_k)I\{X_k \notin B_{bR}(x)\}]P(D_{-k})P(D^*)$$

$$+ 6E[W_b(x-X_k)I\{X_k \notin B_{bR}(x)\}]\theta_{l_n}P(D^*)$$

$$= E[W_b(x-X_k)I\{X_k \notin B_{bR}(x)\}]P(D)P(D^*)$$

$$+ 6E[W_b(x-X_k)I\{X_k \notin B_{bR}(x)\}]P(D^*)\theta_{l_n}$$

$$+ E[W_b(x-X_n)I\{X_k \notin B_{bR}(x)\}]P(D^*)[P(D_{-k}) - P(D)]. \tag{5.8}$$

From (5.7) and (5.8) we see that

$$(I) \geq a_n \int E[(W_b(x-X_k^*)-W_b(x-X_k))I\{X_k \in B_{bR}(x0\}]P(D^*)P(D)dx + a_nT_n, \qquad (5.9)$$

where $T_n$ is the integral over $R^d$ of the sum of the right hand sides of (5.7) and (5.8) minus

$$\int E[(W_b(x-X_k^*) - W_b(x-X_k))I\{X_k \notin B_{bR}(x)\}]P(D^*)P(D)dx,$$

and

$$a_n = n^{-1}(l_n-2s_n)r_n.$$

Now the integral on the right hand side of (5.9) can be written as

$$a_n \int f(y) \int_{B_R(0)} W(z)P(D^{*+})P(D^+)dzdy, \qquad (5.10)$$

where $D^{*+}$ ($D^+$ respectively) stands for $\{X_j^* \notin B_{bR}(y-zb),\ j \in \Psi\}$ ($\{X_j \notin B_{bR}(y-zb),\ j \in \Lambda\}$ respectively). By Lebesgue's density theorem we have

$$\lim nP\{X_j \in B_{bR}(y-zb)\} = \beta R^d v f(y), \qquad (5.11)$$

where V is the volume of the unit ball in $R^d$. Now by means of Lemma 5.A.1 (see Appendix 5.A.), Fatou's lemma and (5.11) we obtain

$$\underline{\lim}\ a_n \int E[(W_b(x-X_k^*) - W_b(x-X_k))I\{X_k \notin B_{bR}(x)\}]P(D^*)P(D)dx$$
$$\geq \int f(y) \int_{B_R(0)} W(z)\exp(-\beta R^d v f(y))dxdy > 0. \qquad (5.12)$$

Also, by a further application of Lebesgue's density theorem it is easily seen that

$$\int E[W_b(x-X_k)I\{X_k \notin B_{bR}(x)\}]P(D^*)[P(D_{-k})-P(D)]dx = o(1).$$

Indeed, the above integrand is majorized by the integrable function $h(x) = E[W_b(x-X_k)]$. On the other hand

$$P(D_{-k}) - P(D) \leq l_nP\{X_1 \in B_{bR}(x)\} = o(1),$$

by Lebeque's density theorem, our assumption that $nb^d = 0(1)$ and $l_n = O(n^{\alpha})$.

The other components of $T_n$ are also readily seen to be o(1) because they are dominated by

$$\theta_{s_n} \int E[W_b(x-X_1)]dx \leq \theta_{s_n} = o(1),$$

where the above inequality was obtained by an application of Young's inequality

$$\int f*g \leq \int f \int g,$$

where * stands for the convolution operator. Thus

$$\lim\inf (I) \geq \int f(y)\exp(-\beta R^d v f(y))dy \int_{B_R(0)} W(z)dz > 0.$$

By a further application of Young's inequality we obtain

$$(III) \leq 3s_n N^{-1} \int EW_b(x-X_k)dx \leq 3s_n N^{-1} = o(1).$$

Now from (5.12), $T_n = o(1)$, (III) = 0(1) and (II) $\geq 0$ we obtain

$$\lim\inf E \int |f_n - f_n^*| > 0,$$

which cannot hold. Therefore no subsequence of $\{nb^d\}$ can have a finite limit and thus $\lim nb^d = \infty$. □

# APPENDIX 5.A

<u>Lemma 5.A.1</u> Let $\{X_n\}$ be a uniform mixing process such that the sequence of mixing weights satisfy

$$\Sigma(\varphi_k)^q < \infty,$$

for some $q > 1$. Consider the following array of events

$$\begin{array}{llll} A_{1b(1)} & & & \\ A_{1b(2)} & A_{2b(2)} & & \\ \cdots & & & \\ A_{1b(n)} & A_{2b(n)} & \cdots & A_{nb(n)} \end{array}$$

and suppose that

(i) For each n and k ($1 \leq k \leq n$) we have $A_{kb(n)} \in \sigma(X_k)$,

(ii) For each n and k we have $P\{A_{1b(n)}\} = P\{A_{kb(n)}\}$,

(iii) $\lim_n nP\{A_{1b(n)}\} = \beta < \infty$.

Let $S_i$, $L_i$, $M_i$ be (for $1 \leq i \leq r_n$)

$$\bigcup_{j\in\Psi_{ni}} A_{jb(n)}, \quad \bigcup_{j\in\Lambda_{ni}} A_{jb(n)}, \quad \bigcup_{j\in\Gamma_{ni}} A_{jb(n)},$$

respectively. Then

(A) $\liminf_n P\{\Sigma' I(S_i) = 0\} = 1$

(B) $\liminf_n P\{\Sigma' I(L_i) = 0\} \geq \exp(-\beta)$

(C) $\liminf_n P\{\Sigma' I(M_i) = 0\} \geq \exp(-\beta)$,

where the summations above are for i in $\{1,\ldots,r_n\}$.

<u>Proof</u>: As the technique of proof is virtually the same for cases (A), (B) and (C) we will only show (B). We have:

$$P(L_i) \leq l_n P\{A_{1b(n)}\} \qquad (5.A.1)$$

and

$$|P\{L_j | L_1 \ldots L_{j-1}\} - P\{L_j\}| \leq \varphi_{s_n}. \tag{5.A.2}$$

From (5.A.1) and (5.A.2) we obtain

$$r_n P\{L_1\} - r_n \varphi_{s_n} \leq \sum_{j=1}^{r_n} P\{L_j | L_1 \ldots L_{j-1}\}$$

$$\leq r_n P\{L_1\} + r_n \varphi_{s_n}. \tag{5.A.3}$$

Also

$$\sum_{j=1}^{r_n} P^2\{L_j | L_1, \ldots, L_{j-1}\} \leq 2r_n (P^2\{L_1\} + \varphi_{s_n}^2)$$

$$\leq 2r_n l_n^2 P^2\{A_{1b(n)}\} + 2r_n \varphi_{s_n}^2. \tag{5.A.4}$$

Let us just recall that $l_n = 0(n^\theta)$, $s_n = 0(n^{(1-\theta)q})$ and $r_n = 0(n^{1-\theta})$, where $\alpha$ is such that $(1-\theta)q < \theta < 1$. Now, as $\Sigma \varphi_k^{\frac{1}{2}} < \infty$ and $\lim_n nP\{A_{1b(n)}\} = \beta$, we have that the right hand side of (5.A.4) goes to zero as n goes to infinity. Assumptions of Theorem 5.2 are in force and we conclude that

$$|P\{I_{L_1} + \ldots + I_{L_r} = 0\} - P\{Z = 0\}|$$

$$\leq \alpha(2r_n l_n^2 P^2\{A_{1b(n)}\} + r_n \varphi_{s_n}^2) + 2r_n \varphi_{s_n} = o(1),$$

where $Z \sim$ Poisson $(r_n P\{L_i\} - r_n \varphi_{s_n})$. Whence

$$P\{I_{L_1} + \ldots + I_{L_{r_n}} = 0\} = \exp(-r_n P\{L_1\} + o(1)) + o(1).$$

Therefore

$$\liminf P\{I_{L_1} + \ldots + I_{L_{r_n}} = 0\}$$

$$= \exp(-\limsup r_n P\{L_1\})$$

$$\geq \exp(-\limsup r_n l_n P\{A_{1b(n)}\}) = \exp(-\beta),$$

where we have used (5.A.1) and assumption (iii). □

<u>Lemma 5.A.2</u> Let $A$, $B$ and $C$ be sub $\sigma$-algebras defined in a p-space $(\Omega, F, P)$.

Let us define (for any two $\sigma$-algebras $M$ and $H$) $M \vee H$ as the $\sigma$-algebra generated by $M$ and $H$. Let us also put

$$\varphi(M,H) = \sup\{|P(H|M) - P(H), H \in H, M \in M\}.$$

Then

$$\varphi(B, A \vee C) \leq 3\max\{\varphi(B \vee C, A), \varphi(B,C), \varphi(A,B)\}.$$

<u>Proof</u>: Let us denote by $D$ the algebra generated by $A$ and $C$. It is easy to see that any $D$-element can be written as

$$D = \bigcup_{i=1}^{n} A_i \cap C_i ,$$

for some sets $A_i \in A$, $C_i \in C$ and $1 \leq n < \infty$, where $(A_i \cap C_i) \cap (A_j \cap C_j) = \varphi$, if $i \mp j$. Let B be a $B$-element. We have

$$\begin{aligned}
P\{A_i \cap C_i \cap B\} &\leq P(A_i)P(B \cap C_i) + \varphi(B \vee C, A)P(B \cap C_i) \\
&\leq P(A_i)[P(B)P(C_i) + P(B).\varphi(B,C)] + \varphi(B \vee C, A)P(B \cap C_i) \\
&\leq P(B)[P(A_i, C_i) + P(A_i)\varphi(A,C)] + P(A_i)P(B)\varphi(B,C) \\
&\quad + \varphi(B \vee C, A)P(B \cap C_i) .
\end{aligned}$$

Summarising over i we obtain

$$P\{\bigcup_{i=1}^{n} A_i \cap C_i \cap B\} \leq P(B)P\{\bigcup_{i=1}^{n} A_i \cap C_i\}$$

$$+ P(B).3\max\{\varphi(B \vee C, A), \varphi(B,C), \varphi(A,B)\}.$$

Similarly we obtain

$$P(B)(\{\bigcup_{i=1}^{n} A_i \cap C_i\} \leq P\{\bigcup_{i=1}^{n} A_i \cap C_i \cap B\}$$

$$+ P(B).3\max\{\varphi(B \vee C, A), \varphi(B,C), \varphi(A,B)\}.$$

Whence

$$|P\{\bigcup_{i=1}^{n} A_i \cap C_i|B\} - P\{\bigcup_{i=1}^{n} A_i \cap C_i\}|$$

$$\leq 3\max\{\varphi(B\vee C,A),\varphi(B,C),\varphi(A,B)\}.$$

Now the class $\mathcal{G}$ of sets G such that

$$|P\{G|B\} - P\{G\}| \leq 3\max\{\varphi(B\vee C,A),\varphi(B,C),\varphi(A,B)\},$$

for all B in $\mathcal{B}$ is clearly a monotone class (see e.g. Dudley (1989)) containing $\mathcal{D}$. Therefore

$$|P\{G|B\} - P\{G\}| \leq 3\max\{\varphi(B\vee C,A),\varphi(B,C),\varphi(A,B)\}$$

holds for every G in $\sigma(\mathcal{D})$ and B $\in$ $\mathcal{B}$. □

## REFERENCES


Abou-Jaoude, S. (1976a), "Sur une condition necessaire et suffisante de $L_1$-convergence presque complete de l'estimateur de la partition fixe pour une densite", C. R. Acad. Sci. Paris Ser. A, 283, 1107-10.

Abou-Jaoude, S. (1976b),"Sur la convergence $L_1$ et $L_\infty$ de l'estimateur de la partition aleatoire pour une densite", Ann. Inst. Henry Poincare, 12, 299-317.

Abou-Jaoude, S. (1976c),"Conditions necessaires et suffisantes de convergence $L_1$ en probabilite de l'histogramme pour une densite" Ann. Inst. Henry Poincare, 12, 213-231.

Abou-Jaoude, S. (1977), "La convergence $L_1$ et $L_\infty$ de certains estimateurs d'une densite de probabilite", These de Doctorat d'Etat, Universite Paris VI, Paris.

Abramovitz, M. and Stegun, I. (1965)."Handbook of Mathematical Functions", Dover.

Adenstedt, E. (1974)."On large-sample estimation for the mean of a stationary random sequence". Ann. Math. Stat., 2, 1095-1107.

Akonom, J. and Gourieroux, C. (1988)."A functional limit theorem for fractional processes". CEPREMAP, D.P. 8801.

Barbour, A.D. (1987), "Asymptotic expansions in the Poisson limit theorem", Ann. Probab., 14, 748-766.

Barbour, A.D. and Hall, P. (1984), "On the rate of Poisson convergence", Math. Proc. Camb. Phil. Soc., 95, 473-480.

Barbour, A.D. and Holst, L. (1989), "Some applications of Stein-Chen method for proving Poisson convergence", Adv. App. Probab., 21, 74-90.

Bean, S. J. and Tsokos, C. P.(1980), "Developments in nonparametric density estimation", Int. Stat. Rev., 48, 267-87.

Berbee, H.C.P. (1979). Random walks with stationary increments and renewal theory. Math. Centre Tracts, 112, Mathematisch Centrum, Amsterdam.

Berkes, I. and Philipp, W. (1979)."Approximation theorems for independent and weakly dependent random vectors". Ann. Probab., 7, 29-54.

Billingsley, P. (1968). Convergence of probability measures. John Wiley: New York.

Billingsley, P. (1979). Probability and measure. John Wiley: New York.

Borovkov, A.A. (1988), "Refinement of Poisson approximation", Theo. Probab. Appl., 33, 343-347.

Box, G.E.P. and Jenkins, M.J. (1970). Time series analysis: Forecasting and control. Holden Day: San Francisco.

Bradley, R.C. (1983)."Approximation theorems for strongly mixing random variables". Mich. Math. J. , 30, 69-81.

Breiman, L. (1967). "On the tail behaviour of sums of independent random variables". Z. Warsch. verw. Gebiete, 46, 1-11.

Cacoulos, T. (1966)."Estimation of a multivariate density". Ann. Inst. Stat. Math., 18, 179-189.

Carbon, M. (1983)."Inegalite de Bernstein pour les processes fortement

melangeants non necessairement stationnaires". Comptes Rendus de l'Academie des Sciences de Paris, serie 1, 297, 303-306.

Chanda, K.C. (1983)."Density estimation for linear processes". Ann. Inst. Stat. Math., 35, 439-446.

Chen, L.H.Y. (1975), "Poisson approximation for dependent trials", Ann. Probab., 3, 534-545.

Chen, L.H.Y. (1979), "Stein's method in limit theorems for dependent random variables". Proceedings of the first Franco-Southeast Asian Mathematical Conference (Nanyang University, Singapore). Southeast Asian Bull. Math., Special issue A, 36-50.

Cheng, B. and Robinson, P.M. (1990)."Density estimation in strongly dependent non-linear time series". Manuscript.

Chow, Y.S. and Teicher, H. (1988). Probability theory (2ns ed.). Springer Verlag: New York.

Chung, K.L. and Williams, R.J. (1983). Introduction to stochastic integration. Birkhauser: Boston.

Csorgo, M. and Revesz, P. (1981). Strong approximations in probability and statistics. Academic Press, New York.

Dabrowski, A. (1983)."A note on a theorem of Berkes and Philipp for dependent sequencies". Stat. Probab. Lett., 1, 53-55.

Davidov, Yu.A. (1968)."Convergence of distributions generated by stationary stochastic processes". Th. Probab. Appl., 13, 691-696.

Davidov, Yu.A. (1970)."The invariance principle for stationary processes". Theory Probab. Appl., 15, 487-498.

Deheuvels, P. and Pfeiffer, D. (1986), Ann. Probab., 14, 665-678.

Dehling, H. (1983)."Limit theorems for sums of weakly dependent B-space valued random variables". Z. Wahrsch. verw. Gebiete, 63, 393-432.

Dehling, H. and Philipp, W. (1982)."Almost sure invariance principles for weakly dependent vector valued random variables". Ann. Probab., 10, 689-701.

Denker, M. and Keller, G. (1983), "On U-statistics and V. Mises' statistics for weakly dependent processes", Z. Wahrsch. Verw. Gebeite, 64, 505-522.

Devroye, L. (1983), "The equivalence of weak, strong and complete convergence in $L_1$ for kernel density estimates", Ann. Stat., 11, pp.896–904.

Devroye, L. (1987), A course in density estimation. Birkhauser, Boston.

Devroye, L. and Gyorfi, L. (1985), Nonparametric Density Estimation, John Wiley, New York.

Devroye, L. and Gyorfi, L. (1985)."Distribution-free exponential bound for the $L_1$ error of partitioning estimates of a regression function". In: Probability and Statistical Decision Theory, Proceedings of the Fourth Pannonian Symposium on Mathematical Statistics. F. Konechy (ed.), Reidel, Dordrecht, 67–76.

Devroye, L. and Wagner, T.J. (1980)."On the $L_1$ convergence of Kernel estimators of regression functions with applications in discrimination". Z. Wahrsch. verw. Gebiete, 51, 15–25.

Devroye, L. and Wagner, T.J. (1980a)."Distribution-free consistency results in nonparametric discrimination and regression function estimation". Ann. Stat., 8, 231–239.

Doukhan, P. and Leon, M. (1980)."Estimation dans le processus $x_{n+1}=f(x_n)+\epsilon_n$". C.R. Acad. Sci., Paris, 297, A, 61–64.

Doukhan, P., Leon, J. and Portal, F. (1984)."Vitesse de convergence dans le theoreme central du limite pour des variables aleatoires a valeurs dans an espace de Hilbert"; Comptes Rendus de l'Academie des Sciences de Paris, 289,305–308.

Dudley, R.M. (1968)."Distances of probability measures and random variables". Ann. Math. Stat., 39, 1563–1572.

Dudley, R.M. (1989). Real analysis and probability. Wadsworth & Brooks.Cole: Belmont.

Eberlein, E. (1986)."On strong invariance principles under dependence assumptions". Ann. Probab., 14, 260–270.

Einmahl, U. (1987)."Strong invariance principles for partial sums of independent random vectors". Ann. Probab., 15, 1419–1440.

Einmahl, U. (1987)."A useful estimate in the multidimensional invariance principle". Probab. Th. Rel. Fields, 70, 81–101.

Einmahl, U. (1989)."Extensions of results of Komlos, Major and Tusnady to the multivariate case". J. Mult. Anal., 28, 20–48.

Engle, R. and Granger, C.W. (1987)."Co-integration and error correction: representation, estimation and testing". Econometrica, 55, 251–276.

Erdos, P. and Kac, M. (1946)."On certain limit theorems of the theory of probability". Bull. Amer. Math. Soc., 52, 292–302.

Fox, R. and Taqqu, M. (1986)."Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series". Ann. Stat., 14, 517–532.

Freedman, D. (1974), "The Poisson approximation for dependent events", Ann. Probab., 2, 256-269.

Fryer, M. J.(1977). "A review of some nonparametric methods of density estimation", J. Inst. Math. Appl., 20, 335-54.

Geweke, J. and Porter-Hudak, S. (1982)."The estimation and application of long memory time series models". J. Time Series Anal., 4, 221-238.

Gorodetskii, V.V. (1977a)."On convergence to semi-stable Gaussian processes". Theory Probab. Appl., 22, 498-508.

Gorodetskii, V.V. (1977b)."On the strong mixing property for linear sequences". Theory Probab. Appl., 22, 411-413.

Granger, C.W. (1978)."New classes of time series models". Statistician, 27, 237-253.

Granger, C.W. (1980)."Long memory relationships and the aggregation of dynamic models". J. Econometrics, 14, 227-238.

Granger, C.W.J. and Andersen, A.P. (1978). An introduction to bilinear time series models. Vandenhoeck & Ruprecht: Gottingen.

Granger, C.W. and Hoyeux, R. (1980)."An introduction to long memory time series models and fractional differencing". J. Time Series Anal., 1, 15-29.

Gyorfi, L. (1987)."Density estimation from dependent sample". In Statistical Data Analysis based on the $L_1$-Norm and Related Topics (Y. Dodge ed.) North-Holland, 393-402.

Gyorfi, L. (1991)."Universal consistencies of a regression estimate for unbounded regression functions".In G. Roussas (ed.), Nonparametric Functional Estimation and Related Topics, Kluwer Academic Press.

Gyorfi, L., Hardle, W., Sarda, P. and Vieu, P. (1989). Nonparametric Curve Estimation from time series. Lecture Notes in Statistics, vol.60, Springer Verlag: New York.

Hagan, V. and Ozaki, T. (1981)."Modelling non-linear vibrations using an amplitude-dependent autoregressive time series model". Biometrika, 68, 189-196.

Haggan, V., Heravi, S.M. and Priestley, M.B. (1984)."A study of the application of state-dependent models in non-linear time series analysis". J. Time Series Anal., 5, 69-102.

Hall, P. and Heyde, C.C. (1980). Martingale limit theory and its applications. Academic Press: New York.

Heyde, C.C. and Scott, D.J. (1973)."Invariance principles for the law of iterated logarithm for martingales and processes with stationary increments". Ann. Probab., 1, 428-436.

Hodges, J.L. and LeCam, L. (1960), "The Poisson approximation to the binomial distribution", Ann. Math. Statist., 31, 737–740.

Hoeffding, W. (1963), "Probability inequalities for the sum of bounded random variables", JASA, 58, 13–30.

Hosking, J. (1981)."Fractional differencing", Biometrika, 68, 165–176.

Ibragimov, I.A. and Linnik, Yu.V. (1970). Independent and stationary sequences of random variables. Wolters-Nordhoff: Groningen.

Komlos, J., Major, P. and Tusnady, G. (1976)."An approximation of partial sums of independent rv's, and the sample df. II". Z. Wahrsch. verw. Gebieht, 34, 33–58.

Kuelbs, J. and Philipp, W. (1980)."Almost sure invariance principles for partial sums of mixing B-valued random variables". Ann. Probab., 8, 1003-1036.

Li, W. and McLeod, A. (1986)."Fractional time series modelling". Biometrika, 73, 217–221.

Loftsgaarden, D.O. and Quesenberry, C.P. (1965)."A nonparametric estimate of a multivariate density function". Ann. Math. Stat., 36, 1049–1051.

Mack, D.S. and Rosenblatt, M. (1979)."Multivariate K–nearest neighbour density estimates". J. Mult. Anal., 9, 1–15.

Major, P. (1976)."The approximation of partial sums of independent r.v.'s." Z. Wahrsch. verw. Gebiete, 35, 213–220.

Major, P. (1976)."Approximation of partial sums of iid r.v.'s when the summands have only two moments". Z. Wahrsch. verw. Gebiete., 35, 221–229.

Major, P. (1979)."An improvement of Strassen's invariance principle". Ann. Probab., 7, 55–61.

Mandelbrot, B. and Van Ness, J. (1968)."Fractional Brownian motions, fractional noises and applications". S.I.A.M. Review, 10, 422–437.

Moore, D.S. and Yackel, J.W. (1977a)."Consistency properties of nearest neighbour density function estimates". Ann. Stat., 5, 143–154.

Moore, D.S. and Yackel, J.W. (1977b)."Large sample properties of nearest neighbour density function estimates"; in "Statistical Decision Theory and Related Topics" (S. Gupta and J. Yackel eds.), vol.II, Academic Press, New York.

Parzen, E. (1962)."On estimation of a probability density function and mode". Ann. Math. Stat., 31, 1065–1079.

Petrov, V.V. (1975). Sums of independent random variables. Springer–Verlag: New York.

Phan, D.T. and Tran, L.T. (1985)."Some mixing properties of time series models". Stoch. Proc. Appl., 19, 297–303.

Philipp, W. and Stout, W.F. (1975)."Almost sure invariance principles for partial sums of weakly dependent random variables". Mem. Amer. Math. Soc., no.161. American Mathematical Society: Providence, Rhode Island.

Pollard, D. (1984). Convergence of stochastic processes. Springer-Verlag: New York.

Prakasa Rao, B.L.S. (1983). Nonparametric functional estimation. Academic Press: New York.

Prakasa Rao, B.L.S. (1987). Asymptotic theory of statistical inference. John Wiley: New York.

Priestley, M.B. (1980)."State-dependent models: a general approach to non-linear time series analysis", J. Time Series Anal., 1, 47-71.

Reznik, M.K. (1968)."The law of iterated logarithm for some classes of stationary processes". Theor. Probab. Appl., 8, 606-621.

Robinson, P.M. (1983)."Nonparametric estimators for time series". J. Time Series Anal., 4, 185-207.

Robinson, P.M. (1986)."Nonparametric methods in specification". Economic Journal, 96, 134-144.

Robinson, P.M. (1988). Nonparametric function estimation for long memory time series. Manuscript.

Robinson, P.M. (1990). Time series with strong dependence. Manuscript.

Rosenblatt, M. (1956)."Remarks on some nonparametric estimates of a density function". Ann. Math. Stat., 27, 832-837.

Rosenblatt, M. (1976)."Fractional integrals of stationary processes and the central limit theorem". J. Appl. Probab., 13, 723-732.

Roussas, G.G. (1988)."Nonparametric estimation in mixing sequences of random variables". J. Stat. Plann. Inference, 18, 135-149.

Roussas, G.G. (1988)."A moment inequality of $S(n,k_n)$ for triangular arrays of random variables under mixing conditions, with applications". In Statistical Theory and Data Analysis II (k. Matusite ed.). North Holland.

Roussas, G.G. and Ioannides, D. (1987)."Moment inequalities for mixing sequences of random variables". Stoch. Anal. Appl., 5, 61-120.

Roussas, G.G. and Ioannides, D. (1988)."Probability bounds for sums in triangular arrays of random variables under mixing conditions". In Statistical Theory and Data Analysis II (K. Matusita (ed.)) North-Holland Press.

Sarda, P. (1991)."Estimating smooth distribution functions". In G. Roussas (ed.), Nonparametric Functional Estimation and Related Topics, Kluwer Academic Press.

Sawyer, S. (1967)."A uniform rate of convergence for the maximum absolute value of partial sums in probability". Comm. Pure Appl. Math., 20, 647-658.

Silverman, B.W. (1986). Density estimation for statistics and data analysis. Chapman and Hall, London.

Skorohod, A.V. (1965). Studies in the theory of random processes. Addison-Wesley, Reading: Massachusetts.

Stone, C.J. (1977)."Consistent nonparametric regression". Ann. Stat., 5, 595-645.

Strassen, V. (1964)."An invariance principle for the law of iterated logarithm". Z. Wahrsch. verw. Gebiete., 3, 211-226.

Strassen, V. (1967)."Almost sure behaviour of sums of independent random variables and martingales". Proc. 5th Berkeley Symposium Math. Stat. Probab., 2, 315-343.

Tapia, R. A. and Johnson, J. R.(1978). Nonparametric probability density estimation. The John Hopkins University Press, Baltimore.

Tong, H. and Lim, K.S. (1980)."Threshold autoregression, limit cycles and cyclical data". J.R.S.S. (B), 42, 245-292.

Tran, L.T. (1989)."The $L_1$ convergence of Kernel density estimates under dependence". Canand. J. Stat., 17, 197-208.

Tran, L.T. (1990). Density estimation for time series by histograms. Manuscript.

Volkonskii, V.A. and Rozanov, Y.A. (1961)."Some limit theorems for random functions II". Th. Probab. Appl., 6, 186-198.

Wertz, W. and Schneider, B.(1979)."Statistical density estimation: a bibliography". Int. Stat. Rev., 47, 155-75.

Yokoyama, R. (1980)."Moment bounds for stationary mixing sequences". Z. Wahrsch. Vwer. Gebiete, 52, 45-57.

Yoshihara, K. (1976)."Limiting behaviour of v-statistics for stationary, absolutely regular processes". Z. Wahrsch. verw. Gebiete., 35, 237-252.

Yoshihara, K. (1978)."Probability inequalities for sums of absolutely regular processes and their applications". Z. Wahrsch. Verw. Gebiete, 43, 319-329.

Yoshihara, K. (1984)."Density estimation for samples satisfying a certain absolute regularity condition". J. Stat. Plann. Infer., 9, 19-32.