GENERAL LINEAR PROCESSES: A PROPERTY OF THE EMPIRICAL PROCESS APPLIED TO DENSITY AND MODE ESTIMATION

by

K.C. Chanda and F.H. Ruymgaart

November 1987

MIMEO SERIES #1736

DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

GENERAL LINEAR PROCESSES: A PROPERTY OF THE EMPIRICAL PROCESS APPLIED TO DENSITY AND MODE ESTIMATION

by

K.C. Chanda *and* F.H. Ruymgaart*

Department of Mathematics
Texas Tech University
Lubbock, Texas 79409
U.S.A.

Department of Mathematics
Kath. Un. Nijmegen
6525 ED Nijmegen
HOLLAND

*Part of this research was done while the author was visiting the University of North Carolina, Chapel Hill.

SUMMARY

General linear processes do not in general satisfy strong mixing conditions (Bradley (1987)). Therefore, we investigate the empirical process based on samples from such a general linear process by using a truncation argument and derive a local fluctuation inequality. It is well-known that such a fluctuation inequality is of basic importance in the study of the empirical process (see e.g. Einmahl (1987)). Here it is applied to obtain a rate of a.s. convergence for certain density estimators in the supremum norm. This extends a result by Chanda (1983). As a direct corollary a rate of a.s. convergence for a mode estimator is obtained.

AMS 1980 subject classification: primary 62M10, secondary 62G99.
key words and phrases: linear processes, empirical processes, density and mode estimation.

1. INTRODUCTION, NOTATION, ASSUMPTIONS

The class of linear processes contains many important examples of time series models like e.g. moving average (MA), autoregressive (AR), and, more generally, ARMA-processes; see e.g. Anderson (1971) or Hannan (1970). In his review paper on strong mixing Bradley (1986) refers to an interesting counterexample by Rosenblatt (1980, p. 267) which shows that even decent processes like AR(1)-processes need not always satisfy a strong mixing condition. Hence, statistical procedures for mixing sample elements do not in general immediately apply to samples constituting a linear process.

This is in particular true for the empirical process and its ramifications like density estimators. Properties of empirical processes under mixing conditions may be found in Mehra and Rao (1975) and Basawa and Prakasa Rao (1980, Chapter 11) for the univariate case and in Harel and Puri (1987) for multivariate sample elements. In this note we give a probability inequality on the local fluctuations of the empirical process based on a vector-valued linear process (Section 2). This inequality is of independent interest and typically provides the most important tool in the study of weak and strong convergence properties of empirical processes; see e.g. Einmahl (1987) for the i.i.d. case.

Restricting ourselves to the univariate case for convenience, we apply this inequality to obtain a speed of strong uniform convergence for a class of density estimators, from which strong consistency of the mode estimator as defined in Chernoff (1964) is derived (Section 3). The present result on density estimation extends the local strong convergence of density estimators in Chanda (1983). Also in the univariate case we give a discussion of our assumptions and some examples of processes that are covered by our set-up (Section 4).

The remainder part of this section is devoted to a specification of the model, the assumptions we need and some notation. For each $j \in \mathbb{Z}$, the set of all integers, Z_j is a d-dimensional random vector ($d \in \mathbb{N}$) defined on a probability space $(\Omega, \mathcal{E}, P)$ and A_j is a given non-random d×d-matrix. The Z_j are referred to as the *error terms*.

ASSUMPTION 1.1. *The Z_j ($j \in \mathbb{Z}$) are independent and identically distributed. For $i=1,\ldots,n$ ($n \in \mathbb{N}$) the series on the right below*

$$(1.1) \quad X_i = \Sigma_{k\in\mathbb{Z}} A_k Z_{i-k} ,$$

converges in probability. Hence the sample elements $X_1,\ldots,X_n$ are well-defined random vectors in $\mathbb{R}^d$ and form a stationary general linear process. The common distribution function F of the X_i has marginals with derivatives $f_j (j=1,\ldots,d)$ that are continuous and uniformly bounded by $M \in (0,\infty)$ on their respective supports in $\mathbb{R}$.

For $x = (x_1,\ldots,x_d)$, $y = (y_1,\ldots,y_d) \in \mathbb{R}^d$ we define $x \leq y$ to mean $x_j \leq y_j$ for all $j=1,\ldots,d$ and $x < y$ is defined by the requirement $x_j \leq y_j$ for all $j=1,\ldots,d$ and strict inequality for at least one j. For real $\xi \in \mathbb{R}$ we write $\xi^* = (\xi,\ldots,\xi) \in \mathbb{R}^d$; similarly we define $\infty^* = (\infty,\ldots,\infty)$ and $-\infty^* = (-\infty,\ldots,-\infty)$. For $m \in \mathbb{N}$ we write

$$(1.2) \quad X_{i,m} = \Sigma_{|k|\leq m} A_k Z_{i-k}, \ \bar{X}_{i,m} = \Sigma_{|k|>m} A_k Z_{i-k}.$$

The distribution function of the $X_{i,m}$ is denoted by F_m and that of the $\bar{X}_{i,m}$ by $\bar{F}_m$. For arbitrary $\epsilon > 0$ we write

$$(1.3) \quad \Omega(m,\epsilon,i) = \{-\epsilon^* \leq \bar{X}_{i,m} \leq \epsilon^*\}, \ \Omega(m,\epsilon) = \cap_{i=1}^{n} \Omega(m,\epsilon,i),$$

$$(1.4) \quad P(\Omega^c(m,\epsilon,i)) = \varphi(m,\epsilon), \ 2\ d\ \varphi(m,\epsilon) + 2\ d(1+2d)M\ \epsilon = \delta(m,\epsilon).$$

It is clear that

$$(1.5)\quad P(\Omega^c(m,\epsilon)) \leq n\,\varphi(m,\epsilon).$$

In the example in Rosenblatt (1980), mentioned above, the dimension d=1, F is the uniform (0,1) distribution and the F_m are discrete. Therefore, we preferred not to impose any smoothness condition on the F_m. The smoothness of F as required by Assumption 1.1, however, entails

$$(1.6)\quad |F(x) - F_m(x)| = |F(x) - [P(\{X_i \leq x + \bar{X}_{i,m}\} \cap \Omega(m,\epsilon,i))$$

$$+ P(\{X_i \leq x + \bar{X}_{i,m}\} \cap \Omega^c(m,\epsilon,i))]| \leq$$

$$\leq \max\{F(x) - F(x-\epsilon^*) + \varphi(m,\epsilon),$$

$$F(x+\epsilon^*) - F(x) + \varphi(m,\epsilon)\} \leq$$

$$\leq \varphi(m,\epsilon) + 2d\,M\,\epsilon, \quad \textit{for all } x \in \mathbb{R}^d.$$

From this we obtain the useful relation

$$(1.7)\quad \sup_{x\in\mathbb{R}^d} |F(x) - F_m(x)| \leq \varphi(m,\epsilon) + 2d\,M\,\epsilon < \delta(m,\epsilon).$$

In asymptotic situations we mostly have $m = m_n \to \infty$ and $\epsilon = \epsilon_n \to 0$, as the sample size $n \to \infty$, and simply write

$$(1.8)\quad \varphi_n = \varphi(m_n,\epsilon_n),\ \delta_n = \delta(m_n,\epsilon_n),\ \Omega_n = \Omega(m_n,\epsilon_n).$$

The following assumption on the orders of magnitude will be needed.

ASSUMPTION 1.2. *As* $n \to \infty$ *there exist* $\epsilon_n = O(n^{-\alpha})$ *for some* $\alpha > ½$ *and* $m_n = O(n^{\beta})$ *for some* $0 < \beta < ½$, *such that*

$$(1.9)\quad \varphi_n = O(n^{-\rho}),\ \textit{for some } \rho > 2,\ \textit{and hence } \delta_n = O(n^{-\alpha}).$$

Let us conclude with some more notation that is used throughout. For any $\Psi : \mathbb{R}^d \to \mathbb{R}$ and $x < y$ we define

$$(1.10)\quad \Psi\{x,y\} = \Delta_x^y\, \Psi,$$

where Δ_x^y is the usual difference operator. Note that in this notation we may e.g. write $P(x < X_i \leq y) = F\{x,y\}$. Throughout this paper the numbers

$$(1.11)\quad A,\ B,\ C \in (0,\infty)$$

will be used as *generic constants* that may only depend on the dimension d. Hence these numbers are in fact independent of all the relevant parameters like in particular the distribution functions F and $F_m (m \in \mathbb{N})$, and the sample size n.

2. A FUNDAMENTAL FLUCTUATION INEQUALITY

The empirical distribution function based on the $X_1,\ldots,X_n$ will as usual be defined by

$$(2.1)\quad \hat{F}_n(x) = \frac{1}{n} \cdot \#\{1 \leq i \leq n : X_i \leq x\},\ x \in \mathbb{R}^d,$$

and the corresponding empirical process by

$$(2.2)\quad U_n = \{U_n(x) = n^{\frac{1}{2}}(\hat{F}_n(x) - F(x)),\ x \in \mathbb{R}^d\}.$$

An important role in our inequality is played by the function

$$(2.3)\quad \psi(\lambda) = 2\,\lambda^{-2} \int_0^\lambda \log(1+x)dx,\ \lambda > 0;\ \psi(0) = 1.$$

This function is continuous on $[0,\infty)$ and $\psi(\lambda) \downarrow 0$, as $\lambda \uparrow \infty$.

THEOREM 2.1. *Let Assumption 1.1 be satisfied. For arbitrary* $m \in \mathbb{N}$ *and* $\epsilon > 0$ *we have*

$$(2.4)\quad P(\sup_{a\le x<y\le b} |U_n\{x,y\}| \ge \lambda) \le$$

$$\le m\, C \exp\left[\frac{-A\,\lambda^2}{m\, F\{a,b\}}\, \psi\left[\frac{B\lambda}{n^{½}\, F\{a,b\}}\right]\right] + n\, \varphi(m,\epsilon),$$

for any $-\infty^* \le a < b \le \infty^*$ *provided that*

$$(2.5)\quad \lambda \ge 4\, n^{½}\, \delta(m,\epsilon),\ F\{a,b\} \ge 2\, \delta(m,\epsilon).$$

PROOF. Without loss of generality we may and will assume that

$$(2.6)\quad \nu = n/(2m+1) \in \mathbb{N},$$

so that $2m+1 = n/\nu$. It is immediate from the definition of the $X_{i,m}$ that

$$(2.7)\quad X_{j+(i-1)n/\nu,m}\ ,\quad i \in \{1,\dots,\nu\},$$

are i.i.d. Let $\hat{F}^{(j)}_{\nu,m}$ denote the empirical distribution function based on the sample of size ν in (2.7), $\hat{F}_{n,m}$ the empirical distribution function based on all of the $X_{1,m},\dots,X_{n,m}$, and

$$(2.8)\quad U^{(j)}_{\nu,m} = \nu^{½}\, (\hat{F}^{(j)}_{\nu,m} - F_m),$$

$$(2.9)\quad U_{n,m} = n^{½}(\hat{F}_{n,m} - F_m).$$

The following relation is obvious

$$(2.10)\quad U_{n,m} = (n/\nu)^{-½}\, \Sigma_{j=1}^{n/\nu}\, U^{(j)}_{\nu,m}\ .$$

It follows that

$$(2.11)\quad P\left[\sup_{a\le x<y\le b} |U_{n,m}\{x,y\}| \ge \lambda\right] \le$$

$$\le \Sigma_{j=1}^{n/\nu}\, P\left[\sup_{a\le x<y\le b} |U^{(j)}_{\nu,m}\{x,y\}| \ge \lambda (n/\nu)^{-½}\right].$$

To each of the probabilities on the right in (2.11) we may apply e.g. Ruymgaart and Wellner (1984, Theorem 1.1), since this inequality is well-known to remain true for independent and identically distributed d-dimensional random vectors with *arbitrary* range and *arbitrary* distribution function (see also e.g. Einmahl (1987, Inequality 2.5 and Section 6.3.c). Application yields

(2.12)
$$P\left[\sup_{a\leq x<y\leq b} |U_{n,m}\{x,y\}| \geq \lambda\right] \leq$$

$$\leq \frac{n}{\nu}\, C \exp\left[\frac{-A\,\lambda^2 \nu}{n\, F_m\{a,b\}}\,\psi\left[\frac{B\lambda}{n^{½}\, F_m\{a,b\}}\right]\right] .$$

We may now pass from the $U_{n,m}$- process to the U_n- process, provided that we restrict the outcomes to the subset $\Omega(m,\epsilon)$ in (1.3). Let us note that

(2.13)
$$\pm 1_{\Omega(m,\epsilon)} \cdot U_n\{x,y\} \leq \pm U_{n,m}\{x \mp \epsilon^*,\ y \pm \epsilon^*\}$$

$$+ n^{½}|F_m\{x \mp \epsilon^*, y \pm \epsilon^*\} - F\{x \mp \epsilon^*, y \pm \epsilon^*\}|$$

$$+ n^{½}|F(x \mp \epsilon^*, y \pm \epsilon^*\} - F\{x,y\}| \leq$$

$$\leq \pm U_{n,m}\{x \mp \epsilon^*, y \pm \epsilon^*\} + n^{½}\,\delta(m,\epsilon), \quad \textit{for all} \quad -\infty^* \leq x < y \leq \infty^*,$$

in view of Assumption 1.1 and (1.7).

It is immediate from (2.13) that

(2.14)
$$\sup_{a\leq x<y\leq b} 1_{\Omega(m,\epsilon)} \cdot |U_n\{x,y\}| \leq$$

$$\leq \sup_{a-\epsilon^*\leq x<y\leq b+\epsilon^*} |U_{n,m}\{x,y\}| + n^{½}\,\delta(m,\epsilon) .$$

and application of (2.12) yields (note (2.5))

$$(2.15)\quad P\left[\sup_{a\le x<y\le b} |U_n\{x,y\}| \ge \lambda\right] \le$$

$$\le \frac{n}{\upsilon}\, C \exp\left[\frac{-A(\lambda-n^{1/2}\,\delta(m,\epsilon))^2\upsilon}{n\, F_m\{a-\epsilon^*,\ b+\epsilon^*\}}\ \psi\left[\frac{B(\lambda-n^{1/2}\,\delta(m,\epsilon))}{n^{1/2}\, F_m\{a-\epsilon^*,\ b+\epsilon^*\}}\right]\right]$$
$$+\ P(\Omega^c(m,\epsilon)).$$

Let us now recall (1.7), note that $F\{a,b\} - \delta(m,\epsilon) \le F_m\{a-\epsilon^*,\ b+\epsilon^*\} \le F\{a,b\} + \delta(m,\epsilon)$ and that $\upsilon/n = 1/(2m+1)$, use the fact that ψ is decreasing on $[0,\infty)$ and exploit the generic character of the numbers A, B, C $\in (0,\infty)$ to arrive at

$$(2.16)\quad P\left[\sup_{a\le x<y\le b} |U_n\{x,y\}| \ge \lambda\right] \le$$

$$\le m\, C \exp\left[\frac{-A(\lambda-n^{1/2}\,\delta(m,\epsilon))^2}{m(F\{a,b\} + \delta(m,\epsilon))}\ \psi\left[\frac{B\lambda}{n^{1/2}(F\{a,b\} - \delta(m,\epsilon))}\right]\right]$$
$$+\ n\,\varphi(m,\epsilon)\ .$$

We finally obtain (2.4) by observing that condition (2.5) implies $(\lambda-n^{1/2}\,\delta(m,\epsilon))^2 \ge \frac{1}{2}\lambda^2$, $F\{a,b\} + \delta(m,\epsilon) \le 3\, F\{a,b\}$ and $F\{a,b\} - \delta(m,\epsilon) \ge \frac{1}{2} F\{a,b\}$. Q.E.D.

Let us now specialize to d=1, introduce the uniform (0,1) random variables $F(X_1) = \xi_1,\ldots,F(X_n) = \xi_n$ and write $\hat{\Gamma}_n$ for their empirical distribution function. The corresponding reduced empirical process is written

$$(2.17)\quad \tilde{U}_n = \{\tilde{U}_n(t) = n^{1/2}(\hat{\Gamma}_n(t)-t),\ t \in [0,1]\}.$$

It is clear that we have the relation

$$(2.18)\quad \tilde{U}_n\{s,t\} \overset{d}{=} U_n\{F^{-1}(s),\ F^{-1}(t)\},\quad \textit{for all } 0 \le s < t \le 1.$$

Although the $\xi_1,\dots,\xi_n$ do *not* form a linear process the following result is nevertheless immediate from Theorem 2.1 and (2.18).

COROLLARY 2.1. *Take* d=1 *and let Assumption 1.1 be fulfilled. For arbitrary* $m \in \mathbb{N}$ *and* $\epsilon > 0$ *the reduced empirical process in* (2.17) *satisfies*

$$(2.19)\quad P\left[\sup_{a\le s<t\le b} |\tilde{U}_n\{s,t\}| \ge \lambda\right] \le$$

$$\le m\ C \exp\left[\frac{-A\ \lambda^2}{m(b-a)}\ \psi\left[\frac{B\lambda}{n^{\frac{1}{2}}(b-a)}\right]\right] + n\ \varphi(m,\epsilon),$$

for any $0 \le a < b \le 1$ *provided that*

$$(2.20)\quad \lambda \ge 4\ n^{\frac{1}{2}}\ \delta(m,\epsilon),\qquad b-a \ge 2\ \delta(m,\epsilon).$$

If in addition Assumption 1.2 is fulfilled and if we also choose $a = a_n$, $b = b_n$ and $\lambda = \lambda_n$ depending on n the upper bound in (2.19) often can be written in a simpler form.

COROLLARY 2.2. *Take* d = 1 *and let both Assumption 1.1 and Assumption 1.2 be satisfied. Let us choose* $a=a_n$, $b=b_n \in [0,1]$ $\lambda=\lambda_n \in (0,\infty)$ *such that* $b_n-a_n = n^{-\gamma}$ *for some* $0 < \gamma < \frac{1}{2}$ *and* $\lambda_n = c\ n^{\sigma}$ *for some* $c \in (0,\infty)$ *and* $\sigma < \frac{1}{2}-\gamma$. *Then there exists* $n(\alpha,\beta,\gamma,\lambda,\sigma) \in \mathbb{N}$ *such that*

$$(2.21)\quad P\left[\sup_{a_n\le s<t\le b_n} |\tilde{U}_n\{s,t\}| \ge c\ n^{\sigma}\right] \le$$

$$\le C\ n^{\beta} \exp\left[-A\ c^2\ n^{2\sigma+\gamma-\beta}\right] + C\ n^{1-\rho},$$

provided that $n \ge n(\alpha,\beta,\gamma,\lambda,\sigma)$.

PROOF. Let us just note that the conditions on the parameters entail that $n^{\sigma-\frac{1}{2}+\gamma} \to 0$, as $n \to \infty$, so that $\psi(B\,\lambda_n/(n^{\frac{1}{2}}(b_n-a_n))) \to 1$, as $n \to \infty$, and hence may be absorbed in the generic constant A for n sufficiently large. It is, moreover, clear that condition (2.20) is automatically fulfilled for n sufficiently large. Q.E.D.

3. DENSITY AND MODE ESTIMATION

Throughout this section we take d=1. We will consider so called naive estimators of $F'=f$, defined by

$$(3.1)\quad \hat{f}_n = \frac{1}{\ell_n}\hat{F}_n\{x-\tfrac{1}{2}\,\ell_n,\ x+\tfrac{1}{2}\,\ell_n\},\ x \in \mathbb{R},$$

for some $\ell_n > 0$. For the expected values we write

$$(3.2)\quad E\,\hat{f}_n(x) = \frac{1}{\ell_n}F\{x-\tfrac{1}{2}\,\ell_n,\ x+\tfrac{1}{2}\,\ell_n\} = f_n(x),\ x \in \mathbb{R}.$$

THEOREM 3.1. *Take d=1, let Assumptions 1.1 and 1.2 be fulfilled and suppose in addition that $F'=f$ is twice continuously differentiable with $|f''|$ bounded by $M \in (0,\infty)$ on $\mathbb{R}$. Then we have*

$$(3.3)\quad n^{\xi}\ \sup_{x\in\mathbb{R}}\ |\hat{f}_n(x) - f(x)| \xrightarrow{a.s.} 0,\ \text{as } n \to \infty,\ \text{for any } 0 < \xi < 2(1-\beta)/5,$$

when we choose $\ell_n=n^{-\gamma}$ with $\gamma = (1-\beta)/5$.

PROOF. The present smoothness conditions entail that

$$(3.4)\quad n^{\xi}\ \sup_{x\in\mathbb{R}}|f_n(x)-f(x)| = \mathcal{O}(n^{\xi-2\gamma}) \to 0,\ \text{as}\ \ n \to \infty,$$

so that it remains to consider $\hat{f}_n-f_n$. Let us write

$$(3.5)\quad F(x \pm \tfrac{1}{2}\,\ell_n) = t^{\pm}_{x,n}\ ,$$

and note that

$$(3.6)\quad |t^{+}_{x,n} - t^{-}_{x,n}| \leq M\,\ell_n, \text{ for all } x \in \mathbb{R}.$$

Choosing $k_n = [1/(M\,\ell_n)] \in \mathbb{N}$, where $[z]$ is the greatest integer $\leq z \in \mathbb{R}$, let us partition $[0,1]$ into the subintervals

$$(3.7)\quad ((j-1)/k_n,\ j/k_n] = (t_{j-1}, t_j],\ j \in \{1,\dots,k_n\}.$$

Let us note that the length of the intervals in (3.7) is of order $\mathcal{O}(n^{-\gamma})$ but at least $M\,\ell_n$, so that any interval $(t^{-}_{x,n},\ t^{+}_{x,n}]$ intersects at most two adjacent subintervals in (3.7).

In terms of the reduced empirical process in (2.17) we have

$$(3.8)\quad P(n^{\xi} \sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f_n(x)| \geq c) =$$

$$= P(n^{\xi+\gamma-\frac{1}{2}} \sup_{x\in\mathbb{R}} |\tilde{U}_n\{t^{-}_{x,n},\ t^{+}_{x,n}\}| \geq c) \leq$$

$$\leq P\left[\max_{j=1,\dots,k_n} \sup_{t_{j-1}\leq s<t\leq t_j} |\tilde{U}_n\{s,t\}| \geq \tfrac{1}{2}\, c\, n^{\frac{1}{2}-\xi-\gamma}\right],$$

for arbitrary $c \in (0,\infty)$. Because $\frac{1}{2}-\xi-\gamma < \frac{1}{2}-\gamma$, $0 < (1-\beta)/5 < \frac{1}{2}$ and $M\,n^{-\gamma} \leq t_j - t_{j-1} \leq 2M\,n^{-\gamma}$ for all $j=1,\dots,k_n$, Corollary 2.2 applies and yields for n sufficiently large

$$(3.9)\quad p_n(c) = P\left[n^{\xi} \sup_{x\in\mathbb{R}} |\hat{f}_n(x) - f_n(x)| \geq c\right] \leq$$

$$\leq (C/M)n^{\beta+\gamma} \exp\left[-A\, c^2 n^{1-2\xi-2\gamma-\beta}\right] + C\, n^{1-\rho}.$$

Because $1-2\xi-2\gamma-\beta > 1-4(1-\beta)/5 - (1-\beta)/5-\beta = 0$ and $\rho > 2$ by Assumption 1.2 it is clear that $\Sigma_{n=1}^{\infty} p_n(c) < \infty$. Q.E.D.

In Chanda (1985) a class of linear estimators of the form $\Sigma_{i=1}^{n} a_i X_i$ for estimating the symmetry point of a symmetric density is considered.

The asymptotics of location estimators that are linear combinations of order statistics of the form $\Sigma_{i=1}^{n} a_i X_{i:n}$ might be investigated in a forthcoming paper. The previous theorem enables us, however, to simply prove a speed of a.s. convergence of the mode estimator that was first considered in Chernoff (1964) and based on naive density estimators.

THEOREM 3.2. *In addition to the conditions of Theorem 3.1 let us assume that* f *has a unique maximum at* θ. *Let* $\hat{f}_n$ *have a maximum at* $\hat{\theta}_n$. *Then we have*

$$(3.10)\quad n^{\xi/2}(\hat{\theta}_n-\theta) \xrightarrow{a.s.} 0, \text{ as } n \to \infty, \textit{ for any } 0 < \xi < 2(1-\beta)/5.$$

PROOF. Let us choose ξ as in (3.10) and let Ω_0 be a subset of Ω with $P(\Omega_0)=1$ on which (3.3) holds true. For any $\omega \in \Omega_0$ and $\epsilon > 0$ there exists $n_1 = n_1(\omega,\epsilon)$ such that

$$(3.11)\quad f(\theta) - \epsilon\, n^{-\xi} \le \hat{f}_{n,\omega}(\hat{\theta}_n(\omega)) \le f(\theta) + \epsilon\, n^{-\xi},$$

for $n \ge n_1$. This entails that

$$(3.12)\quad \hat{\theta}_n(\omega) \in \{x \in \mathbb{R} : f(x) + \epsilon\, n^{-\xi} \ge f(\theta) - \epsilon\, n^{-\xi}\}.$$

Moreover, there exist $\delta > 0$ such that $f''(x) \le -c < 0$ for all $x \in (\theta-\delta,\theta+\delta)$, and $n_2 = n_2(\delta) \ge n_1$ such that the set on the right in (3.12) is contained in $(\theta-\delta,\ \theta+\delta)$ for $n \ge n_2$. Hence for $n \ge n_2$ we have

$$(3.12)\quad \hat{\theta}_n(\omega) \in \{x \in \mathbb{R} : f(\theta) - [f(\theta) - \tfrac{1}{2} c(x-\theta)^2] < 2\,\epsilon\, n^{-\xi}\} =$$

$$= (\theta - 2(\epsilon/c)^{1/2}\, n^{-\frac{1}{2}\xi},\ \theta + 2(\epsilon/c)^{1/2}\, n^{-\frac{1}{2}\xi}).$$

Since $\epsilon > 0$ is arbitrary this proves (3.10). Q.E.D.

4. DISCUSSION AND EXAMPLES

Also in this section we choose d=1. The assumptions in Section 1 will be discussed by showing that they are satisfied for two examples of subclasses of the class of linear processes.

AR(1)-PROCESSES. For $0 < r < 1$ let us consider

$$(4.1)\quad X_i = \Sigma_{k=1}^{\infty} r^k Z_{i-k}.$$

It is usually assumed that the error terms Z_j have $E(Z_j) = 0$ and $Var(Z_j) = \sigma^2 \in (0,\infty)$, so that the convergence in (4.1) is in quadratic mean and consequently in probability as we require. Let us write $\varphi(t) = E(\exp(i\ t\ Z_j))$, $t \in \mathbb{R}$. Assuming that

$$(4.2)\quad \int_{-\infty}^{\infty} |\varphi(t)|dt < \infty,$$

entails that the distribution of the Z_j has a continuous density bounded by the number on the left in (4.2). It is easily seen that (4.2) also implies that the X_i have a distribution with bounded continuous density. Hence Assumption 1.1 is satisfied.

Let us next note that, for any $\epsilon > 0$ and $m \in \mathbb{N}$,

$$(4.3)\quad \varphi(m,\epsilon) = P\left[|\Sigma_{k=m+1}^{\infty} r^k Z_{i-k}| > \epsilon\right] \le \sigma^2 r^{2(m+1)} / (\epsilon^2(1-r^2)).$$

Choosing $\epsilon = \epsilon_n = n^{-\alpha}$ and $m = m_n = [n^\beta]$, for any $\alpha > ½$ and $0 < \beta < ½$, we find that

$$(4.4)\quad \varphi_n \le c_1 r^{2([n^\beta]+1)} n^{2\alpha},$$

for some number $c_1 \in (0,\infty)$. It is obvious that for *each* $\rho > 2$ there exists a number $c_2 = c_2(\rho) \in (0,\infty)$ such that

$$(4.5)\quad \varphi_n \le c_2 n^{-\rho}, \quad \textit{for all } n \in \mathbb{N}.$$

Hence Assumption 1.2 is amply fulfilled.

The counterexample by Rosenblatt (1980), considered in Bradley (1986) and mentioned in the introduction, is the special case where $r = ½$ and the Z_j are $\{0,1\}$- variables with $P(Z_j=0) = P(Z_j=1) = ½$. Althouth the Z_j are discrete the X_i turn out to have the uniform distribution on $(0,1)$, so that the assumptions are still fulfilled. It is interesting that this process is *not strongly mixing*.

ERROR TERMS WITH STABLE D.F. Let us now assume that the Z_j have a symmetric stable distribution with scale parameter 1 and index $0 < \mu < 1$, so that the first moment doesn't even exist. According to Leadbetter, Lindgren and Rootzén (1983, p. 73) the series

$$(4.6)\quad X_i = \Sigma_{k\in\mathbb{Z}}\, a_k\, Z_{i-k}\,,\quad a_k \in \mathbb{R},$$

converges a.s. if and only if

$$(4.7)\quad \Sigma_{k\in\mathbb{Z}}\, |a_k|^{\mu} < \infty$$

Since condition (4.2) is obviously fulfilled in the stable case, Assumption 1.2 is satisfied provided that (4.7) holds true.

Let us now assume that

$$(4.8)\quad |a_k| \le c\, k^{-\tau}\,,\quad \textit{for some } \tau > 1/\mu\,,$$

and $c \in (0,\infty)$. Using Leadbetter, Lindgren and Rootzén (1983, p. 74) we see that for some numbers $c_1 = c_1(\mu)$, $c_2 = c_2(\mu) \in (0,\infty)$, and any $\epsilon > 0$ and $m \in \mathbb{N}$ we have

$$(4.9)\quad \varphi(m,\epsilon) = P\Big[\big|\Sigma_{|k|>m}\, a_k\, Z_{i-k}\big| > \epsilon\Big] \le$$

$$\leq c_1 \left[\Sigma_{|k|>m} \ (c\ k^{-\tau})^{\mu} \right] \epsilon^{-\mu} \leq$$

$$\leq c_2 \ m^{1-\tau\mu} \ \epsilon^{-\mu} \ .$$

Choosing $\epsilon = \epsilon_n = n^{-\alpha}$ and $m = m_n = [n^{\beta}]$ for any $\alpha > ½$ and $0 < \beta < ½$, we obtain

$$(4.10) \quad \varphi_n \leq c_3 \ n^{\beta(1-\tau\mu)} n^{\alpha\mu},$$

where $c_3 = c_3(\mu) \in (0,\infty)$. It follows that

$$(4.11) \quad \varphi_n \leq c_3 \ n^{-\rho} \ , \quad \rho > 2,$$

for each $\tau \in \mathbb{R}$ satisfying

$$(4.12) \quad \tau > (\rho + \alpha\mu + \beta)/(\beta\mu).$$

(Note that (4.12) implies $\tau > 1/\mu$.) Relation (4.12) also implies that for

$$(4.13) \quad \tau > 1 + 5/\mu, \ 0 < \mu \leq 1$$

we can find $\alpha > ½$, $0 < \beta < ½$, and $\rho > 2$ such that (4.12) is fulfilled. Hence we have established simple conditions under which Assumption 1.2 is satisfied.

The purpose of this example was to provide a useful situation in which the assumptions are easily seen to be satisfied, although first moments do not even exist. It is not excluded, however, that in this case a strong mixing condition may be fulfilled.

REFERENCES

[1] ANDERSON, T.W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

[2] BASAWA, I.V. and PRAKASA RAO, B.L.S. (1980). *Statistical Inference for Stochastic Processes*. Ac. Press, New York.

[3] BRADLEY, R.C. (1986). Basic properties of strong mixing conditions. In: *Dependence in Probability and Statistics* (E. Eberlein and M.S. Taqqu, Eds.), pp. 165-192. Birkhäuser, Boston.

[4] CHANDA, K.C. (1983). Density estimation for linear processes. *Ann. Inst. Statist. Math.* 35, 439-446.

[5] CHANDA, K.C. (1985). Sampling distribution for a class of estimators for nonregular linear processes. *Statist. Probab. Letters* 3, 261-268.

[6] CHERNOFF, H. (1964). Estimation of the mode. *Ann. Inst. Statist. Math.* *16*, 31-41.

[7] EINMAHL, J.H.J. (1987). *Multivariate Empirical Processes*. CWI tract 32, Amsterdam.

[8] HANNAN, E.J. (1970). *Multiple Time Series*. Wiley, New York.

[9] HAREL, M. and PURI, M.L. (1987). Convergence faible de la statistique sérielle linéaire de rang en condition de dépendance avec applications aux séries chronologiques et processus de Markov. *C.R. Acad. Sci. Paris, t.304, série 1, no. 19*, 583-586.

[10] LEADBETTER, M.R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

[11] MEHRA, K.L. and RAO, M.S. (1975). Weak convergence of generalized empirical processes relative to d_q under strong mixing. *Ann. Probability* 3, 979-991.

[12] ROSENBLATT, M. (1980). Linear processes and bispectra. *J. Appl. Probability 17*, 265-270.

[13] RUYMGAART, F.H. and WELLNER, J.A. (1984). Some properties of weighted multivariate empirical processes. *Statist. Decisions* 2, 199-223.