COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 664

FIRST ORDER AUTOREGRESSIVE PROCESSES AND STRONG MIXING

Donald W. K. Andrews

Yale University

March 1983

FIRST ORDER AUTOREGRESSIVE PROCESSES AND STRONG MIXING

Donald W. K. Andrews

Yale University

March 1983

Summary

A sufficient condition is given such that first-order autoregressive processes are strong mixing. The condition is specified in terms of the univariate distribution of the independent identically distributed innovation random variables. Normal, exponential, uniform, Cauchy, and many other continuous innovation random variables are shown to satisfy the condition. In addition, an example of a first-order autoregressive process which is not strong mixing is given. This process has Bernoulli (p) innovation random variables and any autoregressive parameter in $(0, 1/2]$.

1. Introduction

Let $\langle X_t \rangle \equiv (\ldots X_{t-1}, X_t, X_{t+1}, \ldots)$ be a sequence of random variables (rv's), and $\mathcal{X}_{a,b}$ be the σ-algebra generated by $(X_a, X_{a+1}, \ldots, X_b)$ for $-\infty \leq a \leq b \leq \infty$. $\langle X_t \rangle$ is said to be strong mixing if

$$\sup_{-\infty<t<\infty} \sup_{A\in\mathcal{X}_{-\infty,t}, B\in\mathcal{X}_{t+s,\infty}} |P(A\cap B) - P(A)P(B)| \equiv \alpha(s) \downarrow 0 \quad \text{as} \quad s \to \infty . \quad (1)$$

The strong mixing condition was introduced by Rosenblatt in 1956 to prove the central limit theorem for "weakly dependent" rv's. It has since assumed a position of considerable importance in probability theory. This is due to its tractability in the derivation of asymptotic properties of various functions of sequences of dependent rv's. Its areas of application include: central limit theorems — Rosenblatt (1956), Ibragimov (1962), Rozanov (1963), Philipp (1969), Ibragimov and Linnik (1971); laws of the iterated logarithm — Oodaira (1975), Philipp (1977); empirical processes — Deo (1973), Yokoyama (1973), Withers (1975), Deo (1975a), Mehra and Rao (1975b), Philipp and Pinzur (1977), Andrews (1982b); order statistics — Loynes (1965), Welsch (1971, 1972), Mehra and Rao (1975a), Deo (1975b); and robust estimators — Koul (1977), Andrews (1982a, 1982c).

While the strong mixing condition has been widely adopted in the literature, the number of well-known processes which satisfy the condition is still somewhat limited. M-dependent processes are trivially strong mixing, since $\alpha(s) = 0$, for all $s > M$. Kolmogorov and Rozanov (1960) proved that Gaussian processes with continuous and positive spectral densities are strong mixing. Also, Ibragimov and Linnik (1971) have shown that stationary Markov processes $\langle Y_t \rangle$ are strong mixing provided their s-step transition probabilities, $p^{(s)}(y,A)$, satisfy

$$\sup_{y \in W} |p^{(s)}(y,A) - P(Y_t \in A)| \leq C_1 \rho^s \quad \text{for all} \quad s = 1, 2, \ldots , \tag{2}$$

for all measurable subsets A of W, where Y_t takes values in W, and $C_1 < \infty$ is some constant.

In this paper we add to the list of strong mixing processes certain <u>first-order autoregressive</u> (AR(1)) processes. By definition, $<X_t>$ is an AR(1) process with innovation rv ε_0 (with corresponding innovation process $<\varepsilon_t>$) and autoregressive parameter ρ if

$$X_t - \rho X_{t-1} = \varepsilon_t \quad \text{for all} \quad t = \ldots -1, 0, 1, \ldots , \tag{3}$$

and $<\varepsilon_t>$ is a sequence of <u>independent identically distributed</u> (iid) rv's. AR(1) processes are one of the most basic stochastic processes. As such, they have had wide applications in statistics and applied probability. For examples, Durbin and Watson (1950, 1951) use AR(1) processes to model the error term in time series regression models. The Box Jenkins (1969) time series approach makes extensive use of AR(1) models (and more complicated autoregressive-moving average models) to represent the distributions of time series from engineering, economics, and other fields of study. Also, AR(1) processes are seeing new applications in modelling dependent processes with gamma marginal distributions (see Gaver and Lewis (1980)).

The condition used in this paper to ensure that an AR(1) process is strong mixing is that the innovation rv ε_0 has a density which satisfies a smoothness condition. Since an AR(1) process is defined in terms of the distribution of ε_0 and the autoregressive parameter ρ, this is a more transparent and natural condition than that given in terms of s-step transition probabilities in (2) above. (Note, condition (2) is also sufficient for an AR(1) process to be strong mixing, since AR(1) processes are stationary

Markov processes.) Section 3 gives several alternative smoothness conditions and numerous examples of distributions of ε_0 which satisfy them.

It is not surprising that many AR(1) processes are strong mixing, since many of the results which use the strong mixing condition have been proved directly for AR(1) processes. However, in Section 4 we show that the smoothness condition placed on the distribution of the innovation rv is not superfluous. It is demonstrated, by explicit construction of sequences of sets $A \in \mathcal{X}_{-\infty,t}$ and $B \in \mathcal{X}_{t+s,\infty}$, $s = 1, 2, \ldots$, that AR(1) processes generated by iid Bernoulli (p) innovations are not strong mixing--no matter how small the autoregressive coefficient is! This (somewhat surprising) result is unfortunate. We would like the mixing condition used for the numerous results listed above to be satisfied by all processes which are sufficiently "weakly dependent" for central limit theorem and related results to hold.

When first introducing the strong mixing condition Rosenblatt (1956) commented that "It would be of very great interest to see how much stronger the notion of a strong mixing condition is than that of an ordinary mixing condition [as defined, e.g., in Hannan (1970)] in the case of a strictly stationary process." The result of Section 4 addresses this question, since not all AR(1) processes are strong mixing, yet it is well known that all (L^2) AR(1) processes are ordinary mixing (e.g., see Hannan (1970, Chap. IV, Thm. 3)). However, while the ordinary mixing condition is sufficient to derive Ergodic-type results, it is not sufficient to derive the central limit theorem and related results which are possible with the strong mixing condition.

2. Strong Mixing AR(1) Processes

In this section we introduce a condition, S_ρ , on the distribution of an innovation rv ε_0 . This condition may depend on ρ , an autoregressive parameter. We show that an AR(1) process $\langle X_t \rangle$ generated by ε_0 and ρ , where ε_0 satisfies S_ρ , is strong mixing. Further, the mixing numbers of $\langle X_t \rangle$ are shown to be dominated by an exponentially declining sequence.
The condition S_ρ is examined in Section 3. There, several alternative conditions are given which involve only the distribution of ε_0 and not the autoregressive parameter ρ . These conditions are much simpler than S_ρ , but stronger. Examples of distributions which satisfy one of these conditions are given.

Define the variational distance between two rv's Y and Z by

$$\Delta_V(Y,Z) \equiv \sup_{D \in \mathcal{B}} |P(Y \in D) - P(Z \in D)| \quad \text{where } \mathcal{B} \text{ is the Borel } \sigma\text{-algebra on } R ,$$

$$= \frac{1}{2}\int |f_Y(\omega) - f_Z(\omega)| d\mu(\omega)$$

where f_Y and f_Z are the densities of Y and Z with respect to some measure μ .

The innovation rv ε_0 (with corresponding iid innovation process $\langle \varepsilon_t \rangle$) satisfies condition S_ρ if:

(S_ρ) For a given constant $\rho \in (-1,1)$, there exist positive constants p , q and C , and a positive integer s_0 such that

$$\varepsilon_0 \in L^p ,$$

and for $s = s_0$,

$$\Delta_V\left(\sum_{\ell=0}^{s-1} \rho^\ell \varepsilon_\ell, \sum_{\ell=0}^{s-1} \rho^\ell \varepsilon_\ell + y\right) \leq C|y|^q \qquad (4)$$

where $\mathcal{B}$ is the Borel σ-algebra on R .

Theorem 1. Suppose the rv ε_0 satisfies condition S_ρ , then the AR(1) process $\langle X_t \rangle$ with innovation rv ε_0 and autoregressive parameter ρ is strong mixing, and the mixing numbers $\alpha(s)$, $s = 1, 2, \ldots$ of $\langle X_t \rangle$ satisfy

$$\alpha(s) \leq \tilde{\alpha}(s) \equiv \begin{cases} 2(C+1)E|X_t|^\nu \cdot |\rho^\nu|^s & \text{for } s \geq s_0 \\ 1 & \text{for } 1 \leq s < s_0 , \end{cases} \tag{5}$$

where $\nu = \min\{p,q,1\}$, and s_0 , C , p , and q are as in S_ρ .

The proof of Theorem 1 uses the following lemmas, whose proofs follow that of the Theorem:

Lemma 1. If the rv ε_0 and the constant ρ are such that condition S_ρ holds for some $s = s_0$, then condition S_ρ holds for all integers $s \geq s_0$ (and the constants p , q , and C are independent of s).

Given an innovation process $\langle \varepsilon_t \rangle$, let $E_{a,b}$ be the σ-algebra generated by $\varepsilon_a, \varepsilon_{a+1}, \ldots, \varepsilon_b$, for $-\infty \leq a \leq b \leq \infty$. Let $X'_{a,b}$ be the class of all finite unions of finite intersections of sets of the form $\{X_t \in D\}$ where $D \in \mathcal{B}$, the Borel σ-algebra on R , and where t is an integer in $(a-1,b+1)$, for $-\infty \leq a \leq b \leq \infty$. Note, $X'_{a,b}$ is an algebra.

Lemma 2. If the distribution of ε_0 satisfies S_ρ , then for any $B \in X'_{t+s,\infty}$, there exists $\tilde{B} \in E_{t+1,\infty}$ such that

$$\sup_{A \in X_{-\infty,t}} |P(A \cap B) - P(A \cap \tilde{B})| \leq \tilde{\alpha}(s) \tag{6}$$

for all $s = 1, 2, \ldots,$ where $\tilde{\alpha}(s)$ is independent of B and t , and is defined in (5).

Lemma 3. Lemma 2 holds with $X'_{t,\infty}$ replaced by $X_{t,\infty}$.

Proof of Theorem 1. Suppose $A \in X_{-\infty,t}$, $B \in X_{t+s,\infty}$, and $\tilde{B}$ is a set in $E_{t+1,\infty}$ such that (6) holds. The existence of $\tilde{B}$ is guaranteed by Lemma 3. Then,

$$|P(A \cap B) - P(A)P(B)|$$

$$\leq |P(A \cap B) - P(A \cap \tilde{B})| + |P(A \cap \tilde{B}) - P(A)P(\tilde{B})| + |P(A)P(\tilde{B}) - P(A)P(B)|$$

$$\leq 2 \cdot \sup_{A \in X_{-\infty,t}} |P(A \cap B) - P(A \cap \tilde{B})|$$

since $A \in X_{-\infty,t} \subset E_{-\infty,t}$ and $\tilde{B} \in E_{t+1,\infty}$ imply A and $\tilde{B}$ are independent,

$\leq \tilde{\alpha}(s)$ for all $s = 1, 2, \ldots,$ by Lemma 3, where $\tilde{\alpha}(s)$ is independent of $B \in X_{t+s,\infty}$ and of t .

Hence,

$$\alpha(s) \equiv \sup_t \sup_{A \in X_{-\infty,t}, B \in X_{t+s,\infty}} |P(A \cap B) - P(A)P(B)|$$

$$\leq \tilde{\alpha}(s)$$

for all $s = 1, 2, \ldots$. □

Proof of Lemma 1. We prove the Lemma by induction. Suppose (4) holds for $s = v$, we show (4) holds for $s = v+1$.

$$\sup_{D\in\mathcal{B}} \left|P\left(\sum_{\ell=0}^{v} \rho^\ell \varepsilon_\ell \in D-y\right) - P\left(\sum_{\ell=0}^{v} \rho^\ell \varepsilon_\ell \in D\right)\right|$$

$$= \sup_{D\in\mathcal{B}} \left|\int_R [P(V+\tilde{e} \in D-y) - P(V+\tilde{e} \in D)]dP_{\tilde{E}}(\tilde{e})\right|$$

where $V \equiv \sum_{\ell=0}^{v-1} \rho^\ell \varepsilon_\ell$ and $\tilde{E} \equiv \rho^v \varepsilon_v$,

$$\leq \int_R \sup_{D\in\mathcal{B}} \left|P(V \in D-\tilde{e}-y) - P(V \in D-\tilde{e})\right| dP_{\tilde{E}}(\tilde{e})$$

$$= \int_R \sup_{D\in\mathcal{B}} \left|P(V \in D-y) - P(V \in D)\right| dP_{\tilde{E}}(\tilde{e})$$

$$= \sup_{D\in\mathcal{B}} \left|P(V \in D-y) - P(V \in D)\right|$$

since the integrand above is independent of $\tilde{e}$,

$\leq C\cdot|y|^q$ for all $|y| \leq 1$, since (4) is assumed to hold for $s = v$.

By assumption of the Lemma, (4) holds for $s = s_0$. Hence, by induction, (4) holds for all integers $s \geq s_0$. □

Proof of Lemma 2. Any set B in $\mathcal{X}'_{t+s,\infty}$ can be written in the form

$$B = \bigcup_{j=1}^{J} \bigcap_{k=1}^{K} (X_{t+s+i_{jk}} \in B_{jk})$$

where i_{jk} is a non-negative integer for all j , k , B_{jk} is a Borel set in R for all j , k , and J , K are finite. For notational simplicity let $r = t+s$, $v = t+s+i_{jk}$, and $i = i_{jk}$, for $j = 1, \ldots, J$ and $k = 1, \ldots, K$. (Note that v and i vary as j and k vary.) Also, let $\bigcup_j \bigcap_k$ denote $\bigcup_{j=1}^{J} \bigcap_{k=1}^{K}$.

By repeated substitution in (3), X_v can be written as

$$X_v = \sum_{\ell=0}^{\infty} \rho^{\ell} \varepsilon_{v-\ell} \ .$$

Define

$$\tilde{X}_{t,v} \equiv \sum_{\ell=0}^{v-t-1} \rho^{\ell} \varepsilon_{v-\ell} \in E_{t+1,v} \ ,$$

and

$$Y_{t,v} \equiv \sum_{\ell=v-t}^{\infty} \rho^{\ell} \varepsilon_{v-\ell} = \rho^{v-t} X_t \in X_{-\infty,t} \subset E_{-\infty,t} \quad \text{for all} \quad i \ .$$

Note,

$$X_v = \tilde{X}_{t,v} + Y_{t,v} \ .$$

Define

$$\tilde{B} \equiv \bigcup_j \bigcap_k (\tilde{X}_{t,v} \in B_{jk}) \in E_{t+1,\infty}$$

where B_{jk} $\forall j,k$ are as in the definition of B .

X_v and $\tilde{X}_{t,v}$ can be written in terms of X_r and $\tilde{X}_{t,r}$, respectively, and $\varepsilon_{r+1}, \varepsilon_{r+2}, \ldots, \varepsilon_v$:

$$X_v = \sum_{\ell=0}^{i-1} \rho^{\ell} \varepsilon_{v-\ell} + \sum_{\ell=i}^{\infty} \rho^{\ell} \varepsilon_{v-\ell} = W_{r,i} + \rho^{i} X_r \ ,$$

and $$\tilde{X}_{t,v} = \sum_{\ell=0}^{i-1} \rho^{\ell} \varepsilon_{v-\ell} + \sum_{\ell=i}^{v-t-1} \rho^{\ell} \varepsilon_{v-\ell} = W_{r,i} + \rho^{i} \tilde{X}_{t,r} \ ,$$

where $$W_{r,i} \equiv \sum_{\ell=0}^{i-1} \rho^{\ell} \varepsilon_{v-\ell} \in E_{t+s+1,\infty} \ .$$

Now, for any set $A \in X_{-\infty,t}$

$$|P(A\cap B) - P(A\cap\widetilde{B})|$$

$$= \left|P\left(A \cap [\bigcup_j \bigcap_k (W_{r,i} + \rho^i X_r \in B_{jk})]\right) - P\left(A \cap [\bigcup_j \bigcap_k (W_{r,i} + \rho^i \widetilde{X}_{t,r} \in B_{jk})]\right)\right|$$

$$= \left|\int_{R^{JK}} \cdots \int \left[P\left(A \cap [\bigcup_j \bigcap_k (w_{r,i} + \rho^i X_r \in B_{jk})]\right) - P\left(A \cap [\bigcup_j \bigcap_k (w_{r,i} + \rho^i \widetilde{X}_{t,r} \in B_{jk})]\right)\right] dP_{\underset{\sim}{W}}(\underset{\sim}{w})\right|$$

where $\underset{\sim}{W} = (W_{r,i_{11}}, \ldots, W_{r,i_{JK}})$ and $\underset{\sim}{w} = (w_{r,i_{11}}, \ldots, w_{r,i_{JK}})$; since X_r and $\widetilde{X}_{r,s}$ are independent of $\underset{\sim}{W}$ by the independence of $E_{-\infty,r}$ and $E_{r+1,\infty}$, and of $E_{t+1,r}$ and $E_{r+1,\infty}$,

$$= \left|\int_{R^{JK}} \cdots \int \left[P(A \cap [X_r \in \frac{1}{\rho^i}(\bigcup_j \bigcap_k B_{jk} - w_{r,i})]) - P(A \cap [\widetilde{X}_{t,r} \in \frac{1}{\rho^i}(\bigcup_j \bigcap_k B_{jk} - w_{r,i})])\right] dP_{\underset{\sim}{W}}(\underset{\sim}{w})\right|$$

$$\leq \int_{R^{JK}} \cdots \int \sup_{D \in \mathcal{B}} |P(A \cap [X_r \in D]) - P(A \cap [\widetilde{X}_{t,r} \in D])| \, dP_{\underset{\sim}{W}}(\underset{\sim}{w})$$

where $\mathcal{B}$ is the Borel σ-algebra on R ,

$$= \sup_{D \in \mathcal{B}} |P(A \cap [\widetilde{X}_{t,r} + \rho^s X_t \in D]) - P(A \cap [\widetilde{X}_{t,r} \in D])|$$

$$\leq \int_R \sup_{D \in \mathcal{B}} \left|P(A \cap [\widetilde{X}_{t,r} + \rho^s x_t \in D] | X_t = x_t) - P(A \cap [\widetilde{X}_{t,r} \in D] | X_t = x_t)\right| dP_{X_t}(x_t)$$

$$= \int_R \sup_{D \in \mathcal{B}} |P(A | X_t = x_t) \cdot [P(\widetilde{X}_{t,r} + \rho^s x_t \in D) - P(\widetilde{X}_{t,r} \in D)]| \, dP_{X_t}(x_t)$$

since 1) conditional on $X_t = x_t$, A and $\widetilde{X}_{t,r} + \rho^s X_t$ are independent, and A and $\widetilde{X}_{t,r}$ are independent, and 2) $\widetilde{X}_{t,r}$ and X_t are independent,

$$\leq \int_R \sup_{D \in \mathcal{B}} |P(\widetilde{X}_{t,r} \in D - \rho^s x_t) - P(\widetilde{X}_{t,r} \in D)| \, dP_{X_t}(x_t)$$

$$\le \int_{[|\rho^s x_t| \ge 1]} dP_{X_t}(x_t) + \int_{[|\rho^s x_t|<1]} \sup_{D\in\mathcal{B}} |P(\tilde{X}_{t,r} \in D - \rho^s x_t) - P(\tilde{X}_{t,r} \in D)| dP_{X_t}(x_t)$$

$$= P(|\rho^s X_t|^\nu \ge 1) + \int_{[|\rho^s x_t|<1]} \sup_{D\in\mathcal{B}} |P(\sum_{\ell=0}^{s-1} \rho^\ell \varepsilon_\ell \in D - \rho^s x_t) - P(\sum_{\ell=0}^{s-1} \rho^\ell \varepsilon_\ell \in D)| dP_{X_t}(x_t)$$

by the stationarity of $\langle\varepsilon_t\rangle$, where $\nu = \min\{p,q,1\}$,

$$\le |\rho^\nu|^s E|X_t|^\nu + \int_{[|\rho^s x_t|<1]} C\cdot|\rho^s x_t|^\nu dP_{X_t}(x_t)$$

for all $s \ge s_0$, the first term by Markov's inequality, and the second by condition S_ρ,

$$\le (C+1)E|X_t|^\nu|\rho^\nu|^s .$$

Thus,

$$2\cdot \sup_{A\in X_{-\infty,t}} |P(A\cap B) - P(A\cap\tilde{B})| \le \begin{cases} 2(C+1)E|X_t|^\nu|\rho^\nu|^s & \text{for } s \ge s_0 \\ 1 & \text{for } s < s_0 , \end{cases}$$

$$\equiv \tilde{\alpha}(s) ,$$

where $\tilde{\alpha}(s)$, $s = 1, 2, \ldots$ are eventually exponentially declining. □

<u>Proof of Lemma 3</u>. Let

$$M_{t+s,\infty} = \{B\in X_{t+s,\infty} : \exists\tilde{B}\in E_{t+1,\infty} \text{ with } 2\cdot \sup_{A\in X_{-\infty,t}} |P(A\cap B) - P(A\cap\tilde{B})| \le \tilde{\alpha}(s)\} .$$

$X'_{t+s,\infty} \subset M_{t+s,\infty}$ by Lemma 2. It is easy to show $X'_{t+s,\infty}$ is an algebra and $M_{t+s,\infty}$ is a monotone class. Hence,

$$\sigma(X'_{t+s,\infty}) \subset M_{t+s,\infty} ,$$

where $\sigma(X'_{t+s,\infty})$ is the σ-algebra generated by $X'_{t+s,\infty}$. Since $\sigma(X'_{t+s,\infty}) = X_{t+s,\infty}$, this gives the desired result. □

3. Conditions on the Innovation Rv for Strong Mixing

In this section we give a number of alternative conditions which imply S_ρ and are easier to verify than S_ρ. In addition, numerous examples of distributions which satisfy these conditions are provided. First, however, the following lemma is used to show that the results of Section 2 only apply to AR(1) processes generated by non-atomic innovation rv's.

Lemma 4. Condition S_ρ is satisfied only if ε_0 is a non-atomic rv.

Proof of Lemma 4. If ε_0 has an atom of probability, then so does $H \equiv \sum_{\ell=0}^{s-1} \rho^\ell \varepsilon_\ell$, and

$$\Delta_V(H, H+y) \geq \sup_{a \in R} |P(H \leq a) - P(H < a-y)| \not\to 0 \quad \text{as} \quad y \downarrow 0 .$$

This contradicts S_ρ. □

Consider the case of an innovation rv ε_0 for which S_ρ holds with $s_0 = 1$. That is, $\varepsilon_0 \in L^p$ for some $p > 0$, and

$$\Delta_V(\varepsilon_0, \varepsilon_0 + y) \leq C|y|^q \tag{7}$$

for all $|y| \leq 1$, for some constants $q > 0$ and $C < \infty$. In this case S_ρ is independent of ρ and depends on the distribution of ε_0 in a very simple way. If ε_0 has a density f with respect to Lebesgue measure,

(7) becomes

$$\int |f(\omega) - f(\omega-y)| d\omega \leq C|y|^q . \tag{8}$$

(8) is implied by the Lipschitz-type condition

$$|f(\omega) - f(\omega-y)| \leq C|y|^q g(\omega) \tag{9}$$

for almost (Lebesgue) all $\omega \in \Omega_{\varepsilon_0} \cup \Omega_{\varepsilon_0+y}$, where Ω_{ε_0} and Ω_{ε_0+y} are the supports of ε_0 and ε_0+y , respectively, and where $g(\omega)$ is some integrable function (with respect to Lebesgue measure). Alternatively, with $q = 1$, (8) is equivalent to the following condition which is related to the differentiability of f :

$$\int \left| \frac{f(\omega) - f(\omega-y)}{y} \right| d\omega \leq C . \tag{10}$$

Any one of the conditions (7), (8), (9), and (10) plus $\varepsilon_0 \in L^p$ imply S^ρ holds.

Define the Kolmogorov distance between two rv's Y and Z as

$$\Delta_K(Y,Z) \equiv \sup_{a \in R} |F_Y(a) - F_Z(a)| \tag{11}$$

where $F_Y(\cdot)$ and $F_Z(\cdot)$ are the distribution functions of Y and Z , respectively. Then condition (7) can be altered by replacing $\Delta_V(\cdot,\cdot)$ by $\Delta_K(\cdot,\cdot)$ provided the density of ε_0 satisfies a certain monotonicity condition. For any two rv's Y and Z , $\Delta_K(Y,Z) \leq \Delta_V(Y,Z)$, so the replacement of Δ_V by Δ_K corresponds to a weakening of the inequality of (7). This is a useful alteration because the inequality with Δ_K is much easier to verify than that with Δ_V , and the monotonicity condition is satisfied by virtually all well-known distributions of continuous rv's.

Theorem 2. Suppose $\varepsilon_0 \in L^p$, for some $p > 0$, and ε_0 satisfies

(K1) $\Delta_K(\varepsilon_0, \varepsilon_0 + y) \leq C|y|^q$ for all $|y| \leq 1$, for some constants $q > 0$, and $C < \infty$, and

(K2) ε_0 has density f with respect to Lebesgue measure which is (non-strictly) monotone on each set in a partition of R into a finite number of (finite or infinite) intervals,

then (7) holds and S_ρ is satisfied.

Remarks: 1. The Kolmogorov distance between ε_0 and $\varepsilon_0 + y$ is simply

$$\Delta_K(\varepsilon_0, \varepsilon_0 + y) = \sup_{a \in R} \int_{[a-|y|,a]} f(\omega)d\omega .$$

2. K1 holds with $q = 1$ if f is bounded, since

$$\Delta_K(\varepsilon_0, \varepsilon_0 + y) = \sup_{a \in R} \int_{[a-|y|,a]} f(\omega)d\omega \leq f_B \cdot |y| ,$$

where f_B is the bound on the density of f.

3. In view of Remark 2 and easy verification of K2, the Theorem shows that innovation rv's with any of the following distributions generate strong mixing AR(1) processes: normal, exponential, uniform, Cauchy (which is in L^p for $0 < p < 1$), χ^2, t, F with parameters $\nu \geq 1$ and $\lambda > 0$, Laplace, extreme value, logistic, and Weibull with parameter $c \geq 1$. In addition, any truncated version of the above distributions satisfies K1 and K2.

Proof of Theorem 2. We will show (8) holds. Denote the intervals in the partition of R given in K2 by $I(a_j, b_j)$, $j = 1, \ldots, k_1$, where the interval endpoints a_j, b_j satisfy $-\infty \leq a_j \leq b_j \leq \infty$, and where f is non-decreasing on $I(a_j, b_j)$ for $j = 1, \ldots, k_0$ and is non-increasing on

$I(a_j, b_j)$ for $j = k_0+1, \ldots, k_1$. Then,

$$\sum_{j=1}^{k_1} \int_{I(a_j,b_j)} |f(\omega) - f(\omega-y)|d\omega$$

$$= \sum_{j=1}^{k_0} \left[\int_{I(a_j+y,b_j)} f(\omega)d\omega - \int_{I(a_j,b_j-y)} f(\omega)d\omega + \int_{I(a_j,a_j+y)} |f(\omega) - f(\omega-y)|d\omega \right]$$

$$+ \sum_{j=k_0+1}^{k_1} \left[\int_{I(a_j,b_j-y)} f(\omega)d\omega - \int_{I(a_j+y,b_j)} f(\omega)d\omega + \int_{I(a_j,a_j+y)} |f(\omega) - f(\omega-y)|d\omega \right]$$

by monotonicity of f on the relevant intervals and change of variables,

$$\leq \sum_{j=1}^{k_0} \left[\int_{I(b_j-y,b_j)} f(\omega)d\omega + \int_{I(a_j,a_j+y)} f(\omega)d\omega + \int_{I(a_j-y,a_j)} f(\omega)d\omega \right]$$

$$+ \sum_{j=k_0+1}^{k_1} \left[\int_{I(a_j,a_j+y)} f(\omega)d\omega + \int_{I(a_j,a_j+y)} f(\omega)d\omega + \int_{I(a_j-y,a_j)} f(\omega)d\omega \right]$$

$$\leq 3k_1 C|y|^q$$

by K1. Hence (8) holds. □

4. Non-Strong Mixing AR(1) Processes

Consider an AR(1) process with Bernoulli (p) innovations and autoregressive parameter $\rho \in (0, 1/2]$. Theorem 3 below uses this example to show that there exist non-strong mixing AR(1) processes, including ones with arbitrarily small autoregressive parameters. This implies condition S_ρ of Section 2 is not redundant. Theorem 3 is proved directly, rather than by contradiction. In consequence, the proof shows clearly how the discreteness of the innovation rv leads to a failure of the strong mixing condition.

Theorem 3. There exist AR(1) processes which are not strong mixing. In particular, AR(1) processes generated by a Bernoulli (p) innovation rv and an autoregressive parameter $\rho \in (0, 1/2]$ are not strong mixing.

Remarks: 1. The proof of Theorem 3 uses Lemmas 5 and 6, whose proofs follow those of the Theorem. The proof shows that for any period $t+s$ there is a set B (which depends on s) such that the probability that X_{t+s} falls in B is bounded away from one (independently of s); yet conditional on $X_t \in (0,\rho)$ the probability is one. This holds for all future time periods $t+s$, and implies that $\langle X_t \rangle$ is not strong mixing.

2. The condition $\rho \leq 1/2$ is necessary for Lemma 5 which is used in the proof of Theorem 3. One might conjecture that $\rho \leq 1/2$ is not a necessary condition for the result of Theorem 3.

Lemma 5. Suppose $\langle X_t \rangle$ is an AR(1) process generated by a Bernoulli (p) innovation rv ε_0 (with corresponding iid innovation process $\langle \varepsilon_t \rangle$) and an autoregressive parameter ρ , where $p \in (0,1)$ and $\rho \in (0, 1/2]$. Define $\hat{X}_{t,s} = \sum_{\ell=0}^{s-1} \rho^{\ell} \varepsilon_{t-\ell}$, and take W_s to be the support of $\hat{X}_{t,s}$. Then, the elements of W_s are at least of distance ρ^{s-1} apart, for all $s = 1, 2, \ldots$.

Lemma 6. For $\langle X_t \rangle$, p , and ρ as in Lemma 5,

$$P(X_t \in (0,\rho)) > 0 \ ,$$

and
$$P(X_t \in [\rho,1]) > 0 \ .$$

Proof of Theorem 3. Let $\langle X_t \rangle$, $\langle \varepsilon_t \rangle$, ρ , $\hat{X}_{t,s}$, and W_s be as in Lemma 5.

Write X_{t+s} as

$$X_{t+s} = \hat{X}_{t,s} + \rho^s X_t \ .$$

W_s can be written as

$$W_s = \bigcup_{j=1}^{J} \{w_j\}$$

where $J \leq 2^s$, and w_j , $j = 1, 2, \ldots$ are the values $\hat{X}_{t,s}$ takes on with positive probability.

Define $A = \{X_t \in A_1\} \in \mathcal{X}_{-\infty,t}$, $B = \{X_{t+s} \in B_1\} \in \mathcal{X}_{t+s,\infty}$, and $D = [\rho,1]$, where $A_1 \equiv (0,\rho)$, and $B_1 \equiv \sum_{j=1}^{J} (w_j, \ w_j + \rho^{s+1})$. Then,

$$\begin{aligned} \alpha(s) &\equiv \sup_{t'>0} \ \sup_{A'\in\mathcal{X}_{-\infty,t'},B'\in\mathcal{X}_{t'+s,\infty}} |P(A'\cap B') - P(A')P(B')| \\ &\geq |P(A\cap B) - P(A)P(B)| \ . \qquad (12) \end{aligned}$$

Now, for $x_t \in A_1 \equiv (0,\rho)$

$$\begin{aligned} B_1 - \rho^s x_t &\equiv \bigcup_{j=1}^{J} (w_j - \rho^s x_t, \ w_j + \rho^s(\rho - x_t)) \\ &\supset \bigcup_{j=1}^{J} \{w_j\} \ , \end{aligned}$$

and

$$P(\hat{X}_{t,s} \in B_1 - \rho^s x_t) \geq P(\hat{X}_{t,s} \in W_s) = 1 \tag{13}$$

Thus,

$$P(A \cap B) = P(X_t \in A_1 \text{ and } \hat{X}_{t,s} + \rho^s X_t \in B_1)$$

$$= \int_{A_1} P(\hat{X}_{t,s} + \rho^s x_t \in B_1) dP_{X_t}(x_t)$$

by independence of $\hat{X}_{t,s}$ and X_t ,

$$= \int_{A_1} dP_{X_t}(x_t) \quad \text{by (13),}$$

$$= P(A) \; . \tag{14}$$

Further, for $x_t \in D \equiv [\rho,1]$

$$B_1 - \rho^s x_t \equiv \bigcup_{j=1}^{J} (w_j - \rho^s x_t,\ w_j + \rho^s(\rho - x_t))$$

$$\subset \bigcup_{j=1}^{J} (w_j - \rho^s,\ w_j)$$

and

$$P(\hat{X}_{t,s} \in B_1 - \rho^s x_t) \leq P(\hat{X}_{t,s} \in \bigcup_{j=1}^{J} (w_j - \rho^s,\ w_j))$$

$$= 0 \tag{15}$$

by Lemma 5, since the values where $\hat{X}_{t,s}$ takes positive probability (viz., w_j , $j = 1, 2, \ldots$) are at least distance ρ^{s-1} apart. Hence,

$$P(B) = P(\hat{X}_{t,s} + \rho^s X_t \in B_1)$$

$$= \int_R P(\hat{X}_{t,s} + \rho^s x_t \in B_1) dP_{X_t}(x_t) \quad \text{by independence of } \hat{X}_{t,s} \text{ and } X_t ,$$

$$\leq \int_{R-D} dP_{X_t}(x_t) + \int_D P(\hat{X}_{t,s} \in B_1 - \rho^s x_t) dP_{X_t}(x_t)$$

$$= 1 - P(X_t \in D) \tag{16}$$

by (15).

Thus,

$$\alpha(s) \geq |P(A \cap B) - P(A)P(B)| \quad \text{by (12),}$$

$$\geq P(A) - P(A)[1 - P(X_t \in D)] \quad \text{by (14) and (16),}$$

$$= P(A)P(X_t \in D)$$

$$> 0$$

by Lemma 6. $P(A)P(X_t \in D)$ is independent of s, so

$$\alpha(s) \not\to 0 \quad \text{as} \quad s \to \infty ,$$

and $\langle X_t \rangle$ is not strong mixing. □

<u>Proof of Lemma 5</u>. The Lemma is proved by induction. For sets G, G_1, and G_2 in R, let

$$d(G) = \inf\{|g_0 - g_1| : g_0, g_1 \in G\}$$

and

$$d(G_1, G_2) = \inf\{|g_1 - g_2| : g_1 \in G_1, g_2 \in G_2\} .$$

For $s = 1$, $W_s = \{0,1\}$, and $d(W_s) = 1 \geq \rho^{s-1}$. Hence, the result of the Lemma holds for $s = 1$.

Suppose

$$d(W_s) \geq \rho^{s-1} \tag{17}$$

for $s = v$. We show (17) holds for $s = v+1$.

Now,

$$W_{v+1} = W_v \cup (W_v + \rho^v)$$

where $W_v + \rho^v \equiv \{w + \rho^v : w \in W_v\}$, and

$$\begin{aligned} d(W_{v+1}) &= d(W_v) \wedge d(W_v + \rho^v) \wedge d(W_v, W_v + \rho^v) \\ &= \rho^{v-1} \wedge \rho^{v-1} \wedge [\rho^v \wedge (\rho^{v-1} - \rho^v)] \text{ since (17) holds for } s = v , \\ &\geq \rho^v \end{aligned}$$

since $\rho^{v-1} - \rho^v \geq \rho^v$ for $\rho \in (0, 1/2]$. □

Proof of Lemma 6. $\sum_{\ell=0}^{2} \varepsilon_{t-\ell} = 0$ and $\varepsilon_{t-3} = 1$ imply

$$\begin{aligned} 0 < X_t \leq \sum_{\ell=3}^{\infty} \rho^{\ell} &= \frac{\rho^3}{1-\rho} \\ &< \rho \quad \text{for all } \rho \in (0, 1/2] . \end{aligned}$$

So,

$$0 < (1-p)^3 p = P(\sum_{\ell=0}^{2} \varepsilon_{t-\ell} = 0, \varepsilon_3 = 1) \leq P(X_t \in (0,\rho)) ,$$

as desired.

Also, $\varepsilon_t = 0$ and $\varepsilon_{t-1} = 1$ imply

$$\rho \leq X_t \leq \sum_{\ell=1}^{\infty} \rho^{\ell} = \frac{\rho}{1-\rho} \leq 1 \quad \text{for} \quad \rho \in (0, 1/2] \ .$$

Thus,

$$0 < (1-p)p = P(\varepsilon_t = 0, \ \varepsilon_{t-1} = 1) \leq P(X_t \in [\rho,1]) \ . \quad \square$$

The result of Theorem 3 has already been shown for the case $\rho = 1/2$ by Ibragimov and Linnik (1971, p. 360). Theorem 3 gives a stronger, and more interesting, result since it applies no matter how small the autoregressive parameter is. Chernick (1981) implicitly shows AR(1) processes with autoregressive parameter $\rho = 1/r$ for some integer r and innovation rv ε_0 where $P(\varepsilon_0 = k\rho) = \rho$ for $k = 0, 1, \ldots, r-1$, are not strong mixing. However, his results and those of Ibragimov and Linnik are all proved by contradiction. Theorem 3 is proved directly, and hence, gives greater insight as to why the strong mixing condition fails.

5. Acknowledgments

I would like to thank David Pollard for reading a draft of this paper and suggesting improvements.

REFERENCES

Andrews, D. W. K. (1982a). A Model for Robustness against Distributional Shape and Dependence over Time, Ph.D. Dissertation, Univ. of California, Berkeley.

Andrews, D. W. K. (1982b). Generalized empirical processes of dependent random variables with discontinuous marginals, submitted to Ann. Probability.

Andrews, D. W. K. (1982c). Robust and asymptotically efficient estimation of location in a stationary strong mixing Gaussian parametric model, submitted to Ann. Statist.

Billingsley, P. (1979). Probability and Measure. Wiley, New York.

Box, G. E. P. and G. M. Jenkins (1969). Time Series Analysis: Forecasting and Control. Holden-Day, San Fransisco.

Chernick, M. R. (1981). A limit theorem for the maximum of autoregressive processes with uniform marginal distributions, Ann. Probability 9 145-149.

Deo, C. M. (1973). A note on empirical processes of strong-mixing sequences, Ann. Probability 1 870-875.

Deo, C. M. (1975a). On empirical processes with random sample sizes, Sankhya A 37 121-128.

Deo, C. M. (1975b). Weak convergence of stochastic processes related to the absolute extremes of a Gaussian sequence, Sankhya A 37 334-344.

Durbin, J. and G. S. Watson (1950). Testing for serial correlation in least squares regression, Biometrica 37 409-428.

Durbin, J. and G. S. Watson (1951). Testing for serial correlation in least squares regression, Biometrica 38 159-178.

Gaver, D. P. and P. A. W. Lewis (1980). First order autoregressive gamma sequences and point processes, Adv. Appl. Prob. 12 727-745.

Hannan, E. J. (1970). Multiple Time Series. Wiley, New York.

Ibragimov, I. A. (1962). Some limit theorems for stationary sequences, Theor. Probability Appl. 4 349-382.

Ibragimov, I. A. and Y. V. Linnik (1971). Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff Pub. Co., The Netherlands. Trans. by J. F. C. Kingman.

Kolmogorov, A. N. and Y. A. Rozanov (1960). On strong mixing conditions for stationary Gaussian processes, Theor. Probability Appl. 5 204-208.

Koul, H. L. (1977). Behavior of robust estimators in the regression model with dependent errors, Ann. Statist. 5 681-699.

Loynes, R. M. (1965). Extreme values in uniformly mixing stationary stochastic processes, Ann. Math. Statist. 36 993-999.

Mehra, K. L. and M. S. Rao (1975a). On functions of order statistics for mixing processes, Ann. Statist. 3 874-883.

Mehra, K. L. and M. S. Rao (1975b). Weak convergence of generalized empirical processes relative to d_q under strong mixing, Ann. Probability 3 979-991.

Oodaira, H. (1975). Some functional laws of the iterated logarithm for dependent random variables, Collo. Math. Soc. Janos Bolyai 11, 253-272.

Philipp, W. (1969). The central limit problem for mixing sequences of random variables, Z. Wahrscheinlichkeitstheorie verw. Geb. 12 155-171.

Philipp, W. (1977). Functional law of the iterated logarithm for empirical distribution functions of weakly dependent random variables, Ann. Probability 5 319-350.

Philipp, W. and L. Pinzur (1980). Almost sure approximation theorems for the multivariate empirical process, Z. Wahrscheinlichkeitstheorie verw. Geb. 54 1-13.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition, Proceedings of the National Academy of Sciences USA 42:1 43-47.

Rozanov, Y. A. (1963). Stationary Random Processes. Holden-Day, London.

Welsch, R. E. (1971). A weak convergence theorem for order statistics from strong-mixing processes, Ann. Math. Statist. 42 1637-1646.

Welsch, R. E. (1972). Limit laws for extreme order statistics from strong-mixing processes, Ann. Math. Statist. 43 439-446.

Withers, C. S. (1975). Convergence of empirical processes of mixing rv's on [0,1], Ann. Math. Statist. 3 1101-1108.

Yokoyama, R. (1973). Weak convergence of empirical processes for strong mixing sequences of random variables, Science Reports of the Tokyo Kyoiku Daigaku Sec. A, Vol. 12, No. 318, pp. 36-39.