The Annals of Probability
1975, Vol. 3, No. 5, 829-839

A MAXIMAL INEQUALITY AND DEPENDENT STRONG LAWS

BY D. L. MCLEISH
York University

This paper contains a general dependent extension of Doob's inequality for martingales, $E(\max_{i \leqq n} S_i^2) \leqq 4ES_n^2$. This inequality is then used to extend the martingale convergence theorem for L_2 bounded variables, and to prove strong laws under dependent assumptions. Strong and φ-mixing variables are shown to satisfy the conditions of these theorems and hence strong laws are proved as well for these.

0. Introduction. Among the most useful tools of probability theory are the many devices used to bound, in L_p norm or in probability, the quantity $\max_{i \leqq n} |S_i|$ where $S_n = \sum_{i=1}^n X_i$. When the X_i are independent random variables, many inequalities such as those of Kolmogorov, Ottaviani, and Bernstein are available; when S_n is a martingale, we have from Doob's inequality (Doob, page 317)

$$(0.1) \qquad ||\max_{i \leqq n} |S_i|\,||_p \leqq q||S_n||_p \qquad \text{for} \quad p > 1 , \qquad p^{-1} + q^{-1} = 1 ,$$

where $||\cdot||_p$ denotes the $L_p(\Omega)$ norm for random variables, and $||U||_\infty$ is defined as ess sup $|U|$.

In this paper we give a comparable result for dependent variables under conditional expectation assumptions. This is used in Section 1 to derive an analogue of the martingale convergence theorem and strong law.

It is well known (cf. Chung (1968), page 119) that when the X_i are independent random variables with $EX_i = 0$, $\sum_{i=1}^\infty E|X_i|^p/i^p < \infty$ for any $1 \leqq p \leqq 2$ is a sufficient condition for the strong law to hold: $S_n/n \to 0$ almost surely. Chow (1967) has proved a similar result when the X_i are martingale differences, and a number of authors, for example Cohn, and Iosifescu have studied the strong law under φ-mixing assumptions. A source for the latter results is the book by Iosifescu and Theodorescu (1969). Chow's martingale result is extended for $p = 2$ by (1.9), and in Section 2, the law of large numbers under mixing conditions weaker than both φ-mixing and strong mixing is investigated.

Finally in Section 3 we briefly treat functions of mixing processes, with an application to functions of the form $f_k(2^k\omega)$.

In none of these theorems is stationarity required. Applications are made to autoregressive time series.

1. Mixingales. Let $\{\mathscr{F}^n; -\infty \leqq n \leqq \infty\}$ be any sequence of subsigma algebras of the probability triple $(\Omega, \mathscr{F}, P)$ which are increasing in n. We will

Received October 29, 1973; revised February 26, 1975.
AMS 1970 *subject classifications.* Primary 60F15; Secondary 60G45.
Key words and phrases. Strong law, martingale convergence theorem, dependent variables, Doob's inequality, mixing.

represent the conditional expectation $E(U\mid\mathscr{F}^m)$ by $E_m U$ and for random variables $X_1, X_2, \cdots$ (specified later) put $S_n = \sum_{i=1}^n X_i$. Recall that $(S_n, \mathscr{F}^n)$ is a martingale if each S_n is integrable and;

(1.1) (a) $E_{n-1}X_n = 0$ a.s. and
(b) Each X_n is measurable $\mathscr{F}^n$.

For martingales, it is no restriction to assume that each of the $\mathscr{F}^n$ is complete, in which case (b) is equivalent to

(b′) $$\|E_n X_n - X_n\|_2 = 0 \quad \text{for all} \quad n\,.$$

(In the following we do not assume the $\mathscr{F}^n$ are complete; (b′) is introduced purely for motivational reasons). We now define the asymptotic analogues to (a) and (b′).

(1.2) DEFINITION. The sequence $(X_n, \mathscr{F}^n)$ is a *mixingale* if, for sequences of finite nonnegative constants c_n and ψ_m where $\psi_m \to 0$ as $m \to \infty$, we have for all $n \geqq 1$, $m \geqq 0$,

(a) $\|E_{n-m}X_n\|_2 \leqq \psi_m c_n$ and
(b) $\|X_n - E_{n+m}X_n\|_2 \leqq \psi_{m+1}c_n$.

For example if $\psi_m = 0$ for all $m \geqq 1$, then this coincides with the definition of an L_2 martingale. Normally c_n will be some measure of the relative size of the random variable X_n such as the norm $\|X_n\|_p$, and in general,

(1.3) $$\|X_n\|_2 \leqq \|E_n X_n\|_2 + \|X_n - E_n X_n\|_2 \leqq (\psi_0 + \psi_1)c_n\,.$$

For fixed n, Lemma 1, page 184 of Billingsley (1968) implies that the left-hand side of (1.2 b) is nonincreasing in m, and by the conditional Jensen's inequality, this holds as well for the left-hand side of (a). Therefore, we may assume ψ_m is nonincreasing, for otherwise we could replace ψ_m by $\psi_m' = \min_{n\leqq m}\psi_n$ in (1.2). By the same authority, $\|E_{-\infty}X_i\|_2 \leqq \psi_m c_i$ for all m, and hence is 0. Similarly, $X_i - E_{+\infty}X_i = 0$ a.s. for all i.

This, then, is the asymptotic analogue to martingales. Often each X_n is $\mathscr{F}^n$-measurable, in which case (1.2 b) is automatic. We will require a specific rate of convergence of ψ_m to 0, to which end we adopt the following definition:

(1.4) DEFINITION. we will call the sequence $\{\psi_k\}$ of size $-p$ if there exists a positive sequence $\{L(k)\}$ such that

(a) $\sum_n 1/nL(n) < \infty$
(b) $L_n - L_{n-1} = O(L(n)/n)$
(c) L_n is eventually nondecreasing and
(d) $\psi_n = O[1/n^{\frac{1}{2}}L(n)]^{2p}$.

REMARK. Observe that condition (b) will follow for any sequence L_n such that $L_n - L_{n-1}$ is regularly varying with exponent -1 (cf. Feller (1971), page 280). For example any sequence which is $O[n^{\frac{1}{2}}\log n(\log\log n)^{1+\delta}]^{-p}$ with $\delta > 0$ is of size $-p/2$. Also summability conditions such as $\sum_{n=1}^{\infty} \varphi_n{}^\theta < \infty$ imply, for monotone sequences φ_n, that $\varphi_n = o(1/n^{1/\theta})$ and hence that φ_n is of size $-q$ for any $q < 1/\theta$ (Knopp (1946), page 124). I am indebted to P. Billingsley for both the information and reference.

For the next lemma, define $Y_{j,k} = \sum_{i=1}^{j} (E_{i+k} X_i - E_{i+k-1} X_i)$ for each k, j. Observe that if the X_j are any square integrable random variables then $(Y_{j,k}, \mathscr{F}^{j+k})$, $j = 1, 2, \cdots$ is a square integrable martingale.

(1.5) LEMMA. *Let $\{X_n\}$ be any sequence of zero-mean, square integrable random variables, $\mathscr{F}^n$ any nondecreasing sequence of sigma algebras such that $E_{-\infty} X_i = X_i - E_{+\infty} X_i = 0$ a.s. for all i. Then the partial sums S_n have a representation as an infinite sum of square integrable martingales $S_n = \sum_{k=-\infty}^{\infty} Y_{n,k}$ a.s. for all n, and for any positive sequence $\{a_k\}$;*

$$E(\max_{m \leq n} S_m{}^2) \leqq 4(\sum_{i=-\infty}^{\infty} a_i)\{\sum_{k=-\infty}^{\infty} a_k{}^{-1} \operatorname{Var}(Y_{n,k})\} .$$

(A number of persons have pointed out since the submission of this paper that this representation is used elsewhere in the literature; in the stationary case it is apparently due to Statulevicius (1969) and Gordin (1969), but also appears in Heyde (1973), Scott (1973), Philipp and Stout (1974) in varying forms.)

PROOF. We show first that $X_i = \sum_{k=-\infty}^{\infty} (E_{i+k} X_i - E_{i+k-1} X_i)$ a.s. Now,

$$\sum_{k=-m}^{n} (E_{i+k} X_i - E_{i+k-1} X_i) = E_{i+n} X_i - E_{i-m-1} X_i .$$

The first term on the right-hand side forms, for fixed i, a martingale sequence, and hence, by Doob (1953), Theorem 4.3 (page 331) converges as $n \to \infty$ to $E(X_i \mid \mathscr{F}^\infty) = X_i$.

Similarly, the second term forms a backwards martingale sequence, and converges to 0 almost surely by the same authority. Therefore $S_j = \sum_{k=-\infty}^{\infty} Y_{j,k}$ a.s.

Observe that by Cauchy–Schwartz, for $j \geqq 1$,

$$S_j{}^2 = \sum_{k=-\infty}^{\infty} a_k{}^{\frac{1}{2}} \frac{Y_{j,k}}{a_k{}^{\frac{1}{2}}} \leqq (\sum_{k=-\infty}^{\infty} a_k)(\sum_{k=-\infty}^{\infty} Y_{j,k}^2/a_k) .$$

If we now take $\max_{j \leq n}$, then take expectations on both sides of this inequality, and finally apply Doob's inequality (Doob, page 317) to bound the right-hand side, we get the desired inequality.

We now prove a mixingale analogue of Doob's inequality.

(1.6) THEOREM. *Let $\{X_i\}$ be a mixingale such that the sequence $\{\psi_n\}$ is of size $-\frac{1}{2}$. Then there exists a finite constant K depending only on $\{\psi_n\}$ such that for all n, $E(\max_{i \leq n} S_i{}^2) \leqq K(\sum_{i=1}^{n} c_i{}^2)$.*

PROOF. Rearranging the terms in Lemma (1.5) yields

$$E(\max_{j \leq n} S_j{}^2) \leqq 4(\sum_{i=-\infty}^{\infty} a_i) \sum_{i=1}^{n} \left\{\frac{EZ_{i,0}^2}{a_1} + \sum_{k=1}^{\infty} EZ_{i,k}^2(a_{k+1}^{-1} - a_k{}^{-1}) + \frac{EE_i{}^2 X_i}{a_0} + \sum_{k=1}^{\infty} EE_{i-k}^2 X_i(a_k{}^{-1} - a_{k-1}^{-1})\right\}$$

if the $a_i = a_{-i}$ is a positive sequence, nonincreasing in $i \geqq 0$, and $Z_{i,k} = X_i - E_{i+k}X_i$. By (1.2), this is less than

$$(1.7) \qquad 4(\sum_{i=-\infty}^{\infty} a_i)(\sum_{i=1}^{n} c_i^2)\left\{\frac{\psi_0^2+\psi_1^2}{a_0} + 2\sum_{k=1}^{\infty}\psi_k^2(a_k^{-1}-a_{k-1}^{-1})\right\}.$$

For $i \geqq 1$, set $a_i = a_{-i} = \min_{j\leqq i} 1/jL(j)$ and $a_0 = a_1$ and observe that a_i is summable, nonincreasing, and for all sufficiently large i, $a_i = 1/iL(i)$ follows from (1.4a) and (1.4c). Therefore, as $k \to \infty$,

$$\begin{aligned} a_k^{-1} - a_{k-1}^{-1} &= L(k-1) + k[L(k) - L(k-1)] \\ &= O(L(k)) \quad \text{by (1.4b)}. \end{aligned}$$

Therefore $\psi_k^2(a_k^{-1} - a_{k-1}^{-1})$ is $O(1/kL(k))$ and by (1.4a) is therefore summable. □

As a consequence of the Cauchy criterion for almost sure convergence, we get the following analogue to the martingale convergence theorem for L_2 bounded random variables.

(1.8) COROLLARY. *Suppose $\{X_i\}$ is a mixingale such that $\{\psi_n\}$ is of size $-\frac{1}{2}$ and $\sum_{i=1}^{\infty} c_i^2 < \infty$. Then S_n converges* a.s.

This corollary, with Kronecker's lemma, leads to the following:

(1.9) COROLLARY. *Suppose $\{X_i\}$ is a mixingale with $\{\psi_n\}$ of size $-\frac{1}{2}$ and $\sum_{i=1}^{\infty} c_i^2/i^2 < \infty$. Then $S_n/n \to 0$* a.s.

Note also that if X_i is a square integrable sequence of martingale differences such that $\sum_i EX_i^2/i^2 < \infty$, then a strong law follows from Corollary (1.9). This result is included in Chow (1967).

EXAMPLE 1. Let $\{\xi_i; i = \cdots -1, 0, 1, 2, \cdots\}$ be a sequence of martingale differences: i.e. setting $\mathscr{F}_m^n = \sigma(\xi_m, \xi_{m+1}, \cdots, \xi_n)$ for $n \geqq m$, assume that $E(\xi_n \mid \mathscr{F}_{-\infty}^{n-1}) = 0$ a.s. and $E\xi_n^2 = 1$. For any doubly infinite array of constants, $\{d_{n,i}\}$ put $X_n = \sum_{i=-\infty}^{\infty} d_{n,i}\xi_i$. We will assume $\sum_{i=-\infty}^{\infty} d_{n,i}^2 < \infty$ for all n, whence X_n is well defined (almost surely) by the martingale (letting the upper limit of the summation approach ∞) and the reverse martingale (letting the lower limit approach $-\infty$) convergence theorems. Then $(X_n, \mathscr{F}_{-\infty}^n)$ is a mixingale with $c_m = 1$ for all m and $\psi_m^2 = \sup_n \sum_{\{i:|i-n|>m\}} d_{n,i}^2$, and so if ψ_m is of size $-\frac{1}{2}$, $1/n \sum_{i=1}^{n} X_i \to_{\text{a.s.}} 0$ follows from Corollary (1.9).

Frequently the sequence $d_{n,i}$ above is of the form f_{i-n} for some square summable sequence f_i, in which case we need only know that the tails $\sum_{|i|>m} f_i^2 = O(1/mL^2(m))$ as $m \to \infty$. This is true if $\{f_i\}$ and $\{f_{-i}\}$ are of size -1 with the function L slowly varying at ∞, for by Feller (1971), 9.5, page 281,

$$\sum_{i=m}^{\infty} \frac{1}{i^2L^2(i)} \sim \frac{1}{mL^2(m)} \qquad \text{as } m \to \infty.$$

EXAMPLE 2. Let X_n be an autoregressive series of the form $X_n = \sum_{i=1}^{\infty} d_i X_{n-i} + Y_n$. Put $\mathscr{F}_n^m = \sigma(X_n, X_{n+1}, \cdots, X_m)$ and suppose Y_n satisfies $E(Y_n \mid \mathscr{F}_{-\infty}^{n-1}) = 0$ a.s. for all n. Suppose also that $\psi_0 = \sup_n \|X_n\|_2 < \infty$, $\sum_{i=1}^{\infty} |d_i| < 1$, $\{d_n\}$ is of size $-\frac{3}{2}$, and the sequence L_n in (1.4) is slowly varying at ∞. Then $S_n/n \to 0$ a.s. as $n \to \infty$.

PROOF. Clearly,

$$\|E(X_n \mid \mathscr{F}_{-\infty}^{n-m})\|_2 \leqq \sum_{i=m+1}^{\infty} |d_i|\,\|X_{n-i}\|_2 + \sum_{i=1}^{m} |d_i|\,\|E(X_{n-i} \mid \mathscr{F}_{-\infty}^{n-m})\|_2$$

for $m \geqq 1$. Taking the supremum over n on both sides

$$\psi_m \leqq \psi_0 \sum_{i=m+1}^{\infty} |d_i| + \sum_{i=1}^{m} |d_i|\psi_{m-i} \qquad \text{for} \quad m \geqq 1\,. \tag{1.10}$$

First we choose a constant C_1 such that $\limsup_{n\to\infty} |d_n| n^{\frac{3}{2}} L(n) < C_1$. Now define for each n, $f_n = \max\{|d_n|, C_1/n^{\frac{3}{2}}L(n)\}$ if $\sum_{j=n}^{\infty} C_1/j^{\frac{3}{2}}L(j) < 1 - \sum_{i=1}^{\infty} |d_i|$ and otherwise $f_n = |d_n|$. Observe that

$$|d_n| \leqq f_n \quad \text{for all} \quad n \qquad \text{and} \qquad \sum_{i=1}^{\infty} f_i < 1\,. \tag{1.11}$$

We also define recursively $v_0 = \psi_0$ and

$$v_n = \psi_0\left(\sum_{i=n+1}^{\infty} f_i\right) + \sum_{i=1}^{n} f_i v_{n-i} \qquad \text{for} \quad n \geqq 1\,. \tag{1.12}$$

Now (1.12) is the renewal equation (cf. Feller (1957), page 290) and therefore if we put $F(s) = \sum_{i=1}^{\infty} f_i s^i$, $V(s) = \sum_{i=0}^{\infty} v_i s^i$,

$$\begin{aligned} b_n &= \psi_0\,, & n &= 0 \\ &= \psi_0 \sum_{i=n+1}^{\infty} f_i & \text{for} \quad n &\geqq 1 \end{aligned}$$

and

$$B(s) = \sum_{i=0}^{\infty} b_i s^i = \psi_0(1 - F(1)) + \psi_0\left(\frac{F(1) - F(s)}{1 - s}\right)$$

then the equation has a solution in terms of generating functions;

$$V(s) = \frac{B(s)}{1 - F(s)}\,. \tag{1.13}$$

Since $f_n \sim C_1/n^{\frac{3}{2}}L(n)$ it follows from Theorem 5, page 447 of Feller (1971) that

$$B(s) \sim \frac{\text{constant}}{(1-s)^{\frac{1}{2}}L(1/(1-s))} \qquad \text{as} \quad s \uparrow 1\,.$$

But this, (1.13) and the fact that $F(1) < 1$ imply that

$$V(s) \sim \frac{\text{constant}}{(1-s)^{\frac{1}{2}}L(1/(1-s))} \qquad \text{as} \quad s \uparrow 1\,. \tag{1.14}$$

But it is easily shown from (1.12) that $v_m - v_{m+1} = \sum_{i=1}^{m} f_i(v_{m-i} - v_{m+1-i})$ for $m \geqq 1$, and $v_0 \geqq v_1$. Therefore, by induction, v_m is monotonic and, again by Theorem 5, page 447 of Feller (1971), and (1.14), $v_n \sim \text{constant}/n^{\frac{1}{2}}L(n)$. But (1.10), (1.11) and (1.12) imply $\psi_m \leqq v_m$ for all m. Therefore ψ_m is of size $-\frac{1}{2}$ and by Corollary 1.9 (with $c_i = 1$ for all i), $S_n/n \to_{\text{a.s.}} 0$. □

The conditions of Example 2 are trivially satisfied if d_i decreases exponentially; e.g., $d_i = ar^i$ for $0 < r < 1$ and $a < (1-r)/r$, or if $d_i = 0$ for $i \geqq n_0$, say and $\sum_{i=1}^{n_0} |d_i| < 1$. MacQueen (1973) considers this case (with only finitely many d_i nonzero) and under additional restrictions, proves a similar result when the sum $\sum_{i=1}^{n_0} |d_i|$ is allowed to be equal to 1.

2. Mixing. We now apply the concept of mixingale to prove strong limit theorems under strong and φ-mixing conditions. Define two measures of dependence between sigma algebras $\mathscr{F}$ and $\mathscr{A}$ by

$$\varphi(\mathscr{F}, \mathscr{A}) = \sup_{\{F \in \mathscr{F}, G \in \mathscr{A}: P(F)>0\}} |P(G \mid F) - P(G)| \qquad \text{and}$$
$$\alpha(\mathscr{F}, \mathscr{A}) = \sup_{F \in \mathscr{F}, G \in \mathscr{A}} |P(FG) - P(F)P(G)| .$$

The following lemma relates the concept of mixing to that of a mixingale. (2.2) is due to Serfling (1968).

(2.1) LEMMA. *Suppose X is a random variable measurable with respect to $\mathscr{A}$, and $1 \leqq p \leqq r \leqq \infty$. Then*

$$\|E(X \mid \mathscr{F}) - EX\|_p \leqq 2\{\varphi(\mathscr{F}, \mathscr{A})\}^{1-1/r}\|X\|_r \qquad \text{and} \tag{2.2}$$

$$\|E(X \mid \mathscr{F}) - EX\|_p \leqq 2(2^{1/p} + 1)\{\alpha(\mathscr{F}, \mathscr{A})\}^{1/p-1/r}\|X\|_r . \tag{2.3}$$

PROOF. For (2.3) put $\alpha = \alpha(\mathscr{F}, \mathscr{A})$, $c = \alpha^{-1/r}\|X\|_r$, and $X_1 = XI(|X| \leqq c)$, where $I(A)$ is the indicator function of the set A, and $X_2 = X - X_1$. Here we have neglected the trivial independent case and assumed $\alpha > 0$. Then

$$\begin{aligned} \|E(X \mid \mathscr{F}) - EX\|_p &\leqq \|E(X_1 \mid \mathscr{F}) - EX_1\|_p + \|E(X_2 \mid \mathscr{F}) - EX_2\|_p \\ &\leqq (2c)^{(p-1)/p}E^{1/p}|E(X_1 \mid \mathscr{F}) - EX_1| + 2\|X_2\|_p \\ &\leqq (2c)^{(p-1)/p}(4\alpha c)^{1/p} + 2\,\frac{\|X_2\|_r^{\,r/p}}{c^{(r-p)/p}} , \end{aligned}$$

where the first term in the last step follows from Lemma (5.2) of Dvoretsky (1972), and the second from the standard inequality

$$E|X|^pI(|X| > c) \leqq \frac{1}{c^{r-p}}\, E|X|^rI(|X| > c) .$$

Substituting for c and using the fact that $\|X_2\|_r \leqq \|X\|_r$, this bound becomes

$$2(2^{1/p} + 1)\alpha^{1/p-1/r}\|X\|_r . \qquad \square$$

In this section we consider a doubly infinite sequence of random variables $\{X_i; -\infty < i < \infty\}$ defined on $(\Omega, \mathscr{F}, P)$ and put $\mathscr{F}_n{}^m = \sigma(X_i; n \leqq i \leqq m)$. (A one-sided sequence may be handled within this framework by defining $X_j = 0$ for all negative j.) Define

$$\varphi_m = \sup_n \varphi(\mathscr{F}_{-\infty}^{\,n}, \mathscr{F}_{n+m}^{\,n+m}) \qquad \text{and} \tag{2.4}$$

$$\alpha_m = \sup_n \alpha(\mathscr{F}_{-\infty}^{\,n}, \mathscr{F}_{n+m}^{\,n+m}) . \tag{2.5}$$

We usually assume these quantities to converge to 0 at a specific rate. Observe that $\varphi_m \to 0$ is a weakening of the φ-mixing condition; it would be equivalent if $\mathscr{F}_{n+m}^{\,n+m}$ were replaced by $\mathscr{F}_{n+m}^{\,\infty}$ in (2.4), and a number of strong limit theorems under conditions on φ_m so defined are found in Iosifescu and Theodorescu (1971). In this case the existence of a zero-one law facilitates the proof of such results. In the same way, $\alpha_m \to 0$ is a weakening of the strong mixing condition introduced by Rosenblatt (1956). Blum, Hanson, and Koopmans (1963) have used $\mathscr{F}_{n+m}^{n+m}$ as we do, but they use the stronger dependence coefficient $\psi_m = \sup |P(AB)/P(A)P(B) - 1|$ where the supremum is taken over all $A \in \mathscr{F}_{-\infty}^{n}$, $B \in \mathscr{F}_{n+m}^{n+m}$ such that $P(A)$ and $P(B) > 0$, and over all n. G. O'Brien (1975), has shown that any decreasing sequence with limit 0 is possible for either α_m or φ_m, and a minor extension of his example verifies that our mixing conditions so defined are weaker than the usual ones (as well as being easier to verify). Clearly $\psi_m \geqq \varphi_m \geqq \alpha_m$ and for this reason we restrict ourselves to the two weaker conditions φ_m or $\alpha_m \to 0$.

(2.6) REMARK. We will assume throughout the remainder of this section that r is a number chosen (we assume such a number exists) with $2 \leqq r \leqq \infty$ such that either;

(a) $\{\varphi_n\}$ is of size $-r/(2r-2)$ or
(b) $r > 2$ and $\{\alpha_n\}$ is of size $-r/(r-2)$.

The following theorem is a dependent analogue of the sufficiency half of the three series theorem.

(2.7) THEOREM. *Suppose there exists a sequence of constants* $0 \leqq d_i \leqq \infty$ *such that, defining* $\bar{X}_i = X_i I[|X_i| \leqq d_i]$ *the series;*

(a) $\sum_i P(|X_i| > d_i)$
(b) $\sum_i ||\bar{X}_i||_r^2$

converge. Then $\sum_i (X_i - E\bar{X}_i)$ *converges* a.s.

PROOF. By (a) it suffices to show

$$\sum_{i=1}^{n} (\bar{X}_i - E\bar{X}_i) \qquad \text{converges as} \quad n \to \infty . \tag{2.8}$$

But,

$$||E_{i-m}(\bar{X}_i - E\bar{X}_i)||_2 \leqq 2\varphi_m^{1-1/r}||\bar{X}_i||_r \qquad \text{by (2.2), or}$$
$$\leqq 5\alpha_m^{\frac{1}{2}-1/r}||\bar{X}_i||_r \qquad \text{by (2.3).}$$

Therefore $\bar{X}_i - E\bar{X}_i$ is a mixingale with $\{\psi_m\}$ of size $-\frac{1}{2}$ and $c_i = ||\bar{X}_i||_r$. Therefore, applying Corollary (1.8) to the sequence proves (2.8). □

We now extend some of the strong laws known in the independent case to mixing variables.

Let $g_n(x)$ be a sequence of nonnegative, measurable functions on $[0, \infty)$ such that $g_n(0) = 0$ for all n and for a sequence of positive numbers $\{d_n\}$;

(a) $\inf_n \inf_{x \leqq d_n} g_n(x)/x^r > 0$ and
(b) $\inf_n \inf_{x > d_n} g_n(x) > 0$.

The following lemma parallels Loève, 16.4(b).

(2.9) LEMMA. *Suppose* $\sum_{n=1}^{\infty} E^{2/r} g_n(|X_n|) < \infty$. *Then* $\sum_n (X_n - E\bar{X}_n)$ *converges* a.s.

PROOF. Using (a) and (b) above, one can find positive numbers K_1, K_2

independent of n, for which

$$E^{2/r}g_n(|X_n|) \geqq \{K_1 P[|X_n| > d_n] + K_2 \textstyle\int |\bar{X}_n|^r\}^{2/r},$$

and therefore the two series in (2.7) converge.

If we now apply Lemma (2.9) to the sequence $Y_n = X_n/n$, $g_n(x) = x^p$ for some $1 \leqq p \leqq r$ and $d_n = 1$ for all n, we obtain the following:

(2.10) THEOREM. *Suppose for some p with $r/2 < p \leqq r < \infty$ one has*

$$\sum_{n=1}^{\infty} E^{2/r} \frac{|X_n|^p}{n^p} < \infty, \tag{2.11}$$

and that each X_n has 0 *mean. Then $\sum_n X_n/n$ converges* a.s. *and $S_n/n \to 0$* a.s.

When the r.v.'s are independent, we may take $r = 2$ and obtain the theorems of Kolmogorov ($p = 2$), and Marcinkiewicz and Zygmund ($p < 2$; cf. Chung page 119).

If (2.11) holds for any p, r such that $1 \leqq p \leqq r/2$, the referee has pointed out that the strong law follows from Kronecker's lemma and the monotone convergence theorem without any restriction on the dependence. Moreover, the condition:

$$\sup_i E|X_i|^{r/2+\delta} < \infty \qquad \text{for any positive} \quad \delta \tag{2.12}$$

is sufficient for (2.11). Again in the independent case (letting $r = 2$), this reduces to Markov's theorem, but as the degree of dependence increases ($r \uparrow$), higher moments are required to be uniformly bounded. The idea here is that we may trade off some of the independence in exchange for knowledge of higher moments.

The following theorem is analogous to an i.i.d. Strong Law, and to Loève's 3°, page 242.

(2.13) THEOREM. *Suppose $r = 2$, $\int_0^\infty \sup_n P[|X_n| > x]\,dx < \infty$, and $E(X_n) = 0$ for all n. Then $S_n/n \to 0$* a.s.

PROOF. Put $q(x) = \sup_n P[|X_n| > x]$; we apply Lemma (2.9) to the sequence $Y_n = X_n/n$ with $g_n(x) = \min(x^2, 1)$ and $d_n = 1$ for all n. Then $\sum_{n=1}^\infty Eg_n(|Y_n|) \leqq \sum_{n=1}^\infty 2\int_0^1 q(nx)x\,dx \leqq 2\int_0^\infty q(x)\,dx$, by the moments Lemma, Loève, page 242. It follows that

$$\frac{1}{n}\sum_{j=1}^n X_j - E\tilde{X}_j \to 0 \quad \text{a.s.},$$

where $\tilde{X}_j = (X_j \vee -j) \wedge j$. But $|E\tilde{X}_j| \leqq \int_j^\infty q(x)\,dx \to 0$ as $j \to \infty$.

3. Functions of mixing processes. Suppose $\{\xi_n; -\infty < n < \infty\}$ is a strong or φ-mixing sequence (for example an appropriate Markov chain). In this section we consider random variables defined by $X_n = f_n\{\xi_j; -\infty < j < \infty\}$ where f_n is a nonrandom function of the whole history, past and future, of the process. Such variables will be called "functions of mixing processes" (cf. Billingsley, Section 21). Our restriction on f_n will be that f_n is "almost" a function only of a neighborhood of the present epoque $(\xi_{n-m}, \xi_{n-m+1}, \cdots, \xi_{n+m})$, that is, the additional knowledge of the distant future and past of the process will have only a marginal effect on an approximation, based on the present, of X_n. We now lend some rigor to this condition.

Put

$$\mathscr{F}_n{}^m = \sigma(\xi_n, \cdots, \xi_m) \qquad \text{for } m \geqq n .$$
$$\varphi_m = \sup_n \varphi(\mathscr{F}_{-\infty}^n, \mathscr{F}_{n+m}^\infty) , \qquad or$$
$$\alpha_m = \sup_n \alpha(\mathscr{F}_{-\infty}^n, \mathscr{F}_{n+m}^\infty) , \qquad \text{and}$$
$$\nu_m = \sup_i ||E(X_i \mid \mathscr{F}_{i-m}^{i+m}) - X_i||_2 \qquad \text{for all } m .$$

(3.1) THEOREM. *Suppose* $EX_i = 0$ *for all* i, $\{\nu_n\}$ *is of size* $-\frac{1}{2}$, *and* $\sum_i ||X_i||_r{}^2/i^2 < \infty$ *for some* r *with* $2 \leqq r \leqq \infty$. *Moreover, assume either*

(a) $\{\varphi_n\}$ *is of size* $-r/(2r-2)$ *or*

(b) $\{\alpha_n\}$ *is of size* $-r/(r-2)$ *with* $r > 2$.

Then $S_n/n \to 0$ a.s.

PROOF. Clearly

$$\begin{aligned} ||E_{i-2m} X_i||_2 &\leqq ||E_{i-2m} E(X_i \mid \mathscr{F}_{i-m}^{i+m})||_2 + ||X_i - E(X_i \mid \mathscr{F}_{i-m}^{i+m})||_2 \\ &\leqq 2\varphi_m{}^{1-1/r} ||E(X_i \mathscr{F}_{i-m}^{i+m})||_2 + \nu_m \quad \text{by (2.2)} , \\ &\leqq 2\varphi_m{}^{1-1/r} ||X_i||_r + \nu_m \quad \text{by the conditional Jensen's inequality} \end{aligned}$$

or

$$\begin{aligned} &\leqq 5\alpha_m{}^{\frac{1}{2}-1/r} ||E(X_i \mid \mathscr{F}_{i-m}^{i+m})||_r + \nu_m \quad \text{by (2.3)} \\ &\leqq 5\alpha_m{}^{\frac{1}{2}-1/r} ||X_i||_r + \nu_m . \end{aligned}$$

Also, by Lemma 1, page 184 of Billingsley, $||E_{i+2m} X_i - X_i||_2 \leqq \nu_{2m}$. Therefore, X_i is a mixingale with $\psi_m = 2\varphi_{[m/2]}^{1-1/r} + \nu_{[m/2]}$ *or* $5\alpha_{[m/2]}^{\frac{1}{2}-1/r} + \nu_{[m/2]}$ (the square brackets denote "the greatest integer contained in"), and $c_i = \max(||X_i||_r, 1)$. Therefore, by (1.9), $S_n/n \to 0$ a.s.

EXAMPLE. Let $(\Omega, \mathscr{F}, P)$ be the unit interval with its Borel sets, and with P representing Borel measure. Consider a sequence f_k of measurable functions on Ω such that $\int_0^1 f_k(\omega)\, dP = 0$ and $\sum_{k=1}^\infty Ef_k{}^2/k^2 < \infty$. Define the random variable $\omega_k(\omega)$ equal to the kth digit in the binary decimal expansion of ω (as usual, possible ambiguity is eliminated by taking a terminating expansion whenever possible). Put $\mathscr{F}_m{}^n = \sigma(\omega_m, \omega_{m+1}, \cdots, \omega_n)$. Let $\hat{g}$ denote the periodic extension of g to $[0, \infty)$ and define $X_k(\omega) = \hat{f}_k(2^k\omega)$. Observe that since each $\hat{f}_k$ has period 1, X_k is $\mathscr{F}_{k+1}^\infty$-measurable. Therefore, for ω in the interval $[j/2^{k+m}, (j+1)/2^{k+m})$,

$$\begin{aligned} E(X_k \mid \mathscr{F}_{k-m}^{k+m})(\omega) &= E(X_k \mid \mathscr{F}_{k+1}^{k+m})(\omega) = 2^m \int_{j/2^m}^{(j+1)/2^m} f_k(y)\, dy \\ &= \widehat{E(f_k \mid \mathscr{F}_1{}^m)}(2^k\omega) . \end{aligned}$$

Therefore,

$$E(X_k - E(X_k \mid \mathscr{F}_{k+1}^{k+m}))^2 = E\{f_k - E(f_k \mid \mathscr{F}_1{}^m)\}^2 .$$

We will assume

(3.2) $$\nu_m{}^2 = \sup_k E\{f_k - E(f_k \mid \mathscr{F}_1{}^m)\}^2 \quad \text{is of size} \quad -1 .$$

In words, this means that the sequence f_k can be approximated within ν_m in L_2 norm by a sequence of step functions which are constant on intervals of the form $[j/2^m, (j+1)/2^m)$. Now the conditions of Theorem (3.1) are satisfied with $r = 2$, for the random variables ω_k are independent; thus $S_n/n \to 0$ a.s.

It is easy to show that (3.2) is a consequence of the following condition.

(3.3) There exists a positive nondecreasing function $g(x)$ defined on $[0, 1]$ and sets $\Omega_k' \subset \Omega$ of P-measure 1 such that $|f_k(\omega) - f_k(\omega')| \leqq g(|\omega - \omega'|)$ for all k and $\omega, \omega' \in \Omega_k'$ and $\{g(2^{-m})\}$ is of size $-\frac{1}{2}$.

Any function $g(y) = O(1/-\ln y)^\theta$ as $y \to 0$ for some $\theta > \frac{1}{2}$ is of size $-\frac{1}{2}$ and hence appropriate in Condition (3.3). We summarize these results with the following theorem.

(3.4) THEOREM. *If* $\{f_k\}$ *is a sequence of zero-mean functions on* $[0, 1]$ *with* $\sum_k Ef_k{}^2/k^2 < \infty$, *such that* (3.2) *or* (3.3) *holds, then* $1/n \sum_{k=1}^n \hat{f}_k(2^k\omega) \to 0$ *almost everywhere.*

(3.5) THE STATIONARY CASE. When the sequence $\{X_n\}$ is stationary and satisfies the usual strong mixing condition, the results of Section 2 are known to hold with the only requirement that $\alpha_n \to 0$.

Similarly if all of the f_k's are the same function f, the conclusions of (3.4) hold for any zero-mean L_1 function f.

Both of these well known results are immediate consequences of the ergodic theorem.

REFERENCES

[1] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] BLUM, J., HANSON, D. and KOOPMANS, L. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** 1-11.

[3] CHOW, Y. S. (1967). On a strong law of large numbers for martingales. *Ann. Math. Statist.* **38** 610.

[4] CHUNG, K. L. (1968). *A Course in Probability Theory.* Harcourt, Brace, New York.

[5] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[6] DVORETSKY, A. Asymptotic normality for sums of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 513–535. Univ. of Calif. Press.

[7] FELLER, W. (1957). *An Introduction to Probability Theory and its Applications* **1**, 2nd ed. Wiley, New York.

[8] FELLER, W. (1971). *An Introduction to Probability Theory and its Applications* **2**, 2nd ed. Wiley, New York.

[9] GORDIN, M. I. (1969). The central limit theorem for stationary processes. *Soviet Math. Dokl.* **10** 1174–1176.

[10] HEYDE, C. C. and SCOTT, D. J. (1973). Invariance principles for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probability* **1** 428–437.

[11] IOSIFESCU, M. and THEODORESCU, R. (1969). *Random Processes and Learning.* Springer, Berlin.

[12] KNOPP, K. (1946). *Theory and Applications of Infinite Series.* Blackie, London.

[13] LOÈVE, M. (1960). *Probability Theory*, 2nd ed. Van Nostrand, Princeton.

[14] MACQUEEN, J. B. (1973). A linear extension of the martingale theorem. *Ann. Probability* **2** 263–271.

[15] MCLEISH, D. (1973). Invariance principles for dependent variables. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete.* (To appear.)

[16] O'BRIEN, G. L. and KESTEN, H. (1975). Examples of mixing sequences. (To appear.)

[17] PHILIPP, W. and STOUT, W. F. (1975). Almost sure invariance principles for sums of weakly dependent random variables. *Mem. Amer. Math. Soc.* **161**.

[18] ROSENBLATT, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* **42** 43–47.

[19] SCOTT, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Appl. Probability* **5** 119–137.

[20] SERFLING, R. J. (1968). Contributions to central limit theory for dependent variables. *Ann. Math. Statist.* **39** 1158–1175.

DEPARTMENT OF MATHEMATICS
YORK UNIVERSITY
DOWNSVIEW 463, ONTARIO
CANADA