LECTURE 5

Measures of Short-Range Dependence

There are a number of ways of setting up conditions under which versions of the classical limit theorems still hold for dependent sequences. We shall be mainly concerned with the central theorem for partial sums of the random variables of the sequence. One of these ways is to consider some type of mixing condition. With this in mind we consider the following measures of dependence for two sub-σ-fields $\mathscr{A} \subset \mathscr{F}$, $\mathscr{B} \subset \mathscr{F}$ of a probability space $(\Omega, \mathscr{F}, P)$. They are

$$\alpha(\mathscr{A}, \mathscr{B}) = \sup |P(A \cap B) - P(A)P(B)|, \qquad A \in \mathscr{A}, B \in \mathscr{B},$$

$$\phi(\mathscr{A}, \mathscr{B}) = \sup |P(B|A) - P(B)|, \qquad A \in \mathscr{A}, B \in \mathscr{B}, P(A) > 0,$$

$$\phi_{\mathrm{rev}}(\mathscr{A}, \mathscr{B}) = \phi(\mathscr{B}, \mathscr{A}),$$

$$\psi(A, B) = \sup \frac{|P(A \cap B) - P(A)P(B)|}{P(A)P(B)}, \qquad A \in \mathscr{A}, B \in \mathscr{B},$$

$$\rho(\mathscr{A}, \mathscr{B}) = \sup |\mathrm{corr}(X, Y)|, \qquad X \in L_2(\mathscr{A}), Y \in L_2(\mathscr{B}), X, Y \text{ real},$$

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

with the supremum taken over partitions $\{A_i, i \in I\}$ and $\{B_j, j \in J\}$ each of Ω but with $A_i \in \mathscr{A}$, $B_j \in \mathscr{B}$. Here $L_2(\mathscr{A})$ denotes the set of square integrable functions measurable with respect to the σ-field $\mathscr{A}$.

Let us now consider a stationary sequence of random variables X_k, $k = \ldots, -1, 0, 1, \ldots$. $\mathscr{F}$ is the σ-field generated by the random variables of the process $\mathscr{B}(X_j, -\infty < j < \infty)$. Set

$$\mathscr{B}_n = \mathscr{B}(X_j, j \le n),$$

$$\mathscr{F}_m = \mathscr{B}(X_j, j \ge m).$$

We then introduce the coefficients

$$\alpha(n) = \alpha(\mathscr{B}_0, \mathscr{F}_n),$$
$$\phi(n) = \phi(\mathscr{B}_0, \mathscr{F}_n),$$
$$\psi(n) = \psi(\mathscr{B}_0, \mathscr{F}_n),$$
$$\rho(n) = \rho(\mathscr{B}_0, \mathscr{F}_n),$$
$$\beta(n) = \beta(\mathscr{B}_0, \mathscr{F}_n).$$

The process (X_k) is said to be strongly mixing if $\lim_{n\to\infty} \alpha(n) = 0$, ϕ-mixing if $\lim_{n\to\infty} \phi(n) = 0$, ψ-mixing if $\lim_{n\to\infty} \psi(n) = 0$, to have asymptotic correlation zero if $\lim_{n\to\infty} \rho(n) = 0$, and to be absolutely regular if $\lim_{n\to\infty} \beta(n) = 0$. The absolutely regular condition is sometimes referred to as a weak Bernoulli condition. An extended discussion of these conditions can be found in the paper of Bradley (1986).

Certain fairly obvious relations among these different mixing conditions are (i) asymptotic correlation zero $\Rightarrow$ strong mixing; (ii) absolute regularity $\Rightarrow$ strong mixing; (iii) ϕ-mixing $\Rightarrow$ asymptotic correlation zero, absolute regularity; (iv) ψ-mixing $\Rightarrow$ ϕ-mixing.

It is of some interest to see what these different mixing conditions amount to in the case of familiar classes of random processes. We shall first consider the class of Gaussian stationary sequences. It will be convenient to let $\sigma(X_i, i \in I)$ stand for the σ-field generated by the random variables X_i, $i \in I$, where I is an index set. First of all, the analysis of the prediction problem implies that a Gaussian stationary sequence (X_k) is purely nondeterministic (in the sense that the backward tail field is trivial) if and only if the spectral distribution function of the process is absolutely continuous with a spectral density $f(\lambda)$ satisfying

$$\int_{-\pi}^{\pi} \log f(\lambda)\, d\lambda > -\infty.$$

The paper of Kolmogorov and Rozanov (1960) implies that the strong mixing condition and asymptotic correlation zero are equivalent in the case of Gaussian stationary sequences. For the general class of stationary sequences, asymptotic correlation zero is a more stringent condition than strong mixing. They showed that a sufficient condition for strong mixing in the Gaussian stationary context is that the spectral density be continuous and positive. Helson and Sarason (1967) showed that a necessary and sufficient condition for Gaussian strong mixing is that the spectral density be expressible as

$$f(\lambda) = |P(e^{i\lambda})|^2 \exp\left[u(e^{i\lambda}) + \tilde{v}(e^{i\lambda})\right], \tag{5.1}$$

where P is polynomial, u and v are continuous real functions on the unit circle in the complex plane and $\tilde{v}$ is the conjugate function of v.

The book of Ibragimov and Rozanov (1978) has a detailed discussion of the Gaussian case. One can find there a demonstration of the equivalence of absolute regularity and the representation of the spectral density $f(\lambda)$ as

$$f(\lambda) = |P(e^{i\lambda})|^2 \exp\left[\sum_{j=-\infty}^{\infty} a_j e^{ij\lambda}\right]$$

with P a polynomial whose possible roots lie on the unit circle and

$$\sum_{j=-\infty}^{\infty} |j| \cdot |a_j|^2 < \infty.$$

Given these results, it is clear that a Gaussian stationary sequence with spectral density

$$f(\lambda) = \exp\left[\sum_{j=1}^{\infty} 2^{-j} \cos(2^{2^j}\lambda)\right]$$

is strongly mixing but not absolutely regular. The spectral density of a strongly mixing process need not be regular as can be seen in the case of the spectral density f, where

$$f(\lambda) = \exp\left\{\sum_{1}^{\infty} \frac{\cos k\lambda}{k \log(k+1)}\right\}.$$

Here the conjugate function of $\log f(\lambda)$,

$$\sum_{1}^{\infty} \frac{\sin k\lambda}{k \log(k+1)},$$

is continuous and so by (5.1) the process is strongly mixing. However,

$$\log f(\lambda) \sim \int_1^{1/\lambda} \frac{dt}{t \log(t+1)} \sim \log\log \frac{1}{\lambda}$$

as $\lambda \downarrow 0$.

One can show that if a stationary Gaussian sequence is ϕ-mixing, it must be a finite moving average of independent identically distributed Gaussian random variables, or equivalently, a finite step dependent sequence. This is a consequence of the observation that if U, V are jointly Gaussian with $\operatorname{corr}(U,V) \neq 0$, then $\phi(\sigma(U), \sigma(V)) = 1$. This follows from the fact [if $\operatorname{corr}(U,V) > 0$] that for $a > 0$,

$$P(U > a | V > b) \to 1$$

as $b \to \infty$.

It is also of interest to consider what can be said of these different mixing conditions in the context of stationary Markov sequences (real-valued). By using the Markov property, it is clear that

$$\alpha(n) = \alpha(\sigma(X_0), \sigma(X_n)),$$
$$\phi(n) = \phi(\sigma(X_0), \sigma(X_n)),$$
$$\psi(n) = \psi(\sigma(X_0), \sigma(X_n)),$$
$$\rho(n) = \rho(\sigma(X_0), \sigma(X_n)),$$
$$\beta(n) = \beta(\sigma(X_0), \sigma(X_n)).$$

In the case of $\rho(n)$, $\phi(n)$, $\psi(n)$, if the mixing coefficient tends to zero as $n \to \infty$, the rate of decrease must be exponential. This does not hold for $\alpha(n)$ or $\beta(n)$. The basic convergence theorem for convergence of transition probabilities of an aperiodic stationary irreducible countable state Markov chain implies that such a chain is absolutely regular (and hence strongly mixing).

Let the one step transition function for the Markov sequence be $P(x, A)$ with $x \in R$ and A a Borel set. We assume that for each Borel set A, $P(x, A)$ is a Borel function. Then the transition function induces a bounded operator T on the bounded Borel functions g

$$(Tg)(x) = \int P(x, dy) g(y)$$

with

$$\sup_x |(Tg)(x)| \leq \sup_x |g(x)|.$$

If there is an invariant probability measure μ on the Borel sets,

$$\int \mu(dx) P(x, A) = \mu(A).$$

An early sufficient condition for the existence of such an invariant measure is the Doeblin condition D. This is the condition that there be a finite measure ϕ $(0 < \phi(R) < \infty)$ [see Doob (1953)] on Borel sets, an integer $n \geq 1$ and a positive ε such that [$P_n(\cdot, \cdot)$ is the n step transition probability]

$$P_n(x, A) \leq 1 - \varepsilon \quad \text{if } \phi(A) \leq \varepsilon$$

for all x. If the Markov sequence satisfies D and has no cyclically moving sets (is purely nondeterministic), we shall say it satisfies the Doeblin condition D_0. One can show that this condition is equivalent to

$$\sup_B |P_n(x, B) - \mu(B)| \to 0$$

for almost all $x(d\mu)$ as $n \to \infty$. This can in turn be shown to be equivalent to an L^∞ condition

$$\sup_{f \perp 1} \frac{\|T^n f\|_\infty}{\|f\|_\infty} \to 0.$$

In an extension of Doeblin's work, Harris (1956) proposed a recurrence condition that implies the existence of an invariant measure (which may be σ-finite). It is the condition that there is a measure ϕ (which one can take to be finite) such that for any set B with $\phi(B) > 0$,

$$P(X_n \in B \text{ for infinitely many positive integers } n | X_0 = x) = 1$$

for all x.

One can show that a strictly stationary real-valued aperiodic Markov sequence satisfying the Harris recurrence condition is absolutely regular [see Bradley (1986)]. It is of some interest to note that if a stationary sequence X_k

is strongly mixing, the partial sums

$$S_n = \sum_{j=1}^{n} X_j$$

when centered and normalized $(S_n - a_n)/b_n$ with $b_n \to \infty$ as $n \to \infty$ can only have a limit law that is stable. A simple example of a bounded function of a Markov chain with a stable but non-Gaussian limit law is given by Davydov (1973). The states of the chain are integers. The transition probabilities are $p_{n,n+1} = p_{-n,n-1} = \alpha_n$, $n \geq 0$, with $p_{n,0} = p_{-n,0} = 1 - \alpha_n$, $n > 0$, $p_{0,0} = 0$, $\alpha_0 = \frac{1}{2}$, $0 < \alpha_n < 1$ for $n \geq 1$. Let f_{00}^n be the probability of first return to zero at time n given that one left zero at time zero. Then $f_{00}^1 = 0$ and for $n \geq 2$, $f_{00}^n = \beta_{n-1} - \beta_n$ with $\beta_0 = \beta_1 = 1$, $\beta_n = \alpha_1 \cdot \alpha_2 \cdots \alpha_{n-1}$. The chain is recurrent if and only if $\beta_n \to 0$ and the stationary distribution is

$$\pi_0 = \left(\sum_{0}^{\infty} \beta_n \right)^{-1}, \qquad \pi_j = \pi_{-j} = \tfrac{1}{2}\pi_0 \beta_{j-1}, \qquad j > 0,$$

which exists only when $\sum_0^\infty \beta_n < \infty$. This construction is similar to that of chains mentioned in Chung (1960). Let $f_{00}^n = An^{-2-\delta}$, $0 < \delta < 1$, and let the function g be given by $g(0) = 0$, $g(j) = g(-j) = 1 + 1/j$, $j > 0$. We consider the partial sums $\sum_{j=1}^n Y_j$ with $Y_j = g(X_j)$ with X_j the Markov chain. One can show that if centered and normalized partial sums converge in distribution for any given initial distribution, they will converge to the same limiting distribution with any other initial distribution. Let the initial distribution have all its mass at zero. Then

$$P\left[\sum_{j=\tau_s+1}^{\tau_{s+1}} Y_j = \sum_{j=1}^{n-1} \left(1 + \frac{1}{j}\right) \right] = P\left[\sum_{j=\tau_s+1}^{\tau_{s+1}} Y_j = -\sum_{j=1}^{n-1} \left(1 + \frac{1}{j}\right) \right] = \frac{1}{2} f_{00}^n$$

if the random variable τ_s is the time of sth return to zero. But then with $Z_s = \sum_{j=\tau_s+1}^{\tau_{s+1}} Y_j$,

$$P[|Z_s| > x] \sim \sum_{n>x} f_{00}^n \sim Cx^{-1-\delta}$$

and because of the independence of the Z's, the partial sums S_n when properly normalized converge in distribution to a symmetric stable law with exponent $1 + \delta$. It should, however, be noted that in considering probability density estimates, we have to deal with triangular arrays rather than with partial sums of a fixed stationary process.

In the general context of a real stationary sequence, one can show that

$$\phi(n) = \operatorname{ess\,sup}\left[\sup |P(B|\mathscr{B}_0) - P(B)|,\ B \in \mathscr{F}_n\right]$$

and

$$\beta(n) = E\left[\sup |P(B|\mathscr{B}_0) - P(B)|,\ B \in \mathscr{F}_n\right].$$

Let us just give a statement and brief sketch of a derivation of a central limit theorem for triangular arrays on the space of a strongly mixing process that is a small generalization of a central limit theorem given in Rosenblatt (1956b).

THEOREM. *Consider* $\{Y_j^{(n)}, j = \dots, -1, 0, 1, \dots\}$, $EY_j^{(n)} \equiv 0$, $n = 1, 2, \dots$, *a sequence of strictly stationary processes defined on the probability space of a strongly mixing stationary process* $\{X_m\}$. *Assume that* $Y_j^{(n)}$ *is measurable with respect to* $\mathscr{F}_{j-c(n)} \cap \mathscr{B}_{j+c(n)}$, *where* $c(n) = o(n)$, $c(n) \uparrow \infty$ *as* $n \to \infty$. *Let*

$$h_n(b-a) = E\left|\sum_{j=a}^{b} Y_j^{(n)}\right|^2.$$

Also given any two sequences $s(n), m(n)$ *with* $c(n) = o(m(n))$, $m(n) \leq n$ *and* $s(n)/m(n) \to 0$, *let*

$$h_n(m(n))/h_n(s(n)) \to \infty.$$

Further assume

$$\{h_n(m(n))\}^{-(2+\delta)/2} E\left|\sum_{k=1}^{m(n)} Y_k^{(n)}\right|^{2+\delta} = O(1)$$

as $m(n) \to \infty$ *for some* $\delta > 0$. *There are then sequences* $k(n), p(n) \to \infty$ *as* $n \to \infty$ *with* $k(n)p(n) \simeq n$ *such that*

$$\sum_{j=1}^{n} Y_j^{(n)} \Big/ \sqrt{k(n) h_n(p(n))}$$

is asymptotically normally distributed with mean zero and variance one. If $k(n)h_n(p(n)) \simeq h_n(n)$, *the normalization of the partial sum can be replaced by* $\sqrt{h_n(n)}$.

The argument proceeds by a big block, small block argument. Such an argument was employed by Bernstein (1927) many years ago and it is still often useful. Let $r = 1, \dots, k(n)$ with $k(n)(p(n) + q(n)) = n$,

$$U_r(n) = \sum_{j=(r-1)(p(n)+q(n))+1}^{rp(n)+(r-1)q(n)} Y_j^{(n)},$$

$$V_r(n) = \sum_{j=rp(n)+(r-1)q(n)+1}^{r(p(n)+q(n))} Y_j^{(n)},$$

$p(n), q(n) \to \infty$ and $q(n)/p(n) \to \infty$. The U_r's are the big blocks and the V_r's are the small blocks. Now

$$E^{1/2}\left|\sum_{r=1}^{k} \frac{V_r(n)}{\sqrt{k(n)h_n(p(n))}}\right|^2 = O(k(n)h_n(q(n))/h_n(p(n)))^{1/2}$$

and $k(n)$, $p(n)$, $q(n)$ can also be chosen so that the expression on the right tends to zero as $n \to \infty$. The normalized sum of the small blocks can therefore be disregarded. Introduce

$$G_{r,n}(x) = P\left[U_r(n)\{k(n)h_n(p(n))\}^{-1/2} \le x\right]$$

and the event

$$\left\{l_r\delta < \frac{U_r(n)}{\sqrt{k(n)h_n(p(n))}} \le (l_r + 1)\delta\right\} = A(r, n, l_r, \delta).$$

Repeated use of the strong mixing condition implies that if $t_k = (k(n)/\varepsilon)^{1/2}$ and $c(n) = o(q(n))$, then

$$\left|\sum_{(l_1+\cdots+l_k)\delta \le x} P\left(\bigcap_{r=1}^{k} A(r, n, l_r, \delta)\right) - \sum_{(l_1+\cdots+l_k)\delta \le x} \prod_{r=1}^{k} P(A(r, n, l_r, \delta))\right|$$
$$\le k\left(\frac{2t_k}{\delta}\right)^k \alpha(q(n) - c(n)) + \varepsilon.$$

This inequality can be used to show that

$$P\left(\sum_{r=1}^{k(n)} U_r \Big/ \sqrt{k(n)h_n(p(n))} \le x\right)$$

can be approximated by $G_{1,n} * \cdots * G_{k,n}(x)$. With appropriate choice of $k(n)$, $p(n)$, $q(n)$ and an application of the Liapounov form of the central limit theorem for independent random variables, as $n \to \infty$ the distribution of $G_{1n} * \cdots * G_{k,n}(x)$ tends to that of a standard normal distribution.

A type of mixing condition can also be formulated in terms of cumulants. Moments can be identified as the coefficients in the Taylor expansion of the joint characteristic function of the random variables X_j about zero. Cumulants are the corresponding coefficients in the Taylor expansion of the logarithm of the joint characteristic function (see the discussion at the beginning of Lecture 7) about zero. Let the kth cumulant

$$c_k(\tau_1, \ldots, \tau_{k-1}) = \operatorname{cum}(X_0, X_{\tau_1}, \ldots, X_{\tau_{k-1}}).$$

A summability condition on the kth cumulant function

$$\sum_{\tau_1, \ldots, \tau_{k-1}} |c_k(\tau_1, \ldots, \tau_{k-1})| < \infty \tag{5.2}$$

for $k = 1, 2, \ldots$ was used in Brillinger (1965) to obtain a central limit theorem. Strengthened versions of such conditions were employed in Brillinger and Rosenblatt (1967) to obtain asymptotic normality for cumulant spectral estimates obtained by taking smoothed and weighted versions of periodogram-like functions (obtained from finite Fourier transforms) computed at values of the argument of the form $2\pi s/N$ with s integral and N the sample size. It would be appropriate to call these polynomial-type mixing conditions. The condition

(5.2) has the disadvantage that all moments are assumed to exist. It can, however, be used to prove asymptotic normality for partial sums of polynomial functions for classes of processes for which all moments exist but that are not strongly mixing and do not satisfy other related conditions of the type discussed earlier. An example of such a class of processes is the set of autoregressive moving average stationary processes with the independent random sequence ξ_k generating the process a sequence of random variables taking on only a finite number of distinct values.