Strong Mixing Conditions

Richard C. Bradley
Department of Mathematics, Indiana University, Bloomington, Indiana, USA

There has been much research on stochastic models that have a well defined, specific structure — for example, Markov chains, Gaussian processes, or linear models, including ARMA (autoregressive – moving average) models. However, it became clear in the middle of the last century that there was a need for a theory of statistical inference (e.g. central limit theory) that could be used in the analysis of time series that did not seem to "fit" any such specific structure but which did seem to have some "asymptotic independence" properties. That motivated the development of a broad theory of "strong mixing conditions" to handle such situations. This note is a brief description of that theory.

The field of strong mixing conditions is a vast area, and a short note such as this cannot even begin to do justice to it. Journal articles (with one exception) will not be cited; and many researchers who made important contributions to this field will not be mentioned here. All that can be done here is to give a narrow snapshot of part of the field.

The strong mixing (α-mixing) condition. Suppose $X := (X_k, k \in \mathbf{Z})$ is a sequence of random variables on a given probability space $(\Omega, \mathcal{F}, P)$. For $-\infty \le j \le \ell \le \infty$, let $\mathcal{F}_j^\ell$ denote the σ-field of events generated by the random variables $X_k,\ j \le k \le \ell$ ($k \in \mathbf{Z}$). For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, define the "measure of dependence"

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|. \tag{1}$$

For the given random sequence X, for any positive integer n, define the dependence coefficient

$$\alpha(n) = \alpha(X, n) := \sup_{j \in \mathbf{Z}} \alpha(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty). \tag{2}$$

By a trivial argument, the sequence of numbers $(\alpha(n), n \in \mathbf{N})$ is nonincreasing. The random sequence X is said to be "strongly mixing", or "α-mixing", if $\alpha(n) \to 0$ as $n \to \infty$. This condition was introduced in 1956 by Rosenblatt [Ro1], and was used in that paper in the proof of a central limit theorem. (The phrase "central limit theorem" will henceforth be abbreviated CLT.)

In the case where the given sequence X is strictly stationary (i.e. its distribution is invariant under a shift of the indices), eq. (2) also has the simpler form

$$\alpha(n) = \alpha(X, n) := \alpha(\mathcal{F}_{-\infty}^0, \mathcal{F}_n^\infty). \tag{3}$$

For simplicity, *in the rest of this note, we shall restrict to strictly stationary sequences*. (Some comments below will have obvious adaptations to nonstationary processes.)

In particular, for strictly stationary sequences, the strong mixing (α-mixing) condition implies Kolmogorov regularity (a trivial "past tail" σ-field), which in turn implies "mixing" (in the ergodic-theoretic sense), which in turn implies ergodicity. (None of the converse implications holds.) For further related information, see e.g. [Br, v1, Chapter 2].

Comments on limit theory under α-mixing. Under α-mixing and other similar conditions (including ones reviewed below), there has been a vast development of limit theory — for example, CLTs, weak invariance principles, laws of the iterated logarithm, almost sure invariance principles, and rates of convergence in the strong law of large numbers. For example, the CLT in [Ro1] evolved through subsequent refinements by several researchers into the following "canonical" form. (For its history and a generously detailed presentation of its proof, see e.g. [Br, v1, Theorems 1.19 and 10.2].)

Theorem 1. *Suppose $(X_k, k \in \mathbf{Z})$ is a strictly stationary sequence of random variables such that $EX_0 = 0$, $EX_0^2 < \infty$, $\sigma_n^2 := ES_n^2 \to \infty$ as $n \to \infty$, and $\alpha(n) \to 0$ as $n \to \infty$. Then the following two conditions (A) and (B) are equivalent:*
(A) The family of random variables $(S_n^2/\sigma_n^2, n \in \mathbf{N})$ is uniformly integrable.
(B) $S_n/\sigma_n \Rightarrow N(0,1)$ as $n \to \infty$.

If (the hypothesis and) these two equivalent conditions (A) and (B) hold, then $\sigma_n^2 = n \cdot h(n)$ for some function $h(t)$, $t \in (0, \infty)$ which is slowly varying as $t \to \infty$.

Here $S_n := X_1 + X_2 + \ldots + X_n$; and $\Rightarrow$ denotes convergence in distribution. The assumption $ES_n^2 \to \infty$ is needed here in order to avoid trivial α-mixing (or even 1-dependent) counterexamples in which a kind of "cancellation" prevents the partial sums S_n from "growing" (in probability) and becoming asymptotically normal.

In the context of Theorem 1, if one wants to obtain asymptotic normality of the partial sums (as in condition (B)) without an explicit uniform integrability assumption on the partial sums (as in condition (A)), then as an alternative, one can impose a combination of assumptions on, say, (i) the (marginal) distribution of X_0 and (ii) the rate of decay of the numbers $\alpha(n)$ to 0 (the "mixing rate"). This involves a "trade-off"; the weaker one assumption is, the stronger the other has to be. One such CLT of Ibragimov in 1962 involved such a "trade-off" in which it is assumed that for some $\delta > 0$, $E|X_0|^{2+\delta} < \infty$ and $\sum_{n=1}^{\infty} [\alpha(n)]^{\delta/(2+\delta)} < \infty$. Counterexamples of Davydov in 1973 (with just slightly weaker properties) showed that that result is quite sharp. However, it is not at the exact "borderline". From a covariance inequality of Rio in 1993 and a CLT (in fact a weak invariance principle) of Doukhan, Massart, and Rio in 1994, it became clear that the "exact borderline" CLTs of this kind have to involve quantiles of the (marginal) distribution of X_0 (rather than just moments). For a generously detailed exposition of such CLTs, see [Br, v1, Chapter 10]; and for further related results, see also Rio [Ri].

Under the hypothesis (first sentence) of Theorem 1 (with just finite second moments), there is no mixing rate, no matter how fast (short of m-dependence), that can insure that a CLT holds. That was shown in 1983 with two different counterexamples, one by the author and the other by Herrndorf. See [Br, v1&3, Theorem 10.25 and Chapter 31].

Several other classic strong mixing conditions. As indicated above, the terms "α-mixing" and "strong mixing condition" (singular) both refer to the condition $\alpha(n) \to 0$.

(A little caution is in order; in ergodic theory, the term "strong mixing" is often used to refer to the condition of "mixing in the ergodic-theoretic sense", which is weaker than α-mixing as noted earlier.) The term "strong mixing conditions" (plural) can reasonably be thought of as referring to all conditions that are at least as strong as (i.e. that imply) α-mixing. In the classical theory, five strong mixing conditions (again, plural) have emerged as the most prominent ones: α-mixing itself and four others that will be defined here.

Recall our probability space $(\Omega, \mathcal{F}, P)$. For any two σ-fields $\mathcal{A}$ and $\mathcal{B} \subset \mathcal{F}$, define the following four "measures of dependence":

$$\phi(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A)>0} |P(B|A) - P(B)|; \tag{4}$$

$$\psi(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A)>0, P(B)>0} |P(B \cap A)/[P(A)P(B)] - 1|; \tag{5}$$

$$\rho(\mathcal{A}, \mathcal{B}) := \sup_{f \in \mathcal{L}^2(\mathcal{A}),\, g \in \mathcal{L}^2(\mathcal{B})} |\mathrm{Corr}(f, g)|; \quad \text{and} \tag{6}$$

$$\beta(\mathcal{A}, \mathcal{B}) := \sup\ (1/2) \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| \tag{7}$$

where the latter supremum is taken over all pairs of finite partitions $(A_1, A_2, \dots, A_I)$ and $(B_1, B_2, \dots, B_J)$ of Ω such that $A_i \in \mathcal{A}$ for each i and $B_j \in \mathcal{B}$ for each j. In (6), for a given σ-field $\mathcal{D} \subset \mathcal{F}$, the notation $\mathcal{L}^2(\mathcal{D})$ refers to the space of (equivalence classes of) square-integrable, $\mathcal{D}$-measurable random variables.

Now suppose $X := (X_k, k \in \mathbf{Z})$ is a strictly stationary sequence of random variables on $(\Omega, \mathcal{F}, P)$. For any positive integer n, analogously to (3), define the dependence coefficient

$$\phi(n) = \phi(X, n) := \phi(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{n}^{\infty}), \tag{8}$$

and define analogously the dependence coefficients $\psi(n)$, $\rho(n)$, and $\beta(n)$. Each of these four sequences of dependence coefficients is trivially nonincreasing. The (strictly stationary) sequence X is said to be
"ϕ-mixing" if $\phi(n) \to 0$ as $n \to \infty$;
"ψ-mixing" if $\psi(n) \to 0$ as $n \to \infty$;
"ρ-mixing" if $\rho(n) \to 0$ as $n \to \infty$; and
"absolutely regular", or "β-mixing", if $\beta(n) \to 0$ as $n \to \infty$.

The ϕ-mixing condition was introduced by Ibragimov in 1959 and was also studied by Cogburn in 1960 . The ψ-mixing condition evolved through papers of Blum, Hanson, and Koopmans in 1963 and Philipp in 1969; and (see e.g. [Io]) it was also implicitly present in earlier work of Doeblin in 1940 involving the metric theory of continued fractions. The ρ-mixing condition was introduced by Kolmogorov and Rozanov 1960. (The "maximal correlation coefficient" $\rho(\mathcal{A}, \mathcal{B})$ itself was first studied by Hirschfeld in 1935 in a statistical context that had no particular connection with "stochastic processes".) The absolute regularity (β-mixing) condition was introduced by Volkonskii and Rozanov in 1959, and in the ergodic theory literature it is also called the "weak Bernoulli" condition.

For the five measures of dependence in (1) and (4)–(7), one has the following well known inequalities:

$$\begin{aligned}
2\alpha(\mathcal{A},\mathcal{B}) &\le \beta(\mathcal{A},\mathcal{B}) \le \phi(\mathcal{A},\mathcal{B}) \le (1/2)\psi(\mathcal{A},\mathcal{B});\\
4\alpha(\mathcal{A},\mathcal{B}) &\le \rho(\mathcal{A},\mathcal{B}) \le \psi(\mathcal{A},\mathcal{B}); \quad \text{and}\\
\rho(\mathcal{A},\mathcal{B}) &\le 2[\phi(\mathcal{A},\mathcal{B})]^{1/2}[\phi(\mathcal{B},\mathcal{A})]^{1/2} \le 2[\phi(\mathcal{A},\mathcal{B})]^{1/2}.
\end{aligned}$$

For a history and proof of these inequalities, see e.g. [Br, v1, Theorem 3.11]. As a consequence of these inequalities and some well known examples, one has the following "hierarchy" of the five strong mixing conditions here:

(i) ψ-mixing implies ϕ-mixing.

(ii) ϕ-mixing implies both ρ-mixing and β-mixing (absolute regularity).

(iii) ρ-mixing and β-mixing each imply α-mixing (strong mixing).

(iv) Aside from "transitivity", there are in general no other implications between these five mixing conditions. In particular, neither of the conditions ρ-mixing and β-mixing implies the other.

For all of these mixing conditions, the "mixing rates" can be essentially arbitrary, and in particular, arbitrarily slow. That general principle was established by Kesten and O'Brien in 1976 with several classes of examples. For further details, see e.g. [Br, v3, Chapter 26].

The various strong mixing conditions above have been used extensively in statistical inference for weakly dependent data. See e.g. [DDLLLP], [DMS], [Ro3], or [Žu].

Ibragimov's conjecture and related material. Suppose (as in Theorem 1) $X := (X_k, k \in \mathbf{Z})$ is a strictly stationary sequence of random variables such that

$$EX_0 = 0, \quad EX_0^2 < \infty, \quad \text{and} \quad ES_n^2 \to \infty \text{ as } n \to \infty. \tag{9}$$

In the 1960s, I.A. Ibragimov conjectured that under these assumptions, if also X is ϕ-mixing, then a CLT holds. Technically, this conjecture remains unsolved. Peligrad showed in 1985 that it holds under the stronger "growth" assumption $\liminf_{n\to\infty} n^{-1}ES_n^2 > 0$. (See e.g. [Br, v2, Theorem 17.7].)

Under (9) and ρ-mixing (which is weaker than ϕ-mixing), a CLT need not hold (see [Br, v3, Chapter 34] for counterexamples). However, if one also imposes either the stronger moment condition $E|X_0|^{2+\delta} < \infty$ for some $\delta > 0$, or else the "logarithmic" mixing rate assumption $\sum_{n=1}^{\infty} \rho(2^n) < \infty$, then a CLT does hold (results of Ibragimov in 1975). For further limit theory under ρ-mixing, see e.g. [LL] or [Br, v1, Chapter 11].

Under (9) and an "interlaced" variant of the ρ-mixing condition (i.e. with the two index sets allowed to be "interlaced" instead of just "past" and "future"), a CLT does hold. For this and related material, see e.g. [Br, v1, Sections 11.18-11.28].

There is a vast literature on central limit theory for random fields satisfying various strong mixing conditions. See e.g. [Ro3], [Žu], [Do], and [Br, v3]. In the formulation of mixing conditions for random fields — and also "interlaced" mixing conditions for random sequences — some caution is needed; see e.g. [Br, v1&3, Theorems 5.11, 5.13, 29.9, and 29.12].

Connections with specific types of models. Now let us return briefly to a theme from the beginning of this write-up: the connection between strong mixing conditions and specific structures.

Markov chains. Suppose $X := (X_k, k \in \mathbf{Z})$ is a strictly stationary Markov chain. In the case where X has finite state space and is irreducible and aperiodic, it is ψ-mixing, with at least exponentially fast mixing rate. In the case where X has countable (but not necessarily finite) state space and is irreducible and aperiodic, it satisfies β-mixing, but the mixing rate can be arbitrarily slow. In the case where X has (say) real (but not necessarily countable) state space, (i) Harris recurrence and "aperiodicity" (suitably defined) together are equivalent to β-mixing, (ii) the "geometric ergodicity" condition is equivalent to β-mixing with at least exponentially fast mixing rate, and (iii) one particular version of "Doeblin's condition" is equivalent to ϕ-mixing (and the mixing rate will then be at least exponentially fast). There exist strictly stationary, countable-state Markov chains that are ϕ-mixing but not "time reversed" ϕ-mixing (note the asymmetry in the definition of $\phi(\mathcal{A}, \mathcal{B})$ in (4)). For this and other information on strong mixing conditions for Markov chains, see e.g. [Ro2, Chapter 7], [Do], [MT], and [Br, v1&2, Chapters 7 and 21].

Stationary Gaussian sequences. For stationary Gaussian sequences $X := (X_k, k \in \mathbf{Z})$, Ibragimov and Rozanov [IR] give characterizations of various strong mixing conditions in terms of properties of spectral density functions. Here are just a couple of comments: For stationary Gaussian sequences, the α- and ρ-mixing conditions are equivalent to each other, and the ϕ- and ψ-mixing conditions are each equivalent to m-dependence. If a stationary Gaussian sequence has a continuous positive spectral density function, then it is ρ-mixing. For some further closely related information on stationary Gaussian sequences, see also [Br, v1&3, Chapters 9 and 27].

Dynamical systems. Many dynamical systems have strong mixing properties. Certain one-dimensional "Gibbs states" processes are ψ-mixing with at least exponentially fast mixing rate. A well known standard "continued fraction" process is ψ-mixing with at least exponentially fast mixing rate (see [Io]). For certain stationary finite-state stochastic processes built on piecewise expanding mappings of the unit interval onto itself, the absolute regularity condition holds with at least exponentially fast mixing rate. For more detains on the mixing properties of these and other dynamical systems, see e.g. Denker [De].

Linear and related processes. There is a large literature on strong mixing properties of strictly stationary linear processes (including strictly stationary ARMA processes and also "non-causal" linear processes and linear random fields) and also of some other related processes such as bilinear, ARCH, or GARCH models. For details on strong mixing properties of these and other related processes, see e.g. Doukhan [Do, Chapter 2].

However, many strictly stationary linear processes *fail* to be α-mixing. A well known classic example is the strictly stationary AR(1) process (autoregressive process of order 1) $X := (X_k, k \in \mathbf{Z})$ of the form $X_k = (1/2)X_{k-1} + \xi_k$ where $(\xi_k, k \in \mathbf{Z})$ is a sequence of independent, identically distributed random variables, each taking the values 0 and 1 with probability 1/2 each. It has long been well known that this random sequence X is not α-mixing. For more on this example, see e.g. [Br, v1, Example 2.15] or [Do, Section 2.3.1].

Further related developments. The AR(1) example spelled out above, together with many other examples that are not α-mixing but seem to have some similar "weak dependence" quality, have motivated the development of more general conditions of weak dependence that have the "spirit" of, and most of the advantages of, strong mixing conditions, but are less restrictive, i.e. applicable to a much broader class of models (including the AR(1) example above). There is a substantial development of central limit theory for strictly stationary sequences under weak dependence assumptions explicitly involving characteristic functions in connection with "block sums"; much of that theory is codified in [Ja]. There is a substantial development of limit theory of various kinds under weak dependence assumptions that involve covariances of certain multivariate Lipschitz functions of random variables from the "past" and "future" (in the spirit of, but much less restrictive than, say, the dependence coefficient $\rho(n)$ defined analogously to (3) and (8)); see e.g. [DDLLLP]. There is a substantial development of limit theory under weak dependence assumptions that involve dependence coefficients similar to $\alpha(n)$ in (3) but in which the classes of events are restricted to intersections of finitely many events of the form $\{X_k > c\}$ for appropriate indices k and appropriate real numbers c; for the use of such conditions in extreme value theory, see e.g. [LLR]. In recent years, there has been a considerable development of central limit theory under "projective" criteria related to martingale theory (motivated by Gordin's martingale-approximation technique — see [HH]); for details, see e.g. [Pe]. There are far too many other types of weak dependence conditions, of the general spirit of strong mixing conditions but less restrictive, to describe here; for more details, see e.g. [DDLLLP] or [Br, v1, Chapter 13].

References

[Br] R.C. Bradley. *Introduction to Strong Mixing Conditions*, Vols. 1, 2, and 3. Kendrick Press, Heber City (Utah), 2007.

[DDLLLP] J. Dedecker, P. Doukhan, G. Lang, J.R. León, S. Louhichi, and C. Prieur. *Weak Dependence: Models, Theory, and Applications*. Lecture Notes in Statistics 190. Springer-Verlag, New York, 2007.

[DMS] H. Dehling, T. Mikosch, and M. Sørensen, eds. *Empirical Process Techniques for Dependent Data*. Birkhäuser, Boston, 2002.

[De] M. Denker. The central limit theorem for dynamical systems. In: *Dynamical Systems and Ergodic Theory*, (K. Krzyzewski, ed.), pp. 33-62. Banach Center Publications, Polish Scientific Publishers, Warsaw, 1989.

[Do] P. Doukhan. *Mixing: Properties and Examples*. Springer-Verlag, New York, 1995.

[HH] P. Hall and C.C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, San Diego, 1980.

[IR] I.A. Ibragimov and Yu.A. Rozanov. *Gaussian Random Processes*. Springer-Verlag, New York, 1978.

[Io] M. Iosifescu. Doeblin and the metric theory of continued fractions: a functional theoretic solution to Gauss' 1812 problem. In: *Doeblin and Modern Probability*, (H. Cohn, ed.), pp. 97-110. Contemporary Mathematics 149, American Mathematical Society, Providence, 1993.

[Ja] A. Jakubowski. *Asymptotic Independent Representations for Sums and Order Statistics of Stationary Sequences*. Uniwersytet Mikołaja Kopernika, Toruń, Poland, 1991.

[LL] Z. Lin and C. Lu. *Limit Theory for Mixing Dependent Random Variables*. Kluwer Academic Publishers, Boston, 1996.

[LLR] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York, 1983.

[MT] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability* (3rd printing). Springer-Verlag, New York, 1996.

[Pe] M. Peligrad. Conditional central limit theorem via martingale approximation. In: *Dependence in Probability, Analysis and Number Theory*, (I. Berkes, R.C. Bradley, H. Dehling, M. Peligrad, and R. Tichy, eds.), pp. 295-309. Kendrick Press, Heber City (Utah), 2010.

[Ri] E. Rio. *Théorie Asymptotique des Processus Aléatoires Faiblement Dépendants*. Mathématiques & Applications 31. Springer, Paris, 2000.

[Ro1] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* 42 (1956) 43-47.

[Ro2] M. Rosenblatt. *Markov Processes, Structure and Asymptotic Behavior*. Springer-Verlag, New York, 1971.

[Ro3] M. Rosenblatt. *Stationary Sequences and Random Fields*. Birkhäuser, Boston, 1985.

[Žu] I.G. Žurbenko. *The Spectral Analysis of Time Series*. North-Holland, Amsterdam, 1986.