Stat 8112 Lecture Notes
Stationary Stochastic Processes
Charles J. Geyer
April 29, 2012

1 Stationary Processes

A sequence of random variables X_1, X_2, ... is called a *time series* in the statistics literature and a (discrete time) *stochastic process* in the probability literature.

A stochastic process is *strictly stationary* if for each fixed positive integer k the distribution of the random vector

$$(X_{n+1}, \ldots, X_{n+k})$$

has the same distribution for all nonnegative integers n.

A stochastic process having second moments is *weakly stationary* or *second order stationary* if the expectation of X_n is the same for all positive integers n and for each nonnegative integer k the covariance of X_n and X_{n+k} is the same for all positive integers n.

2 The Birkhoff Ergodic Theorem

The Birkhoff ergodic theorem is to strictly stationary stochastic processes what the strong law of large numbers (SLLN) is to independent and identically distributed (IID) sequences. In effect, despite the different name, it is the SLLN for stationary stochastic processes.

Suppose X_1, X_2, ... is a strictly stationary stochastic process and X_1 has expectation (so by stationary so do the rest of the variables). Write

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

To introduce the Birkhoff ergodic theorem, it says

$$\overline{X}_n \xrightarrow{\text{a.s.}} Y, \tag{1}$$

where Y is a random variable satisfying $E(Y) = E(X_1)$. More can be said about Y, but we will have to develop some theory first.

The SSLN for IID sequences says the same thing as the Birkhoff ergodic theorem (1) except that in the SLLN for IID sequences the limit $Y = E(X_1)$ is constant. Here are two simple examples where we have convergence to a limit that is not constant almost surely. First, consider a sequence in which $X_n = X_1$ almost surely for all n. Then, of course, $\overline{X}_n = X_1$, for all n, and $\overline{X}_n \xrightarrow{\text{a.s.}} X_1$. Second consider an IID sequence X_1, X_2, $\ldots$ and a single random variable Y that is not constant almost surely. Define $Z_n = X_n + Y$, so $\overline{Z}_n = \overline{X}_n + Y$. Then $\overline{Z}_n \xrightarrow{\text{a.s.}} E(X_1) + Y$. The amazing thing about the Birkhoff ergodic theorem is that the behavior of these simple examples illustrates the general situation.

2.1 Conditional Expectation

To understand the limit random variable in the Birkhoff ergodic theorem (1) we need to understand (i) measure-theoretic conditional expectation, (ii) measure-preserving transformations, and (iii) invariant sigma-algebras. This section and the two following sections explain these ideas.

A family of subsets of the sample space Ω is a *sigma-algebra* or a *sigma-field* if it contains Ω and is closed under complementation and countable unions (in which case it is also closed under countable intersections by De Morgan's laws).

The sigma-algebra concept is fundamental in probability theory because a probability measure is a function on a sigma-algebra, a map $P : \mathcal{A} \to [0, 1]$ that has the familiar axiomatic properties of probability theory (countable additivity and $P(\Omega) = 1$). The triple $(\Omega, \mathcal{A}, P)$ is called a *probability space*.

A random variable is a *measurable function on the sample space*, that is, a function $X : \Omega \to \mathbb{R}$ such that

$$X^{-1}(B) = \{\, \omega \in \Omega : X(\omega) \in B \,\}$$

is an element of $\mathcal{A}$ for every Borel set $B \in \mathcal{B}$, where $\mathcal{B}$ is the Borel sigma-algebra for $\mathbb{R}$, which is the smallest sigma-algebra containing all intervals.

A random variable is $\mathcal{C}$-measurable, where $\mathcal{C}$ is a sub-sigma-algebra of $\mathcal{A}$, if $X^{-1}(B)$ is an element of $\mathcal{C}$ for every $B \in \mathcal{B}$.

The *conditional expectation of X given $\mathcal{C}$*, written $E(X \mid \mathcal{C})$, is a $\mathcal{C}$-measurable random variable Y having the following property

$$E(Y I_C) = E(X I_C), \qquad C \in \mathcal{C}. \tag{2}$$

Such random variables always exist and are unique up to redefinition on sets of measure zero, that is, if Y_1 and Y_2 are two such random variables then $Y_1 = Y_2$ almost surely (see Chapter 23 of Fristedt and Gray, 1996, or Sections 10–12 of Chapter V in Feller, 1971, or Section 9.1 of Chung, 1974).

When we want to emphasize the non-uniqueness of conditional expectations, we say that Y_1 and Y_2 are different *versions* of the conditional expectation of X given $\mathcal{C}$. But usually we don't emphasize the non-uniqueness and just say either is "the" conditional expectation.

The notion of conditioning on a sigma-algebra rather than a random variable or random variables will be new to those without previous exposure to measure theory. How does this relate to the commonplace notation of a conditional expectation, where we condition on a set of random variables rather than a sigma-algebra? The former is a special case of the latter where the sigma-algebra in question is the one *generated* by those random variables, which is defined as follows: if $\mathcal{X}$ is a family of random variables, then the sigma-algebra generated by $\mathcal{X}$ is the smallest sub-sigma-algebra $\mathcal{C}$ of $\mathcal{A}$ having the property that every $X \in \mathcal{X}$ is $\mathcal{C}$-measurable. The notion of conditioning on a sigma-algebra is more general than the notion of conditioning on a family of random variables in that not every sub-sigma-algebra is generated by a family of random variables.

Measure-theoretic conditional expectation satisfies the usual laws of conditional expectation

$$E\{E(X \mid \mathcal{B})\} = E(X) \tag{3}$$

and

$$E(YX \mid \mathcal{C}) = YE(X \mid \mathcal{C}), \qquad \text{almost surely,} \tag{4}$$

for any random variables X and Y such that X has expectation and Y is $\mathcal{C}$-measurable. Also (3) can be improved to

$$E\{E(X \mid \mathcal{C}) \mid \mathcal{D}\} = E(X \mid \mathcal{D}), \qquad \text{almost surely,} \tag{5}$$

whenever $\mathcal{D} \subset \mathcal{C}$.

Measure-theoretic conditional expectation also satisfies the usual laws of for expectations with the exception that one must add "almost surely." For example, monotone convergence: if $X_n \uparrow X$ and all these random variables have expectation, then

$$E(X_n \mid \mathcal{C}) \uparrow E(X \mid \mathcal{C}), \quad \text{almost surely.}$$

As usual, probability is just expectation of indicator functions: the *conditional probability of an event A given a sigma-algebra $\mathcal{C}$* is

$$P(A \mid \mathcal{C}) = E(I_A \mid \mathcal{C}).$$

Measure-theoretic conditional probability also satisfies the usual laws of for expectations with the exception that one must add "almost surely." For example, countable additivity : if A_n, $n = 1, 2, \ldots$ is measurable partition of the sample space, then

$$\sum_{n=1}^{\infty} P(A_n \mid \mathcal{C}) = 1, \quad \text{almost surely.}$$

2.2 Measure-Preserving Transformations

Let $(\Omega, \mathcal{A}, P)$ be a probability space. A measurable function $T : \Omega \to \Omega$ is *measure preserving* if $P \circ T^{-1} = P$, meaning $P(T^{-1}(A)) = P(A)$ for all $A \in \mathcal{A}$.

Any measure-preserving transformation T and random variable Y generate a strictly stationary stochastic process

$$X_n = Y \circ T^{n-1}, \qquad n = 1, 2 \ldots, \tag{6}$$

where the right side means the function

$$\omega \mapsto Y(T(\ldots T(T(\omega)) \ldots)),$$

wherein there are $n - 1$ applications of the function T.

Conversely, the law of any strictly stationary stochastic process X considered as a random element of $\mathbb{R}^\infty$ can be determined from a measure-preserving transformation that is the shift transformation on $\mathbb{R}^\infty$ defined by

$$T((x_1, x_2, \ldots)) = (x_2, x_3, \ldots) \tag{7}$$

(Section 6.2 in Breiman (1968) gives a complete discussion of this).

2.3 The Invariant Sigma-Algebra

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let T be a measure preserving transformation on it. The *invariant* sigma-algebra is

$$\mathcal{S} = \{\, A \in \mathcal{A} : T^{-1}(A) = A \,\}.$$

Let X be a strictly stationary stochastic process defined from this measure-preserving transformation via (6).

Theorem 1 (Birkhoff Ergodic). *If $\overline{X}_n$ is the sample mean for a strictly stationary stochastic process, then*

$$\overline{X}_n \xrightarrow{a.s.} E(X_1 \mid \mathcal{S}), \tag{8}$$

where $\mathcal{S}$ is the invariant sigma-algebra.

For a proof, see Fristedt and Gray (1996, Section 28.3).

2.4 Ergodic Stationary Stochastic Processes

We say a sub-sigma-algebra $\mathcal{C}$ of a probability space is *trivial* if every event in $\mathcal{C}$ has probability zero or one. This is the same thing as saying every $\mathcal{C}$-measurable random variable is almost surely constant, and this implies

$$E(X \mid \mathcal{C}) = E(X), \qquad \text{almost surely,}$$

by the properties of conditional expectation.

A stationary stochastic process is *ergodic* if the invariant sigma-algebra is trivial. Thus for an ergodic strictly stationary stochastic process the Birkhoff ergodic theorem says

$$\overline{X}_n \xrightarrow{\text{a.s.}} E(X_1),$$

which is the same as the conclusion of the SLLN for IID sequences.

3 Mixing and Mixing Coefficients

A strictly stationary stochastic process that is determined by a measure-preserving transformation T is *mixing* if

$$\lim_{k \to \infty} Q\big(A \cap T^{-k}(B)\big) = Q(A)Q(B), \qquad A, B \in \mathcal{B}. \tag{9}$$

Mixing implies ergodicity (Fristedt and Gray, 1996, Theorem 6 in Section 28.5).

For a stochastic process X_1, X_2, $\ldots$, define

$$\mathcal{A}_1^n = \sigma(X_1, \ldots, X_n), \tag{10}$$

the sigma-algebra generated by the first n random variables, and

$$\mathcal{A}_n^\infty = \sigma(X_n, X_{n+1}, \ldots), \tag{11}$$

the sigma-algebra generated by the (infinite) family of random variables that is the subsequence starting at n. And define $L_2(\mathcal{A})$ to be the family of all square integrable $\mathcal{A}$-measurable random variables.

Define

$$\alpha_n = \sup_{\substack{m \in \mathbb{N} \\ A \in \mathcal{A}_1^m \\ B \in \mathcal{A}_{m+n}^\infty}} |P(A \cap B) - P(A)P(B)| \tag{12}$$

$$\beta_n = \sup_{\substack{m \in \mathbb{N} \\ A_1, \ldots, A_I \in \mathcal{A}_1^m \\ A_1, \ldots, A_I \text{ partition } \Omega \\ B_1 \ldots B_J \in \mathcal{A}_{m+n}^\infty \\ B_1, \ldots, B_J \text{ partition } \Omega}} \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)| \tag{13}$$

$$\rho_n = \sup_{\substack{m \in \mathbb{N} \\ X \in L_2(\mathcal{A}_1^m) \\ Y \in L_2(\mathcal{A}_{m+n}^\infty)}} |\operatorname{cor}(X, Y)| \tag{14}$$

$$\phi_n = \sup_{\substack{m \in \mathbb{N} \\ A \in \mathcal{A}_1^m \\ B \in \mathcal{A}_{m+n}^\infty \\ P(B) > 0}} |P(A \mid B) - P(A)| \tag{15}$$

The stochastic process is said to be

- *alpha-mixing* or *strongly mixing* if $\alpha_n \to 0$,
- *beta-mixing* or *absolutely regular* if $\beta_n \to 0$,
- *rho-mixing* if $\rho_n \to 0$, or
- *phi-mixing* if $\phi_n \to 0$.

Bradley (1986) establishes the inequalities

$$2\alpha_n \le \beta_n \le \phi_n$$
$$4\alpha_n \le \rho_n \le 2\sqrt{\phi_n}$$

and also states that alpha-mixing implies mixing. Hence

phi-mixing implies beta-mixing implies alpha-mixing implies mixing
phi-mixing implies rho-mixing implies alpha-mixing implies mixing

4 Cramér-Wold and Centering

The Cramér-Wold lemma says a sequence of random vectors X_1, X_2, ... converges in law to a random vector X if and only if for every nonrandom vector v

$$v^T X_n \xrightarrow{\mathcal{L}} v^T X$$

(proved using characteristic functions). Thus the multivariate central limit theorem (CLT) can be derived from the univariate CLT. This is the reason why only univariate CLT or infinite-dimensional CLT are discussed in the literature.

If X_1, X_2, ... is a (strictly or weakly) stationary stochastic process, then so is Y_1, Y_2, ... defined by $Y_n = X_n - \mu$, where $\mu = E(X_1)$. We say the Y process is *centered*, meaning $E(Y_n) = 0$ for all n. Clearly, saying

$$\tau_n \overline{Y}_n \xrightarrow{w} \text{Normal}(0, \sigma^2)$$

is the same thing as saying

$$\tau_n (\overline{X}_n - \mu) \xrightarrow{w} \text{Normal}(0, \sigma^2)$$

this we can infer the general CLT from the CLT for centered processes. Thus much of the literature only discusses centered processes.

5 Uniform Integrability and the CLT

Uniform integrability was defined and two theorems about it were stated in Section 9 of the handout about the Wilks, Rao, and Wald tests.

Let X_1, X_2, ... be a strictly stationary centered stochastic process having second moments, and define

$$S_n = \sum_{i=1}^{n} X_i$$
$$\sigma_n^2 = \text{var}(S_n)$$

Then S_n / σ_n is a standardized random variable (mean zero, standard deviation one).

Denker (1986, Theorem 3) proves the following theorem.

Theorem 2. *Suppose S_n and σ_n are defined above for a strictly stationary centered stochastic process having second moments and $\sigma_n^2 \to \infty$. Then*

$$\frac{S_n}{\sigma_n} \xrightarrow{w} \text{Normal}(0, 1) \tag{16}$$

holds if and only if S_n^2/σ_n^2 is a uniformly integrable sequence.

This gives us a CLT for stationary processes that is remarkably simple, but at the same time fairly useless because verification of uniform integrability is hard. Denker (1986) does remark that one can use the well known criterion (Billingsley, 1999, pp. 31–32) that a sequence Y_1, Y_2, $\ldots$ is uniformly integrable if there exists a $\delta > 0$ such that

$$\sup_n E\{|Y_n|^{1+\delta}\} < \infty.$$

Applying this to Theorem 2 gives

$$E\{S_n^{2+\delta}\} = O(\sigma_n^{2+\delta})$$

as a sufficient (but not necessary) condition for asymptotic normality of S_n/σ_n, but this condition is not easy to establish either.

The theorem does not give the rate of convergence, that is, we do not know (and do not need to know) how σ_n^2 varies as a function of n. So an interesting question is: if the CLT holds, does it hold at the usual $\sqrt{n}$ rate? Denker (1986, Theorem 2) gives a theorem of Ibragimov.

Theorem 3. *Under the assumptions of Theorem 2, if the CLT* (16) *holds, then*

$$\sigma_n^2 = nh(n),$$

where h is a slowly varying function on $[1, \infty)$*.*

A function h is *slowly varying* (Feller, 1971, Section 8 of Chapter VIII) if

$$\frac{h(tx)}{h(x)} \to 1, \qquad \text{as } x \to \infty,$$

for all $t > 0$. Examples of slowly varying functions are constant functions and $x \mapsto (\log x)^r$, for any $r > 0$ (no matter how large). Non-examples are $x \mapsto x^\epsilon$ for any $\epsilon > 0$ (no matter how small). Thus the CLT for stationary processes requires close to $\sqrt{n}$ rate, but not exactly $\sqrt{n}$ rate.

6 Alpha-Mixing and the CLT

Bradley (1985, Theorem 0) states the following theorem, attributing it to Ibragimov (see also Theorem 1.7 in Peligrad, 1986, and Theorem 18.5.3 in Ibragimov and Linnik, 1971).

Theorem 4. *Suppose X_1, X_2, ... is a strictly stationary centered stochastic process, $\sigma_n^2 \to \infty$, and either of the following two conditions holds.*

(i) *For some $\delta > 0$*

$$E(|X_1|^{2+\delta}) < \infty \quad \textit{and} \quad \sum_{n=1}^{\infty} \alpha_n^{\delta/(2+\delta)} < \infty. \tag{17}$$

(ii) *For some $C < \infty$*

$$|X_1| < C, \textit{ almost surely,} \quad \textit{and} \quad \sum_{n=1}^{\infty} \alpha_n < \infty. \tag{18}$$

Then

$$\sigma^2 = E(X_1^2) + 2 \sum_{k=2}^{\infty} E(X_1 X_k), \tag{19}$$

the infinite series in (19) *being absolutely summable. If $\sigma > 0$, then*

$$n^{-1/2} S_n \xrightarrow{w} \text{Normal}(0, \sigma^2), \tag{20}$$

Peligrad (1986) states that Theorem 4 is about as sharp as possible, citing counterexamples due to Davydov where the mixing is just a little bit slower than Theorem 4 requires and the CLT fails to hold.

Finally we can do an example that is a real application. A stationary stochastic process is *m-dependent* if $a_n = 0$ for all $n > m$. If $E(|X_1|^{2+\delta}) < \infty$ for some $\delta > 0$ and the process is m-dependent, then the CLT holds.

A times series is *moving average of order q*, abbreviated MA(q), if

$$X_n = Y_n - \sum_{i=1}^{q} \theta_i Y_{n-i},$$

(Box, Jenkins, and Reinsel, 1994, Section 3.1.4) where Y_{1-q}, Y_{2-q}, ... are IID. Clearly an MA(q) time series is strictly stationary and q-dependent. Thus, if the Y_n have $2 + \delta$ moments for some $\delta > 0$, the CLT holds.

7 Rho-Mixing and the CLT

Peligrad (1986, Theorem 2.2, Remark 2.2, and Theorem 2.3) gives the following theorem.

Theorem 5. *Suppose X_1, X_2, $\ldots$ is a second order stationary centered stochastic process having second moments, $\sigma_n^2 \to \infty$, and either of the following two conditions holds.*

(i) *For some $\delta > 0$*

$$E(|X_1|^{2+\delta}) < \infty \tag{21}$$

(ii)

$$\sum_{n=1}^{\infty} \frac{\rho_n}{n} < \infty. \tag{22}$$

Then $n^{-1}\sigma_n^2$ converges, and (20) *holds, where $\sigma^2 = \lim_n n^{-1}\sigma_n^2$.*

8 Phi-Mixing and the CLT

Peligrad (1986, Theorem 3.1) gives the following theorem.

Theorem 6. *Suppose X_1, X_2, $\ldots$ is a second order stationary phi-mixing centered stochastic process having second moments, $\sigma_n^2 \to \infty$, and the Lindeberg condition*

$$\lim_{n \to \infty} \sum_{i=1}^{n} E\{X_i^2 I(X_i^2 > \epsilon^2 \sigma_n^2)\} = 0, \qquad \text{for every } \epsilon > 0.$$

holds. Then (16) *holds.*

References

Billingsley, P. (1999). *Convergence of Probability Measures*, second edition. New York: Wiley.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis*, 3rd edition. Englewood Cliffs, NJ: Prentice-Hall.

Bradley, R. C. (1985). On the central limit question under absolute regularity. *Annals of Probability*, **13**, 1314–1325.

Bradley, R. C. (1986). Basic properties of strong mixing conditions. In Eberlein and Taqqu (1986).

Breiman, L. (1968). *Probability.* Addison-Wesley. Republished by SIAM, 1992.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd edition. New York: Academic Press.

Denker, M. (1986). Uniform integrability and the central limit theorem for strongly mixing processes. In Eberlein and Taqqu (1986).

Eberlein, E. and Taqqu, M. S., eds. (1986). *Dependence in Probability and Statistics: A Survey of Recent Results* (Oberwolfach, 1985). Boston: Birkhäuser.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications,* 2nd edition, volume 2. New York: Wiley.

Fristedt, B. E. and Gray, L. F. (1996). *A Modern Approach to Probability Theory*. Boston: Birkhäuser.

Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables* (edited by J. F. C. Kingman). Groningen: Wolters-Noordhoff.

Peligrad, M. (1986). Recent advances in the central limit theorem and its weak invariance principle for mixing sequences of random variables (a survey). In Eberlein and Taqqu (1986).