LECTURE 7

STATIONARITY, ERGODICITY, WEAK DEPENDENCE

The material is adapted from Peter Phillips' Lecture Notes on Stationary and Nonstationary Time Series and White (1999).

Often econometricians have to deal with data sets that come in the form of a time series or stochastic process, a collection of observations on the same variable (or vector of variables) indexed by the date of measurement of each observation. The data is usually collected at equally spaced dates (daily, weekly, monthly and etc.) and indexed by $t = 1, \dots, n$:

$$\{X_t : t = 1, \dots, n\}.$$

The index t measures passage of time, and X_t measures evolution of the process in time. It is usually assumed that the observed sample is only a segment of the process that started in the infinite past and will go on indefinitely:

$$\{\dots, X_{-1}, X_0, X_1, X_2, \dots, X_n, X_{n+1}, \dots\}.$$

Preliminaries

A random experiment is an experiment the outcome of which cannot be predicted with certainty. A sample space, Ω, is a collection (set) of all possible outcomes of the random experiment.

Definition 1 ***(σ-field)*** *A collection $\mathcal{F}$ of subsets of a set Ω is a σ-field provided that*

(i) $\Omega \in \mathcal{F}$.

(ii) *if $F \in \mathcal{F}$ then $F^c \in \mathcal{F}$.*

(iii) *if $F_i \in \mathcal{F}$ for all $i = 1, 2, \dots$ then $\cup_{i=1}^{\infty} F_i \in \mathcal{F}$.*

A simple example of a σ-field is $\{\emptyset, \Omega\}$. Given an event $F \subset \Omega$, the smallest σ-field that includes F is given by $\mathcal{F} = \{F, F^c, \emptyset, \Omega\}$. We say that such $\mathcal{F}$ is generated by F. For example, let $\Omega = \{\omega_1, \omega_2, \omega_3\}$, then the smallest σ-field that includes $\{\omega_1\}$ is $\mathcal{F} = \{\{\omega_1\}, \{\omega_2, \omega_3\}, \emptyset, \{\omega_1, \omega_2, \omega_3\}\}$.

A pair $(\Omega, \mathcal{F})$ is called a measurable space. The events in σ-field are the elements for which it is possible to assign probabilities without violating the axioms of probabilities given below.

Definition 2 ***(Probability)*** *A mapping $P : \mathcal{F} \to [0, 1]$ is called a probability measure provided that*

(i) $P(\Omega) = 1$.

(ii) *For any $F \in \mathcal{F}$, $P(F) \geq 0$.*

(iii) *Let $F_i \in \mathcal{F}$ for all $i = 1, 2, \dots,$ and $F_i \cap F_j = \emptyset$ for $i \neq j$, then $P(\cup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} P(F_i)$.*

The triple $(\Omega, \mathcal{F}, P)$ is called a probability space. Note that probabilities are defined on the elements of $\mathcal{F}$ and not on Ω.

Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be two measurable spaces, and $T : \Omega \to \Lambda$. For $F \subset \Omega$.

$$TF = \{T(\omega) : \omega \in F\} \subset \Lambda. \tag{1}$$

Let $G \subset \Lambda$, then

$$T^{-1}G = \{\omega \in \Omega : T(\omega) \in G\} \subset \Omega. \tag{2}$$

T is said to be measurable if $T^{-1}G \in \mathcal{F}$ for all $G \in \mathcal{G}$. If T is measurable, and $(\Omega, \mathcal{F}, P)$ is a probability space, then the probability of $G \in \mathcal{G}$ is given by $P\left(T^{-1}G\right) = P(\{\omega : T(\omega) \in G\})$.

When the sample space is given by R, the relevant σ-field is a σ-field generated by the collection of all open subsets of R.

Definition 3 ***(Borel σ-field)*** *The Borel σ-field $\mathcal{B}$ is the smallest σ-field that includes all open subsets of R.*

The Borel σ-field is an extremely rich collection of subsets of the real line. However, there exist subsets of real line that are not in $\mathcal{B}$. For such subsets probabilities cannot be defined.

Definition 4 ***(Random variable)*** *A random variable is a measurable function $X : \Omega \to R$.*

The σ-field generated by the random variable X, denoted as $\sigma(X)$, is the smallest σ-field of Ω with respect to which X is measurable. We can think of $\sigma(X)$ as the information contained in or revealed by X.

The Borel σ-field $\mathcal{B}$ can be also generated by the collection of all half-lines $(-\infty, x]$, $x \in R$. Recall that the cumulative distribution function (CDF) of X is $F(x) = P((-\infty, x])$. Consequently, from a CDF we can compute $P(B)$ for any $B \in \mathcal{B}$.

Definition 5 ***(Conditional expectation)*** *Let X be a random variable on $(\Omega, \mathcal{F}, P)$, and $\mathcal{G} \subset \mathcal{F}$ be a σ-field. The conditional expectation of X with respect to $\mathcal{G}$ is any $\mathcal{G}$ measurable random variable denoted as $E(X|\mathcal{G})$ such that $\int_G X dP = \int_G E(X|\mathcal{G})\, dP$ for any $G \in \mathcal{G}$.*

Stationarity

Let $h : \Omega \to R_\infty$. Let x be a sequence in R_∞, i.e. $x = (\ldots, x_{-1}, x_0, x_1, \ldots)$. The probability space is now given by $(R_\infty, \mathcal{B}_\infty, P)$, where $\mathcal{B}_\infty$ is a σ-field generated by the cylinder sets $\times_{j=-\infty}^{0} R \times_{j=1}^{n} B_i \times_{j=n+1}^{\infty} R$, where $B_i \in \mathcal{B}$. Consider the sequence (time-series) of random variables $\{X_t\}_{t=1}^n$, where $X_t(x)$ picks the t-th coordinate of x:

$$X_t(x) = x_t.$$

The sequence $\{X_t\}_{t=1}^n$ is measurable with respect to $\mathcal{B}_\infty$. Note that a single outcome in the sample space determines the entire infinite trajectory of $\{X_t\}$.

Definition 6 ***(Strict stationarity)*** *$\{X_t\}$ is strictly stationary if $(X_{t_1}, \ldots, X_{t_k})$ has the same joint distribution as $(X_{t_1+h}, \ldots, X_{t_k+h})$ for all k, h and $t_1, \ldots t_k$.*

An example of a stationary sequence is a sequence of iid random variables (however, a stationary sequence does not have to be iid). Note that the identically distributed property is not sufficient, since strict stationarity puts restrictions on all joint distributions. It is possible to construct a sequence of random variables having the same marginal distribution for all t, but with different joint distributions for different collections of t's Next, we discuss construction of strictly stationary sequences.

Definition 7 ***(Backshift transformation)*** *$S : R_\infty \to R_\infty$ such that*

$$\begin{aligned} x &= \left(\ldots, \; x_{-1}, \; x_0, \; x_1, \; \ldots \right) \text{ and} \\ S(x) &= \left(\ldots, \; x_0, \; x_1, \; x_2, \; \ldots \right) \end{aligned}$$

is called backshift transformation.

Thus, S transforms x by shifting each coordinate back by one location. Suppose that in $x \in R_\infty$, the value x_1 is located at coordinate 1, x_2 is located at 2, and etc. Using S and X_1 we can generate the entire process $\{X_t\}$:

$$\begin{aligned} X_1(x) &= x_1, \\ X_2(x) &= X_1(S(x)) = x_2, \\ X_3(x) &= X_1\left(S^2(x)\right) = x_3, \\ &\ldots \\ X_n(x) &= X_1\left(S^{n-1}(x)\right) = x_n. \end{aligned}$$

Let's define a set of infinite trajectories x that satisfy a certain condition (this set is denoted by E below). For example, for $B \in \mathcal{B}_\infty$, let

$$E = \{x : (X_n(x), \ldots) \in B\}.$$

We have

$$\begin{aligned} SE &= \{S(x) : x \in E\} \text{ (by the definition in (1))} \\ &= \{S(x) : (X_n(x), \ldots) \in B\} \text{ (by the definition of } E) \\ &= \{x : (X_{n-1}(x), \ldots) \in B\}, \end{aligned}$$

where the equality in the last line holds because S shifts the value that was at coordinate n to coordinate $n-1$ and by the definition of $X_{n-1}(x)$. Similarly, by the definition in (2),

$$\begin{aligned} S^{-1}E &= \{x : S(x) \in E\} \\ &= \{x : (X_n(S(x)), \ldots) \in B\} \text{ (by the definition of } E) \\ &= \{x : (X_{n+1}(x), \ldots) \in B\}. \end{aligned}$$

Definition 8 *Measurable transformation S is measure preserving if $P(E) = P(S^{-1}E)$ for all $E \in \mathcal{B}_\infty$.*

Theorem 1 *If $\{X_t\}$ is strictly stationary then there exists a measure preserving transformation S such that $X_t(x) = X_1(S^{t-1}(x))$.*

Proof. Let S be the backshift transformation. The condition $X_t(x) = X_1(S^{t-1}(x))$ is satisfied by the definition of the backshift transformation. We need to show that S is measure preserving. Pick some k and $t_1, \ldots, t_k$. Let B be an element of the Borel σ-field on R^k , and $E = \{x : (X_{t_1}(x), \ldots, X_{t_k}(x)) \in B\}$ (recall that the definition of strict stationarity involves only finite collections of random variables).

$$\begin{aligned} P(S^{-1}E) &= P\{x : (X_{t_1}(S(x)), \ldots, X_{t_k}(S(x))) \in B\} \\ &= P\{x : (X_{t_1+1}(x), \ldots, X_{t_k+1}(x)) \in B\} \text{ (by the definition of } S) \\ &= P\{x : (X_{t_1}(x), \ldots, X_{t_k}(x)) \in B\} \text{ (due to strict stationarity of } \{X_t\}) \\ &= P(E), \end{aligned}$$

where the equality before last follows from the strict stationarity of $\{X_t\}$. ■

The converse result is also true.

Theorem 2 *Let S be a measure preserving transformation, and $X : R_\infty \to R$. Then $\{X_t\}$ is constructed as $X_t(x) = X(S^{t-1}(x))$ is strictly stationary.*

Proof.

$$\begin{aligned} P(\{x : (X_{t_1}(x), \ldots, X_{t_k}(x)) \in B\}) &= P(E) \\ &= P(S^{-1}E) \\ &= P(S^{-h}E) \\ &= P(\{x : (X_{t_1+h}(x), \ldots, X_{t_k+h}(x)) \in B\}), \end{aligned}$$

where the second and third equalities are due to the fact that S is measure preserving. ■

Theorems 1 and 2 together imply that, if one is interested only in strictly stationary processes, then he can restrict his attention to the processes generated as

$$X_t(x) = X(S^{t-1}(x)), \text{ where } S \text{ is measure preserving.} \tag{3}$$

Now that we know that all strictly stationary processes are generated according to (3), we can investigate the implications of that result. The following theorem says that measurable functions of strictly stationary processes produce strictly stationary processes.

Theorem 3 *Suppose that*

(i) $\{X_t\}$ *is strictly stationary.*

(ii) $\varphi : R_\infty \to R$ *is measurable.*

(iii) $\{Y_t\}$ *is generated as* $Y_t = \varphi(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots)$.

Then $\{Y_t\}$ *is strictly stationary.*

Proof. Choose some arbitrary $t_1, \ldots, t_k$. Since $\{X_t\}$ is strictly stationary, $X_t(x) = X_1\left(S^{t-1}(x)\right)$, where S the backshift transformation and S is measure preserving. Hence, we can view each Y_t as a function of $x \in R_\infty$:

$$\begin{aligned} Y_t &= \varphi(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots) \\ &= \varphi(\ldots, X_1(S^{t-2}(x)), X_1(S^{t-1}(x)), X_1(S^t(x)), \ldots) \\ &\equiv Y_t(x). \end{aligned}$$

Let C be an element of the Borel σ-field on R^k, and consider the following event $(Y_{t_1}, \ldots Y_{t_k}) \in C$. Since Y's are functions of x, we can re-state this event in terms of possible trajectories x (of $\{X_t\}$). Hence,

$$\begin{aligned} P\{(Y_{t_1}, \ldots Y_{t_k}) \in C\} &= P\{x : (Y_{t_1}(x), \ldots Y_{t_k}(x)) \in C\} \\ &= PS^{-h}\{x : (Y_{t_1}(x), \ldots Y_{t_k}(x)) \in C\} \qquad \text{(because } S \text{ is measure preserving)} \\ &= P\left\{x : \left(Y_{t_1}\left(S^h(x)\right), \ldots Y_{t_k}\left(S^h(x)\right)\right) \in C\right\} \\ &= P\{x : (Y_{t_1+h}(x), \ldots Y_{t_k+h}(x)) \in C\} \\ &= P\{(Y_{t_1+h}, \ldots Y_{t_k+h}) \in C\}. \end{aligned}$$

■

Ergodicity

When studying the asymptotic properties of various estimators, we rely on convergence of sample averages. It turns out that strict stationarity by itself is not sufficient for convergence of sample averages. One needs another property: Ergodicity.

Ergodicity is concerned with a question whether temporal averages converge to population averages (expectations). Consider the following example. Let $\{X_t : t = 1, \ldots, n\}$ be a sequence of iid random variables with mean zero. Let Z be another mean zero random variable independent of $\{X_t\}$. Define $Y_t = X_t + Z$. Note that $EY_t = 0$, and $\{Y_t\}$ is stationary:

$$\begin{aligned} P(Y_{t_1} \le a_1, \ldots, Y_{t_k} \le a_k) &= P(X_{t_1} + Z \le a_1, \ldots, X_{t_k} + Z \le a_k) \\ &= P(Z \le a_1 - X_{t_1}, \ldots, Z \le a_k - X_{t_k}) \\ &= EP(Z \le a_1 - X_{t_1}, \ldots, Z \le a_k - X_{t_k} | X_{t_1}, \ldots, X_{t_k}) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(Z \le a_1 - x_{t_1}, \ldots, Z \le a_k - x_{t_k})\, dP_{X_{t_1}, \ldots, X_{t_k}}(x_1, \ldots, x_k) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} P(Z \le a_1 - x_{t_1}, \ldots, Z \le a_k - x_{t_k})\, dP_{X_{t_1+h}, \ldots, X_{t_k+h}}(x_1, \ldots, x_k) \\ &= P(X_{t_1+h} + Z \le a_1, \ldots, X_{t_k+h} + Z \le a_k) \\ &= P(Y_{t_1+h} \le a_1, \ldots, Y_{t_k+h} \le a_k), \end{aligned}$$

where the equality in the fourth line holds by independence of Z and $\{X_t\}$. While the process $\{Y_t\}$ is strictly stationary, at the same time we have

$$\begin{aligned} n^{-1}\sum_{t=1}^{n} Y_t &= n^{-1}\sum_{t=1}^{n} X_t + Z \\ &\to_p Z \\ &\neq EY_t. \end{aligned}$$

The problem with this example is that $\{Y_t\}$ shows too much temporal dependence: no matter how large h is, Y_t and Y_{t+h} have the same non-zero correlation.

Definition 9 ***(Invariance)*** *An event F is invariant under a transformation S if $F = S^{-1}F$.*

According to the definition and when S is the backshift transformation, once an invariant event occurs in some period, it will keep occurring in the future. In the above example, the invariant events are given by $\sigma(Z)$, the σ-field generated by the variable Z.

Definition 10 ***(Ergodic transformation)*** *Transformation S is ergodic if for all invariant events $F \in \mathcal{B}_\infty$, $P(F) = 0$ or $P(F) = 1$.*

Definition 11 ***(Ergodic process)*** *A process $\{X_t\}$ such that $X_t(x) = X_1\left(S^{t-1}(x)\right)$ is ergodic if S is ergodic.*

Ergodicity means that the invariant events are certain (such as the whole sample space) or can be ignored. In the absence of ergodicity, there is some positive probability of hitting an invariant event. As a result, it is impossible to fully sample from the sample space as $n \to \infty$.

Theorem 4 ***(Pointwise Ergodic theorem)*** *Suppose that*

(i) *S is measure preserving.*

(ii) $X_t(x) = X_1\left(S^{t-1}(x)\right)$.

(iii) $E|X_1| < \infty$.

Let $\mathcal{G}$ be the σ-field generated by the collection of all invariant events under S. Then, as $n \to \infty$,

$$P\left(n^{-1}\sum_{t=1}^{n} X_t \to E\left(X_1|\mathcal{G}\right)\right) = 1.$$

In the above example, the σ-field of the invariant events is given by $\sigma(Z)$. Therefore, the Pointwise Ergodic theorem suggests, we have that $n^{-1}\sum_{t=1}^{n} Y_t \to_{a.s.} E\left(Y_1|Z\right) = Z$.

The following result follows from the Pointwise Ergodic theorem.

Corollary 1 ***(SLLN for stationary and ergodic processes)*** *Let $\{X_t : t = 1,\ldots,n\}$ be a strictly stationary and ergodic process such that $E|X_1| < \infty$. Then, as $n \to \infty$,*

$$P\left(n^{-1}\sum_{t=1}^{n} X_t \to E\left(X_1\right)\right) = 1.$$

Proof. Since $\{X_t\}$ is strictly stationary, by the Pointwise Ergodic theorem, $n^{-1}\sum_{t=1}^{n} X_t \to E\left(X_1|\mathcal{G}\right)$ with probability 1, where $\mathcal{G}$ is generated by the collection of invariant events. Note that since the process is ergodic, for all $G \in \mathcal{G}$ we have that $P(G) = 0$ or $P(G) = 1$.

By the definition of conditional expectation, $E\left(X_1|\mathcal{G}\right)$ is a $\mathcal{G}$-measurable random variable such that

$$\int_G X_1 dP = \int_G E\left(X_1|\mathcal{G}\right) dP \text{ for any } G \in \mathcal{G}.$$

If G is such that $P(G) = 0$, then $\int_G X_1 dP = 0$, and we can simply set $E(X_1|\mathcal{G}) = E(X_1)$ on such sets G. If $P(G) = 1$ then using the facts that $P(G^c) = 0$ and $R_\infty = G \cup G^c$ we can write

$$\begin{aligned}
\int_G X_1 dP &= \int_G X_1 dP + 0 \\
&= \int_G X_1 dP + \int_{G^c} X_1 dP \\
&= \int_{R_\infty} X_1 dP \\
&= E(X_1).
\end{aligned}$$

Since $\mathcal{G}$ contains only trivial events (events that have probability zero or one, we conclude that $E(X_1|\mathcal{G}) = E(X_1)$ with probability 1. ■

The following theorem says that measurable functions of strictly stationary and ergodic processes are strictly stationary and ergodic.

Theorem 5 *Suppose that*

(i) $\{X_t\}$ *is strictly stationary and ergodic.*

(ii) $\varphi : R_\infty \to R$ *is measurable.*

(iii) $\{Y_t\}$ *is generated as* $Y_t = \varphi(\ldots, X_{t-1}, X_t, X_{t+1}, \ldots)$.

Then $\{Y_t\}$ *is strictly stationary and ergodic.*

Proof. Stationarity has been shown above. Let C_y be an invariant event for $\{Y_t\}$. We need to show that $P(C_y) = 0$ or 1. Let C be some event in $\mathcal{B}_\infty$, and suppose

$$C_y = \{(\ldots, Y_{t-1}, Y_t, Y_{t+1}, \ldots) \in C\}$$

is an invariant event. Note that we can write $Y_t(x) = \varphi(\ldots, X_{t-1}(x), X_t(x), X_{t+1}(x), \ldots,)$, and that C_y occurs if and only if C_x occurs, where

$$C_x = \{x : (\ldots, Y_{t-1}(x), Y_t(x), Y_{t+1}(x), \ldots) \in C\},$$

so that

$$P(C_y) = P(C_x).$$

Similarly, $S^{-1}C_y$ occurs if and only if $S^{-1}C_x$ occurs. Since C_y is invariant, $S^{-1}C_y = C_y$, and we have that $S^{-1}C_x = C_x$ or C_x is invariant for $\{X_t\}$. However, since $\{X_t\}$ is strictly stationary and ergodic, S must be ergodic and $P(C_x) = 0, 1$. It follows that $P(C_y) = 0, 1$ ■

According to the last theorem, if $\{X_t\}$ is strictly stationary and ergodic, then $\{Y_t\}$ generated as

$$Y_t = \sum_{j=-\infty}^{\infty} \beta_j X_{t-j}, \text{ where } \sum_{j=-\infty}^{\infty} |\beta_j| < \infty,$$

is also strictly stationary and ergodic

In a typical econometric application, it is assumed that $\{(X_t', U_t)\}$ is strictly stationary and ergodic sequence, and $Y_t = X_t'\beta + U_t$, $\beta \in R^k$. Since stationarity and ergodicity are preserved under measurable transformations, $\{(X_t X_t')\}$ and $\{(X_t U_t)\}$ are also strictly stationary and ergodic. Next, if it is assumed that $EX_1U_1 = 0$, $EX_{1j}^2 < \infty$ for all $j = 1, \ldots, k$, and EX_1X_1' is positive definite, it follows that $\widehat{\beta}_n = \beta + \left(n^{-1}\sum_{t=1}^n X_t X_t'\right)^{-1} n^{-1}\sum_{t=1}^n X_t U_t \to_{a.s.} \beta + (EX_1X_1')^{-1} 0 = \beta$.

The following theorem gives the necessary and sufficient conditions for ergodicity.

Theorem 6 *A measure preserving transformation S is ergodic if and only if*

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} P\left(F\cap S^{-i}G\right) = P(F)\,P(G) \tag{4}$$

for all $F, G \in \mathcal{B}_\infty$.

Proof. Suppose that S is ergodic. Define the random variable $1_G(x) = 1$ if $x \in G$ and zero otherwise. Note that $\{X_t\}$ generated as

$$X_t = 1_G\left(S^t(x)\right)$$

is a strictly stationary by Theorem 2, and since S is ergodic, it is also ergodic. By the SLLN for stationary and ergodic processes, with the probability 1,

$$\begin{aligned}\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} 1_G\left(S^i(x)\right) &= E1_G \\ &= P(G).\end{aligned}$$

Multiply both sides of the above equation by 1_F. Then,

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} 1_F 1_G\left(S^i(x)\right) = 1_F P(G).$$

Next taking expectations on both sides of the above equation, we obtain

$$\begin{aligned}P(F)\,P(G) &= E\left(\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} 1_F 1_G\left(S^i(x)\right)\right) \\ &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(1_F 1_G\left(S^i(x)\right)\right) \\ &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} P\left(F\cap S^{-i}G\right),\end{aligned}$$

where the second equality is due to the dominated convergence theorem.

Now, suppose that (4) holds for all $F, G \in \mathcal{B}_\infty$. Let E be an invariant event, so that $E = S^{-i}E$. Since (4) holds for all elements of $\mathcal{B}_\infty$, let's apply it with $F = G = E$:

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} P\left(E\cap S^{-i}E\right) = P(E)\,P(E).$$

On the other hand, since E is invariant,

$$\begin{aligned}\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} P\left(E\cap S^{-i}E\right) &= \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} P(E\cap E) \\ &= P(E).\end{aligned}$$

Therefore, $P(E) = (P(E))^2$ or $P(E)$ is either zero or one. ■

The event $S^{-h}G$ is G shifted h periods into the future. If F and $S^{-h}G$ were independent, we would have $P\left(F\cap S^{-h}G\right) = P(F)\,P\left(S^{-h}G\right) = P(F)\,P(G)$, since S is measure preserving. Thus ergodicity means that F, G and $S^{-h}G$ are on average asymptotically independent.

Mixing and weak dependence

Ergodicity is related to the concept of mixing, the capacity of the transformation S to mix the points in Ω, so that all non-negligible events are sampled from.

Definition 12 *Let S be measure preserving. S is said to be mixing if for all $F, G \in \mathcal{B}_\infty$,*

$$\lim_{n\to\infty} P\left(F \cap S^{-n}G\right) = P(F)\,P(G).$$

The following theorem establishes the relationship between mixing and ergodicity. Note that the converse result is not true. The proof of the above theorem can be based on the same argument as that of the second part of Theorem 6.

Theorem 7 *If S is mixing then it is ergodic.*

The idea of mixing is that distant events are almost independent. The following two measures of dependence are commonly used.

Definition 13 *Let $\mathcal{F}$ and $\mathcal{G}$ be two σ-fields. Define*

$$\begin{aligned}
\alpha(\mathcal{F},\mathcal{G}) &= \sup_{F\in\mathcal{F},G\in\mathcal{G}} \left|P(F\cap G) - P(F)\,P(G)\right|, \\
\phi(\mathcal{F},\mathcal{G}) &= \sup_{F\in\mathcal{F},G\in\mathcal{G}:P(G)>0} \left|P(F|G) - P(F)\right|.
\end{aligned}$$

For a time series $\{X_t\}$ (not necessary stationary), let $\mathcal{B}_r^s = \sigma(X_r,\ldots,X_s)$.

Definition 14 ***(Mixing coefficients)***

(a) ***(Strong** a- **mixing)*** $\alpha(m) = \sup_j \alpha\left(\mathcal{B}_{-\infty}^{j}, \mathcal{B}_{j+m}^{\infty}\right)$.

(b) ***(Uniform** ϕ- **mixing)*** $\phi(m) = \sup_j \phi\left(\mathcal{B}_{-\infty}^{j}, \mathcal{B}_{j+m}^{\infty}\right)$.

The mixing coefficients measure how much dependence exists between events separated by at least m periods.

For example, suppose that $\{X_t\}$ is a sequence of iid random variables. Then, $\phi(m) = 0$ for all $m > 0$. Now, suppose that $\{Y_t\}$ is generated from $\{X_t\}$ as $Y_t = X_t + \theta_1 X_{t-1} + \ldots + \theta_k X_{t-k}$. Then, $\phi(m) = 0$ for all $m > k$.

Definition 15 *A process $\{X_t\}$ is said to be α-mixing (strong mixing) if* $\lim_{m\to\infty}\alpha(m) = 0$. *Similarly, it is ϕ-mixing (uniform mixing) if* $\lim_{m\to\infty}\phi(m) = 0$.

A mixing process is also called *weakly dependent*. In general, a mixing process can be nonstationary: distributions are allowed to change over time.

Note that for $P(G) > 0$,

$$\begin{aligned}
\left|P(F|G) - P(F)\right| &= \left|\frac{P(F\cap G)}{P(G)} - P(F)\right| \\
&= \frac{\left|P(F\cap G) - P(F)\,P(G)\right|}{P(G)} \\
&\geq \left|P(F\cap G) - P(F)\,P(G)\right|,
\end{aligned}$$

where the last equality holds because $0 < P(G) \leq 1$. Therefore, uniform mixing implies strong mixing. Thus, strong mixing is a weaker concept.

Definition 16 *A mixing coefficient is said to be of size* $-a$, $a > 0$ *if for some* $\varepsilon > 0$ *it is* $O\left(m^{-a-\varepsilon}\right)$.

Mixing coefficients explicitly control the amount of dependence. A sequence shows less dependence as a increases.

The mixing property is preserved under transformations.

Theorem 8 *Let* $\{X_t\}$ *be* α*-mixing (*ϕ*-mixing) of size* $-a$, $a > 0$. *Suppose that* $Y_t = g\left(X_t, X_{t-1}, \ldots, X_{t-h}\right)$, *where* g *is measurable and* h *is finite. Then,* $\{Y_t\}$ *is* α*-mixing (*ϕ*-mixing) of size* $-a$ *as well.*

Proof. (Davidson, 1994) Let $\mathcal{B}_r^s = \sigma\left(X_r, \ldots, X_s\right)$, and let $\mathcal{C}_r^s = \sigma\left(Y_r, \ldots, Y_s\right)$. Since Y_t is a function of $X_t, X_{t-1}, \ldots, X_{t-h}$, it is measurable with respect to any σ-field on which $X_t, X_{t-1}, \ldots, X_{t-h}$ is measurable. Thus, $\mathcal{C}_{-\infty}^j \subset \mathcal{B}_{-\infty}^j$, and $\mathcal{C}_{j+m}^{\infty} \subset \mathcal{B}_{j+m-h}^{\infty}$ for all j and $m \geq h$ (Since Y's are functions of X's, a σ-field generated by Y's cannot contain more information than a σ-field generated by X's).

$$\begin{aligned}
\alpha_Y(m) &= \sup_j \sup_{F \in \mathcal{C}_{-\infty}^j, G \in \mathcal{C}_{j+m}^{\infty}} \left|P(F \cap G) - P(F) P(G)\right| \\
&\leq \sup_j \sup_{F \in \mathcal{B}_{-\infty}^j, G \in \mathcal{B}_{j+m-h}^{\infty}} \left|P(F \cap G) - P(F) P(G)\right| \\
&= \alpha_X(m-h) \\
&= O\left(m^{-a-\varepsilon}\right) \text{ for } \varepsilon > 0 \text{ and finite } h < m,
\end{aligned}$$

where the last inequality follows because h is fixed. The proof is identical in the case of uniform mixing. ■

The above theorem implies that if, for example, the vector sequence $\{(X_t', u_t)\}$ is mixing of size $-a$, then $\{X_t u_t\}$, $\{X_t X_t'\}$ are also mixing of size $-a$.

The following results give the SLLN and CLT for weakly dependent sequences. Note that stationarity is not required.

Theorem 9 ***(SLLN)*** *Let* $\{X_t\}$ *be a sequence such that*

(i) ϕ *is of size* $-r/(2r-1)$, $r \geq 1$, *or* α *is of size* $-r/(r-1)$, $r > 1$,

(ii) *for some* $\delta > 0$, $\sup_t E\left|X_t\right|^{r+\delta} < \Delta < \infty$.

Then, as $n \to \infty$,

$$P\left(n^{-1} \sum_{t=1}^n X_t - n^{-1} \sum_{t=1}^n E X_t \to 0\right) = 1.$$

The above theorem does not require that $n^{-1} \sum_{t=1}^n E X_t$ converges. However, if it converges, then, with probability, one $n^{-1} \sum_{t=1}^n X_t$ converges to $\lim_{n \to \infty} n^{-1} \sum_{t=1}^n E X_t$

Theorem 10 ***(CLT)*** *Let* $\{X_t\}$ *be a sequence such that* $E X_t = 0$, *and*

(i) ϕ *is of size* $-r/2(r-1)$, $r \geq 2$ *or* α *of size* $-r/(r-2)$, $r > 2$,

(ii) $\sup_t E\left|X_t\right|^r < \Delta < \infty$.

If $\omega_n = Var\left(n^{-1/2} \sum_{t=1}^n X_t\right) > \delta > 0$ *for all* n *sufficiently large, then, as* $n \to \infty$,

$$n^{-1/2} \sum_{t=1}^n X_t / \omega_n^{1/2} \to_d N(0,1).$$

In the above two results, note the trade-off between the amount of temporal dependence and likelihood of observing outliers (moment restrictions). Larger r means a smaller probability of observing outliers. At the same time, larger r means that the mixing coefficients in the above LLN and CLT theorems are of smaller size and therefore there is more temporal dependence. For example, for $r > 1$ and $h > 0$,

$$\frac{r}{r-1} - \frac{r+h}{r+h-1} = \frac{h}{(r-1)(r+h-1)} > 0.$$

The CLT can be extended to the case of random vectors.

Lemma 1 ***(Cramer-Wold device)*** *Let $\{Z_n\}$ be a sequence of random k-vectors. Suppose that Z is a random k-vector. $Z_n \to_d Z$ if and only if, for all $\lambda \in R^k$ (non-random) such that $\lambda'\lambda = 1$, $\lambda' Z_n \to_d \lambda' Z$.*

Definition 17 *Let $\{M_n\}$ be a sequence of $k \times k$ matrices. Let $\underline{e}_n$ be the smallest eigenvalue of M_n. Then, M_n is said to be uniformly positive definite if, for all n sufficiently large, $\underline{e}_n > \delta > 0$ uniformly in n.*

Corollary 2 *Let $\{X_t\}$ be a sequence of k-vectors such that $EX_t = 0$, and*

(i) *ϕ is of size $-r/2(r-1)$, $r \geq 2$ or α of size $-r/(r-2)$, $r > 2$,*

(ii) *$\sup_t E|X_{tj}|^r < \Delta < \infty$ for all $j = 1, \ldots, k$.*

If $\Omega_n = Var\left(n^{-1/2}\sum_{t=1}^n X_t\right)$ is uniformly positive definite for all n sufficiently large, then, as $n \to \infty$,

$$\Omega_n^{-1/2} n^{-1/2} \sum_{t=1}^n X_t \to_d N(0, I_k).$$

The proof will use the Minkowski's inequality. Suppose that $E|X|^r < \infty$ and $E|Y|^r < \infty$, $r \geq 1$. Then,

$$(E|X+Y|^r)^{1/r} \leq (E|X|^r)^{1/r} + (E|Y|^r)^{1/r}.$$

Proof. We will use the Cramer-Wold device with $Z_n = \Omega_n^{-1/2} n^{-1/2}\sum_{t=1}^n X_t$ and $Z \sim N(0, I_k)$. Let $\lambda \in R^k$ such that $\lambda'\lambda = 1$. Since $\lambda' Z \sim N(0, \lambda'\lambda) =^d N(0,1)$, by the Cramer-Wold device it is sufficient to show that $\lambda' Z_n = \lambda'\Omega_n^{-1/2} n^{-1/2}\sum_{t=1}^n X_t \to_d N(0,1)$ if $\lambda'\lambda = 1$. This can be achieved by verifying the conditions of Theorem 10 for the stochastic process $\left\{\lambda'\Omega_n^{-1/2} X_t\right\}_{t=1}^n$, where $\lambda'\lambda = 1$.

First, $E\left(\lambda'\Omega_n^{-1/2} X_t\right) = 0$. Second, by Theorem 8, $\lambda'\Omega_n^{-1/2} X_t$ is ϕ-mixing of size $-r/2(r-1)$, $r \geq 2$ or α-mixing of size $-r/(r-2)$, $r > 2$. Next, set $c_n = \Omega_n^{'-1/2}\lambda$.

$$\begin{aligned}
E\left|\lambda'\Omega_n^{-1/2} X_t\right|^r &= E\left|\sum_{j=1}^k c_{jn} X_{tj}\right|^r \\
&\leq \left(\sum_{j=1}^k |c_{jn}| (E|X_{jt}|^r)^{1/r}\right)^r \quad \text{(by Minkowski's)} \\
&< \Delta\left(\sum_{j=1}^k |c_{jn}|\right)^r \\
&\leq \Delta\left(k\sum_{j=1}^k |c_{jn}|^2\right)^{r/2} \quad \text{(by Cauchy-Schwartz)} \\
&= \Delta k^{r/2}\left(\lambda'\Omega_n^{-1}\lambda\right)^{r/2}.
\end{aligned}$$

Since Ω_n is symmetric and uniformly positive definite, we can write $\Omega_n = C_n E_n C_n'$, where E_n is the matrix of eigenvalues, and $C_n' C_n = C_n C_n' = I_k$. Set $d_n = C_n'\lambda$. We have $d_n' d_n = \lambda' C_n C_n' \lambda = 1$. Let $\underline{e}_n$ be the smallest eigenvalue of E_n (and, therefore, $\underline{e}_n^{-1}$ is the largest eigenvalue of E_n^{-1}). Since Ω_n is uniformly positive definite, there is some $\delta > 0$ such that $\underline{e}_n > \delta$. Next,

$$\begin{aligned}
\lambda'\Omega_n^{-1}\lambda &= d_n' E_n^{-1} d_n \\
&\leq \underline{e}_n^{-1}\sum_{i=1}^k d_{ni}^2 \\
&< \delta^{-1} < \infty. \qquad (5)
\end{aligned}$$

Thus,

$$\sup_t E\left|\lambda'\Omega_n^{-1/2}X_t\right|^r < \Delta k^{r/2}\delta^{-r/2}.$$

Lastly,

$$\begin{aligned} Var\left(\lambda'\Omega_n^{-1/2}n^{-1/2}\sum_{t=1}^{n}X_t\right) &= \lambda'\Omega_n^{-1/2}Var\left(n^{-1/2}\sum_{t=1}^{n}X_t\right)\Omega_n^{'-1/2}\lambda \\ &= \lambda'\Omega_n^{-1/2}\Omega_n\Omega_n^{'-1/2}\lambda \\ &= \lambda'\lambda \\ &= 1. \end{aligned}$$

Hence, by the univariate CLT, $\lambda'\Omega_n^{-1/2}n^{-1/2}\sum_{t=1}^{n}X_t \to_d N(0,1)$ for any λ such that $\lambda'\lambda = 1$. The conclusion follows. ■

Note that the result in (5) implies if Ω_n is uniformly positive definite then $\Omega_n^{-1} = O(1)$.

Long-run variance

For a scalar process $\{X_t\}$, $\lim_{n\to\infty}\omega_n$ (when exists) is called the long-run variance of $\{X_t\}$ (and it is denoted $\Omega = \lim_{n\to\infty}\Omega_n$ for a vector process). When $EX_t = 0$ for all t,

$$\begin{aligned} Var\left(n^{-1/2}\sum_{t=1}^{n}X_t\right) &= n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}EX_tX_s \\ &= n^{-1}\sum_{t=1}^{n}EX_t^2 + 2\sum_{j=1}^{n-1}n^{-1}\sum_{t=j+1}^{n}EX_tX_{t-j}. \end{aligned} \quad (6)$$

A sufficient condition for the existence of the long-run variance is that, in addition to moments restrictions, the correlation between X_t and X_{t-j} goes to zero sufficiently fast. The covariance (correlation) between X_t and X_{t-j} is called autocovariance (autocorrelation) of order j. The following result establishes a relationship between autocovariances and mixing coefficients.

Lemma 2 *(White (1999), Corollary 6.17, page 155) Let $EX_t = 0$ and $E|X_t|^r < \infty$ for some $r \geq 2$ and all t. Then*

$$\begin{aligned} |EX_tX_{t-j}| &\leq 2\phi(j)^{1-1/r}\left(EX_{t-j}^2\right)^{1/2}\left(E|X_t|^r\right)^{1/r}, \text{ and} \\ |EX_tX_{t-j}| &\leq 2\left(2^{1/2}+1\right)\alpha(j)^{1/2-1/r}\left(EX_{t-j}^2\right)^{1/2}\left(E|X_t|^r\right)^{1/r}. \end{aligned}$$

Using Lemma 2, we can obtain conditions sufficient for the long-run variance to be bounded. This is an important result that will be used repeatedly in Lecture 8.

Lemma 3 *Let $\{X_t\}$ be mixing ϕ of size $-r/(r-1)$, $r \geq 2$ or α of size $-2r/(r-2)$, $r > 2$. Suppose that $EX_t = 0$, and $\sup_t E|X_t|^r \leq \Delta < \infty$. Then, $\Omega_n = O(1)$.*

Proof. By the norm inequality (see Davidson (1994), page 138) and since $r > 2$,

$$\left(EX_t^2\right)^{1/2} \leq \left(E|X_t|^r\right)^{1/r},$$

and therefore,

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}EX_t^2 &\leq \sup_t EX_t^2 \\ &\leq \sup_t\left(E|X_t|^r\right)^{2/r} \\ &< \Delta^{2/r} < \infty. \end{aligned}$$

Next, in the case of ϕ-mixing,

$$\begin{aligned}
\left| n^{-1} \sum_{t=j+1}^{n} EX_t X_{t-j} \right| &\leq n^{-1} \sum_{t=j+1}^{n} \left| EX_t X_{t-j} \right| \\
&\leq n^{-1} \phi\left(j\right)^{1-1/r} \sum_{t=j+1}^{n} 2 \left(EX_{t-j}^2\right)^{1/2} \left(E\left|X_t\right|^r\right)^{1/r} \\
&\leq \frac{n-j}{n} 2 \left(\sup_t EX_t^2\right)^{1/2} \left(\sup_t E\left|X_t\right|^r\right)^{1/r} \phi\left(j\right)^{1-1/r} \\
&\leq \frac{n-j}{n} 2\Delta^{2/r} \phi\left(j\right)^{1-1/r} \\
&\leq 2\Delta^{2/r} \phi\left(j\right)^{1-1/r}.
\end{aligned}$$

However, given the conditions imposed on the mixing coefficients, for some $\varepsilon > 0$, there exists a constant $K > 0$ such that

$$\phi\left(j\right) < Kj^{-r/(r-1)-\varepsilon} \text{ for } j = 1, 2, \ldots,$$

and therefore,

$$\begin{aligned}
\sum_{j=1}^{\infty} \phi\left(j\right)^{1-1/r} &< K \sum_{j=1}^{\infty} j^{-(r/(r-1)+\varepsilon)(1-1/r)} \\
&= K \sum_{j=1}^{\infty} j^{-1-\varepsilon(1-1/r)} \\
&< \infty.
\end{aligned}$$

By a similar argument, $n^{-1} \sum_{t=j+1}^{n} EX_t X_{t-j}$ is summable for $j \geq 1$ in the case of α-mixing. ■

Definition 18 ***(Covariance stationarity)*** *A process $\{X_t\}$ such that $EX_t^2 < \infty$ for all t is covariance stationary if $EX_t = \mu$ and $Cov\left(X_t, X_s\right) = \gamma\left(|s-t|\right)$ for all t, s.*

While strict stationarity requires that all joint finite dimensional distributions of $\{X_t\}$ remain unchanged when shifted in time, for a covariance stationary process only the first two moments remain the same. For a covariance stationary process, the autocovariances are independent of t and depend only on j, the time lag between X_t and X_{t-j}.

For a covariance stationary process, let $\gamma(j)$ be the j-th autocovariance. Note that the variance can be written as $\gamma(0)$. Then the long-run variance is

$$\omega = \gamma(0) + 2 \sum_{j=1}^{\infty} \gamma(j),$$

provided that

$$\sum_{j=1}^{\infty} |\gamma(j)| < \infty.$$

For autocovariances to be summable, $\gamma(j)$ has to go to zero at sufficiently fast rate as $j \to \infty$. Note that $\gamma(j) \to 0$ as $j \to \infty$ is implied by mixing.

In the case of a vector process $\{X_t\}$ such that $EX_t = 0$, (6) becomes

$$n^{-1} \sum_{t=1}^{n} EX_t X_t' + \sum_{j=1}^{n-1} n^{-1} \sum_{t=j+1}^{n} \left(EX_t X_{t-j}' + EX_{t-j} X_t'\right),$$

If $\{X_t\}$ is covariance stationary, its long-run variance-covariance matrix, when exists, is

$$\begin{aligned}\Omega &= EX_0X_0' + \sum_{j=1}^{\infty}\left(EX_jX_0' + EX_0X_j'\right) \\ &= \Gamma(0) + \sum_{j=1}^{\infty}\left(\Gamma(j) + \Gamma(j)'\right),\end{aligned}$$

where

$$\begin{aligned}\Gamma(j) &= EX_jX_0' \\ &= EX_tX_{t-j}' \text{ for all } t.\end{aligned}$$

When the long-run variance exists, the CLT can be stated as

$$n^{-1/2}\sum_{t=1}^{n} X_t \to_d N\left(0, \lim_{n\to\infty} \Omega_n\right),$$

which is $N\left(0, EX_0X_0' + \sum_{j=1}^{\infty}\left(EX_jX_0' + EX_0X_j'\right)\right)$ in the covariance stationary case. Recall that in the iid case, the asymptotic variance is EX_0X_0'. Thus, weak dependency of $\{X_t\}$ is reflected asymptotically through the sum of all autocovariances in the expression for the asymptotic variance of the averages.