Lecture 9

1 Ergodic Theorems

We will prove the mean and the pointwise ergodic theorems for stationary processes, and we will establish the decomposition of a stationary measure into its ergodic components, and apply the theory to stationary Markov processes.

Definition 1.1 [Stationary processes] Let $(\Omega,\mathcal{F})$ be a Polish space with Borel $\sigma$-algebra $\mathcal{F}$. A sequence of $S$-valued random variables $(X_n)_{n \in \mathbb{N}}$ is called a stationary sequence if $(X_1, X_2, \cdots)$ is equally distributed with $(X_{k+1}, X_{k+2}, \cdots)$ for any $k \in \mathbb{N}$. Similarly, a doubly infinite sequence $(X_n)_{n \in \mathbb{Z}}$ is stationary if $(X_n)_{n \in \mathbb{Z}}$ is equally distributed with $(X_{n+k})_{n \in \mathbb{Z}}$ for any $k \in \mathbb{Z}$.

Example 1.2 If $X$ is a Markov chain starting from a stationary distribution, then $(X_1, X_2, \cdots)$ is a stationary process. By Kolmogorov extension theorem, we can even extend it to a doubly infinite stationary sequence $(X_n)_{n \in \mathbb{Z}}$, where for any $n_0 \in \mathbb{Z}$, $(X_{n_0+i})_{i \in \mathbb{Z}}$ is distributed as the Markov starting from its stationary distribution at time $0$. We can think of $(X_n)_{n \in \mathbb{Z}}$ as the Markov starting from equilibrium at time $t = -\infty$.

Example 1.3 Starting from any periodic sequence $x = (x_1, x_2, \cdots)$ of period $k$, i.e., $x_n = x_{n+k}$ for all $n \in \mathbb{N}$, we can define a stationary sequence by letting $\mathbb{P}(X = T^i x) = \frac{1}{k}$ for each $0 \leq i \leq k-1$, where $T^i x = (x_{i+1}, x_{i+2}, \cdots)$.

We can cast stationary processes into a more general framework by regarding the sequence $(X_n)_{n \in \mathbb{N}}$ as a random variable taking values in the product space $\Omega := \mathbb{S}^\mathbb{N}$ with product $\sigma$-field $\mathcal{F}$. If $\mu$ denotes the law of $(X_n)_{n \in \mathbb{N}}$ on $(\Omega, \mathcal{F})$, and let $T$ denote the shift map $T(\omega_1, \omega_2, \cdots) = (\omega_2, \omega_3, \cdots)$ for $\omega = (\omega_1, \omega_2, \cdots) \in \Omega$, then the stationarity of $(X_n)_{n \in \mathbb{N}}$ is equivalent to the invariance of $\mu$ under the transformation $T : \Omega \to \Omega$, i.e., $\mu(A) = \mu(T^{-1} A)$ for all $A \in \mathcal{F}$.

Definition 1.4 [Measure preserving transformation] Let $(\Omega, \mathcal{F}, \mu)$ be a probability space. A measurable map $T : \Omega \to \Omega$ is called a measure preserving transformation if $\mu(A) = \mu(T^{-1} A)$ for all $A \in \mathcal{F}$.

Remark. Normally the definition of a MPT also assumes that $T$ has a measurable inverse $T^{-1}$. But the results we prove here do not require this assumption.

We just saw that every stationary sequence gives a probability measure on the sequence space with the shift map as the canonical measure preserving transformation. Conversely, if $T$ is a measure preserving transformation on $(\Omega, \mathcal{F}, \mu)$, and $f : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$, then $X_n := f(T^n \omega)$ is a stationary sequence taking values in $(S, \mathcal{S})$.

The map $T$ induces an operator $U$ on functions on $(\Omega, \mathcal{F}, \mu)$, namely, $Uf(\omega) = f(T\omega)$. Because $T$ is measure preserving, it is easy to see that

$$
\int_{\Omega} |f(\omega)|^p d\mu = \int_{\Omega} |f(T\omega)|^p d\mu = \int_{\Omega} |Uf(\omega)|^p d\mu,
$$

1
i.e., $U$ is an isometry on $L_p(\Omega, \mathcal{F}, \mu)$ for any $1 \leq p < \infty$. Furthermore, $U$ is a unitary operator on $L_2(\Omega, \mathcal{F}, \mu)$, since for any $f, g \in L_2(\Omega, \mathcal{F}, \mu),$

$$\langle f, g \rangle = \int_\Omega f(\omega)g(\omega)d\mu = \int_\Omega f(T\omega)g(T\omega)d\mu = \langle Uf, Ug \rangle.$$

The collection of sets which are invariant with respect to a measure preserving transformation will play an important role in ergodic theory.

**Definition 1.5 [Invariant sets and $\sigma$-field, ergodicity]** Let $T$ be a measure preserving transformation on $(\Omega, \mathcal{F}, \mu)$. A set $A \in \mathcal{F}$ is called invariant if $\mu(A \Delta T^{-1}(A)) = 0$, where $A \Delta B := (A \setminus B) \cup (B \setminus A)$ denotes the symmetric difference. The class of invariant sets $\mathcal{I}$ form a $\sigma$-field, called the invariant $\sigma$-field. The invariant measure $\mu$ is called ergodic for the transformation $T$, if $\mathcal{I}$ is trivial in the sense that $\mu(A) \in \{0, 1\}$ for all $A \in \mathcal{I}$.

We can now state the following convergence result for functionals of a stationary process, called the ergodic theorems.

**Theorem 1.6 [Ergodic theorems]** Let $T$ be a measure preserving transformation on the probability space $(\Omega, \mathcal{F}, \mu)$, and let $f \in L_p(\Omega, \mathcal{F}, \mu)$ for some $1 \leq p < \infty$. Then

$$\lim_{n \to \infty} A_nf(\omega) := \lim_{n \to \infty} \frac{f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)}{n} = \mathbb{E}_\mu[f|\mathcal{I}],$$

(1.1)

where the convergence is a.s. and in $L_p$. When $\mu$ is ergodic for $T$, $\mathbb{E}_\mu[f|\mathcal{I}] = \mathbb{E}_\mu[f]$ a.s. Conversely, if $A_nf(\omega) \to C_f$ $\mu$-a.s. to some constant $C_f$ for every bounded measurable $f$, then $\mu$ must be ergodic.

**Remark.** The last claim in Theorem 1.6 is easily seen by applying (1.1) to $f(\omega) = 1_E(\omega)$ for an invariant set $E \in \mathcal{I}$ with $\mu(E) = (0, 1)$. The a.s. convergence in Theorem 1.6 is known as the pointwise or individual ergodic theorem, due to Birkhoff. The $L_p$ convergence in Theorem 1.6 is known as mean ergodic theorems, with the $L_2$ version due to von Neumann. When $\mu$ is ergodic, we see that the empirical distribution of $\omega, T\omega, \cdots, T^n\omega$ converge weakly to $\mu$. In Statistical Physics, ergodic theorem is often phrased as time average equals ensemble average, where ensemble average refers to average w.r.t. the ergodic measure $\mu$.

**Proof of the mean ergodic theorems.** We start with the $L_2$ case. Denote $H := L_2(\Omega, \mathcal{F}, \mu)$, and let $H_0 := \{f \in H : Uf = f\}$ be the space of eigen-functions of $U$ with eigenvalue 1, where $Uf(\omega) = f(T\omega)$. Let

$$A_nf = \frac{1}{n} \sum_{i=0}^{n-1} U^i f.$$ 

Then $\|A_nf\|_2 \leq \|f\|_2$ for all $f \in H$, and $A_nf = f$ for all $f \in H_0$. If $f = (I - U)g$ for some $g \in H$, then $A_nf = \frac{2^n - U^n g}{n} \to 0$ as $n \to \infty$. Since $A_n$ is a bounded operator, by approximation, $A_nf \to 0$ for all $f \in \text{Range}(I - U)H$.

We claim that $\text{Range}(I - U)H = H_0^\perp$. Indeed, for any $g \in H$ and $h \in H_0$, since $U$ is unitary, we have

$$\langle (I - U)g, h \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle U^{i-1}(I - U)g, U^{i-1}h \rangle = \frac{1}{n} \sum_{i=1}^{n} \langle U^{i-1}(I - U)g, h \rangle = \frac{1}{n} \langle (I - U^n)g, h \rangle.$$

Since \( n \in \mathbb{N} \) is arbitrary, we have \( \langle (I - U)g, h \rangle = 0 \) for all \( g \in H \) and \( h \in H_0 \), and hence \( \text{Range}(I - U)H \perp H_0 \).

To verify \( \text{Range}(I - U)H = H_0^\perp \), it only remains to show that if \( h \in \overline{\text{Range}(I - U)H}^\perp \), then \( h \in H_0 \). Indeed, if \( \langle f, h \rangle = 0 \) for all \( f \in \overline{\text{Range}(I - U)H} \), then in particular \( \langle (I - U)g, h \rangle = \langle g, (I - U^*)h \rangle = 0 \) for all \( g \in H \). Therefore \( U^*h = h \). Since \( T \) may not have a measurable inverse and hence \( U^{-1} \) may not exist, we only have \( U^*U = 1 \), but not \( UU^* = 1 \). Instead, to deduce \( Uh = h \) from \( U^*h = h \), we note that

\[
\|Uh - h\|_2^2 = \langle Uh - h, Uh - h \rangle = 2\|h\|_2^2 - \langle h, Uh \rangle - \langle Uh, h \rangle = 2\|h\|_2^2 - \langle U^*h, h \rangle - \langle h, U^*h \rangle = 0.
\]

By the above reasoning, \( A_n f \to Pf \), where \( P \) denotes the orthogonal projection onto \( H_0 \). To verify the \( L_2 \) mean ergodic theorem, it only remains to verify that \( Pf = \mathbb{E}_\mu[f|\mathcal{F}] \). Recall that conditional expectation is an orthogonal projection in \( L_2 \) space. A function \( f \in \mathcal{I} \) if and only if its level sets \( \{ \omega : a \leq f(\omega) < b \} \in \mathcal{I} \). It is not difficult to see that such \( f \) must satisfy \( f(\omega) = f(T\omega) \) a.s., i.e., \( f \in H_0 \).

For general \( 1 \leq p < \infty \), we can approximate \( f \in L_p(\Omega, \mathcal{F}, \mu) \) by bounded functions \( f^\varepsilon \) such that \( \|f - f^\varepsilon\|_p \leq \varepsilon \). By the \( L_2 \) mean ergodic theorem, \( \|A_n f^\varepsilon - Pf^\varepsilon\|_2 \to 0 \). Since \( f^\varepsilon \) is bounded, we also have \( \|A_n f^\varepsilon - Pf^\varepsilon\|_p \to 0 \). Therefore

\[
\limsup_{n \to \infty} \|A_n f - Pf\|_p \leq \limsup_{n \to \infty} \|A_n f - A_n f^\varepsilon\|_p + \limsup_{n \to \infty} \|A_n f^\varepsilon - Pf^\varepsilon\|_p + \limsup_{n \to \infty} \|P f - Pf^\varepsilon\|_p \leq 2\varepsilon,
\]

where we used the fact that \( A_n \) and \( P \) both have norm 1 in \( L_p(\Omega, \mathcal{F}, \mu) \). Since \( \varepsilon > 0 \) is arbitrary, we have \( \|A_n f - Pf\|_p \to 0 \).

To prove the pointwise ergodic theorem, we need a maximal inequality analogous to Doob’s inequality in order to control the fluctuation of the time average.

**Lemma 1.7 [Maximal ergodic lemma]** Let \( f \in L_1(\Omega, \mathcal{F}, \mu) \) and for \( n \geq 1 \), let

\[
E_n^0 = \{ \omega : \max_{1 \leq j \leq n} (f(\omega) + f(T\omega) + \cdots + f(T^{j-1}\omega)) \geq 0 \}.
\]

Then

\[
\int_{E_n^0} f(\omega) d\mu \geq 0.
\]

**Proof.** Let

\[
h_n(\omega) = \max_{1 \leq j \leq n} (f(\omega) + f(T\omega) + \cdots + f(T^{j-1}\omega)) = f(\omega) + h_{n-1}^+(T\omega),
\]

where \( h_{n-1}^+(\omega) = \max(0, h_n(\omega)) \). On \( E_n^0 \), \( h_n(\omega) = h_{n-1}^+(\omega) \), therefore

\[
f(\omega) = h_{n-1}^+(\omega) - h_{n-1}^+(T\omega),
\]

and hence

\[
\int_{E_n^0} f(\omega) d\mu = \int_{E_n^0} [h_{n-1}^+(\omega) - h_{n-1}^+(T\omega)] d\mu
\]

\[
\geq \int_{E_n^0} h_{n-1}^+(\omega) - h_{n}^+(T\omega) d\mu
\]

\[
= \int_{E_n^0} h_n^+(\omega) d\mu - \int_\Omega h_n^+(\omega) d\mu
\]

\[
\geq - \int_{\Omega \setminus E_n^0} h_n^+(\omega) d\mu = 0.
\]

\[\square\]
**Corollary 1.8 [Wiener’s maximal inequality]** Let $f \in L_1(\Omega, \mathcal{F}, \mu)$, and for $l > 0$ and $n \in \mathbb{N}$, let

$$E_n = \{ \omega : \max_{1 \leq j \leq n} A_j f \geq l \},$$

where $A_nf = \frac{1}{n} \sum_{i=0}^{n-1} f(T^i\omega)$. Then

$$\mu(E_n) \leq \frac{1}{l} \int_{E_n} f(\omega) d\mu \leq \frac{1}{l} \int_{E_n} |f(\omega)| d\mu.$$  \hspace{1cm} (1.3)

**Proof.** Note that $E_n$ is just the event that $\{ \omega : \max_{1 \leq j \leq n} \sum_{i=1}^j (f(T^{-1}i\omega) - l) \geq 0 \}$. Therefore by Lemma 1.7 applied to $f(\omega) - l$, we have

$$\int_{E_n} (f(\omega) - l) d\mu \geq 0,$$

which implies that

$$\mu(E_n) \leq \frac{1}{l} \int_{E_n} f(\omega) d\mu \leq \frac{1}{l} \int_{E_n} |f(\omega)| d\mu.$$

\hfill \blacksquare

**Proof of the pointwise ergodic theorem.** It suffices to consider $f \in L_1(\Omega, \mathcal{F}, \mu)$. Recall $H, H_0$ from the proof of the mean ergodic theorems. For $f \in H_0$, obviously $A_nf = f \to f$ a.s. For $f = (I - U)g$ with $g$ bounded, $A_nf = (g - U^ng)/n \to 0$ a.s. Therefore the pointwise ergodic theorem holds for all $f = f_1 + f_2$ with $f_1 \in H_0$ and $f_2 = (I - U)g$ for some bounded $g$. This class of functions are dense in $L_1(\Omega, \mathcal{F}, \mu)$, which we denote by $\mathcal{D}$.

Let $f \in L_1(\Omega, \mathcal{F}, \mu)$ and let $f_m \in \mathcal{D}$ such that $\|f - f_m\|_1 \to 0$ as $m \to \infty$. Since $A_nf \to Pf$ in $L_1$ by the mean ergodic theorem, it only remains to verify that

$$\mu(G_\epsilon) := \mu(\omega : \limsup_{n \to \infty} A_nf(\omega) - \liminf_{n \to \infty} A_nf(\omega) > \epsilon) = 0$$

for all $\epsilon > 0$.

Since $A_nf_m \to Pf_m$ a.s. and in $L_1$ for each $m \in \mathbb{N}$, we have

$$\mu(G_\epsilon) \leq \mu\{ \omega : \sup_{n \in \mathbb{N}} |(A_nf)(\omega) - (A_nf_m)(\omega)| \geq \epsilon/2 \}.$$  

Applying Corollary 1.8 to $(f - f_m)$ as well as $(f_m - f)$, and note that $E_n$ is increasing in $n$, we have

$$\mu\{ \omega : \sup_{n \in \mathbb{N}} |(A_nf)(\omega) - (A_nf_m)(\omega)| \geq \epsilon/2 \} \leq \frac{4}{\epsilon} \|f - f_m\|_1.$$  

Since $\|f - f_m\|_1 \to 0$ as $m \to \infty$, we have $\mu(G_\epsilon) = 0$.

\hfill \blacksquare

**Example 1.9 [Strong law of large numbers]** Let $(X_n)_{n \in \mathbb{N}}$ be i.i.d. integrable random variables. On the infinite product space, the law of $(X_1, X_2, \ldots)$ is invariant w.r.t. the shift map $T$. Therefore we can apply the ergodic theorem to the function $f(x_1, x_2, \ldots) = x_1$ and conclude that $\frac{1}{n} \sum_{i=1}^n X_i$ converges a.s. and in $L_1$ to $E[X_1|\mathcal{I}]$. We claim that $\mathcal{I}$ is trivial so that $E[X_1|\mathcal{I}] = E[X_1]$ a.s. Indeed, any invariant set in $\mathcal{I}$ (modulo sets of measure 0) belongs to the tail $\sigma$-field, which by Komogorov’s 0-1 law has either probability 0 or 1.

**Example 1.10 [Irrational rotation on the circle]** Let $S$ be the unit circle parameterized by $[0, 1]$ where 0 and 1 are identified. Let $\theta \in (0, 1)$ be irrational, and define $Tx = (x + \theta) \pmod 1$ for $x \in [0, 1)$. Then Lebesgue measure on $[0, 1)$ is invariant for $T$. By computing the Fourier transform $\phi_A(k) = \int_A e^{2\pi kx} dx$, $k \in \mathbb{Z}$, it can be shown that any invariant set $A$ has Lebesgue measure either 0 or 1. The ergodic theorem implies that the empirical distribution of $x, Tx, T^2x, \cdots$ converge weakly to the Lebesgue measure on $[0, 1)$. 

4
Example 1.11 [Benford’s law] The claim is that as \( m \to \infty \), the leading digit of \( 2^m \) will turn out to be \( k \), \( k \in \{1, \cdots, 9\} \), with an asymptotic frequency of \( \log_{10}(k+1) - \log_{10}k \). This can be understood in terms of irrational rotation on the circle. Note that \( 2^m \) has leading digit \( k \) if and only if \( m \log_{10} 2 \) (mod 1) \( \in [\log_{10} k, \log_{10}(k+1)) \). Since \( \log_{10} 2 \) is irrational, we can apply the ergodic theorem for irrational rotation on the circle.

2 Mixing

The following lemma gives an alternative characterization of ergodicity.

Lemma 2.1 A measure preservation transformation \( T \) on a probability space \((\Omega, \mathcal{F}, \mu)\) is ergodic if and only if for all \( A, B \in \mathcal{F} \),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mu(T^{-k} A \cap B) = \mu(A)\mu(B). \tag{2.4}
\]

Proof. By Theorem 1.6, \( \frac{1}{n} \sum_{i=1}^{n} 1_A(T^i \omega) \to \mu(A|\mathcal{I}) \) a.s. and in \( L_1 \), where \( \mathcal{I} \) is the invariant \( \sigma \)-field. Therefore

\[
\frac{1}{n} \sum_{i=1}^{n} \mu(T^{-k} A \cap B) = \mathbb{E}_\mu \left[ 1_B(\omega) \frac{1}{n} \sum_{i=1}^{n} 1_A(T^i \omega) \right] \to \mathbb{E}_\mu[1_B(\omega)\mu(A|\mathcal{I})].
\]

If \( T \) is ergodic on \((\Omega, \mathcal{F}, \mu)\), then \( \mu(A|\mathcal{I}) = \mu(A) \) a.s., from which (2.4) follows. Conversely, if \( A \) is an invariant set, then \( \mu(T^{-k} A \cap B) = \mu(A \cap B) \) for all \( k \in \mathbb{N} \), which when plugged in (2.4) implies \( \mu(A \cap B) = \mu(A)\mu(B) \) for all \( B \in \mathcal{F} \). Setting \( B = A \) then gives \( \mu(A) = \mu(A)^2 \), which implies \( \mu(A) \in \{0, 1\} \), and hence \( T \) is ergodic for \((\Omega, \mathcal{F}, \mu)\). \( \blacksquare \)

The characterization given in Lemma 2.1 can be strengthened to define the notion of weak and strong mixing.

Definition 2.2 [Weak mixing and strong mixing] Let \( T \) a measure preserving transformation on a probability space \((\Omega, \mathcal{F}, \mu)\). \( T \) is called weak mixing if

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \left| \mu(T^{-k} A \cap B) - \mu(A)\mu(B) \right| = 0 \quad \text{for all } A, B \in \mathcal{F}, \tag{2.5}
\]

and strong mixing if

\[
\lim_{n \to \infty} \left| \mu(T^{-n} A \cap B) - \mu(A)\mu(B) \right| = 0 \quad \text{for all } A, B \in \mathcal{F}. \tag{2.6}
\]

Note that obviously strong mixing implies weak mixing, which in turn implies ergodicity.

Remark. If \((X_n)_{n \geq 0}\) is a Markov chain in equilibrium, i.e., \( X_1 \), and hence \( X_i \), all have stationary distribution \( \mu \), then strong mixing is equivalent to the convergence of \( X_n \) in distribution to \( \mu \) conditional on \( X_0 = x \) for any \( x \) in the state space \( S \). Indeed, if we assume strong mixing and take \( A, B \) in (2.6) to be of the form \( A = 1_F(X_0) \) and \( B = 1_{\{X_0=x\}} \) for any \( x \in S \) and \( F \in \mathcal{S} \), the \( \sigma \)-algebra on the state space \( S \), then (2.6) implies

\[
\mu(X_0 = x, X_n \in F) = \mu(x)\mathbb{P}_x(X_n \in F) \longrightarrow \mu(x)\mu(F).
\]
Since $F \in S$ can be chosen arbitrarily, this implies that conditional on $X_0 = x$, $X_n$ converges in distribution to $\mu$. The converse is also not hard to see.

**Remark.** If for $i = 1, 2$, $T_i$ is a measure preserving transformation on a probability space $(\Omega_i, \mathcal{F}_i, \mu_i)$, it is natural to define a transformation $T$ on the product probability space $(\Omega, \mathcal{F}, \mu)$, where $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F}$ is the product $\sigma$-algebra on $\Omega$ and $\mu = \mu_1 \times \mu_2$ the product measure, with $T(\omega_1, \omega_2) = (T_1 \omega_1, T_2 \omega_2)$ for any $\omega_1 \in \Omega_1$ and $\omega_2 \in \Omega_2$. It is easy to see that $T$ preserves $\mu$. The question is under what conditions is $T$ ergodic? It is not sufficient to assume only the ergodicity of $T_1$ and $T_2$, as easily seen from the example where $T_1 = T_2$ are irrational rotations on the unit circle, and $\mu_1 = \mu_2$ is the Lebesgue measure. If $T_1 = T_2$, then a necessary and sufficient condition for the ergodicity of $T \times T$ is that $T$ is weakly mixing.