**PROBABILITY THEORY: LECTURE NOTES 7**

CHANDRA NAIR

**Disclaimer:** These notes are created from a variety of great references in the field, most notably, Varadhan’s lecture notes and Dembo’s lecture notes.

**Definition 1.** A sequence of random variables \( \{X_n\} \) is said to be *stationary* if the distribution of \( \{X_1, \ldots, X_n\} \) is identical to that of \( \{X_{1+k}, \ldots, X_{n+k}\} \) for all \( n \geq 1 \) and \( k \geq 1 \).

**Definition 2.** Let \((\Omega, \mathcal{F}, P)\) be a probability space. A measurable mapping \( T : \Omega \to \Omega \) is said to be *measure-preserving* if \( P(T^{-1}(A)) = P(A) \) for all \( A \in \mathcal{F} \).

A measure-preserving map naturally gives rise to a stationary sequence as follows: let \( X(\omega) \) be a random variable; define \( X_n(\omega) = X(T^n(\omega)) \), \( n \geq 0 \), where \( T^0(\omega) := \omega \).

To see that the above process is stationary, define \( B := \{ \omega : (X_1, \ldots, X_n) \in A \} \). Then note that \( T^{-k}(B) = \{ \omega : (X_{1+k}, \ldots, X_{n+k}) \in A \} \). Since \( T \) is measure-preserving, we are done.

**Remark 1.** That any stationary process of real-valued random variables is induced by such a measure-preserving transformation is a consequence of Kolmogorov’s extension theorem (why?):

This part of the notes will focus on measure-preserving transformations and hence \( T \) will always be assumed to be measure-preserving.

**Definition 3.** A set \( A \) is said to be *strictly-invariant* if \( T^{-1}(A) = A \), while a set \( A \) is said to be *invariant* if \( P(T^{-1}(A) \Delta A) = 0 \).

**Exercise 7.1.** Show the following basic properties of mappings and sets.

1. \( T^{-1}(\bigcup_i A_i) = \bigcup_i T^{-1}(A_i) \).
2. \( T(T^{-1}(A)) = A, T^{-1}(T(A)) \supseteq A \).
3. \( A \Delta B = A^c \Delta B^c \).
4. \( \bigcup_i (A_i \Delta B_i) \supseteq (\bigcup_i A_i) \Delta (\bigcup_i B_i) \).
5. \( T^{-1}(A^c) = (T^{-1}(A))^c \).
6. \( T^{-1}(A \Delta B) = T^{-1}(A) \Delta T^{-1}(B) \).
7. \( A \Delta (\bigcup_{i=1}^\infty B_i) \subseteq (A \Delta B_1) \cup (\bigcup_{i=1}^\infty (B_i \Delta B_{i+1})) \).

From the above properties, it is immediate that \( \mathcal{I} \) - the collection of invariant sets - is a \( \sigma \)-algebra. We will call this to be the *invariant-\( \sigma \)-algebra*. Similarly, we can also define the *strictly-invariant-\( \sigma \)-algebra*.

Given an invariant set \( A \), let \( B := \bigcup_{n=0}^\infty T^{-n}(A) \). Note that \( A \subseteq B \) and \( T^{-1}(B) = \bigcup_{n=1}^\infty T^{-n}(A) \subseteq B \). Define \( C := \bigcap_{n=0}^\infty T^{-n}(B) \). Note that \( T^{-1}(C) = \bigcap_{n=1}^\infty T^{-n}(B) \); however since \( B \cap T^{-1}(B) = T^{-1}(B) \) we have \( T^{-1}(C) = C \). Argue that \( P(A \Delta C) = 0 \).

*Date:* December 11, 2019.
Lemma 1. If $X$ is $\mathcal{I}$-measurable then $X(T(\omega)) = X(\omega)$ almost surely.

Proof. If $A$ is invariant then $T^{-1}(A)$ is also invariant (use (6) in Exercise along with measure-preserving property). Therefore $X(T(\omega))$ is also $\mathcal{I}$-measurable. Given two rational numbers $p < q$ let $A_{p,q} = \{\omega : X(\omega) < p, X(T(\omega)) > q\}$, and let $B_p = \{\omega : X(\omega) < p\}$. It is clear that $A_{p,q} \subseteq B_p \Delta T^{-1}(B_p)$ and hence $P(A_{p,q}) = 0$. Now the lemma follows immediately. \hfill \Box

Definition 4. A measure-preserving transformation associated with a stationary process is called ergodic if $A \in \mathcal{I}$, the invariant $\sigma$-algebra, implies that $P(A) = 0$ or $P(A) = 1$.

Lemma 2 (Maximal Ergodic Lemma). Let $X_j(\omega) = X(T^k(\omega))$, $S_k = \sum_{i=0}^{k-1} X_i(\omega)$, and $M_k(\omega) = \max(0, S_1(\omega), ..., S_k(\omega))$. Then $E(X_{1_{M_k > 0}}) \geq 0$.

Proof. If $j \leq k$ then $M_k(T(\omega)) \geq S_j(T(\omega))$, implying

$$X(\omega) \geq S_{j+1}(\omega) - M_k(T(\omega)), \ j = 0, ..., k$$

Therefore

$$E(X(\omega)1_{M_k > 0}) \geq \int_{M_k > 0} \max\{S_1(\omega), ..., S_k(\omega)\} - M_k(T(\omega))dP$$

$$\quad = \int_{M_k > 0} M_k(\omega) - M_k(T(\omega))dP \geq 0.$$ 

The last inequality is due to the following observation: $M_k(\omega) = 0$ we have $M_k(T(\omega)) \geq 0$. However since integrals of $M_k(\omega)$ and $M_k(T(\omega))$ are same (measure-preserving property of $T$), the inequality follows. \hfill \Box

Theorem 1 (Birkhoff’s Ergodic Theorem). For any $X \in L^1$, 

$$\frac{1}{n} \sum_{m=0}^{n-1} X(T^m(\omega)) \to E(X|\mathcal{I}) \text{ a.s.}$$

and in $L^1$.

Proof. Since $E(X|\mathcal{I})$ is invariant under $T$ (see lemma above) w.l.o.g. we can center $X$ and assume $E(X|\mathcal{I}) = 0$. Let $X = \limsup \frac{S_n}{n}$ and let $D = \{\omega : X > \varepsilon\}$. Since $X(T(\omega)) = X(\omega)$, we have that $D \in \mathcal{I}$.

Define a new sequence of random variables $Y(\omega) = (X(\omega) - \varepsilon)1_D$ and let $U_n = Y_0 + \cdot \cdot \cdot + Y_{n-1}$. Let $M_n(\omega) = \max(0, U_1(\omega), .., U_n(\omega))$. Observe that $M_n \uparrow$ and $\lim_n M_n > 0$ on $D$. Let $E_n = \{\omega : M_n > 0\}$. Hence $E_n \uparrow D$. Since $|Y| \leq |X| + \varepsilon$ we have

$$0 \leq E(Y1_{E_n}) \to E(Y1_D).$$

where the inequality comes from Maximal ergodic lemma. Hence

$$E((X(\omega) - \varepsilon)1_D) \geq 0 \implies E(E((X(\omega) - \varepsilon)1_D|\mathcal{I})) = E(1_D E(X|\mathcal{I}) - \varepsilon 1_D) \geq 0.$$ 

This shows that $P(D) = 0$, similarly working with $-X$, completes the almost sure convergence.

To show convergence in $L_1$, let $X_M = X1_{|X| < M}$. Almost sure convergence above and bounded convergence theorem says that

$$E \left| \frac{1}{n} \sum_{m=0}^{n-1} X_M(T^m(\omega)) - E(X_M|\mathcal{I}) \right| \to 0.$$
Let \( \hat{X}_M = X - X_M \). Since
\[
E \left| \frac{1}{n} \sum_{m=0}^{n-1} \hat{X}_M(T^m(\omega)) \right| \leq E(|\hat{X}_M|),
\]
and since \( |E(\hat{X}_M|\mathcal{I})| \leq E(|\hat{X}_M|) \), triangle inequality completes the \( L^1 \) convergence. \( \square \)

Given a measurable transformation \( T \), let \( \mathcal{M} \) denote the convex set of all probability measures that is \( T \)-invariant (this could be empty).

**Theorem 2.** A probability measure \( P \in \mathcal{M} \) is ergodic if and only if it is an extreme point of \( \mathcal{M} \).

**Proof.** Assume that \( P \) is ergodic and yet \( P = aP_1 + (1 - a)P_2 \) for \( 0 < a < 1 \). Since \( P \) is ergodic, it implies that \( P_1 = P_2 \) on \( \mathcal{I} \), hence \( P_1 \) and \( P_2 \) are also ergodic. Let \( f \) be any bounded measurable function on \( (\omega, \mathcal{F}) \). Define
\[
h(\omega) = \lim_{n \to \infty} \frac{1}{n} \left( f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega) \right)
\]
when it exists. From Ergodic theorem, we know that the limit exists on a set \( E \) with \( P_1(E) = P_2(E) = 1 \). Further, from bounded convergence theorem we also know that
\[
E_{P_i}(h) = \int f dP_i \quad i = 1, 2.
\]
However since \( h \) is \( \mathcal{I} \) measurable and \( P_1 = P_2 \) on \( \mathcal{I} \), we see that \( \int f dP_1 = \int f dP_2 \) for any bounded measurable function (in particular indicator functions). Hence \( P_1 = P_2 \) on \( \mathcal{F} \).

If \( P \) is an extreme point of \( \mathcal{M} \) and \( P \) is not ergodic, then there exists \( A \in \mathcal{I} \) with \( 0 < P(A) < 1 \). Define \( P_1(E) = \frac{P(E \cap A)}{P(A)} \) and \( P_2(E) = \frac{P(E \cap A^C)}{P(A^C)} \). Note that \( P_1 \), \( P_2 \) belong to \( \mathcal{M} \) and if \( \frac{P_1(E \cap A)}{P_1(A)} = \frac{P_2(E \cap A^C)}{P_2(A^C)} \forall E \in \mathcal{F} \); however this cannot happen for \( E = A \). Hence \( P \) (which can be written as a non-trivial convex combination of \( P_1, P_2 \)) is not extremal in \( \mathcal{M} \). \( \square \)

**Lemma 3.** For any stationary measure \( P \), the regular conditional probability of \( P \) given \( \mathcal{I} \), denoted by \( Q(\omega, \cdot) \), is stationary and ergodic.

**Proof.** We know that almost surely
\[
Q(\omega, A) = E(1_A|\mathcal{I}).
\]
We need to show that \( Q(\omega, A) = Q(\omega, TA) \). Suffices to show that for all \( I \in \mathcal{I} \)
\[
\int_I 1_{AdP} = \int_I 1_{TAdP}
\]
or in other words \( P(A \cap I) = P(TA \cap I) \) which is immediate due to invariance of \( I \).

To show ergodicity, we need to show that \( Q(\omega, I) = 0 \) or 1, for \( I \in \mathcal{I} \) for almost all \( \omega \). This is again immediate. (note that the issue of throwing away too many null sets was covered during definition of regular conditional probabilities). \( \square \)

**Theorem 3.** Any invariant measure \( P \in \mathcal{M} \) can be written as a convex combination of ergodic measures, i.e.
\[
P = \int_{\mathcal{M}_e} Q \mu_P(dQ).
\]
Proof. By regular conditional probabilities

\[ P = \int Q(w, \cdot) dP. \]

By previous lemma \( Q(\omega, \cdot) \in \mathcal{M}_\epsilon \) and hence we have a induced measure \( \mu \) on measures in \( \mathcal{M}_\epsilon \). By changing the integration with respect to that measure, we are done. \( \square \)

8. Backwards Martingales, Exchangeable processes, and de Finetti’s theorem

Definition 5. A backwards martingale is a sequence of random variables \( \{X_n\}, n \leq 0 \), adapted to a filtration, \( \{\mathcal{F}_n\}_{n \leq 0} \) defined by

\[ X_n := \mathbb{E}(X_0 | \mathcal{F}_n), n \leq 0, \]

where \( X_0 \) is an integrable random variable.

Lemma 4. Let \((\Omega, \mathcal{F}, P)\) be a probability space and \( X \) be an integrable random variable. Consider the collection

\[ \{Y : Y = \mathbb{E}(X | \mathcal{G}), \text{ for some } \mathcal{G} \subset \mathcal{F}\}. \]

Then the collection of random variables is uniformly integrable.

Remark: Formally the collection contains versions of conditional expectation.

Proof. Let \( \epsilon > 0 \) be given. Let \( c_\delta = \sup_{A \in \mathcal{F}, P(A) \leq \delta} \mathbb{E}(\mathbb{E}[|X|_A]) \). We know from dominated convergence theorem that \( c_\delta \downarrow 0 \) and \( \delta \downarrow 0 \). Choose \( \delta_0 \) such that \( c_{\delta_0} \leq \epsilon \).

Choose \( M \) such that \( \frac{1}{M} \mathbb{E}(\mathbb{E}[|X|]) \leq \delta_0 \).

Jensen’s inequality says that \( |Y| \leq \mathbb{E}(|X| | \mathcal{G}) \) a.s. In particular \( \mathbb{E}(|Y|) \leq \mathbb{E}(|X|) \).

Hence \( P(|Y| > M) \leq \frac{1}{M} \mathbb{E}(\mathbb{E}[|X|]) \leq \delta_0 \). From the definition of conditional expectation

\[ \mathbb{E}(|Y| | |Y| > M) \leq \mathbb{E}(|X| | |Y| > M) \leq \epsilon. \]

\( \square \)

Theorem 4. The limit \( X_{-\infty} = \lim_{n \to -\infty} X_n \) exists a.s. and in \( L^1 \).

Proof. Doob’s upcrossing inequality for the number of upcrossings, \( U_n \), between \([a, b]\) made by \( X_{-n}, \ldots, X_0 \) yields \( (b-a) \mathbb{E}(U_n) \leq \mathbb{E}(X_0 - a)_{+} < \infty \). Hence the limit exists almost surely. Since the collection is uniformly integrable (above lemma), we have convergence in \( L^1 \) (another lemma established before).

Given a collection of random variables \( X_1, X_2, \ldots \), let \( \mathcal{E}_n \) be the events that are invariant under permutations of \( \{1, 2, \ldots, n\} \) but leave \( n + 1, \ldots \) fixed. Let \( \mathcal{E} = \cap_n \mathcal{E}_n \) be the exchangeable \( \sigma \)-algebra.

Definition 6. A sequence \( X_1, X_2, \ldots \) is said to be exchangeable if for every \( n \) and for every permutation \( \pi \) of \( \{1, \ldots, n\} \) the distribution of \( (X_1, \ldots, x_n) \) and \( X_{\pi(1)}, \ldots, X_{\pi(n)} \) are the same.

Theorem 5 (de Finetti). If \( X_1, X_2, \ldots, X_n \) are exchangeable the conditioned on \( \mathcal{E} \), \( X_1, \ldots \) are independent and identically distributed.
Proof. Let $f$ be a bounded measurable function. Define

$$A_n(f) := \frac{1}{(k)} \sum_{i\in[n]} f(X_{i_1}, \ldots, X_{i_k}).$$

Since $A_n$ is exchangeable, we have

$$A_n(f) = E(A_n(f)|\mathcal{E}_n) = \frac{1}{(n)k} \sum_{i\in[n]} E(f(X_{i_1}, \ldots, X_{i_k})|\mathcal{E}_n) = E(f(X_1, \ldots, X_k)|\mathcal{E}_n).$$

From backwards martingale theorem $A_n(f) \to A_\infty(f) = E(f(X_1, \ldots, X_k)|\mathcal{E}).$

Consider $\phi(x_1, \ldots, x_k) = f(x_1, \ldots, x_{k-1})g(x_k)$ and define $\phi_j(x_1, \ldots, x_{k-1}) = f(x_1, \ldots, x_{k-1})g(x_j)$ for $1 \leq j \leq k - 1$. Then observe that

$$(n)_{k-1}A_n(f)nA_n(g) = (n)_kA_n(\phi) + (n)_{k-1} \sum_{j=1}^{k-1} A_n(\phi_j) \iff$$

$$A_n(f)A_n(g) = \frac{n - k + 1}{n} A_n(\phi) + \frac{1}{n} \sum_{j=1}^{k-1} A_n(\phi_j).$$

Taking $n \to \infty$ we obtain $A_\infty(f)A_\infty(g) = A_\infty(\phi)$, or

$$E(f(X_1, \ldots, X_{k-1})g(X_k)|\mathcal{E}) = E(f(X_1, \ldots, X_{k-1})|\mathcal{E}) E(g(X_k)|\mathcal{E}).$$

$\square$