Lecture 4

With Corollary 3.3.1 in hand, we can now prove the Birkhoff ergodic theorem.

Proof of Birkhoff Ergodic Theorem. By considering real and imaginary parts, we can assume $f \in L^1_{\mathbb{R}}(\mu)$. Set for any $x \in X$,

$$f^*(x) = \overline{\lim_{n\to\infty}} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \text{ and } f_*(x) = \underline{\lim_{n\to\infty}} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x),$$

then $f^*(Tx) = f^*(x)$ and $f_*(Tx) = f_*(x)$. For any $\alpha, \beta \in \mathbb{R}$ with $\alpha < \beta$, define

$$E_{\alpha,\beta} = \{x \in X : f_*(x) < \alpha < \beta < f^*(x)\},$$

then $E_{\alpha,\beta} \in \mathscr{B}$ and $T^{-1}(E_{\alpha,\beta}) = E_{\alpha,\beta}$. We claim that $\mu(E_{\alpha,\beta}) = 0$. To see this, define $B_\beta = \{x \in X : \sup_{n\geq 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) > \beta\}$, clearly $E_{\alpha,\beta} \subset B_\beta$. By Corollary 3.3.1, we have

$$\int_{E_{\alpha,\beta}} f d\mu = \int_{E_{\alpha,\beta}\cap B_\beta} f d\mu \geq \beta\mu(E_{\alpha,\beta} \cap B_\beta) = \beta\mu(E_{\alpha,\beta}). \tag{3.3}$$

Now consider $-f$ instead of f. Let $g = -f$, then $g^*(x) = -f_*(x)$, $g_*(x) = -f^*(x)$. Moreover we have

$$E_{\alpha,\beta} = \{x \in X : g_*(x) < -\beta < -\alpha < g^*(x)\}.$$

Similar to (3.3), we have $\int_{E_{\alpha,\beta}} g d\mu \geq -\alpha\mu(E_{\alpha,\beta})$, that is

$$\int_{E_{\alpha,\beta}} f d\mu \leq \alpha\mu(E_{\alpha,\beta}). \tag{3.4}$$

By (3.3) and (3.4), we have $\beta\mu(E_{\alpha,\beta}) \leq \alpha\mu(E_{\alpha,\beta})$, since $\alpha < \beta$, it forces $\mu(E_{\alpha,\beta})$ to be 0. We observe that

$$\{x \in X : f^*(x) > f_*(x)\} = \bigcup_{\alpha,\beta\in\mathbb{Q} \text{ with } \alpha<\beta} E_{\alpha,\beta},$$

hence $\mu(\{x \in X : f^*(x) > f_*(x)\}) = 0$, therefore $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = f^*(x)$ a.e. It's clear $f^*(Tx) = f^*(x)$.

Next we show that $f^* \in L^1(\mu)$. Let $h(x) = |f(x)|$, then $h \in L^1(\mu)$ and $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} h(T^k x) = h^*(x)$ a.e. Since $f^*(x) \leq h^*(x)$, it suffices to show $h^* \in$

$L^1(\mu)$. By Fatou's lemma and the fact that μ is T-invariant, we have

$$\begin{aligned}\int h^* d\mu &= \int \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} h(T^k x) d\mu \\ &\leq \varliminf_{n\to\infty} \int \frac{1}{n}\sum_{k=0}^{n-1} h(T^k x) d\mu \\ &= \varliminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int h(T^k x) d\mu \\ &= \varliminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int h d\mu = \int h d\mu < \infty.\end{aligned}$$

To the end, we prove that $\int_A f^* d\mu = \int_A f d\mu$ for any $A \in \mathscr{B}$ with $T^{-1}A = A$. Fix $A \in \mathscr{B}$ with $T^{-1}A = A$. Define for $k \in \mathbb{Z}, n \in \mathbb{N}_+$ that $D_{n,k} = \{x \in X : \frac{k}{n} \leq f^*(x) < \frac{k+1}{n}\}$. Then $D_{n,k}$ are T-invariant and $(D_{n,k})_{k\in\mathbb{Z}}$ is a partition of X for each n. Set for $\epsilon > 0$,

$$B_{\frac{k}{n}-\epsilon} = \{x : \sup_{n\geq 1} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) > \frac{k}{n} - \epsilon\},$$

then $D_{n,k} \subset B_{\frac{k}{n}-\epsilon}$. By Corollary 3.3.1 again,

$$\int_{A\cap D_{n,k}} f d\mu = \int_{A\cap D_{n,k}\cap B_{\frac{k}{n}-\epsilon}} f d\mu \geq (\frac{k}{n}-\epsilon)\mu(A\cap D_{n,k}\cap B_{\frac{k}{n}-\epsilon}) = (\frac{k}{n}-\epsilon)\mu(A\cap D_{n,k}),$$

letting $\epsilon \to 0$, we have $\int_{A\cap D_{n,k}} f d\mu \geq \frac{k}{n}\mu(A\cap D_{n,k})$. Notice that $f^* < \frac{k+1}{n}$ on $D_{n,k}$, hence

$$\int_{A\cap D_{n,k}} f^* d\mu \leq \frac{k+1}{n}\mu(A\cap D_{n,k}) \leq \int_{A\cap D_{n,k}} f d\mu + \frac{1}{n}\mu(A\cap D_{n,k}).$$

Summing k over $\mathbb{Z}$, $\int_A f^* d\mu \leq \int_A f d\mu + \frac{\mu(A)}{n}$. Letting $n \to \infty$, we obtain

$$\int_A f^* d\mu \leq \int_A f d\mu. \tag{3.5}$$

Replacing f by $-f$ in (3.5), we have $\int_A (-f)^* d\mu \leq \int_A (-f) d\mu$, that is $\int_A -f_* d\mu \leq \int_A (-f) d\mu$, hence $\int_A f_* d\mu \geq \int_A f d\mu$. Since $f^* = f_*$ a.e., we complete the proof. □

We can understand the result of Birkhoff ergodic theorem from another point of view. Let $\mathcal{I} = \{A \in \mathscr{B} : T^{-1}A = A\}$, i.e. the collection of T-invariant sets in $\mathscr{B}$, it's easy to see $\mathcal{I}$ is a sub-σ-algebra of $\mathscr{B}$. By Birkhoff ergodic theorem, given an integrable function f, the Birkhoff average f^* satisfies $\int_A f^* d\mu = \int_A f d\mu$ for any $A \in \mathcal{I}$, with this property f^* is in fact the conditional expectation of f w.r.t $\mathcal{I}$. We will see this more precisely after we introduce the notion of conditional expectation as follows.

Theorem 3.4 (Radon-Nikodym Theorem)**.** *Let $(X, \mathscr{B}, m)$ be a probability space. Let μ be a finite signed measure on measurable space $(X, \mathscr{B})$ that is absolutely continuous w.r.t m. Then there exists $f \in L^1(X, \mathscr{B}, m)$ such that*

$$\mu(A) = \int_A f dm \text{ for any } A \in \mathscr{B}. \tag{3.6}$$

Moreover f is unique in the sense that if $g \in L^1(X, \mathscr{B}, m)$ is another function satisfying (3.6), then $f = g$ a.e.

f in (3.6) is denoted by $f = \frac{d\mu}{dm}$, and called the Radon-Nikodym derivative of μ w.r.t m.

Definition 3.1 (Conditional expectation)**.** *Let $(X, \mathscr{B}, m)$ be a probability space. Let $\mathscr{C}$ be a sub-σ-algebra of $\mathscr{B}$. Let $f \in L^1(X, \mathscr{B}, m)$. Define μ on $(X, \mathscr{C})$ by $\mu(C) = \int_C f dm$ for $C \in \mathscr{C}$. Then μ is a finite signed measure and absolutely continuous w.r.t m over $(X, \mathscr{C})$. By Radon-Nikodym theorem, there exists a unique $g \in L^1(X, \mathscr{C}, m)$ such that $\mu(C) = \int_C gdm$ for any $C \in \mathscr{C}$, we denote g by $\mathbb{E}(f|\mathscr{C})$ and call it the conditional expectation of f over $\mathscr{C}$.*

In the above definition, $\mathbb{E}(f|\mathscr{C})$ as a $\mathscr{C}$-measurable function, can be understood as the average of f over sub-σ-algebra $\mathscr{C}$, it provides information of f up to $\mathscr{C}$. To illustrate this, let's consider the following example.

Example. Let $(X, \mathscr{B}, m)$ be a probability space and $f \in L^1(X, \mathscr{B}, m)$. A finite collection of elements in $\mathscr{B}$, say $D = \{C_1, C_2, \cdots, C_k\}$, is said to be a finite partition of X if $X = \bigcup_{i=1}^k C_i$ is a mutually disjoint union. Let $\mathscr{C}$ be a sub-σ-algebra of $\mathscr{B}$. If $\mathscr{C} = \mathscr{B}$, clearly $\mathbb{E}(f|\mathscr{C}) = f$, it provides all information of f. If $\mathscr{C} = \mathscr{N} := \{A \in \mathscr{B} : m(A) = 0 \text{ or } 1\}$, i.e. the trivial sub-$\sigma$-algebra consisting of $\mathscr{B}$-measurable sets of full or null measure, then it's easy to see $\mathbb{E}(f|\mathscr{N}) = \int_X f dm$ a constant, it provides no information but the integral of f over the whole space. If we let $\mathscr{C}$ be the sub-σ-algebra generated by the partition D, it' easy to see $\mathscr{C}$ consists of finite unions of elements in D. Let $g(x) = \sum_{i=1}^{k} (\frac{1}{m(C_i)} \int_{C_i} f dm)\chi_{C_i}(x)$, then $g \in L^1(X, \mathscr{C}, m)$. For each $C_i \in D$, we have $\int_{C_i} gdm = \int_{C_i} f dm$, hence $\int_C gdm = \int_C f dm$ for any $C \in \mathscr{C}$, therefore $g = \mathbb{E}(f|\mathscr{C})$. Notice that on each piece C_i of the partition D, $\mathbb{E}(f|\mathscr{C})$ takes constant value $\frac{1}{m(C_i)} \int_{C_i} f dm$.

We collect some properties of conditional expectation that may be used later.

Proposition 3.1. *Let $(X, \mathscr{B}, m)$ be a probability space. Let $\mathscr{C}$ a sub-σ-algebra of $\mathscr{B}$. The operator $\mathbb{E}(\cdot|\mathscr{C}) : L^1(\mathscr{B}) \to L^1(\mathscr{C})$ enjoys the following properties.*

(i) $\mathbb{E}(\alpha f + \beta g|\mathscr{C}) = \alpha\mathbb{E}(f|\mathscr{C}) + \beta\mathbb{E}(g|\mathscr{C})$, for any $\alpha, \beta \in \mathbb{R}$ and $f, g \in L^1(X, \mathscr{B}, m)$.

(ii) If $\mathscr{C}_1 \subset \mathscr{C}_2 \subset \mathscr{B}$, then $\mathbb{E}(\mathbb{E}(f|\mathscr{C}_1)|\mathscr{C}_2) = \mathbb{E}(f|\mathscr{C}_1) = \mathbb{E}(\mathbb{E}(f|\mathscr{C}_2)|\mathscr{C}_1)$, for any $f \in L^1(X, \mathscr{B}, m)$.

(iii) If $f \in L^1(X, \mathscr{B}, m)$ and $g \in L^\infty(X, \mathscr{C}, m)$, then $\mathbb{E}(fg|\mathscr{C}) = g\mathbb{E}(f|\mathscr{C})$.

(iv) $|\mathbb{E}(h|\mathscr{C})| \leq \mathbb{E}(|h||\mathscr{C})$ *for all* $h \in L^1(X, \mathscr{B}, m)$. *If* $p \in (1, \infty)$ *and* $1/p + 1/q = 1$, *then*

$$\mathbb{E}(|fg||\mathscr{C}) \leq \mathbb{E}(|f|^p|\mathscr{C})^{1/p}\mathbb{E}(|g|^q|\mathscr{C})^{1/q},$$

for any $f \in L^p(X, \mathscr{B}, m), g \in L^q(X, \mathscr{B}, m)$.

(v) Let $(f_n)_{n\geq 1}$ *be a nonnegative increasing sequence in* $L^1(X, \mathscr{B}, m)$. *If* $f_n \uparrow f \in L^1(X, \mathscr{B}, m)$ *a.e., then* $\mathbb{E}(f_n|\mathscr{C}) \uparrow \mathbb{E}(f|\mathscr{C})$ *a.e.*

(vi) If T *is a measure-preserving transformation on* X, *then* $\mathbb{E}(f|\mathscr{C}) \circ T = \mathbb{E}(f \circ T|T^{-1}\mathscr{C})$.

Remark: All "=" and "≤" in the above proposition are understood in the sense of "almost everywhere".

Proof. We only prove (vi) since other properties can be found in any text book on probability theory. Recall we have for any $g \in L^1(X, \mathscr{B}, m)$, $\int_X g \circ T dm = \int_X g dm$. Since by definition $\mathbb{E}(f|\mathscr{C}) \in L^1(\mathscr{C})$, we have $\mathbb{E}(f|\mathscr{C}) \circ T \in L^1(T^{-1}\mathscr{C})$. Fix $A \in T^{-1}\mathscr{C}$, let $B \in \mathscr{C}$ such that $A = T^{-1}B$. Then we have

$$\begin{aligned}
\int_A \mathbb{E}(f|\mathscr{C}) \circ T dm &= \int_{T^{-1}B} \mathbb{E}(f|\mathscr{C}) \circ T dm = \int_X \mathbb{E}(f|\mathscr{C}) \circ T \chi_{T^{-1}B} dm \\
&= \int_X \mathbb{E}(f|\mathscr{C}) \circ T \chi_B \circ T dm = \int_X (\mathbb{E}(f|\mathscr{C})\chi_B) \circ T dm \\
&= \int_X \mathbb{E}(f|\mathscr{C})\chi_B dm \xlongequal{by(iii)} \int_X \mathbb{E}(f\chi_B|\mathscr{C}) dm \\
&= \int_X f\chi_B dm = \int_X (f\chi_B) \circ T dm \\
&= \int_X f \circ T \chi_B \circ T dm = \int_{T^{-1}B} f \circ T dm = \int_A f \circ T dm.
\end{aligned}$$

Hence we have proved (vi). $\square$