3.7 Birkhoff Ergodic Theorem and Applications

In this section we will state the Birkhoff Ergodic Theorem, which is one of the key theorems in Ergodic Theory. The motivation came originally from the *Boltzmann Ergodic Hypothesis* formulated by Boltzmann in the 1930s (see below). The concept of ergodicity was developed exactly in order to prove the Boltzmann Ergodic Hypothesis, thus giving birth to the field of Ergodic Theory.

Botzmann Ergodic Hypothesis

Let X be the phase space of a physical system (for example, the points of X could represent configurations of positions and velocities of particles of a gas in a box). A measurable function $f : X \to \mathbb{R}$ represents an *observable* of the physical system, that is a quantity that can be measured, for example velocity, position, temperature and so on. The value $f(x)$ is the measurement of the observable f that one gets when the system is in the state x. Time evolution of the system, if measured in discrete time units, is given by a transformation $T : X \to X$, so that if $x \in X$ is the initial state of the system, then $T(x)$ is the state of the system after one time unit. If the physical system is in equilibrium, the time evolution T is a measure-preserving transformation.

In order to measure a physical quantity, one usually repeats measurements in time and consider their average. If $x \in X$ is the initial state, the measurements of the observable $f : X \to \mathbb{R}$ at successive time units are given by $f(x), f(T(x)), \dots, f(T^k(x)), \dots$. Thus, the *average* of the first n measurements is given by

$$\frac{\sum_{k=0}^{n-1} f(T^k x)}{n} \qquad \text{(time average)}.$$

This quantity is called *time average* of the observable f after time n.

On the other hand, the *space average* of the observable f is simply

$$\int f \mathrm{d}\mu \qquad \text{(space average)}.$$

In physics one would like to know the *space average* of the observable with respect to the invariant measure, but since experimentally one computes easily time averages (just by repeating measurements of the system at successive instant of times), it is natural to ask whether (and hope that) long time averages give a good approximation of the space average. Boltzmann's conjectured the following:

Boltzmann Ergodic Hypothesis: for almost every initial state $x \in X$ the time averages of any observable f converge as time tends to infinity to the space average of f.

Unfortunately, after many efforts to prove this general form of the Boltzmann Ergodic Hypothesis, it turned out that the conclusion is *not true* in general, for any measure-preserving transformation T. On the other hand, *under the assumption that T is ergodic*, the conclusion of the Boltzmann Ergodic Hypothesis holds and this is exactly the content of Birkhoff Ergodic Theorem for ergodic transformations. Finding the right condition under which the Hypothesis holds motivated the definition of ergodicity and gave birth to the study of Ergodic theory.

Two versions of Birkhoff Ergodic Theorem

The first formulation of Birkhoff Ergodic Theorem gives a result which is weaker than the Ergodic Hypothesis, but holds in general for *any measure preserving transformation* that preserves a *finite* measure.

Theorem 3.7.1 (Birkhoff Ergodic Theorem for measure preserving transformations). *Let $(X, \mathscr{B}, \mu)$ be a* finite *measured space. Let $T : X \to X$ be* measure-preserving *transformation. For any $f \in L^1(X, \mathscr{B}, \mu)$, the following limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$$

exists for μ-almost every $x \in X$. *Moreover, if, for the x for which the limit exists we call*

$$\overline{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x))$$

the function $\overline{f}$ *(which is defined almost everywhere) is* invariant, *that is*

$$\overline{f} \circ T = \overline{f} \qquad \text{for } \mu - \text{almost every } x \in X,$$

and furthermore

$$\int \overline{f} \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

Let us stress again that this theorem, as Poincaré Recurrence, follows simply from preserving a finite measure. We will not prove the theorem here. The proof can be found for example in the book by Pollicott and Yuri.

The following version of Birkhoff Ergodic Theorem for ergodic transformations is simply a Corollary of this general Birkhoff Ergodic Theorem:

Theorem 3.7.2 (Birkhoff Ergodic Theorem for ergodic transformations). *Let* $(X, \mathscr{B}, \mu)$ *be a* probability *space. Let* $T : X \to X$ *be an* ergodic *measure-preserving transformation. For any* $f \in L^1(X, \mathscr{B}, \mu)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int f \mathrm{d}\mu \qquad \text{for } \mu - \text{almost every } x \in X.$$

Proof. By Birkhoff Ergodic Theorem for measure preserving transformations for $\mu-$almost every x the limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \overline{f}(x)$$

exists and defines a function $\overline{f}$ such that $\overline{f} \circ T = \overline{f}$ almost everywhere. Since T is ergodic, every function which is invariant almost everywhere is constant almost everywhere. In particular, $\overline{f}$ is constant almost everywhere. If c is the value of this constant, since μ is a probability measure

$$\int \overline{f} \mathrm{d}\mu = c \cdot \mu(X) = c \cdot 1 = c,$$

but since the ergodic theorem for measure preserving transformations also gives that $\int \overline{f} \mathrm{d}\mu = \int f \mathrm{d}\mu$, we conclude that $\int f \mathrm{d}\mu = c$. Thus, for almost every $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = c = \int f \mathrm{d}\mu,$$

which is the conclusion we were looking for. □

Applications of Birkhoff Ergodic Theorem

The version of Birkhoff Ergodic Theorem for ergodic transformations shows that Boltzmann's Ergodic Hypothesis is true if the time evolution is ergodic. Birkhoff ergodic Theorem has many other applications in different areas of mathematics. We will show a few consequences.

1. Frequencies of Visits. Let $(X, \mathscr{B}, \mu)$ be a *probability* space and let $T : X \to X$ be an *ergodic* measure-preserving transformation. Let $A \in \mathscr{B}$ be a measurable set of positive measure $\mu(A) > 0$. Given $x \in X$, the frequencies of visits of x to A up to time n are given by

$$\frac{Card\ \{\ 0 \leq k \leq n-1, \qquad T^k(x) \in A\ \}}{n} = \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k(x)),$$

as we have already seen at the beginning of Chapter 3. If we apply Birkhoff ergodic theorem to the function $f = \chi_A$, which is measurable since $A \in \mathscr{B}$ and integrable since $\int \chi_A \mathrm{d}\mu = \mu(A) \leq \mu(X) < +\infty$, we get that for almost every x

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k(x)) = \int \chi_A \mathrm{d}\mu = \mu(A).$$

Thus, for almost every x in A the limit of the frequencies of visits exists and is equal to $\mu(A)$:

$$\lim_{n\to\infty} \frac{Card\ \{\ 0 \le k \le n-1, \qquad T^k(x) \in A\ \}}{n} = \mu(A) \quad \text{for } \mu-\text{a.e. } x \in X.$$

Example 3.7.1. Let R_α be an irrational rotation. We showed in §3.6 that R_α is ergodic with respect to λ. Thus, if we take as set an interval $[a,b]$, for almost every $x \in [0,1]$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{[a,b]}(R_\alpha^k(x)) = \lambda([a,b]) = b-a.$$

Remark 3.7.1. In the special case of the rotation, one can prove that actually the conclusion of Birkhoff ergodic theorem holds *for all* initial points $x \in X$. In particular, for example, it holds for $x=0$. Thus, since $R_\alpha^k(0) = \{k\alpha\}$ where $\{\cdot\}$ denotes the fractional part, we have

$$\lim_{n\to\infty} \frac{Card\{0 \le k < n,\ \{k\alpha\} \in [a,b]\}}{n} = \lambda([a,b]) = b-a.$$

We say that the sequence $(\{k\alpha\})_{k\in\mathbb{N}}$ is *equidistributed* in $[0,1]$.

*** Exercise 3.7.1.** Let α be irrational. Show that if

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \chi_{[a,b]} \left((R_\alpha^k(x) \right)$$

exists for almost every point $x \in [0,1]$, then it exists *for all* points y and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{[a,b]} \left((R_\alpha^k(y) \right) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{[a,b]} \left((R_\alpha^k(x) \right).$$

2. Borel Normal Numbers. Let $x \in [0,1]$ and consider its *binary expansion*, that is

$$x = \sum_{i=1}^{\infty} \frac{a_i}{2^i},$$

where the $a_i \in \{0,1\}$ are the digits of the binary expansion of x. Remark that the binary expansion is unique for *almost every*[1] $x \in X$.

Definition 3.7.1. A number $x \in [0,1]$ is called *normal in base* 2 if the frequency of occurrence of the digit 0 is the binary expansion and the frequency of occurrence of the digit 1 both exist and equal 1/2.

Theorem 3.7.3 (Borel theorem on normal numbers). *Almost every $x \in [0,1]$ is normal in base* 2.

Proof. Let us prove the Theorem using Birkhoff ergodic theorem. Consider the doubling map $T(x) = 2x$ mod 1. We proved that T preserves the probability measure λ on $X = [0,1]$ and is ergodic with respect to λ. Recall that we showed in §1.4.2 that

$$x = \sum_{i=1}^{\infty} \frac{a_i}{2^i} \quad \Rightarrow \quad T^k(x) = \sum_{i=1}^{\infty} \frac{a_{k+i}}{2^i},$$

that is, the doubling map act as a *shift* on the digits of the binary expansion of x. Since the first digit a_1 of the expansion is clearly $a_1 = 0$ if and only if $x \in [0,1/2)$ and $a_1 = 0$ if and only if $x \in [1/2,1]$, this shows that, since a_{k+1} is the first digit of the expansion of $T^k(x)$,

$$a_{k+1} = \begin{cases} 0 & \text{iff } T^k x \in [0,1/2) \\ 1 & \text{iff } T^k x \in [1/2,1] \end{cases}$$

[1] The numbers for which it is not unique are exactly the ones of the form $k/2^n$, for which one has two expansions, one with a tail of 0 and one with a tail of 1 in the digits. Numbers of this form are clearly countable and thus have Lebesgue measure zero.

Thus,

$$\frac{Card\{\, 1 \le k \le n \quad a_k = 0\,\}}{n} = \frac{Card\{\, 0 \le k < n \quad a_{k+1} = 0\,\}}{n}$$
$$= \frac{Card\{\, 0 \le k < n \quad T^k(x) \in [0,1/2)\,\}}{n}.$$

Since T is ergodic, by Birhoff ergodic theorem applied to $f = \chi_{[0,1/2)}$, for almost every $x \in [0,1]$

$$\lim_{n\to\infty} \frac{Card\{\, 1 \le k \le n \quad a_k = 0\,\}}{n} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \chi_{[0,1/2)}(T^k(x)) = \lambda([0,1/2)) = 1/2,$$

thus the frequency of occurrence of 0 is 1/2. Similarly, for almost every $x \in [0,1]$

$$\lim_{n\to\infty} \frac{Card\{1 \le k \le n \quad a_k = 1\}}{n} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \chi_{[1/2,1]}(T^k(x)) = \lambda([1/2,1]) = 1/2.$$

Remark that the intersection of two full measure sets has full measure (since the complement is the union of two measure zero sets, which has measure zero). We conclude that for almost every $x \in [0,1]$ the frequency of both 0 and 1 exists and equals 1/2, thus almost every x is normal in base 2. □

Exercise 3.7.2. Consider the unit interval $[0,1]$ with the Lebesgue measure. Let $r \ge 2$ be an integer.

(a) Give a similar definition of a number which is *normal in base* r;

(b) Show that almost every $x \in [0,1]$ is normal base r;

(c) Deduce that almost every $x \in [0,1]$ is simultaneously normal with respect to any base $r = 2, 3, \dots, n, \dots$.

3. Leading digits of powers of two. Consider the sequence $(2^n)_{n\in\mathbb{N}}$ of powers of two:

$$1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, \dots$$

The *leading digit* of a number is simply the first digit in the decimal expansion. For example, the leading digit of 512 is 5. The leading digits of the previous sequence are written in bold font:

1, **2**, **4**, **8**, **1**6, **3**2, **6**4, **1**28, **2**56, **5**12, **1**024, **2**048, ...

Consider the sequence of the leading digits:

$$1, 2, 4, 8, 1, 3, 6, 1, 2, 5, 1, 2, \dots$$

Exercise 3.7.3. What is the frequency of the digit 1 in the sequence of leading digits of $(2^n)_{n\in\mathbb{N}}$?

We will use Birhoff Ergodic Theorem to answer this question. More in general, we will show that the frequency of the digit k as leading digit in the sequence $(2^n)_{n\in\mathbb{N}}$ is given by

$$\log_{10}\left(1 + \frac{1}{k}\right)$$

where $\log_{10}$ denotes the logarithm in base 10 (that is, $\log_{10}(a) = b$ if and only if $10^a = b$). In particular, the frequency of occurrency of the digit 1 in the leading digits of $(2^n)_{n\in\mathbb{N}}$ is $\log_{10} 2$.

Notice that the leading digit of 2^n is k if and only if there exists an integer $r \ge 0$ such that

$$k10^r \le 2^n < (k+1)10^r.$$

For example, $2 \cdot 100 \le 256 < 3 \cdot 100$ shows that the leading digit of 256 is 2.

Taking logarithms in base 10 and using the properties of logarithms (as $\log_{10}(ab) = \log_{10}(a) + \log_{10}(b)$ and $\log_{10} 10^r = r$), this shows that

$$\begin{aligned}\log_{10}(k10^r) &\leq \log_{10} 2^n < \log_{10}((k+1)10^r),\\ \log_{10} k + r &\leq n \log_{10} 2 < \log_{10}(k+1) + r.\end{aligned}$$

Thus, equivalently,

$$(n \log_{10} 2 \mod 1) \in I_k = [\log_{10} k, \log_{10}(k+1)]\,.$$

Notice that if we call $\alpha = \log_{10} 2$, the sequence

$$\begin{aligned}(n \log_{10} 2 \mod 1)_{n\in\mathbb{N}} &= 0, \quad \log_{10} 2 \mod 1, \quad 2\log_{10} 2 \mod 1, \quad 3\log_{10} \mod 1, \dots\\ &= 0, \quad \log_{10} 2 \mod 1, \quad \log_{10} 2 + \log_{10} 2 \mod 1, \quad 2\log_{10} 2 + \log_{10} 2 \mod 1, \dots\end{aligned}$$

is the orbit $\mathcal{O}^+_{R_\alpha}(0)$ of 0 under the rotation by α. Thus,

$$\begin{aligned}&\frac{Card\{\, 0 \leq n < N \quad \text{such that the leading digit of } 2^n \text{ is } k\,\}}{N} =\\ &\frac{Card\{\, 0 \leq n < N \quad \text{such that } (n\log_{10} 2 \mod 1) \in I_k\,\}}{N} =\\ &\frac{Card\{\, 0 \leq n < N \quad \text{such that } R_\alpha^n(0) \in I_k\,\}}{N} = \frac{1}{N}\sum_{n=0}^{N-1} \chi_{I_k}(R_\alpha^n(0)).\end{aligned}$$

One can show that $\log_{10} 2$ is irrational, thus R_α is an irrational rotation and hence it is ergodic with respect to the Lebesgue measure. By Remark 3.7.1, the Birkhoff sums of an ergodic rotation converge for all points to the integral, so

$$\begin{aligned}\lim_{N\to\infty} \frac{Card\{0 \leq n < N \quad \text{s.t. the leading digit of } 2^n \text{ is } k\}}{N} &= \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N-1} \chi_{I_k}(R_\alpha^n(0))\\ = \lambda(I_k) = \log_{10}(k+1) - \log_{10} k = \log_{10}\left(1 + \frac{1}{k}\right).\end{aligned}$$

Exercise 3.7.4. Consider the sequence $\{3^n\}_{n\in\mathbb{N}}$ of powers of 3:

$$1, 3, 9, 81, 243, 729, 2187, 6561, \dots$$

The *second* leading digits in the expansion in base 10, starting from $n \geq 3$ (so that there is a second digit), are the digits in bold font:

$$8\mathbf{1}, 2\mathbf{4}3, 7\mathbf{2}9, 2\mathbf{1}87, 6\mathbf{5}61, \dots$$

Consider the sequence of *second* leading digits in the expansion in base 10, starting from $n \geq 3$:

$$1, 4, 2, 1, 5, \dots$$

What is the frequency of occurrence of the digit k as *second* leading digit of $\{3^n\}_{n\geq 3}$?

4. Continued Fractions. Let $x \in [0,1]$ and let us express it as a continued fraction $[a_0, a_1, \dots, a_n, \dots]$ where a_i are the entries of the CF expansion. Let us show that for almost every $x \in [0,1]$ the frequency of occurrence of the digit k as entry of the continued fraction of x is given by

$$\frac{1}{\log 2} \log\left(\frac{(k+1)^2}{k(k+2)}\right). \tag{3.1}$$

We showed in §1.7 that the entries of the continued fraction expansion of x are given by the itinerary of $\mathcal{O}^+_G(x)$ with respect to the partition $P_k = (1/(k+1), 1/k]$, that is the entry $a_i = k$ if and only if $G^i(x) \in P_k$ (see Theorem 1.7.1). Thus,

$$\frac{Card\{0 \leq j < n \quad \text{such that} \quad a_j = k\}}{n} = \frac{1}{n}\sum_{j=0}^{n-1} \chi_{P_k}(G^j(x)).$$

Since G is ergodic with respect to the Gauss measure μ, for $\mu-$almost every $x \in [0,1]$ the limit of the previous quantity as $n \to \infty$ exists and is given by

$$\lim_{n\to\infty} \frac{Card\{0 \le j < n \quad \text{such that} \quad a_j = k\}}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \chi_{P_k}(G^j(x))$$

$$= \mu(P_k) = \int_{\frac{1}{k+1}}^{\frac{1}{k}} \frac{1}{\log 2} \frac{\mathrm{d}x}{1+x} = \left.\frac{\log(1+x)}{\log 2}\right|_{\frac{1}{k+1}}^{\frac{1}{k}} = \frac{1}{\log 2} \log\left(\frac{1+\frac{1}{k}}{1+\frac{1}{k+1}}\right) = \frac{1}{\log 2} \ln\left(\frac{\frac{1+k}{k}}{\frac{k+2}{k+1}}\right),$$

which, simplifying, gives (3.1).

One can show that the same conclusion holds for Lebesgue a.e.$x \in X$, since if it failed for a set of λ-positive measure A, it would fail for a set of μ-positive measure, since

$$\mu(A) = \int_A \frac{1}{(1+x)\log 2} \ge \frac{1}{2\ln 2}\lambda(A) > 0.$$

[More in general, one can show that the measure μ and the measure λ have the same sets of measure zero. Measures with these property are called *absolutely continuous* with respect to each other and if a property holds for almost every point according to one such measure, it holds also for almost every point for the other.]

Exercise 3.7.5. This was covered in class

(a) Show that the function $f : [0,1] \to \mathbb{R}$ **defined by**

$$f(x) = \ln(n), \qquad \textbf{if } x \in P_n = \left(\frac{1}{n+1}, \frac{1}{n}\right]$$

is in $L^1(\mu)$ **and that**

$$\int f \mathrm{d}\mu = \sum_{n=1}^{\infty} \frac{\log n}{\log 2} \ln\left(\frac{(n+1)^2}{n(n+2)}\right) < +\infty;$$

(b) Show that for almost every point $x \in [0,1]$

$$\frac{1}{n}\sum_{i=0}^{n-1} \log a_i = \int f(x)\mathrm{d}\mu;$$

(c) Deduce that for almost every point $x \in [0,1]$ **the geometric mean (which is the expression in (3.2)) of the entries of the CF has a limit and**

$$\lim_{N\to\infty} (a_0 a_2 \dots a_{N-1})^{\frac{1}{N}} = \prod_{n=1}^{\infty} \left(\frac{(n+1)^2}{n(n+2)}\right)^{\frac{\log n}{\log 2}}. \tag{3.2}$$

Ergodicity and Birkhoff Ergodic Theorem

The second form of Birkhoff Ergodic theorem shows that ergodicity is sufficient for Boltzmann ergodic Hypothesis to hold. It turns out that it is also necessary: if the conclusion of Birkhoff ergodic theorem holds, that is the time averages converge to the space averages for almost every point and all observables, then the transformation T has to be ergodic. We show this in the Theorem 3.7.4 below. In the same Theorem 3.7.4 we also show how Birkohff ergodic Theorem can be rephrased in terms of measures of sets (see Part (3) in Theorem 3.7.4) to give another useful characterization of ergodicity.

Theorem 3.7.4. *Let* $(X, \mathscr{B}, \mu)$ *be a* probability *space and* $T : X \to X$ *a measure-preserving transformation. The following are equivalent:*

(1) T *is* ergodic*;*

(2) for any $f \in L^1(X, \mathscr{B}, \mu)$ and μ-almost every $x \in X$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int f \mathrm{d}\mu; \tag{3.3}$$

(3) for any $A, B \in \mathscr{B}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \mu(A)\mu(B). \tag{3.4}$$

Saying that (1), (2) and (3) are *equivalent* means one holds if and only if any of the others hold. In particular, (1) equivalent to (2) shows that the conclusion of the second form of Birkhoff Ergodic theorem (Boltzmann ergodic Hypothesis) holds if and only if T is ergodic.

The equivalence between (1) and (3) gives another characterization of ergodicity. We defined ergodicity in terms of triviality of invariant sets ($T^{-1}(A) = A$ implies $\mu(A) = 0$ or 1) and we already saw that equivalently invariant functions are constant ($f \circ T = f$ a.e. implies f constant a.e.). Equivalently, one can define ergodicity by requiring that any two measurable sets A, B

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \mu(A)\mu(B).$$

(Compare this characterization with the definition of mixing in the next section §3.8 and see the comments after (3.43) in §3.8.)

Proof of Theorem 3.7.4. We will show that (1) ⇒ (2), (2) ⇒ (3) and (3) ⇒ (1). This will prove the equivalence.

The implication (1) ⇒ (2) is simply the statement of Birkhoff Ergodic Theorem for ergodic transformations: if T is ergodic, the convergence of time averages to space averages stated in (3.3) holds for all $f \in L^1(\mu)$ and almost every point.

Let us show that (2) ⇒ (3). Assume that (3.3) holds for all $f \in L^1(\mu)$ and almost every point. To show that (3) holds, take any two measurable sets $A, B \in \mathscr{B}$. Consider the characteristic function χ_A. Since $\int \chi_A \mathrm{d}\mu = \mu(A) \leq \mu(X) < \infty$, $\chi_A \in L^1(\mu)$ and we can apply (3.3) to $f = \chi_A$. Thus,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k(x)) = \int \chi_A \mathrm{d}\mu = \mu(A), \qquad \text{for } a.e.\ x \in X.$$

Multiplying both sides by $\chi_B(x)$ we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k(x))\chi_B(x) = \mu(A)\chi_B(x), \qquad \text{for } a.e.\ x \in X. \tag{3.5}$$

Recall that we showed that $\chi_A \circ T = \chi_{T^{-1}(A)}$ (see equation (3.14) in §3.4), thus $\chi_A \circ T^k = \chi_{T^{-k}(A)}$. Let us show now that

$$\chi_A \chi_B = \chi_{A\cap B}.$$

This holds since characteristic functions take only 0 or 1 as values, so the product $\chi_A\chi_B(x)$ is equal to 1 if and only if both $\chi_A(x) = 1$ and $\chi_B(x) = 1$ (otherwise, if one of the two is 0, the product is 0 also). Thus, $\chi_A\chi_B(x) = 1$ if and only if $x \in A$ and $x \in B$, which equivalently means that $x \in A \cap B$. But a function which is 1 on $A \cap B$ and 0 otherwise is exactly the characteristic function $\chi_{A\cap B}$. Thus

$$\chi_A \circ T^k \chi_B = \chi_{T^{-k}(A)}\chi_B = \chi_{T^{-k}(A)\cap B},$$

and (3.5) can be rewritten as

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{T^{-k}(A)\cap B}(x) = \mu(A)\chi_B(x), \qquad \text{for } a.e.\ x \in X. \tag{3.6}$$

Let us integrate both sides of this equation:

$$\int \frac{1}{n}\sum_{k=0}^{n-1}\chi_{T^{-k}(A)\cap B}(x)\mathrm{d}\mu = \frac{1}{n}\sum_{k=0}^{n-1}\int \chi_{T^{-k}(A)\cap B}(x)\mathrm{d}\mu = \frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}A\cap B), \tag{3.7}$$

$$\int \mu(A)\chi_B(x)\mathrm{d}\mu = \mu(A)\int \chi_B(x)\mathrm{d}\mu = \mu(A)\mu(B). \tag{3.8}$$

Thus, the conclusion follows if we can exchange the sign of limit with the sign of integration and show that the limit of the integrals is the integral of the limits:

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}A\cap B) &= \lim_{n\to\infty}\int \frac{1}{n}\sum_{k=0}^{n-1}\chi_{T^{-k}(A)\cap B}(x)\mathrm{d}\mu \qquad \text{(by (3.7))}\\
&= \int \left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\chi_{T^{-k}(A)\cap B}(x)\right)\mathrm{d}\mu \qquad \text{(if one can exchange)}\\
&= \int \mu(A)\chi_B(x)\mathrm{d}\mu \qquad \text{(by (3.6))}\\
&= \mu(A)\mu(B) \qquad \text{(by (3.8))}.
\end{aligned}$$

The step of exchanging the sign of limit with the sign of integration can be justified by using the Dominated Convergence Theorem (see the Extra 3 in §3.4). Thus, we proved (3).

Let us show that (3) $\Rightarrow$ (1). Assume that (3.4) holds for any $A, B \in \mathscr{B}$. Let us show that T is ergodic by using the definition. Let $A \in \mathscr{B}$ be an invariant set. Apply (3.4) to A taking also $B = A$, so that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}A\cap A) = \mu(A)^2. \tag{3.9}$$

Remark that since A is invariant under T, $T^{-k}(A) = A$, so that $T^{-k}A\cap A = A\cap A = A$. Since if we sum n terms equal to $\mu(A)$ and divide by n we get $\mu(A)$

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}A\cap A) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(A) = \lim_{n\to\infty}\mu(A) = \mu(A),$$

equation (3.9) implies that $\mu(A) = \mu(A)^2$. But the only positive real numbers such that $x = x^2$ are $x = 0, 1$. Thus either $\mu(A) = 0$ or $\mu(A) = 1$. This shows that T is ergodic. $\square$