Birkhoff's Ergodic Theorem, Equidistribution and Generic Points

Seminar on Homogeneous Dynamics and Applications
Tel-Aviv University
Roei Raveh

January 2024

1 Introduction

The goal of this talk is to prove Birkhoff's pointwise ergodic theorem and to introduce the notion of equidistribution and generic points. We give a brief overview of the needed background in ergodic theory, as well as some examples and applications in number theory. We begin with a few basic definitions in ergodic theory:

Definition 1.1. Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{F}, \nu)$ be a probability spaces.

1. Let $f : X \to Y$ be a measurable map. Define $f_*\mu(A) = \mu(f^{-1}(A))$ for $A \in \mathcal{F}$, then $f_*\mu$ is a measure on $(Y, \mathcal{F})$ and is called the *pull-back measure of f*.
2. A measurable map $f : X \to Y$ is *measure preserving* if $\mu(f^{-1}(A)) = \nu(A)$ for any $A \in \mathcal{F}$, i.e. if $f_*\mu = \nu$.
3. Let $T : X \to X$ be measure-preserving, then the measure μ is said to be *T-invariant*, $(X, \mathcal{B}, \mu, T)$ is called a *measure-preserving system*, and T a *measure-preserving transformation*

Proposition 1.1. *A measure μ on X is T-invariant if and only if*

$$\int_X f \, \mathrm{d}\mu = \int_X f \circ T \, \mathrm{d}\mu \tag{1.1}$$

for all $f \in L^\infty$. Furthermore, if μ is T-invariant, then (1.1) *holds for all $f \in L^1(\mu)$*

Example 1.1. Consider $(\mathbb{T}, \mathcal{B}, m)$ where $\mathbb{T} = \mathbb{R}/\mathbb{Z} \cong [0, 1)$, $\mathcal{B}$ is the Borel σ-algebra, and m the Lebesgue measure. Define $S : \mathbb{T} \to \mathbb{T}$ as $S(x) = x^2$, then S is not measure-preserving. Clearly,

$$S^{-1}[0, 1/4) = [0, 1/2),$$

Hence,

$$m\left(S^{-1}[0,1/4)\right) = \frac{1}{2} \neq \frac{1}{4} = m\left([0,1/4)\right).$$

Example 1.2. Take $(\mathbb{T},\mathcal{B},m)$ as in the previous example. Let $d \geq 2$ be an integer, and define $T:\mathbb{T}\to\mathbb{T}$ as $T(x)=\{dx\}$, then T is measure-preserving. Indeed, we need to show $T_*m=m$, and it is enough to show for all intervals $[a,b)\subseteq[0,1)$.

$$T^{-1}[a,b) = \dot{\bigcup_{k=0}^{d-1}} \left[\frac{k+a}{d},\frac{k+b}{d}\right).$$

Hence,

$$T_*m\left([a,b)\right) = \sum_{k=0}^{d-1} m\left(\left[\frac{k+a}{d},\frac{k+b}{d}\right)\right) = \sum_{k=0}^{d-1}\frac{b-a}{d} = b-a = m\left([a,b)\right).$$

Definition 1.2. A measure-preserving transformation $T:X\to X$ of a probability space $(X,\mathcal{B},\mu)$ is *ergodic* if for any $B\in\mathcal{B}$, if B is T-invariant, i.e. $T^{-1}B=B$ then $\mu(B)\in\{0,1\}$.

Example 1.3. Consider the map $T:\mathbb{T}\to\mathbb{T}$ as defined in Example 1.2, i.e. $Tx=\{dx\}$, then T is ergodic. For all $n\geq 1$ and $0\leq k\leq d^n-1$ denote $I_d(n,k)=\left[\frac{k}{d^n},\frac{k+1}{d^n}\right)$. Let $E\in\mathcal{B}$,

$$T^{-n}E\cap I_d(n,k) = \left\{\frac{k+x}{d^n} : x\in E\right\},$$

which yields

$$m(T^{-n}E\cap I_d(n,k)) = d^{-n}m(E).$$

Denote $\mathcal{F}_d=\{\varnothing\}\cup\{I_d(n,k):0\leq k\leq d^n-1,\ n\geq 1\}$, then $\mathcal{F}_d$ is a π-system and $\sigma(\mathcal{F}_d)=\mathcal{B}$ (where $\sigma(\mathcal{F}_d)$ is the σ-algebra generated by $\mathcal{F}_d$). Now, let $A\in\mathcal{B}$ be T-invariant and suppose $m(A)\neq 0$. Define $\mu(E)=\frac{m(E\cap A)}{m(A)}$, then μ is a probability measure on $(\mathbb{T},\mathcal{B})$. Since for all $n\geq 1$

$$m(A\cap I_d(n,k)) = d^{-n}m(A),$$

then for any $E\in\mathcal{F}_d$ we have $m(E)=\mu(E)$. Thus, $m=\mu$ on $\sigma(\mathcal{F}_d)=\mathcal{B}$. Thus, for any $B\in\mathcal{B}$,

$$m(A\cap B) = m(A)m(B).$$

In particular $A\in\mathcal{B}$ and thus,

$$m(A) = m(A)^2.$$

Which shows $m(A)=1$.

Exercise 1.1. *Let $(X,\mathcal{B},\mu,T)$ be a measure-preserving system and let $f:X\to\mathbb{R}$ be a measurable function. Suppose that $f=f\circ T$ and T is ergodic, then f is constant μ-almost everywhere.*

2 Birkhoff's Ergodic Theorem

Theorem 2.1 (Birkhoff's Ergodic Theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system, and let $f \in L^1(\mu)$. There exists a T-invariant function $f^* \in L^1(\mu)$, such that*

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

μ-almost everywhere and in $L^1(\mu)$, and

$$\int_X f \, \mathrm{d}\mu = \int_X f^* \, \mathrm{d}\mu \, .$$

If T is also ergodic, then

$$f^*(x) = \int_X f \, \mathrm{d}\mu$$

μ-almost everywhere.

2.1 Application: Normal Numbers

Definition 2.1. Let $\theta \in [0, 1)$ and let $\theta = \sum_{n=1}^{\infty} \frac{a_n}{b^n}$ be its expansion in base b (i.e. $a_n \in \{0, 1, \dots, b-1\}$). θ is said to be *simply normal in base* b if for any $k \in \{0, 1, \dots, b-1\}$,

$$\lim_{n\to\infty} \frac{\#\{1 \le j \le n : a_j = k\}}{n} = \frac{1}{b}.$$

θ is said to be *normal in base* b if for any $k_1, \dots, k_i \in \{0, 1, \dots, b-1\}$,

$$\lim_{n\to\infty} \frac{\#\{1 \le j \le n - i + 1 : a_j = k_1, \dots, a_{j+i-1} = k_i\}}{n} = \frac{1}{b^i}.$$

Finally, θ is called *normal* (sometimes *absolutely normal* or *completely normal*) if it is normal in base b, for all $b \ge 2$.

Example 2.1. The number

$$\frac{123,456,789}{999,999,999} = 0.\overline{0123456789}$$

is simply normal in base 10. However, any rational number is not normal in any base. The number

$$0.12345678910111213141516171819202122232425262728 29...$$

is normal in base 10, as well as the number

$$0.23571113171923293137414347535961677173 79...$$

that was proven to be normal (in base 10) by Copeland and Erdős.

It is fairly easy to construct a number that is normal in a given base; while it is incredibly difficult to construct an absolutely normal number. Furthermore, there is no proof to the normality of $\sqrt{2}, e, \pi$ or numbers similar to them.

Theorem 2.2 (Borel)**.** *Let $\mathcal{N}$ be the set of absolutely normal numbers, then $m(\mathcal{N}) = 1$.*

Proof. Let $b \geq 2$ an integer and let $\mathcal{N}_b$ be the set of normal numbers in base b. Define $T_b : \mathbb{T} \to \mathbb{T}$ as $T_b x = \{bx\}$. For $x \in [0,1)$ let $x = \sum_{n=1}^{\infty} \frac{a_n}{b^n}$ be its base b expansion.
Let $k_1, \ldots, k_i \in \{0, 1, \ldots, b-1\}$, and denote $p = k_1 b^{i-1} + \ldots + k_{i-1} b + k_i$. Since

$$Tx = \sum_{n=1}^{\infty} \frac{a_{n+1}}{b^n},$$

for any $1 \leq j \leq n - i + 1$ we have $a_j = k_1, \ldots, a_{j+i-1} = k_i$ if and only if

$$T^{j-1} x \in \left[\frac{p}{b^i}, \frac{p+1}{b^i}\right) =: A.$$

Hence,

$$\#\{1 \leq j \leq n-i+1 : a_j = k_1, \ldots, a_{j+i-1} = k_i\} = \sum_{j=1}^{n-i+1} \mathbb{1}_A(T^{j-1}x) = \sum_{j=0}^{n-i} \mathbb{1}_A(T^j x).$$

By Theorem 2.1,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-i} \mathbb{1}_A(T^j x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_A(T^j x) = \int_{\mathbb{T}} \mathbb{1}_A \, \mathrm{d}m = \frac{1}{b^i}, \quad m\text{-a.e.}$$

since T_b is ergodic. Thus,

$$\lim_{n\to\infty} \frac{\#\{1 \leq j \leq n-i+1 : a_j(x) = k_1, \ldots, a_{j+i-1}(x) = k_i\}}{n} = \frac{1}{b^i}, \quad m\text{-a.e.}$$

Meaning $m(\mathbb{T} \setminus \mathcal{N}_b) = 0$ for all $b \geq 2$. Therefore

$$\mathbb{T} \setminus \mathcal{N} = \bigcup_{b=2}^{\infty} \mathbb{T} \setminus \mathcal{N}_b,$$

is a set of measure zero. Hence, $m(\mathcal{N}) = 1$. □

Remark. Although the set of non-normal numbers is of measure zero, it is uncountable. For instance, every element of the middle-thirds Cantor set is non-normal.

In order to prove Theorem 2.1 we will need Theorem 2.3.

2.2 Maximal Ergodic Theorem

Theorem 2.3 (Maximal Ergodic Theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a probability space and let $g \in L^1(\mu)$ be a real-valued function. For any $\alpha \in \mathbb{R}$, define*

$$E_\alpha = \left\{ x \in X : \sup_{n \geq 1} \left(\frac{1}{n} \sum_{k=0}^{n-1} g\left(T^k x\right) \right) > \alpha \right\}.$$

Then

$$\alpha \mu\left(E_\alpha\right) \leq \int_{E_\alpha} g \, \mathrm{d}\mu \leq \|g\|_1.$$

Moreover, $\alpha \mu\left(E_\alpha \cap A\right) \leq \int_{E_\alpha \cap A} g \, \mathrm{d}\mu$ for any T-invariant set A, i.e., $T^{-1}A = A$.

To prove Theorem 2.3 we will need the following proposition:

Proposition 2.1 (Maximal Inequality)**.** *Let $U : L^1(\mu) \to L^1(\mu)$ be a positive linear operator with $\|U\| \leq 1$. Define*

$$f_0 = 0, \quad f_n = \sum_{i=0}^{n-1} U^i f \quad \forall n \geq 1$$

and $F_N = \max\{f_n : 0 \leq n \leq N\}$. Then for all $N \geq 1$,

$$\int_{\{F_N > 0\}} f \, \mathrm{d}\mu \geq 0.$$

Proof of Theorem 2.3. Let $U : L^1(\mu) \to L^1(\mu)$ be the operator $Uf = f \circ T$. Clearly, U is a positive linear operator with $\|U\| \leq 1$. Let $g \in L^1(\mu)$, $\alpha \in \mathbb{R}$ and A a T-invariant set, and denote $f = \mathbb{1}_A \cdot (g - \alpha)$. Now, define $\{f_n\}_{n=0}^\infty$ and $\{F_N\}_{N=0}^\infty$ as stated in Proposition 2.1. Then,

$$E_\alpha = \bigcup_{N=0}^\infty \{F_N > 0\}.$$

Therefore, $\int_{E_\alpha} f \, \mathrm{d}\mu \geq 0$ which means $\int_{E_\alpha \cap A} g \, \mathrm{d}\mu \geq \alpha \mu(E_\alpha \cap A)$. □

Proof of Proposition 2.1. Let $N \geq 1$, clearly $F_N \in L^1(\mu)$. Since U is positive and linear and because $F_N \geq f_n$ for all $0 \leq n \leq N$, we have

$$UF_N + f \geq Uf_n + f = f_{n+1}.$$

Hence,

$$UF_N + f \geq \max_{1 \leq n \leq N} f_n.$$

Denote $P = \{x \in X : F_N(x) > 0\}$. Since $f_0 = 0$, for all $x \in P$ we have

$$F_N(x) = \max_{0 \leq n \leq N} f_n(x) = \max_{1 \leq n \leq N} f_n(x).$$

Therefore, for all $x \in P$

$$UF_N(x) + f(x) \geq F_N(x).$$

We have $F_N \geq 0$ and thus $UF_N \geq 0$. Hence,

$$\begin{aligned}
\int_P f \, \mathrm{d}\mu &\geq \int_P F_N \, \mathrm{d}\mu - \int_P UF_N \, \mathrm{d}\mu \\
&= \int_X F_N \, \mathrm{d}\mu - \int_P UF_N \, \mathrm{d}\mu && (F_N(x) = 0 \text{ for all } x \notin P) \\
&\geq \int_X F_N \, \mathrm{d}\mu - \int_X UF_N \, \mathrm{d}\mu \\
&= \|F_N\|_1 - \|UF_N\|_1 \geq 0 && (\text{since } \|U\| \leq 1).
\end{aligned}$$

□

It would be beneficial to state a similar result for a lower bound:

Corollary 2.1. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a probability space and let $g \in L^1(\mu)$ be a real-valued function. For any $\beta \in \mathbb{R}$, define*

$$E^\beta = \left\{ x \in X : \sup_{n \geq 1} \left(\frac{1}{n} \sum_{k=0}^{n-1} g\left(T^k x\right) \right) < \beta \right\}.$$

Then

$$\beta \mu\left(E^\beta\right) \geq \int_{E^\beta} g \, \mathrm{d}\mu \, .$$

Moreover, $\beta \mu\left(E^\beta \cap A\right) \geq \int_{E^\beta \cap A} g \, \mathrm{d}\mu$ for any T-invariant set A.

We are now ready to prove Birkhoff's ergodic theorem:

Proof of Theorem 2.1. Let $f \in L^1(\mu)$ and WLOG assume that $f \geq 0$. For all $x \in X$, define

$$f^*(x) = \limsup_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f\left(T^k x\right),$$

$$f_*(x) = \liminf_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f\left(T^k x\right).$$

For all $n \geq 1$ and $x \in X$,

$$\frac{n+1}{n} \left(\frac{1}{n+1} \sum_{k=0}^{n} f\left(T^k x\right) \right) = \frac{1}{n} \sum_{k=0}^{n} f\left(T^k x\right) = \frac{1}{n} \sum_{k=0}^{n-1} f\left(T^k \left(Tx\right)\right) + \frac{1}{n} f(x) \tag{2.1}$$

By taking the limit along a subsequence for which the LHS of (2.1) converges to the lim sup, we can deduce $f^* \leq f^* \circ T$. In the same way, taking the limit along a subsequence for which the RHS of (2.1) converges to the lim sup, we can deduce $f^* \geq f^* \circ T$. A similar argument for f_* shows that

$$f^* = f^* \circ T, \quad f_* = f_* \circ T \tag{2.2}$$

Now fix rationals $\alpha > \beta$, and write

$$E_\alpha^\beta \{x \in X : f_* < \beta \text{ and } f^*(x) > \alpha\}.$$

We have $T^{-1} E_\alpha^\beta = E_\alpha^\beta$ and $E_\alpha^\beta \subseteq E_\alpha$. By Theorem 2.3,

$$\int_{E_\alpha^\beta} f \, \mathrm{d}\mu \geq \alpha \mu(E_\alpha^\beta). \tag{2.3}$$

And by Corollary 2.1,

$$\int_{E_\alpha^\beta} f \, \mathrm{d}\mu \leq \beta \mu(E_\alpha^\beta). \tag{2.4}$$

The inequalities (2.3) and (2.4) show that $\mu(E_\alpha^\beta) = 0$. Now, since

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

we have $\mu(N) = 0$. Hence,

$$f^*(x) = f_*(x) \quad \mu\text{-a.e.}$$

Denote

$$g_n(x) := \frac{1}{n} \sum_{k=0}^{n-1} f\left(T^k x\right).$$

By Fatou's lemma,

$$\int_X f_* \, \mathrm{d}\mu \leq \liminf_{n \to \infty} \int_X g_n \, \mathrm{d}\mu = \liminf_{n \to \infty} \int_X f \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \, .$$

Using the reverse Fatou lemma,

$$\int_X f^* \, \mathrm{d}\mu \geq \limsup_{n \to \infty} \int_X g_n \, \mathrm{d}\mu = \limsup_{n \to \infty} \int_X f \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \, .$$

Meaning,

$$\int_X f \, \mathrm{d}\mu = \int_X f^* \, \mathrm{d}\mu \, .$$

Furthermore, since $g_n \xrightarrow[\mu\text{-a.e.}]{n \to \infty} f^*$ and $\|g_n\|_1 \xrightarrow{n \to \infty} \|f^*\|_1$ we can deduce

$$g_n \xrightarrow[L^1(\mu)]{n \to \infty} f^*.$$

□

Remark. Our use of Fatou's lemma and the reverse Fatou lemma was possible due to the assumption that $f \geq 0$. This ensures that the integral always has value (however, it may be infinite).

3 Equidistribution and Generic Points

Throughout this section we assume that $(X, \mathcal{B}, \mu, T)$ is a measure-preserving system, X an LCSC topological space, $\mathcal{B}$ the Borel σ-algebra, μ a probability measure on X, and $T : X \to X$ continuous.

Recall. A topological space X is LCSC if it is Hausdorff, locally compact, and second-countable. In addition, $C_c(X)$ (the set of continuous functions with compact support) is a separable metric space with respect to the uniform norm,

$$\|f\|_\infty = \sup\{|f(x)| : x \in X\}.$$

However, $C_c(X)$ is not a complete metric space and its completion is the space of continuous functions f that tends to zero outside of compact sets, i.e. for every $\varepsilon > 0$ there exists $K \subseteq X$ compact, such that $\sup\{|f(x)| : x \in X \setminus K\} < \varepsilon$.

Definition 3.1. A sequence of elements (x_n) is *equidistributed with respect to* μ if for any $f \in C_c(X)$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(x_j) = \int_X f \, \mathrm{d}\mu \, . \tag{3.1}$$

Equivalently, (x_n) is equidistributed if

$$\frac{1}{n} \sum_{j=1}^{n} \delta_{x_j} \to \mu$$

in the weak*-topology.

Remark. When dealing with $X = [a, b] \subseteq \mathbb{R}$ and the Lebesgue measure it is common to replace $C_c([a, b])$ with the Riemann integrable functions on $[a, b]$. When $X = \mathbb{T}$ we sometimes say that the sequence (x_n) is *uniformly distributed modulo* 1.

The notion of equidistribution strengthens the notion of topological-density. We want our sequence to have enough information about the measure to reconstruct it. The following result gives us two different ways to think about equidistribution in the particular case of $([0, 1], \mathcal{B}, m)$:

Theorem 3.1 (Weyl's criterion)**.** *Let $(x_n) \subseteq [0, 1]$, the following are equivalent:*

(I) The sequence (x_n) is equidistributed.

(II) For all $k \in \mathbb{Z} \setminus \{0\}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} e^{2\pi i k x_j} = 0.$$

(III) For any $[a, b] \subseteq [0, 1]$,

$$\lim_{n\to\infty} \frac{\#\{1 \le j \le n : x_j \in [a, b]\}}{n} = b - a.$$

Example 3.1. For all $\alpha \in \mathbb{T} \setminus \mathbb{Q}$, the sequence $(n\alpha)_{n\in\mathbb{N}}$ is equidistributed. By Weyl's criterion it suffices to prove (II), and indeed, for all $k \in \mathbb{Z} \setminus \{0\}$,

$$\frac{1}{n}\sum_{j=1}^{n} e^{2\pi i k j\alpha} = \frac{1}{n}\sum_{j=1}^{n}\left(e^{2\pi i k\alpha}\right)^{j} = \frac{e^{2\pi i k\alpha}}{n}\frac{1-e^{2\pi i k n\alpha}}{1-e^{2\pi i k\alpha}} \xrightarrow[n\to\infty]{} 0.$$

Proof of Theorem 3.1. (I) $\iff$ (II): (I) implies (II) from the definition of equidistribution. Conversely, (II) implies that (3.1) holds for trigonometric polynomials, and since they are dense in $C([0,1])$ this implies (I).

(I) $\iff$ (III): Assume (I) and let $[a,b] \subseteq [0,1]$. Let $\varepsilon > 0$ and define

$$f^{+}(x) = \begin{cases} 1 & x \in [a,b], \\ \frac{x-(a-\varepsilon)}{\varepsilon} & x \in [\max 0, a-\varepsilon, a), \\ \frac{(b+\varepsilon)-x}{\varepsilon} & x \in (b, \min b+\varepsilon, 1], \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f^{-}(x) = \begin{cases} 1 & x \in [a+\varepsilon, b-\varepsilon], \\ \frac{x-a}{\varepsilon} & x \in [a, a+\varepsilon], \\ \frac{b-x}{\varepsilon} & x \in [b-\varepsilon, b], \\ 0 & \text{otherwise.} \end{cases}$$

Then $f^{-}(x) \le \mathbb{1}_{[a,b]}(x) \le f^{+}(x)$ for all $x \in [0,1]$, and

$$\int_{[0,1]} \left(f^{+} - f^{-}\right) \mathrm{d}m \le 2\varepsilon.$$

Thus,

$$\frac{1}{n}\sum_{j=1}^{n} f^{-}(x_j) \le \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[a,b]}(x_j) \le \frac{1}{n}\sum_{j=1}^{n} f^{+}(x_j).$$

Since $f^{+}, f^{-} \in C([0,1])$, by equidistribution we get

$$b-a-2\varepsilon \le \int_{[0,1]} f^{-} dm \le \liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[a,b]}(x_j)$$
$$\le \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[a,b]}(x_j) \le \int_{[0,1]} f^{-} dm \le b-a+2\varepsilon$$

Thus,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[a,b]}(x_j) = \limsup_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{[a,b]}(x_j) = b-a$$

as required. Conversely, approximate f with simple functions. □

Definition 3.2. A point $x \in X$ is called *generic* (with respect to μ and T) if the sequence of points along the orbit $(T^n x)_{n\in\mathbb{N}}$ is equidistributed with respect to μ.

Remark. If μ and ν are T-invariant probability measures and $x \in X$ is generic with respect to both μ and ν, then $\mu = \nu$. Since for any $f \in C_c(X)$,

$$\int_X f \,\mathrm{d}\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f\left(T^j x\right) = \int_X f \,\mathrm{d}\nu\,.$$

The notion of a generic point is closely related to Birkhoff's ergodic theorem. The main difference being that Birkhoff's ergodic theorem fixes a function, while generic points allows us to use the "ergodic property" for a large family of functions as the next proposition shows:

Proposition 3.1. *Suppose T is ergodic, then μ-almost all $x \in X$ are generic with respect to μ and T.*

Proof. Let $\{f_n\}_{n=1}^{\infty}$ be a dense sequence in $C_c(X)$. Let $n \in \mathbb{N}$, by Theorem 2.1, there exists a set of measure zero E_n, such that for any $x \in X \setminus E_n$ we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} f_n(T^j x) = \int_X f_n \,\mathrm{d}\mu\,.$$

Denote $X' = X \setminus \bigcup_{n=1}^{\infty} E_n$, then $\mu(X') = 1$. We claim that every $x \in X'$ is generic. Indeed, let $x_0 \in X'$, $f \in C_c(X)$ and $\varepsilon > 0$, there exists $n \in \mathbb{N}$ such that

$$|f_n(x) - f(x)| < \varepsilon,$$

for all $x \in X$. Hence,

$$\int_X f \,\mathrm{d}\mu - \varepsilon \le \liminf_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} f_n(T^j x_0) \le \limsup_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} f_n(T^j x_0) \le \int_X f \,\mathrm{d}\mu + \varepsilon.$$

Taking $\varepsilon \to 0$ we obtain,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x_0) = \int_X f \,\mathrm{d}\mu\,.$$

□

While Proposition 3.1 proves that almost every point is generic, can we to construct a generic point for a given T?

Example 3.2. Any normal number in base b, is a generic point with respect to m and T_b. Let x be a normal number in base b, when proving Theorem 2.2 we showed that for any interval $I_b(n,k) = \left[\frac{k}{b^n}, \frac{k+1}{b^n}\right)$ condition (III) in Weyl's criterion holds for $(T_b^n x)_{n\in\mathbb{N}}$. Thus, it holds for any interval of the form $\left[\frac{k}{b^n}, \frac{\ell}{b^m}\right)$ with $k < \ell$. We can use those intervals to approximate all other intervals and show condition (III) which would imply the equidistribution of $(T_b^n x)_{n\in\mathbb{N}}$.