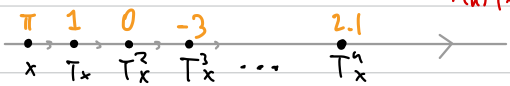
# Measure Theory with Ergodic Horizons
# Lecture 23

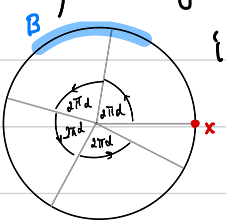
**Classical Pointwise Ergodic Theorem (Birkhoff 1931).** Let $(X, \mathcal{B}, \mu)$ be a probability space. A $(\mathcal{B},\mathcal{B})$-measurable $\mu$-preserving transformation $T$ is ergodic iff for each $f \in L^1(X,\mu)$ and for a.e. $x \in X$,

$$\lim_n \underbrace{\left(\text{average } f \text{ over } \{x, Tx, T^2x, \dots, T^n x\}\right)}_{A_n f(x) := \frac{1}{n+1}\sum_{i=0}^{n} f\circ T^i(x)} = \int f \, d\mu.$$




## Applications.

**(a) Irrational rotations.** Let $\alpha \in [0,1)$ be irrational and $T_\alpha : S^1 \to S^1$ be the rotation by the angle $2\pi\alpha$. It's clear that $T_\alpha$ preserves the Haar measure on $S^1$, i.e. defined by arc-lengths (= push-forward of Lebesgue from $[0,1)$ by $x \mapsto e^{2\pi x}$). We also know from the 99% lemma that $T_\alpha$ is ergodic. Let's apply the ptwise erg. thm. to $\mathbb{1}_B$ for some set $B \subseteq S^1$.

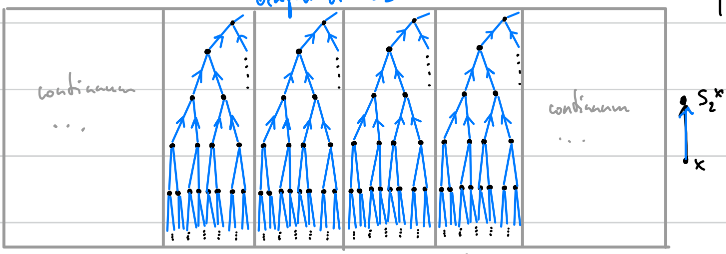



Then $\int \mathbb{1}_B \, d\mu = \mu(B)$, while $A_n \mathbb{1}_B(x) = \frac{1}{n+1} \sum \mathbb{1}_B(T_\alpha^i x) =$
$= \frac{1}{n+1} \left| \{x, T_\alpha x, T_\alpha^2 x, \dots, T_\alpha^n x\} \cap B \right| =$ the density of $B$ in $I_n(x) := \{x, T_\alpha x, \dots, T_\alpha^n x\}$.

The theorem says that the frequency of encountering a point of $B$ as $x$ moves by $T_\alpha$ converges to the proportion of the whole space $S^1$ occupied by $B$, i.e. the number $\mu(B)$.

(b) Let $k \in \mathbb{N}^+$ and let $\nu$ be a prob. measure on $k := \{0, 1, \dots, k-1\}$. Let $\mu := \nu^{\mathbb{N}}$, so $\mu$ is a Bernoulli measure on $k^{\mathbb{N}}$. Let $S_k : k^{\mathbb{N}} \to k^{\mathbb{N}}$, $(x_n) \mapsto (x_{n+1})$ denote the shift, i.e.
$S_k(x_0, x_1, x_2, \dots) = (x_1, x_2, \dots)$.

Graph of $S_2$




orbits of $S_2$

This map is $\mu$-preserving, which we verify for $k=2$ and $\mu = \mu_p$.

Proof. Enough to show on cylinders: Let $w \in 2^n$. Then

$$S_2^{-1}([w]) = [*w], \text{ so}$$

$$\mu_p(S_2^{-1}[w]) = \mu_p([*w]) = \mu_p([0w]) + \mu_p([1w]) = (1-p)\mu_p([w]) + p\,\mu_p([w]) = \mu_p([w]). \square$$

Prop. $S_k$ is mixing, i.e. for any measurable sets $A, B \subseteq k^{\mathbb{N}}$, we have $\lim_{n\to\infty} \mu(S_k^{-n}(A) \cap B) = \mu(A) \cdot \mu(B)$.

Proof. Do first for cylinders, then approximate. Left as HW. $\square$

Cor. $S_k$ is ergodic.

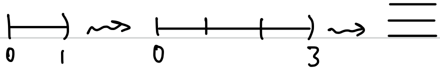
Proof. Let $A$ be any measurable $S_k$-invariant set. Then $\lim_{n\to\infty} \mu(S_k^{-n}(A) \cap A) = \mu(A)^2$. But $S_k^{-n}(A) = A$ by invariance, so $\mu(S_k^{-n}(A) \cap A) = \mu(A \cap A) = \mu(A)$, thus $\mu(A) = \mu(A)^2$. Hence $\mu(A) = 0$ or $1$. $\square$

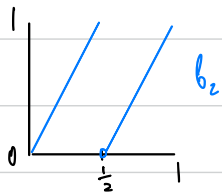
We apply the ptwise ergodic theorem to $\mathbb{1}_{[a]}$, for $a \in k$. Then $\int \mathbb{1}_{[a]}\, d\mu = \mu([a]) = \nu(\{a\})$. On the other hand, $A_n \mathbb{1}_{[a]}(x) = \frac{1}{n+1} \sum \mathbb{1}_{[a]}(S_k^n(x)) = \frac{1}{n+1} \# \text{ of } a \text{ in } \{x_0, x_1, \dots, x_n\} =$

the frequency of the letter $a$ among first $n+1$ letters. Then the theorem says that the frequency of $a$ converges to the "weight" of $a$, i.e. $\nu(a)$.

If instead of $k$ we took an arbitrary probability space $(Y,\nu)$ and apply the ptwise erg. thm to the shift $s: Y^{\mathbb{N}} \to Y^{\mathbb{N}}$, we get the law of large numbers, a most used result in probability theory.

(c) Let $k \in \mathbb{N}$, $k \geq 2$ and define the baker's map $b_k : [0,1) \to [0,1)$.
$$x \mapsto k \cdot x \bmod 1$$

For $k=3$, 


Note that if we take the decimal representation $x = 0.x_0x_1x_2\ldots$ then $b_{10}(x) = 0.x_1x_2x_3\ldots$, so $b_{10}$ shifts the decimal representation. Same for $k$, $b_k$ shift the $k$-ary representation of $x$.

Indeed, let $\varphi : [0,1) \to k^{\mathbb{N}}$ be defined by taking each $x$ to $x_0x_1x_2\ldots$ where $0.x_0x_1x_2\ldots$ is the $k$-ary representation of $x$. This is well-defined on irrationals, and we ignore the rationals since they form a null set. Letting $\lambda$ be the Lebesgue measure on $[0,1)$, what is $\varphi_*\lambda$? It is not hard to check that $\varphi_*\lambda = \nu_u^{\mathbb{N}}$, where $\nu_u$ is the uniform $\frac{1}{k}$ probability measure on $k$, HW. Thus, $\varphi$ is a measure isomorphism from $([0,1),\lambda)$ to $(k^{\mathbb{N}}, \nu_u^{\mathbb{N}})$. Finally, $\varphi$ is $(b_k, s_k)$-equivariant, i.e. $\varphi \circ b_k = s_k \circ \varphi$. Thus, $b_k$ on $([0,1),\lambda)$ is isomorphic/conjugate to $s_k$ on $(k^{\mathbb{N}}, \nu_u^{\mathbb{N}})$.
Hence $b_k$ is $\lambda$-preserving and ergodic.

Applying the ptwise erg. theorem to $\mathbb{1}_{[\frac{i}{10}, \frac{i+1}{10})}$ for $i \in \{0,1,\dots,9\}$, we see that $\lim_{n\to\infty}$ (frequency of $i$ among the first $n+1$ digits of $x$) $= \lambda\left([\frac{i}{10}, \frac{i+1}{10})\right) = \frac{1}{10}$ for a.e. $x \in [0,1)$, as expected.

Now we prove the pointwise ergodic theorem. Recall the following lemma from HW.

**Local-global bridge.** Let $T$ be a measure preserving transformation on a probability space $(X,\mu)$. Let $f \in L^1(X,\mu)$ and $n \in \mathbb{N}$. Then

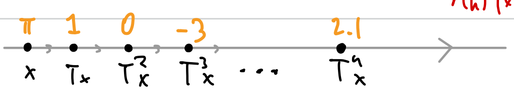
(a) $\int f \, d\mu = \int A_n f \, d\mu$, where $A_n f = \frac{1}{n+1} \sum_{i=0}^{n} f \circ T^i$.

(b) The $\varepsilon \cdot \delta$ trick: Fix $\varepsilon, \delta > 0$ (e.g. $\delta = \varepsilon$). If $Z \subseteq X$ has measure $\leq \varepsilon \cdot \delta$, then on a set $X' \subseteq X$ of measure $\geq 1 - \delta$, the average $A_n \mathbb{1}_Z(x) < \varepsilon$ for all $x \in X'$!

**Proof.** (a) was immediate from $\int f \, d\mu = \int f \circ T \, d\mu$, done in HW, and (b) follows from (a) as follows: by (a), we have $\varepsilon \cdot \delta \geq \mu(Z) = \int \mathbb{1}_Z \, d\mu = \int A_n \mathbb{1}_Z \, d\mu \geq \varepsilon \cdot \mu(\{x \in X : A_n \mathbb{1}_Z(x) \geq \varepsilon\})$ where the last inequality is Chebyshev. Hence $\mu(\{x \in X : A_n \mathbb{1}_Z(x) \geq \varepsilon\}) \leq \delta$, so take $X' := \{x \in X : A_n \mathbb{1}_Z < \varepsilon\}$. $\square$

**Classical Pointwise Ergodic Theorem (Birkhoff 1931).** Let $(X, \mathcal{B}, \mu)$ be a probability space. A $(\mathcal{B},\mathcal{B})$-measurable $\mu$-preserving transformation $T$ is ergodic iff for each $f \in L^1(X,\mu)$ and for a.e. $x \in X$,

$$\lim_n \underbrace{\left(\text{average } f \text{ over } \{x, Tx, T^2x, \dots, T^n x\}\right)}_{A_n f(x) := \frac{1}{n+1} \sum_{i=0}^{n} f \circ T^i(x)} = \int f \, d\mu.$$



Proof (b.) Invariance of limit, & tiling + local-global bridge.