CHAPTER II ERGODIC THEOREMS

II.1 THE MEAN ERGODIC THEOREM

2.1 VON NEUMANN'S MEAN ERGODIC THEOREM

Thm: Let $(X, \mathcal{B}, \mu, T)$ be a ppt. If $f \in L^2(X,\mu)$, then

$$\frac{1}{N}\sum_{k=0}^{N-1} f \circ T^k \xrightarrow[L^2]{} \overline{f} \quad \text{as } N \to \infty,$$

where $\overline{f}$ is invariant. If μ is ergodic, $\overline{f}$ is a.e. constant.

Proof: Note that $f \mapsto f \circ T$ is an L^2-isometry, i.e. $\|f\|_2 = \|f\circ T\|_2$

Step 1: Assume $f = g - g\circ T$ for some $g \in L^2$.

Then

$$\left\| \frac{1}{N}\sum_{k=0}^{N-1} f\circ T^k \right\|_2 = \left\| \frac{1}{N}\left(g - g\circ T^N\right)\right\|_2 \leq \frac{1}{N}\left(\|g\|_2 + \|g\circ T^N\|_2\right)$$

$$= \frac{2}{N}\|g\|_2 \longrightarrow 0 \quad \text{as } N \to \infty$$

$\Longrightarrow$ claim holds for $f \in \mathcal{C} := \{ f = g - g\circ T \mid g \in L^2\}$ with $\overline{f} = 0$.

Step 2: Claim holds for $f \in \overline{\mathcal{C}}$ (closure of $\mathcal{C}$)

exercise.

Importance: we will write $L^2 = \overline{\mathcal{C}} \oplus \overline{\mathcal{C}}^{\perp}$, and for this we need a closed subspace.

Step 3: $f \in \overline{\mathcal{C}}^{\perp} \Longrightarrow f$ is invariant

$$\|f - f\circ T\|_2^2 = \|f\|_2^2 - 2\langle f, f\circ T\rangle + \|f\circ T\|_2^2$$

$$= 2\|f\|_2^2 - 2\langle f, f\circ T - f + f\rangle$$

$$= 2\langle f, f - f\circ T\rangle = 0 \text{ by assumption}$$

Hence, for every $f \in L^2$ we have $f_1 \in \overline{\mathcal{C}}$ and f_2 invariant s.t. $f = f_1 + f_2$ $\Longrightarrow$

$$\frac{1}{N}\sum_{k=0}^{N-1} f \circ T^k \xrightarrow[N\to\infty]{} \underbrace{\overline{f_1}}_{=0} + \underbrace{\overline{f_2}}_{=f_2\text{, since } f_2 = f_2\circ T \text{ a.e.}} = f_2$$

The final claim ("if μ ergodic...") follows from Thm 1.13 ∎

2.2 CONSEQUENCES OF THE MET

- Proof shows: $\overline{f}$ = "L^2-projection of f on the space of invariant functions"
- Cor.: A ppt $(X, \mathcal{B}, \mu, T)$ is ergodic if and only if for all $A, B \in \mathcal{B}$:

$$\frac{1}{N}\sum_{k=0}^{N-1} \mu(A \cap T^{-k}B) \longrightarrow \mu(A)\mu(B) \quad \text{as } N\to\infty$$

"Ergodicity = mixing on average"

Proof: "$\Longrightarrow$" Let $f_n = \frac{1}{n}\sum_{k=0}^{n-1} \chi_B \circ T^k$ and $g = \chi_A$

If $f_n \xrightarrow[n\to\infty]{L^2} \overline{f}$, also $\langle f_n, g\rangle \xrightarrow[n\to\infty]{} \langle \overline{f}, g\rangle$.

From the MET: $\overline{f} = \text{const} = \int \chi_B \, d\mu = \mu(B)$, and the claim follows

"$\Longleftarrow$" Take $A = B$ and assume it is an invariant set. We get $\mu(A) = \mu(A)^2$, hence ergodicity. ∎

II.2 THE BIRKHOFF ERGODIC THEOREM

2.3 THE MAXIMAL ERGODIC THEOREM

Def: If $a_1, \dots, a_n \in \mathbb{R}$ and $1 \le m \le n$, then we say that a_k is an m-leader if $a_k + \dots + a_{k+p-1} \geq 0$ for some $1 \le p \le m$.

Lemma: For every n, $1 \le n \le m$, the sum of all n-leaders is ≥ 0.

Proof: No n-leader $\Rightarrow$ claim holds. Let a_k be the first n-leader, and $p \ge 1$ the smallest number s.t. $a_k + \dots + a_{k+p-1} \ge 0$. Then, for all j, $k \le j \le k+p-1$, we have $a_j + \dots + a_{k+p-1} \ge 0$ (otherwise $a_k + \dots + a_{j-1} \ge 0$, and p is not minimal) $\longrightarrow$ a_j is also an n-leader $\Rightarrow$ the sum of all n-leaders from a_1 to a_{k+p-1} is $a_k + \dots + a_{k+p-1} \ge 0$. Apply the same argument to the rest of the sequence, $a_{k+p}, \dots, a_m$, to get the claim ∎

Thm: Let $(X, \mathcal{B}, \mu, T)$ be a mpt, and $f \in L^1(X, \mu)$. Define

$$A = \left\{ x \in X \,\middle|\, \sup_{n \ge 0} \sum_{k=0}^{n} f(T^k x) \ge 0 \right\} = \left\{ x \,\middle|\, \sum_{k=0}^{n} f(T^k x) \ge 0 \text{ for some } n \right\}$$

Then

$$\int_A f \, d\mu \ge 0$$

Proof: Let $A_n := \left\{ x \in X \,\middle|\, \sum_{i=0}^{k} f(T^i x) \ge 0 \text{ for some } 0 \le k \le n \right\}$. Then

$A_n \subseteq A_{n+1}$, and $A_n \uparrow A$. By the dominated conv. thm., it suffices to show that $\int_{A_n} f \, d\mu \ge 0$ $\forall n$.

Fix $n \in \mathbb{N}$. Let $s_m(x)$ denote the sum of n-leaders in the sequence $f(x), f(Tx), \dots, f(T^{n+m+1} x)$ (1)
Let $B_k \subset X$ be the set of points for which $f(T^k x)$ is an n-leader of (1). By the lemma:

$$0 \le \int s_m(x) \, d\mu(x) = \sum_{k=0}^{m+n-1} \int_{B_k} f(T^k x) \, d\mu(x) \qquad (2)$$

Since $f(T^k x) = f(T^{k-1} Tx)$, $x \in B_k \Leftrightarrow Tx \in B_{k-1}$, but only for $k \le m-1$ (check this!)

$\Rightarrow B_k = T^{-1}(B_{k-1}) = \dots = T^{-k}(B_0)$ for $k \le m-1$.

$$\Rightarrow \int_{B_k} f(T^k x)\, d\mu = \int_{T^{-k}(B_0)} f(T^k x)\, d\mu \underset{\uparrow}{=} \int_{B_0} f\, d\mu$$

μ is invariant

$\Rightarrow$ In (2) the first m terms are equal, thus

$$0 \leq m \cdot \int_{B_0} f\, d\mu + \sum_{k=m}^{m+n-1} \underbrace{\int_{B_k} f(T^k x)\, d\mu}_{\leq \int_X |f|\, d\mu} \leq m \int_{A_m} f\, d\mu + n \int |f|\, d\mu$$

Note $B_0 = A_m$

m arbitrary $\Rightarrow \int_{A_m} f\, d\mu \geq -\frac{n}{m} \int |f|\, d\mu \xrightarrow[m\to\infty]{} 0$

this yields the claim. ∎

2.4 THE BIRKHOFF ERGODIC THEOREM

Thm: Let $(X, \mathcal{B}, \mu, T)$ be a ppt and $f \in L^1(X,\mu)$. Then

(a) The limit

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

exists for a.e. $x \in X$;

(b) If $f \in L^p(X,\mu)$, $1 \leq p < \infty$, then $\bar{f} \in L^p(X,\mu)$ and $f_n \xrightarrow{L^p} \bar{f}$;

(c) $\bar{f} = \bar{f} \circ T$ (a.e.);

(d) $\int f\, d\mu = \int \bar{f}\, d\mu$

Proof: (a): For $a, b \in \mathbb{R}$, $a < b$, consider

$$X(a,b) = \left\{ x \in X \;\middle|\; \liminf_{n\to\infty} f_n(x) < a < b < \limsup f_n(x) \right\}$$

Note: this set is measurable and T-invariant, and T preserves $\mu|_{X(a,b)}$. For every $x \in X(a,b)$ we have

$$\sup_{m \geq 0} \sum_{k=0}^{m} \left[f(T^k x) - b \right] \geq 0$$

$\Rightarrow$ Apply Thm 2.3 (max. erg. thm.) to $T: X(a,b) \to X(a,b)$ and $f-b$, yields

$$\left.\begin{array}{l} \int_{X(a,b)} f-b \, d\mu \geq 0 \\ \text{Similarly for } a-f: \\ \int_{X(a,b)} a-f \, d\mu \geq 0 \end{array}\right\} \Rightarrow \int_{X(a,b)} a-b \, d\mu \geq 0 \underset{a<b}{\Rightarrow} \mu(X(a,b)) = 0$$

a,b arbitrary $\Rightarrow$ f_n converges a.e. Let $\bar{f}(x) = \liminf_{n\to\infty} f_n(x)$

(b): $\bar{f}$ is measurable and $f_n \to \bar{f}$ a.e., and

$$\int |\bar{f}| \, d\mu = \int |\liminf_n f_n| \, d\mu \overset{\text{a.e. conv.}}{=} \int \liminf_n |f_n| \, d\mu$$

$$\overset{\text{Fatou}}{\leq} \liminf_n \int |f_n| \, d\mu \leq \liminf_n \frac{1}{n} \sum_{k=0}^{n-1} \int |f \circ T^k| \, d\mu$$

$$\overset{\mu \text{ invariant}}{=} \liminf_n \frac{1}{n} \sum_{k=0}^{n-1} \int |f| \, d\mu = \int |f| \, d\mu$$

$\to \|\bar{f}\|_p \leq \|f\|_p$ ⊛

Thus $\bar{f} \in L^1$. Analogous argumentation shows $\bar{f} \in L^p$, and (one uses the triangle inequality for $\|\cdot\|_p$ in the above computation)

Convergence in norm:

- $f \in L^\infty$. Hence $f \in L^1$ too, and $|\bar{f} - f_n|^p \to 0$ a.e. for $p < \infty$.

  Also: $$|\bar{f}(x)| \leq \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |f(T^k x)| \leq \|f\|_\infty \quad \text{a.e.}$$

  $$\Rightarrow |f_n(x) - \bar{f}(x)|^p \leq 2^p \|f\|_\infty^p \underset{\text{thm.}}{\overset{\text{dom. conv.}}{\Longrightarrow}} f_n \xrightarrow{L^p} \bar{f}$$

- Fix $f \in L^p$, $\varepsilon > 0$. Let $f_0 \in L^\infty$ s.t. $\|f - f_0\|_p \leq \varepsilon/3$, and $N > 0$ s.t.

  $$\|\bar{f}_0 - \frac{1}{n} \sum_{k=0}^{n-1} f_0 \circ T^k \|_p \leq \varepsilon/3 \quad \forall n \geq N$$

  Then

  $$\|\bar{f} - f_n\|_p \leq \underbrace{\|\bar{f} - \bar{f}_0\|_p}_{\mathrm{I}} + \underbrace{\|\bar{f}_0 - \frac{1}{n}\sum_{k=0}^{n-1} f_0 \circ T^k \|_p}_{\mathrm{II}} + \underbrace{\|\frac{1}{n}\sum_{k=0}^{n-1} (f - f_0) \circ T^k \|_p}_{\mathrm{III}}$$

To I: $\overline{f}-\overline{f_0}=\overline{(f-f_0)}$, hence by ⊛

$$\|\overline{f}-\overline{f_0}\|_p \le \|\overline{(f-f_0)}\|_p \le \|f-f_0\|_p \le \varepsilon/3$$

To II: For $m \geq N$, II $\le \varepsilon/3$ by the assumption above

To III: $\|\frac{1}{m}\sum_{k=0}^{m-1}(f-f_0)\circ T^k\|_p \overset{\mu \text{ inv.}}{=} \|\frac{1}{m}\sum_{k=0}^{m-1} f-f_0\|_p \le \|f-f_0\|_p \le \varepsilon/3$

Adding up proves $f_m \xrightarrow{L^p} \overline{f}$.

(c): Simple consequence of

$$\|\overline{f}-\overline{f}\circ T\|_p \le \|\overline{f}-f_m\|_p + \|f_m - f_m\circ T\|_p + \|f_m\circ T-\overline{f}\circ T\|_p$$

for every $m\in\mathbb{N}$ and $f_m \to \overline{f}$ in L^p.

(d): Follows from $f_m \to \overline{f}$ in L^1. ∎

Remark: The BET is also called individual ergodic theorem, because it shows the a.e. convergence of the means $f_m(x)$ for individual paths of the system (property (a) in the thm), while von Neumann's MET only shows L^2 convergence.

2.5 CONSEQUENCES OF THE BET

Corollary: If $(X,\mathcal{B},\mu,T)$ is an ergodic ppt $f\in L^1(X,\mu)$, then

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

Examples:

1) If $f=\chi_A$, then $\mu(A)$ is the relative amount of time the trajectory $x, Tx, T^2x, \dots$ spends in A (holds for a.e. x)

2) Borel's normal number theorem: $x\in[0,1)$ is called normal, if $x=0.d_1d_2d_3\dots$, and

$$\lim_{m\to\infty}\frac{1}{m}\#\{0\le i\le m-1 \mid d_i=k\} = \frac{1}{10} \quad \text{for every } k=0,1,\dots,9$$

Almost every (w.r.t. Lebesgue) $x \in [0,1)$ is normal.

Proof: $T(x) = 10x \mod 1$, preserves Lebesgue meas. μ.

Let $f = \chi_{[\frac{k}{10}, \frac{k+1}{10})}$, then

$$f(T^n x) = 1 \iff d_n = k$$

Use the BET. ∎

3) $T: [0,1] \to [0,1]$, $x \mapsto \sqrt{x}$

- $\mu = \delta_1$ invariant. f continuous, then

$$\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \frac{1}{n}\sum_{k=0}^{n-1} f\left(x^{2^{-k}}\right) \xrightarrow[n\to\infty]{} f(1) = \int f \, d\mu \quad \text{for } x \neq 0,$$

since $x^{2^{-n}} \longrightarrow 1$ for $x \neq 0$.

- $\mu = \delta_0$ also invariant! $\int f \, \delta_0 = f(0) \neq f(1)$ in general!

Still: BET holds for δ_0-a.e. x, i.e. only for $x = 0$

⇝ restricted classes of invariant measures are interesting!
↳ (more informative)

2.6 APPLICATION TO MARKOV CHAINS (cf. [Wa])

A stochastic matrix is irreducible, if for $\forall i,j \in S$ exists $n \in \mathbb{N}$ s.t.

$$0 < p_{ij}^{(n)}, \quad \text{where} \quad P^n = \left(p_{ij}^{(n)}\right)_{i,j \in S}$$

Lemma: P stoch. matrix, $p > 0$ prob. vector with $p^T P = p^T$. Then

$$Q = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} P^k \quad \text{exists. Also}$$

(i) Q is stochastic

(ii) $QP = PQ = Q$

(iii) $v^T P = v^T \Rightarrow v^T Q = v^T$

(iv) $Q^2 = Q$

<u>Proof</u>: Let μ be the (p,P)-Markov measure, T the one-sided shift.

Let $\chi_i := \chi_{[i]}$, $[i] = \{\{x_k\} \mid x_0 = i\}$ cylinder

BET $\Rightarrow$ $\frac{1}{n}\sum_{k=0}^{n-1} \chi_j(T^k x) \xrightarrow[n\to\infty]{} \overline{\chi_j}(x)$ μ-a.e.

Multiply by χ_i, and integrate ($|\chi_{i,j}| < 1$, use mon. conv. thm)

$$\int \frac{1}{n}\sum_{k=0}^{n-1} \chi_i(x)\chi_j(T^k x)\, d\mu \longrightarrow \int \chi_i \overline{\chi_j}\, d\mu$$

$$\parallel$$

$$\frac{1}{n}\sum_{k=0}^{n} \mu([i] \cap T^{-k}[j]) = \frac{1}{n}\sum_k p_i p_{ij}^{(k)}$$

$$\Rightarrow \left(\frac{1}{n}\sum_k P^k\right)_{ij} = \frac{1}{n}\sum_k p_{ij}^{(k)} \longrightarrow \frac{1}{p_i}\int \chi_i \overline{\chi_j}\, d\mu =: q_{ij}$$

(i) – (iv) follows by direct computation.

<u>Thm</u>: Let T denote the (p,P)-Markov shift, $p > 0$. Let Q be as in the Lemma above. Then the following are equivalent:

(i) T is ergodic

(ii) All rows of Q are identical

(iii) Every entry in Q is strictly positive

(iv) P is irreducible

(v) 1 is a simple eigenvalue of P

<u>Proof</u>: (i) $\Rightarrow$ (ii) As in Lemma: $\frac{1}{n}\sum_{k=0}^{n-1} \mu([i] \cap T^{-k}[j]) = p_i q_{ij}$

ergodicity $\parallel$ Cor. 2.2

$\mu[i]\mu[j] = p_i p_j$ $\Rightarrow q_{ij} = p_j$

(ii) $\Rightarrow$ (iii) $p^T P = p^T \overset{\text{Lemma}}{\Longrightarrow} p^T Q = p^T \underset{\text{ident}}{\overset{\text{rows}}{\Longrightarrow}} q_{ij} = p_j > 0 \quad \forall i,j$

(iii) $\Rightarrow$ (iv) $\frac{1}{n}\sum_{k=0}^{n-1} p_{ij}^{(k)} \longrightarrow q_{ij} > 0 \Rightarrow \exists k \in \mathbb{N}: p_{ij}^{(k)} > 0$

(iv) $\Rightarrow$ (iii) Lemma: $QP = Q$. Then: $QP^m = Q \ \forall m$.

Let $S_i = \{ j \in S \mid q_{ij} > 0 \}$. Then $QP^m = Q \Leftrightarrow q_{ij} = \sum_\ell q_{i\ell} p_{\ell j}^{(m)}$

$\Rightarrow q_{i\ell} p_{\ell j}^{(m)} \le q_{ij} \ \forall \ell$, hence if $\ell \in S_i \wedge p_{\ell j}^{(m)} > 0 \Rightarrow q_{ij} > 0$

Now for fixed ij, pick any $\ell \in S_i$ (S_i not empty, since Q stoch.) and choose m s.t. $p_{\ell j}^{(m)} > 0$. Thus $q_{ij} > 0$.

(iii) $\Rightarrow$ (ii) Fix j and let $q_j := \max_i q_{ij} > 0$. If $q_{ij} < q_j$ for some i, then by $Q^2 = Q$:

$$q_{\ell j} = \sum_m q_{\ell m} q_{mj} < \sum_m q_{\ell m} q_j = q_j \quad \forall \ell, \text{ a contradiction}$$

(ii) $\Rightarrow$ (i) Since cylinders generate the whole Borel-σ-algebra, by Cor. 2.2 it suffices to show that

$$\frac{1}{m} \sum_k \mu(A \cap T^{-k} B) = \mu(A)\mu(B), \qquad \begin{aligned} A &= [a_0, \dots, a_i] \\ B &= [b_0, \dots, b_j] \end{aligned}$$

For k suff. large (i.e. $k > i$)

$$\mu(A \cap T^{-k} B) = p_{a_0} p_{a_0 a_1} \cdots p_{a_{i-1} a_i} p_{a_i b_0}^{(k-i)} p_{b_0 b_1} \cdots p_{b_{j-1} b_j}$$

$\Rightarrow$

$$\lim_{m \to \infty} \frac{1}{m} \sum_k \mu(A \cap T^{-k} B) = \underbrace{p_{a_0} p_{a_0 a_1} \cdots p_{a_{i-1} a_i}}_{= \mu(A)} \cdot \underbrace{p_{b_0 b_1} \cdots p_{b_{j-1} b_j}}_{= \mu(B)/p_{b_0}} \cdot \underbrace{\frac{1}{m} \sum_k p_{a_i b_0}^{(k-i)}}_{\to q_{a_i b_0} \overset{(ii)}{=} p_{b_0}}$$

$$= \mu(A)\mu(B)$$

(ii) $\Rightarrow$ (v) Since (ii) $\Rightarrow q_{ij} = p_j \Rightarrow v^T Q = v^T$ iff $v = \lambda p$, $\lambda \in \mathbb{C}$.

Lemma $\Rightarrow$ every left eigenvector of P at eigenvalue 1 is collinear with p.

(v) $\Rightarrow$ (ii) $QP = Q$, hence each row of Q is left eigenvector of P at 1.

Q stochastic and P has a simple eigenvalue 1

$\Rightarrow$ all rows of Q are identical

■

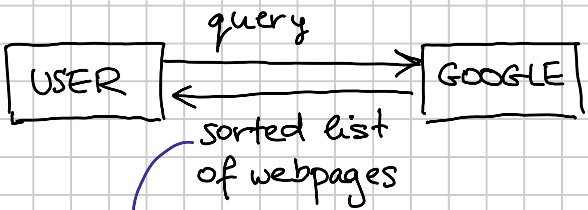
Remark: If P is aperiodic, the (p, P)-Markov shift is mixing

$\hookrightarrow \; :\Longleftrightarrow \; \forall i: \; \gcd\{n \in \mathbb{N} \mid p_{ii}^{(n)} > 0\} = 1$

$\hookrightarrow$ greatest common divisor

2.7 INTERNET SEARCH

Search engine: Google



sorted by: • relevance (frequency and position of keywords, ...)

• PageRank

PageRank:

Webpages are nodes, enumerated $1, 2, \dots, N$

Node 0: "Google node"

Edge from i to j $(i \to j)$, iff webpage i links to webpage j. Add $i \to 0$ and $0 \to i$ $\forall i = 1, \dots, N$.

Let d_i denote the number of outgoing edges from node i; $0 < \varrho < 1$ param.

Define $P \in \mathbb{R}^{(N+1)\times(N+1)}$ by

$$P_{0i} = \frac{1}{N} \qquad P_{i0} = \begin{cases} 1 & , d_i = 1 \\ 1-\varrho & , d_i > 1 \end{cases} \qquad P_{ij} = \begin{cases} \frac{\varrho}{d_i - 1} & , i \to j \quad (i \neq j) \\ 0 & , \text{otherwise} \end{cases}$$

The PageRank of webpage i is p_i, where $p^T P = p^T$ (*)

BET: The rank of a page is the relative time a random walker on the world wide web spends on that page. $1-\varrho$ is the probability to restart the walk in node 0.

Direct computation of (*) unthinkable: $\sim 50 \cdot 10^9$ indexed webpages

http://www.worldwidewebsize.com/

By dynamical systems (BET) approach $\rightsquigarrow$ vector iteration

(cf. numerical eigenvalue computation)

II.3 THE ERGODIC DECOMPOSITION

2.8 CONDITIONAL EXPECTATION

$(X, \mathcal{B}, \mu)$ probability space, $\mathcal{C} \subset \mathcal{B}$ sub-σ-algebra

Def: Let $f \in L^1(X, \mathcal{B}, \mu)$, then the conditional expectation of f w.r.t. $\mathcal{C}$ is the unique function $\mathbb{E}(f|\mathcal{C}) \in L^1(X, \mathcal{C}, \mu)$ s.t.

$$\int \varphi f \, d\mu = \int \varphi \mathbb{E}(f|\mathcal{C}) \, d\mu \qquad \forall \varphi \in L^\infty(X, \mathcal{C}, \mu),$$

or equivalently

$$\int_A f \, d\mu = \int_A \mathbb{E}(f|\mathcal{C}) \, d\mu \qquad \forall A \in \mathcal{C}.$$

Proof of existence and uniqueness:

For $A \in \mathcal{C}$, $\nu_f(A) := \int_A f \, d\mu$ is a finite measure $\nu_f : \mathcal{C} \to [0, \infty)$, with $\nu_f \ll \mu$

$\Rightarrow$ by Radon-Nikodym: $\exists! \, g \in L^1(X, \mathcal{C}, \mu)$ s.t. $g = \frac{d\nu_f}{d\mu}$ $\blacksquare$

Example: Fix $A \in \mathcal{B}$, and let $\mathcal{C} = \{\emptyset, A, A^c, X\}$, μ prob. measure

Then $g \in L^\infty(X, \mathcal{C}, \mu) \iff g = c_1 \chi_A + c_2 \chi_{A^c}$ for $c_1, c_2 \in \mathbb{R}$

$$\Rightarrow \mathbb{E}(f|\mathcal{C}) = \left[\int_A f \, d\mu\right] \chi_A + \left[\int_{A^c} f \, d\mu\right] \chi_{A^c}$$

2.9 ERGODIC LIMIT

- MET: $f \in L^2 \Rightarrow \bar{f}$ is L^2-projection of f on the space of inv. functions
- BET: $\frac{1}{n} \sum_k f \circ T^k \longrightarrow \bar{f}$ in L^1

Q: How does $f \in L^1$ relate to the invariant function $\bar{f}$?

Observe:

μ ergodic $\Rightarrow \bar{f} = \text{const}$, and the const. fcns are the only inv. fcns

$f \equiv c$ m'ble wrt $\{\emptyset, X\}$, the only "inv. sets" if μ ergodic

$\mathcal{J}_T := \{ E \in \mathcal{B} \mid T^{-1}E = E \}$ the σ-alg of inv. sets

Note: $\varphi \in L^1(X, \mathcal{J}_T, \mu)$:

$$\int_E \varphi \circ T \, d\mu \underset{E \text{ inv.}}{=} \int_{T^{-1}E} \varphi \circ T \, d\mu \underset{\mu \text{ inv.}}{=} \int_E \varphi \, d\mu \iff \varphi = \varphi \circ T$$

Step 1: $\bar{f}$ from BET: $\bar{f} = \bar{f} \circ T \implies \bar{f}$ is $\mathcal{J}_T$-measurable and clearly integrable

Step 2: $\varphi \in L^\infty(X, \mathcal{J}_T, \mu)$:

$$\int \varphi \underbrace{\frac{1}{n} \sum_{k}^{n-1} f \circ T^k}_{\xrightarrow[n \to \infty]{L^1} \bar{f}} \, d\mu \underset{\varphi \text{ inv.}}{=} \frac{1}{n} \sum_{k}^{n-1} \int (\varphi f) \circ T^k \, d\mu \underset{T \text{ mpt}}{=} \frac{1}{n} \sum_{k}^{n-1} \int \varphi f \, d\mu = \int \varphi f \, d\mu$$

$$\xrightarrow[n \to \infty]{} \int \varphi \bar{f} \, d\mu$$

$\bar{f} \in L^1(X, \mathcal{J}_T, \mu)$

$$\implies \boxed{\bar{f} = \mathbb{E}(f \mid \mathcal{J}_T)} \quad \text{"projection"}$$

Note:

If there is $A \in \mathcal{J}_T$ s.t. $\nexists E \in \mathcal{J}_T$ with $E \subset A$ and $\mu(A \setminus E) > 0$, then $\bar{f}|_A$ constant, and A invariant $\implies T|_A$ ergodic

Q: Can we decompose μ into ergodic parts?

2.10 CONDITIONAL PROBABILITIES

Standard prob. space: $(X, \mathcal{B}, \mu)$

Complete metric separable

Borel

Thm: $(X, \mathcal{B}, \mu)$: std. prob. space, and $\mathcal{C} \subset \mathcal{B}$ a σ-algebra.

Then there exists a family of prob-measures $\{\mu_x\}_{x \in X}$ s.t.:

(a) $x \longmapsto \mu_x(A)$ is $\mathcal{C}$-measurable for every $A \in \mathcal{B}$;

(b) if f is μ-integrable, then $x \longmapsto \int f \, d\mu_x$ is μ-integrable, and

$$\int \int f \, d\mu_x \, d\mu = \int f \, d\mu$$

(c) if f is μ-integrable then $\int f \, d\mu_x = \mathbb{E}(f | \mathcal{C})(x)$ for μ-a.e. x

If $\mathcal{C} = \mathcal{J}_T$ for a mpt T, then

(d) μ_x is invariant and ergodic for μ-a.e. x

<u>Proof</u>: See [Sa, Section 2.3]

The idea: $f \in C^0(X) \rightsquigarrow$ define $L_x(f) = \bar{f}(x)$

$\overset{\text{Riesz}}{\Longrightarrow} \exists! \, \mu_x$ prob. measure with $L_x(f) = \int f \, d\mu_x$

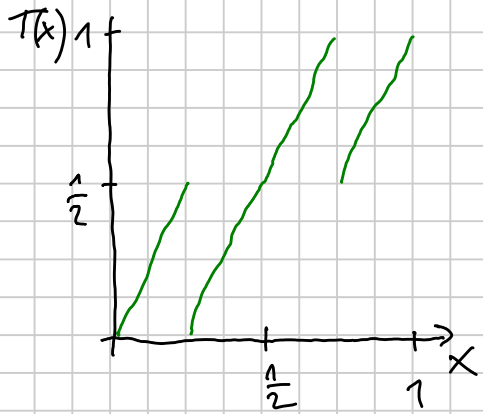
<u>Examples</u>:

1) $(X, \mathcal{B}, \mu, T)$ ergodic $\Rightarrow L_x(f) = \bar{f}(x) \underset{\text{BET}}{\overset{\text{a.e.}}{=}} \int f \, d\mu =: \int f \, d\mu_x$

$\Rightarrow \mu_x = \mu$ a.e.

2) $T = id \Rightarrow L_x(f) = f(x) = \int f \, d\delta_x \longrightarrow \mu_x = \delta_x$

Dirac measure $\longrightarrow \delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$

3) 

two indep. angle doubling maps

$$L_x(f) = \begin{cases} \int_0^{1/2} f(x) \, dx, & x \in [0, 1/2) \\ \int_{1/2}^1 f(x) \, dx, & x \in [1/2, 1) \end{cases}$$

$$\Rightarrow \mu_x = \begin{cases} 2 m_{[0,1/2)}, & x \in [0, 1/2) \\ 2 m_{[1/2,1)}, & x \in [1/2, 1) \end{cases}$$

m: Lebesgue measure

2.11 RELATED TOPICS

1) When does a map have a unique ergodic measure?
$\rightsquigarrow$ "unique ergodicity", cf. [Ma], [BS]

2) Given a measure ν with $\nu \neq \nu \circ T^{-1}$, will $\nu \circ T^{-n}$ converge in some weak or strong sense? (Note: with the right notion of convergence, the limit — if it exists — will be an inv. measure)

Ex: ergodicity is not enough! $\rightsquigarrow$ Markov chains: irreducibility $\Rightarrow \exists_1$ stationary distribution p, but $q^T P^n \longrightarrow p^T$ does not have to hold prob. vectors q. Take $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Aperiodicity of P ($\Rightarrow$ mixingness of the corr. Markov shift) guarantees $q^T P^n \longrightarrow p^T$ $(n \to \infty)$.

II.4 THE SUBADDITIVE ERGODIC THEOREM

2.12 RANDOM WALKS ON GROUPS

Example 1: $\underline{p} = (p_1, \dots, p_d)$ probability vector, $(\Sigma, \mathcal{B}, \mu, \sigma)$ the $\underline{p}$-Bern. shift. Let $v_1, \dots, v_d \in \mathbb{Z}^2$, and $f: \Sigma \longrightarrow \mathbb{Z}^2$, $f(\{x_k\}) = v_{x_0}$

$$f_n(\{x_k\}) := f(\{x_k\}) + f(\sigma\{x_k\}) + \dots + f(\sigma^{n-1}\{x_k\})$$

is a random walk on $\mathbb{Z}^2$

BET (componentwise) $\Longrightarrow$ $\frac{1}{n} f_n(x) \xrightarrow[n\to\infty]{} \sum p_i v_i$ a.e.

$\rightsquigarrow$ Avg. velocity on the long run: $\sum p_i v_i$

Example 2: $X \subset \mathbb{R}^d$ open, $T: X \to X$ diffeomorphism (i.e. both T and T^{-1} are C^1-functions)

$DT(x) \in GL(d)$: derivative

$$DT^n(x) = DT(T^{n-1}x)\,DT^{n-1}(x) = \ldots = DT(T^{n-1}x)\cdot \ldots \cdot DT(Tx)\cdot DT(x)$$

is random walk on $GL(d)$ (this time non-commutative!)

Average stretching by T along trajectory?

$$\rightarrow \underbrace{\frac{1}{n}\log \| DT^n(x)\|}_{=:\, g^{(n)}(x)}$$

cf. Lyapunov exponent

Observe:

$$g^{(n+m)}(x) = \log \| DT^{n+m}(x)\| = \log \| DT^m(T^n x)\, DT^n(x)\|$$

$$\leq \log \| DT^m(T^n x)\| + \log \| DT^n(x)\|$$

$$= g^{(m)}(T^n x) + g^{(n)}(x)$$

"sub-additivity"

2.13 KINGMAN'S SUBADDITIVE ERGODIC THEOREM

Thm: Let $(X, \mathcal{B}, \mu, T)$ be a ppt, and suppose $g^{(n)}: X \to \mathbb{R}$ is a sequence of integrable functions such that $g^{(n+m)}(x) \leq g^{(m)}(T^n x) + g^{(n)}(x)$ $\forall x \in X$ $\forall n, m \geq 0$

Then $\lim\limits_{n\to\infty} \frac{1}{n} g^{(n)}(x)$ exists a.e. (it may be $-\infty$) and it is an invariant function.

Corollary: $T: X \to X$ is a measure pres. diffeomorphism with $\int \|DT\|\, d\mu < \infty$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \| DT^n(x)\| \qquad (*)$$

exists a.e. and is an invariant function.

If, in addition, μ is ergodic, then $(*)$ is constant a.e.

"Asymptotic growth rate in the derivative cocycle"