Ergodicity and Mixing

For the next equivalent condition to ergodicity we need to recall the Dominated convergence theorem.

<u>Theorem</u>: Say α, α_n and ϕ are measurable in $(X, \mathcal{B}, \mu)$ with $|\alpha_n| \le \phi$ a.e. $\forall n$ and $\phi \in L^1(\mu)$ and $\alpha_n \to \alpha$ a.e. then $\int \alpha_n \, d\mu \to \int \alpha \, d\mu$ and $\|\alpha_n - \alpha\|_1 \to 0$

<u>CORR</u> (Bounded convergence theorem) Let $\phi \equiv M$ a constant

<u>Theorem</u> $(X, \mathcal{B}, \mu)$ is probability space and $f:(X)\circlearrowleft$ is a mpT. f is ergodic $\Leftrightarrow$ $\forall A, B \in \mathcal{B}$

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap B) \to \mu(A)\mu(B)$$

<u>Interpretation</u>: on <u>average</u>, $f^{-i}(A)$ and B become independent events so probability of both occurring is the product.

<u>Proof</u> Let $\alpha = \chi_A$ the pointwise ergodic theorem

then $$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(f^i(x)) \to \int \chi_A \, d\mu = \mu(A) \text{ a.e.}$$

multiply by χ_B yields

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_A(f^i(x))\chi_B(x) \to \mu(A)\chi_B \text{ a.e.}$$

Now certainly $0 \le \chi_A(f^i(x))\chi_B(x) \le 1$ so

by the bounded convergence theorem.

$$\frac{1}{n}\sum_{i=0}^{n-1} \int \chi_A(f^i(x))\chi_B(x)\,d\mu \to \int \mu(A)\chi_B\,d\mu$$

but $\chi_A(f^i(x))\chi_B = \chi_{f^{-i}(A)\cap B}(x)$

and so $\int \chi_A(f^i(x))\chi_B\,d\mu = \mu(f^{-i}(A)\cap B)$

and so $\frac{1}{n}\sum_{i=0}^{n-1} \mu(f^{-i}(A)\cap B) \to \mu(A)\mu(B)$

For the converse,

Say the convergence holds and $f^{-1}(E)=E$.

Put $A=B=E$ in the convergence formula

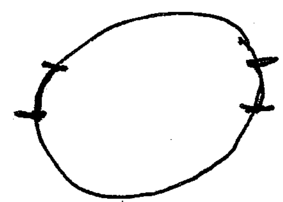
$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(f^{-i}(E)\cap E) \to (\mu(E))^2$$

but $f^{-1}(E)=E$, so $\mu(f^{-i}(E)\cap E)=\mu(E)$, $\forall i$

and so $\mu(E)=(\mu(E))^2$ so $\mu(E)=0$, or 1 and

so f is ergodic

<u>Example</u>: $R_\alpha : S^1 \circlearrowleft$ rigid rotation by $\alpha \notin \mathbb{Q}$. Pick two small intervals I_1 and I_2



$\Rightarrow$ $\mu(R_\alpha^{-i}(I_1)\cap I_2)$ convergences to

$\mu(I_1)\cap\mu(I_2)$ on <u>average</u>.

Now note that ergodicity is equivalent to

$$\frac{1}{n}\sum_{i=0}^{n-1} \mu(f^{-i}(A) \cap B) \to \mu(A)\mu(B)$$

so the convergence is on <u>average</u>. A stronger condition is <u>strong mixing</u>

<u>Def</u>: $f:(X,\mathcal{B},\mu)\circlearrowleft$ a mpt is called strong mixing if $\forall A, B \in \mathcal{B}$

$$\lim_{n\to\infty} \mu(f^{-i}(A)\cap B) \to \mu(A)\mu(B)$$

<u>Interpretation</u> so $f^{-i}(A)$ and B simultaneous occurrence has a probability that tends towards independence. OR $\frac{\mu(f^{-i}(A)\cap B)}{\mu(B)} \to \mu(A)$, proportion of B occupied by $f^{-i}(A)$ converges to proportion of A in X $(\mu(X)=1)$

Remarks (a) Sometimes strong mixing is just called mixing.

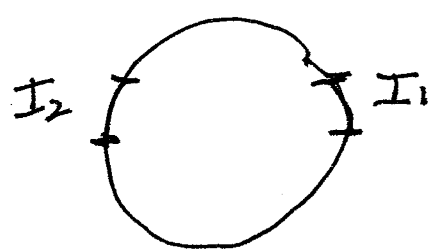
(b) There is an intermediate condition ~~we~~ which we won't consider; weak mixing, which requires

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| \mu(f^{-i}(A) \cap B) - \mu(A)\mu(B) \right| = 0$$

(c) strong mixing $\Rightarrow$ weak mixing $\Rightarrow$ ergodic

---

Example: $R_\alpha : S^1 \circlearrowleft$, $\alpha \notin \mathbb{Q}$ is ergodic but not mixing



the sequence $\mu(R_\alpha^{-i} I_1 \cap I_2)$ has zeros interspersed with nonzeros for arbitrarily large i, so

$\mu(R_\alpha^{i} I_1 \cap I_2)$ doesn't converge.

On the other hand, (Σ_n, σ) with a Bernoulli measure $\mu_{\vec{p}}$ with all $p_i > 0$ is strong mixing. To show this we need to following we state without proof. It says we just need to check the conditions for ergodicity and mixing on a generating semi-Algebra

Theorem: Let $(X, \mathcal{B}, \mu)$ be a probability space and $\mathcal{S}$ is a semialgebra that generates $\mathcal{B}$ and f is a mpt.

(a) f is ergodic $\Leftrightarrow$ $\forall A, B \in \mathcal{S}$,

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

(b) f is mixing $\Leftrightarrow$ $\forall A, B \in \mathcal{S}$

$$\lim_{n\to\infty} \mu(f^{-i}(A) \cap B) \to \mu(A)\mu(B)$$

<u>Example:</u> Bernoulli measures are mixing on (Σ_n, σ)

Recall that the algebra of finite union of cylinder sets generate the Borels. We just give the arguments for cylinder sets: the extension to finite unions is straight forward. The proof is almost identical to that in ET6 Lecture for ergodicity.

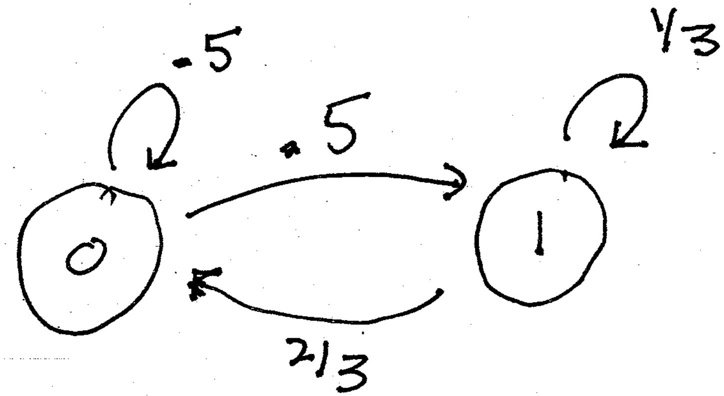
For A and B cylinder sets, we may find N large enough that $\sigma^{-N}(A)$ and B restrict different coordinates of the shift. Then using the definition of the product measure,

$$\mu\left(\sigma^{-i}(A) \cap B\right) = \mu(\sigma^{-i}(A))\,\mu(B)$$
$$= \mu(A)\,\mu(B)$$

all $i \geq N$ since $\sigma^{-i}(A)$ restricts different coordinates than B and the measure is invariant.

MARKOV Measures

We model a 2 state Markov process



- The arrow indicates a possible transition and the number on the arrow indicates the probability of that transion

- So if you are in state 1 there is 1/3 change your next step keeps you there and a 2/3 chance your next step takes you to state 0

- This information is encoded in a stochastic or transition matrix

$$A = \begin{bmatrix} 1/2 & 1/2 \\ 2/3 & 1/3 \end{bmatrix}$$

- So A_{ij} = probability of transitioning from state i to state j.
- We want to use this to get a measure which models the Markov process.
- The measure is on Σ_2, the two-sided full shift.
- Since $\Sigma_2 = \{0,1\}^{\mathbb{Z}}$, we use indices $i = 0,1$, $j = 0,1$ on A.

- We need one more ingredient - an initial probability distribution. In this case

$\vec{p}$: [4/7 3/7] and then $\vec{p}A = \vec{p}$

- This says the probability distribution is invariant under the transition probabilities

- The role of $\vec{p}$ is a little mysterious at first, but we will see the essential role it plays in getting a measure - it also corresponds to an asymptotic, steady state of the system.

- Recall a cylinder set for a block $B = b_0 b_1 \ldots b_k$ starting at place m is

$$_m[B] = \{ \underline{s} \in \Sigma_2 : s_m = b_0, s_{m+1} = b_1, \ldots, s_{m+k} = b_k \}$$

- Define the Markov measure determined by $\vec{p}, A$ on cylinder sets as

$$\mu(_m[B]) = p_{b_0} A_{b_0 b_1} A_{b_1 b_2} \cdots A_{b_{k-1} b_k}$$

- Think of this as a finite sequence of steps
  p_{b_0} is probability you start at b_0, $A_{b_0 b_1}$ is probability you transition to b_1, etc.
- These steps are exactly those determined by B
- Since $\mu(_m[B])$ doesn't depend on m, it is automatically invariant (checking on semi-Algebra..)

Example

$$\mu|_3[01101]) = p_0 A_{01} A_{11} A_{10} A_{01}$$
$$= \frac{4}{7} \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{2}$$

Next time we check additivity which justifies the choice of $\vec{p}_0$