Markov Processes and Random Transformations

Thomas Hille

September 1, 2012

Contents

1 Ergodicity

1.1 Measure Preserving Systems

Definition 1.1. *Let $(X, \mathcal{B}, \mu)$ be a probability space. A measurable map $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is called measure-preserving if $\mu(T^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}$. The measure μ is said to be T-invariant and $(X, \mathcal{B}, \mu, T)$ is called a measure-preserving system.*

Theorem 1.2. *Let $(X, \mathcal{B}, \mu)$ be a probability space and $\mathscr{S}$ a semi-algebra generating $\mathcal{B}$. A measurable map $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ is measure-preserving if and only if $\mu(T^{-1}(B)) = \mu(B)$ for all $B \in \mathscr{S}$*

Proof. Define $\mathcal{D} := \{B \in \mathcal{B} \mid \mu(T^{-1}(B)) = \mu(B)\}$. $\mathcal{D}$ is a Dynkin-System, $\mathscr{S}$ is in particular a π-System and $\mathscr{S} \subseteq \mathcal{D}$. Hence, by Dynkin's Lemma we get that $\mathcal{B} = \sigma(\mathscr{S}) \subseteq \mathcal{D}$.

□

Example 1.1. *(Circle rotation)* Let $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ be the 1-torus, $\alpha \in \mathbb{R}$ and λ the Lebesgue-measure on $\mathbb{T}$. Define

$$R_\alpha : \mathbb{T} \to \mathbb{T}, x \mapsto x + \alpha \mod 1.$$

$\mathscr{S} := \{[a, b) \subseteq \mathbb{T} \mid a \leq b \mod 1\}$ is a semi-algebra generating the Borel σ-algebra $\mathcal{B}$. Let $[a, b) \in \mathscr{S}$, then

$$\lambda(R_\alpha^{-1}([a, b))) = b - a = \lambda([a, b)).$$

So $(\mathbb{T}, \mathcal{B}, \lambda, R_\alpha)$ is a measure preserving system.

Example 1.2. *(Circle-doubling map)* Consider the 1-torus together with the Lebesgue measure $(\mathbb{T}, \mathcal{B}, \lambda)$ and define

$$T_2 : \mathbb{T} \to \mathbb{T}, x \mapsto 2x \mod 1.$$

Let $[a, b) \in \mathscr{S} := \{[a, b) \subseteq \mathbb{T} \mid a \leq b \mod 1\}$ as above. Then

$$T_2^{-1}([a, b)) = [\frac{a}{2}, \frac{b}{2}) \cup [\frac{a}{2} + \frac{1}{2}, \frac{b}{2} + \frac{1}{2})$$

is a disjoint union, so

$$\lambda(T_2^{-1}([a, b))) = b - a = \lambda([a, b)).$$

We conclude that $(\mathbb{T}, \mathcal{B}, \lambda, T_2)$ is a measure preserving system.

Example 1.3. *(Gauss map)* Consider $X = [0,1] \setminus \mathbb{Q}$ together with the induced Borel σ-algebra. Define

$$T : X \to X, x \mapsto \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor.$$

T(x) is the fractional part of $\frac{1}{x}$. Let

$$d\mu(x) := \frac{1}{\log 2} \frac{d\lambda(x)}{1+x},$$

where λ denotes the Lebesgue measure. Take $[a,b] \in \mathcal{B}$, then

$$\mu(T^{-1}([a,b])) = \sum_{n=1}^{\infty} \mu([\frac{1}{b+n}, \frac{1}{a+n}]) = \sum_{n=1}^{\infty} \frac{1}{\log 2} \int_{\frac{1}{b+n}}^{\frac{1}{a+n}} \frac{d\lambda(x)}{1+x} =$$

$$= \frac{1}{\log 2} \Bigg(\underbrace{\sum_{n=1}^{\infty} \log(1 + \frac{1}{a+n})}_{= \lim_{n\to\infty}(-\log(a+1) + \log(1+a+n))} - \sum_{n=1}^{\infty} \log(1 + \frac{1}{b+n}) \Bigg) =$$

$$= \frac{1}{\log 2} \lim_{n\to\infty} \left(\log(\frac{b+1}{a+1}) + \log(\frac{1+a+n}{1+b+n}) \right) = \mu([a,b]).$$

So, $(X, \mathcal{B}, \mu, T)$ is a measure preserving system.

Lemma 1.3. *Let $(X, \mathcal{B})$ be a measurable space and $T : (X, \mathcal{B}) \to (X, \mathcal{B})$ a measurable map. A probability measure $\mu \in \mathcal{P}(X)$ is T-invariant if and only if*

$$\int_X f d\mu = \int_X f \circ T d\mu$$

for all bounded functions $f \in \mathcal{L}^\infty$.

Proof. Suppose the equation

$$\int_X f d\mu = \int_X f \circ T d\mu$$

holds for all $f \in \mathcal{L}^\infty$ and let $B \in \mathcal{B}$ be a measurable set, then

$$\mu(B) = \int_X \mathbb{1}_B d\mu = \int_X \mathbb{1}_B \circ T d\mu = \int_X \mathbb{1}_{T^{-1}(B)} d\mu = \mu_X(T^{-1}(B)).$$

So μ is T-invariant.
Assume μ is T-invariant. By measure theoretic induction it is enough to show

$$\int_X f d\mu = \int_X f \circ T d\mu$$

for functions of the form $f = \mathbb{1}_B$, where $B \in \mathcal{B}$,

$$\int_X \mathbb{1}_B d\mu = \mu(B) = \mu(T^{-1}(B)) = \int_X \mathbb{1}_B \circ T d\mu.$$

□

Example 1.4. *(Bernoulli shift)* Let $X = \{1, 2, \ldots, n\}^{\mathbb{Z}}$ together with the σ-algebra $\mathcal{B}$ generated by finite cylinders of the form

$$\mathcal{C}_{a,F} := \{x \in X \mid x \mid_F = a(F)\},$$

where $F \subseteq \mathbb{Z}$ is a finite set and $a : F \to \{1, \ldots, n\}$ a map.
Define on X the infinite product measure

$$\mu := (\sum_{k=1}^{n} p_k \delta_k)^{\otimes \mathbb{Z}},$$

where $(p_1, \ldots, p_n)$ is a probability vector and the left-shift on X:

$$\theta : X \to X, x \mapsto (x_{j+1})_{j \in \mathbb{Z}}.$$

Since μ is preserved by θ on finite cylinders and these generate $\mathcal{B}$ we may conclude that $(X, \mathcal{B}, \mu, \theta)$ is a measure preserving system.

1.2 Ergodicity

Definition 1.4. *A measure preserving system $(X, \mathcal{B}, \mu, T)$ is called ergodic if the only T-invariant measurable sets $B \in \mathcal{B}$ are μ-trivial.*

Proposition 1.5. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. The following conditions are equivalent:*
i)$(X, \mathcal{B}, \mu, T)$ is an ergodic system.
ii) For $B \in \mathcal{B}$, $\mu(B \Delta T^{-1} B) = 0$ implies $\mu(B) \in \{0, 1\}$.
iii)For $B \in \mathcal{B}$, $\mu(B) > 0$ implies

$$\mu(\bigcup_{n \in \mathbb{N}} T^{-n} B) = 1.$$

iv)For $A, B \in \mathcal{B}$, $\mu(A), \mu(B) > 0$ implies the existence of some $n \in \mathbb{N}$ such that

$$\mu(T^{-n}A \cap B) = 1.$$

v)For $f \in \mathcal{L}^\infty$, $f \circ T = f$ μ-a.e. implies that f is constant μ-a.e.

Proof. *i)* $\Rightarrow$ *ii)* Let $B \in \mathcal{B}$ be a measurable set such that $\mu(B \Delta T^{-1}B) = 0$. Starting with B, we want to construct a T-invariant measurable set $C \in \mathcal{B}$ such that $\mu(B) = \mu(C)$. Define

$$C := \limsup_{n\to\infty} T^{-n}B = \bigcap_{n\in\mathbb{N}_0} \bigcup_{k\geq n} T^{-k}B.$$

By construction C is measurable and T-invariant, so $\mu(C) \in \{0, 1\}$ by ergodicity of the system. The sequence $(\bigcup_{k\geq n} T^{-k}B)_{n\geq 0}$ is decreasing and for each $n \geq 0$

$$B\Delta \bigcup_{k\geq n} T^{-k}B \subseteq \bigcup_{k\geq n} B\Delta T^{-k}B.$$

So

$$\mu(C\Delta B) \leq \lim_{n\to\infty} \mu(\bigcup_{k\geq n} B\Delta T^{-k}B) \leq \lim_{n\to\infty} \sum_{k\geq n} \mu(B\Delta T^{-k}B) \leq$$

$$\leq \lim_{n\to\infty} \sum_{k\geq n} \underbrace{\mu(\bigcup_{l=0}^{k-1} T^{-l}B\Delta T^{-l-1}B)}_{=0} = 0.$$

This implies that $\mu(C) = \mu(B)$.
ii) $\Rightarrow$ *iii)* Let $B \in \mathcal{B}$ such that $\mu(B) > 0$. Define

$$A := \bigcup_{n\in\mathbb{N}} T^{-n}B.$$

Then $T^{-1}A \subseteq A$ and $\mu(A) = \mu(T^{-1}A)$, since T is a measure preserving transformation. Hence $\mu(A\Delta T^{-1}A) = 0$, so $\mu(A) \in \{0, 1\}$. Since $B \subseteq A$ and $\mu(B) > 0$ we conclude $\mu(A) = 1$.
iii) $\Rightarrow$ *iv)* Let $A, B \in \mathcal{B}$, $\mu(A), \mu(B) > 0$. Then

$$\mu(\bigcup_{n\in\mathbb{N}} T^{-n}A) = 1.$$

So

$$0 < \mu(B) = \mu(B \cap \bigcup_{n\in\mathbb{N}} T^{-n}A) \leq \sum_{n\in\mathbb{N}} \mu(B \cap T^{-n}A).$$

This implies the existence of some $n \in \mathbb{N}$ such that $\mu(B \cap T^{-n}A) > 0$.
$iv) \Rightarrow i)$ Let $A \in \mathcal{B}$ be a T-invariant set. Let $n \in \mathbb{N}$, then

$$\mu(\underbrace{T^{-n}(A)}_{=A} \cap A^c) = 0.$$

So $\mu(A) = 0$ or $\mu(A^c) = 0$.
$ii) \Rightarrow v)$ Let $f \in \mathcal{L}^\infty$ so that $f \circ T = f$ μ-a.e. Without loss of generality we can assume f to be real-valued. Let $n \in \mathbb{N}$, then

$$X = \bigcup_{k \in \mathbb{Z}} f^{-1}([\frac{k}{n}, \frac{k+1}{n})) = \bigcup_{k \in \mathbb{Z}} A_n^{k(n)}$$

is a disjoint partition of X and

$$T^{-1}A_n^k(n) \Delta A_n^k(n) \subseteq \{x \in X \mid f \circ T(x) \neq f(x)\},$$

is a null-set for all $k(n) \in \mathbb{Z}$. For any $n \in \mathbb{N}$ we obtain $\mu(A_n^{k(n)}) \in \{0, 1\}$. So, for all $n \in \mathbb{N}$ there exists exactly one $k(n) \in \mathbb{Z}$ such that $\mu(A_n^{k(n)}) = 1$. Define

$$Y := \bigcap_{n \in \mathbb{N}} A_n^{k(n)}.$$

Then, $f\mid_Y$ is constant μ-a.e. on Y. So f is constant μ-a.e.
$v) \Rightarrow ii)$ Let $B \in \mathcal{B}$ such that $\mu(B \Delta T^{-1}B) = 0$. Define $f := \mathbb{1}_B$, which is T-invariant since

$$f \circ T = \mathbb{1}_{T^{-1}B} = \mathbb{1}_B = f.$$

So $f = const.$ μ-a.e. Hence $\mu(B) \in \{0, 1\}$.

$\square$

Example 1.5. *(Circle rotation)* Let $(\mathbb{T}, \mathcal{B}, \lambda, R_\alpha)$ as above.
Suppose first $\alpha = \frac{p}{q} \in \mathbb{Q}$ is rational, then $f := \chi_q$ is not constant and $f \circ R_\alpha = f$. So $(\mathbb{T}, \mathcal{B}, \lambda, R_\alpha)$ is not ergodic if α is rational.
Assume $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is irrational and $f \in L^2(\mathbb{T}, \mu)$ is a T-invariant function. The Fourier series corresponding to f

$$f = \sum_{n \in \mathbb{Z}} a_n \chi_n$$

converges in L^2. Moreover $g \mapsto g \circ R_\alpha$ is an L^2-isometry. Hence

$$f \circ R_\alpha = \sum_{n \in \mathbb{Z}} a_n \chi_n \circ R_\alpha.$$

Uniqueness of the Fourier coefficients implies that for any $n \in \mathbb{N}$

$$a_n = a_n e^{2\pi i n\alpha}.$$

Therefore, $a_n = 0$ if $n \neq 0$ and $f = a_0$ μ-a.e. So $(\mathbb{T}, \mathcal{B}, \lambda, R_\alpha)$ is ergodic if α is irrational.

Example 1.6. *(Bernoulli shift)* Let $(X, \mathcal{B}, \mu, \theta)$ be defined as above and let $B \in \mathcal{B}$ be a θ-invariant set. For any $\epsilon > 0$ there exists a finite cylinder $\mathcal{C}_\epsilon := \mathcal{C}_{a(\epsilon),F(\epsilon)}$ with $\mu(\mathcal{C}_\epsilon \Delta B) < \epsilon$. Without loss of generality we may write $F(\epsilon) = [-N_\epsilon, N_\epsilon] \cap \mathbb{Z}$. Take any $M_\epsilon > 2N_\epsilon$. Then $\theta^{-M_\epsilon}(\mathcal{C}_\epsilon)$ and $\mathcal{C}_\epsilon$ are two independent events. So

$$\mu(\theta^{-M_\epsilon}(\mathcal{C}_\epsilon) \setminus \mathcal{C}_\epsilon) = \underbrace{\mu(\theta^{-M_\epsilon}(\mathcal{C}_\epsilon))}_{=\mu(\mathcal{C}_\epsilon)} \mu(\mathcal{C}_\epsilon^c) = \mu(\mathcal{C}_\epsilon)\mu(\mathcal{C}_\epsilon^c).$$

B is θ-invariant. Hence,

$$\mu(\theta^{-M}\mathcal{C}_\epsilon \Delta B) = \mu(\theta^{-M}\mathcal{C}_\epsilon \Delta \theta^{-M} B) = \mu(\mathcal{C}_\epsilon \Delta B) < \epsilon.$$

Therefore $\mu(\theta^{-M}\mathcal{C}_\epsilon \Delta \mathcal{C}_\epsilon) < 2\epsilon$ and

$$\mu(B)\mu(B^c) \leq (\mu(\mathcal{C}_\epsilon) + \epsilon)(\mu(\mathcal{C}_\epsilon^c) + \epsilon) = \underbrace{\mu(\mathcal{C}_\epsilon)\mu(\mathcal{C}_\epsilon^c)}_{2\epsilon} + \epsilon + \epsilon^2 < 5\epsilon.$$

This holds for each $\epsilon > 0$. Therefore $\mu(B) \in \{0, 1\}$ and we can conclude that $(X, \mathcal{B}, \mu, \theta)$ is ergodic.

1.3 The Mean and Pointwise Ergodic Theorem

Theorem 1.6 (Mean Ergodic Theorem). *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Denote by P_I the orthogonal projection onto the closed subspace $I := \{g \in L^2(X, \mu) \mid g \circ T = g\}$. Then, for any $f \in L^2(X, \mu)$*

$$\frac{1}{n} \sum_{k=0}^{N-1} f \circ T^k \xrightarrow[n\to\infty]{L^2} P_I f.$$

Proof. The basic idea is to see that if $f \in I$, then the theorem holds, so we just need to prove this for elements on the orthogonal complement.
A candidate for a dense subset of the orthogonal complement of I is

$$B := \{g \in L^2(X, \mu) \mid g \circ T - g\}.$$

It suffices to verify that $B^{\perp} = I$. Suppose now that $f = g \circ T - g \in B$, then

$$\|\frac{1}{n}\sum_{k=0}^{n-1}(g \circ T - g) \circ T^k\|_{L^2} = \frac{1}{n}\|g \circ T^n - g\|_{L^2} \xrightarrow[n\to\infty]{} 0.$$

$\square$

Corollary 1.7. *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. Then, for any $f \in L^1_\mu(X)$ there exists a T-invariant function $f' \in L^1_\mu(X)$ so that*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow[n\to\infty]{L^1} f'.$$

In particular f' is given by

$$f' = \mathbb{E}(f \mid \mathcal{F}),$$

where $\mathcal{F} := \{B \in \mathbb{B} \mid \mu(T^{-1}B\Delta B) = 0\}$.

Proof. We want to show that for any $f \in L^\infty(X, \mu) \subseteq L^2(X, \mu)$ the convergence to $P_I f$ is in L^1. By density of $L^\infty(X, \mu)$ in $L^1(X, \mu)$ we can extend the same conclusion to any $f \in L^1(X, \mu)$.
Let $f \in L^\infty(X, \mu)$, then $\|P_I f\|_\infty \leq \|f\|_\infty$. Moreover $\|\cdot\|_{L^1} \leq \|\cdot\|_{L^2}$, so the convergence to $P_I f$ is in L^1.

$\square$

Theorem 1.8 (Birkhoff). *Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. If $f \in L^1_\mu(X)$, then there is a T-invariant function $f^* \in L^1_\mu(X)$ so that for μ-a.e. $x \in X$*

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k(x) \xrightarrow[n\to\infty]{L^1} f^*(x).$$

Moreover,

$$\int_X f^* d\mu = \int_X f d\mu,$$

and if the system is ergodic, then for μ-a.e. $x \in X$

$$f^*(x) = \int_X f d\mu.$$

Proof. Without loss of generality we may assume that f is real valued. For any $x \in X$ define following two functions,

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

and

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

We want to show that the set on which these two functions differ is a null set. For this, note that

$$\{f^* > f_*\} \subseteq \bigcup_{p,q\in\mathbb{Q}} \{f^* \geq p > q \geq f_*\}.$$

It is sufficient to show that for any rationals $p > q$, the set $\{f^* \geq p > q \geq f_*\}$ is a null set. This follows from the maximal ergodic theorem, since

$$\int_{\{f^* \geq p > q \geq f_*\}} f d\mu \geq p\mu(\{f^* \geq p > q \geq f_*\})$$

and

$$\int_{\{f^* \geq p > q \geq f_*\}} f d\mu \leq q\mu(\{f^* \geq p > q \geq f_*\}),$$

implies that $\{f^* \geq p > q \geq f_*\}$ is a null set.
To prove convergence in L^1 we know from the previous corollary, that

$$f_n := \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \xrightarrow[n\to\infty]{L^1} f'.$$

Hence, we can find a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ such that $\lim_{k\to\infty}(f_{n_k})_{k\in\mathbb{N}} = f'$ μ-a.e. So $f' = f^*$ and the convergence

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) \xrightarrow[n\to\infty]{} f^*(x).$$

is in L^1.

□

2 Markov Processes and G-Spaces

2.1 Markov Processes and Convergence Results

Let X be a compact metric space. Denote by $\mathcal{P}(X)$ the space of probability measures defined on the Borel σ-algebra $\mathcal{B}$ of X. By the Tychonoff-Alaoglu theorem $\mathcal{P}(X)$ forms a non-empty compact metric space in the weak*-topology.

Definition 2.1. *Let X be a compact metric space and*

$$\Psi : X \to \mathcal{P}(X),\ x \mapsto \mu_x,$$

a continuous map. The Markov operator acting on $\mathcal{C}(X)$ is the map

$$P : \mathcal{C}(X) \times X \to \mathcal{C}(X),\ (f, x) \mapsto Pf(x) := \int_X f(y) d\mu_x(y).$$

The measures μ_x are called transition probabilities of P. The dual operator is defined on $\mathcal{P}(X)$ by

$$P^* : \mathcal{P}(X) \times \mathcal{B} \to \mathcal{P}(X),\ (\mu, B) \mapsto P^*\mu(B) := \int_X \mu_x(B) d\mu(x).$$

Remark. A Markov operator defines in a natural way a stochastic kernel from $(X, \mathcal{B})$ to itself, namely by setting

$$K : X \times \mathcal{B} \to [0, 1],\ (x, B) \mapsto P\mathbb{1}_B(x) = \mu_x(B).$$

Example 2.1. *(Random walk on the n-cycle)* Let $X = \mathbb{Z}/n\mathbb{Z}$ and define

$$\Psi : X \to \mathcal{P}(M),\ x \mapsto \mu_x := \frac{1}{2}(\delta_{x-1} + \delta_{x+1}).$$

The corresponding Markov operator is given by

$$P : \mathcal{C}(X) \times X \to \mathcal{C}(X),\ (f, x) \mapsto Pf(x) := \frac{1}{2}(f(x + 1) + f(x - 1)).$$

Alternatively if we identify $\mathcal{C}(X) \cong \mathbb{R}^n$ we may represent P in the canonical basis as

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \cdots & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \ldots & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Remark. For any $x \in X$, the transition probability μ_x can be interpreted as the probability of the process to move from x to a measurable set $B \in \mathcal{B}$. The corresponding Markov operator is the description of all these possible moves.

Definition 2.2. *Let X be a compact metric space, P a Markov operator acting on $\mathcal{C}(X)$ and $\mu_0 \in \mathcal{P}(X)$ a probability distribution on X. A stochastic process $(X_n)_{n\in\mathbb{N}_0}$ with state space X is called a Markov process with respect to P and μ_0 if for any bounded measurable function $f \in \mathcal{L}^\infty(X)$ and any $n \in \mathbb{N}_0$ we have for μ_0-a.e. $x \in X$*

$$\mathbb{E}(f(X_{n+1}) \mid X_0, \dots, X_n) = \mathbb{E}(f(X_{n+1}) \mid X_n) = Pf(X_n),$$

and the distribution of X_0 is given by μ_0.

Remark (*Canonical space of Markov processes*). Define on $X^{\mathbb{N}_0}$ the projections

$$X_n : X^{\mathbb{N}_0} \to X,\ (x_k)_{k\geq 0} \mapsto x_n.$$

and the σ-algebra $\mathcal{B}_0 := \sigma(\bigcup_{n\in\mathbb{N}_0} X_n)$ generated by these. The measure space $(X^{\mathbb{N}_0}, \mathcal{B}_0)$ is the canonical space of Markov processes.

Lemma 2.3. *Let X be a compact metric space, P a Markov operator acting on $\mathcal{C}(X)$ and $\mu_0 \in \mathcal{P}(X)$ a probability distribution on X. Then, there exists a unique probability measure $\mathbb{P}^{\mu_0}$ on $(X^{\mathbb{N}_0}, \mathcal{B}_0)$ such that $(X_n)_{n\geq 0}$ is a Markov process with respect to P and μ_0. Moreover, for any $n \in \mathbb{N}_0$ and any bounded measurable function $f \in \mathcal{L}^\infty(X^{n+1})$ on X^{n+1},*

$$\mathbb{E}^{\mathbb{P}^\mu}(f(X_0, \dots, X_n)) = \int_M \cdots \int_M f(x_0, \dots, x_n) d\mu_{x_{n-1}}(x_n) \dots d\mu_{x_0}(x_1) d\mu_0(x_0).$$

Proof. This is an application of the Ionescu-Tulcea theorem taking as stochastic kernels copies of the stochastic kernel induced by P.

□

Theorem 2.4. *Let X be a compact metric space, P a Markov operator acting on $\mathcal{C}(X)$, $\mu_0 \in \mathcal{P}(X)$ a probability measure on X, $(X_n)_{n\in\mathbb{N}_0}$ a Markov process with respect to P and μ_0 . For any $f \in \mathcal{C}(X)$ and $\mathbb{P}^{\mu_0}$-a.s.*

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) \leq \sup\{\int f d\mu \mid \mu \in \mathcal{P}(X) : P^*\mu = \mu\}.$$

Proof. Let $f \in \mathcal{C}(X)$ be a continuous function on X. The theorem follows from the following two observations

Claim 1. Suppose there exists some $g \in \mathcal{C}(X)$ such that $f = Pg - g$. Then, $\mathbb{P}^{\mu_0}$-a.s.

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = 0.$$

Claim 2. For any $\epsilon > 0$ there exist $g, h \in \mathcal{C}(X)$ so that $f = Pg - g + h$ and

$$\|h\|_\infty \le \sup\{\int f d\mu \mid \mu \in \mathcal{P}(X) : P^*\mu = \mu\} + \epsilon.$$

Without loss of generality we may assume that f is positive. For any $\epsilon > 0$ we can find $g, h \in \mathcal{C}(X)$ so that $f = Pg - g + h$ and

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = \limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \big(Pg(X_k) - g(X_k) + h(X_k)\big)$$

$$\le \limsup_{n\to\infty} \|h\|_\infty \le \sup\{\int f d\mu \mid \mu \in \mathcal{P}(M) : P^*\mu = \mu\} + \epsilon.$$

Proof (Claim 1). Without loss of generality we may assume that $f \ge 0$, so $Pg \ge g$. Then, for any $n \in \mathbb{N}_0$

$$\frac{1}{n+1} \sum_{k=0}^{n} f(X_k) \le \sum_{k=0}^{n} \frac{1}{k+1} f(X_k) =: M_n.$$

We want to construct a martingale that converges in L^2 and behaves like $(M_n)_{n\in\mathbb{N}_0}$. Define for $n \in \mathbb{N}_0$ the following process

$$N_n := \sum_{k=1}^{n} \frac{1}{k+1} \big(Pg(X_{k-1}) + g(X_k)\big).$$

$(N_n)_{n\in\mathbb{N}_0}$ is a martingale with respect to the filtration generated by the Markov process since

$$\mathbb{E}(N_{n+1} \mid X_0, \dots X_n) = \mathbb{E}(N_n) + \frac{1}{n+1} \underbrace{\mathbb{E}(Pg(X_n) - g(X_{n+1}) \mid X_0, \dots X_n)}_{= Pg(X_n) - Pg(X_n) = 0}.$$

Moreover

$$\sup_{n\in\mathbb{N}} \|N_n\|_{L2} \le 2\|g\|_\infty \sum_{k=1}^{\infty} \frac{1}{k^2} \le \infty.$$

So by Doob's martingale convergence theorem $(N_n)_{n\in\mathbb{N}}$ converges in L^2 and $\mathbb{P}^{\mu_0}$-a.s. By rearranging the sum to reconstruct M_n we get $\mathbb{P}^{\mu_0}$-a.s. convergence of the process $(M_n)_{n\in\mathbb{N}_0}$. Applying Kronecker's lemma to $(M_n)_{n\in\mathbb{N}_0}$ we conclude, that $\mathbb{P}^{\mu_0}$-a.s.

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = 0.$$

□

Proof (Claim 2). Define the linear subspace $S := \{Pg - g \mid g \in \mathcal{P}(X)\}$ and denote the distance of f to S by

$$\delta := \inf_{g\in\mathcal{C}(\mathcal{X})} \|f - (Pg - g)\|_\infty.$$

By the Hahn-Banach theorem, there exists a continuous linear functional $\Lambda \in \mathcal{C}(X)^*$ so that $\|\Lambda\|_{op} = 1$, $\Lambda(f) = \delta$ and $\Lambda \mid_S = 0$. By the Riesz-representation theorem there is a signed measure $|\mu|$ representing Λ with $\||\mu|\| = 1$, $P^*|\mu| = |\mu|$ and $\int f d|\mu| = \delta$.
Decompose $|\mu|$ into its positive and negative parts $|\mu| = |\mu|^+ - |\mu|^-$. Then, $P^*|\mu|^\pm = |\mu|^\pm$.
Define $\mu := \frac{|\mu|^+}{\||\mu|^+\|} \in \mathcal{P}(X)$. Then $P^*\mu = \mu$ and $\int f d\mu \geq \int f d|\mu| = \delta$.
Let $\epsilon > 0$, choose $g \in \mathcal{C}(X)$ so that $\|f - (Pg - g)\|_\infty \leq \delta + \epsilon$ and let $h := f - (Pg - g) \in \mathcal{C}(X)$. Then $\|h\|_\infty \leq \delta + \epsilon$ and

$$\delta \leq \sup\{\int f d\mu \mid \mu \in \mathcal{P}(X) : P^*\mu = \mu\},$$

since for $\mu := \frac{|\mu|^+}{\||\mu|^+\|} \in \mathcal{P}(X)$ we have $\int f d\mu \geq \delta$. □

□

Corollary 2.5. *Let X be a compact metric space and P a Markov operator acting on $\mathcal{C}(X)$. The space of P^*-invariant probability measures*

$$\mathcal{P}^{P^*}(X) := \{\mu \in \mathcal{P}(X) \mid P^*\mu = \mu\},$$

is not empty.

Corollary 2.6. *Let X be a compact metric space, P a Markov operator acting on $\mathcal{C}(X)$, $\mu_0 \in \mathcal{P}(X)$ a measure on X, $(X_n)_{n\in\mathbb{N}_0}$ a Markov process with respect to P and μ_0 . Let $f \in \mathcal{C}(X)$ so that for all $\mu \in \mathcal{P}^{P^*}(X)$*

$$\int_X f d\mu = \bar{f},$$

where $\bar{f}$ is a constant. Then, $\mathbb{P}^{\mu_0}$-a.s.

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = \bar{f}.$$

Proof. This follows directly from the theorem applied to f and $-f$:

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) \leq \sup\{\int f d\mu \mid \mu \in \mathcal{P}^{P^*}(X)\} = \bar{f},$$

and

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} -f(X_k) \leq \sup\{\int -f d\mu \mid \mu \in \mathcal{P}^{P^*}(X)\} = -\bar{f}.$$

So,

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) \leq \bar{f} \leq \liminf_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k).$$

□

Example 2.2 (*Bernoulli process*). Let $(X_n)_{n\in\mathbb{N}}$ be a Markov process on $X = \{-1, 1\}$ with transition probabilities

$$\mu_1 = \mu_{-1} = p\delta_1 + (1-p)\delta_{-1}.$$

Consider a continuous function $f \in \mathcal{C}(X)$ on X. Note that any P^*-invariant probability measure $\mu \in \mathcal{P}^{P^*}(X)$ on X is equal to μ_1, since for any $B \in 2^X$ we have

$$P^*\mu(B) = \int_X \mu_x(B) d\mu(x) = \mu_1(B) = \mu(B).$$

Hence, $\mathcal{P}^{P^*}(X) = \{\mu_1\}$. Moreover,

$$\int_X f d\mu_1 = pf(1) + (1-p)f(-1).$$

So we can deduce, that

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = pf(1) + (1-p)f(-1),$$

for any $f \in \mathcal{C}(M)$. Consider $f = \mathbb{1}_{\{1\}} \in \mathcal{C}(M)$. Then for any $n \in \mathbb{N}$, $\frac{1}{n+1}\sum_{k=0}^{n} \mathbb{1}_{\{1\}}(X_k)$ can be interpreted as the relative number of successes of the Markov process and converges to

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{1}_{\{1\}}(X_k) = p.$$

2.2 Stationary Measures

Definition 2.7. *Let X be a compact metric space and $\mu \in \mathcal{P}(X)$ a probability measure on X. A sequence $(x_n)_{n\in\mathbb{N}_0} \in X^{\mathbb{N}_0}$ is equidistributed with respect to μ if*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} \delta_{x_k} = \mu$$

converges in the weak-topology.*

Example 2.3 (*Random equidistribution on $\mathbb{T}$*). Let $X = \mathbb{T}$ be the 1-torus and $\lambda \in \mathcal{P}(\mathbb{T})$ the Lebesgue measure on $\mathbb{T}$. Take as transition probability $\mu_x = \lambda$ for any $x \in \mathbb{T}$, i.e. the probability to move from x to any other point in $\mathbb{T}$ is uniformly distributed. Note that $\mathcal{P}^{P^*}(\mathbb{T}) = \{\lambda\}$, since for any $B \in \mathcal{B}$ we have

$$P^*\mu(B) = \int_{\mathbb{T}} \lambda(B) d\mu(x) = \lambda(B) = \mu(B).$$

Hence, for any $f \in \mathcal{C}(\mathbb{T})$ we get

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = \int_{\mathbb{T}} f d\lambda.$$

So the random sequence $\mu_{n+1} = \frac{1}{n+1}(\delta_{X_0} + \cdots + \delta_{X_n})$ equidistributes with respect to the Lebesgue measure λ.
In general, for any $\mu \in \mathcal{P}(\mathbb{T})$ it is possible to construct a random equidistributed sequence by choosing as transition probabilities $\mu_x = \mu$.

Remark. Last example shows that it is simple to construct a random equidistributed sequence if the set of P^*-invariant probability measures $\mathcal{P}^{P^*}(X)$ is uniquely determined.

Proposition 2.8. *Let X be a compact metric space and $T : X \to X$ a measurable map. The space*

$$\mathcal{P}^T(X) := \{\mu \in \mathcal{P}(X) \mid \mu \textit{ is } T\textit{-invariant } \}$$

of T-invariant probability measures $\mu \in \mathcal{P}(X)$ is a weak-compact convex subset of $\mathcal{P}(X)$. The extremal points of $\mathcal{P}^T(X)$ are precisely the ergodic measures in $\mathcal{P}^T(X)$.*

Corollary 2.9. *Let X be a compact metric space, P a Markov operator acting on $\mathcal{C}(X)$ and $\mu \in \mathcal{P}^*(X)$ a P^*-invariant probability measure on X. Then $\mathbb{P}^\mu$ is an ergodic measure for the shift $\theta : X^{\mathbb{N}_0} \to X^{\mathbb{N}_0}$ if and only if μ is an extremal point of $\mathcal{P}^{P^*}(X)$.*

Proof. Suppose μ is not an extremal point of of $\mathcal{P}^{P^*}(X)$. Then, we can find two distinct extremal points $\mu_1, \mu_2 \in \mathcal{P}^{P^*}(X)$ and some $t \in (0,1)$ so that

$$\mu = t\mu_1 + (1-t)\mu_2.$$

By uniqueness of the measures $\mathbb{P}^\mu, \mathbb{P}^\mu_1, \mathbb{P}^\mu_2 \in \mathcal{P}(X^{\mathbb{N}_0})$ we obtain

$$\mathbb{P}^\mu = t\mathbb{P}^\mu_1 + (1-t)\mathbb{P}^\mu_2.$$

Assume by contradiction that $\mathbb{P}^\mu$ is an ergodic measure. Let $B \in \mathcal{B}_0$ be a θ-invariant set. Then, $\mathbb{P}^\mu_1$ and $\mathbb{P}^\mu_2$ are also ergodic measures, since $\mathbb{P}^\mu_1(B), \mathbb{P}^\mu_2(B) \in \{0,1\}$. Let $f \in \mathcal{L}^\infty(X^{\mathbb{N}_0})$ be a bounded measurable function. By Birkhoff's ergodic theorem we conclude, that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ \theta^k(\omega) = \int f d\mathbb{P}^\mu_1 = \int f d\mathbb{P}^\mu_2,$$

for $\mathbb{P}^\mu_1, \mathbb{P}^\mu_2$-a.e. $\omega \in X^{\mathbb{N}_0}$. Therefore we get that $\mathbb{P}^\mu_1 = \mathbb{P}^\mu_2$, which is a contradiction.
Suppose $\mathbb{P}^\mu$ is not ergodic. We want to find distinct measures $\mu_1, \mu_2 \in \mathbb{P}^{P^*}(X)$ and some $t \in (0,1)$ so that

$$\mu = t\mu_1 + (1-t)\mu_2.$$

Since $\mathbb{P}^\mu$ is not ergodic there exists a θ-invariant measurable set $B_0 \in \mathcal{B}_0$ so that $\mathbb{P}^\mu(B_0) =: t \in (0,1)$. Moreover since $\mathbb{P}^\mu$ is θ-invariant we can find some measurable set $B \in \mathcal{B}$ such that $\mu(B) = \mathbb{P}^\mu(B_0)$. Define

$$\mu_1(A) := \frac{1}{t}\mu(A \cap B), \;\; \mu_2(A) := \frac{1}{1-t}\mu(A \cap B^c).$$

Note that $\mu = t\mu_1 + (1-t)\mu_2$ and μ_1, μ_2 are distinct measures. So, μ is not extremal.

□

Corollary 2.10. *Let X be a compact metric space and P a Markov operator acting on $\mathcal{C}(X)$. If the space $\mathcal{P}^{P^*}(X)$ of P^*-invariant probability measures on X contains more than one element, then there are two mutually singular measures $\mu_1, \mu_2 \in \mathcal{P}^{P^*}(X)$.*

Proof. Without loss of generality suppose $\mu_1, \mu_2 \in \mathcal{P}^{P^*}(X)$ are two distinct extremal probability measures. Let $f \in \mathcal{C}(X)$ be a continuous function so that

$$\int_X f d\mu_1 \neq \int_X f d\mu_2.$$

Since $\mathbb{P}_1^{\mu}$ and $\mathbb{P}_2^{\mu}$ are ergodic measures for θ we conclude for the Markov process $(X_n)_{n\in\mathbb{N}_0}$ with respect to P and δ_x using Birhoff's ergodic theorem that

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} f(X_k) = \int_X f d\mu_1$$

for μ_1-a.e. $x \in X$ and

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} f(X_k) = \int_X f d\mu_2$$

for μ_2-a.e. $x \in X$. It follows that the set

$$A := \{x \in X \mid \lim_{n\to\infty} \sum_{k=0}^{n-1} f(X_k) = \int f d\mu_1 \text{ holds } \}$$

is measurable and has $\mu_1(A) = 1$ but $\mu_2(A) = 0$.

□

Corollary 2.11. *Let X be a compact metric space and P a Markov operator acting on $\mathcal{C}(X)$. If $\mathcal{P}^{P^*}(X)$ has a unique P^*-invariant measure $\mu \in \mathcal{P}(X)$, then $\mathbb{P}^{\mu} \in \mathcal{P}(X^{\mathbb{N}_0})$ is an ergodic measure with respect to the shift θ.*

Remark. If $\mathcal{P}^{P^*}(X)$ has more than one invariant measure, then we can find a measurable disjoint partition $B_1, B_2 \in \mathcal{B}$ of X so that if the Markov process starts in B_1 then it will stay in B_1 almost surely.

2.3 Random Walks on G-Spaces

Definition 2.12. *A compact metric space X is called a G-space if there is a locally compact group G acting on X so that*

$$G \times X \to X,\ (g, x) \mapsto g \cdot x$$

is continuous. Given $\mu \in \mathcal{P}(X)$ and $m \in \mathcal{P}(G)$, the convolution probability measure $m * \mu \in \mathcal{P}(X)$ on X is the image of $m \times \mu$ under the action map. For $f \in \mathcal{C}(X)$ we have

$$\int_X f d(m * \mu) = \int_G \int_X f(g \cdot x) d\mu(x) dm(g).$$

Moreover, μ is called m-stationary if $m * \mu = \mu$.

Remark (*Induced random walk*)*.* The continuous map

$$\Psi : X \to \mathcal{P}(X),\ x \mapsto \mu_x = m * \delta_x,$$

induces a Markov operator by setting

$$P : \mathcal{C}(X) \times X \to \mathcal{C}(X),\ (f, x) \mapsto Pf(x) := \int_X f(y) d\mu_x = \int_G f(g \cdot x) dm(g).$$

Moreover, if for any $x \in X$, the associated measure $m * \delta_x$ is m-stationary. Then any probability measure $\mu \in \mathcal{P}(X)$ is P^*-invariant

$$P^* \nu(B) = \int_X m * \delta_x(B) d\nu(x) = \int_X \delta_x(B) d\nu(x) = \nu(B).$$

Remark. A probability measure $m \in \mathcal{P}(G)$ induces a random walk on X, where X_{n+1} arises by applying some $g \in G$ to X_n and g is chosen with respect to the probability measure m.
The random walk may be transient or recurrent. In the case where the random walk is recurrent and the expected return time is finite, we talk about positive recurrence, this is the case if and only if the Markov process has an m-invariant measure $\mu \in \mathcal{P}^m(X)$.

Proposition 2.13. *Let G be a locally compact group and $m \in \mathcal{P}(G)$ a probability measure on G. If X is a G-space, then the space*

$$\mathcal{P}^m(X) := \{\mu \in \mathcal{P}(X) \mid m * \mu = \mu\}$$

of m-stationary measures is not empty.

Proof. Note that $m \in \mathcal{P}(G)$ induces a Markov operator P acting on $\mathcal{C}(X)$, so a measure $\mu \in \mathcal{P}(X)$ is P^*-invariant if and only if it is m-stationary, and we already know that the space $\mathcal{P}^{P^*}(X)$ is not empty. Nevertheless the following proof is more constructive. Let $\mu \in \mathcal{P}(X)$ be any probability measure on X. Define the averages

$$\mu_n := \frac{1}{n}\sum_{k=0}^{n-1} m^n * \mu,$$

where $\mu^n * m := m * (m * (\ldots (m * \mu) \ldots))$. $(\mu_n)_{n\in\mathbb{N}} \in \mathcal{P}(X)^{\mathbb{N}}$ is a sequence of probability measures on X and $\mathcal{P}(X)$ is a weak*-compact space. Therefore, we can find a subsequence $(\mu_{n_l})_{l\in\mathbb{N}} \subseteq (\mu_n)_{n\in\mathbb{N}}$ so that $\mu_{n_l} \xrightarrow[l\to\infty]{} \mu^* \in \mathcal{P}(X)$. We want to show that the limit μ^* is m-stationary. Let $f \in \mathcal{C}(X)$ be a continuous function. Then, for any $l \in \mathbb{N}$ we can approximate the difference of $m * \mu_{n_l}(f)$ from $\mu_{n_l}(f)$ by

$$|\int_X f d(m * \mu_{n_l}) - \int_X f d\mu_{n_l}| \leq \frac{2}{n_l}\|f\|_\infty.$$

The triangle inequality yields the desired result.

□

Example 2.4 (*Random walk on* $\mathbb{Z}$). Let $p \in [0, \frac{1}{2})$ and consider the probability distribution on $\mathbb{Z}$ given by

$$m := p\delta_1 + (1-p)\delta_{-1}.$$

The induced Markov operator on $\mathbb{Z}$ is then

$$P : \mathcal{C}(\mathbb{Z}) \times \mathbb{Z} \to \mathcal{C}(\mathbb{Z}),\ (f, x) \mapsto pf(x+1) + (1-p)f(x+1).$$

We want to describe the space $\mathcal{P}^m(\mathbb{Z})$ of m-stationary probability measures. Let $\mu \in \mathcal{P}^m(\mathbb{Z})$, for any $x \in \mathbb{Z}$ the measure μ must satisfy

$$\mu(x) = m * \mu(x) = \int_{\mathbb{Z}}\int_{\mathbb{Z}} \mathbb{1}_x(y+z) d\mu(y) dm(z) =$$

$$= \int_{\mathbb{Z}} \Big(p\mathbb{1}_x(y+1) + (1-p)\mathbb{1}_x(y-1)\Big) d\mu(y) = p\mu(x-1) + (1-p)\mu(x+1).$$

Solutions of this linear recurrence are of the form

$$\mu(x) = \alpha + \beta\left(\frac{p}{1-p}\right)^x.$$

So $\mathcal{P}^m(\mathbb{Z}) = \varnothing$. This shows that X has to be compact.

Example 2.5 (*Lazy random walk on the hypercube*). Consider as state space the hypercube $X = \{0,1\}^d$. Two vertices $x, y \in X$ are said to be neighbors, $x \sim y$, if and only if they differ in exactly one coordinate or equivalently if the (euclidean) distance is one Each vertex has exactly d neighbors. The group $G = (\mathbb{Z}/2\mathbb{Z})^d$ acts on X by addition. Denote by $e_k = (0, \dots, 1, \dots, 0) \in G$ the basis elements and introduce the following probability measure on G,

$$m = \frac{1}{2d}(\delta_{e_1} + \cdots + \delta_{e_d}) + \frac{1}{2}\delta_0.$$

The corresponding Markov operator is the map

$$P : \mathcal{C}(X) \times X \to \mathcal{C}(X),\ (f, x) \mapsto \frac{1}{2d}\sum_{k=1}^{d} f(e_k + x) + \frac{1}{2}f(x).$$

The Markov process moves at each step to one neighbor or stays put with with equal probability $\frac{1}{2}$. This can be seen by looking at the transition probabilities

$$\mu_x(y) = P\mathbb{1}_{\{y\}}(x) = \begin{cases} \frac{1}{2d} & ,\text{if } x \sim y \\ \frac{1}{2} & ,\text{if } x = y \\ 0 & , else \end{cases}.$$

Note that the uniform distribution on X is P^*-invariant (resp. m-stationary).

3 Examples of G-Spaces

3.1 Law of Large Numbers for Matrix Products

An application of the last section is the following case: $G = GL_n(\mathbb{R})$ is a locally compact group. Denote by $\mathbb{P}^{n-1}(\mathbb{R})$ the $(n-1)$-dimensional projective space. Suppose $m \in \mathcal{P}(G)$ is a probability measure on G with compact support $supp(m) \subset\subset G$. $\mathbb{P}^{n-1}(\mathbb{R})$ is a G-space with respect to G and the group acts continuously on $\mathbb{P}^{n-1}(\mathbb{R})$ by multiplication

$$G \times \mathbb{P}^{n-1}(\mathbb{R}) \to \mathbb{P}^{n-1}(\mathbb{R}),\ (g, [x]) \mapsto [gx],$$

where $[x] \in \mathbb{P}^{n-1}(\mathbb{R})$ denotes the equivalence class of $x \in \mathbb{R}^n \setminus \{0\}$. Introduce the Markov operator induced on $\mathcal{C}(\mathbb{P}^{n-1}(\mathbb{R})$ by G,

$$P_G : \mathcal{C}(\mathbb{P}^{n-1}(\mathbb{R})) \times \mathbb{P}^{n-1}(\mathbb{R}) \to \mathcal{C}(\mathbb{P}^{n-1}(\mathbb{R})),\ (f, [x]) \mapsto \int_G f(g \cdot [x]) dm(g).$$

Fix some $[x] \in \mathbb{P}^{n-1}(\mathbb{R})$, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of G-valued independent indentically m-distributed random variables and consider the random walk on $\mathbb{P}^{n-1}(\mathbb{R})$ defined by

$$S_n = \begin{cases} X_n \dots X_1 \cdot [x] & \text{if } n \geq 1 \\ [x] & \text{if } n = 0 \end{cases}.$$

Note that for any $f \in \mathcal{C}(\mathbb{P}^{n-1}(\mathbb{R}))$ we have

$$\mathbb{E}(f(S_n) \mid S_0, \dots, S_{n-1}) = \mathbb{E}(f(X_n \dots X_1 \cdot [x]) \mid X_1, \dots, X_{n-1}) =$$

$$= \int_G f\big(g \cdot (X_{n-1} \dots X_1 \cdot [x])\big) dm(g) = Pf(S_{n-1}).$$

So $(S_n)_{n\in\mathbb{N}}$ defines a Markov process on $\mathbb{P}^{n-1}(\mathbb{R})$ with respect to P_G. The aim of this example is to construct a non-commutative analogue of the Law of Large Numbers. In order to do this, we will analyze the following expression

$$\lim_{n\to\infty} \frac{1}{n} \log \|S_n\| = \lim_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1 \cdot [x]\|.$$

Theorem 3.1. *Let $m \in \mathcal{P}(G)$ be a probability measure with compact support and $(X_n)_{n\in\mathbb{N}}$ a G-valued independent identically m-distributed sequence of random variables. Then, for any $x \in \mathbb{R}^n \setminus \{0\}$ with probability one*

$$\limsup_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1 \cdot x\| \leq$$

$$\leq \sup\{\int\int \log \frac{\|g[y]\|}{\|[y]\|} dm(g)d\nu([y]) \mid \mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))\}.$$

Moreover, if there is a constant β so that for any m-invariant measure $\mu \in \mathcal{P}(\mathbb{P}^{n-1}(\mathbb{R}))$,

$$\int\int \log \frac{\|g[y]\|}{\|[y]\|} dm(g)d\nu([y]) = \beta.$$

Then, with probability one

$$\lim_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1\| = \beta.$$

Proof. Define $X := supp(m) \times \mathbb{P}^{n-1}(\mathbb{R})$

$$P : \mathcal{C}(X) \times X \to \mathcal{C}(M),\ (f, (g, [x])) \mapsto \int_G f(h, h \cdot [x])dm(h).$$

Note that if a function $f \in \mathcal{C}(X)$ is of the form $f(g, [x]) = f([x])$, then P coincides with P_G.
For any $g \in G$ and $[x] \in \mathbb{P}^{n-1}(\mathbb{R})$ the expression $\frac{\|x\|}{\|g^{-1}x\|}$ is independent of the choice of $x \in [x]$. Therefore, the map

$$f : M \to \mathbb{R},\ (g, [x]) \mapsto \log\left(\frac{\|x\|}{\|g^{-1}x\|}\right)$$

defines a well-defined and continuous function on M. The process

$$Y_n = \begin{cases} (X_n, S_n) & \text{, if } n \geq 1 \\ (id, [x]) & \text{, if } n = 0 \end{cases},$$

is as above a Markov process with respect to P. Then, for any $n \in \mathbb{N}$

$$\frac{1}{n+1}\sum_{k=0}^{n} \log(Y_k) = \frac{1}{n+1}\sum_{k=1}^{n} \log\left(\frac{\|X_k \dots X_1 \cdot x\|}{\|X_{k-1} \dots X_1 \cdot x\|}\right)$$

$$= \frac{1}{n+1}\log(\|X_n \dots X_1 \cdot x\|) - \frac{1}{n+1}\log(\|x\|).$$

So,

$$\limsup_{n\to\infty} \frac{1}{n} \log(\|X_n \dots X_1 \cdot x\|) \leq \sup\{\int \log \frac{\|y\|}{\|g^{-1}y\|} d\nu(g,[y]) \mid \nu \in \mathcal{P}^{P^*}(M)\}.$$

We want to show that there is a bijection between P^*-invariant measures and m-invariant measures.

Suppose $\nu \in \mathcal{P}^{P^*}(X)$ is a P^*-invariant measure on X. Denote by $\mu \in \mathcal{P}(\mathbb{P}^{n-1}(\mathbb{R}))$ the projection of ν on $\mathbb{P}^{n-1}(\mathbb{R})$. If $f \in \mathcal{C}(X)$ is a continuous function of the form $f(g,[x]) = f([x])$, then

$$\int f([x]) d\mu([x]) = \int f([x]) d\nu(g,[x]) = \int f([x]) d(P^*\nu)(g,[x]) =$$

$$= \int Pf([x]) d\nu(g,[x]) = \int\int f(g\cdot[x]) dm(g) d\mu([x]).$$

So μ is m-invariant. Conversely suppose that $\mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))$ is an m-invariant measure. Define a measure $\nu \in \mathcal{P}(X)$ on X by

$$\int f(g,[x]) d\nu(g,[x]) := \int\int f(g, g\cdot[x]) dm(g) d\mu([x]),$$

for $f \in \mathcal{C}(X)$. Then $\nu \in \mathcal{P}^{P^*}(X)$ is a P^*-invariant measure, since

$$\int f d(P^*\nu) = \int Pf d\nu = \int\int Pf(g, g\cdot[x]) dm(g) d\mu([x]) =$$

$$= \int\int f(h, h\cdot[x]) dm(h) d\mu([x]) = \int f d\nu.$$

Let $\nu \in \mathcal{P}^{P^*}(X)$ be a P^*-invariant measure on X and $\mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))$ the corresponding m-invariant measure on $\mathbb{P}^{n-1}(\mathbb{R})$. Then

$$\int \log \frac{\|y\|}{\|g^{-1}y\|} d\nu(g,[y]) = \int\int \log \frac{\|gy\|}{\|y\|} dm(g) d\mu([y]).$$

Hence, we can rewrite the above inequality as

$$\limsup_{n\to\infty} \frac{1}{n} \log(\|X_n \dots X_1 \cdot x\|) \leq$$

$$\leq \sup\{\int\int \log \frac{\|gy\|}{\|y\|} dm(g) d\mu([y]) \mid \mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))\}.$$

□

Corollary 3.2. *Let $m \in \mathcal{P}(G)$ be a probability measure with compact support and $(X_n)_{n\in\mathbb{N}}$ a G-valued independent identically m-distributed sequence of random variables. With probability one*

$$\lim_{n\to\infty} \frac{1}{n} \log \|X_n \cdots X_1\| =$$

$$= \sup\{\int\int \log\left(\frac{\|gx\|}{\|x\|}\right) dm(g)d\mu([x]) \mid \mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))\}.$$

Proof. Note that the previous theorem implies directly that

$$\limsup_{n\to\infty} \frac{1}{n} \log(\|X_n \dots X_1\|) \leq$$

$$\leq \sup\{\int\int \log \frac{\|gx\|}{\|x\|} dm(g)d\mu([x]) \mid \mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))\}$$

holds with probability one, independent of the choice of norm. We want to show the inverse inequality for the $\liminf$.
Let $\mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))$ be an m-invariant probability measure and $[U_0]$ a μ-distributed random variable independent of the sequence $(X_n)_{n\in\mathbb{N}}$. Consider the random walk on $\mathbb{P}^{n-1}(\mathbb{R})$ defined by

$$S_n = \begin{cases} X_n \dots X_1 \cdot [U_0] & \text{, if } n \geq 1 \\ [U_0] & \text{, if } n = 0 \end{cases},$$

and the process

$$Y_n = \begin{cases} (X_n, S_n) & \text{, if } n \geq 1 \\ (id, [U_0]) & \text{, if } n = 0 \end{cases},$$

on $M := supp(m) \times \mathbb{P}^{n-1}(\mathbb{R})$. The latter process is a Markov process with respect to the Markov operator

$$P : \mathcal{C}(M) \times M \to \mathcal{C}(M),\ (f, (g, [x])) \mapsto \int_G f(h, h \cdot [x])dm(h).$$

The process $(S_n)_{n\geq 0}$ is stationary: For $n = 0$ this is clear, for $n \geq 1$ this follows from the fact that X_n and U_{n-1} are independent. So the process $(Y_n)_{n\geq 0}$ is stationary. Hence

$$\lim_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1 U_0\| = f^*,$$

where $f^* \in L^1(X)$ and

$$\mathbb{E}(f^*) = \int \int \log\left(\frac{\|gx\|}{\|x\|}\right) dm(g) d\mu([x]).$$

Note that

$$\liminf_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1 U_0\|$$

is a random variable, which is measurable with respect to the tail σ-algebra $\mathcal{F}_\infty := \bigcap_{n\geq 0} \sigma(X_n, X_{n+1}, \dots)$. So by Kolmogorov's zero-one law it is a.e. constant. Moreover, there must be a measurable set B of positive probability so that

$$\liminf_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1 U_0\| \geq \int \int \log\left(\frac{\|gx\|}{\|x\|}\right) dm(g) d\mu([x]).$$

Hence B is a set of full measure. Since μ was arbitrary, we conclude that

$$\liminf_{n\to\infty} \frac{1}{n} \log \|X_n \dots X_1 U_0\| \geq$$

$$\sup\{\int \int \log\left(\frac{\|gx\|}{\|x\|}\right) dm(g) d\mu(x) \mid \mu \in \mathcal{P}^m(\mathbb{P}^{n-1}(\mathbb{R}))\}$$

holds with probability one. The previous theorem gives the remaining inequality.

□

3.2 Rotations on the 2-Sphere

Let $\mathbb{S}^2 := \{x \in \mathbb{R}^3 \mid \|x\| = 1\}$ be the 2-sphere embedded in $\mathbb{R}^3$ together with its Borel σ-algebra. Parametrize $\mathbb{S}^2$ in spherical coordinates,

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \cos(\varphi)\sin(\theta) \\ \sin(\varphi)\sin(\theta) \\ \cos(\theta) \end{pmatrix},$$

where $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$. The Haar measure on $\mathbb{S}^2$ is the usual Lebesgue measure given by

$$d\mu = \frac{1}{4\pi} \sin(\theta) d\theta d\varphi.$$

Let $G = SO(3) = \{g \in O(3) \mid det(g) = 1\}$ be the special orthogonal group. G acts continuously on $\mathbb{S}^2$ by left-multiplication,

$$G \times \mathbb{S}^2 \to \mathbb{S}^2,\ (g, x) \mapsto g \cdot x.$$

$SO(3)$ is a compact subgroup, each element $g \in SO(3)$ can be parametrized (up to measure zero) by the Euler angles as

$$g = g_\varphi h_\theta g_\vartheta$$

where $\theta \in [0, \pi]$, $\varphi \in [0, 2\pi]$, $\vartheta \in [0, 2\pi]$ and

$$g_\varphi = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad h_\theta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

The normalized Haar measure on $SO(3)$ is then given by

$$\int_{SO(3)} f(g)dm(g) = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} f(g_\varphi h_\theta g_\vartheta) \sin(\theta) d\varphi d\theta d\vartheta,$$

for $f \in \mathcal{C}(SO(3))$. The main observation for this example is that $SO(3)$ acts as measure preserving transformations on $\mathbb{S}^2$.

Lemma 3.3. *$SO(3)$ is a family of measure preserving transformations of the sphere $\mathbb{S}^2$.*

Proof. Let $f \in \mathcal{L}^\infty$ be a bounded function on the sphere and $g \in SO(3)$ an arbitrary element. Then

$$\int_{\mathbb{S}^2} f \circ g \, d\mu = \int_{\mathbb{S}^2} f d\mu,$$

since $\det g = 1$. Since $g \in SO(3)$ was arbitrary we conclude the claim. □

We want to construct a Markov process with a unique P^*-invariant probability measure. The idea is to use the left-invariance property of the Haar measure. Take any $g \in SO(3) \setminus \{id\}$ and define

$$\Psi_1 : \mathbb{S}^2 \to \mathcal{P}(\mathbb{S}^2), \ x \mapsto \delta_{g \cdot x}.$$

The corresponding Markov operator acting on $\mathcal{C}(\mathbb{S}^2)$ is the map

$$P_1 : \mathcal{C}(\mathbb{S}^2) \times \mathbb{S}^2 \to \mathcal{C}(\mathbb{S}^2), \ (f, x) \mapsto \int_{\mathbb{S}^2} f(y) d\mu_x(y) = f(g \cdot x).$$

Lemma 3.4. *Let $(X_n)_{n \geq 0}$ be a Markov process with respect to P_1 and some initial distribution $\mu_0 \in \mathcal{P}(\mathbb{S}^2)$. Then*

$$X_n \in g^n \cdot supp(\mu_0)$$

for $\mathbb{P}^{\mu_0}$-a.e. $x \in \mathbb{S}^2$.

Proof. For X_0 it is clear by definition of the initial distribution. Consider first $n = 1$,

$$\mathbb{P}\big(X_1 \in g \cdot supp(\mu_0)\big) = \mathbb{P}\big(\mathbb{P}\big(X_1 \in g \cdot supp(\mu_0) \mid X_0\big)\big) =$$

$$= \mathbb{E}\big(P_1 \mathbb{1}_{\{X_0 \in g \cdot supp(\mu_0)\}}(X_0)\big) = \mathbb{P}\big(g \cdot X_0 \in g \cdot supp(\mu_0)\big) = 1.$$

Let $n \in \mathbb{N}$ be arbitrary, then by conditioning on the first $n-1$ steps and using the Markov property we get

$$\mathbb{P}\big(X_n \in g^n \cdot supp(\mu_0)\big) = \mathbb{P}\big(\mathbb{P}\big(X_n \in g^n \cdot supp(\mu_0) \mid X_0, \ldots, X_{n-1}\big)\big) =$$

$$= \mathbb{E}\big(P_1 \mathbb{1}_{\{X_{n-1} \in g^n \cdot supp(\mu_0)\}}(X_0)\big) = \mathbb{P}\big(g^{n-1} \cdot X_{n-1} \in g^{n-1} \cdot supp(\mu_0)\big) = 1.$$

□

Corollary 3.5. *Suppose $\mu_0 = \delta_{x_0}$ is the initial distribution for the Markov chain with respect to P_1 for some $x_0 \in \mathbb{S}^2$. Then, for any $n \in \mathbb{N}_0$ we have $X_n = g^n \cdot x_0$ for $\mathbb{P}^{\mu_0}$-a.e. $x \in \mathbb{S}^2$.*
Moreover $\{g^n \cdot x_0\}_{n \geq 0} \subseteq \Gamma$, where $\Gamma \subset \mathbb{S}^2$ is a circle radial to the axis of rotation and the sequence equidistributes with respect to the line element along Γ induced by μ if and only if the rotation angle along the axis of rotation is irrational.

Proof. Without loss of generality we may assume that $g \in SO(3)$ is a rotation of angle $\alpha \in [0, 2\pi]$ about the z-axis and that $\{g^n \cdot x_0\}_{n \geq 0}$ has more than 3 three elements.
Let $\Gamma = E \cap \mathbb{S}^2$, where E is the unique plane through the orbit of x_0. In particular Γ is a circle radial to the axis of rotation.
If $\alpha \in \mathbb{Q}$ is rational then $\{g^n \cdot x_0\}_{n \geq 0}$ is finite. If $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ is irrational then the map $g\mid_\Gamma : \Gamma \to \Gamma$ is an ergodic transformation of Γ with respect to the line element induced by μ *(Circle rotation)*. Hence, the sequence $\{g^n \cdot x_0\}_{n \geq 0}$ equidistributes on Γ if and only if α is irrational.

□

Lemma 3.6. *$\mu \in \mathcal{P}(\mathbb{S}^2)$ is the unique non-atomic P_1^*-invariant probability measure.*

Proof. Suppose $\nu \in \mathcal{P}(\mathbb{S}^2)$ is a P_1^*-invariant probability measure on $\mathbb{S}^2$. For any $B \in \mathcal{B}$ we get

$$P_1^* \nu(B) = \int_{\mathbb{S}^2} \mu_x(B) d\nu(x) = \int_{\mathbb{S}^2} \delta_{g \cdot x}(B) d\nu(x) = \nu(g \cdot B) = \nu(B).$$

So ν is invariant under g. The only rotation-invariant non-atomic probability measure on the sphere is μ. Hence $\nu = \mu$.

□

Remark. If $x, y \in \mathbb{S}^2$ are the unique points on the sphere fixed by g, then $\delta_x, \delta_y \in \mathcal{P}(\mathbb{S}^2)$ are the only atomic P_1^*-invariant probability measures.

Corollary 3.7. *Let $(X_n)_{n\geq 0}$ be a Markov process with respect to P_1 and $\mu_0 \in \mathcal{P}(\mathbb{S}^2)$. For any continuous function $f \in \mathcal{C}(\mathbb{S}^2)$, $\mathbb{P}^{\mu_0}$-a.s.*

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) \leq \sup\{\int_{\mathbb{S}^2} f d\nu \mid \nu \in \{\mu, \delta_x, \delta_y\}\}.$$

Lemma 3.8. *For any $f \in \mathcal{C}(\mathbb{S}^2)$, $P_1 f = f$ μ-a.e. implies that f is constant μ-a.e.*

Proof. Suppose $f \in \mathcal{C}(\mathbb{S}^2)$ and $P_1 f = f$ μ-a.e. For any $x \in \mathbb{S}^2$ we must have

$$P_1 f(x) = f(g \cdot x) = f(x).$$

This holds if and only if f is constant, since $g \neq id$.

□

The second Markov process that we want to describe is the Markov process induced by the natural map

$$\Psi_2 : \mathbb{S}^2 \to \mathcal{P}(\mathbb{S}^2),\ x \mapsto \delta_x * m.$$

The corresponding Markov operator acting on $\mathcal{C}(\mathbb{S}^2)$ is the map

$$P_2 : \mathcal{C}(\mathbb{S}^2) \times \mathbb{S}^2 \to \mathcal{C}(\mathbb{S}^2),\ (f, x) \mapsto \int_{SO(3)} f(g \cdot x) dm(g).$$

The dynamics of this Markov process is the following: If at time n the Markov chain is at the point $X_n = x_n \in \mathbb{S}^2$, then it moves to some $x \in B \subset \mathcal{B}$ in the next step with probability $P\mathbb{1}_B(x)$. Informally, the Markov process moves from x_n to x with probabililty $dm(g)$ if there is some $g \in SO(3)$ so that $x = g \cdot x_n$.

Lemma 3.9. *The space of m-invariant probability measures on $\mathbb{S}^2$ contains only $\mu \in \mathcal{P}(\mathbb{S}^2)$.*

Proof. Let $f \in \mathcal{C}(\mathbb{S}^2)$ be a continuous function and $\nu \in \mathcal{P}^m(\mathbb{S}^2)$ an m-invariant probability measure. Let $g \in SO(3)$, then

$$\int_{\mathbb{S}^2} f(g\cdot x) d\nu(x) = \int_{\mathbb{S}^2} f(g\cdot x) d(m*\nu)(x) = \int_{SO(3)} \int_{\mathbb{S}^2} f(h\cdot g\cdot x) d\nu(x) dm(h) =$$

$$= \int_{\mathbb{S}^2} \int_{SO(3)} f(h \cdot x) dm(h) d\nu(x) = \int_{\mathbb{S}^2} f(x) d(m * \nu)(x) = \int_{\mathbb{S}^2} f(x) d\nu(x).$$

So any m-invariant probability measure is $SO(3)$-invariant. Since $\mathcal{P}^m(\mathbb{S}^2)$ is not empty and μ is the only $SO(3)$-invariant probability measure on the sphere, we may conclude that $\mathcal{P}^m(\mathbb{S}^2) = \{\mu\}$.

□

Corollary 3.10. *Let $\mu_0 \in \mathbb{S}^2$ be an initial probability distribution for the Markov process $(X_n)_{n\geq 0}$ with respect to P_2 and $f \in \mathcal{C}(\mathbb{S}^2)$ a continuous function on the sphere, then $\mathbb{P}^{\mu_0}$-a.s.*

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f(X_k) = \int_{\mathbb{S}^2} f d\mu.$$

Lemma 3.11. *For any $B \in \mathcal{B}$, $\mu(g^{-1}B\Delta B) = 0$ for m-a.e. $g \in SO(3)$ implies $\mu(B) \in \{0, 1\}$.*

Proof. Let $B \in \mathcal{B}$ be a measurable set so that $gB = B$ for m-a.e. $g \in SO(3)$ and $\mu(B) \in (0, 1)$.
Since $B \neq \mathbb{S}^2$, we can find $(\varphi_1, \varphi_2) \times (\theta_1, \theta_2) \times (\vartheta_1, \vartheta_2) \in [0, 2\pi] \times [0, \pi] \times [0, 2\pi]$ so that for any $g_{\varphi,\theta,\vartheta} \in SO(3)$, where $(\varphi, \theta, \vartheta) \in (\varphi_1, \varphi_2) \times (\theta_1, \theta_2) \times (\vartheta_1, \vartheta_2)$ we have

$$g_{\varphi,\theta,\vartheta} \cap B \neq B,$$

and

$$m(\{g \in SO(3) \mid g = g_{\varphi,\theta,\vartheta}, (\varphi, \theta, \vartheta) \in (\varphi_1, \varphi_2) \times (\theta_1, \theta_2) \times (\vartheta_1, \vartheta_2)\}) > 0$$

which is a contradiction.

□

3.3 Stationary Measures on the Torus

Let $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ be the d-torus and μ the Lebesgue measure on $\mathbb{T}^d$. Denote by $SL_d(\mathbb{Z}) := \{g \in GL_d(\mathbb{Z}) \mid det\, g = 1\}$ the special linear group of integer matrices.

Proposition 3.12. *Let $g \in SL_d(\mathbb{Z})$. The action of g on $\mathbb{Z}^d$ is ergodic if and only if no eigenvalue of g is a root of unity.*

Proof. The proposition follows directly from the following claim,

Claim. g is ergodic with respect to μ if and only if $e^{2\pi i\langle n, A^l \cdot x\rangle} = e^{2\pi i\langle n,x\rangle}$ μ-a.e. for some $l > 0$ implies $n = 0$.

Proof (Claim). Suppose that g is ergodic. Let $n \in \mathbb{Z}^d$ and $l > 0$ so that μ-a.e.

$$e^{2\pi i \langle n, g^l \cdot x\rangle} = e^{2\pi i \langle n, x\rangle}.$$

Without loss of generality we may assume that $l > 0$ is the smallest exponent so that the above equality holds. Define

$$f(x) := \sum_{k=0}^{l-1} e^{2\pi i \langle n, g^k \cdot x\rangle}.$$

Note that $f \circ g = f$ μ-a.e. By assumption g is an ergodic transformation, so f is constant μ-a.e. Hence, $n = 0$.
Conversely suppose that $e^{2\pi i \langle n, g^l \cdot x\rangle} = e^{2\pi i \langle n, x\rangle}$ μ-a.e. for some $l > 0$ implies $n = 0$. Let $f \in L^2(\mathbb{T}^d, \mu)$ and suppose that $f \circ g = f$ μ-a.e. Since $f \circ g^l = f$ μ-a.e. for all $l > 0$, we get from the Fourier series for both functions following expression,

$$\sum_{k \in \mathbb{Z}^d} a_k e^{2\pi i \langle k, g^l \cdot x\rangle} = \sum_{k \in \mathbb{Z}^d} a_k e^{2\pi i \langle k, x\rangle}.$$

By the uniqueness of the Fourier coefficients we obtain for any $k \in \mathbb{Z}^d$,

$$a_k = a_{gk} = \cdots = a_{g^l k} = \ldots.$$

Note that if $a_k \neq 0$, then there must be some $l > 0$ so that $k = g^l k$. So $e^{2\pi i \langle k, g^l \cdot x\rangle} = e^{2\pi i \langle k, x\rangle}$ μ-a.e., hence $k = 0$. Therefore $f = a_0$ μ-a.e.

$\square$

Suppose g were no ergodic, then there is some $n \in \mathbb{Z}^d \setminus \{0\}$ and some $l > 0$ such that $e^{2\pi i \langle n, g^l \cdot x\rangle} = e^{2\pi i \langle n, x\rangle}$ μ-a.e. So $(g^{tr})^l n = n$. Hence, g^l has an eigenvalue 1 and g an l-th root of unity as eigenvalue.

Suppose g has an l-th root of unity as eigenvalue. Then, g^l has 1 as an eigenvalue. So $n(g^l - id) = 0$ for some $n \in \mathbb{R}^d \setminus \{0\}$. Since $g \in \mathbb{Z}^d$ we can take $n \in \mathbb{Z}^d$. Moreover $e^{2\pi i \langle n, g^l \cdot x\rangle} = e^{2\pi i \langle n, x\rangle}$. So g is not ergodic.

$\square$

The first Markov process we will analyze on the torus is the following trivial Markov process: Let $g \in SL_d(\mathbb{Z})$ so that no eigenvalue of g is a root of unity. Consider the following probability distribution on $G = \langle g \rangle < SL_d(\mathbb{Z})$, $m_1 := \delta_g \in \mathcal{P}(G)$. The corresponding Markov operator acting on $\mathcal{C}(\mathbb{T}^d)$ is

$$P_1 : \mathcal{C}(\mathbb{T}^d) \times \mathbb{T}^d \to \mathcal{C}(\mathbb{T}^d),\ (f, x) \to f(g \cdot x).$$

Remark. The space $\mathcal{P}^{m_1}(\mathbb{T}^d)$ contains the *Lebesgue* measure μ, and Dirac measures $(\delta_x)_{x\in\{0,\sigma(g)\}}$, where $\sigma(g)$ denotes the spectrum of g.
Lebesgue measure is invariant with respect to the action of g, since $det\, g = 1$, and g fixes any $x \in \{0, \sigma(g)\}$.
This is the usual behavior of invariant distributions, convex combination of the Haar measure and atomic measures. Following proposition explains one case.

Proposition 3.13. *Let $G<SL_d(\mathbb{Z})$ be a finite index subgroup and $\mu \in \mathcal{P}^G(\mathbb{T}^d)$ a G-invariant probability measure on $\mathbb{T}^d$. Then, μ is a convex combination of the Haar measure and atomic measures on finite orbits.*

Proof. Suppose $\nu \in \mathcal{P}^G(\mathbb{T}^d)$ is not the Haar measure. Then there exists some $n \in \mathbb{Z}^d \setminus \{0\}$ so that $|\hat{\nu}(n)| > 0$. For any $g \in G$,

$$\hat{\nu}(n) = \int_{\mathbb{T}^d} e^{2\pi i\langle n, g\cdot x\rangle} d\nu(x) = \int_{\mathbb{T}^d} e^{2\pi i\langle g^{tr}\cdot n, x\rangle} d\nu(x) = \hat{\nu}(g^{tr}\cdot n),$$

since ν is G-invariant. By Wiener's lemma,

$$\sum_{x\in\mathbb{T}^d} \nu(\{x\})^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{|B_n|} \sum_{b\in B_n} \hat{\nu}(b)^2.$$

where the sum is taken over all atoms and $B_n = \{b \in \mathbb{Z}^d \mid \max |b_k| \leq n\}$ It follows that ν has atoms, since it is not the Haar measure. Moreover, any orbit $G^{tr} n \subset \mathbb{Z}^d \setminus \{0\}$ has positive measure. So the atoms of ν must lie on finite orbits.

□

Example 3.1 (*Finite index subgroups*). For any $n \in \mathbb{N}$ define the reduction map

$$\Phi : SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/n\mathbb{Z}), \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mod n.$$

Φ is an epimorphism and $SL_2(\mathbb{Z})/ker(\Phi)$ is a subgroup of index n. This can be seen by considering the generators

$$A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},\ B := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

of $SL_2(\mathbb{Z})$ and observing that $SL_2(\mathbb{Z}/n\mathbb{Z})$ is generated by $\Phi(A)$ and $\Phi(B)$, which have both order n.

Proposition 3.14 (**Benoist & Quint**). *Suppose $G < SL_d(\mathbb{Z})$ is a subgroup, whose action is strongly irreducible and proximal, and $m \in \mathcal{P}(G)$ is a generating probability measure on G so that supp(G) is finite. Then the Haar measure μ is the unique m-stationary, non-atomic probability distribution.*

For the sketch of the proof we introduce a new dynamical system, we define $\Omega = G^{\mathbb{N}}$, $\mathbb{P} = m^{\mathbb{N}}$ the product measure on Ω and θ the Bernoulli shift $\theta(g_1, g_2, \dots) = (g_2, g_3, \dots)$, where $\omega = (g_1, g_2, \dots) \in \Omega$.
We say that a measure $\nu \in \mathcal{P}^m(\mathbb{T}^d)$ is m-ergodic if it is an extremal point of $\mathcal{P}^m(\mathbb{T}^d)$.

Lemma 3.15. *If $\nu \in \mathcal{P}^m(\mathbb{T}^d)$ is a m-stationary and m-ergodic probability measure on $\mathbb{T}^d$, such that its support supp(ν) is countable. Then ν is G-invariant and its support is finite.*

Proof. Choose some $x \in \mathbb{T}^d$ such that it has maximal measure, say $\nu(\{x\}) = p > 0$. Then

$$\nu(\{x\}) = \int_G \nu(\{g^{-1}x\})dm(g).$$

For m-a.e. $g \in G$, $g^{-1}x$ has also maximal measure. So the support of ν has to be finite.

□

Lemma 3.16 (**Furstenberg**). *Let $\nu \in \mathcal{P}^m(\mathbb{T}^d)$ be a m-stationary measure on $\mathbb{T}^d$. For $\mathbb{P}$-a.e. $\omega \in \Omega$ the following limit exists*

$$\nu_\omega := \lim_{n\to\infty} (g_1 \dots g_n)_* \nu,$$

it satisfies the equivariance condition

$$\nu_\omega = (g_1)_* \nu_{\theta\omega},$$

and ν can be recovered as the average

$$\nu = \int_\Omega \nu_\omega d\mathbb{P}(\omega).$$

Proof. Define the σ-algebra $\mathcal{F}_n := \sigma(\{g_1 \dots g_n\})$. The process $(M_n)_{n\in\mathbb{N}}$

$$M_n : \Omega \to \mathcal{P}(\mathbb{T}^d),\ \omega \to (g_1 \dots g_n)_* \nu$$

is a bounded martingale with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. By Doob's martingale convergence theorem the limit exists and is a probability measure on $\mathbb{T}^d$. The remaining properties follow for continuous functions on $\mathbb{T}^d$, hence also for the measure ν_ω.

□

Lemma 3.17. *For $\mathbb{P}$-a.e. $\omega \in \Omega$ there is a line $V_\omega \in \mathbb{P}^{d-1}(\mathbb{R})$ so that any cluster point g of $\frac{g_1 \cdots g_n}{\|g_1 \cdots g_n\|}$ has image $im(g) = V_\omega$ and satisfies the equivariance property*

$$V_\omega = g_1 V_{\theta\omega}.$$

Proof. See Theorem 4.3 in [3].

□

Furthermore we choose some $v_\omega \in V_\omega$ such that $\|v_\omega\| = 1$ and $\kappa(\omega) := \log \|g_1 v_{T\omega}\|$, so that for $\mathbb{P}$-a.e. $\omega \in \Omega$ we have

$$g_1 v_{\theta\omega} = e^{\kappa(\omega)} v_\omega$$

(we assume that we can choose it to be positive). The main step in the proof of the proposition is the following observation:

Lemma 3.18. *For $\mathbb{P}$-a.e. $\omega \in \Omega$, the limit probability ν_ω is V_ω-invariant; that is ν_ω is translation invariant, $x \mapsto x + v$, with $v \in V_\omega$.*

Proof (Sketch of Proof of Proposition). Let $\nu \in \mathcal{P}(\mathbb{T}^d)$ be a non-atomic probability measure. We want to show that $\nu = \mu$.
The stabilizer of ν

$$G_\nu = \{g \in G \mid g_* \nu = \nu\}$$

is a closed subgroup. The measure ν_ω is translation invariant, so we can consider C_ω, the connected component of V_ω inside of $\mathbb{T}^d$, which satisfies the equivariance condition

$$C_\omega = g_1 C_{\theta\omega}.$$

C_ω is by construcion a non-zero subtorus. The push-forward measure $\mathbb{P}_*$ of $\mathbb{P}$ under the map $\omega \mapsto C_\omega$ is a m-stationary and m-ergodic measure on the countable set $\mathcal{T}$ of non-zero subtori of $\mathbb{T}^d$. By Lemma 3.15 this implies that the support of $\mathbb{P}_*$ is finite and G-invariant. Since the action of G is strongly irreducible we may conclude that $C_\omega = \mathbb{T}^d$. In particular this conclusion, together with the first part of Lemma 3.16, implies that ν_ω is the unique Haar measure on $\mathbb{T}^d$, hence $\nu_\omega = \mu$. The second part of Lemma 3.16 recovers ν as the average over all ν_ω, so $\nu = \mu$.

□

Remark. *(Conditional measures)* Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure space and $\mathcal{A} \subset \mathcal{B}$ a σ-subalgebra. Let $Y \subset X$ be a measurable subset of finite positive measure $\mu(Y) > 0$ and consider the restriction of the measure μ to Y,

$$\mu_Y := \frac{1}{\mu(Y)} \mu|_Y.$$

Formally we define the inclusion $\iota : Y \longrightarrow X$ and denote by $\mathcal{A}_Y := \iota^{-1}\mathcal{A}$ the σ-algebra generated by $\mathcal{A}$ on Y, so that μ_Y is a measure on $\mathcal{A}_Y$. If $f \in L^\infty_\mu(X, \mathcal{B})$, then for μ_Y-a.e. $y \in Y$,

$$\mathbb{E}(\mathbb{1}_Y|\mathcal{A})(y) \neq 0,$$

and

$$\mathbb{E}(f \circ \iota|\mathcal{A}_Y)(y) = \frac{\mathbb{E}(f\mathbb{1}_Y|\mathcal{A})(y)}{\mathbb{E}(\mathbb{1}_Y|\mathcal{A})(y)}.$$

To check this it is enough to note that the set $D := \{\mathbb{E}(\mathbb{1}_Y|\mathcal{A}) = 0\}$ is $\mathcal{A}$-measurable and satisfies $\mu(Y \cap D) = 0$. The equality follows directly from the defining property of the conditional expectation.

Proof (Sketch of Proof of Lemma 3.18). Introduce the dynamical system on $\Omega \times \mathbb{T}^d$,

$$\bar{\mathbb{P}} := \int_\Omega \delta_\omega \otimes \nu_\omega d\mathbb{P}(\omega),$$

$$\bar{T} : \Omega \times \mathbb{T}^d \to \Omega \times \mathbb{T}^d,\ (\omega, x) \mapsto (\theta(\omega), g_1^{-1}x).$$

The probability measure $\bar{\mathbb{P}}$ is $\bar{T}$-invariant. Let $\omega \in \Omega$ and define the parametrization of the leaves by

$$\pi_\omega : \mathbb{R} \to x + V_\omega,\ t \mapsto x + tv_\omega.$$

For $\bar{\mathbb{P}}$-a.e. $(\omega, x) \in \Omega \times \mathbb{T}^d$ the conditional measure along the leaves $K(\omega, x)$ is a Radon measure on $\mathbb{R}$. Lemma 3.18 can be reformulated as translation invariance of these measures $K(\omega, x)$. Define the translation on $\mathbb{R}$ by

$$\tau_t : \mathbb{R} \to \mathbb{R},\ s \mapsto s + t.$$

We need to show that for $\bar{\mathbb{P}}$-a.e. $(\omega, x) \in \Omega \times \mathbb{T}^d$ and $\epsilon > 0$ there is some $t \in (0, \epsilon)$, such that

$$(\tau_t)_* K(\omega, x) = K(\omega, x).$$

The map K satisfies the following two properties:

i) There is a measurable set $A \subset \Omega \times \mathbb{T}^d$ of full measure so that if $(\omega, x) \in A$ and $(\omega, x + tv_\omega) \in A$ then $K(\omega, x) = (\tau_t)_* K(\omega, x + tv_\omega)$.

ii) For $\bar{\mathbb{P}}$-a.e. $(\omega, x) \in \Omega \times \mathbb{T}^d$ we have $K(\omega, x) = (e^{\kappa(\omega)})_* K(\bar{T}(\omega, x))$.

Property *i)* implies that we need only to find enough points $(\omega, x) \in A$ and $(\omega, x + tv_\omega) \in A$ on the same leaf such that K takes the same value on them. Property *ii)* implies that for all $n \in \mathbb{N}$

$$K(\omega, x) = (e^{\kappa_n(\omega)})_* K(\bar{T}^n(\omega, x)),$$

where $\kappa_n(\omega) := \kappa(\omega) + \cdots + \kappa(\theta^{n-1}\omega)$.

The strategy is first to find a measurable set, which is large enough and contains a measurable set containing a large proportion of elements satisfying the first property of the map K. The next idea is to find a set on which we can control the map K, by the second property we can actually control it on $\{|\theta_n(h_{n,\omega}(a)) - \theta_n(\omega)| < 1\}$, where $a \in \mathit{supp}(m)$. We just need to show that the set of elements satisfying this inequality is almost of full measure. The limit measure ν_ω has no atoms for *a.e.* $\omega \in \Omega$. This will allow us to contruct on a set of almost full measure for any point $(\omega, x) \in \Omega \times \mathbb{T}^d$ a sequence (see below) which will be *near* to the line V_ω. The sequence has a limit point and by the first property above the cluster point is invariant under translation, which is what we want to show on a set of full measure up to an ϵ-set.

Let $\epsilon > 0$, then by Lusin's theorem we can find a compact set $C \subset A$ such that $\bar{\mathbb{P}}(C) > 1 - \epsilon$ and on which the functions κ, K, and $\omega \mapsto V_\omega$ are continuous.
To simplify notations, assume that $|\mathit{supp}(m)| = 2$. Define

$$h_{n,\omega,x} : \mathit{supp}(m)^n \to \bar{T}^{-n}(\bar{T}^n(\omega, x)),\ a \mapsto (h_{n,\omega}(a), a_1 \dots a_n g_n^{-1} \dots g_1^{-1} x),$$

where $h_{n,\omega}(a) := (a_1, \dots, a_n, g_{n+1}, g_{n+2}, \dots)$ and $\omega = (g_1, g_2, \dots) \in \Omega$. As a varies, $h_{n,\omega}(a)$ parametrizes the fiber $T^{-n}(T^n\omega) \in \Omega \times \mathbb{T}^d$ and $h_{n,\omega,x}$ the fibers of $\bar{T}^n$ (which contain 2^n elements). Introduce

$$A_{n,\omega} := \{a \in \mathit{supp}(m)^n \mid |\theta_n(h_{n,\omega}(a)) - \theta_n(\omega)| < 1\}.$$

The following Lemma implies that up to an ϵ-negligible set, all elements in $h_{n,\omega,x}(a)$ lie inside of the Lusin set C.

Lemma 3.19. *Let $C \subset \Omega\times\mathbb{T}^d$ be a measurable set. For $\bar{\mathbb{P}}$-a.e. $(\omega, x) \in \Omega\times\mathbb{T}^d$ the following limit exists*

$$\psi_{\omega,x} = \lim_{n\to\infty} \frac{1}{|A_{n,\omega}|} \sum_{a\in A_{n,\omega}} \mathbb{1}_C(h_{n,\omega,x}(a)),$$

and satisfies

$$\int_{\Omega\times\mathbb{T}^d} \psi_{\omega,x} d\mathbb{P}(\omega, x) = \bar{\mathbb{P}}(C)$$

Pick C as in Lusin's theorem, then by Egoroff's theorem we can also find a compact subset $L \subset A$ such that $\bar{\mathbb{P}}(L) > 1 - \epsilon$ and on which the averages above are larger than $1 - \epsilon$ uniformly for $n \geq N$.
The last part uses a so-called exponential drift argument:

Lemma 3.20. *For $\mathbb{P}$-a.e. $\omega \in \Omega$ and for any $x \in \mathbb{T}^d$ we get $\nu_\omega(x + V_\omega) = 0$.*

This lemma implies that we can find a set of full measure such that the above condition holds. Hence if we take the Lusin set L we can find (conditioned) on L a set of full measure such that for $\bar{\mathbb{P}}$-a.e. $(\omega, x) \in \Omega \times \mathbb{T}^d$ there is a sequence $y_k := x + v_k \in \mathbb{T}^d$ so that $(\omega, y_k) \in L$, $v_k \to 0$ as $k \to \infty$ and $v_k \notin V_\omega$ for all $k \in \mathbb{N}$.
For each $k \in \mathbb{N}$ we can find some $n_k \in \mathbb{N}$ such that

$$e^{\kappa_{n_k}(\omega)} \|g_{n_k}^{-1} \dots g_1^{-1} v_k\| \ll_k 1.$$

At least $1 - 8\epsilon$ of all elements $a \in A_{n,\omega}$ parametrize elements of both fibers which belong to C,

$$h_{n_k,\omega,x}(a) = (\omega', x'),\ h_{n_k,\omega,y_k} = (\omega', y') \in C,$$

where we can write $y' = x' + v'$ and $v' = a_1 \dots a_{n_k} g_{n_k}^{-1} \dots g_1^{-1} v_k$. This drift vector v' can be controlled by

$$\|v'\| \ll \|a_1 \dots a_{n_k}\| \|g_{n_k}^{-1} \dots g_1^{-1} v_k\| \ll \|e^{\kappa_{n_k}(h_{n_k,\omega(a)}}\| \|g_{n_k}^{-1} \dots g_1^{-1} v_k\| \ll$$
$$\ll e^{\kappa_{n_k}(\omega)} \|g_{n_k}^{-1} \dots g_1^{-1} v_k\| \ll 1.$$

After taking a subsequence we may assume that $\theta_n(h_{n,\omega}(a)) - \theta_n(\omega) \to \theta_\infty$. Without loss of generality we can assume that $\theta_\infty = 0$. Taking the limit (with respect to the subsequence) of the above sequences we find

$$(\omega'_\infty, x'_\infty),\ (\omega'_\infty, y'_\infty) \in C,$$

with $y'_\infty = x'_\infty + v'_\infty$, $v'_\infty = t_\infty v_{\omega'_\infty}$ and $t_\infty \ll 1$. Up to a set of ϵ-measure we stay in the Lusin set L, so

$$K(\omega, x) = K(\omega'_\infty, x'_\infty) = (\tau_{t_\infty})_* K(\omega'_\infty, x'_\infty + v'_\infty) = (\tau_{t_\infty})_* K(\omega, x).$$

$\square$

4 Random Transformations

4.1 Random Ergodic Theorem

Let $(X, \mathcal{B}, \mu)$ be a probability space and $\Phi \subseteq X^X$ a familiy of measure-preserving transformations of X. Suppose $m \in \mathcal{P}(\Phi)$ is a probability measure on Φ and consider a sequence $(X_n)_{n\in\mathbb{N}}$ of m-distributed Φ-valued random variables defined on the product space $(\Omega, \mathbb{P}) := (\Phi^{\mathbb{N}}, m^{\mathbb{N}})$. As in the example of products of random matrices we will be interested in the random walk on Φ. For this define for $n \in \mathbb{N}$ the product of random transformations

$$S_n := X_n \circ \cdots \circ X_1.$$

Furthermore we will assume that the action of Φ on X

$$\Phi \times X \to X,\ (\varphi, x) \mapsto \varphi(x)$$

is measurable with respect to the measurable structure on Φ and on X. Note, that if Φ is a family of invertible transformations, then we can talk about the group G of transformations of X generated by Φ.
Denote by θ the Bernoulli shift on Ω,

$$\theta : \Omega \to \Omega,\ (\varphi_n)_{n\in\mathbb{N}} \mapsto (\varphi_{n+1})_{n\in\mathbb{N}},$$

and consider the following transformation of the product space $\Omega \times X$,

$$T : \Omega \times X \to \Omega \times X,\ ((\varphi_n)_{n\in\mathbb{N}}, x) \mapsto (\theta((\varphi_n)_{n\in\mathbb{N}}), \varphi_1(x)).$$

Since T is the composition of measure preserving transformations, it is also measure-preserving. Consider the Markov operator acting on $L^p(X, \mu)$, for $p \in [1, \infty]$, defined by

$$P : L^p(X, \mu) \times X \to L^p(X, \mu),\ (f, x) \mapsto \int_\Phi f(\varphi(x))dm(\varphi).$$

Note that if X is a compact metric space, then it is sufficient to define P on the space of continuous functions $\mathcal{C}(X)$ on X.

Proposition 4.1. *Let G be a locally compact group, $m \in \mathcal{P}(G)$ a probability measure on G, X a compact G-space and $\mu \in \mathcal{P}(X)$ a probability measure on X. Then,*
i) The product measure $\mathbb{P} \times \mu$ is T-invariant if and only if μ is m-invariant.
ii) The product measure $\mathbb{P} \times \mu$ is an ergodic measure if and only if μ is an extremal point of $\mathcal{P}^m(X)$.

Proof. It is sufficient to note that any continuous function $f \in \mathcal{C}(X)$ can be identified with a continuous function on $\Omega \times X$ by setting $f(\omega, x) := f(x)$. Let $f \in \mathcal{C}(X)$ be a continuous function on X. If the product measure $\mathbb{P} \times \mu$ is T-invariant then

$$\underbrace{\int_{\Omega \times X} fd(\mathbb{P} \times \mu)}_{= \int_X fd\mu} = \underbrace{\int_{\Omega \times X} f \circ Td(\mathbb{P} \times \mu)}_{= \int_X \int_G f(g \cdot x) dm(g) d\mu(x)}.$$

So μ is m-invariant. Conversely if μ is not m-invariant, the same computation shows that $\mathbb{P} \times \mu$ is not T-invariant.

$\square$

Theorem 4.2 (**Random Ergodic Theorem**). *The following conditions are equivalent:*
i) For any $B \in \mathcal{B}$, $\mu(\varphi^{-1} B \Delta B) = 0$ for m-a.e. $\varphi \in \Phi$ implies $\mu(B) \in \{0, 1\}$.
ii) For $f \in L^p(X, \mu)$, $Pf = f$ μ-a.e. implies that f is constant μ-a.e.
iii) The measure preserving system $(\Omega \times X, \mathbb{P} \otimes \mu, T)$ is ergodic.
Moreover, if any of these conditions is satisfied, then for any $f \in L^1(X, \mu)$ and for $\mathbb{P} \otimes \mu$-a.e. $(\omega, x) \in \Omega \times X$,

$$\frac{1}{n} \sum_{k=0}^{n-1} f \circ S_n(\omega)(x) \xrightarrow[n \to \infty]{L^1} \int_X fd\mu.$$

Proof. $i) \Rightarrow ii)$ Let $f \in L^p(X, \mu)$ so that $Pf = f$ μ-a.e. We want to show that sets of the form $f^{-1}\big([r, \infty)\big) \in \mathcal{B}$ are μ-trivial for any rational $r \in \mathbb{Q}$. Note that $|f| = |Pf| \leq P|f|$, so $|Pf| - |f| \in L^p(X, \mu)$ is non-negative. Moreover we have

$$\int_X (P|f| - |f|) d\mu = \underbrace{\int_X \int_\Phi |f|(\varphi(x)) dm(\varphi) d\mu(x)}_{= \int_X |f| dP^* \mu} - \int_X |f| d\mu = 0$$

Since $P^* \mu = \mu$:

$$P^* \mu(B) = \int_X \int_\Phi \mathbb{1}_{\varphi^{-1} B}(x) dm(\varphi)\, d\mu(x) = \int_\Phi \underbrace{\mu(\varphi^{-1} B)}_{= \mu(B)} dm(x) = \mu(B).$$

So $|f| = P|f|$ μ-a.e., that is $|f| \in L^p(X, \mu)$ is P-invariant. Hence,

$$f^+ = f \vee 0 = \frac{1}{2}(|f| + f),$$

is also P-invariant, so the set $E_0{:=}f^{-1}\big([0,\infty)\big)$ satisfies $\mu(\varphi^{-1}E_0\Delta E_0){=}0$ for m-a.e. $\varphi \in \Phi$. Therefore E_0 is μ-trivial.
Repeat the argument for the function $f - r$, where $r \in \mathbb{Q}$. Then, all sets of the form $f^{-1}\big([r,\infty)\big) \in \mathcal{B}$ are μ-trivial, so f is constant μ-a.e.
ii) $\Rightarrow$ *iii*) We will show that any T-invariant function $f \in L^\infty(\Omega \times X, \mathbb{P} \otimes \mu)$ is constant $\mathbb{P} \otimes \mu$-a.e. Let $f \in L^2(\Omega \times X, \mathbb{P} \otimes \mu)$ be a T-invariant function. For $\mathbb{P} \otimes \mu$-a.e. $(\omega, x) \in \Omega \times X$ we have

$$f \circ T(\omega, x) = f(\theta\omega, \varphi_1(x)) = f(\omega, x).$$

Fix $x \in X$. Consider the projections

$$\omega_n : \Omega \to \Phi,\ \omega = (\varphi_n)_{n\in\mathbb{N}} \mapsto \varphi_n$$

and the martingale generated by these projections and x,

$$M_n(x) := \mathbb{E}(f(\omega, x) \mid \omega_1 \dots \omega_n).$$

Since f is T-invariant we can rewrite $M_n(x)$ in the following way:

$$M_n(x) := \mathbb{E}(\underbrace{f(\omega, x)}_{= f(\theta\omega, \omega_1(x))} \mid \omega_1 \dots \omega_n) = \mathbb{E}(f(\theta\omega, \omega_1(x)) \mid \omega_2 \dots \omega_n) =$$

$$= \cdots = \mathbb{E}(f(\theta^n\omega, \omega_n \dots \omega_1(x)) = M_0(\omega_n \dots \omega_1 x).$$

In particular, for $n = 0$ we obtain for any $x \in X$

$$M_0(x) = \mathbb{E}(f(\omega, x)) = \mathbb{E}(\mathbb{E}(f(\omega, x) \mid \omega_1)) = \mathbb{E}(f(\theta\omega, \omega_1 x)) = PM_0(x),$$

since

$$PM_0(x) = \int_\Phi \underbrace{M_0(\phi(x))}_{= \mathbb{E}(f(\omega, x))}\ dm(\phi) = \mathbb{E}(f(\omega, x)).$$

So M_0 is P-invariant. Hence it is constant μ-a.e. In particular we get that for any $n \in \mathbb{N}$, $M_n = M_0$ is constant μ-a.e. Denote by $\mathcal{F}_n := \sigma((\omega_1, \dots \omega_n), id)$ the σ-algebra generated by the projections and x.
The sequence $L^2(\mathcal{F}_n) \subseteq L^2(\Omega \times X, \mathbb{P} \otimes \mu)$ increases to $L^2(\Omega \times X, \mathbb{P} \otimes \mu)$, so the limit function $f(\omega, x)$ has to be constant and equal to M_0 $\mathbb{P} \otimes \mu$-a.e., since the orthogonal projection of $f(\omega, x)$ onto each $L^2(\mathcal{F}_n)$ is constant.

$\square$

Corollary 4.3 (**Kakutani**). *Let G be a locally compact group and $m \in \mathcal{P}(G)$ a generating probability measure for G. Moreover, suppose that the action on X is ergodic; that is, for any $g \in G$ the system $(X, \mathcal{B}, \mu, g)$ is ergodic. Then, for any $f \in L^1(X, \mu)$ and for $\mathbb{P} \otimes \mu$-a.e. $(\omega, x) \in \Omega \times X$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} f \circ S_k(\omega)(x) = \int_X f d\mu.$$

Proof. We want to show that for any $B \in \mathcal{B}$, $\mu(gB\Delta B) = 0$ for m-a.e. $g \in G$ implies $\mu(B) \in \{0, 1\}$, the previous theorem implies the rest. Take any such $B \in \mathcal{B}$. The unitary representation of G on $L^2(X, \mu)$,

$$\pi : G \to L^2(X, \mu),\ g \mapsto \big(x \mapsto f(g^{-1}x)\big)$$

is continuous in the weak topology. Moreover, $\pi(g)(\mathbb{1}_B) = \mathbb{1}_B$ for m-a.e. $g \in G$. So by continuity, $\pi(g)(\mathbb{1}_B) = \mathbb{1}_B$ for any $g \in supp(m)$. Since for any $g \in G$ the system $(X, \mathcal{B}, \mu, g)$ is ergodic, $\pi(g)(\mathbb{1}_B) = \mathbb{1}_B$ implies that $\mathbb{1}_B$ is constant. Hence, $\mu(B) \in \{0, 1\}$.

□

4.2 Examples

Example 4.1. Let $(X, \mathcal{B}, \mu, T)$ be an ergodic system. Then, the group $G = \{id\}$ acts trivially on X and the above theorem is just the ergodic theorem from section 1.

Example 4.2 (*Simple rotation on the 2-sphere*). Consider the 2-sphere $\mathbb{S}^2$ with the Lebesgue measure μ and the special orthogonal group $SO(3)$ with measure $m = \delta_g$, where $g \in SO(3) \setminus \{id\}$. The corresponding Markov operator is the map,

$$P : \mathcal{C}(\mathbb{S}^2) \times \mathbb{S}^2 \to \mathcal{C}(\mathbb{S}^2),\ (f, x) \mapsto \int_{\mathbb{S}^2} f(y) d\mu_x(y) = f(g \cdot x).$$

From Lemma 3.9. we already know that for any $f \in \mathcal{C}(\mathbb{S}^2)$, $Pf = f$ μ-a.e. implies that f is constant μ-a.e.

Example 4.3 (*Rotation on the 2-sphere*). Consider as above the 2-sphere with the Lebesgue measure and $SO(3)$ acting on the sphere with the Haar measure m. The induced Markov operator is the map,

$$P : \mathcal{C}(\mathbb{S}^2) \times \mathbb{S}^2 \to \mathcal{C}(\mathbb{S}^2),\ (f, x) \mapsto \int_{SO(3)} f(g \cdot x) dm(g).$$

In Lemma 3.12. we proved that for any $B \in \mathcal{B}$, $\mu(gB\Delta B) = 0$ for m-a.e. $g \in SO(3)$ implies $\mu(B) \in \{0, 1\}$.

Example 4.4 (*Automorphisms on the torus*). As above, let $g \in SL_d(\mathbb{Z})$ so that no eigenvalue of g is a root of unity and consider the probability distribution $m := \delta_g \in \mathcal{P}(G)$ on $G = \langle g \rangle < SL_d(\mathbb{Z})$. The corresponding Markov operator acting on $\mathcal{C}(\mathbb{T}^d)$ is

$$P : \mathcal{C}(\mathbb{T}^d) \times \mathbb{T}^d \to \mathcal{C}(\mathbb{T}^d),\ (f, x) \to f(g \cdot x).$$

Since no eigenvalue of g is a root of unity, the action of G on $\mathbb{T}^d$ is ergodic.

References

[1] Y. Benoist, J.-F. Quint, *Stationary measures and invariant subsets of homogeneous spaces (II)*, CRAS 349, 2011, pp. 341-345, 2011.

[2] Y. Benoist, J.-F. Quint, *Introduction to random walks on homogeneous spaces*, 10th Takagi Lectures, 2012.

[3] P. Bougerol, J. Lacroix, *Products of random matrices with applications to Schrödinger operators*, Birkhäuser, 1985.

[4] J. Bourgain, A. Furman, E. Lindenstrauss, S. Mozes, *Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus* J. Amer. Math. Soc. 24, 2011, pp. 231 - 280.

[5] R. Durrett, *Probability: Theory and examples*, fourth edition, Cambridge University Press, 2010.

[6] M. Einsiedler and T. Ward, *Ergodic Theory: with a view towards Number Theory*, Springer-Verlag, London, 2011.

[7] M. Einsiedler and T. Ward, *Functional Analysis Notes*, draft, 2012

[8] A. Furman, *Random walks on groups and random transformations. Handbook of dynamical systems, Vol. 1A, 931-1014* , North-Holland, Amsterdam, 2002.

[9] A. Furman, Y. Shalom, *Sharp ergodic theorems and strong ergodicity*, Ergod. Th. Dynam. Syst., 1999, No 19, 1037-1061.

[10] H. Furstenberg, Y. Kifer, *Random matrix products*, Israel J. Math. 46, 1983, No. 1, 12-32.

[11] D. A. Levin, Y. Peres and E. L. Wilmer, *Markov chains and mixing times*, AMS, 2009.

[12] D. Revuz, *Markov Chains*, second edition, North-Holland, Amsterdam, 1984