THE ERGODIC THEOREM FOR COMPACT GROUPS

These notes contain a proof of the ergodic theorem for compact groups, that is to say the necessary and sufficient condition for a simple random walk on a compact group to converge together with a description of the limit. The document was written for third-year students but they nevertheless assume knowledge concerning the Fourier transform on compact groups.

1 THE SETTING

We are interested in the following basic problem : let G be a compact Hausdorff group and let μ be a probability measure on G. Under what condition does the sequence of measures $(\mu^{*k})_{k\in\mathbf{N}}$ converge ? Before discussing details, let us make the terms more precise.

Given two probability measures μ and ν on a compact group G, one may define their *convolution product* through the formula

$$(\mu * \nu)(A) = (\mu\otimes\nu)\left(\{(g_1,g_2)\in G^2 \mid g_1g_2\in A\}\right)$$

where A is any Borel subset of G. Equivalently, given a Borel measurable function $f : G\to\mathbf{C}$, we have

$$\int_G f(g)\mathrm{d}(\mu*\nu)(g) = \int_G\int_G f(gh)\mathrm{d}\mu(g)\mathrm{d}\mu(h).$$

It follows from the associativity of the group law that convolution itslef is associative. Therefore, iterating the convolution product of a single probability measure μ with itself yields a well-defined sequence of convolution powers μ^{*k} which form a sequence of measures and we want to study its potential convergence. In the present text, "convergence" for a sequence of measures $(\mu_k)_{k\in\mathbf{N}}$ will always mean the following.

DEFINITION 1.1. A sequence $(\mu_k)_{k\in\mathbf{N}}$ of measures on a compact group G is said to *converge to a measure* μ if for any essentially bounded measurable function $f : G\to\mathbf{C}$,

$$\int_G f(g)\mathrm{d}\mu_k(g) \underset{k\to+\infty}{\longrightarrow} \int_G f(g)\mathrm{d}\mu(g).$$

2 THE CONVERGENCE CRITERION

2.1 A NECESSARY CONDITION

We will start by giving a necessary condition for convergence which is very simple and almost sufficient. Before explaining it, let us introduce some vocabulary.

DEFINITION 2.1. A measure is said to be *deterministic* if it is equal to δ_g for some $g\in G$.

Let us start with a simple computation which we isolate in a separate lemma for clarity.

Lemma 2.2. *Let $g\in\mathbb{G}$. Then, $\delta_g^{*k} = \delta_{g^k}$.*

Proof. We proceed by induction on k. The result is true for $k = 1$ and if it holds for some $k\geqslant 1$, then for any Borel subset $A\subset G\times G$,

$$\begin{aligned}\delta_{g^{k+1}}(A) &= \delta_g * \delta_{g^k}(A)\\ &= (\delta_g\otimes\delta_{g^k})\left(\{(g_1,g_2)\in G^2 \mid g_1g_2\in A\}\right)\\ &= \delta_{gg^k\in A}\\ &= \delta_{g^{k+1}}(A).\end{aligned}$$

■

For deterministic measures, convergence is easily characterized.

Proposition 2.3. *Let μ be a deterministic probability measure on a finite group G. Then, the sequence $(\mu^{*k})_{k\in\mathbf{N}}$ converges if and only if $\mu = \delta_e$.*

Proof. By Lemma 2.2, we are investigating the convergence of the sequence $(\delta_{g^k})_{k\in\mathbf{N}}$. Let us prove that this implies the convergence of the sequence $(g^k)_{k\in\mathbf{N}}$. By contradiction, assume that the sequence does not converge. Because G is compact, there must be at least two cluster points, say g_1 and g_2. Let now A_1 and A_2 be open subsets of G such that $g_i \in A_i$ for $i = 1,2$ and $A_1 \cap A_2 = \emptyset$ (they exist because G is Hausdorff). By construction, the sequence $(g^k)_{k\in\mathbf{N}}$ takes infinitely many times values in A_1 and infinitely many times values in A_2. Therefore,

$$\int_G \mathbf{1}_{A_1}(h)\mathrm{d}\delta_{g^k}(h)$$

takes infinitely many times the values 0 and 1. This contradicts the convergence of the sequence $(\mathbf{1}_{A_1}(g^k))_{k\in\mathbf{N}}$, hence the result.

Let now h be the limit of that sequence and observe that $(g^{k+1})_{k\in\mathbf{N}}$ converges both to h and to gh. It follows from uniqueness of the limit that $h = gh$, which yields $g = e$. Conversely, it follows from the beginning of the proof that $\delta_e^{*k} = \delta_{e^k} = \delta_e$ so that the sequence is constant hence convergent in that case. ■

The previous, apparently harmless statement is the key to the convergence criterion. To see it, let us consider the situation where we have a quotient group

$$q : G \to H.$$

We can then naturally push the measure μ forward to a measure $q_*\mu$ on H through the formula

$$q_*\mu(A) = \mu(q^{-1}(A))$$

for A a Borel subset of H. Otherwise said, for an essentially bounded Borel measurable function f on H,

$$\int_H f(h)q_*\mu(h) = \int_G f\circ q(g)\mu(g).$$

The definition moreover behaves nicely with respect to convolution.

Lemma 2.4. *With the previous notations, if μ,ν are probability measure on G, then*

$$\left(q_*\mu\right) * \left(q_*\nu\right) = q_*\left(\mu * \nu\right).$$

Proof. Let $B \subset H$ be a Borel subset. Then,

$$\begin{aligned}(q_*\mu * q_*\nu)(B) &= \left((q_*\mu)\otimes(q_*\nu)\right)\left(\{(h_1,h_2)\in H^2 \mid h_1h_2\in B\}\right)\\ &= (\mu\otimes\nu)\left(\{(g_1,g_2)\in G^2 \mid q(g_1g_2)\in B\}\right)\\ &= \mu*\nu(q^{-1}(B))\\ &= q_*(\mu*\nu).\end{aligned}$$

■

From this a necessary condition for convergence follows. Let us define the *support* of μ to be the intersection of all the closed subsets of G with measure one. This is a closed subset denoted by $\mathrm{Supp}(\mu)$.

Proposition 2.5. *If the sequence of measures $(\mu^{*k})_{k\in\mathbf{N}}$ is convergent, then the support of μ is not contained in a non-trivial coset with respect to a normal subgroup of G.*

Proof. We proceed by contradiction. Let $N < G$ be a normal subgroup and let $g_0 \in G \setminus N$ be such that the support of μ is contained in $g_0 N$. Set $H = G/N$ and let q be the canonical quotient map. We claim that $q_*\mu$ is deterministic. Indeed, for any Borel subset $A \subset G$,

$$\begin{aligned} q_*\mu(A) &= \mu(q^{-1}(A) \\ &= \mu\left(q^{-1}(A) \cap (g_0 N)\right) \end{aligned}$$

Because $q^{-1}(A)$ is a union of left cosets, either $g_0 N \subset A$ or $A \cap g_0 N = \emptyset$. In the first case, $q_*\mu(A) = 1$ and in the second case $q_*\mu(A) = 0$. In other words,

$$q_*\mu = \delta_{g_0 N}.$$

We know by Lemma 2.3 that the sequence $(q_*\mu^{*k})_{k\in\mathbf{N}}$ converges if and only if it is supported on the neutral element, that is to say if $g_0 \in N$. Therefore, the sequence does not converge, which contradicts the convergence of $(\mu^{*k})_{k\in\mathbf{N}}$ since for any function $f : H \to \mathbf{C}$ we have

$$\begin{aligned} \int_H f(h)\mathrm{d}(q_*\mu)^{*k}(h) &= \int_G f\circ q(g)\mathrm{d}\mu^{*k}(g) \\ &\underset{k\to+\infty}{\longrightarrow} \int_G f\circ q(g)\mathrm{d}\nu(g) \\ &= \int_H f(h)\mathrm{d}(q_*\nu)(h), \end{aligned}$$

where ν denotes the limit of $(\mu^{*k})_{k\in\mathbf{N}}$. ∎

Note that if the support is contained in a proper closed subgroup, even not a normal one, then the random walk may very well converge. However, its limit will obviously not be the Haar measure on G since the latter has support equal to G.

2.2 THE COMPLETE CHARACTERIZATION

The necessary and sufficient criterion for convergence of the sequence $(\mu^{*k})_{k\in\mathbf{N}}$ is a slight refinement of Proposition 2.5 and requires two additional features : first, to take into account the fact that the support of μ may not generate G as a topological group and second, that we need to be able to prove the converse.

The first point will be dealt with directly in the proof of Theorem 2.9. As for the second one, it is easily understood using the *Fourier transform* [1]. Recall that if π denotes a finite-dimensional complex representation of G, then its Fourier transform under μ is the matrix

$$\widehat{\mu}(\pi) = \int_G \pi(g)\mathrm{d}\mu(g) \in M_{\dim(\pi)}(\mathbf{C}).$$

If $\mathrm{Irr}(G)$ denotes the set of equivalence classes of irreducible representations of G and if π_α is a fixed representative of the class $\alpha \in \mathrm{Irr}(G)$, then μ is completely determined by the data of the matrices $(\widehat{\mu}(\pi_\alpha))_{\alpha\in\mathrm{Irr}(G)}$. Moreover, convolution and Fourier transform are closely related.

Lemma 2.6. *For a finite-dimensional representation π and two Borel probability measure μ, ν on G, we have*

$$\widehat{\mu * \nu}(\pi) = \widehat{\mu}(\pi)\widehat{\nu}(\pi).$$

In particular, $\widehat{(\mu^{*k})}(\pi) = \widehat{\mu}(\pi)^k$.

1. The reader unfamiliar with the subject may refer for instance to A. Robert, *Introduction to the representation theory of compact and locally compact groups*, London Mathematical Society Lecture Note Series **80**, Cambridge University Press, 1983 for a comprehensive exposition of the subject. The reader already familiar with the representation theory of finite groups (as treated for instance in J.-P. Serre, *Linear representations of finite groups*, Graduate Texts in Mathematics **42**, Springer, 1977) may simply trust our word that things work the same in the compact setting.

Proof. By definition,

$$\begin{aligned}\widehat{\mu * \nu}(\pi) &= \int_G \pi(g)\mathrm{d}(\mu * \nu)(g)\\ &= \int_G\int_G \pi(gh)\mathrm{d}\mu(g)\mathrm{d}\nu(h)\\ &= \int_G\int_G \pi(g)\pi(h)\mathrm{d}\mu(g)\mathrm{d}\nu(h)\\ &= \left(\int_G \pi(g)\mathrm{d}\mu(g)\right)\left(\int_G \pi(h)\mathrm{d}\nu(h)\right)\\ &= \widehat{\mu}(\pi)\widehat{\nu}(\pi).\end{aligned}$$

■

We will need at some point an elementary result from linear algebra that we state and prove separately for completeness. In what follows, $\|\cdot\|$ denotes the norm on $M_n(\mathbf{C})$ subordinated to the standard Hilbert space norm on $\mathbf{C}^n$.

Lemma 2.7. *Let $M \in M_n(\mathbf{C})$ be a matrix with $\|M\| = 1$ and assume that the sequence M^k does not converge. Then, M has an eigenvalue of modulus one.*

Proof. Because of the fact that M has norm one, we know that all its eigenvalues have modulus less than one. Moreover, we can use its Dunford decomposition $M = D + N$, where D is diagonal with coefficients being eigenvalues of M and N is nilpotent and commutes with D. We therefore have

$$M^k = \sum_{i=0}^{k} \binom{k}{i} N^i D^{k-i}.$$

If k_0 is such that $N^{k_0} = 0$, then for $k \geqslant k_0$ we have

$$M^k = \sum_{i=0}^{k_0} N^i D^{k-i}.$$

If all the eigenvalues of M have modulus strictly less than one, then D^{k-i} tends to 0 as k goes to infinity for all $0 \leqslant i \leqslant k_0$, hence M^k also tends to 0, a contradiction. ■

Using this we can give a sufficient condition for convergence.

Lemma 2.8. *Assume that μ is not supported on a coset of a subgroup of G. Then, the sequence $(\mu^{*k})_{k\in\mathbf{N}}$ converges.*

Proof. By continuity of the Fourier transform and its inverse, the sequence converges if and only if the sequence $(\widehat{\mu}(\pi)^k)_{k\in\mathbf{N}}$ converges for all irreducible representations π of G. Let us assume the existence of a π such that the sequence does not converge and recall that it is equivalent to a unitary one, hence we may assume without loss of generality π to be unitary. Because

$$\|\widehat{\mu}(\pi)\| \leqslant 1,$$

we must have $\|\widehat{\mu}(\pi)\| = 1$ so that by Lemma 2.7 there exists a vector ξ of norm one in the carrier space of π and a complex number λ of modulus one such that $\widehat{\mu}(\pi)\xi = \lambda\xi$. This means that the following inequalities are all equalities :

$$\begin{aligned}\left\|\widehat{\mu}(\pi)\xi\right\| &= \left\|\int_G \pi(g)\xi\mathrm{d}\mu(g)\right\|\\ &\leqslant \int_G \|\pi(g)\xi\|\mathrm{d}\mu(g)\\ &\leqslant 1\end{aligned}$$

The equality in the first inequality implies that the vectors $\pi(g)\xi$ are colinear for almost all g, i.e. there exists η and complex numbers $(\lambda_g)_{g\in G}$ such that

$$\pi(g)\xi = \lambda_g\eta$$

for almost all g in the support of μ. Because furthermore

$$\begin{aligned}\left(\int_G \lambda_g \mathrm{d}\mu(g)\right)\eta &= \int_G \lambda_g\eta \mathrm{d}\mu(g)\\ &= \left(\int_G \pi(g)\mathrm{d}\mu(g)\right)\xi\\ &= \widehat{\mu}(\pi)\xi\\ &= \lambda\xi,\end{aligned}$$

we can chose $\eta = \xi$. Because $\pi(g)$ is unitary, this forces $|\lambda_g| = 1$ for almost all g. But then, we have

$$\begin{aligned}\left|\int_G \lambda_g \mathrm{d}\mu(g)\right| &= \left\|\widehat{\mu}(\pi)\xi\right\|\\ &= |\lambda|\\ &= 1\end{aligned}$$

so that the triangle inequality is an equality. This means that all the numbers λ_g are equal up to a positive real number. Because $|\lambda_g| = 1$, this implies that they are all equal. In conclusion, there exists a unit vector ξ and a complex number λ of modulus one such that

$$\pi(g)\xi = \lambda\xi$$

for almost all $g \in \mathrm{Supp}(\mu)$ and since π is a continuous function, the equality in fact holds for all $g \in \mathrm{Supp}(\mu)$.

Let us now consider the two following subsets of G :

$$H_\lambda = \{g \in G \mid \pi(g)\xi = \lambda\xi\} \quad \& \quad H_1 = \{g \in G \mid \pi(g)\xi = \xi\}.$$

Then, H_1 is a obviously a subgroup and it is moreover proper (i.e. not equal to G) because ξ is a fixed vector for the restriction of π to H_1 and since α is irreducible on G, it cannot have a fixed vector [2]. Moreover, H_λ is a coset of H_1 and $\mathrm{Supp}(\mu) \subset H_\lambda$, hence the result. ■

We are now ready for a complete characterization. To simplify the notations, we will write G_μ for the closed subgroup of G generated by the support of μ. It is intuitively clear that this is the important object and it is often assumed that $G_\mu = G$. We will however give the statement in full generality, because we do not want to make an assumption on the potential limit of the sequence for the moment.

THEOREM 2.9 The sequence $(\mu^{*k})_{k\in\mathbf{N}}$ converges if and only if the support of μ is not contained in a non-trivial coset of a normal closed subgroup of G_μ.

Proof. Assume that the sequence converges. Note that because by definition $\mu(G_\mu) = 1$, the restriction $\widetilde{\mu}$ of μ to G_μ is again a probability measure and that

$$\widetilde{\mu^{*k}} = \widetilde{\mu}^{*k}.$$

As a consequence, the latter sequence converges so that by Lemma 2.5, the support of $\widetilde{\mu}$ is not contained in a non-trivial coset of a normal subgroup of G_μ. This implies that the same holds for μ, proving the "only if" part.

2. If $\pi(g)\xi = \xi$ for all $g \in G$, then $\mathbf{C}.\xi$ is a subspace which is stable under the representation, hence a subrepresentation.

As for the "if" part, with the notations of Lemma 2.8 it is enough to see that the closed subgroup H_1 is normal in G_μ. Let us set

$$K = \{g \in G \mid gH_1 = H_1 g\},$$

which is the largest subgroup containing H_1 as a normal subgroup. If we can prove that K contains the support of μ, then because it is a subgroup we will have $G_\mu \subset K$, concluding the proof. But we have, for any $g \in H_\lambda$,

$$gH_1 = H_\lambda = H_1 g$$

and since the support of μ is contained in H_λ, the proof is complete. ■

3 THE LIMIT

We have now answered our original question, namely deciding whether the sequence of convolution powers of a probability measure converges or not. This naturally leads us to a second question which is that of describing the limit. An immediate observation is that if the sequence $(\mu^{*k})_{k\in\mathbf{N}}$ converges to a measure ν, then $(\mu^{*2k})_{k\in\mathbf{N}}$ converges to both ν and $\nu * \nu$, so that these two measures are equal. Hence, the next definition is meaningful for our problem.

DEFINITION 3.1. A measure ν is said to be *idempotent* if $\nu * \nu = \nu$.

Conversely, any idempotent measure is the limit of the constant sequence of its own convolution powers, so that we are really looking for a description of all idempotent measures. To give it, let us first point out a useful fact.

Lemma 3.2. *Let ν be an idempotent measure on G. Then, for any representation π of G, $\widehat{\nu}(\pi)$ is the matrix of a projection.*

Proof. This follows from an easy computation :

$$\begin{aligned}\widehat{\nu}(\pi)^2 &= \widehat{\nu}(\pi)\widehat{\nu}(\pi)\\ &= \widehat{\nu * \nu}(\pi)\\ &= \widehat{\nu}(\pi).\end{aligned}$$

■

We are now ready for the final word of the story.

THEOREM 3.3 Let ν be an idempotent measure. Then, ν is the uniform measure on G_ν.

Proof. By contradiction, assume that there exists a non-trivial irreducible representation π of G such that $\widehat{\nu}(\pi) \neq 0$. Because this is the matrix of a projection by Lemma 3.2, there exists $\xi \in \mathbf{C}^{\dim(\pi)}$ such that

$$\widehat{\nu}(\pi)\xi = \xi$$

and it follows from the same argument as in the proof of Lemma 2.8 that the support of ν is contained in the proper closed subgroup

$$H = \{g \in G \mid \pi(g)\xi = \xi\}.$$

But H contains the closed subgroup generated by $\mathrm{Supp}(\nu)$, and the fact that $H \neq G_\nu$ yields a contradiction. We conclude that $\widehat{\nu}(\pi) = 0$. But then, ν has the same Fourier transform as the Haar measure, hence it is the Haar measure. ■

We can now combine Theorem 2.9 with Theorem 3.3 to recover the classical formulation of the ergodic theorem for compact groups.

Corollary 3.4. *Let G be a compact group and let μ be a probability measure on G. Then, the sequence $(\mu^{*k})_{k\in\mathbf{N}}$ converges to the Haar measure if and only if it satisfies the following two conditions :*

1. *Its support is not contained in a coset of a proper closed normal subgroup ;*
2. *Its support is not contained in a proper closed subgroup.*