Mean ergodic theorem

Ergodic theory studies the long time average behavior of a system. The basic principle is that all systems must tend on average to equilibrium in some sense, and we wish to understand in what sense.

We begin with a version of this in Hilbert space.

Theorem. *Let $\mathcal{H}$ be a Hilbert space, and let $U\colon \mathcal{H} \to \mathcal{H}$ be a linear map such that $\|Uf\| \le \|f\|$ for all $f \in \mathcal{H}$. Let*

$$S_N f = \frac{1}{N} \sum_{n=0}^{N-1} U^n f,$$

and let $P\colon \mathcal{H} \to \mathcal{H}$ be the orthogonal projection onto the kernel of $(I - U)$. Then

$$\lim_{N\to\infty} \|S_N f - Pf\| = 0, \qquad \text{for all } f \in \mathcal{H}. \tag{$*$}$$

Here f describes the state of a system, and the map U represents time evolution. Thus, as $N \to \infty$, S_N is the long-time average behavior of the system. The theorem asserts that the long time average converges to a particular projection.

The most important case is $\mathcal{H} = L^2(X, \mu)$ and $Uf = f \circ T$, where $T\colon X \to X$ obeys $\int f \circ T \, d\mu = \int f \, d\mu$ for every $f \in L^1(X, \mu)$; we say that μ is an *invariant measure* for T. Then $\|Uf\| = \|f\|$ for every $f \in \mathcal{H}$.

Proof. The proof proceeds by orthogonal decomposition.

(1) If $f \in \operatorname{Ran}(I - U)$, i.e. if $f = g - Ug$ for some $g \in \mathcal{H}$, then the series telescopes and we have
$$\|S_N f\| = \|g - U^N g\|/N \le 2\|g\|/N \to 0.$$

(2) Furthermore, $\|S_N f\| \to 0$ for $f \in \overline{\operatorname{Ran}(I - U)}$ by continuity. Indeed, given any linear operator S on a normed vector space V, to show that $\|S_N f\|_V \to 0$ for all $f \in V$, it is enough to check it for f ranging over a dense subset of V and then check also $\sup_N \|S_N\|_{V\to V} < +\infty$. (This is exercise I.27 in Reed and Simon)

(3) If $f \in \overline{\operatorname{Ran}(I - U)}^\perp$, then in particular $\langle f, (I-U)f\rangle = 0$ and so $\langle f, Uf\rangle = \|f\|^2$. Consequently,
$$\|f - Uf\|^2 = \|f\|^2 - 2\operatorname{Re}\langle f, Uf\rangle + \|Uf\|^2 = \|Uf\|^2 - \|f\|^2 \le 0,$$
and hence $Uf = f$.

(4) The above proves ($*$) with P equal to orthogonal projection onto $\overline{\operatorname{Ran}(I - U)}^\perp$, and also that $\overline{\operatorname{Ran}(I - U)}^\perp \subset \operatorname{Ker}(I - U)$. It remains to show that $\operatorname{Ker}(I - U) \subset \overline{\operatorname{Ran}(I - U)}^\perp$. But if $Uf = f$, then $S_N f = f$, which implies $Pf = f$ by ($*$), which implies $f \in \overline{\operatorname{Ran}(I - U)}^\perp$ because P is a projection. □

This beautiful proof comes from *Sur la théorie ergodique* by Frédéric Riesz, 1944. It generalizes J. V. Neumann's 1931 *Proof of the Quasi-Ergodic Hypothesis*.

Exercises.

Kiril Datchev, March 31, 2025. Questions, comments, and corrections are gratefully received at kdatchev@purdue.edu.

(1) Adapt the proof above to show that if $f \in \mathcal{H}$, then
$$\lim_{N\to\infty} \sup_{M\in\mathbb{N}_0} \left\|S_N U^M f - Pf\right\| = 0.$$
Hint: Show that if $f = g - Ug$, then $\|S^N U^M \tilde{f}\| \leq \frac{2}{N}\|g\| + \|f - \tilde{f}\|$ for all N, M, and $\tilde{f}$.

(2) Let $U\colon \mathbb{F}^n \to \mathbb{F}^n$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, be a linear map whose eigenvectors span $\mathbb{F}^n$ and whose eigenvalues all have magnitude at most one. Prove that $\lim_{N\to\infty} |S_N v - Pv| = 0$ for all $v \in \mathbb{F}^n$, where P is a projection (not necessarily orthogonal) onto the kernel of $(I - U)$.

(3) Prove that if $\mathcal{H} = L^2(\mathbb{R}^d, dx)$ and $Uf(x) = f(x - p)$ for some fixed p, then $P = 0$ in ($*$).

If μ is finite, then constants belong to the range of P. If the range of P consists only of constants, then we say μ is *ergodic* and the Mean Ergodic Theorem becomes
$$S_N f \to \frac{1}{\mu(X)} \int_X f\, d\mu, \quad \text{in the } L^2 \text{ sense.}$$
In words, we say the time averages converge to the space average.

This convergence result has many variants, analogous in some ways to the different convergence results for Fourier series and their generalizations. The most important is Birkhoff's Ergodic Theorem (BET), which asserts convergence almost everywhere for $f \in L^1(X, \mu)$. For a more substantial introduction to the subject, including a proof of BET, see (in ascending order of scope):

- Walkden's Notes on Ergodic Theory for a thorough presentation of BET, with some simple and striking applications of it,
- Pollicott and Yuri's small book *Dynamical Systems and Ergodic Theory* for more connections to other mathematical problems, particularly in number theory,
- Katok and Hasselblatt's big book *Introduction to the Modern Theory of Dynamical Systems* for a broad introduction to dynamical systems.

Below we develop some examples and applications of the Mean Ergodic Theorem, following the references above.

Our main examples are ones in which X is the unit circle in $\mathbb{C}$, μ is arclength measure, and T is either a rotation or the doubling map $T(e^{i\theta}) = e^{2i\theta}$.

Consider first the case where T is rotation by an angle α. Then $Uf = f$ if and only if both α and 2π are periods of f. If $\alpha/\pi \notin \mathbb{Q}$ this occurs if and only if f is constant, so μ is ergodic. If $\alpha/\pi \in \mathbb{Q}$ then there are many other solutions to $Uf = f$. In this case we can get ergodicity by using a sum of delta masses for μ instead of arclength measure. (More invariant measures can be constructed by taking nonnegative linear combinations (or more general superpositions) of these examples.) We say that the dynamical system *decomposes* into independent dynamical systems, one for each periodic orbit.

Consider now the case where T is the doubling map. Writing $f(e^{i\theta}) = \sum c_n e^{in\theta}$ we find that $Uf = f$ if and only if $\sum a_n e^{in\theta} = \sum a_{2n} e^{in\theta}$, i.e. if and only if $a_n = a_{n/2}$ for even n and $a_n = 0$ for odd n, which implies again $Uf = f$ if and only if f is constant. Thus arclength measure is ergodic for the doubling map. In this case there are also other invariant measures, like δ_0 and $\delta_{2\pi/3} + \delta_{4\pi/3}$.

Exercise. (4) Let $T\colon X \to X$ be given, and suppose T has a periodic orbit of period n (i.e. there is $x \in X$ such that $n = \min\{m \in \mathbb{N}\colon T^m x = x\}$). Prove there is a unique invariant probability measure of T supported on this orbit, and it is ergodic. Give some examples in the setting of rotations on the circle and the doubling map on the circle.

We now prove that irrational rotation is uniquely ergodic.

Theorem. *Let X be the unit circle in $\mathbb{C}$, let $T(\theta) = \theta + \alpha$ with $\alpha/\pi \notin \mathbb{Q}$, and let μ be a finite Borel measure on X such that*

$$S_N f \to \int_X f\,d\mu, \quad \text{in the } L^2 \text{ sense.}$$

Then μ is normalized arclength measure, i.e. $d\mu = d\theta/2\pi$.

Proof. If $k \neq 0$, then

$$S_N e^{ik\theta} = \frac{1}{N}\sum_{n=0}^{N-1} e^{ik(\theta+n\alpha)} = \frac{e^{ik\theta}}{N}\frac{1-e^{ikN\alpha}}{1-e^{ik\alpha}}.$$

The denominator is nonzero because $\alpha/\pi \notin \mathbb{Q}$, so $\|S_N e^{ik\theta}\|_{C(K)} \to 0$ as $N \to \infty$. Since $S_N 1 = 1$, $\|S_N\|_{C(X)\to C(X)} = 1$ and since the span of $\{e^{ik\theta}\}_{k\in\mathbb{Z}}$ is dense in $C(X)$, it follows that, for all f in $C(X)$, $\|S_N f - c_f\|_{C(X)} \to 0$ as $N \to \infty$, where $c_f = \frac{1}{2\pi}\int f d\theta$.

Since $\|g\|_{L^2(X,\mu)} \leq \sqrt{\mu(X)}\|g\|_{C(X)}$ for all $g \in C(X)$, it follows that $\|S_N f - c_f\|_{L^2(X,\mu)} \to 0$ for all $f \in C(X)$. By uniqueness of limits, $\frac{1}{2\pi}\int f d\theta = \int f d\mu$. Since a Borel measure is uniquely determined by the linear functional it induces on $C(X)$, it follows that $d\mu = d\theta/2\pi$. [1] □

We end with a fun example. Let X be the space of sequences of letters; this represents the result of a monkey typing randomly. Define a *cylinder of length* n to be the set of sequences starting with a given string. Thus, the sequences starting with the letter H are a cylinder of length 1, and the sequences starting with the text of Hamlet are a cylinder of length 133,834. Given any n, X is the disjoint union of the 26^n cylinders of length n. Define a probability measure μ on X by letting $\mu(A) = 1/26^n$ whenever $A \subset X$ is cylinder of length n, and extending μ to the σ-algebra generated by the cylinders.

It is easy to see that sequences without the letter 'H' have measure zero, i.e. the letter 'H' appears with probability 1. This is because the chances of 'H' not appearing in the first N letters is $(25/26)^N$, which tends to 0 as $N \to \infty$. One can similarly show that any given string appears with probability 1, and further it appears infinitely often with probability 1. Thus, for any N, the monkey typing randomly will eventually write Hamlet N times.

Now let us use ergodicity to show that every string shows up on a regular basis. Let $T\colon X \to X$ be the map which deletes the first letter in the sequence. If we can show that T is ergodic, then we can let I be any cylinder of length n, and apply the Mean Ergodic Theorem to obtain $S_N 1_I \to 1/26^n$ in $L^2(X,\mu)$. Thus, for instance, the letter A shows up 1/26 of the time on average, and the text of Hamlet shows up about $1/26^{133{,}834}$ of the time on average.

μ is invariant because $\mu(T^{-1}A) = \mu(A)$ for any cylinder A. It is ergodic because it is *mixing*: this means that $\mu(T^{-N}A \cap B) \to \mu(A)\mu(B)$ as $N \to \infty$ for all measurable A and B. In words, we say μ is

[1] In more detail, if two Borel measures μ and ν on a compact metric X space differ on a measurable set, then by inner regularity they must differ on a compact set $K \subset X$, and since $f_n(x) = \exp(-nd(x,K))$ defines a sequence of continuous functions converging monotonically to the indicator function of K, it follows that there is N such that $\int f_n d\mu \neq \int f_n d\nu$ for $n \geq N$.

mixing if the chances of being now in B and N steps from now in A converge to the product of the chances of being in A and B separately; i.e. in the limit the events 'being now in B' and 'being N steps from now in A' decouple and approach being independent events.

To see that mixing implies ergodic, observe that mixing implies $\mu(A) = \mu(A)^2$ when $T^{-1}A = A$. To check that our T is mixing, note that if A is a cylinder of length m and B is a cylinder of length n, then $\mu(T^{-N}A) \cap \mu(B) = \mu(A)\mu(B)$ when $N \geq n$.

In general, mixing is a stronger property than ergodicity. For example, if A and B are intervals on the unit circle, T is an irrational rotation, and μ is arclength measure, then $\mu(T^{-N}A \cap B)$ has no limit as $N \to \infty$. The point is that ergodicity is a more general property of systems: see Section 4.1.e of Katok and Hasselblatt for more.