Ergodic theorems

Hannes Gernandt

04.02.2020

This talk is based on the Lecture notes of the 22nd Internet Seminar. The lectures are available online and if anybody is interested in the topic of this talk, one should have a look at these.

1 Introduction

The Ergodenhypothese (ergodic hypothesis) was formulated around 1880 by Boltzmann, where he claimed that for an ideal gas, over time goes through every physically feasible state. This used by Boltzmann to deduce that independently of the initial state the average number of visits of a region in the phase space is proportional to the volume of the region. Rephrased and simplified this means time mean equals space mean. However this was doubted to be true by Lord Kelvin and Poincare and disproven by Plancherel and Rosenthal in 1913. Reformulation as quasi ergodic hypothesis which means that the for almost every initial value this property holds.

This was finally be proven independently by von Neumann (mean ergodic theorem) and by Birkhoff (pointwise ergodic theorem).

Lets do some mathematics:

Definition 1.1. Let (X, Σ, μ) be a probability space and let $T : X \to X$ be a measurable. Then T is called *measure-preserving transformation* iff

$$\mu(T^{-1}(B)) = \mu(B), \quad \text{for all } B \in \Sigma.$$

$A \in \Sigma$ is called T*-invariant* if $A = T^{-1}A$ up to a set of zero measure. We call T *ergodic* if $\mu(A) \in \{0, 1\}$ for each invariant set $A \in \Sigma$.

T-invariance is equivalent to $T^{-1}(A) \subseteq A$

measure theoretic definition of ergodicity

The intuition about this is that X can be thought of the phase space of a physical system and $x \in X$ is a particular state. The transformation T describes the dynamics, i.e. what happens for a given initial state x within one time step. The sequence $x, Tx, T^2x, \dots$ is the path or the phase space motion. In particular we consider only discrete time steps.

Important examples of ergodic transformations are shifts on measure spaces (Y, Σ, μ), then one can construct the product measure space $Y^{\mathbb{N}}$ with the σ-Algebra induced by the cylindrical sets

$$Y \times \dots \times A_1 \times \dots \times A_N \times Y \times \dots$$

and measure $\mu_\times(A) := \prod_{i=1}^N \mu(A_i)$ then the left shift is defined by

$$T(x_n)_{n\in\mathbb{N}} = (x_{n+1})_{n\in\mathbb{N}}.$$

One can show that this is indeed ergodic, see Proposition 3.17 (Lecture 3 Internet Seminar).

We formulate the von-Neumann ergodic theorem.

Theorem 1.2. *For a probability space (X, Σ, μ) consider the Hilbert space $L^2(X, \mu)$. Let $T : X \to X$ be measure-preserving and P the orthogonal projection onto the fixed space*

Note that F is closed

$$F := \{f \in L^2(X, \mu) : f \circ T = f\}.$$

Then for each $f \in L^2(X, \mu)$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f \circ T^n = Pf, \tag{1}$$

where the convergence is in L^2-norm.

The averages on the left hand side are also called *Cesàro means*

Assumption: F contains only constant functions then the right hand side is equal to $\int_X f d\mu 1(\cdot)$. Therefore we have time mean equals space mean. Maybe for very "wild" dynamics, as I (as a mathematician) would expect it for an ideal gas, it might be the case that F contains only constant functions. Later in the application, we have that T is the left shift and then the only invariant elements are constant functions or sequences.

In the following, we will prove the von Neumann mean ergodic theorem using a little bit operator theory. Since $T : X \to X$ is only measure preserving and measurable it might be nonlinear, therefore we want to associate a suitable linear operator S_T on the measurable functions by

$$f \mapsto S_T f := f \circ T$$

This operator is called *Koopman-operator* in the literature. Clearly $S_T f$ is again a measurable function, it is linear and we can rewrite $f \circ T^n = S_T^n f$. Moreover, since T is measure-preserving, we have for all measurable A

$$\int_A f d\mu = \int_A f \circ T d\mu = \int_A S_T d\mu$$

and therefore S_T is an isometry between L^p for all $p \in [1, \infty]$ (split up in positive and negative part) and it is multiplicative and the constant functions are fixed points.

In the following, we will prove the von Neumann theorem in Hilbert-space in a more general form. As we have seen above the Koopman operator is in this case a contraction between L^2.

2 Von Neumann's and Birkhoff's theorems

We translate the measure preserving property in the language of operator theory: We consider contractions $S : H \to H$ in Hilbert space H which satisfy $\|Sx\| \leq \|x\|$ for all $x \in H$, i.e. $\|S\| \leq 1$. Recall that $S^* : H \to H$ is the Hilbert space adjoint of S which is uniquely given by $(Sx, y) = (x, S^*y)$ for all $x, y \in H$. Recall that a subspace M of a Hilbert space H is S-invariant if $SM \subseteq M$ and it is called S-reducing if it is S-invariant and S^*-invariant. Equivalent to the S^*-invariance is that $SM^\perp \subseteq M^\perp$.

also Banach spaces are possible!

Below we derive a decomposition of H into S-reducing subspaces, which will be of importance in the proof of the ergodic theorem.

Proposition 2.1. *The aim will be to derive the von Neumann decomposition for contractions*

$$H = \mathrm{Fix}\,(S) \oplus \overline{\mathrm{ran}\,(I - S)} \tag{2}$$

into closed and reducing subspaces holds, where

$$\mathrm{Fix}\,(S) := \{x \in H : Sx = x\}.$$

Proof. We prove $\mathrm{Fix}\,(S) = \mathrm{Fix}\,(S^*)$. We show that $Sx = x$ is equivalent to $S^*x = x$ is equivalent to $(Sx, x)_H = \|x\|^2$ for some x. Clearly both of the first conditions imply the third condition. Furthermore, we have from the inner product and the third condition and contraction

$$\|Sx - x\|^2 = \|Sx\|^2 - 2\mathrm{Re}\,(Sx, x)_H + \|x\|^2 = \|Sx\|^2 - \|x\|^2 \leq 0$$

which proves $Sx = x$. Rewriting the third condition with adjoint and using that it is a real number, we obtain the second condition. Therefore $\mathrm{Fix}\,(S)$ is not only S-invariant, but also S^*-invariant. The subspace $\mathrm{Fix}\,(S)$ is also closed and hence by definition reducing and so is the orthogonal complement (if we use the characterization that $SM^\perp \subseteq M^\perp$). Next, we use

$$\begin{aligned} H = \mathrm{Fix}\,(S) \oplus \mathrm{Fix}\,(S)^\perp &= \mathrm{Fix}\,(S) \oplus \mathrm{Fix}\,(S^*)^\perp \\ &= \mathrm{Fix}\,(S) \oplus \ker(I - S^*)^\perp \\ &= \mathrm{Fix}\,(S) \oplus \overline{\mathrm{ran}\,(I - S)}, \end{aligned}$$

where we used the kernel image formula. This proves the von Neumann decomposition. ☐

We formulate a version of von Neumann's theorem in terms of contractions.

Theorem 2.2. *Let S be a contraction in a Hilbert space H then for each $x \in H$*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} S^n x = P_{\mathrm{Fix}\,(S)} x, \tag{3}$$

where $P_{\mathrm{Fix}\,(S)}$ is the orthogonal projection onto the (closed) fixed space $\mathrm{Fix}\,(S)$.

The result will essentially follow from the proposition below.

Lemma 2.3. *Let S be a contraction on a Hilbert space then $x \perp \operatorname{Fix}(S)$, i.e. $x \in \overline{\operatorname{ran}(I-S)}$ if and only if*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} S^n x = 0. \tag{4}$$

Proof. $\Rightarrow$: For all $N \in \mathbb{N}$ we have

$$(I-S)\sum_{n=0}^{N-1} S^n = \sum_{n=0}^{N-1} S^n(I-S) = \sum_{n=0}^{N-1} S^n - S^{n+1} = I - S^N. \tag{5}$$

Since $\|S^N\| \leq 1$ for all $N \in \mathbb{N}$, we have for $x = (I-S)y$ for some $y \in H$ that

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} S^n \right\| \leq \frac{2\|y\|}{N}$$

which implies (4). The same can be concluded for $x \in \overline{\operatorname{ran}(I-S)}$, by taking a sequence $x_k \to x$ as $k \to \infty$. Let $\varepsilon > 0$ be given then $\|x - x_k\| \leq \frac{\varepsilon}{2}$ for all sufficiently large k and

$$\begin{aligned}\left\| \frac{1}{N} \sum_{n=0}^{N-1} S^n x \right\| &= \left\| \frac{1}{N} \sum_{n=0}^{N-1} S^n (x - x_k) + \frac{1}{N} \sum_{n=0}^{N-1} S^n x_k \right\| \\ &\leq \left\| \frac{1}{N} \sum_{n=0}^{N-1} S^n \right\| \|(x - x_k)\| + \left\| \frac{1}{N} \sum_{n=0}^{N-1} S^n x_k \right\| \\ &\leq \frac{\varepsilon}{2} + \frac{2\|y_k\|}{N},\end{aligned}$$

where we used that the Cesàro-mean is again a contraction. By choosing N sufficiently large, we obtain $\|y_k\|/N \leq \frac{\varepsilon}{2}$. This proves the convergence of the norm on the left hand side to zero as $N \to \infty$.

$\Leftarrow$: From (5) applied for all $n = 0, \dots N-1$ we have

Thanks to Dymitro Strelnikov for this remark!

$$I - \frac{1}{N}\sum_{n=0}^{N-1} S^n = \frac{1}{N}\sum_{n=0}^{N-1} I - S^n = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{j=0}^{n-1} S^j(I-S) = (I-S)\frac{1}{N}\sum_{n=0}^{N-1}\sum_{j=0}^{n-1} S^j.$$

Therefore, for all $x \in H$ satisfying (4) we obtain from the pointwise limit in the above equation that

$$x = \lim_{N\to\infty} x - \frac{1}{N}\sum_{n=0}^{N-1} S^n x = (I-S)\frac{1}{N}\sum_{n=0}^{N-1}\sum_{j=0}^{n-1} S^j$$

and therefore $x \in \overline{\operatorname{ran}(I-S)}$. □

We continue with the proof of Theorem 2.2.

Proof. The convergence in (3) for all $x \in \mathrm{Fix}\,(S)$ is trivial. By the (reducing) von Neumann decomposition (2) and linearity, it remains to prove (3) for all $x \in \overline{\mathrm{ran}\,(I-S)}$ with limit zero. But this is precisely Lemma 2.3. □

As a consequence, we can apply Theorem 2.2 to the Koopman operator S_T of a measure preserving system and obtain the mean ergodic theorem from the introduction.

We present now the ergodic theorem of Birkhoff which looks slightly different, as it provides pointwise convergence of the averaged paths for almost every initial value.

Theorem 2.4. *Let (X, T, μ) be a measure-preserving system and let $f \in L^1(X, \mu)$. Then the limit*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^n x)$$

exists for almost every $x \in X$.

Moreover the system is ergodic if and only if the above limit equals $\int_X f d\mu$.

Proof. The idea of the proof is that $D := \mathrm{Fix}\,(S) \oplus (I-S)L^\infty(X, \mu)$ is dense in $L^2(X, \mu)$ (by von Neumann decomposition) and hence in $L^1(X, \mu)$. Moreover for all D die pointwise-limit exists almost everywhere. It remains to prove the closedness of the set of functions for which the limit exists almost everywhere (since it contains the dense subset D). A detailed proof can be found in Lecture 7 of the Internet seminar. □

The original formulation of Birkhoff was for the time continuous case.

If the fixed space contains only constant functions then we obtain with $f = 1_A$ and von Neumann's theorem

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} 1_A(T^n x) = \int_X 1_A d\mu = \mu(A),$$

for almost every $x \in X$. This means that time mean equals space mean and thus, we have proven the quasi-ergodic hypothesis for ergodic systems and almost every initial value.

3 An Application to number theory

We apply the pointwise ergodic theorem to prove a result in number theory. For another possible application, the proof of Kolmogorov's law of large numbers, we refer to Lecture 9 of the Internet seminar.

In 1909, Borel introduced the following property of real numbers.

Definition 3.1. We say that a number x is normal in base b if for every finite word $S = a_1 \dots a_d \in \{0, \dots, b-1\}^d$ the number of appearances of S among the first n positions $N_b(S, n)$ satisfies

$$\lim_{n\to\infty} \frac{N_b(S, n)}{n} = \frac{1}{b^d}.$$

Furthermore, x is normal, if it is normal with respect to all bases $b \geq 2$.

It is not known whether $\sqrt{2}$, e or π are normal, however there is the following result.

Theorem 3.2 (Borel)**.** *Lebesgue almost every number is normal.*

For the proof involves the pointwise ergodic theorem. We need some technical lemma

Lemma 3.3. *Let (X, μ), (Y, ν) be probability spaces and $\theta : X \to Y$ measure preserving. Assume that $(g_n)_n$ is a sequence of measurable functions in Y which converges ν-almost everywhere to a measurable function g. Then $S_\theta g_n$ converges μ-almost everywhere, to $S_\theta g$.*

We to do the obvious extension of "measure-preserving" for two different measure spaces

Proof. The definition of the Koopman operator implies

$$S_\theta \sup_n g_n = \sup_n S_\theta g_n, \quad S_\theta \inf_n g_n = \inf_n S_\theta g_n,$$

hence for lim sup and lim inf and therefore the limits coincide. □

Recall S_θ is the Koopman operator

Proof. (of Theorem 3.2 Since the set of bases b is countable, it is enough to prove the result for a fixed basis b (countable union of zero sets is again a zero set). We have to show that every fixed finite word $S = a_1 \ldots a_d$ has an asymptotic occurrence of $(1/b)^d$ for almost all $y \in [0, 1]$. Recall that the Bernoulli shift $T = B(1/b, \ldots, 1/b)$ (i.e. uniform distribution) in $X = \{0, \ldots, b-1\}^{\mathbb{N}}$ is ergodic with respect to the product measure. If we apply the pointwise ergodic theorem to the cylinder set

$$A = \{(x_1, x_2, \ldots) \,:\, x_1 = a_1, \ldots, x_d = a_d\},$$

then we obtain

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} 1_A(T^n x) = \mu(A) = \frac{1}{b^d}. \tag{6}$$

The expression on the left hand side is the number of appearances of the word S in the first N positions of the sequence $(x_n)_n$ divided by N. Now we have to relate the sequence $(x_n)_n$ with elements in $Y = [0, 1]$ via the measure preserving map

$$\theta : X \to Y, (x_n)_n \mapsto y = \sum_{n=0}^{\infty} \frac{x_n}{b^n}.$$

The convergence (6) translates by the Lemma 3.3 to Y. □

To see that it is measure preserving I think, we need the product space metric $\sum_{i=0}^{\infty} b^{-i} d(x_n, y_n)$

There are various more applications of ergodic theorem's. They play an important role in the proof of the Green-Tao theorem that the prime numbers contain arbitrarily long arithmetic progressions.