Pointwise Ergodic Theorem

Alexander dca877

May 2nd, 2020

1 Overview

We'll be introducing the subject of measure preserving dynamical systems which is a core component of ergodic theory and providing interesting results concerning many of the involved components. This should be understandable by anyone who has learned some measure theory.

2 Measure Preserving Dynamical Systems

A dynamical system is some space X along with a transformation function

$$T : X \to X.$$

We think of the action of T on X as representing the passage of time. As often comes up in application, we impose extra structure on X and T that X is a measure space and T preserves the measure.

Essentially, we have $(X, \mathscr{B}, \mu)$ where $\mu : \mathscr{B} \to \mathbb{R}_{\geqslant 0}$ assigns a notion of "size" to an element of $\mathscr{B}$ which is a family of subsets of X called the σ-algebra of events of X. We can interpret $\mathscr{B}$ as the events one can find the probability of and μ as a way of assigning relative probability to an event.

I find it helpful to consider X the space of configurations of a physical system, instead of as the physical location of a particle. From Hamiltonian Mechanics, the position of a particle in phase space completely determines the dynamics whereas the location alone is often not enough. The same principle applies here where X represents the phase space (the space whose coordinates are the positions in a physical system and the momenta of the system).

Often to correspond with the case where μ in fact does represent a probability of events, we let $\mu(X) = 1$.

To make this a dynamical system, we take the measure space along with the transformation law T to get $(X, \mathscr{B}, \mu, T)$ and in the case we often care about, for all $B \in \mathscr{B}$, $\mu(B) = \mu(T(B))$. Additionally we assume that T is an invertible transformation. If this is so, we call this a measure preserving dynamical system, since T takes sets events to events of the same size.

3 Examples of Measure Preserving Systems

Consider the circle S^1 with $d\mu = \frac{1}{2\pi}|d\theta|$. We often specify the infinitesimal change in the measure μ with respect to the coordinates of the system for ease of notation. With $d\mu$ specified, we define $\mu(B) = \int_B d\mu$ for B in $\mathscr{B}$. One can then check that $\mu(S^1) = 1$. Hence μ is a probability measure.

We can then define a transformation as $T(\theta) = \theta + 2\pi\alpha \mod 2\pi$ where α is a fixed real number in $[0, 1]$. We then easily see that T preserves μ and hence $(S^1, \mathscr{B}, \mu, T)$ is a measure preserving dynamical system. This system corresponds to rotating the circle counterclockwise by $2\pi\alpha$ radians.

An example of a system whose measure is not a probability space is the configuration space determined by a simple harmonic oscillator in one dimension with unit mass and force constant. The Hamiltonian for this system is $H(q, p) = \frac{1}{2}p^2 + \frac{1}{2}q^2$. Then the equations of motion for this system are $\dot{q} = p$ and $\dot{p} = -q$. Let our space be $\mathbb{R}^2$ where the first coordinate is our position and the second coordinate is momentum. Then in fact, we can associate a T for this system to take a pair (p, q) to a new pair (p', q') where we get this new pair by rotating (p, q) by a 1 radian in the clockwise direction. More interestingly, we can get a measure for this system where $d\mu = dxdy$. This measure is preserved by Liouville's Theorem in classical dynamics. This example exemplifies the idea that measure preserving dynamical systems need not be over a space of finite measure. However, in the below results we'll restrict ourselves to spaces of finite measure.

4 Functions in $L^1(X)$

We say that integrable, real valued functions of X are in $L^1(X)$. We interpret these functions as a physical measurement that an observer can make regarding a point along trajectory (a trajectory is essentially the "path" a point makes under evolution corresponding to T).

We wish to signal out a specific type of $L^1(X)$ function immediately: the invariant function. These are function f such that f is constant along trajectories. I.e. $f(T(x)) = f(x)$. If we view T as representing time evolution, then invariant functions are measurements that are stationary across time. We note that we only require this holds for almost every $x \in X$.

5 Pointwise Ergodic Theorem Introduction

We want to show a result regarding averages across a trajectory. We define a partial average

$$A_N(f)(x) = \frac{1}{N}\sum_{n=0}^{N-1} f(T^n(x)).$$

The question is, when does the partial averages approach, in a limiting sense, an average over the entire trajectory? I.e. when does the below $A(f)$ converge?

Let

$$A(f)(x) = \lim_{N\to\infty} A_N(f)(x)$$

.

If f is an invariant function, $fT = f$ hence we have $A_N(f)(x) = f(x)$ for almost every x. Then in this case, $A(f)$ exists at almost every point in X.

Another interesting type of function whose behavior under A_N we can completely analyze is the "coboundary" $f = gT - g$ where $g \in L^\infty(X)$ where $L^\infty(X)$ is the class of bounded functions on X. We can check that f is in $L^1(X)$ by computing its norm.

Let M be a bound on g. Then,

$$\begin{aligned} ||f||_1 = \int_X |f(x)|d\mu &\leqslant \int_X |gT(x)|d\mu + \int_X |g(x)|d\mu \\ &= \int_X |g(x)|d\mu + \int_X |g(x)|d\mu \\ &= 2M\int_X d\mu \\ &= 2M\mu(\mathrm{X}) < \infty \end{aligned}$$

In the above we used the triangle inequality and that T is a measure preserving transformation which allows us to equate $\int |gT(x)|$ with $\int |g(x)|$. This is a form of change of variables.

Returning to the problem of averaging, $A_N(f)$ becomes a telescoping sum equal to $A_N(f)(x) = \frac{gT^{N+1}(x)-g(x)}{N}$. Since g is bounded by M, $|A_N(f)(x)| \leqslant 2M/N$ which tends to 0 as N approaches ∞. Hence, $A(f)(x) = 0$.

We were able to provide a sufficient condition for the existence of averages in the case of invariant functions and the difference of L^∞ functions.

Before answering the question of when the average exists in full, we'll provide a rough interpretation for the class of "coboundary" functions. If we view a L^∞ function g as a bounded measurement on a space, we can interpret the value of $f = gT - g$ as denoting the amount of change in the measurement over one time interval but since the measurement is bounded, the change is bounded the average change over n time intervals is the change over n time intervals divided by n but the numerator is bounded and hence this quantity approaches zero.

6 Existence of Averages Proof

The broad idea for this proof is to show that the set containing sums of L^1 invariant functions and coboundary functions is dense in L^1 and that the set of functions where this theorem holds is closed thus since the set of functions where the theorem holds is closed and contains a dense set, the theorem must hold over all of L^1.

Let $\mathcal{D} = \{f + gT - g | f \in L^1, f \text{ invariant}, g \in L^\infty\}$. We'll show $\mathcal{D}$ is dense in L^1.

First we note that $||h||_1 \leqslant ||h||_2 * \mu(X)^{1/2}$. This is simply the Cauchy-Schwarz inequality applied to $|h|$ and 1. Since we're working in a space of finite measure, this means that $||h||_1 \leqslant c||h||_2$ where c is a constant. Hence, L^2 functions are L^1. We'll show the restriction of $\mathcal{D}$ to L^2 is dense in L^2.

Let $\mathcal{P} = \{gT - g | g \in L^\infty\}$. $\overline{\mathcal{P}}$ denotes the closure of $\mathcal{P}$ in L^2. By a similar argument to showing that $gT - g$ is in L^1, $\mathcal{P}$ is in L^2.

Let $\mathcal{I} = \{f | f \in L^2, f \text{ invariant}\}$.

We'll show that $L^2 = \overline{\mathcal{P}} \oplus \mathcal{I}$. I.e. every function in L^2 is uniquely a sum of an invariant function and a function that is arbitrarily close to a coboundary. It suffices to compute the set of L^2 functions that are orthogonal to every element of $\mathcal{P}$ and then since the inner product is continuous (since L^2 is a Hilbert space) this set is orthogonal to every element of $\overline{\mathcal{P}}$. If f is invariant, $(f, gT - g) = (f, gT) - (f, g)$. But, $(f, gT) = \int_X \overline{f(x)} gT(x) d\mu = \int_X \overline{fT^{-1}(x)} g(x) d\mu = (fT^{-1}, g) = (f, g)$. Hence, $(f, gT - g) = 0$. Hence $\mathcal{I}$ lies in the space of orthogonal function to $\mathcal{P}$. To show the converse, suppose f is orthogonal to every element of $\mathcal{P}$. Then, $(f, gT - g) = 0$. Thus, $(f, gT) = (f, g)$. By a similar arguement to before, $(f, gT) = (fT^{-1}, g)$. We're left with the conclusion, $(f, g) = (fT^{-1}, g)$ and $(f - fT^{-1}, g) = 0$. Let g be the characteristic function of the set where $f(x) - fT^{-1}(x) > 0$. This set is measurable since we presume f is and this g is bounded by 1. Plugging in, $0 = (f - fT^{-1}, g) = \int_{\{f - fT^{-1} > 0\}} (f(x) - fT^{-1}(x)) d\mu$. But this is then the integral of a positive quantitiy, and thus the integral is non-negative. But the integral is 0, hence the integrand is positive over a set of measure zero. Similarly, for the reverse inequality, the integrand is negative over a set of measure zero. Combined, $f(x) = fT^{-1}(x)$ except on a set of measure zero. Hence f is invariant.

Thus the orthogonal space to $\mathcal{P}$ is precisely the invariant functions. Thus we see our direct sum decomposition of L^2 as $\overline{\mathcal{P}} \oplus \mathcal{I}$.

Hence the set $\{f + gT - g | f \in L^2, fT = f, g \in L^\infty\}$ is dense in L^2.

Since $\mathcal{D} \cap L^2$ is dense in L^2, and since simple functions are dense in both L^1 and L^2 (in the corresponding metrics), to show $\mathcal{D}$ is dense in L^1 we fix an arbitrary $f \in L^1$ and an arbitrary $\epsilon > 0$.. Then we can find a simple function $s \in L^1$ such that $||f - s||_1 < \epsilon$. But all simple functions in L^1 are bounded L^∞ and hence in L^2. Since $s \in L^2$, we can find an $h \in \mathcal{D}$ such that $||s - h||_2 < \epsilon$. But by our above use of Cauchy-Schwarz, $h \in L^1$ and $||s - h||_1 \leqslant c||s - h||_2 < c\epsilon$. Combing these, $||f - h||_1 \leqslant ||f - s||_1 + ||s - h||_1 < (1 + c)\epsilon$. Since ϵ was arbitrary, we establish the desired estimate.

Hence, every L^1 function is arbitrarily close to the sum of an invariant function and a coboundary.

Let's pause for a moment to interpret this result in a kinda rough manner:

every measurement on a dynamical system (an L^1 function) is arbitrarily close to a measurement that's constant along trajectories and a measurement that resembles a bounded perturbation or error from the function being invariant that goes to zero in the average.

Now we need to show that the subset of L^1 where A_N converges is closed (since we showed that A_N converges over a dense set of L^1, if we know convergence holds over a closed set, then we're done).

Suppose $f_n \to f$ in L^1 and for each f_n, $A_N(f_n)$ converges. Since for each n, $A_N(f)$ converges, we can choose N sufficiently large to make $\|A_N(f_n) - A_M(f_n)\|_1 < \epsilon$

We note that A_N is linear. We have the result (presented here without proof, that $\mu(\{\sup_N |A_N(g)| > \alpha\}) \leqslant \frac{1}{\alpha}\|g\|_1$. This is the maximal ergodic theorem in a weaker form.

We want to find the measure of the set
$E = \{\limsup_{N\to\infty, M\to\infty} |A_N(f)(x) - A_M(f)(x)| > 0\}$. If $\mu(E) = 0$ then for almost every x, $A_N(f)(x)$ is a Cauchy sequence of real numbers and hence $A_N(f)(x)$ is convergent and the average is defined. Let $E_\alpha = \{\limsup_{N\to\infty, M\to\infty} |A_N(f)(x) - A_M(f)(x)| > \alpha\}$

$|A_N(f) - A_M(f)| \leqslant |A_N(f) - A_N(f_n)| + |A_N(f_n) - A_M(f_n)| + |A_M(f_n) - A_M(f)|$.

Since we assume that on f_n the average exists, the middle term tends to 0 as N, M approach ∞.

Hence $E_\alpha \subset \{\limsup_{N\to\infty, M\to\infty} |A_N(f) - A_N(f_n)| + |A_M(f_n) - A_M(f)| > \alpha\}$. For this inequality to hold, one of the summands must be greater than $\frac{\alpha}{2}$. Hence $E_\alpha \subset \{\limsup_{N\to\infty} |A_N(f) - A_N(f_n)| > \frac{\alpha}{2}\}$. We can then use the maximal ergodic theorem to estimate the size of this set.

$$\mu(E_\alpha) \leqslant \mu(\{\sup_N |A_N(f - f_n)| > \frac{\alpha}{2}\}) \leqslant \frac{2}{\alpha}\|f - f_n\|_1$$

But $f_n \to f$ in L^1. Hence we can give $\mu(E_\alpha)$ an arbitrarily small upper bound. Hence $\mu(E_\alpha) = 0$ for every α.

Since $E_{1/n} \nearrow E$, $\mu(E) = 0$. Hence the average exists for f almost everywhere. This shows that the set of L^1 functions with well defined averages is closed. Since this set contains a set that is dense in L^1, this set is necessarily all of L^1.

Thus we've established the result that the average of any L^1 function exists at almost every point of space.

7 Towards the Pointwise Ergodic Theorem

So now we know that $A(f)(x)$ is well defined at almost every $x \in X$. We want to connect this average, which is essentially a time average, to an average over the space X. First we note, $A_N(f)(Tx) = \frac{1}{N}(f(x)) + \frac{N+1}{N}A_{N+1}(f)(x)$.

If x is such a point that the average of at x and Tx exists (which is all X except the union of two measure zero sets), $A(f)(Tx) = A(f)(x)\lim_{N\to\infty}\frac{N+1}{N} + \lim_{N\to\infty}\frac{f(x)}{N} = A(f)(x)$. Hence $A(f)$ is an invariant function except on a set of measure zero. This has the interpretation that averaging along trajectories essentially does not depend on the initial measurement(s).

We can further show that $A(f)$ is in L^1 by noting that $A_N(f) \leqslant A_N(|f|)$, hence $A(f) \leqslant A(|f|)$ and

$$\begin{aligned}
\|A(f)\|_1 \leqslant \|A(|f|)\|_1 &= \int_X \lim \frac{1}{N}\sum_{n=0}^{N-1} |f|T^n(x)d\mu(x) \\
&= \lim \frac{1}{N}\sum_{n=0}^{N-1}\int_X |f|T^n(x)d\mu(x) \\
&= \lim \frac{1}{N}\sum_{n=0}^{N-1}\int_X |f|(x)d\mu(x) \\
&= \lim \frac{1}{N}\sum_{n=0}^{N-1}\|f\|_1 \\
&= \|f\|_1
\end{aligned}$$

where the second equality follows by noting that since the limit converges almost everywhere and is non-negative, we can apply Fatou's Lemma noting that almost everywhere convergence implies the limit is the limit infimum. The third equality follows by changing variables.

Hence $A(f)$ is defined almost everywhere and L^1 and constant along trajectories. That's about as much as we can say at the moment.

Now, we can impose the further restriction on our dynamical system that all invariant functions must be constant almost everywhere. This restriction makes the system ergodic (this is taken as a definition). An interpretation for this condition that is one could theoretically use invariant functions to separate orbits, but an ergodic dynamical system essentially cannot do that since every orbit "explores the whole space".

Hence if our system is ergodic, since $A(f)$ is invariant, $A(f)$ must be constant.

If we pass to the Lebesgue decomposition of f, $f = f_+ - f_-$, noting that $A_N(f) = A_N(f_+) - A_N(f_-)$, so $A(f) = A(f_+) - A(f_-)$ almost everywhere. Working with just $A(f_+)$, by a similar calculation to showing $A(f)$ was L^1, we see $\int_X A(f_+)(x)d\mu(x) = \|f_+\|_1 = \int_X f_+ d\mu$ and similarly for $A(f_-)$. By linearity, $\int_X A(f)d\mu = \|f_+\|_1 - \|f_-\|_1 = \int_X (f_+ - f_-)d\mu = \int_X f d\mu$. But $A(f)$ is constant here, so $\int_X A(f)d\mu = \overline{f}\int_X d\mu = \overline{f}$ where $\overline{f}$ denotes the almost everywhere constant value of $A(f)$ and we take μ to be a probability measure so $\mu(X) = \int_X d\mu = 1$.

So we see, $\overline{f} = \int_X f(x)d\mu(x)$.

Hence we see

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} fT^n(x) = \int_X f d\mu$$

for almost every $x \in X$. This is the pointwise ergodic theorem and it states that the time average over an orbit converge almost everywhere to the average over the entire space. In fact, that this theorem holds for a dynamical system is equivalent to the system being ergodic (to see this consider an invariant function f, then the ergodic theorem says that $f(x)$ is constant for almost every x (and in fact equals the space average), thus reproducing our original definition).

8 Closing Summary

We see a powerful structural theorem here, that the average operator is in fact well defined over almost every point and that if the system is ergodic, the average over a trajectory ("in time") in fact equals the space average, justifying the interpretation of an ergodic dynamical system as a system where almost every trajectory explores all of the configuration space.