

Tune-Up: Ergodic Theorem

Dynamical systems are difficult to discuss because they are literally describe “everything”. Anything that “happens” is an orbit in a dynamical system. Turning those observations into a meaningful discussion is another matter.¹ A physical system might display randomness yet be quite decisive. The elements of chance could emerge through specific determined behavior. E.g. a baseball game, water escaping a fire hydrant, cars driving on a street. Or more theoretical, E.g. gas particles moving in a box, the billiard ball flow. We might try to ask about *periodic* behavior in these circumstances.

So we need to be able to choose one example or a class of examples as a way to read the text. Or we would study “generic” behavior. Let’s see on example:

Ergodic theory in its broadest sense is the study of group actions on measure spaces. Historically the discipline has tended to concentrate on the framework of integer actions, in line with its formal origins in the work of John von Neumann and George David Birkhoff in the early 1930s. To a considerable extent this continues to hold today, not least due to a variety of deep interactions with number theory and smooth dynamics.

Even though the dynamical systems we study are physical, quite general, they have no obvious arithmetical properties. Spaces like \mathbb{R}^2 or \mathbb{H} are what happen *after* uniformation, so this is after we try to force Nature to behave for us. So there is obviously going to be a lot of noise between the generic models we choose and nature.

It also looks like computer programs could be a source of examples for us. What is a computer anyway? There is some kind of memory storage – a tray that holds things – and some computing process – a dynamical system and then a result – something on that tray which we take out and move someplace else. So, in Von Neumann’s time we had a big question *what is computation?* This is already being discussed in babbages time yet all cultures have had to develop bookkeeping systems.

Example 5 people arranged in a circle, one person has \$1 and other people have nothing. Then, gives $1/3$ of his savings to the person on his left and $2/3$ of his savings to the person on the right. Show that eventually, everyone ends up with 20 cents.

Example 10 people walk around a house [floor plan], show that no matter how people walk around the house all arrangements are equally likely. This result also doesn’t depend on the shape of the room and how they’re connected. And yet in many technical cases it’s false. Let’s try “a generic random walk in a connected domain is ergodic”. It’s even true, overwhelmingly true.

That’s why these arguments need illustrations, no amount of theory or discussion is going to substitute us going out there and seeing the result forself.

Thm Let (X, \mathcal{B}, μ, T) be a measure-preserving dynamical system, let P_T be the orthogonal projection onto the closed subspace.

$$I = \{g \in L^2_\mu : U_T g = g\} \subseteq L^2_\mu$$

¹Example: one textbook prove the **central limit theorem** as a consequence of Birkhoff’s **ergodic theorem**.

Then for any $f \in L^2_\mu$,

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \xrightarrow{L^2_\mu} P_T f$$

All of this abstract nonsense measures a big zero. It's very important since it offers us a roadmap to the stuff we actually want to say. Ergodicity will be important when $N \gg 1$ such as $N = 10^{30}$ the number of particles in a room or $N = 10^8$ the number of people in a city and we don't have the resources to check every single example and we want to be sure the result is "natural" or "generic" or close to random.

X is our space of possibilities and \mathcal{B} is our space of "measurable" events.

- Just so we can think about why measure theory could be hard, any news article is all about events that our difficult to measure. E.g. what are the odds it will rain next week? It is tempting to gather all the information we have about the weather today and all information about how the weather changes. And since the weather report is usually inaccurate, **next week's weather is non-measurable**.
- We could have conditional probability on some rare event (e.g. "during an earthquake, it is always sunny"). Or we could try to define conditional probability on an algebra \mathcal{B} without specifying any kind of event. So the details of **measure theory** will be very useful.

μ the stationary measure could be hard to calculate. In many cases, we just want a perfectly random thing from a vast set of possibilities. In other case we want the "noise" around the "typical" behavior of a setting.

- In fact, our schedules are not random. Four of us live in this house and two of us always leave in the morning and sometimes I leave home before lunch or early in the morning or a little bit afternoon. Then we return home during rushour except when we go out and get home much later. Therefore, we are almost always here in the evening, and we all sleep here at night. Except for those times when we don't.
- If we don't know the stationary measure μ we will try to *estimate* the stationary measure as the average of a counting measure of some kind. So the ergodic theorem seems to be run on some kind of loop. [counting measure] \rightarrow [stationary measure]. The ergodic theorem shows this might not depend our choice of counting functions.

T is one-step of the process happening. So literally this could be anything. That's why this is such a difficult set-up to use.

- a general example for T is shuffling a deck of cards, and there $N = 52! \approx 10^{70}$ ways of shuffling a deck of cards, how is it possible to be sure we get a perfectly random deck after just 5 or 8 shuffles?

$U_T : L^2_\mu \rightarrow L^2_\mu$ is defined by $U_T(f) = f \circ T$. We have that L^2_μ is a Hilbert space (just an an example of of how odd measures could be, what if the stationary measure were $d\mu = \frac{1}{2}(d\theta + \delta_{\theta=0})$) This is an example of "lifting" from category theory, where since the base thing gets shuffled around, so do the observations we're making about the room.

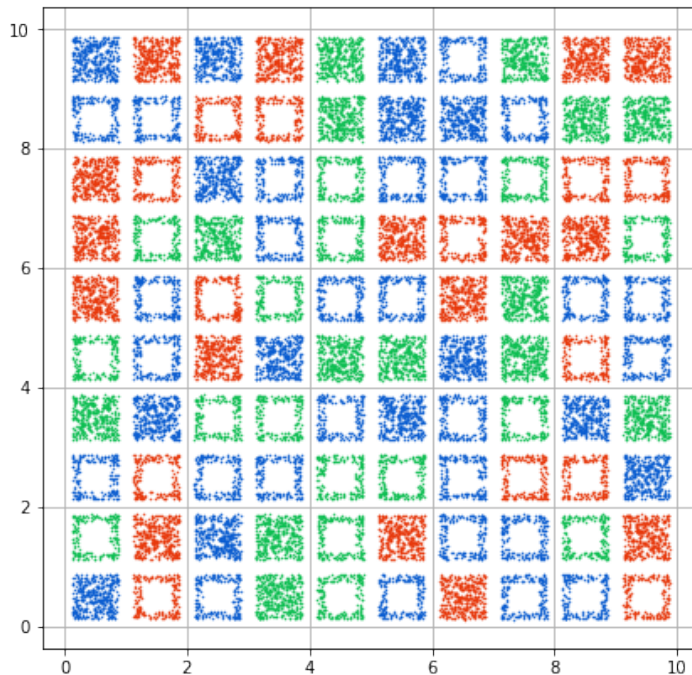
Example U_T is an **isometry**.

$$\begin{aligned} \langle U_T f_1, U_T f_2 \rangle &= \int f_1 \circ T \overline{f_2 \circ T} d\mu \\ &= \int f_1 \overline{f_2} d\mu \\ &= \langle f_1, f_2 \rangle \end{aligned}$$

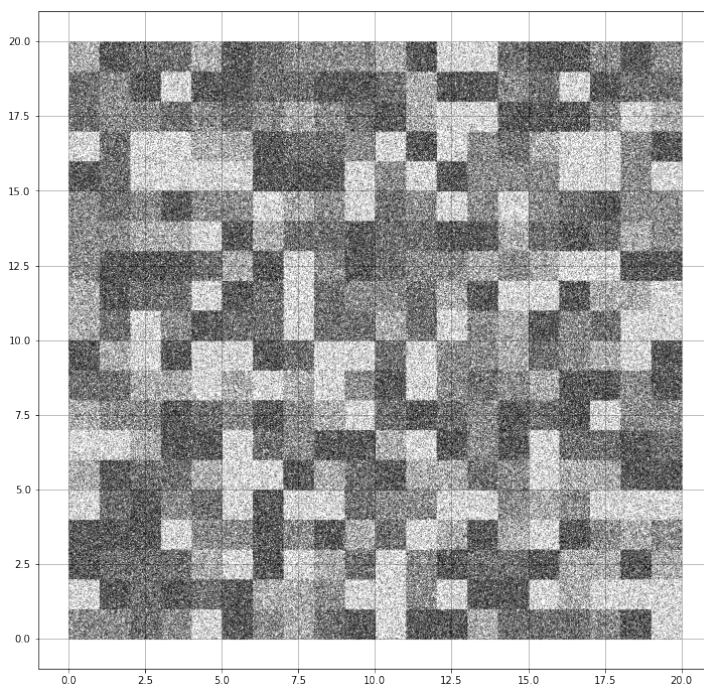
Length-preserving function on this Hilbert space. If we rotate an object, the pictures on the object also get rotated.

So the author of the book as done quite enough, we *need* more illustrations. These were never obvious as soon as we try to turn these example something useful there is an exponential amount of work.

Illustration This looks random in some way but there are lots of patterns, so let's use the word "ergodic". In fact, at each step I just randomly chose a color and randomly filled in a square.



In this example, I randomly filled in a square pattern. The average of this distribution would be a uniform distribution on the square (a constant function at value $f(x, y) = 0.5$) there is clearly a lot more going on.



The proof tells us almost nothing. It is reduced to an exercise in Linear algebra. In Monte-Carlo simulations we are looking for "typical" or "random behaviors" of somewhat hard to describe objects. Features that are easy for humans to distinguish and yet difficult for computers.

Step 1 Let $B = \{U_T g - g | g \in L_\mu^2\}$. We claim that $B^\perp = I = \{g \in L_\mu^2 : U_T g = g\}$.

I take it there's a reciprocity between solving the equation $U_T g = g$ and looking at values of $U_T g - g$ as $g \in L_\mu^2$ varies in the Hilbert space.

If $U_T f = f$ then $f \in B^\perp$. If $f \in B^\perp$ then $U_T^* f = f$. So, $\|U_T f - f\|_2 = 0$ and $f = U_T f$.

Therefore $L_\mu^2 = I \oplus \overline{B}$. Any $f \in L_\mu^2$ splits as $f = P_T f + h$ with $h \in \overline{B}$. We claim that:

$$\frac{1}{N} \sum U_T^n h \xrightarrow{L_\mu^2} 0$$

This would be clear for $h = U_T g - g \in B$, yet all we know is that $h \in \overline{B}$.

Let $g_i \in L_\mu^2$ be a sequence of $h_i = U_T g_i - g_i \rightarrow h$. By the **triangle inequality**:

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 \leq \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n (h - h_i) \right\|_2 + \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h_i \right\|_2$$

We can conveniently find $h_i \approx h$ and once we've found h_i , run the process U_T for sufficiently long so that:

- $\|h - h_i\|_2 < \epsilon$
- $\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h_i \right\|_2 < \epsilon$

Then using the example of triangle inequality we've just found (notice twice the error):

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h_i \right\|_2 < 2 \times \epsilon$$

so

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n h_i \rightarrow 0$$

for any $h \in \overline{B}$ and $N \rightarrow \infty$. □

We've learned nothing. OK, the **Mean Ergodic Theorem** says the "average" of the mixing of a function f by the linear operator $U_T f$ tends to the invariant subspace (even though T itself was highly *non-linear*, in fact "chaotic" or "mixing"), basically a mix of cyclic behaviors. Maybe also try to visualize what convergence in L_μ^2 means, e.g. definitely **not** point-wise convergence.

References

[1] ...