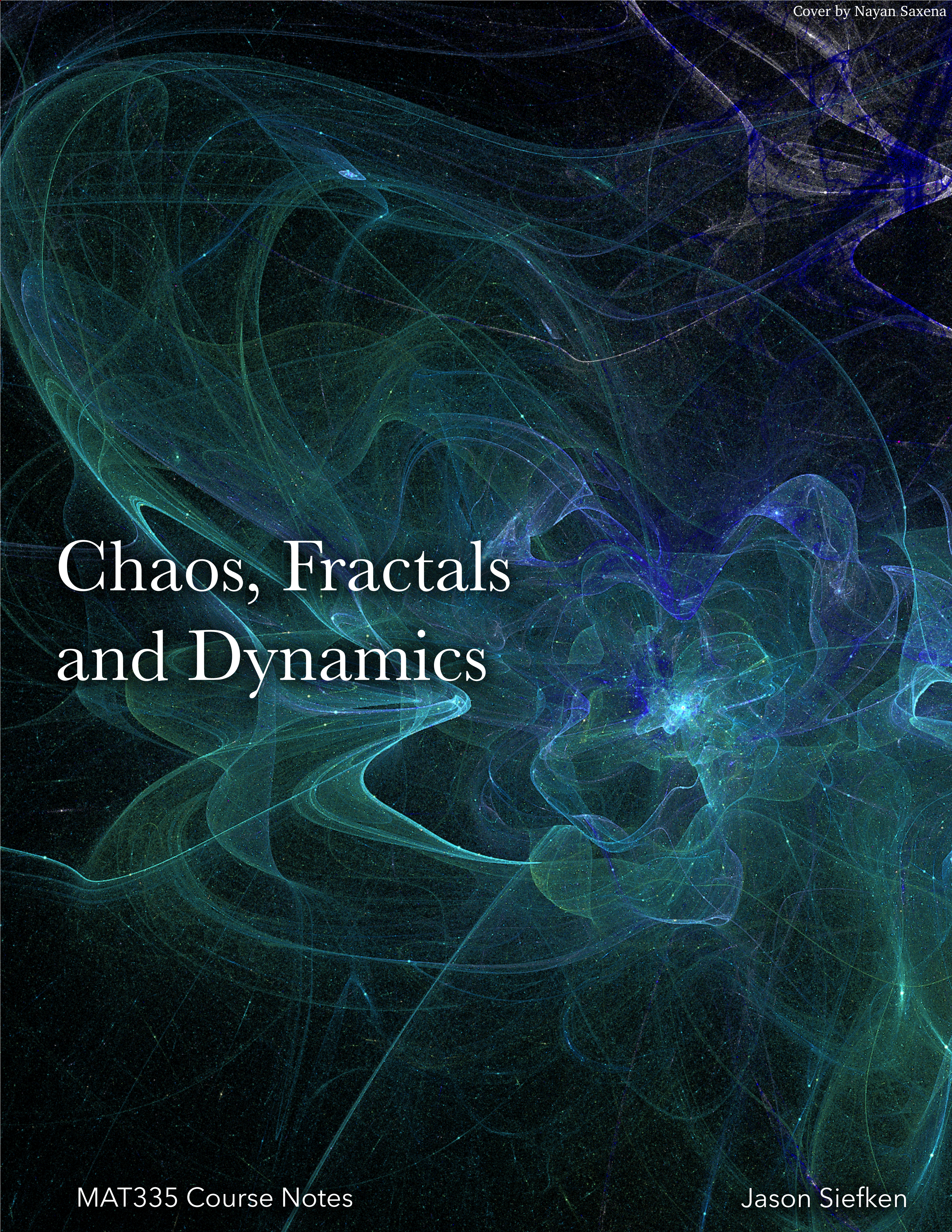


Cover by Nayan Saxena

Chaos, Fractals and Dynamics

MAT335 Course Notes

Jason Siefken

Dynamical Systems

Dynamical System

DEF A pair (T,X) is called a ***dynamical system*** if X is a set and $T: X \to X$ is a function. In the context of a dynamical system, T is often called a ***transformation***.

This definition is very general, and most things you encounter could be considered a dynamical system. For example, if $X = \{\text{air molecules and their positions on earth}\}$ and $T: X \to X$ is the result of the wind blowing for one second, then (T,X) is a dynamical system. Alternatively, we could take the state of your computer's memory (RAM) to be a set and your processor executing a single instruction to be a transformation.

It's hard to say much about general dynamical systems. However, throughout the course, we will find ways to classify dynamical systems. Once we "narrow the field", we'll be able to say lots of interesting things.

Newton's Method

Newton's method is a way of using tangent-line approximations to functions to estimate their roots. It is an iterative procedure.[1]

---

1 Let $f: \mathbb{R} \to \mathbb{R}$ be a differentiable function and let $T_f : \{\text{guesses}\} \to \{\text{guesses}\}$ be a single application of Newton's method.

1.1 Find a general formula for T_f.

1.2 Let $f(x) = x(x-2)(x-3)$. Compute $T_f^n(4)$ for $n = 0,1,2,3$.

1.3 Do you think

$$\lim_{n\to\infty} T_f^n(4)$$

converges? If so, what does it converge to? Can you prove your answer?

Fixed Point

DEF Let (T,X) be a dynamical system. A point $a \in X$ is called a ***fixed point*** if $T(a) = a$.

Basin of Attraction

DEF Let (T,X) be a dynamical system and let $x \in X$. The ***basin of attraction*** of x is the set

$$A_x = \{y \in X : \lim_{n\to\infty} T^n y = x\}.$$

Eventually, we will talk about more general *basins of attraction*, but for now we will limit ourselves to that of a single point.

---

2 Let $f(x) = x(x-2)(x-3)$ and let T_f be the function that applies a single iteration of Newton's method (as before).

2.1 Is $[3,4) \subseteq A_3$ for T_f? What about $[100,1000]$? $(2,3)$?

2.2 Describe A_3.

2.3 Is A_3 connected?

[1] Whenever something is iterative, you should think dynamics!

Inverse Image

DEF

Let $f : A \to B$ be a function and let $X \subseteq B$. The ***inverse image*** of X under f, denoted $f^{-1}(X)$, is

$$f^{-1}(X) = \{x \in A : f(x) \in X\}.$$

Note: a function *need not* be invertible to have inverse images. In fact, the idea of inverse images applies to *every* function.

3

3.1 Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = x^2$. Find $g^{-1}(\{1\})$, $g^{-1}(\{4\})$, $g^{-1}(\{0\})$, $g^{-1}(\{-1\})$, and $g^{-1}([3,4])$.

3.2 Let f and T_f be as before. (Recall, $f(x) = x(x-2)(x-3)$). Find $T_f^{-1}([3,4))$.

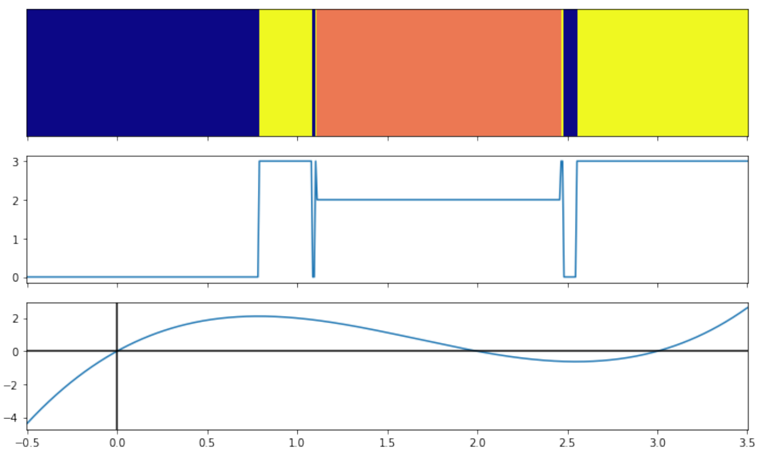
3.3 Define

$$Q = \bigcup_{n \geq 0} T_f^{-n}([3, 3.1))$$

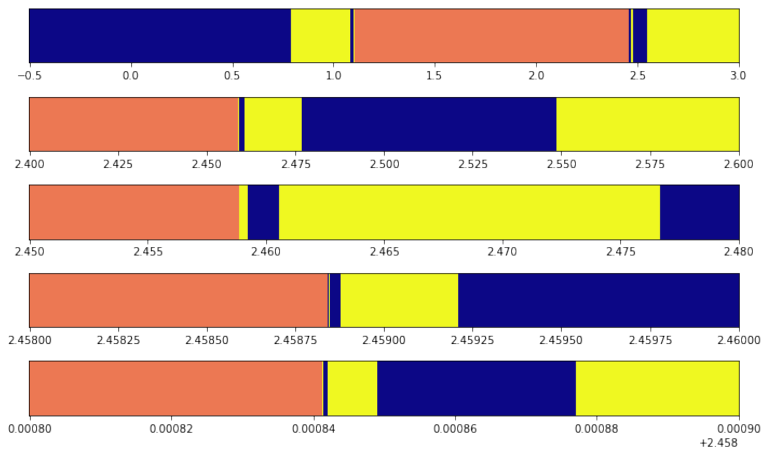
where T_f^0 is the identity function.

Is $Q = A_3$? Why or why not?

Using a computer, we can graph A_0, A_2, and A_3.



Zooming in around $x \approx 2.4588$:



We've just seen our first **fractal**! For now, we will define a *fractal* as a set with repeated patterns at all scales.

Fractals

Let's construct some famous fractals.

4 Let K_0 be an equilateral triangle with sides of length 1. Let K_1 be the result of applying $\to$  to each side of K_0. Repeat this process to get K_2 from K_1, etc. and define

$$K_\infty = \lim_{n\to\infty} K_n.$$

4.1 Draw K_0, K_1, and K_2.

4.2 Find the perimeter of K_0, K_1, and K_2. Find a general formula for the perimeter of K_n.

4.3 What is the perimeter of K_∞? What is the area enclosed by K_∞?

K_∞ is called the *Koch Snowflake*.

5 Let T_0 be a filled-in equilateral triangle. To get T_1, T_2, etc., apply the substitution rule  $\to$  to each (sub)triangle of T_0, T_1, etc.. Define T_∞ to be the limit of this process.

5.1 Draw T_0, T_1, and T_2.

5.2 Find a formula for the area of T_n.

5.3 Compute the area of T_∞.

5.4 Is T_∞ the empty set? Why or why not?

T_∞ is called *Sierpinski's Triangle*.

6 Let $C_0 = [0,1]$ be the unit interval. Recursively define C_i by the substitution rule $\to$, which removes the middle $1/3$ of every interval. Define C_∞ to be the limit of this process.

6.1 Compute the length of C_n.

6.2 Compute the length of C_∞.

C_∞ is called the *Cantor set*.

Our normal sense of measurement fails when it comes to these fractals. We need a new idea: *similarity dimension*.

Dimension

Dimension can be thought of as a relationship between scale and content.[2]

7

7.1 Let $\ell = [0,1)$, $2\ell = [0,2)$, $3\ell = [0,3)$, etc.. How many disjoint copies of ℓ does it take to cover $n\ell$?

7.2 Let $S = [0,1)^2$, $2S = [0,2)^2$, etc.. How many disjoint copies of S does it take to cover nS?

7.3 Let $C = [0,1)^3$, $2C = [0,2)^3$, etc.. How many disjoint copies of C does it take to cover nC?

7.4 Based on the patterns you see, describe an algorithm that can be used to find the dimensions of ℓ, S, and C.

7.5 Let T be the filled-in equilateral triangle. Apply your algorithm to $2T$.

7.6 Let T_∞ be the Sierpinski triangle. Apply your algorithm to $2T_\infty$. Does the number you get make sense?

Similarity Dimension

DEFINITION

A set $Q \subseteq \mathbb{R}^n$ has ***similarity dimension*** d if there exists a $c \in \mathbb{Z}$ and $s \in \mathbb{R}_{\geq 1}$ satisfying

$$d = \log_s(c)$$

and sQ (Q scaled up by a factor of s) is covered by c copies of Q (at most overlapping on their boundaries).

8

8.1 Compute the similarity dimension of (a) a line segment, (b) the Cantor set, and (c) Sierpinski's triangle.

8.2 Compute the similarity dimension of the Koch snowflake.

What about sets that aren't self-similar?

9

Let K'_∞ be the "Koch snowflake" obtained with the substitution rule $\to$.

9.1 Find the perimeter and dimension of K'_∞.

9.2 Let K_{strange} be the "Koch snowflake" obtained by the rule $\to$ or chosen randomly at each stage. What should the dimension of K_{strange} be? Can you compute it's similarity dimension?

[2]Here "content" refers to the "stuff inside" of an object.

We need a way to define dimension for shapes that aren't self-similar. Let's again work from sets whose dimension we know: cubes.

Box Covering

DEFINITION

A d-dimensional ***box covering*** of $X \subseteq \mathbb{R}^k$ is a collection $C = \{B_i\}$ of d-dimensional cubes which satisfy

1. if $i \neq j$, B_i and B_j intersect at most on their boundaries;
2. $B_i \cap X \neq \{\}$ for all i;
3. $X \subseteq \bigcup_i B_i$.

Outer Measure

DEFINITION

The ***d-dimensional outer measure*** of $X \subseteq \mathbb{R}^k$ is

$$\lim_{n\to\infty} \inf_{C_n} \operatorname{volume}(C_n)$$

where C_n is a d-dimensional box covering of X with cubes of side-length $1/n$.

You should think of $\inf_{C_n} \operatorname{volume}(C_n)$ as the "smallest possible box covering of size $1/n$ that still covers the set".

---

10 Let $\ell \subseteq \mathbb{R}^3$ be the line segment from $\vec{0}$ to $(1, 0, 0)$.

10.1 Find the 1, 2, and 3-dimensional outer measures of ℓ.

10.2 Does ℓ have a 0-dimensional outer measure?

10.3 Let $T \subseteq \mathbb{R}^3$ be the filled in triangle with vertices $(0, 0, 0)$, $(1, 0, 0)$, and $(0, 1, 0)$. Find the 1, 2, and 3-dimensional outer measures of T.

10.4 Find the 1-dimensional outer measure of the Cantor set.

What would it mean to have a fractional-dimensional outer measure? Let B be a d-dimensional box with side lengths k. Its volume is k^d. Divide the box in half along every dimension and each sub-box has volume $(k/2)^d = (1/2)^d k^d$, and so there must be 2^d sub-boxes.

What if there were fewer "sub-boxes"?

---

11 Let C_α be the Cantor-like set obtained by removing the middle α of each subinterval. (I.e., the standard Cantor set is $C_{1/3}$.)

11.1 Find the number of boxes in a 1-dimensional box-covering of C_0 (the interval) and $C_{1/3}$ where the width of each box is 1/3, 1/9, 1/27, etc..

11.2 Based on what you know about how many width-k boxes it takes to fill d-dimensional space, find a formula relating d, the number of boxes, and the width of the boxes.

11.3 Use your formula to estimate d for $C_{1/3}$. How does this compare to the similarity-dimension of $C_{1/3}$?

Box-counting Dimension

DEFINITION

Let $X \subseteq \mathbb{R}^m$ and let $B \subseteq \mathbb{R}^m$ and let B be a minimal-dimensional, minimally-sized box such that $X \subseteq B$. The ***box-counting dimension*** of X is

$$d = \lim_{n\to\infty} \frac{\log(\#\text{ of sub-boxes of } B_n \text{ that intersect } X)}{\log n},$$

where B_n is B "cut" along each axis into n equally-spaced slices.

11.4 Find the box-counting dimension of $C_{1/3}$.

11.5 Find the box-counting dimension of the unit simplex in $\mathbb{R}^2$. I.e. $\{\vec{v} \in \mathbb{R}^2 \ : \ \vec{v} = \alpha\vec{e}_1 + \beta\vec{e}_2$ for some $\alpha, \beta \geq 0$ satisfying $\alpha + \beta \leq 1\}$

11.6 Intuitively, what should $\lim_{\alpha\to 0} \dim(C_\alpha)$ be? Find the box-counting and similarity dimension of C_α and verify.

Box-counting dimension is difficult to compute exactly, but it's useful for approximations. Computers are pretty good at counting boxes!

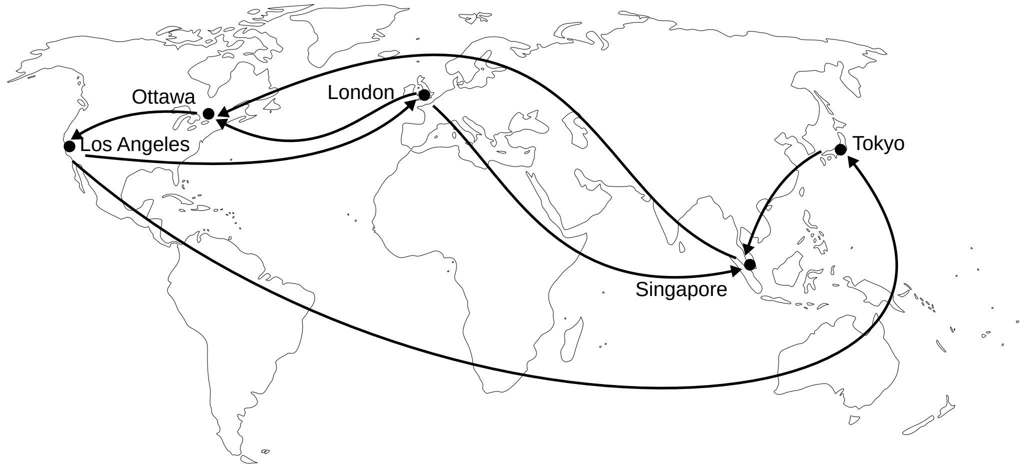
Transition Matrix

DEFINITION

Let $\mathcal{G}$ be a directed graph with vertices $\{1, \ldots, n\}$. A ***transition matrix*** for $\mathcal{G}$ is an $n \times n$ matrix $A = [a_{ij}]$ where

$$a_{ij} = \begin{cases} 1 & \text{if there is an edge from } j \text{ to } i \\ 0 & \text{otherwise} \end{cases}.$$

12 The map shows the direct, one-way flights offered by the Pacific Rim air shipping company.



We can think of their flight map as a graph (in the graph-theory sense).

12.1 Write down a *transition matrix* A for the above graph.

(a) What do the diagonal entries tell you about the available flights?

(b) Should $a_{ij} = a_{ji}$? Explain.

12.2 Write down a matrix $B = [b_{ij}]$ where the entry b_{ij} indicates the number of ways to go from city j to city i using exactly two flights.

(a) Compute A^2 and compare with B.

(b) What information does the 1st row of A give you about flights?

(c) What information does the 2nd column of A give you about flights?

(d) Based upon your last two answers what does the $1, 2$ entry of A^2 tell you about flights?

12.3 Compute A^3. What does it tell you about shipping routes?

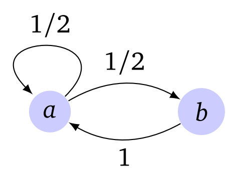
12.4 A package with a lost tracking number is getting kicked around from route to route! Each time it lands, it is randomly (and with equal probability) put on another flight. After several weeks (and 100s of flights), the package is finally noticed. Where is it most likely to be?

Markov Chain

DEF Given a graph G, a ***stationary Markov chain on G***, is a random process, denoted $X_1, X_2, \ldots$ where X_i indicates the location on the graph at the ith step and the probability distribution of X_i is *completely determined* by the value of X_{i-1}.

Some people call stationary Markov chains "memoryless" processes because what happened more than one step prior has no affect on what happens in the next step.

13 Consider the following directed graph with transition probabilities labeled.



13.1 Find a transition matrix, T, for the graph.

13.2 Find a matrix P_1 whose i, j entry represents the probability of transitioning from state j to state i in exactly one step

.

13.3 Find a matrix P_2 whose i, j entry represents the probability of transitioning from state j to state i in exactly two steps.

13.4 How do P_2 and P_1^2 relate?

13.5 Does $\lim_{n\to\infty} P_1^n$ exist? If so, what is it?

Stochastic Matrix

DEF A vector $\vec{v} \in \mathbb{R}^n$ is called a ***probability vector*** if its entries are non-negative and sum to one. A matrix is called a ***stochastic matrix*** if its columns are probability vectors.

In the context of Markov chains, we also refer to stochastic matrices as *transition* matrices.[3]

---

14 You're a picky eater. You have meals of healthy food and dessert, but you never have healthy food twice in a row. Each time you eat dessert, you flip a weighted coin to decide what meal to eat next. Let p represent the weight of the coin.

14.1 Are your eating habits described by a Markov chain? Why or why not?

14.2 Draw a graph representing this situation.

14.3 Your friend visits you for New Years and sees your having a healthy dinner. You lose touch after that, but bump into each other at a restaurant six years later. What type of meal are you most likely to be eating? Does this depend on p?

---

15 15.1 Is a Markov chain a dynamical system? Why or why not?

15.2 Consider the Markov chain with states $\{a, b\}$ given by the transition matrix $\begin{bmatrix} 1/2 & 1 \\ 1/2 & 0 \end{bmatrix}$. Given that you start in state a, give probability vectors indicating the probability of being in a particular state after 0, 1, 2, and 3 steps along the Markov chain.

15.3 Can this Markov chain be *modeled* by a dynamical system? If so, describe such a model.

[3] Yes, this word is doing double duty!

DEFINITION

Realization of a Markov Chain

Given a Markov Chain $\mathcal{M} = (X_1, X_2, \ldots)$ with state space S, a ***realization*** of $\mathcal{M}$ is a sequence of states whose transitions are allowed by the Markov chain. I.e., an allowed element of $S^{\mathbb{N}}$.

A realization $\vec{r} \in S^{\mathbb{N}}$ is called ***generic*** for $\mathcal{M}$ if the transition probabilities for $\mathcal{M}$ can be recovered from $\vec{r}$ by averaging.

16 Consider the Markov chain $\mathcal{M}$ with state space $[0,1]$ and the transition rule

$$X_{i+1} = \begin{cases} X_i/3 & \text{with probability } 1/2 \\ X_i/3 + 2/3 & \text{with probability } 1/2 \end{cases}$$

16.1 Using a random number generator, write down the initial segment (up to 4 transitions) of two different realizations of $\mathcal{M}$.

16.2 Let $\vec{r} = (r_0, r_1, \ldots)$ be a realization of $\mathcal{M}$. Is it possible that $\lim_{i\to\infty} r_i$ exists? Why or why not?

16.3 Let $\vec{s} = (s_0, s_1, \ldots)$ be a realization for $\mathcal{M}$ that is generic. Is it possible that $\lim_{i\to\infty} s_i$ exists? Why or why not?

16.4 Can $\mathcal{M}$ be modeled by a dynamical system? If so, describe the model.

16.5 Suppose we start off with a uniform distribution on $[0,1]$. Draw the resulting distribution after 1, 2, and 3 steps along $\mathcal{M}$.

DEFINITION

Iterated Function System (IFS)

An ***iterated function system*** with functions $(f_1, \dots, f_n)$ and transition probabilities $(p_1, \dots, p_n)$ is a stationary Markov chain where transitions are given by the rule

$$X_{i+1} = \begin{cases} f_1(X_i) & \text{with probability } p_1 \\ f_2(X_i) & \text{with probability } p_2 \\ & \vdots \end{cases}$$

DEFINITION

Invariant Set

Let $\mathcal{I}$ be an iterated function system with functions $(f_1, \dots, f_n)$ and non-zero transition probabilities $(p_1, \dots, p_n)$. The set X is called an ***invariant set*** for $\mathcal{I}$ if

$$X = \bigcup_i f_i(X).$$

---

17 Let $\mathcal{I}$ be the iterated function system with transition probabilities $(1/2, 1/2)$ and functions $(f_1 : [0,1] \to [0,1], f_2 : [0,1] \to [0,1])$ given by $f_1(x) = x/3$ and $f_2(x) = x/3 + 2/3$.

17.1 Find an invariant set for $\mathcal{I}$.

17.2 Is the Cantor set an invariant set for $\mathcal{I}$?

17.3 Is there a larger invariant set for $\mathcal{I}$ than the Cantor set? Why or why not?

17.4 If the probabilities for f_1 and f_2 change, will that affect the invariant sets?

---

18 Define $f_{\vec{a}} : [0,1]^2 \to [0,1]^2$ by $f_{\vec{a}}(\vec{x}) = \vec{x}/2 + \vec{a}$. Let $\mathcal{F}$ be the iterated function system with functions $(f_{\vec{0}}, f_{(1/2,0)}, f_{(0,1/2)})$ and equal probabilities.

18.1 Draw a maximal invariant set for $\mathcal{F}$.

18.2 Can you create an iterated function system so that the Sierpinski triangle is an invariant set? If so, do it!

Continuous-time Dynamical Systems

The dynamical systems we've seen so far involve taking one "step" at a time. However, as we experience life, it seems like time advances continuously, not discretely. Can we capture that in a dynamical system?

Continuous-time Dynamical System

DEF

Let X be a set and let $(T^a : X \to X)_{a\in\mathbb{R}}$ be a family of functions satisfying $T^a \circ T^b = T^{a+b}$. Then, (T^a, X) is called a ***continuous time dynamical system***.

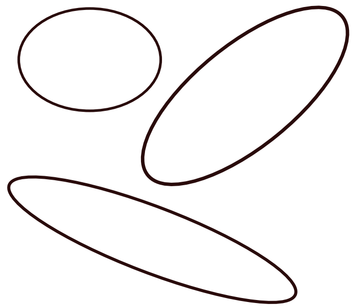
19

19.1 Let $S^a : \mathbb{R} \to \mathbb{R}$ be defined by $S^a(x) = ax$. Is $(S^a, \mathbb{R})$ a continuous time dynamical system?

19.2 Let $T^a : \mathbb{R} \to \mathbb{R}$ be defined by $T^a(x) = 2^a x$. Is $(T^a, \mathbb{R})$ a continuous time dynamical system?

19.3 Let $\mathcal{F} = \{\text{functions from } \mathbb{R} \text{ to } \mathbb{R}\}$ and define $E^t : \mathcal{F} \to \mathcal{F}$ by $E^t(f(x)) = f(x+t)$. Is $(E^t, \mathcal{F})$ a continuous time dynamical system?

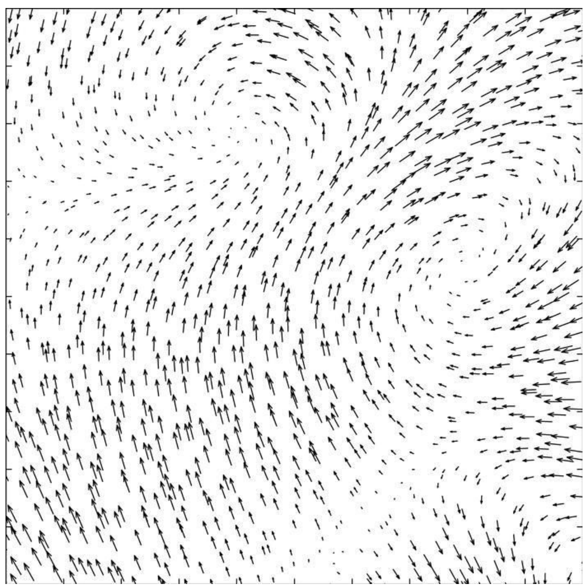
20.1 Consider the following garden paths.



Let P represent the set of points on the paths and define $W^t : P \to P$ to be the result of walking at unit speed counter-clockwise around your path for t seconds.

Is (W^t, P) a continuous time dynamical system?

20.2 You're watching the surface of a lake and carefully map out the velocity of the water at each point. You produce the following picture of velocities.



Let S be the points on the surface of the lake and let $\Phi^t : S \to S$ be the result of "flowing" along the surface for t seconds. Is (Φ^t, S) a continuous time dynamical system?

Systems like (Φ^t, S) come up often because they are described by *local* rules. That is, recorded in the picture is information about where you flow after tiny time increments—finding where you end up after bigger time increments takes work.

Vector Field

DEF

Given a set X, an ***n-dimensional vector field on X*** is a function $F : X \to \mathbb{R}^n$ which assigns a vector to each point in X.

Vector fields show up a lot in physics because *velocities* and *accelerations* of particles in a fluid naturally produce vector fields.

---

21 Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field and let $\Phi^t : \mathbb{R}^2 \to \mathbb{R}^2$ be the function which "flows" points along the vector field with velocity at a given point equal to the vector at that point.

21.1 Find a V so that all points flow vertically at unit speed.

21.2 Find a V so that all points flow radially outward and increase in speed.

21.3 Find a V so that all points flow around in a circle.

21.4 Find a V so that all points flow around in a circle at the same speed.

21.5 Find a V so that all points flow around in a circle with the same period (that is, it takes the same amount of time for any given point to go around the circle).

Flow & Orbit

DEFINITION

Let (W^t, X) be a continuous time dynamical system and let $x \in X$. The ***flow of W^t starting at x*** is the function $f : \mathbb{R} \to X$ defined by

$$f(t) = W^t(x).$$

The ***orbit of x under W^t*** is the set $\{f(t) : t \in \mathbb{R}\}$, where f is the flow of W^t starting at x. We call the set $\{f(t) : t \in \mathbb{R}^+\}$ the ***forward orbit*** of x.

We will often notate the flow of W^t starting at x by $\phi_t(x)$. As a flow, we think of t as the variable.

---

22 Let $V : \mathbb{R}^2 \to \mathbb{R}^2$ be a vector field and let $\Phi^t : \mathbb{R}^2 \to \mathbb{R}^2$ be the function that flows points along the vector field.

22.1 If possible, find a V so that some orbits are infinitely long.

22.2 If possible, find a V so that some orbits are not infinitely long.

22.3 If possible, find a V so that some orbits are infinitely long and others are not.

22.4 If possible, find a V so that all orbits are finite line segments.

22.5 If possible, find a V so that the forward orbit of every point ends up at $\vec{e}_1$.

22.6 If possible, find a V so that the forward orbit of every point ends up at $\vec{0}$ or $\vec{e}_1$ (and at least one orbit heads towards $\vec{0}$ and one towards $\vec{e}_1$).

DEFINITION

Stable & Unstable Points

Let $(W^t, \mathbb{R}^n)$ be a continuous time dynamical system and let $\phi_t(\vec{w})$ represent the flow of W^t starting at $\vec{w} \in \mathbb{R}^n$. We call the point $\vec{x} \in \mathbb{R}^n$ ***stable*** if for all $\varepsilon > 0$, there exists a $\delta > 0$ so that for all $\vec{y} \in \mathbb{R}^n$,

$$\|\vec{x} - \vec{y}\| < \delta \qquad \text{implies} \qquad \|\phi_t(\vec{x}) - \phi_t(\vec{y})\| < \varepsilon \qquad \text{for all} \qquad t > 0.$$

Otherwise, $\vec{x}$ is called ***unstable***.

---

23 Given a 2×2 matrix M, define a vector field $V_M : \mathbb{R}^2 \to \mathbb{R}^2$ by $V_M(\vec{x}) = M\vec{x}$. Consider the following matrices and their associated vector fields:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad D = \begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

23.1 Explain in plain language what it means for x to be a *stable* point of a continuous dynamical system.

23.2 Use an online plotter (for example `https://www.desmos.com/calculator/eijhparfmd` or `https://anvaka.github.io/fieldplay`) to plot V_X for $X \in \{I, A, B, C, D, E\}$ and determine which points are stable/unstable for each.

23.3 Compute the eigenvalues for each of the matrices. Can you relate the eigenvalues to stable/unstable points?

Differential Equations

To really get a handle on what's going on, let's think about some differential equations!

24

24.1 Let a be a constant. Find a non-trivial solution to the differential equation $f'(t) = af(t)$. For a constant $x \in \mathbb{R}$, find a solution to $f'(t) = af(t)$ that satisfies $f(0) = x$.

24.2 Consider the (boring) vector field $V : \mathbb{R}^1 \to \mathbb{R}^1$ defined by $V(x) = ax$. Find a non-trivial flow for the corresponding dynamical system. Is V given by matrix multiplication? If so, what's the matrix?

24.3 Let $D = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ and define the vector field $W : \mathbb{R}^2 \to \mathbb{R}^2$ by $W(\vec{x}) = D\vec{x}$. Find a non-trivial flow for the corresponding dynamical system.

Can you give conditions on a and b so that $\vec{0}$ is stable/unstable with respect to the corresponding dynamical system? If a and b are complex numbers, does that change your answer?

24.4 Suppose $(W^t, \mathbb{R}^n)$ is a continuous dynamical system. Define what the *derivative of W^t with respect to t* should mean. How does this relate to vector fields and flows?

24.5 Prove that if $f : \mathbb{R} \to \mathbb{R}^n$ is a function and M is a matrix, then $(Mf)' = M(f')$.

24.6 Suppose A is a matrix and $A = PDP^{-1}$ where D is diagonal. Further, suppose f is a solution to $f'(t) = Df(t)$. Find a solution to the differential equation $\Phi'(t) = A\Phi(t)$.

24.7 Let M be a matrix and consider the continuous dynamical system coming from the vector field $\vec{x} \mapsto M\vec{x}$. Classify the behaviour near $\vec{0}$ based on the eigenvalues of M.

25

Consider the continuous dynamical systems $(W_1^t, \mathbb{R}^2)$, $(W_2^t, \mathbb{R}^2)$, $(W_3^t, \mathbb{R}^2)$, and $(W_4^t, \mathbb{R}^2)$ given by flows along the vector fields

$$V_1\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix} \qquad V_2\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + y^2 \\ -y \end{bmatrix} \qquad V_3\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x + y^2 \end{bmatrix} \qquad V_4\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x + y^2 \\ -y \end{bmatrix}$$

25.1 Classify $(0, 0)$ as stable or unstable for each system.

25.2 Classify $(0.01, 0.01)$ as stable or unstable for each system.

25.3 Conjecture: under what conditions will a first-order approximation about at a fixed point tell you about stability *near* that fixed point.

DEFINITION

Structural Stability

A continuous dynamical system (W^t, X) is called ***structurally stable at $\vec{x} \in X$ with respect to fixed points/stable points/periodic points/etc.*** if the topology of the set of fixed points/stable points/periodic points/etc. near $\vec{x}$ remains invariant with respect to small perturbations of W^t.

26

26.1 What should a *small perturbation* of $(W^t, \mathbb{R}^n)$ mean?

26.2 What should it mean for the *topology* of a set to be *invariant*?

26.3 Out of the examples in 25, which are structurally stable at $\vec{0}$?

26.4 Revisit your conjecture from 25.3. Can you rephrase it in terms of structural stability?

DEFINITION

Time-1 Map

Let (W^t, X) be a continuous dynamical system. The ***time-1 map*** associated with (W^t, X) is the dynamical system (T, X) where

$$T(x) = W^1(x)$$

for all $x \in X$.

---

27 Let S be a circle of circumference 1 and let (W_k^t, S) be the continuous dynamical system that flows points counter-clockwise along S at speed k.

27.1 For which k is the time-1 map associated with (W_k^t, S) trivial?

27.2 For which k does the time-1 map associated with (W_k^t, S) have periodic points?

Circle Maps

Let S be a circle of circumference 1. We can associate S with the unit interval $[0,1]$ provided we "glue" the endpoints together. We write $[0,1]/0 \sim 1$ to notate the set $[0,1]$ where 0 and 1 are considered the same point. That is, if you move to the right starting at 0.99, you'll wrap around and end up at 0.01.

Given a function $f : [0,1) \to \mathbb{R}$, we can create a function $g : [0,1) \to [0,1)$ via the formula

$$g(x) = f(x) \mod 1.$$

That is, compute the value $f(x)$ with "wrap around", and that's what $g(x)$ is.

Symbolic Coding

DEFINITION

Let (T, X) be a discrete dynamical system and let $\mathcal{P} = \{P_a, P_b, \ldots\}$ be a partition of X. A ***symbolic coding*** of $x \in X$ relative to the partition $\mathcal{P}$ is the sequence

$$\mathbb{C}(x) = (c_0, c_1, \ldots)$$

where

$$c_i = \begin{cases} a & \text{if} \quad T^i x \in P_a \\ b & \text{if} \quad T^i x \in P_b \\ \vdots & \end{cases}$$

28 Let (W_k^t, S) be as in 27. We can describe this dynamical system by

$$W_k^t(x) = x + kt \mod 1.$$

Let $P_a = [0, 1/2)$ and $P_b = [1/2, 1)$ be a partition of $[0,1]/0 \sim 1$, and let (T_k, S) be the time-1 map for (W_k^t, S).

28.1 Let $k = 0.25$. Find $\mathbb{C}(0)$, $\mathbb{C}(1/5)$, and $\mathbb{C}(1/3)$ for $T_k, \{P_a, P_b\}$.

28.2 Suppose $\mathbb{C}(x) = (a, a, b, b, a, a, b, b, \ldots)$. What can you say about x and k?

28.3 Is $(a, b, a, a, b, b, a, a, a, b, b, b, \ldots)$ the symbolic coding of any point? Why or why not?

28.4 For which k can x be *exactly* recovered from $\mathbb{C}(x)$?

28.5 Let $\{a, b\}^{\mathbb{N}}$ be the set of sequences of a's and b's. We can think of $\mathbb{C} : S \to \{a, b\}^{\mathbb{N}}$ as a function. For which k is $\mathbb{C}$ one-to-one? Onto?

Orbit

DEFINITION

Let (T,X) be a discrete dynamical system. The ***orbit*** of a point $x \in X$ under T is the set

$$\mathcal{O}(x) = \{T^i x \ : \ i \in \mathbb{Z}\}$$

if T is invertible and

$$\mathcal{O}(x) = \left\{T^i x \ : \ i \in \mathbb{N} \cup \{0\}\right\}$$

if T is non-invertible.

The *doubling map* is the function $T : [0,1) \to [0,1)$ defined by $x \mapsto 2x \mod 1$.

29 Let $\mathcal{P} = \{\mathcal{P}_0, \mathcal{P}_1\} = \{[0,1/2),[1/2,1)\}$ be a partition of $[0,1)$ And let $\mathbb{C} : [0,1) \to \{0,1\}^{\mathbb{N}}$ be the coding function arising from the doubling map and the partition $\mathcal{P}$.

29.1 Is the map $Q : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto 2x$ invertible? Is the doubling map invertible? Explain.

29.2 Find $\mathcal{O}(1/3)$, $\mathcal{O}(4/5)$, and $\mathcal{O}(0)$ under the doubling map.

29.3 Find $\mathbb{C}(1/3)$, $\mathbb{C}(4/5)$, and $\mathbb{C}(0)$.

29.4 Suppose $\mathcal{O}(x)$ is a finite set. What can you say about $\mathbb{C}(x)$?

29.5 Suppose $x \neq y$. Is it possible that $\mathbb{C}(x) = \mathbb{C}(y)$?

29.6 Is $\mathbb{C}$ invertible?

29.7 Define $\mathbb{E} : \{0,1\}^{\mathbb{N}} \to [0,1)$ by

$$(a_1, a_2, a_3, \ldots) \mapsto \sum \frac{a_i}{2^i}.$$

Compute $\mathbb{E}(\mathbb{C}(1/2))$, $\mathbb{E}(\mathbb{C}(1/3))$, and $\mathbb{E}(\mathbb{C}(4/5))$.

29.8 What is $\mathbb{E} \circ \mathbb{C}$? How about $\mathbb{C} \circ \mathbb{E}$?

Base-n Expansion

DEF

A number $x \in [0,1)$ has a ***base-n expansion*** $0.d_1 d_2 d_3 \cdots$ if

$$x = \sum \frac{d_i}{n^i}.$$

30

30.1 Explain how base-2 representations relate to dynamical systems.

30.2 Find a dynamical system and a partition so that $\mathbb{C}$ produces base-10 representations.

30.3 Is the map from a number to its base-n representation invertible? Explain.

DEFINITION

Full Shift

Let $\Omega = \{0, 1, \ldots, n-1\}^{\mathbb{N}}$ and define $\sigma : \Omega \to \Omega$ by

$$\sigma(a_0, a_1, a_2, \ldots) = (a_1, a_2, \ldots).$$

The dynamical system (σ, Ω) is called the ***full shift on n symbols***.

The dynamical system (σ, Ω) is also called a *shift space* and is equipped with a topology coming from the *standard metric* on Ω. This metric determines which sequences converge.

DEFINITION

Standard Metric on Ω

The ***standard metric*** on Ω is the function $d : \Omega \times \Omega \to [0, \infty)$ defined by

$$d(x, y) = \begin{cases} 0 & \text{if } x = y \\ 2^{-i} & \text{where } i = \min_j \{j : x_j \neq y_j\} \end{cases}$$

where $x = (x_0, x_1, \ldots)$ and $y = (y_0, y_1, \ldots)$ are points in Ω.

---

31 Let (σ, Ω) be the full shift on two symbols.

31.1 Let $x = (1, 0, 1, 0, 1, 0, \ldots)$ and $y = (0, 0, 0, 0, 1, 0, 1, 0, \ldots)$. Find $d(x, y)$, $d(\sigma(x), \sigma(y))$, $d(\sigma^2(x), \sigma^2(y))$, ...

31.2 Let $\vec{1} = (1, 1, 1, \ldots)$. Find points $a, b \in \Omega$ so that

$$d(\sigma^n(a), \vec{1}) \nearrow 1 \qquad \text{and} \qquad d(\sigma^n(b), \vec{1}) \searrow 0.$$

31.3 Let $(T, [0, 1))$ be the doubling map and let $\mathbb{C}$ be the coding function for the partition $\{[0, 1/2), [1/2, 1)\}$. You can view $\mathbb{C}$ as a function from $[0, 1) \to \Omega$.

For $t = 1/3$, compute $\mathbb{C}(t)$, $\mathbb{C} \circ T(t)$, and $\sigma \circ \mathbb{C}(t)$. What do you notice? Will this always happen?

31.4 Is $\mathbb{C}([0, 1)) \subseteq \Omega$ closed? (I.e., does it include all its limit points?)

31.5 Let $t, t' \in [0, 1)$ and let $a = \mathbb{C}(t)$ and $a' = \mathbb{C}(t')$.

(a) If $d(a, a')$ is small, what can you say about $|t - t'|$?

(b) If $|t - t'|$ is small, what can you say about $d(a, a')$?

The full 2-shift isn't a perfect representation of the doubling map, but it's close! And, it turns out to be much easier to study.

DEFINITION

Subshift

Let (σ,Ω) be the full shift on n symbols. The system (σ,X) is called a ***subshift*** provided

1. $X \subseteq \Omega$ is a closed set (i.e., it contains all its limit points); and
2. X is σ-invariant (i.e., $\sigma(X)=X$).

---

32 Consider the circle rotation with $k = 1/4$. Let $\mathcal{O}(0)$ be the *orbit* of 0 under this map, and define $X = \mathbb{C}(\mathcal{O}(0))$ with the usual partition.

32.1 Write down X.

32.2 Is (σ,X) a subshift?

32.3 For each $x \in X$, find a corresponding $x' \in [0,1)$ so that $x = \mathbb{C}_T(x')$ where $\mathbb{C}_T$ is the coding with respect to the *doubling map* and the standard partition.

32.4 Can you model a circle rotation with the doubling map/subshift?

33 Let $\mathcal{M}$ be the Markov chain with transition matrix $\begin{bmatrix} .5 & 1 \\ .5 & 0 \end{bmatrix}$ and let $\mathcal{M}'$ be the Markov chain with transition matrix $\begin{bmatrix} .5 & .9 \\ .5 & .1 \end{bmatrix}$.

Let $X = \{\text{realizations of } \mathcal{M}\}$ and let $X' = \{\text{realizations of } \mathcal{M}'\}$.

33.1 Is (σ, X) a subshift? Why or why not? What does σ mean in terms of the original Markov chain?

33.2 Is (σ, X') a subshift? What does it "look" like?

33.3 Let Y be the set of *generic* realizations of $\mathcal{M}$. Is (σ, Y) a subshift?

Word

DEFINITION

Fix $\Omega = \{0, \ldots, n-1\}^{\mathbb{N}}$. The finite list $w = (w_0, \ldots, w_{k-1}) \in \{0, \ldots, n-1\}^k$ is called a ***word of length k*** in Ω.

If $x = (x_0, x_1, \ldots) \in \Omega$, we say that ***$x$ contains the word w*** if for some j, $w_i = x_{j+i}$ for all $i = 0, \ldots, k-1$.

If $X \subseteq \Omega$, we say ***X contains the word w*** if there exists $x \in X$ where x contains the word w.

Normalized Topological Entropy

DEFINITION

Let (σ, X) be a subshift of the full shift (σ, Ω). The ***normalized topological entropy*** of X is

$$\mathcal{H}(X) = \lim_{k\to\infty} \frac{\log(\#\text{ words of length } k \text{ in } X)}{\log(\#\text{ words of length } k \text{ in } \Omega)}.$$

34

Let $\mathcal{C}$ be the 3-state Markov chain with transition matrix $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ and let $X \subseteq \{0,1,2\}^{\mathbb{N}} = \Omega$ be the set of realizations of $\mathcal{C}$ starting at 0 or 2.

34.1 Draw the graph associated with $\mathcal{C}$.

34.2 Find the (normalized topological) entropy of (σ, X).

34.3 Find the (normalized topological) entropy of (σ, Ω).

34.4 Draw $\mathbb{E}(X)$ as a subset of $[0,1]$. What is its fractal dimension?

Entropy gives a way of classifying subshifts based on how "big" they are. This can be interpreted as how densely a subshift can store information.

---

35 Let $\Omega = \{0,1\}^{\mathbb{N}}$ and let $X \subseteq \Omega$ be the set of points with no two 1's in a row.

35.1 How many words of length $1,2,3,4$ are there in Ω? How about in X?

35.2 Is (σ, X) a subshift? Why or why not?

35.3 Can (σ, X) be modeled by a Markov chain? If so, write down its transition matrix (not the stochastic one, the other one).

35.4 How do $A, A^2, A^3, \ldots$ relate to the number of words in X?

35.5 Find the entropy of (σ, X).

35.6 If you wanted to transmit n bits of information but you were only allowed to send subwords of X, how long of a subword would you need?

Entropy relates to how much information you can get about a point by asking a series of questions.

---

36 Let $C \subseteq [0,1]$ be the standard Cantor set. Every "piece" of C has a left and a right. You are trying to locate a point $x \in C$ by asking a series of questions like:

- Is x in the left or right half of C?
- Is x in the left or right half of the left/right half of C?

  $\vdots$

36.1 What can you say about x if you know

(a) it's in the left half of C?

(b) it's in the left half of the left half of C?

(c) it's in the right half of the left half of the left half of C?

36.2 Let $A_n = (a_1, \ldots, a_n) \in \{\text{left}, \text{right}\}^n$ be the answers to the first n questions you ask.

Given A_n, what is the maximum error you could have in guessing x?

36.3 Playing the same game trying to find the point $y \in [0,1]$, you ask a series of left-right questions and get a sequence of answers A'_n. Given A'_n what is the maximum error you could have in guessing y?

36.4 Let E_n be the maximum error you have in guessing x given the answers A_n and let E'_n be the maximum error you have in guessing y given the answers A'_n. Compute

$$\lim_{n\to\infty} \frac{\log(E_n)}{\log(E'_n)}.$$

36.5 If you want to locate a point in $[0,1]$ with error less than ε, how many left-right questions do you need to ask?

36.6 If you want to locate a point in C with error less than ε, how many left-right questions do you need to ask?

37 Let $X \subseteq \Omega = \{0, 1, 2\}^{\mathbb{N}}$ be the set of all sequences without 1's.

37.1 Given that w is a length-n word in X, how many yes-no questions do you need to ask to determine w?

37.2 Given that w is a length-n word in Ω, how many yes-no questions do you need to ask to determine w?

37.3 How does entropy relate to the number of yes-no questions needed to determine a word?

DEFINITION

Expansive

A dynamical system (T,X) is ***expansive*** at the point $x \in X$ if there exists $\varepsilon > 0$ such that for all $y \neq x$,

$$d(T^n x, T^n y) > \varepsilon$$

for some $n \geq 0$. The system (T,X) is called ***expansive*** if it is expansive at every point in X.

38

38.1 What's the difference between being expansive and not being *stable*?

38.2 Let (T,S) be a circle rotation by $1/4$. Is (T,S) expansive?

38.3 Let (σ,Ω) be the full two-shift. Is (σ,Ω) expansive?

38.4 Let $(T_f,\mathbb{R})$ be the Newton's method map from the first day of class. Is $(T_f,\mathbb{R})$ expansive?

38.5 Suppose (T,X) is an expansive dynamical system.

(a) Can (T,X) have fixed points?

(b) Can (T,X) have a periodic point?

(c) Can (T,X) have an attracting point?

Transitivity

DEF Let (T, X) be a dynamical system. The point $x \in X$ is called ***transitive*** if $\overline{\mathcal{O}(x)} = X$. The system (T, X) is called ***transitive*** if there exists a transitive point in X.

Here, $\overline{\mathcal{O}(x)} = X$ denotes the *closure* of the orbit of x. That is, all points that are limits of points in $\mathcal{O}(x)$.

39

39.1 Let $(T_{1/4}, S)$ be a circle rotation with angle $1/4$. Find $\overline{\mathcal{O}(0)}$. Is $(T_{1/4}, S)$ transitive?

39.2 Let $(T_{\sqrt{2}}, S)$ be a circle rotation with angle $\sqrt{2}$. Find $\overline{\mathcal{O}(0)}$. Is $(T_{\sqrt{2}}, S)$ transitive?

39.3 Let (σ, Ω) be the full two-shift and let $x = 10100100010000100\cdots$. What is $\overline{\mathcal{O}(x)}$.

39.4 Is (σ, Ω) transitive?

Chaos

DEF

A dynamical system (T, X) is called ***chaotic*** if it is both expansive and transitive.

40

40.1 Let (T_θ, S) be a circle rotation with angle θ. Is (T_θ, S) chaotic?

40.2 Let (σ, Ω) be the full two-shift. Is (σ, Ω) chaotic?

40.3 Let (σ, X) be a subshift of the full two-shift. Is (σ, X) chaotic?

Chaos is a battle between order and disorder. Expansivity says close points eventually get far, so you cannot "approximate" a chaotic system into the future. Transitivity says that points (and open neighborhoods in particular) get smeared all over the space.

We see systems like this in nature, and have a *theory of probability* that we use to describe unpredictable events. Can we do something similar to "predict" chaotic systems?

Indicator Function

DEFINITION

Given a set $A \subseteq X$, the ***indicator of A*** is the function $\mathbb{I}_A : X \to \{0,1\}$ defined by

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}.$$

Empirical Measure

DEFINITION

Let (T, X) be a dynamical system and let $x \in X$. The ***empirical measure generated by x*** is the function μ_x : {open or closed subsets of X} $\to [0,1]$ given by

$$\mu_x(A) = \lim_{n\to\infty} \frac{1}{n} \sum_{0\le i<n} \mathbb{I}_A(T^i x) = \text{\% of time } \mathcal{O}(x) \text{ is in } A,$$

provided this limit exists (for all open/closed $A \subseteq X$).

An empirical measure can be extended to more than just open subsets. Typically, the domain of an empirical measure is extended to the *Borel* subset, which you'd learn about in an analysis class.

---

41 Let $(T, [0,1))$ be the doubling map.

41.1 Find $\mu_x([0,1/2))$ and $\mu_x([1/2,1))$ for $x = 0$.

41.2 Find $\mu_x([0,1/2))$ and $\mu_x([1/2,1))$ for $x = 1/5$.

41.3 Find $\mu_x([0,1/2))$ and $\mu_x([1/2,1))$ for $x = \sqrt{2}/2$.

41.4 Find $\mu_x([0,1/2))$ and $\mu_x([1/2,1))$ when $x \in [0,1]$ is such that the binary expansion of x has 0's 2/3rds of the time.

41.5 Find an x such that μ_x doesn't exist.

Chaotic systems can often be "coupled" to probability spaces. When this coupling is possible, *almost every* empirical measure is the same and is equal to the underlying probability measure. This means, despite being deterministic, a chaotic system can be thought of as interchangeable with a probability space.

Probability Measure

DEFINITION

Let Ω be a set and let $\mathcal{B} = \{\text{open or closed subsets of } \Omega\}$. A ***Probability measure on*** Ω is a function $\mu : \mathcal{B} \to \mathbb{R}_{\geq 0}$ that satisfies

1. $\mu(\emptyset) = 0$,
2. $\mu(\Omega) = 1$,
3. If $\{A_n\} \subseteq \mathcal{B}$ is a *countable* collection of disjoint sets such that $\bigcup A_n \in \mathcal{B}$, then

$$\mu\Big(\bigcup A_n\Big) = \sum \mu(A_n).$$

Sets in domain of μ are called ***measurable sets***.

Invariant Measure

DEFINITION

Let (T, Ω) be a dynamical system and let μ be a probability measure on Ω. The measure μ is called ***invariant with respect to*** T if for any measurable set $A \subseteq \Omega$,

$$\mu(A) = \mu\Big(T^{-1}(A)\Big).$$

The *Carathédory Extension Theorem* says that any property of a probability measure on $[0,1]$ can be verified by testing on intervals.

---

42 Let $(T, \Omega = [0,1))$ be the doubling map. Throughout this question, all measures are defined on $\Omega = [0,1)$.

42.1 Compute $T^{-1}(\Omega)$, $T^{-1}([1/2, 1))$, and $T^{-2}([1/2, 1))$.

42.2 Let μ be defined by $\mu(A) = \begin{cases} 1 & \text{if } \frac{1}{3} \in A \\ 0 & \text{otherwise} \end{cases}$. Is μ a probability measure? Is μ invariant?

42.3 Let μ be defined by $\mu(A) = \begin{cases} 1 & \text{if } \frac{1}{3} \in A \text{ or } \frac{2}{3} \in A \\ 0 & \text{otherwise} \end{cases}$. Is μ a probability measure? Is μ invariant?

42.4 Let μ be defined by $\mu(A) = \dfrac{\left|\left\{\frac{1}{3}, \frac{2}{3}\right\} \cap A\right|}{2}$, where $|\cdot|$ is set cardinality. Is μ a probability measure? Is μ invariant?

42.5 Let λ be defined so that $\lambda([a,b]) = b - a$ (provided $b \geq a$). λ is called *Lebesgue measure*. Is λ invariant?

42.6 Let μ_x be an emperical measure. Is μ_x invariant?

Probability measures solve the issues of talking about probability when there are an infinite number of possible outcomes. Instead of talking about the probability of *events*, probability measures assign probabilities to collections of events.

Generic Point

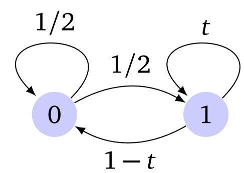
DEF

If (T,Ω) is a dynamical system and μ an invariant probability measure for that system, the point $x \in \Omega$ is called ***generic*** if

$$\mu_x = \mu.$$

Using measures, we can address earlier questions with a more rigorous footing.

43 Consider the family of Markov chains, X_t, given by the following graph.



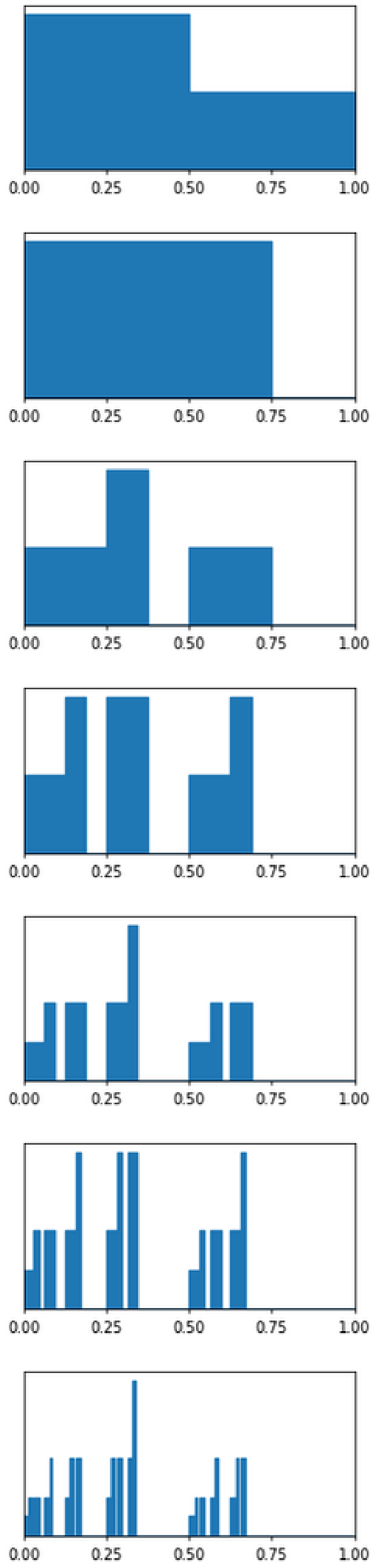
Generic realizations of X_t have 0's $\dfrac{1-t}{3/2-t}$ portion of the time and 1's $\dfrac{1}{2(3/2-t)}$ portion of the time.

For a realization x of X_t, we will interpret it as the binary expansion of a number in $[0,1]$.

43.1 Let T be the doubling map. Find $\mu_x([0,1/2))$ and $\mu_x([1/2,1))$ where x is a generic realization for X_t with $t = 0$, $1/4$, $1/2$, 1.

43.2 For each of the x's from the previous part, make a "graph" of μ_x. That is, divide $[0,1]$ into 2^n equal-length sub intervals and make a bar chart showing the measure of each sub interval. Do this for $n = 1,2,3$.

Progressively more detailed estimations of μ_x:



44 Let X_t be as in 43.

44.1 Compute the topological entropy for realization of X_t for $t = 0, 1/10, 1/2, 1$.

44.2 Let x be a generic realization of X_t.

(a) If $t = 1/2$, how many subwords of length 5 do you expect to see in the first 1000 bits of x?

(b) If $t = 1/10^5$, how many subwords of length 5 do you expect to see in the first 1000 bits of x?

44.3 Is topological entropy a good measure of how much information realizations of X_t contain? Why or why not?

44.4 How many "typical" words of length n are there in realizations of X_t? You may use the approximation, which can be derived via Sterling's formula:

$$\log\binom{n}{\alpha n} \approx -n\Big(\alpha\log\alpha + (1-\alpha)\log(1-\alpha)\Big)$$

for $\alpha \in (0, 1)$.

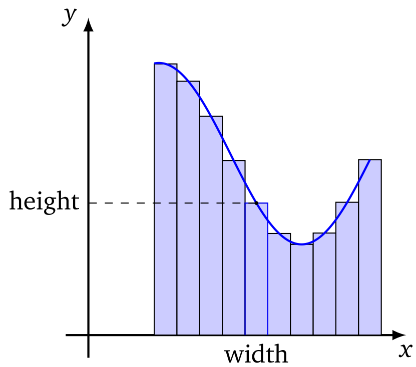
44.5 What should the *relative information entropy* of typical realizations of X_t be? That is, how much information can a "typcial" length n word of X_t contain vs. a length n word with no restrictions at all?

Applications

Ergodic Theory

One of the uses of measures is that they allow for another way to integrate functions.

Normally, we integrate functions using a Riemann sum.



We compute the area of each little rectangle as width of base × height, add them up, and take the limit as the bases shrink to zero. However, if μ is a measure, we can use it to measure the "size" of the base and instead add up $\mu(\text{base}) \times \text{height}$. When using a measure in this way to integrate a function f, we write $\int f \, d\mu$.

If μ_x is an *empirical* measure generated by x, then $\mu(\text{base})$ is just the percent of time $T^i x$ spends in *base*. Thus, we have another way to compute the integral!

$$\int f \, d\mu_x = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N-1} f(T^i x).$$

The orbit of x is used to compute an integral. This is sometimes called "swapping the space average for the time average". The expression $\int f \, d\mu_x$ is the integral of f over *space*. Since the measure of the entire space is 1, this is an average. On the other hand, $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N-1} f(T^i x)$ is the average of f over the orbit of x, and so is a *time average*, where time corresponds to applications of T.

If we have an empirical measure, this all works out well. But, how hard is it to get a hold of an empirical measure? This question is answered by *ergodic theory*: if your dynamical system is ergodic, it's really easy!

Ergodic

DEFINITION

Let (T, Ω) be a dynamical system and let μ be an invariant probability measure. The system (T, Ω, μ) is called ***ergodic*** if every T-invariant set has measure 1 or 0. That is,

$$T^{-1}(A) = A \qquad \implies \qquad \mu(A) = 1 \quad \text{or} \quad \mu(A) = 0.$$

An ergodic system is one that cannot be "broken up into smaller pieces". That is, if you found an invariant subsystem, it would either be effectively the whole space (measure 1) or effectively nothing (measure 0).

Given an ergodic system, we can employ the powerful Birkhoff Ergodic Theorem.

Birkhoff's Ergodic Theorem

THEOREM

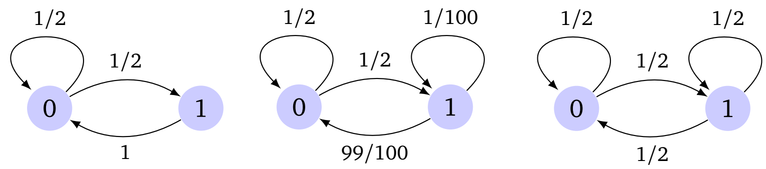
Let (T, Ω, μ) be an ergodic dynamical system and let $f : \Omega \to \mathbb{R}$ be a continuous function. Then, there exists a set $G \subseteq \Omega$ satisfying $\mu(G) = 1$ such that for all $x \in G$,

$$\int f \, d\mu = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N-1} f(T^i x).$$

The Birkhoff Ergodic Theorem can be interpreted as saying that almost all points in Ω will generate the *same* empirical measure. Thus long-term time statistics can almost always be used in place of space averages (for ergodic systems).

Measure-theoretic Entropy

Measures also solve an issue with topological entropy. Consider the following systems



X $\qquad$ Y $\qquad$ Z

We can compute

$$\mathcal{H}(X) = \log_2(\phi) \approx .69 \quad < \quad \mathcal{H}(Y) \quad = \quad \mathcal{H}(Z) = 1.$$

But clearly system Y carries similar amounts of information to system X, since, although two 1's in a row is *allowed*, it is very uncommon.

We can solve this issue by attempting to measure the number of "typical" words in a generic point instead of the total number of words, and we can use measures to define what "typical" means.

Consider asking questions about a point in $x \in [0,1]$. If you ask yes/no questions, your best bet is to ask questions that bisect the set where x may be. By doing this, after asking n questions, your maximum uncertainty of where x may be is $1/2^n$. In other words,

$$\#\text{ bits of information gathered} = -\log_2(\textit{uncertainty}).$$

Using this formula and applying measures, we can say that if we're convinced that $x \in I \subseteq [0,1]$, we've gathered the equivalent of $-\log_2(\mu(I))$ bits of information.

Now, suppose $\mathcal{P} = \{P_1, \ldots, P_n\}$ is a partition of $[0,1]$. In order to figure out which partition of $\mathcal{P}$ our point is in, on average, we need

$$h(\mathcal{P}) = -\sum \mu(P_i)\log_2(\mu(P_i)).$$

That is, $h(\mathcal{P})$ is the weighted average of the amount of information we'd get from knowing that $x \in P_i$, weighted by the size of P_i. We define this quantity to be the entropy of the partition $\mathcal{P}$.

Now, let $(T, \Omega = [0,1))$ be the doubling map and let μ be a T-invariant probability measure. Further, let $\mathcal{P} = \{P_0 = [0,1/2), P_1 = [1/2,1)\}$ to be a partition of Ω.

For a point x, knowing which partition of $\mathcal{P}$ that x is in would tell us the first digit in the binary expansion of x. Knowing which element of $T^{-1}\mathcal{P} = \{T^{-1}P_0, T^{-1}P_1\}$ that x is in would tell us the second digit in the binary expansion of x.

For two partitions $\mathcal{A}$ and $\mathcal{B}$, define $\mathcal{A} \vee \mathcal{B}$ to be the common refinement of $\mathcal{A}$ and $\mathcal{B}$ (i.e., intersect all the elements of $\mathcal{A}$ and $\mathcal{B}$ in every way possible and disregard any resulting empty sets). The operation $\vee$ is called *join*.

The partition elements of $\mathcal{P} \vee T^{-1}\mathcal{P}$ will tell us the first *two* digits of the binary expansion of a point. Similarly,

$$\mathcal{P}^n = \bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$$

gives us information about the first n digits of the binary expansion of a point. We can now make a new definition

Measure-theoretic Entropy

DEFINITION

Let (T, Ω) be a dynamical system, let μ be an invariant probability measure, and let $\mathcal{P}$ be a partition of Ω. Further, let $\mathcal{P}^n = \bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$. The ***measure theoretic entropy of*** (T, Ω, μ) ***with respect to*** $\mathcal{P}$ is

$$h(\mu \mid \mathcal{P}) = \lim_{n\to\infty} h(\mathcal{P}^n).$$

The ***measure theoretic entropy of*** (T, Ω, μ) is

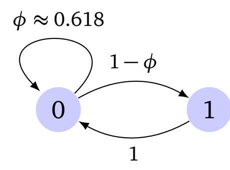
$$h^\infty(\mu) = \sup_{\text{partitions } \mathcal{Q}} h(\mu \mid \mathcal{Q}).$$

Sinai's Theorem states that the measure theoretic entropy of (T, Ω, μ) is actually achieved by a particular partition of Ω. And in the case of the doubling map, the partition $\mathcal{P} = \{[0, 1/2), [1/2, 1)\}$ is such a partition.

Recall the Markov chains X, Y, and Z from before. Let x be a generic point for X, y a generic point for Y, and z a generic point for Z. Numerically computing we have

$$h^{\infty}(\mu_x) \approx 0.68 \qquad h^{\infty}(\mu_y) \approx 0.70 \qquad h^{\infty}(\mu_z) = 1$$

which is much closer to what we'd expect. However, you may notice that $h^{\infty}(\mu_x) \approx 0.68 < 0.69 \approx \mathcal{H}(X)$. Indeed, we haven't achieved the maximal entropy possible for this system! If we instead consider



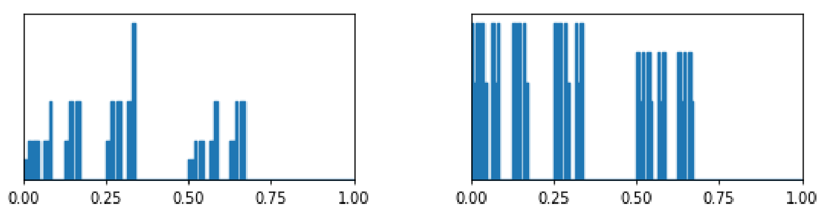
W

and let w be a generic point for W, we have

$$h^{\infty}(\mu_w) = \mathcal{H}(W) = \mathcal{H}(X) \approx 0.69.$$

Why should this be? There are several reasons: fist, when walking along the graph for X, if you choose to step to 0, the next step you have complete freedom to choose 0 or 1. However, if you choose to step to 1, the following step is 100% determined. Thus, stepping to 0 more often gives an entropy boost. However, if you step to 0 too often, you'll miss out on potential entropy because you'll be packing in too many runs of 0's.

Here's another way to see the same thing:



On the left is a graph approximating the density of μ_x and on the right is a graph approximating the density of μ_w. Notice how the μ_w graph is as constant as possible. μ_w *must* give zero mass to sets of points with consecutive 1's in their binary expansion, but other than giving zero mass to those points, μ_w gives equal weight to everything that's left (as equal as possible subject to the constraint of being an invariant measure). On the other hand, μ_x gives much more weight to some points than to others. In particular, long sequences of 0's are given much less weight than needed to maximize entropy.