Notes for 160.734

Part VIII: Symbolic Dynamics, Measure Theory, and Ergodic Theory

D.J.W. Simpson
SFS, Massey University
June 24, 2019

Here we look at some advanced approaches for studying maps with the over-riding aim of obtaining a deeper understanding of chaos. This is the most technical part of the course; we will be using techniques from functional analysis and statistics. For additional guidance on the topics that follow, as well as the library books listed in the course outline, you may find *Scholarpedia* and *Wikipedia* particularly helpful.

---

1 The shift map

- The shift map, defined precisely in a moment, is a map that acts on a space of sequences. Here we consider only binary sequences: sequences involving two different symbols. We will denote these symbols L and R, for reasons that will become apparent in §2.
- An example of one of our sequences is $\mathcal{S} = LLRLRL\ldots$ and we write $s_0 = L$, $s_1 = L$, $s_2 = R$, and so on. Let Σ_2 denote the set of all such sequences.
- A periodic sequence of period n is the infinite concatenation of a *word* of length n, and we denote it as the $\infty^{\rm th}$ power of the word. For example, we write $LRLRLRLR\ldots = (LR)^\infty$ (a period-2 sequence).
- We can also use this notation to abbreviate sequences that are *eventually periodic*, e.g. $RRLLLLLL\ldots = RRL^\infty$.

Definition 1.1. The *distance* between any $\mathcal{S}, \mathcal{T} \in \Sigma_2$ is

$$d(\mathcal{S}, \mathcal{T}) = \sum_{i=0}^{\infty} \frac{1}{2^i} \left\{ \begin{array}{ll} 0\,, & s_i = t_i \\ 1\,, & s_i \neq t_i \end{array} \right\}. \tag{1.1}$$

- Notice how differences in earlier symbols give greater value to the distance. Notice also that the *diameter* of Σ_2 is 2.

Exercise 1.1. Show that the distance between L^∞ and $(LR)^\infty$ is $\frac{2}{3}$.

Exercise 1.2. Find two sequences a distance $\frac{13}{16}$ from L^∞.

- Proofs of the following two results are relatively straight-forward and left to the reader.

Lemma 1.1. *Let $\mathcal{S}, \mathcal{T} \in \Sigma_2$ and $n \ge 0$. If $s_i = t_i$ for all $i = 0, 1, \ldots, n$, then $d(\mathcal{S}, \mathcal{T}) \le \frac{1}{2^n}$.*

Lemma 1.2. *The distance d is a metric on the space Σ_2 (hence (Σ_2, d) is a metric space).*

Definition 1.2. The *shift map* $\sigma : \Sigma_2 \to \Sigma_2$ is defined by $\sigma(\mathcal{S}) = s_1 s_2 \ldots$ (that is, all symbols shift one place to the left and the first symbol is dropped).

- The shift map has two fixed points: L^∞ and R^∞, and two period-2 solutions: $(LR)^\infty$ and $(RL)^\infty$. More generally, if $\mathcal{W}$ is a *primitive* word (that is, it cannot be written as a power) of length n, then $\mathcal{W}^\infty$ is a period-n solution of the shift map.

Lemma 1.3. *The shift map is a continuous function on Σ_2 (with respect to the metric d).*

Proof. Let $\varepsilon > 0$ and choose any $\mathcal{S} \in \Sigma_2$.

CLAIM: There exists $\delta > 0$ such that for all $\mathcal{T} \in B_\delta(\mathcal{S})$ we have $\sigma(\mathcal{T}) \in B_\varepsilon(\sigma(\mathcal{S}))$.

Let $\delta = \frac{\varepsilon}{2}$. Choose any $\mathcal{T} \in B_\delta(\mathcal{S})$. We can assume $s_0 = t_0$ (say by requiring $\varepsilon < 1$), then

$$d(\sigma(\mathcal{S}), \sigma(\mathcal{T})) = 2d(\mathcal{S}, \mathcal{T}) < 2\delta = \varepsilon.$$

That is, $\sigma(\mathcal{T}) \in B_\varepsilon(\sigma(\mathcal{S}))$, as required. □

Lemma 1.4. *The shift map is transitive.*

Proof. Let $\varepsilon > 0$ and choose any $\mathcal{S}, \mathcal{T} \in \Sigma_2$.

CLAIM: There exists $\mathcal{U} \in \Sigma_2$ and $i_1, i_2 \ge 0$ such that $\sigma^{i_1}(\mathcal{U}) \in B_\varepsilon(\mathcal{S})$ and $\sigma^{i_2}(\mathcal{U}) \in B_\varepsilon(\mathcal{T})$.

We define $\mathcal{U}$ rather cleverly as follows: let the first two symbols of $\mathcal{U}$ be L and R, let the next eight symbols of $\mathcal{U}$ consist of all four words of length 2, let the next 24 symbols of $\mathcal{U}$ consist of all eight words of length 3, and so on. For example

$$\mathcal{U} = LRLLLRRLRRLLLLLR\ldots$$

Let $n \geq 0$ be such that $\frac{1}{2^n} < \varepsilon$. By construction, the first $n+1$ elements of $\mathcal{S}$ appear somewhere in $\mathcal{U}$ as a block. That is, the first $n+1$ elements of $\mathcal{S}$ and $\sigma^{i_1}(\mathcal{U})$ are the same, for some $i_1 \geq 0$. Then by Lemma 1.1,

$$d\big(\mathcal{S}, \sigma^{i_1}(\mathcal{U})\big) \leq \frac{1}{2^n} < \varepsilon,$$

and similarly there exists such an i_2. □

Lemma 1.5. *The shift map exhibits sensitive dependence on initial conditions on Σ_2.*

Proof.

> Claim: There exists $\beta > 0$ such that for all $\mathcal{S} \in \Sigma_2$ and $\varepsilon > 0$ there exists $\mathcal{T} \in B_\varepsilon(\mathcal{S})$ and $n \geq 0$ such that $d(\sigma^n(\mathcal{S}), \sigma^n(\mathcal{T})) \geq \beta$.

Let $\beta = 1$. Let $\varepsilon > 0$ and choose any $\mathcal{S} \in \Sigma_2$. Let $n \geq 0$ be such that $\frac{1}{2^n} < \varepsilon$. Define $\mathcal{T}$ by $t_i = s_i$ for all $i \neq n$. Then $d(\mathcal{S}, \mathcal{T}) = \frac{1}{2^n}$, hence $\mathcal{T} \in B_\varepsilon(\mathcal{S})$. The sequences $\sigma^n(\mathcal{S})$ and $\sigma^n(\mathcal{T})$ differ only in their first symbols, hence

$$d(\sigma^n(\mathcal{S}), \sigma^n(\mathcal{T})) = 1 \geq \beta,$$

as required. □

- We conclude that σ is chaotic in the sense that it is transitive and has sensitive dependence on initial conditions.

2 The tent map

- Here we study the *tent map*:

$$T(x) = \begin{cases} 2x, & x \leq \frac{1}{2}, \\ 2-2x, & x \geq \frac{1}{2}, \end{cases} \tag{2.1}$$

  on the invariant interval $[0,1]$.

Exercise 2.1. Use the function $h(x) = \sin^2\left(\frac{\pi x}{2}\right)$ (a homeomorphism on $[0,1]$) to show that T is conjugate to f_4 (the logistic map with $a = 4$).

Exercise 2.2. Find a period-3 solution of T. Hint: There are two of them. Explain how this shows that f_4 is chaotic.

- Now we define a function $h : [0,1] \to \Sigma_2$ as follows. For any $x \in [0,1]$ and $i \geq 0$, let

$$s_i = \begin{cases} L, & T^i(x) \leq \frac{1}{2}, \\ R, & T^i(x) > \frac{1}{2}. \end{cases} \tag{2.2}$$

  Then set $\mathcal{S} = h(x)$, where $\mathcal{S} = s_0 s_1 \ldots$. The sequence $h(x)$ is the symbolic representation of the forward orbit of x relative to $x = \frac{1}{2}$.
- More generally, almost any one-dimensional map can be assigned symbols in this way, where two symbols suffice for unimodal maps (although sometimes it is helpful to assign a third symbol for the critical point), and more symbols are needed for maps with two or more critical points.

Exercise 2.3. Show that

$$h(0.11) = LLL(RLLRLLLRRR)^\infty.$$

Exercise 2.4. Find $x \in [0,1]$ such that $d(h(x), h(0.4)) = 2$.

- Notice that

$$h \circ T = \sigma \circ h, \tag{2.3}$$

  on $[0,1]$. It may be shown that h^{-1} is continuous and onto but not one-to-one [1]. Thus h does not quite show that T and σ are conjugate. In this instance we say that σ is *semi-conjugate* to T, which, roughly speaking, means that T exhibits all of the dynamics of σ.

3 An introduction to kneading theory

- Here we introduce *kneading theory* (pioneered by Milnor and Thurston in the late 1970's, see [2]) which allows us to make remarkably strong statements about the dynamics of a one-dimensional map based purely on the symbol sequences associated with its critical points.
- Let $f : [0,1] \to [0,1]$ be a unimodal map with maximum $c \in (0,1)$.

- For any x that is not a *preimage* of c (meaning there does not exist $i \geq 0$ such that $f^i(x) = c$), let $h(x) = \mathcal{S} \in \Sigma_2$ where

$$s_i = \begin{cases} L, & f^i(x) < c, \\ R, & f^i(x) > c. \end{cases} \tag{3.1}$$

- For all $x \neq c$, let

$$a(x) = \begin{cases} 1, & x < c, \\ -1, & x > c. \end{cases} \tag{3.2}$$

- For any x that is not a *preimage* of c and all $n \geq 0$, let

$$\theta_n(x) = \prod_{i=0}^{n} a\big(f^i(x)\big). \tag{3.3}$$

Exercise 3.1. Show that $\theta_i(x)$ is the sign of the slope of f^i at x.

- The sequence $\theta_0(x), \theta_1(x), \ldots$ is similar to the symbol sequence $h(x)$, but $\theta_i(x) \in \{-1, 1\}$ whereas $s_i \in \{L, R\}$, and each $\theta_i(x)$ is given by a product involving previous iterates whereas s_i is given by the sign of $x - c$.
- Instead of working with sequences $\phi_0, \phi_1, \ldots$ (where $\phi_i = \theta_i(x)$ for each i), we find it more elegant to work with power series $\Phi(t) = \sum_{i=0}^{\infty} \phi_i t^i$.

Definition 3.1. Analogous to (1.1), the *distance* between any two power series $\Phi(t) = \sum_{i=0}^{\infty} \phi_i t^i$ and $\Psi(t) = \sum_{i=0}^{\infty} \psi_i t^i$ with $\phi_i, \psi_i \in \{-1, 1\}$ for all i, is

$$d(\Phi, \Psi) = \sum_{i=0}^{\infty} \frac{1}{2^i} \left\{ \begin{array}{ll} 0\,, & \phi_i = \psi_i \\ 1\,, & \phi_i \neq \psi_i \end{array} \right\}. \tag{3.4}$$

Definition 3.2. We define a *lexiographical ordering* on power series with coefficients of 1 or -1 as follows. For any distinct Φ and Ψ, let i be the smallest index for which $\phi_i \neq \psi_i$. If $\phi_i < \psi_i$ we write $\Phi \prec \Psi$, otherwise we write $\Phi \succ \Psi$.

Definition 3.3. The *kneading series* of any x that is not a preimage of c is

$$k(x, t) = \sum_{i=0}^{\infty} \theta_i(x) t^i. \tag{3.5}$$

The *kneading series* (plural!) of any x that is a preimage of c are

$$k(x_-, t) = \lim_{y \to x^-} k(y, t), \tag{3.6}$$

$$k(x_+, t) = \lim_{y \to x^+} k(y, t), \tag{3.7}$$

where the limits are taken through points y that are not preimages of c (this can always be achieved).

Exercise 3.2. Show that $k(c_-, t) = -k(c_+, t)$.

Exercise 3.3. Show that if x is not a preimage of c, then

$$k(f(x), t) = a(x)\sigma(k(x, t)).$$

Exercise 3.4. Show that as a function of x, $k(x, t)$ is decreasing in lexiographical order.

Definition 3.4. The *kneading invariant* is

$$k_f(t) = k(c_-, t). \tag{3.8}$$

- The following result highlights the convenience of working with power series over sequences.

Theorem 3.1. *The topological entropy*[1] *of f is positive if and only if $k_f(t) = 0$ for some $t \in (-1, 1)$.*

- The next result shows us how we can use the kneading invariant to determine which symbol sequences are possible for f.

Theorem 3.2. *Suppose $f(c) > c$. A series $\Phi(t)$ is equal to some kneading series of f if and only if, for all $i \geq 0$, either $k_f(t) \preceq \sigma^i(\Phi(t))$ or $-k_f(t) \succeq \sigma^i(\Phi(t))$.*

- Equivalently, $\Phi(t)$ is not a kneading series of f if and only if there exists $i \geq 0$ such that

$$-k_f(t) \prec \sigma^i(\Phi(t)) \prec k_f(t). \tag{3.9}$$

[1] A set $\Omega \subset [0, 1]$ is said to be (n, ε)-*separated* if for every distinct $x, y \in \Omega$ there exists $i \in \{0, 1, \ldots, n-1\}$ such that $|f^i(x) - f^i(y)| \geq \varepsilon$. Let $s(n, \varepsilon)$ denote the number of elements (i.e. the cardinality) of the largest (n, ε)-separated subset of $[0, 1]$. The *topological entropy* is

$$h(f) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\ln(s(n, \varepsilon))}{n},$$

and is a measure of the growth rate of the number of 'distinguishable' orbits under iteration by f.

Example 3.1. Consider

$$f_4(x) = 4x(1-x). \tag{3.10}$$

Here $c = \frac{1}{2}$, $f(c) = 1$, $f^2(c) = f^3(c) = \cdots = 0$. Consequently, $h(c_-) = LRL^\infty$. Thus $k_f(t) = 1 - t - t^2 - t^3 - \cdots$. Notice that this kneading invariant is special in the sense that there does not exist $\Psi(t)$ such that $-k_f(t) \prec \Psi(t) \prec k_f(t)$. Therefore every series $\Phi(t)$ is a kneading series of f. In other words, for every $\mathcal{S} \in \Sigma_2$ there exists $x \in [0,1]$ such that $h(x) = \mathcal{S}$.

Example 3.2. Consider $f_a = ax(1-x)$ with $a = 3.84$. Here f_a has an attracting period-3 cycle and

$$\begin{aligned} k_f(t) &= 1 - t - t^2 - t^3 + t^4 + t^5 + t^6 - t^7 + \cdots \\ &= \frac{1-t-t^2}{1+t^3}\,. \end{aligned} \tag{3.11}$$

Suppose (for a contradiction, as we will see) that there exists $x \in [0,1]$ such that $h(x) = (LRLRL)^\infty$. Then

$$\begin{aligned} k(x,t) &= 1 - t - t^2 + t^3 + t^4 + t^5 - t^6 - t^7 + \cdots \\ &= \frac{1-t-t^2+t^3+t^4}{1-t^5}\,. \end{aligned}$$

But

$$\sigma^2(k(x,t)) = -1 + t + t^2 + t^3 - t^4 - t^5 + t^6 + \cdots\ ,$$

from which we can see that (3.9) is satisfied with $i = 2$. Hence $k(x,t)$ is not a kneading series, which is a contradiction.

Exercise 3.5. Analogous to the previous example, use Theorem 3.2 to show that there does exist $x \in [0,1]$ such that $h(x) = (LRLRR)^\infty$ for the map $f_{3.84}$.

4 An introduction to measure theory

Definition 4.1. Let $\mathcal{X}$ be a set. A collection Σ of subsets of $\mathcal{X}$ is a *σ-algebra* if it

i) includes $\varnothing$,
ii) is closed under complement, and
iii) is closed under the union or intersection of countably many subsets.

Definition 4.2. A function $\mu : \Sigma \to \mathbb{R}$ is a *measure* if

i) $\mu(\varnothing) = 0$,
ii) $\mu(E) \geq 0$, for all $E \in \Sigma$, and
iii) $\sum_k \mu(E_k) = \mu(\cup_k E_k)$, for all countable disjoint collections $E_k \in \Sigma$.

Definition 4.3. Let Σ be a σ-algebra of $\mathcal{X}$ and $\mu : \Sigma \to \mathbb{R}$ a measure. The triple $(\mathcal{X}, \Sigma, \mu)$ is called a *measure space* and the elements of Σ are called *measurable sets*.

Definition 4.4. Let $(\mathcal{X}, \Sigma, \mu)$ be a measure space. A property of $\mathcal{X}$ is said to hold for μ *almost all* $\mathbf{x} \in \mathcal{X}$ (or just almost all $\mathbf{x} \in \mathcal{X}$, if it is clear which measure is being referred to) if it holds for all $\mathbf{x}$ in some subset $E \subset \mathcal{X}$ with $\mu(\mathcal{X} \setminus E) = 0$.

Definition 4.5. Let $\mathcal{X}$ be a topological space (i.e. open sets are defined). A subset $E \subset \mathcal{X}$ is called a *Borel set* if it can be formed from open sets by a countable combination of unions, intersections, and complements.

Lemma 4.1. *The collection of all Borel sets is a σ-algebra (called the Borel σ-algebra).*

Definition 4.6.

i) The *Lebesgue measure* of a box $B = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n$ is $\lambda(B) = \prod_j (b_j - a_j)$.
ii) Let $E \subset \mathbb{R}^n$ be a Borel set (this ensures that following limit exists). The *Lebesgue measure* of E is

$$\lambda(E) = \inf_{\text{covering boxes } B_k} \sum_k \lambda(B_k).$$

Lemma 4.2. *The Lebesgue measure is a measure on the Borel σ-algebra.*

Definition 4.7. Let $(\mathcal{X}, \Sigma, \mu)$ be a measure space. A function $\varphi : \mathcal{X} \to \mathbb{R}$ is said to be *measurable* if the preimage of every interval (t, ∞) belongs to Σ.

Definition 4.8. Let $(\mathcal{X}, \Sigma, \mu)$ be a measure space.

i) The *Lebesgue integral* of a simple function $s = \sum_{k=1}^m a_k \chi_{E_k}$, where $a_k > 0$ and $E_k \in \Sigma$ for each k, and χ_E is the indicator function, is

$$\int s\, d\mu = \sum_k a_k \mu(E_k).$$

ii) The *Lebesgue integral* of a non-negative measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ is

$$\int \varphi \, d\mu = \sup_{\text{non-negative simple } s \leq \varphi} \int s \, d\mu.$$

iii) The *Lebesgue integral* of a measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ is

$$\int \varphi \, d\mu = \int \varphi^+ \, d\mu - \int \varphi^- \, d\mu,$$

where

$$\varphi^+(\mathbf{x}) = \begin{cases} 0, & \varphi(\mathbf{x}) < 0, \\ \varphi(x), & \varphi(\mathbf{x}) > 0, \end{cases}$$

and $\varphi^-(\mathbf{x}) = \varphi^+(\mathbf{x}) - \varphi(\mathbf{x})$.

Definition 4.9. A measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ is said to be *Lebesgue integrable* if $\int |\varphi| \, d\mu < \infty$.

Definition 4.10. Let μ and ν be measures on a σ-algebra Σ. Then μ is said to be *absolutely continuous* (with respect to ν) if $\mu(E) = 0$ for every $E \in \Sigma$ for which $\nu(E) = 0$.

Theorem 4.3 (Radon-Nikodym)**.** *Let μ and ν be measures on a σ-algebra Σ. There exists a Lebesgue integrable function $g : \mathcal{X} \to \mathbb{R}$ such that $\mu(E) = \int_E g \, d\nu$ for all $E \in \Sigma$, if and only if μ is absolutely continuous with respect to ν.*

Definition 4.11. A *Radon-Nikodym derivative* of μ (with respect to ν) is a function $g : \mathcal{X} \to \mathbb{R}$ for which $\mu(E) = \int_E g \, d\nu$ for all $E \in \Sigma$.

- Radon-Nikodym derivatives are unique up to zero measure (with respect to ν), so we usually refer to 'the' Radon-Nikodym derivative.
- The Radon-Nikodyn derivative gives us a practical way to evaluate a Lebesgue integral. For instance if $\mathcal{X} = \mathbb{R}$ and $\mu(E) = \int_E g \, d\lambda$ for all E, then
$$\int_E \varphi \, d\mu = \int_E \varphi(x) g(x) \, dx. \tag{4.1}$$

Definition 4.12. A measure μ is said to be a *probability measure* if $\mu(\mathcal{X}) = 1$.

Example 4.1. Consider $\mathcal{X} = \mathbb{R}$ and let

$$g(x) = \frac{1}{\sqrt{2\pi\beta^2}} \, \mathrm{e}^{\frac{-(x-\alpha)^2}{2\beta^2}}, \tag{4.2}$$

be the probability density function for a normal (or Gaussian) distribution with mean α and standard deviation β. We can define a probability measure μ by $\mu(E) = \int_E g \, d\lambda$. The function g is the Radon-Nikodym derivative of μ. If $x \in \mathbb{R}$ is chosen randomly according to (4.2), then the probability that $x \in E$ is $\mu(E)$, for any measurable $E \subset \mathbb{R}$. If E is an interval $[a, b]$, then $\mu(E) = \int_a^b g(x) \, dx$.

Example 4.2. Consider $\mathcal{X} = \mathbb{R}$ and let

$$g(x) = \delta(x - \alpha), \tag{4.3}$$

be the δ-function (or Dirac δ-function) centred at α.

Strictly speaking we cannot use g to define a measure via $\mu(E) = \int_E g \, d\lambda$ because g is not a function (it is a *distribution*[2]). Above the Lebesgue integral was only defined for functions. Nevertheless we can define the intended probability measure δ_α (known as the *Dirac measure* centred at α) by

$$\delta_\alpha(E) = \begin{cases} 0, & \alpha \notin E, \\ 1, & \alpha \in E. \end{cases}$$

For any continuous $\varphi : \mathbb{R} \to \mathbb{R}$, we have $\int_{\mathbb{R}} \varphi \, d\delta_\alpha = \varphi(\alpha)$ by Definition 4.8. This is the same as

$$\int_{-\infty}^{\infty} \varphi(x) \delta(x - \alpha) \, dx = \varphi(\alpha),$$

which should be familiar to you from previous encounters with the δ-function.

Observe that δ_α is not absolutely continuous with respect to λ: with $E = \{\alpha\}$ we have $\delta_\alpha(E) = 1$ and $\lambda(E) = 0$. Thus by Theorem 4.3 there does not exist Lebesgue integrable $g : \mathbb{R} \to \mathbb{R}$ such that $\delta_\alpha(E) = \int_E g \, d\lambda$ for all measurable E.

[2] A distribution is a type of 'generalised function' not to be confused with a probability distribution.

5 Invariant measures of maps

- Given a map $f : \mathcal{X} \to \mathcal{X}$, rather than iterating a point $\mathbf{x} \in \mathcal{X}$ under f, we may wish to iterate a measure μ under f.
- That is, if $\mathbf{x}$ is a random variable distributed according to μ, we would like to determine the statistical properties of iterates of $\mathbf{x}$ under f.
- We write $f_*\mu$ to denote the *push-forward measure* given by iterating μ under f. It is straightforward to see that for any measurable $E \subset \mathcal{X}$, we have[3]

$$(f_*\mu)(E) = \mu\big(f^{-1}(E)\big), \tag{5.2}$$

Definition 5.1. Let $(\mathcal{X}, \Sigma, \mu)$ be a measure space. Let $f : \mathcal{X} \to \mathcal{X}$ be measurable[4]. Then μ is an *invariant density* of f if $\mu\big(f^{-1}(E)\big) = \mu(E)$ for all $E \in \Sigma$. We also say that f *preserves* μ.

- We can also study how the Radon-Nikodym derivative of a probability measure maps under f, which is essentially asking how a probability density function, we will just say *density*, maps under f. We write the image of a density g under f as $P_f g$, where P_f is the *Frobenius-Perron operator*. For simplicity, for the remainder of this section we consider $\mathcal{X} = \mathbb{R}$.
- By (5.2) we have

$$\int_E (P_f g)(y)\, dy = \int_{f^{-1}(E)} g(y)\, dy,$$

  for all measurable $E \subset \mathbb{R}$. Putting $E = [a, x]$ gives

$$\int_a^x (P_f g)(y)\, dy = \int_{f^{-1}([a,x])} g(y)\, dy.$$

  Then by differentiating we obtain

$$(P_f g)(x) = \frac{d}{dx} \int_{f^{-1}([a,x])} g(y)\, dy. \tag{5.3}$$

- If g is invertible and increasing then $f^{-1}([a,x]) = \big[f^{-1}(a), f^{-1}(x)\big]$ and so

$$(P_f g)(x) = g\big(f^{-1}(x)\big) \frac{d}{dx}\big(f^{-1}(x)\big). \tag{5.4}$$

Exercise 5.1. Consider the map $f(x) = x^2$ on $[0, 1]$ and the density $g(x) = 1$ (corresponding to uniform distribution). Use (5.4) to compute $(P_f g)(x)$ and $(P_f^2 g)(x)$.

- We now have the mathematical tools to make more detailed remarks about the logistic family that we first encountered in Part VII:

$$f_a(x) = ax(1 - x). \tag{5.5}$$

Exercise 5.2. Let g be a probability density on $[0, 1]$, that is $g : [0, 1] \to [0, \infty)$ is Lebesgue integrable and $\int_0^1 g(x)\, dx = 1$. Show that

$$\begin{aligned}(P_{f_4} g)(x) = \tfrac{1}{4\sqrt{1-x}} \Big(&g\big(\tfrac{1}{2} - \tfrac{1}{2}\sqrt{1-x}\big) \\ &+ g\big(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1-x}\big)\Big).\end{aligned}$$

Hint: Use (5.3).

Exercise 5.3. Show that

$$g(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \tag{5.6}$$

is an invariant probability density of f_4. Hint: To show that $g(x)$ is normalised observe that $\frac{d}{dx}\sin^{-1}(\sqrt{x}) = \frac{1}{2\sqrt{x(1-x)}}$.

Exercise 5.4. Show that

$$g(x) = \frac{1}{2}\delta\Big(x - \tfrac{5+\sqrt{5}}{8}\Big) + \frac{1}{2}\delta\Big(x - \tfrac{5-\sqrt{5}}{8}\Big),$$

is an invariant probability density of f_4 (ignoring the fact that strictly speaking g is not Lebesgue integrable).

- Recall from the bifurcation diagram of f_a that the apparently chaotic regime is interpersed with 'periodic windows' (where f_a is not chaotic). The next result (which was first proved only recently by applying quite difficult complex variables techniques to f_a as an analytic function on $\mathbb{C}$, see [3, 4]) tells us that these windows are dense.

[3]The inverse is defined as

$$f^{-1}(E) = \{\mathbf{x} \in \mathcal{X} \mid f(\mathbf{x}) \in E\}. \tag{5.1}$$

[4]This means that $f^{-1}(E) \in \Sigma$ for all $E \in \Sigma$.

Theorem 5.1. *There exists an open dense subset of values of $a \in [0,4]$ for which f_a has a periodic orbit attracting λ almost all $x \in [0,1]$.*

- Theorem 5.1 tells us that chaos in the logistic family is not 'robust'.
- It is tempting to infer from Theorem 5.1 and the fact that every periodic window covers a subset of $[0,4]$ with non-zero Lebesgue measure, that if we were to choose $a \in [0,4]$ randomly, then the probability that f_a is chaotic is zero. But this is not the case:

Theorem 5.2 (Jakobson [5])**.** *The Lebesgue measure of the set of $a \in [0,4]$ for which f_a has a positive Lyapunov exponent is positive.*

6 An introduction to ergodic theory

- This section mostly follows [6], see also [7, 8, 9, 10].
- Invariant measures define a distribution in space. However, we are often more interested in the distribution of the points of an orbit $\{f^n(\mathbf{x})\}$, which is a distribution in time. Ergodic theory connects spatial and temporal distributions and shows that, rather remarkably, they are often the same. In this section we also introduce SRB measures which provide a rigorous foundation for describing chaotic attractors.

- Throughout this section $\mathcal{X}$ denotes a compact metric space and Σ is the collection of Borel subsets of $\mathcal{X}$. We let $C(\mathcal{X})$ denote the set of all continuous functions $\varphi : \mathcal{X} \to \mathbb{R}$ and $M(\mathcal{X})$ denote the set of all probability measures on $\mathcal{X}$.

Exercise 6.1. Show that $M(\mathcal{X})$ is a convex set.

Lemma 6.1. *For any $\mu \in M(\mathcal{X})$, define a functional*[5] *$J : C(\mathcal{X}) \to \mathbb{R}$ by*

$$J(\varphi) = \int_{\mathcal{X}} \varphi \, d\mu. \tag{6.1}$$

Then J is continuous, linear[6]*, positive*[7] *and normalised*[8]*.*

Theorem 6.2 (Riesz representation theorem)**.** *Let $J : C(\mathcal{X}) \to \mathbb{R}$ be continuous, linear, positive and normalised. Then there exists unique $\mu \in M(\mathcal{X})$ such that $J(\varphi) = \int_{\mathcal{X}} \varphi \, d\mu$ for all $\varphi \in C(\mathcal{X})$.*

- We conclude from Lemma 6.1 and the Riesz representation theorem that there is a one-to-one correspondence between probability measures $\mu \in M(\mathcal{X})$ and continuous, linear, positive, normalised $J : C(\mathcal{X}) \to \mathbb{R}$.
- Let $f : \mathcal{X} \to \mathcal{X}$ be continuous. Let $\Omega \in \Sigma$ be such that $f(\Omega) \subset \Omega$ and $f(\mathcal{X} \setminus \Omega) \subset \mathcal{X} \setminus \Omega$. Then for any $\mathbf{x} \in \Omega$, the forward orbit of $\mathbf{x}$ and any backwards orbit of $\mathbf{x}$ must be contained in Ω. Similarly for any $\mathbf{x} \in \mathcal{X} \setminus \Omega$, the forward orbit of $\mathbf{x}$ and any backwards orbit of $\mathbf{x}$ must be contained in $\mathcal{X} \setminus \Omega$. In this way Ω and $\mathcal{X} \setminus \Omega$ partition $\mathcal{X}$ into two sets that do not interact under f.

Exercise 6.2. Consider the map on $\mathcal{X} = [0,1]$ given by

$$f(x) = \tfrac{1}{2} + \tfrac{3}{2}\left(x - \tfrac{1}{2}\right) - 2\left(x - \tfrac{1}{2}\right)^3. \tag{6.2}$$

Show that $[0,1]$ can be partitioned in the sense discussed above for the sets $\Omega_1 = \left[0, \tfrac{1}{2}\right)$, $\Omega_2 = \{0\}$ and $\Omega_3 = \left\{\tfrac{1}{2}\right\}$.

Exercise 6.3. Show that $f(\Omega) \subset \Omega$ and $f(\mathcal{X} \setminus \Omega) \subset \mathcal{X} \setminus \Omega$ if and only if $f^{-1}(\Omega) = \Omega$.

Definition 6.1. Let $f : \mathcal{X} \to \mathcal{X}$ be continuous. Suppose $\mu \in M(\mathcal{X})$ is invariant under f. Then μ is *ergodic* if $\mu(\Omega) \in \{0,1\}$ for every $\Omega \in \Sigma$ for which $f^{-1}(\Omega) = \Omega$.

- Given a continuous function $f : \mathcal{X} \to \mathcal{X}$, we let $M(\mathcal{X}, f)$ denote the set of all $\mu \in M(\mathcal{X})$ that are invariant under f and $E(\mathcal{X}, f)$ denote the set of all ergodic $\mu \in M(\mathcal{X}, f)$.
- Ergodic measures are those that cannot be decomposed: $\mu \in M(\mathcal{X}, f)$ is ergodic if and only if there do not exist $\nu_1, \nu_2 \in M(\mathcal{X}, f)$ and $a \in (0,1)$ such that $\mu = (1-a)\nu_1 + a\nu_2$[9].

[5] A functional is a function of functions!

[6] Meaning $J(a\varphi + b\psi) = aJ(\varphi) + bJ(\psi)$ for all $\varphi, \psi \in C(\mathcal{X})$ and all $a, b \in \mathbb{R}$.

[7] Meaning $J(\varphi) \geq 0$ whenever $\varphi \geq 0$.

[8] Meaning $J(\mathbf{1}) = 1$.

[9] This can be visualised by treating $M(\mathcal{X})$ as a metric space (not done here as it is relatively technical: one can first treat $M(\mathcal{X})$ as a topological space by using the *weak* topology* and then constructing a metric from this). Then $M(\mathcal{X}, f)$ is a non-empty, compact, convex subset of $M(\mathcal{X})$, and $E(\mathcal{X}, f)$ is the boundary of $M(\mathcal{X}, f)$.

- The next result shows that any $\mu \in M(\mathcal{X}, f)$ can be decomposed into ergodic measures.

Theorem 6.3 (Ergodic decomposition)**.** *For all $\mu \in M(\mathcal{X}, f)$, there exists a unique measure τ on $M(\mathcal{X}, f)$ with $\tau(E(\mathcal{X}, f)) = 1$ such that*

$$\int_{\mathcal{X}} \varphi \, d\mu = \int_{E(\mathcal{X},f)} \int_{\mathcal{X}} \varphi \, d\sigma \, d\tau(\sigma), \tag{6.3}$$

for all $\varphi \in C(\mathcal{X})$.

- Let $\mu \in M(\mathcal{X}, f)$ and $\varphi \in C(\mathcal{X})$. Then $\int_{\mathcal{X}} \varphi \, d\mu$ is the 'space average' of φ with respect to μ. In contrast, for any $\mathbf{x} \in \mathcal{X}$, the quantity $\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi\big(f^i(\mathbf{x})\big)$ is the 'time average' of φ under iteration by f.

Theorem 6.4 (Birkhoff's ergodic theorem)**.** *For any $\mu \in E(\mathcal{X}, f)$ and μ almost all $\mathbf{x} \in \mathcal{X}$, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi\big(f^i(\mathbf{x})\big) = \int_{\mathcal{X}} \varphi \, d\mu, \tag{6.4}$$

for all $\varphi \in C(\mathcal{X})$.

Example 6.1. Consider again $f(x) = 4x(1-x)$. Here we will use Birkhoff's ergodic theorem to show that the Lyapunov exponent $\chi(x)$ of λ almost all $x \in [0,1]$ is $\ln(2)$ (the calculation can also be achieved in a less computationally intensive way by converting to the tent map).

Lyapunov exponents for N-dimensional maps were defined in Part VII. In one-dimension there is only one Lyapunov exponent:

$$\begin{aligned} \chi(x) &= \lim_{n\to\infty} \frac{1}{n} \ln\big(\big|(f^n)'(x)\big|\big) \\ &= \lim_{n\to\infty} \frac{1}{n} \ln\left(\prod_{i=0}^{n-1} \big|f'\big(f^i(x)\big)\big| \right), \end{aligned} \tag{6.5}$$

where the primes denote differentiation with respect to x. Equation (6.5) can be rewritten as

$$\chi(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi\big(f^i(x)\big), \tag{6.6}$$

where $\varphi(y) = \ln|f'(y)|$. Let $g(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ be the invariant density of Exercise 5.3. It is straightforward to see that the corresponding measure, call it μ, is ergodic. Also here λ almost all $x \in [0,1]$ is the same as μ almost all $x \in [0,1]$. Thus we can apply Birkhoff's ergodic theorem to say that for λ almost all $x \in [0,1]$,

$$\begin{aligned} \chi(x) &= \int_{[0,1]} \varphi \, d\mu \\ &= \int_0^1 \varphi(y) g(y) \, dy \\ &= \frac{2}{\pi} \int_0^{\frac{1}{2}} \frac{\ln(4-8y)}{\sqrt{y(1-y)}} \, dy. \end{aligned}$$

Under the integral substitution $z = \sin^{-1}(1-2y)$ this becomes

$$\chi(x) = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \ln(4\sin(z)) \, dz.$$

This can be rewritten as $\chi(x) = 2\ln(2) + I$, where $I = \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \ln(\sin(z)) \, dz$. With a little algebraic wizardry[10] one can show that $I = -\ln(2)$ and thus the Lyapunov exponent is $\chi(x) = \ln(2)$.

Definition 6.2. The *basin* of $\mu \in M(\mathcal{X}, f)$, denoted $\mathscr{B}(\mu)$, is the set of all $\mathbf{x} \in \mathcal{X}$ for which (6.4) holds for all $\varphi \in C(\mathcal{X})$.

Definition 6.3. A measure $\mu \in M(\mathcal{X}, f)$ on a measure space $(\mathcal{X}, \Sigma, \nu)$ is said to be *physical* if $\nu(\mathscr{B}(\mu)) > 0$.

- Often a map has more than one ergodic measure (in fact possibly uncountably many [11]). Ergodic physical measures are the ones that we will 'see' in practice and are analogous to Milnor attractors.
- The next definition follows that given in [12].

[10] The integral I is the same as $\frac{2}{\pi} \int_0^{\frac{\pi}{2}} \ln(\cos(z)) \, dz$, thus we can write $I = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} (\ln(\sin(z)) + \ln(\cos(z))) \, dz$. We have

$$\ln(\sin(z)) + \ln(\cos(z)) = \ln(\sin(z)\cos(z)) = \ln\left(\frac{1}{2}\sin(2z)\right) = \ln(\sin(2z)) - \ln(2),$$

and thus

$$I = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \ln(\sin(2z) \, dz - \frac{1}{2}\ln(2) = \frac{1}{2\pi} \int_0^{\pi} \ln(\sin(z) \, dz - \frac{1}{2}\ln(2) = \frac{1}{2} I - \frac{1}{2}\ln(2).$$

By solving for I we obtain $I = -\ln(2)$.

Definition 6.4. A measure $\mu \in M(\mathcal{X}, f)$ is called an *SRB (Sinai-Ruelle-Bowen) measure* if it has a positive Lyapunov exponent (almost everywhere) and has absolutely continuous conditional measures on its unstable manifolds.

- If $\mathbf{x}^*$ is an asymptotically stable fixed point of f, then the corresponding Dirac measure $\delta_{\mathbf{x}^*}$ is physical. However $\delta_{\mathbf{x}^*}$ is not absolutely continuous and therefore not an SRB measure.

Theorem 6.5. *Every ergodic SRB measure with no zero Lyapunov exponent is physical.*

Definition 6.5. A measure $\mu \in M(\mathcal{X}, f)$ is said to be *mixing* if

$$\mu\big(\Omega \cap f^{-n}(\Psi)\big) \to \mu(\Omega)\mu(\Psi) \text{ as } n \to \infty, \quad (6.7)$$

for all $\Omega, \Psi \in \Sigma$.

- Roughly speaking, $\frac{\mu(\Omega \cap f^{-n}(\Psi))}{\mu(\Omega)}$ is the fraction of points in Ω that map to Ψ under f^n, assuming $\mu(\Omega) \neq 0$. The measure μ is mixing if this fraction converges to the fraction of points in $\mathcal{X}$ that belong to Ψ.

Lemma 6.6. *If $\mu \in M(\mathcal{X}, f)$ is mixing then it is ergodic.*

Proof. Choose any $\Omega \in \Sigma$ for which $f^{-1}(\Omega) = \Omega$. Then for any $n \geq 1$,

$$(\mathcal{X} \setminus \Omega) \cap f^{-n}(\Omega) = (\mathcal{X} \setminus \Omega) \cap \Omega = \varnothing\,.$$

Thus if μ is mixing then $\mu(\mathcal{X} \setminus \Omega)\mu(\Omega) = 0$. Thus either $\mu(\mathcal{X} \setminus \Omega) = 0$ or $\mu(\Omega) = 0$, hence μ is ergodic. □

Theorem 6.7. *Every ergodic SRB measure with no zero Lyapunov exponent is mixing.*

Definition 6.6. Let $\varphi \in C(\mathcal{X})$ and $\mathbf{x}$ be a random variable distributed according to some $\mu \in M(\mathcal{X}, f)$. The *mean* and *variance* of $\varphi(\mathbf{x})$ are

$$\mathbb{E}[\varphi(\mathbf{x})] = \int_{\mathcal{X}} \varphi\, d\mu\,, \quad (6.8)$$

$$\mathrm{Var}(\varphi(\mathbf{x})) = \mathbb{E}[\varphi^2(\mathbf{x})] - \mathbb{E}[\varphi(\mathbf{x})]^2\,. \quad (6.9)$$

Exercise 6.4. Consider $\varphi(x) = \sqrt{x}$ for $\mathcal{X} = [0, 1]$.

i) Show that $\mathbb{E}[\varphi(x)] = \frac{2}{3}$ and $\mathrm{Var}(\varphi(x)) = \frac{1}{18}$ for the Lebesgue measure λ.

ii) Show that $\mathbb{E}[\varphi(x)] = \frac{2}{\pi}$ and $\mathrm{Var}(\varphi(x)) = \frac{\pi^2-8}{2\pi^2}$ for the measure with density $g(x) = \frac{1}{\pi\sqrt{x(1-x)}}$.

- If $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ are statistically *independent* random variables distributed according to $\mu \in M(\mathcal{X}, f)$, then

$$\mathbb{E}[\varphi(\mathbf{x})\varphi(\mathbf{y})] = \mathbb{E}[\varphi(\mathbf{x})]^2. \quad (6.10)$$

  This helps motivate the following definition.

Definition 6.7. Let $\varphi \in C(\mathcal{X})$ and $\mathbf{x}$ be a random variable distributed according to some $\mu \in M(\mathcal{X}, f)$. The *autocorrelations* of $\varphi(\mathbf{x})$ are

$$R(n) = \frac{\mathbb{E}[\varphi(\mathbf{x})\varphi(f^n(\mathbf{x}))] - \mathbb{E}[\varphi(\mathbf{x})]^2}{\mathrm{Var}(\varphi(\mathbf{x}))}, \quad (6.11)$$

where $n \geq 0$.

- Notice $R(0) = 1$. Also, if μ is mixing then $R(n) \to 0$ as $n \to \infty$.

Exercise 6.5. Let $\varphi \in C(\mathcal{X})$ and $\mathbf{x}$ be a random variable distributed according to $\mu \in M(\mathcal{X}, f)$. Suppose μ is mixing. Let

$$D_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=0}^{n-1}\varphi\big(f^i(\mathbf{x})\big) - \mathbb{E}[\varphi(\mathbf{x})]\right), \quad (6.12)$$

for all $n \geq 0$. Notice that $\frac{1}{\sqrt{n}}D_n \to 0$ as $n \to \infty$ by Birkhoff's ergodic theorem. Here we determine the basic statistical properties of D_n in order understand the rate at which $\frac{1}{\sqrt{n}}D_n \to 0$[11].

i) By directly expanding terms, show that

$$\mathbb{E}\big[D_n^2\big] = \mathrm{Var}(\varphi(\mathbf{x}))\left(R(0) + \frac{2}{n}\sum_{i=1}^{n-1}(n-i)R(i)\right).$$

ii) Let $\xi^2 = R(0) + 2\sum_{i=1}^{\infty} R(i)$, assuming that this series converges. Show that

$$\mathbb{E}\big[D_n^2\big] \to \mathrm{Var}(\varphi(\mathbf{x}))\xi^2,$$

as $n \to \infty$.

iii) Use the central limit theorem to argue that D_n converges in distribution to a normal distribution with mean 0 and variance $\mathrm{Var}(\varphi(\mathbf{x}))\xi^2$.

[11]See also *http://www.scholarpedia.org/article/Decay_of_correlations*.

References

[1] R.C. Robinson. *An Introduction to Dynamical Systems. Continuous and Discrete.* Prentice Hall, Upper Saddle River, NJ, 2004.

[2] J. Milnor and W. Thurston. On iterated maps of the interval. In A. Dold and B. Eckmann, editors, *Lect. Notes. in Math.*, volume 1342, pages 465–563, New York, 1988. Springer-Verlag.

[3] J. Graczyk and G. Swiatek. Generic hyperbolicity in the logistic family. *Ann. Math.*, 146(1):1–52, 1997.

[4] M. Lyubich. Dynamics of quadratic polynomials, I-II. *Acta. Math.*, 178:185–297, 1997.

[5] M.V. Jakobson. Absolutely continuous invariant measures for one-parameter families of one-dimensional maps. *Commun. Math. Phys.*, 81:39–88, 1981.

[6] P. Walters. *An Introduction to Ergodic Theory.* Springer-Verlag, New York, 1982.

[7] R. Mañé. *Ergodic Theory and Differentiable Dynamics.* Springer-Verlag, New York, 1987.

[8] S. Luzzatto. Introduction to smooth ergodic theory. *Unpublished.*, 2014.

[9] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems.* Cambridge University Press, New York, 1995.

[10] S. Kalikow and R. McCutcheon. *An Outline of Ergodic Theory.*, volume 122 of *Cambridge studies in advanced mathematics.* Cambridge University Press, Cambridge, 2010.

[11] J.-P. Eckmann and D. Ruelle. Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.*, 57(3):617–656, 1985.

[12] L.-S. Young. What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.*, 108:733–754, 2002.