Chapter 4

Introduction to Dynamics

4.1 Deterministic Dynamical Systems

Having covered programming and metric spaces in some depth, we now possess ample tools for analysis of dynamics. After starting with deterministic dynamical systems, setting up the basic theory and the notion of stability, we turn to stochastic models, where evolution of the state variable is affected by noise. While deterministic systems are clearly a kind of stochastic system (with zero-variance noise), we will see that the converse is also true: Stochastic models can be embedded in the deterministic framework. Through this embedding we can study the dynamic properties of stochastic systems using our knowledge of the deterministic model.

4.1.1 The Basic Model

Suppose that we are observing the time path of some variable x in a metric space S. At t, the current *state* of the system is denoted by x_t. Assume that from the current state x_t we can compute the time $t+1$ value x_{t+1} by applying a map h. That is, $x_{t+1} = h(x_t)$. The two primitives that make up this system are S and h:

Definition 4.1.1 A *dynamical system* is a pair (S, h), where $S = (S, \rho)$ is an arbitrary metric space and h is a map from S into itself.

By the *n-th iterate* of $x \in S$ under h we mean $h^n(x)$. It is conventional to set $h^0(x) := x$. The *trajectory* of $x \in S$ under h is the sequence $(h^t(x))_{t \geq 0}$. As before, $x^* \in S$ is a fixed point of h in S if $h(x^*) = x^*$. Fixed points are also said to be *stationary* or *invariant* under h.[1]

[1]Similar terminology applies to sets. For example, if $h(A) \subset A$, then A is said to be *invariant* under h.

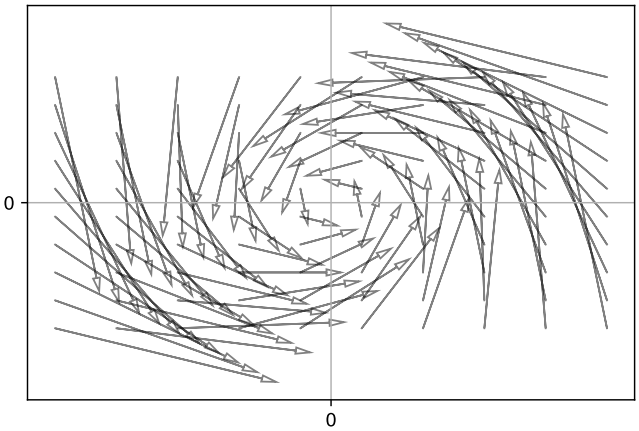

Figure 4.1 The result of mapping $x \mapsto h(x)$ for a grid of x values

Figure 4.1 illustrates the dynamics of one particular map h on $S := \mathbb{R}^2$ by showing an arrow from x to $h(x)$ for $x \in$ a grid of points. Details on the map are in the section of the Jupyter code book corresponding to this chapter.

Exercise 4.1 Show that if (S, h) is a dynamical system, if $x' \in S$ is the limit of some trajectory (i.e., $h^t(x) \to x'$ as $t \to \infty$ for some $x \in S$), and if h is continuous at x', then x' is a fixed point of h.

Exercise 4.2 Prove that if h is continuous on S and $h(A) \subset A$ (i.e., h maps $A \to A$), then $h(\operatorname{cl} A) \subset \operatorname{cl} A$.

Let x^* be a fixed point of h on S. By the *stable set* $\Lambda(x^*)$ of x^* we refer to all $x \in S$ such that $\lim_{t\to\infty} h^t(x) = x^*$. Clearly, $\Lambda(x^*)$ is nonempty. (Why?) The fixed point x^* is said to be *locally stable*, or an *attractor*, whenever there exists an open set G with $x^* \in G \subset \Lambda(x^*)$. Equivalently x^* is locally stable whenever there exists an ϵ-ball around x^* such that every trajectory starting in that ball converges to x^*:

Exercise 4.3 Prove that x^* is locally stable if and only if there exists an $\epsilon > 0$ such that $B(\epsilon, x^*) \subset \Lambda(x^*)$.

In this book we will be interested primarily in *global* stability:

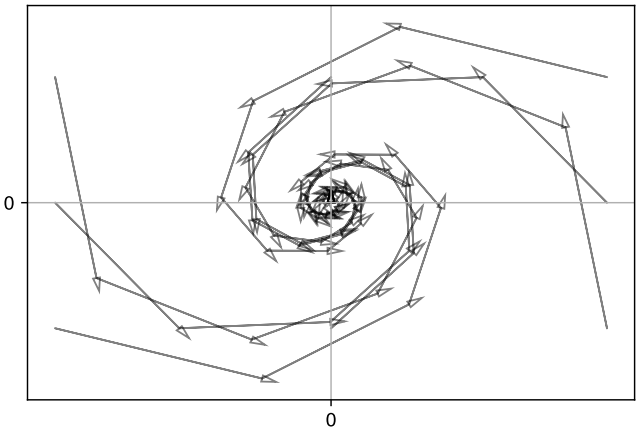

Figure 4.2 Global stability

Definition 4.1.2 A dynamical system (S, h) is called *globally stable* or *asymptotically stable* if

1. h has a *unique* fixed point x^* in S, and
2. $h^t(x) \to x^*$ as $t \to \infty$ for all $x \in S$.

Exercise 4.4 Prove that if x^* is a fixed point of (S, h) to which every trajectory converges, then x^* is the only fixed point of (S, h).

Figure 4.2 helps to visualize the concept of global stability, plotting 9 individual trajectories of a stable map h on $\mathbb{R}^2$. The details are in the Jupyter code book (see page xiv).

Figure 4.3 also illustrates global stability, in this case for the one-dimensional system (S, h), where $S := (0, \infty)$ and $h(k) := sAk^\alpha$ with $s \in (0, 1]$, $A > 0$ and $\alpha \in (0, 1)$. The system represents a simple Solow–Swan growth model, where next period's capital stock $h(k)$ is the savings rate s times current output Ak^α. The value A is a productivity parameter and α is the capital intensity. Figure 4.3 is called a 45 degree diagram. When the curve h lies above (resp., below) the 45 degree line we have $h(k) > k$ (resp., $h(k) < k$), and hence the trajectory moves to the right (resp., left). Two trajectories are shown, converging to the unique fixed point k^*.

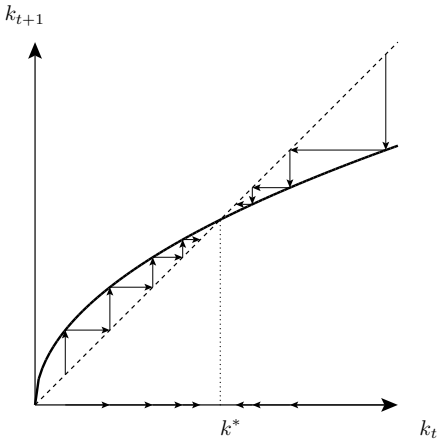

Figure 4.3 45 degree diagram

Regarding local stability of (S, h) when S is an open subset of $\mathbb{R}$, it is well-known that

Lemma 4.1.3 *If h is a map with continuous derivative h' and x^* is a fixed point of h with $|h'(x^*)| < 1$, then x^* is locally stable.*

Intuitively, when the condition holds, $h(x) \approx h(x^*) + h'(x^*)(x - x^*)$ is locally uniformly contracting in the neighborhood of x^*.

Example 4.1.4 Consider a growth model with "threshold" nonconvexities of the form $k_{t+1} = sA(k_t)k_t^\alpha$, where $s, \alpha \in (0,1)$ and $k \mapsto A(k)$ is some increasing function with $A(k) > 0$ when $k > 0$. Suppose, for example, that A is a step function of the form

$$A(k) = \begin{cases} A_1 & \text{if } 0 < k < k_b \\ A_2 & \text{if } k_b \leq k < \infty \end{cases}$$

Here k_b is a "threshold" value of capital stock, and $0 < A_1 < A_2$. Let k_i^* be the solution to $k = sA_i k^\alpha$ for $i = 1, 2$ when it exists. A plot is given in figure 4.4 for the case where $k_1^* < k_b < k_2^*$. The two fixed points k_1^* and k_2^* are local attractors, as can be verified from lemma 4.1.3. Long-run outcomes depend on initial conditions, and for this reason the model is said to exhibit *path dependence*.

Exercise 4.5 A dynamical system (S, h) is called *Lagrange stable* if every trajectory is precompact in S. In other words, the set $\{h^n(x) : n \in \mathbb{N}\}$ is precompact for every

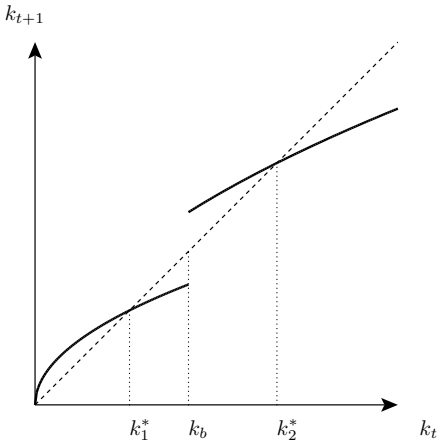

Figure 4.4 Threshold externalities

$x \in S$.[2] Show that if S is a closed and bounded subset of $\mathbb{R}^k$, then (S, h) is Lagrange stable for any choice of h.

Exercise 4.6 Let $S = \mathbb{R}$, and let $h\colon \mathbb{R} \to \mathbb{R}$ be an increasing function, in the sense that if $x \leq y$, then $h(x) \leq h(y)$. Show that every trajectory of h is a monotone sequence in $\mathbb{R}$ (either increasing or decreasing).

Exercise 4.7 Now order points in $\mathbb{R}^n$ by setting $x \leq y$ whenever $x_i \leq y_i$ for i in $\{1, \ldots, n\}$ (i.e., each component of x is dominated by the corresponding component of y). Let $S = \mathbb{R}^n$, and let $h\colon S \to S$ be monotone increasing. (The definition is the same.) Show that the same result no longer holds—h does not necessarily generate monotone trajectories.

4.1.2 Global Stability

Global stability will be a key concept for the remainder of the text. Let's start our investigation of global stability by looking at linear (more correctly, affine) systems in one dimension.

[2] Equivalently every subsequence of the trajectory has a convergent subsubsequence.

Exercise 4.8 Let $S = (\mathbb{R}, |\cdot|)$ and $h(x) = ax + b$. Prove that

$$h^t(x) = a^t x + b \sum_{i=0}^{t-1} a^i \qquad (x \in S,\ t \in \mathbb{N})$$

From this expression, prove that (S, h) is globally stable whenever $|a| < 1$, and exhibit the fixed point.

Exercise 4.9 Show that the condition $|a| < 1$ is also necessary, in the sense that if $|a| \geq 1$, then (S, h) is not globally stable. Show, in particular, that $h^t(x_0)$ converges to $x^* := b/(1-a)$ only if $x_0 = x^*$.

In exercise 4.8 we found a direct proof of global stability for our affine system when $|a| < 1$. For more complex systems direct methods are usually unavailable, and we must deploy more powerful machinery, such as Banach's fixed point theorem (theorem 3.2.16 on page 57).

Exercise 4.10 Let (S, h) be as in exercise 4.8. Using theorem 3.2.16, prove that (S, h) is globally stable whenever $|a| < 1$.

Exercise 4.11 Let $S := (0, \infty)$ with $\rho(x, y) := |\ln x - \ln y|$. Prove that ρ is a metric on S and that (S, ρ) is a complete metric space. Consider the growth model $k_{t+1} = h(k_t) = sAk_t^\alpha$ in figure 4.3, where $s \in (0, 1]$, $A > 0$ and $\alpha \in (0, 1)$. Convert this into a dynamical system on (S, ρ), and prove global stability using theorem 3.2.16.

Next we consider linear systems in $\mathbb{R}^n$. In general, a function $h\colon \mathbb{R}^n \to \mathbb{R}^n$ is called linear if

$$h(\alpha x + \beta y) = \alpha h(x) + \beta h(y) \qquad \forall\, x, y \in \mathbb{R}^n \quad \forall\, \alpha, \beta \in \mathbb{R} \tag{4.1}$$

It can be shown that every such h is continuous. If E is an $n \times n$ matrix, then the map on $\mathbb{R}^n$ defined by $x \mapsto Ex$ is linear. In fact it can be shown that for *all* linear maps $h\colon \mathbb{R}^n \to \mathbb{R}^n$ there exists a matrix E_h with $h(x) = E_h x$ for all $x \in \mathbb{R}^n$. An *affine* system on $\mathbb{R}^n$ is a map $h\colon \mathbb{R}^n \to \mathbb{R}^n$ given by

$$h(x) = Ex + b \ \text{ where } E \text{ is an } n \times n \text{ matrix and } b \in \mathbb{R}^n$$

To investigate this system, let $\|\cdot\|$ be any norm on $\mathbb{R}^n$, and define

$$\lambda := \max\{\|Ex\| : x \in \mathbb{R}^n,\ \|x\| = 1\} \tag{4.2}$$

Exercise 4.12 If you can, prove that the maximum exists. Using the properties of norms and linearity of E, show that $\|Ex\| \leq \lambda \|x\|$ for all $x \in \mathbb{R}^n$. Show in addition that if $\lambda < 1$, then $(\mathbb{R}^n, h)$ is globally stable.

Let's look at an application of these ideas. Carvalho and Tahbaz-Salehi (2019) study production networks by building on earlier work due to Long and Plosser (1983), who study business cycles using multisector growth models. Long and Plosser solve their model to obtain a system for log output given by $y_{t+1} = Ay_t + b$. Here $A = (a_{ij})$ is a matrix of input/output elasticities across sectors, and y_t is a 6×1 vector recording output in agriculture, mining, construction, manufacturing, transportation and services. Using cost share data and the hypothesis of perfect competition, the authors calculate A to be given by

$$A = (a_{ij}) = \begin{pmatrix} 0.45 & 0.00 & 0.01 & 0.21 & 0.10 & 0.16 \\ 0.00 & 0.09 & 0.04 & 0.17 & 0.05 & 0.49 \\ 0.00 & 0.01 & 0.00 & 0.42 & 0.12 & 0.09 \\ 0.06 & 0.03 & 0.01 & 0.46 & 0.06 & 0.13 \\ 0.00 & 0.00 & 0.02 & 0.12 & 0.10 & 0.32 \\ 0.02 & 0.02 & 0.06 & 0.20 & 0.09 & 0.38 \end{pmatrix}$$

Exercise 4.13 Prove that Long and Plosser's system is stable in the following way: Let $A = (a_{ij})$ be an $n \times n$ matrix where the sum of any of the rows of A is strictly less than 1 (i.e., $\max_i \alpha_i < 1$, where $\alpha_i := \sum_j |a_{ij}|$). Using the norm $\|\cdot\|_\infty$ in (4.2), show that for A we have $\lambda < 1$. Now argue that in Long and Plosser's model, (y_t) converges to a limit y^*, which is independent of initial output y_0, and, moreover, is the unique solution to the equation $y^* = Ay^* + b$.[3]

Exercise 4.14 Let $B = (b_{ij})$ be an $n \times n$ matrix where the sum of any of the *columns* of B is strictly less than 1 (i.e., $\max_j \beta_j < 1$, where $\beta_j := \sum_i |b_{ij}|$). Using the norm $\|\cdot\|_1$ in (4.2), show that for B we have $\lambda < 1$. Conclude that if $h(x) = Bx + b$, then $(\mathbb{R}^n, h)$ is globally stable.

The following results will be needed later in the text:

Exercise 4.15 Suppose that h is uniformly contracting on complete space S, so (S, h) is globally stable. Prove that if $A \subset S$ is nonempty, closed and invariant under h (i.e., $h(A) \subset A$), then the fixed point of h lies in A.

Lemma 4.1.5 *Let (S, h) be a dynamical system. If h is nonexpansive and (S, h^N) is globally stable for some $N \in \mathbb{N}$, then (S, h) is globally stable.*

Proof. By hypothesis, h^N has a unique fixed point x^* in S, and $h^{kN}(x) \to x^*$ as $k \to \infty$ for all $x \in S$. Pick any $\epsilon > 0$, and choose $k \in \mathbb{N}$ so that $\rho(h^{kN}(h(x^*)), x^*) < \epsilon$. Then

$$\rho(h(x^*), x^*) = \rho(h(h^{kN}(x^*)), x^*) = \rho(h^{kN}(h(x^*)), x^*) < \epsilon$$

[3] You are proving d_∞-convergence of trajectories, but this is equivalent to d_2-convergence by theorem 3.2.14.

It follows that x^* is a fixed point of h. (Why?)

Stability: Fix $x \in S$ and $\epsilon > 0$. Choose $k \in \mathbb{N}$ so that $\rho(h^{kN}(x), x^*) < \epsilon$. Then nonexpansiveness implies that, for each $n \geq kN$,

$$\rho(h^n(x), x^*) = \rho(h^{n-kN}(h^{kN}(x)), x^*) \leq \rho(h^{kN}(x), x^*) < \epsilon$$

In other words, (S, h) is globally stable. □

4.1.3 Chaotic Dynamic Systems

In this section we make a brief foray into chaotic (or complex) dynamical systems. Chaotic dynamics is a field that initially benefited and then suffered from excessive hype. Nonetheless, it retains great practical significance in various branches of science.

To begin, consider first the dynamical system (S, h) defined by

$$h(x) = 4x(1 - x) \qquad (x \in S := [0, 1]) \tag{4.3}$$

The function h is called the quadratic (or logistic) map and is often found in biological models related to population dynamics. Readers can check that h maps S into itself.

In the previous section we defined global stability. For these systems all trajectories converge to a single point, so long series will have an average value close to that point. Other systems can have several attractors, so the point where the trajectory settles down to depends on the initial condition. We will see that for (4.3) dynamics are still more complicated.

Figure 4.5 shows one trajectory starting at initial condition 0.11. Code used to generate the figure can be found in the Jupyter code book.

Notice that in figure 4.5 the trajectory continues to traverse through the space without settling down. Some experimentation shows that this happens for many initial conditions (but not all—does the map have any fixed points?). Moreover a slight variation in the initial condition typically leads to a time series that bears no clear resemblance to the previous one.

Science and mathematics are all about simplification and reduction. For example, with a globally stable system we can usually focus our attention on the steady state. (How does this state fit with the data?) From this perspective figure 4.5 is a little distressing. Unless the initial conditions are very special and can be known exactly, it seems that long run outcomes cannot be predicted.[4] However, this conclusion is too pessimistic, as the next exercise illustrates.

[4]Which is problematic for a scientific study—what falsifiable implications can be drawn from these models?

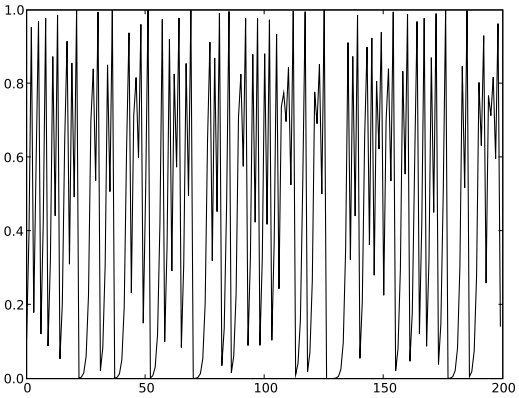

Figure 4.5 Trajectory of the quadratic map

Exercise 4.16 Using your preferred plotting tool, histogram some trajectories generated by the quadratic map, starting at different initial conditions. Use relatively long trajectories (e.g., around 5,000 points), and a fine histogram (about 40 bins). What regularities can you observe?

Incidentally, the time series in figure 4.5 looks quite random, and in exercise 4.16 we treated the trajectory in a "statistical" way, by computing its histogram. Is there in fact any formal difference between this kind of complex dynamics and the dynamics produced in systems perturbed by random variables?

One answer was proposed by Kolmogorov, who suggested measuring the "randomness" of a string of numbers by the size of the shortest computer program that can replicate it.[5] The upper bound of this measure is the size of the string itself because, if necessary, one can simply enumerate the string. This upper bound is associated with complete randomness. On the other hand, our code used to produce the time series for the quadratic map was only a few lines, and therefore has a low Kolmogorov score. In this sense we can differentiate it from a random string.

How does the quadratic map behave when we let the multiplicative parameter take values other than 4? Consider the more general map $x \mapsto rx(1-x)$, where $0 \leq r \leq 4$. A subset of these maps is plotted in figure 4.6, along with a 45 degree line. More curvature corresponds to greater r. It turns out that for some values of r this system is globally stable. For others, like 4, the behavior is highly complex.

[5] Put differently, by how much can we *compress* such a string of numbers?

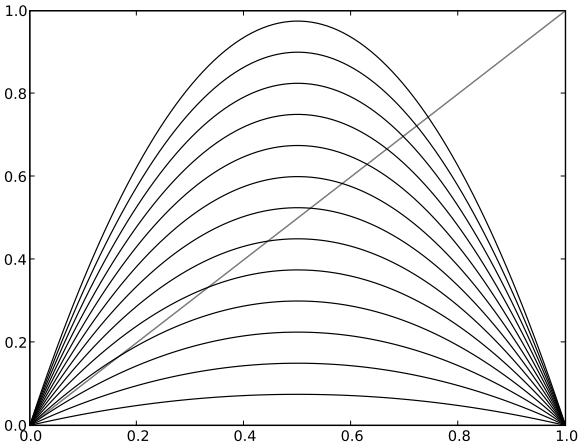

Figure 4.6 Quadratic maps, $r \in [0, 4]$

The *bifurcation diagram* shown in figure 4.7 helps to give an understanding of the dynamics. On the x-axis the parameter r ranges from 2.7 to 4. The y-axis corresponds to the state space S. For each value r in a grid over $[2.7, 4]$, a trajectory of length 1000 was generated. The first 950 points were discarded, and the last 50 were plotted. For $r \leq 3$, interior points converge to a unique interior steady state. For $r \in (3, 1 + \sqrt{6}]$, the state eventually oscillates between two "periodic attractors." From there the number of periodic attractors increases rapidly, and the behavior of the system becomes correspondingly more "chaotic."

Exercise 4.17 Reproduce figure 4.7 using your preferred computing environment.

4.1.4 Equivalent Dynamics and Linearization

In general, nonlinear models are much more difficult to analyze than linear models, leading researchers to approximate nonlinear models with linearized versions. The latter are usually obtained by a first-order Taylor expansion. Since fixed points are the natural focus of analysis, it is standard to take expansions around fixed points.

Let us see how this is done in the one-dimensional case. Let (S, h) be a dynamical system where S is an open subset of $\mathbb{R}$, and h is continuously differentiable, with derivative h' on S. Pick any $a \in S$. The first-order Taylor expansion around a is the

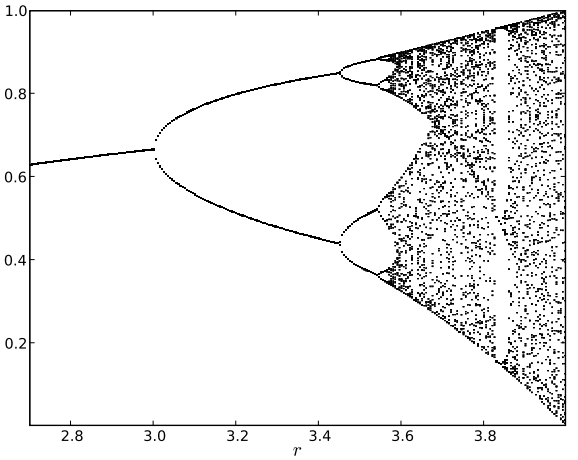

Figure 4.7 Bifurcation diagram

map h_1 defined by

$$h_1(x) = h(a) + h'(a)(x - a) \tag{4.4}$$

Notice that h_1 is an affine function on $\mathbb{R}$ with $h(a) = h_1(a)$. Clearly, h_1 approximates h in some sense when $|x - a|$ is small. For this reason it is regarded as a "linear" approximation to h around a.

Now let x^* be a fixed point of h, so

$$h_1(x) = x^* + h'(x^*)(x - x^*) \tag{4.5}$$

You can check that x^* is also a fixed point of the approximating map h_1. Note also that x^* will be stable for h_1 whenever $|h'(x^*)| < 1$. But this is precisely the condition for x^* to be a local attractor for h (lemma 4.1.3). So it seems that we can learn something about how $h^t(x)$ will behave when $|x - x^*|$ is small by studying the simple affine map h_1 and the trajectory $h_1^t(x)$ that it generates.

The well-known Hartman–Grobman theorem formalizes this idea. To state the theorem, it is necessary to introduce the abstract but valuable notion of topological conjugacy. First, let S and $\hat{S}$ be two metric spaces. A function τ from S to $\hat{S}$ is called a *homeomorphism* if it is continuous, a bijection, and its inverse τ^{-1} is also continuous. Two dynamical systems (S, g) and $(\hat{S}, \hat{g})$ are said to be *topologically conjugate* if there exists a homeomorphism τ from S into $\hat{S}$ such that g and $\hat{g}$ commute with τ in the sense that $\hat{g} = \tau \circ g \circ \tau^{-1}$ everywhere on $\hat{S}$. In other words, shifting a point $\hat{x} \in \hat{S}$ to

$\hat{g}(\hat{x})$ using the map $\hat{g}$ is equivalent to moving $\hat{x}$ into S with τ^{-1}, applying g, and then moving the result back using τ:

$$\begin{array}{ccc} x & \xrightarrow{\;g\;} & g(x) \\ \uparrow{\scriptstyle \tau^{-1}} & & \downarrow{\scriptstyle \tau} \\ \hat{x} & \xrightarrow{\;\hat{g}\;} & \hat{g}(\hat{x}) \end{array}$$

Exercise 4.18 Let $S := ((0,\infty), |\cdot|)$ and $\hat{S} := (\mathbb{R}, |\cdot|)$. Let $g(x) = Ax^{\alpha}$, where $A > 0$ and $\alpha \in \mathbb{R}$, and let $\hat{g}(\hat{x}) = \ln A + \alpha\hat{x}$. Show that g and $\hat{g}$ are topologically conjugate under $\tau := \ln$.

Exercise 4.19 Show that if (S, g) and $(\hat{S}, \hat{g})$ are topologically conjugate, then $x \in S$ is a fixed point of g on S if and only if $\tau(x) \in \hat{S}$ is a fixed point of $\hat{g}$ on $\hat{S}$.

Exercise 4.20 Let $x^* \in S$ be a fixed point of g and let x be any point in S. Show, in addition, that $\lim_{t\to\infty} g^t(x) = x^*$ if and only if $\lim_{t\to\infty} \hat{g}^t(\tau(x)) = \tau(x^*)$.

Exercise 4.21 Let $x^* \in S$ be a fixed point of g. Show that if x^* is a local attractor for (S, g), then $\tau(x^*)$ is a local attractor for $(\hat{S}, \hat{g})$. Show that if (S, g) is globally stable, then $(\hat{S}, \hat{g})$ is globally stable.

We can now state the theorem of Hartman and Grobman. In the statement of the theorem, S is an open subset of $\mathbb{R}$ and $h\colon S \to S$. In this setting, h is called a C^1-*diffeomorphism* if both h and its inverse h^{-1} are continuously differentiable on S. A fixed point x^* of h in S is called *hyperbolic* if $|h'(x^*)| \neq 1$.

Theorem 4.1.6 (Hartman–Grobman) *Let h be a diffeomorphism, let $x^* \in S$ be a fixed point of h in S, and let h_1 be the Taylor approximation in (4.5). If x^* is hyperbolic, then there exists an open set G containing x^* such that h and h_1 are topologically conjugate on G.*[6]

Be careful when applying this theorem, which is one of the most misused mathematical results in all of economics. It provides only a *neighborhood* of S such that behavior of the approximation is *qualitatively* similar to that of the original system. As it stands, the Hartman–Grobman theorem provides no basis for *quantitative* analysis.[7]

[6] To see why $|h'(x^*)| \neq 1$ is important, consider the case of $h(x) = \arctan(x)$.

[7] For a discussion of some of the problems associated with applying linearization to quantitative models with significant nonlinearities, see, for example, Boneva et al. (2016) or Pohl et al. (2018).

4.2 Finite State Markov Chains

Next we start our journey into the world of stochastic dynamics. We begin with finite state Markov chains, which were mentioned briefly in chapter 1. Finite state Markov chains are employed routinely in almost every field of science and form a core part of quantitative modeling in economics, finance and operations research. Our treatment of finite state stochastic dynamics is also geared toward building intuition, notation and tools that will be used in the general state case.

4.2.1 Definition

Let $S = \{x_1, \ldots, x_N\}$. A typical element of S is usually denoted by x, rather than a symbol such as x_i or x_n, in order to make our notation more consistent with the continuous state theory developed below. The set S will be called the *state space*. The set of *distributions* on S will be denoted $\mathscr{P}(S)$, and consists of all functions $\phi\colon S \to \mathbb{R}$ with $\phi(x) \geq 0$ for all $x \in S$, and $\sum_{x\in S} \phi(x) = 1$. In general, $\phi(x)$ will correspond to the probability attached to the point x in the state space under some given scenario.[8]

A quick digression: Although ϕ has been introduced as a function from S to $\mathbb{R}$, one can also think of it as a *vector* under the one-to-one correspondence

$$\mathscr{P}(S) \ni \phi \leftrightarrow (\phi(x))_{x\in S} := (\phi(x_1), \ldots, \phi(x_N)) \in \mathbb{R}^N \tag{4.6}$$

Under the correspondence (4.6), the collection of functions $\mathscr{P}(S)$ becomes a subset of the vector space $\mathbb{R}^N$—in particular, the elements of $\mathbb{R}^N$ that are nonnegative and sum to one. This set is called the unit simplex, and is illustrated for the case of $N = 3$ in figure 4.8.

The basic primitive for a discrete time Markov process on S is a *stochastic kernel*, the definition of which is as follows.

Definition 4.2.1 A *stochastic kernel* p is a function from $S \times S$ into $[0, 1]$ such that

1. $p(x, y) \geq 0$ for each (x, y) in $S \times S$, and
2. $\sum_{y\in S} p(x, y) = 1$ for each $x \in S$.

In other words, the function $S \ni y \mapsto p(x, y) \in \mathbb{R}$ is an element of $\mathscr{P}(S)$ for all $x \in S$. This distribution is represented by the symbol $p(x, dy)$ in what follows.

As well as being a function, the distribution $p(x, dy)$ can be viewed as a row[9] vector $(p(x, x_1), \ldots, p(x, x_N))$ in $\mathbb{R}^N$, located in the unit simplex, and these rows can

[8] What we call a distribution here is also referred to as a probability mass function.

[9] When treating distributions as vectors it is traditional in the Markov chain literature to regard them as row vectors.

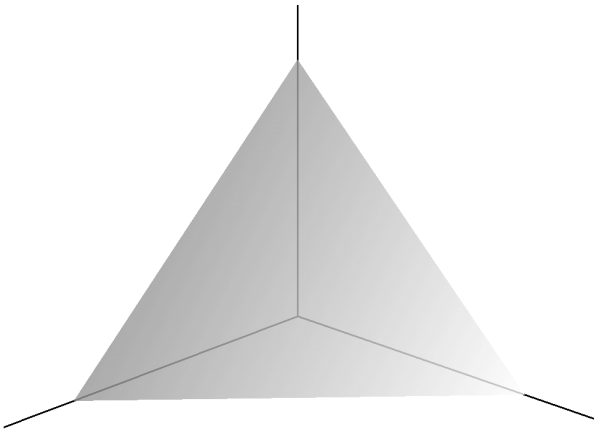

Figure 4.8 The unit simplex with $N = 3$

be stacked horizontally to produce an $N \times N$ matrix with the property that each row is nonnegative and sums to one:

$$p = \begin{pmatrix} p(x_1, dy) \\ \vdots \\ p(x_N, dy) \end{pmatrix} = \begin{pmatrix} p(x_1, x_1) & \cdots & p(x_1, x_N) \\ \vdots & & \vdots \\ p(x_N, x_1) & \cdots & p(x_N, x_N) \end{pmatrix} \tag{4.7}$$

Conversely, any square $N \times N$ matrix that is nonnegative and has all rows summing to one defines a stochastic kernel. However, when we move on to infinite state spaces there is no concept of matrices, and hence most of the theory is stated in terms of kernels.

In this chapter we are going to study a sequence of random variables $(X_t)_{t \geq 0}$, where each X_t takes values in S. The sequence updates according to the following rule: If $X_t = x$, then, in the following period X_{t+1} takes the value y with probability $p(x, y)$. In other words, once the current state X_t is realized, the probabilities for X_{t+1} are given by $p(X_t, dy)$. Figure 4.9 depicts an example of a simple Markov system, where $S = \{x_1, x_2, x_3\}$, and $p(x_i, x_j)$ is the probability that X_t moves from state x_i at time t to x_j at time $t + 1$.

The transition probabilities at each time depend on nothing other than the *current* location of the state. This is the "Markov" assumption. Moreover the transition probabilities do not depend on time. This is called time homogeneity. While these assumptions might seem strict, it turns out that, with some manipulation, a large class of systems can be embedded in the basic Markov framework. Typically this is achieved

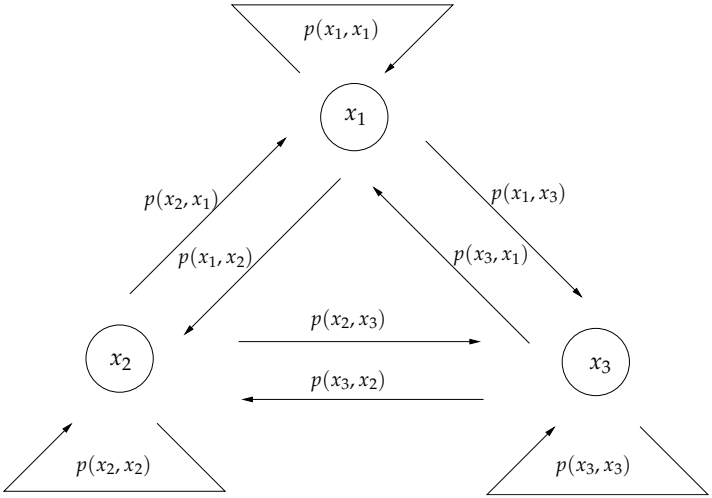

Figure 4.9 Finite Markov chain

by enlarging the state space until it contains all the information required to update the state.

A simple example of a stochastic kernel is the one used in Hamilton (2005), who investigates a nonlinear statistical model of the business cycle based on US unemployment data. As part of his calculation he estimates the kernel

$$p_H := \begin{pmatrix} 0.971 & 0.029 & 0 \\ 0.145 & 0.778 & 0.077 \\ 0 & 0.508 & 0.492 \end{pmatrix} \tag{4.8}$$

Here $S = \{x_1, x_2, x_3\} = \{NG, MR, SR\}$, where NG corresponds to normal growth, MR to mild recession, and SR to severe recession. For example, the probability of transitioning from severe recession to mild recession in one period is 0.508. The length of each period is one month.

For another example of a Markov model, consider the growth dynamics study of Quah (1993), who analyzes the evolution of real GDP per capita relative to the world average for a "typical" country (e.g., $X_t = 2$ implies that income per capita for the country in question is twice the world average at time t). A natural state space is $\mathbb{R}_+$, but to simplify matters Quah discretizes this space into five bins that correspond to values for relative GDP of 0 to 0.25, 0.25 to 0.5, 0.5 to 1, 1 to 2 and 2 to ∞ respectively. He then calculates the stochastic kernel by setting $p(x, y)$ equal to the fraction of times

that a country, finding itself in state x, subsequently makes the transition to state y.[10] The result of this calculation is

$$p_Q := \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix} \tag{4.9}$$

For example, the probability of our typical country transitioning from the lowest bin to the second lowest bin in one year is 0.03.

Algorithm 4.1: Simulation of a Markov chain

```
draw X_0 ~ ψ and set t = 0
while True do                          // "while True" means repeat forever
    draw X_{t+1} ~ p(X_t, dy)
    set t = t + 1
end
```

Let us now clarify the definition of a Markov chain $(X_t)_{t\geq 0}$ corresponding to a given stochastic kernel p. It is helpful to imagine that we wish to simulate $(X_t)_{t\geq 0}$ on a computer. First we draw X_0 from some predetermined *initial condition* $\psi \in \mathscr{P}(S)$. As $p(x, dy)$ gives the transition probabilities for X_{t+1} conditional on $X_t = x$, we now draw X_1 from $p(X_0, dy)$. Taking the result X_1, we then draw X_2 from $p(X_1, dy)$, and so on. This is the content of algorithm 4.1, as well as the next definition.

Definition 4.2.2 Let $\psi \in \mathscr{P}(S)$. A sequence of S-valued random variables $(X_t)_{t\geq 0}$ is called *Markov-*(p, ψ) if

1. at time $t = 0$, the initial state X_0 is drawn from ψ, and
2. at each $t \geq 1$, X_t is drawn from $p(X_{t-1}, dy)$.

If $\psi = \delta_x$ for some $x \in S$, then $(X_t)_{t\geq 0}$ is called *Markov-*(p, x).

4.2.2 From MCs to SRSs

Let's think carefully about the mechanics of simulating Markov chains. How exactly should we implement algorithm 4.1 on a computer? Considering this problem leads

[10] His data span the period 1962 to 1984, and have a sample of 118 countries. The transitions are over a one year period. The model is assumed to be stationary (transition probabilities do not vary with time), so all of the transitions (1962 to 1963, 1963 to 1964, etc.) can be pooled when calculating transition probabilities.

us to investigate the connection between Markov chains generated by stochastic kernels on one hand and stochastic recursive sequences (SRSs, also called stochastic difference equations) on the other. Stochastic recursive sequences lie at the heart of many economic models.

A typical SRS has the form

$$X_{t+1} = F(X_t, W_{t+1}), \quad X_0 \sim \psi \in \mathscr{P}(S), \quad F\colon S \times Z \to S \tag{4.10}$$

where $(W_t)_{t\geq 1}$ is a sequence of IID shocks taking values in arbitrary set Z. The shocks W_t are to be thought of as functions on a common set Ω, called the *probability space*. The system now evolves as follows:

1. At the start of time, nature selects an $\omega \in \Omega$ according to some probability $\mathbb{P}$.
2. The draw ω determines a complete realization of the path $(W_t(\omega))_{t\geq 1}$.
3. The draw ω also determines X_0, with $\mathbb{P}\{\omega : X_0(\omega) = x\} = \psi(x)$.
4. Given $(W_t(\omega))_{t\geq 1}$ and $X_0(\omega)$, we construct the time path $(X_t(\omega))_{t\geq 0}$ via

$$X_1(\omega) = F(X_0(\omega), W_1(\omega)), \quad X_2(\omega) = F(X_1(\omega), W_2(\omega)), \quad \text{etc.}$$

The idea that all uncertainty is realized at the start of time by a single observation ω from Ω is a convenient mathematical fiction. It does, however, have an close analogy with what happens on a machine when we run a simulation. In particular, a sequence of "random" numbers produced by a computer is in fact only *pseudorandom*, meaning that the sequence is produced deterministically, according to a particular rule and initialized by a particular *seed*, while at the same time mimicking the properties of independent draws.

From this perspective, you can think of ω as the seed that is fixed at the beginning of our simulation, which then determines the whole path via steps 1–4 above.

The SRS (4.10) induces a stochastic kernel p on S by

$$p(x, y) = \mathbb{P}\{F(x, W_t) = y\} := \mathbb{P}\{\omega \in \Omega : F(x, W_t(\omega)) = y\}$$

We now show it is possible to go the other way, representing *any* Markov-(p, ψ) process by an SRS such as (4.10). One we have the SRS representation, we will have another way to view Markov chains, which is helpful for concepts and theory, as well as a natural way to simulate paths from a given kernel.

To this end, let p be a stochastic kernel on S and fix $\psi \in \mathscr{P}(S)$. To generate a Markov-(p, ψ) chain, we take $(W_t)_{t\geq 0}$ to be IID uniform on $(0, 1]$ and let

$$X_0 = \tau(W_0; \psi), \quad X_{t+1} = \tau(W_{t+1}; p(X_t, dy)) \tag{4.11}$$

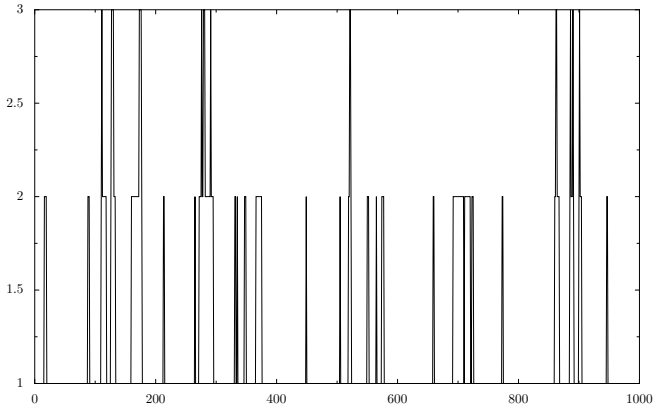

Figure 4.10 Simulation of the Hamilton Markov chain

where τ is the function discussed at length in §2.2.1. The second equality can be rewritten as

$$X_{t+1} = F(X_t, W_{t+1}) \quad \text{where} \quad F(x,z) := \tau(z; p(x,dy)) \tag{4.12}$$

The discussion of the inverse transform method in §2.2.1 tells is that, since W is uniform on $(0,1]$, the random variable $F(x,W)$ has distribution $p(x,dy)$. As a result, the sequence $(X_t)_{t\geq 0}$ generated by (4.11) and (4.12) obeys $X_0 \sim \psi$ and $X_{t+1} \sim p(X_t, dy)$ for $t \geq 0$. In other words, $(X_t)_{t\geq 0}$ is a Markov-(p,ψ) chain.

Exercise 4.22 Using an implementation of the function τ from §2.2.1, or your own version of the inverse transform method in your preferred language, combined with the SRS formulation in (4.11), simulate and plot a time series from Hamilton's Markov chain. You can identify the state space $S = \{NG, MR, SR\}$ with the integers $\{0,1,2\}$.

The Jupyter code book contains multiple solutions to exercise 4.22. One is coded to replicate the mathematical description as closely as possible. Another uses existing (and highly efficient) code from the QuantEcon library. Figure 4.10 shows one time series generated in this exercise.

Incidentally, SRSs are sometimes referred to as *iterated function systems* (IFSs). In this framework one thinks of updating the state from X_t to X_{t+1} by the *random function* $F_{W_{t+1}} := F(\cdot, W_{t+1})$. In practice the only change is a notational one: $X_{t+1} = F_{W_{t+1}}(X_t)$ as compared to (4.10). The main advantage is that we can now write

$$X_t = F_{W_t} \circ F_{W_{t-1}} \circ \cdots \circ F_{W_1}(X_0) = F_{W_t} \circ F_{W_{t-1}} \circ \cdots \circ F_{W_1}(\tau(W_0; \psi))$$

We see more clearly that X_t is just a fixed function of the initial condition and shocks up to time t.

4.2.3 Marginal Distributions

Let $(X_t)_{t\geq 0}$ be Markov-(p, ψ). For every $t \in \mathbb{N}$, let $\psi_t \in \mathscr{P}(S)$ denote the distribution of X_t. That is, $\psi_t(y)$ is the probability that $X_t = y$, given that X_0 is drawn from initial distribution ψ, and that the chain subsequently follows $X_{t+1} \sim p(X_t, dy)$. This distribution is sometimes called the *marginal* or *unconditional* distribution of X_t. We can understand it as follows: Generate n independent realizations of X_t, and calculate the fraction that takes the value y. Call this number $\psi_t^n(y)$. The probability $\psi_t(y)$ can be thought of as the limit of $\psi_t^n(y)$ as $n \to \infty$.

A method for computing the fraction $\psi_t^n(y)$ is given in algorithm 4.2. In the algorithm, the instruction draw $X \sim p(X, dy)$ should be interpreted as: Draw a random variable Y according to the distribution $p(X, dy)$ and then set $X = Y$. Also, $\mathbb{1}\{X_t^i = y\}$ is an indicator function, equal to one when $X_t^i = y$ and zero otherwise.

Algorithm 4.2: Approximate marginal distribution

```
for i in 1 to n do
    draw X ~ ψ
    for j in 1 to t do
        draw X ~ p(X, dy)
    end
    set X_t^i = X
end
return (1/n) Σ_{i=1}^n 𝟙{X_t^i = y}
```

Exercise 4.23 Implement algorithm 4.2 for Hamilton's Markov chain. You can identify the state space $S = \{NG, MR, SR\}$ with the integers $\{0, 1, 2\}$. Set $\psi = (0, 0, 1)$, so the economy starts in severe recession with probability one. Compute an approximation to $\psi_t(y)$, where $t = 10$ and $y = 0$. For sufficiently large n you should get an answer close to 0.6.

Exercise 4.24 Rewrite algorithm 4.2 using a counter that increments by one whenever the output of the inner loop produces a value equal to y instead of recording the value of each X_t^i.

Now consider again a Markov-(p, ψ) chain $(X_t)_{t\geq 0}$ for arbitrary stochastic kernel p and initial condition ψ. As above, let $\psi_t \in \mathscr{P}(S)$ be the marginal distribution of X_t.

From ψ_t and our complete description of the dynamics in p, it seems possible that we will be able to calculate the distribution of X_{t+1}. That is to say, we might be able to link ψ_t and ψ_{t+1} using p. That we can in fact construct such a recursion is one of the most fundamental and important properties of Markov chains.

To begin, pick any $y \in S$. Using the law of total probability (see §A.1.3), we can decompose the probability that $X_{t+1} = y$ into conditional parts as follows:

$$\mathbb{P}\{X_{t+1} = y\} = \sum_{x \in S} \mathbb{P}\{X_{t+1} = y \mid X_t = x\} \cdot \mathbb{P}\{X_t = x\}$$

Rewriting this statement in terms of our marginal and conditional probabilities gives

$$\psi_{t+1}(y) = \sum_{x \in S} p(x, y)\psi_t(x) \qquad (y \in S) \tag{4.13}$$

This is precisely the kind of recursion we are looking for. Let's introduce some additional notation to help manipulate this expression.

Definition 4.2.3 Given stochastic kernel p, the *Markov operator* corresponding to p is the map $\mathbf{M}$ sending $\mathscr{P}(S) \ni \psi \mapsto \psi\mathbf{M} \in \mathscr{P}(S)$, where $\psi\mathbf{M}$ is defined by

$$\psi\mathbf{M}(y) = \sum_{x \in S} p(x, y)\psi(x) \qquad (y \in S) \tag{4.14}$$

The notation appears unusual, in the sense that we normally write $\mathbf{M}(\psi)$ instead of $\psi\mathbf{M}$ for the image of ψ under a mapping $\mathbf{M}$. However, such notation is traditional in the Markov literature. It reminds us that applying the Markov operator to a distribution $\psi \in \mathscr{P}(S)$ is just postmultiplication of the row vector $(\psi(x))_{x \in S}$ by the stochastic kernel (viewed as a matrix).

Combining (4.13) and (4.14), we obtain the fundamental recursion

$$\psi_{t+1} = \psi_t\mathbf{M} \tag{4.15}$$

Check this carefully until you feel comfortable with the notation.

This representation (4.15) is easy to manipulate. For example, suppose that we want to calculate ψ_{j+k} from ψ_j. Clearly,

$$\psi_{j+k} = \psi_{j+k-1}\mathbf{M} = (\psi_{j+k-2}\mathbf{M})\mathbf{M} = \psi_{j+k-2}\mathbf{M}^2 = \cdots = \psi_j\mathbf{M}^k$$

where $\mathbf{M}^m$ is the m-th composition of the map $\mathbf{M}$ with itself. In particular, setting $j = 0$ and $k = t$, we have $X_t \sim \psi\mathbf{M}^t$ when $X_0 \sim \psi$. Let's state these results as a theorem:

Theorem 4.2.4 *Let $(X_t)_{t \geq 0}$ be Markov-(p, ψ), and let $\mathbf{M}$ be the Markov operator corresponding to p. If ψ_t is the marginal distribution of X_t for each t, then $\psi_{t+1} = \psi_t\mathbf{M}$ and $\psi_t = \psi\mathbf{M}^t$.*

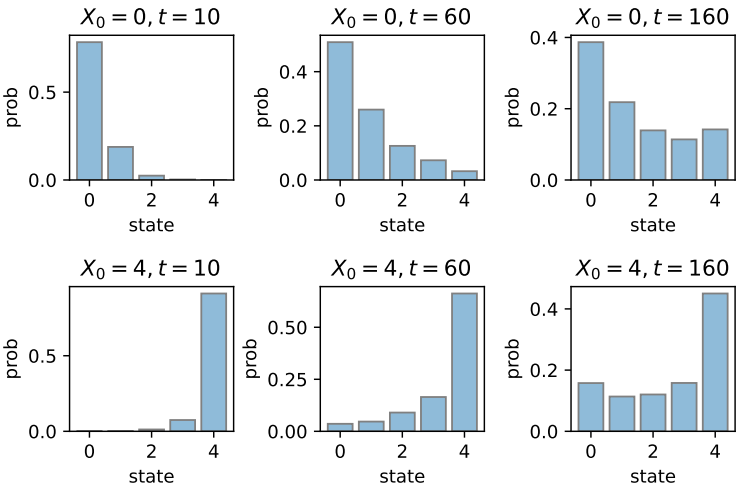

Figure 4.11 Top: $X_0 = 0$. Bottom: $X_0 = 4$

To illustrate these ideas, consider again the kernel p_Q calculated by Danny Quah, and let $\mathbf{M}_Q$ be the Markov operator. The states are enumerated as $S = \{0, 1, 2, 3, 4\}$. We can evaluate probabilities of different outcomes for a given country over time by iteratively applying $\mathbf{M}_Q$ to an initial condition ψ, generating the sequence $(\psi \mathbf{M}_Q^t)$. Figure 4.11 shows the elements $\psi \mathbf{M}_Q^{10}$, $\psi \mathbf{M}_Q^{60}$, and $\psi \mathbf{M}_Q^{160}$ of this sequence. In the top graph, the country in question is initially in the poorest group, so $\psi = (1, 0, 0, 0, 0)$. The bottom graph shows the corresponding elements when the initial condition is reset to $\psi = (0, 0, 0, 1, 0)$.

4.2.4 Other Identities

Let's think a bit more about the iterates of the Markov operator $\mathbf{M}$. To begin, fix a kernel p with Markov operator $\mathbf{M}$ and define the t-th order kernel p^t by

$$p^1 := p, \quad p^t(x, y) := \sum_{z \in S} p^{t-1}(x, z) p(z, y) \qquad ((x, y) \in S \times S, \ t \in \mathbb{N})$$

Exercise 4.25 Show that p^t is in fact a stochastic kernel on S for each $t \in \mathbb{N}$.

Exercise 4.26 Let $t \in \mathbb{N}$. Show that if p is interpreted as the matrix in (4.7), then $p^t(x, y)$ is the (x, y)-th element of its t-th power.

To interpret p^t, we can use the following lemma:

Lemma 4.2.5 *If* $\mathbf{M}$ *is the Markov operator defined by stochastic kernel* p *on* S*, then its* t*-th iterate* $\mathbf{M}^t$ *is the Markov operator defined by* p^t*, the* t*-th order kernel of* p*. In other words, for any* $\psi \in \mathscr{P}(S)$ *we have*

$$\psi \mathbf{M}^t(y) = \sum_{x \in S} p^t(x, y)\psi(x) \qquad (y \in S)$$

We prove only the case $t = 2$ here, and leave the full proof for the reader. (Hint: Use induction.) Pick any $\psi \in \mathscr{P}(S)$ and any y in S. Then

$$\begin{aligned} \psi \mathbf{M}^2(y) = ((\psi \mathbf{M})\mathbf{M})(y) &= \sum_{z \in S} p(z, y)\psi \mathbf{M}(z) \\ &= \sum_{z \in S} p(z, y) \sum_{x \in S} p(x, z)\psi(x) \\ &= \sum_{x \in S} \sum_{z \in S} p(x, z)p(z, y)\psi(x) = \sum_{x \in S} p^2(x, y)\psi(x) \end{aligned}$$

Now let $\delta_x \in \mathscr{P}(S)$ be the distribution that puts all mass on $x \in S$ (i.e., $\delta_x(y) = 1$ if $y = x$ and zero otherwise). Applying lemma 4.2.5 with $\psi = \delta_x$, we obtain $\delta_x \mathbf{M}^t(y) = p^t(x, y)$ for all $y \in S$. In other words, the distribution $p^t(x, dy)$ is precisely $\delta_x \mathbf{M}^t$, which we know is the distribution of X_t when $X_0 = x$. More generally, $p^k(x, y)$ is the probability that the state moves from x now to y in k steps:

$$p^k(x, y) = \mathbb{P}\{X_{t+k} = y \mid X_t = x\} \qquad (x, y \in S,\ k \in \mathbb{N})$$

and $p^k(x, dy)$ is the conditional distribution of X_{t+k} given $X_t = x$.

Exercise 4.27 Confirm the *Chapman–Kolmogorov equation*, which states that for any $j, k \in \mathbb{N}$,

$$p^{j+k}(x, y) = \sum_{z \in S} p^j(x, z)p^k(z, y) \qquad ((x, y) \in S \times S)$$

Now let's introduce another operation for the Markov operator $\mathbf{M}$. So far we have $\mathbf{M}$ acting on distributions to the left, as in $\psi \mathbf{M}(y) = \sum_x p(x, y)\psi(x)$. We also let $\mathbf{M}$ act on functions to the right, as in

$$\mathbf{M}h(x) = \sum_{y \in S} p(x, y)h(y) \qquad (x \in S) \tag{4.16}$$

where $h \colon S \to \mathbb{R}$ is any function. Thus $\mathbf{M}$ takes a given function h on S and sends it into a new function $\mathbf{M}h$ on S. In terms of matrix algebra, this is pre-multiplication of the column vector $(h(y))_{y \in S}$ by the matrix (4.7).

To understand (4.16), recall that if Y is a random variable on S with distribution $\phi \in \mathscr{P}(S)$ (i.e., $\mathbb{P}\{Y = y\} = \phi(y)$ for all $y \in S$) and h is a real-valued function on S, then the expectation $\mathbb{E}h(Y)$ of $h(Y)$ is the sum of all values $h(y)$ weighted by the probabilities $\mathbb{P}\{Y = y\}$:

$$\mathbb{E}h(Y) := \sum_{y \in S} h(y)\mathbb{P}\{Y = y\} = \sum_{y \in S} \phi(y)h(y)$$

In terms of vectors we are just computing inner products.

It is now clear that $\mathbf{M}h(x) = \sum_{y \in S} p(x,y)h(y)$ should be interpreted as the expectation of $h(X_{t+1})$ given $X_t = x$. Analogous to the result in lemma 4.2.5, we have

$$\mathbf{M}^t h(x) = \sum_{y \in S} p^t(x,y)h(y) \qquad (x \in S,\ t \in \mathbb{N}) \tag{4.17}$$

Since $p^t(x, dy)$ is the distribution of X_t given $X_0 = x$, it follows that $\mathbf{M}^t h(x)$ is the expectation of $h(X_t)$ given $X_0 = x$.

Exercise 4.28 Confirm the claim in (4.17).

Now the finishing touches. Fix an initial condition $\psi \in \mathscr{P}(S)$, a function h as above and a $k \in \mathbb{N}$. Define

$$\psi\mathbf{M}^k h := \sum_{y \in S} \sum_{x \in S} p^k(x,y)\psi(x)h(y) \tag{4.18}$$

In terms of linear algebra, this expression can be thought of as the inner product of $\psi\mathbf{M}^k$ and h. Since $\psi\mathbf{M}^k$ is the distribution of X_{t+k} when $X_t \sim \psi$, (4.18) gives us the expectation of $h(X_{t+k})$ given $X_t \sim \psi$. In symbols,

$$\psi\mathbf{M}^k h = \mathbb{E}[h(X_{t+k}) \mid X_t \sim \psi] \tag{4.19}$$

Exercise 4.29 Suppose that the business cycle evolves according to Hamilton's kernel p_H on $S = \{NG, MR, SR\}$, and that a firm makes profits $\{1000, 0, -1000\}$ in these three states. Compute expected profits at $t = 5$, given that the economy starts in NG. How much do profits change when the economy starts in SR?

Exercise 4.30 Compute expected profits at $t = 1000$ for each of the three possible initial states. What do you notice?

Exercise 4.31 Suppose now that the initial state will be drawn according to $\psi = (0.2, 0.2, 0.6)$. Compute expected profits at $t = 5$ using (4.19).

4.2.5 Constructing Joint Distributions

Let's now consider the joint distributions of a Markov-(p, ψ) process $(X_t)_{t\geq 0}$. We would like to understand more about probabilities not just for individual elements of the sequence such as X_t, but rather for a collection of elements. For example, how do we compute the probability that $(X_t, X_{t+1}) = (x, y)$, or that $X_j \leq x$ for $j \leq t$?

Consider first the pair (X_0, X_1), which can be thought of as a single bivariate random variable taking values in $S^2 := S \times S$. Thus the joint distribution is an element of $\mathscr{P}(S^2)$. A typical element of S^2 is a pair (x^0, x^1), where $x^i \in S$.[11] We wish to find the probability $\mathbb{P}\{X_0 = x^0, X_1 = x^1\}$.

To begin, pick any $(x^0, x^1) \in S^2$, and let

$$q_2(x^0, x^1) := \mathbb{P}\{X_0 = x^0, X_1 = x^1\} = \mathbb{P}\{X_0 = x^0\} \cap \{X_1 = x^1\}$$

From (A.2) on page 327, for any events A and B we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B \mid A)$. It follows that

$$q_2(x^0, x^1) = \mathbb{P}\{X_0 = x^0\}\mathbb{P}\{X_1 = x^1 \mid X_0 = x^0\} = \psi(x^0)p(x^0, x^1)$$

Similarly, the distribution $q_3 \in \mathscr{P}(S^3)$ of (X_0, X_1, X_2) is

$$\begin{aligned} q_3(x^0, x^1, x^2) &= \mathbb{P}\{X_0 = x^0, X_1 = x^1, X_2 = x^2\} \\ &= \mathbb{P}\{X_0 = x^0, X_1 = x^1\}\mathbb{P}\{X_2 = x^2 \mid X_0 = x^0, X_1 = x^1\} \\ &= \psi(x^0)p(x^0, x^1)p(x^1, x^2) \end{aligned}$$

Notice that we are using $\mathbb{P}\{X_2 = x^2 \mid X_0 = x^0, X_1 = x^1\} = p(x^1, x^2)$. This is reasonable because, if $X_1 = x^1$, then $X_2 \sim p(x^1, dy)$.

Continuing along the same lines yields the general expression

$$q_{T+1}(x^0, \ldots, x^T) = \psi(x^0) \prod_{t=0}^{T-1} p(x^t, x^{t+1}) \tag{4.20}$$

To evaluate (4.20) we can use the function given in algorithm 4.3.

Exercise 4.32 Show that for Hamilton's kernel p_H and $\psi = (0.2, 0.2, 0.6)$, the probability of path (NG, MR, NG) is 0.000841.

A solution to this and other computational exercises below can be found in the Jupyter code book.

[11] A word on notation: Superscripts represent time, so $x^0 \in S$ is a typical realization of X_0, $x^1 \in S$ is a typical realization of X_1, and so on.

Algorithm 4.3: A function to compute the probability of path $(x^0, x^1, \ldots, x^T)$

Data: a stochastic kernel p and initial distribution ψ on S
Function $q(x^0, x^1, \ldots, x^T)$
  set $s = \psi(x^0)$
  for t in $0, \ldots, T-1$ **do**
    set $s = s \cdot p(x^t, x^{t+1})$
  end
  return s

From our expression for q_{T+1} in (4.20) we can also compute the probabilities of more complex events. By an event is meant any subset B of S^{T+1}. For example,

$$B := \{(x^0, \ldots, x^T) \in S^{T+1} : x^t \leq x^{t+1} \text{ for } t = 0, \ldots, T-1\}$$

is an event. It consists of all paths $(x^0, \ldots, x^T)$ in S^{T+1} that are increasing (i.e., nondecreasing). To obtain the probability of any such event B we just sum $q_{T+1}(x^0, \ldots, x^T)$ over all distinct paths in B.

One important special case is events of the form

$$D^0 \times \cdots \times D^T = \{(x^0, \ldots, x^T) \in S^{T+1} : x^t \in D^t \text{ for } t = 0, \ldots, T\}$$

where $D^t \subset S$ for each t. Then $\mathbb{P}\{(X_0, \ldots, X_T) \in D^0 \times \cdots \times D^T\} = \mathbb{P} \cap_{t \leq T} \{X_t \in D^t\}$, and for this kind of event the following lemma applies:

Lemma 4.2.6 *If $D^0, \ldots, D^T$ is any collection of subsets of S, then*

$$\mathbb{P} \cap_{t \leq T} \{X_t \in D^t\} = \sum_{x^0 \in D^0} \psi(x^0) \sum_{x^1 \in D^1} p(x^0, x^1) \cdots \sum_{x^T \in D^T} p(x^{T-1}, x^T)$$

Proof. For any such sets D^t, the probability $\mathbb{P} \cap_{t \leq T} \{X_t \in D^t\}$ can be computed by summing over distinct paths:

$$\begin{aligned} \mathbb{P} \cap_{t \leq T} \{X_t \in D^t\} &= \sum_{(x^0, \ldots, x^T) \in D^0 \times \cdots \times D^T} q_{T+1}(x^0, \ldots, x^T) \\ &= \sum_{x^0 \in D^0} \cdots \sum_{x^T \in D^T} q_{T+1}(x^0, \ldots, x^T) \end{aligned}$$

The last step now follows from the expression for q_{T+1} in (4.20). □

Exercise 4.33 Returning to Hamilton's kernel p_H, and using the same initial condition $\psi = (0.2, 0.2, 0.6)$ as in exercise 4.32, compute the probability that the economy starts and remains in recession through periods 0, 1, 2.

Another way to compute this probability is via Monte Carlo:

Exercise 4.34 Generate 10,000 observations of (X_0, X_1, X_2), starting at the same initial condition $\psi = (0.2, 0.2, 0.6)$. Count the number of paths that do not enter state NG and divide by 10,000 to get the fraction of paths that remain in recession. This fraction converges to the probability of the event, so you should get approximately the same number as you found in exercise 4.33.

Now let's think a little bit about computing expectations. Recall the firm in exercise 4.29. If the firm operates up until period T, and if the interest rate is equal to r, then the net present value (NPV) of the firm is the expected sum of discounted profits

$$\mathbb{E}\,\Pi(X_0, \ldots, X_T) \quad \text{where} \quad \Pi(X_0, \ldots, X_T) := \sum_{t=0}^{T} \rho^t h(X_t)$$

and $\rho := 1/(1+r)$. Expectations for finite state spaces are found by summing values weighted by probabilities. In this case,

$$\mathbb{E}\,\Pi(X_0, \ldots, X_T) = \sum \Pi(x^0, \ldots, x^T) q_{T+1}(x^0, \ldots, x^T) =: \sum \Pi(\mathbf{x}) q_{T+1}(\mathbf{x})$$

where the sum is over all $\mathbf{x} \in S^{T+1}$.

For large T and S this kind of computation is problematic. For example, if S has ten elements and $T = 100$, then we must sum $\Pi(\mathbf{x})q_{T+1}(\mathbf{x})$ over 10^{100} paths.

Exercise 4.35 If the computer can evaluate one billion (10^9) paths per second, how may years will it take to evaluate all of the paths? Compare this with current estimates of the age of the universe.

Fortunately, the computational problem can be greatly simplified in this particular case by linearity of expectations, which gives

$$\mathbb{E}\Pi = \mathbb{E}\left[\sum_{t=0}^{T} \rho^t h(X_t)\right] = \sum_{t=0}^{T} \rho^t \mathbb{E}h(X_t) = \sum_{t=0}^{T} \rho^t \psi \mathbf{M}^t h$$

The second equality (linearity of $\mathbb{E}$) can be proved from the definition of the joint distribution, but we treat it in much greater generality below. The third equality follows from (4.19) on page 83.

Exercise 4.36 Compute NPV when $r = 0.05$. Take the same initial condition as in exercise 4.32. Plot expected profits against T. For what values of T will the firm's expected profits be positive?

4.3 Stability of Finite State MCs

In chapter 1 we investigated a Markovian model where the distribution for log income converges to a unique distribution $N(\mu^*, v^*)$, independent of initial conditions. This

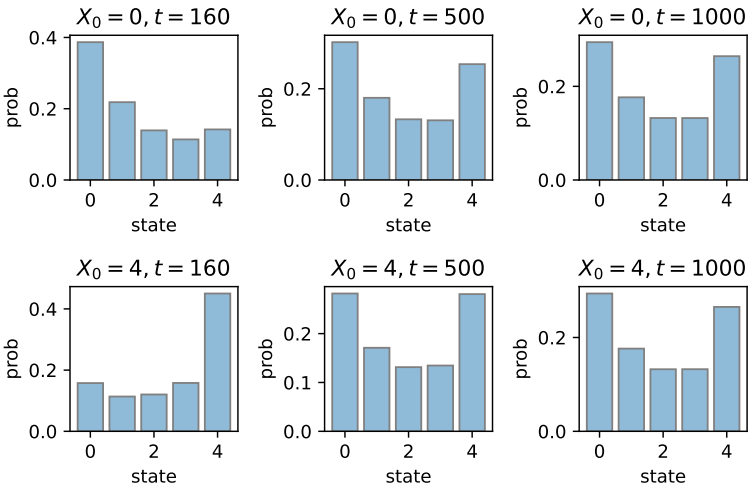

Figure 4.12 Top: $X_0 = 1$. Bottom: $X_0 = 4$

behavior means that knowledge of the limiting distribution gives us a great deal of predictive power in terms of likelihoods for long-run outcomes. In fact stability also gives us a number of statistical properties that are central to time series econometrics. As a result, we are motivated to study when one does observe stability, beginning with the case of finite state Markov chains.

To start the ball rolling, consider again the sequences of distributions in figure 4.11 (page 81). What happens if we extend the time horizon? In other words, what sort of limiting properties, if any, do these sequences possess? Figure 4.12 repeats the same distribution projections, but this time for dates $t = 160$, $t = 500$, and $t = 1,000$. Looking at the top graph for starters, note that after about $t = 500$ there seems to be very little change in ψ_t. In other words, it appears that the sequence (ψ_t) is converging. Interestingly, the sequence in the bottom graph seems to be converging to the same limit.

Perhaps we are again observing a form of global stability? It turns out that we are, but to show this we must first define stability for Markov chains and derive theorems that allow us to establish this property.

4.3.1 Stationary Distributions

Recall that a dynamical system (U, h) consists of a metric space U and a map $h\colon U \to U$. Recall also the definition of the Markov operator $\mathbf{M}$ corresponding to a given

stochastic kernel p: Given $\psi \in \mathscr{P}(S)$, the operator $\mathbf{M}$ is a map sending ψ into $\psi\mathbf{M}$, where $\psi\mathbf{M}(y) = \sum_{x \in S} p(x,y)\psi(x)$ for each $y \in S$. What we are going to do now is view $(\mathscr{P}(S), \mathbf{M})$ as a dynamical system in its own right (recalling that trajectories of the form $(\psi\mathbf{M}^t)_{t\geq 0}$ correspond to the sequence of marginal distributions for a Markov-(p,ψ) process $(X_t)_{t\geq 0}$; see page 80). To do this, we need to introduce a metric on $\mathscr{P}(S)$, and also establish that $\mathbf{M}$ does indeed send $\mathscr{P}(S)$ into itself.

Exercise 4.37 Confirm that $\psi\mathbf{M} \in \mathscr{P}(S)$ whenever $\psi \in \mathscr{P}(S)$.

To set $\mathscr{P}(S)$ up as a metric space, we define

$$\|\psi\|_1 := \sum_{x \in S} |\psi(x)| \text{ for each } \psi \in \mathscr{P}(S), \quad \text{and} \quad d_1(\psi, \psi') := \|\psi - \psi'\|_1$$

If one views $\mathscr{P}(S)$ as the unit simplex in $\mathbb{R}^N$ rather than as a space of functions (see the correspondence (4.6) on page 73), then our norm and distance are just the regular $\|\cdot\|_1$ norm (see page 41) and d_1 distance on $\mathbb{R}^N$. Viewed in this way, $\mathscr{P}(S)$ is a closed and bounded subset of $(\mathbb{R}^N, d_1)$, and therefore both compact and complete.[12]

The next exercise introduces another way to view the distance imposed by d_1.

Exercise 4.38 Let $\psi_1, \psi_2 \in \mathscr{P}(S)$, and for each $A \subset S$ let $\Psi_i(A) := \sum_{x \in A} \psi_i(x) =$ the probability of $A \subset S$ under distribution ψ_i. Let $s(\psi_1, \psi_2) = \sup_{A \subset S} |\Psi_1(A) - \Psi_2(A)|$. Show that

1. the supremum is achieved by $D = \{x \in S : \psi_1(x) \geq \psi_2(x)\}$ and
2. the norm $\|\cdot\|_1$ and s are connected by $\|\psi_1 - \psi_2\|_1 = 2s(\psi_1, \psi_2)$.

To illustrate the dynamical system $(\mathscr{P}(S), \mathbf{M})$ and its trajectories, consider Hamilton's kernel p_H and the corresponding operator $\mathbf{M}_H$. Here $\mathscr{P}(S)$ can be visualized as the unit simplex in $\mathbb{R}^3$. Figure 4.13 shows four trajectories $(\psi\mathbf{M}_H^t)$ generated by iterating $\mathbf{M}_H$ on four different initial conditions ψ. All trajectories converge toward the bottom right-hand corner. Indeed, we will prove below that $(\mathscr{P}(S), \mathbf{M}_H)$ is globally stable.

Exercise 4.39 Let $\mathbf{M}$ be the Markov operator determined by an arbitrary stochastic kernel p. Show that $\mathbf{M}$ is d_1-nonexpansive on $\mathscr{P}(S)$, in the sense that for any $\psi, \psi' \in \mathscr{P}(S)$ we have $d_1(\psi\mathbf{M}, \psi'\mathbf{M}) \leq d_1(\psi, \psi')$.

[12] Interested readers are invited to supply the details of the argument. The connection between the function space $(\mathscr{P}(S), d_1)$ and the unit simplex in $(\mathbb{R}^N, d_1)$ can be made precise using the concept of isomorphisms. Metric spaces (S, ρ) and (S', ρ') are said to be *isometrically isomorphic* if there exists a bijection $\tau\colon S \to S'$ such that $\rho(x,y) = \rho'(\tau(x), \tau(y))$ for all $x, y \in S$. In our case, the bijection in question is (4.6) on page 73. If (S, ρ) and (S', ρ') are isometrically isomorphic, then (S, ρ) is complete if and only if (S', ρ') is complete, and compact if and only if (S', ρ') is compact.

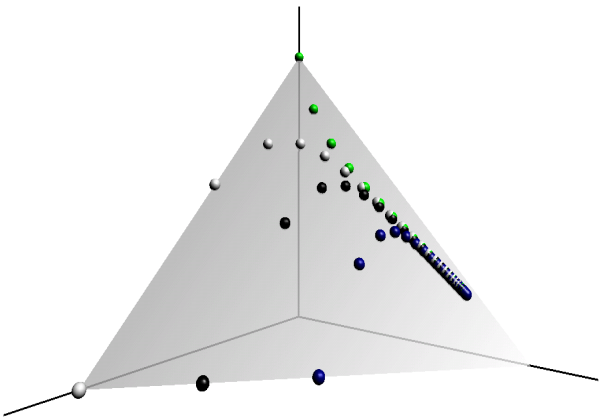

Figure 4.13 Trajectories of $(\mathscr{P}(S), \mathbf{M}_H)$

Let us now turn to the existence of fixed points for the system $(\mathscr{P}(S), \mathbf{M})$. For Markov chains, fixed points are referred to as stationary distributions:

Definition 4.3.1 A distribution $\psi^* \in \mathscr{P}(S)$ is called *stationary* or *invariant* for $\mathbf{M}$ if $\psi^*\mathbf{M} = \psi^*$. In other words, ψ^* is a stationary distribution for $\mathbf{M}$ if it is a fixed point of the dynamical system $(\mathscr{P}(S), \mathbf{M})$.

If ψ^* is stationary for $\mathbf{M}$, if $\mathbf{M}$ corresponds to kernel p, if $(X_t)_{t\geq 0}$ is Markov-(p, ψ), and if X_t has distribution ψ^* for some t, then X_{t+1} has distribution $\psi_{t+1} = \psi_t\mathbf{M} = \psi^*\mathbf{M} = \psi^*$. In fact iteration shows that X_{t+k} has distribution ψ^* for every $k \in \mathbb{N}$, so probabilities are stationary over time. Moreover if $(X_t)_{t\geq 0}$ is Markov-(p, ψ^*), then $X_t \sim \psi^*$ for all t, and the random variables $(X_t)_{t\geq 0}$ are identically distributed (but not IID—why?).

On the other hand, stationary distributions are just fixed points of a dynamical system $(\mathscr{P}(S), \mathbf{M})$. This is convenient for analysis because we already know various techniques for studying fixed points and stability properties of deterministic dynamical systems. For example, suppose that we view $\mathscr{P}(S)$ as the unit simplex in $\mathbb{R}^N$, and $\psi \mapsto \psi\mathbf{M}$ as postmultiplication of vector $\psi \in \mathbb{R}^N$ by the matrix corresponding to p. This mapping is d_1-nonexpansive (recall exercise 4.39), and hence d_1-continuous (exercise 3.44, page 57). The unit simplex is a compact, convex subset of $(\mathbb{R}^N, d_1)$. (Proof?) Applying Brouwer's theorem (theorem 3.2.15, page 56) we obtain our first major result for Markov chains:

Listing 4.1 (fphamilton.py) Computing stationary distributions

```
import numpy as np
from numpy.linalg import solve

pH = ((0.971, 0.029, 0.000),
      (0.145, 0.778, 0.077),
      (0.000, 0.508, 0.492))

I = np.identity(3)
Q, b = np.ones((3, 3)), np.ones((3, 1))
A = np.transpose(I - pH + Q)
print(solve(A, b))
```

Theorem 4.3.2 *Every Markov operator defined over a finite state space has at least one stationary distribution.*

There may, of course, be many stationary distributions, just as other dynamical systems can have many fixed points.

Exercise 4.40 For which kernel p is every $\psi \in \mathscr{P}(S)$ stationary?

Let's consider a technique for computing fixed points using matrix inversion. In terms of linear algebra, row vector $\psi \in \mathscr{P}(S)$ is stationary if and only if $\psi(\mathbf{I}_N - p) = 0$, where $\mathbf{I}_N$ is the $N \times N$ identity matrix, and p is the matrix in (4.7). One idea would be to try to invert $(\mathbf{I}_N - p)$. However, this does not impose the restriction that the solution ψ is an element of $\mathscr{P}(S)$. That restriction can be imposed in the following way.

Exercise 4.41 Let $\mathbb{1}_N$ be the $1 \times N$ row vector $(1, \ldots, 1)$. Let $\mathbb{1}_{N\times N}$ be the $N \times N$ matrix of ones. Show that if ψ is stationary, then

$$\mathbb{1}_N = \psi(\mathbf{I}_N - p + \mathbb{1}_{N\times N}) \tag{4.21}$$

Explain how this imposes the restriction that the elements of ψ sum to 1.

Taking the transpose of (4.21) we get $(\mathbf{I}_N - p + \mathbb{1}_{N\times N})^\top \psi^\top = \mathbb{1}_N^\top$. This is a linear system of the form $Ax = b$, which can be solved for $x = A^{-1}b$. (The solution is not necessarily unique. We return to the issue of uniqueness below.) Listing 4.1 shows how to do this in Python using NumPy.

Exercise 4.42 Use this technique to solve for the stationary distribution of Quah's

kernel p_Q.[13] Plot it as a bar plot, and compare with the $t = 1000$ distributions in figure 4.12.

Exercise 4.43 Recall the firm introduced on page 83. Compute expected profits at the stationary distribution. Compare it with profits at $t = 1000$, as computed in exercise 4.30, from a range of initial states. Interpret your results.

4.3.2 The Dobrushin Coefficient

Now let's consider convergence to the stationary distribution. We continue to impose on $\mathscr{P}(S)$ the distance d_1 and study the dynamical system $(\mathscr{P}(S), \mathbf{M})$. By definition 4.1.2, the system $(\mathscr{P}(S), \mathbf{M})$ is globally stable if

1. it has a unique fixed point (stationary distribution) $\psi^* \in \mathscr{P}(S)$, and
2. $d_1(\psi \mathbf{M}^t, \psi^*) := \|\psi \mathbf{M}^t - \psi^*\|_1 \to 0$ as $t \to \infty$ for all $\psi \in \mathscr{P}(S)$.

The second condition implies that if $(X_t)_{t \geq 0}$ is Markov-(p, ψ) for some $\psi \in \mathscr{P}(S)$, then the distribution of X_t converges to ψ^*.

Exercise 4.44 Exercise 4.40 asked you to provide an example of a kernel where global stability fails. Another is the "periodic" Markov chain

$$p = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Show that $\psi^* := (1/2, 1/2)$ is the unique stationary distribution. Give a counterexample to the claim $\|\psi \mathbf{M}^t - \psi^*\|_1 \to 0$ as $t \to \infty$, $\forall \psi \in \mathscr{P}(S)$.

How might one check for stability of a given kernel p and associated dynamical system $(\mathscr{P}(S), \mathbf{M})$? Exercise 4.39 suggests the way forward: $\mathbf{M}$ is nonexpansive on $\mathscr{P}(S)$, and if we can upgrade this to a uniform contraction then Banach's fixed point theorem (page 57) implies that $(\mathscr{P}(S), \mathbf{M})$ is globally stable, and that convergence to equilibrium takes place at a geometric rate.

Which kernels will we be able to upgrade? Intuitively, stable kernels are those where current states have little influence on future states. An extreme example is where the distributions $p(x, dy)$ are all equal: $p(x, dy) = q \in \mathscr{P}(S)$ for all $x \in S$. In this case the current state has no influence on tomorrow's state—indeed, the resulting process is IID with $X_t \sim q$ for all t. The Markov operator satisfies $\psi \mathbf{M} = q$ for all $\psi \in \mathscr{P}(S)$ (check it), and $(\mathscr{P}(S), \mathbf{M})$ is globally stable.

A less extreme case is when the distributions $p(x, dy)$ are "similar" across $x \in S$. One similarity measure for two distributions $p(x, dy)$ and $p(x', dy)$ is $\sum_y p(x, y) \wedge$

[13] We prove below that the fixed point is unique.

$p(x', y)$, where $a \wedge b := \min\{a, b\}$. If $p(x, dy) = p(x', dy)$ then the value is one. If the supports[14] of $p(x, dy)$ and $p(x', dy)$ are disjoint, then the value is zero. This leads us to the Dobrushin coefficient, which measures the stability properties of a given kernel p.

Definition 4.3.3 Given stochastic kernel p, the *Dobrushin coefficient* $\alpha(p)$ is defined by

$$\alpha(p) := \min \left\{ \sum_{y \in S} p(x, y) \wedge p(x', y) \ : \ (x, x') \in S \times S \right\} \tag{4.22}$$

Exercise 4.45 Prove that $0 \leq \alpha(p) \leq 1$ always holds.

Exercise 4.46 Show that $\alpha(p) = 1$ if and only if $p(x, dy)$ is equal to a constant distribution $q \in \mathscr{P}(S)$ for every $x \in S$.

Exercise 4.47 Show that $\alpha(p) = 0$ for the periodic kernel in exercise 4.44, and for p corresponding to the identity matrix.

Exercise 4.48 Distributions ϕ and ψ are said to *overlap* if there exists a y such that $\phi(y) > 0$ and $\psi(y) > 0$. Show that $\alpha(p) > 0$ if and only if for each pair $(x, x') \in S \times S$ the distributions $p(x, dy)$ and $p(x', dy)$ overlap.

The following result links the Dobrushin coefficient to stability via Banach's fixed point theorem (page 57).

Theorem 4.3.4 *If p is a stochastic kernel on S with Markov operator $\mathbf{M}$, then*

$$\|\phi \mathbf{M} - \psi \mathbf{M}\|_1 \leq (1 - \alpha(p)) \|\phi - \psi\|_1 \qquad \forall \, \phi, \psi \in \mathscr{P}(S)$$

Moreover this bound is the best available, in the sense that if $\lambda < 1 - \alpha(p)$, then there exists a pair ϕ, ψ in $\mathscr{P}(S)$ such that $\|\phi \mathbf{M} - \psi \mathbf{M}\|_1 > \lambda \|\phi - \psi\|_1$.

The first half of the theorem says that if $\alpha(p) > 0$, then $\mathbf{M}$ is uniformly contracting (for the definition see page 57) with modulus $1 - \alpha(p)$. Since $(\mathscr{P}(S), d_1)$ is complete, Banach's fixed point theorem then implies global stability of $(\mathscr{P}(S), \mathbf{M})$. The second part of the theorem says that this rate $1 - \alpha(p)$ *is the best available*, which in turn suggests that the Dobrushin coefficient is a good measure of the stability properties of $\mathbf{M}$. For example, if $\alpha(p) = 0$, then we can be certain $\mathbf{M}$ is not a uniform contraction.

Some intuition for theorem 4.3.4 and it's stability implications was discussed above. The coefficient is large (close to one) when all distributions $p(x, dy)$ are similar across

[14] The support of $\phi \in \mathscr{P}(S)$ is $\{y \in S : \phi(y) > 0\}$.

x, and the current state has little influence on future states. This is the stable case. The coefficient is zero when there exists states x and x' such that $p(x, dy)$ and $p(x', dy)$ have disjoint support, as with the identity kernel and the periodic kernel. More intuition on the link between positivity of $\alpha(p)$ and stability is given in the next section.

The proof of theorem 4.3.4 is given in the appendix to this chapter. The fact that $1 - \alpha(p)$ is the best rate possible may suggest to you that the proof is not entirely trivial. Indeed this is the case. We have to do better than crude inequalities. All but the most enthusiastic readers are encouraged to skip the proof and move to the next section.

4.3.3 Stability

Let p be a stochastic kernel on S. If $\alpha(p) > 0$, then $(\mathscr{P}(S), \mathbf{M})$ is globally stable by Banach's fixed point theorem. In fact we can say a bit more. We now present our main stability result for finite chains, which clarifies the relationship between the Dobrushin coefficient and stability.

Theorem 4.3.5 *Let p be a stochastic kernel on S with Markov operator $\mathbf{M}$. The following statements are equivalent:*

1. *The dynamical system $(\mathscr{P}(S), \mathbf{M})$ is globally stable.*
2. *There exists a $t \in \mathbb{N}$ such that $\alpha(p^t) > 0$.*

Another way to phrase the theorem is that $(\mathscr{P}(S), \mathbf{M})$ is globally stable if and only if there is a $t \in \mathbb{N}$ such that, given any pair of states x, x', one can find at least one state y such that $p^t(x, y)$ and $p^t(x', y)$ are both positive. Thus, if we run two Markov chains from any two starting points x and x', there is a positive probability that the chains will meet. This is connected with global stability because it rules out the kind of behavior seen in example 4.1.4 (page 64), where initial conditions determine long-run outcomes.

Exercise 4.49 Consider the periodic kernel in exercise 4.44. Show that $\alpha(p^t) = 0$ for every $t \in \mathbb{N}$.

Exercise 4.50 Prove that if $\min_{x \in S} p^t(x, \bar{y}) =: \epsilon > 0$ for some $\bar{y} \in S$, then $(\mathscr{P}(S), \mathbf{M})$ is globally stable.

Exercise 4.51 Stokey and Lucas (1989, thm. 11.4) prove that $(\mathscr{P}(S), \mathbf{M})$ is globally stable if there exists a $t \in \mathbb{N}$ such that $\sum_{y \in S} \min_{x \in S} p^t(x, y) > 0$. Show how this result is implied by theorem 4.3.5.

Exercise 4.52 Prove theorem 4.3.5 using results from earlier in the text.

Let's consider how to apply theorem 4.3.5. In view of exercise 4.50, if there exists a y with $p(x, y) > 0$ for all $x \in S$, then $\alpha(p) > 0$ and global stability holds. A case in point is Hamilton's kernel (4.8) on page 75, which is globally stable as a result of the strict positivity of column two.

Next consider Quah's kernel p_Q (page 76). We know from theorem 4.3.2 that at least one stationary distribution exists, and we calculated a stationary distribution in exercise 4.42. We should now check that there are not many stationary distributions—otherwise exhibiting one of them is not very interesting. Also, the stationary distribution becomes a better predictor of outcomes if we know that all trajectories converge to it.

Exercise 4.53 Show that the Dobrushin coefficient $\alpha(p_Q)$ is zero.

Since $\alpha(p_Q) = 0$, let's look at the higher order iterates. In his study Quah calculates the 23rd-order kernel

$$p_Q^{23} = \begin{pmatrix} 0.61 & 0.27 & 0.09 & 0.03 & 0.00 \\ 0.37 & 0.32 & 0.20 & 0.09 & 0.02 \\ 0.14 & 0.23 & 0.31 & 0.25 & 0.07 \\ 0.04 & 0.11 & 0.25 & 0.39 & 0.22 \\ 0.00 & 0.01 & 0.04 & 0.12 & 0.82 \end{pmatrix} \tag{4.23}$$

Exercise 4.54 Show that $\alpha(p_q^{23}) > 0$.

Exercise 4.55 As $(\mathscr{P}(S), \mathbf{M}_Q)$ is globally stable, we can iterate $\mathbf{M}_Q$ on any initial condition ψ to calculate an approximate fixed point ψ^*. Take $\psi = (1, 0, 0, 0, 0)$ as your initial condition and iterate until $d_1(\psi \mathbf{M}_Q^t, \psi \mathbf{M}_Q^{t+1}) < 0.0001$. Compare your result with that of exercise 4.42.

Exercise 4.56 Code a function that takes a kernel p as an argument and returns $\alpha(p)$. Write another function that repeatedly calls the first function to compute the smallest $t \geq 1$ such that $\alpha(p^t) > 0$, and prints that t along with the value $\alpha(p^t)$. Include a maximum value T such that if t reaches T the function terminates with a message that $\alpha(p^t) = 0$ for all $t \leq T$. Now show that the first t such that $\alpha(p_Q^t) > 0$ is 2.

One interesting fact regarding stationary distributions is as follows: Let p be a kernel such that $(\mathscr{P}(S), \mathbf{M})$ is globally stable, and let ψ^* be the unique stationary distribution. Let $(X_t)_{t \geq 0}$ be Markov-(p, x), where $\psi^*(x) > 0$. The *return time to* x is defined as the random variable

$$\tau(x) := \inf\{t \geq 1 : X_t = x\}$$

It turns out that for $\tau(x)$ so defined we have $\mathbb{E}\tau(x) = 1/\psi^*(x)$. We will skip the proof (see Norris, 1997, thm. 1.7.7), but let's try running a simulation. The pseudocode in algorithm 4.4 indicates how one might go about estimating $\mathbb{E}\tau(x)$.[15]

Algorithm 4.4: Computing the mean return time

```
for i in 1 to n do                          // n is the number of replications
    set t = 0
    set X = x
    repeat
        draw X ~ p(X, dy)
        set t = t + 1
    until X = x
    set τ_i = t
end
return n^{-1} Σ_{i=1}^n τ_i
```

Exercise 4.57 Implement algorithm 4.4 for Hamilton's Markov chain. Examine whether for fixed $x \in S$ the output converges to $1/\psi^*(x)$ as $n \to \infty$.

Finally, let's consider a slightly more elaborate application, which concerns so-called (s, S) inventory dynamics. Inventory management is a major topic in operations research that also plays a role in macroeconomics due to the impact of inventories on aggregate demand. The discrete choice flavor of (s, S) models accord well with the data on capital investment dynamics.

Let $q, Q \in \{0\} \cup \mathbb{N}$ with $q \leq Q$, and consider a firm that, at the start of time t, has inventory $X_t \in \{0, \ldots, Q\}$. Here Q is the maximum level of inventory that the firm is capable of storing. (We are studying (q, Q) inventory dynamics because the symbol S is taken.) If $X_t \leq q$, then the firm orders inventory $Q - X_t$, bringing the current stock to Q. If $X_t > q$ then the firm orders nothing. At the end of the period t demand D_{t+1} is observed, and the firm meets this demand up to its current stock level. Any remaining inventory is carried over to the next period. Thus

$$X_{t+1} = \begin{cases} \max\{Q - D_{t+1}, 0\} & \text{if } X_t \leq q \\ \max\{X_t - D_{t+1}, 0\} & \text{if } X_t > q \end{cases}$$

If we adopt the notation $x^+ := \max\{x, 0\}$ and let $\mathbb{1}\{x \leq q\}$ be one when $x \leq q$ and

[15] If $\psi^*(x) > 0$, then $(X_t)_{t \geq 0}$ returns to x (infinitely often) with probability one, so the algorithm terminates in finite time with probability one.

zero otherwise, then this can be rewritten more simply as

$$X_{t+1} = (X_t + (Q - X_t)\mathbb{1}\{X_t \leq q\} - D_{t+1})^+$$

or, if $h_q(x) := x + (Q - x)\mathbb{1}\{x \leq q\}$ is the stock on hand after orders for inventory are completed, as

$$X_{t+1} = (h_q(X_t) - D_{t+1})^+$$

We assume throughout that $(D_t)_{t \geq 1}$ is an IID sequence taking values in $\{0\} \cup \mathbb{N}$ according to distribution $b(d) := \mathbb{P}\{D_t = d\} = (1/2)^{d+1}$.

Exercise 4.58 Let $S = \{0, 1, \ldots, Q\}$ and let $\mathbf{M}_q$ be the Markov operator on S corresponding to these dynamics. Show that $(\mathscr{P}(S), \mathbf{M}_q)$ is always globally stable *independent* of the precise values of q and Q.

In what follows we let ψ_q^* denote the stationary distribution corresponding to threshold q.

Exercise 4.59 Show numerically that if $Q = 5$, then

$$\psi_2^* = (0.0625, 0.0625, 0.125, 0.25, 0.25, 0.25)$$

Now consider profits of the firm. To minimize the number of parameters, suppose that the firm buys units of the product for zero dollars and marks them up by one dollar. Revenue in period t is $\min\{h_q(X_t), D_{t+1}\}$. Placing an order for inventory incurs fixed cost C. As a result profits for the firm at time t are given by

$$\pi_q(X_t, D_{t+1}) = \min\{h_q(X_t), D_{t+1}\} - C\mathbb{1}\{X_t \leq q\}$$

If we now sum across outcomes for D_{t+1} taking $X_t = x$ as given, then we get

$$g_q(x) := \mathbb{E}[\pi_q(x, D_{t+1})] = \sum_{d=0}^{\infty} \pi_q(x, d) b(d) = \sum_{d=0}^{\infty} \frac{\pi_q(x, d)}{2^{d+1}}$$

which is interpreted as expected profits in the current period when the inventory state X_t is equal to x.

Exercise 4.60 One common performance measure for an inventory strategy (in this case, a choice of q) is long-run average profits, which is defined here as $\mathbb{E}g_q(X)$ when $X \sim \psi_q^*$ (i.e., $\sum_{x \in S} g_q(x)\psi_q^*(x)$). Show numerically that according to this performance measure, when $Q = 20$ and $C = 0.1$, the optimal policy is $q = 7$.

4.3.4 The Law of Large Numbers

In this section we continue our discussion of stability by investigating some probabilistic properties of sample paths. In particular, we reinforce our informal discussion of ergodicity in chapter 1 by analyzing the law of large numbers in the context of Markov chains.

In algorithm 4.2 (page 79) we computed an approximation to the marginal distribution ψ_t via Monte Carlo. The basis of Monte Carlo is that if we sample independently from a fixed probability distribution and count the fraction of times that an event happens, that fraction converges to the probability of the event (as determined by this probability distribution). This is more or less the frequentist definition of probabilities, but it can also be proved from the axioms of probability theory. The theorem in question is the law of large numbers (LLN), a variation of which is as follows:

Theorem 4.3.6 *If F is a cumulative distribution function on $\mathbb{R}$, $(X_t)_{t\geq 1} \overset{\text{IID}}{\sim} F$, and $h\colon \mathbb{R} \to \mathbb{R}$ is a measurable function with $\int |h(x)|F(dx) < \infty$, then*

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to \mathbb{E}h(X_1) := \int h(x)F(dx) \quad \text{as } n \to \infty \text{ with probability one} \tag{4.24}$$

This result is fundamental to statistics. It states that for IID sequences, sample means converge to means as the sample size gets large. Later we will give a formal definition of independence and prove a version of the theorem. At that time the term "measurable function" and the nature of probability one convergence will be discussed. Suffice to know that measurability of h is never a binding restriction for the problems we consider.

Example 4.3.7 If $(X_i)_{i=1}^n$ are independent standard normal random variates, then according to theorem 4.3.6 we should find that $n^{-1}\sum_{i=1}^n X_i^2 \to 1$. (Why?) You might like to check this by simulation.

Another use of the LLN: Suppose that we wish to compute $\mathbb{E}h(X)$, where h is some real function. One approach would be to use pen and paper plus our knowledge of calculus to solve the integral $\int_{-\infty}^{\infty} h(x)F(dx)$. In some situations, however, this is not so easy. If instead we have access to a random number generator that can generate independent draws $X_1, X_2, \ldots$ from F, then we can produce a large number of draws, take the mean of the $h(X_i)$ terms, and appeal to (4.24).

In (4.24) the sequence of random variables is IID. In some situations the LLN extends to sequences that are neither independent nor identically distributed. For example, we have the following result concerning stable Markov chains:

Theorem 4.3.8 *Let S be finite, let $\psi \in \mathscr{P}(S)$, let p be a stochastic kernel on S with $\alpha(p^t) > 0$ for some $t \in \mathbb{N}$, and let $h\colon S \to \mathbb{R}$. If $(X_t)_{t\geq 0}$ is Markov-(p, ψ), then*

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t) \to \sum_{x\in S} h(x)\psi^*(x) \quad \text{as } n \to \infty \text{ with probability one} \tag{4.25}$$

where ψ^* is the unique stationary distribution of p.

The left-hand side is the average value of $h(X_t)$, and the right-hand side is the expectation of $h(X)$ when $X \sim \psi^*$. Note that the result holds for *every* initial condition $\psi \in \mathscr{P}(S)$.

The proof of theorem 4.3.8 requires more tools than we currently have in hand.[16] The intuition is that when the chain is globally stable, X_t is approximately distributed according to ψ^* for large t. In addition the stability property implies that initial conditions are unimportant, and for the same reason X_t has little influence on X_{t+k} for large k. Hence there is a kind of asymptotic independence in the chain. Together, these two facts mean that our chain approximates the IID property that drives the LLN.

If $h(x) = 1$ if $x = y$ and zero otherwise (i.e., $h(x) = \mathbb{1}\{x = y\}$), then (4.25) becomes

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t) = \frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{X_t = y\} \to \psi^*(y) \quad \text{as } n \to \infty \tag{4.26}$$

This provides a new technique for computing the stationary distribution, via Monte Carlo. Exercise 4.61 illustrates.

Exercise 4.61 Let p_H be Hamilton's kernel, and let $h(x) = 1$ if $x = NG$ and zero otherwise. Take any initial condition, and draw a series of length 10^6. Compute the left-hand side of (4.25). Compare it with the right-hand side, calculated via the algebraic method shown in listing 4.1.

When the state space is small, this is a less efficient technique for computing the stationary distribution than the algebraic method used in listing 4.1. However, the computational burden of the algebraic method increases rapidly with the size of the state space. For large or infinite state spaces, a variation of the LLN technique used in exercise 4.61 moves to center stage. See §6.1.3 for details.[17]

The importance of theorem 4.3.8 extends beyond this new technique for computing stationary distributions. It provides a new *interpretation* for the stationary distribution: If we turn (4.26) around, we get

$$\psi^*(y) \cong \text{ fraction of time that } (X_t) \text{ spends in state } y$$

[16] A version of theorem 4.3.8 is proved in §11.1.1.

[17] The look-ahead method introduced in §6.1.3 concerns infinite state spaces, but it can be applied to finite state spaces with the obvious modifications.

This is indeed a new interpretation of ψ^*, although it is not generally valid unless the chain in question is stable (in which case the LLN applies).

Exercise 4.62 Give an example of a kernel p and initial condition ψ where this interpretation fails.

In the preceding discussion, h was an indicator function, which reduced the discussion of expectations to one of probabilities. Now let's consider more general expectations.

Exercise 4.63 Recall the firm introduced on page 83. Extending exercise 4.61, approximate expected profits at the stationary distribution using theorem 4.3.8. Compare your results to those of exercise 4.43.

Thus the LLN provides a new way to compute expectations with respect to stationary distributions. However, as was the case with probabilities above, it also provides a new interpretation of these expectations when the Markov chain is stationary. For example, if h denotes profits as above, then we have

$$\sum_{x \in S} h(x)\psi^*(x) \cong \text{ long-run average profits}$$

Again, this interpretation is valid when the chain in question is stationary, but may not be valid otherwise.

4.4 Commentary

Regarding deterministic, discrete-time dynamical systems, good mathematical introductions are provided by Holmgren (1996) and Wiggins (2003), who treat elementary theory, topological conjugacy, and chaotic dynamics. For dynamics from an economic perspective, see, for example, Stokey and Lucas (1989), Azariadis (1993), de la Fuente (2000), Shone (2003), Caputo (2005), Gandolfo (2005) or Ljungqvist and Sargent (2018).

The threshold externality model in example 4.1.4 is a simplified version of Azariadis and Drazen (1990). See Durlauf (1993) for a stochastic model with multiple equilibria. Dosi et al. (2019) study convergence and divergence in a large, agent-based model using simulation. Johnson and Papageorgiou (2020) review the evidence on economic development and cross-country convergence.

Our discussion of chaotic dynamics lacked economic applications, but plenty exist. The Solow–Swan model produces chaotic dynamics with some minor modifications (e.g., Böhm and Kaas 2000). Moreover rational behavior in infinite-horizon, optimizing models can lead to chaos, cycles, and complex dynamics. See, for example, Benhabib and Nishimura (1985), Venditti (1998), or Mitra and Sorger (1999). For more dis-

cussion of complex economic dynamics, see Medio (1995), Brock and Hommes (1998), Kikuchi (2008), or Matsuyama et al. (2016).

For a general discussion of the relationship between complexity theory and economics, see Arthur (2010).

Good references on finite state Markov chains include Norris (1997), Häggström (2002), and Bremaud (2020). These texts provide a more traditional approach to stability of Markov chains based on irreducibility and aperiodicity. It can be shown that every irreducible and aperiodic Markov chain is globally stable, and as a result satisfies the conditions of theorem 4.3.5 (in particular, $\alpha(p^t) > 0$ for some $t \in \mathbb{N}$). The converse is not true, so theorem 4.3.5 is more general.

The Dobrushin coefficient was introduced by Dobrushin (1956). For an alternative discussion of the Dobrushin coefficient in the context of finite state Markov chains, see Bremaud (2020).

The treatment of (s, S) dynamics in §4.3.3 is loosely based on Norris (1997). For another discussion of inventory dynamics see Stokey and Lucas (1989, sec. 5.14). An interesting analysis of aggregate implications is Nirei (2008). A treatment of discrete adjustment models can be found in Stokey (2008). Beare (2012) studies stability of Markov chains generated by Archimedean copulas.