Chapter 1

Introduction

Modeling of dynamics combines beautiful theory, challenging computational problems and great practical significance. What more could we ask for when seeking an exciting field?

As an example of practical significance, at the time of writing, the COVID-19 pandemic is ongoing, with restrictions on movement and activities slowly being relaxed as vaccination rates increase. The question being debated in policy circles is: how fast should restrictions be eased relative to expansion of vaccinations across the population? Is an 80% vaccination rate for adults sufficient for the end of personal restrictions, or should we wait for higher? How many fatalities will we have to tolerate?

In addressing this question, accurate modeling is, without exaggeration, a matter of life and death.

The debate on how to act is intense partly because predictions vary with the assumptions and simplifications inserted into each model. Different research teams tend to approach modeling in different ways. This is not necessarily a failing of the modeling process, since creating a distribution of beliefs across outcomes based on an ensemble of models is reasonable strategy for evaluating proposed policies.[1]

Modeling pandemics is difficult because it involves both evolution of the pathogen and, more critically, assumptions about human behavior. The choices individuals make in the way they live their social and work lives have enormous impact on the spread of infection. Predicting human behavior is hard, particularly since those humans making decisions are themselves basing their choices on forecasts of the future consequences of individual and collective decisions. On top of these problems, we also must contend with shocks to the system, such as mutations of the virus, changes

[1] The only problem with this idea is that politicians seem to put full probability mass on model predictions that they believe will bolster their electoral prospects.

in the policies of other countries, progress in treatments, and so on.

While modeling pandemics involves a specific form of dynamics, the discussion above sounds much like a vast range of economic modeling problems. The outcomes we observe in economic systems depend on individual choices and the aggregate impact of those choices. Moreover, there is feedback not just from individual choices to aggregate outcomes, but also from aggregate outcomes to individual choices (e.g., asset prices depend on investment decisions and investment decisions depend on asset prices).

Significantly, individual choices are made on the basis of both current conditions and beliefs held by the individuals in question over future conditions. Hence beliefs are part of the feedback loop. At the same time, individual and aggregate conditions are influenced by external shocks, which affect both current outcomes and beliefs about the future.

To handle such complexity, we need computational muscle power and careful, well-constructed theory. Theory guides us not just in building models, but also in designing algorithms to facilitate efficient computation. This algorithmic theory is becoming more important every year, since clever algorithms can revolutionize what is possible on the existing set of hardware.

In this first chapter we explore some of the foundations of dynamic modeling from a high-level perspective. In subsequent chapters we will return to all of the themes raised here and analyze them in detail.

1.1 Stochastic Dynamics

This section introduces finite Markov models and hints at their vast range of applications.

1.1.1 Markov Dynamics

To understand what I am about to discuss—and what most of the book is about—you need to be familiar with Markov models. For now I will restrict attention to "finite" Markov models, often called Markov chains. Some of what follows might be familiar to you but I recommend you skim it anyway.

Markov chains are the simplest nontrivial class of stochastic dynamic systems. At the same time, the theory of Markov chains is far from trivial. Markov chains play central roles in fields as diverse as quantum mechanics, biology, artificial intelligence, management science, finance, sociology and, of course, economics. An enormous number of dynamic systems can be accurately replicated with Markov chains.

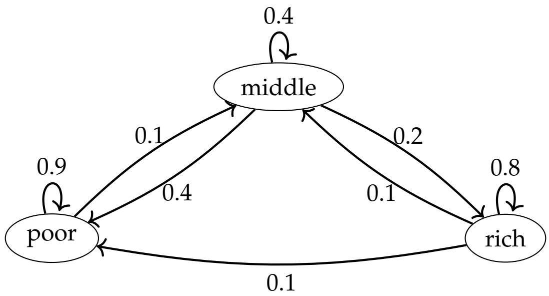

Figure 1.1 A simple Markov model

Later we will cover the formal theory of Markov chains, as well as general state Markov processes, in great detail. For now, let's look at an example. Consider figure 1.1. We imagine that a household can be in one of three states: poor, rich or middle class. The arrows show the transition probabilities over one year. For example, a rich household has a 10% probability of becoming poor in one year, while a poor household has a 90% chance of remaining poor.

What could we deduce from this simple model if we were to take its numbers seriously? One way to address this question is to consider what would happen to a large population of households that follows these dynamics. In particular, we can run a simulation where we assign households to states according to some specified initial distribution (the fraction of households in each state at the start of the simulation run) and then update each one independently according to the probabilities in figure 1.1.

Figure 1.2 shows the results of such a simulation, with 1,000 households. The distribution $\psi_0 = (p_1, p_2, p_3)$ in the title of each subfigure indicates the share of households in each state (poor, middle, rich) at the start of time. The bar graph below the title shows the distribution (i.e., share of population in each state) at the end of the simulation run, after updating each household 100 times.

The most striking result of the simulation is that the final distribution is independent of the initial distribution. Later we will prove that this result is exactly what we should expect, given the dynamics specified in figure 1.1: for this model, there is a unique distribution ψ, called the stationary distribution of the model, such that the distribution of the population across states *always* converges to ψ as the population size and time go to $+\infty$, regardless of the initial distribution.

One reason convergence to the unique stationary distribution is important is that it provides a firm prediction from the model. If the system we are observing has been evolving for some time, then we expect the observed distribution across states at the current time to match ψ. While this might or might not hold when we observe the

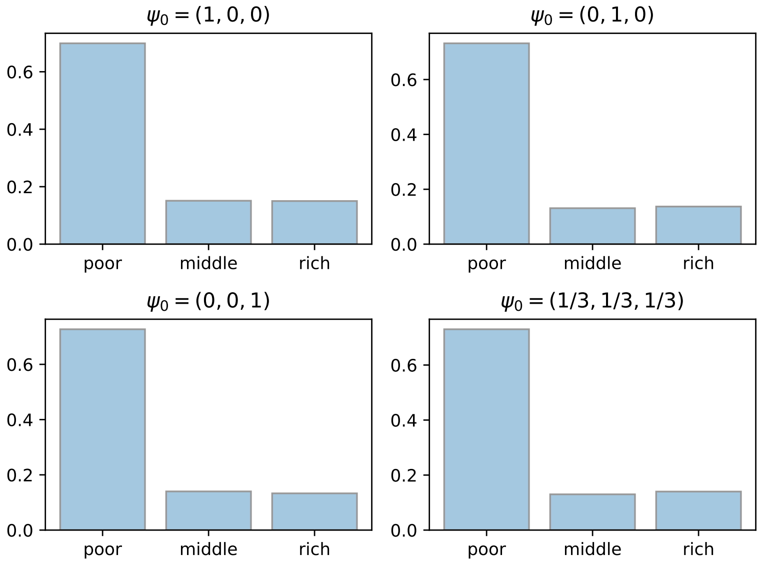

Figure 1.2 Distribution of population after 100 periods

system, we still value the fact that the model makes a firm prediction. Models that make strong predictions are falsifiable, and this property lies at the heart of scientific analysis.[2]

Of course, models that predict a unique long run outcome, independent of initial conditions, are not the only models of interest in economics. Sometimes dynamics are "path dependent," meaning that initial conditions never cease to exert influence on future outcomes. One commonly cited example in the popular science literature is the standard Latin script keyboard layout, called the QWERTY keyboard, which was designed in 1873. While more efficient layouts have been proposed since, all have failed to capture significant market share.

We can modify the class transition model discussed above to generate path dependence. For example, figure 1.3 shows another version of the model, but now there is no route out of poverty. In the language of Markov chains, we call poor an *absorbing state*. We will see other examples of absorbing states soon, in a classic model of segregation.

[2]For example, Wikipedia asserts that "Astrology is a pseudoscience that claims to divine information about human affairs and terrestrial events by studying the movements and relative positions of celestial objects." On what basis case we assert that astrology is a pseudoscience? The answer is that astrology makes no falsifiable predictions.

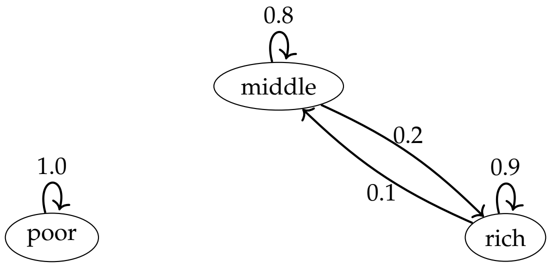

Figure 1.3 No route out of poverty

Before moving on to larger and more interesting models, there is another idea we can introduce here, which is one of the core topics of stochastic dynamic modeling. I am referring to the concept of *ergodicity*, which was formalized in part by Ludwig Boltzmann (1844–1906) in his work on statistical mechanics. Loosely speaking, for a dynamic system, ergodicity is said to hold if time series averages coincide with cross-sectional averages.

To clarify this concept in the setting of Markov chains, consider again the model of class dynamics in figure 1.1. This is the first set of probabilities we studied, where the long run cross-sectional distribution was calculated by simulation in figure 1.2. There is another way to calculate this distribution: if we take a single household and record the fraction of time that it spends in each state over a very long simulation run, the distribution, shown in figure 1.4, is identical in the limit to the one we obtained in figure 1.2.

In the setting of this model, ergodicity means that a long-lived household will experience the different states of the model in proportion to their probability under the cross-sectional (stationary) distribution. In contrast, the path dependent model in figure 1.3 is not ergodic. For example, the experience of an initially poor household will be continuous poverty, even when the cross-sectional distribution indicates a large fraction of the population is either middle class or rich.

Because expectations are computed from probabilities, when ergodicity holds we can also recover cross-sectional expectations from time series. For example, for some arbitrary function h, we have

$$\frac{1}{T}\sum_{t=1}^{T} h(X_t) \approx \frac{1}{M}\sum_{i=1}^{M} h(X_i')$$

Here $(X_t)_{t=1}^T$ is a time series generated by the model and $(X_i')_{i=1}^M$ is a large population simulated according to the model, at some fixed point in time (that is, a cross-section).

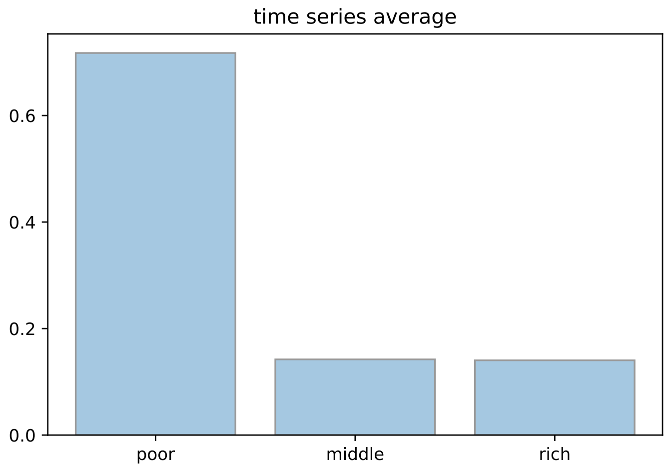

Figure 1.4 Fraction of time spent in each state by a single household

For example, if $h(x) = x$, then the claim is that the time series sample mean can be used to compute the population mean and vice versa.

Ergodicity is a fundamental concept that plays a key role in economics, finance and econometrics, as well as many other fields. We return to the study of ergodicity in chapter 4. As well as presenting theory, we will also run simulations that show ergodicity in action.

1.1.2 Interacting Particle Systems

One unrealistic feature of the model of class transitions discussed in §1.1.1 is that households do not interact. Instead, each one updates completely independently of all others. In practice households interact both directly—for example, by influencing each other's choices—and indirectly, by contributing to the determination of aggregate quantities and prices. Economists and statisticians are steadily building tools to help us understand these interactions.

In this section, we begin to consider interaction between agents in a Markov setting. However, the first model we consider, called the Ising model, is from statistical mechanics rather than economics. It is one of a family of models called "interacting particle systems." In these systems, individual entities, or particles, influence the behavior of neighboring particles, and these local interactions shape macro-level outcomes. Even though the model is not from economics, you might already have a sense

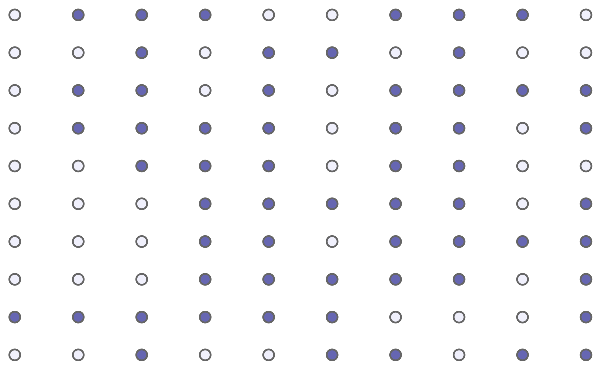

Figure 1.5 A spin configuration in the plane (light $= -1$ and dark = $+1$)

of its relevance: we can start to identify individuals, households and firms with the "particles" in the model.

The Ising model is one of the foundational models of ferromagnetics in particular and phase transitions more generally. Evolution within the basic model takes place on a lattice of points in the plane, denoted below by L. Magnetization at each point in the lattice is in one of two spin states, up or down, which we identify with $+1$ and -1 respectively. Thus the state of the system at any given point of time is a particular configuration of spins across lattice points. Mathematically, such a configuration is a map from the lattice to the set $\{+1, -1\}$. The *state space* for the model, denoted here by S, is the set of all such maps. A typical element of S is usually denoted by σ and called a *spin configuration*. A small example is shown in figure 1.5.

The main systematic force governing dynamics within the model is that magnets that are close prefer to be aligned in the same direction. In thermodynamics, however, the model is never completely at rest at normal temperatures, due to continuous fluctuation of individual molecules. Hence physicists, in describing the equilibrium of the system, refer not to a fixed spin configuration, but rather to a distribution over S, the set of all configurations. The equilibrium distribution tells the frequency at which different spin configurations are likely to be observed as we continue to view the system.

Under standard assumptions, the equilibrium distribution takes the form

$$\psi(\sigma) = c \exp\left(\frac{J}{2} \sum_{i \in L} \sum_{j \sim i} \sigma(i)\sigma(j)\right) \qquad (\sigma \in S)$$

Here $\psi(\sigma)$ is the probability assigned to the configuration σ in equilibrium, $\sigma(i) \in \{-1, +1\}$ is the spin on lattice point $i \in L$, while $i \sim j$ indicates that i and j are neighbors. The value c is a positive constant and J is inverse to the temperature.[3] In line with our discussion above, ψ puts large mass on configurations where many neighbors have the same sign (in which case $\sigma(i)\sigma(j)$ is positive).

Most of the challenges associated with the Ising model are due to the fact that the state space is very large. There are 2^m possible spin configurations, where m is the number of points in the lattice. When the lattice is even moderately large, this number is enormous. So if we want to compute an expectation such as

$$\mathbb{E}_\psi h := \sum_{\sigma \in S} h(\sigma)\psi(\sigma) \tag{1.1}$$

where ψ is the equilibrium distribution and h is some function of interest, the sum cannot be calculated directly, even with massively powerful computers.

As a result, mathematicians and physicists have developed other approaches to evaluating these kinds of expectations. The most important family of methods is those that are based on Markov chain Monte Carlo (MCMC). The idea behind MCMC is to *design a Markov chain such that ψ, the distribution of interest, is the stationary distribution of the chain.* The next step is to generate a long time series $(\sigma_1, \sigma_2, \ldots, \sigma_T)$ from the Markov chain and then approximate the expectation $\mathbb{E}_\psi$ via the sample mean

$$\mathbb{E}_\psi \approx \frac{1}{T} \sum_{t=1}^{T} h(\sigma_t)$$

The key concept that connects the cross-sectional average in (1.1) and this time series average is ergodicity. Thus, the Markov chain Monte Carlo scheme must be designed in order to produce ergodicity.

These kinds of probabilistic methods have been revolutionary not just in statistical mechanics, but also in Bayesian statistics and machine learning. While we will not cover the Monte Carlo methods related to the Ising model in detail, the core ideas of ergodicity, simulation and statistical methods in a Markov setting pervade the pages of this textbook.[4]

[3] If $J = 0$ then the spins are called noninteracting, which returns us to a setting of the independently evolving entities, such as the households studied in §1.1.1.

[4] It is worth mentioning that many of the probabilistic arguments for the Ising model, including the beautiful idea of perfect sampling via coupling from the past, draw on coupling methods, which form a core part of our stability arguments for Markov chains.

1.1.3 A Model of Segregation

What does all this have to do with economics? To give one example, let us now describe a small part of the work of Thomas Schelling, who received the Nobel Prize in Economic Sciences in 2005.[5] The part I refer to is Schelling's model of residential segregation (Schelling 1969, 1971), which seems to attract more attention every year that passes.

Schelling designed his model to help explain the rise and prevalence of segregated neighborhoods in US cities. In particular, beginning in the 1950s, US cities witnessed large population movements along racial lines. For example, white middle class households shifted out of inner city areas in cities such as Chicago, Detroit and Cleveland. At the same time, black households shifted into these areas, often from the rural South. By the 2010 census, Chicago's Washington Park was recorded as 97% black. These changes to the structure of neighborhoods have had large and lasting impacts on the distribution of tax revenue, provision of social services, and other social phenomena.

Schelling's main insight was that, even if people are comfortable living in mixed neighborhoods, which contain roughly even quantities of people of both colors, such neighborhoods are inherently unstable once the model becomes dynamic. Configurations that are unstable are less likely to be observed than stable ones, just as a pendulum pointing straight up (an unstable equilibrium) is observed in the real world less often than a pendulum pointing straight down (a stable equilibrium). Below we explore, through modeling and simulation, Schelling's idea that mixed neighborhoods are unstable.

In the version of the model we analyze here, the two races are imaginary and will be called "light people" and "dark people." The terminology refers only to the shade of the circles that we use in the figures. Thus, segregation can be along any recognizable division, including class, education, skin color, etc. Within the model, the agents—or households—are located on a two-dimensional lattice, like the interacting particle system discussed in §1.1.2.

Following Schelling's initial specification, we assume that each household's satisfaction with their current location depends on the color of their neighbors. Specifically, the household will be regarded as happy in their current location whenever at least half of their neighbors are of the same color as they are. If less than half of their neighbors are of the same color, the household becomes unhappy and seeks to move.

[5]Thomas Schelling passed away in 2016, at the age of 95. (It seems that the life expectancy of academic economists is strongly positively correlated with intelligence and creativity.) My favorite quote regarding Schelling is this one, by Rajiv Sethi: "[Schelling's] lack of concern with professional methodological norms allowed him to generate new knowledge with great freedom, and to make innovations in method that may end up being even more significant than his specific insights into economic and social life."

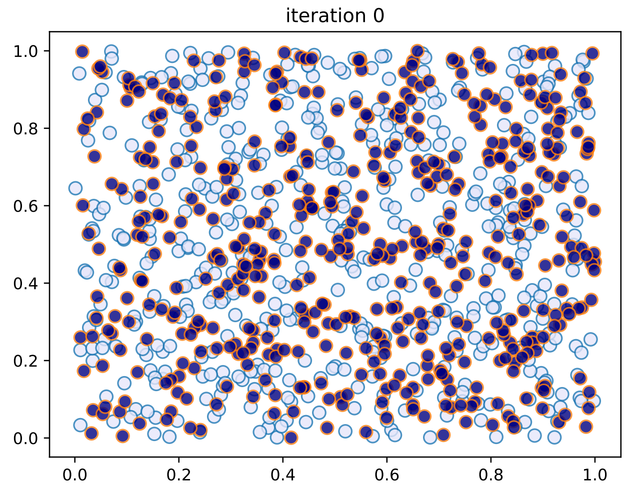

Figure 1.6 An initial configuration of households

Schelling emphasized the fact that the preferences of each household, as just described, are not overtly racist in the following sense: households are perfectly comfortable living in a mixed neighborhood. Only when they start to feel isolated do they wish to move. Hence the assumptions do not rule out prevalence of mixed neighborhoods directly. Survey data and empirical evidence collected over the past few decades have supported the preferences posed by Schelling.[6]

Schelling ran his simulation manually, using a chessboard. In our version of the model, we take the lattice to be all pairs (x, y) of 64 bit floating point numbers in the unit square $[0, 1] \times [0, 1]$. An initial configuration over the lattice is formed by randomly assigning colors to n households and then randomly assigning each household to a location (x, y), using an independent bivariate draw from the uniform distribution. Any locations not selected in this process are regarded as unoccupied. Figure 1.6 shows one typical realization.

At each turn, one of the n agents is selected randomly, with uniform probability. If the household is happy, no change occurs. If the household is unhappy, a new location (x', y') is selected randomly, with x' and y' being drawn independently from a uniform distribution. If the household is happy at (x', y'), the turn stops. If not, a new location is selected and the process repeats until the household is happy.

Note the similarity to the Ising model. In the latter, local interactions are through

[6]See, for example, Clark and Fossett (2008) or Card et al. (2008).

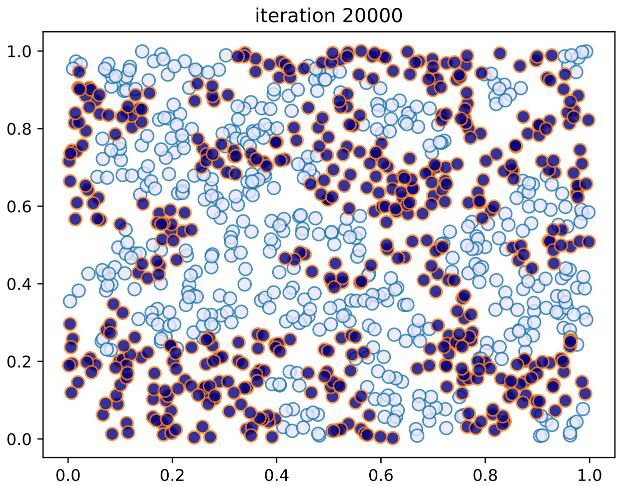

Figure 1.7 An absorbing state with significant segregation

magnetic effects. Magnets that are close prefer to be aligned in the same direction. Similarly, in the Schelling model, household that are close prefer to be of the same race. (In other related models, such as voter models, agents prefer to be close to those who share the same opinions.)

Let's now look at how the system evolves when run according to the dynamics specified above. The set of neighbors for a given household is specified as the closest 10 households, measured by Euclidean distance. Thus, a household is happy if five or more neighbors are of the same color. Households are randomly selected and updated, as described above. Figure 1.7 shows one realization after 10,000 such updates. Testing the happiness of households at this point, we find that all are happy. Hence, the system has reached a completely stable configuration: no further movement occurs.

The most interesting result is that the residential pattern has gone from completely mixed (figure 1.6) to significantly segregated. Moreover, repeating the simulation any number of times produces a similar result (as you can verify using the code in the accompanying Jupyter code book). Hence, mixed neighborhoods are unstable and segregated neighborhoods are stable.

What is the intuition behind this result? In essence, there is a positive feedback effect associated with each move. When a light household moves from an unhappy location to a happy one, it makes the neighborhood that it left more dark and its new location more light. Dark households in the new location might now find themselves outnumbered and hence shift to a darker location. This chain reaction continues, with every move destabilizing mixed neighborhoods and reinforcing segregation.

Unlike many economic research exercises, Schelling is not just rationalizing what we already observe with a mathematical model. In fact the segregation produced by the Schelling model is a classic example of an emergent phenomenon: a macro-level pattern not inherent in individual choices, as a result of interactions between these individuals. Schelling himself emphasized the importance of such phenomena within economics (Schelling 1969):

> Economists are familiar with systems that lead to aggregate results that the individual neither intends nor needs to be aware of, the results sometimes having no recognizable counterpart at the level of the individual. The creation of money by a commercial banking system is one; the way that savings decisions cause depressions or inflations is another.

Schelling's model shows how the decisions of many agents, combined with resource constraints, can aggregate in surprising and important ways.[7]

1.1.4 A Markov Perspective

At this point, let us turn back to Markov chains and try to provide a more formal interpretation of the dynamics in the Schelling model. The model as described is indeed a Markov chain. For the state space S, we take the set of all configurations of households across the unit square. We can express S as the set of all mappings σ from $L :=$ all pairs of 64 bit floating point numbers in $[0,1] \times [0,1]$ to $E := \{0,1,2\}$. Here 0 represents light, 1 represents dark and 2 represents unoccupied.

This state space is astronomically large—larger than the number of atoms in the known universe. Nevertheless, it is finite, and the process we used to update from current state σ_t, which is the current configuration of households, to next period state σ_{t+1}, depends only on the current state and independent draws of random numbers. This is the essence of the Markov property.

As we simulated the system, we noticed that it soon converges to a state where all households are happy. In the language of Markov chains, such convergence indicates that we have reached an absorbing state. As we saw in our discussion of class transitions in §1.1.1, existence of an absorbing state means that the dynamics of the model fail to be ergodic.

[7] Another example, which is related to savings, recessions, and inflation, is the story of the Capitol Hill baby sitting co-op, originally related by Sweeney and Sweeney (1977), and popularized in a series of articles by Paul Krugman.

On reflection, the fact that the neighborhood structure becomes fixed and unchanging, due to arrival at an absorbing state, contradicts what we observe in real life. Neighborhoods are constantly in flux. This is reminiscent of the Ising model, where, at normal temperature ranges, fluctuations at the atomic level continue to perturb the system.

To allow for constant flux, let us now make a small modification to the model: every time a household is updated, the process runs as before but, in addition, once the update has occurred, the color of the household is flipped with small probability ϵ. This loosely captures the idea that, when households move across cities—or perhaps out of cities while others move in—the move can be for reasons other than homophilic (same-race) preference.

The most significant aspect of this change is that the model is now ergodic. How can we be certain of this fact given the enormous size of the state space? The reasoning is from the theory of Markov chains. In essence, there is now sufficient "mixing" to ensure that the current state can evolve into any other possible state once enough time has elapsed. We will study exactly how mixing generates convergence and ergodicity in chapter 4.

Figures 1.8–1.9 show some results generated by simulating under the modified update rule, with ϵ set to 0.01. Figure 1.8 is the initial configuration and figure 1.9 is the result of 500,000 updates. In interpreting figure 1.9, it is important to remember that the displayed configuration is not an absorbing state, since continuous mixing implies that the neighborhoods always shift. Nonetheless, the general pattern is representative of other simulations under the same update rule: the additional mixing introduced by the flip modification leads to *more* segregation, rather than less. (Compare figure 1.9 with figure 1.7, which exhibits less severe segregation under the original Schelling dynamics.)

Why would additional mixing lead to more segregation? Doesn't mixing tend to break up segregated neighborhoods? The basic intuition is that mixing shocks—and hence destabilizes—configurations that are only partially stable. The segregation generated under the original Schelling dynamics is not particularly extreme. Hence it can still be destabilized by shocks to the system.

(Numerous optimization algorithms use some form of randomization for essentially the same purpose—to continue to explore the whole domain of the objective function rather only follow local dynamics. Only following local dynamics leads to local optimizers that might be a long way from the global optimizer—just as the principle of "always walking uphill" might not lead to the top of the highest mountain in a fixed geographical area containing many hills.)

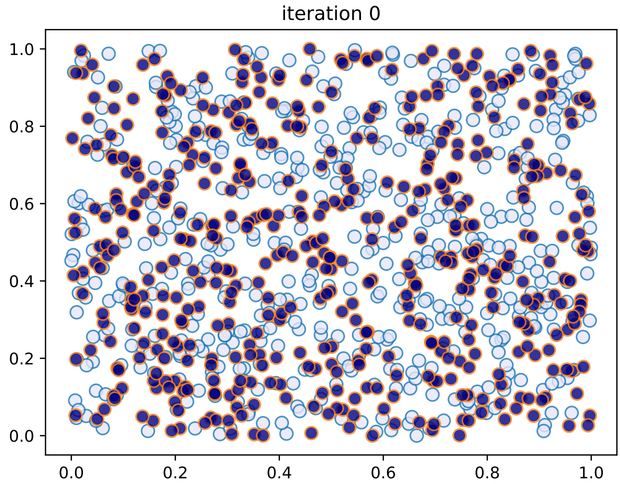

Figure 1.8 Another initial configuration of households

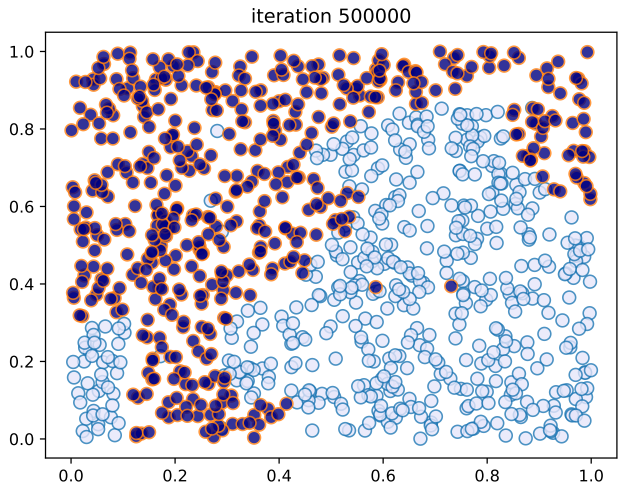

Figure 1.9 The result of iteration with low level mixing

1.2 Where to From Here?

We will not return to the Schelling model, since it is only one of many interesting models that we wish to understand. However, the underlying concepts and the set of questions that the model raises will continue to direct us as we expand our knowledge of stochastic dynamics and computing. In the rest of this section we provide further guidance on the next steps in our journey.

1.2.1 General State Space

All of the Markov models we have dealt with so far have a finite state space. We also need to consider Markov models where the state space is infinite. While many concepts and principles persist across this transition, there are some major differences. Hence we must invest effort in learning about both.

Let's begin with a very simple Markov system on the real line $\mathbb{R}$. It takes the form

$$X_{t+1} = aX_t + b + W_{t+1}, \quad \text{where} \quad W_{t+1} \overset{\text{IID}}{\sim} N(0,1) \tag{1.2}$$

and X_0 is a constant. This system is typically called the Gaussian AR(1) model. It is important to note for what follows that X_t and W_{t+j} are independent for all $j \geq 1$, since W_{t+j} only affects X_{t+j} and after.

Despite the uncountable state space, the system in (1.2) is easy to analyze. For starters, every X_t is normally distributed.

Exercise 1.1 Prove this. (Note: solutions to exercises can be found in appendix C.)

One of the many nice things about normal distributions is that they are determined by only two parameters, the mean and the variance. If we can find these parameters, then we know the distribution. So suppose that $X_t \sim N(\mu_t, v_t)$, where the constants μ_t and v_t are given. If you are familiar with manipulating means and variances, you will be able to deduce from (1.2) that $X_{t+1} \sim N(\mu_{t+1}, v_{t+1})$, where

$$\mu_{t+1} = a\mu_t + b \quad \text{and} \quad v_{t+1} = a^2 v_t + 1 \tag{1.3}$$

Paired with initial conditions μ_0 and v_0, these laws of motion pin down the sequences $(\mu_t)_{t\geq 0}$ and $(v_t)_{t\geq 0}$, and hence the distribution $N(\mu_t, v_t)$ of X_t at each point in time. A sequence of distributions starting from $X_t \sim N(1.0, 1.0)$ is shown in figure 1.10. The parameters are $a = 0.9$ and $b = 1.0$.

In the figure it appears that the distributions are converging to some kind of limiting distribution. This is due to the fact that $|a| < 1$, which implies that the sequences in (1.3) are convergent. The limits are

$$\mu^* := \lim_{t\to\infty} \mu_t = \frac{b}{1-a} \quad \text{and} \quad v^* := \lim_{t\to\infty} v_t = \frac{1}{1-a^2} \tag{1.4}$$

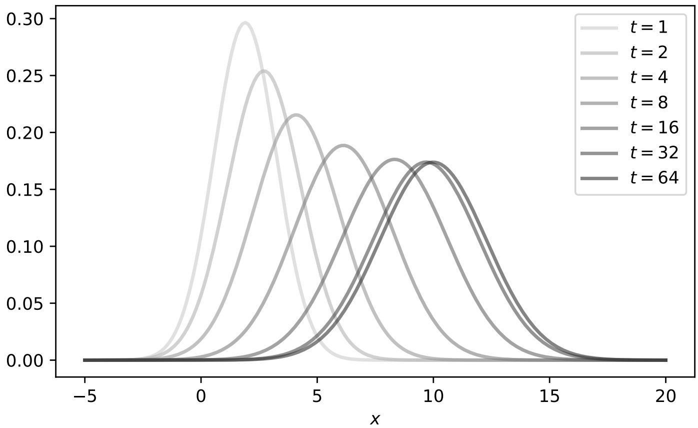

Figure 1.10 Sequence of marginal distributions

Hence the distribution $N(\mu_t, v_t)$ of X_t converges to $N(\mu^*, v^*)$.[8] Note that this "equilibrium" is a distribution rather than a single point—just like the Ising model, as well as the Schelling model with added mixing that we discussed in §1.1.4.

All this analysis depends, of course, on the law of motion (1.2) being linear, and the shocks being normally distributed. How important are these two assumptions in facilitating the simple techniques we employed? The answer is that they are both critical, and without either one we must start again from scratch.

To illustrate this point, let's briefly consider the threshold autoregression model

$$X_{t+1} = \begin{cases} A_1 X_t + b_1 + W_{t+1} & \text{if } X_t \in B \subset \mathbb{R}^n \\ A_2 X_t + b_2 + W_{t+1} & \text{otherwise} \end{cases} \tag{1.5}$$

Here X_t is $n \times 1$, A_i is $n \times n$, b_i is $n \times 1$, and $(W_t)_{t \geq 1}$ is an IID sequence of normally distributed random $n \times 1$ vectors. Although, for this system, the departure from linearity is relatively small (in the sense that the law of motion is at least piecewise linear), analysis of dynamics is far more complex. Through the text we will build a set of tools that permit us to analyze nonlinear systems such as (1.5), including conditions used to test whether the distributions of $(X_t)_{t \geq 0}$ converge to some stationary (i.e., limiting) distribution. We also discuss how one should go about computing the stationary distribu-

[8] What do we really mean by "convergence" here? We are talking about convergence of a sequence of *functions* to a given function. But how to define this? There are many possible ways, leading to different notions of equilibria, and we will need to develop some understanding of the definitions and the differences.

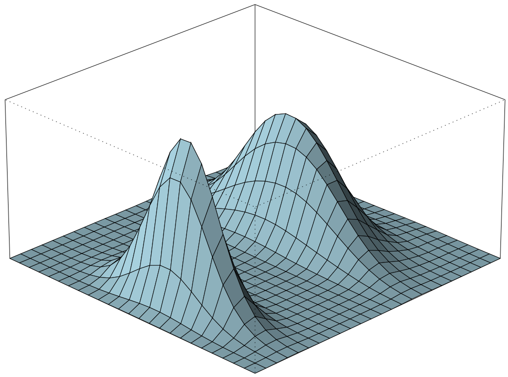

Figure 1.11 Stationary distribution

tions of nonlinear stochastic models. Figure 1.11 shows the stationary distribution of (1.5) for a given set of parameters, based on such a computation.

Now let's return to the linear model (1.2) and investigate its sample paths. Figure 1.12 shows a simulated time series over 250 periods. The initial condition is $X_0 = 14$, and the parameters are as before. The horizontal line is the mean μ^* of the stationary distribution. The sequence is obviously correlated, and not surprisingly, shows no tendency to settle down to a constant value. On the other hand, the sample mean $\bar{X}_t := \frac{1}{t}\sum_{i=1}^{t} X_i$ seems to converge to μ^*, as shown in figure 1.13.

The convergence of $\bar{X}_t$ certainly does not follow from the classical law of large numbers, since $(X_t)_{t\geq 0}$ is neither independent nor identically distributed. Instead, it follows from ergodicity, which we discussed previously in the context of finite Markov chains. We will prove this fact later in the text.

To give a sense of why ergodicity matters here, suppose that our simple model is being used to represent a given economy over a given period of time. Suppose further that the precise values of the underlying parameters a and b are unknown, and that we wish to estimate them from the data. The method of moments technique proposes that we do this by identifying the first and second moments with their sample counterparts. That is, we set

$$\text{first moment } = \mu^*(a,b) = \frac{1}{t}\sum_{i=1}^{t} X_t$$

$$\text{second moment } = v^*(a,b) + \mu^*(a,b)^2 = \frac{1}{t}\sum_{i=1}^{t} x_t^2$$

The right-hand side components $\frac{1}{t}\sum_{i=1}^{t} X_t$ and $\frac{1}{t}\sum_{i=1}^{t} X_t^2$ are collected from data, and

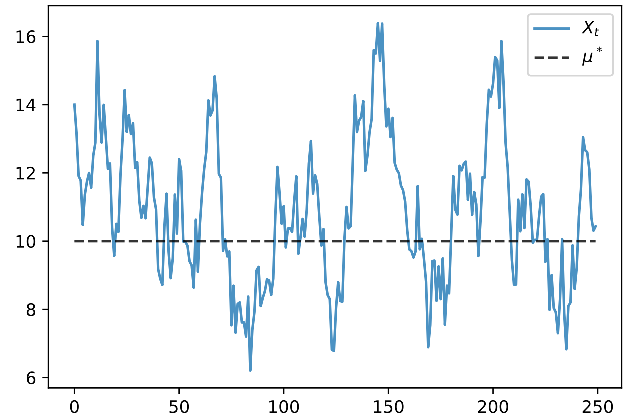

Figure 1.12 Time series

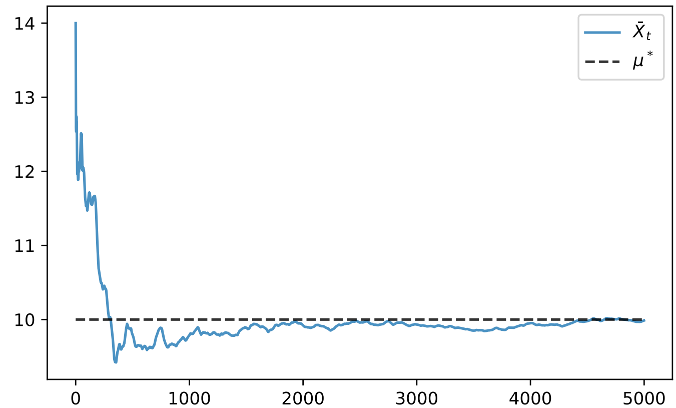

Figure 1.13 Sample mean of time series

the two equalities are solved simultaneously to calculate values for a and b.

The underlying assumption that underpins this whole technique is ergodicity. We will need to think hard about how to establish this property, especially when we go beyond the linear Gaussian model. As part of this journey, we will invest in learning some of the foundations of probability theory.

1.2.2 Forward Looking Agents

The behavioral rule in the Schelling model is very simple: each household chooses to stay or move depending on the current relative payoff of these two actions. There is no forward looking aspect to the decision process. Agents do not concern themselves with dynamics.

This assumption seems unrealistic. Real estate agents and the media often refer to "up-and-coming" neighborhoods, or neighborhoods that are "gentrifying". Both of these terms are inherently dynamic. Both buyers and sellers make some estimate of how prices and characteristics in a given area are likely to change.

In other economic settings, expectations over future outcomes are just as important. The purchase of any asset involves a consideration of likely future payoffs. The same is true of accepting or rejecting a job offer. Businesses forecast future revenue and costs when making investment decisions.

Our baseline assumption in these kinds of scenarios will be that agents act in order to optimize some kind of objective function. Optimization involving present and future outcomes, subject to constraints on resources, information and processing capabilities, is both reasonable for many types of actors and sufficiently broad to allow for a vast range of circumstances and assumptions. Since the objective function and constraints can include many factors, setting optimization as the baseline is not the same as insisting that economic agents are hyper-rational or completely selfish.

As such, we will need to consider methods aimed at optimizing various criteria in stochastic dynamic settings that are typically Markov. These kinds of problems are called Markov control problems or dynamic programs. They will be one of the main topics of the text.

When considering forward looking agents, there is also the issue of rational expectations, which is currently the mainstream paradigm in macroeconomics. To explain the basic idea in the context of the Schelling model, a rational expectations equilibrium would be one with the following properties. First, households make a guess of the Markov process that drives the entire residential configuration (σ_t) over time. Now they choose how to act on the basis of that guess. A rational expectations equilibrium is a set of decision rules that verifies their guess, in the sense that, under the choices that obey these decision rules, the macro configuration (σ_t) does in fact evolve as they predicted.

Rational expectations is not as crazy as it sounds. There must be some degree of consistency between people's beliefs about aggregate outcomes and what actually occurs. At the same time, it is wise to be skeptical. For example, in the Schelling model, the idea of imposing rational expectations seems ridiculous, given the enormous complexity of the system. Moreover, in the real world there are many more determinants of housing choices than just race.

The story is similar for the macroeconomy, which is not only massively complex but also nonstationary. Institutions, technology and social norms all change. Large shocks occur. Financial crises suggest that positive feedback loops are important, which in turn implies that dynamics can be strongly nonlinear. In these settings, it seems more likely that economic actors who need to forecast aggregate variables simply extrapolate based on recent experience or follow opinions in their social network.

As such, this textbook focuses less on rational expectations macroeconomics and more on providing foundational mathematical and computational skills, as well as other essential tools for dynamic modeling.

1.3 Commentary

Good sources of information on the Ising model include Lindvall (1992) and Kendall et al. (2005). For a discussion of the connection between the Ising model and the Schelling model, see Stauffer and Schulze (2007). Our modified Schelling simulation with added mixing was partly inspired by Zhang (2004), who also studies an ergodic version of the model. Some innovative recent work on neighborhood dynamics can be found in Knaap et al. (2019).

The problem of finding the set of neighbors of a given household in the Schelling problem is closely related to the k-nearest neighbors algorithm, a popular technique for classification and prediction. I have not pursued this connection, although it does occur to me that this or one of several other machine learning routines could be employed to determine where a given household will be happy.

The remarkable paper of Propp and Wilson (1996) proposed a method for exact sampling from the stationary distribution of the Ising model, using what is now known as "coupling from the past." Similar ideas could probably be applied to the ergodic version of Schelling's model studied above, although I haven't investigated this idea.

The style of modeling used by Schelling, combining simulation, simple decision rules, aggregation and the study of emergent phenomena, is now called "agent-based computational economics." Background reading in this field can be found in Tesfatsion and Judd (2006), Gallegati et al. (2017), Hommes and LeBaron (2018), Dosi and Roventini (2019) and Grabner et al. (2019).

For more on the rational expectations debate, see Akerlof and Shiller (2010).

Other high level material on computational economics includes Kendrick et al. (2005), Heer and Maussner (2009), Afonso and Vasconcelos (2015), and Fehr and Kindermann (2018).

Code and other information relevant to this chapter can be found at the author's website. See page xiv for more information.