1

Introduction and Preliminaries

Conceptual Outline

1.1 A deceptively simple model of the dynamics of a system is a deterministic iterative map applied to a single real variable. We characterize the dynamics by looking at its limiting behavior and the approach to this limiting behavior. Fixed points that attract or repel the dynamics, and cycles, are conventional limiting behaviors of a simple dynamic system. However, changing a parameter in a quadratic iterative map causes it to undergo a sequence of cycle doublings (bifurcations) until it reaches a regime of chaotic behavior which cannot be characterized in this way. This deterministic chaos reveals the potential importance of the influence of fine-scale details on large-scale behavior in the dynamics of systems.

1.2 A system that is subject to complex (external) influences has a dynamics that may be modeled statistically. The statistical treatment simplifies the complex unpredictable stochastic dynamics of a single system, to the simple predictable dynamics of an ensemble of systems subject to all possible influences. A random walk on a line is the prototype stochastic process. Over time, the random influence causes the ensemble of walkers to spread in space and form a Gaussian distribution. When there is a bias in the random walk, the walkers have a constant velocity superimposed on the spreading of the distribution.

1.3 While the microscopic dynamics of physical systems is rapid and complex, the macroscopic behavior of many materials is simple, even static. Before we can understand how complex systems have complex behaviors, we must understand why materials can be simple. The origin of simplicity is an averaging over the fast microscopic dynamics on the time scale of macroscopic observations (the ergodic theorem) and an averaging over microscopic spatial variations. The averaging can be performed theoretically using an ensemble representation of the physical system that assumes all microscopic states are realized. Using this as an assumption, a statistical treatment of microscopic states describes the macroscopic equilibrium behavior of systems. The final part of Section 1.3 introduces concepts that play a central role in the rest of the book. It discusses the differences between equilibrium and complex systems. Equilibrium systems are divisible and satisfy the ergodic theorem. Complex systems

are composed out of interdependent parts and violate the ergodic theorem. They have many degrees of freedom whose time dependence is very slow on a microscopic scale.

1.4 To understand the separation of time scales between fast and slow degrees of freedom, a two-well system is a useful model. The description of a particle traveling in two wells can be simplified to the dynamics of a two-state (binary variable) system. The fast dynamics of the motion within a well is averaged by assuming that the system visits all states, represented as an ensemble. After taking the average, the dynamics of hopping between the wells is represented explicitly by the dynamics of a binary variable. The hopping rate depends exponentially on the ratio of the energy barrier and the temperature. When the temperature is low enough, the hopping is frozen. Even though the two wells are not in equilibrium with each other, equilibrium continues to hold within a well. The cooling of a two-state system serves as a simple model of a glass transition, where many microscopic degrees of freedom become frozen at the glass transition temperature.

1.5 Cellular automata are a general approach to modeling the dynamics of spatially distributed systems. Expanding the notion of an iterative map of a single variable, the variables that are updated are distributed on a lattice in space. The influence between variables is assumed to rely upon local interactions, and is homogeneous. Space and time are both discretized, and the variables are often simplified to include only a few possible states at each site. Various cellular automata can be designed to model key properties of physical and biological systems.

1.6 The equilibrium state of spatially distributed systems can be modeled by fields that are treated using statistical ensembles. The simplest is the Ising model, which captures the simple cooperative behavior found in magnets and many other systems. Cooperative behavior is a mechanism by which microscopic fast degrees of freedom can become slow collective degrees of freedom that violate the ergodic theorem and are visible macroscopically. Macroscopic phase transitions are the dynamics of the cooperative degrees of freedom. Cooperative behavior of many interacting elements is an important aspect of the behavior of complex systems. This should be contrasted to the two-state model (Section 1.4), where the slow dynamics occurs microscopically.

1.7 Computer simulations of models such as molecular dynamics or cellular automata provide important tools for the study of complex systems. Monte Carlo simulations enable the study of ensemble averages without necessarily describing the dynamics of a system. However, they can also be used to study random-walk dynamics. Minimization methods that use iterative progress to find a local minimum are often an important aspect of computer simulations. Simulated annealing is a method that can help find low energy states on complex energy surfaces.

1.8 We have treated systems using models without acknowledging explicitly that our objective is to describe them. All our efforts are designed to map a system onto a description of the system. For complex systems the description must be quite long, and the study of descriptions becomes essential. With this recognition, we turn

to information theory. The information contained in a communication, typically a string of characters, may be defined quantitatively as the logarithm of the number of possible messages. When different messages have distinct probabilities P in an ensemble, then the information can be identified as $-\ln(P)$ and the average information is defined accordingly. Long messages can be modeled using the same concepts as a random walk, and we can use such models to estimate the information contained in human languages such as English.

1.9 In order to understand the relationship of information to systems, we must also understand what we can infer from information that is provided. The theory of logic is concerned with inference. It is directly linked to computation theory, which is concerned with the possible (deterministic) operations that can be performed on a string of characters. All operations on character strings can be constructed out of elementary logical (Boolean) operations on binary variables. Using Turing's model of computation, it is further shown that all computations can be performed by a universal Turing machine, as long as its input character string is suitably constructed. Computation theory is also related to our concern with the dynamics of physical systems because it explores the set of possible outcomes of discrete deterministic dynamic systems.

1.10 We return to issues of structure on microscopic and macroscopic scales by studying fractals that are self-similar geometric objects that embody the concept of progressively increasing structure on finer and finer length scales. A general approach to the scale dependence of system properties is described by scaling theory. The renormalization group methodology enables the study of scaling properties by relating a model of a system on one scale with a model of the system on another scale. Its use is illustrated by application to the Ising model (Section 1.6), and to the bifurcation route to chaos (Section 1.1). Renormalization helps us understand the basic concept of modeling systems, and formalizes the distinction between relevant and irrelevant microscopic parameters. Relevant parameters are the microscopic parameters that can affect the macroscopic behavior. The concept of universality is the notion that a whole class of microscopic models will give rise to the same macroscopic behavior, because many parameters are irrelevant. A conceptually related computational technique, the multigrid method, is based upon representing a problem on multiple scales.

The study of complex systems begins from a set of models that capture aspects of the dynamics of simple or complex systems. These models should be sufficiently general to encompass a wide range of possibilities but have sufficient structure to capture interesting features. An exciting bonus is that even the apparently simple models discussed in this chapter introduce features that are not typically treated in the conventional science of simple systems, but are appropriate introductions to the dynamics of complex systems.Our treatment of dynamics will often consider discrete rather than continuous time. Analytic treatments are often convenient to formulate in continu-

ous variables and differential equations;however, computer simulations are often best formulated in discrete space-time variables with well-defined intervals. Moreover, the assumption of a smooth continuum at small scales is not usually a convenient starting point for the study of complex systems. We are also generally interested not only in one example of a system but rather in a class of systems that differ from each other but share a characteristic structure. The elements of such a class of systems are collectively known as an ensemble. As we introduce and study mathematical models, we should recognize that our primary objective is to represent properties of real systems. We must therefore develop an understanding of the nature of models and modeling, and how they can pertain to either simple or complex systems.

1.1 Iterative Maps (and Chaos)

An iterative map f is a function that evolves the state of a system s in discrete time

$$s(t) = f(s(t - \delta t)) \tag{1.1.1}$$

where $s(t)$ describes the state of the system at time t. For convenience we will generally measure time in units of δt which then has the value 1,and time takes integral values starting from the initial condition at $t = 0$.

Many of the complex systems we will consider in this text are of the form of Eq.(1.1.1),if we allow s to be a general variable of arbitrary dimension. The generality of iterative maps is discussed at the end of this section. We start by considering several examples of iterative maps where s is a single variable. We discuss briefly the binary variable case, $s = \pm 1$. Then we discuss in greater detail two types of maps with s a real variable, s , linear maps and quadratic maps. The quadratic iterative map is a simple model that can display complex dynamics. We assume that an iterative map may be started at any initial condition allowed by a specified domain of its system variable.

1.1.1 *Binary iterative maps*

There are only a few binary iterative maps.Question 1.1.1 is a complete enumeration of them.*

Question 1.1.1 Enumerate all possible iterative maps where the system is described by a single binary variable, $s = \pm 1$.

Solution 1.1.1 There are only four possibilities:

$$\begin{aligned} s(t) &= 1 \\ s(t) &= -1 \\ s(t) &= s(t-1) \\ s(t) &= -s(t-1) \end{aligned} \tag{1.1.2}$$

*Questions are an integral part of the text. They are designed to promote independent thought. The reader is encouraged to read the question, contemplate or work out an answer and then read the solution provided in the text. The continuation of the text assumes that solutions to questions have been read.

It is instructive to consider these possibilities in some detail. The main reason there are so few possibilities is that the form of the iterative map we are using depends,at most, on the value of the system in the previous time. The first two examples are constants and don't even depend on the value of the system at the previous time. The third map can only be distinguished from the first two by observation of its behavior when presented with two different initial conditions.

The last of the four maps is the only map that has any sustained dynamics. It cycles between two values in perpetuity. We can think about this as representing an oscillator. ∎

Question 1.1.2

a. In what way can the map $s(t) = -s(t-1)$ represent a physical oscillator?

b. How can we think of the static map, $s(t) = s(t-1)$, as an oscillator?

c. Can we do the same for the constant maps $s(t) = 1$ and $s(t) = -1$?

Solution 1.1.2 (*a*) By looking at the oscillator displacement with a strobe at half-cycle intervals,our measured values can be represented by this map. (*b*) By looking at an oscillator with a strobe at cycle intervals. (*c*) You might think we could, by picking a definite starting phase of the strobe with respect to the oscillator. However, the constant map ignores the first value, the oscillator does not. ∎

1.1.2 *Linear iterative maps: free motion, oscillation, decay and growth*

The simplest example of an iterative map with s real, s , is a constant map:

$$s(t) = s_0 \tag{1.1.3}$$

No matter what the initial value,this system always takes the particular value s_0. The constant map may seem trivial,however it will be useful to compare the constant map with the next class of maps.

A linear iterative map with unit coefficient is a model of free motion or propagation in space:

$$s(t) = s(t-1) + v \tag{1.1.4}$$

at successive times the values of s are separated by v, which plays the role of the velocity.

Question 1.1.3 Consider the case of zero velocity

$$s(t) = s(t-1) \tag{1.1.5}$$

How is this different from the constant map?

Solution 1.1.3 The two maps differ in their dependence on the initial value. ∎

Runaway growth or decay is a multiplicative iterative map:

$$s(t) = gs(t-1) \tag{1.1.6}$$

We can generate the values of this iterative map at all times by using the equivalent expression

$$s(t) = g^t s_0 = e^{\ln(g)t} s_0 \tag{1.1.7}$$

which is exponential growth or decay. The iterative map can be thought of as a sequence of snapshots of Eq.(1.1.7) at integral time. $g = 1$ reduces this map to the previous case.

Question 1.1.4 We have seen the case of free motion, and now jumped to the case of growth. What happened to accelerated motion? Usually we would consider accelerated motion as the next step after motion with a constant velocity. How can we write accelerated motion as an iterative map?

Solution 1.1.4 The description of accelerated motion requires two variables: position and velocity. The iterative map would look like:

$$\begin{aligned} x(t) &= x(t-1) + v(t-1) \\ v(t) &= v(t-1) + a \end{aligned} \tag{1.1.8}$$

This is a two-variable iterative map. To write this in the notation of Eq.(1.1.1) we would define s as a vector $s(t) = (x(t), v(t))$. ∎

Question 1.1.5 What happens in the rightmost exponential expression in Eq. (1.1.7) when g is negative?

Solution 1.1.5 The logarithm of a negative number results in a phase $i\pi$. The term $i\pi t$ in the exponent alternates sign every time step as one would expect from Eq. (1.1.6). ∎

At this point,it is convenient to introduce two graphical methods for describing an iterative map. The first is the usual way of plotting the value of s as a function of time. This is shown in the left panels of Fig. 1.1.1. The second type of plot,shown in the right panels, has a different purpose. This is a plot of the iterative relation $s(t)$ as a function of $s(t-1)$. On the same axis we also draw the line for the identity map $s(t) = s(t-1)$. These two plots enable us to graphically obtain the successive values of s as follows. Pick a starting value of s, which we can call $s(0)$. Mark this value on the abscissa. Mark the point on the graph of $s(t)$ that corresponds to the point whose abscissa is $s(0)$,i.e.,the point $(s(0), s(1))$.Draw a horizontal line to intersect the identity map. The intersection point is $(s(1), s(1))$. Draw a vertical line back to the iterative map. This is the point $(s(1), s(2))$. Successive values of $s(t)$ are obtained by iterating this graphical procedure. A few examples are plotted in the right panels of Fig. 1.1.1.

In order to discuss the iterative maps it is helpful to recognize several features of these maps.First,intersection points of the identity map and the iterative map are the fixed points of the iterative map:

$$s_0 = f(s_0) \tag{1.1.9}$$

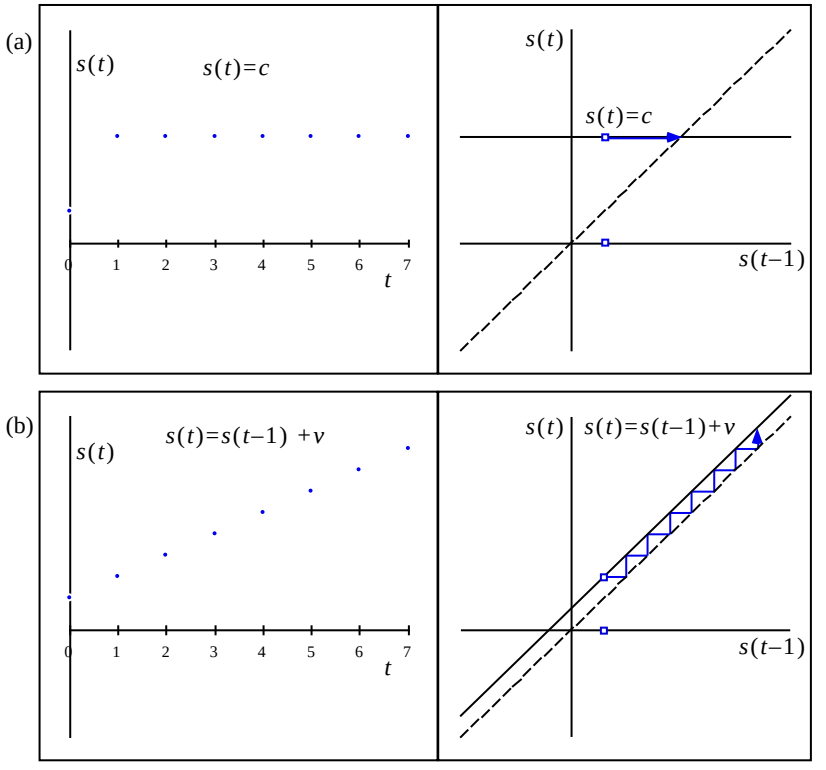
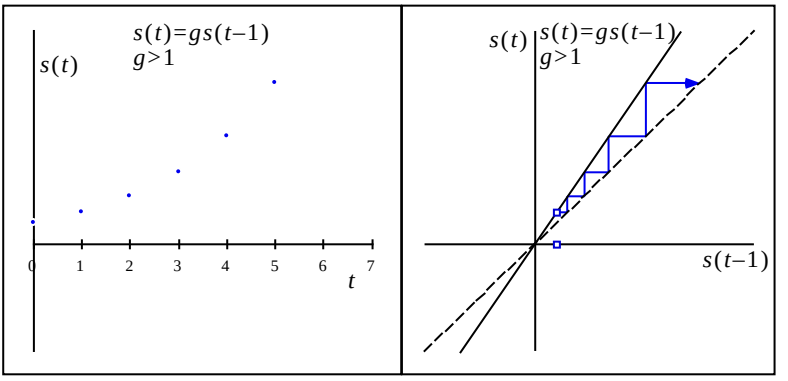


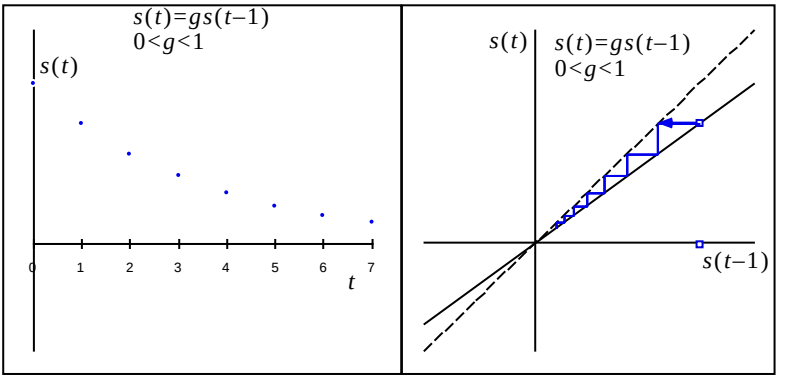
Figure 1.1.1 The left panels show the time-dependent value of the system variable $s(t)$ resulting from iterative maps. The first panel (a) shows the result of iterating the constant map; (b) shows the result of adding v to the previous value during each time interval; (c)–(f) show the result of multiplying by a constant g, where each figure shows the behavior for a different range of g values: (c) $g > 1$, (d) $0 < g < 1$, (e) $-1 < g < 0$, and (f) $g < -1$. The right panels are a different way of showing graphically the results of iterations and are constructed as follows. First plot the function $f(s)$ (solid line), where $s(t) = f(s(t-1))$. This can be thought of as plotting $s(t)$ vs. $s(t-1)$. Second, plot the identity map $s(t) = s(t-1)$ (dashed line). Mark the initial value $s(0)$ on the horizontal axis, and the point on the graph of $s(t)$ that corresponds to the point whose abscissa is $s(0)$, i.e. the point $(s(0), s(1))$. These are shown as squares. From the point $(s(0), s(1))$ draw a horizontal line to intersect the identity map. The intersection point is $(s(1), s(1))$. Draw a vertical line back to the iterative map. This is the point $(s(1), s(2))$. Successive values of $s(t)$ are obtained by iterating this graphical procedure. ∎

Fixed points,not surprisingly, play an important role in iterative maps. They help us describe the state and behavior of the system after many iterations. There are two kinds of fixed points—stable and unstable. Stable fixed points are characterized by "attracting" the result of iteration of points that are nearby. More precisely, there exists

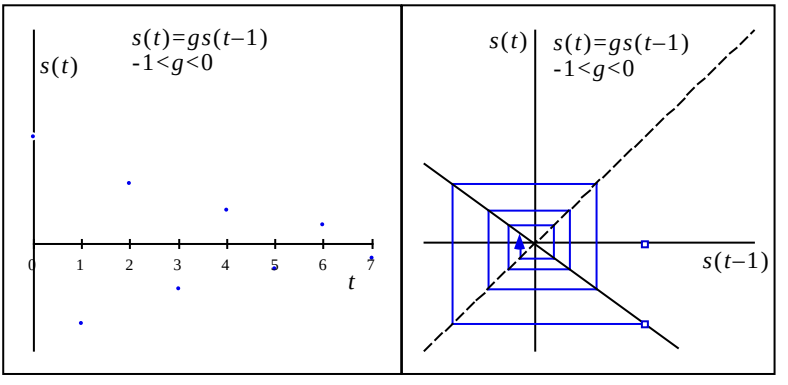
(c)


s(t)=gs(t−1)
g>1
s(t)
0 1 2 3 4 5 6 7
t
s(t)
s(t)=gs(t−1)
g>1
s(t−1)

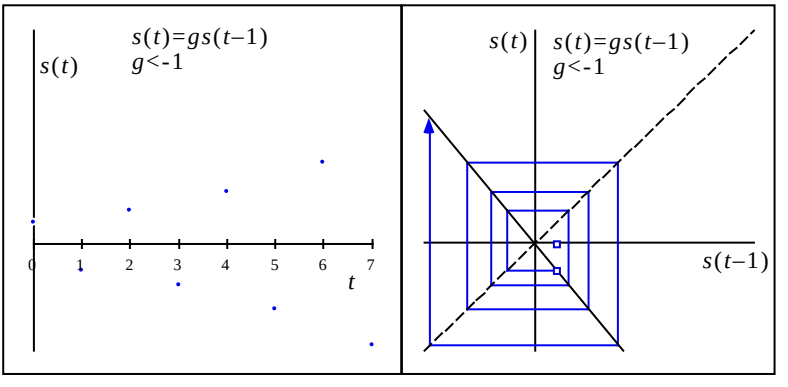
(d)


s(t)=gs(t−1)
0<g<1
s(t)
0 1 2 3 4 5 6 7
t
s(t)
s(t)=gs(t−1)
0<g<1
s(t−1)

(e)


s(t)=gs(t−1)
-1<g<0
s(t)
0 1 2 3 4 5 6 7
t
s(t)
s(t)=gs(t−1)
-1<g<0
s(t−1)

(f)


s(t)=gs(t−1)
g<-1
s(t)
0 1 2 3 4 5 6 7
t
s(t)
s(t)=gs(t−1)
g<-1
s(t−1)

a neighborhood of points of s_0 such that for any s in this neighborhood the sequence of points

$$\{s, f(s), f^2(s), f^3(s), \quad \} \tag{1.1.10}$$

converges to s_0. We are using the notation $f^2(s) = f(f(s))$ for the second iteration,and similar notation for higher iterations. This sequence is just the time series of the iterative map for the initial condition s. Unstable fixed points have the opposite behavior, in that iteration causes the system to leave the neighborhood of s_0. The two types of fixed points are also called attracting and repelling fixed points.

The family of multiplicative iterative maps in Eq.(1.1.6) all have a fixed point at $s_0 = 0$. Graphically from the figures, or analytically from Eq. (1.1.7), we see that the fixed point is stable for $|g| < 1$ and is unstable for $|g| > 1$. There is also distinct behavior of the system depending on whether g is positive or negative. For $g < 0$ the iterations alternate from one side to the other of the fixed point, whether it is attracted to or repelled from the fixed point. Specifically, if $s < s_0$ then $f(s) > s_0$ and vice versa, or $\text{sign}(s - s_0) = -\text{sign}(f(s) - s_0)$. For $g > 0$ the iteration does not alternate.

Question 1.1.6 Consider the iterative map.

$$s(t) = gs(t-1) + v \tag{1.1.11}$$

convince yourself that v does not affect the nature of the fixed point, only shifts its position.

Question 1.1.7 Consider an arbitrary iterative map of the form Eq.(1.1.1), with a fixed point s_0 (Eq.(1.1.9)). If the iterative map can be expanded in a Taylor series around s_0 show that the first derivative

$$g = \left.\frac{df(s)}{ds}\right|_{s_0} \tag{1.1.12}$$

characterizes the fixed point as follows:

For $|g| < 1$, s_0 is an attracting fixed point.

For $|g| > 1$, s_0 is a repelling fixed point.

For $g < 0$, iterations alternate sides in a sufficiently small neighborhood of s_0.

For $g > 0$,iterations remain on one side in a sufficiently small neighborhood of s_0.

Extra credit: Prove the same theorem for a differentiable function (no Taylor expansion needed) using the mean value theorem.

Solution 1.1.7 If the iterative map can be expanded in a Taylor series we write that

$$f(s) = f(s_0) + g(s - s_0) + h(s - s_0)^2 + \tag{1.1.13}$$

where g is the first derivative at s_0, and h is one-half of the second derivative at s_0. Since s_0 is a fixed point $f(s_0) = s_0$ we can rewrite this as:

$$\frac{f(s) - s_0}{s - s_0} = g + h(s - s_0) + \qquad (1.1.14)$$

If we did not have any higher-order terms beyond g, then by inspection each of the four conditions that we have to prove would follow from this expression without restrictions on s. For example, if $|g| > 1$, then taking the magnitude of both sides shows that $f(s) - s_0$ is larger than $s - s_0$ and the iterations take the point s away from s_0. If $g > 0$,then this expression says that $f(s)$ stays on the same side of s_0. The other conditions follow similarly.

To generalize this argument to include the higher-order terms of the expansion, we must guarantee that whichever domain g is in ($g > 1$, $0 < g < 1$, $-1 < g < 0$, or $g < -1$), the same is also true of the whole right side. For a Taylor expansion, by choosing a small enough neighborhood $|s - s_0| < \delta$, we can guarantee the higher-order terms are less than any number ε we choose. We choose ε to be half of the minimum of $|g - 1|$, $|g - 0|$ and $|g + 1|$. Then $g + \varepsilon$ is in the same domain as g. This provides the desired guarantee and the proof is complete.

We have proven that in the vicinity of a fixed point the iterative map may be completely characterized by its first-order expansion (with the exception of the special points $g = \pm 1, 0$). ∎

Thus far we have not considered the special cases $g = \pm 1, 0$. The special cases $g = 0$ and $g = 1$ have already been treated as simpler iterative maps. When $g = 0$, the fixed point at $s = 0$ is so attractive that it is the result of any iteration. When $g = 1$ all points are fixed points.

The new special case $g = -1$ has a different significance. In this case all points alternate between positive and negative values, repeating every other iteration. Such repetition is a generalization of the fixed point. Whereas in the fixed-point case we repeat every iteration, here we repeat after every two iterations. This is called a 2-cycle, and we can immediately consider the more general case of an n-cycle. In this terminology a fixed point is a 1-cycle.One way to describe an n-cycle is to say that iterating n times gives back the same result, or equivalently, that a new iterative map which is the nth fold composition of the original map $h = f^n$ has a fixed point. This description would include also fixed points of f and all points that are m-cycles, where m is a divisor of n. These are excluded from the definition of the n-cycles. While we have introduced cycles using a map where all points are 2-cycles,more general iterative maps have specific sets of points that are n-cycles. The set of points of an n-cycle is called an orbit. There are a variety of properties of fixed points and cycles that can be proven for an arbitrary map. One of these is discussed in Question 1.1.8.

Question 1.1.8 Prove that there is a fixed point between any two points of a 2-cycle if the iterating function f is continuous.

Solution 1.1.8 Let the 2-cycle be written as

$$\begin{aligned} s_2 &= f(s_1) \\ s_1 &= f(s_2) \end{aligned} \tag{1.1.15}$$

Consider the function $h(s) = f(s) - s$, $h(s_1)$ and $h(s_2)$ have opposite signs and therefore there must be an s_0 between s_1 and s_2 such that $h(s_0) = 0$—the fixed point. ∎

We can also generalize the definition of attracting and repelling fixed points to consider attracting and repelling n-cycles. Attraction and repulsion for the cycle is equivalent to the attraction and repulsion of the fixed point of f^n.

1.1.3 *Quadratic iterative maps: cycles and chaos*

The next iterative map we will consider describes the effect of nonlinearity (self-action):

$$s(t) = as(t-1)(1 - s(t-1)) \tag{1.1.16}$$

or equivalently

$$f(s) = as(1 - s) \tag{1.1.17}$$

This map has played a significant role in development of the theory of dynamical systems because even though it looks quite innocent,it has a dynamical behavior that is not described in the conventional science of simple systems. Instead, Eq. (1.1.16) is the basis of significant work on chaotic behavior, and the transition of behavior from simple to chaotic. We have chosen this form of quadratic map because it simplifies somewhat the discussion. Question 1.1.11 describes the relationship between this family of quadratic maps,parameterized by a, and what might otherwise appear to be a different family of quadratic maps.

We will focus on a values in the range $4 > a > 0$. For this range, any value of s in the interval $s \in [0,1]$ stays within this interval. The minimum value $f(s) = 0$ occurs for $s = 0,1$ and the maximal value occurs for $s = 1/2$. For all values of a there is a fixed point at $s = 0$ and there can be at most two fixed points, since a quadratic can only intersect a line (Eq. (1.1.9)) in two points.

Taking the first derivative of the iterative map gives

$$\frac{df}{ds} = a(1-2s) \tag{1.1.18}$$

At $s = 0$ the derivative is a which, by Question 1.1.7,shows that $s = 0$ is a stable fixed point for $a < 1$ and an unstable fixed point for $a > 1$. The switching of the stability of the fixed point at $s = 0$ coincides with the introduction of a second fixed point in the interval [0,1] (when the slope at $s = 0$ is greater than one, $f(s) > s$ for small s, and since

$f(1) = 0$, we have that $f(s_1) = s_1$ for some s_1 in [0,1] by the same construction as in Question 1.1.8). We find s_1 by solving the equation

$$s_1 = as_1(1-s_1) \tag{1.1.19}$$

$$s_1 = (a-1)/a \tag{1.1.20}$$

Substituting this into Eq. (1.1.18) gives

$$\left.\frac{df}{ds}\right|_{s_1} = 2-a \tag{1.1.21}$$

This shows that for $1 < a < 3$,the new fixed point is stable by Question 1.1.7. Moreover, the derivative is positive for $1 < a < 2$,so s_1 is stable and convergence is from one side. The derivative is negative for $2 < a < 3$, so s_1 is stable and alternating.

Fig. 1.1.2(a)–(c) shows the three cases: $a = 0.5$, $a = 1.5$ and $a = 2.8$. For $a = 0.5$, starting from anywhere within [0,1] leads to convergence to $s = 0$. When $s(0) > 0.5$ the first iteration takes the system to $s(1) < 0.5$. The closer we start to $s(0) = 1$ the closer to $s = 0$ we get in the first jump. At $s(0) = 1$ the convergence to 0 occurs in the first jump. A similar behavior would be found for any value of $0 < a < 1$. For $a = 1.5$ the behavior is more complicated. Except for the points $s = 0,1$,the convergence is always to the fixed point $s_1 = (a - 1)/a$ between 0 and 1. For $a = 2.8$ the iterations converge to the same point;however, the convergence is alternating. Because there can be at most two fixed points for the quadratic map, one might think that this behavior would be all that would happen for $1 < a < 4$.One would be wrong. The first indication that this is not the case is the instability of the fixed point at s_1 starting from $a = 3$.

What happens for $a > 3$? Both of the fixed points that we have found,and the only ones that can exist for the quadratic map, are now unstable. We know that the iteration of the map has to go somewhere, and only within [0,1]. The only possibility, within our experience, is that there is an attracting n-cycle to which the fixed points are unstable. Let us then consider the map $f^2(s)$ whose fixed points are 2-cycles of the original map. $f^2(s)$ is shown in the right panels of Fig. 1.1.2 for increasing values of a. The fixed points of $f(s)$ are also fixed points of $f^2(s)$. However, we see that two additional fixed points exist for $a > 3$. We can also show analytically that two fixed points are introduced at exactly $a = 3$:

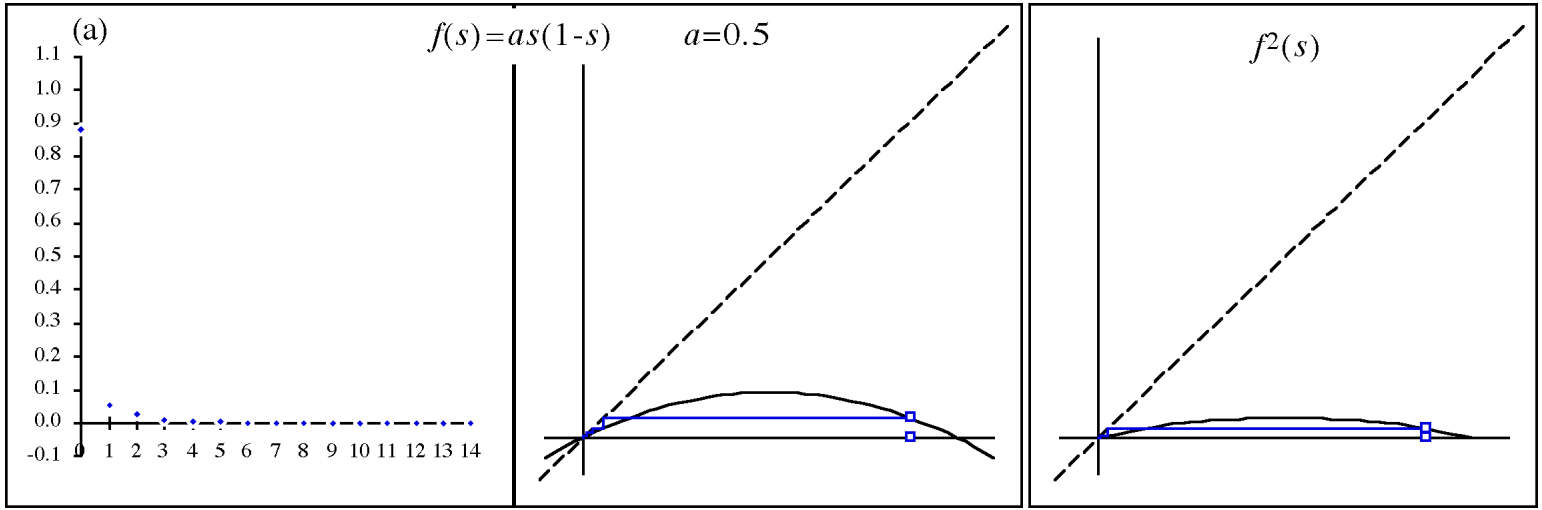
$$f^2(s) = a^2s(1-s)(1-as(1-s)) \tag{1.1.22}$$

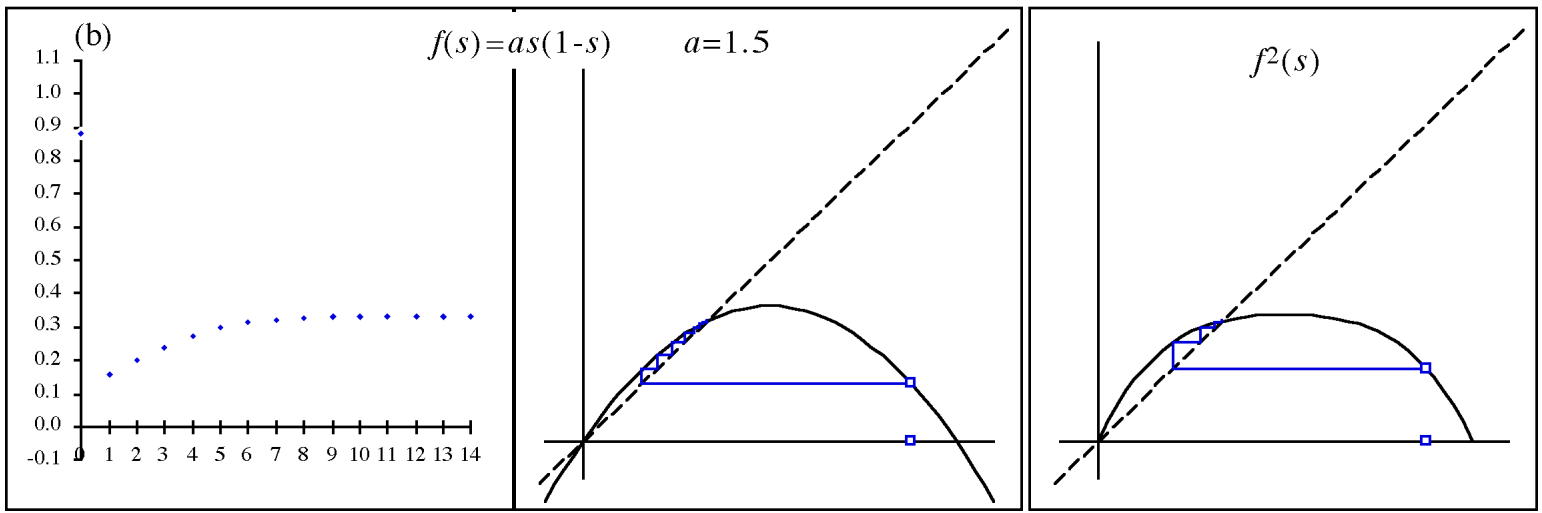
To find the fixed point we solve:

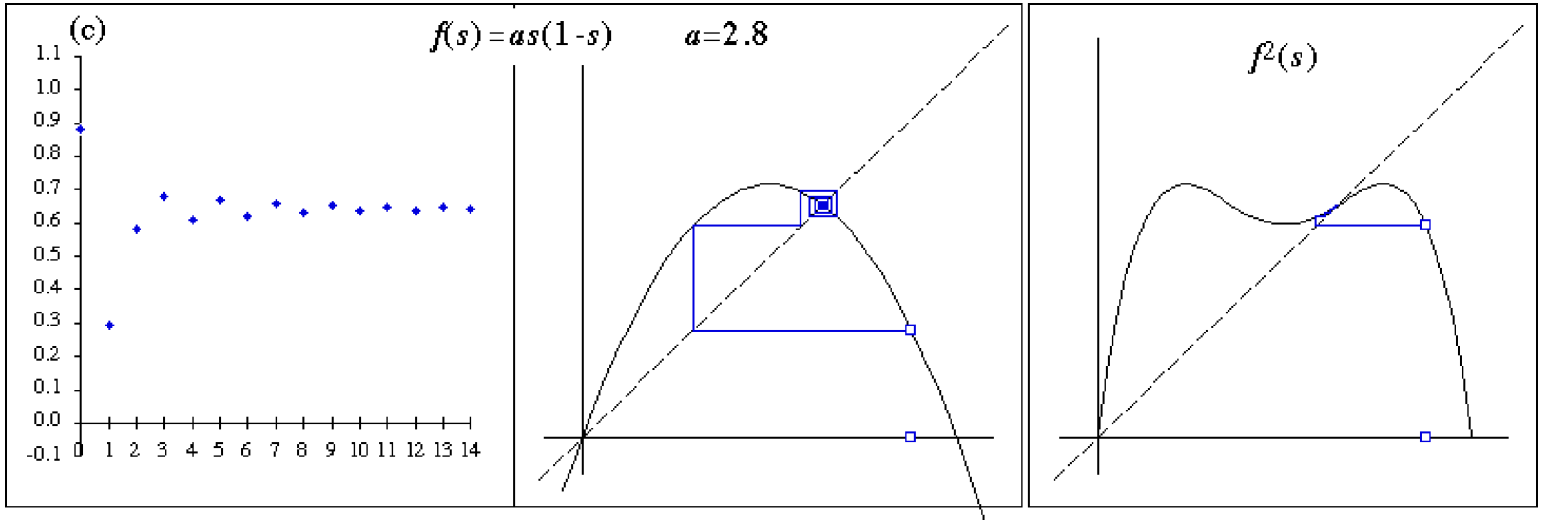
$$s = a^2s(1-s)(1-as(1-s)) \tag{1.1.23}$$

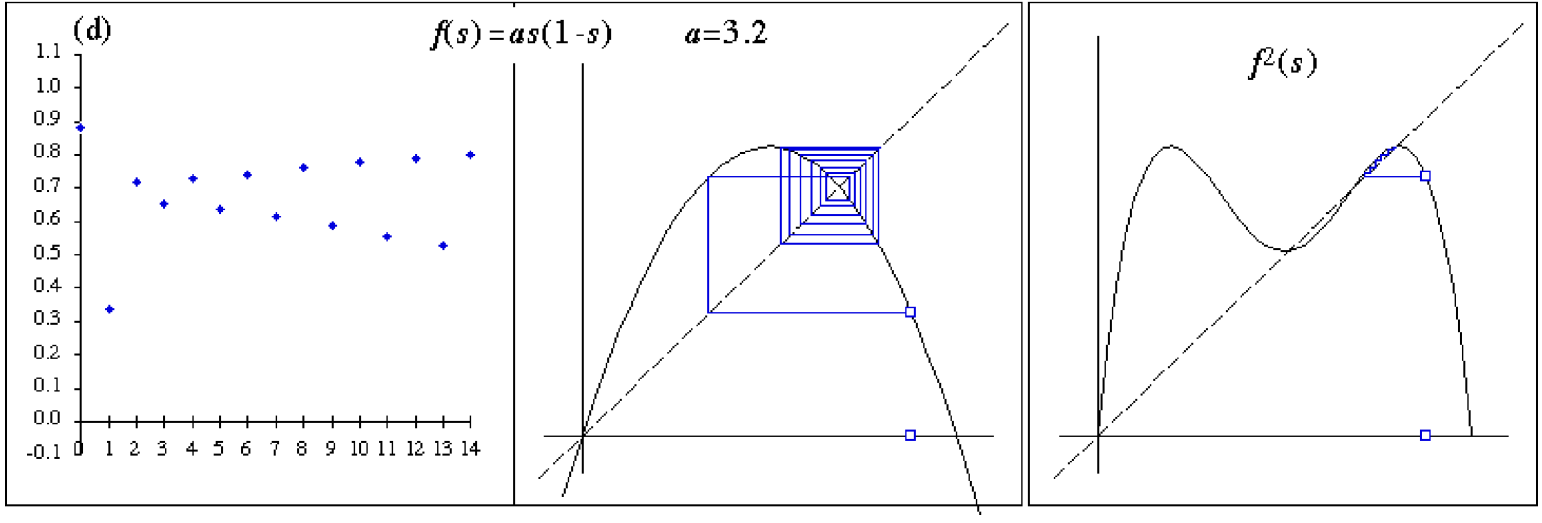
We already know two solutions of this quartic equation—the fixed points of the map f. One of these at $s = 0$ is obvious. Dividing by s we have a cubic equation:

$$a^3s^3 - 2a^3s^2 + a^2(1+a)s + (1-a^2) = 0 \tag{1.1.24}$$


(a)
$f(s)=as(1-s)$
$a=0.5$
$f^2(s)$
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14


(b)
$f(s)=as(1-s)$
$a=1.5$
$f^2(s)$
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14


(c)
$f(s) = as(1-s)$
a=2.8
$f^2(s)$
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14


(d)
$f(s) = as(1-s)$
a=3.2
$f^2(s)$
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14

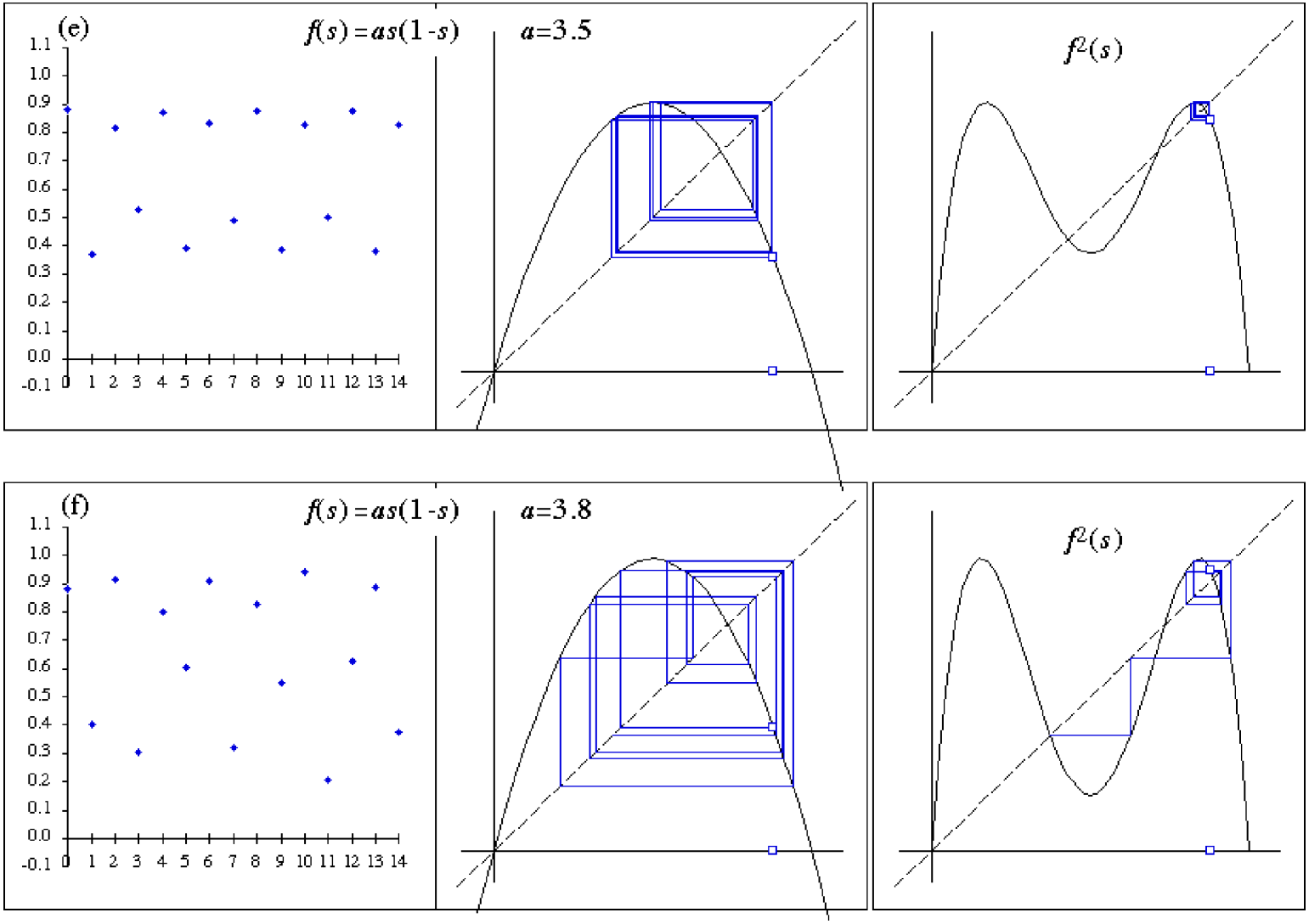

(e)
$f(s) = as(1-s)$
a=3.5
$f^2(s)$
(f)
$f(s) = as(1-s)$
a=3.8
$f^2(s)$

Figure 1.1.2 (pp. 28-30) Plots of the result of iterating the quadratic map $f(s) = as(1 - s)$ for different values of a. The left and center panels are similar to the left and right panels of Fig. 1.1.1. The left panels plot $s(t)$. The center panels describe the iteration of the map $f(s)$ on axes corresponding to $s(t)$ and $s(t - 1)$. The right panels are similar to the center panels but are for the function $f^2(s)$. The different values of a are indicated on the panels and show the changes from (a) convergence to $s = 0$ for $a = 0.5$, (b) convergence to $s = (a - 1)/a$ for $a = 1.5$, (c) alternating convergence to $s = (a - 1)/a$ for $a = 2.8$, (d) bifurcation — convergence to a 2-cycle for $a = 3.2$, (e) second bifurcation — convergence to a 4-cycle for $a = 3.5$, (f) chaotic behavior for $a = 3.8$. ∎

We can reduce the equation to a quadratic by dividing by $(s - s_1)$ as follows (we simplify the algebra by dividing by $a(s - s_1) = (as - (a - 1))$):

$$
\begin{array}{r}
a^2s^2 - a(a+1)s + (a+1) \\
\hline
(as - (a-1)) \;\big)\; a^3s^3 - 2a^3s^2 + a^2(1+a)s + (1-a^2) \\
\underline{a^3s^3 - (a-1)a^2s^2} \\
-(a+1)a^2s^2 + a^2(1+a)s + (1-a^2) \\
\underline{-(a+1)a^2s^2 + a(1+a)(a-1)s} \\
+a(1+a)s + (1-a^2)
\end{array}
\tag{1.1.25}
$$

Now we can obtain the roots to the quadratic:

$$a^2s^2 - a(a+1)s + (a+1) = 0 \tag{1.1.26}$$

$$s_2 = \frac{(a+1) \pm \sqrt{(a+1)(a-3)}}{2a} \tag{1.1.27}$$

This has two solutions (as it must for a 2-cycle) for $a < -1$ or for $a > 3$. The former case is not of interest to us since we have assumed $0 < a < 4$. The latter case is the two roots that are promised. Notice that for exactly $a = 3$ the two roots that are the new 2-cycle are the same as the fixed point we have already found s_1. The 2-cycle splits off from the fixed point at $a = 3$ when the fixed point becomes unstable. The two attracting points continue to separate as a increases. For $a > 3$ we expect that the result of iteration eventually settles down to the 2-cycle. The system state alternates between the two roots Eq. (1.1.27). This is shown in Fig. 1.1.2(d).

As we continue to increase a beyond 3, the 2-cycle will itself become unstable at a point that can be calculated by setting

$$\left.\frac{df^2}{ds}\right|_{s_2} = -1 \tag{1.1.28}$$

to be $a = 1 + \sqrt{6} = 3.44949$. At this value of a the 2-cycle splits into a 4-cycle (Fig. 1.1.2(e)).Each of the fixed points of $f^2(s)$ simultaneously split into 2-cycles that together form a 4-cycle for the original map.

Question 1.1.9 Show that when f has a 2-cycle, both of the fixed points of f^2 must split simultaneously.

Solution 1.1.9 The split occurs when the fixed points become unstable—the derivative of f^2 equals –1. We can show that the derivative is equal at the two fixed points of Eq. (1.1.27), which we call $s_2^{\pm}$:

$$\left.\frac{df^2}{ds}\right|_{s_2} = \left.\frac{df(f(s))}{ds}\right|_{s_2} = \left.\frac{df(s)}{ds}\right|_{f(s_2)} \left.\frac{df(s)}{ds}\right|_{s_2} \tag{1.1.29}$$

where we have made use of the chain rule. Since $f(s_2^+) = s_2^-$ and vice versa, we have shown this expression is the same whether $s_2 = s_2^+$ or $s_2 = s_2^-$.

Note: This can be generalized to show that the derivative of f^k is the same at all of its k fixed points corresponding to a k-cycle of f. ∎

The process of taking an n-cycle into a $2n$-cycle is called bifurcation. Bifurcation continues to replace the limiting behavior of the iterative map with progressively longer cycles of length 2^k. The bifurcations can be simulated. They occur at smaller and smaller intervals and there is a limit point to the bifurcations at $a_c = 3.56994567$. Fig. 1.1.3 shows the values that are reached by the iterative map at long times—the stable cycles—as a function of $a < a_c$. We will discuss an algebraic treatment of the bifurcation regime in Section 1.10.

Beyond the bifurcation regime $a > a_c$ (Fig. 1.1.2(f)) the behavior of the iterative map can no longer be described using simple cycles that attract the iterations. The behavior in this regime has been identified with chaos. Chaos has been characterized in many ways, but one property is quite generally agreed upon—the inherent lack of predictability of the system dynamics. This is often expressed more precisely by describing the sensitivity of the system's fate to the initial conditions. A possible definition is: There exists a distance d such that for any neighborhood V of any point s it is possible to find a point s' within the neighborhood and a number of iterations k so that $f^k(s')$ is further than d away from $f^k(s)$. This means that arbitrarily close to any point is a point that will be displaced a significant distance away by iteration. Qualitatively, there are two missing aspects of this definition,first that the points that move far away must not be too unlikely (otherwise the system is essentially predictable) and second that d is not too small (in which case the divergence of the dynamics may not be significant).

If we look at the definition of chaotic behavior, we see that the concept of scale plays an important role. A small distance between s and s' turns into a large distance between $f^k(s)$ and $f^k(s')$. Thus a fine-scale difference eventually becomes a large-scale difference. This is the essence of chaos as a model of complex system behavior. To understand it more fully, we can think about the state variable s not as one real variable,

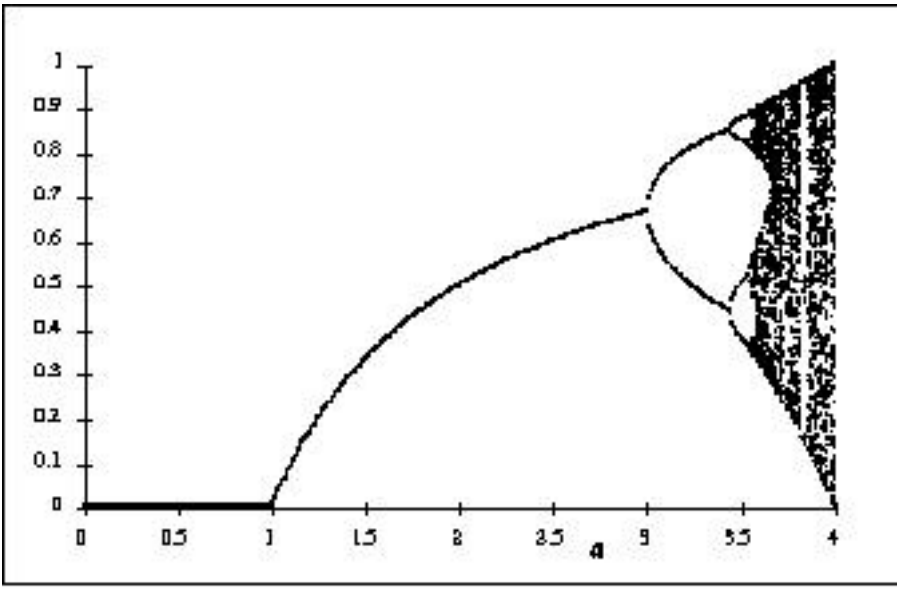


Figure 1.1.3 A plot of values of s visited by the quadratic map $f(s) = as(1 - s)$ after many iterations as a function of a, including stable points, cycles and chaotic behavior. The different regimes are readily apparent. For $a < 1$ the stable point is $s = 0$. For $1 < a < 3$ the stable point is at $s_0 = (a - 1)/a$. For $3 < a < a_c$ with $a_c = 3.56994567$, there is a bifurcation cascade with 2-cycles then 4-cycles, etc. 2^k-cycles for all values of k appear in progressively narrower regions of a. Beyond 4-cycles they cannot be seen in this plot. For $a > a_c$ there is chaotic behavior. There are regions of s values that are not visited and regions that are visited in the long time behavior of the quadratic map in the chaotic regime which this figure does not fully illustrate. ∎

but as an infinite sequence of binary variables that form its binary representation $s = 0.r_1r_2r_3r_4\ldots$ Each of these binary variables represents the state of the system—the value of some quantity we can measure about the system—on a particular length scale. The higher order bits represent the larger scales and the lower order ones represent the finer scales. Chaotic behavior implies that the state of the first few binary variables, r_1r_2, at a particular time are determined by the value of fine scale variables at an earlier time. The farther back in time we look, the finer scale variables we have to consider in order to know the present values of r_1r_2. Because many different variables are relevant to the behavior of the system, we say that the system has a complex behavior. We will return to these issues in Chapter 8.

The influence of fine length scales on coarse ones makes iterative maps difficult to simulate by computer. Computer representations of real numbers always have finite precision. This must be taken into account if simulations of iterative maps or chaotic complex systems are performed.

Another significant point about the iterative map as a model of a complex system is that there is nothing outside of the system that is influencing it. All of the information we need to describe the behavior is contained in the precise value of s. The complex behavior arises from the way the different parts of the system—the fine and course scales—affect each other.

Question 1.1.10: Why isn't the iterative map in the chaotic regime equivalent to picking a number at random?

Solution 1.1.10: We can still predict the behavior of the iterative map over a few iterations. It is only when we iterate long enough that the map becomes unpredictable. More specifically, the continuity of the function $f(s)$ guarantees that for s and s' close together $f(s)$ and $f(s')$ will also be close together. Specifically, given an ε it is possible to find a δ such that for $|s - s'| < \delta$, $|f(s) - f(s')| < \varepsilon$. For the family of functions we have been considering, we only need to set $\delta < \varepsilon/a$, since then we have:

$$|f(s) - f(s')| = a|s(1-s) - s'(1-s')| = a|s - s'||1 - (s + s')| < a|s - s'| < \varepsilon \tag{1.1.30}$$

Thus if we fix the number of cycles to be k, we can always find two points close enough so that $|f^k(s') - f^k(s)| < \varepsilon$ by setting $|s - s'| < \varepsilon/a^k$. ∎

The tuning of the parameter a leading from simple convergent behavior through cycle bifurcation to chaos has been identified as a universal description of the appearance of chaotic behavior from simple behavior of many systems. How do we take a complicated real system and map it onto a discrete time iterative map? We must define a system variable and then take snapshots of it at fixed intervals (or at least well-defined intervals). The snapshots correspond to an iterative map. Often there is a natural choice for the interval that simplifies the iterative behavior. We can then check to see if there is bifurcation and chaos in the real system when parameters that control the system behavior are varied.

One of the earliest examples of the application of iterative maps is to the study of heart attacks. Heart attacks occur in many different ways. One kind of heart attack is known as fibrillation. Fibrillation is characterized by chaotic and ineffective heart muscle contractions. It has been suggested that bifurcation may be observed in heartbeats as a period doubling (two heartbeats that are inequivalent). If correct, this may serve as a warning that the heart structure, due to various changes in heart tissue parameters, may be approaching fibrillation. Another system where more detailed studies have suggested that bifurcation occurs as a route to chaotic behavior is that of turbulent flows in hydrodynamic systems. A subtlety in the application of the ideas of bifurcation and chaos to physical systems is that physical systems are better modeled as having an increasing number of degrees of freedom at finer scales. This is to be contrasted with a system modeled by a single real number, which has the same number of degrees of freedom (represented by the binary variables above) at each length scale.

1.1.4 *Are all dynamical systems iterative maps?*

How general is the iterative map as a tool for describing the dynamics of systems? There are three apparent limitations of iterative maps that we will consider modifying later, Eq. (1.1.1):

a. describes the homogeneous evolution of a system since *f* itself does not depend on time,

b. describes a system where the state of the system at time *t* depends only on the state of the system at time $t - \delta t$, and

c. describes a deterministic evolution of a system.

We can, however, bypass these limitations and keep the same form of the iterative map if we are willing to let *s* describe not just the present state of the system but also

a. the state of the system and all other factors that might affect its evolution in time,

b. the state of the system at the present time and sufficiently many previous times, and

c. the probability that the system is in a particular state.

Taking these caveats together, all of the systems we will consider are iterative maps, which therefore appear to be quite general. Generality, however, can be quite useless, since we want to discard as much information as possible when describing a system.

Another way to argue the generality of the iterative map is through the laws of classical or quantum dynamics. If we consider *s* to be a variable that describes the positions and velocities of all particles in a system, all closed systems described by classical mechanics can be described as deterministic iterative maps. Quantum evolution of a closed system may also be described by an iterative map if *s* describes the wave function of the system. However, our intent is not necessarily to describe microscopic dynamics, but rather the dynamics of variables that we consider to be relevant in describing a system. In this case we are not always guaranteed that a deterministic iterative map is sufficient. We will discuss relevant generalizations, first to stochastic maps, in Section 1.2.

Extra Credit Question 1.1.11 Show that the system of quadratic iterative maps

$$s(t) = s(t-1)^2 + k \tag{1.1.31}$$

is essentially equivalent in its dynamical properties to the iterative maps we have considered in Eq. (1.1.16).

Solution 1.1.11 Two iterative maps are equivalent in their properties if we can perform a time-independent one-to-one map of the time-dependent system states from one case to the other. We will attempt to transform the family of quadratic maps given in this problem to the one of Eq. (1.1.16) using a linear map valid at all times

$$s(t) = ms'(t) + b \tag{1.1.32}$$

By direct substitution this leads to:

$$ms'(t)+b=(ms'(t-1)+b)^2+k \tag{1.1.33}$$

We must now choose the values of m and b so as to obtain the form of Eq. (1.1.16).

$$s'(t)=ms'(t-1)(s'(t-1)+\frac{2b}{m})+\frac{1}{m}(b^2+k-b) \tag{1.1.34}$$

For a correct placement of minus signs in the parenthesis we need:

$$s'(t)=(-m)s'(t-1)(\left(-\frac{2b}{m}\right)-s'(t-1))+\frac{1}{m}(b^2+k-b) \tag{1.1.35}$$

or

$$b^2-b+k=0 \tag{1.1.36}$$

$$\frac{2b}{m}=-1 \tag{1.1.37}$$

giving

$$b=(1\pm\sqrt{1-4k})/2 \tag{1.1.38}$$

$$a=-m=2b=(1\pm\sqrt{1-4k}) \tag{1.1.39}$$

We see that for $k < 1/4$ we have two solutions. These solutions give all possible (positive and negative) values of a.

What about $k > 1/4$? It turns out that this case is not very interesting compared to the rich behavior for $k < 1/4$, since there are no finite fixed points,and therefore by Question 1.1.8 no 2-cycles (it is not hard to generalize this to n-cycles). To confirm this, verify that iterations diverge to $+\infty$ from any initial condition.

Note: The system of equations of this question are the ones extensively analyzed by Devaney in his excellent textbook *A First Course in Chaotic Dynamical Systems.* ∎

Extra Credit Question 1.1.12 You are given a problem to solve which when reduced to mathematical form looks like

$$s=f_c(s) \tag{1.1.40}$$

where f is a complicated function that depends on a parameter c. You know that there is a solution of this equation in the vicinity of s_0. To solve this equation you try to iterate it (Newton's method) and it works,since you find that $f^k(s_0)$ converges nicely to a solution. Now, however, you realize that you need to solve this problem for a slightly different value of the parameter c, and when you try to iterate the equation you can't get the value of s to converge. Instead the values start to oscillate and then behave in a completely erratic

way. Suggest a solution for this problem and see if it works for the function $f_c(s) = cs(1 - s)$, $c = 3.8$, $s_0 = 0.5$. A solution is given in stages (a) - (c) below.

Solution 1.1.12(a) A common resolution of this problem is to consider iterating the function:

$$h_c(s) = \alpha s + (1-\alpha) f_c(s) \tag{1.1.41}$$

where we can adjust α to obtain rapid convergence. Note that solutions of

$$s = h_c(s) \tag{1.1.42}$$

are the same as solutions of the original problem.

Question 1.1.12(b) Explain why this could work.

Solution 1.1.12(b) The derivative of this function at a fixed point can be controlled by the value of α. It is a linear interpolation between the fixed point derivative of f_c and 1. If the fixed point is unstable and oscillating, the derivative of f_c must be less than -1 and the interpolation should help.

We can also explain this result without appealing to our work on iterative maps by noting that if the iteration is causing us to overshoot the mark, it makes sense to mix the value s we start from with the value we get from $f_c(s)$ to get a better estimate.

Question 1.1.12(c) Explain how to pick α.

Solution 1.1.12(c) If the solution is oscillating, then it makes sense to assume that the fixed point is in between successive values and the distance is revealed by how much further it gets each time;i.e., we assume that the iteration is essentially a linear map near the fixed point and we adjust α so that we compensate exactly for the overshoot of f_c.

Using two trial iterations, a linear approximation to f_c at s_0 looks like:

$$\begin{aligned} s_2 &= f_c(s_1) \approx g(s_1 - s_0) + s_0 \\ s_3 &= f_c(s_2) \approx g(s_2 - s_0) + s_0 \end{aligned} \tag{1.1.43}$$

Adopting the linear approximation as a definition of g we have:

$$g \equiv (s_3 - s_2)/(s_2 - s_1) \tag{1.1.44}$$

Set up α so that the first iteration of the modified system will take you to the desired answer:

$$s_0 = \alpha s_1 + (1-\alpha) f_c(s_1) \tag{1.1.45}$$

or

$$s_0 - s_1 = (1-\alpha)(f_c(s_1) - s_1) = (1-\alpha)(s_2 - s_1) \tag{1.1.46}$$

$$(1-\alpha) = (s_0 - s_1)/(s_2 - s_1) \tag{1.1.47}$$

To eliminate the unknown s_0 we use Eq. (1.1.43) to obtain:

$$(s_2 - s_1) = g(s_1 - s_0) + (s_0 - s_1) \tag{1.1.48}$$

$$(s_0 - s_1) = (s_2 - s_1)/(1 - g) \tag{1.1.49}$$

or

$$1 - \alpha = 1/(1 - g) \tag{1.1.50}$$

$$\alpha = -g/(1 - g) = (s_2 - s_3)/(2s_2 - s_1 - s_3) \tag{1.1.51}$$

It is easy to check, using the formula in terms of g, that the modified iteration has a zero derivative at s_0 when we use the approximate linear forms for f_c. This means we have the best convergence possible using the information from two iterations of f_c. We then use the value of α to iterate to convergence. Try it! ∎

1.2 Stochastic Iterative Maps

Many of the systems we would like to consider are described by system variables whose value at the next time step we cannot predict with complete certainty. The uncertainty may arise from many sources, including the existence of interactions and parameters that are too complicated or not very relevant to our problem. We are then faced with describing a system in which the outcome of an iteration is probabilistic and not deterministic. Such systems are called stochastic systems. There are several ways to describe such systems mathematically. One of them is to consider the outcome of a particular update to be selected from a set of possible values. The probability of each of the possible values must be specified. This description is not really a model of a single system, because each realization of the system will do something different. Instead, this is a model of a collection of systems—an ensemble. Our task is to study the properties of this ensemble.

A stochastic system is generally described by the time evolution of random variables. We begin the discussion by defining a random variable. A random variable s is defined by its probability distribution $P_s(s')$, which describes the likelihood that s has the value s'. If s is a continuous variable, then $P_s(s')ds'$ is the probability that s resides between s' and $s' + ds'$. Note that the subscript is the variable name rather than an index. For example, s might be a binary variable that can have the value +1 or −1. $P_s(1)$ is the probability that $s = 1$ and $P_s(-1)$ is the probability that $s = -1$. If s is the outcome of an unbiased coin toss, with heads called 1 and tails called −1, both of these values are 1/2. When no confusion can arise, the notation $P_s(s')$ is abbreviated to $P(s)$, where s may be either the variable or the value. The sum over all possible values of the probability must be 1.

$$\sum_{s'} P_s(s') = 1 \tag{1.2.1}$$

In the discussion of a system described by random variables, we often would like to know the average value of some quantity $Q(s)$ that depends in a definite way on the value of the stochastic variable s. This average is given by:

$$< Q(s) >= \sum_{s'} P_s(s')Q(s') \tag{1.2.2}$$

Note that the average is a linear operation.

We now consider the case of a time-dependent random variable. Rather than describing the time dependence of the variable $s(t)$, we describe the time dependence of the probability distribution $P_s(s';t)$. Similar to the iterative map, we can consider the case where the outcome only depends on the value of the system variable at a previous time,and the transition probabilities do not depend explicitly on time. Such systems are called Markov chains. The transition probabilities from a state at a particular time to the next discrete time are written:

$$P_s(s'(t)|s'(t-1)) \tag{1.2.3}$$

P_s is used as the notation for the transition probability, since it is also the probability distribution of s at time t, given a particular value $s'(t-1)$ at the previous time. The use of a time index for the arguments illustrates the use of the transition probability. $P_s(1|1)$ is the probability that when $s = 1$ at time $t-1$ then $s = 1$ at time t. $P_s(-1|1)$ is the probability that when $s = 1$ at time $t-1$ then $s = -1$ at time t. The transition probabilities,along with the initial probability distribution of the system $P_s(s';\, t = 0)$, determine the time-dependent ensemble that we are interested in. Assuming that we don't lose systems on the way, the transition probabilities of Eq. (1.2.3) must satisfy:

$$\sum_{s''} P_s(s''|s') = 1 \tag{1.2.4}$$

This states that no matter what the value of the system variable is at a particular time, it must reach some value at the next time.

The stochastic system described by transition probabilities can be written as an iterative map on the probability distribution $P(s)$

$$P_s(s';t) = \sum_{s'} P_s(s'|s'')P_s(s'';t-1) \tag{1.2.5}$$

It may be more intuitive to write this using the notation

$$P_s(s'(t);t) = \sum_{s'(t-1)} P_s(s'(t)|s'(t-1))P_s(s'(t-1);t-1) \tag{1.2.6}$$

in which case it may be sufficient, though hazardous, to write the abbreviated form

$$P(s(t)) = \sum_{s(t-1)} P(s(t)|s(t-1))P(s(t-1)) \tag{1.2.7}$$

It is important to recognize that the time evolution equation for the probability is linear. The linear evolution of this system (Eq. (1.2.5)) guarantees that superposition applies. If we start with an initial distribution $P(s;0) = \frac{1}{2}P^1(s;0) + \frac{1}{2}P^2(s;0)$ at time $t = 0$, then we could find the result at time t by separately looking at the evolution of each of the probabilities $P^1(s;0)$ and $P^2(s;0)$. Explicitly we can write $P(s;t) = \frac{1}{2}P^1(s;t) + \frac{1}{2}P^2(s;t)$. The meaning of this equation should be well noted. The right side of the equation is the sum of the evolved probabilities $P^1(s;t)$ and $P^2(s;t)$. This linearity is a direct consequence of the independence of different members of the ensemble and says nothing about the complexity of the dynamics.

We note that ultimately we are interested in the behavior of a particular system $s(t)$ that only has one value of s at every time t. The ensemble describes how many such systems will behave. Analytically it is easier to describe the ensemble as a whole, however, simulations may also be used to observe the behavior of a single system.

1.2.1 *Random walk*

Stochastic systems with only one binary variable might seem to be trivial, but we will devote quite a bit of attention to this problem. We begin by considering the simplest possible binary stochastic system. This is the system which corresponds to a coin toss. Ideally, for each toss there is equal probability of heads ($s = +1$) or tails ($s = -1$), and there is no memory from one toss to the next. The ensemble at each time is independent of time and has an equal probability of ± 1:

$$P(s;t) = \frac{1}{2}\delta_{s,1} + \frac{1}{2}\delta_{s,-1} \tag{1.2.8}$$

where the discrete delta function is defined by

$$\delta_{i,j} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{1.2.9}$$

Since Eq. (1.2.8) is independent of what happens at all previous times, the evolution of the state variable is given by the same expression

$$P(s'|s) = \frac{1}{2}\delta_{s',1} + \frac{1}{2}\delta_{s',-1} \tag{1.2.10}$$

We can illustrate the evaluation of the average of a function of s at time t:

$$\langle Q(s) \rangle_t = \sum_{s'=\pm 1} Q(s')P_s(s';t) = \sum_{s'=\pm 1} Q(s')\left(\frac{1}{2}\delta_{s',1} + \frac{1}{2}\delta_{s',-1}\right) = \frac{1}{2}\sum_{s'=\pm 1} Q(s') \tag{1.2.11}$$

For example, if we just take $Q(s)$ to be s itself we have the average of the system variable:

$$\langle s \rangle_t = \frac{1}{2}\sum_{s'=\pm 1} s' = 0 \tag{1.2.12}$$

Question 1.2.1 Will you win more fair coin tosses if (*a*) you pick heads every time, or if (*b*) you alternate heads and tails, or if (*c*) you pick heads or tails at random or if (*d*) you pick heads and tails by some other system? Explain why.

Solution 1.2.1 In general, we cannot predict the number of coin tosses that will be won, we can only estimate it based on the chance of winning. Assuming a fair coin means that this is the best that can be done. Any of the possibilities (*a*)–(*c*) give the same chance of winning. In none of these ways of gambling does the choice you make correlate with the result of the coin toss. The only system (*d*) that can help is if you have some information about what the result of the toss will be,like betting on the known result *after* the coin is tossed.A way to write this formally is to write the probability distribution of the choice that you are making. This choice is also a stochastic process. Calling the choice $c(t)$, the four possibilities mentioned are:

$$(a)\quad P(c;t) = \delta_{c,1} \tag{1.2.13}$$

$$(b)\quad P(c;t) = \frac{1+(-1)^t}{2}\delta_{c,1} + \frac{1-(-1)^t}{2}\delta_{c,-1} = \mathrm{mod}_2(t)\delta_{c,1} + \mathrm{mod}_2(t+1)\delta_{c,-1} \tag{1.2.14}$$

$$(c)\quad P(c;t) = \frac{1}{2}\delta_{c,1} + \frac{1}{2}\delta_{c,-1} \tag{1.2.15}$$

$$(d)\quad P(c;t) = \delta_{c,s(t)} \tag{1.2.16}$$

It is sufficient to show that the average probability of winning is the same in each of (*a*)–(*c*) and is just 1/2. We follow through the manipulations in order to illustrate some concepts in the treatment of more than one stochastic variable. We have to sum over the probabilities of each of the possible values of the coin toss and each of the values of the choices, adding up the probability that they coincide at a particular time t:

$$<\delta_{c,s}> = \sum_{s'}\sum_{c'} \delta_{c',s'} P_s(s';t) P_c(c',t) \tag{1.2.17}$$

This expression assumes that the values of the coin toss and the value of the choice are independent,so that the joint probability of having a particular value of s and a particular value of c is the product of the probabilities of each of the variables independently:

$$P_{s,c}(s',c';t) = P_s(s';t) P_c(c';t) \tag{1.2.18}$$

—the probabilities-of-independent-variables factor. This is valid in cases (*a*)–(*c*) and not in case (*d*), where the probability of c occurring is explicitly a function of the value of s.

We evaluate the probability of winning in each case (*a*) through (*c*) using

$$\begin{aligned} <\delta_{c,s}> &= \sum_{s'}\sum_{c'} \delta_{c',s'}\left(\tfrac{1}{2}\delta_{s',1} + \tfrac{1}{2}\delta_{s',-1}\right)P_c(c';t) \\ &= \sum_{c'}\left(\tfrac{1}{2}\delta_{c',1} + \tfrac{1}{2}\delta_{c',-1}\right)P_c(c';t) \\ &= \sum_{c'}\left(\tfrac{1}{2}P_c(1;t) + \tfrac{1}{2}P_c(-1;t)\right) = \tfrac{1}{2} \end{aligned} \tag{1.2.19}$$

where the last equality follows from the normalization of the probability (the sum over all possibilities must be 1, Eq. (1.2.1)) and does not depend at all on the distribution. This shows that the independence of the variables guarantees that the probability of a win is just 1/2.

For the last case (*d*) the trivial answer, that a win is guaranteed by this method of gambling, can be arrived at formally by evaluating

$$<\delta_{c,s}> = \sum_{s'}\sum_{c'} \delta_{c',s'} P_{s,c}(s',c';t) \tag{1.2.20}$$

The value of s at time t is independent of the value of c, but the value of c depends on the value of s. The joint probability $P_{s,c}(s',c';t)$ may be written as the product of the probability of a particular value of $s = s'$ times the conditional probability $P_c(c'|s';t)$ of a particular value of $c = c'$ given the assumed value of s:

$$\begin{aligned} <\delta_{c,s}> &= \sum_{s'}\sum_{c'} \delta_{c',s'} P_s(s';t) P_c(c'\,|\,s';t) \\ &= \sum_{s'}\sum_{c'} \delta_{c',s'} P_s(s';t)\delta_{c',s'} = \sum_{s'} P_s(s';t) = 1 \end{aligned} \tag{1.2.21}$$ ∎

The next step in our analysis of the binary stochastic system is to consider the behavior of the sum of $s(t)$ over a particular number of time steps. This sum is the difference between the total number of heads and the total number of tails. It is equivalent to asking how much you will win or lose if you gamble an equal amount of money on each coin toss after a certain number of bets. This problem is known as a random walk, and we will define it as a consideration of the state variable

$$d(t) = \sum_{t'=1}^{t} s(t') \tag{1.2.22}$$

The way to write the evolution of the state variable is:

$$P(d'\,|\,d) = \tfrac{1}{2}\delta_{d',d+1} + \tfrac{1}{2}\delta_{d',d-1} \tag{1.2.23}$$

Thus a random walk considers a state variable d that can take integer values $d \in \{\ldots, -1,0,1,\ldots\}$. At every time step, $d(t)$ can only move to a value one higher or one lower than where it is. We assume that the probability of a step to the right (higher) is equal to that of a step to the left (lower). For convenience, we assume (with no loss of gener-

ality) that the system starts at position $d(0) = 0$. This is built into Eq.(1.2.22). Because of the symmetry of the system under a shift of the origin, this is equivalent to considering any other starting point. Once we solve for the probability distribution of d at time t, because of superposition we can also find the result of evolving any initial probability distribution $P(d;t = 0)$.

We can picture the random walk as that of a drunk who has difficulty consistently moving forward. Our model of this walk assumes that the drunk is equally likely to take a step forward or backward. Starting at position 0, he moves to either +1 or −1. Let's say it was +1. Next he moves to +2 or back to 0. Let's say it was 0. Next to +1 or −1. Let's say it was +1. Next to +2 or 0. Let's say +2. Next to +3 or +1. Let's say +1. And so on.

What is the value of system variable $d(t)$ at time t? This is equivalent to asking how far has the walk progressed after t steps. Of course there is no way to know how far a particular system goes without watching it. The average distance over the ensemble of systems is the average over all possible values of $s(t)$. This average is given by applying Eq. (1.2.2) or Eq. (1.2.11) to all of the variables $s(t)$:

$$\begin{aligned} <d(t)> &= \frac{1}{2}\sum_{s(t)=\pm 1} \cdots \frac{1}{2}\sum_{s(3)=\pm 1} \frac{1}{2}\sum_{s(2)=\pm 1} \frac{1}{2}\sum_{s(1)=\pm 1} d(t) \\ &= \sum_{t'=1}^{t} <s(t')> = 0 \end{aligned} \tag{1.2.24}$$

The average is written out explicitly on the first line using Eq.(1.2.11). The second line expression can be arrived at either directly or from the linearity of the average. The final answer is clear, since it is equally likely for the walker to move to the right as to the left.

We can also ask what is a typical distance traveled by a particular walker. By typical distance we mean how far from the starting point. This can either be defined by the average absolute value of the distance, or as is more commonly accepted, the root mean square (RMS) distance:

$$\sigma(t) = \sqrt{<d(t)^2>} \tag{1.2.25}$$

$$<d(t)^2> = <\left(\sum_{t'=1}^{t} s(t')\right)^2> = <\sum_{t',t''=1}^{t} s(t')s(t'')> = \sum_{t',t''=1}^{t} <s(t')s(t'')> \tag{1.2.26}$$

To evaluate the average of the product of the two steps, we treat differently the case in which they are the same step and when they are different steps. When the two steps are the same one we use $s(t) = \pm 1$ to obtain:

$$<s(t)^2> = <1> = 1 \tag{1.2.27}$$

Which follows from the normalization of the probability (or is obvious). To evaluate the average of the product of two steps at different times we need the joint probability of $s(t)$ and $s(t')$. This is the probability that each of them will take a particular

value. Because we have assumed that the steps are *independent*, the joint probability is the product of the probabilities for each one separately:

$$P(s(t), s(t')) = P(s(t))P(s(t')) \qquad t \neq t' \tag{1.2.28}$$

so that for example there is 1/4 chance that $s(t) = +1$ and $s(t) = -1$. The independence of the two steps leads the average of the product of the two steps to factor:

$$\begin{aligned} <s(t)s(t')> &= \sum_{s(t),s(t')} P(s(t), s(t'))s(t)s(t') \\ &= \sum_{s(t),s(t')} P(s(t))P(s(t'))s(t)s(t') \qquad t \neq t' \\ &= <s(t)><s(t')> = 0 \end{aligned} \tag{1.2.29}$$

This is zero, since either of the averages are zero. We have the combined result:

$$<s(t)s(t')> = \delta_{t,t'} \tag{1.2.30}$$

and finally:

$$<d(t)^2> = \sum_{t',t''=1}^{t} <s(t')s(t'')> = \sum_{t',t''=1}^{t} \delta_{t',t''} = \sum_{t'=1}^{t} 1 = t \tag{1.2.31}$$

This gives the classic and important result that a random walk travels a typical distance that grows as the square root of the number of steps taken: $\sigma(t) = \sqrt{t}$.

We can now consider more completely the probability distribution of the position of the walker at time t. The probability distribution at $t = 0$ may be written:

$$P(d;0) = \delta_{d,0} \tag{1.2.32}$$

After the first time step the probability distribution changes to

$$P(d;1) = \tfrac{1}{2}\delta_{d,1} + \tfrac{1}{2}\delta_{d,-1} \tag{1.2.33}$$

this results from the definition $d(1) = s(1)$. After the second step $d(2) = s(1) + s(2)$ it is:

$$P(d;2) = \tfrac{1}{4}\delta_{d,2} + \tfrac{1}{2}\delta_{d,0} + \tfrac{1}{4}\delta_{d,-2} \tag{1.2.34}$$

More generally it is not difficult to see that the probabilities are given by normalized binomial coefficients,since the number of ones chosen out of t steps is equivalent to the number of powers of x in $(1 + x)^t$. To reach a position d after t steps we must take $(t + d)/2$ steps to the right and $(t - d)/2$ steps to the left. The sum of these is the number of steps t and their difference is d. Since each choice has 1/2 probability we have:

$$P(d,t) = \frac{1}{2^t}\left(\begin{array}{c} t \\ (d+t)/2 \end{array}\right)\delta_{t,d}^{oddeven} = \frac{1}{2^t}\frac{t!}{[(d+t)/2]![(t-d)/2]!}\delta_{t,d}^{oddeven}$$
$$\delta_{t,d}^{oddeven} = \frac{(1+(-1)^{t+d})}{2} \tag{1.2.35}$$

where the unusual delta function imposes the condition that d takes only odd or only even values depending on whether t is odd or even.

Let us now consider what happens after a long time. The probability distribution spreads out,and a single step is a small distance compared to the typical distance traveled. We can consider s and t to be continuous variables where both conditions $d,t >> 1$ are satisfied. Moreover, we can also consider $|d| << t$, because the chance that all steps will be taken in one direction becomes very small. This enables us to use Sterling's approximation to the factorial

$$x! \sim \sqrt{2\pi x}\, e^{-x} x^x$$
$$\ln(x!) \sim x(\ln x - 1) + \ln(\sqrt{2\pi x}) \tag{1.2.36}$$

For large t it also makes sense not to restrict d to be either odd or even. In order to allow both, we,in effect, interpolate and then take only half of the probability we have in Eq. (1.2.35). This leads to the expression:

$$P(d,t) = \frac{\sqrt{t}}{\sqrt{2\pi(t-d)(t+d)}2^t}\frac{t^t e^{-t}}{[(d+t)/2]^{[(d+t)/2]}[(t-d)/2]^{[(t-d)/2]}e^{-(d+t)/2-(t-d)/2}}$$
$$= \frac{(2\pi t(1-x^2))^{-1/2}}{(1+x)^{[(1+x)t/2]}(1-x)^{[(1-x)t/2]}} \tag{1.2.37}$$

where we have defined $x = d/t$. To approximate this expression it is easier to consider it in logarithmic form:

$$\ln(P(d,t)) = -(t/2)[(1+x)\ln(1+x) + (1-x)\ln(1-x)] - (1/2)\ln(2\pi t(1-x^2))$$
$$\approx -(t/2)[(1+x)(x - x^2/2 + \ldots) + (1-x)(-x - x^2/2 + \ldots)] - (1/2)\ln(2\pi t + \ldots)$$
$$= -tx^2/2 - \ln(\sqrt{2\pi t}) \tag{1.2.38}$$

or exponentiating:

$$P(d,t) = \frac{1}{\sqrt{2\pi t}} e^{-d^2/2t} = \frac{1}{\sqrt{2\pi}\sigma} e^{-d^2/2\sigma^2} \tag{1.2.39}$$

The prefactor of the exponential, $1/\sqrt{2\pi}\sigma$, originates from the factor $\sqrt{2\pi x}$ in Eq. (1.2.36). It is independent of d and takes care of the normalization of the probability. The result is a Gaussian distribution. Questions 1.2.2–1.2.5 investigate higher-order corrections to the Gaussian distribution.

Question 1.2.2 In order to obtain a correction to the Gaussian distribution we must add a correction term to Sterling's approximation:

$$\begin{aligned} x! &\sim \sqrt{2\pi x}\, e^{-x} x^x \left(1 + \frac{1}{12x} + \quad\right) \\ \ln(x!) &\sim x(\ln x - 1) + \ln\left(\sqrt{2\pi x}\right) + \ln\left(1 + \frac{1}{12x} + \quad\right) \end{aligned} \tag{1.2.40}$$

Using this expression, find the first correction term to Eq. (1.2.37).

Solution 1.2.2 The correction term in Sterling's approximation contributes a factor to Eq. (1.2.37) which is (for convenience we write here $c = 1/12$):

$$\frac{(1+c/t)}{(1+2c/(t+d))(1+2c/(t-d))} = \left(1 - \frac{3c}{t} + \quad\right) = \left(1 - \frac{1}{4t} + \quad\right) \tag{1.2.41}$$

where we have only kept the largest correction term, neglecting d compared to t. Note that the correction term vanishes as t becomes large. ∎

Question 1.2.3 Keeping additional terms of the expansion in Eq.(1.2.38), and the result of Question 1.2.2,find the first order correction terms to the Gaussian distribution.

Solution 1.2.3 Correction terms in Eq. (1.2.38) arise from several places. We want to keep all terms that are of order $1/t$. To do this we must keep in mind that a typical distance traveled is $d \sim \sqrt{t}$, so that $x \sim 1/\sqrt{t}$. The next terms are obtained from:

$$\begin{aligned} \ln(P(d,t)) &= -(t/2)[(1+x)\ln(1+x) + (1-x)\ln(1-x)] \\ &\quad - (1/2)\ln(2\pi t(1-x^2)) + \ln(1 - 1/4t) \\ &\approx -(t/2)[(1+x)(x - \tfrac{1}{2}x^2 + \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 \ldots) \\ &\quad + (1-x)(-x - \tfrac{1}{2}x^2 - \tfrac{1}{3}x^3 - \tfrac{1}{4}x^4 \ldots)] \\ &\quad - \ln\left(\sqrt{2\pi t}\right) - (1/2)\ln(1-x^2) + \ln(1 - 1/4t) \\ &\approx -(t/2)[(x + x^2 - \tfrac{1}{2}x^2 - \tfrac{1}{2}x^3 + \tfrac{1}{3}x^3 + \tfrac{1}{3}x^4 - \tfrac{1}{4}x^4 \ldots) \\ &\quad + (-x + x^2 - \tfrac{1}{2}x^2 + \tfrac{1}{2}x^3 - \tfrac{1}{3}x^3 + \tfrac{1}{3}x^4 - \tfrac{1}{4}x^4 \ldots)] \\ &\quad - \ln\left(\sqrt{2\pi t}\right) + (x^2/2 + \quad) + (-1/4t + \quad) \\ &= -tx^2/2 - \ln\left(\sqrt{2\pi t}\right) - tx^4/12 + x^2/2 - 1/4t \end{aligned} \tag{1.2.42}$$

This gives us a distribution:

$$P(d,t) = \sqrt{\frac{1}{2\pi t}} e^{-d^2/2t} e^{-d^4/12t^3 + d^2/2t^2 - 1/4t} \qquad (1.2.43)$$ ∎

Question 1.2.4 What is the size of the additional factor? Estimate the size of this term as t becomes large.

Solution 1.2.4 The typical value of the variable d is its root mean square value $\sigma = \sqrt{t}$. At this value the additional term gives a factor

$$e^{1/6t} \qquad (1.2.44)$$

which approaches 1 as time increases. ∎

Question 1.2.5 What is the fraction error that we will make if we neglect this term after one hundred steps? After ten thousand steps?

Solution 1.2.5 After one hundred time steps the walker has traveled a typical distance of ten steps. We generally approximate the probability of arriving at this distance using Eq.(1.2.39). The fractional error in the probability of arriving at this distance according to Eq. (1.2.44) is $1 - e^{1/6t} \approx -1/6t = -0.00167$. So already at a distance of ten steps the error is less than 0.2%.

It is much less likely for the walker to arrive at the distance $2\sigma = 20$. The ratio of the probability to arrive at 20 compared to 10 is $e^{-2}/e^{-0.5}$ 0.22. If we want to know the error of this smaller probability case we would write $(1 - e^{-16/12t + 4/2t - 1/4t}) = (1 - e^{5/12t}) \approx -0.0042$, which is a larger but still small error.

After ten thousand steps the errors are smaller than the errors at one hundred steps by a factor of one hundred. ∎

1.2.2 *Generalized random walk and the central limit theorem*

We can generalize the random walk by allowing a variety of steps from the current location of the walker to sites nearby, not only to the adjacent sites and not only to integer locations. If we restrict ourselves to steps that on average are balanced left and right and are not too long ranged, we can show that all such systems have the same behavior as the simplest random walk at long enough times (and characteristically not even for very long times). This is the content of the central limit theorem. It says that summing any set of independent random variables eventually leads to a Gaussian distribution of probabilities, which is the same distribution as the one we arrived at for the random walk. The reason that the same distribution arises is that successive iteration of the probability update equation, Eq.(1.2.7),smoothes out the distribution, and the only relevant information that survives is the width of the distribution which is given by $\sigma(t)$. The proof given below makes use of a Fourier transform and can be skipped by readers who are not well acquainted with transforms. In the next section we will also include a bias in the random walk. For long times this can be described as

an average motion superimposed on the unbiased random walk. We start with the unbiased random walk.

Each step of the random walk is described by the state variable $s(t)$ at time t. The probability of a particular step size is an unspecified function that is independent of time:

$$P(s;t) = f(s) \tag{1.2.45}$$

We treat the case of integer values of s. The continuum case is Question 1.2.6. The absence of bias in the random walk is described by setting the average displacement in a single step to zero:

$$< s > = \sum_s s f(s) = 0 \tag{1.2.46}$$

The statement above that each step is not too long ranged,is mathematically just that the mean square displacement in a single step has a well-defined value (i.e., is not infinite):

$$< s^2 > = \sum_s s^2 f(s) = \sigma_0^2 \tag{1.2.47}$$

Eqs. (1.2.45)–(1.2.47) hold at all times.

We can still evaluate the average of $d(t)$ and the RMS value of $d(t)$ directly using the linearity of the average:

$$< d(t) > = < \sum_{t'=1}^{t} s(t') > = t < s > = 0 \tag{1.2.48}$$

$$< d(t)^2 > = < \left(\sum_{t'=1}^{t} s(t') \right)^2 > = \sum_{t',t''=1}^{t} < s(t') s(t'') > \tag{1.2.49}$$

Since $s(t')$ and $s(t'')$ are independent for $t' \neq t''$, as in Eq. (1.2.29), the average factors:

$$< s(t') s(t'') > = < s(t') > < s(t'') > = 0 \qquad t' \neq t'' \tag{1.2.50}$$

Thus, all terms $t' \neq t''$ are zero by Eq. (1.2.46). We have:

$$< d(t)^2 > = \sum_{t'=1}^{t} < s(t')^2 > = t \sigma_0^2 \tag{1.2.51}$$

This means that the typical value of $d(t)$ is $\sigma_0 \sqrt{t}$.

To obtain the full distribution of the random walk state variable $d(t)$ we have to sum the stochastic variables $s(t)$. Since $d(t) = d(t-1) + s(t)$ the probability of transition from $d(t-1)$ to $d(t)$ is $f(d(t) - d(t-1))$ or:

$$P(d'|d) = f(d'-d) \tag{1.2.52}$$

We can now write the time evolution equation and iterate it t times to get $P(d;t)$.

$$P(d;t) = \sum_{d'} P(d|d')P(d';t-1) = \sum_{d'} f(d-d')P(d';t-1) \tag{1.2.53}$$

This is a convolution, so the most convenient way to effect a t fold iteration is in Fourier space. The Fourier representation of the probability and transition functions for integral d is:

$$\begin{aligned} \tilde{P}(k;t) &\equiv \sum_{d} e^{-ikd} P(d;t) \\ \tilde{f}(k) &\equiv \sum_{s} e^{-iks} f(s) \end{aligned} \tag{1.2.54}$$

We use a Fourier series because of the restriction to integer values of d. Once we solve the problem using the Fourier representation, the probability distribution is recovered from the inverse formula:

$$P(d;t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{ikd}\tilde{P}(k;t) \tag{1.2.55}$$

which is proved

$$\begin{aligned} &\frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{ikd}\tilde{P}(k;t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{ikd}\sum_{d'} e^{-ikd'}P(d';t) \\ &= \frac{1}{2\pi}\sum_{d'} P(d';t)\int_{-\pi}^{\pi} dk e^{ik(d-d')} = \sum_{d'} P(d';t)\delta_{d,d'} = P(d;t) \end{aligned} \tag{1.2.56}$$

using the expression:

$$\delta_{d,d'} = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{ik(d-d')} \tag{1.2.57}$$

Applying Eq. (1.2.54) to Eq. (1.2.53):

$$\begin{aligned} \tilde{P}(k;t) &= \sum_{d} e^{-ikd}\sum_{d'} f(d-d')P(d';t-1) \\ &= \sum_{d'}\sum_{d} e^{-ik(d-d')}e^{-ikd'} f(d-d')P(d';t-1) \\ &= \sum_{d'}\sum_{d''} e^{-ikd''}e^{-ikd'} f(d'')P(d';t-1) \\ &= \sum_{d'} e^{-ikd''} f(d'')\sum_{d'} e^{-ikd'}P(d';t-1) = \tilde{f}(k)\tilde{P}(k;t-1) \end{aligned} \tag{1.2.58}$$

we can iterate the equation to obtain:

$$\tilde{P}(k;t) = \tilde{f}(k)\tilde{P}(k;t-1) = \tilde{f}(k)^t \tag{1.2.59}$$

where we use the definition $d(1) = s(1)$ that ensures that $P(d;1) = P(s;1) = f(d)$.

For large t the walker has traveled a large distance,so we are interested in variations of the probability $P(d;t)$ over large distances. Thus,in Fourier space we are concerned with small values of k. To simplify Eq.(1.2.59) for large t we expand $\tilde{f}(k)$ near $k = 0$. From Eq.(1.2.54) we can directly evaluate the derivatives of $\tilde{f}(k)$ at $k = 0$ in terms of averages:

$$\left.\frac{d^n \tilde{f}(k)}{d^n k}\right|_{k=0} = \sum_s (-is)^n f(s) = (-i)^n \langle s^n \rangle \tag{1.2.60}$$

We can use this expression to evaluate the terms of a Taylor expansion of $\tilde{f}(k)$:

$$\tilde{f}(k) = \tilde{f}(0) + \left.\frac{\partial \tilde{f}(k)}{\partial k}\right|_{k=0} k + \frac{1}{2}\left.\frac{\partial^2 \tilde{f}(k)}{\partial k^2}\right|_{k=0} k^2 + \ldots \tag{1.2.61}$$

$$\tilde{f}(k) = \langle 1 \rangle - i\langle s \rangle k - \frac{1}{2}\langle s^2 \rangle k^2 + \ldots \tag{1.2.62}$$

Using the normalization of the probability ($\langle 1 \rangle = 1$),and Eqs.(1.2.46) and (1.2.47), gives us:

$$\tilde{P}(k;t) = \left(1 - \tfrac{1}{2}\sigma_0^2 k^2 + \ldots\right)^t \tag{1.2.63}$$

We must now remember that a typical value of $d(t)$,from its RMS value, is $\sigma_0\sqrt{t}$. By the properties of the Fourier transform,this implies that a typical value of k that we must consider in Eq.(1.2.63) varies with time as $1/\sqrt{t}$. The next term in the expansion,cubic in k, would give rise to a term that is smaller by this factor, and therefore becomes unimportant at long times. If we write $k = q/\sqrt{t}$, then it becomes clearer how to write Eq. (1.2.63) using a limiting expression for large t:

$$\tilde{P}(k;t) = \left(1 - \tfrac{1}{2}\frac{\sigma_0^2 q^2}{t} + \ldots\right)^t \sim e^{-\sigma_0^2 q^2/2} = e^{-t\sigma_0^2 k^2/2} \tag{1.2.64}$$

This Gaussian, when Fourier transformed back to an expression in d, gives us a Gaussian as follows:

$$P(d;t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{ikd} e^{-t\sigma_0^2 k^2/2} \quad \frac{1}{2\pi}\int_{-\infty}^{\infty} dk e^{ikd} e^{-t\sigma_0^2 k^2/2} \tag{1.2.65}$$

We have extended the integral because the decaying exponential becomes narrow as t increases. The integral is performed by completing the square in the exponent, giving:

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk e^{-d^2/2t\sigma_0^2} e^{-(t\sigma_0^2 k^2 - 2ikd - d^2/t\sigma_0^2)/2} = \frac{1}{\sqrt{2\pi t\sigma_0^2}} e^{-d^2/2t\sigma_0^2} \tag{1.2.66}$$

or equivalently:

$$P(d;t) = \frac{1}{\sqrt{2\pi\sigma(t)^2}} e^{-d^2/2\sigma(t)^2} \tag{1.2.67}$$

which is the same as Eq. (1.2.39).

Question 1.2.6 Prove the central limit theorem when s takes a continuum of values.

Solution 1.2.6 The proof follows the same course as the integer valued case. We must define the appropriate averages,and the transform. The average of s is still zero, and the mean square displacement is defined similarly:

$$< s > = \int ds\, s f(s) = 0 \tag{1.2.46′}$$

$$< s^2 > = \int ds\, s^2 f(s) = \sigma_0^2 \tag{1.2.47′}$$

To avoid problems of notation we substitute the variable x for the state variable d:

$$< x(t) > = < \sum_{t'=1}^{t} s(t') > = t < s > = 0 \tag{1.2.48′}$$

Skipping steps that are the same we find:

$$< x(t)^2 > = < \left(\sum_{t'=1}^{t} s(t') \right)^2 > = \sum_{t'=1}^{t} < s(t')^2 > = t\sigma_0^2 \tag{1.2.51′}$$

since $s(t')$ and $s(t'')$ are still independent for $t' \neq t''$. Eq. (1.2.53) is also essentially unchanged:

$$P(x;t) = \int dx'\, f(x - x') P(x';t-1) \tag{1.2.53′}$$

The transform and inverse transform must now be defined using

$$\begin{aligned} \tilde{P}(k;t) &\equiv \int dx\, e^{-ikx} P(x;t) \\ \tilde{f}(k) &\equiv \int ds\, e^{-iks} f(s) \end{aligned} \tag{1.2.54′}$$

$$P(d;t) = \frac{1}{2\pi}\int dk e^{ikd}\tilde{P}(k;t) \qquad (1.2.55')$$

The latter is proved using the properties of the Dirac (continuum) delta function:

$$\delta(x-x') = \frac{1}{2\pi}\int dk e^{ik(x-x')}$$
$$\int dx'\delta(x-x')g(x') = g(x) \qquad (1.2.56')$$

where the latter equation holds for an arbitrary function $g(x)$.

The remainder of the derivation carries forward unchanged. ∎

1.2.3 *Biased random walk*

We now return to the simple random walk with binary steps of ±1. The model we consider is a random walk that is biased in one direction.Each time a step is taken there is a probability P_+ for a step of +1, that is different from the probability P_- for a step of –1, or:

$$P(s;t) = P_+\delta_{s,1} + P_-\delta_{s,-1} \qquad (1.2.68)$$

$$P(d'|d) = P_+\delta_{d',d+1} + P_-\delta_{d',d-} \qquad (1.2.69)$$

where

$$P_+ + P_- = 1 \qquad (1.2.70)$$

What is the average distance traveled in time t?

$$<d(t)> = \sum_{t'=1}^{t} <s(t')> = \sum_{t'=1}^{t}(P_+ - P_-) = t(P_+ - P_-) \qquad (1.2.71)$$

This equation justifies defining the mean velocity as

$$v = P_+ - P_- \qquad (1.2.72)$$

Since we already have an average displacement it doesn't make sense to also ask for a typical displacement,as we did with the random walk—the typical displacement is the average one. However, we can ask about the spread of the displacements around the average displacement

$$\sigma(t)^2 = <(d(t) - <d(t)>)^2> = <d(t)^2> - 2<d(t)>^2 + <d(t)>^2$$
$$= <d(t)^2> - <d(t)>^2 \qquad (1.2.73)$$

This is called the standard deviation and it reduces to the RMS distance in the unbiased case. For many purposes $\sigma(t)$ plays the same role in the biased random walk as in the unbiased random walk. From Eq. (1.2.71) and Eq. (1.2.72) the second term is $(vt)^2$. The first term is:

$$\begin{aligned} <d(t)^2> &= <\left(\sum_{t'=1}^{t} s(t')\right)^2> = \sum_{t',t''=1}^{t} <s(t')s(t'')> \\ &= \sum_{t',t''=1}^{t} \left(\delta_{t',t''} + (1-\delta_{t',t''})(P_+^2 + P_-^2 - 2P_+P_-)\right) \\ &= t + t(t-1)v^2 = t^2v^2 + t(1-v^2) \end{aligned} \tag{1.2.74}$$

Substituting in Eq. (1.2.73):

$$\sigma^2 = t(1-v^2) \tag{1.2.75}$$

It is interesting to consider this expression in the two limits $v = 1$ and $v = 0$. For $v = 1$ the walk is deterministic, $P_+ = 1$ and $P_- = 0$, and there is no element of chance; the walker always walks to the right. This is equivalent to the iterative map Eq.(1.1.4).Our result Eq.(1.2.66) is that $\sigma = 0$, as it must be for a deterministic system. However, for smaller velocities,the spreading of the systems σ increases until at $v = 0$ we recover the case of the unbiased random walk.

The complete probability distribution is given by:

$$P(d;t) = P_+^{(d+t)/2} P_-^{(d-t)/2} \left(\begin{array}{c} t \\ (d+t)/2 \end{array}\right) \delta_{t,d}^{oddeven} \tag{1.2.76}$$

For large t the distribution can be found as we did for the unbiased random walk. The work is left to Question 1.2.7.

Question 1.2.7 Find the long time (continuum) distribution for the biased random walk.

Solution 1.2.7 We use the Sterling approximation as before and take the logarithm of the probability. In addition to the expression from the first line of Eq. (1.2.38) we have an additional factor due to the coefficient of Eq.(1.2.76) which appears in place of the factor of $1/2^t$. We again define $x = d/t$, and divide by 2 to allow both odd and even integers. We obtain the expression:

$$\begin{aligned} \ln(P(d,t)) &= (t/2)[(1+x)\ln 2P_+ + (1-x)\ln 2P_-] \\ &-(t/2)[(1+x)\ln(1+x) + (1-x)\ln(1-x)] - (1/2)\ln(2\pi t(1-x^2)) \end{aligned} \tag{1.2.77}$$

It makes the most sense to expand this around the mean of x, $<x> = v$. To simplify the notation we can use Eq. (1.2.70) and Eq. (1.2.72) to write:

$$\begin{aligned} P_+ &= (1+v)/2 \\ P_- &= (1-v)/2 \end{aligned} \tag{1.2.78}$$

With these substitutions we have:

$$\begin{aligned} \ln(P(d,t)) &= (t/2)[(1+x)\ln(1+v) + (1-x)\ln(1-v)] \\ &-(t/2)[(1+x)\ln(1+x) + (1-x)\ln(1-x)] - (1/2)\ln(2\pi t(1-x^2)) \end{aligned} \tag{1.2.79}$$

We expand the first two terms in a Taylor expansion around the mean of x and expand the third term inside the logarithm. The first term of Eq.(1.2.79) has only a constant and linear term in a Taylor expansion. These cancel the constant and the first derivative of the Taylor expansion of the second term of Eq. (1.2.79) at $x = v$. Higher derivatives arise only from the second term:

$$\begin{aligned}\ln(P(d,t)) &= -(t/2)\left[\frac{1}{(1-v^2)}(x-v)^2 + \frac{2}{3(1-v^2)^2}(x-v)^3 + \ldots\right] \\ &\quad -(1/2)\ln(2\pi t[(1-v^2) - 2v(x-v) + \ldots]) \qquad (1.2.80) \\ &= -\left[\frac{(d-vt)^2}{2\sigma(t)^2} + \frac{(d-vt)^3}{3\sigma(t)^4} + \ldots\right] - (1/2)\ln(2\pi(\sigma(t)^2 - 2v(d-vt) + \ldots))\end{aligned}$$

In the last line we have restored d and used Eq.(1.2.75). Keeping only the first terms in both expansions gives us:

$$P(d;t) = \frac{1}{\sqrt{2\pi\sigma(t)^2}} e^{-(d-vt)^2/2\sigma(t)^2} \qquad (1.2.81)$$

which is a Gaussian distribution around the mean we obtained before. This implies that aside from the constant velocity, and a slightly modified standard deviation, the distribution remains unchanged.

The second term in both expansions in Eq.(1.2.80) become small in the limit of large t, as long as we are not interested in the tail of the distribution. Values of $(d - vt)$ relevant to the main part of the distribution are given by the standard deviation, $\sigma(t)$. The second terms in Eq. (1.2.80) are thus reduced by a factor of $\sigma(t)$ compared to the first terms in the series. Since $\sigma(t)$ grows as the square root of the time, they become insignificant for long times. The convergence is slower, however, than in the unbiased random walk (Questions 1.2.2–1.2.5). ∎

Question 1.2.8 You are a manager of a casino and are told by the owner that you have a cash flow problem. In order to survive, you have to make sure that nine out of ten working days you have a profit. Assume that the only game in your casino is a roulette wheel. Bets are limited to only red or black with a 2:1 payoff. The roulette wheel has an equal number of red numbers and black numbers and one green number (the house always wins on green). Assume that people make a fixed number of 10^6 total \$1 bets on the roulette wheel in each day.

a. What is the maximum number of red numbers on the roulette wheel that will still allow you to achieve your objective?

b. With this number of red numbers, how much money do you make on average in each day?

Solution 1.2.8 The casino wins \$1 for every wrong bet and loses \$1 for every right bet. The results of bets at the casino are equivalent to a random walk with a bias given by:

$$P_+ = (N_{red} + 1)/(N_{red} + N_{black} + 1) \tag{1.2.82}$$

$$P_- = N_{black}/(N_{red} + N_{black} + 1) \tag{1.2.83}$$

where,as the manager, we consider positive the wins of the casino. The color subscripts can be used interchangeably, since the number of red and black is equal. The velocity of the random walk is given by:

$$v = 1/(2N_{red} + 1) \tag{1.2.84}$$

To calculate the probability that the casino will lose on a particular day we must sum the probability that the random walk after 10^6 steps will result in a negative number. We approximate the sum by an integral over the distribution of Eq. (1.2.81). To avoid problems of notation we replace d with y:

$$\begin{aligned} P_{loss} &= \int_{-\infty}^{0} dy P(y; t = 10^6) = \frac{1}{\sqrt{2\pi\sigma(t)^2}} \int_{-\infty}^{0} dy e^{-(y-vt)^2/2\sigma(t)^2} \\ &= \frac{1}{\sqrt{2\pi\sigma(t)^2}} \int_{-\infty}^{-vt} dy' e^{-(y')^2/2\sigma(t)^2} \\ &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z_0} dz e^{-z^2} = \frac{1}{2}(1 - \mathrm{erf}(z_0)) \end{aligned} \tag{1.2.85}$$

$$\begin{aligned} z &= y'/\sqrt{2}\sigma(t) \\ z_0 &= -vt/\sqrt{2\sigma(t)^2} = -vt/\sqrt{2t(1-v^2)} \end{aligned} \tag{1.2.86}$$

We have written the probability of loss in a day in terms of the error function erf(x)—the integral of a Gaussian defined by

$$\mathrm{erf}(z_0) \equiv \frac{2}{\sqrt{\pi}} \int_0^{z_0} dz e^{-z^2} \tag{1.2.87}$$

Since

$$\mathrm{erf}(\infty) = 1 \tag{1.2.88}$$

we have the expression

$$(1 - \mathrm{erf}(z_0)) \equiv \frac{2}{\sqrt{\pi}} \int_{z_0}^{\infty} dz e^{-z^2} \tag{1.2.89}$$

which is also known as the complementary error function erfc(x).

To obtain the desired constraint on the number of red numbers, or equivalently on the velocity, we invert Eq. (1.2.85) to find a value of v that gives the desired $P_{loss} = 0.1$, or $\mathrm{erf}(z_0) = 0.8$. Looking up the error function or using iterative guessing on an appropriate computer gives $z_0 = 0.9062$. Inverting Eq. (1.2.86) gives:

$$v = \frac{1}{\sqrt{t/2z_0 - 1}} \approx \sqrt{2z_0/t} \tag{1.2.90}$$

The approximation holds because t is large. The numerical result is $v = 0.0013$. This gives us the desired number of each color (inverting Eq. (1.2.84)) of $N_{red} = 371$. Of course the result is a very large number and the problem of winning nine out of ten days is a very conservative problem for a casino. Even if we insist on winning ninety-nine out of one hundred days we would have $\mathrm{erf}(z_0) = 0.98$, $z_0 = 1.645$, $v = 0.0018$ and $N_{red} = 275$. The profits per day in each case are given by vt, which is approximately \$1,300 and \$1,800 respectively. Of course this is much less than for bets on a more realistic roulette wheel. Eventually as we reduce the chance of the casino losing and z_0 becomes larger, we might become concerned that we are describing the properties of the tail of the distribution when we calculate the fraction of days the casino might lose, and Eq. (1.2.85) will not be very accurate. However, it is not difficult to see that casinos do not have cash flow problems. ∎

In order to generalize the proof of the central limit theorem to the case of a biased random walk, we can treat the continuum case most simply by considering the system variable $\hat{x}$, where (using $d \to x$ for the continuum case):

$$\hat{x} = x - \langle x \rangle_t = x - t\langle s \rangle = x - vt \tag{1.2.91}$$

Only x is a stochastic variable on the right side, v and t are numbers. Since iterations of this variable would satisfy the conditions for the generalized random walk, the generalization of the Gaussian distribution to Eq. (1.2.81) is proved. The discrete case is more difficult to prove because we cannot shift the variable d by arbitrary amounts and continue to consider it as discrete. We can argue the discrete case to be valid on the basis of the result for the continuum case, but a separate proof can be constructed as well.

1.2.4 *Master equation approach*

The Master equation is an alternative approach to stochastic systems, an alternative to Eq. (1.2.5), that is usually applied when time is continuous. We develop it starting from the discrete time case. We can rewrite Eq. (1.2.5) in the form of a difference equation for a particular probability $P(s)$. Beginning from:

$$P(s;t) = P(s;t-1) + \left[\sum_{s'} P(s|s')P(s';t-1) - P(s;t-1)\right] \tag{1.2.92}$$

we extract the term where the system remains in the same state:

$$P(s;t) = P(s;t-1) + \left[\sum_{s' \neq s} P(s|s')P(s';t-1) + P(s|s)P(s;t-1) - P(s;t-1)\right] \quad (1.2.93)$$

We use the normalization of probability to write it in terms of the transitions away from this site:

$$P(s;t) = P(s;t-1) + \left[\sum_{s' \neq s} P(s|s')P(s';t-1) + \left[1 - \sum_{s' \neq s} P(s'|s)\right]P(s;t-1) - P(s;t-1)\right] \quad (1.2.94)$$

Canceling the terms in the bracket that refer only to the probability $P(s;t-1)$ we write this as a difference equation. On the right appear only the probabilities at different values of the state variable ($s' \neq s$):

$$P(s,t) - P(s;t-1) = \sum_{s' \neq s} \Big(P(s|s')P(s';t-1) - P(s'|s)P(s;t-1) \Big) \quad (1.2.95)$$

To write the continuum form we reintroduce the time difference between steps Δt.

$$\frac{P(s,t) - P(s;t-\Delta t)}{\Delta t} = \sum_{s' \neq s} \left[\frac{P(s|s')}{\Delta t} P(s';t-\Delta t) - \frac{P(s'|s)}{\Delta t} P(s;t-\Delta t) \right] \quad (1.2.96)$$

When the limit of $\Delta t \rightarrow 0$ is meaningful, it is possible to make the change to the equation

$$\dot{P}(s,t) = \sum_{s' \neq s} \Big(R(s|s')P(s';t) - R(s'|s)P(s;t) \Big) \quad (1.2.97)$$

Where the ratio $P(s|s')/\Delta t$ has been replaced by the rate of transition $R(s|s')$. Eq.(1.2.97) is called the Master equation and we can consider Eq.(1.2.95) as the discrete time analog.

The Master equation has a simple interpretation: The rate of change of the probability of a particular state is the total rate at which probability is being added into that state from all other states,minus the total rate at which probability is leaving the state. Probability is acting like a fluid that is flowing to or from a particular state and is being conserved,as it must be. Eq.(1.2.97) is very much like the continuity equation of fluid flow, where the density of the fluid at a particular place changes according to how much is flowing to that location or from it.We will construct and use the Master equation approach to discuss the problem of relaxation in activated processes in Section 1.4.