Chapter 1

Background Material

To explore the biological world in both space and time, several mathematical tools are necessary. We need to be facile with multivariable calculus, with the qualitative theory of ordinary differential equations and their numerical simulation, and with basic stochastic processes. These are the topics to which we now give our attention.

1.1. Multivariable Calculus

1.1.1. Partial Derivatives. In what is to come, we will be dealing with functions of both space and time. We may be interested in the electrolyte balance within muscle tissue, or the distribution of microorganisms occupying a lake; either way, we are studying how something occupying a region in space evolves over time. We often describe position in this space by using the Cartesian coordinates x, y, and z, and time by the variable t (although other coordinate representations like polar or spherical coordinates are sometimes useful).

The quantity of interest may be the concentration (= number per unit volume) of calcium ions in a cell or the concentration of microorganisms in the lake, but it is typically denoted by some scalar function $u = u(x, y, z, t)$. If u changes smoothly in time, then it has a time derivative $\frac{\partial u}{\partial t}$ defined by[1]

$$\frac{\partial u}{\partial t} = \lim_{\Delta t \to 0} \frac{u(x, y, z, t + \Delta t) - u(x, y, z, t)}{\Delta t}. \tag{1.1}$$

The fundamental theorem of calculus states that

$$u(x, y, z, b) - u(x, y, z, a) = \int_a^b \frac{\partial u}{\partial t} dt. \tag{1.2}$$

In words, the cumulative change in u over an interval of time can be measured by observing the difference between u at the end and the beginning of the interval.

[1] For notation, the objects $\frac{\partial u}{\partial t}$ and u_t are exactly the same thing, as are $\frac{\partial u}{\partial x}$ and u_x.

Similarly, if u varies smoothly in space, spatial derivatives can be defined, such as

$$\frac{\partial u}{\partial x} = \lim_{\Delta x \to 0} \frac{u(x+\Delta x, y, z, t) - u(x, y, z, t)}{\Delta x}, \tag{1.3}$$

and this is one component of the *gradient* of u, the vector-valued function

$$\nabla u = (\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}). \tag{1.4}$$

The gradient holds all of the information we need about how u changes in space, but since there are an infinite number of directions in which we could move from a particular point, to find the derivative of u in a particular direction, $\mathbf{v}$, where $\mathbf{v}$ is a unit vector, we define the *directional derivative*,

$$\frac{\partial u}{\partial \mathbf{v}} = \nabla u \cdot \mathbf{v}. \tag{1.5}$$

One important object that uses the gradient is

$$\mathbf{n} = \frac{\nabla u}{|\nabla u|}, \tag{1.6}$$

which, provided $|\nabla u| \neq 0$, is a unit vector pointing in the direction of the greatest increase of the function u. The importance of this to a skier or snowboarder is obvious, pointing in the direction parallel to the "fall-line". It is also noteworthy that $\mathbf{n}$ is perpendicular (orthogonal) to level surfaces of the function u. We can verify this by noting that if $(x(s), y(s), z(s))$ is a curve in space parametrized by s, the tangent direction of the curve is the vector $(\dot{x}(s), \dot{y}(s), \dot{z}(s))$. If the function $u(x, y, z)$ is a constant on this curve, $u(x(s), y(s), z(s)) = C$, then differentiating this with respect to s, we find that

$$0 = \frac{\partial u}{\partial x}\dot{x}(s) + \frac{\partial u}{\partial y}\dot{y}(s) + \frac{\partial u}{\partial z}\dot{z}(s) \equiv \nabla u \cdot \begin{pmatrix} \dot{x}(s) \\ \dot{y}(s) \\ \dot{z}(s) \end{pmatrix}, \tag{1.7}$$

as claimed.

1.1.2. Vector Fields. It could be that the quantity of interest is a vector valued function, for example, the velocity of the water in a river or the velocity of the blood in an artery, given by $\mathbf{v} = (v_1, v_2, v_3)$ where each of the components of the vector $\mathbf{v}$ is a function of x, y, z, and t. Of course, this vector valued function could be the gradient of some scalar function u. One important quantity for any vector valued function is its *divergence*, denoted

$$\nabla \cdot \mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}. \tag{1.8}$$

The most important theorem regarding the divergence, called the *divergence theorem*, is stated as

$$\int_\Omega \nabla \cdot \mathbf{v} dV = \int_{\partial\Omega} \mathbf{v} \cdot \mathbf{n} dS, \tag{1.9}$$

where Ω is a region of interest with dimension $d = 1, 2$ or 3, $\partial\Omega$ is its $(d-1)$-dimensional boundary, and $\mathbf{n}$ is the unit outward normal vector to the boundary. Here dV is the volume element for the space, with units (length)^d and dS is its surface element, with

units (length)$^{d-1}$. This theorem gives an understanding to the physical meaning of divergence. If $\mathbf{v}$ represents a flow field of some material, then $\int_{\partial\Omega} \mathbf{v}\cdot\mathbf{n}dS$ represents the total flux of that material across the boundary out of the region Ω. The divergence of $\mathbf{v}$ represents the source density (sink, if negative) of the material at each point inside the domain Ω. The divergence theorem states that the net flux of material across the boundary is the cumulation (i.e., the integral) of all the sources in the interior. Consequently, if the net flux is outward, it is because there are more sources inside than there are sinks, and vice versa.

The divergence theorem is valid for one, two, or three dimensional regions Ω. In one-dimensional space, the divergence theorem is the same as the fundamental theorem of calculus

$$v(b) - v(a) = \int_a^b \frac{dv}{dx}dx. \tag{1.10}$$

The divergence theorem can be used to derive another important identity, namely,

$$\int_\Omega \nabla u dV = \int_{\partial\Omega} u\mathbf{n}dS, \tag{1.11}$$

where u is a scalar quantity. The proof follows from the identity

$$\int_{\partial\Omega} (u\mathbf{v})\cdot\mathbf{n}dS = \int_\Omega \nabla\cdot(u\mathbf{v})dV = \int_\Omega u\nabla\cdot\mathbf{v}dV + \int_\Omega \mathbf{v}\cdot\nabla u dV, \tag{1.12}$$

which, if $\mathbf{v}$ is a constant vector, reduces to

$$\mathbf{v}\cdot\int_{\partial\Omega} u\mathbf{n}dS = \mathbf{v}\cdot\int_\Omega \nabla u dV. \tag{1.13}$$

Since $\mathbf{v}$ is arbitrary, (1.11) follows.

Notice that the gradient operator produces a vector quantity from a scalar quantity, and the divergence operator produces a scalar quantity from a vector quantity. These operations can be composed to form the *Laplacian operator* $\nabla\cdot\nabla u$, which in Cartesian coordinates is

$$\nabla^2 u \equiv \nabla\cdot\nabla u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}. \tag{1.14}$$

The Laplacian operator has important physical significance, in that $\nabla^2 u$ is a measure of the "average" convexity of u. To see this, suppose that $x = x_0$ is a local minimum for the function $u(x)$. This means that for all x in some neighborhood of x_0, $u(x) \geq u(x_0)$. Thus, the second derivative of u with respect to x at x_0 is

$$\begin{aligned} \frac{\partial^2 u}{\partial x^2} &= \lim_{\Delta x\to 0} \frac{u(x_0+\Delta x) - 2u(x_0) + u(x_0-\Delta x)}{\Delta x^2} \\ &= \lim_{\Delta x\to 0} \frac{(u(x_0+\Delta x) - u(x_0)) + (u(x_0-\Delta x) - u(x_0))}{\Delta x^2} \geq 0. \end{aligned} \tag{1.15}$$

Similarly, $\frac{\partial^2 u}{\partial y^2} \geq 0$ and $\frac{\partial^2 u}{\partial z^2} \geq 0$. Therefore $\nabla^2 u \geq 0$ at a point which is a local minimum for u.

An important identity involving the Laplacian is

$$\int_{\partial\Omega} v\nabla u \cdot \mathbf{n} dS = \int_{\Omega} \nabla \cdot (v\nabla u) dV = \int_{\Omega} \nabla v \cdot \nabla u dV + \int_{\Omega} v\nabla^2 u dV, \tag{1.16}$$

which has use as an integration by parts formula. Setting $v = 1$, we find that

$$\int_{\Omega} \nabla^2 u dV = \int_{\Omega} \nabla \cdot \nabla u dV = \int_{\partial\Omega} \nabla u \cdot \mathbf{n} dS, \tag{1.17}$$

another very useful identity.

There are some other facts from vector calculus needed for this book, namely, what gradient, divergence, and Laplacian operators look like in different coordinate systems. The two most important coordinate systems here, other than Cartesian coordinates, are polar and spherical coordinates.

The relationship between polar and Cartesian coordinates is given by

$$x = r\cos\theta, \qquad y = r\sin\theta. \tag{1.18}$$

In polar coordinates, the Laplacian operator is

$$\nabla^2 u = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}. \tag{1.19}$$

The relationship between spherical and Cartesian coordinates is given by

$$x = r\cos\theta\sin\phi, \qquad y = r\sin\theta\sin\phi, \qquad z = r\cos\phi. \tag{1.20}$$

In spherical coordinates, the Laplacian operator is

$$\nabla^2 u = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial u}{\partial r}\right) + \frac{1}{r^2\sin^2\phi}\frac{\partial^2 u}{\partial \theta^2} + \frac{1}{r^2\sin\phi}\frac{\partial}{\partial\theta}\left(\sin\phi\frac{\partial u}{\partial\phi}\right). \tag{1.21}$$

1.1.3. Taylor's Theorem. A second result from calculus that is used often in this book is Taylor's theorem. For a function $f(x)$ of a single variable x which has N continuous derivatives in the vicinity of a point, say x_0, Taylor's theorem states that

$$f(x) = \sum_{j=0}^{N} \frac{d^j f(x_0)}{dx^j}\frac{(x-x_0)^j}{j!} + O\Big((x-x_0)^{N+1}\Big) \tag{1.22}$$

in some open neighborhood of $x = x_0$. It is possible to let N be large and even let $N \to \infty$, however, in this book we almost never (see Exercise 1.24) use Taylor's theorem for N larger than four.[2]

[2]The "Big-Oh" notation $O(x^n)$ carries the meaning that $f(x) = O(x^n)$ if

$$|f(x)| \le K|x|^n, \tag{1.23}$$

for some positive constant K in some nonzero neighborhood of $x = 0$. In words, $f(x)$ goes to zero at least as fast as the monomial x^n as x goes to zero.

1.2. Ordinary Differential Equations

1.2.1. First Order Equations. An ordinary differential equation specifies a relationship between the (time) derivative of some quantity u and its values through, say,

$$\frac{du}{dt} = f(u,t). \tag{1.24}$$

This equation is *autonomous* if f is independent of t, so that

$$\frac{du}{dt} = f(u). \tag{1.25}$$

Many of the problems discussed in this book are autonomous in time.

If u is a scalar quantity, the solution of equation (1.25) can be readily understood using graphical means, i.e., by plotting $\frac{du}{dt}$ vs. u. An example is shown in Figure 1.1.

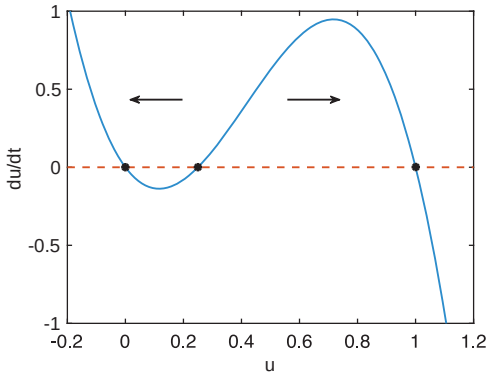


Figure 1.1. Plot of $\frac{du}{dt}$ vs. u for the bistable function $f(u) = au(1-u)(u-\alpha)$ with $\alpha = 0.25$, $a = 10$.

The first things to notice are the zeros of $f(u)$, i.e., the *equilibria*. For the example $f(u) = au(u-1)(\alpha - u)$, shown in Figure 1.1, the equilibria are at $u_0 = 0$, $u_0 = \alpha$, and $u_0 = 1$. Next, one can determine the direction of movement if u is not at an equilibrium. These are shown with arrows in Figure 1.1. For example, if $0 < u < \alpha$, $\frac{du}{dt} < 0$ indicating that u is decreasing there, while if $\alpha < u < 1$, $\frac{du}{dt} > 0$ so that u is increasing there. This is our first indication that $u_0 = 0$ and $u_0 = 1$ are *stable* equilibria, while $u_0 = \alpha$ is *unstable*.

The next thing to do is to linearize the equations about the equilibria. Linearization is a very important procedure by which one reduces a nonlinear equation to a linear equation.[3] It is a good idea to understand it thoroughly, because it is used often in this text.

[3] A linear operator is an operator L for which $L(u+v) = Lu + Lv$ and $L(au) = aL(u)$ for any scalar quantity a. In words, the operation on the sum of operands is the same as the sum of the operation on the operands and the operation on a scalar times an operand is the same as the scalar times the operation on the operand. A linear equation is one consisting only of the sum of linear operators. For example, both $\frac{\partial u}{\partial t}$ and $\frac{\partial^2 u}{\partial x^2}$ are linear operators.

The linearization of any differentiable function or operator $G(u)$ about u_0 is defined as

$$\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} G(u_0 + \epsilon U), \tag{1.26}$$

so the linearization of the differential equation (1.25) about any of its equilibria is

$$\lim_{\epsilon \to 0} \frac{\partial}{\partial \epsilon} \left(\frac{d}{dt}(u_0 + \epsilon U) - f(u_0 + \epsilon U) \right), \tag{1.27}$$

which reduces to

$$\frac{dU}{dt} = f'(u_0)U. \tag{1.28}$$

The solution of the linearized problem is the exponential function

$$U(t) = U_0 \exp(f'(u_0)t), \tag{1.29}$$

and it is now obvious that $U(t)$ grows if $f'(u_0) > 0$ and decays if $f'(u_0) < 0$. Hence, for our example here, the equilibria $u_0 = 0$ and $u_0 = 1$ are *linearly stable* while the equilibrium $u_0 = \alpha$ is unstable. This agrees with our graphical stability analysis.

Finally, it is noteworthy that the equation (1.25) is *separable* and can be rewritten as

$$\frac{du}{f(u)} = dt, \tag{1.30}$$

which, after integrating both sides of the equation, enables us to write

$$F(u) - F(u(0)) = t, \tag{1.31}$$

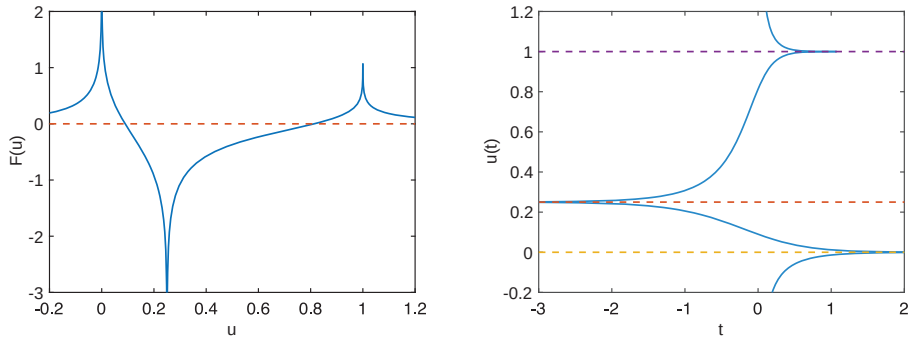
where $F(u) = \int^u \frac{du}{f(u)}$ and $u = u(0)$ at $t = 0$. In most situations, this is not a particularly useful representation of the solution, since analytical inversion of the function $F(u)$ to find $u(t)$ explicitly is usually impossible. However, through the wonders of Matlab, it is easy to graph this solution. That is, plot t as a function of u and then reverse the axes.

As an example, for the function $f(u) = au(1-u)(u-\alpha)$,

$$F(u) = \frac{1}{a\alpha(\alpha - 1)} \left(\alpha \ln(1-u) - \ln(|u-\alpha|) + (1-\alpha)\ln(u) \right), \tag{1.32}$$

a plot of which is shown in Figure 1.2(a), and then, reversing the axes gives the plot of $u(t)$ as a function of t, shown in Figure 1.2(b). This plot illustrates the fact that the solution has a different outcome as $t \to \infty$ depending on the initial condition. Clearly (as we already knew), if $0 < u(0) < \alpha$, $u(t) \to 0$ as $t \to \infty$, whereas, if $\alpha < u(0) < 1$, then $u(t) \to 1$ as $t \to \infty$. Only if $u(0) \equiv \alpha$ does $u(t) \to \alpha$ as $t \to \infty$, since then $u(t)$ is identically equal to α for all time.

This "trick" to invert functions by plotting the function and then reversing the axes is extremely useful and is used many times in this book.



(a) **(b)**

Figure 1.2. (a) Plot of $F(u)$ from (1.32)vs. u, and (b) plot of $u(t)$ as a function of t, for the function (1.32), with $\alpha = 0.25$.

1.2.2. Systems of first order equations. We now turn our attention to systems of first order equations, which can still be written in the form of (1.24) provided we recognize that u is a vector, rather than a scalar, quantity. The most important example for this text is when there are two unknown scalar functions $u(t)$ and $v(t)$ and the equations describing their evolution are in the form

$$\frac{du}{dt} = f(u, v), \tag{1.33}$$

$$\frac{dv}{dt} = g(u, v). \tag{1.34}$$

As with first order equations, a useful way to proceed is with a graphical, or *phase plane*, analysis. The first step of this analysis is to plot the *nullclines*, the curves in the $u - v$ plane along which either u or v do not change, i.e., $\frac{du}{dt} = 0$ or $\frac{dv}{dt} = 0$.

There are many examples of this procedure in this book, however, for purposes of illustration, let's look at solutions of the second order differential equation

$$\frac{d^2u}{dt^2} + f(u) = 0, \tag{1.35}$$

where $f(u) = au(1-u)(u-\alpha)$, the same function as used above. To write this equation as a first order system, we set $v = \frac{du}{dt}$, and then the equations are

$$\frac{du}{dt} = v, \tag{1.36}$$

$$\frac{dv}{dt} = -f(u). \tag{1.37}$$

The nullclines for this system are easily determined, being the line $v = 0$ for the u nullcline, and $f(u) = 0$ for the v nullclines, i.e., the lines $u = 0$, $u = \alpha$, and $u = 1$. These are shown plotted in Figure 1.3 as dashed lines.

The next step is to identify all the *critical points*, i.e., the points at which $\frac{du}{dt}$ and $\frac{dv}{dt}$ are simultaneously zero, hence, points of equilibrium. These are, of course, all the

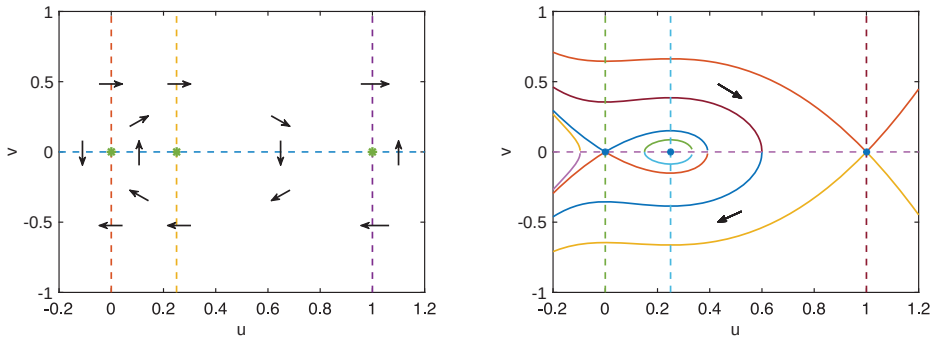


Figure 1.3. Phase portrait for the equations (1.36)–(1.37), showing nullclines and direction arrows in (a) and trajectories in (b).

intersections of the u and v nullclines. For this example, they are the points with $v = 0$ and $u = 0, \alpha$ and 1.

Next, we determine the direction of the flow in regions bounded by the nullclines. For this example, u increases if $v > 0$ and decreases if $v < 0$, while v increases if $0 < u < \alpha$ and decreases if $\alpha < u < 1$. It is also possible at this point to sketch a few typical trajectories by following the vector of flow directions. It quickly becomes apparent that the equilibria at $u = 0$ and $u = 1$ are saddle points, but the nature of the critical point at $u = \alpha$ cannot be decided by graphical means alone.

How do we classify critical points in general? The answer is using linear stability analysis, which proceeds as follows: one considers the linear system that approximates the full system of equations in the vicinity of a given point in phase space, and studies how that linear system behaves.

For the general system (1.33)–(1.34), the linearization is (in all of its glory)

$$\begin{aligned} \lim_{\epsilon\to 0}\frac{\partial}{\partial\epsilon}\begin{pmatrix} \frac{d}{dt}(u_0+\epsilon U) - f(u_0+\epsilon U, v_0+\epsilon V) \\ \frac{d}{dt}(v_0+\epsilon V) - g(u_0+\epsilon U, v_0+\epsilon V) \end{pmatrix} \\ = \frac{d}{dt}\begin{pmatrix} U \\ V \end{pmatrix} - \begin{pmatrix} \frac{\partial f_0}{\partial u} & \frac{\partial f_0}{\partial v} \\ \frac{\partial g_0}{\partial u} & \frac{\partial g_0}{\partial v} \end{pmatrix}\begin{pmatrix} U \\ V \end{pmatrix}. \end{aligned} \tag{1.38}$$

Consequently, the linearized system is

$$\frac{d}{dt}\begin{pmatrix} U \\ V \end{pmatrix} = A\begin{pmatrix} U \\ V \end{pmatrix}, \tag{1.39}$$

where the matrix A is the *Jacobian matrix* for this system,

$$A = \begin{pmatrix} \frac{\partial f_0}{\partial u} & \frac{\partial f_0}{\partial v} \\ \frac{\partial g_0}{\partial u} & \frac{\partial g_0}{\partial v} \end{pmatrix}, \tag{1.40}$$

and f_0, g_0 denote evaluation at the equilibria u_0, and v_0.

To find the solutions of the linearized system (1.39), we try an exponential solution of the form

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} U_0 \\ V_0 \end{pmatrix} \exp(\lambda t), \tag{1.41}$$

and determine that it must be that

$$(A - \lambda I) \begin{pmatrix} U_0 \\ V_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{1.42}$$

where I is the 2×2 identity matrix. The values of λ for which this equation has a solution are known as the eigenvalues of A, satisfying $\det(A-\lambda I) = 0$. For the 2×2 case at hand, this is the quadratic polynomial (also known as the *characteristic polynomial*) $\lambda^2 - \mathrm{Tr}(A)\lambda + \det(A) = 0$, with $\mathrm{Tr}(A)$ and $\det(A)$ representing the trace and determinant of A, respectively.

> For this text, it is assumed that you have some basic familiarity with linear algebra, including what is a matrix, and what is the determinant of a square matrix. Also, for Exercise 1.29, you will need to know how to use row reduction (i.e., Gaussian elimination) to reduce a matrix equation to an upper triangular system. As a reminder, the eigenvalues and eigenvectors of an $n \times n$ matrix A are defined as the numbers λ and vectors ϕ for which $A\phi = \lambda\phi$. Necessarily, the eigenvalues are roots of the characteristic polynomial $\det(A - \lambda I) = 0$, an nth order polynomial. Thus, there are always n eigenvalues, counting possible multiplicities. The *algebraic multiplicity* of an eigenvalue is the multiplicity with which that eigenvalue is a root of the characteristic polynomial, and the number of linearly independent eigenvectors associated with that eigenvalue is its *geometric multiplicity*. The geometric multiplicity is always less than or equal to the algebraic multiplicity, but never less than one. Thus, if the eigenvalues of a matrix are distinct, it has n linearly independent eigenvectors.

If $\det(A) < 0$, then there are two real roots, of opposite sign; the equilibrium is a *saddle point*. A typical phase portrait for a saddle point is shown in Figure 1.4. A saddle point has four special trajectories, two of which leave the saddle point, i.e., approach the saddle point in backwards time, as $t \to -\infty$, and two of which approach the saddle point as $t \to \infty$. These are identified as the unstable and stable manifolds, respectively. The unstable manifold has the same direction as the eigenvector of A corresponding to the positive eigenvalue, and the stable manifold has the same direction as the eigenvector of A corresponding to the negative eigenvalue.

If $\det(A) > 0$, there are four possible outcomes, depending on the sign of the discriminant, $\mathrm{disc} = \mathrm{Tr}(A)^2 - 4\det(A)$, and the sign of $\mathrm{Tr}(A)$. If $\mathrm{disc} > 0$, the two roots are real both with the same sign as $\mathrm{Tr}(A)$, and if $\mathrm{disc} < 0$, the two roots are a complex conjugate pair with the sign of the real part the same as the sign of $\mathrm{Tr}(A)$. If the roots are real, the equilibrium is called a *node*, and if the roots are complex, it is called a *spiral point*. If the real parts are positive, the equilibrium is unstable, while if they are negative, the equilibrium is stable. The intermediate case with $\mathrm{Tr}(A) = 0$ has neutral stability and is called a *center*. Thus, the four cases with $\det(A) > 0$ are stable node, stable spiral, unstable node, unstable spiral. These four are summarized in Table 1.1.

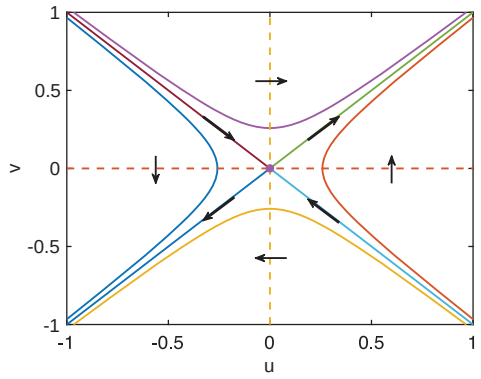


Figure 1.4. Typical phase portrait for a saddle point.

Table 1.1. Summary of stability criteria for $\det(A) > 0$.

Tr(A)\ disc	< 0 (complex roots)	> 0 (real roots)
> 0	unstable spiral	unstable node
< 0	stable spiral	stable node

Typical phase portraits for a stable node and a stable spiral are shown in Figure 1.5.

For the example problem (1.36)–(1.37), the Jacobian matrix is

$$A = \begin{pmatrix} 0 & 1 \\ -f'(u_0) & 0 \end{pmatrix}, \tag{1.43}$$

and since $\mathrm{Tr}(A) = 0$, the determining feature is the sign of $f'(u_0)$. In particular, the critical points at $u_0 = 0$ and $u_0 = 1$ are both saddle points, while the critical point at $u_0 = \alpha$ is a neutral center.

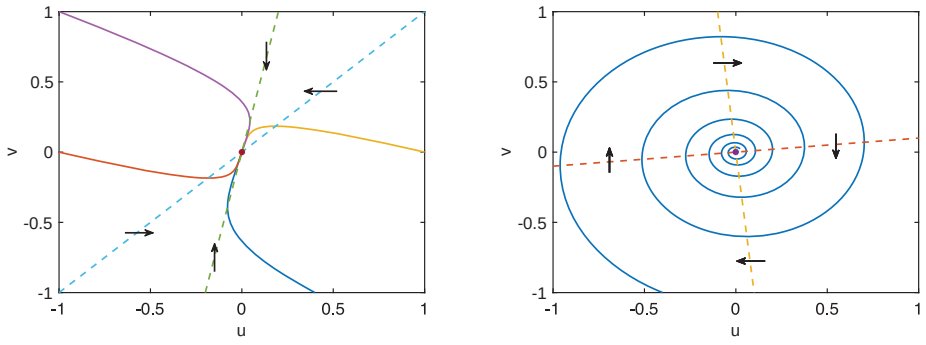


Figure 1.5. Typical phase portraits for (a) a stable node and (b) a stable spiral.

With this information, it is now usually possible to get an understanding of the behavior of the solutions. However, there are situations where this information is not sufficient to tell the whole story, and one such case is when there are isolated, closed orbits, i.e., limit cycles. We do not discuss limit cycles any further here.

Another important feature of these equations is that sometimes, but not often, it is possible to find expressions for some of the solution curves. For the current example problem, the slope of trajectories in the phase plane is given by

$$\frac{dv}{du} = -\frac{f(u)}{v}, \tag{1.44}$$

which is separable, yielding

$$-f(u)du = vdv. \tag{1.45}$$

Integrating both sides of this equation, we find that

$$F(u) + \frac{1}{2}v^2 = F(u_0), \tag{1.46}$$

where $F(u) = \int^u f(u)du$, and u_0, $v_0 = 0$ is a point on the trajectory. The expression (1.46) is called an integral of the motion. These trajectories can easily be plotted in the $u-v$ phase plane by plotting v as a function of u, wherever it is defined. Some examples are shown in Figure 1.3(b).

1.2.2.1. *Higher Order Systems.* Consider now the first order system of differential equations

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(\mathbf{u}), \tag{1.47}$$

where $\mathbf{u} \in R^n$, and $\mathbf{f} : R^n \to R^n$. Although we cannot readily draw it for $n > 2$, according to (1.47) the flow of $\mathbf{u}$ is in the direction of the vector field $\mathbf{f}$. Equilibria are any points $\mathbf{u_0}$ for which $\mathbf{f}(\mathbf{u_0}) = 0$. The stability of an equilibrium $\mathbf{u_0}$ is determined by the linearized system

$$\frac{d\mathbf{u}}{dt} = A\mathbf{u}, \tag{1.48}$$

where $A = \frac{\partial \mathbf{f}(\mathbf{u_0})}{\partial \mathbf{u}}$ is the Jacobian matrix. Solutions of (1.48) are linear combinations of exponentials $\exp(\lambda t)$ where λ are eigenvalues of A, roots of the characteristic polynomial $\det(A - \lambda I) = 0$. Then, an equilibrium solution $\mathbf{u_0}$ is linearly stable if all the eigenvalues have negative real part, but unstable if any of the eigenvalues have positive real part.

1.2.3. Numerical Simulation. It is extremely important to be able to numerically simulate differential equations efficiently and accurately. A number of algorithms to do this have been devised and implemented in Matlab, so that it is rarely necessary to write your own differential equation numerical solver. However, there are a few basic ideas of numerical simulation that are useful for this text, which are summarized here.

Suppose we wish to simulate the differential equation

$$\frac{du}{dt} = f(u). \tag{1.49}$$

We set $u^n = u(n\Delta t)$ where Δt is the discrete time step. The simplest discretization of the derivative is

$$\frac{du}{dt} \approx \frac{1}{\Delta t}(u^{n+1} - u^n), \tag{1.50}$$

and this suggests the algorithm

$$\frac{1}{\Delta t}(u^{n+1} - u^n) = f(u^n). \tag{1.51}$$

With this algorithm, we can determine u^{n+1} if u^n is known using

$$u^{n+1} = u^n + \Delta t f(u^n). \tag{1.52}$$

This method is called the *forward Euler method*.

Other possibilities exist. For example, we could try

$$\frac{1}{\Delta t}(u^{n+1} - u^n) = f(u^{n+1}), \tag{1.53}$$

called the *backward Euler method*. This method is implicit, because to find u^{n+1} given u^n, we must solve the equation

$$u^{n+1} - \Delta t f(u^{n+1}) = u^n, \tag{1.54}$$

and this typically involves additional approximations. In spite of this additional complexity, this method is recommended for *stiff* differential equations, in which one or more of the variables changes rapidly compared to others.

Both of these methods are first order accurate, meaning that the error of the numerical solution vanishes at a rate that is linear in Δt. A more accurate approximation would be to use a centered difference such as

$$\frac{1}{\Delta t}(u^{n+1} - u^n) = f(u^{n+\frac{1}{2}}), \tag{1.55}$$

but this is more complicated than the backward Euler method as it involves the additional unknown $u^{n+\frac{1}{2}}$. However, the approximation

$$\frac{1}{\Delta t}(u^{n+1} - u^n) = \frac{1}{2}(f(u^n) + f(u^{n+1})), \tag{1.56}$$

avoids this complexity. For this, finding u^{n+1} given u^n requires solution of the equation

$$u^{n+1} - \frac{\Delta t}{2} f(u^{n+1}) = u^n + \frac{\Delta t}{2} f(u^n), \tag{1.57}$$

which, from a computational perspective is no more difficult than solving (1.54). The additional advantage of this algorithm is that it is second order accurate, meaning that the error of the numerical solution vanishes at a rate that is quadratic in Δt.

1.2.4. Modeling Chemical Reactions. One of the important uses of differential equations, at least in this book, is to model the dynamics of chemical reactions. The two elementary reactions that are of most importance here are conversion between species, denoted

$$A \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} B, \tag{1.58}$$

called a first order reaction, and formation and degradation of a product from two component species, denoted

$$A + B \underset{\delta}{\overset{\gamma}{\rightleftharpoons}} C, \tag{1.59}$$

called a second order reaction.

The differential equations describing the first of these are

$$\frac{da}{dt} = \beta b - \alpha a, \qquad \frac{db}{dt} = -\beta b + \alpha a, \tag{1.60}$$

where $a = [A]$ and $b = [B]$, is the statement in math symbols that B is created from A at rate $\alpha[A]$ and A is created from B at rate $\beta[B]$. Of course, the total of A and B is a conserved quantity, since $\frac{d}{dt}(a + b) = 0$.

The second of these reactions is described by the three differential equations

$$\frac{da}{dt} = -\gamma ab + \delta c, \qquad \frac{db}{dt} = -\gamma ab + \delta c, \qquad \frac{dc}{dt} = \gamma ab - \delta c, \tag{1.61}$$

where $c = [C]$, which puts into math symbols the fact that C is created from the combination of A and B at a rate that is proportional to the product $[A][B]$, called the law of mass action. Notice that the units of γ are different ((time)$^{-1}$ (concentration)$^{-1}$) than those for first order reactions ((time)$^{-1}$). The degradation of C into A and B is a first order reaction. For this reaction there are two conserved quantities, namely $[A] + [C]$ and $[B] + [C]$.

An important example of reaction kinetics occurs in the study of epidemics, with the so-called SIR epidemic. Here S represents susceptible individuals, I represents infected individuals, and R represents recovered or removed individuals. We represent the disease process by the reaction scheme

$$S + I \overset{\alpha}{\longrightarrow} 2I, \quad I \overset{\beta}{\longrightarrow} R. \tag{1.62}$$

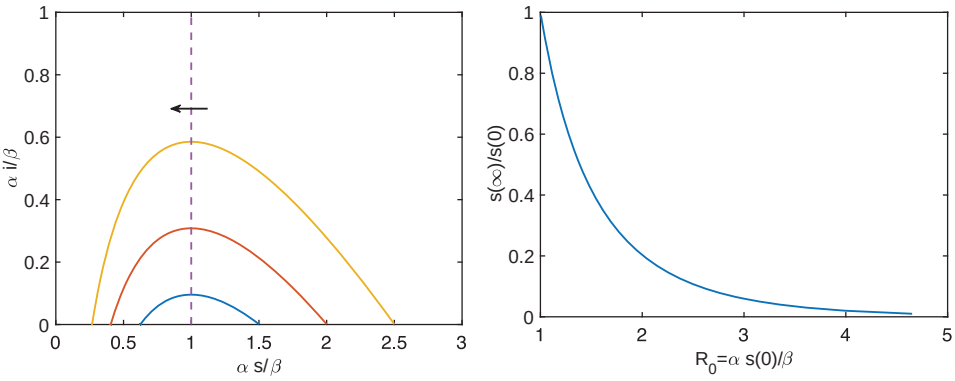
This implies that a susceptible individual can become infected following contact with an infected individual, and that infected individuals recover at an exponential rate.

Using the law of mass action described above, the deterministic differential equations for these reactions are

$$\frac{ds}{dt} = -\alpha si, \quad \frac{di}{dt} = \alpha si - \beta i, \quad \frac{dr}{dt} = \beta i. \tag{1.63}$$

Analysis of these equations is readily accomplished using the $s - i$ phase plane, shown in Figure 1.6(a). The lines $s = 0$ (vertical) and $i = 0$ (horizontal) are nullclines for s, while the vertical line $\frac{\alpha s}{\beta} = 1$, shown dashed, is the nullcline for i. s is everywhere decreasing, while i increases if $\frac{\alpha s}{\beta} > 1$ and decreases if $\frac{\alpha s}{\beta} < 1$.

It is clear from the phase portrait that there is a threshold phenomenon. That is, starting from a very small initial value of i, the infected population will grow only if $s(0)$, the initial susceptible population, is greater than $\frac{\beta}{\alpha}$. If the initial susceptible population is smaller than $\frac{\beta}{\alpha}$, the infection will die without spreading.



(a) **(b)**

Figure 1.6. (a) Phase portrait for the SIR dynamics with $\frac{\alpha s}{\beta}$ on the horizontal axis and $\frac{\alpha i}{\beta}$ on the vertical axis. (b) Plot of $\frac{s(\infty)}{s(0)}$, the fraction of susceptible population remaining following an epidemic as a function of $R_0 = \frac{\alpha s(0)}{\beta}$.

We can find a relationship between the initial and final susceptible population as follows. Note that

$$\frac{di}{ds} = -1 + \frac{\beta}{\alpha s}, \tag{1.64}$$

which is separable, so that integral curves are given by

$$i = s(0) - s + \frac{\beta}{\alpha} \ln\Big(\frac{s}{s(0)}\Big), \tag{1.65}$$

taking $i(0) = 0$. Thus,

$$0 = 1 - \frac{s(\infty)}{s(0)} + \frac{\beta}{\alpha s(0)} \ln\Big(\frac{s(\infty)}{s(0)}\Big), \tag{1.66}$$

specifies the relationship between the initial and final susceptible populations, $s(0)$ and $s(\infty)$. We cannot solve this explicitly for $s(\infty)$, but it is easy to solve for $\frac{\alpha s(0)}{\beta}$ as a function of $\frac{s(\infty)}{s(0)}$, i.e.,

$$\frac{\alpha s(0)}{\beta} = \frac{\ln\Big(\frac{s(\infty)}{s(0)}\Big)}{\frac{s(\infty)}{s(0)} - 1}, \tag{1.67}$$

and then use this to plot $\frac{s(\infty)}{s(0)}$ as a function of $\frac{\alpha s(0)}{\beta}$, as shown in Figure 1.6(b). The quantity $\frac{\alpha s(0)}{\beta}$ is usually referred to as R_0, the *basic reproduction number* in the epidemiology literature and is a measure of how easily spread a disease is. For example, $R_0 = 18$ for measles which implies that almost nobody will remain susceptible after a measles outbreak. For COVID-19, R_0 is estimated to be between 2 and 3, which implies that, without a vaccine, between 80% and 90% of the susceptible population will become infected before the epidemic runs its course.

Figure 1.6(b) confirms what we intuitively expect, namely, that the more easily spread a disease is (larger R_0), the more individuals will be infected during the course of an epidemic. It also suggests the effect of control measures, including vaccination which reduces the initial density of susceptibles $s(0)$, quarantine of infectives, which reduces the number of contacts of susceptibles with infectives (reduce α), social distancing and mask wearing, which reduces the infectiousness of a contact (reduce α), and shortening the recovery time with medical treatments (increase β).

The calculations to make these plots were done using the Matlab code SIR_pp.m.

1.3. Stochastic Processes

1.3.1. Decay Processes. Now that we have the review of differential equations behind us, we must face the fact that differential equation descriptions of biological processes are at best, highly idealized. This is because biological processes, and in fact many physical processes, are not deterministic, but noisy, or stochastic. This noise, or randomness, could be because, while the process actually is deterministic, we do not have the ability or the patience to accurately calculate the outcome of the process. For example, the flipping of a coin or the spin of a roulette wheel has a deterministic result, in that, if initial conditions were known with sufficient accuracy, an accurate calculation of the end result could be made. However, this is so impractical that it is not worth pursuing. Similarly, the motion of water vapor molecules in the air is by completely deterministic process (following Newton's Second Law, no quantum physics required) but determining the behavior of a gas by solving the governing differential equations for the position of each particle is completely out of the question.

There are other processes for which deterministic laws are not even known. This is because they are governed by quantum dynamics, having possible changes of state that cannot be described by a deterministic equation. For example, the decay of a radioactive particle and the change of conformation of a protein molecule, such as an ion channel, cannot, as far as we know, be described by a deterministic process. Similarly, the mistakes made by the reproductive machinery of a cell when duplicating its DNA (i.e., the mutations) cannot, as far as we currently know, be described by a deterministic process.

Given this reality, we are forced to come up with another way to describe interesting processes. And this is by keeping track of various *statistics* as time proceeds. For example, it may not be possible to exactly track the numbers of people who get the flu every year, but an understanding of how the average number changes over several years may be sufficient for health care policy makers. Similarly, with carbon dating techniques, it is not necessary to know exactly how many carbon-14 molecules there are in a particular painting at a particular time, but an estimate of an average or expected number of molecules can be sufficient to decide if the painting is genuine or a forgery.

1.3.1.1. *Probability Theory.* To make some progress in this way of describing things, we must define some terms. First, there must be some object that we wish to measure or quantify, also called a *random variable*, and the collection of all possible outcomes of this measurement is called its *state space*, or *sample space*. For example,

the flip of a coin can result in it landing with head or tail up, and these two outcomes constitute the state space. Similarly, an ion channel may at any given time be either open or closed, and this also constitutes its state space. The random variable could be a discrete or continuous variable taking on only integer values if it is discrete or a real valued number or vector if it is continuous.

The idea of a *probability* is intuitively clear, defined as the percentage of time a particular state occurs after a very large, i.e., infinite, number of observations of the state of the object have been made. Of course, this is a somewhat unsatisfactory definition, because no experiment can be repeated an infinite number of times. However, this definition is useful, even if it can never be checked. So, we easily understand that the probability that a coin toss will result in heads rather than tails should be one half. However, it is also clear that one should not expect the outcome of a large number of coin flips to result in *exactly* one half heads and one half tails. In fact, the probability that after 1,000 coin flips you will observe *exactly* 500 heads and 500 tails is rather small, at 0.025. (See Exercise 1.19). One goal of probability theory is to understand random variables by determining their *probability distribution functions* or *probability density functions*. For a discrete random variable, this would be some nonnegative quantity p_j which is the probability that the random variable is the value j, while for a continuous random variable x this is a nonnegative function $p(\xi)$ of the real variable ξ, having the feature that $P(x \in \Omega) = \int_\Omega p(\xi)d\xi$ is the probability that x is in the set Ω.

The two most famous and most commonly used probability distribution functions are the *binomial distribution* and the *normal distribution*. The binomial distribution answers the question of the probability of k successful outcomes in N independent experiments, denoted $p(k|N)$, where the probability of an individual successful outcome is p. The answer is

$$p(k|N) = \left(\begin{array}{c} N \\ \kappa \end{array} \right) p^k (1-p)^{N-k}, \tag{1.68}$$

where

$$\left(\begin{array}{c} N \\ \kappa \end{array} \right) = \frac{N!}{k!\,(N-k)!}. \tag{1.69}$$

The normal distribution (also known as the Gaussian distribution) is given by

$$p(\xi) = \frac{1}{\sqrt{2\pi}} \exp(\frac{-\xi^2}{2}), \tag{1.70}$$

and in this text, we also denote this as the distribution $\mathcal{N}(0,1)$.[4]

An important measure of a probability density function is called an *expected value*. Specifically, if $f(x)$ is some function of the random variable x, then the expected value of f is defined as

$$E(f) = \int f(\xi)p(\xi)d\xi, \tag{1.71}$$

[4]The notation $\mathcal{N}(\mu,\sigma^2)$ refers to the normal distribution with mean μ and variance σ^2, given by $p(\xi) = \frac{1}{\sqrt{2\pi}\sigma} \exp(\frac{-(\xi-\mu)^2}{2\sigma^2})$.

if x is a continuous random variable, or

$$E(f) = \sum f(k)p_k, \tag{1.72}$$

if it is a discrete random variable. In both cases, the integral or sum is taken over all possible values of the random variable.

This definition leads to three important measures of a random variable, namely its *mean value*, its *variance* and its *standard deviation*. The mean value of x is the expected value of x, i.e., the expected value of $f(x) = x$, $E(x)$. The variance of x is the expected value of the square of x minus its mean, i.e.,

$$\sigma^2 = \mathrm{var}(x) = E((x - E(x))^2) = E(x^2) - (E(x))^2. \tag{1.73}$$

The standard deviation is $\sigma = \sqrt{\mathrm{var}(x)}$. Some obvious identities are that $E(ax) = aE(x)$ for any scalar a, and $\mathrm{var}(ax) = a^2\,\mathrm{var}(x)$ for any scalar a.

It is a general fact that if y and z are independent random variables (meaning that their joint probability density function[5] $p(\xi, \eta)$ satisfies $p(\xi, \eta) = p_y(\xi)p_z(\eta)$, where $p_y(\xi)$ and $p_z(\eta)$ are the individual probability density functions for y and z, respectively), then

$$\mathrm{var}(y + z) = \mathrm{var}(y) + \mathrm{var}(z). \tag{1.74}$$

(This is a statement that you, the reader, should verify. See Exercise 1.13.)

It is straightforward to verify that for the normal distribution, $E(x) = 0$ and $E(x^2) = 1$. One can also use Matlab[6] to determine that for a random variable x that is normally distributed, the probability that x is within one standard deviation of zero is

$$P(-1 < x < 1) = \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} \exp(\frac{-\xi^2}{2}) d\xi = 0.6827. \tag{1.77}$$

One can also calculate that $P(-1.96 < x < 1.96) = 0.95$ and $P(-2.576 < x < 2.576) = 0.99$. In the field of statistics, these are referred to as the 95% and 99% confidence intervals, respectively. In words, this means that 68% of the time a normally distributed random variable will be within one standard deviation of the mean, while 95% or 99% percent of the time it will be within 1.96 or 2.58 standard deviations, respectively, of the mean.

The normal distribution shows up often, as described by the

Central Limit Theorem. Suppose $m_1, m_2, \ldots,$ are independent, identically distributed, random variables, with mean $\mu = E(m_i)$ and variance $\sigma^2 = \mathrm{var}(m_i)$. Then, the

[5]The joint probability density function has exactly the meaning one would expect, namely that the probability that the random variable ordered pair (y, z) is in some set Ω is $P((y, z) \in \Omega) = \int_\Omega p(\xi, \eta) d\xi d\eta$.

[6]The Matlab function $\mathrm{erf}(x)$ is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-\xi^2) d\xi, \tag{1.75}$$

and this is known as the *error function*. It follows that for a normally distributed random variable x,

$$P(-X < x < X) = \frac{1}{\sqrt{2\pi}} \int_{-X}^{X} \exp(-\frac{\xi^2}{2}) d\xi = \frac{2}{\sqrt{\pi}} \int_0^{\frac{X}{\sqrt{2}}} \exp(-\xi^2) d\xi = \mathrm{erf}(\frac{X}{\sqrt{2}}). \tag{1.76}$$

random variable

$$s_N = \sum_{j=1}^{N} m_j \tag{1.78}$$

with N large, is approximately normally distributed with mean $\mu_N = N\mu$ and variance $\sigma_N^2 = N\sigma^2$. In other words, the distribution for x_N is well approximated by

$$f_N(s) = \frac{1}{\sqrt{2\pi}\sigma_N} \exp\Big(- \frac{(s-\mu_N)^2}{2\sigma_N^2}\Big) = \frac{1}{\sqrt{2\pi N}\sigma} \exp\Big(- \frac{(s-N\mu)^2}{2N\sigma^2}\Big). \tag{1.79}$$

This distribution is also denoted as $\mathcal{N}(N\mu, N\sigma^2)$.

In a similar vein, the random variable,

$$y_N = \frac{1}{N} \sum_{j=1}^{N} m_j \tag{1.80}$$

with N large, is approximately normally distributed with mean $\mu_N = \mu$ and variance $\sigma_N^2 = \frac{1}{N}\sigma^2$, i.e., the distribution for y_N is well approximated by

$$f_N(y) = \frac{\sqrt{N}}{\sqrt{2\pi}\sigma} \exp\Big(- \frac{N(y-\mu)^2}{2\sigma^2}\Big), \tag{1.81}$$

denoted $\mathcal{N}(\mu, \frac{1}{N}\sigma^2)$.

A *stochastic process* or a random process is a collection of random variables, parametrized by some ordered index set. If the index set is continuous time it is a continuous time process, whereas if the index set consists of discrete points (of time) it is a discrete time stochastic process. For example, consider a carbon atom that can be in several isotopic states, including carbon-14, carbon-13 or carbon-12.[7] The state of carbon-14 can change at any time to nitrogen-14 by the process known as radioactive decay (or beta decay) in which one of the neutrons of carbon-14 becomes a proton. Or consider the state of an ion channel which can flip back and forth between its two states, open and closed, at different times.

Since the transitions between states may occur at any time, and those times cannot be known exactly, the best we can do is to track the probability that the object is in a particular state as a function of time. To study these processes, we need one more assumption, that the probability of the object changing its state in some small time interval is independent of how long it has been in its current state. Such a process is called a *Markov process*. For example, the probability that a carbon-14 atom decays to a nitrogen-14 atom in the next 30 seconds is completely independent of how long the carbon-14 molecule has been in existence. Similarly, suppose you have flipped a coin 1000 times and amazingly, it has come up heads every time. What is the probability that it will be heads on the 1001st time? Answer, one half: The probability of landing heads is independent of the history of previous flips (if it is a fair coin and unless someone is cheating).

[7] A carbon atom has 6 protons and 6 (carbon-12), 7 (carbon-13), or 8 (carbon-14) neutrons. Carbon-12 and carbon-13 are both stable. Nitrogen has 7 protons and 7 neutrons.

Let's consider first a discrete state space, continuous time process. To track the probability that the object is in a given state, we make use of the *law of total probability*. In words, the probability that the object is in a particular state at some time $t + \Delta t$ is equal to the probability that it was in the same state at time t *times* the probability of not switching out of that state in time Δt, *plus* the probability that it was *not* in that state at time t *times* the probability of switching into that state in time Δt. In mathematical language, if we denote the probability that the object is in state j at time t by $p_j(t)$, and the probability of switching from state i to state j in time Δt by $s_{ji}(\Delta t)$, then

$$p_j(t + \Delta t) = p_j(t)(1 - \sum_{i \neq j} s_{ij}(\Delta t)) + \sum_{i \neq j} p_i(t) s_{ji}(\Delta t). \tag{1.82}$$

Now we also make the definition that

$$\lambda_{ji} = \lim_{\Delta t \to 0} \frac{s_{ji}(\Delta t)}{\Delta t}, \tag{1.83}$$

then we can divide the equation (1.82) by Δt and take the limit $\Delta t \to 0$ and find

$$\frac{dp_j}{dt} = -\sum_{i \neq j} \lambda_{ij} p_j(t) + \sum_{i \neq j} \lambda_{ji} p_i(t). \tag{1.84}$$

It is useful to notice that (1.84) can be written in matrix form. That is, let $\mathbf{p}$ be the vector with elements $p_j(t)$. Then,

$$\frac{d\mathbf{p}}{dt} = A\mathbf{p}, \tag{1.85}$$

where the matrix A has elements $a_{j,k} = \lambda_{j,k}$ if $k \neq j$ and $a_{j,j} = -\sum_{k \neq j} \lambda_{k,j}$. Then, the matrix A has zero column sums, $\sum_j a_{j,k} = 0$ for all k.

As a first example, let's think about the obviously stochastic process of catching fish. Here the random variable is the number of fish caught at time t since you began fishing, and we suppose that at any given time the probability of catching a fish in a small window of time Δt is $\alpha \Delta t$, where α is a constant independent of time. This assumption is obviously not correct, as any fisherman can attest, but it is useful for this example. Following the above arguments, the equations describing the dynamics of $p_j(t)$, the probability of having caught exactly j fish at time t, are

$$\frac{dp_j}{dt} = -\alpha p_j(t) + \alpha p_{j-1}(t), \tag{1.86}$$

for $j = 0, 1, 2, \ldots$, and $p_{-1} = 0$. This follows since there are only two ways that the state can change, namely from state $j - 1$ to state j by catching a fish, or from state j to state $j + 1$, also by catching a fish. And since the probability of catching a fish is independent of how many fish have been caught previously (in a big enough body of water), the rates for both of these events are the same, namely α.

How long will it be to catch the first fish? Since

$$\frac{dp_0}{dt} = -\alpha p_0(t), \tag{1.87}$$

we find that

$$p_0(t) = \exp(-\alpha t), \tag{1.88}$$

and the probability that you have caught at least one fish by time t is $1 - p_0(t) = 1 - \exp(-\alpha t)$. The function $-\frac{dp_0}{dt} = \alpha p_0(t) = \alpha \exp(-\alpha t)$ represents the *probability density function* (pdf) for the fish catching event and the function $1 - \exp(-\alpha t)$ is called the *cumulative distribution function* (cdf) because it represents the probability of having caught at least one fish in the time interval $[0, t]$. Accordingly, this event is said to be *exponentially distributed* and is identified as an exponential process.. The expected value of an exponential distribution is

$$E(t) = \int_0^\infty \tau\alpha \exp(-\alpha\tau) d\tau = \frac{1}{\alpha}, \tag{1.89}$$

and the expected value of t^2 is

$$E(t^2) = \int_0^\infty \tau^2\alpha \exp(-\alpha\tau) d\tau = \frac{2}{\alpha^2}, \tag{1.90}$$

so that the variance is $\mathrm{var}(t) = E(t^2) - E(t)^2 = \frac{1}{\alpha^2}$.

What more can we learn from these equations? We can check that total probability is conserved, since

$$\frac{d}{dt}\sum_{j=0}^{\infty} p_j = -\alpha \sum_{j=0}^{\infty} p_j(t) + \alpha \sum_{j=1}^{\infty} p_{j-1}(t) = 0. \tag{1.91}$$

Of course, $\sum_{j=0}^{\infty} p_j = 1$, since at time zero you have caught no fish, $p_0(0) = 1$ and $p_j(0) = 0$ for all $j > 0$. We can also find that the expected value of j satisfies

$$\begin{aligned} \frac{d}{dt}E(j) &= \frac{d}{dt}\sum_{j=0}^{\infty} j p_j = -\alpha \sum_{j=0}^{\infty} j p_j(t) + \alpha \sum_{j=1}^{\infty} j p_{j-1}(t) \\ &= -\alpha \sum_{j=0}^{\infty} j p_j(t) + \alpha \sum_{j=0}^{\infty} (j+1) p_j(t) = \alpha, \end{aligned} \tag{1.92}$$

so that $E(j) = \alpha t$. Similarly,

$$\begin{aligned} \frac{d}{dt}E(j^2) &= \frac{d}{dt}\sum_{j=0}^{\infty} j^2 p_j = -\alpha \sum_{j=0}^{\infty} j^2 p_j(t) + \alpha \sum_{j=1}^{\infty} j^2 p_{j-1}(t) \\ &= -\alpha \sum_{j=0}^{\infty} j^2 p_j(t) + \alpha \sum_{j=0}^{\infty} (j+1)^2 p_j(t) \\ &= \alpha \sum_{j=0}^{\infty} (2j+1) p_j(t) = 2\alpha E(j) + \alpha, \end{aligned} \tag{1.93}$$

so that $E(j^2) = \alpha^2 t^2 + \alpha t$. It follows that $\mathrm{var}(j) = \alpha t$.

These equations can be solved analytically (in Exercise 1.24), finding

$$p_j(t) = \frac{(\alpha t)^j}{j!} \exp(-\alpha t), \tag{1.94}$$

and plots of the first five of these ($j = 0, 1, 2, 3, 4$) are shown in Figure 1.7. It is apparent from these plots and easily verified that the maximum of $p_j(t)$ occurs at $\alpha t = j$. Notice

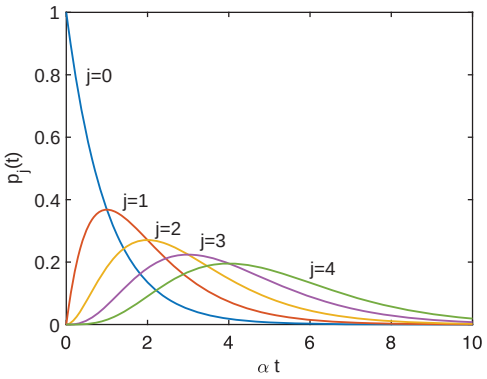


Figure 1.7. Plots of the Poisson distribution $p_j(t)$ for $j = 0, 1, 2, 3, 4$.

that α is a rate, having units of 1/time. Consequently, $\frac{1}{\alpha}$ is called the *time constant* for this process. Notice, also, that $p_4(t)$ looks similar to a normal distribution, which, according to the central limit theorem, it should. In fact, according to the central limit theorem, $p_j(t) \approx \mathcal{N}(\frac{j}{\alpha}, \frac{j}{\alpha^2})$ for j large.

This process is called a *Poisson process*, named after its inventor, French mathematician Siméon Denis Poisson, and not after the French meaning of his name (*poisson* means fish in French).

Now, let's consider the conversion of one chemical species to another. This we describe by a chemical reaction in which some species, S, is converted to a different species I at rate α, and this is denoted as

$$S \xrightarrow{\alpha} I. \tag{1.95}$$

For example, consider the decay of carbon-14 to nitrogen-14.

Suppose that at some time denoted $t = 0$, there is one molecule of carbon-14. What is the probability as a function of time that the carbon-14 molecule has not undergone decay by time $t > 0$? If we let $p_j(t)$ denote the probability of having exactly j molecules of carbon-14, then, following the same argument just given, the probability that there is exactly one carbon-14 molecule at time t satisfies the differential equation

$$\frac{dp_1}{dt} = -\alpha p_1, \tag{1.96}$$

where α is the *rate of decay*, known to be 1.21×10^{-4} per year for carbon-14. Since $p_1(0) = 1$, the solution of this equation is

$$p_1(t) = \exp(-\alpha t). \tag{1.97}$$

The function $1 - p_1(t) = 1 - \exp(-\alpha t)$ represents the probability that the decay event has taken place by time t, (i.e., the cdf) and $-\frac{dp_1}{dt} = \alpha p_1 = \alpha \exp(-\alpha t)$ represents the probability distribution function for the decay event. Accordingly, the decay event is *exponentially distributed*. As we learned from our fishing experience, the expected value of an exponential distribution is $E(t) = \frac{1}{\alpha}$, and the variance is $\text{var}(t) = \frac{1}{\alpha^2}$.

Now suppose we start with many, say n, carbon-14 molecules at time $t = 0$. What is the probability that we still have exactly n molecules at some time $t > 0$? Since the behavior of each atom of carbon-14 is independent of how many particles there are, it must be that $p_n(t)$ satisfies

$$p_n(t) = (p_1(t))^n = \exp(-n\alpha t). \tag{1.98}$$

This follows from the law of independence, namely, that the probability of a collection of independent events is the product of the probabilities of the individual events. Consequently,

$$\frac{dp_n}{dt} = -n\alpha p_n. \tag{1.99}$$

Now, suppose we start at time $t = 0$ with n molecules, but want to know the probability of having exactly $k < n$ molecules at some later time. For example, when will there be no molecules left, or, in mathematical language, what is $p_0(t)$?

Following the arguments given above (see (1.84)), we can write a differential equation for $p_k(t)$ as

$$\frac{dp_k}{dt} = (k+1)\alpha p_{k+1} - k\alpha p_k. \tag{1.100}$$

This follows since the state with k molecules is entered from the state with $k+1$ molecules at rate $(k+1)\alpha$ when one of the $k+1$ molecules decays and the state with k molecules is left at rate $k\alpha$ when one of the k molecules decays.

The equation for p_k is one of $n+1$ equations (since $k = 0, 1, \ldots, n$), and while it is easy to determine $p_n(t)$ and not too hard to find $p_{n-1}(t)$, finding the general solution for $p_k(t)$ is more complicated (see Exercise 1.30). It is not difficult to simulate this system of equations, and this simulation is done in the Matlab code exponential_decay_via_Gillespie.m.

We can, however, get some information about this process without solving the full system of equations. For example, summing all of the equations, we find that

$$\frac{d}{dt}\sum_{k=0}^{n} p_k(t) = 0. \tag{1.101}$$

In other words, the total probability does not change, that is, total probability is a conserved quantity. We can also find the expected value of k, which is defined as

$$u = E(k) \equiv \sum_{k=0}^{n} k p_k(t), \tag{1.102}$$

by multiplying equation (1.100) by k and adding all the equations together. We find

$$\frac{du}{dt} = -\alpha u, \tag{1.103}$$

and again, we find one of our favorite first order differential equations.

Notice that u is not an integer, even though it describes a process that only takes on integer values. In fact, u is a continuous function of time, and since the equation (1.103) is linear in u, the function u can be scaled to be the percentage of remaining molecules, or the concentration of molecules in a specified volume.

This equation has an important interpretation. If we were to watch a single particle to see when it decays, it would be extremely hard, essentially impossible, to predict when decay will occur. However, if we start with a large number of particles and record the times that each individual particle decays and then make a plot of the percentage of particles remaining at time t, we will see a curve that is quite close to the curve $\exp(-\alpha t)$. Further, if we make a histogram of the times that the individual decay events took place, it would be quite close to the curve $-\frac{du}{dt} = \alpha \exp(-\alpha t)$. This is to say that the decay times are distributed along the real number line, and the density with which they are distributed is the pdf.

Now let's use this information to numerically simulate this stochastic decay process. Suppose we start with n particles. When will the first decay event occur? We do not know this exactly, but we do know that this next reaction time should be distributed like $\exp(-n\alpha t)$, since the probability that decay occurs after time t is $\exp(-n\alpha t)$. The function $1 - \exp(-n\alpha t)$ is the cdf for the probability that at least one decay event has taken place in the interval $[0, t]$. In fact, the cdf $r = 1 - \exp(-n\alpha t)$ is a uniformly distributed random variable (see Exercise 1.28). So, we pick a uniformly distributed random number R between zero and one, $0 < R < 1$ (rather, let Matlab pick a number for you), and then take the next decay time increment δt_n to be the time at which the cdf is equal to $1 - R$, i.e., we invert the cdf to be such that

$$R = \exp(-n\alpha \delta t_n), \tag{1.104}$$

or

$$\delta t_n = \frac{-1}{n\alpha} \ln R. \tag{1.105}$$

Record this time increment, and repeat, next for $n - 1$ particles, and so on until all n particles have decayed.

The Matlab code that carries out this algorithm, called the *Gillespie algorithm*, or *next reaction time algorithm*, is titled exponential_decay_via_Gillespie.m, and an example of a simulation starting with 25 initial particles is shown in Figure 1.8.

There is an interesting and important observation to make about this decay process, and that is that the process always terminates in finite time, although the exponential curve that approximates it does not. To determine the expected time for the process to terminate, notice that the expected time to go from k to $k - 1$ molecules is

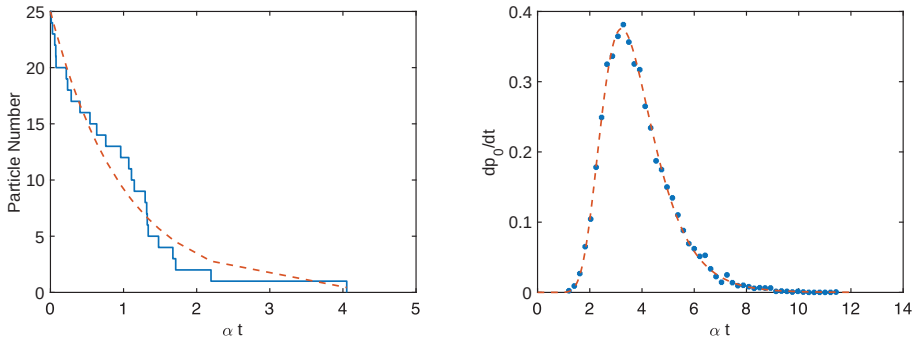
$$t_k = \frac{1}{k\alpha}, \tag{1.106}$$

and consequently, the expected time to go from n molecules to extinction is

$$E(t) = \frac{1}{\alpha} \sum_{k=1}^{n} \frac{1}{k}. \tag{1.107}$$

Unfortunately, there is no closed form formula for this expression, although for large n it is well approximated by

$$\sum_{k=1}^{n} \frac{1}{k} = \ln(n) + \gamma + O(\frac{1}{n}), \tag{1.108}$$



(a) **(b)**

Figure 1.8. (a) Result of Gillespie simulation of decay (solid curve) compared to the function $N\exp(-\alpha t)$ (dashed curve), plotted as a function of αt, for $N = 25$ initial particles. (b) Histogram of extinction times for a simulation of 10,000 trials for decay of $N = 25$ initial particles, normalized to have total area one (shown with dots), compared with the function $\frac{dp_0}{dt}$ (dashed) coming from the numerical solution of the system (1.100).

where γ is the Euler–Mascheroni constant $\gamma = 0.57721\cdots$. However, the validity of this formula can be verified by numerical simulation using the code exponential_decay_via_Gillespie.m. An example of this verification is shown in Figure 1.8(b) where a histogram of extinction times following 10,000 trials starting with 25 particles is shown. For this plot, the mean extinction time was computed to be 3.7969 and the theoretical value from (1.107) is $\alpha E(t) = 3.8160$.

This example of decay is only one of several possibilities. The death process could be *bimolecular*, as in

$$S + S \xrightarrow{\beta} S. \tag{1.109}$$

The differential equation describing this reaction is

$$\frac{ds}{dt} = -\beta s^2, \tag{1.110}$$

and for a stochastic simulation, one takes

$$\lambda_{k-1,k} = \frac{k(k-1)}{2}\frac{\beta}{\text{vol}}. \tag{1.111}$$

The motivation for this formula is that in a container with k S particles, the number of ways that a pair of particles can interact is $\frac{k(k-1)}{2}$. Then, the rate of interaction must be modified to take into account that the units of β for this reaction are $(\text{concentration})^{-1}\cdot(\text{time})^{-1}$, whereas $\lambda_{k-1,k}$ needs to be in units of $(\text{time})^{-1}$. The factor vol has units of volume, and is the conversion factor between concentration and particle numbers, so that the factor $\frac{\beta}{\text{vol}}$ has the correct units. Specifically, since concentration is number per volume, then the volume vol is number per concentration.

The spread of an infection might be described by the reaction

$$S + I \xrightarrow{\gamma} I + I, \tag{1.112}$$

in which case

$$\lambda_{k-1,k} = k(N-k)\frac{\gamma}{\text{vol}}, \tag{1.113}$$

assuming the total number of particles $S + I$ is the constant N.

1.3.2. Several Reactions. In the example of particle decay there was only one reaction possible. However, this is not typical as most chemical reactions involve a range of possible reactions. For example, suppose a particle (like a bacterium) may reproduce at some rate or it may die at a different rate. The question addressed here is how to do a stochastic simulation of this process.

Suppose the state S_j can transition to the state S_k at rate λ_{kj}. To do a stochastic simulation of this process, we must decide when the next reaction takes place and which reaction it is that takes place.

To decide when the next reaction takes place, we use the fact that the probability that the next reaction has taken place by time t is 1 minus the probability that the next reaction has not taken place by time t. Furthermore, the probability that the reaction from state j to state k has not taken place by time t is $\exp(-\lambda_{kj}t)$. So, the probability that no reaction has taken place by time t (since these reactions are assumed to be independent) is

$$\prod_k \exp(-\lambda_{kj}t) = \exp(-\sum_k \lambda_{kj}t). \tag{1.114}$$

It follows that the cdf for the next reaction is

$$1 - \exp(-\sum_k \lambda_{kj}t) = 1 - \exp(-rt), \tag{1.115}$$

where $r = \sum_k \lambda_{kj}$. In other words, the next reaction is an exponential process with rate r.

Next, the probability that the next reaction is the ith reaction $S_j \to S_i$ is

$$p_{ij} = \frac{\lambda_{ij}}{\sum_k \lambda_{kj}} = \frac{\lambda_{ij}}{r}. \tag{1.116}$$

To be convinced of this, apply the results of Exercise 1.26 to the case where either the $S_j \to S_i$ reaction occurs first or another reaction occurs first.

With these facts in hand, as we did above, we pick the next reaction time increment to be

$$\delta t = \frac{-1}{r}\ln R_1, \tag{1.117}$$

where $0 < R_1 < 1$ is a uniformly distributed random number. Next, to decide which of the reactions to implement, construct the vector $x_k = \frac{1}{r}\sum_{i=1}^{k}\lambda_{ij}$, the scaled vector of cumulative sums of λ_{ij}. Notice that the vector x_k is ordered with $0 \le x_1 \le x_2 \le \cdots \le x_N = 1$, where N is the total number of states. Now, pick a second random number R_2, uniformly distributed between zero and one, and pick the next reaction to be $S_j \to S_k$ where

$$k = \min_j\{R_2 \le x_j\}. \tag{1.118}$$

With this formula, the proportion of the times the kth reaction is picked is $\frac{\lambda_{kj}}{r}$, as it must be.

Let's consider the specific example of the death and birth process with

$$S \xrightarrow{\alpha} I, \quad S \xrightarrow{\beta} 2S. \tag{1.119}$$

The rate at which one particle is removed is

$$\lambda_{k-1,k} = k\alpha, \tag{1.120}$$

and the rate at which one particle is added is

$$\lambda_{k+1,k} = k\beta. \tag{1.121}$$

The deterministic differential equation governing the mean of the population is

$$\frac{ds}{dt} = (\beta - \alpha)s, \tag{1.122}$$

and if $\alpha > \beta$, the population will go extinct. However, in contrast to the pure death process, the population will not decay monotonically. The equations governing the probability of having k particles at time t are

$$\frac{dp_k}{dt} = -(\alpha k + \beta k)p_k + \alpha(k+1)p_{k+1} + \beta(k-1)p_{k-1}, \tag{1.123}$$

and this is an infinite system of equations, since there is no a priori bound on the number of particles at any given time. Consequently, to solve this system numerically it must be truncated at some relatively large value of k. (For an analytical solution, see Exercise 9.9.)

The stochastic simulation for this process uses the Matlab code stochastic_birth_death.m and an example of a sample trajectory is shown plotted in Figure 1.9.

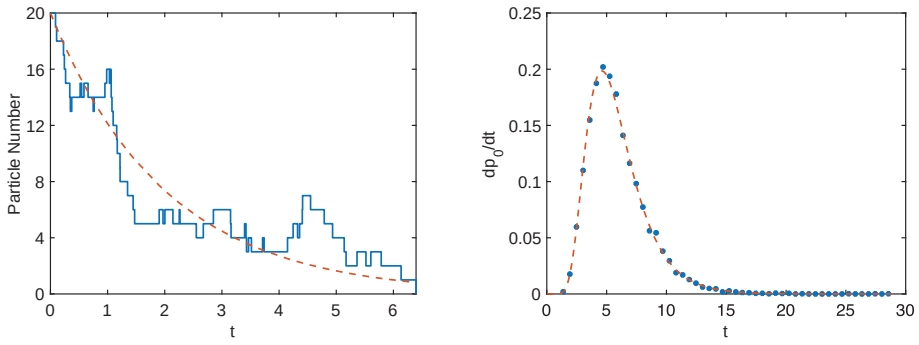


Figure 1.9. (a) Trajectory of a simulated birth-death process with death rate $\alpha = 1$ and birth rate $\beta = 0.5$ with the curve $N\exp(-(\alpha-\beta)t)$ (dashed line). (b) Histogram of extinction times for a simulation of 10,000 trials for decay of $N = 20$ initial particles, normalized to have total area one (shown with dots), compared with the function $\frac{dp_0}{dt}$ (dashed line) coming from the solution of the system (1.123).

1.3.3. Multiple Species. This method of simulation and analysis generalizes readily to the situation where there are multiple species and multiple reactions. At any given time, the state vector is the vector of integers $\mathcal{S} = (n_1, n_2, \ldots, n_K)$, and there are J reactions with rates r_j that depend on the state of the system $\mathcal{S}$. For each reaction there is a change in the state vector $c(j,k)$, meaning that if reaction j occurs, the kth integer n_k changes by the amount $c(j,k)$.

As an example, consider the SIR reactions

$$S + I \xrightarrow{\alpha} 2I, \quad I \xrightarrow{\beta} R. \tag{1.124}$$

Here S represents susceptible individuals in a population, I represents the infected and contagious individuals, and R represents those individuals who are removed and no longer contagious. The deterministic differential equations for these reactions are given by (1.63), however, as we all know from experience with COVID-19, the evolution of an epidemic is highly stochastic.

An interesting question to ask is how many individuals have been infected and how many susceptibles remain (or survive) after an infection has run its course, and we can address this question using a stochastic simulation. The setup for this stochastic simulation is straightforward. The state space is identified by the three integers n_s, n_i, and n_r, and the two reactions are at rates

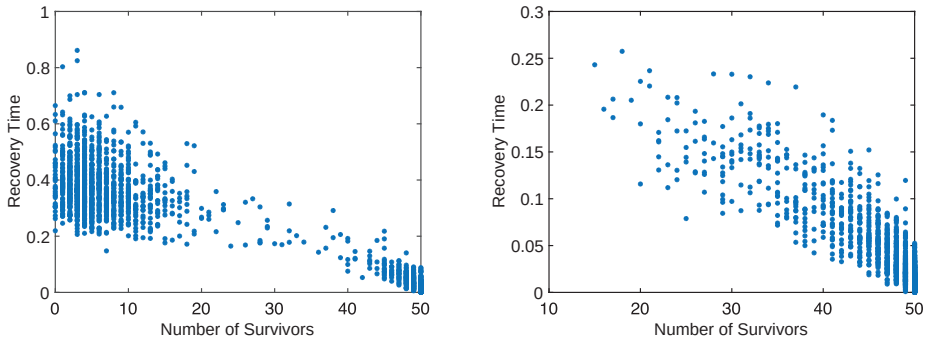
$$r_1 = \alpha n_s n_i, \quad r_2 = \beta n_i, \tag{1.125}$$

and the change matrix $C = c(j,k)$ is

$$C = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}. \tag{1.126}$$

This is easily implemented in Matlab code and in fact, the code that does this is titled stochastic_SIR.m.

Scatter plots of recovery times vs. number of survivors for the SIR stochastic process shown in Figure 1.10 are surprising, and are certainly different than what is predicted by the deterministic model. (Recovery time refers to the first time at which there are no more infected individuals.) The deterministic model predicts a unique outcome (recall (1.67)), with an epidemic spreading if $R_0 = \frac{\alpha s(0)}{\beta} > 1$ and not spreading if $R_0 = \frac{\alpha s(0)}{\beta} < 1$. However, in Figure 1.10(a), where $R_0 = 2.5$, the results of the stochastic simulation show a biphasic outcome, with many of the trials, as expected, having a large epidemic with few survivors and long recovery times, but also with a significant number of trials with little spread of the infection, a large percentage of survivors, and a short recovery time. Similarly, in Figure 1.10(b), where $R_0 = 0.9$, most of the trials result in a short-lived epidemic with a high percentage of survivors. However, there are nonetheless quite a few trials showing a substantial epidemic with few survivors and long recovery times, noticeably different than the prediction of the deterministic model.



(a) **(b)**

Figure 1.10. Scatter plot of recovery times vs. number of survivors for 2000 trials of the SIR stochastic process starting with $N = 50$ S individuals and one I (infected) initially, running the process until there are no I individuals, with parameters $\alpha = 1$, and (a) $\beta = 20$ ($R_0 = 2.5$), (b) $\beta = 55$ ($R_0 = 0.9$).

There is an important lesson to be learned here. In particular, the prediction of the deterministic model is not necessarily correct. For this problem, if $R_0 > 1$, it is not *necessarily* correct that there will be a substantial spread of the infection, and if $R_0 < 1$ it does not *necessarily* mean that there will no spread of the infection.

There are many other examples illustrating this kind of cautionary tale (for example, see Exercises 1.36–1.39). These are, however, important to help us understand the strengths, as well as the limitations, of mathematical models. And since forewarned is forearmed, let's embark on the task ahead, of modeling biological processes in both time and space.

Exercises

1.1. Calculate $\frac{\partial f}{\partial t}$ and $\frac{\partial^2 f}{\partial x^2}$ for $f = \frac{1}{\sqrt{t}} \exp(-\frac{x^2}{4t})$.

1.2. For the functions
(i) $f = x^2 + y^2$,
(ii) $f = xy$:
(a) Find $\frac{\partial f}{\partial x}$, $\frac{\partial^2 f}{\partial x \partial y}$, ∇f.
(b) Determine if there are critical points. Which, if any, are local maxima?
(c) Visualize the surface and the level curves for the function $z = f(x, y)$ using Matlab. Explore different Matlab functions for plotting: mesh, surf, contour, contour3, and see what they do. Type
`help mesh`
or similar to see the available options.

1.3. Determine which, if any, of the following vector fields are gradient fields. If it is a gradient field, find ϕ such that $\mathbf{F} = \nabla\phi$.
(a) $\mathbf{F} = (x + y, x - y)$,

(b) $\mathbf{F} = (x^2y, xy^2)$.

1.4. Suppose $u(x)$ is a smooth function. Use Taylor's theorem to verify that

$$u(x - \Delta x) - 2u(x) + u(x + \Delta x) = \frac{d^2u(x)}{dx^2}\Delta x^2 + O(\Delta x^4).$$

1.5. Use Taylor's theorem and the fact that e^x satisfies $\frac{d}{dx}(e^x) = e^x$ and $e^0 = 1$ to find the power series representation of e^x.

1.6. Use graphical analysis to find the solution of the differential equation (often called the *logistic equation*)

$$\frac{du}{dt} = u(1 - u)$$

for all positive initial data.

1.7. Determine the type of the critical point at the origin and sketch phase portraits for the systems

$$\frac{d}{dt}\begin{pmatrix} u \\ v \end{pmatrix} = A\begin{pmatrix} u \\ v \end{pmatrix},$$

with

(a)
$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

(b)
$$A = \begin{pmatrix} 0.1 & 1 \\ -1 & 0.1 \end{pmatrix},$$

(c)
$$A = \begin{pmatrix} -0.1 & 1 \\ -1 & -0.1 \end{pmatrix},$$

(d)
$$A = \begin{pmatrix} -1 & 0.2 \\ 0.3 & -0.3 \end{pmatrix}.$$

1.8. Give a phase plane analysis for the equation $u_{xx} + f(u) = 0$ where $f(u) = u(1 - u)$.

1.9. What fraction of a population of susceptible individuals s_0 should be vaccinated in order to prevent the spread of an SIR epidemic with parameters α (rate of infection) and β (rate of recovery)?

1.10. Suppose the rate of susceptible infection is given by the enzymatic rate

$$\frac{ds}{dt} = -\frac{\alpha si}{1 + \frac{s}{K}}.$$

(a) Why is this likely to be more reasonable than the law of mass action reaction rate? (Think about what happens when s is very large.)

(b) Give a phase plane analysis of these SIR dynamics including recovery of infected individuals at rate β.

(c) What is the threshold for $s_0 = s(0)$ for there to be an epidemic? Is this larger or smaller than for a law of mass action (i.e., $K \to \infty$) epidemic?

(d) Find a relationship between $s(\infty)$ and $s(0)$ following an epidemic. Is $s(\infty)$ larger or smaller than that for a law of mass action (i.e., $K \to \infty$) epidemic?

1.11. Simulate the Hodgkin–Huxley equations

$$C_m \frac{dV}{dt} + I_{\text{ion}} = I_{stim},$$

where $I_{\text{ion}} = \bar{g}_{Na} m^3 h(V - V_{\text{Na}}) + \bar{g}_K n^4 (V - V_{\text{K}}) + g_L(V - V_{\text{L}})$, and m, n, and the h are solutions of the ordinary differential equation

$$\frac{dj}{dt} = \alpha_j(V)(1 - j) - \beta_j(V) j,$$

for $j = m, n, h$, and

$$\alpha_m = 0.1 \frac{-25 - V}{\exp\left(\frac{25-v}{10}\right) - 1}, \quad \beta_m = 4 \exp\left(\frac{-V}{18}\right),$$

$$\alpha_h = 0.07 \exp\left(\frac{-V}{20}\right), \quad \beta_h = \frac{1}{\exp(\frac{30-V}{10}) + 1},$$

$$\alpha_n = 0.01 \frac{10 - V}{\exp(\frac{10-V}{10}) - 1}, \quad \beta_n = 0.125 \exp\left(\frac{-V}{80}\right).$$

For these expressions, the potential V is measured in units of mV, current density is in units of μA/cm^2, conductances $\bar{g}$ are in units of mS/cm^2, and capacitance is in units of μF/cm^2. The remaining parameters are

$$\bar{g}_{\text{Na}} = 120, \qquad \bar{g}_{\text{K}} = 36, \qquad \bar{g}_{\text{L}} = 0.3, \qquad C_m = 1,$$

with Nernst potentials $V_{\text{Na}} = 115$ mV, $V_{\text{K}} = -12$ mV, $V_{\text{L}} = -10.6$ mV.

(a) Show that the solution goes to a steady state if $I_{stim} = 0$.

(b) Show that for a range of values of $I_{stim} > 0$, the solution is periodic.

1.12. Simulate and give a phase plane analysis for the reduced Hodgkin–Huxley equations

$$C_m \frac{dV}{dt} + I_{\text{ion}} = I_{stim},$$

where $I_{\text{ion}} = \bar{g}_{Na} m_\infty^3 h(V - V_{\text{Na}}) + \bar{g}_K n^4 (V - V_{\text{K}}) + g_L(V - V_{\text{L}})$, and n is the solution of the ordinary differential equation

$$\frac{dn}{dt} = \alpha_n(V)(1 - n) - \beta_n(V) n,$$

and

$$m_\infty(V) = \frac{\alpha_m(V)}{\alpha_m(V) + \beta_m(V)},$$

where

$$\alpha_m = 0.1 \frac{-25 - V}{\exp\left(\frac{25-v}{10}\right) - 1}, \quad \beta_m = 4 \exp\left(\frac{-V}{18}\right),$$

$$\alpha_n = 0.01 \frac{10 - V}{\exp(\frac{10-V}{10}) - 1}, \quad \beta_n = 0.125 \exp\left(\frac{-V}{80}\right),$$

and $n + h = n_0 = 0.8$.

For these expressions, the potential V is measured in units of mV, current density is in units of μA/cm^2, conductances $\bar{g}$ are in units of mS/cm^2, and capacitance is in units of μF/cm^2. The remaining parameters are

$$\bar{g}_{\mathrm{Na}} = 120, \qquad \bar{g}_{\mathrm{K}} = 36, \qquad \bar{g}_{\mathrm{L}} = 0.3, \qquad C_m = 1,$$

with Nernst potentials $V_{\mathrm{Na}} = 115$ mV, $V_{\mathrm{K}} = -12$ mV, $V_{\mathrm{L}} = -10.6$ mV.

(a) Show that the solution goes to a steady state if $I_{stim} = 0$.

(b) Show that for a range of values of $I_{stim} > 0$, the solution is periodic.

Hint. To determine the nullcline $\frac{dV}{dt} = 0$, use bisection to find the n value for each value of V.

> As a reminder, the bisection method works as follows: Suppose you wish to find the zero of a function $f(u)$, and you know two values of u, say u_L and u_U, for which $f(u_L)f(u_U) < 0$. Set $u_c = \frac{1}{2}(u_L + u_U)$ and evaluate $f(u_c)$. If $f(u_L)f(u_c) > 0$, then replace u_L by u_c, whereas if $f(u_L)f(u_c) < 0$, then replace u_U by u_c. Repeat this until the difference between u_L and u_U is satisfactorily small. A Matlab code that does this and is readily adapted to work for your favorite function is found with the codes in Appendix A and is titled bisect_function.m.

1.13. (a) Suppose that y and z are independent random variables. Verify that $\mathrm{var}(y+z) = \mathrm{var}(y)+\mathrm{var}(z)$. *Remark.* Use that the joint probability function for y and z satisfies $p(\xi,\eta) = p_y(\xi)p_z(\eta)$.

(b) Show that this may fail if y and z are not independent, for example with the random variable $x + x$.

1.14. Suppose m_1,m_2,…,m_N are independent, identically distributed, random variables, with mean $\mu = E(m_i)$ and variance $\sigma^2 = \mathrm{var}(m_i)$. Find the mean and the variance of the random variable

$$x_N = \sum_{j=1}^{N} m_j.$$

1.15. Suppose $m_1, m_2, \ldots,$ are independent, identically distributed, random variables, with mean $\mu = E(m_i)$ and variance $\sigma^2 = \mathrm{var}(m_i)$. Find the mean and variance of the random variable

$$x_N = \frac{1}{N}\sum_{j=1}^{N} m_j.$$

1.16. What are the mean and variance of the uniform distribution on the interval $[0,1]$, $p(x) = 1$, for $0 \le x \le 1$?

1.17. What are the mean and variance of the discrete random variable {0,1} with each occurring with equal probability?

1.18. What are the mean and variance of the roll of a die {1,2,3,4,5,6} with equal probability?

1.19. Suppose you flip a coin N times, where N is a large even number.

(a) Use the central limit theorem to estimate the probability that there will be exactly $\frac{N}{2}$ heads.

(b) Do a numerical experiment (using the Matlab function randi) to check the validity of this formula.

1.20. Use Matlab to explore the validity of the central limit theorem, as follows. Use the Matlab function randi to generate N random integers m_i, $i = 1, 2, \ldots, N$, and then find the average $x_N = \frac{1}{N}\sum_{j=1}^{N} m_j$. Collect many samples of x_N and plot the distribution (the normalized histogram) and compare it to the appropriate normal distribution as predicted by the central limit theorem.

1.21. Use Matlab to explore the validity of the central limit theorem, as follows. Use the Matlab function rand to generate N uniformly distributed random numbers m_i, $i = 1, 2, \ldots, N$, $0 \le m_i \le 1$, and then find the average $x_N = \frac{1}{N}\sum_{j=1}^{N} m_j$. Collect many samples of x_N and plot the distribution (the normalized histogram) and compare it to the appropriate normal distribution as predicted by the central limit theorem.

1.22. For the differential equation

$$\frac{dc}{dt} = -\frac{vc}{K+c},$$

(a) What are the units of the parameters v and K?

(b) Show that by introducing appropriately scaled variables for c and t, one can transform this equation into the equation with no free parameters

$$\frac{du}{d\tau} = -\frac{u}{1+u}.$$

(c) Find and plot the solution of this equation for several different initial values including $u(0) \gg 1$ and $u(0) \ll 1$.
Hint. Find τ as a function of u, but plot it as u as a function of τ.

1.23. Suppose a fisherman fishes for four hours at a lake with a fishing success rate of 1/hour. What is the probability that he gets skunked (no fish), and what is the probability that he catches eight or more fish?

1.24. Find the solution of the system of differential equations (1.86), subject to initial conditions $p_0(0) = 1$, $p_j(0) = 0$ for $j > 0$, as follows:

(a) Let $g(z,t) = \sum_{j=0}^{\infty} z^j p_j(t)$. Find a differential equation for g.

(b) Solve the differential equation subject to initial data $g(z,0) = 1$.

(c) Expand $g(z,t)$ into its power series in z, thereby determining $p_j(t)$.

1.25. The decay rate of carbon-14 is 1.21×10^{-4} per year.

(a) Suppose you have one molecule of carbon-14 in a flask. What is the probability that it will decay within the next 30 seconds?

(b) The half-life of a radioactive particle is the time it takes for half of some initial quantity to decay. What is the half life of carbon-14?

1.26. (a) Suppose you go fishing with a friend who usually catches more fish than you do. (An experience I know all too well!) After much observation, you determine that, on average, he catches twice as many fish as you do. Suppose you both start fishing at the same time. What is the probability that you catch a fish before he does?

(b) More generally, suppose your friend catches fish at rate α_1 and you with rate α_2. What is the probability that you will catch a fish before he does?

Hint. The joint probability distribution function for two independent random variables t_1 and t_2 with probability density functions $f_1(\tau_1)$ and $f_2(\tau_2)$, respectively, is the product $f(\tau_1, \tau_2) = f_1(\tau_1)f_2(\tau_2)$. The probability that t_1, t_2 are in some domain Ω is $P((t_1, t_2) \in \Omega) = \int_\Omega f(\tau_1, \tau_2)d\tau_1 d\tau_2$. Suppose $f_1(\tau_1)$ and $f_2(\tau_2)$ represent the pdf's for catching a fish for fisherman 1 and 2, respectively, find the probability that $t_2 < t_1$, i.e., $P(t_2 < t_1)$.

1.27. An infection spreads in a neighborhood from person to person through nearest neighbor interaction only. Suppose the rate of spread is α meaning that if an individual is infected, her neighbor will become infected at rate α. However, recovery from infection is at rate β, so it might be that an infected individual recovers before her neighbor is infected. What is the probability that exactly n individuals will be infected after the neighbor at the end of the street becomes infected? What is the expected value and variance of n? *Hint.* Use the probability that passing the infection before recovery is $\frac{\alpha}{\alpha+\beta}$ and the probability of recovery before passing the infection is $\frac{\beta}{\alpha+\beta}$.

1.28. Suppose t is a random variable $0 < t < \infty$ with probability density function $f(\tau)$ with $\int_0^\infty f(\tau)d\tau = 1$. Show that the cumulative distribution function $r = \int_0^t f(\tau)d\tau$ is a uniformly distributed random variable on the interval $[0, 1]$. *Hint.* Calculate that the probability $P(r_0 < r < r_1) = r_1 - r_0$ for any $r_0 < r_1$ with $0 < r_0 < r_1 < 1$.

1.29. A finite birth-death process is identified by the reactions

$$S_1 \underset{\beta_2}{\overset{\alpha_1}{\rightleftarrows}} S_2 \underset{\beta_3}{\overset{\alpha_2}{\rightleftarrows}} \cdots S_{N-1} \underset{\beta_N}{\overset{\alpha_{N-1}}{\rightleftarrows}} S_N.$$

(a) What is the matrix A defined by (1.85) for this process?

(b) Suppose all of the reaction rates α_j, β_{j+1}, $j = 1, 2, \ldots, N-1$ are nonzero. Show that A, an $N \times N$ matrix, has rank $= N - 1$, i.e., has rank deficiency $= 1$. *Hint.* Use Gaussian elimination to row reduce the matrix to an upper diagonal matrix.

(c) Suppose $\alpha_1 = 0$ and all other reaction rates are nonzero. Show that the nullspace of A, i.e., the nontrivial solution of $A\mathbf{p_0} = 0$, is one dimensional and is spanned by the vector $\mathbf{p_0} = (1, 0, 0, \ldots, 0)^T$. *Remark.* The state S_1 is called an *absorbing state*.

(d) Realizing that A has $N-1$ eigenvalues with negative real part and one zero eigenvalue (you do not need to prove this) show that, for *any* initial data $\mathbf{p}(0)$ with $\mathbf{1}^T \cdot \mathbf{p}(0) = 1$, $\lim_{t\to\infty} \mathbf{p}(t) = \mathbf{p_0}$ where $\mathbf{p_0}$ is the unique element of the nullspace of A.

1.30. Use induction to verify that the solution of the equations (1.100) is given by

$$p_k(t) = \begin{pmatrix} N \\ \kappa \end{pmatrix} \exp(-\alpha k t)(1 - \exp(-\alpha t))^{N-k}$$

so that the extinction probability is $p_0(t) = (1 - \exp(-\alpha t))^N$. How does this compare to the numerically computed solution found using the Matlab code exponential_decay_via_Gillespie.m?

1.31. Suppose the bimolecular decay process (1.109) has a rate constant β with units $(\text{time})^{-1}(\text{mM})^{-1}$. What is the conversion factor vol?

1.32. Suppose that some chemical species S is degraded by the bimolecular process

$$S + S \xrightarrow{\beta} S.$$

(a) What is the deterministic differential equation governing this decay process? What is the solution of this differential equation and what prediction does it make about the time course of decay for this process?
(b) There can never be less than one molecule of S. Do a stochastic simulation of this decay process and find the time distribution for arriving at one molecule for this process.
(c) What is the expected time for n molecules to decay into $n-1$ molecules? What is the expected time for N molecules to decay to one molecule? *Hint.* Use that $\sum_{k=2}^{N} \frac{1}{k(k-1)} = 1 - \frac{1}{N}$.
(d) What are the equations governing $p_k(t)$, the probability of having k molecules at time t. Simulate these equations and compare $\frac{dp_1}{dt}$ with the arrival time distribution found from your stochastic simulation.

Hint. Use the Matlab code exponential_decay_via_Gillespie.m as a template for this solution.

1.33. Suppose that there is a chemical species S that is degraded into species I by the reaction

$$S + I \xrightarrow{\gamma} 2I.$$

(a) Supposing that $[S]+[I] = S_0$, determine the deterministic differential equation governing the decay of S and its solution.
(b) Do a stochastic simulation of this decay process and find the time distribution of extinction of S, starting from 1 I and $N-1$ S particles.
(c) What are the equations governing $p_k(t)$, the probability of having k molecules at time t. Simulate these equations and compare $\frac{dp_0}{dt}$ with the arrival time distribution found from your stochastic simulation.

Hint. Use the Matlab code exponential_decay_via_Gillespie.m as a template for this solution.

1.34. Suppose that some chemical species S degrades into another species I at some rate α, via the bimolecular reaction

$$S + S \xrightarrow{\alpha} S + I,$$

and that S can reproduce via the reaction

$$S \xrightarrow{\beta} 2S.$$

(a) What is the deterministic differential equation governing the dynamics of S, and what are its stable and unstable equilibria?

(b) Write the equations governing the probability $p_k(t)$ that there are k molecules of S at time t. Do a numerical simulation of these equations and show that the solution goes to an equilibrium. What is the relationship between this equilibrium solution and $\frac{\alpha}{\beta}$?

(c) Do a stochastic simulation of this process. The population cannot go extinct. What does the population do? Compare this to the equilibrium solution of $p_k(t)$. *Hint.* Use the Matlab code stochastic_birth_death.m as a template for this stochastic simulation.

1.35. An enzyme E converts a substrate S molecule into product P by the reactions

$$S + E \underset{\beta}{\overset{\alpha}{\rightleftarrows}} C \xrightarrow{\gamma} P + E.$$

(a) Do a stochastic simulation of this decay process, assuming there is a single enzyme molecule. How does the rate of this decay process compare with that of an exponential decay process? *Warning.* The simulation is slower as $\frac{\beta}{\gamma}$ gets larger. Why? *Hint.* Use the Matlab code stochastic_SIR.m as a template for this stochastic simulation.

(b) Calculate the extinction time distribution for this process, and compare it with a Gaussian distribution with mean and variance computed from the data.

(c) What are the equations governing $p_{k,1}(t)$ and $p_{k,0}(t)$, where $p_{k,1}$ is the probability of having k substrate molecules and one free enzyme molecule, and $p_{k,0}$ is the probability of having k substrate molecules and no free enzyme molecules. Solve these equations numerically and compare $\frac{dp_{0,1}}{dt}$ with the data for exit time distribution computed from your stochastic simulation.

1.36. Consider the chemical reactions

$$S + U \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} 2U, \quad U \xrightarrow{k_2} P.$$

Assume that $[S] = s$ is constant.

(a) Do a stochastic simulation of these reactions. Find the extinction time distribution starting with about $\frac{k_1 s - k_2}{k_{-1}}$ molecules of U. (Choose $\frac{k_1 s - k_2}{k_{-1}}$ to be of order 10. What happens to the simulation time if $\frac{k_1 s - k_2}{k_{-1}}$ is much larger than this?)

(b) Find the deterministic differential equation governing the concentration of species U. Give a complete analysis of the differential equation under the assumption that $k_1 s > k_{-1}$. What do you observe about the deterministic differential equation that is in contradiction with the stochastic simulation? *Remark.* This contradiction is referred to as *Keizer's paradox* [**36**],[**10**].

(c) Find the equation governing the probability $p_n(t)$ that there are n molecules of U at time t, and simulate this equation to find the probability of going extinct as a function of time. Observe that $p_0(t) \to 1$ as $t \to \infty$. How does $\frac{dp_0}{dt}$ compare with the extinction distribution found in your simulation?

1.37. Suppose that some chemical species S converts into another species I at some rate α, via the bimolecular reaction

$$S + I \xrightarrow{\alpha} 2I, \tag{1.127}$$

and that I degrades back to S via the reaction

$$I \xrightarrow{\beta} S. \tag{1.128}$$

Remark. If S represents susceptibles and I infectives, then this is called an *SIS* disease.

(a) What is the deterministic differential equation describing the dynamics of S? Notice that $[S] + [I] = S_0$ is a constant. Under what conditions on parameters is there an endemic stable steady state for which $I > 0$?

(b) Do a stochastic simulation of this reaction. Show that I *always* goes extinct, even when the deterministic model predicts a stable endemic state. Explain this apparently paradoxical behavior. (See Exercise 1.29.)

(c) Find the equation governing the probability $p_n(t)$ that there are n individuals of I at time t, and simulate this equation to find the probability of going extinct as a function of time. Observe that $p_0(t) \to 1$ as $t \to \infty$. How does $\frac{dp_0}{dt}$ compare with the extinction distribution found in your simulation?

1.38. An *SIRS* disease is a disease for which recovery and immunity is only temporary, and is described by the reaction scheme

$$S + I \xrightarrow{\alpha} 2I, \quad I \xrightarrow{\beta} R, \quad R \xrightarrow{\gamma} S. \tag{1.129}$$

(a) Write the deterministic differential equations for these reactions. Give a phase plane analysis of this system and show that if $\frac{\beta}{\alpha s_0} < 1$, where $s + i + r = s_0$, then there is an endemic state, i.e., a nontrivial, stable steady state solution with $i > 0$. Show also that the steady state solution $s = s_0$, $i = 0$ is unstable.

(b) Do a stochastic simulation of this system and show that the infection *always* goes extinct. *Hint.* Assume that a concentration of one corresponds to 100 individuals. Choose parameter values for which the steady state of I is not too large. Why is this helpful?

(c) Explain this paradoxical result.

1.39. Consider the chemical reactions

$$U + V \xrightarrow{\alpha} 2V, \quad U \xrightarrow{\beta} 2U, \quad V \xrightarrow{\gamma} \emptyset.$$

(a) Find the deterministic differential equations governing the concentrations of species U and V. Give a complete phase plane analysis of this system, including plotting some typical trajectories. Find a first integral of the motion. (These equations are known as the Lotka–Volterra equations.)

(b) Do a stochastic simulation of these reactions with the same parameter values. Assume that a concentration of one corresponds to 100 individuals. Make a plot of typical stochastic trajectories in $u - v$ concentration space to compare with the deterministic phase portrait. Notice that one of the

populations always goes extinct; explain why this is in contradiction to the deterministic solution. Find the distribution of extinction times.

Hint. Without loss of generality, $\alpha = \beta = 1$, so the only free parameter is γ. Choose "easy" values of γ for this exercise.