

Mathematical Modeling

Preliminary Lecture Notes

Adolfo J. Rumbos

© *Draft date March 23, 2012*

Contents

1	Preface	5
2	Modeling Process	7
2.1	Constructing Models	7
2.1.1	Conservation Principles	8
2.1.2	Constitutive Equations	8
2.2	Example: Bacterial Growth in a Chemostat	8
2.3	Analysis of Models	11
2.3.1	Nondimensionalization	11
3	Continuous Deterministic Models	17
3.1	Example: Modeling Traffic Flow	17
3.2	Analysis of the Traffic Flow Equation	20
4	Stochastic Models	35
4.1	Random Variables	36
4.1.1	A Brief Excursion into Probability	37
4.1.2	Discrete Random Variables	40
4.1.3	The Binomial Distribution	42
4.1.4	Expected Value	48
4.1.5	The Poisson Distribution	49
4.1.6	Estimating Mutation Rates in Bacterial Populations	52
4.1.7	Another Brief Excursion into Probability	53
4.1.8	Continuous Random Variables	56
4.2	Random Processes	61

Chapter 1

Preface

The main goal of this course is to provide opportunities for students to construct and analyze mathematical models that arise in the physical, biological and social sciences. Mathematical models are usually created in order to obtain understanding of problems and situations arising in the real world; other times, the main goal is to make predictions or to control certain processes; finally, some models are created in order to aid in decision making.

Construction of a mathematical model consists of translating a real world problem into a mathematical problem involving parameters, variables, functions, equations and/or inequalities. Analysis of the model involves the solution (if possible) of the mathematical problem through logical, algebraic, analytical or computational means, and assessing what the solutions imply about the real situation under study. If an analytical or computational solution is not possible, computer simulations can sometimes be used in order to study various scenarios implied or predicted by the model.

Analysis techniques can be drawn from many areas of mathematics. In this course, it is assumed that students have a good working knowledge of Calculus, Linear Algebra and Ordinary Differential Equations. These areas are adequate for the analysis of some models. However, many modeling situations require the use of some probability theory and linear programming. These mathematical topics will be covered in the course. In calculus and differential equations courses, students have been exposed to some continuous models. In this course, we will also introduce students to discrete models as well.

Chapter 2

The Process of Modeling

In this course we identify three stages in the process of mathematical modeling:

1. Construction of the model.
2. Analysis of the model.
3. Testing of the model.

We will get a chance to go through at least the first two stages listed above in a variety of case studies or examples.

Of course, the modeling process always begins with a question that we want to answer, or a problem we have to solve. Often, asking the right questions and posing the right problems can be the hardest part in the modeling process. This part of the process involves getting acquainted with the intricacies of the science involved in the particular question at hand. It is unrealistic to expect that a mathematical modeling course will teach students to do this in a systematic way. The best we can do is to present many case studies and examples of real life modeling situations that mathematicians have analyzed in various situations. One of the goals of the course is to have students grapple with this issue when working with a specific problem in a term project that will involve a large portion of the course.

2.1 Constructing Models

Model construction involves the translation of a scientific question or problem into a mathematical one. The hope here is that answering the mathematical question, or solving the mathematical problem, if possible, might shed some light in the understanding of the situation being studied. In a physical science, this process is usually attained through the use of established scientific principles or laws that can be stated in mathematical terms. In general, though, we might not have the advantage of having at our disposal a large body of scientific principles. This is particularly the case if scientific principles have not been discovered yet

(in fact, the reason we might be resorting to mathematical modeling is that, perhaps, mathematics can aid in the discovery of those principles).

2.1.1 Conservation Principles

There are, however, a few general and simple principles that can be applied in a variety of situations. For instance, in this course we'll have several opportunities to apply **conservation principles**. These are rather general principles that can be applied in situations in which the evolution in time of the quantity of a certain entity within a certain system is studied. For instance, suppose the quantity of a certain substance confined within a system is given by a continuous function of time, t , and is denoted by $Q(t)$ (the assumption of continuity is one that needs to be justified by the situation at hand). A conservation principle states that the rate at which a the quantity $Q(t)$ changes has to be accounted by how much of the substance goes into the system and how much of it goes out of the system. For the case in which Q is also assumed to be differentiable (again, this is a mathematical assumption that would need some justification), the conservation principle can be succinctly stated as

$$\frac{dQ}{dt} = \text{Rate of } Q \text{ in} - \text{Rate of } Q \text{ out.} \quad (2.1)$$

In this case, the conservation principle might lead to a differential equation, or a system of differential equations, and so the theory of differential equations can be used to help in the analysis of the model.

2.1.2 Constitutive Equations

The right-hand side of the equation in (2.1) requires further modeling; in other words, we need to postulate a kind of functional form for the rates in the right-hand side of (2.1). This might take the general form of rewriting the equation in (2.1) as

$$\frac{dQ}{dt} = f(t, Q, \lambda_1, \lambda_2, \dots, \lambda_p), \quad (2.2)$$

where $\{\lambda_1, \lambda_2, \dots, \lambda_p\}$ is a collection of parameters that are relevant to the real-life problem being modeled. The functional form of the right-hand side in (2.2) may be obtained from empirical or theoretical of relations between variables usually called constitutive equations.

In the next subsection we present the first case study of the course in which we see the two elements outlined above in the construction of models.

2.2 Example: Bacterial Growth in a Chemostat

This example presented in this subsection is discussed on page 121 of [EK88]. The diagram in Figure 2.2.1 shows schematically what goes on in a chemostat that is used to harvest bacteria at a constant rate. The box in the top-left

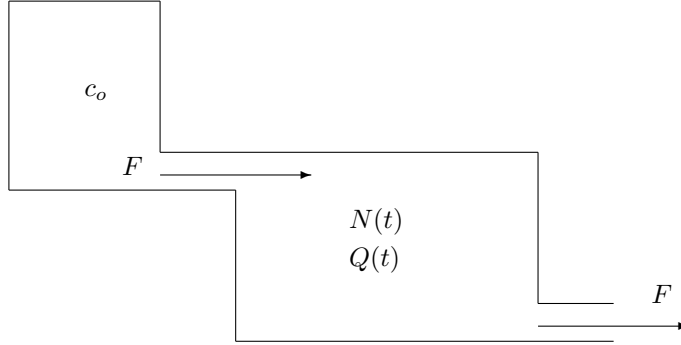


Figure 2.2.1: Chemostat

portion of the diagram in Figure 2.2.1 represents a stock of nutrient at a fixed concentration c_o , in units of mass per volume. Nutrient flows into the bacterial culture chamber at a constant rate F , in units of volume per time. The chamber contains $N(t)$ bacteria at time t . The chamber also contains an amount $Q(t)$ of nutrient, in units of mass, at time t . If we assume that the culture in the chamber is kept well-stirred, so that there are no spatial variations in the concentration of nutrient and bacteria, we have that the nutrient concentration is a function of time given by

$$c(t) = \frac{Q(t)}{V}, \quad (2.3)$$

where V is the volume of the culture chamber. If we assume that the culture in the chamber is harvested at a constant rate F , as depicted in the bottom-right portion of the diagram in Figure 2.2.1, the volume, V , of the culture in equation (2.3) is fixed.

We will later make use of the bacterial density,

$$n(t) = \frac{N(t)}{V}, \quad (2.4)$$

in the culture at time t .

The parameters, c_o , F and V , introduced so far can be chosen or adjusted. The problem at hand, then, is to design a chemostat system so that

1. The flow rate, F , will not be so high that the bacteria in the culture will be washed out, and
2. the nutrient replenishment, c_o , is sufficient to maintain the growth of the colony.

In addition to assuming that the culture in the chamber is kept well-stirred and that the rate of flow into and out of the chamber are the same, we will also make the following assumptions:

1. The bacterial colony depends on only one nutrient for growth;
2. The growth rate of the bacterial population is a function of the nutrient concentration; in other words, the *per-capita* growth rate, $K(c)$, is a function of c .

We will apply a conservation principles to the quantities $N(t)$ and $Q(t)$ in the growth chamber. For the number of bacteria in the culture, the conservation principle in (2.1) reads:

$$\frac{dN}{dt} = \text{Rate of } N \text{ in} - \text{Rate of } N \text{ out.} \quad (2.5)$$

We are assuming here that N is a differentiable function of time. This assumption is justified if

- (i) we are dealing with populations of very large size so that the addition (or removal) of a few individuals is not very significant; for example, in the case of a bacterial colony, N is of the order of 10^6 cells per milliliter;
- (ii) "there are no distinct population changes that occur at timed intervals," see [EK88, pg. 117].

Using the constitutive assumption stated previously, we have that

$$\text{Rate of } N \text{ in} = K(c)N, \quad (2.6)$$

since $K(c)$ is the *per-capita* growth rate of the bacterial population.

Since culture is taken out of the chamber at a rate F , we have that

$$\text{Rate of } N \text{ out} = Fn, \quad (2.7)$$

where n is the bacterial density defined in (2.4). We can therefore re-write (2.5) as

$$\frac{dN}{dt} = K(c)N - \frac{F}{V}N. \quad (2.8)$$

Next, apply the conservation principle (2.1) to the amount of nutrient, $Q(t)$, in the chamber, where

$$\text{Rate of } Q \text{ in} = Fc_o, \quad (2.9)$$

and

$$\text{Rate of } Q \text{ out} = Fc + \alpha K(c)N, \quad (2.10)$$

where we have introduced another parameter α , which measures the fraction of nutrient that is being consumed as a result of bacterial growth. The reciprocal of the parameter α ,

$$Y = \frac{1}{\alpha}, \quad (2.11)$$

measures the number of cells produced due to consumption of one unit of nutrient, and is usually referred to as the yield.

Combining (2.10), (2.9) and (2.1) we see that the conservation principle for Q takes the form

$$\frac{dQ}{dt} = Fc_o - Fc - \alpha K(c)N. \quad (2.12)$$

Using the definition of c in (2.3) we can re-write (2.12) as

$$\frac{dQ}{dt} = Fc_o - \frac{F}{V}Q - \alpha K(c)N. \quad (2.13)$$

The differential equations in (2.8) and (2.13) yield the system of differential equations

$$\begin{cases} \frac{dN}{dt} = K(c)N - \frac{F}{V}N; \\ \frac{dQ}{dt} = Fc_o - \frac{F}{V}Q - \alpha K(c)N. \end{cases} \quad (2.14)$$

Thus, application of conservation principles and a few constitutive assumptions has yielded a system of ordinary differential equations (2.14) for the variables N and Q in the chemostat system. We have therefore constructed a preliminary mathematical model for bacterial growth in a chemostat.

Dividing the equations in (2.14) by the fixed volume, V , of the culture in the chamber, we obtain the following system of ordinary differential equations for the bacterial population density, $n(t)$, and the nutrient concentration, $c(t)$.

$$\begin{cases} \frac{dn}{dt} = K(c)n - \frac{F}{V}n; \\ \frac{dc}{dt} = \frac{F}{V}c_o - \frac{F}{V}c - \alpha K(c)n. \end{cases} \quad (2.15)$$

Thus, we have arrived at a mathematical model that describes the evolution in time of the population density and nutrient concentration in a chemostat system. We will analyze the system in (2.15) in subsequent sections.

2.3 Analysis of Models

In the process of constructing the differential equations model expressed in the system in (2.15) we made several simplifying assumptions; for instance, we assumed that the mixture in the culture is well-stirred and that the volume V is fixed, so that the bacterial density and nutrient concentration are functions of a single variable, t . We also assumed that these are differentiable functions. Simplification is an important part of the modeling process; otherwise the mathematical problem might be intractable.

2.3.1 Nondimensionalization

In this section we illustrate yet another way to simplify the problem which consists of introducing dimensionless variables (variables without units). This

process is known as nondimensionalization, and it has the added benefit of decreasing the number of parameters in the system, reducing thereby the complexity of the problem. We illustrate this procedure in the analysis of the chemostat system in (2.15).

Note that the system in (2.15) has four parameters; namely, c_o , F , V and α . Before proceeding with the analysis of system (2.15), we will use the constitutive equation

$$K(c) = \frac{bc}{a+c}, \quad (2.16)$$

where a and b are two additional positive parameters. Thus, the system in (2.15) becomes

$$\begin{cases} \frac{dn}{dt} = \frac{bnc}{a+c} - \frac{F}{V}n; \\ \frac{dc}{dt} = \frac{F}{V}c_o - \frac{F}{V}c - \frac{abc}{a+c}, \end{cases} \quad (2.17)$$

with six parameters. A procedure that consolidates the set of parameters into a smaller set will greatly simplify the analysis.

The constitutive equation in (2.16) is borrowed from the Michaelis–Menten theory of enzyme kinetics. It models the *per-capita* growth rate, $K(c)$, of the bacterial population as an increasing function of the nutrient concentration with a limiting value b ; hence, b has the units of a *per-capita* growth rate, namely, 1/time. The parameter a has units of concentration (mass/volume), and it represents the value of the nutrient concentration at which the *per-capita* growth rate, $K(c)$, is half of its limiting value. Figure 2.3.2 shows a sketch of the graph of K as a function of c as given by (2.16).

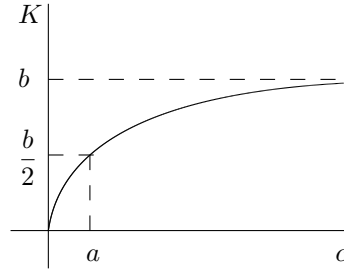


Figure 2.3.2: Sketch of Graph of $K(c)$

Nondimensionalizing the system in (2.17) consists of introducing new variables, \hat{n} , \hat{c} and τ , to replace n , c and t , respectively, in such a way that \hat{n} , \hat{c} and τ have no units, or dimensions. This can be achieved by scaling the variables n , c and t by appropriate scaling factors so that the units will cancel out. For instance, we can scale c by the parameter a , since they have the same units, to

get

$$\widehat{c} = \frac{c}{a}. \quad (2.18)$$

It is not clear at the moment what the scaling factor for n and t should be, so we shall denote them by μ and λ , respectively. We then have that,

$$\widehat{n} = \frac{n}{\mu}, \quad (2.19)$$

and

$$\tau = \frac{t}{\lambda}, \quad (2.20)$$

where μ has units of bacterial density (cells/volume), and λ has units of time.

Next, we find expressions for the derivatives

$$\frac{d\widehat{n}}{d\tau} \quad \text{and} \quad \frac{d\widehat{c}}{d\tau}. \quad (2.21)$$

To find the expressions in (2.21) we need to apply the Chain Rule; for instance,

$$\frac{d\widehat{n}}{d\tau} = \frac{d\widehat{n}}{dt} \cdot \frac{dt}{d\tau} \quad (2.22)$$

To compute the right-hand side of (2.22), we use (2.19) and (2.20) to obtain from (2.22) that

$$\frac{d\widehat{n}}{d\tau} = \frac{\lambda}{\mu} \frac{dn}{dt}. \quad (2.23)$$

Next, substitute the expression for $\frac{dn}{dt}$ in the first equation in (2.17) into the right-hand side of (2.23) to obtain

$$\frac{d\widehat{n}}{d\tau} = \frac{\lambda}{\mu} \left[\frac{bn\widehat{c}}{1+\widehat{c}} - \frac{F}{V}n \right], \quad (2.24)$$

where we have also used the expression for \widehat{c} in (2.18). Distributing on the right-hand side of (2.24) we obtain

$$\frac{d\widehat{n}}{d\tau} = \lambda b \frac{\widehat{n}\widehat{c}}{1+\widehat{c}} - \frac{\lambda F}{V} \widehat{n}, \quad (2.25)$$

where we have used (2.19).

We will now choose λ so that

$$\frac{\lambda F}{V} = 1, \quad (2.26)$$

from which we get that

$$\lambda = \frac{V}{F} \quad (2.27)$$

is our scaling factor for t ; observe that the parameter λ in (2.27) has units of time.

Next, we consolidate the parameters b and λ into a single parameter, which we will call α_1 , by the formula

$$\lambda b = \alpha_1, \quad (2.28)$$

which yields

$$\alpha_1 = \frac{bV}{F}, \quad (2.29)$$

by the use of the definition of λ in (2.27). Note that α_1 in (2.32) is dimensionless since b has units of 1/time. Combining (2.25), (2.26) and (2.28), we obtain the dimensionless differential equation

$$\frac{d\hat{n}}{d\tau} = \alpha_1 \frac{\hat{n}\hat{c}}{1+\hat{c}} - \hat{n}. \quad (2.30)$$

Similar calculations (see Problem 1 in Assignment 2) show that

$$\frac{d\hat{c}}{d\tau} = \alpha_2 - \frac{\hat{n}\hat{c}}{1+\hat{c}} - \hat{c}, \quad (2.31)$$

where we have set

$$\alpha_2 = \frac{c_o}{a} \quad (2.32)$$

and

$$\frac{\alpha b \lambda \mu}{a} = 1,$$

so that

$$\mu = \frac{a}{\alpha b \lambda}. \quad (2.33)$$

Note that the parameter α_2 in (2.32) is dimensionless and that the units of μ defined in (2.33) are cells/volume.

Putting together the equations in (2.30) and (2.33) we obtain the system

$$\begin{cases} \frac{d\hat{n}}{d\tau} = \alpha_1 \frac{\hat{n}\hat{c}}{1+\hat{c}} - \hat{n}; \\ \frac{d\hat{c}}{d\tau} = \alpha_2 - \frac{\hat{n}\hat{c}}{1+\hat{c}} - \hat{c}, \end{cases} \quad (2.34)$$

in the nondimensional variables \hat{n} , \hat{c} and τ defined in (2.19), (2.18) and (2.20), respectively. Observe that the system in (2.34) contains two dimensionless parameters, α_1 and α_2 , as opposed to the six parameters in the original system in (2.17). This reduction in the number of parameters greatly simplifies the problem in two aspects:

1. the mathematical calculations are simpler to perform;
2. the dimensionless parameters, α_1 and α_2 , consolidate the information contained in the six original parameters, and this makes the analysis easier to carry out.

For instance, the equilibrium points of the system in (2.34) are expressed in terms of the parameters α_1 and α_2 as follows

$$(0, \alpha_2) \quad \text{and} \quad \left(\alpha_1 \left(\alpha_2 - \frac{1}{\alpha_1 - 1} \right), \frac{1}{\alpha_1 - 1} \right). \quad (2.35)$$

In order to obtain biologically feasible equilibrium solutions, we must require that

$$\alpha_1 > 1 \quad (2.36)$$

and

$$\alpha_2 > \frac{1}{\alpha_1 - 1}. \quad (2.37)$$

In terms of the original parameters, conditions (2.36) and (2.37) translate into

$$F < bV$$

and

$$c_o > \frac{aF}{bV - F},$$

respectively.

The equilibrium solution $(0, \alpha_2)$ in (2.35) is referred to as the “washout” solution, since all the bacteria washed out because of the flow; while the second solution in (2.35) is the “survival” solution.

Stability analysis of the dimensionless system in (2.34) will reveal further conditions that determine whether the chemostat system will yield a sustainable crop of bacteria. Some of this analysis is carried out in Assignment 2.

Chapter 3

Continuous Deterministic Models

The chemostat model discussed in the previous chapter is an example of a continuous model—all the variables in question, N , Q and t , are assumed to be continuous variables (in fact, we assumed that N and Q are differentiable functions of t). Application of conservation principles, in that case, led to a system of ordinary differential equations, which also makes the chemostat model discussed in the previous chapter a deterministic model—the solutions to the model are completely determined by the parameters in the model and by the initial conditions; in particular, a given set of initial conditions and parameters give rise to the same predictions every time the model is run.

In this chapter we present another kind of deterministic model that involves variables that depend continuously on more than one variable. We will again apply a conservation principle; however, in this case we will get a partial differential equation model. In subsequent sections in this chapter we will present an approach to analysis of the equations that result from this process. In particular, we will see an application to modeling traffic flow.

3.1 Example: Modeling Traffic Flow

Consider the unidirectional flow of traffic in a one-lane, straight road depicted in Figure 3.1.1. In this idealized road, vehicles are modeled by moving points. The location, x , of a point-vehicle is measured from some reference point along an axis parallel to the road. We postulate a traffic density, $\rho(x, t)$, measured in units of number of cars per unit length of road at location x and time t . We interpret $\rho(x, t)$ as follows: Consider a section of the road from x to $x + \Delta x$ at time t . Let $\Delta N([x, x + \Delta x], t)$ denote the number of cars in the section $[x, x + \Delta x]$ at time t . We define $\rho(x, t)$ by the expression

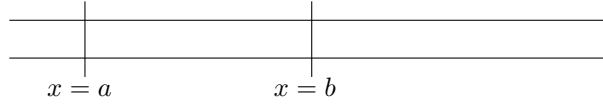
$$\rho(x, t) = \lim_{\Delta x \rightarrow 0} \frac{\Delta N([x, x + \Delta x], t)}{\Delta x}, \quad (3.1)$$



Figure 3.1.1: One-lane unidirectional flow

provided that the limit on the right-hand side of (3.1) exists.

Next, consider a section of the road from $x = a$ to $x = b$ pictured in Figure 3.1.2. Using the traffic density, $\rho(x, t)$, we calculate the number of vehicles in

Figure 3.1.2: Section of road from $x = a$ to $x = b$

the section $[a, b]$ at time t to be given by the integral

$$N(t) = \int_a^b \rho(x, t) dx. \quad (3.2)$$

If we assume that density, $\rho(x, t)$, is a differentiable function, we can state a conservation principle for the number of cars in section $[a, b]$ as follows:

$$\frac{dN}{dt} = \text{Number of cars entering at } a - \text{Number of cars leaving at } b. \quad (3.3)$$

We can re-write the conservation principle in (3.3) more succinctly by postulating a **traffic flux** function, $q(x, t)$, which measures the number of cars crossing location x per unit of time at time t . Using the traffic flux function we can then re-write (3.3) as

$$\frac{dN}{dt} = q(a, t) - q(b, t),$$

or

$$\frac{dN}{dt} = -[q(b, t) - q(a, t)]. \quad (3.4)$$

Assuming that the flux function is differentiable and that its partial derivatives are continuous, we can invoke the Fundamental Theorem of Calculus to re-write (3.4) as

$$\frac{dN}{dt} = - \int_a^b \frac{\partial}{\partial x} [q(x, t)] dx. \quad (3.5)$$

Next, assume that the traffic density, ρ , has continuous partial derivatives to obtain from (3.2) that

$$\frac{dN}{dt} = \int_a^b \frac{\partial}{\partial t} [\rho(x, t)] dx. \quad (3.6)$$

Combining (3.6) and (3.5) we then see that the conservation principle in (3.4) now takes the form

$$\int_a^b \frac{\partial}{\partial t} [\rho(x, t)] dx = - \int_a^b \frac{\partial}{\partial x} [q(x, t)] dx. \quad (3.7)$$

Rewrite the equation in (3.7) as

$$\int_a^b \frac{\partial}{\partial t} [\rho(x, t)] dx + \int_a^b \frac{\partial}{\partial x} [q(x, t)] dx = 0,$$

or

$$\int_a^b \left[\frac{\partial}{\partial t} [\rho(x, t)] + \frac{\partial}{\partial x} [q(x, t)] \right] dx = 0, \quad (3.8)$$

and observe that the points a and b are arbitrary. Since, we are assuming that the partial derivatives of ρ and q are continuous, we can show (see Assignment 4) that, given that (3.8) holds true for all intervals $[a, b]$, then we must have that

$$\frac{\partial}{\partial t} [\rho(x, t)] + \frac{\partial}{\partial x} [q(x, t)] = 0. \quad (3.9)$$

The equation in (3.9) is an example of a **partial differential equation** or PDE. It is usually written in a more compact form

$$\frac{\partial \rho}{\partial t} + \frac{\partial q}{\partial x} = 0, \quad (3.10)$$

or

$$\rho_t + q_x = 0, \quad (3.11)$$

where the subscripts in (3.11) indicate partial derivatives with respect to the subscripted variables.

Ideally, we would like to find a solution, ρ , to (3.10) subject to some initial condition

$$\rho(x, 0) = \rho_o(x), \quad (3.12)$$

for some initial traffic density profile, ρ_o , along the road.

Before we proceed any further, we need to model the traffic flux, $q(x, t)$. Imagine a vehicle at x is moving with a velocity v . Then, for a short time interval, $[t, t + \Delta t]$ the car moves a distance approximately give by

$$\Delta x \approx v \Delta t.$$

The number of cars in the section $[x, x + \Delta x]$ is then, approximately, given by

$$\Delta N \approx \rho(x, t) v \Delta t. \quad (3.13)$$

Dividing (3.13) by Δt and letting $\Delta t \rightarrow 0$, we obtain from (3.13) the rate of cars crossing the location x at time t ; in other words, the traffic flux, $q(x, t)$. We therefore obtain the constitutive equation

$$q = v\rho. \quad (3.14)$$

Substituting the expression for q into (3.10) we obtain

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}[v\rho] = 0, \quad (3.15)$$

The next step is to model the velocity v in (3.15). It is reasonable to assume that v is a function of traffic density—the higher the density, the lower the traffic speed. We may therefore write

$$v = f(\rho, \Lambda), \quad (3.16)$$

where f is a continuous function of ρ and a set of parameters, Λ . Some of the parameters might be a maximum density, ρ_{\max} , dictated by bumper to bumper traffic, and a maximum speed, v_{\max} ; for instance, v_{\max} is a speed limit. Given the parameters ρ_{\max} and v_{\max} , the simplest model for the relationship between v and ρ is the constitutive equation

$$v = v_{\max} \left(1 - \frac{\rho}{\rho_{\max}} \right). \quad (3.17)$$

Later in this course we shall see how to derive expressions like (3.17) relating traffic velocity to traffic density from theoretical considerations; for now, we simply note that it models the intuitive notion that we stated above regarding the traffic velocity being low for high traffic density.

The partial differential equation model for traffic flow presented in this section, based on the conservation equation in (3.15) and a constitutive relation for the traffic velocity, v , and the traffic density ρ (of which (3.17) is just an example), was first introduced by Lighthill and Whitman in 1955 (see [LW55]); it was also treated by Richards in 1956, [Ric56].

3.2 Analysis of the Traffic Flow Equation

In this section we outline an analysis of the equation in (3.15) with v as given in (3.17); namely,

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x} \left[v_{\max} \left(1 - \frac{\rho}{\rho_{\max}} \right) \rho \right] = 0. \quad (3.18)$$

We are interested in answering the question of whether the partial differential equation in (3.18) has a solution satisfying an initial condition of the form in (3.12); namely,

$$\rho(x, 0) = \rho_o(x), \quad (3.19)$$

where ρ_o is some given continuous function of x .

Before we proceed with the analysis, we nondimensionalize the equation in (3.18) by introducing dimensionless variables

$$u = \frac{\rho}{\rho_{\max}}, \quad \widehat{x} = \frac{x}{L}, \quad \text{and} \quad \widehat{t} = \frac{t}{\tau}, \quad (3.20)$$

where L and τ are some characteristic lengths and time scaling factors, respectively. It follows from the Chain Rule and the last two equations in (3.20) that

$$\frac{\partial \rho}{\partial x} = \frac{\partial \rho}{\partial \widehat{x}} \frac{\partial \widehat{x}}{\partial x},$$

or

$$\frac{\partial \rho}{\partial x} = \frac{1}{L} \frac{\partial \rho}{\partial \widehat{x}}, \quad (3.21)$$

and

$$\frac{\partial \rho}{\partial t} = \frac{1}{\tau} \frac{\partial \rho}{\partial \widehat{t}}. \quad (3.22)$$

Substituting the equations in (3.21) and (3.22) into (3.18) and using the first equation in (3.20), we obtain

$$\frac{1}{\tau} \frac{\partial \rho}{\partial \widehat{t}} + \frac{v_{\max}}{L} \frac{\partial}{\partial \widehat{x}} [(1-u)\rho] = 0. \quad (3.23)$$

Next, divide the equation in (3.23) and use the first equation in (3.20) to get

$$\frac{\partial u}{\partial \widehat{t}} + \frac{v_{\max} \tau}{L} \frac{\partial}{\partial \widehat{x}} [(1-u)u] = 0, \quad (3.24)$$

where we have also multiplied by τ . Setting

$$\frac{v_{\max} \tau}{L} = 1,$$

or, equivalently, choosing the time scale τ to be L/v_{\max} , we see that we can re-write (3.24) as

$$\frac{\partial u}{\partial \widehat{t}} + \frac{\partial}{\partial \widehat{x}} [(1-u)u] = 0. \quad (3.25)$$

The equation in (3.25) is now in dimensionless form. If the original time and space variables, t and x , are assumed to be in units of τ and L , respectively, we can then rewrite the equation in (3.25) as

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} [(1-u)u] = 0, \quad (3.26)$$

where u , x and t represent real (dimensionless) variables. The equation in (3.26) is the one we'll be analyzing for the remainder of this section.

Set

$$g(u) = u(1-u); \quad (3.27)$$

then, the partial differential equation in (3.26) can be written in the form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x} [g(u)] = 0, \quad (3.28)$$

or

$$\frac{\partial u}{\partial t} + g'(u) \frac{\partial u}{\partial x} = 0. \quad (3.29)$$

We will describe here a general procedure for analyzing the equation in (3.28) for the case in which g is a differentiable function. We begin by presenting the simplest example of a linear function

$$g(u) = cu,$$

for some constant c . The equation in (3.29) then becomes the linear first order partial differential equation

$$\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0. \quad (3.30)$$

We will show how to find a solution to the differential equation in (3.30) subject to the initial condition

$$u(x, 0) = f(x), \quad (3.31)$$

where f is some differentiable function of a single variable.

The problem of determining a solution of the differential equation in (3.30) subject to the condition in (3.31) is an example of initial value problem and is usually written in a the more compact form

$$\begin{cases} \frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} = 0, & x \in \mathbb{R}, t > 0; \\ u(x, 0) = f(x), & x \in \mathbb{R}. \end{cases} \quad (3.32)$$

Example 3.2.1 (The Method of Characteristic Curves). We can arrive at a solution to the initial value problem in (3.32) by using the method of characteristic curves.

A characteristic curve in the xt -plane, parametrized by

$$t \mapsto (x(t), t), \quad \text{for } t \in \mathbb{R}, \quad (3.33)$$

is obtained as follows. First, evaluate the function u on the curve in (3.33) to obtain a real-valued function of a single variable

$$t \mapsto u(x(t), t), \quad \text{for } t \in \mathbb{R}. \quad (3.34)$$

Differentiate with respect to t the function defined in (3.34), using the Chain Rule, to obtain

$$\frac{d}{dt} [u(x(t), t)] = \frac{\partial u}{\partial x} \frac{dx}{dt} + \frac{\partial u}{\partial t} \frac{dt}{dt},$$

which we can re-write as

$$\frac{du}{dt} = \frac{\partial u}{\partial t} + \frac{dx}{dt} \frac{\partial u}{\partial x}. \quad (3.35)$$

Comparing the right-hand side of the equation in (3.35) with the left-hand side of the partial differential equation in (3.32), we see that if we choose the curve in (3.33) so that

$$\frac{dx}{dt} = c, \quad (3.36)$$

then the equation in (3.35) turns into

$$\frac{du}{dt} = 0. \quad (3.37)$$

The ordinary differential equation in (3.36) defines a family of curves in the xt -plane given by

$$x = ct + k, \quad (3.38)$$

where k is a real parameter. The curves in (3.38) are straight lines of slope $1/c$ in the xt -plane; some of the curves for the case $c > 0$ are pictured in Figure 3.2.3.

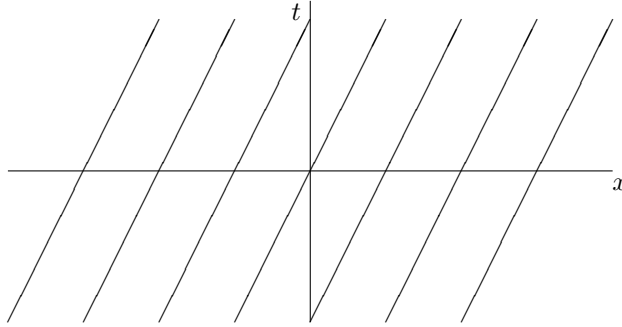


Figure 3.2.3: Characteristic Curves for (3.32)

The curves in (3.38) are called the **characteristic curves** of the partial differential equation in (3.32) and are defined by the ordinary differential equation in (3.36).

Since the equation in (3.37) was derived by differentiating u along a characteristic curve, it implies that u is constant along characteristics. We can therefore conclude from (3.37) that

$$u = \text{constant along characteristics.} \quad (3.39)$$

The equation in (3.39) allows us to obtain a formula for $u(x, t)$, where (x, t) lies along the characteristic indexed by k in (3.38) as follows

$$u(x, t) = \varphi(k), \quad (3.40)$$

where $\varphi(k)$ denotes the constant value of u along the characteristic in (3.38) indexed by k .

Solving for k in (3.38) and substituting into (3.40) yields a general formula for computing a solution to the partial differential equation in (3.32):

$$u(x, t) = \varphi(x - ct). \quad (3.41)$$

Substituting the initial condition (3.32) into (3.41) yields

$$\varphi(x) = f(x), \quad \text{for all } x \in \mathbb{R},$$

so that

$$u(x, t) = f(x - ct), \quad \text{for } x \in \mathbb{R}, t \in \mathbb{R}, \quad (3.42)$$

solves the initial value problem (3.32). The expression in (3.42) says that the solution to (3.32) is a **traveling wave**, which moves to the right with speed c if $c > 0$, or to the left if $c < 0$. In other words, the initial profile, $f(x)$, moves propagates without distortion with velocity c . The solution in (3.42) is also known as an **advection wave** and the partial differential equation in (3.32) is known as the **advection equation**.

The method of characteristic curves illustrated in Example 3.2.1 also applies to nonlinear equations. The analysis is more involved, but in some cases it yields a lot of information about solutions to the equation. In the next example we consider the case in which

$$g(u) = \frac{1}{2}u^2 \quad (3.43)$$

in (3.28). Thus, in view of (3.29) and (3.43), the partial differential equation in (3.28) becomes

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0. \quad (3.44)$$

The partial differential equation in (3.44) is known as the inviscid Burgers' equation.

Example 3.2.2 (Inviscid Burgers' Equation). Solve the equation in (3.44) subject to the initial condition

$$u(x, 0) = f(x),$$

where f is some given continuous function; in other words, solve the initial value problem

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = 0, & x \in \mathbb{R}, t > 0; \\ u(x, 0) = f(x), & x \in \mathbb{R}. \end{cases} \quad (3.45)$$

We proceed as in Example 3.2.1 by first obtaining the equation for the characteristic curves

$$\frac{dx}{dt} = u. \quad (3.46)$$

In this case we cannot solve directly for the characteristic curves. However, as in Example 3.2.1, the a solution, u , to the partial differential equation in (3.45) must solve the ordinary differential equation

$$\frac{du}{dt} = 0; \quad (3.47)$$

Thus, u must be constant along the characteristic curves given by (3.46). Thus, we can solve (3.47) to yield

$$u(x, t) = \varphi(k), \quad (3.48)$$

where $\varphi(k)$ denotes the constant value of u along the characteristic curve given by (3.46) indexed by k . We can then re-write (3.46) as

$$\frac{dx}{dt} = \varphi(k),$$

which can be solved to yield the characteristic curves

$$x = \varphi(k)t + k. \quad (3.49)$$

Hence, according to (3.49), the characteristic curves of the partial differential equation in (3.45) are straight lines in the xt -plane whose slopes depend on the value of u on the them.

Solving for the right-most k in (3.49) and substituting into (3.48), we obtain an expression that defines u implicitly

$$u(x, t) = \varphi(x - u(x, t)t), \quad (3.50)$$

where we have used the expression for $\varphi(k)$ in (3.48).

Finally, employing the initial condition in (3.45), we obtain from (3.50) that

$$u(x, t) = f(x - u(x, t)t). \quad (3.51)$$

In subsequent examples we will see realizations of the implicit formula in (3.51) for specific initial conditions f .

Example 3.2.3 (Inviscid Burgers' Equation, Continued). Solve the initial value problem in (3.45) where the initial condition is given by

$$f(x) = x, \quad \text{for all } x \in \mathbb{R}. \quad (3.52)$$

In this case the expression in (3.51) defining u implicitly reads

$$u(x, t) = x - u(x, t)t, \quad (3.53)$$

in view of (3.52). Solving the expression in (3.53) for $u(x, t)$ yields

$$u(x, t) = \frac{x}{1+t}, \quad \text{for all } x \in \mathbb{R} \text{ and } t \geq 0. \quad (3.54)$$

Example 3.2.4 (Inviscid Burgers' Equation, Continued). Solve the initial value problem in (3.45) where the initial condition is given by

$$f(x) = 1 - x, \quad \text{for all } x \in \mathbb{R}. \quad (3.55)$$

In this case the expression in (3.51) defining u implicitly reads

$$u(x, t) = 1 - (x - u(x, t)t), \quad (3.56)$$

in view of (3.55). Solving the expression in (3.56) for $u(x, t)$ yields

$$u(x, t) = \frac{1 - x}{1 - t}, \quad \text{for all } x \in \mathbb{R} \text{ and } 0 \leq t < 1. \quad (3.57)$$

It is interesting to notice the strikingly different results obtained in Examples 3.2.3 and 3.2.4. In Example (3.2.4), the solutions, $u(x, t)$ in (3.57) ceases to exist at $t = 1$; while the solution in Example 3.2.3 exists for all positive values of t , according to (3.54). In subsequent examples we try to explain why the two examples display vastly different results by examining the method of characteristic curves more carefully.

Example 3.2.5 (Inviscid Burgers' Equation, Continued). Solve the initial value problem in (3.45) where the initial condition is given by

$$f(x) = \begin{cases} 0, & \text{if } x < 0; \\ x, & \text{if } 0 \leq x < 1; \\ 1, & \text{if } x \geq 1. \end{cases} \quad (3.58)$$

Figure 3.2.4 shows a sketch of the initial condition, f . The characteristic curves

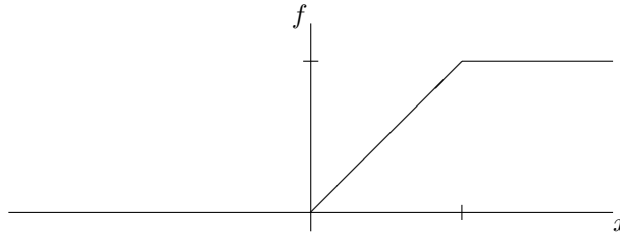


Figure 3.2.4: Sketch of the graph of f in (3.58)

for the differential equation in (3.45) are solutions to the ordinary differential equation

$$\frac{dx}{dt} = u. \quad (3.59)$$

Along the characteristic curves, u solves the ordinary differential equation

$$\frac{du}{dt} = 0, \quad (3.60)$$

which implies that u is constant along characteristics so that

$$u(x, t) = \varphi(k), \quad (3.61)$$

for some parameter k , depending on the particular characteristic curve, and φ is some arbitrary real-valued function. Solving (3.59) for u given by (3.61) we obtain the equation for the characteristic curves:

$$x = \varphi(k)t + k. \quad (3.62)$$

Next, solve for k in (3.62) and substitute in (3.61) to obtain that u is given implicitly by

$$u(x, t) = \varphi(x - u(x, t)t). \quad (3.63)$$

Using the initial condition in (3.45), we obtain from (3.63) that

$$\varphi(x) = f(x), \quad \text{for all } x \in \mathbb{R}, \quad (3.64)$$

so that

$$u(x, t) = f(x - u(x, t)t). \quad (3.65)$$

It follows from (3.62) and (3.64) that the characteristic curves are given by the equations

$$x = f(k)t + k; \quad (3.66)$$

thus, according to (3.66), the characteristic curves of the partial differential equation in (3.45) are straight lines of slope $1/f(k)$ going through the point $(k, 0)$ on the x -axis. In particular, for $k \leq 0$ the characteristic curves are the vertical lines

$$x = k,$$

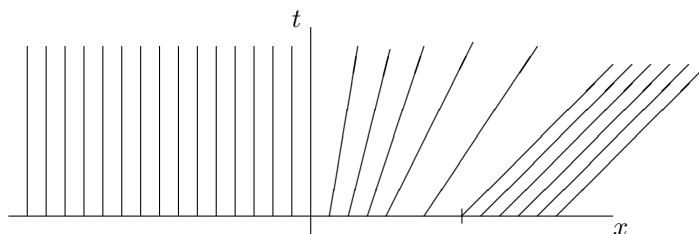
since $f(k) = 0$ for $k \leq 0$; for $0 < k < 1$, the characteristic curves are the straight lines

$$x = k(t + 1);$$

and, for $k \geq 1$, the characteristic lines are the straight lines of slope 1

$$x = t + k.$$

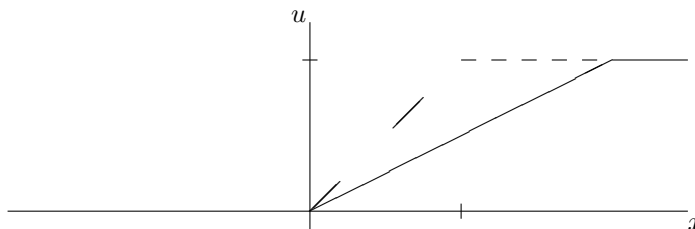
A sketch of The characteristic curves is shown in Figure 3.2.5. Notice that the characteristic curves for $0 \leq k \leq 1$ fan out from the t -axis to the line $x = t + 1$. Since the solution, u , to the initial value problem in (3.45) is constant along characteristic curves (see the equations in (3.60) and (3.61)), the sketch in Figure 3.2.5 shows that u can be computed by traveling back along the characteristic curves to the initial time, $t = 0$, and reading off value of $f(k)$ for the particular value point k on the x -axis. Thus, in theory, the initial value problem in (3.45) with initial condition given in (3.58) can be solved, and the solution is unique.

Figure 3.2.5: Sketch of the characteristic curves for (3.45) with f given in (3.58)

In this case we can write down a formula for $u(x, t)$:

$$u(x, t) = \begin{cases} 0, & \text{if } x \leq 0, t \geq 0; \\ \frac{x}{1+t}, & \text{if } 0 < x < t+1, t \geq 0; \\ 1, & \text{if } x \geq t+1, t \geq 0. \end{cases} \quad (3.67)$$

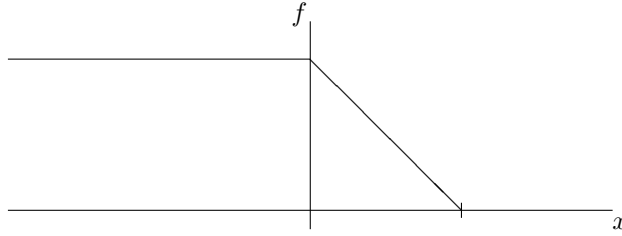
The fanning out of the characteristic curves pictured in Figure 3.2.5 has the effect of stretching the initial profile for u in the x direction. This is shown in Figure 3.2.6, where sketch of $u(x, t)$ as given in (3.67) is shown for $t = 0$ and $t = 1$; the initial profile is shown in dashed lines.

Figure 3.2.6: Sketch of the graph of $u(x, t)$ for $t = 0$ and $t = 1$

The nature of the solutions to the initial value problem in (3.45) changes dramatically when the following initial profile is used.

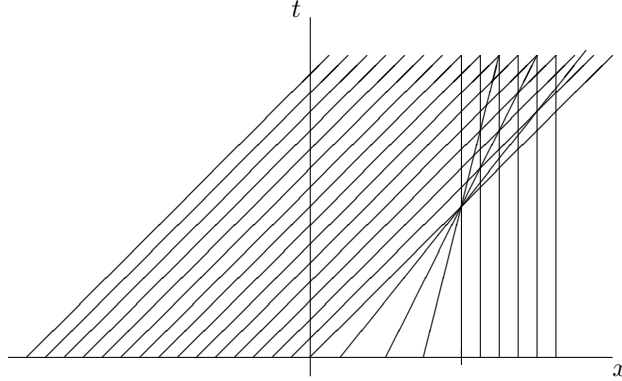
$$f(x) = \begin{cases} 1, & \text{if } x < 0; \\ 1-x, & \text{if } 0 \leq x < 1; \\ 0, & \text{if } x \geq 1. \end{cases} \quad (3.68)$$

A sketch of the graph of f in (3.68) is shown in Figure 3.2.7.

Figure 3.2.7: Sketch of the graph of f in (3.68)

Example 3.2.6 (Inviscid Burgers' Equation, Continued). Solve the initial value problem in (3.45) where f is as given in (3.68).

Proceeding as in Example 3.2.6, we sketch the characteristic curves in Figure 3.2.8. In Figure 3.2.8 we see that the characteristic curves,

Figure 3.2.8: Sketch of the characteristic curves for (3.45) with f given in (3.68)

$$x = f(k)t + k, \quad (3.69)$$

for $0 < k < 1$, instead of fanning out, bunch in and all meet at the single point with coordinates $(1, 1)$ in the xt -plane. To see why this is the case, take two of the characteristic curves in (3.69) with equations

$$x = f(k_1)t + k_1, \quad (3.70)$$

and

$$x = f(k_2)t + k_2, \quad (3.71)$$

with $0 < k_1 < k_2 \leq 1$. To find the intersection of the lines in (3.70) and (3.71), set the equations equal to each other and use the definition of f in (3.68), so that $f(k) = 1 - k$, for $0 < k \leq 1$, to get that

$$(1 - k_1)t + k_1 = (1 - k_2)t + k_2,$$

from which we get that $t = 1$. Thus, $u(x, t)$ ceases to exist in the usual sense at $t = 1$.

As in Example 3.2.5, we can obtain a formula for computing $u(x, t)$, at least for $t < 1$:

$$u(x, t) = \begin{cases} 1, & \text{if } x < t < 1; \\ \frac{1-x}{1-t}, & \text{if } t < x \leq 1; \\ 0, & \text{if } x > 1, t < 1. \end{cases} \quad (3.72)$$

Figure 3.2.9 shows a picture of the graph of $u(x, t)$, for $t = 1/2$, as a function x . As t approaches 1 from the left, we see from (3.72) that the profile of $u(x, t)$

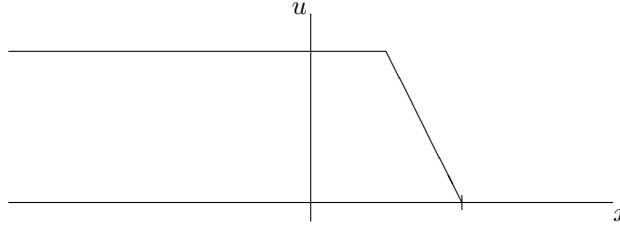


Figure 3.2.9: Sketch of the graph of $u(x, t)$ in (3.72) for $t = 1/2$

approaches that shown in Figure 3.2.10. Thus, as $t \rightarrow 1^-$, $u(x, t)$ develops a

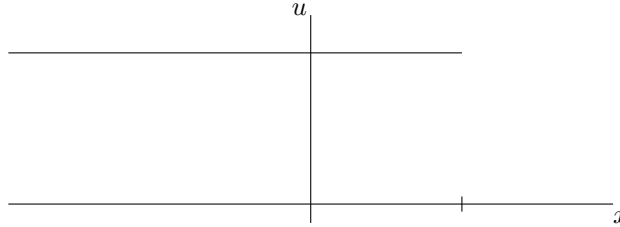


Figure 3.2.10: Sketch of the graph of $u(x, t)$ in (3.72) as $t \rightarrow 1^-$

discontinuity.

As seen in Example 3.2.6, a solution to the initial value problem in (3.45), where f is as given in (3.68), ceases to exist in the usual sense. However, some sort of solution, known as a **shock wave**, does exist in a generalized sense. This generalized solution, or weak solutions as it is also called, will not solve the partial differential equation, but it will solve the integral equation formulation of the conservation principle that led to the partial differential equation. For

the case of the inviscid Burgers' equation, the conservation principle is

$$\frac{d}{dt} \int_a^b u(x, t) dx = \text{Flux at } a - \text{Flux at } b, \quad (3.73)$$

for all $a, b \in \mathbb{R}$ with $a < b$. In the case of Burgers' equation, the flux is given by

$$F(x, t) = \frac{1}{2}[u(x, t)]^2 \quad (3.74)$$

Combining (3.73) and (3.74), the equation in (3.73) can be written as

$$\frac{d}{dt} \int_a^b u(x, t) dx = \frac{1}{2}[u(a, t)]^2 - \frac{1}{2}[u(b, t)]^2, \quad (3.75)$$

for all $a, b \in \mathbb{R}$ with $a < b$.

The conservation principle in (3.75) can be used to describe what happens to the solution after a shock forms; for instance, after $t = 1$ in Example 3.2.6. We saw in that example that a discontinuity develops. The discontinuity will continue to travel along some curve in the xt -plane parametrized by a path of the form

$$t \mapsto (\sigma(t), t). \quad (3.76)$$

We would like to describe the path in (3.76). In order to do this, we assume that the path in (3.76) is differentiable with continuous derivative (in other words, the path in (3.76) is a C^1 path). We also assume that u has a jump discontinuity along the path in (3.76), so that the one sided limits

$$u^-(t) = \lim_{x \rightarrow \sigma(t)^-} u(x, t) \quad \text{and} \quad u^+(t) = \lim_{x \rightarrow \sigma(t)^+} u(x, t) \quad (3.77)$$

For instance, in Example 3.2.6 we saw that the solution to the inviscid Burgers' equation with the initial condition in (3.68) has a jump discontinuity at $t = 1$ with

$$u^-(1) = 1 \quad \text{and} \quad u^+(1) = 0. \quad (3.78)$$

The conservation expression in (3.73) states that the quantity

$$Q(a, b, t) = \int_a^b u(x, t) dx, \quad (3.79)$$

is conserved. Using (3.79), we can re-write (3.75)

$$\frac{d}{dt}[Q(a, b, t)] = \frac{1}{2}[u(a, t)]^2 - \frac{1}{2}[u(b, t)]^2. \quad (3.80)$$

Consider the quantity $Q(a, b, t)$ over a short time interval $[t, t + \Delta t]$, where $a = \sigma(t)$ and $b = \sigma(t + \Delta t)$. We have by the definition of Q in (3.79) that

$$Q(\sigma(t), \sigma(t + \Delta t), t + \Delta t) = \int_{\sigma(t)}^{\sigma(t + \Delta t)} u(x, t + \Delta t) dx \quad (3.81)$$

Using (3.77) we can approximate (3.81) as follows

$$Q(\sigma(t), \sigma(t + \Delta t), t + \Delta t) \approx u^-(t)(\sigma(t + \Delta t) - \sigma(t)) \quad (3.82)$$

since $u(x, t + \Delta t)$ is to the left of the shock for $\Delta t > 0$ (see Figure 3.2.11). Similarly, we get from (3.79) and (3.77) that

$$Q(\sigma(t), \sigma(t + \Delta t), t) \approx u^+(t)(\sigma(t + \Delta t) - \sigma(t)) \quad (3.83)$$

It follows from (3.82) and (3.83) that

$$\begin{aligned} & \frac{Q(\sigma(t), \sigma(t + \Delta t), t + \Delta t) - Q(\sigma(t), \sigma(t + \Delta t), t)}{\Delta t} \\ & \approx [u^-(t) - u^+(t)] \frac{\sigma(t + \Delta t) - \sigma(t)}{\Delta t}. \end{aligned} \quad (3.84)$$

Letting $\Delta t \rightarrow 0$ in (3.84) we obtain

$$\frac{dQ}{dt} = [u^-(t) - u^+(t)] \frac{d\sigma}{dt} \quad (3.85)$$

across the shock.

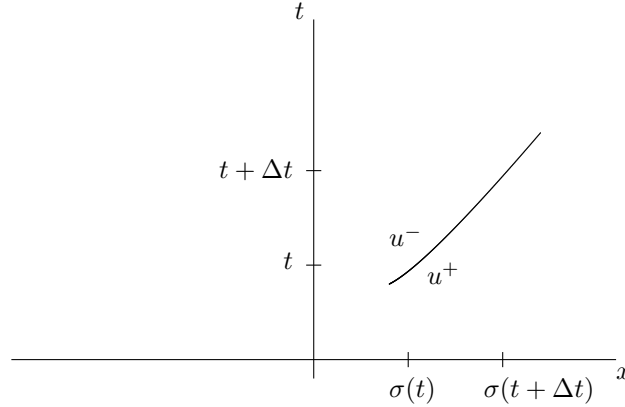


Figure 3.2.11: Path of a shock solution for (3.45) with f given in (3.68)

On the other hand, for any $\tau \in (t, t + \Delta t)$ we have that

$$\lim_{\Delta t \rightarrow 0} \left[\frac{1}{2} [u(\sigma(t), \tau)]^2 - \frac{1}{2} [u(\sigma(t + \Delta t), \tau)]^2 \right] = \frac{1}{2} [u^-(t)]^2 - \frac{1}{2} [u^-(t)]^2 \quad (3.86)$$

(see Figure 3.2.11).

Next, combine (3.85), (3.86) and the conservation principle in (3.75) to get

$$[u^-(t) - u^+(t)] \frac{d\sigma}{dt} = \frac{1}{2} [u^-(t)]^2 - \frac{1}{2} [u^-(t)]^2,$$

from which we get that

$$\frac{d\sigma}{dt} = \frac{u^-(t) + u^+(t)}{2} \quad (3.87)$$

The differential equation in (3.87) gives the path that the jump discontinuity will follow after the shock forms.

Example 3.2.7 (Inviscid Burgers' Equation, Continued). We saw in Example 3.2.6 that the solution to the initial value problem in (3.45) where f is as given in (3.68) develops a shock at $t = 1$. The solution after $t = 1$ has a jump discontinuity that travels along the path

$$t \mapsto (\sigma(t), t)$$

given by (3.87) where u^+ and u^- are given in (3.78). Consequently,

$$\frac{d\sigma}{dt} = \frac{1}{2},$$

so that

$$\sigma(t) = \frac{t}{2} + c. \quad (3.88)$$

Since $\sigma(1) = 1$, we get from (3.88) that

$$c = \frac{1}{2}$$

in (3.88). We therefore get the following formula for the weak solution to initial value problem discussed in Example 3.2.6:

$$u(x, t) = \begin{cases} 1, & \text{if } x < \frac{t+1}{2}, t \geq 1; \\ 0, & \text{if } x > \frac{t+1}{2}, t \geq 1. \end{cases} \quad (3.89)$$

Figure 3.2.12 shows a picture of some of the characteristic curves for (3.45) with f given in (3.68), which also incorporates the shock wave solution that we derived in (3.89).

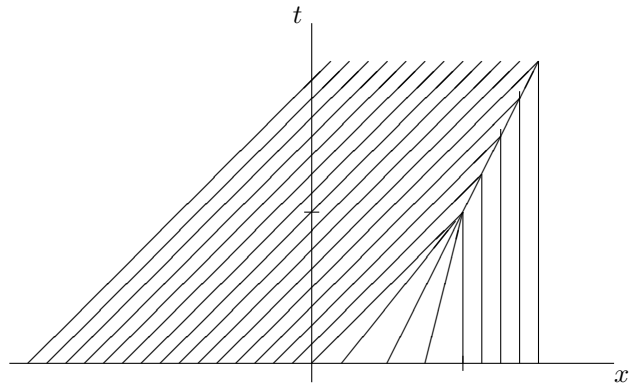


Figure 3.2.12: Sketch of the characteristic curves for (3.45) with f given in (3.68) with shock wave solution

Chapter 4

Stochastic Models

The models discussed in the previous chapter have been deterministic—the variables used in the models are completely determined by the values of a set of parameters and the values of the variables at some initial point or curve. For instance, when modeling bacterial growth, the number of bacteria in a culture at time, t , might be modeled by a continuous variable, $N(t)$, which satisfies the initial value problem

$$\begin{cases} \frac{dN}{dt} = rN \left(1 - \frac{N}{K}\right); \\ N(0) = N_o, \end{cases} \quad (4.1)$$

known as the logistic model. The model in (4.1) was derived in Assignment #1 as a special case of the bacterial growth in a chemostat model presented in Section 2.2 of these notes. The parameters in the equation in (4.1) are the intrinsic growth rate, r , and the carrying capacity of the growth medium, K . Given values for the parameters r and K , and the size of the bacterial population, N_o , at time $t = 0$, the values of population size, $N(t)$, for later values of t , are completely determined by the formula

$$N(t) = \frac{N_o K}{N_o + (K - N_o)e^{-rt}}, \quad \text{for } t \geq 0. \quad (4.2)$$

Thus, if an experiment is performed in which N_o bacteria, whose intrinsic growth rate is r , are placed in a growth medium with carrying capacity, K , at time $t = 0$, then the number of bacteria, $N(t)$, in the culture at time $t > 0$ will be given by (4.2). The experiment may be repeated many times; if the same initial condition holds, and the values of the parameters are the same, the same population value, $N(t)$, at some time later will be that given by (4.2).

The Logistic growth scenario described in the previous paragraphs is to be contrasted with the situation in which, in addition to finding out how many bacteria, $N(t)$, are present at time t , we want to know how many of those

bacteria develop a certain type of mutation. Here, an element of chance needs to be introduced in the model. Each of the $N(t)$ bacteria present at time t might or might not develop of a mutation in a short time interval $[t, t + \Delta t]$; it is not possible to predict with certainty whether a given bacterium will develop a mutation. Thus, if we let $M(t)$ be the number of bacteria that develop a mutation during the time interval $[0, t]$, then every time we run the experiment of placing N_o bacteria in a culture at time $t = 0$, under the same conditions, and count the number of bacteria that have developed the mutation at time t , we will not get the same value for $M(t)$. Thus, $M(t)$ is not a *function* in the usual sense that we understand that term. After a time interval of length t , $M(t)$ can take on a range of values, and each value has a certain likelihood or probability of occurring. This notion of a “function” $M(t)$ whose values cannot be predicted, but for which we can obtain a measure of their likelihood is what is known as a *random variable*.

4.1 Random Variables

If $M(t)$ denotes the number of bacteria that develop a mutation from an initial number N_o in a time interval $[0, t]$, it is reasonable to model it as a random variable. Roughly speaking, random variables are quantities that are determined from outcomes of a *random experiment*. A random experiment is a process which can be repeated indefinitely under the same set of conditions, but whose outcome cannot be predicted with certainty before the experiment is performed. For instance, suppose you start with one bacterium in a medium conducive to growth; t units of time later we count how many out of the $N(t)$ bacteria have developed a mutation. The number of bacteria, $M(t)$, that have developed the mutation is a random variable.

Even though we are not able to predict with certainty what specific value the random variable, $M(t)$, will take on at time t , in many modeling situations we are able to ascertain the likelihood, or probability, that $M(t)$ will take on a range of values.

Example 4.1.1. Suppose that two bacteria, a and b , can randomly develop a mutation in a unit of time. Assume that each bacterium can mutate at most once in the unit of time of the experiment. Let M denote the number of bacteria out of the two that develop mutations after one unit of time. Then M can take on the values 0, 1, or 2. We cannot predict precisely what value M will take on. Any time we run the experiment of placing the two bacteria under observation and counting the number of mutations we may get any of the possible values. M is thus an example of a random variable. The best we can hope for is an estimate of the probabilities that M can take on any of the possible values; in symbols, we want to estimate

$$\Pr[M = k], \quad \text{for } k = 0, 1, 2, \quad (4.3)$$

4.1.1 A Brief Excursion into Probability

The expression $\Pr[M = k]$ in (4.3) denotes the probability of the *event* $[M = k]$; that is, $\Pr[M = k]$ gives a measure of the likelihood that k bacteria out of the two in Example 4.1.1 will develop a mutation.

An event is a possible outcome, or set of outcomes, of a random experiment. In Example 4.1.1, the event denoted by $[M = k]$ represents a set of outcomes in which k of the two bacteria have developed a mutation.

A probability function assigns a real value between 0 and 1 to an event. A probability of 0 means an impossible event, and a probability of 1 means that the event will surely happen. The assignments of probability between 0 and 1 will depend on assumptions made about the experiment at hand.¹

In order to compute the probabilities of the events $[M = k]$, for $k = 0, 1, 2$, in Example 4.1.1, we need to make some assumptions regarding how mutations occur. Let A denote the event that bacterium a develops a mutation and B the event that bacterium B develops a mutation in one unit of time. Suppose we are told that the probability that a bacterium will develop a mutation is p , where $0 < p < 1$ (p is called the *mutation rate*). We then have that

$$\Pr(A) = p \quad \text{and} \quad \Pr(B) = p. \quad (4.4)$$

We assume that the event that A occurs will not affect the probability of event B . We say that A and B are stochastically independent.

Definition 4.1.2 (Stochastic Independence). We say that events A and B are stochastically independent if the probability of the joint occurrence of A and B is the product of the individual probabilities. In symbols,

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B), \quad (4.5)$$

where $A \cap B$ denotes the event that both A and B happen jointly.

In Example 4.1.1, $A \cap B$ corresponds to the event that both bacteria develop a mutation in a unit of time. Thus,

$$A \cap B = [M = 2].$$

Thus, the independence assumption implies that

$$\Pr[M = 2] = \Pr(A) \cdot \Pr(B) = p \cdot p = p^2, \quad (4.6)$$

where we have used (4.5) and (4.4).

We next see how to compute $\Pr[M = 0]$ and $\Pr[M = 1]$ in Example 4.1.1.

Definition 4.1.3 (Complement of an Event). Given an event, A , the event that A does not occur is called the complement of A and is denoted by A^c .

¹For example, in the experiment of tossing a “fair die,” it is assumed that all faces of the die are equally likely; thus, the probability of any given face is $1/6$.

Thus, in Example 4.1.1, A^c is the event that bacterium a does not develop a mutation in one unit of time. Observe that A and A^c are mutually exclusive; that is, if A occurs then A^c cannot occur.

Definition 4.1.4 (Mutually Exclusive Events). Events A and B are said to be mutually exclusive if and only if $A \cap B = \emptyset$.

Definition 4.1.5. Given events A and B , the symbol $A \cup B$ denotes the event that either A or B occurs.

Definition 4.1.6 (Probability of Mutually Exclusive Events). If A and B are mutually exclusive, then

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \quad (4.7)$$

For example, since A and A^c are mutually exclusive, it follows that

$$\Pr(A \cup A^c) = \Pr(A) + \Pr(A^c).$$

On the other hand, $\Pr(A \cup A^c) = 1$, since $A \cup A^c$ is a sure event. It then follows that

$$\Pr(A) + \Pr(A^c) = 1,$$

from which we get the following property of the probability function.

Proposition 4.1.7 (Probability of Complementary Event).

$$\Pr(A^c) = 1 - \Pr(A). \quad (4.8)$$

Definition 4.1.8 (Set Difference). Given any events A and B , we define the set

$$A \setminus B = \{x \in A \mid x \notin B\}.$$

Note that $A \setminus B$ and B are mutually exclusive, by Definition 4.1.8. Furthermore,

$$A = B \cup (A \setminus B). \quad (4.9)$$

We therefore have the following proposition.

Proposition 4.1.9 (Probability of Difference of Events). Given events A and B , with $B \subseteq A$,

$$\Pr(A \setminus B) = \Pr(A) - \Pr(B). \quad (4.10)$$

Proof: It follows from (4.9) and Definition 4.1.6 that

$$\Pr(A) = \Pr(B) + \Pr(A \setminus B),$$

which implies (4.10). ■

As a consequence of the properties of probability that we have discussed so far, we have the following proposition.

Proposition 4.1.10 (Probability of Union of Events). For any two events, A and B ,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \quad (4.11)$$

Proof: Note that $A \cap B \subseteq A$. Thus, applying Proposition 4.1.9,

$$\Pr(A \setminus (A \cap B)) = \Pr(A) - \Pr(A \cap B), \quad (4.12)$$

where

$$\begin{aligned} x \in A \setminus (A \cap B) & \text{ iff } x \in A \text{ and } x \notin A \cap B \\ & \text{ iff } x \in A \text{ and } x \notin B \\ & \text{ iff } x \in A \cap B^c, \end{aligned}$$

so that

$$A \setminus (A \cap B) = A \cap B^c. \quad (4.13)$$

Substituting (4.13) into (4.12) then yields

$$\Pr(A \cap B^c) = \Pr(A) - \Pr(A \cap B). \quad (4.14)$$

Similar calculations show that

$$\Pr(B \cap A^c) = \Pr(B) - \Pr(A \cap B). \quad (4.15)$$

Observing that

$$A \cup B = (A \cap B^c) \cup (B \cap A^c) \cup (A \cap B), \quad (4.16)$$

where $A \cap B^c$, $B \cap A^c$ and $A \cap B$ are mutually exclusive, we get from (4.16) and Definition 4.1.6 that

$$\Pr(A \cup B) = \Pr(A \cap B^c) + \Pr(B \cap A^c) + \Pr(A \cap B). \quad (4.17)$$

Combining (4.14), (4.15) and (4.17) yields (4.11). ■

Example 4.1.11 (Continuation of Example 4.1.1.). Let A , B and M be as in Example 4.1.1. In this example we compute $\Pr[M = 0]$ and $\Pr[M = 1]$.

The probability that bacterium a will not mutate is

$$\Pr(A^c) = 1 - p, \quad (4.18)$$

where we have used (4.8) and (4.4). Likewise, the probability that bacterium b will not mutate is

$$\Pr(B^c) = 1 - p. \quad (4.19)$$

Since A^c and B^c are independent, it follows from (4.5) and (4.19) that

$$\Pr(A^c \cap B^c) = P(A^c) \cdot P(B^c) = (1 - p) \cdot (1 - p) = (1 - p)^2.$$

In other words, the probability that no mutation occurs is $(1 - p)^2$. We therefore have that

$$\Pr[M = 0] = P(A^c \cap B^c) = (1 - p)^2. \quad (4.20)$$

To compute $P[M = 1]$, first observe that

$$[M = 1] = (A \cap B^c) \cup (A^c \cap B), \quad (4.21)$$

where the events $A \cap B^c$ and $A^c \cap B$ are mutually exclusive. It then follows from (4.7) and (4.21) that

$$P[M = 1] = P(A \cap B^c) + P(A^c \cap B). \quad (4.22)$$

Next, use the independence of the events A and B^c (see Problem 1 in Assignment #7) to compute

$$\begin{aligned} P(A \cap B^c) &= \Pr(A) \cdot \Pr(B^c) \\ &= \Pr(A)(1 - \Pr(B)) \\ &= p(1 - p), \end{aligned} \quad (4.23)$$

where we have used Definition 4.1.2, Proposition 4.1.7 and (4.4). Similarly,

$$P(A^c \cap B) = (1 - p)p. \quad (4.24)$$

Combining (4.22), (4.23) and (4.24) then yields

$$\Pr[M = 1] = 2(1 - p)p. \quad (4.25)$$

In order to compute $\Pr[M = 0]$, first observe that

$$[M = 0] = A^c \cap B^c;$$

so that, by virtue of the independence of A^c and B^c (see Problem 1 in Assignment #7),

$$\begin{aligned} \Pr[M = 0] &= \Pr(A^c) \cdot \Pr(B^c) \\ &= (1 - p)^2. \end{aligned} \quad (4.26)$$

where we have used (4.18) and (4.19).

Putting together the results in (4.6), (4.25) and (4.26) we obtain

$$\Pr[M = k] = \begin{cases} (1 - p)^2 & \text{if } k = 0; \\ 2p(1 - p) & \text{if } k = 1; \\ p^2 & \text{if } k = 2; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.27)$$

4.1.2 Discrete Random Variables

The random variable, M , in Example 4.1.1 is an instance of a discrete random variable.

Definition 4.1.12 (Discrete Random Variable). A random variable, X , which takes on a finite, or countable set of values, x_1, x_2, x_3, \dots , is said to be discrete. The set of probability values

$$p_X(x_k) = \Pr[X = x_k], \quad \text{for } k = 1, 2, 3, \dots$$

is called the probability mass function of X , or simply the probability distribution of X . Observe that

$$\sum_{k=1}^{\infty} p_X(x_k) = 1. \quad (4.28)$$

Example 4.1.13 (Continuation of Example 4.1.1.). The expression in (4.27) is called the probability distribution of the random variable, M , defined in Example 4.1.1. We therefore have that

$$p_M(k) = \begin{cases} (1-p)^2 & \text{if } k = 0; \\ 2(1-p)p & \text{if } k = 1; \\ p^2 & \text{if } k = 2; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.29)$$

Observe that

$$\begin{aligned} \sum_{k=0}^2 p_M(k) &= (1-p)^2 + 2(1-p)p + p^2 \\ &= [(1-p) + p]^2 \\ &= 1, \end{aligned}$$

so that the function defined in (4.29) satisfies the condition (4.28) in the definition of a the distribution of a random variable (Definition 4.1.12).

Definition 4.1.14 (Bernoulli Trials). A random experiment with two mutually exclusive outcomes, one called a “success” and the other a “failure,” is called a Bernoulli trial. We associate a random variable, X , with a Bernoulli as follows: $X = 1$ if the outcome is a success, and $X = 0$ if the outcome is a failure. If the probability of a success is p , then then distribution of X is

$$p_X(k) = \begin{cases} 1-p & \text{if } k = 0; \\ p & \text{if } k = 1; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.30)$$

The random variable X is said to have a Bernoulli distribution with parameter p . We write $X \sim \text{Bernoulli}(p)$.

Example 4.1.15 (Continuation of Example 4.1.1.). The bacterial mutation situation described in Example 4.1.1 may be modeled by using two Bernoulli trials, X_1 and X_2 , with parameter p , where p is the probability that a bacterium will develop a mutation in the time interval $[0, 1]$. Event A is then $[X_1 = 1]$, while event B is the event $[X_2 = 1]$.

Definition 4.1.16 (Independent Discrete Random Variables). Two discrete random variables, X and Y , are said to be stochastically independent if and only if

$$\Pr(X = a, Y = b) = \Pr(X = a) \cdot \Pr(Y = b), \quad (4.31)$$

for all values of a and b .

Example 4.1.17 (Continuation of Example 4.1.1.). If we assume that the Bernoulli random variables, X_1 and X_2 , postulated in Example 4.1.17 as a model for the bacterial mutation situation described in Example 4.1.1, are also stochastically independent, then events $A = [X_1 = 1]$ and $B = [X_2 = 1]$ are independent by virtue of (4.31). We also see that the random variable M , the number of mutations in one unit of time, in the two-bacteria culture, is given by

$$M = X_1 + X_2.$$

Thus, the calculations leading to (4.1.13) in Example 4.1.13 show that, if X_1 and X_2 are independent Bernoulli(p) random variables and $Y = X_1 + X_2$, then the random variable Y has the distribution function

$$p_Y(k) = \begin{cases} (1-p)^2 & \text{if } k = 0; \\ 2(1-p)p & \text{if } k = 1; \\ p^2 & \text{if } k = 2; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.32)$$

4.1.3 The Binomial Distribution

Consider now three bacteria, labeled 1, 2 and 3, and ask the question: How many mutations will there be in a unit of time? As we did in Example 4.1.17, we may postulate three Bernoulli(p) random variables, X_1 , X_2 , and X_3 , where p is the mutation rate. Thus, the event $[X_i = 1]$ is the event that bacterium i will develop a mutation, for $i = 1, 2, 3$. This time, in addition to assuming that X_1 , X_2 and X_3 are pairwise independent, we also need to assume that

$$\Pr(X_1 = a, X_2 = b, X_3 = c) = \Pr(X_1 = a) \cdot \Pr(X_2 = b) \cdot \Pr(X_3 = c),$$

for all values a , b and c . This is the concept of mutual independence.

Definition 4.1.18 (Mutual Independence). Three discrete random variables, X_1 , X_2 and X_3 , are said to be mutually independent if

$$\Pr(X_i = a, X_j = b) = \Pr(X_i = a) \cdot \Pr(X_j = b), \quad \text{for } i \neq j, \quad (4.33)$$

for all values of a and b ; that is, X_1 , X_2 and X_3 are pairwise stochastically independent, and

$$\Pr(X_1 = a, X_2 = b, X_3 = c) = \Pr(X_1 = a) \cdot \Pr(X_2 = b) \cdot \Pr(X_3 = c), \quad (4.34)$$

for all values of a , b and c .

Lemma 4.1.19. *Let X_1 , X_2 and X_3 be mutually independent, discrete random variables and define $Y_2 = X_1 + X_2$. Then, Y_2 and X_3 are stochastically independent.*

Proof: Compute

$$\begin{aligned} \Pr(Y_2 = w, X_3 = z) &= \Pr(X_1 + X_2 = w, X_3 = z) \\ &= \sum_x \Pr(X_1 = x, X_2 = w - x, X_3 = z), \end{aligned}$$

where the summation is taken over all possible values of X_1 . Thus, using (4.34) in Definition 4.1.18,

$$\begin{aligned} \Pr(Y_2 = w, X_3 = z) &= \sum_x \Pr(X_1 = x) \cdot \Pr(X_2 = w - x) \cdot \Pr(X_3 = z) \\ &= \left(\sum_x \Pr(X_1 = x) \cdot \Pr(X_2 = w - x) \right) \cdot \Pr(X_3 = z) \end{aligned}$$

Hence, by pairwise independence, (see (4.33) in Definition 4.1.18),

$$\begin{aligned} \Pr(Y_2 = w, X_3 = z) &= \left(\sum_x \Pr(X_1 = x, X_2 = w - x) \right) \cdot \Pr(X_3 = z) \\ &= \Pr(X_1 + X_2 = w) \cdot \Pr(X_3 = z) \\ &= \Pr(Y_2 = w) \cdot \Pr(X_3 = z), \end{aligned}$$

which shows the independence of Y_2 and X_3 . ■

Example 4.1.20. Suppose X_1 , X_2 and X_3 be three mutually independent Bernoulli random variables with parameter p , where $0 < p < 1$. Define $Y_3 = X_1 + X_2 + X_3$. Find the probability distribution for Y_3 .

Solution: Observe that Y_3 takes on the values 0, 1, 2 and 3, and that

$$Y_3 = Y_2 + X_3,$$

where the probability distribution for Y_2 is given in (4.32).

We compute

$$\begin{aligned} \Pr(Y_3 = 0) &= \Pr(Y_2 = 0, X_3 = 0) \\ &= \Pr(Y_2 = 0) \cdot \Pr(X_3 = 0), \text{ by independence (Lemma 4.1.19),} \\ &= (1 - p)^2 \cdot (1 - p) \\ &= (1 - p)^3. \end{aligned}$$

Next, since the event $(Y_3 = 1)$ consists of the disjoint union of the events $(Y_2 = 1, X_3 = 0)$ and $(Y_2 = 0, X_3 = 1)$,

$$\begin{aligned}\Pr(Y_3 = 1) &= \Pr(Y_2 = 1, X_3 = 0) + \Pr(Y_2 = 0, X_3 = 1) \\ &= \Pr(Y_2 = 1) \cdot \Pr(X_3 = 0) + \Pr(Y_2 = 0) \cdot \Pr(X_3 = 1) \\ &= 2p(1-p)(1-p) + (1-p)^2p \\ &= 3p(1-p)^2,\end{aligned}$$

where we have used Lemma 4.1.19 and the definition of the probability distribution of Y_2 in (4.32). Similarly,

$$\begin{aligned}\Pr(Y_3 = 2) &= \Pr(Y_2 = 2, X_3 = 0) + \Pr(Y_2 = 1, X_3 = 1) \\ &= \Pr(Y_2 = 2) \cdot \Pr(X_3 = 0) + \Pr(Y_2 = 1) \cdot \Pr(X_3 = 1) \\ &= p^2(1-p) + 2p(1-p)p \\ &= 3p^2(1-p),\end{aligned}$$

and

$$\begin{aligned}\Pr(Y_3 = 3) &= \Pr(Y_2 = 2, X_3 = 1) \\ &= \Pr(Y_2 = 0) \cdot \Pr(X_3 = 0) \\ &= p^2 \cdot p \\ &= p^3.\end{aligned}$$

We then have that the probability distribution of Y_3 is given by

$$p_{Y_3}(k) = \begin{cases} (1-p)^3 & \text{if } k = 0, \\ 3p(1-p)^2 & \text{if } k = 1, \\ 3p^2(1-p) & \text{if } k = 2, \\ p^3 & \text{if } k = 3, \\ 0 & \text{elsewhere.} \end{cases} \quad (4.35)$$

□

If we go through the calculations in Examples 4.1.11 and 4.1.20 for the case of four mutually independent² Bernoulli trials with parameter p , where $0 < p < 1$, X_1, X_2, X_3 and X_4 , we obtain that for $Y_4 = X_1 + X_2 + X_3 + X_4$,

$$p_{Y_4}(k) = \begin{cases} (1-p)^4 & \text{if } k = 0, \\ 4p(1-p)^3 & \text{if } k = 1, \\ 6p^2(1-p)^2 & \text{if } k = 2, \\ 4p^3(1-p) & \text{if } k = 3, \\ p^4 & \text{if } k = 4, \\ 0 & \text{elsewhere.} \end{cases} \quad (4.36)$$

²Here, not only do we require that the random variable be pairwise independent, but also that for any group of $k \geq 2$ events $(X_j = x_j)$, the probability of their intersection is the product of their probabilities.

Observe that the terms in the expressions for $p_{Y_2}(y)$, $p_{Y_3}(y)$ and $p_{Y_4}(y)$ in (4.32), (4.35) and (4.36), respectively, are the terms in the expansion of $[(1-p) + p]^n$ for $n = 2, 3$ and 4 , respectively. By the Binomial Expansion Theorem,

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}, \quad \text{for } a, b \in \mathbb{R}, n \in \mathbf{N}, \quad (4.37)$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad k = 0, 1, 2, \dots, n, \quad (4.38)$$

are called the **binomial coefficients**, we obtain that

$$[(1-p) + p]^n = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k}.$$

This suggests that if

$$Y_n = X_1 + X_2 + \dots + X_n,$$

where X_1, X_2, \dots, X_n are n mutually independent Bernoulli trials with parameter p , for $0 < p < 1$, then

$$p_{Y_n}(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n,$$

and $p_{Y_n}(k) = 0$ elsewhere. In fact, this statement is true, and will be proved as the following theorem. We shall establish this as the following theorem:

Theorem 4.1.21. Assume that X_1, X_2, \dots, X_n are mutually independent Bernoulli trials with parameter p , where $0 < p < 1$. Define

$$Y_n = X_1 + X_2 + \dots + X_n.$$

Then the probability distribution of Y_n is

$$p_{Y_n}(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \dots, n; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.39)$$

Proof: We prove this result by induction on n .

For $n = 1$ we have that $Y_1 = X_1$, and therefore

$$p_{Y_1}(0) = \Pr(X_1 = 0) = 1 - p$$

and

$$p_{Y_1}(1) = \Pr(X_1 = 1) = p.$$

Thus,

$$p_{Y_1}(k) = \begin{cases} 1 - p & \text{if } k = 0, \\ p & \text{if } k = 1. \end{cases}$$

Observe that $\binom{1}{0} = \binom{1}{1} = 1$ and therefore the result in (4.39) holds true for $n = 1$.

Next, assume the theorem is true for n ; that is, suppose that

$$p_{Y_n}(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \dots, n, \quad (4.40)$$

and $p_{Y_n}(k) = 0$ elsewhere. We show then show that the result also holds true for $n + 1$. In other words, we show that if $X_1, X_2, \dots, X_n, X_{n+1}$ are mutually independent Bernoulli trials with parameter p , with $0 < p < 1$, and

$$Y_{n+1} = X_1 + X_2 + \dots + X_n + X_{n+1}, \quad (4.41)$$

then, the pmf of Y_{n+1} is

$$p_{Y_{n+1}}(k) = \binom{n+1}{k} p^k (1-p)^{n+1-k} \quad \text{for } k = 0, 1, 2, \dots, n, n+1, \quad (4.42)$$

and $p_{Y_{n+1}}(k) = 0$ elsewhere.

From (4.41) we see that

$$Y_{n+1} = Y_n + X_{n+1},$$

where Y_n and X_{n+1} are independent random variables, by an argument similar to the one in the proof of Lemma 4.1.19 since the X_k 's are mutually independent. Therefore, the following calculations are justified:

(i) for $k \leq n$, or $k < n + 1$,

$$\begin{aligned} \Pr(Y_{n+1} = k) &= \Pr(Y_n = k, X_{n+1} = 0) + \Pr(Y_n = k - 1, X_{n+1} = 1) \\ &= \Pr(Y_n = k) \cdot \Pr(X_{n+1} = 0) \\ &\quad + \Pr(Y_n = k - 1) \cdot \Pr(X_{n+1} = 1) \\ &= \binom{n}{k} p^k (1-p)^{n-k} (1-p) \\ &\quad + \binom{n}{k-1} p^{k-1} (1-p)^{n-k+1} p, \end{aligned}$$

where we have used the inductive hypothesis (4.40). Thus,

$$\Pr(Y_{n+1} = k) = \left[\binom{n}{k} + \binom{n}{k-1} \right] p^k (1-p)^{n+1-k}. \quad (4.43)$$

The expression in (4.42) will follow from (4.43) the fact that

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}, \quad (4.44)$$

which can be established by the following counting argument:

Imagine $n + 1$ balls in a bag, n of which are blue and one is red. We consider the collection of all groups of k balls that can be formed out of the $n + 1$ balls in the bag. This collection is made up of two disjoint sub-collections: the ones with the red ball and the ones without the red ball. The number of elements in the collection with the one red ball is

$$\binom{n}{k-1} \cdot \binom{1}{1} = \binom{n}{k-1}, \quad (4.45)$$

while the number of elements in the collection of groups without the red ball are

$$\binom{n}{k}. \quad (4.46)$$

Adding the amounts in (4.45) and (4.46) must yield $\binom{n+1}{k}$, which proves (4.44).

Thus, combining (4.43) and (4.44), we obtain (4.42) for the case $k < n + 1$.

(ii) If $k = n + 1$, then, using again the independence of Y_n and X_{n+1} ,

$$\begin{aligned} \Pr(Y_{n+1} = k) &= \Pr(Y_n = n, X_{n+1} = 1) \\ &= \Pr(Y_n = n) \cdot \Pr(X_{n+1} = 1) \\ &= p^n p \\ &= p^{n+1} \\ &= \binom{n+1}{k} p^k (1-p)^{n+1-k}, \end{aligned}$$

since $k = n + 1$, and so (4.42) is established for $k = n + 1$.

The proof is now complete. ■

Definition 4.1.22 (Binomial Distribution). A discrete random variable, Y , which counts the number of successes in n independent Bernoulli(p) trials, and having the distribution

$$p_Y(k) = \begin{cases} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \dots, n, \\ 0 & \text{elsewhere.} \end{cases} \quad (4.47)$$

is called a binomial random variable with parameters n and p , where p is the probability of each success. We write

$$Y \sim \text{Binomial}(n, p).$$

Remark 4.1.23. Theorem 4.1.21 shows that a Binomial(n, p) distribution is the sum of n independent Bernoulli(p) trials.

Example 4.1.24 (Bacterial Mutations). Consider a culture containing N bacteria. Let M denote the number of mutations in the culture that develop in the culture in a unit of time. Then, assuming a mutation rate of p (the probability that a given bacterium will develop a mutation in a unit of time), M can be modeled by a binomial random variable with parameters N and p ; that is, M has a probability distribution given by (4.47), where $n = N$:

$$p_M(k) = \begin{cases} \binom{N}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \dots, N; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.48)$$

4.1.4 Expected Value

Definition 4.1.25 (Expected Value of a Discrete Random Variable). Given a discrete random variable, X , with values x_1, x_2, \dots, x_n , and distribution p_X , the weighted average of the values of X ,

$$x_1 p_X(x_1) + x_2 p_X(x_2) + \dots + x_n p_X(x_n),$$

is called the **expected value** of X and is denoted by $E(X)$. We then have that

$$E(X) = \sum_{k=1}^n x_k p_X(x_k) \quad (4.49)$$

Example 4.1.26 (Expected Value of a Bernoulli Trial). Let X denote a Bernoulli random variable with parameter p . Then, the expected value of X is

$$E(X) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p. \quad (4.50)$$

Example 4.1.27 (Expected Value of a Binomial Random Variable). Let

$$X_1, X_2, \dots, X_n$$

denote n independent Bernoulli trials with parameter p , and put $Y = X_1 + X_2 + \dots + X_n$. Then, using the result of Problem 1 in Assignment #8, we have that

$$E(Y) = E(X_1) + E(X_2) + \dots + E(X_n) = np,$$

where we have used (4.50). Thus, the expected value of a binomial random variable, Y , with parameters n and p is

$$E(Y) = np. \quad (4.51)$$

Alternatively, we could have used the definition of p_Y in (4.47) and the formula for the expected value in (4.49) to get

$$E(Y) = \sum_{k=1}^n \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \quad (4.52)$$

Next, make the change of variables $m = k - 1$ in (4.52) to get

$$\begin{aligned} E(Y) &= \sum_{m=0}^{n-1} \frac{n(n-1)!}{m!(n-1-m)!} p^{m+1} (1-p)^{n-1-m} \\ &= np \cdot \sum_{m=0}^{n-1} \binom{n-1}{m} p^m (1-p)^{n-1-m} \end{aligned} \quad (4.53)$$

Thus, applying the formula for the binomial theorem in (4.37) for $n-1$ in place of n and $a = p$, $b = 1-p$, we obtain from the result in (4.53) that

$$E(Y) = np \cdot (p + 1 - p)^{n-1},$$

which yields (4.51).

4.1.5 The Poisson Distribution

We have seen that, if $Y \sim \text{Binomial}(n, p)$, then Y_n has a probability distribution given by

$$p_{Y_n}(k) = \begin{cases} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} & \text{for } k = 0, 1, 2, \dots, n, \\ 0 & \text{elsewhere,} \end{cases} \quad (4.54)$$

and the expected value of Y_n is given by

$$E(Y_n) = np, \quad \text{for all } n. \quad (4.55)$$

In this section we explore what happens to the distribution of Y_n as $n \rightarrow \infty$, while $E(Y_n)$ is kept at a constant value, λ . In other words, we would like explore the limit

$$\lim_{n \rightarrow \infty} p_{Y_n}(k) \quad \text{while} \quad np = \lambda, \quad \text{for all } n, \quad (4.56)$$

where λ is a constant.

Note that from

$$np = \lambda,$$

we get that

$$p = \frac{\lambda}{n}. \quad (4.57)$$

It follows from (4.57) that

$$p \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty,$$

so that the limiting setting described in (4.56) is relevant in modeling situations in which there are large number of independent trials with a very small probability of success. This is precisely the situation of a bacterial culture of size N in the order of millions, and mutation rates typically of the order of 10^{-8} . Thus, the limiting distribution in (4.56), if it exists, can be used too approximate the distribution of mutations in a large colony of bacteria when the mutation rate is very small.

Fix k in (4.54). Then, for $n > k$ we may write

$$p_{Y_n}(k) = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad (4.58)$$

where we have used (4.57) to replace p by $\frac{\lambda}{n}$. Next, rewrite (4.58) as

$$p_{Y_n}(k) = \frac{\lambda^k}{k!} \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n,$$

which can in turn be rewritten as

$$p_{Y_n}(k) = \frac{\lambda^k}{k!} \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k} \left(1 - \frac{\lambda}{n}\right)^n. \quad (4.59)$$

Now, since k and λ are fixed,

$$\lim_{n \rightarrow \infty} \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k} = 1. \quad (4.60)$$

Next, use the well-known limit

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x, \quad \text{for all } x \in \mathbb{R}, \quad (4.61)$$

to obtain that

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}. \quad (4.62)$$

Thus, using the limits in (4.62) and (4.60), we obtain from (4.59) that

$$\lim_{n \rightarrow \infty} p_{Y_n}(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \quad (4.63)$$

The limit expression in (4.63) shows that the sequence of random variables

$$Y_n \sim \text{Binomial}\left(n, \frac{\lambda}{n}\right), \quad \text{for } n = 1, 2, 3, \dots,$$

has a limiting distribution given by

$$p_Y(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{for } k = 0, 1, 2, 3, \dots \quad (4.64)$$

To see that the expression in (4.64) does indeed define a probability distribution observe that

$$\begin{aligned} \sum_{k=0}^{\infty} p_Y(k) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}. \end{aligned} \quad (4.65)$$

Thus, using the well known series expansion result

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x, \quad \text{for all } x \in \mathbb{R}, \quad (4.66)$$

we obtain from (4.65) that

$$\sum_{k=0}^{\infty} p_Y(k) = e^{-\lambda} e^{\lambda} = 1.$$

Note that the expressions in (4.61) and (4.66) are well known realizations of the exponential function $x \mapsto e^x$.

Definition 4.1.28 (Poisson Distribution). A discrete random variable, Y , which can take on the values $k = 0, 1, 2, \dots$, is said to have a Poisson distribution with parameter λ , if

$$p_Y(k) = \begin{cases} \frac{\lambda^k}{k!} e^{-\lambda} & \text{for } k = 0, 1, 2, \dots; \\ 0 & \text{elsewhere.} \end{cases} \quad (4.67)$$

We write

$$Y \sim \text{Poisson}(\lambda).$$

Example 4.1.29 (Expected Value of the Poisson Distribution). Let $Y \sim \text{Poisson}(\lambda)$. Compute $E(Y)$.

Solution: Since Y takes on a countable number of values, the expected value of Y is given by the series

$$E(Y) = \sum_{m=0}^{\infty} m p_Y(m), \quad (4.68)$$

where p_Y is given in (4.67). Thus, noting that the first term in the series in (4.68) is zero, we obtain from (4.68) and (4.67) that

$$\begin{aligned} E(Y) &= \sum_{m=1}^{\infty} m \cdot \frac{\lambda^m}{m!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{m=1}^{\infty} \frac{\lambda^m}{(m-1)!}. \end{aligned} \tag{4.69}$$

Next, make the change of variables $k = m - 1$ in (4.69) to obtain

$$\begin{aligned} E(Y) &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k+1}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k \cdot \lambda}{k!}, \end{aligned}$$

so that

$$E(Y) = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}. \tag{4.70}$$

Finally, use the series expansion for e^x in (4.66) to obtain from (4.70) that

$$E(Y) = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

□

Thus, we have shown that

$$Y \sim \text{Poisson}(\lambda) \Rightarrow E(Y) = \lambda; \tag{4.71}$$

in other words, the expected value of a Poisson random variable with parameter λ is λ . In Assignment #9 you'll be asked to show that the variance, $\text{Var}(Y)$ of $Y \sim \text{Poisson}(\lambda)$ is also λ ; where,

$$\text{Var}(Y) = E(Y^2) - [E(Y)]^2.$$

4.1.6 Estimating Mutation Rates in Bacterial Populations

In the early 1940s, Luria and Delbrück [LD43] devised the following procedure (known as the *fluctuation test*) to estimate the *mutation rate*, p , for certain bacteria:

Imagine that you start with a single normal bacterium (with no mutations) and allow it to grow to produce several bacteria. Place each of these bacteria in test-tubes each with a medium conducive to growth. Suppose the bacteria in the test-tubes are allowed to reproduce for n division cycles. After the n^{th} division cycle, the content of each test-tube is placed onto a agar plate

containing a virus population which is lethal to the bacteria which have not developed resistance. Those bacteria which have mutated into resistant strains will continue to replicate, while those that are sensitive to the virus will die. After certain time, the resistant bacteria will develop visible colonies on the plates. The number of these colonies will then correspond to the number of resistant cells in each test tube at the time they were exposed to the virus. This number corresponds to the number of bacteria in the colony that developed a mutation which led to resistance. We denote this number by Y_N , where N is the size of the colony after the n^{th} division cycle. Assuming that the bacteria may develop mutation to resistance after exposure to the virus, if N is very large, according to the result in Section 4.1.5, the distribution of Y_N can be approximated by a Poisson distribution with parameter $\lambda = pN$, where p is the mutation rate and N is the size of the colony. It then follows that the probability of no mutations occurring in one division cycle is

$$\Pr(Y_N = 0) \approx e^{-\lambda}, \quad (4.72)$$

according to (4.67). This probability can also be estimated experimentally as Luria and Delbrück showed in their 1943 paper. In one of the experiments described in that paper, out of 87 cultures of 2.4×10^8 bacteria, 29 showed not resistant bacteria (i.e., none of the bacteria in the culture mutated to resistance and therefore all perished after exposure to the virus). We therefore have that

$$\Pr(Y_N = 0) \approx \frac{29}{87}.$$

Comparing this to the expression in Equation (4.72), we obtain that

$$e^{-\lambda} \approx \frac{29}{87},$$

which can be solved for λ to obtain

$$\lambda \approx -\ln\left(\frac{29}{87}\right)$$

or

$$\lambda \approx 1.12.$$

The mutation rate, p , can then be estimated from $\lambda = pN$:

$$p = \frac{\lambda}{N} \approx \frac{1.12}{2.4 \times 10^8} \approx 4.7 \times 10^{-9}.$$

4.1.7 Another Brief Excursion into Probability

We have seen that, if A and B are independent events, then

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B); \quad (4.73)$$

in other words, the probability of the joint occurrence of two independent events is the product of their probabilities. In many situations, however, the occurrence of an event will have an effect on the probability of the occurrence of the other event. Here is a simple example that illustrates how this can happen.

Example 4.1.30. Put four marbles in a bag. Two of the marbles are red and the other two are blue. Pick two marbles at random and without replacement. Labeling the marbles R_1, R_2, B_1 and B_2 , for the two red marbles and the two blue marbles, respectively, we see that the sample space for the experiment (the set of all possible outcomes of the experiment) can be written as

$$\mathcal{C} = \{R_1R_2, R_1B_1, R_1B_2, R_2B_1, R_2B_2, B_1B_2\}.$$

The assumption of randomness in the picking of the two marbles implies that each of the outcomes in \mathcal{C} is equally likely; so that

$$\Pr(c) = \frac{1}{6}, \quad \text{for all } c \in \mathcal{C}. \quad (4.74)$$

Let A denote the event that at least one of the marbles is red and B denote the event that the two marbles have the same color; then,

$$A = \{R_1R_2, R_1B_1, R_1B_2, R_2B_1, R_2B_2\},$$

so that, in view of (4.74),

$$\Pr(A) = \frac{5}{6}. \quad (4.75)$$

Similarly, since

$$B = \{R_1R_2, B_1B_2\},$$

it follows from (4.74) that Then,

$$\Pr(B) = \frac{2}{6} = \frac{1}{3}. \quad (4.76)$$

On the other hand, since

$$A \cap B = \{R_1R_2\},$$

that is, the joint event $A \cap B$ is the event that both marbles are red,

$$\Pr(A \cap B) = \frac{1}{6}. \quad (4.77)$$

It follows from (4.75) and (4.76) that

$$\Pr(A) \cdot \Pr(B) = \frac{5}{6} \cdot \frac{1}{3} = \frac{5}{18}. \quad (4.78)$$

Comparing (4.77) and (4.78) we see that (4.73) does not hold true in this example. Thus, A and B are not stochastically independent.

Definition 4.1.31 (Conditional Probability). Given two events, A and B , with $\Pr(B) > 0$, we define the conditional probability of A given B , denoted $\Pr(A | B)$, by

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)}. \quad (4.79)$$

Similarly, if $\Pr(A) > 0$, the conditional probability of B given A , denoted $\Pr(B | A)$, by

$$\Pr(B | A) = \frac{\Pr(A \cap B)}{\Pr(A)}. \quad (4.80)$$

Remark 4.1.32. It follows from (4.80) that, if $\Pr(A) > 0$, then

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B | A); \quad (4.81)$$

thus, by introducing the concept of conditional probability, in some sense we recover (4.73), which, as seen in Example 4.1.30, does not hold in general. Observe that (4.81) holds true for all events A and B such that $\Pr(A) > 0$.

Proposition 4.1.33. Assume that $\Pr(A) > 0$ and $\Pr(B) > 0$. Then, events A and B are independent if and only if

$$\Pr(A | B) = \Pr(A), \quad \text{and} \quad \Pr(B | A) = \Pr(B). \quad (4.82)$$

Proof: Assume that A and B are independent. Then, (4.73) holds true. Since $\Pr(B) > 0$, it follows from (4.73) and (4.79) that

$$\Pr(A | B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A) \cdot \Pr(B)}{\Pr(B)} = \Pr(A).$$

Similarly, since $\Pr(A) > 0$, it follows from (4.73) and (4.80) that

$$\Pr(B | A) = \Pr(B).$$

Thus, (4.82) holds true if A and B are independent.

Conversely, suppose that (4.82) holds true. It then follows from (4.81) that

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B | A) = \Pr(A) \cdot \Pr(B),$$

which shows that A and B are independent. ■

Proposition 4.1.34 (Properties of Conditional Probabilities). Let $A, B, E_1, E_2, \dots, E_n$ denote subsets of a sample space \mathcal{C} on which a probability function, \Pr , is defined.

1. If $\Pr(B) > 0$, then $\Pr(A^c | B) = 1 - \Pr(A | B)$.
2. (*Law of Total Probability*). Suppose $E_1, E_2, E_3, \dots, E_n$ are mutually exclusive such that $\Pr(E_i) > 0$ for $i = 1, 2, 3, \dots, n$, and

$$\mathcal{C} = \bigcup_{k=1}^n E_k.$$

Then,

$$\Pr(B) = \sum_{k=1}^n \Pr(E_k) \cdot \Pr(B | E_k) \quad (4.83)$$

4.1.8 Continuous Random Variables

We begin with an example regarding modeling the time, T , that people spend at a check-out counter in a supermarket, for instance. In this case, T is a random variable that can take on a continuum of values; that is, any real value within an interval of real numbers; more specifically, in this case, T could take on any value in the interval $(0, \infty)$. Thus, T is an example of a continuous random variable.

For a continuous random variable, T , we are interested in its probability density function, f_T , or pdf. Once the pdf is known, we can compute the probability that T will take on a certain range of values by integration; for example,

$$\Pr(a < T \leq b) = \int_a^b f_T(t) dt. \quad (4.84)$$

In the next example, we model the service time T and show how to compute f_T by first computing the cumulative distribution function, F_T , defined by

$$F_T(t) = \Pr(T \leq t), \quad \text{for all } t \in \mathbb{R}. \quad (4.85)$$

It follows from (4.84) and (4.85) that

$$F_T(t) = \int_{-\infty}^t f_T(\tau) d\tau, \quad \text{for all } t \in \mathbb{R},$$

so that, by the Fundamental Theorem of Calculus,

$$f_T(t) = F_T'(t), \quad (4.86)$$

for values of t at which f_T is continuous.

Example 4.1.35 (Service time at a checkout counter). Suppose you sit by a checkout counter at a supermarket and measure the time, T , it takes for each customer to be served. This is a continuous random variable that takes on values in a time continuum. We would like to compute the cumulative distribution function given by (4.85), $F_T(t) = \Pr(T \leq t)$, for $t > 0$.

Let $N(t)$ denote the number of customers being served at a checkout counter (not in line) at time t , and assume that either $N(t) = 1$ or $N(t) = 0$; in other words, $N(t)$ is a Bernoulli random variable for each t . Here, we are also assuming that, once service is completed, no new customer will walk up to the checkout counter.

Set

$$p(t) = \Pr[N(t) = 1], \quad \text{for } t \geq 0, \quad (4.87)$$

and assume that $p(0) = 1$; that is, at the start of the observation, one person is being served.

We consider what happens to the probability $p(t)$ for t in a short time interval $[t, t + \Delta t]$. We would like to estimate $p(t + \Delta t)$, where Δt is very small; i.e., the probability that a person is being served at time $t + \Delta t$.

We assume that the the probability that service will be completed in the short time interval $[t, t + \Delta t]$ is proportional to Δt ; say $\mu\Delta t$, where $\mu > 0$ is a proportionality constant. Then, the probability that service will not be completed at $t + \Delta t$ is $1 - \mu\Delta t$. This situation is illustrated in the *state diagram* pictured in Figure 4.1.1: The circles in the state diagram represent the possible

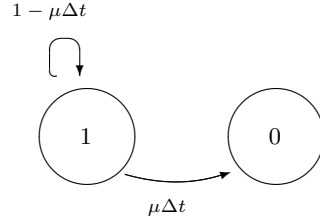


Figure 4.1.1: State diagram for $N(t)$

values of $N(t)$, or *states*. In this case, the states are 1 or 0, corresponding to one person being served and no person being served, respectively. The arrows represent *transition probabilities* from one state to another (or the same) in the interval from t to $t + \Delta t$. Thus the probability of going from state $N(t) = 1$ to state $N(t) = 0$ in that interval (that is, service is completed) is approximately $\mu\Delta t$, while the probability that the person will still be at the counter at the end of the interval is $1 - \mu\Delta t$.

By the law of total probability in (4.83)

$$\begin{aligned} \Pr(N(t + \Delta t) = 1) &= \Pr(N(t) = 1) \cdot \Pr(N(t + \Delta t) = 1 \mid N(t) = 1) \\ &\quad + \Pr(N(t) = 0) \cdot \Pr(N(t + \Delta t) = 1 \mid N(t) = 0), \end{aligned}$$

where $\Pr(N(t + \Delta t) = 1 \mid N(t) = 1)$ is the probability that service is not completed in the time interval $[t, t + \Delta t]$; so that

$$\Pr(N(t + \Delta t) = 1 \mid N(t) = 1) \approx 1 - \mu\Delta t,$$

by the previous consideration. We can therefore write

$$\begin{aligned} p(t + \Delta t) &\approx p(t)(1 - \mu\Delta t) \\ &\quad + (1 - p(t)) \cdot \Pr(N(t + \Delta t) = 1 \mid N(t) = 0), \end{aligned} \tag{4.88}$$

Since we are also assuming that

$$\Pr(N(t + \Delta t) = 1 \mid N(t) = 0) = 0,$$

for Δt small enough (see also the diagram in Figure 4.1.1), we therefore get from (4.88) that

$$p(t + \Delta t) \approx p(t)(1 - \mu\Delta t),$$

or

$$p(t + \Delta t) - p(t) \approx -\mu\Delta t. \quad (4.89)$$

Dividing both sides of (4.89) by $\Delta t \neq 0$ and taking the limit as $\Delta t \rightarrow 0$ we obtain that

$$\lim_{\Delta t \rightarrow 0} \frac{p(t + \Delta t) - p(t)}{\Delta t} = -\mu p(t) \quad (4.90)$$

It follows from (4.90) that p is differentiable and satisfies the differential equation.

$$\frac{dp}{dt} = -\mu p(t). \quad (4.91)$$

The first order differential equation in (4.91) can be solved subject to the initial condition $p(0) = 1$ to yield

$$p(t) = e^{-\mu t}, \quad \text{for } t \geq 0. \quad (4.92)$$

Recall that T denotes the time it takes for service to be completed, or the service time at the checkout counter. Thus, it is the case that

$$\Pr[T > t] = \Pr[N(t) = 1], \quad \text{for } t \geq 0; \quad (4.93)$$

that is, $T > t$ if and only if at time t the person is still at the checkout counter. It follows from (4.93) and (4.87) that

$$\Pr[T > t] = p(t), \quad \text{for } t \geq 0,$$

which can be written as

$$\Pr[T > t] = e^{-\mu t}, \quad \text{for } t \geq 0,$$

in view of (4.92). We then get that

$$\Pr[T \leq t] = 1 - \Pr[T > t] = 1 - e^{-\mu t}, \quad \text{for } t \geq 0,$$

so that the cumulative distribution function (cdf) of T is

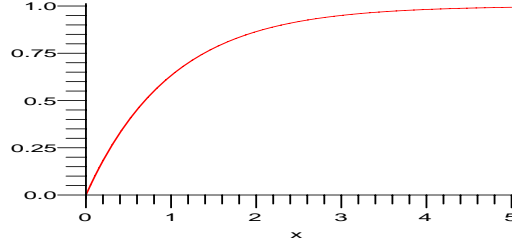
$$F_T(t) = \begin{cases} 1 - e^{-\mu t}, & \text{for } t \geq 0; \\ 0 & \text{for } t < 0. \end{cases} \quad (4.94)$$

A portion of the graph of this cumulative distribution function is shown in Figure 4.1.2. It follows from (4.86) and (4.94) that the probability density function for the service time, T , is given by

$$f_T(t) = \begin{cases} \mu e^{-\mu t}, & \text{for } t \geq 0; \\ 0 & \text{for } t < 0. \end{cases} \quad (4.95)$$

To see that the function in (4.95) indeed defines a probability density function, compute

$$\int_{-\infty}^{\infty} f_T(t) dt = \lim_{b \rightarrow \infty} \int_0^b \mu e^{-\mu t} dt, \quad (4.96)$$

Figure 4.1.2: Cumulative Distribution Function of T for $t \geq 0$

where

$$\int_0^b \mu e^{-\mu t} dt = [-e^{-\mu t}]_0^b = 1 - e^{-\mu b} \quad (4.97)$$

It follows from (4.97) and (4.96) that

$$\int_{-\infty}^{\infty} f_T(t) dt = \lim_{b \rightarrow \infty} [1 - e^{-\mu b}] = 1, \quad (4.98)$$

since $\mu > 0$.

Definition 4.1.36 (Probability Density Function). An integrable function, $f: \mathbb{R} \rightarrow \mathbb{R}$, is said to be a probability density function if

(i) $f(x) \geq 0$ for all $x \in \mathbb{R}$, and

(ii) $\int_{-\infty}^{\infty} f(x) dx = 1$.

Remark 4.1.37. It follows from (4.98) in Example 4.1.35 that the function f_T defined in (4.95) is a probability density function. We say that f_T is the probability density function of the random variable T .

Definition 4.1.38 (Cumulative Distribution Function). Let X be a random variable with probability density function f_X . Then, for any real numbers, a and b , with $a < b$;

$$\Pr[a < X \leq b] = \int_a^b f_X(x) dx.$$

The function, $F_X: \mathbb{R} \rightarrow \mathbb{R}$, defined by

$$F_X(x) = \Pr[X \leq x] = \int_{-\infty}^x f_X(t) dt, \quad \text{for all } x \in \mathbb{R},$$

is called the cumulative distribution function of X .

Remark 4.1.39. The function F_T defined in (4.94) in Example 4.1.35 is the cumulative distribution function of the service time T .

Definition 4.1.40 (Expected Value of a Continuous Random Variable). Let X be a continuous random variable with probability density function f_X . If

$$\int_{-\infty}^{\infty} |x|f_X(x) \, dx < \infty,$$

we define the **expected value of X** , denoted $E(X)$, by

$$E(X) = \int_{-\infty}^{\infty} xf_X(x) \, dx.$$

Example 4.1.41 (Average Service Time). In the service time example, Example 4.1.35, we showed that the time, T , that it takes for service to be completed at a checkout counter has an exponential distribution with probability density function given in (4.95),

$$f_T(t) = \begin{cases} \mu e^{-\mu t} & \text{for } t > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (4.99)$$

where μ is a positive parameter. Note that for the expression in (4.99) to make sense, the parameter μ has to have units of 1/time.

Observe that

$$\begin{aligned} \int_{-\infty}^{\infty} |t|f_T(t) \, dt &= \int_0^{\infty} t\mu e^{-\mu t} \, dt \\ &= \lim_{b \rightarrow \infty} \int_0^b t \mu e^{-\mu t} \, dt \\ &= \lim_{b \rightarrow \infty} \left[-te^{-\mu t} - \frac{1}{\mu} e^{-\mu t} \right]_0^b \\ &= \lim_{b \rightarrow \infty} \left[\frac{1}{\mu} - be^{-\mu b} - \frac{1}{\mu} e^{-\mu b} \right] \\ &= \frac{1}{\mu}, \end{aligned}$$

where we have used integration by parts and L'Hospital's rule. It then follows that

$$\int_{-\infty}^{\infty} |t|f_T(t) \, dt = \frac{1}{\mu} < \infty$$

and therefore the expected value of T exists and

$$E(T) = \int_{-\infty}^{\infty} tf_T(t) \, dt = \int_0^{\infty} t\mu e^{-\mu t} \, dt = \frac{1}{\mu}.$$

Thus, the parameter μ is the reciprocal of the expected service time, or **average service time**, at the checkout counter.

Example 4.1.42. Suppose the average service time, or **mean service time**, at a checkout counter is 5 minutes. Compute the probability that a given person will spend at least 6 minutes at the checkout counter.

Solution: By the result of Example 4.1.35, we assume that the service time, T , has a probability density function given in (4.95) with $\mu = 1/5$. We then have that

$$\Pr(T \geq 6) = \int_6^{\infty} f_T(t) dt = \int_6^{\infty} \frac{1}{5} e^{-t/5} dt = e^{-6/5} \approx 0.30.$$

Thus, there is a 30% chance that a person will spend 6 minutes or more at the checkout counter. \square

Definition 4.1.43 (Exponential Distribution). A continuous random variable, X , is said to be exponentially distributed with parameter $\beta > 0$, written

$$X \sim \text{Exponential}(\beta),$$

if it has a probability density function given by

$$f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & \text{for } x > 0, \\ 0 & \text{otherwise.} \end{cases}$$

The expected value of $X \sim \text{Exponential}(\beta)$, for $\beta > 0$, is $E(X) = \beta$.

4.2 Random Processes

In this section we come back to the problem of determining the distribution of the number of mutations of a certain type that occur in a bacterial colony. We analyzed this problem in Section 4.1 by considering the limit of Binomial(N, p) distributions when N is very large while $Np = \lambda$ is a constant. This approach led to a Poisson(λ) distribution. Here we take into account the fact that the bacterial population size is changing with time, t . Accordingly, we define $M(t)$ to be the number of mutations of a certain type that occur in the time interval $[0, t]$, for $t > 0$. Note that, for each t , $M(t)$ is a random variable that can take on any of the values

$$0, 1, 2, 3, \dots$$

We are interested in computing the probability that $M(t)$ attains each of those values for each time t . In symbols, we would like to compute

$$\Pr[M(t) = m] \quad \text{for } m = 0, 1, 2, \dots \quad \text{and } t > 0.$$

We shall denote $\Pr[M(t) = m]$ by $P_m(t)$.

We would like to compute $P_m(t)$, for each $m = 0, 1, 2, 3, \dots$ and $t > 0$, under the following assumptions:

- (i) $P_0(0) = \Pr[M(0) = 0] = 1$; that is, initially no bacterium has mutated into a strain of the particular type under study. It then follows that

$$P_m(0) = 0, \quad \text{for } m \geq 1 \quad (4.100)$$

- (ii) The probability that any bacterium develops a mutation in a short time interval $[t, t + \Delta t]$ depends only on Δt and not on the number of mutant bacteria at previous times.
- (iii) The probability of a new mutation in the short interval $[t, t + \Delta t]$ is proportional to Δt ; in symbols

$$\Pr(\text{new mutation in } [t, t + \Delta t]) \approx \lambda \Delta t,$$

where $\lambda > 0$ is a constant of proportionality.

- (iv) Δt is so small that the probability of two or more mutations occurring in the short time interval $[t, t + \Delta t]$ is zero.

In order to determine $P_m(t)$ for each $m = 1, 2, 3, \dots$ and $t > 0$, first we need to estimate $P_m(t + \Delta t)$ for Δt small enough. Thus, we need to model the process of going from time t to the time $t + \Delta t$. As in Example 4.1.35, examination of a state diagram for $M(t)$ can be used to aid us in this process. The state diagram is pictured in Figure 4.1.1. Each of the circles in the state diagram represents the number of mutations at any given stage.

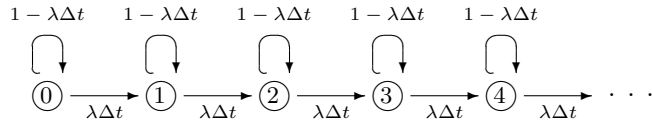


Figure 4.2.3: State diagram for $M(t)$

The arrows in Figure 4.2.3 indicate the transition probabilities of going from one state to the next, or those of remaining in the same state, in a short time interval $[t, t + \Delta t]$. For instance, if at time t there are no mutants in the colony (i.e., the system is in state 0 at that time), then at time $t + \Delta t$ there might a bacterium that has developed a mutation. The system would then go from state 0 to state 1 in the time interval $[t, t + \Delta t]$; the probability of this occurrence is approximately $\lambda \Delta t$ by assumption (iii); this is indicated by the arrow in the diagram that goes from state 0 to state 1. On the other hand, there might not be a new mutation in the time interval $[t, t + \Delta t]$; the probability of this occurring is approximately $1 - \lambda \Delta t$ (why?), and this is shown by the arrow that starts at state 0 and which winds back again to 0. Observe that assumption (iv) is implicit in the state diagram in Figure 4.2.3 since the states can only increase by 1 and not by 2 or more; thus, arrows from a given state either return to that state or go to the next one.

The state diagram in Figure 4.2.3 can be used to estimate $P_m(t + \Delta t)$, given that we know $P_m(t)$, for very small values of Δt . We start out with the case $m = 0$ as follows:

$$P_0(t + \Delta t) = P_0(t) \cdot \Pr(\text{no new mutations in } [t, t + \Delta t] \mid M(t) = 0).$$

Using assumption (ii) we have

$$P_0(t + \Delta t) = P_0(t) \cdot P(\text{no new mutations in } [t, t + \Delta t]), \quad (4.101)$$

since

$$\Pr(\text{no new mutations in } [t, t + \Delta t] \mid M(t) = 0) = \Pr(\text{no new mutations in } [t, t + \Delta t])$$

by stochastic independence. It then follows from (4.101) and assumption (iii) that

$$P_0(t + \Delta t) \approx P_0(t) \cdot (1 - \lambda\Delta t),$$

or

$$P_0(t + \Delta t) \approx P_0(t) - \lambda\Delta t P_0(t). \quad (4.102)$$

Rearranging equation (4.102) and dividing by Δt , we obtain

$$\frac{P_0(t + \Delta t) - P_0(t)}{\Delta t} \approx -\lambda P_0(t). \quad (4.103)$$

Next, let $\Delta t \rightarrow 0$ in (4.103) to conclude that $P_0(t)$ is differentiable and

$$\frac{dP_0}{dt} = -\lambda P_0; \quad (4.104)$$

that is, $P_0(t)$ satisfies a first order differential equation. The differential equation in (4.104) can be solved by separation of variables to yield

$$P_0(t) = C e^{-\lambda t}, \quad (4.105)$$

for some constant C . Since $P_0(0) = 1$ by assumption (i), it follows that $C = 1$ in (4.105), and so the probability of no mutations in the colony at time t is given by

$$P_0(t) = e^{-\lambda t}, \quad \text{for } t \geq 0. \quad (4.106)$$

We next proceed to compute $P_1(t)$. Using the state diagram in Figure 4.2.3 we obtain that

$$P_1(t + \Delta t) \approx P_0(t) \cdot \lambda\Delta t + P_1(t) \cdot (1 - \lambda\Delta t), \quad (4.107)$$

since, according to the state diagram in Figure 4.2.3, the system can get to state 1 at $t + \Delta t$ via two routes: (i) from state 0 through a new mutation which occurs with probability $\lambda\Delta t$, approximately, or (ii) from state 1 if no new mutation occurs in the time interval $[t, t + \Delta t]$, and the approximate probability

of this occurrence is $1 - \lambda\Delta t$. Here we have also used the law of total probability in (4.83) and the independence assumption (ii).

Rearranging equation (4.107) and dividing by $\Delta t \neq 0$, we obtain

$$\frac{P_1(t + \Delta t) - P_1(t)}{\Delta t} \approx -\lambda P_1(t) + \lambda P_0(t). \quad (4.108)$$

Next, let $\Delta t \rightarrow 0$ in (4.108) to conclude that P_1 is differentiable and satisfies the differential equation

$$\frac{dP_1}{dt} = -\lambda P_1 + \lambda P_0(t)$$

or, using (4.106),

$$\frac{dP_1}{dt} = -\lambda P_1 + \lambda e^{-\lambda t}. \quad (4.109)$$

The differential equation (4.109) can be solved as follows: Rewrite the equation as

$$\frac{dP_1}{dt} + \lambda P_1 = \lambda e^{-\lambda t} \quad (4.110)$$

and multiply both sides of (4.110) by $e^{\lambda t}$ to get

$$e^{\lambda t} \frac{dP_1}{dt} + \lambda e^{\lambda t} P_1 = \lambda \quad (4.111)$$

Observe that, by the Product Rule,

$$\frac{d}{dt}(e^{\lambda t} P_1) = e^{\lambda t} \frac{dP_1}{dt} + \lambda e^{\lambda t} P_1,$$

and so the differential equation in (4.111) reduces to

$$\frac{d}{dt}(e^{\lambda t} P_1) = \lambda. \quad (4.112)$$

The equation in (4.112) can be integrated to yield

$$e^{\lambda t} P_1 = \lambda t + C,$$

for some arbitrary constant C , and therefore

$$P_1(t) = \lambda t e^{-\lambda t} + C e^{-\lambda t} \quad (4.113)$$

for $t \geq 0$.

Next, use the initial condition $P_1(0) = 0$ in (4.100), which follows from assumption (i), to get that $C = 0$, and therefore

$$P_1(t) = \lambda t e^{-\lambda t}, \quad \text{for } t \geq 0 \quad (4.114)$$

In order to compute $P_2(t)$, we proceed in a way similar to that used to compute $P_1(t)$. From the state diagram in Figure (4.2.3) we get that

$$P_2(t + \Delta t) = P_1(t) \cdot \lambda \Delta t + P_2(t) \cdot (1 - \lambda \Delta t),$$

from which we are led to the differential equation

$$\frac{dP_2}{dt} = -\lambda P_2 + \lambda P_1(t)$$

or, using (4.114),

$$\frac{dP_2}{dt} = -\lambda P_2 + \lambda^2 t e^{-\lambda t}. \quad (4.115)$$

We can solve this differential equation as we solved (4.115), by first rearranging and multiplying by $e^{\lambda t}$ to get

$$e^{\lambda t} \frac{dP_2}{dt} + \lambda e^{\lambda t} P_2 = \lambda^2 t, \quad (4.116)$$

and then re-writing the left-hand side of (4.116), so that

$$\frac{d}{dt}(e^{\lambda t} P_2) = \lambda^2 t. \quad (4.117)$$

Next, integrate the equation in (4.117) and use the initial condition $P_2(0) = 0$ in (4.100) to get

$$P_2(t) = \frac{(\lambda t)^2}{2} e^{-\lambda t}, \quad \text{for } t \geq 0. \quad (4.118)$$

One can go through the same procedure leading to (4.118) to obtain the formula

$$P_3(t) = \frac{(\lambda t)^3}{3!} e^{-\lambda t}$$

for $P_3(t)$, and this suggests the general formula for $P_m(t)$, $m = 0, 1, 2, \dots$, to be

$$P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}, \quad \text{for } t \geq 0. \quad (4.119)$$

We will establish the formula in (4.119) by induction on m . Observe that we have already established the basic case $m = 0$ in (4.106). Next, for the inductive step, assume that the formula (4.119) holds for m , and we seek to show that it also holds for $m + 1$. Using the state diagram 4.2.3 we see that

$$P_{m+1}(t + \Delta t) \approx P_m(t) \cdot \lambda \Delta t + P_{m+1}(t) \cdot (1 - \lambda \Delta t),$$

from which we are led to the differential equation

$$\frac{d}{dt}(P_{m+1}) = -\lambda P_{m+1} + \lambda P_m(t)$$

or, using the inductive hypothesis (4.119),

$$\frac{d}{dt}(P_{m+1}) = -\lambda P_{m+1} + \frac{\lambda^{m+1} t^m}{m!} e^{-\lambda t}. \quad (4.120)$$

We can solve the differential equation in (4.120) as we solved (4.115); that is, first rearrange the equation and multiply by $e^{\lambda t}$ to get

$$e^{\lambda t} \frac{d}{dt}(P_{m+1}) + \lambda e^{\lambda t} P_{m+1} = \frac{\lambda^{m+1} t^m}{m!}; \quad (4.121)$$

then, re-write the left-hand side of the equation in (4.121) to get

$$\frac{d}{dt}(e^{\lambda t} P_{m+1}) = \frac{\lambda^{m+1} t^m}{m!}. \quad (4.122)$$

Integrating (4.122) and using the initial condition $P_{m+1}(0) = 0$ in (4.100), we obtain

$$P_{m+1}(t) = \frac{(\lambda t)^{m+1}}{(m+1)!} e^{-\lambda t}$$

for all $t \geq 0$, since $(m+1)! = (m+1)m!$. This establishes the formula (4.119) for the case $m+1$, and therefore formula (4.119) is now proved for all $m = 0, 1, 2, \dots$ by induction on m .

Note that the formula in (4.119),

$$P_m(t) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}, \quad \text{for } m = 0, 1, 2, 3, \dots \text{ and } t \geq 0, \quad (4.123)$$

is the probability distribution of a Poisson(λt) random variable. We have therefore demonstrated that assumptions (i)–(iv) imply that, for each $t > 0$, $M(t)$ is a Poisson random variable with parameter λt ,

$$M(t) \sim \text{Poisson}(\lambda t), \quad \text{for } t > 0. \quad (4.124)$$

We say that $M(t)$ is a *Poisson random process*. This particular random process is characterized by the assumptions (i)–(iv) that we made on $M(t)$.

In general, a random process is a collection, $\{X(t) \mid t \in \mathcal{I}\}$, of random variables, $X(t)$, for t in some indexing set \mathcal{I} . If \mathcal{I} is countable, for instance, $\mathcal{I} = \mathbf{N}$ or $\mathcal{I} = \mathbf{Z}$, the random process $\{X(t) \mid t \in \mathcal{I}\}$ is called a discrete-time random process. If \mathcal{I} is some interval of real numbers, then $\{X(t) \mid t \in \mathcal{I}\}$ is called a continuous-time random process. For the case of the Poisson random process defined in (4.123), $\mathcal{I} = [0, \infty)$. Therefore, the Poisson random process is a continuous-time random process.

Given a random process, $\{X(t) \mid t \in \mathcal{I}\}$, the mean of the random process is given by

$$E[X(t)], \quad \text{for } t \in \mathcal{I}.$$

It follows from (4.124), or (4.123), and the calculations in Section 4.1.5 that

$$E[M(t)] = \lambda t, \quad \text{for } t > 0. \quad (4.125)$$

Thus, assumptions (i)–(iv) imply that the expected number of mutations in the interval $[0, t]$ is proportional to the length of the time interval. The constant of proportionality, λ , represents the average number of mutations per unit time.

Bibliography

- [EK88] L. Edelstein-Keshet. *Mathematical Models in Biology*. Random House/Birkhäuser, 1988.
- [LD43] S. E. Luria and M. Delbrück. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28:491–511, 1943.
- [LW55] M. J. Lighthill and G. B. Whitham. On kinematic waves II: A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond.*, 229:317–345, 1955.
- [Ric56] P. Richards. Shock waves on the highways. *Operations Research*, 4(1):42–51, 1956.