Chapter 9
Epidemic Models

9.1 Introduction to Epidemic Models

Communicable diseases such as measles, influenza, and tuberculosis are a fact of life. We will be concerned with both epidemics, which are sudden outbreaks of a disease, and endemic situations, in which a disease is always present. The AIDS epidemic, the recent SARS epidemic, recurring influenza pandemics, and outbursts of diseases such as the Ebola virus are events of concern and interest to many people. The prevalence and effects of many diseases in less-developed countries are probably not as well known but may be of even more importance. Every year millions, of people die of measles, respiratory infections, diarrhea, and other diseases that are easily treated and not considered dangerous in the Western world. Diseases such as malaria, typhus, cholera, schistosomiasis, and sleeping sickness are endemic in many parts of the world. The effects of high disease mortality on mean life span and of disease debilitation and mortality on the economy in afflicted countries are considerable.

We give a brief introduction to the modeling of epidemics; more thorough descriptions may be found in such references as [Anderson & May (1991), Diekmann & Heesterbeek (2000)]. This chapter will describe models for epidemics, and the next chapter will deal with models for endemic situations, but we begin with some general ideas about disease transmission.

The idea of invisible living creatures as agents of disease goes back at least to the writings of Aristotle (384 BC–322 BC). It developed as a theory in the sixteenth century. The existence of microorganisms was demonstrated by van Leeuwenhoek (1632–1723) with the aid of the first microscopes. The first expression of the germ theory of disease by Jacob Henle (1809–1885) came in 1840 and was developed by Robert Koch (1843–1910), Joseph Lister (1827–1912), and Louis Pasteur (1822–1875) in the latter part of the nineteenth century and the early part of the twentie/th century.

The mechanism of transmission of infections is now known for most diseases. Generally, diseases transmitted by viral agents, such as influenza, measles, rubella

F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, Texts in Applied Mathematics 40, DOI 10.1007/978-1-4614-1686-9_9,

(German measles), and chicken pox, confer immunity against reinfection, while diseases transmitted by bacteria , such as tuberculosis, meningitis, and gonorrhea, confer no immunity against reinfection. Other human diseases, such as malaria, are transmitted not directly from human to human but by vectors, agents (usually insects) that are infected by humans and who then transmit the disease to humans. There are also diseases such as West Nile virus, that are transmitted back and forth between animals and vectors. Heterosexual transmission of HIV/AIDS is also a vector process in which transmission goes back and forth between males and females.

We will focus on the transmission dynamics of an infection from individual to individual in a population, but many of the same ideas arise in transmission of a genetic characteristic, such as gender, race, genetic diseases, a cultural "characteristic," such as language or religion, an addictive activity, such as drug use, and the gain or loss of information communicated through gossip, rumors, and so on.

Similarly, many of the ideas arise also with different characterizations of what is meant by an individual, including the types of cells in the study of disease dynamics of the immune system. In the study of *Chagas* disease, a "house" (infested houses may correspond to "infected" individuals) may be chosen as an epidemiological unit; in tuberculosis, a household or community or a group of strongly linked individuals ("cluster") may be the chosen unit.

An epidemic, which acts on a short temporal scale, may be described as a sudden outbreak of a disease that infects a substantial portion of the population in a region before it disappears. Epidemics usually leave many members untouched. Often these attacks recur with intervals of several years between outbreaks, possibly diminishing in severity as populations develop some immunity. This is an important aspect of the connection between epidemics and disease evolution.

The historian W.H. McNeill argues, especially in his book *Plagues and Peoples* (1976), that the spread of communicable diseases frequently has been an important influence in history. For example, there was a sharp population increase throughout the world in the eighteenth century; the population of China increased from 150 million in 1760 to 313 million in 1794, and the population of Europe increased from 118 million in 1700 to 187 million in 1800. There were many factors involved in this increase, including changes in marriage age and technological improvements leading to increased food supplies, but these factors are not sufficient to explain the increase. Demographic studies indicate that a satisfactory explanation requires recognition of a decrease in the mortality caused by periodic epidemic infections. This decrease came about partly through improvements in medicine, but a more important influence was probably the fact that more people developed immunities against infection as increased travel intensified the circulation and cocirculation of diseases.

There are many biblical references to diseases as historical influences. The Book of Exodus describes the plagues that were brought down upon Egypt in the time of Moses. Another example is the decision of Sennacherib, the king of Assyria, to abandon his attempt to capture Jerusalem about 700 BC because of the illness of his soldiers (Isaiah 37, 36-38), and there are several other biblical descriptions of epidemic outbreaks.

The fall of empires has been attributed directly or indirectly to epidemic diseases. In the second century AD, the so-called Antonine plagues (possibly measles and) invaded the Roman Empire, causing drastic population reductions and economic hardships leading to disintegration of the empire because of disorganization, which facilitated invasions of barbarians. The Han Empire in China collapsed in the third century AD after a very similar sequence of events. The defeat of a population of millions of Aztecs by Cortez and his 600 followers can be explained, in part, by a smallpox epidemic that devastated the Aztecs but had almost no effect on the invading Spaniards, thanks to their built-in immunities. The Aztecs were not only weakened by disease but also confounded by what they interpreted as a divine force favoring the invaders. Smallpox then spread southward to the Incas in Peru and was an important factor in the success of Pizarro's invasion a few years later. smallpox was followed by other diseases such as measles and diphtheria imported from Europe to North America. In some regions, the indigenous populations were reduced to one-tenth of their previous levels by these diseases: Between 1519 and 1530 the Indian population of Mexico was reduced from 30 million to 3 million.

The Black Death (probably bubonic plague) spread from Asia throughout Europe in several waves during the fourteenth century, beginning in 1346, and is estimated to have caused the death of as much as one-third of the population of Europe between 1346 and 1350. The disease recurred regularly in various parts of Europe for more than 300 years, notably as the Great Plague of London of 1665–1666. It then gradually withdrew from Europe. Since the plague struck some regions harshly while avoiding others, it had a profound effect on political and economic developments in medieval times. In the last bubonic plague epidemic in France (1720–1722), half the population of Marseilles, 60 percent of the population in nearby Toulon, 44 per cent of the population of Arles, and 30 percent of the population of Aix and Avignon died, but the epidemic did not spread beyond Provence. Expansions and interpretations of these historical remarks may be found in McNeill (1976), which was our primary source on the history of the spread and effects of diseases.

The above examples depict the *sudden* dramatic impact that diseases have had on the demography of human populations via disease-induced mortality. In considering the combined role of diseases, war, and natural disasters on mortality rates, one may conclude that historically humans who are more likely to survive and reproduce have either a good immune system, a propensity to avoid war and disasters, or, nowadays, excellent medical care and/or health insurance.

Descriptions of epidemics in ancient and medieval times frequently used the term "plague" because of a general belief that epidemics represented divine retribution for sinful living. More recently, some have described AIDS as punishment for sinful activities. Such views have often hampered or delayed attempts to control this modern epidemic.

There are many questions of interest to public health physicians confronted with a possible epidemic. For example, how severe will an epidemic be? This question may be interpreted in a variety of ways. For example, how many individuals will be affected and require treatment? What is the maximum number of people needing

care at any particular time? How long will the epidemic last? How much good would quarantine of victims do in reducing the severity of the epidemic?

Scientific experiments usually are designed to obtain information and to test hypotheses. Experiments in epidemiology with controls are often difficult or impossible to design, and even if it is possible to arrange an experiment, there are serious ethical questions involved in withholding treatment from a control group. Sometimes data may be obtained after the fact from reports of epidemics or of endemic disease levels, but the data may be incomplete or inaccurate. In addition, data may contain enough irregularities to raise serious questions of interpretation, such as whether there is evidence of chaotic behavior [Ellner, Gallant, and Theiler (1995)]. Hence, parameter estimation and model fitting are very difficult. These issues raise the question whether mathematical modeling in epidemiology is of value.

Mathematical modeling in epidemiology provides understanding of the underlying mechanisms that influence the spread of disease, and in the process, it suggests control strategies. In fact, models often identify behaviors that are unclear in experimental data–often because data are nonreproducible and the number of data points is limited and subject to errors in measurement. For example, one of the fundamental results in mathematical epidemiology is that *most* mathematical epidemic models, including those that include a high degree of heterogeneity, usually exhibit "threshold" behavior, which in epidemiological terms can be stated as follows: *If the average number of secondary infections caused by an average infective is less than one, a disease will die out, while if it exceeds one there will be an epidemic*. This broad principle, consistent with observations and quantified via epidemiological models, has been used routinely to estimate the effectiveness of vaccination policies and the likelihood that a disease may be eliminated or eradicated. Hence, even if it is not possible to verify hypotheses accurately, agreement with hypotheses of a qualitative nature is often valuable. Expressions for the basic reproductive number for HIV in various populations is being used to test the possible effectiveness of vaccines that may provide temporary protection by reducing either HIV-infectiousness or susceptibility to HIV. Models are used to estimate how widespread a vaccination plan must be to prevent or reduce the spread of HIV.

In the mathematical modeling of disease transmission, as in most other areas of mathematical modeling, there is always a trade-off between simple models, which omit most details and are designed only to highlight general qualitative behavior, and detailed models, usually designed for specific situations including short-term quantitative predictions. Detailed models are generally difficult or impossible to solve analytically and hence their usefulness for theoretical purposes is limited, although their strategic value may be high. For public health professionals, who are faced with the need to make recommendations for strategies to deal with a specific situation, simple models are inadequate and numerical simulation of detailed models is necessary. In this chapter, we concentrate on simple models in order to establish broad principles. Furthermore, these simple models have additional value since they are the building blocks of models that include detailed structure. A specific goal is to compare the dynamics of simple and slightly more detailed models primarily

to see whether slightly different assumptions can lead to significant differences in qualitative behavior.

Many of the early developments in the mathematical modeling of communicable diseases are due to public health physicians. The first known result in mathematical epidemiology is a defense of the practice of inoculation against smallpox in 1760 by Daniel Bernouilli, a member of a famous family of mathematicians (eight spread over three generations) who had been trained as a physician. The first contributions to modern mathematical epidemiology are due to P.D. En'ko between 1873 and 1894 [Dietz (1988)], and the foundations of the entire approach to epidemiology based on compartmental models were laid by public health physicians such as Sir R.A. Ross, W.H. Hamer, A.G. McKendrick, and W.O. Kermack between 1900 and 1935, along with important contributions from a statistical perspective by J. Brownlee. A particularly instructive example is the work of Ross on malaria. Dr. Ross was awarded the second Nobel Prize in Medicine for his demonstration of the dynamics of the transmission of malaria between mosquitoes and humans. Although his work received immediate acceptance in the medical community, his conclusion that malaria could be controlled by controlling mosquitoes was dismissed on the grounds that it would be impossible to rid a region of mosquitoes completely and that in any case, mosquitoes would soon reinvade the region. After Ross formulated a mathematical model that predicted that malaria outbreaks could be avoided if the mosquito population could be reduced below a critical threshold level, field trials supported his conclusions and led to sometimes brilliant successes in malaria control. Unfortunately, the Garki project provides a dramatic counterexample. This project worked to eradicate malaria from a region temporarily. However, people who have recovered from an attack of malaria have a temporary immunity against reinfection. Thus elimination of malaria from a region leaves the inhabitants of this region without immunity when the campaign ends, and the result can be a serious outbreak of malaria.

We formulate our descriptions as *compartmental models*, with the population under study being divided into compartments and with assumptions about the nature and time rate of transfer from one compartment to another. Diseases that confer immunity have a different compartmental structure from diseases without immunity and from diseases transmitted by vectors. The rates of transfer between compartments are expressed mathematically as derivatives with respect to time of the sizes of the compartments, and as a result our models are formulated initially as *differential equations*. Models in which the rates of transfer depend on the sizes of compartments over the past as well as at the instant of transfer lead to more general types of functional equations, such as differential–difference equations and integral equations.

In this chapter we describe models for epidemics, acting on a sufficiently rapid time scale that demographic effects, such as births, natural deaths, immigration into and emigration out of a population may be ignored. In the next chapter we will describe models in which demographic effects are included.

9.2 The Simple Kermack–McKendrick Epidemic Model

Throughout history, epidemics have had major effects on the course of events. For example, the Black Death, now identified as probably having been the bubonic plague which had actually invaded Europe as early as the sixth century, spread from Asia throughout Europe in several waves during the fourteenth century, beginning in 1346, and is estimated to have caused the death of as much as one third of the population of Europe between 1346 and 1350. The disease recurred regularly in various parts of Europe for more than 300 years, notably as the Great Plague of London of 1665–1666. It then gradually withdrew from Europe.

More than 15% of the population of London died in the Great Plague (1665–1666). It appeared quite suddenly, grew in intensity, and then disappeared, leaving part of the population untouched. One of the early triumphs of mathematical epidemiology was the formulation of a simple model by Kermack and McKendrick (1927) whose predictions are very similar to this behavior, observed in countless epidemics. The Kermack–McKendrick model is a compartmental model based on relatively simple assumptions on the rates of flow between different classes of members of the population.

In order to model such an epidemic we divide the population being studied into three classes labeled S, I, and R. We let $S(t)$ denote the number of individuals who are susceptible to the disease, that is, who are not (yet) infected at time t. $I(t)$ denotes the number of infected individuals, assumed infectious and able to spread the disease by contact with susceptibles. $R(t)$ denotes the number of individuals who have been infected and then removed from the possibility of being infected again or of spreading infection. Removal is carried out through isolation from the rest of the population, through immunization against infection, through recovery from the disease with full immunity against reinfection, or through death caused by the disease. These characterizations of removed members are different from an epidemiological perspective but are often equivalent from a modeling point of view that takes into account only the state of an individual with respect to the disease.

We will use the terminology SIR to describe a disease that confers immunity against reinfection, to indicate that the passage of individuals is from the susceptible class S to the infective class I to the removed class R. Epidemics are usually diseases of this type. We would use the terminology SIS to describe a disease with no immunity against re-infection, to indicate that the passage of individuals is from the susceptible class to the infective class and then back to the susceptible class. Usually, diseases caused by a virus are of SIR type, while diseases caused by bacteria are of SIS type.

In addition to the basic distinction between diseases for which recovery confers immunity against reinfection and diseases for which recovered members are susceptible to reinfection, and the intermediate possibility of temporary immunity signified by a model of SIRS type, more complicated compartmental structure is possible. For example, there are SEIR and SEIS models, with an exposed period between being infected and becoming infective.

When there are only a few infected members, the start of a disease outbreak depends on random contacts between small numbers of individuals. In the next section we will use this to describe an approach to the study of the beginning of a disease outbreak by means of branching processes, but we begin with a description of deterministic compartmental models.

The independent variable in our compartmental models is the time t, and the rates of transfer between compartments are expressed mathematically as derivatives with respect to time of the sizes of the compartments, and as a result our models are formulated initially as *differential equations*.

We are assuming that the epidemic process is *deterministic*, that is, that the behavior of a population is determined completely by its history and by the rules that describe the model. In formulating models in terms of the derivatives of the sizes of each compartment we are also assuming that the number of members in a compartment is a differentiable function of time. This assumption is plausible once a disease outbreak has become established but is not valid at the beginning of a disease outbreak when there are only a few infectives. In the next section we will describe a different approach for the initial stage of a disease outbreak.

The basic compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by W.O. Kermack and A.G. McKendrick in 1927, 1932, and 1933. The first of these papers described epidemic models. What is often called the Kermack–McKendrick epidemic model is actually a special case of the general model introduced in this paper. The general model included dependence on age of infection, that is, the time since becoming infected, and can be used to provide a unified approach to compartmental epidemic models.

The special case of the model proposed by Kermack and McKendrick in 1927, which is the starting point for our study of epidemic models, is

$$\begin{aligned} S' &= -\beta SI, \\ I' &= \beta SI - \alpha I, \\ R' &= \alpha I\,. \end{aligned} \tag{9.1}$$

A flow chart is shown in Figure 9.1. It is based on the following assumptions:

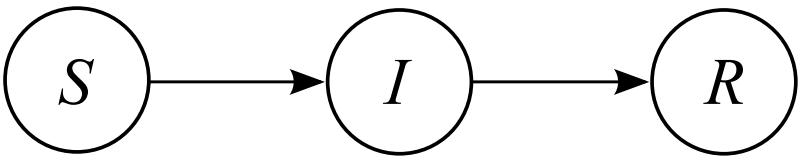

Fig. 9.1 Flow chart for the SIR model.

(i) An average member of the population makes contact sufficient to transmit infection with βN others per unit time, where N represents total population size (mass action incidence).

(ii) Infectives leave the infective class at rate αI per unit time.
(iii) There is no entry into or departure from the population, except possibly through death from the disease.
(iv) There are no disease deaths, and the total population size is a constant N.

According to (i), since the probability that a random contact by an infective is with a susceptible, who can then transmit infection, is S/N, the number of new infections in unit time per infective is $(\beta N)(S/N)$, giving a rate of new infections $(\beta N)(S/N)I = \beta SI$. Alternatively, we may argue that for a contact by a susceptible the probability that this contact is with an infective is I/N and thus the rate of new infections per susceptible is $(\beta N)(I/N)$, giving a rate of new infections $(\beta N)(I/N)S = \beta SI$. Note that both approaches give the same rate of new infections; in models with more complicated compartmental structure one may be more appropriate than the other.

We need not give an algebraic expression for N, since it cancels out of the final model, but we should note that for an *SIR* disease model, $N = S + I + R$. Later, we will allow the possibility that some infectives recover while others die of the disease. The hypothesis (iii) really says that the time scale of the disease is much faster than the time scale of births and deaths, so that demographic effects on the population may be ignored. An alternative view is that we are interested only in studying the dynamics of a single epidemic outbreak.

The assumption (ii) requires a fuller mathematical explanation, since the assumption of a recovery rate proportional to the number of infectives has no clear epidemiological meaning. We consider the "cohort" of members who were all infected at one time and let $u(s)$ denote the number of these who are still infective s time units after having been infected. If a fraction α of these leave the infective class in unit time, then

$$u' = -\alpha u\,,$$

and the solution of this elementary differential equation is

$$u(s) = u(0)\,e^{-\alpha s}\,.$$

Thus, the fraction of infectives remaining infective s time units after having become infective is $e^{-\alpha s}$, so that the length of the infective period is distributed exponentially with mean $\int_0^\infty e^{-\alpha s} ds = 1/\alpha$, and this is what (ii) really assumes. If we assume, instead of (ii), that the fraction of infectives remaining infective a time τ after having become infective is $P(\tau)$, the second equation of (9.1) would be replaced by the integral equation

$$I(t) = I_0(t) + \int_0^\infty \beta S(t-\tau) I(t-\tau) P(\tau) d\tau,$$

where $I_0(t)$ represents the members of the population who were infective at time $t = 0$ and are still infective at time t.

The assumptions of a rate of contacts proportional to population size N with constant of proportionality β and of an exponentially distributed recovery rate are

unrealistically simple. More general models can be constructed and analyzed, but our goal here is to show what may be deduced from extremely simple models. It will turn out that that many more realistic models exhibit very similar qualitative behaviors.

In our model R is determined once S and I are known, and we can drop the R equation from our model, leaving the system of two equations

$$\begin{aligned} S' &= -\beta SI, \\ I' &= (\beta S - \alpha)I\,, \end{aligned} \tag{9.2}$$

together with initial conditions

$$S(0) = S_0, \quad I(0) = I_0, \quad S_0 + I_0 = N.$$

We think of introducing a small number of infectives into a population of susceptibles and ask whether there will be an epidemic. We remark that the model makes sense only so long as $S(t)$ and $I(t)$ remain nonnegative. Thus if either $S(t)$ or $I(t)$ reaches zero, we consider the system to have terminated. We observe that $S' < 0$ for all t and $I' > 0$ if and only if $S > \alpha/\beta$. Thus I increases so long as $S > \alpha/\beta$, but since S decreases for all t, I ultimately decreases and approaches zero. If $S_0 < \alpha/\beta, I$ decreases to zero (no epidemic), while if $S_0 > \alpha/\beta$, I first increases to a maximum attained when $S = \alpha/\beta$ and then decreases to zero (epidemic).

The quantity $\beta S_0/\alpha$ is a threshold quantity, called the *basic reproduction number* [Heesterbeek (1996)] and denoted by $\mathscr{R}_0$, which determines whether there is an epidemict. If $\mathscr{R}_0 < 1$, the infection dies out, while if $\mathscr{R}_0 > 1$, there is an epidemic. The definition of the basic reproduction number $\mathscr{R}_0$ is that it is the number of secondary infections caused by a single infective introduced into a wholly susceptible population of size $N \approx S_0$ over the course of the infection of this single infective. In this situation, an infective makes βN contacts in unit time, all of which are with susceptibles and thus produce new infections, and the mean infective period is $1/\alpha$; thus the basic reproduction number is actually $\beta N/\alpha$ rather than $\beta S_0/\alpha$. Another way to view this apparent discrepancy is to consider two ways in which an epidemic may begin. One way is an epidemic started by a member of the population being studied, for example by returning from travel with an infection acquired away from home. In this case we would have $I_0 > 0, S_0 + I_0 = N$. A second way is for an epidemic to be started by a visitor from outside the population. In this case, we would have $S_0 = N$.

Since (9.2) is a two-dimensional autonomous system of differential equations, the natural approach would be to find equilibria and linearize about each equilibrium to determine its stability. However, since every point with $I = 0$ is an equilibrium, the system (9.2) has a line of equilibria, and this approach is not applicable (the linearization matrix at each equilibrium has a zero eigenvalue).

Fortunately, there is an alternative approach that enables us to analyze the system (9.2). The sum of the two equations of (9.2) is

$$(S+I)' = -\alpha I.$$

Thus $S+I$ is a nonnegative smooth decreasing function and therefore tends to a limit as $t \to \infty$. Also, it is not difficult to prove that the derivative of a nonnegative smooth decreasing function must tend to zero, and this shows that

$$I_\infty = \lim_{t\to\infty} I(t) = 0.$$

Thus $S+I$ has limit S_∞.

Integration of the sum of the two equations of (9.2) from 0 to ∞ gives

$$\alpha \int_0^\infty (S(t)+I(t))dt = S_0 + I_0 - S_\infty = N - S_\infty.$$

Division of the first equation of (9.2) by S and integration from 0 to ∞ gives

$$\log \frac{S_0}{S_\infty} = \beta \int_0^\infty I(t)dt = \frac{\beta}{\alpha}[N - S_\infty] = \mathscr{R}_0 \left[1 - \frac{S_\infty}{N}\right]. \tag{9.3}$$

Equation (9.3) is called the *final size relation*. It gives a relationship between the basic reproduction number and the size of the epidemic. Note that the final size of the epidemic, the number of members of the population who are infected over the course of the epidemic, is $N - S_\infty$. This is often described in terms of the *attack rate* $(1 - S_\infty/N)$. [Technically, the attack rate should be called an attack ratio, since it is dimensionless and is not a rate.]

The final size relation (9.3) can be generalized to epidemic models with more complicated compartmental structure than the simple SIR model (9.2), including models with exposed periods, treatment models, and models including quarantine of suspected individuals and isolation of diagnosed infectives. The original Kermack–McKendrick model (1927) included dependence on the time since becoming infected (age of infection), and this includes such models.

Integration of the first equation of (9.2) from 0 to t gives

$$\log \frac{S_0}{S(t)} = \beta \int_0^t I(t)dt = \frac{\beta}{\alpha}[N - S(t) - I(t)],$$

and this leads to the form

$$I(t) + S(t) - \frac{\alpha}{\beta}\log S(t) = N - \frac{\alpha}{\beta}\log S_0. \tag{9.4}$$

This implicit relation between S and I describes the orbits of solutions of (9.2) in the (S,I) plane.

In addition, since the right side of (9.3) is finite, the left side is also finite, and this shows that $S_\infty > 0$. The final size relation (9.3) is valid for a large variety of epidemic models, as we shall see in later sections.

It is not difficult to prove that there is a unique solution of the final size relation (9.3). To see this, we define the function

$$g(x) = \log \frac{S_0}{x} - \mathscr{R}_0 \left[1 - \frac{x}{N}\right].$$

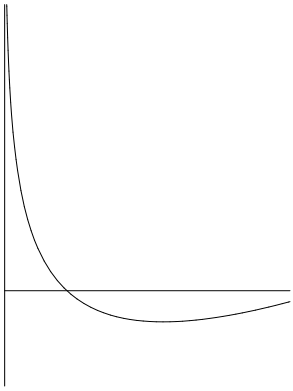

Fig. 9.2 The function g(x).

Then, as shown in Figure 9.2,

$$g(0+) > 0, \quad g(N) < 0,$$

and $g'(x) < 0$ if and only if

$$0 < x < \frac{N}{\mathscr{R}_0}.$$

If $\mathscr{R}_0 \leq 1, g(x)$ is monotone decreasing from a positive value at $x = 0+$ to a negative value at $x = N$. Thus there is a unique zero S_∞ of $g(x)$ with $S_\infty < N$.

If $\mathscr{R}_0 > 1, g(x)$ is monotone decreasing from a positive value at $x = 0+$ to a minimum at $x = N/\mathscr{R}_0$ and then increases to a negative value at $x = N_0$. Thus there is a unique zero S_∞ of $g(x)$ with

$$S_\infty < \frac{N}{\mathscr{R}_0}.$$

In fact,

$$g\left(\frac{S_0}{\mathscr{R}_0}\right) = \log \mathscr{R}_0 - \mathscr{R}_0 + \frac{S_0}{N} \leq \log \mathscr{R}_0 - \mathscr{R}_0 + 1.$$

Since $\log \mathscr{R}_0 < \mathscr{R}_0 - 1$ for $\mathscr{R}_0 > 0$, we actually have

$$g\left(\frac{S_0}{\mathscr{R}_0}\right) < 0,$$

and

$$S_\infty < \frac{S_0}{\mathscr{R}_0}.$$

It is generally difficult to estimate the contact rate β, which depends on the particular disease being studied but may also depend on social and behavioral factors. The quantities S_0 and S_∞ may be estimated by serological studies (measurements of immune responses in blood samples) before and after an epidemic, and from these data the basic reproduction number $\mathscr{R}_0$ may be estimated using (9.3). This estimate, however, is a retrospective one, which can be derived only after the epidemic has run its course.

The maximum number of infectives at any time is the number of infectives when the derivative of I is zero, that is, when $S = \alpha/\beta$. This maximum is given by

$$I_{max} = S_0 + I_0 - \frac{\alpha}{\beta}\log S_0 - \frac{\alpha}{\beta} + \frac{\alpha}{\beta}\log\frac{\alpha}{\beta}, \tag{9.5}$$

obtained by substituting $S = \alpha/\beta$, $I = I_{max}$ into (9.4).

Example 1. A study of Yale University freshmen [Evans (1982), reported by Hethcote(1989)] described an influenza epidemic with $S_0 = 0.911$, $S_\infty = 0.513$. Here we are measuring the number of susceptibles as a fraction of the total population size, or using the population size K as the unit of size. Substitution into the final size relation gives the estimate $\beta/\alpha = 1.18$ and $R_0 = 1.18$. Since we know that τ is approximately 3 days for influenza, we see that β is approximately 0.48 contacts per day per member of the population.

Example 2. (The Great Plague in Eyam) The village of Eyam near Sheffield, England, suffered an outbreak of bubonic plague in 1665–1666 the source of which is generally believed to be the Great Plague of London. The Eyam plague was survived by only 83 of an initial population of 350 persons. Since detailed records were preserved and the community was persuaded to quarantine itself to try to prevent the spread of disease to other communities, the disease in Eyam has been used as a case study for modeling [Raggett (1982)]. Detailed examination of the data indicates that there were actually two outbreaks, of which the first was relatively mild. Thus we shall try to fit the model (9.2) over the period from mid-May to mid-October 1666, measuring time in months with an initial population of 7 infectives and 254 susceptibles, and a final population of 83. Raggett (1982) gives values of susceptibles and infectives in Eyam on various dates, beginning with $S(0) = 254$, $I(0) = 7$, shown in Table 9.1.

The final size relation with $S_0 = 254$, $I_0 = 7$, $S_\infty = 83$ gives $\beta/\alpha = 6.54 \times 10^{-3}$, $\alpha/\beta = 153$. The infective period was 11 days, or 0.3667 month, so that $\alpha = 2.73$. Then $\beta = 0.0178$. The relation (9.5) gives an estimate of 30.4 for the maximum number of infectives. We use the values obtained here for the parameters β and τ in the model (9.2) for simulations of both the phase plane, here the (S,I)-plane, and for graphs of S and I as functions of t (Figures 9.3, 9.4, 9.5). Figure 9.6 plots these data points together with the phase portrait given in Figure 9.3 for the model (9.2).

The actual data for the Eyam epidemic are remarkably close to the predictions of this very simple model. However, the model is really too good to be true. Our model assumes that infection is transmitted directly between people. While this is

Date (1666)	Susceptibles	Infectives
July 3/4	235	14.5
July 19	201	22
August 3/4	153.5	29
August 19	121	21
September 3/4	108	8
September 19	97	8
October 4/5	Unknown	Unknown
October 20	83	0

Table 9.1 Eyam Plague data.

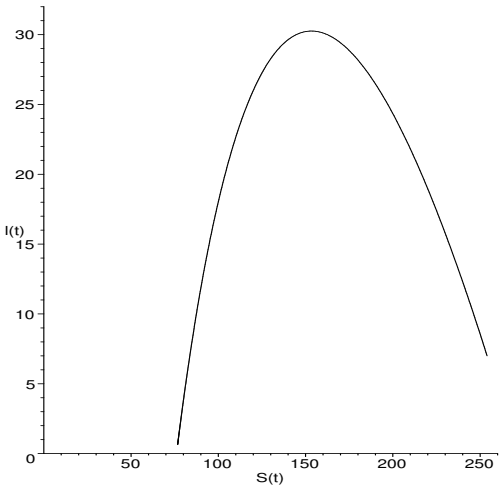

Fig. 9.3 The *S*-*I* plane.

possible, bubonic plague is transmitted mainly by rat fleas. When an infected rat is bitten by a flea, the flea becomes extremely hungry and bites the host rat repeatedly, spreading the infection in the rat. When the host rat dies, its fleas move on to other rats, spreading the disease further. As the number of available rats decreases, the fleas move to human hosts, and this is how plague starts in a human population (although the second phase of the epidemic may have been the pneumonic form of bubonic plague, which can be spread from person to person). One of the main reasons for the spread of plague from Asia into Europe was the passage of many trading ships; in medieval times ships were invariably infested with rats. An accurate model of plague transmission would have to include flea and rat populations, as well as movement in space. Such a model would be extremely complicated, and its predictions might well not be any closer to observations than our simple unrealistic model. Very recent study of the data from Eyam suggests that the rat population may not have been large enough to support the epidemic and human to human transmission

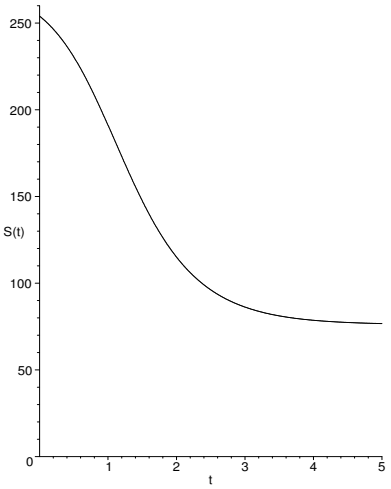

Fig. 9.4 *S* as a function of *t*.

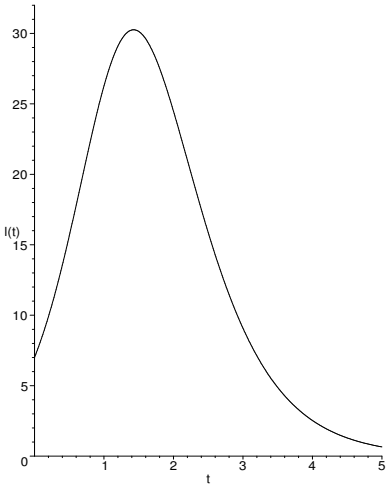

Fig. 9.5 *I* as a function of *t*.

was also a factor. Raggett (1982) also used a stochastic model to fit the data, but the fit was rather poorer than the fit for the simple deterministic model(9.2).

In the village of Eyam the rector persuaded the entire community to quarantine itself to prevent the spread of disease to other communities. One effect of this policy was to increase the infection rate in the village by keeping fleas, rats, and people in close contact with one another, and the mortality rate from bubonic plague was much higher in Eyam than in London. Further, the quarantine could do nothing to prevent the travel of rats and thus did little to prevent the spread of disease to other communities. One message this suggests to mathematical modelers is that control

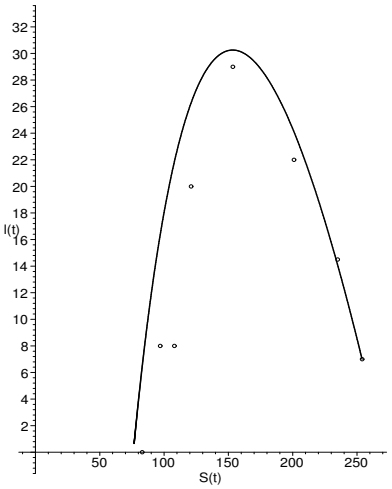

Fig. 9.6 The *S-I* plane, model and data.

strategies based on false models may be harmful, and it is essential to distinguish between assumptions that simplify but do not alter the predicted effects substantially, and wrong assumptions that make an important difference.

In order to prevent the occurrence of an epidemic if infectives are introduced into a population, it is necessary to reduce the basic reproductive number R_0 below one. This may sometimes be achieved by immunization, which has the effect of transferring members of the population from the susceptible class to the removed class and thus of reducing $S(0)$. If a fraction p of the population is successfully immunized, the effect is to replace $S(0)$ by $S(0)(1-p)$, and thus to reduce the basic reproductive number to $\beta S(0)(1-p)/\alpha$. The requirement $\beta S(0)(1-p)/\alpha < 1$ gives $1-p < \alpha/\beta S(0)$, or

$$p > 1 - \frac{\alpha}{\beta S(0)} = 1 - \frac{1}{R_0}.$$

A large basic reproductive number means that the fraction that must be immunized to prevent the spread of infection is large. This relation is connected to the idea of herd immunity, which we shall introduce in the next chapter.

Initially, the number of infectives grows exponentially because the equation for I may be approximated by

$$I' = (\beta N - \alpha)I$$

and the initial growth rate is

$$r = \beta N - \alpha = \alpha(\mathscr{R}_0 - 1)\,.$$

This initial growth rate r may be determined experimentally when an epidemic begins. Then since N and α may be measured, β may be calculated as

$$\beta = \frac{r+\alpha}{N} .$$

However, because of incomplete data and underreporting of cases, this estimate may not be very accurate. This inaccuracy is even more pronounced for an outbreak of a previously unknown disease, where early cases are likely to be misdiagnosed. Because of the final size relation, estimation of β or $\mathscr{R}_0$ is an important question that has been studied by a variety of approaches.

There are serious shortcomings in the simple Kermack–McKendrick model as a description of the beginning of a disease outbreak, and a very different kind of model is required.

Exercises

1. The same survey of Yale students described in Example 1 reported that 91.1 percent were susceptible to influenza at the beginning of the year and 51.4 percent were susceptible at the end of the year. Estimate the basic reproductive number β/α and decide whether there was an epidemic.
2. What fraction of Yale students in Exercise 1 would have had to be immunized to prevent an epidemic?
3. What was the maximum number of Yale students in Exercises 1 and 2 suffering from influenza at any time?
4. An influenza epidemic was reported at an English boarding school in 1978 that spread to 512 of the 763 students. Estimate the basic reproductive number β/α.
5. What fraction of the boarding school students in Exercise 4 would have had to be immunized to prevent an epidemic?
6. What was the maximum number of boarding school students in Exercises 4 and 5 suffering from influenza at any time?
7. A disease is introduced by two visitors into a town with 1200 inhabitants. An average infective is in contact with 0.4 inhabitants per day. The average duration of the infective period is 6 days, and recovered infectives are immune against reinfection. How many inhabitants would have to be immunized to avoid an epidemic?
8. Consider a disease with $\beta = 1/3000$, $1/\alpha = 6$ days in a population of 1200 members. Suppose the disease conferred immunity on recovered infectives. How many members would have to be immunized to avoid an epidemic?
9. A disease begins to spread in a population of 800. The infective period has an average duration of 14 days and the average infective is in contact with 0.1 persons per day. What is the basic reproductive number? To what level must the average rate of contact be reduced so that the disease will die out?
10. European fox rabies is estimated to have a transmission coefficient β of 80 km^2 years/fox and an average infective period of 5 days. There is a critical carrying capacity K_c measured in foxes per km^2, such that in regions with fox density less than K_c, rabies tends to die out, while in regions with fox density greater

than K_c, rabies tends to persist. Estimate K_c. [Remark: It has been suggested in Great Britain that hunting to reduce the density of foxes below the critical carrying capacity would be a way to control the spread of rabies.]

11. A large English estate has a population of foxes with a density of 1.3 foxes/km^2. A large fox hunt is planned to reduce the fox population enough to prevent an outbreak of rabies. Assuming that the contact number β/α is 1 km^2/fox, find what fraction of the fox population must be caught.
12. Following a complaint from the SPCA, organizers decide to replace the fox hunt of Exercise 1 by a mass inoculation of foxes for rabies. What fraction of the fox population must be inoculated to prevent a rabies outbreak?
13. What actually occurs on the estate of these exercises is that 10 percent of the foxes are killed and 15 percent are inoculated. Is there danger of a rabies outbreak.
14. Here is another approach to the analysis of the SIR model (9.2).

    (i) Divide the two equations of the model to give

$$\frac{I'}{S'} = \frac{dI}{dS} = \frac{(\beta S - \alpha)I}{-\beta SI} = -1 + \frac{\alpha}{\beta S}.$$

    (ii) Integrate to find the orbits in the (S,I)-plane,

$$I = -S + \frac{\alpha}{\beta}\log S + c\,,$$

    with c an arbitrary constant of integration.

    (iii) Define the function

$$V(S,I) = S + I - \frac{\alpha}{\beta}\log S$$

    and show that each orbit is given implicitly by the equation $V(S,I) = c$ for some choice of the constant c.

    (iv) Show that no orbit reaches the I-axis and deduce that $S_\infty = \lim_{t\to\infty} S(t) > 0$, which implies that part of the population escapes infection.

9.3 A Branching-Process Disease-Outbreak Model

The Kermack–McKendrick compartmental epidemic model assumes that the sizes of the compartments are large enough that the mixing of members is homogeneous, or at least that there is homogeneous mixing in each subgroup if the population is stratified by activity levels. However, at the beginning of a disease outbreak, there is a very small number of infective individuals, and the transmission of infection is a stochastic event depending on the pattern of contacts between members of the population; a description should take this pattern into account.

Our approach will be to give a stochastic-branching process description of the beginning of a disease outbreak to be applied as long as the number of infectives remains small, distinguishing a (minor) disease outbreak confined to this stage from a (major) epidemic, which occurs if the number of infectives begins to grow at an exponential rate. Once an epidemic has started, we may switch to a deterministic compartmental model, arguing that in a major epidemic, contacts would tend to be more homogeneously distributed. Implicitly, we are thinking of an infinite population, and by a major epidemic we mean a situation in which a nonzero fraction of the population is infected, and by a minor outbreak we mean a situation in which the infected population may grow but remains a negligible fraction of the population.

There is an important difference between the behavior of branching process models and the behavior of models of Kermack–McKendrick type, namely, as we shall see in this section that for a stochastic disease outbreak model if $\mathscr{R}_0 < 1$, the probability that the infection will die out is 1, but if $\mathscr{R}_0 > 1$, there is a positive probability that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic.

We describe the network of contacts between individuals by a graph with members of the population represented by vertices and with contacts between individuals represented by edges. The study of graphs originated with the abstract theory of Erdős and Rényi of the 1950s and 1960s [Erdős and Rényi (1959, 1960, 1961)]. It has become important in many areas of application, including social contacts and computer networks, as well as the spread of communicable diseases. We will think of networks as bidirectional, with disease transmission possible in either direction along an edge.

An edge is a contact between vertices that can transmit infection. The number of edges of a graph at a vertex is called the *degree* of the vertex. The degree distribution of a graph is $\{p_k\}$, where p_k is the fraction of vertices having degree k. The degree distribution is fundamental in the description of the spread of disease.

We think of a small number of infectives in a population of susceptibles large enough that in the initial stage, we may neglect the decrease in the size of the susceptible population. Our development begins along the lines of that of [Diekmann and Heesterbeek (2000)] and then develops along the lines of [Callaway, Newman, Strogatz, and Watts (2000), Newman (2002), Newman, Strogatz, and Watts (2002)]. We assume that the infectives make contacts independently of one another and let p_k denote the probability that the number of contacts by a randomly chosen individual is exactly k, with $\sum_{k=0}^{\infty} p_k = 1$. In other words, $\{p_k\}$ is the degree distribution of the vertices of the graph corresponding to the population network. For the moment, we assume that every contact leads to an infection, but we will relax this assumption later.

It is convenient to define the *probability generating function*

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Since $\sum_{k=0}^{\infty} p_k = 1$, this power series converges for $0 \le z \le 1$, and may be differentiated term by term. Thus

$$p_k = \frac{G_0^{(k)}(0)}{k!}, \quad k = 0, 1, 2, \ldots.$$

It is easy to verify that the generating function has the properties

$$G_0(0) = p_0, \quad G_0(1) = 1, \quad G_0'(z) > 0, \quad G_0''(z) > 0.$$

The mean degree, which we denote by $\langle k \rangle$ or z_1, is

$$\langle k > \rangle = \sum_{k=1}^{\infty} k p_k = G_0'(1).$$

More generally, we define the moments

$$\langle k^j \rangle = \sum_{k=1}^{\infty} k^j p_k, \quad j = 1, 2, \ldots \infty.$$

When a disease is introduced into a network, we think of it as starting at a vertex (patient zero) that transmits infection to every individual to whom this individual is connected, that is, along every edge of the graph from the vertex corresponding to this individual. We may think of this individual as being inside the population, as when a member of a population returns from travel after being infected, or as being outside the population, as when someone visits a population and brings an infection. For transmission of disease after this initial contact we need to use the *excess degree* of a vertex. If we follow an edge to a vertex, the excess degree of this vertex is one less than the degree. We use the excess degree because infection cannot be transmitted back along the edge whence it came. The probability of reaching a vertex of degree k, or excess degree $(k-1)$, by following a random edge is proportional to k, and thus the probability that a vertex at the end of a random edge has excess degree $(k-1)$ is a constant multiple of kp_k with the constant chosen to make the sum over k of the probabilities equal to 1. Then the probability that a vertex has excess degree $(k-1)$ is

$$q_{k-1} = \frac{k p_k}{\langle k \rangle}.$$

This leads to a generating function $G_1(z)$ for the excess degree,

$$G_1(z) = \sum_{k=1}^{\infty} q_{k-1} z^{k-1} = \sum_{k=1}^{\infty} \frac{k p_k}{\langle k \rangle} z^{k-1} = \frac{1}{\langle k \rangle} G_0'(z),$$

and the mean excess degree, which we denote by $\langle k_e \rangle$, is

$$\begin{aligned}\langle k_e\rangle &= \frac{1}{\langle k\rangle}\sum_{k=1}^{\infty} k(k-1)p_k \\ &= \frac{1}{\langle k\rangle}\sum_{k=1}^{\infty} k^2 p_k - \frac{1}{\langle k\rangle}\sum_{k=1}^{\infty} kp_k \\ &= \frac{\langle k^2\rangle}{\langle k\rangle} - 1 = G_1'(1).\end{aligned}$$

We let $\mathscr{R}_0 = G_1'(1)$, the mean excess degree. This is the mean number of secondary cases infected by patient zero and is the basic reproduction number as usually defined; the threshold for an epidemic is determined by $\mathscr{R}_0$. The quantity $\langle k_e\rangle = G_1'(1)$ is sometimes written in the form

$$\langle k_e\rangle = G_1'(1) = \frac{z_2}{z_1},$$

where $z_2 = \sum_{k=1}^{\infty} k(k-1)p_k = \langle k^2\rangle - \langle k\rangle$ is the mean number of second neighbors of a random vertex.

Our next goal is to calculate the probability that the infection will die out and will not develop into a major epidemic, proceeding in two steps. First we find the probability that a secondary infected vertex (a vertex that has been infected by another vertex in the population) will not spark a major epidemic.

Suppose that the secondary infected vertex has excess degree j. We let z_n denote the probability that this infection dies out within the next n generations. For the infection to die out in n generations, each of the j secondary infections coming from the initial secondary infected vertex must die out in $(n-1)$ generations. The probability of this is z_{n-1} for each secondary infection, and the probability that all secondary infections will die out in $(n-1)$ generations is z_{n-1}^j. Now z_n is the sum over j of these probabilities, weighted by the probability q_j of j secondary infections. Thus

$$z_n = \sum_{j=0}^{\infty} q_j z_{n-1}^j = G_1(z_{n-1}).$$

Since $G_1(z)$ is an increasing function, the sequence z_n is an increasing sequence and has a limit z_∞, which is the probability that this infection will die out eventually. Then z_∞ is the limit as $n \to \infty$ of the solution of the difference equation

$$z_n = G_1(z_{n-1}), \quad z_0 = 0.$$

Thus z_∞ must be an equilibrium of this difference equation, that is, a solution of $z = G_1(z)$. Let w be the smallest positive solution of $z = G_1(z)$. Then, because $G_1(z)$ is an increasing function of z, $z \le G_1(z) \le G_1(w) = w$ for $0 \le z \le w$. Since $z_0 = 0 < w$ and $z_{n-1} \le w$ implies

$$z_n = G_1(z_{n-1}) \le G_1(w) = w,$$

it follows by induction that

$$z_n \le w, \quad n = 0, 1, \ldots \infty.$$

From this we deduce that

$$z_\infty = w.$$

The equation $G_1(z) = z$ has a root $z = 1$, since $G_1(1) = 1$. Because the function $G_1(z) - z$ has a positive second derivative, its derivative $G_1'(z) - 1$ is increasing and can have at most one zero. This implies that the equation $G_1(z) = z$ has at most two roots in $0 \le z \le 1$. If $\mathscr{R}_0 < 1$, the function $G_1(z) - z$ has a negative first derivative

$$G_1'(z) - 1 \le G_1'(1) - 1 = \mathscr{R}_0 - 1 < 0,$$

and the equation $G_1(z) = z$ has only one root, namely $z = 1$. On the other hand, if $\mathscr{R}_0 > 1$, the function $G_1(z) - z$ is positive for $z = 0$ and negative near $z = 1$ since it is zero at $z = 1$, and its derivative is positive for $z < 1$ and z near 1. Thus in this case the equation $G_1(z) = z$ has a second root $z_\infty < 1$.

This root z_∞ is the probability that an infection transmitted along one of the edges at the initial secondary vertex will die out, and this probability is independent of the excess degree of the initial secondary vertex. It is also the probability that an infection originating outside the population, such as an infection brought from outside into the population under study, will die out.

Next, we calculate the probability that an infection originating at a primary infected vertex, such as an infection introduced by a visitor from outside the population under study, will die out. The probability that the disease outbreak will die out eventually is the sum over k of the probabilities that the initial infection in a vertex of degree k will die out, weighted by the degree distribution $\{p_k\}$ of the original infection, and this is

$$\sum_{k=0}^{\infty} p_k z_\infty^k = G_0(z_\infty).$$

To summarize this analysis, we see that if $\mathscr{R}_0 < 1$, the probability that the infection will die out is 1. On the other hand, if $\mathscr{R}_0 > 1$, there is a solution $z_\infty < 1$ of

$$G_1(z) = z,$$

and there is a probability $1 - G_0(z_\infty) > 0$ that the infection will persist, and will lead to an epidemic. However, there is a positive probability $G_0(z_\infty)$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic. This distinction between a minor outbreak and a major epidemic, and the result that if $\mathscr{R}_0 > 1$ there may be only a minor outbreak and not a major epidemic, are aspects of stochastic models not reflected in deterministic models.

If contacts between members of the population are random, corresponding to the assumption of mass action in the transmission of disease, then the probabilities p_k are given by the *Poisson distribution*

$$p_k = \frac{e^{-c}c^k}{k!}$$

for some constant c [Brauer, van den Driessche, and Wu (2008), pp. 142–143]. The generating function for the Poisson distribution is $e^{c(z-1)}$. Then $G_1(z) = G_0(z)$, and $\mathscr{R}_0 = c$, so that

$$G_1(z) = G_0(z) = e^{\mathscr{R}_0(z-1)}.$$

The commonly observed situation that most infectives do not pass on infection but there are a few "superspreading events" [Riley et al. (2003)] corresponds to a probability distribution that is quite different from a Poisson distribution, and could give a quite different probability that an epidemic will occur. For example, if $\mathscr{R}_0 = 2.5$, the assumption of a Poisson distribution gives $z_\infty = 0.107$ and $G_0(z_\infty) = 0.107$, so that the probability of an epidemic is 0.893. The assumption that nine out of ten infectives do not transmit infection while the tenth transmits 25 infections gives

$$G_0(z) = (z^{25} + 9)/10, \quad G_1(z) = z^{24}, \quad z_\infty = 0, \quad G_0(z_\infty) = 0.9,$$

from which we see that the probability of an epidemic is 0.1. Another example, possibly more realistic, is to assume that a fraction $(1-p)$ of the population follows a Poisson distribution with constant r, while the remaining fraction p consists of superspreaders each of whom makes L contacts. This would give a generating function

$$\begin{aligned} G_0(z) &= (1-p)e^{r(z-1)} + pz^L, \\ G_1(z) &= \frac{r(1-p)e^{r(z-1)} + pLz^{L-1}}{r(1-p) + pL}, \end{aligned}$$

and

$$\mathscr{R}_0 = \frac{r^2(1-p) + pL(L-1)}{r(1-p) + pL}.$$

For example, if $r = 2.2$, $L = 10$, $p = 0.01$, numerical simulation gives

$$\mathscr{R}_0 = 2.5, \quad z_\infty = 0.146,$$

so that the probability of an epidemic is 0.849.

These examples demonstrate that the probability of a major epidemic depends strongly on the nature of the contact network. Simulations suggest that for a given value of the basic reproduction number, the Poisson distribution is the one with the maximum probability of a major epidemic.

It has been observed that in many situations there is a small number of long-range connections in the graph, allowing rapid spread of infection. There is a high degree of clustering (some vertices with many edges), and there are short path lengths. Such a situation may arise if a disease is spread to a distant location by an air traveler. This type of network is called a *small-world* network. Long range connections in a network can increase the likelihood of an epidemic dramatically.

These examples indicate that the probability of an epidemic depends strongly on the contact network at the beginning of a disease outbreak. We will not explore network models further here, but we point out that this is an actively developing field of science. Some basic references are [Newman (2003, Strogatz (2001)].

9.3.1 Transmissibility

Contacts do not necessarily transmit infection. For each contact between individuals of whom one has been infected and the other is susceptible, there is a probability that infection will actually be transmitted. This probability depends on such factors as the closeness of the contact, the infectivity of the member who has been infected, and the susceptibility of the susceptible member. We assume that there is a mean probability T, called the *transmissibility*, of transmission of infection. The transmissibility depends on the rate of contacts, the probability that a contact will transmit infection, the duration time of the infection, and the susceptibility. The development in Section 9.2 assumed that all contacts transmit infection, that is, that $T = 1$.

In this section, we will continue to assume that there is a network describing the contacts between members of the population whose degree distribution is given by the generating function $G_0(z)$, but we will assume in addition that there is a mean transmissibility T.

When disease begins in a network, it spreads to some of the vertices of the network. Edges that are infected during a disease outbreak are called *occupied*, and the size of the disease outbreak is the cluster of vertices connected to the initial vertex by a continuous chain of occupied edges.

The probability that exactly m infections are transmitted by an infective vertex of degree k is

$$\binom{k}{m} T^m (1-T)^{k-m}.$$

We define $\Gamma_0(z,T)$to be the generating function for the distribution of the number of occupied edges attached to a randomly chosen vertex, which is the same as the distribution of the infections transmitted by a randomly chosen individual for any (fixed) transmissibility T. Then

$$\begin{aligned}
\Gamma_0(z,T) &= \sum_{m=0}^{\infty} \left[\sum_{k=m}^{\infty} p_k \binom{k}{m} T^m (1-T)^{(k-m)} \right] z^m \\
&= \sum_{k=0}^{\infty} p_k \left[\sum_{m=0}^{k} \binom{k}{m} (zT)^m (1-T)^{(k-m)} \right] \qquad (9.6) \\
&= \sum_{k=0}^{\infty} p_k [zT + (1-T)]^k = G_0(1 + (z-1)T).
\end{aligned}$$

In this calculation we have used the binomial theorem to see that

$$\sum_{m=0}^{k} \binom{k}{m} (zT)^m (1-T)^{(k-m)} = [zT + (1-T)]^k.$$

Note that

$$\Gamma_0(0,T) = G_0(1-T), \quad \Gamma_0(1,T) = G_0(1) = 1, \quad \Gamma_0'(z,T) = TG_0'(1+(z-1)T).$$

For secondary infections we need the generating function $\Gamma_1(z,T)$ for the distribution of occupied edges leaving a vertex reached by following a randomly chosen edge. This is obtained from the excess degree distribution in the same way,

$$\Gamma_1(z,T) = G_1(1+(z-1)T)$$

and

$$\Gamma_1(0,T) = G_1(1-T), \quad \Gamma_1(1,T) = G_1(1) = 1, \quad \Gamma_1'(z,T) = TG_1'(1+(z-1)T).$$

The basic reproduction number is now

$$\mathscr{R}_0 = \Gamma_1'(1,T) = TG_1'(1).$$

The calculation of the probability that the infection will die out and will not develop into a major epidemic follows the same lines as the argument for $T = 1$. The result is that if $\mathscr{R}_0 = TG_1'(1) < 1$, the probability that the infection will die out is 1. If $\mathscr{R}_0 > 1$, there is a solution $z_\infty(T) < 1$ of

$$\Gamma_1(z,T) = z,$$

and a probability $1 - \Gamma_0(z_\infty(T),T) > 0$ that the infection will persist, and will lead to an epidemic. However, there is a positive probability $\Gamma_1(z_\infty(T),T)$ that the infection will increase initially but will produce only a minor outbreak and will die out before triggering a major epidemic.

Another interpretation of the basic reproduction number is that there is a *critical transmissibility* T_c defined by

$$T_c G_1'(1) = 1.$$

In other words, the critical transmissibility is the transmissibility that makes the basic reproduction number equal to 1. If the mean transmissibility can be decreased below the critical transmissibility, then an epidemic can be prevented.

The measures used to try to control an epidemic may include contact interventions, that is, measures affecting the network such as avoidance of public gatherings and rearrangement of the patterns of interaction between caregivers and patients in a hospital, and transmission interventions such as careful hand washing or face masks to decrease the probability that a contact will lead to disease transmission.

Exercises

In each exercise, assume that the transmissibility is 1.

1. Show that it is not possible for a major epidemic to develop unless at least one member of the contact network has degree at least 3.
2. What is the probability of a major epidemic if every member of the contact network has degree 3.
3. Consider a truncated Poisson distribution, with

$$p_k = \begin{cases} \frac{e^{-c}c^k}{k!}, & k \le 10, \\ 0, & k > 10. \end{cases}$$

   Estimate (numerically) the probability of a major epidemic if $c = 1.5$.
4. Show that the probability generating function for an exponential distribution, given by

$$p_k = (1 - e^{-1/r})e^{-k/r},)$$

   is

$$G_0(z) = \frac{1 - e^{-1/r}}{1 - ze^{-1/r}}.$$

5. A power law distribution is given by

$$p_k = Ck^{-\alpha}.$$

   For what values of α is it possible to normalize this (i.e., choose C to make $\sum p_k = 1$?

9.4 Network and Compartmental Epidemic Models

Compartmental models for epidemics are not suitable for describing the beginning of a disease outbreak because they assume that all members of a population are equally likely to make contact with a very small number of infectives. Thus, as we have seen in the preceding section, stochastic branching process models are better descriptions of the beginning of an epidemic. They allow the possibility that even if a disease outbreak has a reproduction number greater than 1, it may be only a minor outbreak and may not develop into a major epidemic. One possible approach to a more realistic description of an epidemic would be to use a branching process model initially and then make a transition to a compartmental model when the epidemic has become established and there are enough infectives that mass action mixing in the population is a reasonable approximation. Another approach would be to continue to use a network model throughout the course of the epidemic. In this section we shall indicate how a compartmental approach and a network approach are related.

The development is taken from Volz (2008), Miller (2010), and Miller and Volz (2011).

We assume that there is a known static *configuration model* (CM) network in which the probability that a node u has degree k_u is $P(k_u)$). We let $G_0(z)$ denote the probability generating function of the degree distribution,

$$G_0(z) = \sum_{k=0}^{\infty} p_k z^k,$$

with mean degree $\langle\rangle >= G_0'(1)$.

The per-edge from an infected node is assumed to be β, and it is assumed that infected nodes recover at a rate α. We use an edge-based compartmental model because the probability that a random neighbor is infected is not necessarily the same as the probability that a random individual is infected. We let $S(t)$ denote the fraction of nodes that are susceptible at time t, $I(t)$ the fraction of nodes that are infective at time t, and $R(t)$ the fraction of nodes that are recovered at time t. It is easy to write an equation for R', the rate at which infectives recover. If we know $S(t)$, we can find $I(t)$, because a decrease in S gives a corresponding increase in I. Since

$$S(t) + I(t) + R(t) = 1,$$

we need only find the probability that a randomly selected node is susceptible.

We assume that the hazard of infection for a susceptible node u is proportional to the degree k_u of the node. Each contact is represented by an edge of the network joining u to a neighboring node. We let φ_I denote the probability that this neighbor is infective. Then the per-edge hazard of infection is

$$\lambda_E = \beta \varphi_I.$$

Assuming that edges are independent, u's hazard of infection at time t is

$$\lambda_u(t) = k_u \lambda_E(t) = k_u \beta \varphi_I(t).$$

Consider a randomly selected node u and let $\theta(t)$ be the probability that a random neighbor has not transmitted infection to u. Then the probability that u is susceptible is θ^{k_u}. Averaging over all nodes, we see that the probability that a random node u is susceptible is

$$S(t) = \sum_{k=0}^{\infty} P(k)[\theta(t)]^k = G_0(\theta(t)). \tag{9.7}$$

We break θ into three parts,

$$\theta = \varphi_S + \varphi_I + \varphi_R,$$

with φ_S the probability that a random neighbor v of u is susceptible, φ_I the probability that a random neighbor v of u is infective but has not transmitted infection to u,

and φ_R the probability that a random neighbor v has recovered without transmitting infection to u. Then the probability that v has transmitted infection to u is $1-\theta$.

Since infected neighbors recover at rate α, the flux from φ_I to φ_R is $\alpha\varphi_I$. Thus

$$\varphi_R' = \alpha\varphi_I.$$

It is easy to see from this that

$$R' = \alpha I. \tag{9.8}$$

Since edges from infected neighbors transmit infection at rate β, the flux from φ_I to $(1-\theta)$ is $\beta\varphi_I$. Thus

$$\theta' = -\beta\varphi_I. \tag{9.9}$$

To obtain φ_I' we need the flux into and out of the φ_I compartment. The incoming flux from φ_S results from infection of the neighbor. The outgoing flux to φ_R corresponds to recovery of the neighbor without having transmitted infection, and the outgoing flux to $(1-\theta)$ corresponds to transmission without recovery. The total outgoing flux is $(\alpha+\beta)\varphi_I$.

To determine the flux from φ_S to φ_I, we need the rate at which a neighbor changes from susceptible to infective. Consider a random neighbor v of u; the probability that v has degree k is$kp(k)/\langle k\rangle$. Since there are $(k-1)$ neighbors of v that could have infected v, the probability that v is susceptible is θ^{k-1}. Averaging over all k, we see that the probability that a random neighbor v of u is susceptible is

$$\varphi_S = \sum_{k=0}^{\infty} \frac{kp(k)}{\langle k\rangle}\theta^{k-1} = \frac{G_0'(\theta)}{G_0'(1)}. \tag{9.10}$$

To calculate φ_R, we note that the flux from φ_I to φ_R and the flux from φ_I to $(1-\theta)$ are proportional with proportionality constant α/β. Since both φ_R and $(1-\theta)$ start at zero,

$$\varphi_R = \frac{\alpha}{\beta}(1-\theta). \tag{9.11}$$

Now, using (9.9), (9.10), (9.11), and

$$\varphi_I = \theta - \varphi_S - \varphi_R,$$

we obtain

$$\theta' = -\beta\varphi_I = -\beta\theta + \beta\varphi_S + \beta\varphi_R = -\beta\theta + \beta\frac{G_0'(\theta)}{G_0'(1)} + \alpha(1-\theta). \tag{9.12}$$

We now have a dynamic model consisting of equations (9.12), (9.7), (9.8), and $S+I+R=1$. We wish to show a relationship between this set of equations and the simple Kermack–McKendrick compartmental model (9.2). In order to accomplish this, we need only show under what conditions we would have $S' = -\beta SI$.

Differentiating (9.7) and using (9.9), we obtain

$$S' = G_0'(\theta)\theta' = -G_0'(\theta)\beta\varphi_I.$$

Consider a large population with N members, each making $C \leq N-1$ contacts, so that

$$S = \theta^C, \quad G_0'(\theta) = \frac{CS}{\theta}S'(\theta),$$

and

$$S' = -\beta CS\frac{\varphi_I}{\theta}.$$

We now let $C \to \infty$ (which implies $N \to \infty$) in such a way that

$$\hat{\beta} = \beta C$$

remains constant. Then

$$S' = -\hat{\beta}\frac{\varphi_I}{\theta}.$$

We will now show that

$$\frac{\varphi_I}{\theta} \approx 1,$$

and this will yield the desired approximation

$$S' = -\hat{\beta}SI. \tag{9.13}$$

The probability that an edge to a randomly chosen node has not transmitted infection is θ (assuming that the given target node cannot transmit infection), and the probability that in addition it is connected to an infected node is φ_I. Because $\hat{\beta} = \beta C$ is constant and therefore bounded as C grows, only a fraction no greater than a constant multiple of I/C of edges to the target node may have transmitted infection from a node that is still infected. For large values of C, φ_I is approximately I. Similarly, θ is approximately 1 as $C \to \infty$. Thus $\varphi_I/\theta \approx I$ as $C \to \infty$. This gives the desired approximate equation for S. The result remains valid if all degrees are close to the average degree as the average degree grows.

The edge-based compartmental modeling approach that we have used can be generalized in several ways. For example, heterogeneity of mixing can be included. In general, one would expect that early infections would be in individuals having more contacts, and thus that an epidemic would develop more rapidly than a mass action compartmental model would predict. When contact duration is significant, as would be the case in sexually transmitted diseases, an individual with a contact would play no further role in disease transmission until a new contact is made, and this can be incorporated in a network model.

The network approach to disease modeling is a rapidly developing field of study, and there will undoubtedly be fundamental developments in our understanding of the modeling of disease transmission. Some useful references are [Bansal, Read, Pourbohloul, and Meyers (2010), Meyers (2007), Meyers et al. (2006), Meyers et al. (2005), Newman(2001), Newman (2002), Newman, Strogatz, and Watts (2001), Strogatz (2001)].

In the remainder of this chapter, we assume that we are in an epidemic situation following a disease outbreak that has been modeled initially by a branching process. Thus we return to the study of compartmental models.

9.5 More Complicated Epidemic Models

We have established that the simple Kermack–McKendrick epidemic model (9.2) has the following basic properties:

1. There is a basic reproduction number $\mathscr{R}_0$ such that if $\mathscr{R}_0 < 1$, the disease dies out while if $\mathscr{R}_0 > 1$, there is an epidemic.
2. The number of infectives always approaches zero and the number of susceptibles always approaches a positive limit as $t \to \infty$.
3. There is a relationship between the reproduction number and the final size of the epidemic, which is an equality if there are no disease deaths.

In fact, these properties hold for epidemic models with more complicated compartmental structure. We will describe some common epidemic models as examples.

9.5.1 Exposed Periods

In many infectious diseases there is an exposed period after the transmission of infection from susceptibles to potentially infective members but before these potential infectives develop symptoms and can transmit infection. To incorporate an exposed period with mean exposed period $1/\kappa$, we add an exposed class E and use compartments S, E, I, R and total population size $N = S + E + I + R$ to give a generalization of the epidemic model (9.2)

$$\begin{aligned} S' &= -\beta SI, \\ E' &= \beta SI - \kappa E, \\ I' &= \kappa E - \alpha I. \end{aligned} \tag{9.14}$$

A flow chart is shown in Figure 9.7.

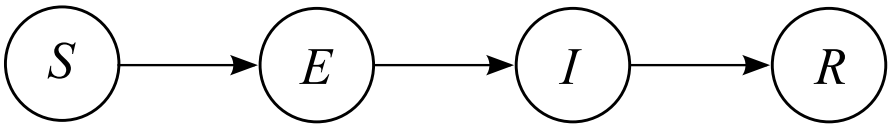

Fig. 9.7 Flow chart for the SEIR model.

The analysis of this model is the same as the analysis of (9.2), but with I replaced by $E+I$. That is, instead of using the number of infectives as one of the variables, we use the total number of infected members, whether or not they are capable of transmitting infection.

In some diseases there is some infectivity during the exposed period. This may be modeled by assuming infectivity reduced by a factor ε during the exposed period. A calculation of the rate of new infections per susceptible leads to a model

$$\begin{aligned} S' &= -\beta S(I+\varepsilon E), \\ E' &= \beta S(I+\varepsilon E) - \kappa E, \\ I' &= \kappa E - \alpha I\ . \end{aligned} \tag{9.15}$$

We take initial conditions

$$S(0) = S_0, \quad E(0) = E_0, \quad I(0) = I_0.$$

For this model,

$$\mathscr{R}_0 = \frac{\beta N}{\alpha} + \varepsilon \frac{\beta N}{\kappa}\ .$$

Integration of the sum of the equations of (9.14) from 0 to ∞ gives

$$N - S_\infty = \alpha \int_0^\infty I(s)ds.$$

Integration of the third equation of (9.15) gives

$$\kappa \int_0^\infty E(s)ds = \alpha \int_0^\infty I(s)ds - I_0,$$

and division of the first equation of (9.15) by S followed by integration from 0 to ∞ gives

$$\begin{aligned} \log \frac{S_0}{S_\infty} &= \int_0^\infty \beta[I(s) + \varepsilon E(s)ds \\ &= \beta \int_0^\infty [I(s) + \varepsilon E(s)ds \\ &= \beta \left[\varepsilon + \frac{\kappa}{\alpha}\right] \int_0^\infty E(s)ds - \frac{\varepsilon\beta I_0}{\kappa} \\ &= \mathscr{R}_0 \left[1 - \frac{S_\infty}{N}\right] - \frac{\varepsilon\beta I_0}{\kappa}. \end{aligned}$$

In this final size relation there is an initial term $\beta I_0/\alpha$, caused by the assumption that there are individuals infected originally who are beyond the exposed stage in which they would have had some infectivity. In order to obtain a final size relation without such an initial term it is necessary to assume $I(0) = 0$, that initial infectives

are in the first stage in which they can transmit infection. If $I(0) = 0$, the final size relation has the form (9.3).

9.5.2 Treatment Models

One form of treatment that is possible for some diseases is vaccination to protect against infection before the beginning of an epidemic. For example, this approach is commonly used for protection against annual influenza outbreaks. A simple way to model this would be to reduce the total population size by the fraction of the population protected against infection.

In reality, such inoculations are only partly effective, decreasing the rate of infection and also decreasing infectivity if a vaccinated person does become infected. This may be modeled by dividing the population into two groups with different model parameters, which would require some assumptions about the mixing between the two groups. This is not difficult, but we will not explore this direction here.

If there is a treatment for infection once a person has been infected, this may be modeled by supposing that a fraction γ per unit time of infectives is selected for treatment, and that treatment reduces infectivity by a fraction δ. Suppose that the rate of removal from the treated class is η. This leads to the SITR model, where T is the treatment class, given by

$$\begin{aligned} S' &= -\beta S[I + \delta T], \\ I' &= \beta S[I + \delta T] - (\alpha + \gamma), I \\ T' &= \gamma I - \eta T. \end{aligned} \tag{9.16}$$

A flow chart is shown in Figure 9.8.

It is not difficult to prove, much as was done for the model (9.2), that

$$S_\infty = \lim_{t\to\infty} S(t) > 0, \quad \lim_{t\to\infty} I(t) = \lim_{t\to\infty} T(t) = 0.$$

In order to calculate the basic reproduction number, we may argue that an infective in a totally susceptible population causes βN new infections in unit time, and the mean time spent in the infective compartment is $1/(\alpha+\gamma)$. In addition, a fraction $\gamma/(\alpha+\gamma)$ of infectives are treated. While in the treatment stage the number of new infections caused in unit time is $\delta\beta N$, and the mean time in the treatment class is $1/\eta$. Thus $\mathscr{R}_0$ is

$$\mathscr{R}_0 = \frac{\beta N}{\alpha+\gamma} + \frac{\gamma}{\alpha+\gamma}\frac{\delta\beta N}{\eta}. \tag{9.17}$$

It is also possible to establish the final size relation (9.3) by means very similar to those used for the simple model (9.2). We integrate the first equation of (9.16) to obtain

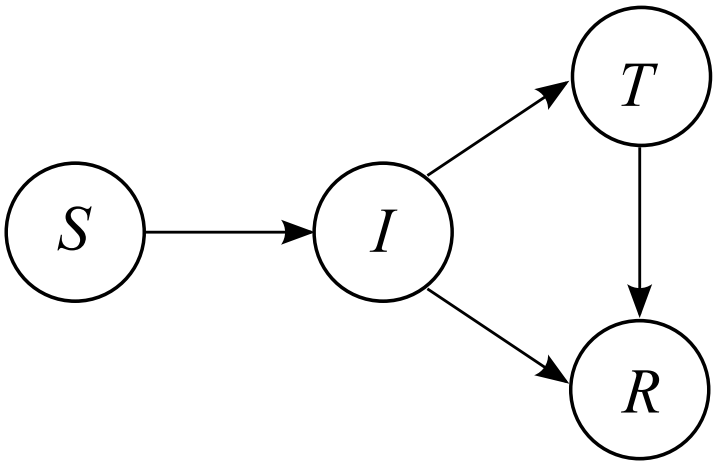

Fig. 9.8 Flow chart for the SITR model.

$$\log \frac{S_0}{S_\infty} = \int_0^\infty \beta[I(t) + \delta T(t)]dt = \beta \int_0^\infty [I(t) + \delta T(t)]dt.$$

Integration of the third equation of (9.16) gives

$$\gamma \int_0^\infty I(t)dt = \eta \int_0^\infty T(t)dt.$$

Integration of the sum of the first two equations of (9.16) gives

$$N - S_\infty = (\alpha + \gamma) \int_0^\infty I(t)dt.$$

Combination of these three equations and (9.17) gives (9.3).

9.5.3 An Influenza Model

In some diseases, such as influenza, at the end of a stage individuals may proceed to one of two stages. There is a latent period after which a fraction p of latent individuals L proceeds to an infective stage I, while the remaining fraction $(1-p)$ proceeds to an asymptomatic stage A, with infectivity reduced by a factor δ and a different period $1/\eta$. The influenza model of [Arino et al. (2006, 2007)] is

$$\begin{aligned} S' &= -\beta S[I + \delta A], \\ L' &= \beta S[I + \delta A] - \kappa L, \\ I' &= p\kappa L - \alpha I, \\ A' &= (1-p)\kappa L - \eta A, \end{aligned} \tag{9.18}$$

and

$$\mathscr{R}_0 = \beta N \left[\frac{p}{\alpha} + \frac{\delta(1-p)}{\eta} \right].$$

A flow chart is shown in Figure 9.9.

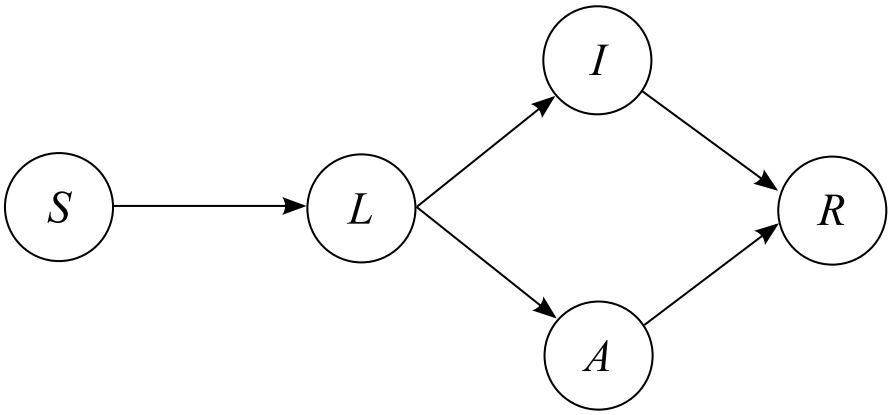

Fig. 9.9 Influenza model flowchart.

The same approach used in earlier examples leads to the same final size relation (9.3).

The model (9.18) is an example of a differential infectivity model. In such models, also used in the study of HIV/AIDS [Hyman, Li and Stanley (1999)], individuals enter a specific group when they become infected and stay in that group over the course of the infection. Different groups may have different parameter values. For example, for influenza infective and asymptomatic members may have different infectivities and different periods of stay in the respective stages.

9.5.4 A Quarantine-Isolation Model

For an outbreak of a new disease, where no vaccine is available, isolation of diagnosed infectives and quarantine of people who are suspected of having been infected (usually by tracing of contacts of diagnosed infectives) are the only control measures available. We formulate a model to describe the course of an epidemic, originally introduced for modeling the SARS epidemic of 2002-20033 [Gumel et al. (2004)], when control measures are begun under the followingassumptions:

1. Exposed members may be infective with infectivity reduced by a factor ε_E, $0 \leq \varepsilon_E < 1$.
2. Exposed members who are not isolated become infective at rate κ_E.

3. We introduce a class Q of quarantined members and a class J of isolated (hospitalized) members, and exposed members are quarantined at a proportional rate γ_Q in unit time (in practice, a quarantine will also be applied to many susceptibles, but we ignore this in the model). Quarantine is not perfect, but it reduces the contact rate by a factor ε_Q. The effect of this assumption is that some susceptibles make fewer contacts than the model assumes.
4. Infectives are diagnosed at a proportional rate γ_I per unit time and isolated. Isolation is imperfect, and there may be transmission of disease by isolated members, with an infectivity factor of ε_J.
5. Quarantined members are monitored, and when they develop symptoms at rate κ_Q they are isolated immediately.
6. Infectives leave the infective class at rate α_I and isolated members leave the isolated class at rate α_J.

These assumptions lead to the SEQIJR model [Gumel et al. (2004)]:

$$\begin{aligned} S' &= -\beta S[\varepsilon_E E + \varepsilon_E \varepsilon_Q Q + I + \varepsilon_J J], \\ E' &= \beta S[\varepsilon_E E + \varepsilon_E \varepsilon_Q Q + I + \varepsilon_J J] - (\kappa_E + \gamma_Q)E, \\ Q' &= \gamma_Q E - \kappa_J Q, \\ I' &= \kappa_E E - (\alpha_I + \gamma_J)I, \\ J' &= \kappa_Q Q + \gamma_J I - \alpha_J J\ . \end{aligned} \tag{9.19}$$

The model before control measures are begun is the special case

$$\gamma_Q = \gamma_J = \kappa_Q = \alpha_J = 0, \quad Q = J = 0$$

of (9.19). It is the same as (9.15).

A flow chart is shown in Figure 9.10.

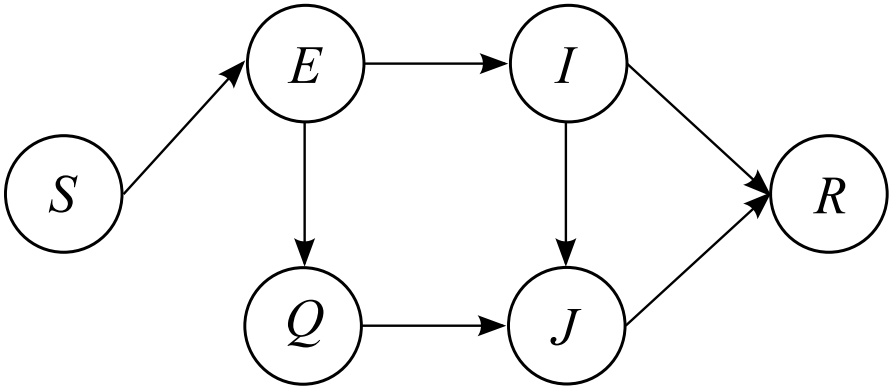

Fig. 9.10 Flow chart for the SEQIJR model.

We define the *control reproduction number* $\mathscr{R}_c$ to be the number of secondary infections caused by a single infective in a population consisting essentially only of susceptibles with the control measures in place. It is analogous to the basic reproduction number, but instead of describing the very beginning of the disease outbreak it describes the beginning of the recognition of the epidemic. The basic reproduction number is the value of the control reproduction number with

$$\gamma_Q = \gamma_J = \kappa_Q = \alpha_J = 0 \, .$$

We have already calculated $\mathscr{R}_0$ for (9.15), and we may calculate $\mathscr{R}_c$ in the same way but using the full model with quarantined and isolated classes. We obtain

$$\mathscr{R}_c = \frac{\varepsilon_E \beta N}{D_1} + \frac{\beta N \kappa_E}{D_1 D_2} + \frac{\varepsilon_Q \varepsilon_E \beta N \gamma_Q}{D_1 \kappa_Q} + \frac{\varepsilon_J \beta N \kappa_E \gamma_J}{\alpha_J D_1 D_2} + \frac{\varepsilon_J \beta N \gamma_Q}{\alpha_J D_1} \, ,$$

where $D_1 = \gamma_Q + \kappa_E, D_2 = \gamma_J + \alpha_I$.

Each term of $\mathscr{R}_c$ has an epidemiological interpretation. The mean duration in E is $1/D_1$ with contact rate $\varepsilon_E \beta$, giving a contribution to $\mathscr{R}_c$ of $\varepsilon_E \beta N / D_1$. A fraction κ_E / D_1 goes from E to I, with contact rate β and mean duration $1/D_2$, giving a contribution of $\beta N \kappa_E / D_1 D_2$. A fraction γ_Q / D_1 goes from E to Q, with contact rate $\varepsilon_E \varepsilon_Q \beta$ and mean duration $1/\kappa_Q$, giving a contribution of $\varepsilon_E \varepsilon_Q \beta N \gamma_Q / D_1 \kappa_Q$. A fraction $\kappa_E \gamma_J / D_1 D_2$ goes from E to I to J, with a contact rate of $\varepsilon_J \beta$ and a mean duration of $1/\alpha_J$, giving a contribution of $\varepsilon_J \beta N \kappa_E \gamma_J / \alpha_J D_1 D_2$. Finally, a fraction γ_Q / D_1 goes from E to Q to J with a contact rate of $\varepsilon_J \beta$ and a mean duration of $1/\alpha_J$, giving a contribution of $\varepsilon_J \beta N \gamma_Q / D_1 \alpha_J$. The sum of these individual contributions gives $\mathscr{R}_c$.

In the model (9.19) the parameters γ_Q and γ_J are *control* parameters, which may be chosen in the attempt to manage the epidemic. The parameters ε_Q and ε_J depend on the strictness of the quarantine and isolation processes and are thus also control measures in a sense. The other parameters of the model are specific to the disease being studied. While they are not variable, their measurements are subject to experimental error.

The linearization of (9.19) at the disease-free equilibrium $(N, 0, 0, 0, 0)$ has matrix

$$\begin{bmatrix} \varepsilon_E \beta N - (\kappa_E + \gamma_Q) & \varepsilon_E \varepsilon_Q \beta & \beta N & \varepsilon_J \beta N \\ \gamma_Q & -\kappa_Q & 0 & 0 \\ \kappa_E & 0 & -(\alpha_I + \gamma_J) & 0 \\ 0 & \kappa_Q & \gamma_J & -\alpha_J \end{bmatrix} .$$

The corresponding characteristic equation is a fourth-degree polynomial equation whose leading coefficient is 1 and whose constant term is a positive constant multiple of $1 - \mathscr{R}_c$, thus positive if $\mathscr{R}_c < 1$ and negative if $\mathscr{R}_c > 1$. If $\mathscr{R}_c > 1$, there is a positive eigenvalue, corresponding to an initial exponential growth rate of solutions of (9.19). If $\mathscr{R}_c < 1$, it is possible to show that all eigenvalues of the coefficient matrix have negative real part, and thus solutions of (9.19) die out exponentially [van den Driessche and Watmough (2002)].

In order to show that analogues of the relation (9.3) and $S_\infty > 0$ derived for the model (9.2) are valid for the management model (9.19), we begin by integrating the equations for $S+E, Q, I, J$, of (9.19) with respect to t from $t = 0$ to $t = \infty$, using the initial conditions

$$S(0)+E(0) = N(0) = N, \qquad Q(0) = I(0) = J(0) = 0\,.$$

We continue by integrating the equation for S, and then an argument similar to the one used for (9.2) but technically more complicated may be used to show that $S_\infty > 0$ for the treatment model (9.19) and also to establish the final size relation

$$\log \frac{S_0}{S_\infty} = \mathscr{R}_c \left[1 - \frac{S_\infty}{N}\right].$$

Thus the asymptotic behavior of the management model (9.19) is the same as that of the simpler model (9.2).

In the various compartmental models that we have studied, there are significant common features. This suggests that compartmental models can be put into a more general framework. In fact, this general framework is the age of infection epidemic model originally introduced by Kermack and McKendrick in 1927. However, we will not explore this generalization here.

Exercises

1. Compare the qualitative behaviors of the models

   $$S' = -\beta SI, \quad I' = \beta SI - \alpha I,$$

   and

   $$S' = -\beta SI, \quad E' = \beta SI - \kappa E, \quad I' = \kappa E - \alpha I,$$

   with

   $$\beta = 1/3000, \quad \alpha = 1/6, \quad \kappa = 1/2, \quad S(0) = 999, \quad I(0) = 1.$$

   These models represent an SIR epidemic model and an SEIR epidemic model respectively with a mean infective period of 6 days and a mean exposed period of 2 days. Do numerical simulations to decide whether the exposed period affects the behavior of the model noticeably.
2. Consider three basic epidemic models–the simple SIR model,

   $$S' = -\beta SI,$$
   $$I' = \beta SI - \alpha I,$$

   the SEIR model with some infectivity in the exposed period,

$$\begin{aligned} S' &= -\beta S(I+\varepsilon E), \\ E' &= \beta S(I+\varepsilon E) - \kappa E, \\ I' &= \kappa E - \alpha I, \end{aligned}$$

and the SIR model with treatment,

$$\begin{aligned} S' &= -\beta S(I+\delta T), \\ I' &= \beta S(I+\delta T) - (\alpha+\varphi)I, \\ T' &= \varphi I - \eta T. \end{aligned}$$

Use the parameter values

$$\beta = \frac{1}{3000}, \quad \alpha = \frac{1}{4}, \quad \varepsilon = \frac{1}{2}, \quad \kappa = \frac{1}{2}, \quad \delta = \frac{1}{2}, \quad \eta = \frac{1}{4}, \quad \varphi = 1,$$

and the initial values

$$S(0) = 995, \quad E(0) = 0, \quad I(0) = 5, \quad T(0) = 0.$$

For each model,

(i) Calculate the reproduction number and the epidemic size.
(ii) Do some numerical simulations to obtain the epidemic size by determining the change in S, the maximum number of infectives by measuring I, and the duration of the epidemic.

If you feel really ambitious, formulate and analyze an SEIR model with infectivity in the exposed period and treatment.

3. Consider an SIR model in which a fraction θ of infectives is isolated in a perfectly quarantined class Q with standard incidence (meaning that individuals make a contacts in unit time of which a fraction $I/(N-Q)$ are infective), given by the system

$$\begin{aligned} S' &= -aS\frac{I}{N-Q}, \\ I' &= aS\frac{I}{N-Q} - (\theta+\alpha)I, \\ Q' &= \theta I - \gamma Q, \\ R' &= \alpha I + \gamma Q. \end{aligned}$$

(i) Find the equilibria.
(ii) Find the basic reproduction number $\mathscr{R}_0$.
(iii) For influenza-like parameters, take $\alpha = 0.5, \theta = 1, 2, 4, \gamma = 0.4$, and $\mathscr{R}_0 = 2.5$, sketch the phase plane of the system and observe what is happening.

4. Isolation/quarantine is a complicated process because we don't live in a perfect world. In hospitals, patients may inadvertently or deliberately break from isolation and in the process have casual contacts with others including medical personnel and visitors. Taking this into account, we are led to the model

$$\begin{aligned} S' &= -aS\frac{[I+\rho\tau Q]}{N-\sigma Q}, \\ I' &= aS\frac{[I+\rho\tau Q]}{N-\sigma Q} - (\theta+\alpha)I, \\ Q' &= \theta I - \gamma Q, \\ R' &= \alpha I + \gamma Q. \end{aligned}$$

   (i) Determine all the parameters in the system and define each parameter.
   (ii) Show that the population is constant.
   (iii) Find all equilibria.
   (iv) Find the reproductive number $\mathscr{R}_0$.
   (v) Describe the asymptotic behavior of the model, including its dependence on the basic reproduction number.

5. Formulate a model analogous to (9.16) for which treatment is not started immediately, but begins at time $\tau > 0$. Can you say anything about the dependence of the reproduction number on τ?

9.6 An SIR Model with a General Infectious Period Distribution

In the simple model (9.2) studied in Section 9.2 we have assumed that the infective period is exponentially distributed. Now let us consider an SIR epidemic model in a population of constant size N with mass action incidence in which $P(\tau)$ is the fraction of individuals who are still infective a time τ after having become infected. The model is

$$\begin{aligned} S' &= -\beta S(t)I(t), \\ I(t) &= I_0(t) + \int_0^t [-S'(t-\tau)]P(\tau)\,d\tau. \end{aligned} \tag{9.20}$$

Here, $I_0(t)$ is the number of individuals who were infective initially at $t = 0$ who are still infective at time t. Then

$$I_0(t) \le (N - S_0)P(t),$$

because if all initial infectives were newly infected we would have equality in this relation, and if some initial infectives had been infected before the starting time $t = 0$, they would recover earlier.

We assume that $P(\tau)$ is a nonnegative, nonincreasing function with $P(0) = 1$. We assume also that the mean infective period $\int_0^\infty P(\tau)d\tau$ is finite. Since a single infective causes βN new infections in unit time and $\int_0^\infty P(\tau)d\tau$ is the mean infective period, it is easy to calculate

$$\mathscr{R}_0 = \beta N \int_0^\infty P(s)\,ds.$$

Since S is a nonnegative decreasing function, it follows as for (9.2) that $S(t)$ deceases to a limit S_∞ as $t \to \infty$, but we must proceed differently to show that $I(t) \to 0$. This will follow if we can prove that $\int_0^t I(s)\,ds$ is bounded as $t \to \infty$. We have

$$\begin{aligned}
\int_0^t I(s)\,ds &= \int_0^t I_0(\tau)\,ds + \int_0^t \int_0^s [-S'(s-\tau)]P(\tau)d\tau ds \\
&\le (N - S_0)\int_0^t P(\tau)d\tau + \int_0^t \int_\tau^t [-S'(s-\tau)]ds P(\tau)d\tau \\
&\le (N - S_0)\int_0^t P(\tau)ds + \int_0^t S_0 - S(t-\tau)]P(\tau)d\tau \\
&\le N\int_0^t P(\tau)d\tau.
\end{aligned}$$

Since $\int_0^\infty P(\tau)d\tau$ is assumed to be finite, it follows that $\int_0^t I(s)\,ds$ is bounded, and thence that $I(t) \to 0$.

Now integration of the first equation in (9.20) from 0 to ∞ gives

$$\log\frac{S_0}{S_\infty} = \beta\int_0^\infty I(\tau)d\tau < \infty,$$

and this shows that $S_\infty > 0$.

If all initially infected individuals are newly infected, so that $I_0(t) = (N - S_0)P(t)$, integration of the second equation of (9.20) gives

$$\begin{aligned}
\int_0^\infty I(s)\,ds &= \int_0^\infty I_0(s)\,ds + \int_0^\infty \int_0^s [-S'(s-\tau)]P(\tau)d\tau ds \\
&= (N - S_0)\int_0^\infty P(\tau)d\tau + \int_0^\infty \int_\tau^\infty [-S'(t-\tau)]ds P(\tau)d\tau \\
&= (N - S_0)\int_0^\infty P(\tau)d\tau + \int_0^\infty [S_0 - S_\infty]P(\tau)d\tau \\
&= (N - S_\infty)\int_0^\infty P(\tau)d\tau \\
&= \mathscr{R}_0\left[1 - \frac{S_\infty}{N}\right],
\end{aligned}$$

and this is the final size relation, identical to (9.3). If there are individuals who were infected before time $t = 0$, a positive term

$$(N - S_0)\int_0^\infty P(t)dt - \int_0^\infty I_0(t)dt$$

must be subtracted from the right side of this equation.

The generalization to arbitrary infectious periods in this section is a component of the age of infection epidemic model of Kermack and McKendrick (1927), which also incorporates general compartmental structures. The examples of this section and the previous section are all special cases of the age of infection model.

Exercises

1. Formulate a description of the model (9.20) with an infective period of fixed length σ and calculate its basic reproduction number.

9.7 The Age of Infection Epidemic Model

The general epidemic model described by Kermack and McKendrick (1927) included a dependence of infectivity on the time since becoming infected (age of infection). We let $S(t)$ denote the number of susceptibles at time t and let $\varphi(t)$ be the total infectivity at time t, defined as the sum of products of the number of infected members with each infection age and the mean infectivity for that infection age. We assume that on average, members of the population make a constant number a of contacts in unit time. We let $B(\tau)$ be the fraction of infected members remaining infected at infection age τ and let $\pi(\tau)$ with $0 \le \pi(\tau) \le 1$ be the mean infectivity at infection age τ. Then we let

$$A(\tau) = \pi(\tau)B(\tau),$$

the mean infectivity of members of the population with infection age τ. We assume that there are no disease deaths, so that the total population size is a constant N.

The age of infection epidemic model is

$$\begin{aligned} S' &= -\beta S\varphi, \\ \varphi(t) &= \varphi_0(t) + \int_0^t \beta S(t-\tau)\varphi(t-\tau)A(\tau)d\tau \\ &= \varphi_0(t) + \int_0^t [-S'(t-\tau)]A(\tau)d\tau. \end{aligned} \tag{9.21}$$

The basic reproduction number is

$$\mathscr{R}_0 = \beta N \int_0^\infty A(\tau)d\tau.$$

We write

$$-\frac{S'(t)}{S(t)} = \beta\varphi_0(t) + \beta\int_0^t [-S'(t-\tau)]A(\tau)d\tau.$$

Integration with respect to t from 0 to ∞ gives

$$\begin{aligned}
\log\frac{S_0}{S_\infty} &= \beta\int_0^\infty \varphi_0(t)dt + \beta\int_0^\infty\int_0^t [-S'(t-\tau)]A(\tau)d\tau dt \\
&= \beta\int_0^\infty \varphi_0(t)dt + \beta\int_0^\infty A(\tau)\int_\tau^\infty [-S'(t-\tau)]dtd\tau \\
&= \beta\int_0^\infty \varphi_0(t)dt + [S_0 - S_\infty]\int_0^\infty A(\tau)d\tau \qquad (9.22)\\
&= \beta[N-S_\infty]\int_0^\infty A(\tau)d\tau + \beta\int_0^\infty [\varphi_0(t) - (N-S_0)A(\tau)d\tau \\
&= \mathscr{R}_0\left[1-\frac{S_\infty}{N}\right] - \beta\int_0^\infty [(N-S_0)A(t) - \varphi_0(t)]dt.
\end{aligned}$$

Here, $\varphi_0(t)$ is the total infectivity of the initial infectives when they reach age of infection t. If all initial infectives have infection age zero at $t=0$, then $\varphi_0(t) = [N-S_0]A(t)$, and

$$\int_0^\infty [\varphi_0(t) - (N-S_0)A(t)]dt = 0.$$

Then (9.22) takes the form

$$\log\frac{S_0}{S_\infty} = \mathscr{R}_0\left(1-\frac{S_\infty}{N}\right), \qquad (9.23)$$

and this is the general final size relation. If there are initial infectives with infection age greater than zero, let $u(\tau)$ be the fraction of these individuals with infection age τ, $\int_0^\infty u(\tau)d\tau = 1$. At time t these individuals have infection age $t+\tau$ and mean infectivity $A(t+\tau)$. Thus

$$\varphi_0(t) = (N-S_\infty)\int_0^\infty u(\tau)A(t+\tau)d\tau,$$

and

$$\begin{aligned}
\int_0^\infty \varphi_0(t)dt &= (N-S_\infty)\int_0^\infty\int_0^\infty u(\tau)A(t+\tau)d\tau dt \\
&= (N-S_\infty)\int_0^\infty u(\tau)\Big[\int_\tau^\infty A(v)dv\Big]d\tau \\
&= (N-S_\infty)\int_0^\infty A(v)\Big[\int_0^v u(\tau)d\tau\Big]dv] \\
&\le (N-S_\infty)\int_0^\infty A(v)dv,
\end{aligned}$$

since $\int_0^v u(\tau)d\tau \le 1$.

Thus, the initial term satisfies

$$\int_0^\infty [(N - S_0)A(t) - \varphi_0(t)]dt \geq 0.$$

The final size relation is sometimes presented in the form

$$\log \frac{S_0}{S_\infty} = \mathscr{R}_0 \left(1 - \frac{S_\infty}{S_0}\right); \tag{9.24}$$

see, for example [Arino et al. (2007), Heffernan, Smith?, and Wahl (2005)]. Without the initial term this form would represent the final size relation for an epidemic started by someone outside the population under study, so that $S_0 = N, I_0 = 0$.

Example 1. The SEIR model (9.15) can be viewed as an age of infection model with $\varphi = \varepsilon E + I$. To use the age of infection interpretation, we need to determine the kernel $A(\tau)$ in order to calculate its integral. We let $u(\tau)$ be the fraction of infected members with infection age τ who are not yet infective and $v(\tau)$ the fraction of infected members who are infective. Then the rate at which members become infective at infection age τ is $\kappa u(\tau)$, and we have

$$\begin{aligned} u'(\tau) &= -\kappa u(\tau), \qquad u(0) = 1, \\ v'(\tau) &= \kappa u(\tau) - \alpha v(\tau), \quad v(0) = 0. \end{aligned}$$

The solution of this system is

$$u(\tau) = e^{-\kappa\tau}, \quad v(\tau) = \frac{\kappa}{\kappa - \alpha}[e^{-\alpha\tau} - e^{-\kappa\tau}].$$

Thus we have

$$A(\tau) = \varepsilon e^{-\kappa\tau} + \frac{\kappa}{\kappa - \alpha}[e^{-\alpha\tau} - e^{-\kappa\tau}],$$

and it is easy to calculate

$$\int_0^\infty A(\tau)d\tau = \frac{1}{\alpha} + \frac{\varepsilon}{\kappa}.$$

This gives the same value for $\mathscr{R}_0$ as was calculated directly.

The age of infection model also includes the possibility of disease stages with distributions that are not exponential [Feng (2007), Feng, Xu, and Zhao (2007)].

Example 2. Consider an *SEIR* model in which the exposed stage has an exponential distribution but the infective stage has a period distribution given by a function P,

$$\begin{aligned} S' &= -\beta SI, \\ E' &= \beta SI - \kappa E, \\ I(t) &= \int_0^t \kappa E(s)P(t-s)ds, \end{aligned} \tag{9.25}$$

with initial conditions

$$S(0) = S_0, \quad E(0) = E_0, \quad I(0) = 0.$$

If we define $u(\tau), v(\tau)$ as in Example 1, we again obtain $u(\tau) = e^{-\kappa\tau}$, and v satisfies

$$A(\tau) = v(\tau) = \int_0^\tau \kappa u(s)P(\tau - s)ds = \int_0^\tau \kappa e^{-\kappa s}P(\tau - s)ds.$$

Now,

$$\begin{aligned}\int_0^\infty A(\tau)d\tau &= \int_0^\infty \int_0^\tau \kappa e^{-\kappa s}P(\tau - s)dsd\tau \\ &= \int_0^\infty \left[\int_s^\infty P(\tau - s)d\tau\right] \kappa e^{-\kappa s}ds \\ &= \int_0^\infty \left[\int_0^\infty P(u)du\right] \kappa e^{-\kappa s}ds \\ &= \int_0^\infty P(u)du.\end{aligned}$$

For period distributions that are not exponential, it is possible to calculate

$$\int_0^\infty A(\tau)d\tau$$

without having to calculate the function $A(\tau)$ explicitly.

Example 3. Consider an SEIR model in which the exposed period has a distribution given by a function Q and the infective period has a distribution given by a function P. Then

$$\begin{aligned}S' &= -\beta SI, \\ E(t) &= E_0Q(t) + \int_0^t [-S'(s)]Q(t - s)ds.\end{aligned}$$

In order to obtain an equation for I, we differentiate the equation for E, obtaining

$$E'(t) = E_0Q'(t) - S'(t) + E_0 \int_0^t [-S'(s)]Q'(t - s)ds.$$

Thus the input to I at time t is

$$E_0Q'(t) + E_0 \int_0^t [-S'(s)]Q'(t - s)ds,$$

and

$$I(t) = E_0 \int_0^t Q'(u)P(t - u)du + E_0 \int_0^t [-S'(s)]Q'(u - s)dsP(t - u)du.$$

The first term in this expression may be written as $I_0(t)$, and the second term may be simplified, using interchange of the order of integration in the iterated integral, to yield

$$\int_0^t \int_0^u [-S'(s)]Q'(u-s)dsP(t-u)du = \int_0^t \int_s^t Q'(u-s)duP(t-u)[-S'(s)]ds.$$

If we define

$$A(t-s) = \int_s^t Q'(u-s)P(t-u)du = \int_0^{t-s} Q'(t-s-v)P(v)dv,$$

we obtain

$$I(t) = I_0(t) + \int_0^t [-S'(s)]A(t-s)ds.$$

Then the model is

$$\begin{aligned} S' &= -\beta SI, \\ E(t) &= E_0Q(t) + \int_0^t [-S'(s)]Q(t-s)ds, \\ I(t) &= I_0(t) + \int_0^t [-S'(s)]A(t-s)ds, \end{aligned} \tag{9.26}$$

which is in age of infection form with $\varphi = I$, and we have an explicit expression for $A(\tau)$.

Exercises

1. Interpret the models (9.16), (9.18), and (9.19) introduced earlier as age of infection models and use this interpretation to calculate their reproduction numbers.
2. Calculate the basic reproduction number for the model (9.26) but with infectivity in the exposed class having a reduction factor ε.

9.8 Models with Disease Deaths

The assumption in the model (9.2) of a rate of contacts per infective that is proportional to population size N, called *mass action incidence* or bilinear incidence, was used in all the early epidemic models. However, it is quite unrealistic, except possibly in the early stages of an epidemic in a population of moderate size. It is more realistic to assume a contact rate that is a nonincreasing function of total population size. For example, a situation in which the number of contacts per infective in unit time is constant, called *standard incidence*, is a more accurate description for sexu-

ally transmitted diseases. If there are no disease deaths, so that the total population size remains constant, such a distinction is unnecessary.

We generalize the model (9.2) by dropping assumption (iv) and replacing assumption (i) by the assumption that an average member of the population makes $C(N)$ contacts in unit time with $C'(N) \geq 0$ [Castillo-Chavez, Cooke, Huang, and Levin (1989a), Dietz (1982)], and we define

$$\beta(N) = \frac{C(N)}{N} .$$

It is reasonable to assume $\beta'(N) \leq 0$ to express the idea of saturation in the number of contacts. Then mass action incidence corresponds to the choice $C(N) = \beta N$, and standard incidence corresponds to the choice $C(N) = \lambda$. The assumptions $C(N) = N\beta(N)$, $C'(N) \geq 0$ imply that

$$\beta(N) + N\beta'(N) \geq 0 . \tag{9.27}$$

Some epidemic models [Dietz (1982)] have used a Michaelis–Menten type of interaction of the form

$$C(N) = \frac{aN}{1+bN} .$$

Another form based on a mechanistic derivation for pair formation [Heesterbeek and Metz (1993] leads to an expression of the form

$$C(N) = \frac{aN}{1+bN+\sqrt{1+2bN}} .$$

Data for diseases transmitted by contact in cities of moderate size [Mena-Lorca and Hethcote (1992)] suggests that data fit the assumption of the form

$$C(N) = \lambda N^a$$

with $a = 0.05$ quite well. All of these forms satisfy the conditions $C'(N) \geq 0$, $\beta'(N) \leq 0$.

Because the total population size is now present in the model, we must include an equation for total population size in the model. This forces us to make a distinction between members of the population who die of the disease and members of the population who recover with immunity against reinfection. We assume that a fraction f of the αI members leaving the infective class at time t recover and the remaining fraction $(1-f)$ die of disease. We use S, I, and N as variables, with $N = S+I+R$. We now obtain a three-dimensional model

$$\begin{aligned} S' &= -\beta(N)SI, \\ I' &= \beta(N)SI - \alpha I, \\ N' &= -(1-f)\alpha I . \end{aligned} \tag{9.28}$$

Since N is now a decreasing function, we define $N(0) = N_0 = S_0 + I_0$. We also have the equation $R' = -f\alpha I$, but we need not include it in the model, since R is determined when S, I, and N are known. We should note that if $f = 1$, the total population size remains equal to the constant N, and the model (9.28) reduces to the simpler model (9.2) with β replaced by the constant $\beta(N_0)$.

We wish to show that the model (9.28) has the same qualitative behavior as the model (9.2), namely that there is a basic reproduction number that distinguishes between disappearance of the disease and an epidemic outbreak, and that some members of the population are left untouched when the epidemic passes. These two properties are the central features of all epidemic models.

For the model (9.28) the basic reproduction number is given by

$$\mathscr{R}_0 = \frac{N_0\beta(N_0)}{\alpha}$$

because a single infective introduced into a wholly susceptible population makes $C(N_0) = N_0\beta(N_0)$ contacts in unit time, all of which are with susceptibles and thus produce new infections, and the mean infective period is $1/\alpha$.

We assume that $\beta(0)$ is finite, thus ruling out standard incidence (standard incidence does not appear to be realistic if the total population N approaches zero, and it would be more natural to assume that $C(N)$ grows linearly with N for small N). If we let $t \to \infty$ in the sum of the first two equations of (9.28), we obtain

$$\alpha \int_0^\infty I(s)ds = S_0 + I_0 - S_\infty = N - S_\infty.$$

The first equation of (9.28) may be written as

$$-\frac{S'(t)}{S(t)} = \beta(N(t))I(t).$$

Since

$$\beta(N) \geq \beta(N_0),$$

integration from 0 to ∞ gives

$$\log\frac{S_0}{S_\infty} = \int_0^\infty \beta(N(t))I(t)dt \geq \beta(N_0)\int_0^\infty I(t)dt = \frac{\beta(N_0)(N_0 - S_\infty)}{\alpha N_0}.$$

We now obtain a final size inequality

$$\log\frac{S_0}{S_\infty} = \int_0^\infty \beta(N(t))I(t)dt \geq \beta(N_0)\int_0^\infty I(t)dt = \mathscr{R}_0\left[1 - \frac{S_\infty}{N_0}\right].$$

If the disease death rate is small, the final size inequality is an approximate equality.

It is not difficult to show that $N(t) \geq fN_0$, and then a similar calculation using the inequality $\beta(N) \leq \beta(fN_0) < \infty$ shows that

$$\log \frac{S_0}{S_\infty} \leq \beta(fN_0) \int_0^\infty I(t)dt,$$

from which we may deduce that $S_\infty > 0$.

Exercises

1. For the model (9.28) show that the final total population size is given by

$$N_\infty = fN_0 + (1-f)S_\infty.$$

9.9 A Vaccination Model

To cope with annual seasonal influenza epidemics there is a program of vaccination before the "flu" season begins. Each year, a vaccine is produced aimed at protecting against the three influenza strains considered most dangerous for the coming season. We formulate a model to add vaccination to the simple SIR model (9.2) under the assumption that vaccination reduces susceptibility (the probability of infection if a contact with an infected member of the population is made).

We consider a population of total size N and assume that a fraction γ of this population is vaccinated prior to a disease outbreak. Thus we have a subpopulation of size $N_U = (1-\gamma)N$ of unvaccinated members and a subpopulation of size $N_V = \gamma N$ of vaccinated members. We assume that vaccinated members have susceptibility to infection reduced by a factor σ, $0 \leq \sigma \leq 1$, with $\sigma = 0$ describing a perfectly effective vaccine and $\sigma = 1$ describing a vaccine that has no effect. We assume also that vaccinated individuals who are infected have infectivity reduced by a factor δ and may also have a recovery rate α_V that is different from the recovery rate of infected unvaccinated individuals α_U.

We let S_U, S_V, I_U, I_V denote the number of unvaccinated susceptibles, the number of vaccinated susceptibles, the number of unvaccinated infectives, and the number of vaccianted infectives respectively.

The resulting model is

$$\begin{aligned} S_U' &= -\beta S_U(I_U + \delta I_V), \\ S_V' &= -\sigma\beta S_V(I_U + \delta I_V), \\ I_U' &= \beta S_U(I_U + \delta I_V) - \alpha_U I_U, \\ I_V' &= \sigma\beta S_V(I_U + \delta I_V) - \alpha_V I_V. \end{aligned} \tag{9.29}$$

The initial conditions prescribe $S_U(0), S_V(0), I_U(0), I_V(0)$, with

$$S_U(0) + I_U(0) = N_U, \quad S_V(0) + I_V(0) = N_V.$$

Since the infection now is beginning in a population that is not fully susceptible, we speak of the *c*ontrol reproduction number $\mathscr{R}_c$ rather than the basic reproduction number. However, as we will soon see, calculation of the control reproduction number will require a more general definition and a considerable amount of technical computation. The computation method is applicable to both basic and control reproduction numbers. We will use the term *reproduction number* to denote either a basic reproduction number or a control reproduction number. We are able to obtain final size relations without knowledge of the reproduction number, but these final size relations do contain information about the reproduction number, and more.

Since S_U and S_V are decreasing nonnegative functions they have limits $S_U(\infty)$ and $S_V(\infty)$ respectively as $t \to \infty$. The sum of the equations for S_U and I_U in (9.29) is

$$(S_U + I_U)' = -\alpha_U I_U,$$

from which we conclude, just as in the analysis of (9.2), that $I_U(t) \to 0$ as $t \to \infty$, and that

$$\alpha \int_0^\infty I_U(t)dt = N_U - S_U(\infty). \tag{9.30}$$

Similarly, using the sum of the equations for S_V and I_V, we see that $I_V(t) \to 0$ as $t \to \infty$, and that

$$\alpha \int_0^\infty I_V(t)dt = N_V - S_V(\infty). \tag{9.31}$$

Integration of the equation for S_U in (9.29) and use of (**??**) gives

$$\begin{aligned} \log \frac{S_U(0)}{S_U(\infty)} &= \beta[\int_0^\infty I_U(t)dt + \delta \int_0^\infty I_V(t)dt] \\ &= \frac{\beta N_U}{\alpha}\left[1 - \frac{S_U(\infty)}{N_U}\right] + \frac{\delta\beta N_V}{\alpha}\left[1 - \frac{S_V(\infty)}{N_V}\right]. \end{aligned} \tag{9.32}$$

A similar calculation using the equation for S_V gives

$$\log \frac{S_V(0)}{S_V(\infty)} = \frac{\sigma\beta N_U}{\alpha}\left[1 - \frac{S_U(\infty)}{N_U}\right] + \frac{\delta\sigma\beta N_V}{\alpha}\left[1 - \frac{S_V(\infty)}{N_V}\right]. \tag{9.33}$$

This pair of equations (9.32), (9.33) are the final size relations. They make it possible to calculate $S_U(\infty), S_V(\infty)$ if the parameters of the model are known.

It is convenient to define the matrix

$$K = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = \begin{bmatrix} \frac{\beta N_U}{\alpha_U} & \frac{\delta\beta N_V}{\alpha_V} \\ \frac{\sigma\beta N_U}{\alpha_U} & \frac{\delta\sigma\beta N_V}{\alpha_V} \end{bmatrix}.$$

Then the final size relations (9.32), (9.33) may be written

$$\log \frac{S_U(0)}{S_U(\infty)} = K_{11}\left[1 - \frac{S_U(\infty)}{N_U}\right] + K_{12}\left[1 - \frac{S_V(\infty)}{N_V}\right], \tag{9.34}$$
$$\log \frac{S_V(0)}{S_V(\infty)} = K_{21}\left[1 - \frac{S_U(\infty)}{N_U}\right] + K_{22}\left[1 - \frac{S_V(\infty)}{N_V}\right].$$

The matrix K is closely related to the reproduction number. In the next section we describe a general method for calculating reproduction numbers that will involve this matrix.

Exercises

1. Suppose we want to model the spread of influenza in a city using an SLIAR model (susceptible-latent-infectious-asympotmatic-recovered, respectively). Then our system of equation would be

$$\begin{aligned} S' &= -\beta(I + \delta A)S, \\ L' &= \beta(I + \delta A)S - \kappa L, \\ I' &= p\kappa L - \gamma I, \\ A' &= (1-p)\kappa L - \eta A, \\ R' &= \eta A + \gamma I, \end{aligned}$$

   where β is the transmission coefficient, δ is the reduced transmissibility factor from asymptomatic contacts, κ is the rate of disease progression from the latent class, p is the proportion of individuals that are clinically diagnosed, η is the recovery rate from the asymptomatic class, γ is the recovery rate from the infectious (clinically diagnosed) class, and N is the total population size.

   (i) Add a vaccination class to the model. Assume that the vaccine imparts partial protection until it becomes fully effective. Is the population of the new system constant? Are there any endemic equilibria?
   (ii) Vary the vaccination rate from 0.2 to 0.8 and determine how the number of infected individuals changes compared with the model without vaccination. Does vaccination prevent the outbreak?

9.10 The Next Generation Matrix

Up to this point, we have calculated reproduction numbers by following the secondary cases caused by a single infective introduced into a population. However, if there are subpopulations with different susceptibilities to infection, as in the vaccination model introduced in Section 9.9, it is necessary to follow the secondary infections in the subpopulations separately, and this approach will not yield the re-

production number. It is necessary to give a more general approach to the meaning of the reproduction number, and this is done through the *next generation matrix* [Diekmann and Heesterbeek (2000), Diekmann, Heesterbeek, and Metz (1990), van den Driessche and Watmough (2002)]. The underlying idea is that we must calculate the matrix whose (i, j) entry is the number of secondary infections caused in compartment i by an infected individual in compartment j. The procedure applies both to epidemic models, as studied in this chapter, and to models with demographics for endemic diseases, to be studied in the next chapter.

In a compartmental disease transmission model we sort individuals into compartments based on a single, discrete state variable. A compartment is called a *disease compartment* if the individuals therein are infected. Note that this use of the term disease is broader than the clinical definition and includes stages of infection such as exposed stages in which infected individuals are not necessarily infective. Suppose there are n disease compartments and m nondisease compartments, and let $x \in \mathrm{R}^{\mathrm{n}}$ and $y \in \mathrm{R}^{\mathrm{m}}$ be the subpopulations in each of these compartments. Further, we denote by $\mathscr{F}_i$ the rate at which secondary infections increase the $i-th$ disease compartment and by $\mathscr{V}_i$ the rate at which disease progression, death, and recovery decrease the $i-th$ compartment. The compartmental model can then be written in the form

$$\begin{aligned} x_i' &= \mathscr{F}_i(x,y) - \mathscr{V}_i(x,y)\,, \quad i = 1,\ldots,n, \\ y_j' &= g_j(x,y)\,, \quad j = 1,\ldots,m. \end{aligned} \tag{9.35}$$

Note that the decomposition of the dynamics into $\mathscr{F}$ and $\mathscr{V}$ and the designation of compartments as infected or uninfected may not be unique; different decompositions correspond to different epidemiological interpretations of the model. The definitions of $\mathscr{F}$ and $\mathscr{V}$ used here differ slightly from those in [van den Driessche and Watmough (2002)].

The derivation of the basic reproduction number is based on the linearization of the ODE model about a disease-free equilibrium. For an epidemic model with a line of equilibria, it is customary to use the equilibrium with all members of the population susceptible. We assume:

- $\mathscr{F}_i(0,y) = 0$ and $\mathscr{V}_i(0,y) = 0$ for all $y \geq 0$ and $i = 1,\ldots,n$.
- The disease-free system $y' = g(0,y)$ has a unique equilibrium that is asymptotically stable, that is, all solutions with initial conditions of the form $(0,y)$ approach a point $(0,y_o)$ as $t \to \infty$. We refer to this point as the disease-free equilibrium.

The first assumption says that all new infections are secondary infections arising from infected hosts; there is no immigration of individuals into the disease compartments. It ensures that the disease-free set, which consists of all points of the form $(0,y)$, is invariant. That is, any solution with no infected individuals at some point in time will be free of infection for all time. The second assumption ensures that the disease-free equilibrium is also an equilibrium of the full system. The uniqueness of the disease-free equilibrium in the second assumption is required for models

with demographics, to be studied in the next chapter. Although it is not satisfied in epidemic models, the specification of a specific disease-free equilibrium with all memebers of the population susceptible is sufficient to validate the results.

Next, we assume:

- $\mathscr{F}_i(x,y) \geq 0$ for all nonnegative x and y and $i = 1,\ldots,n$.
- $\mathscr{V}_i(x,y) \leq 0$ whenever $x_i = 0$, $i = 1,\ldots,n$.
- $\sum_{i=1}^{n} \mathscr{V}_i(x,y) \geq 0$ for all nonnegative x and y.

The reasons for these assumptions are that the function $\mathscr{F}$ represents new infections and cannot be negative, each component $\mathscr{V}_i$ represents a net outflow from compartment i and must be negative (inflow only) whenever the compartment is empty, and the sum $\sum_{i=1}^{n} \mathscr{V}_i(x,y)$ represents the total outflow from all infected compartments. Terms in the model leading to increases in $\sum_{i=1}^{n} x_i$ are assumed to represent secondary infections and therefore belong in $\mathscr{F}$.

Suppose that a single infected person is introduced into a population originally free of disease. The initial ability of the disease to spread through the population is determined by an examination of the linearization of (9.35) about the disease-free equilibrium $(0,y_0)$. It is easy to see that the assumption $\mathscr{F}_i(0,y) = 0, \mathscr{V}_i(0,y) = 0$ implies

$$\frac{\partial \mathscr{F}_i}{\partial y_j}(0,y_o) = \frac{\partial \mathscr{V}_i}{\partial y_j}(0,y_o) = 0$$

for every pair (i,j). This implies that the linearized equations for the disease compartments x are decoupled from the remaining equations and can be written as

$$x' = (F - V)x, \tag{9.36}$$

where F and V are the $n \times n$ matrices with entries

$$F = \frac{\partial \mathscr{F}_i}{\partial x_j}(0,y_o) \qquad \text{and} \qquad V = \frac{\partial \mathscr{V}_i}{\partial x_j}(0,y_o).$$

Because of the assumption that the disease-free system $y' = g(0,y)$ has a unique asymptotically stable equilibrium, the linear stability of the system (9.35) is completely determined by the linear stability of the matrix $(F - V)$ in (9.36).

The number of secondary infections produced by a single infected individual can be expressed as the product of the expected duration of the infectious period and the rate at which secondary infections occur. For the general model with n disease compartments, these are computed for each compartment for a hypothetical index case. The expected time the index case spends in each compartment is given by the integral $\int_0^\infty \phi(t,x_0)\, dt$, where $\phi(t,x_0)$ is the solution of (9.36) with $F = 0$ (no secondary infections) and nonnegative initial conditions x_0 representing an infected index case:

$$x' = -Vx, \qquad x(0) = x_0. \tag{9.37}$$

In effect, this solution shows the path of the index case through the disease compartments from the initial exposure through to death or recovery with the $i - th$

component of $\varphi(t,x_0)$ interpreted as the probability that the index case (introduced at time $t=0$) is in disease state i at time t. The solution of (9.37) is $\phi(t,x_0)=e^{-Vt}x_0$, where the exponential of a matrix is defined by the Taylor series

$$e^A = I + A + \frac{A^2}{2} + \frac{A^3}{3!} + \cdots + \frac{A^k}{k!} + \cdots .$$

This series converges for all t (see, for example, [Hirsch and Smale (1974)]. Thus $\int_0^\infty \varphi(t,x_0)\,dt = V^{-1}x_0$, and the (i,j) entry of the matrix V^{-1} can be interpreted as the expected time an individual initially introduced into disease compartment j spends in disease compartment i.

The (i,j) entry of the matrix F is the rate at which secondary infections are produced in compartment i by an index case in compartment j. Hence, the expected number of secondary infections produced by the index case is given by

$$\int_0^\infty Fe^{-Vt}x_0\,dt = FV^{-1}x_0.$$

Following Diekmann and Heesterbeek (2000), the matrix $K=FV^{-1}$ is referred to as the *next generation matrix* for the system at the disease-free equilibrium. The (i,j) entry of K is the expected number of secondary infections in compartment i produced by individuals initially in compartment j, assuming, of course, that the environment experienced by the individual remains homogeneous for the duration of its infection.

Shortly, we will describe some results from matrix theory that imply that the matrix $K_L=FV^{-1}$, called the *next generation matrix with small domain*, is nonnegative and therefore has a nonnegative eigenvalue, $\mathscr{R}_0=\rho(FV^{-1})$, such that there are no other eigenvalues of K with modulus greater than $\mathscr{R}_0$ and there is a nonnegative eigenvector ω associated with $\mathscr{R}_0$ [Berman and Plemmons (1970), Theorem 1.3.2]. This eigenvector is in a sense the distribution of infected individuals that produces the greatest number $\mathscr{R}_0$ of secondary infections per generation. Thus, $\mathscr{R}_0$ and the associated eigenvector ω suitably define a "typical" infective, and the basic reproduction number can be rigorously defined as the spectral radius of the matrix K_L. The *s*pectral radius of a matrix K_L, denoted by $\rho(K_L)$, is the maximum of the moduli of the eigenvalues of K_L. If K_L is irreducible, then $\mathscr{R}_0$ is a simple eigenvalue of K_L and is strictly larger in modulus than all other eigenvalues of K_L. However, if K_L is reducible, which is often the case for diseases with multiple strains, then K_L may have several positive real eigenvectors corresponding to reproduction numbers for each competing strain of the disease.

We have interpreted the reproduction number for a disease as the number of secondary infections produced by an infected individual in a population of susceptible individuals. If the reproduction number $\mathscr{R}_0=\rho(FV^{-1})$ is consistent with the differential equation model, then it should follow that the disease-free equilibrium is asymptotically stable if $\mathscr{R}_0<1$ and unstable if $\mathscr{R}_0>1$.

This is shown through a sequence of lemmas.

The *s*pectral bound (or abscissa) of a matrix A is the maximum real part of all eigenvalues of A. If each entry of a matrix T is nonnegative, we write $T \geq 0$ and refer to T as a nonnegative matrix. A matrix of the form $A = sI - B$, with $B \geq 0$, is said to have the Z sign pattern. These are matrices whose off-diagonal entries are negative or zero. If in addition, $s \geq \rho(B)$, then A is called an M-matrix. Note that in this section, I denotes an identity matrix, not a population of infectious individuals. The following lemma is a standard result from [Berman and Plemmons (1970)].

Lemma 9.1. *If A has the Z sign pattern, then $A^{-1} \geq 0$ if and only if A is a nonsingular M-matrix.*

The assumptions we have made imply that each entry of F is nonnegative and that the off-diagonal entries of V are negative or zero. Thus V has the Z sign pattern. Also, the column sums of V are positive or zero, which, together with the Z sign pattern, implies that V is a (possibly singular) M-matrix [Berman and Plemmons (1970), condition M_{35} of Theorem 6.2.3]. In what follows, it is assumed that V is nonsingular. In this case, $V^{-1} \geq 0$, by Lemma 9.1. Hence, $K_L = FV^{-1}$ is also nonnegative.

Lemma 9.2. *If F is nonnegative and V is a nonsingular M-matrix, then $\mathscr{R}_0 = \rho(FV^{-1}) < 1$ if and only if all eigenvalues of $(F-V)$ have negative real parts.*

Proof. Suppose $F \geq 0$ and V is a nonsingular M-matrix. By Lemma 9.1, $V^{-1} \geq 0$. Thus, $(I - FV^{-1})$ has the Z sign pattern, and by Lemma 1, $(I - FV^{-1})^{-1} \geq 0$ if and only if $\rho(FV^{-1}) < 1$. From the equalities $(V-F)^{-1} = V^{-1}(I - FV^{-1})^{-1}$ and $V(V-F)^{-1} = I + F(V-F)^{-1}$, it follows that $(V-F)^{-1} \geq 0$ if and only if $(I - FV^{-1})^{-1} \geq 0$. Finally, $(V-F)$ has the Z sign pattern, so by Lemma 9.1, $(V-F)^{-1} \geq 0$ if and only if $(V-F)$ is a nonsingular M-matrix. Since the eigenvalues of a nonsingular M-matrix all have positive real parts, this completes the proof. □

Theorem 9.1. *Consider the disease transmission model given by* (9.35). *The disease-free equilibrium of* (9.35) *is locally asymptotically stable if $\mathscr{R}_0 < 1$, but unstable if $\mathscr{R}_0 > 1$.*

Proof. Let F and V be as defined as above, and let J_{21} and J_{22} be the matrices of partial derivatives of g with respect to x and y evaluated at the disease-free equilibrium. The Jacobian matrix for the linearization of the system about the disease-free equilibrium has the block structure

$$J = \begin{bmatrix} F-V & 0 \\ J_{21} & J_{22} \end{bmatrix}.$$

The disease-free equilibrium is locally asymptotically stable if the eigenvalues of the Jacobian matrix all have negative real parts. Since the eigenvalues of J are those of $(F-V)$ and J_{22}, and the latter all have negative real parts by assumption, the disease-free equilibrium is locally asymptotically stable if all eigenvalues of $(F-V)$ have negative real parts. By the assumptions on $\mathscr{F}$ and $\mathscr{V}$, F is nonnegative and V is a nonsingular M-matrix. Hence, by Lemma 2 all eigenvalues of $(F-V)$ have negative

real parts if and only if $\rho(FV^{-1}) < 1$. It follows that the disease-free equilibrium is locally asymptotically stable if $\mathscr{R}_0 = \rho(FV^{-1}) < 1$.

Instability for $\mathscr{R}_0 > 1$ can be established by a continuity argument. If $\mathscr{R}_0 \leq 1$, then for any $\varepsilon > 0$, $((1+\varepsilon)I - FV^{-1})$ is a nonsingular M-matrix and by Lemma 9.1, $((1+\varepsilon)I - FV^{-1})^{-1} \geq 0$. By Lemma 9.2, all eigenvalues of $((1+\varepsilon)V - F)$ have positive real parts. Since $\varepsilon > 0$ is arbitrary, and eigenvalues are continuous functions of the entries of the matrix, it follows that all eigenvalues of $(V - F)$ have nonnegative real parts. To reverse the argument, suppose all the eigenvalues of $(V - F)$ have nonnegative real parts. For any positive ε, $(V + \varepsilon I - F)$ is a nonsingular M-matrix, and by Lemma 9.2, $\rho(F(V+\varepsilon I)^{-1}) < 1$. Again, since $\varepsilon > 0$ is arbitrary, it follows that $\rho(FV^{-1}) \leq 1$. Thus, $(F - V)$ has at least one eigenvalue with positive real part if and only if $\rho(FV^{-1}) > 1$, and the disease-free equilibrium is unstable whenever $\mathscr{R}_0 > 1$.

These results validate the extension of the definition of the reproduction number to more general situations. In the vaccination model (9.29) of the previous section we calculated a pair of final size relations that contained the elements of a matrix K. This matrix is precisely the next generation matrix with large domain $K_L = FV^{-1}$ that has been introduced in this section.

Example 1. Consider the *SEIR* model with infectivity in the exposed stage,

$$\begin{aligned} S' &= -\beta S(I + \varepsilon E), \\ E' &= \beta S(I + \varepsilon E) - \kappa E, \\ I' &= \kappa E - \alpha I, \\ R' &= \alpha I. \end{aligned} \tag{9.38}$$

Here the disease states are E and I,

$$\mathscr{F} = \begin{bmatrix} \varepsilon E \beta N + I \beta N \\ 0 \end{bmatrix},$$

and

$$F = \begin{bmatrix} \varepsilon\beta N & \beta N \\ 0 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \kappa & 0 \\ -\kappa & \alpha \end{bmatrix}, \quad V^{-1} = \begin{bmatrix} \frac{1}{\kappa} & 0 \\ \frac{1}{\alpha} & \frac{1}{\alpha} \end{bmatrix}.$$

Then we may calculate

$$K_L = FV^{-1} = \begin{bmatrix} \frac{\varepsilon\beta N}{\kappa} + \frac{\beta N}{\alpha} & \frac{\beta N}{\alpha} \\ 0 & 0 \end{bmatrix}.$$

Since FV^{-1} has rank 1, it has only one nonzero eigenvalue, and since the trace of the matrix is equal to the sum of the eigenvalues, it is easy to see that

$$\mathscr{R}_0 = \frac{\varepsilon\beta N}{\kappa} + \frac{\beta N}{\alpha},$$

the element in the first row and first column of FV^{-1}. If all new infections are in a single compartment, as is the case here, the basic reproduction number is the trace of the matrix FV^{-1}.

In general, it is possible to reduce the size of the next generation matrix to the number of *states at infection* [Diekmann and Heesterbeek (2000)]. The states at infection are those disease states in which there can be new infections. Suppose that there are n disease states and k states at infection with $k < n$. Then we may define an auxiliary $n \times k$ matrix E in which each column corresponds to a state at infection and has 1 in the corresponding row and 0 elsewhere. Then the next generation matrix is the $k \times k$ matrix

$$K = E^T K_L E.$$

It is easy to show, using the fact that $EE^T K_L = K_L$, that the $n \times n$ matrix K_L and the $k \times k$ matrix K have the same nonzero eigenvalues and therefore the same spectral radius. Construction of the next generation matrix that has lower dimension than the next generation matrix with large domain may simplify the calculation of the basic reproduction number.

In Example 1 above, the only disease state at infection is E, the matrix A is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

and the next generation matrix K is the 1×1 matrix

$$K = \left[\frac{\varepsilon\beta N}{\kappa} + \frac{\beta N}{\alpha} \right].$$

Example 2. Consider the vaccination model (9.29). The disease states are I_U and I_V. Then

$$\mathscr{F} = \begin{bmatrix} \beta N_U (I_U + \delta I_V) \\ \sigma\beta N_V (I_U + \delta I_V \end{bmatrix},$$

and

$$F = \begin{bmatrix} \beta N_U & \delta\beta N_U \\ \sigma\beta N_V & \sigma\delta\beta N_V \end{bmatrix} \quad V = \begin{bmatrix} \alpha_U & 0 \\ 0 & \alpha_V \end{bmatrix}.$$

It is easy to see that the next generation matrix with large domain is the matrix K calculated in Section 3.3. Since each disease state is a disease state at infection, the next generation matrix is K, the same as the next generation matrix with large domain. As in Example 1, the determinant of K is zero and K has rank 1. Thus the control reproduction number is the trace of K,

$$\mathscr{R}_c = \frac{\beta N_U}{\alpha_U} + \frac{\delta\sigma N_V}{\alpha_V}.$$

The next two examples include demographics (births and deaths).

Example 3. (A multi-strain model of gonorrhea) The following example comes from Lima and Torres (1997). The system of equations is

$$\begin{aligned}
\frac{ds}{dt} &= \rho - \mu S(t) - \frac{c\lambda_1 S(t) I_1(t)}{N(t)} - \frac{c\lambda_2 S(t) I_2(t)}{N(t)} + (1-p)\gamma_1 I_1(t) + \gamma_2 I_2(t), \\
\frac{dI_1}{dt} &= \frac{c\lambda_1 S(t) I_1(t)}{N(t)} - (\mu + \gamma_1) I_1(t), \\
\frac{dI_2}{dt} &= \frac{c\lambda_2 S(t) I_2(t)}{N(t)} - (\mu + \gamma_2) I_2(t) + p\gamma_1 I_1(t),
\end{aligned}$$

where S is the susceptible class, I_1 is the class infected with strain 1, and I_2 is the class of individuals infected with a mutated strain. The "birth" rate of the population is ρ, μ is the natural mortality rate, c is the probability of successful contact, λ_i is the of strain i, γ_i is the recovery rate of strain i, and p is the proportion of the original infected population that become infected by the mutated strain.

The disease-free equilibrium for this model is $[S = N, I_1 = 0, I_2 = 0]$. Next we will reorder our variables: $\left[\begin{array}{cc} \frac{dI_1}{dt} & \frac{dI_2}{dt} \end{array}\right]^T$ and note that we need only the infected classes to calculate $\mathscr{R}_0$. Then the new infection terms are $\frac{c\lambda_1 S(t) I_1(t)}{N(t)}$ in the $\frac{dI_1}{dt}$ equation and $\frac{c\lambda_2 S(t) I_2(t)}{N(t)}$ in the $\frac{dI_2}{dt}$ equation, $p\gamma_1 I_1(t)$ enter the I_2 class, but only *after* they have been infected with strain 1. Then

$$\mathscr{F} = \begin{bmatrix} \frac{c\lambda_1 S(t) I_1(t)}{N(t)} \\ \frac{c\lambda_2 S(t) I_2(t)}{N(t)} \end{bmatrix} \quad \text{and} \quad \mathscr{V} = \begin{bmatrix} (\mu+\gamma_1) I_1(t) \\ (\mu+\gamma_2) I_2(t) - p\gamma_1 I_1(t) \end{bmatrix}.$$

Since we have only two infected classes, $n = 2$, and our Jacobian matrices are

$$\begin{aligned}
\mathbf{F} &= \left[\begin{array}{cc} \frac{c\lambda_1 S(t)}{N(t)} & 0 \\ 0 & \frac{c\lambda_2 S(t)}{N(t)} \end{array}\right]\Bigg|_{DFE} = \begin{bmatrix} c\lambda_1 & 0 \\ 0 & c\lambda_2 \end{bmatrix} \\
\mathbf{V} &= \begin{bmatrix} \mu+\gamma_1 & 0 \\ -p\gamma_1 & \mu+\gamma_2 \end{bmatrix}\Bigg|_{DFE} = \begin{bmatrix} \mu+\gamma_1 & 0 \\ -p\gamma_1 & \mu+\gamma_2 \end{bmatrix}.
\end{aligned}$$

Then we can calculate the inverse of $\mathbf{V}$,

$$\mathbf{V}^{-1} = \begin{bmatrix} \frac{1}{\mu+\gamma_1} & 0 \\ \frac{p\gamma_1}{(\mu+\gamma_1)(\mu+\gamma_2)} & \frac{1}{\mu+\gamma_2} \end{bmatrix},$$

and

$$\mathbf{FV}^{-1} = \begin{bmatrix} \frac{c\gamma_1}{\mu+\gamma_1} & 0 \\ \frac{c\gamma_2 p\gamma_1}{(\mu+\gamma_1)(\mu+\gamma_2)} & \frac{c\gamma_2}{\mu+\gamma_2} \end{bmatrix}.$$

To calculate the spectral radius of $\mathbf{FV}^{-1}$ we find the eigenvalues of the matrix:

$$\rho(\mathbf{FV}^{-1}) = \max\left\{\frac{c\lambda_1}{\mu+\gamma_1}, \frac{c\lambda_2}{\mu+\gamma_2}\right\}.$$

We often call $\mathscr{R}_1 = \frac{c\lambda_1}{\mu+\gamma_1}$ the reproductive number for strain 1 and $\mathscr{R}_2 = \frac{c\lambda_2}{\mu+\gamma_2}$ the reproductive number for strain 2. Then the basic reproductive number is

$$\mathscr{R}_0 = \max[\mathscr{R}_1, \mathscr{R}_2].$$

Example 4. (A three population model of West Nile virus) The following example comes from Chowell-Puente et al. (2004b). The system of equations is

$$\begin{aligned}
S'_M &= \mu_M N_M - \beta_H S_M b \frac{N_H}{N_M}\frac{P_H N_M}{N_H+N_B}\frac{I_H}{N_H} - \beta_B S_M b \frac{N_B}{N_M}\frac{P_B N_M}{N_H+N_B}\frac{I_B}{N_B} - S_M\mu_M,\\
I'_M &= \beta_H S_M b \frac{N_H}{N_M}\frac{P_H N_M}{N_H+N_B}\frac{I_H}{N_H} + \beta_B S_M b \frac{N_B}{N_M}\frac{P_B N_M}{N_H+N_B}\frac{I_B}{N_B} - I_M\mu_M,\\
S'_B &= \Lambda - \beta_M S_B b \frac{N_M}{N_B}\frac{N_B}{N_H+N_B}\frac{I_M}{N_M} - S_B\mu_B,\\
I'_B &= \beta_M S_B b \frac{N_M}{N_B}\frac{N_B}{N_H+N_B}\frac{I_M}{N_M} - I_B\mu_B,\\
S'_H &= \mu_H N_H - \beta_M S_H b \frac{N_M}{N_H}\frac{N_H}{N_H+N_B}\frac{I_M}{N_M} - S_H\mu_H,\\
I'_H &= \beta_M S_H b \frac{N_M}{N_H}\frac{N_H}{N_H+N_B}\frac{I_M}{N_M} - (\mu+\theta)I_H,\\
R'_H &= \theta I_H - \mu_H R_H,
\end{aligned}$$

where S_i refers to the susceptible class of species i, and I_i refers to the infected class of species i for $i = M, B, H$, mosquitoes, birds, and humans, respectively. Then P_i is the mosquito biting preference for species i, μ_i is the natural mortality rate of species i, b is the number of bites per mosquito per unit time, θ is the human recovery rate, β_M is the transmission probability from mosquito to host per bite, β_B is the transmission probability from birds to mosquito, and β_H is the transmission probability from humans to mosquito.

The disease-free equilibrium for this model is $[S_M = N_M, I_M = 0, S_B = \frac{\Lambda}{\mu_B}, I_B = 0, S_H = N_H, I_H = 0, R_H = 0]$. Next we will reorder our variables

$$[I_M, I_B, I_H, S_M, S_B, S_H, R_H]$$

and again note that we need only the infected classes. Then the new infections enter our $\mathscr{F}$ vector as

$$\mathscr{F} = \begin{bmatrix} \beta_H S_M b \frac{N_H}{N_M} \frac{P_H N_M}{N_H+N_B} \frac{I_H}{N_H} + \beta_B S_M b \frac{N_B}{N_M} \frac{P_B N_M}{N_H+N_B} \frac{I_B}{N_B} \\ \beta_M S_B b \frac{N_M}{N_B} \frac{N_B}{N_H+N_B} \frac{I_M}{N_M} \\ \beta_M S_H b \frac{N_M}{N_H} \frac{N_H}{N_H+N_B} \frac{I_M}{N_M} \end{bmatrix},$$

and our $\mathscr{V}$ vector is the remaining terms,

$$\mathscr{V} = \begin{bmatrix} I_M \mu_M \\ I_B \mu_B \\ (\mu+\theta) I_H \end{bmatrix}.$$

Then our Jacobian matrices are

$$\mathbf{F}|_{DFE} = \begin{bmatrix} 0 & \frac{\beta_B b P_B N_M}{N_H+\frac{\Lambda}{\mu_B}} & \frac{\beta_H b P_H N_M}{N_H+\frac{\Lambda}{\mu_B}} \\ \frac{\beta_M b \Lambda}{\mu_B (N_H+\frac{\Lambda}{\mu_B})} & 0 & 0 \\ \frac{\beta_M b N_H}{N_H+\frac{\Lambda}{\mu_B}} & 0 & 0 \end{bmatrix},$$

$$\mathbf{V}|_{DFE} = \begin{bmatrix} \mu_M & 0 & 0 \\ 0 & \mu_B & 0 \\ 0 & 0 & \mu_H+\theta \end{bmatrix},$$

and

$$\mathbf{V}^{-1} = \begin{bmatrix} \frac{1}{\mu_M} & 0 & 0 \\ 0 & \frac{1}{\mu_B} & 0 \\ 0 & 0 & \frac{1}{\mu_H+\theta} \end{bmatrix},$$

$$\mathbf{F}\mathbf{V}^{-1} = \begin{bmatrix} 0 & \frac{\beta_B b P_B N_M}{(N_H+\frac{\Lambda}{\mu_B})\mu_B} & \frac{\beta_H b P_H N_M}{(N_H+\frac{\Lambda}{\mu_B})(\mu_H+\theta)} \\ \frac{\beta_M b \Lambda}{\mu_B (N_H+\frac{\Lambda}{\mu_B \mu_M}} & 0 & 0 \\ \frac{\beta_M b N_H}{N_H+\frac{\Lambda}{\mu_B \mu_M}} & 0 & 0 \end{bmatrix},$$

from which we calculate the eigenvalues and determine that the spectral radius is

$$\mathscr{R}_0 = \sqrt{\frac{bN_H}{\frac{\Lambda}{\mu_B}+N_H} \frac{\beta_M}{\mu_M} \frac{bP_H N_M}{\frac{\Lambda}{\mu_B}+N_H} \frac{\beta_H}{\mu_H+\theta} + \frac{b\frac{\Lambda}{\mu_B}}{\frac{\Lambda}{\mu_B}+N_H} \frac{\beta_M}{\mu_M} \frac{bP_B N_M}{\frac{\Lambda}{\mu_B}+N_H} \frac{\beta_B}{\mu_B}}$$

We have described the next generation matrix method for continuous models. There is an analogous theory for discrete systems, described in [Allen and van den Driessche (2008)].

9.10.1 A Global Asymptotic Stability Result

There are some situations in which $\mathscr{R}_0 < 1$ in which it is possible to show that the asymptotic stability of the disease - free equilibrium is global, that is, all solutions approach the disease - free equilibrium, not only those with initial values sufficiently close to this equilibrium.

We will say that a vector is nonnegative if each of its components is nonnegative, and that a matrix is nonnegative if each of its entries is non - negative. We rewrite the system (9.35) as

$$\begin{aligned} x' &= -Ax - \hat{f}(x,y), \\ y'_j &= g_j(x,y)\,, \quad j = 1,\ldots,m. \end{aligned} \tag{9.39}$$

If $\mathscr{R}_0 < 1$, we have shown that the disease - free equilibrium is asymptotically stable, and that $-A = -(F-V)$ is a non - singular M - matrix.

Theorem 9.2 (Castillo-Chavez, Feng, and Huang (2002)). *If $-A$ is a nonsingular M-matrix and $\hat{f} \geq 0$, if the assumptions on the model* (9.35) *made earlier in this section are satisfied, and if $\mathscr{R}_0 < 1$, then the disease-free equilibrium of* (9.39) *is globally asymptotically stable.*

Proof. The variation of constants formula for the first equation of (9.39) gives

$$x(t) = e^{-tA}x(0) - \int_0^t e^{-(t-s)A}\hat{f}(x(s),y(s))ds.$$

It can be shown that $e^{-tA} \geq 0$ if $-A$ is an M-matrix, because

$$-A = B - sI$$

with $B \geq 0$,

$$e^{-tA} = e^{tB}e^{-stI} = e^{tB}e^{-st}I = e^{tB}e^{-st},$$

and $e^{tB} \geq 0$, since $B \geq 0$. This, together with the assumption that $\hat{f} \geq 0$, implies that

$$0 \leq x(t) \leq e^{-tA}x(0),$$

and since $e^{-tA}x(0) \to 0$ as $t \to \infty$ it follows that $x(t) \to 0$ as $t \to \infty$.

There are examples to show that the disease-free equilibrium may not be globally asymptotically stable if the condition $\hat{f} \geq 0$ is not satisfied.

Exercises

1. For each of the models (9.16), (9.18), and (9.19) use the next generation approach to calculate their reproduction numbers.

2. Use the next generation approach to calculate the basic reproduction number for the model (9.26) but with infectivity in the exposed class having a reduction factor ε.
3. Formulate an SEITR model and calculate its reproduction number.
4. For each of the examples in this section determine whether the disease-free equilibrium is globally asymptotically stable when $\mathscr{R}_0 < 1$.

9.11 Directions for Generalization

A fundamental assumption in the model (9.2) is homogeneous mixing, that all individuals are equivalent in contacts. A more realistic approach would include separation of the population into subgroups with differences in behavior. For example, in many childhood diseases the contacts that transmit infection depend on the ages of the individuals, and a model should include a description of the rate of contact between individuals of different ages. Other heterogeneities that may be important include activity levels of different groups and spatial distribution of populations. Network models may be formulated to include heterogeneity of mixing, or more complicated compartmental models can be developed.

An important question that should be kept in mind in the formulation of epidemic models is the extent to which the fundamental properties of the simple model (9.2) carry over to more elaborate models.

An epidemic model for a disease in which recovery from infection brings only temporary immunity cannot be described by the models of this chapter because of the flow of new susceptibles into the population. This effectively includes demographics in the model, and such models will be described in the next chapter.

Many of the important underlying ideas of mathematical epidemiology arose in the study of malaria begun by Sir R.A. Ross (1911). Malaria is one example of a disease with vector transmission, the infection being transmitted back and forth between vectors (mosquitoes) and hosts (humans). Other vector diseases include West Nile virus and HIV with heterosexual transmission. Vector transmitted diseases require models that include both vectors and hosts.

9.12 Some Warnings

An actual epidemic differs considerably from the idealized models (9.2) and (9.28). Some notable differences are these:

1. When it is realized that an epidemic has begun, individuals are likely to modify their behavior by avoiding crowds to reduce their contacts and by being more careful about hygiene to reduce the risk that a contact will produce infection.
2. If a vaccine is available for the disease that has broken out, public health measures will include vaccination of part of the population. Various vaccination

strategies are possible, including vaccination of health care workers and other first line responders to the epidemic, vaccination of members of the population who have been in contact with diagnosed infectives, and vaccination of members of the population who live in close proximity to diagnosed infectives.
3. Isolation may be imperfect; in-hospital transmission of infection was a major problem in the SARS epidemic.
4. In the SARS epidemic of 2002–2003, in-hospital transmission of disease from patients to health care workers or visitors because of imperfect isolation accounted for many of the cases. This points to an essential heterogeneity in disease transmission that must be included whenever there is any risk of such transmission.

9.13 Project: Discrete Epidemic Models

The discrete analogue of the continuous-time epidemic model (9.2) is

$$\begin{aligned} S_{j+1} &= S_j G_j, \\ I_{j+1} &= S_j\big(1-G_j\big)+\sigma I_j, \\ G_j &= e^{-\beta I_j/N}, \quad j=1,2,\ldots, \end{aligned} \tag{9.40}$$

where S_j and I_j denote the numbers of susceptible and infective individuals at time j, respectively, G_j is the probability that a susceptible individual at time j will remain susceptible to time $j+1$, and $\sigma = e^{-\alpha}$ is the probability that an infected individual at time j will remain infected to time $j+1$. Assume that the initial conditions are $S(0)=S_0>0$, $I(0)=I_0>0$, and $S_0+I_0=N$.

Exercise 1. Consider the system (9.40).

(a) Show that the sequence $\{S_j+I_j\}$ has a limit

$$S_\infty + I_\infty = \lim_{j\to\infty}(S_j+I_j).$$

.

(b) Show that

$$I_\infty = \lim_{j\to\infty} I_j = 0.$$

(c) Show that

$$\log\frac{S_0}{S_\infty} = \beta\sum_{m=0}^{\infty}\frac{I_m}{N}.$$

(d) Show that

$$\log\frac{S_0}{S_\infty} = \mathscr{R}_0\left[1-\frac{S_\infty}{N}\right],$$

with $\mathscr{R}_0 = \frac{\beta}{1-\sigma}$.

Next, consider the case that there are k infected stages and there is treatment in some stages, with treatment rates that can be different in different stages. Assume that selection of members for treatment occurs only at the beginning of a stage. Let $I_j^{(i)}$ and $T_j^{(i)}$ denote the numbers of infected and treated individuals respectively in stage i ($i = 1,2,\ldots,k$) at time j. Let σ_i^I denote the probability that an infected individual in the $I^{(i)}$ stage continues on to the next stage, either treated or untreated, and let σ_i^T denote the probability that an individual in the $T^{(i)}$ stage continues on to the next treated stage. In addition, of the members leaving an infected stage $I^{(i)}$, a fraction p_i enters treatment in $T^{(i+1)}$, while the remaining fraction q_i continues to $I^{(i+1)}$. Let m_i denote the fraction of infected members who go through the stage $I^{(i)}$, and n_i the fraction of infected members who go through the stage $T^{(i)}$. Then,

$$\begin{array}{llll} m_1 = q_1, & m_2 = q_1q_2, & \ldots, & m_k = q_1q_2\cdots q_k, \\ n_1 = p_1, & n_2 = p_1 + q_1p_2, & \ldots, & n_k = p_1 + q_1p_2 + \ldots + q_1q_2\cdots q_{k-1}p_k. \end{array}$$

The discrete system with treatment is

$$\begin{aligned} S_{j+1} &= S_jG_j, \\ I_{j+1}^{(1)} &= q_1S_j(1-G_j) + \sigma_1^I I_j^{(1)}, \\ T_{j+1}^{(1)} &= p_1S_j(1-G_j) + \sigma_1^T T_j^{(1)} \\ I_{j+1}^{(i)} &= q_i(1-\sigma_{i-1}^I)I_j^{(i-1)} + \sigma_i^I \eta_i I_j^{(i)}, \\ T_{j+1}^{(i)} &= p_i(1-\sigma_{i-1}^I)I_j^{(i-1)} + (1-\sigma_{i-1}^T)T_j^{(i-1)} + \sigma_i^T T_j^{(i)}, \end{aligned} \tag{9.41}$$

$[i = 2,\ldots,k, j \geq 0]$, with

$$G_j = e^{-\beta\sum_{i=1}^k \left(\varepsilon_i I_j^{(i)}/N + \delta_i T_j^{(i)}/N\right)},$$

where ε_i is the relative infectivity of untreated individuals at stage i and δ_i is the relative infectivity of treated individuals at stage i. Consider the initial conditions

$$I_0^{(1)}(0) = q_1I_0, \quad T_0^{(1)}(0) = p_1I_0, \quad I_0^{(i)}(0) = T_0^{(i)}(0) = 0, \quad i \geq 2, \quad S_0 + I_0 = N.$$

Exercise 2. Consider the system (9.41). Show that

$$\log\frac{S_0}{S_\infty} = \mathscr{R}_c\left[1 - \frac{S_\infty}{N}\right], \tag{9.42}$$

with

$$\mathscr{R}_c = \beta\sum_{i=1}^k\left[\frac{\varepsilon_i m_i}{1-\sigma_i^I} + \frac{\delta_i n_i}{1-\sigma_i^T}\right].$$

Hint. Equation (9.42) can be proved by showing the following equalities first:

$$\log\frac{S_0}{S_\infty} = \beta\sum_{i=1}^{k}\left[\varepsilon_i\sum_{j=1}^{\infty}\frac{I_j^{(i)}}{N} + \delta_i\sum_{j=1}^{\infty}\frac{T_j^{(i)}}{N}\right],$$

$$m_i(N-S_\infty) = (1-\sigma_i^I)\sum_{j=0}^{\infty} I_j^{(i)}, \quad i \geq 2,$$

$$\frac{n_i(N-S_\infty)}{1-\sigma_i^T} = \sum_{j=0}^{\infty} T_j^{(i)}, \quad i \geq 2.$$

[References: Feng (2007), Feng, Xu, & Zhao (2007).]

9.14 Project: Fitting Data for an Influenza Model

Consider an SIR model (9.2) with basic reproduction number 1.5.

1. Describe the qualitative changes in (S,I,R) as a function of time for different values of β and α with $\beta \in \{0.0001, 0.0002, \ldots, 0.0009\}$, for the initial condition $(S,I,R) = (10^6, 1, 0)$.
2. Discuss the result of part (a) in terms of the basic reproductive number (what is β/γ?). Use a specific disease such as influenza to provide simple interpretations for the different time courses of the disease for the different choices of β and γ.
3. Repeat the steps in part (a) for values of $R_0 \in \{1.75, 2, 2.5\}$, and for each value of R_0, choose the best pair of values (β, α) that fits the slope before the first peak in the data found in the table for reported H1N1 influenza cases in México below. (Hint: normalize the data so that the peak is 1, and then multiply the data by the size of the peak in the simulations.)

9.15 Project: Social Interactions

Suppose we have a system with multiple classes of mathematical biology teachers (MBT) at time t. The classes roughly capture the MBT individual attitudes toward learning new stuff. "Reluctant" means the class of MBTs that come into the door as new hires without a disposition to learn new stuff; the positive class corresponds to those who join the MBTs with the right attitude; and the rest of the classes should be self-explanatory.

Day	Cases	Day	Cases	Day	Cases	Day	Cases	Day	Cases	Day	Cases
75	2	95	4	115	318	135	83	155	152	175	328
76	1	96	11	116	399	136	75	156	138	176	298
77	3	97	5	117	412	137	87	157	159	177	335
78	2	98	7	118	305	138	98	158	186	178	330
79	3	99	4	119	282	139	71	159	222	179	375
80	3	100	4	120	227	140	73	160	204	180	366
81	4	101	4	121	212	141	78	161	257	181	291
82	4	102	11	122	187	142	67	162	208	182	251
83	5	103	17	123	212	143	68	163	198	183	215
84	7	104	26	124	237	144	69	164	193	184	242
85	3	105	20	125	231	145	65	165	243	185	223
86	1	106	12	126	237	146	85	166	231	186	317
87	2	107	26	127	176	147	55	167	225	187	305
88	5	108	33	128	167	148	67	168	239	188	228
89	7	109	44	129	139	149	75	169	219	189	251
90	4	110	107	130	142	150	71	170	199	190	207
91	10	111	114	131	162	151	97	171	215	191	159
92	11	112	155	132	138	152	168	172	309	192	155
93	13	113	227	133	117	153	126	173	346	193	214
94	4	114	280	134	100	154	148	174	332	194	237

Table 9.2 Reported casec for H1N1-pandemic in Mexico.

$R(t)$: reluctant MBTs
$P(t)$: positive MBTs
$M(t)$: masterful MBTs
$U(t)$: unchangeable (that is, negative) MBTs
$I(t)$: inactive MBTs

Assume that $N(t) = R(t) + P(t) + M(t) + U(t) + I(t)$ and that the total number of MBTs is constant, that is, $N(t) = \frac{K}{\mu}$ for all t, where K is a constant. The model is

$$\begin{aligned}
\frac{dP}{dt} &= qK - \beta P\frac{M}{K} + \delta R - \mu P,\\
\frac{dR}{dt} &= (1-q)K - (\delta+\mu)R - \alpha R,\\
\frac{dM}{dt} &= \beta P\frac{M}{K} - (\gamma+\mu)M,\\
\frac{dU}{dt} &= -\mu U + \alpha R,\\
\frac{dI}{dt} &= \gamma M - \mu I,
\end{aligned}$$

where q, β, δ, μ, γ, and α are constants and $0 \le q \le 1$.

1. Interpret the parameters.

2. Look at the stability of the simplest equilibrium (your choice).
3. From $\mathfrak{R}_0$, discuss what would be the impact of changing parameters q, γ, and δ.
4. What are your conclusions from this model?