Chapter Six

Stochastic Dynamics

All the models considered thus far have been deterministic. This means they are essentially fixed "clockwork" systems; given the same starting conditions, exactly the same trajectory is always observed. Such a Newtonian view of the world does not apply to the dynamics of real pathogens. If it were possible to "re-run" a real-world epidemic, we would not expect to observe exactly the same people becoming infected at exactly the same times. Clearly, there is an important element of chance. Stochastic models are concerned with approximating or mimicking this random or probabilistic element. In general, the role played by chance will be most important whenever the number of infectious individuals is relatively small, which can be when the population size is small, when an infectious disease has just invaded, when control measures are successfully applied, or during the trough phase of an epidemic cycle. In such circumstances, it is especially important that stochasticity is taken into account and incorporated into models.

This chapter details three distinct methods of approximating the chance element in disease transmission and recovery: (1) introducing chance directly into the population variables, (2) by random parameter variation, and (3) individual-level, explicit modeling of the random events (Bartlett 1957). This third method is generally the most popular (Mollison et al. 1994; Levin and Durrett 1996) and will be the predominant focus of this chapter, illustrated by examples from the recent literature. All these examples have two elements in common. First, they predict different outcomes from the same initial conditions and, as such, multiple simulations are required to determine the expected range of behavior. Second, they all require the use of a random number-generating routine—this can be thought of as a rather clever computational trick that can deliver an apparently random sequence of numbers in some prescribed range (Box 6.1).

Box 6.1 Random Number Generation

The question of how we can use a computer algorithm to generate a truly random series of numbers has received considerable attention from mathematicians and computer scientists. Ultimately, because computers are deterministic devices, any random number generator (RNG) must also be deterministic. Hence, the set of results given by a RNG must at some point repeat and exhibit some form of pattern (albeit a very complex one) due to the underlying computer code. Some researchers have attempted to overcome these limitations by sampling and processing sources of entropy outside the computer, such as atmospheric noise, radioactive decay, or even lava lamps! In the absence of these more elaborate methods, others have focused on writing RNGs with cycles as long as possible and the patterns too complex to be detected or introduce spurious results. As such, RNGs are inherently quite complex and slow, often being the slowest step of a computer program; therefore, having an efficient routine is a must.

Fortunately, such routines already exist and are widely available. Many computer languages have RNGs built in as standard, or sample code can be found on the Web or in Numerical Recipes books (Press et al. 1988). This chapter assumes that the RNG, $RAND$, returns a

high-precision number uniformly distributed and strictly between 0 and 1—values of exactly 0 or 1 should be discounted because they can produce spurious results. Some RNGs produce integers between zero and some maximum value, in which case the random number must be rescaled before use.

One element all RNGs share is their need for an initial seed value to start the processes. Due to their deterministic nature, given the same seed, the same sequence of "random" numbers will be generated. For this reason, the seed is usually chosen to be the current time (to the nearest second) because this is an ever-changing value.

Other routines exist that will generate random numbers from different distributions, such as normal, Poisson, or binomial (see, for example, Press et al. 1988). Because it is frequently used in the first two sections of this chapter, the following pseudo-code (Box-Mueller 1958) will pick random numbers from a normal distribution, with mean $MEAN$ and variance VAR.

1. $R1 = 2 \times RAND - 1$.
2. $R2 = 2 \times RAND - 1$.
3. $R3 = R1 \times \mathrm{R1} + \mathrm{R2} \times \mathrm{R2}$.
4. if $\mathrm{R3} \geq 1$ or $\mathrm{R3} = 0$, go to line 1.
5. $\mathrm{R4} = \sqrt{-2 \times VAR \times \log(\mathrm{R3})/\mathrm{R3}}$.
6. $NORM = MEAN + \mathrm{R1} \times \mathrm{R4}$.

To enable a fair comparison of the different methods of implementing stochasticity, we primarily focus on the SIR model with births and deaths, which, in the deterministic framework, always converges to an equilibrium in an oscillatory manner (Chapter 2). We also utilize a constant set of parameters: $R_0 = 10$, $1/\gamma = 10$ days, and $\nu = \mu = 0.02$ per year, which corresponds to a mean life expectancy of 50 years and a constant population size. For stochastic models, we need to consider both the long-term behavior and the short-term "transient" dynamics, as well as study both the mean and variance of the number of infected cases. Because stochastic models have close relationships to integer-valued (or individual-based) models, we generally deal with the number (as opposed to the proportion) of susceptibles, infecteds, and recovereds, labeled X, Y, and Z.

Five key features distinguish stochastic models from their deterministic counterparts. Although not evident for all forms of stochastic models, these features can have a profound impact on epidemiological dynamics and will be a consistent theme throughout this chapter. In brief, these distinguishing features are:

1. *Variability between simulations*. The most obvious element of any stochastic model is that different simulations give rise to different outcomes. This implies that although the general statistical properties (such as the mean and the variance) may be accurately predicted, it is generally impossible to predetermine the precise disease prevalence at any given point in the future.
2. *Variances and covariances*. The continual perturbations caused by the random nature of stochastic equations leads to variation in the prevalence of disease and the number of susceptibles. Additionally, the interaction between stochasticity and deterministic dynamics generally causes negative covariance between the numbers of infectious and susceptible individuals, which in turn can cause the mean population levels $(\overline{X}, \overline{Y})$ to deviate from the deterministic equilibria.
3. *Increased transients*. Stochastic perturbations away from the equilibrium solution are countered by the generally convergent behavior of the underlying deterministic dynamics. When far from the endemic equilibrium point, these restorative forces are usually strong and dominate, so that the model acts much like the deterministic

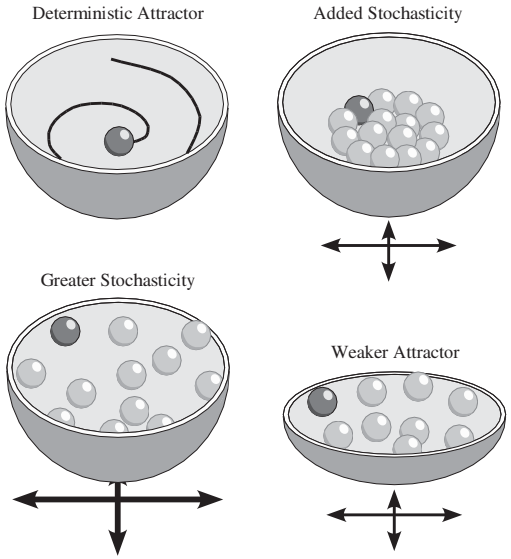

Figure 6.1. Four examples showing how the effects of stochasticity can be conceptualized. Top left: deterministic attractor with no stochasticity. Top right: moderate amounts of stochasticity and a strong underlying attractor. Bottom left: much higher levels of stochasticity lead to greater variation. Bottom right: moderate stochasticity but with a much weaker attractor; this also produces greater variation.

equations. Thus, the dynamics of stochastic models can be conceptualized as resulting from random perturbations away from, and transient-like return toward, the deterministic attractor.

4. *Stochastic resonance*. As most SIR-type disease models approach equilibrium in a series of decaying epidemics, the increased transient-like dynamics of stochastic models leads to oscillations close to the natural frequency (Chapter 2). Thus, stochasticity can *excite* epidemic oscillations around the normal endemic prevalence, leading to sustained cycles.
5. *Extinctions*. Those probabilistic models that are integer-valued may additionally suffer from stochastically driven extinctions. In closed populations, chance fluctuations will always result in eventual disease extinction, irrespective of population size. Long-term persistence requires the import of pathogens from an external source. Similar extinctions may occur during the early stages of invasion.

Figure 6.1 shows an intuitive way to conceptualize the effects of stochasticity. The bowl shape represents the underlying deterministic attractor, which in turn completely defines the dynamics of a ball placed inside the bowl. Stochastic forces can then be visualized as the effects of shaking the bowl. When there are no stochastic forces (top left), the ball follows the deterministic attractor and eventually settles in the bottom of the bowl. When a moderate amount of stochasticity or shaking is added (top right), the position of the ball is, in the long run, less certain although it generally lies close to the base of the bowl. Substantially greater deviation from the deterministically predicted position of the ball occurs when either there is more shaking (greater stochasticity) or a shallower bowl (a weaker underlying attractor), as shown in the two lower panels. Although there are clear

phenomenological discrepancies between the gravity-driven dynamics of the ball and the epidemiological dynamics of a disease, this is a useful pictorial analogy that can help us conceptualize the interactions between stochasticity and determinism.

6.1. OBSERVATIONAL NOISE

The simplest form of noise that can be introduced into disease modeling is observational noise. For this approach, the underlying epidemic dynamics remain the standard deterministic differential equations, but there is assumed to be some uncertainty in the recorded data. Information on notifiable diseases, such as measles in England and Wales (Finkenstädt and Grenfell 1998), foot-and-mouth (Keeling et al. 2001b), or classical swine fever (Elbers et al. 1999), tends to be highly detailed and fairly accurate. In contrast, infectious diseases that are largely asymptomatic such as meningocaccol infection (Ranta et al. 1999) or diseases of wildlife (Hudson et al. 2001) may be subject to greater uncertainty in their reporting.

For any specific infectious disease, the stochasticity due to observational noise can most readily be quantified as either incomplete reporting of true cases, or mis-diagnosis (both type I and type II errors). For the measles data from England and Wales, underreporting is well documented, with an estimated 60% of the actual cases being reported (Fine and Clarkson 1982; Finkenstädt and Grenfell 2000). In contrast, during the 2001 foot-and-mouth outbreak in the United Kingdom, misdiagnoses were frequent—some animals infected with other diseases were incorrectly diagnosed as having foot-and-mouth, although it is unlikely that animals actually infected with foot-and-mouth could have escaped detection.

The recorded number of infected individuals, Y_{recorded}, can be represented as the sum of two binomial distributions:

$$Y_{\text{recorded}} = \text{Bin}(p_r, Y_{\text{true}}) + \text{Bin}(p_m, N - Y_{\text{true}}),$$

where p_r is the probability that a case is correctly reported and p_m is the probability that a healthy individual is misdiagnosed; N is the total population size. Although such considerations are obviously important when interpreting case-report data (Finkenstädt and Grenfell 1998; Rohani et al. 2003), they do not interact with or influence the underlying epidemiological dynamics. The true number of cases always tends to the deterministic fixed point and observations are purely determined by the two probabilities, p_r and p_m. Even in recurrent epidemics, it has been shown that observational error affects the amplitude of data, not the periodicity (see, for example, Miller et al. 1992).

Observational noise does not impact epidemiological dynamics, it modifies only the reported data.

6.2. PROCESS NOISE

A more intuitive and fundamentally different way to incorporate noise is to introduce it directly into the deterministic equations (Bjørnstad, Finkenstädt, and Grenfell 2002). As such, the dynamics at each point in time are subject to some random variability and this variability is propagated forward in time by the underlying equations. We are, therefore,

Box 6.2 Inclusion of Noise in Differential Equations

Consider the stochastic differential equation model:

$$\frac{dx}{dt} = \text{Noise}$$

The most simple means of solving such equations is using the Euler method of integration, breaking time into small components, δt.

$$x_{t+\delta t} = x_t + \delta t \frac{dx}{dt} = x_t + \delta t \text{Noise} = x_0 + \delta t \sum_{1}^{\frac{t}{\delta t}} \text{Noise}$$

Thus, the equations progress as the summation of many small noise terms, with mean zero.

If successive noise terms are independent, then the variance of x at any time decays to zero as the step size is made smaller. Thus, in the limit $\delta t \to 0$, when the updating method is exact, all the noise terms effectively cancel. This is a reflection of the well-known problem that there is no simple mathematical method of expressing the noise term.

The simplest solution to this problem is to scale the noise term with respect to the integration step. Throughout this chapter we shall assume that Noise = RANDN—independent Gaussian with mean 0 and variance 1—and that:

$$\xi = \frac{\text{Noise}}{\sqrt{\delta t}},$$

such that as the time step of integration decreases, the amplitude of the noise used in each step, ξ, increases. If this scaling is used with higher-order integration methods (such a Runge-Kutta; see Press et al. 1988), then the noise should be calculated *before* each integration step, though it is often safer to use forward Euler.

This new definition of ξ has the properties that we require. Such that if

$$\frac{dx}{dt} = f\xi,$$

then, averaged over multiple simulations, the mean of x is zero while the standard deviation grows like $f\sqrt{(t)}$, and therefore the dynamics correspond to a random walk.

concerned with the interplay between deterministic and stochastic forces—how they cancel out or amplify each other.

As a starting point, let us incorporate noise into only the transmission term (Marcus 1991). Let $\xi(t)$ be a time series of random deviates derived from the normal distribution with mean zero and unit variance (see Box 6.2). The basic equations, assuming frequency-dependent (mass-action) transmission, are transformed to:

$$\begin{aligned}
\frac{dX}{dt} &= \nu N - [\beta XY/N + f(X,Y)\xi] - \mu X, \\
\frac{dY}{dt} &= [\beta XY/N + f(X,Y)\xi] - \gamma Y - \mu Y, \\
\frac{dZ}{dt} &= \gamma Y - \mu Z.
\end{aligned} \tag{6.1}$$

This is online program 6.1

For generality, a function $f(X, Y)$ has been included to scale the randomness in response to the current variable sizes.

6.2.1. Constant Noise

Figure 6.2 shows examples of long-term dynamics as the value of f and hence the amount of noise is varied. As expected, when more noise is added, the dynamics deviate further from the deterministic equilibrium. However, rather than these deviations being completely random, there is a distinct oscillatory component to the dynamics. This is caused by the interaction between deterministic and stochastic forces—a common feature in all dynamically stochastic models (Bartlett 1957; Renshaw 1991; Rohani et al. 2002; McKane and Newman 2005). The noise terms allow trajectories to wander away from the equilibrium, but the underlying deterministic clockwork forces them back toward the equilibrium point. This return movement closely matches the deterministic prediction of decaying oscillations with a natural period (see Chapter 2). As pointed out by Grossman (1980), in general, the SIR equilibrium is weakly stable (because the real part of the dominant eigenvalue is small relative to the imaginary part). As a result, the underlying oscillatory dynamics may be "excited" by noise. This process is referred to as stochastic resonance and is a well-known phenomenon in physics. We can therefore see noise (and most forms of dynamic stochasticity) as overcoming deterministic forces when these are weak and close to the equilibrium point, but dominated largely by the deterministic behavior far from the equilibrium point. This tension between deterministic and stochastic forces means that the transient dynamics on the approach to the deterministic equilibrium can dominate the observed stochastic behavior (Rohani et al. 2002; Bauch and Earn 2003). However, if the noise term is very large (greater than shown), it can completely swamp any deterministic component and a pure random-walk is observed.

Noise in the transmission term can cause the number of cases to resonate at (or near) the natural frequency of the system. Hence, the observed stochastic dynamics more closely match the transient behavior of the deterministic equations.

Figure 6.2 also shows how the aggregate properties of the dynamics vary with the noise level. The variance in the number of infectious individuals increases almost linearly with the variance in the noise. For the very highest levels of noise, however, the variance does not increase linearly; this is due to the action of strong nonlinear forces that operate when the amplitude of the resonant epidemics becomes large and the number of infecteds is occasionally forced close to zero. Perhaps more surprisingly, there is a corresponding decrease in the covariance between susceptible and infected individuals. This is because if a particular value of the noise is "good" for the infected population ($\xi \geq 0$), it is "bad" for the susceptibles; hence, increasing noise levels generate increasingly large amplitude fluctuations with a strong negative covariance between X and Y.

Noise in the transmission term causes variance in the number of infecteds and in the number of susceptibles, and negative covariance between them. The magnitude of these values increases almost linearly with the variance of the noise.

This negative covariance has implications for the average dynamics. The rate at which cases arise ($\beta XY/N$) is now reduced, because the product of X and Y is smaller than expected due to the negative covariance. This reduction in transmission (and therefore a reduction in the effective reproductive ratio, $R = R_0 X/N$), is reflected by a slight increase in the mean number of susceptibles. This is a very interesting concept: Background stochasticity does not just cause variations about the deterministically predicted fixed

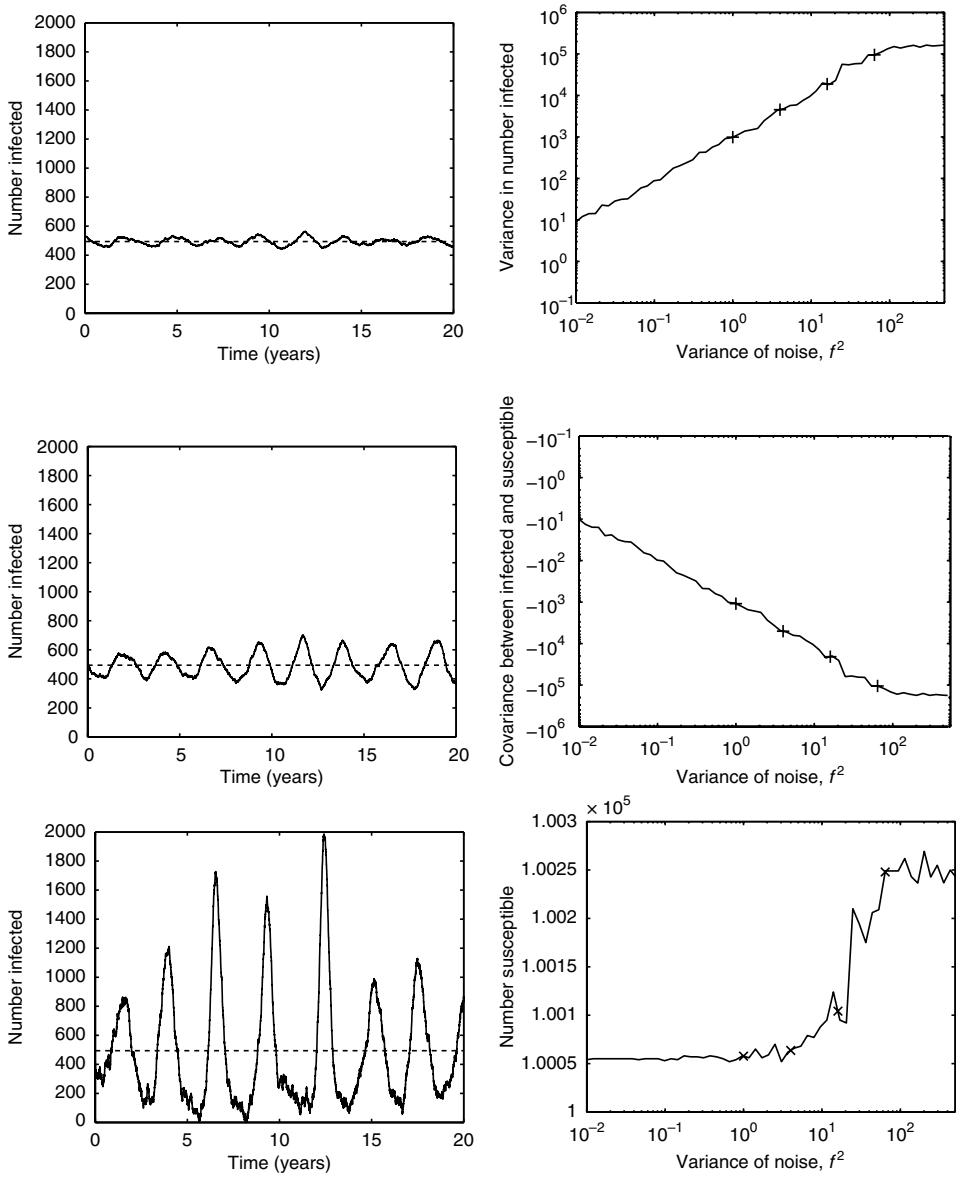

Figure 6.2. The left column gives examples of the dynamics of the SIR model with births and deaths ($\mu = \nu = 0.02$ per year, $R_0 = 10$, $1/\gamma = 10$ days, $N = 10^6$). The amount of noise added to the transmission terms increases from top to bottom ($f = 1$, 2, and 8). The deterministic equilibrium is depicted by the horizontal dashed line. The right column shows how the variance in the number infected, the covariance between susceptible and infecteds, and the average number of susceptibles change with the variance of the noise. From biological considerations, if noise ever forced the number of infected individuals below zero, the number was reset to zero—this only occurs for the largest levels of noise. Dots on the right-hand graphs mark the values of f used to generate the left-hand graphs.

points, but can actually cause significant changes to the mean values (Coulson et al. 2004). The nonlinear nature of the disease equations (in particular the product of X and Y in the transmission term) allows stochasticity to modify the mean values. From the graph it is clear that stochasticity has little effect on the mean values, primarily because the oscillations are relatively small and therefore the localized dynamics are almost linear. However, for small population sizes or diseases that show frequent localized extinctions or strong oscillatory resonance, the impact of the negative covariance on the mean can be far more substantial.

Stochasticity can introduce changes to the mean values as well as causing variation about the means.

6.2.2. Scaled Noise

Although the simple method of including noise outlined above provides variability, the amplitude of such random terms has so far been assumed to remain constant. In practice, this is rarely the case. Generally, the absolute magnitude of the variability increases with increasing population size, although the relative magnitude usually decreases (Keeling and Grenfell 1999; Bjørnstad et al. 2002). If events, such as transmission, occur at random, then in any short time interval the number of events will be Poisson distributed and hence their variance will equal the mean (Kendall 1949). This provides a direct means of determining the magnitude of the noise term, denoted f in the original equation (6.1), such that $f = \sqrt{\text{rate}}$. This concept can be extended still further, such that each event is associated with a noise term:

$$\begin{aligned}
\frac{dX}{dt} &= [\nu N + \sqrt{\nu N}\,\xi_1] - [\beta XY/N + \sqrt{\beta XY/N}\,\xi_2] - [\mu X + \sqrt{\mu X}\,\xi_3], \\
\frac{dY}{dt} &= [\beta XY/N + \sqrt{\beta XY/N}\,\xi_2] - [\gamma Y + \sqrt{\gamma Y}\,\xi_4] - [\mu Y + \sqrt{\mu Y}\,\xi_5], \\
\frac{dZ}{dt} &= [\gamma Y + \sqrt{\gamma Y}\,\xi_4] - [\mu Z + \sqrt{\mu Z}\,\xi_6].
\end{aligned} \qquad (6.2)$$

This is online program 6.2

This set of equations now contains six distinct noise terms ($\xi_1, \ldots \xi_6$), one for each event type; the same noise is used for each event even though it may appear in more than one equation. Although at equilibrium the event rates cancel each other (such that $\frac{d}{dt} = 0$), the noise terms add together. Thus, the amount of noise is determined by the absolute magnitude of opposing rates (such as births, which replenishes susceptibles, and death and infection, which decreases susceptibles), rather than by the rates of change themselves.

The effects of noise from events add together, even though the effects of the events themselves may cancel each other.

In Figure 6.3, we show the typical dynamics of models with this more complete form of scaled noise. Large populations (10 million or 100 million) are affected only slightly by such noise terms, with dynamics that are close to those predicted by the deterministic model. In contrast, smaller population sizes (smaller than 100,000) experience proportionally more noise and their behavior lies further from the deterministic ideal (Keeling and Grenfell 1999; Rohani et al. 2002). Once again, the presence of noise can be seen to induce cycles close to the natural epidemic period in the smaller population sizes (focusing on years 15–20 for $N = 10^6$). A log-log plot of variance against the mean number

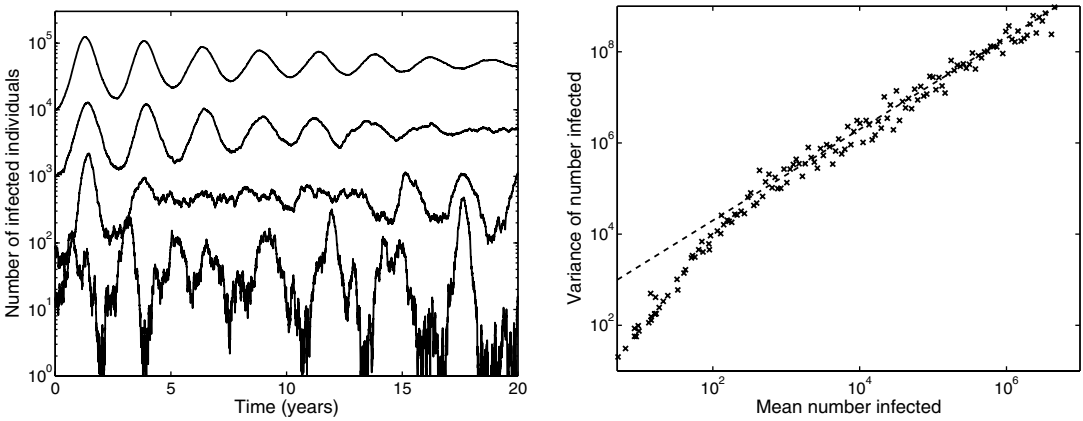

Figure 6.3. The dynamics of SIR epidemics ($\mu = \nu = 5.5 \times 10^{-5}$ per day, $R_0 = 10$, $1/\gamma = 10$ days) with scaled noise. The left-hand graph shows how the expected oscillatory behaviour becomes disrupted by noise in smaller populations, whereas large populations conform close to the deterministic ideal ($N = 10^5, 10^6, 10^7, 10^8$). The right-hand graph show the meanvariance relationship for a range of population sizes from 10,000 to 100 million. The dashed line represents $var = 100 \times mean$, clearly showing how the scaling operates at large population sizes.

infected, Figure 6.3 (right-hand graph), provides a concise method of summarizing the variability across a range of population sizes. For scaled noise, and large population sizes (with more that 200 infected individuals on average), the variance scales linearly with the mean and therefore with the population size as well. This linear behavior is expected for a wide range of stochastic epidemiological (as well as ecological) models, irrespective of the detailed dynamics or the way that stochasticity is introduced (Keeling and Grenfell 1999; Keeling 2000a). Thus, although large populations behave more like the deterministic equations, the absolute level of variation is greater than in small populations. For very small populations, this linear relationship between the mean and variance is destroyed due to the strong nonlinearities that operate when relatively large amplitude epidemics are triggered.

A log-log plot of mean against variance is a useful method of summarizing the variability with population size. For large population sizes, we expect the variance in the number of cases to be proportional to the mean.

6.2.3. Random Parameters

An associated modeling technique comparable to the introduction of noise terms is to consider the governing parameters to be noisy (Rand and Wilson 1991; Roberts and Saha 1999; Mahul and Durand 2000; Dexter 2003). Although mathematically equivalent, this alternative mechanism of including randomness has a very different underlying premise. The previous models assumed that the population levels were subject to random fluctuations, whereas the models in this section consider the basic parameters to vary. This can be for one of two reasons. First, there is variability in the parameters due to some external, unpredictable force. For example, the transmission rate β may be modified

by climatic conditions; many viruses survive better in cold, damp conditions with little ultraviolet light (Ansari et al. 1991; Abad et al. 1994; Albina 1997). Fluctuations in temperature or humidity will therefore be translated to fluctuations in β; such variability is independent of disease prevalence or the number of susceptibles. A second source of parameter fluctuations is individual variation, both in terms of the pattern of contacts and the immunological response to infection (resulting in variable transmission rates). In this case, the fluctuations will be determined by which individuals are infected at that time. As such, this form of stochasticity would be expected to decrease when more individuals are infected because the variability between them will be averaged out. We can consider this as a special case of the structured population models discussed in Chapter 3, where mixing is random and all individuals experience the same force (risk) of infection.

Focusing on variability in the transmission rate, the basic equations for the SIR model become:

$$\begin{aligned} \frac{dX}{dt} &= \nu N - \beta(1+F\xi)XY/N - \mu X, \\ \frac{dY}{dt} &= \beta(1+F\xi)XY/N - \gamma Y - \mu Y. \end{aligned} \tag{6.3}$$

When the variability is due to an external force, F is generally constant. However, when the noise comes from population heterogeneity in transmission we expect $F \propto Y^{-\frac{1}{2}}$—this is because the variability in the *average* transmission rate becomes smoothed out when many individuals are infected. (If the stochasticity is due solely to heterogeneity in susceptibility, then $F \propto X^{-\frac{1}{2}}$ and in general the effects will be weaker.) The proportionality constant in either case is determined by the magnitude of the external fluctuations or by the variability of the population, respectively.

Parameter noise can either be constant (if due to external factors) or decrease with the population level (if due to differences between individuals).

6.2.4. Summary

A range of models have now been described that can include stochasticity by modifying the basic form of the differential equations. This modification is achieved by inserting appropriately scaled noise, thereby mimicking a variety of random effects. We can now contrast these approaches, and comment on their general suitability, advantages, and disadvantages.

6.2.4.1. Contrasting Types of Noise

Let us contrast the differences between the models (equations (6.2) and (6.3)) and the early versions with simple additive noise (equation (6.1)). We do this by noting that f in the original formulation (equation (6.1)) is equal to $\beta F XY/N$ in the new model (equation (6.2)). Hence, we have four distinct noise scenarios:

1. Plain Additive Noise. f is constant.
2. Scaled Additive Noise. $f = \sqrt{\beta XY/N}$.
3. External Parameter Noise. F is constant, so $f \propto \beta XY/N$.
4. Heterogeneous Parameter Noise. $F \propto Y^{-\frac{1}{2}}$, so $f \propto \beta X\sqrt{Y}/N$.

From this comparison, we may consider how these noise terms (f) scale as the population size, N, changes, noting that in general X and Y are proportional to N. Scenario 1 is independent of the population size, scenarios 2 and 4 scale with the square root of the population size, and scenario 3 is proportional to the population size. Thus, when the population size is large the dominant error term is due to externally driven fluctuations in the parameters (scenario 3). Additionally, when there is considerable variability in an individual's response to infection, the parameter noise due to this factor (scenario 4) can easily exceed the scaled additive noise (scenario 2) due to dynamic variability. However, despite the potential importance of parameter noise, it has received relatively little attention in the epidemiology literature.

Noise can be generated from a variety of sources. The relative magnitude of these noise terms depends on the population size. External parameter noise dominates when the population is large.

6.2.4.2. Advantages and Disadvantages

Before we consider event-driven approaches, it is important to reflect on the relative merits of the methods outlined in this section. Including noise into the standard differential equations has three main benefits. First, the modification to the basic equations is relatively straight forward and therefore (with a little care), the same techniques can be used to iterate the equations and determine the dynamics. Second, the correspondence between the standard deterministic models and the stochastic equations is clear—as the noise terms are reduced to zero, we regain the deterministic dynamics. Finally, the computational overheads associated with this form of stochastic model are small in comparison to the techniques used for demographic stochasticity; in particular, simulation times are independent of population size.

These stochastic equations suffer from one major drawback: They do not incorporate the discrete, individual nature of populations and are therefore not a suitable model when population levels (e.g., the number of infectious individuals) are small. Techniques such as scaled noise (Section 6.2.2) or random parameters due to population heterogeneity (Section 6.2.3) are therefore trapped between two opposing elements. Such models differ substantially from their deterministic counter-part only when population levels are low, but are accurate representations of the dynamics only when population levels are large. Thus, although these noise-based models allow for the extinction of pathogen, their accuracy for such individual-based processes is questionable. For this reason the individual-based nature of demographic stochasticity is the predominant method of capturing stochasticity.

In summary, these noise-based modifications to the standard differential equations are most appropriate when considering the effects on large populations of parameter variation due to external sources. They can also help shed some analytical insights into the effects of stochasticity in general (see Section 6.6.1). However, when dealing with low numbers of individuals, an alternative approach is required.

6.3. EVENT-DRIVEN APPROACHES

Modeling approaches that incorporate demographic stochasticity are becoming increasingly popular. This is in large part due to their highly mechanistic approach to including randomness and the individual nature of their formulation (Bartlett 1956; Tille et al. 1991;

Garner and Lack 1995; Barlow et al. 1997; Smith et al. 2001). Demographic stochasticity is defined as fluctuations in population processes that arise from the random nature of events at the level of the individual. Therefore, even though the baseline probability associated with each event is fixed, individuals experience differing fates due to chance. Additionally, in contrast to the previous methods, the number of infectious, susceptible, and recovered individuals is now required to be an integer—we deal with whole numbers of people, animals, or other organisms.

6.3.1. Basic Methodology

Event-driven methods require explicit consideration of events. For the standard SIR model, we need to consider the six events that can occur, each causing the numbers in the relative classes to increase or decrease by one:

- Births occur at rate μN. Result: $X \to X + 1$.
- Transmission occurs at rate $\beta \frac{XY}{N}$. Result: $Y \to Y + 1$ and $X \to X - 1$.
- Recovery occurs at rate γY. Result: $Z \to Z + 1$ and $Y \to Y - 1$.
- Deaths of X, Y, or Z (three independent events) occur at rate μX, μY, and μZ. Result: $X \to X - 1$, $Y \to Y - 1$ or $Z \to Z - 1$.

There are different ways of implementing this framework, though most practitioners use Gillespie's Direct Method (Gillespie 1977). This scheme first estimates the time until the next event, based on the cumulative rates of all possible events. Then, by converting event rates into probabilities, it randomly selects one of these events. The time and numbers in each class are then updated according to which event is chosen. This process is repeated to iterate the model through time (see Section 6.4.1.1 and Box 6.3). Here, noise affects only the probabilities associated with the fates of individuals and the updating of each consecutive event is independent—there is no assumption concerning environmental stochasticity (e.g., "good" versus "bad" years).

Box 6.3 Event-Driven Approaches: Gillespie's Direct Algorithm

In his seminal 1977 paper, Gillespie outlined two alternative, mathematically equivalent, "stochastic simulation algorithms" (SSAs): the direct method (outlined below) and the first reaction method (Box 6.4). The following pseudo-code provides an efficient and accurate implementation of demographic stochasticity.

1. Label all possible events $E_1, \ldots, E_n$.
2. For each event determine the rate at which it occurs, $R_1, \ldots, R_n$.
3. The rate at which any event occurs is $R_{total} = \sum_{m=1}^{n} R_m$.
4. The time until the next event is $\delta t = \dfrac{-1}{R_{total}} log(RAND_1)$.
5. Generate a new random number, $RAND_2$. Set $P = RAND_2 \times R_{total}$.
6. Event p occurs if

$$\sum_{m=1}^{p-1} R_m < P \leq \sum_{m=1}^{p} R_m.$$

7. The time is now updated, $t \to t + \delta t$, and event p is performed.
8. Return to Step 2.

Event-driven approaches require integer-valued variables and a probabilistic fate of individuals.

6.3.1.1. *The* SIS *Model*

Before discussing a general algorithm (see Box 6.3), we consider the particular problem of an SIS disease without births or deaths, which is a common model for the dynamics of sexually transmitted diseases (Jacque and Simon 1993; Allen and Burgin 2000; Welte et al. 2000; Chapters 2 and 3). The iteration of this model is simplified because it has only two possible events: transmission ($X \to X-1$, $Y \to Y+1$) and recovery ($X \to X+1$, $Y \to Y-1$) . Suppose that at time t we have X susceptibles and Y infecteds, which make up the entire population ($N = X+Y$). The underlying differential equations that inform about the rates at which events occur are:

$$\frac{dX}{dt} = -\beta XY/N + \gamma Y,$$
$$\frac{dY}{dt} = \beta XY/N - \gamma Y.$$

Hence, transmission occurs at rate $\beta XY/N$, recovery occurs at rate γY, and so either event occurs at total rate ($\beta XY/N + \gamma Y$). Letting $RAND_1$ denote a uniform random number between (but not including) 0 and 1, the time until either event occurs is given by:

$$\delta t = \frac{-\log(RAND_1)}{\beta XY/N + \gamma Y},$$

which is equivalent to the assumption used previously (Section 6.2.2) that in a small time period the number of events that occur is Poisson distributed (for a detailed derivation, see for example Bartlett 1957; Gillespie 1977; Renshaw 1991). All that remains is to determine which event occurs. We choose the event randomly, weighted by the relative event rates. Thus, transmission occurs if

$$RAND_2 < \frac{\beta XY/N}{\beta XY/N + \gamma Y},$$

This is online program 6.3

otherwise an infectious individual recovers. Here, $RAND_2$ is another random number between 0 and 1. Finally, time is updated ($t \to t + \delta t$) and the appropriate changes are made to X and Y. The process is then iterated for the prescribed period of time we wish to simulate.

The dynamics of the SIS model serve as an interesting example of the differences between deterministic and stochastic systems. Transmission is frequency dependent whereas recovery occurs at a fixed rate and is independent of the number of infecteds. For the logistic growth model commonly used in ecological problems ($\frac{dx}{dt} = rx(1-x/K)$), the reverse is often considered true, that births are independent of population size whereas deaths are density dependent. In the deterministic setting, the SIS and the logistic growth equations are identical (Chapter 2); therefore, the same dynamics are predicted for both models, asymptoting to a fixed prevalence/population level. However, in a stochastic system, this similarity is lost as the density dependence enters in different ways. The most obvious effect of this is that whereas disease prevalence in the SIS model has an absolute upper bound (the number infectious can never be bigger than the population size), there is no such limit for logistic populations. This is an example of a much more general

phenomenon; although there is a single set of deterministic equations underlying each stochastic model, more than one stochastic model may correspond to each deterministic system (Keeling 2000a). In general, it is usually clear from the context what stochastic model is implied by the deterministic equations, although this does not have to be the case.

6.3.2. The General Approach

Box 6.3 generalizes the event-driven approach using the Direct Gillespie Algorithm (Gillespie 1976, 1977; Renshaw 1991), which can easily be adapted to any disease model. In general, the Direct Algorithm has been used in all event-drive simulations given in this chapter, unless explicitly stated otherwise.

Again, this form of stochastic model shows many of the attributes that we have come to expect: the excitation of oscillations at the natural frequency (Bartlett 1957; Renshaw 1991; Rohani et al. 2002; McKane and Newman 2005; Dushoff et al. 2004); variability in the prevalence, which closely mimics the observed patterns and can have a profound impact on the dynamics (Tille et al. 1991; Keeling and Grenfell 1999; Gibson et al. 1999; Smith et al. 2001); and negative covariances between the numbers of susceptible and infectious individuals. In addition, several elements are peculiar to such event-driven models and their method of iteration, which are discussed below.

This is online program 6.4

6.3.2.1. Simulation Time

Using Gillespie's Direct Method (Box 6.3) leads to a scaling of simulation time with population size. Here, the absolute rates of events increase linearly with population size, N. Intuitively, in a large population, individual transmission and recovery events should be much more frequent than in a small population, reflecting the fact that the per capita rates remain constant. This increase in the transition rates leads to a decrease in the time to the next event (δt) and hence an increase in the number of iterations needed to advance the model a specified period of time. As discussed in Box 6.4, an alternative implementation of this model is Gillespie's First Reaction Method, which is perhaps easier to program than the Direct Method, but substantially slower to use (Figure 6.4).

Box 6.4 Event-Driven Approaches: Gillespie's First Reaction Method

The following pseudo-code provides a slower, but often more intuitive, means of modeling demographic stochasticity; this and the Direct Method (Box 6.3) are equivalent and are both exact analogs of the underlying ODE system.

1. Label all possible events $E_1, \ldots, E_n$.
2. For each event determine the rate at which it occurs $R_1, \ldots, R_n$.
3. For each event, m, calculate the time until it next occurs, $\delta t_m = \dfrac{-1}{R_m} log(RAND_m)$.
4. Find the event, p, that happens first (has the smallest δt).
5. The time is now updated, $t \rightarrow t + \delta t_p$, and event p is performed.
6. Return to Step 2.

With either of these popular implementations of stochasticity, the amount of computer time needed to simulate a particular disease scenario increases linearly with the population size (Figure 6.4). Clearly, this has practical implications for the size of population that can be easily studied with this method, especially when multiple simulations may be needed.

Similarly, simulation of a large epidemic with many cases is slower than simulating a disease close to its endemic level, as many more events occur in the same time period. Some approaches to increasing the computational efficiency of such models is discussed in Box 6.5. As shown in Figure 6.4, Gillespie's recently proposed "τ-leap" method is substantially faster than the previous approaches, with simulation time much less affected by population size.

This is online program 6.5

Box 6.5 Event-Driven Approaches: Efficient Algorithms

The overwhelming drawback associated with using the First-Reaction (Box 6.4) and Direct (Box 6.3) Methods of Gillespie is their slow simulation times for large population sizes (see Figure 6.4). Recently, the first reaction method has been modified by Gibson and Bruck (2000). Their scheme (called the Next Reaction method) is substantially more challenging to program but is significantly faster than even the direct method when there are a large number of different event types. The Direct, First Reaction, and Next Reaction methods are all *exact* stochastic analogs of the underlying ODEs. Gillespie (2001) has recently proposed minor sacrifices in simulation accuracy in order to obtain substantial gains in simulation speed. The new algorithm is called the "τ-leap method" and may be explained as follows using the simple SIS model from Section 6.4.1.1:

1. Let the time increment between steps, δt, be "small" and fixed (discussed below).
2. Let $M_T(t)$ and $M_R(t)$ represent the number of transmission and recovery events by time t.
3. Defining $\delta M_i = M_i(t+\delta t) - M_i(t)$ $(i = T, R)$, then

$$\mathrm{P}(\delta M_T = 1 \mid X, Y) = \frac{\beta XY}{N}\delta t + o(\delta t),$$
$$\mathrm{P}(\delta M_R = 1 \mid Y) = \gamma Y \delta t + o(\delta t)$$

   define the transition probabilities for transmission and recovery events occurring in the time interval δt.
4. For small δt, the increments δM_i are approximately Poisson, such that:

$$\delta M_T \approx \mathrm{Poisson}(\frac{\beta XY}{N}\delta t), \qquad \delta M_R \approx \mathrm{Poisson}(\gamma Y \delta t).$$

5. Now, the variables can be updated:

$$X(t+\delta t) = X(t) - \delta M_T + \delta M_R, \qquad Y(t+\delta t) = Y(t) + \delta M_T - \delta M_R.$$

6. Time is updated, $t = t + \delta t$. Return to Step 4.

The key issue here clearly concerns the size of the fixed integration step, δt. Specifically, how small is "small"? For a range of epidemiological systems we have explored, a "good" value for δt appears to be around $1/10$ day, though this will clearly be inefficient (too small) for very small population sizes and perhaps too large when N is of the order of millions. The leap-size selection question is discussed in some depth by Gillespie and Petzold (2003). Problems arise with this method when multiple events associated with the same individual (infection followed by recovery, or infection followed by transmission to someone else) are likely to occur in the same step; for this reason models with more realistic assumptions about the progression of infection ($SEIR$ or multi-compartmental models, see Section 3.3.1) are more suited to simulation by this technique.

Some event-driven algorithms run much slower as the population size (and in particular the number of infectious individuals) becomes large.

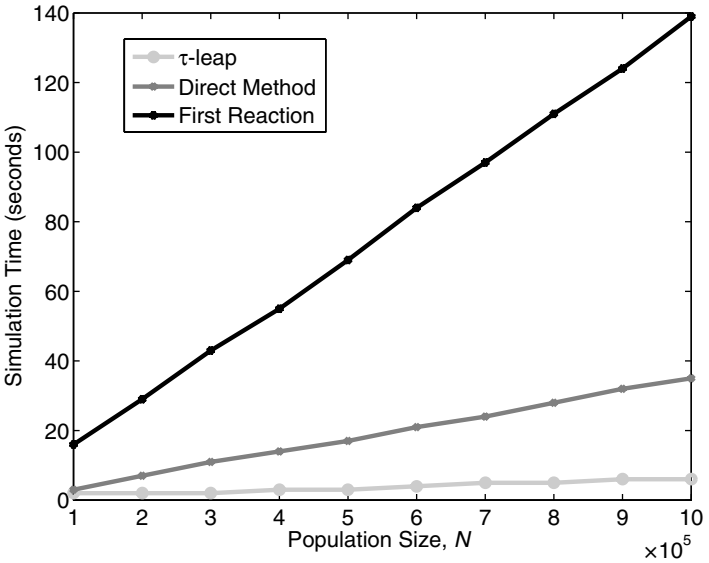

Figure 6.4. The time (in seconds) to simulate 1,000 years of $SEIR$ epidemics ($\nu = \mu = 0.02$ per year, $R_0 = 10$, $1/\sigma = 8$ days, $1/\gamma = 10$ days) on a 3.4 GHz Pentium PC. For Gillespie's Direct (gray line) and First Reaction (black line) methods, as the population size increases so does the simulation time, which, for very large populations, could become prohibitive. In contrast, the τ-leap method (light gray) is fast and largely unaffected by population size.

6.3.3. Stochastic Extinctions and the Critical Community Size

Although extinction of pathogens can happen in noise-based models (Section 6.2), this occurs in a continuous manner and is due to small numbers of infecteds experiencing a run of "bad luck." In contrast, real infectious diseases go extinct because the number of infected individuals drops from one to zero—a stochastic process in which the integer-valued nature of the population is key (Bartlett 1956, 1957; Bolker and Grenfell 1995; Keeling 1997; McKenzie et al. 2001; Dexter 2003; Bozzette et al. 2003).

In deterministic models, when the prevalence of the pathogen becomes low, the infection has a large average growth rate. The same is true for models with demographic stochasticity, except that there is also a chance that the number of infecteds will reach zero. Whether or not a disease is at a high risk of extinction, therefore, depends on several factors. If R_0 is large, then in general there is a large restoring force whenever the number of infectious individuals is low, thus reducing the chance of extinction. Disease extinctions are also much less likely when the average prevalence of the infection is high. Thus, diseases in small populations (Grenfell 1992; Finkenstädt and Grenfell 1998; Bozzette et al. 2003), ones that undergo large amplitude oscillations (see Chapter 5) (Grenfell 1992; Keeling 1997; McKenzie et al. 2001), or diseases with low R_0 (Hagenaars et al. 2001) are the most prone to extinction.

The same reasoning also holds for invading diseases (Bunn et al. 1998). There is a chance that an invading infected individual will recover before passing on the disease to any secondary cases. We can make this argument more mathematically explicit, using what is known as a branching process (Harris 1974; Bartlett 1956; Jacquez and O'Neill

1991; Farrington and Grant 1999; Hagenaars et al. 2001). Suppose that one infectious individual arrives in a large totally susceptible population and let P_{ext} be the probability that the infectious disease goes extinct before it causes a major epidemic (defined as one in which a significant fraction of the population is infected). Initially, one of two events can happen: Either the infectious individual recovers (rate γ) or it causes a secondary case (rate $\beta X/N = \beta$). If the event is recovery, then extinction is guaranteed, otherwise we need to consider the probability of extinction (before a major epidemic) given that two individuals are now infectious. This probability is P_{ext}^2, because it requires the lineages from both transmission events to go extinct independently. The reason the lineages act independently is because neither have yet caused a major epidemic, so the population of susceptibles has not been significantly depleted. Thus:

$$P_{ext} = \frac{\gamma}{\beta+\gamma} \times 1 + \frac{\beta}{\beta+\gamma} \times P_{ext}^2, \tag{6.4}$$

$$\Rightarrow P_{ext} = \frac{\gamma}{\beta} = \frac{1}{R_0}.$$

Therefore, not only is an invading pathogen with a high R_0 difficult to control (Chapters 2 and 8), it also has a low probability of chance extinction.

Similar reasoning and calculations show that if the outbreak starts with n infectious individuals in a population with some immunity, then:

$$P_{ext} = \frac{1}{R^n},$$

where $R = R_0 S = R_0 X/N$ is the effective reproductive ratio. Finally, if the infectious period is a constant interval, rather than the standard assumption that individuals recover at random (Section 3.3), then a similar but more mathematically involved argument reveals that in a totally susceptible population the risk of extinction is:

$$P_{ext} = 1 - R(\infty) < \frac{1}{R_0},$$

where $R(\infty)$ is the final size of the epidemic as predicted by the deterministic equations (Chapter 2). This is an example of a more general phenomenon: Diseases with a more variable infectious period have greater variability in the number of secondary cases, and therefore a greater risk of extinction when infectious numbers are low (Keeling and Grenfell 1997a,b).

From a single infectious individual in a totally susceptible population, the probability of extinction before a major epidemic ensues is equal to $1/R_0$.

A partially resistant population or a highly variable infectious period increases the likelihood of extinction, whereas multiple introductions of pathogen decrease the likelihood.

Although this form of calculation works well for invading diseases when the (initial) growth is exponential, the only way to assess the risk of extinction for endemic diseases is through repeated simulation of the stochastic model. The left-hand graph of Figure 6.5 shows extinction results from demographic (event-driven) stochastic models for an SIR-type infection for various population sizes. As comparison, the right-hand graph gives the corresponding results for measles and whooping cough epidemics in England and Wales

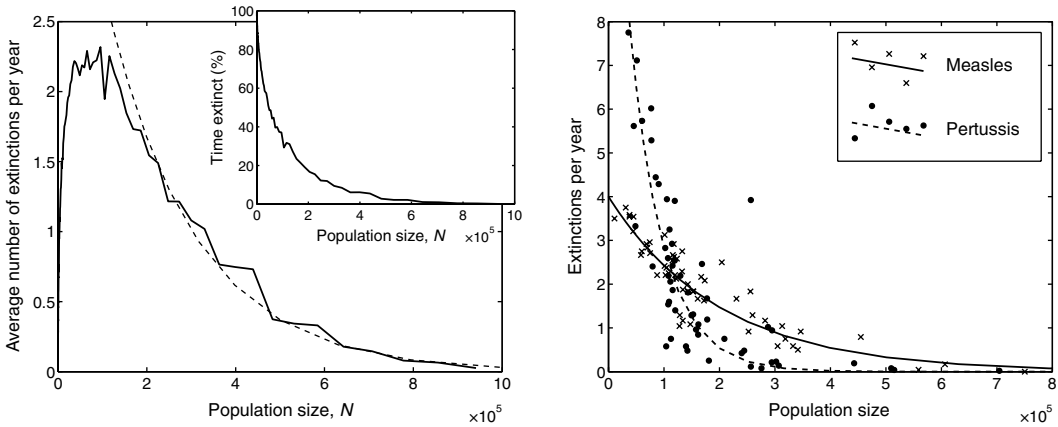

Figure 6.5. The number of extinctions per year from a simple SIR stochastic model (left-hand graph) and from measles and pertussis (whooping cough) data in England and Wales before vaccination (right-hand graph). Exponential functions, $k \sim \exp(-aN)$, are fitted to the point data. The inset graph in the left-hand figure shows how the amount of time the population is disease-free scales with the population size. ($\mu = \nu = 5.5 \times 10^{-5}$ per day, $R_0 = 10$, $\gamma = 1/10$ per day, import rate $\delta = 0.02\sqrt{N}$ per year; see Section 6.3.3.1).

before vaccination. Data and model results all approximate an exponential curve, such that larger populations are less susceptible to stochastic extinctions. The England and Wales data encompasses the effects of age structure (Chapter 3) and seasonality (Chapter 5), which are very strong for these two diseases, whereas the simulated stochastic model ignores such heterogeneities.

Historically, there has been considerable effort in matching the level of persistence predicted by stochastic models to that observed in diseases, with the forced $SEIR$ models suffering much higher levels of extinction (Grenfell et al. 1995; Bolker and Grenfell 1996; Keeling 1997). The principle aim of many stochastic modeling approaches has been to capture the Critical Community Size (see Section 6.3.3.2 for other measures of persistence). The Critical Community Size (CCS) is defined as the smallest population size that does not suffer disease extinction. For measles the CCS has been found to be remarkably consistent, estimated at between 300,000 and 500,000 for England and Wales, the United States, and isolated island communities (Bartlett 1957, 1960; Black 1966). In theory, no population size is ever completely immune from stochastic extinction of the pathogen, but in reality large populations (with substantial imports of infection, see Section 6.3.3.1) tend not to experience disease extinction.

Figure 6.6 shows the approximate CCS (estimated as the population size that experiences one extinction event per year) for the standard SIR model with demographic stochasticity. The infectious period ($1/\gamma$) has the greatest impact on the CCS; diseases with long infectious periods have a much reduced risk of extinction and hence need a much smaller CCS. The basic reproductive ratio, R_0, has a much weaker effect. When the infectious period is short, there is a slight requirement for a larger CCS with increasing R_0; however, when the infectious period is long, smaller R_0 values are associated with a larger CCS. However, to capture the true CCS, the associated stochastic model must also approximate the observed disease dynamics, which for measles or pertussis requires the inclusion of seasonal forcing (Chapter 5) mimicking the opening and closing of schools. This seasonal forcing, and the

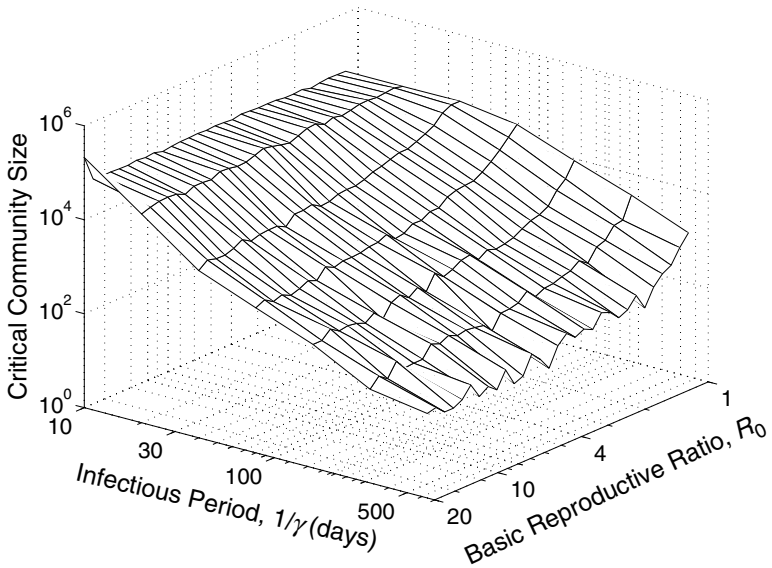

Figure 6.6. The Critical Community Size, approximated as the population size that experiences one extinction event per year, for the stochastic SIR model ($\mu = \nu = 0.02$ per year, import rate $\varepsilon = 0.0213(R_0 - 1)/\sqrt{N}$ per year, this mimics how imports are expected to change with both population size and the average prevalence of infection; see Section 6.3.3.1). Results are from simulations of 100 years after transients.

large-scale epidemics that are generated, leads to frequent extinctions in most stochastic models; the stochastic seasonally forced $SEIR$ model still suffers disease extinctions in populations of 10 million, substantially more than the observed CCS (Bolker and Grenfell 1993). Several modifications have been proposed to increase the persistence of stochastic seasonally forced models, usually involving adding extra levels of heterogeneity such as age structure, spatial structure, or more discrete infectious periods (Bolker and Grenfell 1993, 1995, 1996; Keeling and Grenfell 1997a; Ferguson et al. 1997a; Lloyd 2001).

The question of which ingredients are needed in order to capture the observed measles CCS deserves a little more attention. In Figure 6.7, we plot the extinction frequency as a function of population size for different distributions of latent and infectious periods (denoted by the parameters n and m) in the $SEIR$ model (described in detail in Section 3.3). As shown in the left-hand graph, when identical parameter values are used (notably the amplitude of seasonality, $b_1 = 0.25$), the precise distribution underlying these epidemiological classes does not substantially affect the CCS (cf. Lloyd 2001; Keeling and Grenfell 2002). The figure demonstrates subtle differences when population sizes are small, highlighting the greater extinction frequency of models assuming an exponentially distributed infectious period. As population size increases, however, it is difficult to distinguish between the predictions of models assuming different distributions. Keeling and Grenfell (2002) argued, however, that in order to capture the mechanisms underlying disease persistence, and the CCS in particular, we need to compare "best-fit" models in each case. This means that for any assumed distribution of the latent/infectious period, in order to compare like with like, we need to estimate key parameters such as the amplitude of seasonality by fitting the model to data before examining the predicted CCS. As shown

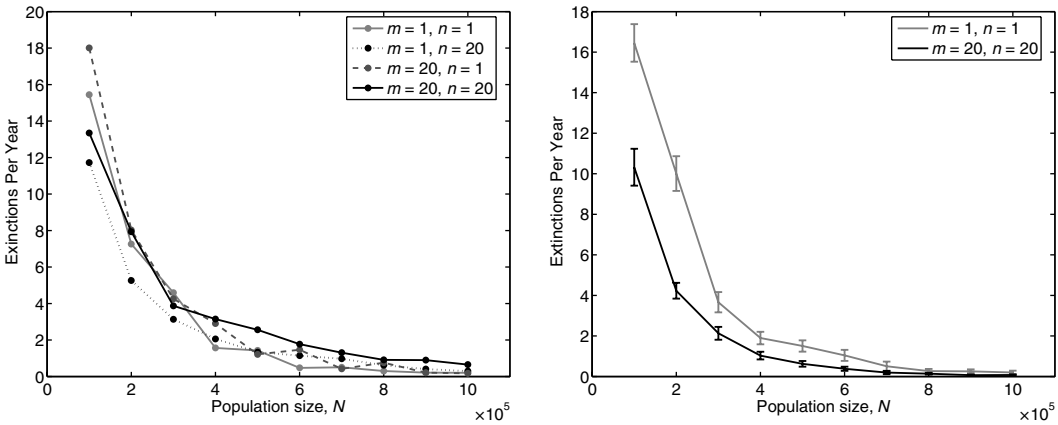

Figure 6.7. The mean number of extinctions per year from 50 stochastic realizations of the $SEIR$ model with different distributions of latent and infectious periods. In the left-hand graph, we plot the mean number of annual extinctions as the number of compartments in the latent and infectious periods are altered. Specifically, we assume both classes conform to a gamma distribution, with different values of the shape parameter, m and n, respectively—simulated using multiple subclasses (Section 3.3). The figure shows that for identical parameter values (specifically the amplitude of seasonality, $b_1 = 0.25$), there is little to distinguish between a purely exponentially distributed model and one containing different distributions. In the right-hand graph, we show that predicted extinction dynamics are significantly different when the best fit model values are used (specifically, $b_1 = 0.29$ for the exponential model and $b_1 = 0.13$ for the gamma distributed model). Error bars demonstrate standard errors. (Other parameter values were $\mu = \nu = 0.02$ per year, $R_0 = 17$, $1/\sigma = 8$ days, $1/\gamma = 5$ days, and the import rate $\delta = 0.02\sqrt{N}$ per year).

in the right-hand graph of Figure 6.7, such an exercise reveals that the exponentially distributed $SEIR$ models are significantly more prone to extinction than models assuming a gamma distribution.

The Critical Community Size (CCS) is the smallest population size observed not to suffer disease extinctions; for measles the CCS is around 400,000. Simple stochastic models that capture the observed seasonally induced measles epidemics fail to show adequate levels of persistence which can be captured only with a greater heterogeneity within the models.

6.3.3.1. The Importance of Imports

Although disease incidence in real populations may frequently undergo stochastically driven fade-outs, imports of pathogen from outside the population can prevent permanent extinction. In human populations, such imports usually take the form of visitors who remain for a limited period but, in doing so, re-introduce the pathogen (Keeling and Rohani 2002). Conceptually, we often think of these visitors as commuters, whose movements between communities prevent the populations from being totally isolated. For wildlife diseases, imports are more likely due to the permanent movement of animals into the group, herd, or area (Clancy 1996), or by exposure to an external source

(Rhodes et al. 1998). Finally, for livestock or crop diseases, imports are due to purchasing infected stock, the mechanical transfer of infected material by humans or vehicles, or by airborne plumes of infection (Cannon and Garner 1999; Ferguson et al. 2001a; Keeling et al. 2001b).

Despite the wide variety of mechanisms, imports are commonly modeled by just two methods. First, by assuming that there is a probability (independent of disease dynamics but dependent on the population size) of an infected individual joining the population:

$$Y \to Y + 1 \qquad \text{Rate } \delta(N).$$

If we also wish to maintain a constant population size, each import can be countered by the loss of either a susceptible or recovered individual. We can equate this mechanism to the permanent immigration of an infected individual.

This is online program 6.6

Second, we may assume there is an external source (again, dependent on the population size) that adds to the force of infection:

$$X \to X - 1, \;\; Y \to Y + 1 \qquad \text{Rate } \varepsilon(N)X.$$

This formulation corresponds either to susceptibles moving to another location and picking up the pathogen prior to returning to their natal population, or to an infectious transient visitor (commuter) transmitting the pathogen in the community but leaving sufficiently rapidly so as not to affect the population size. For this latter method, the rate of imports is governed by the state of the population. Following an extinction, as the number of susceptibles increases, so does the probabilistic risk of imports triggering a new case. Generally, these two modeling approaches produce similar results when correctly scaled ($\varepsilon X \approx \delta \Rightarrow \varepsilon \approx \delta R_0/N$) because the fluctuations in the number of susceptibles are usually relatively small. Interestingly, in human populations, it has been estimated that the number of infectious imports generally scales with the square root of the population size (Bartlett 1957), so that small populations have relatively more imports than large ones.

For human populations, imports of pathogen scale with the square root of the population size.

For measles in England and Wales, by looking at the times between extinctions and invasions, it has been estimated that:

$$\text{Rate of imports} \;\approx\; 5.5 \times 10^{-5}\sqrt{N},$$

so that a population size of one million experiences about 20 infectious imports per year. If we assume that these imports occur due to the movement of infectious individuals (either permanent movements or commuters), we should expect this rate to scale with both the equilibrium prevalence of infection and the transmission rate. By matching to the known parameters and equilibrium levels for measles in England and Wales, we can therefore estimate the expected level of imports for different diseases, although import rates will to some extent depend upon ease of travel, disease morbidity, and the average age of infection.

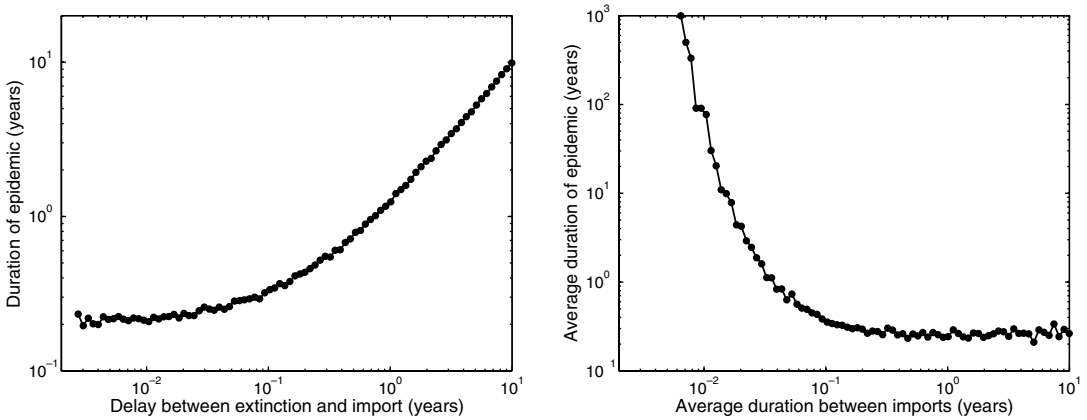

Figure 6.8. Left-hand graph: the average duration of an epidemic caused by an import when there is a fixed delay between extinction and reseeding of infection. Right-hand graph: the average duration of an epidemic caused by an import, when imports are continual and random. ($\nu = \mu = 5.5 \times 10^{-5}$ per day, $R_0 = 10$, $1/\gamma = 10$ days, $N = 10^5$).

$$\begin{aligned}
\text{Rate of imports, measles} &= \delta_{\text{measles}}, \\
\approx\ 5.5 \times 10^{-5}\sqrt{N} &= k\beta I^*_{\text{measles}}, \\
&= k\beta \frac{\mu}{\gamma}\left(1 - \frac{1}{R_0}\right), \\
k &\approx \frac{5.5 \times 10^{-5}\sqrt{N}}{\mu(R_0 - 1)}, \\
&\approx 0.0625\sqrt{N}, \\
\delta_{\text{general}} = k\beta I^* &= 0.0625\mu(R_0 - 1)\sqrt{N}.
\end{aligned}$$

and we therefore expect to see an increase in imports with increasing R_0. We can perform a similar operation assuming commuter-type imports:

$$\varepsilon_{\text{general}} = \frac{1.06\mu(R_0 - 1)}{\sqrt{N}},$$

in which case the expected number of imports ($\varepsilon X \approx \varepsilon N / R_0$) shows little variation with R_0.

Extinctions are much more likely to occur if the number of susceptibles has been reduced by recent stochastic fluctuations, such that the effective reproductive ratio, R, drops below one. In such cases, an immediate reseeding of infection is also very likely to fail. Thus, the *frequency* of imports has a profound effect on the success of the subsequent epidemics. Figure 6.8 (left-hand graph) shows how the average duration of the ensuing epidemic increases with the delay between extinction and reseeding, because this allows the level of susceptibles to rise. The increase in epidemic duration with delay is slightly less than linear, and even with a very rapid reintroduction of infection sustained epidemics can still arise.

Although imports are noticed only once the disease is extinct, they are in fact a continual process happening even when the disease is present. These continual imports can substantially lessen the risk of extinction by increasing the average level of infection and rescuing the disease from times of low prevalence. Continual imports can also interfere with the disease dynamics, damping any fluctuations (Ferguson et al. 1996a) or even supporting infection when it would normally die out (Rhodes et al. 1998). The right-hand graph of Figure 6.8 illustrates this point. If a population has many imports (e.g., when a small town is tightly coupled to a large city (Finkenstädt and Grenfell 1998)), extinctions are rare. When imports are less frequent, the risk of extinction during an epidemic quickly reduces to an asymptotic level. Thus, although infrequent imports may allow the susceptible population to recover between epidemics, they do little to support the persistence of infection.

Imports of infection into a community can significantly modify the extinction process in a complex manner. Frequent imports may prevent extinctions, whereas long delays between extinctions and imports improve the chance that an import successfully triggers a large epidemic.

In many structured models (see Chapters 4 and 7), the early dynamics of infection in a new population or subpopulation are defined by two attributes; the rate that infection arrives in the population, and the subsequent growth within that population. In deterministic (continuous population) models, these imports occur continually at very low levels. Therefore, minute fractional imports can be rapidly increased by the within-population dynamics, leading to a large epidemic despite the fact that less than one infectious individual has arrived in the population. The stochastic framework overcomes this problem, because an entire individual needs to be imported before an epidemic can occur. Thus, when the coupling between populations is low, the stochastic formulation can substantially increase the delay between an epidemic in one population and the epidemics that are triggered in other coupled populations (see Chapter 7 and in particular Figure 7.8).

When dealing with very low levels of imports between populations, it is essential that this is modeled stochastically.

6.3.3.2. Measures of Persistence

Measuring persistence, or the lack of it, in models is far from straightforward. There are three main measures, each of which has its advantages and pitfalls:

1. *Extinctions with Imports*. The most biologically plausible measure of persistence is to simulate the stochastic dynamics of the population, including the random import of infection from external sources, and count the average number of extinctions during a given period (Bartlett 1957; Grenfell 1992; Keeling 1997). Figure 6.5 uses this approach. The main advantage is that it corresponds to biological reality, and therefore the results can be compared to observations. However, as shown above, the persistence of diseases is strongly influenced by the pattern of imports and in general these are very difficult to measure or parameterize.
2. *Time to Extinction*. A more mathematically pleasing approach is to start the stochastic simulations at, or near to, the deterministic equilibrium and measure the average time until the population goes extinct—called the "first passage time." Alternatively,

one could measure the proportion of simulations that have gone extinct after a given time (Farrington and Grant 1999). The appeal of this method is that imports are not required—populations act in isolation. However, this measure does not conform to any observed biological process and the results are sensitive to the initial conditions.

3. *Conditional Extinctions*. An alternative measure is to look at the asymptotic rate of extinctions, conditional on the disease currently being present in populations without imports. In practice, this is calculated by first simulating populations for many generations and then discarding those in which the pathogen has already died out. The remaining simulations, which still possess infection, can then be simulated further to ascertain the rate of extinction. The initial run-in time followed by discarding means that the calculations start from a natural population distribution due to the stochastic dynamics, rather than the somewhat artificial starting condition of Method 2. This measure is theoretically appealing and is often used whenever analytical results are feasible (Nasell 1999; Andersson and Britton 2000).

6.3.3.3. Vaccination in a Stochastic Environment

One of the primary aims of epidemiological modeling is to inform public health intervention measures. The relationship

$$V_C = 1 - \frac{1}{R_0}$$

between the critical level of vaccination needed to eradicate an infectious disease, V_C, and the basic reproductive ratio, R_0, is a fundamental tenet of modern epidemiology (Chapter 8). For stochastic populations this threshold still applies; with vaccination coverage at or above V_C, the disease cannot invade or cause an epidemic. However, stochastic dynamics can work against the pathogen by causing extinctions even though the vaccination level is below the required threshold (Bolker and Grenfell 1996; Smith et al. 2001).

Figure 6.9 considers the effects of vaccinating a proportion of children at birth, such that the underlying differential equations would be:

$$\begin{aligned}
\frac{dX}{dt} &= \nu N(1-V) - \beta XY/N - \mu X, \\
\frac{dY}{dt} &= \beta XY/N - \gamma Y - \mu Y, \\
\frac{dZ}{dt} &= \nu NV + \gamma Y - \mu R.
\end{aligned}$$

Vaccination (below the eradication threshold) has two main effects. As predicted by the deterministic model, vaccination reduces the average level of infection within the population, which in itself has clear benefits. In addition, associated with the lower prevalence is an increased risk of extinction. Thus, as vaccination level is increased, the disease is likely to be eliminated (as least within the local population) by stochastic extinctions before the deterministic threshold is reached. Hence, stochasticity actually benefits control programs, allowing diseases to be eradicated at lower vaccination levels (Smith et al. 2001); however, see Bolker and Grenfell (1996) for a plausible counterexample. Note, however, that the threshold level of vaccination must be achieved if reinvasion of infection is to be prevented.

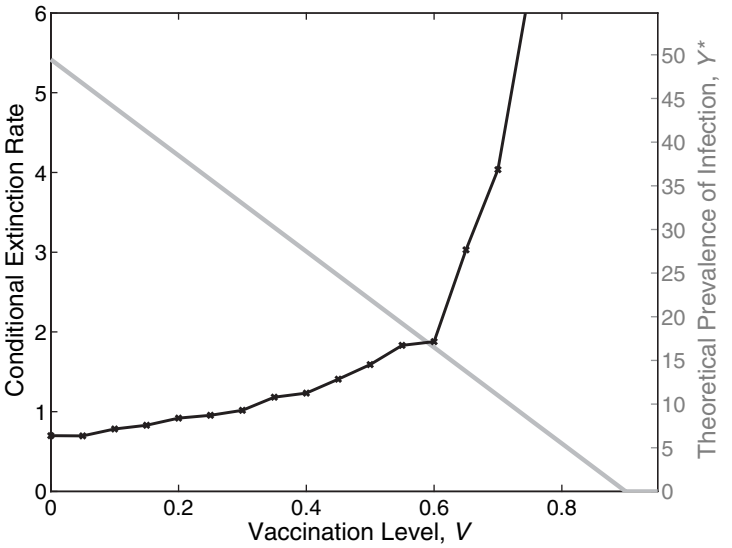

Figure 6.9. The dual effects of vaccination. The conditional extinction rate per year (black line, see Section 6.3.3.2) shows a sharp increase as the deterministically predicted level of infection (gray line) decreases. ($\nu = \mu = 5.5 \times 10^{-5}$ per day, $R_0 = 10$, $1/\gamma = 10$ days, $N = 10^5$.)

In many situations, stochasticity may be beneficial, leading to extinction of a disease well before the vaccination threshold is reached.

6.3.4. Application: Porcine Reproductive and Respiratory Syndrome

One application where population sizes are often relatively small and therefore stochasticity can play a major role is for epidemics within a farm (Stark et al. 2000; MacKenzie and Bishop 2001). Although some farms may have tens of thousands of livestock, even at these numbers stochasticity and chance extinctions can be important. More often, the number of animals is far lower and therefore stochasticity all the more important. Livestock epidemics are becoming an increasingly active area of research, motivated in part by the 2001 foot-and-mouth epidemic in the United Kingdom, the 1997 classical swine fever outbreak in the Netherlands, and the risk of agro-terrorism. For diseases, such as foot-and-mouth, that spread rapidly within the farm it may be plausible to ignore the within-farm dynamics treating the farm as a single unit (Keeling et al. 2001b). However, for diseases with a more prolonged infection and slower spread, the dynamics within a farm may play a critical role in the wider transmission of pathogen.

The example discussed here is Porcine Reproductive and Respiratory Syndrome (PRRS), a viral disease of pigs (Albina 1997; Wills et al. 1997). PRRS was first observed in the United States in 1986, but was only identified following an outbreak in the Netherlands in 1991 when it spread rapidly through the pig-producing areas of the country (Terpstra et al. 1991). Since then, PRRS has been endemic within both the United States (prevalence 60–80%) and the Netherlands. PRRS has economic and welfare implications for the pig industry because it may reduce the reproductive success of sows and cause respiratory problems in piglets (Terpstra et al. 1991). Clinical signs may include

anorexia, fever, or lethargy together with cyanotic (blue) ears, vulva, tails, abdomens, or snouts—giving rise to the disease's alternative name, blue-eared pig disease. Reproductive failure is characterized by late-term abortions, increased numbers of stillborn fetuses, and/or premature weak pigs.

The initial PRRS epidemic spreads slowly within a farm, after which the infection may persist at low levels for many years. This slow spread and low prevalence means that within-farm modeling and a stochastic approach are vital. Nodelijk et al. (2000) modeled this disease based upon the observed dynamics on a closed breeding-to-finish herd of around 115 sows belonging to the University of Utrecht. Blood samples were taken from all sows just before the 1991 outbreak, again during the initial epidemic, and then twice yearly after that; all sera were tested for antibodies to the PRRS virus. It was shown that the observed pattern of seropositives matched SIR-type dynamics with waning immunity, given that imports and exports of livestock were included.

$$\begin{aligned} \frac{dX}{dt} &= \mu N + wZ - \mu X - \beta XY, \\ \frac{dY}{dt} &= \beta XY - \gamma Y - \mu Y, \\ \frac{dZ}{dt} &= \gamma Y - wZ - \mu Z, \end{aligned} \tag{6.5}$$

where, as usual, β is the contact parameter and γ is the recovery rate. (Note that although we have chosen to model transmission as density dependent, the fact that the number of pigs on a farm remains constant makes this assumption irrelevant.) Other parameters are μ, the replacement rate of livestock and w, the rate at which immunity wanes and the pigs revert to being susceptible. The longitudinal data collected were sufficient to parameterize this model; most notably R_0 was estimated to be 3.0 (95% confidence interval 1.5 to 6.0), such that in a deterministic setting the disease would always take off and readily persist.

Breeding-to-finish pig farms, such as the one in the study by Nodelijk et al. (2000), possess additional structure in that sows and rearing pigs are often physically separated. Nodelijk et al. (2000) incorporated this extra detail into their model; however, for clarity we take the simpler approach that the farm consists of one freely mixing herd. For other farm types that specialize in a single age range of pigs, this simplifying assumption is more likely to be true. Figure 6.10 (top row of graphs) shows the distribution of the number of cases (left-hand graph) and duration (right-hand graph) of a series of simulated outbreaks on a farm with 115 sows ($N = 115$). The number of cases (including multiple reinfections of a given pig) is bimodal in nature because some seed infections fail to trigger a significant outbreak. As stated earlier, for three initial infections we would expect a fraction $R_0^{-3} \approx 0.037$ to fail—this is lower than the value from simulations (≈ 0.057), which is attributable to the small population size. There is also a significant probability that the epidemic will be both very large and protracted, because the distributions of both the size and duration have long tails. It is important to recognize that no single measure can encompass the risks described by these distributions, thus it is vitally important that, at the very least, both means and some measure of the variation are always stated.

This model can easily be extended to consider the impact of herd size on the stochastic epidemic dynamics. As expected from the theoretical results shown above and the work on extinctions of measles in human communities (Bartlett 1956, 1957), the average duration and hence the size of epidemics increases dramatically with the number of pigs. For farms

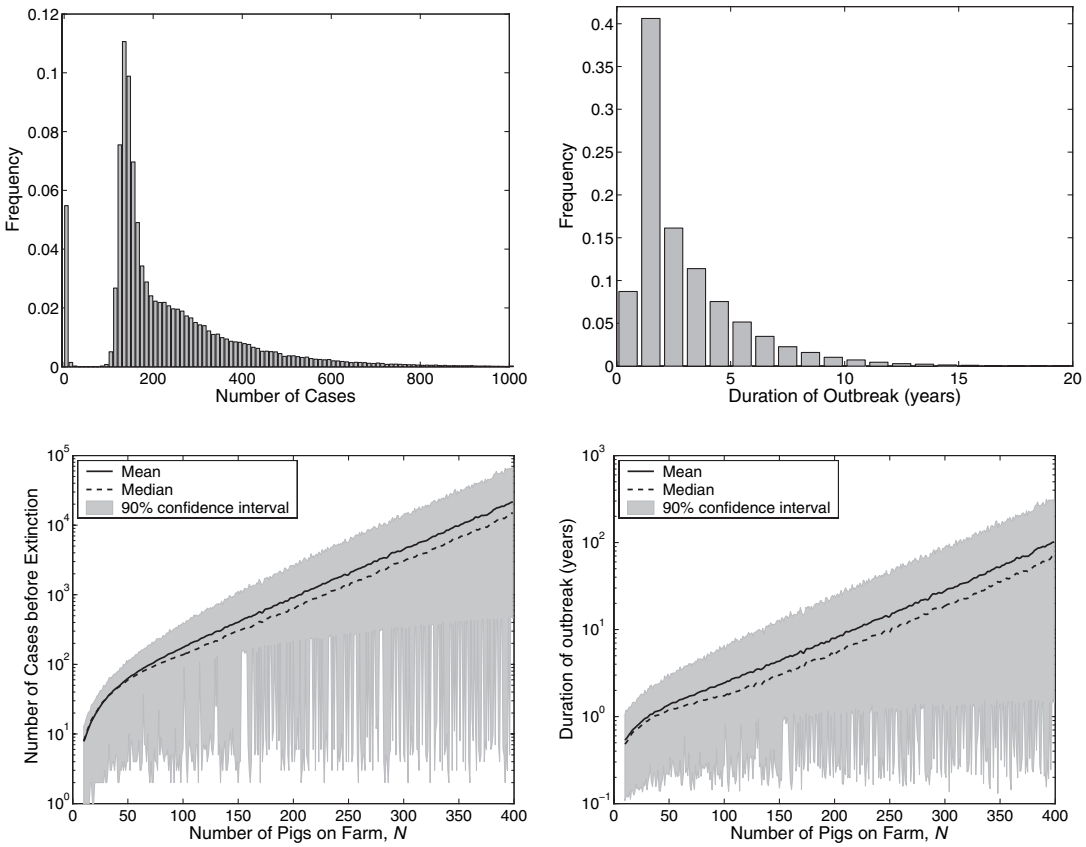

Figure 6.10. The simulated behaviour of PRRS epidemics with event-driven (demographic) stochasticity. The top left-hand graph shows the distribution of the total number of cases, including reinfections, from an introduction of 3 infected animals in a farm with 115 sows. The top right-hand graph shows the duration for a similar scenario. Results are from 100,000 simulations. The lower graphs consider how these quantities change with the number of animals on the farm; both show an exponential increase with the number of sows. ($R_0 = 3$, $\gamma = 6.517$ per year, $\mu = 0.6257$ per year, and $w = 0.2607$ per year). Shaded regions give the 90% confidence intervals associated with a single epidemic.

with more than 50 pigs, the increase in both cases and duration is exponential, suggesting that very large farms can act as persistent reservoirs of infection. In contrast, very small farms cannot sustain the epidemic and the size and duration of the outbreak is even smaller than would be expected by extrapolating from larger farms. Again, the distribution of epidemic sizes and hence the confidence intervals on the model predictions provide useful additional information. It appears that the mean number of cases and duration of the outbreak are dominated by the long tail of both distributions, whereas the lower 5% bound is very noisy but shows little variation with the number of livestock.

These results have clear implications for the farming community. Even very large pig farms may often have very short, limited epidemics that last for less than a year and cause few cases. Such epidemics are extinguished rapidly by stochastic effects due to the small number of infectious animals that seed the epidemic. However, once the epidemic becomes

established it can persist for many years in the larger farms. Thus, these farms should be the target of control measures, because smaller farms are much less likely to sustain the virus and consequently less likely to spread PRRS. Such considerations may also be appropriate for diseases such as bovine tuberculosis, which spreads slowly between cattle with a low R_0, where farm-level dynamics and the persistence in large premises may be key to the national pattern of cases.

6.3.5. Individual-Based Models

Although all the models of demographic (event-driven) stochasticity consider the population to be composed of individuals, they are not strictly individual-based models.This type of model is dealt with in some detail in Section 7.5, though here we focus on the differences between stochastic and individual-based models as well as highlighting a few "tricks" to speed the computation of such models.

Models with demographic stochasticity force the population to be made up of individuals, but all individuals of the same type are compartmentalized and measured by a single parameter—there is no distinction between individuals in the same class. In contrast, individual-based models monitor the state of each individual in the population. For example, in models with demographic stochasticity we know how many individuals are in the infectious class but not *who* they are, whereas in individual-based models we know the state of each individual and hence *which* individuals are infectious. Individual-based models are therefore often more computationally intensive (especially in terms of computer memory) because each individual in the population must have its status recorded.

Although individual-based models can be readily iterated with any of the standard algorithms (see Boxes 6.3, 6.4, and 6.5), there can be computational difficulties due to the vast number of different events—we need a different event for each individual concerned, especially for the fast τ-leap method where a Poisson distribution needs to be realized for each event. A variety of methods can be used to overcome this difficulty; here we outline two that are readily usable.

One solution is to first use either the Gillespie direct method (Box 6.3) or the τ-leap method (Box 6.5) with the total rate of each event (i.e., combining the rates of all infection events), and then determining which individuals were involved in the event. For the SIS model outlined in Section 6.3.1.1, the individual-based equivalent would be:

$$\begin{aligned}\delta t &= \frac{-\log(RAND_1)}{\text{sum of all infection and recovery rates}},\\ &= \frac{-\log(RAND_1)}{\sum_{i=S} T_i + \sum_{i=I} G_i},\end{aligned}$$

which gives the time to the next event. Here T_i and G_i are the transmission and recovery rates of individuals at that moment, and the two sums select only those individuals who are susceptible and infectious, respectively. As before, we determine that the next event would be transmission if:

$$RAND_2 < \frac{\sum_{i=S} T_i}{\sum_{i=S} T_i + \sum_{i=I} G_i}.$$

Suppose the next event is recovery; we then let individual p recover if:

$$\sum_{i=I}^{i\leq p-1} G_i < RAND_3 \times \left(\sum_{i=I} G_i\right) \leq \sum_{i=I}^{i\leq p} G_i. \tag{6.6}$$

This latter step can be greatly simplified if all the recovery (or transmission) events occur at the same rate. Although this modification is formulated for Gillespie's direct method, it can be equally applied to the τ-leap method, with the individual involved in each event determined by equation (6.6). One potential difficulty with this method is that for each event, an additional random number needs to be calculated to determine which individual is involved.

As an alternative—which works particularly well when the time since infection is important (Section 3.3)—we can combine the Gillespie direct and first reaction methods (Boxes 6.3 and 6.4). We again use the example of the SIS model, but suppose that the times to recovery, δG, for each individual are determined as soon as they are infected. The time to the next recovery event is therefore $\min(\delta G_i) = \delta G_{\min_G}$ (where $\min_G$ refers to the individual that will recover first), whereas the time to the next infection is:

$$\delta t = \frac{-\log(RAND_1)}{\sum_{i=S} T_i}.$$

The time to the next event is then the minimum of these two values. If recovery happens first, the individual with the shortest time ($\min_G$) recovers, all the recovery times are reduced by $\delta G_{\min_G}$, and a new minimum recovery time is calculated. If infection happens first, then the individual involved can be calculated as above, all the recovery times are reduced by δt, and a new minimum recovery time is calculated *if* the newly infected individual is predicted to recover before any others.

This concept of storing a recovery time for each individual is very powerful, and can be readily extended to a death time, both of which allow for more realistic assumptions than the standard exponential distribution that arises from constant rates. However, we can also utilize such a method to predict transmission. For each individual we can set a cumulative level of pathogen, C_i, that they need to encounter before they are infected. In a model with frequency-dependent transmission (and equal levels of susceptibility and transmissibility per individual), the cumulative pathogen levels would be chosen from an exponential distribution:

$$C_i = -\log(RAND)$$

and at each time step, say δt, these values get reduced by $\delta t \beta Y/N$. Transmission then occurs when an individual's C_i value drops below zero. Using this approach, the fate of an individual can be determined by the selection of random numbers (from appropriate distributions) at their birth—which may limit the number of times a random number generation routine is needed. However, there is an additional benefit. For normal methods of updating, even when starting with the same random number seed, a small deviation in the epidemic patten (due to slightly different parameters) can initiate major changes because the same stream of random numbers will inevitably be applied to different processes. In contrast, when the fate of an individual is set at birth, the use of the random numbers is independent of the epidemic dynamics. This provides us with a fascinating opportunity to test the effects of parameters or controls on the *same* stochastic epidemic process. Rather than having to phrase all comparisons in terms of averages, we can now test controls on

replicated epidemics. Hence, we may provide definitive answers to questions such as, "If this exact epidemic occurred again, would we be better to vaccinate"?

6.4. PARAMETERIZATION OF STOCHASTIC MODELS

As with all epidemiological problems, accurate parameterization is the key to useful and applied modeling. For stochastic systems, the parameterization approach can often be made much more statistically rigorous. In deterministic settings, parameters are generally chosen that minimize the deviation between the observed and simulated epidemics. However, this minimization requires many arbitrary choices (e.g., the weighting between errors in susceptibles and errors in infecteds, or whether the absolute or proportional errors are used), which can influence the choice of parameters. The simplest fitting procedure chooses parameters so that the deterministic equilibrium predictions match the average observed dynamics. In general, this may often give a reasonable approximation of the true parameter values; however, as we have shown in this chapter already (Section 6.2.1.1), stochasticity can have a significant effect on the mean and thus bias even this simple form of estimation. For example, the negative correlations that develop between susceptible and infected individuals would lead to an underestimation of the basic reproductive ratio, R_0, if the standard deterministic parameterization was used. This problem becomes significantly worse when localized extinctions are frequent.

By including stochasticity in our models, an alternative methodology can be used. Parameters can be found that maximize the *likelihood* (or probability) of the model generating the observed data (Finkenstädt and Grenfell 2000). Such maximum likelihood estimates are a well-defined and well-understood statistical tool. In general, the likelihood is calculated sequentially. Thus the probability, P_t, of observing the data recorded at time t is calculated given that the model agrees with the observations recorded at time $t-1$. The total likelihood is then found from multiplying together all the probabilities. Often these likelihoods will be very small, because the probability of seeing exactly the same number of cases at exactly the same times is highly unlikely. These sort of likelihood calculations lend themselves naturally to an MCMC (Monte Carlo Markov Chain) approach (Ranta et al. 1999; O'Neill et al. 2000; Neal and Roberts 2004), which considers all the plausible stochastic pathways between two observations.

6.5. INTERACTION OF NOISE WITH HETEROGENEITIES

In the above examples, the differences between the deterministic and stochastic models were due either to the noise itself, or the interaction of noise with the nonlinear transmission dynamics. This section considers the interaction of noise with other components of disease models, such as temporal forcing, age, or risk structure and spatial heterogeneities.

6.5.1. Temporal Forcing

As discussed more fully in Chapter 5, temporal forcing can occur for a variety of reasons—generally due to climatic variations or changes in social mixing (such as the differences between school terms and holidays). The main effect of such regular forcing is to induce periodic cycles in disease prevalence. Even when the forcing is annual, multi-annual cycles are common (Keeling et al. 2001a).

However, as discussed in this chapter, the inclusion of stochasticity can often cause the disease to resonate at its natural frequency due to the greater role of transient dynamics. In many cases, a clash can occur between the deterministic period due to forcing and the natural resonant period due to stochasticity (Rohani et al. 2002). Both measles and whooping cough experience strong seasonal forcing due to the school year. The deterministic attractor for measles is biennial, which is very close to the natural resonant period; hence, there is no conflict between the two frequencies. In contrast, whooping cough possesses an annual deterministic attractor whereas the resonant period is closer to three years. This tension between the two periods is observed in the number of reported cases; large populations (which experience relatively little demographic stochasticity) display predominantly annual epidemics, whereas small populations are more likely to experience epidemics every 2–3 years in keeping with the stochastic concept of increased transient-like dynamics (Rohani et al. 2002). This phenomenon is discussed in more detail in Chapter 5 on temporal forcing.

6.5.2. Risk Structure

Risk- and age-structured models subdivide the population into groups of individuals, such that all individuals within a group experience a similar risk of infection (see Chapter 3). The classic example is of sexually transmitted diseases where some individuals, due to their lifestyle, are at a much greater risk of catching and transmitting infection. Such core groups will also be subject to different levels of stochasticity. Although all the models and methods developed in this section transfer directly to such structured populations (Sleeman and Mode 1997; MacKenzie and Bishop 2001), the subtle differences compared to unstructured models needs to be addressed.

We have already seen that in a stochastic environment new infections can fail to invade and endemic infections can be driven extinct. How does structuring the population into risk-groups affect this? The high-risk population is smaller, suggesting that it experiences more stochastic effects, but the reproductive ratio within this group is larger suggesting reduced chance of extinction. Simulation of structured and unstructured SIS-type models with event-driven stochasticity is shown in Figure 6.11. When the models share the same R_0, the structured model shows greater variability relative to the disease prevalence. However, if the two models were parameterized to have the same equilibrium prevalence of infection, then the reverse is true and the unstructured model is more variable. This illustrates a crucial difficulty when comparing models of different forms; the comparison between the models is sensitive to which characteristics are matched. In practical situations, the available data should always constrain our choice of model and parameters (Keeling and Grenfell 2002). If data exists on both the early growth rate (hence R_0) and the equilibrium prevalence, it may be impossible to fit an unstructured model, in which case no fair comparison can be made.

In some cases, the presence of extra structure will always increases the level of variability. Consider a population that has a few very rare super-spreaders or super-shedders. On average, these super-infectious individuals are not likely to be infected, but in those rare simulations where one of them does become infected, the prevalence of the disease can increase dramatically. Such rare events with large consequences will clearly increase the variability between model simulations and reduce our ability to predict the outcome and hence control the disease in real situations.

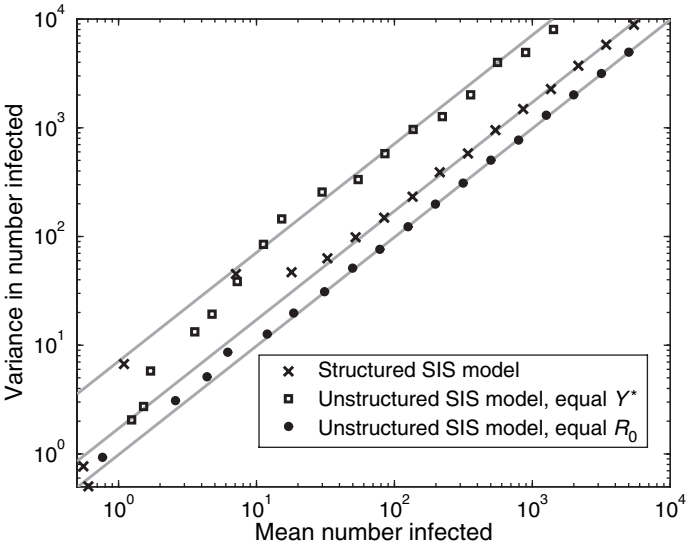

Figure 6.11. A log-log plot of the variance against the mean number of infected individuals for a SIS disease in various population sizes. Randomness was included as individual-based event-driven demographic stochasticity. The structured population has 20% of its individuals in the high-risk group, and assumes the following transmission matrix which was used throughout Chapter 3: $\boldsymbol{\beta} = \begin{pmatrix} 10 & 0.1 \\ 0.1 & 1 \end{pmatrix}$, $\gamma = 1, \delta = 10^{-3} \times N$. The unstructured population has the same basic reproductive ratio (circles $\beta \approx 2.0$), or the same equilibrium infection level (squares $\beta \approx 1.158$), as the structured model.

6.5.3. Spatial Structure

Spatial structure (Chapter 7), the subdivision of the population due to geographical location, is another situation where stochasticity plays a vital and dominant role. Without underlying heterogeneities in the parameters at different locations, most deterministic models asymptote to a uniform solution at all locations, thereby negating the necessity of a spatial model. However, in a stochastic setting, different spatial locations experience different random effects, and spatial heterogeneity may therefore be maintained.

The dynamics of spatially structured models is discussed more fully in Chapter 7. However, one facet of the interaction between spacial partitioning and stochasticity is worth emphasizing here. For the same total population size, does one large homogeneously mixed population suffer more or fewer stochastic effects than many smaller weakly interacting populations? The answer is that it very much depends on the nature of the disease and the host population (Keeling 2000b). However, in many situations a spatially structured population suffers less stochastic effects (and extinctions) than a large well-mixed population, despite the fact that each spatial subunit experiences a high level of stochasticity due to its small size (Keeling et al. 2004a; Section 7.2.3).

6.6. ANALYTICAL METHODS

Although the methods given above provide a robust means to simulate the stochastic behavior of a disease, they lack the analytical tractability that provides an intuitive understanding of the behavior. The following three formulations provide this analytical link, but require a certain level of mathematical sophistication. Hence, some readers may wish to skip the details of this section and read only the highlighted conclusions. In addition, all three methods provide a deterministic description of stochastic processes by considering the distribution or average of many stochastic realizations. In all the following examples, we shall focus primarily on the SIS model (without births or deaths) because this model is only one-dimensional (as $X + Y = N$), which makes the mathematics simpler and more transparent.

6.6.1. Fokker-Plank Equations

These equations are closely linked to the addition of process noise to the standard equations introduced in Section 6.2 of this chapter. Rather than simulate multiple realizations of the same set of noisy equations, it is possible to derive equations for the probability distribution of values (Bartlett 1956; van Herwaarden 1997; Andersson and Britton 2000). Thus $P(X, Y, t)$ is the probability of a stochastic simulation having X susceptible individuals and Y infectious individuals at time t. In deterministic models, the probability distribution is zero everywhere except a single point, and the dynamics of this point obey the standard differential equations. The key question is how noise should be introduced. Intuitively, noise should lead to the probability distribution spreading out, such that the more noise that is experienced, the greater the spread of values around the deterministic solution that are likely to be encountered.

Standard mathematical theory allows us to make this relationship more exact (Øksendal 1998). For a very general differential equation together with additive noise:

$$\frac{dx}{dt} = F(x) + f(x)\xi,$$

the probability distribution $P(x, t)$ is given by the partial differential equation:

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}(F(x)P(x,t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}(f(x)^2 P(x,t)).$$

Hence, noise in the stochastic simulations is replaced by diffusion in the partial differential equations.

The use of such partial differential equations has three advantages over noisy simulations: (1) it needs to be solved only once to capture the entire range of possible dynamics, (2) by setting $\frac{\partial P}{\partial t} = 0$ we can readily find the long-term stationary distribution of values, and (3) the model is deterministic and therefore does not rely upon a time-consuming random number generation routine. However, these three advantages must be offset against the additional difficulty of solving partial differential equations, which in itself is a computationally demanding task.

The Fokker-Plank equation provides an *exact* solution to multiple simulations of a model with noise. Noise in the equations translates to diffusion of the probability distribution.

We now extend this concept to a more epidemiologically relevant situation, the dynamics of an SIS model of the sort commonly used to describe the behavior of sexually transmitted diseases; a similar form of model was used by Roberts and Saha (1999) to study bovine tuberculosis in the possum population of New Zealand. The differential equation for the number of infected individuals, with scaled additive noise, becomes:

$$\begin{aligned}\frac{dY}{dt} &= [\beta XY/N + \sqrt{\beta XY/N}\xi_1] - [\gamma Y + \sqrt{\gamma Y}\xi_2] \\ &= \beta XY/N - \gamma Y + \sqrt{\beta XY/N + \gamma Y}\xi.\end{aligned}$$

Here the two noise terms have been combined by noting that the sum of two normal distributions with variances v_1 and v_2 is another normal distribution with variance $v_1 + v_2$. An equation for X is unnecessary because $X = N - Y$. The probability, P, of having exactly Y infectious individuals is given by:

$$\frac{\partial P(Y)}{\partial t} = -\frac{\partial}{\partial Y}([\beta XY/N - \gamma Y]P(Y)) + \frac{1}{2}\frac{\partial^2}{\partial Y^2}([\beta XY/N + \gamma Y]P(Y)).$$

This expression provides an exact solution to the distribution obtained from infinitely many simulations of the SIS model with scaled additive noise (Figure 6.12).

If we are primarily concerned with the long-term behavior of the pathogen, rather than initial dynamics, then an analytical expression for the equilibrium distribution $P^*(Y)$ can be found by setting $\frac{\partial P}{\partial t} = 0$:

$$\begin{aligned}\Rightarrow [\beta XY/N - \gamma Y]P &= \frac{1}{2}\frac{\partial}{\partial Y}([\beta XY/N + \gamma Y]P), \\ &= \frac{1}{2}\frac{\partial}{\partial Y}([\beta(N-Y)Y/N + \gamma Y]P), \\ \Rightarrow \qquad \frac{\partial P}{\partial Y} &= \left[\frac{2\beta XY - 2\gamma YN - \beta(N-2Y) + \gamma N}{\beta XY + \gamma YN}\right]P.\end{aligned}$$

This is a linear equation in P, which although complex can be solved numerically with considerable ease, Figure 6.12 (lower graphs).

Although this method of understanding the role of noise can be extended to the traditional SIR-type models (Clancy and French 2001), it necessitates a two-dimensional probability distribution (specifying both X and Y), which requires much more sophisticated tools to integrate. Therefore, we will leave the formulation and solution of such models as an exercise for the keen and mathematically able. However, as computational power continues to advance, it may be preferable to solve the exact PDE equations—which capture even the rarest events—rather than perform multiple simulations of the differential equations with noise.

6.6.2. Master Equations

Master equations (also known as Ensemble or Kolmogorov-forward equations) are the integer-valued and event-driven equivalent of the Fokker-Plank equations described above. Essentially, they require the formulation of a separate differential equation for the probability of finding the population in every possible state (Stollenwerk and Briggs 2000; Stollenwerk and Jansen 2003; Viet and Medley 2006). For example, $P_{X,Y}(t)$ is the probability of having X susceptibles and Y infected, where X and Y are both integers,

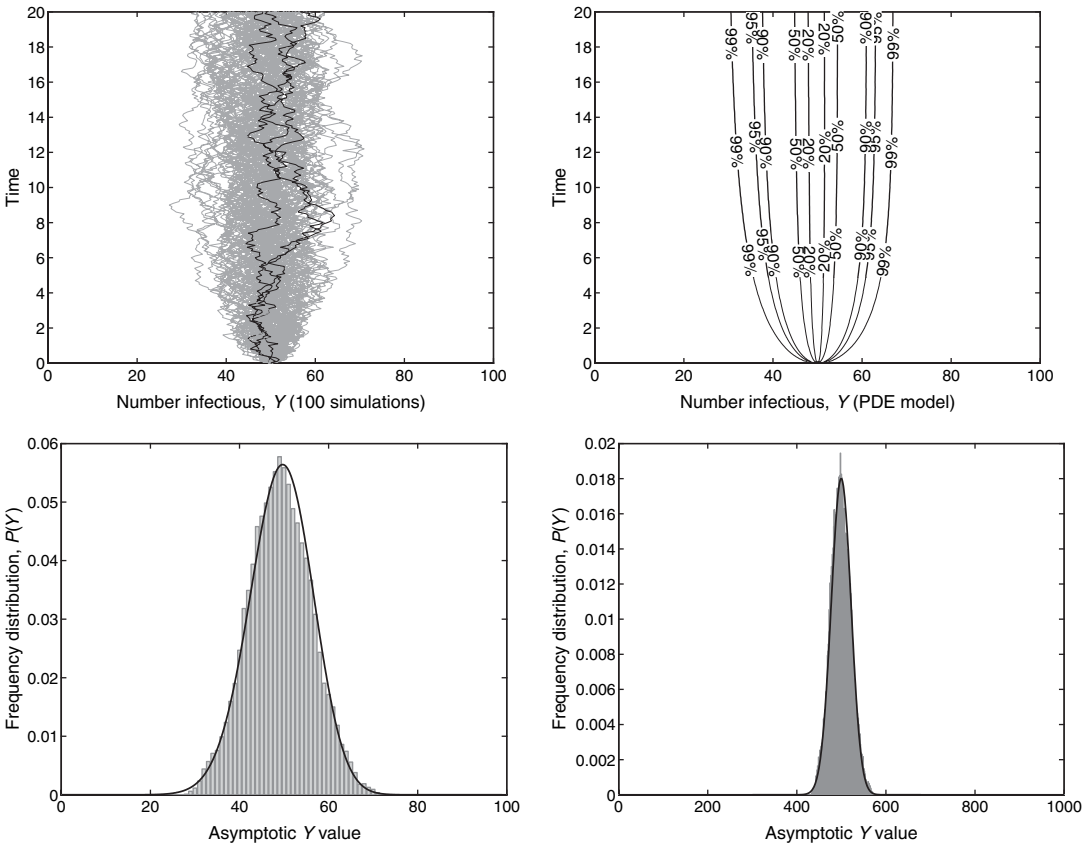

Figure 6.12. Comparison of 100 stochastic simulations and the Fokker-Plank PDE model, for the SIS model starting at the equilibrium point ($Y^* = N/2$). The top-left graph shows the time series of 100 simulations, with three random trajectories highlighted for greater clarity—note that time is plotted on the y-axis. The top-right graph gives the probability contours for the associated PDE model, corresponding to the probability of Y within a given range. The bottom graphs compare the long-term asymptotic results from both simulations and the PDE model for population sizes $N = 100$ and $N = 1000$; the results from the simulations (gray bars) being the average of 100 replicates between times 80 and 100 and the results from the PDE model (solid line) being the solution to $\frac{\partial P}{\partial t} = 0$. (All results are for $R_0 = 2$, $\gamma = 0.1$, $\mu = \nu = 0$).

at time t. Events, such as birth, infection, or recovery, move populations between various states and this is reflected in the equations for the probabilities.

The master equation provides an exact solution to multiple simulations of a model with demographic (event-driven) stochasticity.

Again, because of its simplicity and the fact that only two events can occur, we focus on the SIS equation. We let $P_Y(t)$ be the probability that Y individuals are infectious, noting again that there is no need to explicitly keep track of the susceptibles. It is often conceptually easier to think of a large (infinite) number of simulations, and to consider P_Y to be the proportion of simulations that have Y infecteds. Four processes can occur that

modify the proportion of simulations in state Y:

1. A simulation in state Y can have an infected individual recover at rate γY, such that there are now only $Y-1$ infecteds.
2. A simulation in state Y can have a susceptible individual become infected at rate $\beta XY/N = \beta(N-Y)Y/N$, such that there are now $Y+1$ infected.
3. A simulation in state $Y+1$ can have an infected individual recover at rate $\gamma(Y+1)$, such that Y infected remain.
4. A simulation in state $Y-1$ can have a susceptible individual become infected at rate $\beta(X+1)(Y-1)/N = \beta(N-Y+1)(Y-1)/N$, such that there are now Y infecteds.

Processes 1 and 2 cause the loss of a simulation in state Y, whereas processes 3 and 4 are associated with the gain of a state Y. Formulating an explicit equation for these processes, we obtain:

$$\begin{aligned}\frac{dP_Y}{dt} &= -P_Y[\gamma Y] - P_Y[\beta(N-Y)Y/N] + P_{Y+1}[\gamma(Y+1)] \\ &\quad + P_{Y-1}[\beta(N-Y+1)(Y-1)/N]. \end{aligned} \tag{6.7}$$

This generates $N+1$ differential equations ($Y = 0 \ldots N$), each of which is coupled to the two nearest probabilities. We insist that P_{-1} and P_{N+1} are zero, for obvious biological reasons.

As well as being able to iterate the probabilities forward from any particular initial conditions, one of the great benefits of these models is in understanding how stochasticity affects the final distribution of disease prevalence and the expected variation. However, a significant difficulty exists with this model, which also occurs for many models that use event-driven stochasticity—the fact that extinction events are permanent. Let us consider the dynamics of P_0, noting that P_{-1} is zero by definition:

$$\frac{dP_0}{dt} = -P_0[\gamma \times 0] - P_0[\beta N \times 0/N] + P_1[\gamma \times 1] = \gamma P_1.$$

Hence, P_0 keeps increasing, there is no escape from extinction, and eventually every simulation falls into this absorbing state, although this may take a very long time. So, the modeler is often trapped between the mathematical certainty of extinction and the epidemiological observation that extinctions in large populations are rare. Two solutions exist: either modify the equations such that a low level of infectious imports arrive in the population, or ignore those simulations that have gone extinct and find the distribution of the disease conditional on pathogen persistence (see Section 6.3.3.2). This latter method is often preferable because it retains the simplicity of the standard equations.

If we are interested in the final distribution of cases, then rather than having to iterate all the equations forward it is often far simpler to solve for when $\frac{dP_Y}{dt} = 0$. For this one-dimensional SIS model, the equilibrium solution can be found by equating the proportion of simulations moving from Y to $Y+1$ with the proportion moving from $Y+1$ to Y. If the model is at equilibrium, then these two "movements" must be equal or else some of the P_Y would be changing. This observation leads to iterative equations for the equilibrium

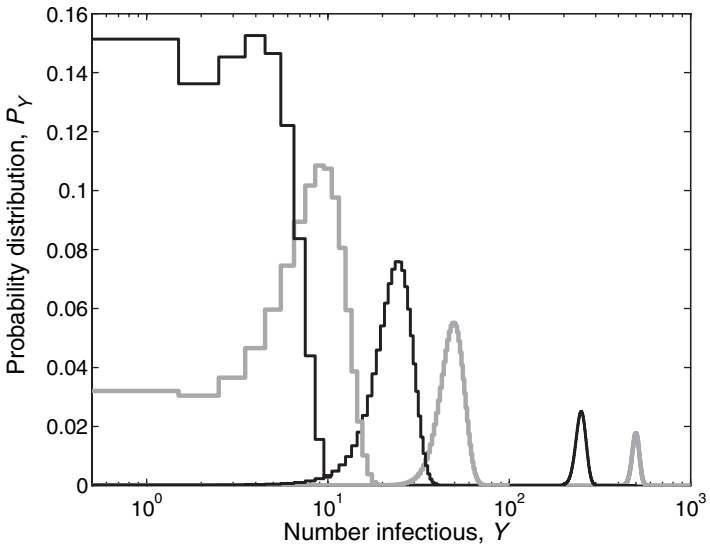

Figure 6.13. Final equilibrium probability P_Y^*, from the master equations for the SIS model, conditional on the infection being present. $R_0 = 2$, $N = 10, 20, 50, 100, 500$, and 1000. Note that the number of infectious individuals is plotted on a log scale. The integrals beneath all the curves are equal to one, even though the log scale makes the distributions for large populations appear to occupy a smaller area.

distribution:

$$P_{Y+1}^*[\gamma(Y+1)] = P_Y^*[\beta(N-Y)Y/N] \qquad \text{such that } \sum_{Y=1}^{N} P_Y^* = 1,$$

$$\Rightarrow \qquad P_{Y+1}^* = P_1^* \prod_{J=1}^{Y} \frac{\beta(N-J)J}{N\gamma(J+1)},$$

$$P_Y^* = P_1^* \frac{(N-1)!}{(N-Y)!Y} \left(\frac{\beta}{\gamma N}\right)^{Y-1}. \tag{6.8}$$

Hence, from either the initial differential equations (6.7) or the final explicit form (6.8), the equilibrium distribution of infection (conditional on nonextinction) can be calculated (cf Andersson and Britton 2000). Such conditional distributions are shown in Figure 6.13 from which several key factors emerge. For large population sizes (N greater than 500), the conditional equilibrium distribution is approximately Gaussian. This is because over the range of likely values the deterministic dynamics are approximately linear and the integer-valued nature of the population is largely irrelevant—in such cases the Fokker-Plank approximations may be very effective. (The smaller values of P_Y^* for these large population sizes occurs because the distribution is spread over a wider range of Y values). For smaller population sizes the distributions are much more skewed because the nonlinear behavior further from the fixed point plays a more dominant role and individuals are more important. For the very smallest population sizes, the distribution becomes bimodal with a peak at $Y = 1$ as well as near the expected mean. These results are equivalent to the earlier observation (Section 6.2.4.2) that for large population sizes the addition of noise to the

differential equations may be an adequate means of capturing stochasticity, whereas for small population sizes an individual-based event-driven approach is required.

This mechanism can be extended to examine the stochastic dynamics of the SIR equation; this is conceptually straightforward, but algebraically awkward. The probability is now two-dimensional, depending on both the number of infected and susceptible individuals. In a population of N individuals, the master equations for the SIR model has $\frac{1}{2}(N+1)(N+2)$ equations, whereas for the SIS model there were only $N+1$ equations, therefore iterating the SIR master equations is far more computationally intensive. However, when trying to understand the stochastic disease dynamics within farms (or small communities) with only a few hundred individuals, the calculation may be feasible. The equations are also slightly more complex to formulate because there are now six possible events (infection, recovery, birth, death of susceptible, death of infected, and death of recovered) that need to be considered. Results from this equation with a population size of $N = 10,000$ are shown in Figure 6.14. There are clearly many differences between the average dynamics of the master equation (black) and the behavior of the standard deterministic equations (gray) (top-left graph). The master equation has a higher level of susceptibles at equilibrium as expected, due to the negative correlations that develop between the numbers of susceptibles and infecteds. Interestingly, due to the higher dimensionality of the master equations, the trajectories of the average quantities can "cross"—this is not observed for the standard model. We also notice that in the master equations the mean values of X and Y asymptote far quicker to their equilibrium values compared to the deterministic model—thus, in some sense stochasticity is acting to increase convergence. Looking at the distributions in more detail, the first snapshot (top-right) shows the initial development of negative correlations such that high prevalence is associated with low levels of susceptibles. At later times (lower graphs), there is a much wider distribution of values and the effects of disease extinction can be observed leading to high levels of susceptibles with few or no infecteds.

Such master equations are clearly a powerful but computationally intensive tool for understanding the dynamics of stochastic disease models. In many respects they should act as the template against which other modeling strategies are judged. With the ever-increasing power of computers, the simulation of such large sets of differential equations will become increasingly feasible. The more mathematically able reader may like to reflect on the fact that equation (6.7) and master equations in general are linear in terms of the probability distributions and therefore can be written in matrix form ($\frac{d\underline{P}}{dt} = \boldsymbol{M}\underline{P}$), which substantially simplifies their analysis and computation.

6.6.3. Moment Equations

A convenient way to calculate the effect of noise or stochasticity on the dynamics of a disease is to develop moment equations (Nasell 1991,1996; Keeling 2000b). This approach has also been highly successful at studying the worm burden associated with a variety of macro-parasitic infections (Michael et al. 1998; Herbert and Isham 2000). The process starts by considering the effects of second-order moments (variances and covariances) on first-order moments (mean values). We introduce the convenient notation that $\langle . \rangle$ defines the average value of a quantity across infinitely many stochastic simulations. Looking at

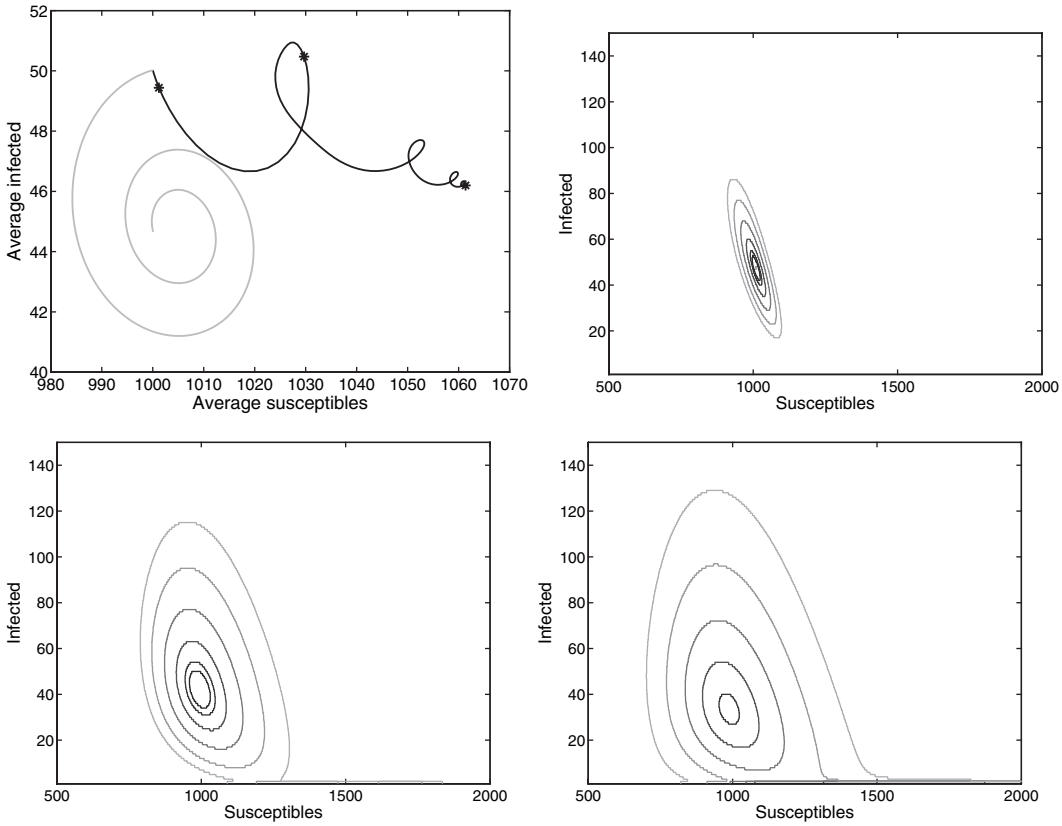

Figure 6.14. Dynamics of the SIR master equation ($R_0 = 10$, $\gamma = 0.1$ per day, $\nu = \mu = 5 \times 10^{-4}$ per day, $N = 10^4$, imports of infection occur at the rate of two per year, $\varepsilon \approx 0.055$ per day). Note that the values of ν and μ lead to a much faster demographic turnover than we have previously assumed; the average life expectancy is around 5.5 years. This has been done to increase the prevalence and hence improve disease persistence. The top-left graph shows the dynamics of the average number susceptible and infected predicted by the master equations (black) and predicted by standard differential equations (gray). The remaining graphs show the predicted distribution of infected and susceptibles at three times (indicated with stars on the top-left graph). The contours correspond to 95%, 90%, 75%, 50%, 25%, and 10% confidence intervals. The master equation was initialized with $P_{1000,50} = 1$ and all other terms zero. The equations were simulated within the rectangle $500 \leq X \leq 3000$, $0 \leq Y \leq 500$, and hence contain 1.25 million individual equations.

the equation for the number of infected individuals:

$$\begin{aligned} \frac{d\langle Y\rangle}{dt} &= \langle \beta XY/N - \gamma Y\rangle, \\ &= \beta\langle XY\rangle/N - \gamma\langle Y\rangle, \\ &= \beta\langle X\rangle\langle Y\rangle/N + \beta Cov_{XY}/N - \gamma\langle Y\rangle. \end{aligned}$$

Hence, the covariance between X and Y enters the equations and modifies the average transmission rate, which in turn modifies the mean values. Occasionally there are clear biological reasons to assume that the covariance has a given form (Keeling et al. 2000),

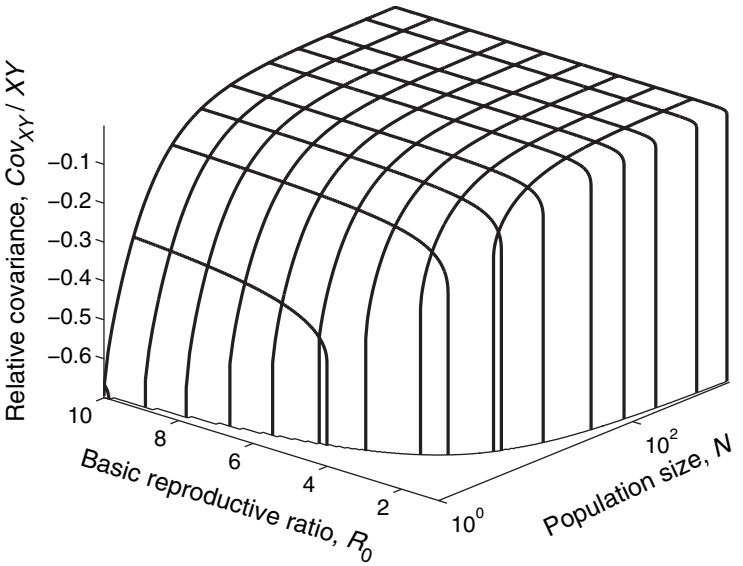

Figure 6.15. The relative magnitude of the covariance compared to the product XY at equilibrium for the SIS model as estimated from the moment closure equations. For small population sizes and small R_0 the disease cannot persist due to the strong negative correlation effects.

or we may wish to leave the covariance terms as a free parameter. However, frequently the next step is to use the disease dynamics to formulate a differential equation that describes the behavior of the covariance (Keeling 2000b). For the SIS model, after a fair amount of algebra, we find that the covariance rapidly converges to a value that depends on X and Y:

$$Cov_{XY} = -\frac{\beta XY + \gamma YN}{2\gamma N + 2\beta Y + \beta - 2\beta X}, \tag{6.9}$$

where closure has been achieved by assuming a normal distribution (Keeling 2000b). Therefore, we find that the covariance is large relative to the standard product (XY) whenever R_0 or the population size N is small (Figure 6.15). In fact, for very small populations or very small R_0, the effect of the covariance is so strong that the disease is always forced extinct—this is partly due to the normality assumption that was used to construct the covariance equation.

Although the amount of algebra involved in determining the moment equations and equilibria is somewhat daunting, there are some significant benefits. In particular, we now have an explicit relationship between the variance, the covariance, and the disease parameters which can often provide a more intuitive understanding than simulations alone. In the example provided by Figure 6.15, it is clear where the covariance is relatively large and hence where stochasticity is likely to affect both the effective transmission of infection and also the mean values. These insights can often provide novel interpretations of the results of stochastic simulations which would have been difficult to extract from the simulation results alone.

6.7. FUTURE DIRECTIONS

Almost all natural populations experience some degree of stochasticity. This is being increasingly recognized by epidemiologists and modelers who seek to account for the random nature of transmission and recovery. In recent years there has been a huge increase in the use of stochastic models in the literature; event-driven stochasticity is proving by far the most influential method due to its mechanistic underpinning and the ability to incorporate complex rules. This trend is likely to continue, as making large individual-based stochastic models become ever more computationally feasible (Levin et al. 1997).

Future work may also consider whether the stochastic nature of disease transmission can be turned to a public health advantage. We have clearly seen that stochasticity can lead to the rapid, and premature, extinction of infection. Harnessing and even amplifying this effect would be highly beneficial and could improve the efficacy of vaccination programs (see Chapter 8).

Stochasticity offers a new form of uncertainty that should be quantified by researchers. It is as important to consider the sensitivity of model results to stochasticity as well as the more common sensitivity to parameters. Associated with this is the difficulty of quantifying and explaining stochastic results. Mean results can often be biased by rare (but still stochasticity feasible) events; thus, at the very least, confidence intervals or distributions should always be specified when describing the outcome of stochastic models.

The analytical models addressed in Section 6.6 offer a bridge between the purely deterministic approaches and computer simulations. Thus, although mathematically daunting, these methods offer an analytical, and therefore often more intuitive, insight into the role of stochasticity. It is very likely that the development of such models, by skilled mathematicians, will play a pivotal role in the way epidemiologists think about stochasticity and stochastic modeling in the future.

6.8. SUMMARY

In this chapter we have shown how the random nature of transmission events and the individual nature of populations can affect disease dynamics. Although there are several methods of simulating noisy dynamics, we have concentrated on event-driven demographic stochasticity as a robust means of capturing random events and the discrete nature of individuals.

➤ Observational noise does not impact the epidemiological dynamics; it modifies only the reported data.

➤ Stochasticity in the transmission term can cause the number of cases to resonate at (or near) the natural frequency of the system. Hence, the observed stochastic dynamics more closely match the transient behavior of the deterministic equations.

➤ Stochasticity in the transmission term causes variance in the number of infecteds and in the number of susceptibles, and the negative covariance between them. The magnitude of these values increases almost linearly with the variance of the noise. Stochasticity can also introduce changes to the mean values as well as cause variation about the means.

➤ A log-log plot of mean against variance is a useful method of summarizing the variability with population size. For large population sizes, we expect the variance in the number of cases to be proportional to the mean.

➤ For large population sizes, where the individual nature of the population is unimportant, noise can be added to the differential equations to mimic the impact of stochasticity.

➤ Noise can be generated from a variety of sources. The relative magnitude of these noise terms depends on the population size. External parameter noise dominates when the population is large.

➤ Demographic stochasticity requires integer-valued variables and a probabilistic fate of individuals.

➤ Models using demographic stochasticity run much slower as the population size (and in particular the number of infectious individuals) becomes large; the τ-leap method provides a possible solution to this difficulty.

➤ From a single infected individual in a totally susceptible population, the probability of extinction before a major epidemic ensues is equal to $\frac{1}{R_0}$. A partially resistant population or a highly variable infectious period increases the likelihood of extinction, whereas multiple introductions of infection decrease the likelihood.

➤ The Critical Community Size (CCS) is the smallest population size observed not to suffer disease extinctions; for measles the CCS is around 400,000. Simple stochastic models that capture the observed seasonally induced measles epidemics fail to show adequate levels of persistence which can be captured only with a greater number of classes within the models and appropriate parameterization.

➤ Imports of infection into a community can significantly modify the extinction process in a complex manner. Frequent imports may prevent extinctions, whereas long delays between extinctions and imports improve the chance that an import successfully triggers an epidemic. When dealing with very low levels of imports between populations it is essential that this is modeled stochastically.

➤ In many situations, stochasticity may be beneficial, leading to the extinction of a disease well before the vaccination threshold is reached. Optimal vaccination strategies should take advantage of such extinctions.

➤ The Fokker-Plank equation provides an *exact* solution to multiple simulations of a model with noise. Noise in the equations translates to diffusion of the probability distribution.

➤ The master equation provides an exact solution to multiple simulations of a model with demographic (event-driven) stochasticity.