Epidemic Models

Introduction

The importance of epidemiology cannot be overstated. In 14th century Europe the Black Death killed 25 million people out of a population of 100 million; the Aztecs lost half their population of 3.5 million from smallpox; whilst around 20 million people died in the world pandemic of influenza in 1919 (Bailey [2]). Today the over-riding concern is the spread of HIV/AIDS, to the considerable detriment of the vast numbers of people suffering from less media-conscious diseases such as malaria, schistosomiasis, filariasis and hookworm disease. Parallel problems in marine and agriculturally based environments take a similar toll on plant, animal and fish populations. Human attempts to control such epidemiological disasters can themselves lead to further problems through, for example, improper use of pesticides and management strategies. Increasing understanding of the underlying processes involved is therefore one of the major problems of our age, and the best way forward is through the careful use of mathematical modelling.

In general, basic model structures allow for the development of theoretically based results, whilst acceptance of epidemiological reality forces attention back onto simulation based procedures since any mathematical results that might be derived may be too opaque to be useful. Fortunately, the modelling of infectious diseases involves is a direct paradigm of the study of ecological processes; this can be used to considerable advantage since all the approaches developed for one can be carried over to the other. For example, whilst the 'death' of a susceptible automatically gives rise to the 'birth' of an infective, since it involves a transfer of state for the same individual, in the parallel biological scenario prey do not become predators but merely act as food for them. Though most epidemic (and biological) models are by necessity naive descriptors of true infection (and population) development, when carefully constructed they can generate extremely useful qualitative predictions not only about possible modes of behaviour, but also about the likely effects of postulated control strategies.

Though a wide range of approaches are available, modellers often remain loyal to their own area of speciality to the detriment of subsequent analyses and the inferences that stem from them (see Renshaw [11]). Deterministic analyses are the simplest to construct, and numerical solutions can easily be obtained via standard packages such as Mathcad or Matlab. Stochastic analyses, however, may well exhibit substantially different behaviour, and although the derivation of theoretically based results can involve the use of specialised approximation techniques, simulated realisations are usually trivial to construct. If the latter lie close to the deterministic trajectory, then a deterministic analysis will be sufficient; otherwise pursuance of the stochastic approach is imperative. Similarly, many real-life epidemiological processes involve a strong spatial component, and again stochastic simulations are easily produced. Finally, because data are often both expensive and logistically difficult to collect, discrete-time processes may be used to model continuous-time scenarios, with local transition rates being replaced by binomial-type transition distributions. In practice, it is vital to consider all such possibilities before selecting the most appropriate model structure for a given situation.

Simple epidemics

In the simplest type of epidemic model infection spreads by contact between members of the community, and infected individuals are not removed from circulation by recovery, isolation or death. Thus all individuals susceptible to the disease must ultimately become infected. Suppose we have a homogeneously mixing group of $n+1$ individuals, and that at time $t=0$ the epidemic starts with

just one infected individual. The remaining n individuals are all assumed to be susceptible to the infection. Then since there are $x(t)y(t)$ possible contact-pairs between the $x(t)$ susceptibles and $y(t)$ infectives at time t, each of which can turn a susceptible into an infective at rate β, it follows that the deterministic rate of decline of susceptibles is given by

$$dx(t)/dt = -\beta x(t)y(t) \ , \tag{1}$$

with $y(t) = n + 1 - x(t)$. Although this system can be easily solved to yield

$$y(t) = (n+1)/[1 + n\exp\{-(n+1)\beta t\}] \ , \tag{2}$$

such mathematical tractability is a rarity. Moreover, from a stochastic viewpoint the apparently simple nature of this process proves to be dangerously deceptive. For not only does the quadratic infective birth rate, $\beta y(t)(n+1-y(t)$, produce considerable algebraic hassle in determining the probabilities $\Pr(y(t) = i)$ $(i = 1, \ldots, n+1)$, but the stochastic mean (Haskey [6]) differs from the deterministic value. This difference is exacerbated when we consider the rate at which new infectives occur. For the deterministic form of the 'epidemic curve', $dy(t)/dt$, rises higher than the stochastic one (based on expected numbers), and then falls away more quickly. So working with the former can lead to wrong interpretation about the progress of the epidemic. For this simplest of models, both exact and asymptotic values for the mean and variance of the duration time (T) of the epidemic can be easily constructed (e.g. Renshaw [11]). In particular, the coefficient of variation $CV(T) \sim \pi/[2\sqrt{3\ln(n)}]$ remains moderately large even for sizeable values of n, so quite substantial differences in epidemiological behaviour can occur between separate, but otherwise identical, groups of individuals. Considerable care is therefore required when attributing unexpected results to abnormal virulence or infectiousness, for this could be due purely to chance fluctuations.

General epidemics

Suppose we now allow infectives to be removed from circulation at rate γ by isolation or death. Although this does not lead to a completely general epidemic that can take account of migration, geography of infection sites, loss of immunity, latent period of infection, variable parameter values, etc., it does give rise to a model which is sufficiently realistic to be useful. Denoting $z(t)$ to be the number of removed infectives by time t and $\rho = \gamma/\beta$ to be the *relative removal rate*, with $x(t) + y(t) + z(t) = n$, the deterministic equation (1) becomes

$$\begin{aligned} dx(t)/dt &= -\beta x(t)y(t) \\ dy(t)/dt &= \beta x(t)y(t) - \gamma y(t) \\ dz(t)/dt &= \gamma y(t) \ . \end{aligned} \tag{3}$$

So an epidemic can build up (i.e. $dy(t)/dt > 0$) only if $x(0) > \rho$. Thus $x(0) = \rho$ defines a (deterministic) threshold density of susceptibles below which an epidemic cannot develop, since infectives are removed at a faster rate than new infectives are produced. Even for this basic model the epidemic curve cannot be derived exactly, though a good approximation can be obtained provided z/ρ is small. Indeed, the celebrated Threshold Theorem of Kermack and McKendrick [8] shows that an initial number of susceptibles $\rho + v$ is eventually reduced to $\rho - v$. i.e. to a value as far below the threshold as it was initially above.

This idea that an outbreak can or cannot occur as the population size switches through n is clearly fanciful, and highlights the danger of relying on a deterministic approach. What happens is that the *probability* that an outbreak occurs will change. In the small time interval $(t, t+h)$ we have the transition probabilities

$$\Pr[(x,y) \to (x-1, y+1)] = \beta xyh \ \text{(infection)} \quad \text{and} \quad \Pr[(x,y) \to (x, y-1)] = \gamma yh \ \text{(removal)} \ ,$$

from which equations can be constructed for the probabilities $p_{ij}(t) = \Pr[x(t) = i, y(t) = j]$. Such equations are extremely difficult to work with, though methods of deriving solutions through the saddlepoint approximation (see Renshaw [12]) are currently under investigation. However, early success was achieved (Whittle [15]) by observing that in the opening stages of an epidemic, $y(t)$ behaves like a simple birth-death process with parameters $n\beta$ and γ. For $y(0) = a$, this gives rise to a basic stochastic threshold theorem: if $n \leq \rho$ then a major outbreak cannot occur; if $n > \rho$ then a minor or major outbreak occurs with probability $(\rho/n)^a$ and $1 - (\rho/n)^a$, respectively. Moreover, the average size of the epidemic is $an/(\rho - n)$ $(n < \rho)$ and $(\rho/n)^a[a\rho/(n - \rho)] + [1 - (\rho/n)^a](r - a)$ $(n > \rho)$ where r is the unique positive root of the equation $a - r + n[1 - \exp(-r/\rho)] = 0$. If n is large then the average duration time is approximately $\gamma^{-1} \ln(a + n)$. Though such results provide useful indicators of epidemic development, they are limited on two counts. First, they provide little information on how *individual realisations* will develop, since probabilities reflect 'average' behaviour over all possible outcomes, and in practice we usually observe a single time-series of events. Second, diseases which are modelled by equations more complex than (3) are likely to be mathematically intractable, and so interest has to centre around the derivation of approximation procedures; even then, solutions may be too opaque to yield practical insight. Fortunately, any stochastic epidemic model, no matter how complex, easily yields to computer simulation, with stochastic trajectories being produced by generating successive event-time pairs (see Renshaw [11]).

In general, denote infection and removal rates by $a(x, y)$ and $b(y)$, respectively; so here $a(x, y) = \beta xy$ and $b(y) = \gamma y$. Then for $\{Z\}$ a sequence of uniform(0,1) pseudo-random numbers, the next event is an infection if $a/(a + b) \leq Z$, else it is a removal, and the time interval to the next event is $-[\ln(Z)]/(a + b)$. Simulating say 100 realisations not only demonstrates the type and variability of behaviour to be seen in major and minor outbreaks, but plotting specific features, such as extinction times and the maximum number of infectives, enables empirical p.d.f.'s to be constructed which encapsulate the underlying statistical features of the process. Ball and O'Neill [3], for example, model the HIV/AIDS epidemic by taking $a = xy/(x + y)$. Indeed, such an approach works just as easily with highly complex forms for the rates $a(x, y)$ and $b(y)$ as with simple ones, and appropriate precision can be achieved by increasing the number of runs.

Recurrent epidemics

That the general epidemic produces either a minor outbreak which then swiftly dies away, or else a major build-up of infectives which then slowly subsides, is clearly useful for rare diseases since any outbreak that does occur can be regarded as a single phenomenon. However, more common diseases like measles, chicken-pox, influenza, diptheria, etc., exhibit periodic flare-ups with infection being sustained at a low level inbetween times by a gradual spread to new susceptibles. The development of processes which can explain such cyclic behaviour has captivated modellers ever since the pioneering work of Bartlett [2]. Not only are such periodic phenomena universal, but the persistence of the oscillations depends on the the total population size.

Suppose that new susceptibles enter the population at rate α. Then the basic deterministic equations (3) become

$$\begin{aligned} dx(t)/dt &= -\beta x(t)y(t) + \alpha \\ dy(t)/dt &= \beta x(t)y(t) - \gamma y(t) \, . \end{aligned} \tag{4}$$

Although these nonlinear equations do not yield a simple solution, considerable progress can be made by examining small departures from the equilibrium values $x^* = \gamma/\beta$ and $y^* = \alpha/\gamma$ by writing $x(t) = x^*(1 + u(t))$ and $y(t) = y^*(1 + v(t))$. For with $\sigma = \gamma/\alpha\beta$, this (local) linearisation

shows that for $\psi = \sqrt{\{(\gamma/\sigma) - (1/4\sigma^2)\}}$,

$$\begin{aligned} v(t) &= v(0)\exp(-t/2\sigma)\cos(\psi t) \\ u(t) &= -v(0)(\gamma\sigma)^{-1/2}\exp(-t/2\sigma)\cos(\psi t - \theta)\ . \end{aligned} \tag{5}$$

Thus both the infective and susceptible populations undergo damped cosine waves with period $T = 2\pi/\psi$ and phase difference $\theta = \cos^{-1}(1/\sqrt{(4\gamma\sigma)}$. Whilst for measles outbreaks (for example) this gives rise to epidemic outbreaks with roughly the right period, the peak-to-peak damping factor of $\exp(T/2\sigma) \simeq 0.6$ of the infective cycles towards a steady endemic state contradicts the epidemiological facts.

It is an unfortunate aspect of modelling that such deterministic methods usually get only part of the story right, more often in the opening (i.e. transient) stages of population development than in the later (generally persistent) phases. The only way of discovering which behavioural features are relevant is to compare deterministic and stochastic realisations. Denote the birth and death rates of susceptibles (S) and infectives (I) by $\lambda_S(x,y) = \alpha$, $\mu_S(x,y) = \lambda_I(x,y) = \beta(x,y)$ and $\mu_I(x,y) = \gamma y$. Then for a given sequence of pseudo-random numbers $\{Z\}$, the (general) algorithm for generating a stochastic realisation is as follows.

(1) evaluate $D(x,y) = \lambda_S(x,y) + \mu_S(x,y) + \mu_I(x,y)$
(2) obtain next Z
(3) evaluate time of next event $t - \ln(Z)/D$
(4) obtain next Z
  (a) if $\lambda_S/D \leq Z$ then $(x,y) \to (x+1,y)$
  (b) else if $(\lambda_S + \mu_S)/D \leq Z$ then $(x,y) \to (x-1,y+1)$
  (c) else $(x,y) \to (x,y-1)$
(5) output new x, y and t and return to (1).

Simulations with this model can produce sustained cyclic flare-ups with the required frequency, with widely varying amplitudes mimicking real-life epidemic behaviour. Indeed, this simple process even has the potential for flip-flopping between the epidemic and endemic states. The only modification needed to ensure long-term persistence is to allow the occasional infective to enter the system (at some small rate δ) to ensure that the epidemic restarts should the infective population die out. Note that only trivial coding changes are required to facilitate the simulation of general multi-type processes.

Natural extensions

Although all the above models take simple mathematical forms, possibilities abound for making them more realistic. For example, considerable attention has been focussed on processes which describe the flow of host individuals between various states of infection (e.g. latent, infectious, immune, carrier, death). These states can be thought of as 'compartments', with an individual's change of state being viewed as a migration from one compartment to another (e.g. Jacquez [7]). Provided that for each compartment the 'birth', 'death' and migration rates are chosen delicately, with careful consideration being paid to biological, epidemiological and environmental reality, then such representations can provide a powerful insight into quite complex diseases which involve both multi-stage development and host-vector interactions. Cvjetanović [5], for example, describes compartmental systems for tetanus, typhoid, cholera and diptheria, using 9, 10, 11 and 10 states, respectively. In such complex situations a strong partnership between numerical solution of the deterministic equations and stochastic simulations is vital if meaningful conclusions concerning population control and development are to be drawn.

A good illustration of how simple models can be usefully extended in this way is provided by Aron

and May's [1] account of the transmission and maintenance of malaria. This can be modelled as a two-state system which focusses on the role played by the basic reproductive rate of the parasite and on the dynamics of the prevalence of infection between humans and mosquitoes. The basic deterministic model may be written in the form

$$\begin{aligned} dx/dt &= \sigma y(1-x) - rx \qquad (\sigma = abM/N) \\ dy/dt &= ax(1-y) - \mu y \end{aligned}$$

where: x and y are the proportions of the human and female mosquito populations that are infected; N and M are the (constant) sizes of the human and female mosquito populations; a is the bite rate of a single mosquito; b is the proportion of infected bites that produce an infection; r is the individual recovery rate for humans; and μ is the individual death rate for mosquitoes. Although this model fails to take account of the different developmental stages of the parasite, it does contain the essentials of the transmission process and enables distinctions to be made between patterns found in various sets of data from different geographic regions. Denoting $\omega = \alpha/\mu$, the equilibrium values

$$x^* = (R-1)/(R+\omega) \quad \text{and} \quad y^* = (R-1)\omega/R(1+\omega)$$

involve the 'basic reproductive rate' R of the parasite, which is essentially the number of secondary cases of infection generated by a single individual. If $R < 1$ then the disease will be unable to sustain itself; whilst the more R exceeds 1 the greater is the resistance of the infection to eradication (see Renshaw [11]). However, on using their parameter values, stochastic realisations do not reproduce observed severe upsurges in new malaria cases, and a greater degree of realism is needed for this to be achieved. One way is to incorporate a fluctuating environment into the process, by letting the total mosquito population vary seasonally with random amplitude. Another is to incorporate the incubation period of the parasite within the mosquito, thereby invoking a time-lag situation.

Given the potentially large number of such possibilities, it is vital to be able to construct a general theoretical framework which accounts for both demographic and environmental stochasticity. Marion et al. [9] construct such an approach by considering some simple models of environmental stochasticity in terms of a non-linear model for nematode infections in ruminants (proposed by Roberts and Grenfell [14]). This is particularly suitable since it captures the essence of more complicated formulations of parasite demography and herd immunity found in the literature. Demographic stochasticity is shown to be important in terms of extinction events and equilibrium model behaviour (the *endemic regime*), but muted in a transitory *managed regime* where the system is periodically perturbed. Whilst allowing for deterministic and stochastic fluctuations in the model parameters shows the crucial importance of environmental fluctuations. Analytic tools explored include moment closure, stochastic local linearisation, the use of the mean-reverting Ornstein-Uhlenbeck process, stochastic differential equations, and the incorporation of weather data to investigate real effects of micro-climatic fluctuations.

Moreover, if we are to gain proper understanding of the mechanics and control of infectious diseases then we have to progress beyond the convenient mathematical assumption that individuals mix homogeneously over the whole region available to them. Disease is spread by two different mechanisms (see Renshaw [12]): by migration to a different location (e.g. movement of rabid foxes, road transport of infected cattle); or, through cross-infection, either locally (e.g. between neighbouring plants) or globally (e.g. via aerosol dispersion). To date the severity of the mathematical problems associated with combining recent non-spatial stochastic developments, as outlined above, with spatial interaction has prevented 'real-world' success, most investigations being based on simple idealisations. However, not only are space-time data now becoming increasingly plentiful over a huge range of spatial scale, but there is a fast-growing public awareness of the importance of tackling huge transmission systems, as exemplified by the great gerbil/flea transmission of plague

in Kazakhstan (Marshall et al. [10]). There is therefore a strong incentive to effect a rapid injection of progress into the development of associated space-time stochastic theory, large system simulation procedures, sampling procedures to gather well-defined space-time data, statistical Markov chain Monte Carlo techniques to cope with missing data, not to mention much greater input from biologists and applied epidemiologists to determine far more precise mechanisms for infective growth and spread.

References

[1] Aron, J.L. & May, R.M. (1982) The population dynamics of malaria. In *The Population Dynamics of Infectious Diseases: Theory and Application*, R.M. Anderson, ed. London: Chapman and Hall, pp.139-179.

[2] Bailey, N.T.J. (1975) *The Mathematical Theory of Infectious Diseases and its Applications.* (2nd edition). London: Griffin.

[3] Ball, F. & O'Neill, P. (1993) A modification of the general epidemic motivated by AIDS modelling. *Advances in Applied Probability* **25**, 39-62.

[4] Bartlett, M.S. (1957) Measles periodicity and community size. *Journal of the Royal Statistical Society* **A 120**, 48-70.

[5] Cvjetanović (1982) The dynamics of bacterial infections. In *The Population Dynamics of Infectious Diseases: Theory and Application*, R.M. Anderson, ed. London: Chapman and Hall, pp. 38-66.

[6] Haskey, H.W. (1954) A general expression for the mean in a simple stochastic epidemic. *Biometrika* **41**, 272-275.

[7] Jacquez, J.A. (1996) *Compartmental Analysis in Biology and Medicine* (3rd edition). Biomedware, Ann Arbor, MI.

[8] Kermack, W.O. & McKendrick, A.G. (1927) Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society of London* **A 115**, 700-721.

[9] Marion, G., Renshaw, E & Gibson, G. (2000) Stochastic modelling of environmental variation for biological populations. *Theoretical Population Biology* **57**, 197-217.

[10] Marshall. E.C., Frigessi, A., Stenseth, N.C., Holden, M., Ageyev, V.S. & Klassovskiy, N.L. (2000) Plague in Kazakhstan: A Bayesian model for the temporal dynamics of a vector-transmitted infectious disease (submitted).

[11] Renshaw, E. (1991) *Modelling Biological Populations in Space and Time.* Cambridge University Press.

[12] Renshaw, E. (1998) Epidemic Models, Spatial. In *Encyclopaedia of Biostatistics*, Wiley.

[13] Renshaw, E. (2000) Applying the saddlepoint approximation to bivariate stochastic processes. *Mathematical Biosciences* (to appear).

[14] Roberts, M.G. & Grenfell, B.T. (1991) The population dynamics of nematode infections of ruminants: periodic perturbations as a model for management. *IMA Journal of Mathematics Applied in Medicine and Biology* **9**, 29-41.

[15] Whittle, P. (1955) The outcome of a stochastic epidemic – a note on Bailey's paper. *Biometrika* **42**, 116-122.