Dieckmann, Heesterbeek, Britton
Mathematical Tools for Understanding
Infectious Disease Dynamics
Princeton University Press, 2013

Chapter 10

Spatial spread

10.1 Posing the problem

As an example of a situation where spatially structured models are relevant, think of a fungal pathogen affecting an agricultural crop. A farmer having ascertained that his field is affected wants to know: How fast is the infection spreading? What fraction of the yield do I stand to lose if I do not spray with fungicides? The trade-off here could be that spraying is expensive and bad for the environment. Suppose that harvest is three months away. Do I take the loss of plants or do I invest in fungicide and accept the concomitant pollution?

So, the key question is: How fast is the infection spreading?

A student of the preceding chapters might be inclined to answer the farmer by first drawing Figure 10.1 and next, in an attempt to be pragmatic rather than scrupulous, saying that the relevant part of the curve is to a good approximation described by Ce^{rt}, with C determined from the current situation and r from (one hopes) known data about spore production and dispersal. The aim of this chapter is to provide the student with better ingredients for an answer. In particular, we will explain that it is much more likely that the fraction of the crop affected will grow as a quadratic function of time.

The main point is that the infestation is localized in patches, called foci, which expand more or less radially. The population growth parameter r, however, describes population growth that is uniform (in space), as can be concluded from the eigenfunction, which is constant (in fact also when we consider R_0). This will be elaborated below.

Thus we arrive at the following set of questions: Do models predict radial expansion of epidemic fronts? If so, how can we determine the *speed* of the front from the model ingredients? What are the conditions that promote wave-like expansion?

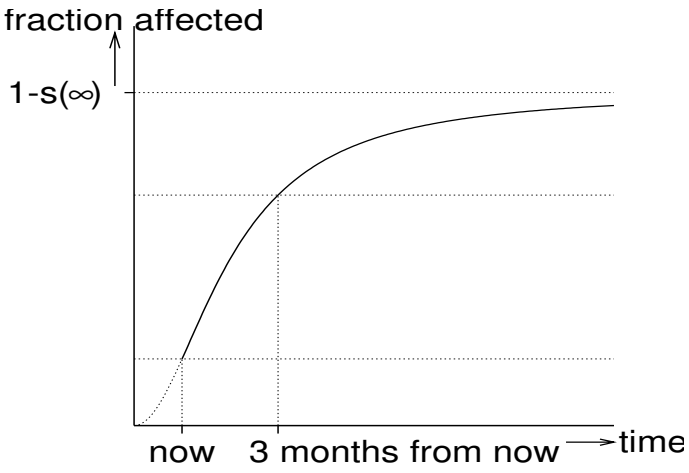

Figure 10.1: Pragmatic sketch of increase in fraction of a plant crop affected by a fungal pathogen as a function of time. See Section 10.1 for context.

10.2 Warming up: the linear diffusion equation

Let u denote the density of a pest species. Let this species inhabit a very large domain without structure (no hedges, roads, canals, rivers, mountains, ...). We take the plane $\mathbb{R}^2$ as an idealized representation of this domain and let $x \in \mathbb{R}^2$ denote a spatial position. Note that $\mathbb{R}^2$ is homogeneous (i.e. translation-invariant: points are interchangeable) and isotropic (the same structural properties in every direction).

We assume that the species grows at a net per capita rate κ. In addition, it disperses. If movement is completely random, we may use the diffusion equation to describe dispersal (the underlying assumption being that the flux is proportional to the gradient). Hence we postulate that

$$u_t = D\triangle u + \kappa u, \tag{10.1}$$

where $u_t = \frac{\partial u}{\partial t} = \frac{\partial}{\partial t}u(t,x)$ is the partial derivative of u with respect to time, where D is a species-dependent constant (called the diffusion constant) representing the mean square distance covered per unit of time, and where

$$\triangle u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}$$

is shorthand for the sum of the second partial derivatives of u with respect to the two coordinate directions.

The chief advantage of (10.1), and the reason to consider it here, is that we can solve this equation explicitly. The explicit expression allows us to pinpoint certain quantitative and qualitative properties, in particular the phenomenon of an *asymptotic*

speed of propagation c_0, which equals the *minimal speed* for which *plane travelling-wave solutions* exist. Armed with this knowledge, we then study, in subsequent sections, the robustness of the conclusions: Do the qualitative phenomena extend to the nonlinear realm? And to models involving kernels $A(\tau, x, \xi)$?

The solution of (10.1),

$$u(t,x) = \frac{1}{4\pi D t} e^{-\frac{|x|^2}{4Dt} + \kappa t}, \tag{10.2}$$

describes the effect of a localized (at $x = 0$) disturbance (at $t = 0$) of the unstable steady state $u \equiv 0$ (read as: u is identically zero).

Exercise 10.1 Verify that u given in (10.2) satisfies the equation (10.1).

Exercise 10.2 Verify that $\int_{\mathbb{R}^2} u(t,x)\, dx = e^{\kappa t}$ for $t > 0$.

Exercise 10.3 Show that uniformly for $|x| \geq \varepsilon$ we have

$$\lim_{t \downarrow 0} u(t,x) = 0$$

and conclude that u is indeed concentrated in $x = 0$ at $t = 0$. (In the language of distributions and measures, we have $\lim_{t\downarrow 0} u(t,x) = \delta(x)$, with δ the Dirac 'function' (measure/distribution) concentrated in $x = 0$.)

What can be said about the behaviour of $u(t,x)$ for large t? Of course the factor $\frac{1}{t}$ goes to zero, but $e^{\kappa t}$ goes to infinity much faster. So, if, on the one hand, we fix x and let t tend to infinity, we find that u grows exponentially with rate κ. If, on the other hand, we fix t and let $|x| \to \infty$ (that is, we observe far ahead in space) we find that u is negligibly small. Thus it appears that the limits $t \to \infty$ and $|x| \to \infty$ cannot be interchanged: the order matters. In such situations of non-uniform convergence, one expects to see a *transition layer* (in which transition from one extreme, zero, to the other extreme, infinity, is made) once we approach infinity in (t,x) space in a tailor-made fashion. We can immediately infer from the explicit expression what this 'tailor-made fashion' is: we have to avoid the exponent in (10.2) going to either $+\infty$ or $-\infty$ (the factor $1/t$ necessitates that we be a little more precise about the first of these possibilities, viz. we have to avoid approaching $+\infty$ too quickly; it is indeed this factor $1/t$ that makes the precise characterization of the transition layer rather subtle). We refrain from a more precise study, and restrict ourselves to the observation that, for any $\varepsilon > 0$, for $t \to \infty$

$$u(t,x) \to \begin{cases} 0 & \text{if } |x|^2 > (4D\kappa + \varepsilon)t^2, \\ \infty & \text{if } |x|^2 < (4D\kappa - \varepsilon)t^2. \end{cases}$$

In suggestive words, we could say that we distinguish the 'not yet' region, being the exterior of a disc, the radius of which grows like $2t\sqrt{D\kappa + \frac{\varepsilon}{4}}$, and the 'already over' region, being the interior of a disc, the radius of which grows like $2t\sqrt{D\kappa - \frac{\varepsilon}{4}}$. (Here ε is a positive number that can be taken arbitrarily small. It relieves us from

going into the details of the subtle limiting behaviour that occurs when x grows like $2t\sqrt{D\kappa} + O(\ln t)$.) Yet another way of expressing this result is to state that

$$c_0 := 2\sqrt{D\kappa}$$

is the *asymptotic speed of propagation* of the disturbance. We conclude that the solution of the linear diffusion equation (10.1) displays radial expansion of a disturbance with a well-defined speed that can be easily computed from the parameters.

Our next aim is to characterize c_0 in a completely different way, viz. as the minimal speed of plane travelling waves. The point is that this characterization carries over much more easily to other situations, in which explicit calculations are often impossible.

A plane wave travelling in the direction specified by a given unit vector ν is described by a solution of the form

$$u(t, x) = w(x \cdot \nu - ct), \tag{10.3}$$

(here $x \cdot \nu$ is the inner product of the vectors x and ν, i.e. $x \cdot \nu = x_1\nu_1 + x_2\nu_2$). We call w the *profile* of the wave, ν the direction of the wave, and c its *speed*.

Exercise 10.4 Show that, in order for (10.3) to define a solution of (10.1), the profile w should satisfy

$$Dw'' + cw' + \kappa w = 0. \tag{10.4}$$

Next argue that this requires w to be of the form $w(\xi) = \exp(\lambda\xi)$, with λ satisfying the characteristic equation

$$D\lambda^2 + c\lambda + \kappa = 0. \tag{10.5}$$

Conclude that

$$\lambda = \lambda_\pm = \frac{-c \pm \sqrt{c^2 - 4D\kappa}}{2D}. \tag{10.6}$$

We want *positive* solutions, because of the interpretation. Oscillating solutions are characterized by complex λ. So we should require $\lambda_\pm$ to be real; that is, we should have $c^2 - 4D\kappa \geq 0$, i.e. $c \geq 2\sqrt{D\kappa} = c_0$. We conclude that plane travelling wave solutions exist for all speeds c that exceed a threshold c_0 and that the minimal plane-wave speed c_0 coincides with the asymptotic speed of propagation.

Exercise 10.5 For $c = c_0$ we have $\lambda_\pm = -\frac{c_0}{2D}$ and

$$w(\xi) = e^{\lambda_\pm(x\cdot\nu - c_0 t)} \sim e^{-\lambda_\pm c_0 t} = e^{\frac{c_0^2}{2D}t} = e^{2\kappa t}$$

for large t. What do you conclude from this?

10.3 Verbal reflections suggesting robustness

Consider a steady state (zero/infection-free) that is unstable and such that any physically feasible (in particular *positive*) perturbation triggers a transition towards another steady state ('infinity' in the case of the linear diffusion equation, i.e. 'after the epidemic'). To this local dynamics, add a spatial component and coupling, which means that perturbations at some point generate perturbations at nearby points. How fast do perturbations spread?

Imagine space to be the same in every point in all directions (i.e. homogeneous and isotropic). Travelling plane waves are uniform in all directions but one. So they manifest how disturbances travel in one direction (although this direction is arbitrary because of the isotropy). And the speed tells us how fast the spread will be.

Plane waves do not come with a unique speed, but rather with a continuum of possible speeds, bounded only at one side. Why should the minimal speed be the truly relevant one?

To fix a unique solution, the partial differential equation (PDE) (10.1) has to be supplemented with an initial condition. At least as a thought experiment, we can therefore manipulate the solution[1]. Imagine a series of fireworks placed in a row, with fuses of varying lengths. By lighting the fuses, one can create a 'travelling wave' of explosions. By choosing the lengths of the fuses appropriately, one can achieve any speed one wants. If, however, the fireworks also have the tendency to kindle their nearest neighbours, this process of self-kindling will dominate as soon as one tries to achieve, by manipulation of the fuses, a speed that is too low. Therefore, the minimal plane-wave speed corresponds to the inherent speed of the self-infection mechanism!

Travelling plane-wave solutions are examples of similarity solutions, i.e. solutions depending only on a certain combination of the independent variables. They show up whenever the dynamics are equivariant under a group of transformations (in this case translations and rotations). They are often the quintessence of intermediate asymptotic behaviour, when the transients reflecting the initial conditions have died out, but where the final state has not yet been achieved everywhere and boundary conditions (every real domain is finite!) do not yet impinge upon the natural dynamics[2].

We conclude that whenever

- local dynamics consist of a transition from an unstable steady state to a stable one,
- perturbations spread, i.e. there is some form of coupling of local dynamics,
- space is homogeneous and isotropic,

we are bound to find that

- travelling plane-wave solutions exist for all speeds c exceeding a minimal speed c_0,

[1] The thought experiment is due to J.A.J. Metz.

[2] G.I. Barenblatt, *Similarity, Self-similarity and Intermediate Asymptotics*. Plenum, New York, 1979.

- the minimal wave speed c_0 is the asymptotic speed of propagation associated with the self-triggering mechanism.

Many of these conclusions remain valid if there is homogeneity but no isotropy[3]. Of course, the minimal plane-wave speed will then depend on the direction ν. When defining what asymptotic propagation means, one should then blow up not a disc but rather another (convex) set defined on the basis of the function $c_0(\nu)$.

Exercise 10.6 Suppose a species spreads by wind-borne propagules. Assume there is a prevailing wind direction. Let the unit vector σ point to this wind direction and let θ be the wind velocity. Then (10.1) should be replaced by

$$u_t = D\triangle u - \theta\sigma \cdot \nabla u + \kappa u,$$

where ∇u is the gradient (i.e. $\nabla u = (\partial u/\partial x_1, \partial u/\partial x_2)^\top$, the vector of partial derivatives), and so $\sigma \cdot \nabla u = \sigma_1 \frac{\partial u}{\partial x_1} + \sigma_2 \frac{\partial u}{\partial x_2}$ is the directional derivative in the direction σ. In case you wonder about the minus sign, put $D = 0$ and $\kappa = 0$ and check that $u(t,x) = \phi(\sigma \cdot x - \theta t)$ satisfies the equation for any function ϕ, and that such solutions correspond to plane waves travelling in the direction σ. Look at Figure 10.2, turn it into an animation of a travelling wave and verify that the direction of propagation is to the right.

Now let us look for wave solutions

$$u(t,x) = w(\sigma \cdot x - ct)$$

travelling with speed c in the direction σ. Show that such solutions exist for all $c \geq c_0 + \theta$ with $c_0 = 2\sqrt{D\kappa}$ as before. Next look for travelling waves in the opposite direction. What do you conclude?

This seems an appropriate place for the following side-remark. Within the general framework, it is possible that a population grows while continuously moving in space to ever more removed regions (i.e. to 'infinity'). The spectral radius criterion would say $R_0 > 1$, so the population grows. A local observer, however, would only notice a very temporary growth, followed by local extinction. Within a more measure-theory-oriented approach, growth is defined in terms of a local indicator, the *Perron root*. In the situation just described, it follows that the Perron root is less than one and this is interpreted as 'the population does not grow'. We refer to Jagers (1995) for a precise definition of the Perron root and note that Shurenkov (1992) has shown that the spectral radius and the Perron root coincide whenever the state space is compact[4].

[3] H.F. Weinberger: Long-time behaviour of a class of biological models. *SIAM J. Math. Anal.* **13** (1982), 353-396; F. van den Bosch, O. Diekmann & J.A.J. Metz: The velocity of spatial population expansion. *J. Math. Biol.* **28** (1990), 529-565.

[4] P. Jagers: The deterministic evolution of general branching populations. In: O. Arino, D. Axelrod & M. Kimmel (eds.), *Mathematical Population Dynamics.* Wuerz, Winnipeg, 1995; V.M. Shurenkov: On the relationship between spectral radii and Perron roots. Preprint 1992-17, Department of Mathematics, Chalmers University Göteborg, 1992.

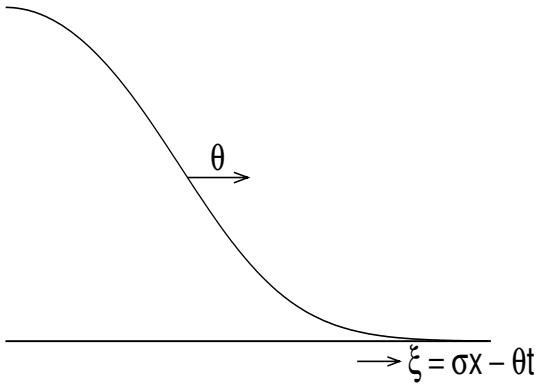

Figure 10.2: Schematic rendering of a travelling wave solution of a diffusion system, see Exercise 10.6.

10.4 Linear structured population models

Let $B(\tau, x, \xi)$ be the expected rate at which an individual born at position ξ in space produces offspring (per unit of space) at position x at age τ. Then straightforward bookkeeping considerations suggest that (10.1) be replaced by the time-translation invariant integral equation

$$u(t,x) = \int_0^\infty \int_\Omega B(\tau, x, \xi) u(t-\tau, \xi)\, d\xi\, d\tau, \tag{10.7}$$

where Ω represents the area in which the species lives (readers are invited to read the interpretation of the right-hand side of (10.7) aloud to themselves, to check their understanding of it). When $\Omega = \mathbb{R}^2$ and we assume homogeneity, B should be a function of the *relative position* $x - \xi$ rather than of x and ξ separately. When, in addition, we assume isotropy, we can consider B as a function of the *distance* $|x - \xi|$ only. Here we restrict our attention to that situation, that is, we consider (with slight notational abuse of the symbol B)

$$u(t,x) = \int_0^\infty \int_{\mathbb{R}^2} B(\tau, |x-\xi|) u(t-\tau, \xi)\, d\xi\, d\tau. \tag{10.8}$$

Tempted by our analysis of the linear diffusion equation and the robustness considerations of the preceding section, we look for travelling plane-wave solutions, i.e. we put

$$u(t,x) = w(x \cdot \nu - ct) \tag{10.9}$$

and deduce that w should satisfy

$$w(\theta) = \int_{-\infty}^\infty V_c(\zeta) w(\theta - \zeta)\, d\zeta, \tag{10.10}$$

where, by definition,

$$V_c(\zeta) = \int_0^\infty \int_{-\infty}^\infty B(\tau, \sqrt{(\zeta + c\tau)^2 + \sigma^2})\, d\sigma\, d\tau, \tag{10.11}$$

which, nota bene, does not depend on ν.

Exercise 10.7 Derive (10.10) in detail.

Exercise 10.8 Derive the characteristic equation

$$1 = \int_{-\infty}^\infty e^{-\lambda\zeta} V_c(\zeta)\, d\zeta \tag{10.12}$$

by inserting the trial solution $w(\theta) = e^{\lambda\theta}$ into (10.10).

Exercise 10.9 We give the right-hand side of (10.12) a name, i.e. we write $1 = L_c(\lambda)$, where

$$L_c(\lambda) := \int_{-\infty}^\infty e^{-\lambda\zeta} V_c(\zeta)\, d\zeta. \tag{10.13}$$

Show that

$$L_c(\lambda) = \int_{-\infty}^\infty e^{-\lambda\alpha} \int_0^\infty \int_{-\infty}^\infty e^{\lambda c\tau} B(\tau, \sqrt{\alpha^2 + \sigma^2})\, d\sigma\, d\tau\, d\alpha, \tag{10.14}$$

and show from this that

$$L_c(0) = \int_0^\infty \int_{\mathbb{R}^2} B(\tau, |\eta|)\, d\eta\, d\tau = R_0, \tag{10.15}$$

$$\frac{dL_c}{d\lambda}(0) = c \int_0^\infty \int_{\mathbb{R}^2} \tau B(\tau, |\eta|)\, d\eta\, d\tau > 0 \tag{10.16}$$

for $c > 0$, and that

$$\frac{d^2 L_c}{d\lambda^2}(\lambda) > 0 \tag{10.17}$$

for all λ and all $c \geq 0$, and finally that for every $\lambda < 0$, $L_c(\lambda)$ is a monotonically decreasing function of c with limit zero for $c \to \infty$.

Conclude from all this that, whenever $R_0 > 1$, the set $\{c$: there exists $\lambda < 0$ such that $L_c(\lambda) < 1\}$ consists of a half-line (c_0, ∞).

Exercise 10.10 Establish that c_0 can be characterized, together with the corresponding value of λ, say λ_0, as the solution of the pair of equations (10.12) and

$$\frac{dL_c}{d\lambda}(\lambda) = 0. \tag{10.18}$$

We conclude that, starting from the modelling ingredient $B(\tau, |\eta|)$, one can constructively define a minimal plane-wave speed c_0 by the pair of equations (10.12) and (10.18). It remains to ascertain that c_0 thus defined is also the asymptotic speed of propagation of disturbances. We postpone remarks on this issue to the next section, where we deal with the nonlinear problem.

Exercise 10.11 When the species considered is actually an infectious agent exploiting a host population, and if we assume mass-action contacts, we have

$$B(\tau, x, \xi) = S_0(x)A(\tau, x, \xi), \tag{10.19}$$

where A is our familiar epidemic model ingredient and $S_0(x)$ is the host density as a function of position. Do you agree? If so, check that travelling front solutions require a uniform host density S_0 (as may be an appropriate assumption within fields of agricultural crops, or, if we think of fields as host individuals, for fields within a region).

We refer to Thieme (1979)[5] for estimates of the speed of propagation using only lower bounds for S_0 and to the book by Shigesada and Kawasaki (1997; see footnote 7) for numerical studies of the speed of propagation when high- and low-density host population patches alternate.

10.5 The nonlinear situation

Nonlinearity leads to boundedness, but, under suitable assumptions, nothing much changes otherwise.

Models in population genetics, combustion and population dynamics lead to nonlinear diffusion equations

$$u_t = D\triangle u + f(u) \tag{10.20}$$

with a nonlinear function f having properties as displayed graphically in Figure 10.3.

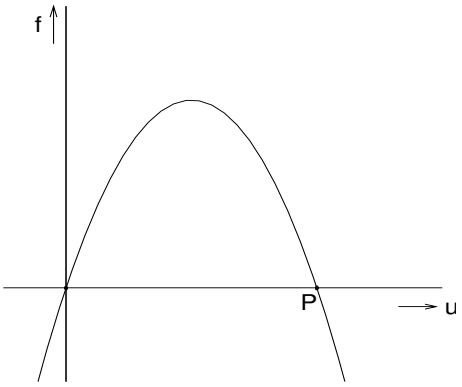

Figure 10.3: Typical shape of a function f in a nonlinear diffusion system in population biology, see Equation (10.20).

[5] H.R. Thieme: Density-dependent regulation of spatially distributed populations and their asymptotic speed of spread. *J. Math. Biol.*, **8** (1979), 173-187.

So, forgetting for a moment about space, zero is an unstable steady state and any positive perturbation ultimately leads to some stable steady state $\bar{u}$, say. Hence the considerations of Section 10.3 apply. But how should one compute c_0?

If we linearize (10.20) at $u \equiv 0$, we obtain equation (10.1) with $\kappa = f'(0)$, to which we can associate a speed c_0. Is this the right one?

To show that (10.20) has travelling plane-wave solutions for every $c \geq c_0$ is a matter of phase-plane analysis, for which we refer to the literature[6]. To show that $u(t,x)$ tends to zero outside a ball that expands with speed larger than c_0, while tending to the stable steady state $\bar{u}$ inside a ball that expands with speed smaller than c_0, is not easy but also not impossible. The intricate proof, for which we refer to Aronson & Weinberger and Diekmann & Temme (see footnote 5), involves comparison arguments based on the maximum principle. Both results involve the condition

$$f(u) \leq f'(0)u, \tag{10.21}$$

which reflects that the living conditions are optimal at very low densities. When the condition is not satisfied, for example due to an Allee effect (such as when it is more difficult to find suitable mates at low population densities), one may have so-called 'pulled' waves, the speed of which is not determined by the low-density situation. Also for this case many results are known (see the references already quoted).

We conclude that, provided the nonlinearity f satisfies certain interpretable and reasonable conditions, there is a well-defined asymptotic speed of propagation that can be calculated from the appropriate linearization.

In the epidemic context, the starting point is equation (8.23) with $S(-\infty, x) = S_0$, independent of x. Introducing (cf. the elaboration of Exercise 8.26)

$$u(t,x) = -\ln \frac{S(t,x)}{S_0} \tag{10.22}$$

we arrive at the nonlinear integral equation

$$u(t,x) = \int_0^\infty \int_{\mathbb{R}^2} S_0 A(\tau, x, \xi) g(u(t-\tau, \xi))\, d\xi\, d\tau \tag{10.23}$$

with

$$g(u) = 1 - e^{-u}. \tag{10.24}$$

We observe two things:

- the linearization at $u \equiv 0$ is of the form (10.7);
- $g(u) \leq g'(0)u$ (the 'virgin' situation is optimal for the infective agent).

[6] K.P. Hadeler & F. Rothe: Travelling fronts in nonlinear diffusion equations. *J. Math. Biol.*, **2** (1975), 251-263; A.I. Volpert, V.A. Volpert & V.A. Volpert: *Travelling Wave Solutions of Parabolic Systems.* AMS Translations of Mathematical Monographs, Vol. 140, 1994; D.G. Aronson & H.F. Weinberger: Nonlinear diffusion in population genetics, combustion, and nerve pulse propagation. In: *Partial Differential Equations and Related Topics.* J.A. Goldstein (ed.), Lecture Notes in Mathematics, Vol. 446, Springer-Verlag, Berlin, 1975, pp 5-49; O. Diekmann & N.M. Temme (eds.): *Nonlinear Diffusion Problems.* Mathematisch Centrum, Amsterdam, 1976.

J. Radcliffe & L. Rass, Spatial Deterministic Epidemics, AMS, 2003

H.R. Thieme, X-Q Zhao, Asymptotic speeds of spread and travelling waves for integral equations and delayed reaction - diffusion models, J. Diff. Equ. (2003) 195: 430 -470

Motivated by these observations, we expect that one can prove for u satisfying (10.23) that c_0 is the asymptotic speed of propagation, where

- c_0 is calculated from S_0 and A (which is, of course, assumed to depend on $|x - \xi|$ only) as in the preceding section;
- u tends to the familiar final size within any ball that expands with speed less than c_0.

That these expectations are warranted was shown in detail by Thieme and Diekmann, using comparison methods in the spirit of Aronson & Weinberger; more complicated situations were dealt with extensively in the book by Radcliffe & Rass (2003). Pioneering work had been done much earlier by Fisher, Skellam, Kolmogorov-Petrovski-Piscounov, Kendall and later Mollison (see the references in Metz & Van den Bosch (1995)[7]).

We conclude that for our familiar epidemic model, one can compute the asymptotic speed of propagation from the model ingredients S_0 and A, viz. by solving (10.12) and (10.18).

In our top-down approach we are still far from the bottom. As a next step, one should introduce parametrized families of kernels A on the basis of a mixture of mechanistic and pragmatic considerations. In addition, it is useful to derive approximation formulae for c_0 involving moments (both of time-type and of space-type) of the kernel A. Using experimental data to estimate parameters (preferably in an independent manner) and to verify predictions, one can then assess the theory[8].

10.6 Summary: The speed of propagation

Within the context of idealized models, we have unambiguously defined the (asymptotic) speed c_0 of the spatial propagation of an infection, and we have characterized c_0 in terms of the basic model ingredients in such a way that the computation of c_0 from the ingredients is rather simple. Thus we added one more indicator of the infectiousness of an agent to the list (consisting so far of R_0, r, the probability of a major outbreak, the size of a major outbreak, and the endemic level). For many ecological or agricultural systems, this is actually the most relevant indicator!

10.7 Addendum on local finiteness

Spatial position creates heterogeneity and, in particular, the contact 'intensity' of a pair of individuals depends on the spatial position of both of them, often by way of

[7] J.A.J. Metz & F. Van den Bosch: Velocities of epidemic spread. In: Mollison (1995), pp. 150-186.

[8] See, for example, F. Van den Bosch, J.A.J. Metz & O. Diekmann: The velocity of spatial population expansion. *J. Math. Biol.*, **28** (1990), 529-565; M.A. Lewis & S. Pacala: Modelling and analysis of stochastic invasion processes. *J. Math. Biol.* **41** (2000), 387-429; U. Dieckmann, R. Law & J.A.J. Metz (eds.): *The Geometry of Ecological Interactions.* Cambridge University Press, Cambridge (2000); N. Shigesada & K. Kawasaki: *Biological invasions: theory and practice.* Oxford University Press, Oxford (1997); J. McGlade (ed.): *Advanced ecological theory: principles and applications.* Blackwell Science, Oxford (1999).

the distance between them. It is this feature of spatial structure that took centre stage in the present chapter. We systematically ignored another important feature: even if a population as a whole is very, say infinitely, large, it may very well be locally small, in the sense that any individual comes into contact with only a fixed finite set of neighbouring individuals. If that is the case, we can ignore neither demographic stochasticity nor dependence (recall Section 2.3), not even in the initial stages of spatial spread.

As always, it helps to look at caricatures first. Imagine that individuals occupy the positions of an integer lattice on a line

o o o o o o o o o o o o o o o o o o o

like, for instance, cows lined up in a row in a very long stable. Assume that an individual can only infect its two immediate neighbours and that, if infected itself, it does infect a susceptible neighbour with transmission probability $p, 0 < p < 1$. Assume one cow is miraculously infected from outside. Its expected number of offspring equals $2p$ which exceeds 1 if $p > 0.5$. But from then on in every generation, an infected cow has precisely 1 susceptible neighbour, so $R_0 = p < 1$ and, we expect, only minor outbreaks can occur.

And indeed, because $p < 1$ sooner or later the right moving boundary between infected and susceptible cows comes to a halt because transmission fails. Similarly the left moving boundary comes to a halt, leaving us with a finite connected patch of cows that became infected.

The next step is to look at a regular lattice in a plane, such as the square lattice

o o o o o o o o o
o o o o o o o o o
o o o o o o o o o
o o o o o o o o o

where each individual has $N = 4$ neighbours. If an individual is infected by a neighbour, it has at most $N - 1$ susceptible neighbours. But now, in fact, it may very well happen that: i) it has even fewer than $N - 1$ susceptible neighbours; and ii) it has to compete with other nearby infected individuals of the same generation to make susceptible neighbours into offspring by infecting them. The key point is that there are multiple pathways in the lattice that connect two given points.

As a result, it makes no sense whatsoever to speak about R_0. There does not exist a 'typical' spatial configuration of already infected and still susceptible individuals, unlike in the one-dimensional lattice with only nearest neighbour transmission, where the order structure gave rise to a well-defined boundary. In the case of spatial structure, $R_0 > 1$ is locally still a necessary condition for initial spread, but it is not sufficient to also guarantee such spread.

Even if R_0 doesn't make sense, the distinction between minor and major outbreaks does. Any finite outbreak is called minor and any infinite outbreak major. If again p denotes the transmission probability, we expect that for small values of p only minor outbreaks occur, but that for p close to 1 major outbreaks are possible and

do indeed occur. That this intuition is right is proved in a rich mathematical field[9] called Percolation Theory. Moreover, in that field a key problem is to characterize the critical value p_c such that for $p > p_c$ major outbreaks are possible while for $p < p_c$ they are not (for the square lattice $p_c = 0.5$). So the idea of a threshold does survive, but we lose the easy way to characterize the threshold in terms of a quantity R_0 that is both well-defined and allows a clear interpretation. This phenomenon has been observed in a natural infection-host system[10].

In Section 12.6 we shall return to network models and discuss a variant that doesn't take the form of a regular spatial lattice. In Section 12.7 we look at approximations that, in a sense, attempt to capture the 'typical' spatial configuration of already infected and still susceptible individuals (these are usually called 'Pair Approximations').

The aim of the present short section has been twofold: i) to draw attention to the fact that local finiteness enhances demographic stochasticity as well as dependence and that, as a result, the approach based on linearization and the concept of R_0 is doomed to fail for certain classes of models; and ii) to refer to the percolation literature for an alternative approach that restores the idea of a threshold between the regime in which all outbreaks are minor and a regime in which major outbreaks do occur.

For network models see

K.Y. Leung, O. Diekmann, Dangerous connections: on binding site models of infectious disease dynamics, J. Math. Biol. (2017) 74: 619-671

I.Z. Kiss, J.C. Miller, P.L. Simon, Mathematics of Epidemics on Networks, Springer, 2017.

[9] There is an equally rich literature on the subject. See B. Bollobás & O. Riordan: *Percolation theory*. Cambridge University Press (2006), and also the many references given there.

[10] S.A. Davis, J.P. Trapman, H. Leirs, M. Begon & J.A.P. Heesterbeek: The abundance threshold for plague as a critical percolation phenomenon. *Nature*, **454** (2008), 634-637.

Putting $\phi(a) = \Lambda(a)N(a)$ and $(K\phi)(a) = N(a)(\widetilde{K}\Lambda)(a)$, we see that K thus defined is indeed the next-generation operator K with the kernel from Exercise 9.5-i. (Note that the relation between Λ and ϕ makes sense, since Λ is a force of infection and ϕ a number density.) Equality of left- and right-hand sides for the linearization would imply $R_0 = 1$. We refer to (the elaboration of) Exercises 8.27-v and 8.26 for a sketch of the proof that $R_0 > 1$ guarantees that a non-trivial solution $\Lambda > 0$ exists.

Exercise 9.14 In the formula (9.2) we replace $N(a)$ by $N(a)\mathcal{F}_v(a)$, to incorporate that only non-vaccinated individuals, and those for whom vaccination failed, are susceptible.

Exercise 9.15 If $\mathcal{F}_v(a) = 1 - v$ for $a > 0$ then $R_v = (1 - v)R_0$ and so $R_v < 1$ corresponds to $v > 1 - \frac{1}{R_0}$.

Exercise 9.16 Next let $\mathcal{F}_v(a) = 1$ for $0 < a < a_v$; and $\mathcal{F}_v(a) = 1 - v$ for $a > a_v$, then

$$R_v = R_0 - v \int_{a_v}^{\infty} \psi(\alpha) f(\alpha) N(\alpha)\, d\alpha,$$

where $\psi(\alpha)$ is defined by (9.3), and so $R_v < 1$ corresponds to

$$v > \frac{R_0 - 1}{\int_{a_v}^{\infty} \psi(\alpha) f(\alpha) N(\alpha)\, d\alpha}.$$

Exercise 9.17 i) This is Exercise 13.7-i, the special case $n = 1$, but now with $N(a)$ replaced by $N(a)\mathcal{F}_v(a)$.

ii) The short-disease approximation amounts to replacing $\psi(\alpha)$, given by (9.3), by $H(\alpha)g(\alpha)$.

17.4 Elaborations for Chapter 10

Exercise 10.1 Define $u(t, x) = w(t, x)e^{\kappa t}$; then $u_t = w_t e^{\kappa t} + \kappa u$, and consequently (10.1) can be rewritten in the form

$$w_t = D\triangle w$$

(the terms κu cancel, and then $e^{\kappa t}$ can be factored out). As you can find in any text book on partial differential equations (PDE; see e.g. Courant & Hilbert (1962)), the so-called fundamental solution of this equation is

$$w(t, x) = \frac{1}{4\pi Dt} \exp\left(-\frac{|x|^2}{4Dt}\right).$$

For completeness, however, we verify that w thus defined satisfies the PDE. To do so, we first compute $\frac{\partial}{\partial x_1} w(t, x)$:

$$\frac{\partial}{\partial x_1} \frac{1}{4\pi Dt} \exp\left(-\frac{x_1^2 + x_2^2}{4Dt}\right) = -\frac{2x_1}{4Dt} w(t, x).$$

Hence

$$\frac{\partial^2}{\partial x_1^2}\frac{1}{4\pi Dt}\exp\left(-\frac{x_1^2+x_2^2}{4Dt}\right) = -\frac{2}{4Dt}w(t,x) + \frac{4x_1^2}{(4Dt)^2}w(t,x).$$

By symmetry, we have

$$\frac{\partial^2}{\partial x_2^2}\frac{1}{4\pi Dt}\exp\left(-\frac{x_1^2+x_2^2}{4Dt}\right) = -\frac{2}{4Dt}w(t,x) + \frac{4x_2^2}{(4Dt)^2}w(t,x),$$

and consequently

$$D\triangle w = -\frac{1}{t}w + \frac{|x|^2}{4Dt^2}w.$$

Taking the derivative with respect to t we obtain

$$\begin{aligned}\frac{\partial}{\partial t}\frac{1}{4\pi Dt}\exp\left(-\frac{|x|^2}{4Dt}\right) &= -\frac{1}{4\pi Dt^2}\exp\left(-\frac{|x|^2}{4Dt}\right) + \frac{|x|^2}{4Dt^2}\frac{1}{4\pi Dt}\exp\left(-\frac{|x|^2}{4Dt}\right)\\ &= -\frac{1}{t}w + \frac{|x|^2}{4Dt^2}w,\end{aligned}$$

which upon combination with the expression for $D\triangle w$ establishes that w satisfies the PDE.

Exercise 10.2 This is most easily achieved by applying a transformation to polar coordinates:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(x_1^2+x_2^2)}\,dx_1\,dx_2 = \int_0^{\infty}\int_0^{2\pi} e^{-ar^2} r\,d\varphi\,dr = 2\pi\frac{-1}{2a}e^{-ar^2}|_0^{\infty} = \frac{\pi}{a}.$$

So when $a = \frac{1}{4Dt}$ this yields $4\pi Dt$. But w has an additional factor $\frac{1}{4\pi Dt}$ and we conclude that $\int_{\mathbb{R}^2} w(t,x)\,dx = 1$. Hence $\int_{\mathbb{R}^2} u(t,x)\,dx = e^{\kappa t}$.

Exercise 10.3 For $t \downarrow 0$ we have $\frac{1}{4\pi Dt} \uparrow \infty$. However, when $|x| \geq \varepsilon$, the factor $\exp(-\frac{|x|^2}{4Dt})$ goes to zero much faster than $\frac{1}{4\pi Dt}$ goes to infinity, and hence $\lim_{t\downarrow 0} u(t,x) = 0$ for $|x| \geq \varepsilon$.

Exercise 10.4 Define $\xi = \xi(t,x) = x\cdot\nu - ct = x_1\nu_1 + x_2\nu_2 - ct$; then $\frac{\partial\xi}{\partial t} = -c$ and $\frac{\partial\xi}{\partial x_1} = \nu_1$, $\frac{\partial\xi}{\partial x_2} = \nu_2$. Hence, by the chain rule,

$$\frac{\partial}{\partial t}w(\xi) = -cw'(\xi), \qquad \frac{\partial}{\partial x_1}w(\xi) = \nu_1 w'(\xi), \qquad \frac{\partial^2}{\partial x_1^2}w(\xi) = \nu_1^2 w''(\xi),$$

and by symmetry

$$\frac{\partial^2}{\partial x_2^2}w(\xi) = \nu_2^2 w''(\xi).$$

Inserting (10.3) into (10.1) and using these identities, we find $-cw' = Dw'' + kw$ (since, ν being a unit vector, $\nu_1^2 + \nu_2^2 = 1$), which is (10.4). Equation (10.4) is a second-order, linear, homogeneous differential equation. Standard theory (see e.g. Hale (1969)) tells us that we should look for two (linearly independent) solutions.

Substitution of $w(\xi) = e^{\lambda\xi}$ into (10.4) yields (10.5), as the factor $e^{\lambda\xi}$ cancels. Hence (10.6) holds.

Exercise 10.5 For uniform (i.e. x-independent) solutions, (10.1) predicts exponential growth at rate k. If we consider a travelling wave with minimal velocity, however, and measure the growth rate at an arbitrary position, we find that it is twice as large. The mechanism is spillover: population density is substantially larger to the left than it is to the right, and so diffusion contributes to net growth rate at any position.

Exercise 10.6 For w we find the equation

$$Dw'' + (c - \theta)w' + kw = 0,$$

and so we should have $c - \theta \geq c_0$, which amounts to $c \geq c_0 + \theta$. If we replace σ by $-\sigma$, that is, if we look for solutions of the form

$$u(t, x) = w(-\sigma \cdot x - ct),$$

we find for w the equation

$$Dw'' + (c + \theta)w' + kw = 0,$$

and so we should have that $c + \theta \geq c_0$, which amounts to $c \geq c_0 - \theta$. If $\theta > c_0$, the minimal wave speed in the direction $-\sigma$ is *negative*! This has to be interpreted as lack of spread in this direction. In other words, the growth of the species is confined to a region that is blown off towards infinity in the σ-direction.

Exercise 10.7 We first perform the transformation $\eta = x - \xi$ of the integration variable in (10.8):

$$u(t, x) = \int_0^\infty \int_{\mathbb{R}^2} B(\tau, |\eta|)\, u(t - \tau, x - \eta)\, d\eta\, d\tau$$

(here we have used that $(-1)^2 = 1$). Now we substitute (10.9) and obtain

$$w(x \cdot \nu - ct) = \int_0^\infty \int_{\mathbb{R}^2} B(\tau, |\eta|)\, w(x \cdot \nu - \eta \cdot \nu - ct + c\tau)\, d\eta\, d\tau.$$

Next, call $x \cdot \nu - ct = \theta$ and write

$$w(\theta) = \int_0^\infty \int_{\mathbb{R}^2} B(\tau, |\eta|)\, w(\theta - \eta \cdot \nu + c\tau) d\eta d\tau.$$

The final step consists of introducing a tailor-made coordinate system. We supplement the unit vector ν by the orthogonal unit vector $\nu^\perp$, which is chosen such that the determinant of the matrix $(\nu\nu^\perp)$ equals one. Let α and σ be coordinates relative to the basis defined by ν and $\nu^\perp$, i.e. let

$$\eta = \alpha\nu + \sigma\nu^\perp;$$

then

$$w(\theta) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty B(\tau, \sqrt{\alpha^2 + \sigma^2})\, w(\theta - \alpha + c\tau)\, d\alpha\, d\sigma\, d\tau.$$

With the change of variable $\zeta = \alpha - c\tau$, we obtain from this

$$w(\theta) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty B(\tau, \sqrt{(\zeta + c\tau)^2 + \sigma^2})\, w(\theta - \zeta)\, d\zeta\, d\sigma\, d\tau,$$

which leads to

$$w(\theta) = \int_{-\infty}^\infty V_c(\zeta)\, w(\theta - \zeta)\, d\zeta$$

after interchanging the order of integration.

Exercise 10.8 All you have to do is to cancel the common factor $e^{\lambda\theta}$ on both sides of the equality.

Exercise 10.9 Substituting (10.11) into (10.13), we find

$$L_c(\lambda) = \int_{-\infty}^\infty e^{-\lambda\zeta} \int_0^\infty \int_{-\infty}^\infty B(\tau, \sqrt{(\zeta + c\tau)^2 + \sigma^2})\, d\sigma\, d\tau\, d\zeta,$$

which, upon reversing the substitution $\alpha = \zeta + c\tau$ made earlier, can be written as

$$L_c(\lambda) = \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-\lambda\alpha} e^{\lambda c\tau} B(\tau, \sqrt{\alpha^2 + \sigma^2})\, d\sigma\, d\alpha\, d\tau,$$

which, essentially, is (10.14). By putting $\lambda = 0$ and reversing the coordinate transformation $\eta = \alpha\nu + \sigma\nu^\perp$, we arrive at the first identity of (10.15), while the second is a direct consequence of the interpretation of B and R_0. If we differentiate with respect to λ, we get two terms, one involving an additional factor $-\alpha$ and the other involving an additional factor $c\tau$. The first of these terms vanishes by symmetry considerations (whatever the function f, $\int_{-\infty}^\infty \alpha f(\alpha^2)\, d\alpha = 0$, since $\int_{-\infty}^0 \alpha f(\alpha^2)\, d\alpha = \int_\infty^0 -\sigma f(\sigma^2)\, d(-\sigma) = -\int_0^\infty \sigma f(\sigma^2)\, d\sigma$). These observations, together with those leading to (10.15), give the equality in (10.16), and the inequality (10.17) is then a direct consequence of the non-negativity of B and the (implicitly!) assumed non-negativity of c.

If we differentiate twice with respect to λ, we obtain three terms involving additional factors α^2, $-2\alpha c\tau$ and $c^2\tau^2$, respectively. For reasons of symmetry, the middle one of these factors vanishes, and we arrive at the conclusion that this second derivative is positive.

The assertion (10.18) is a straightforward consequence of the fact that all the c dependence is concentrated in the factor $\exp(\lambda c\tau)$.

For $c = 0$, the convex function $\lambda \mapsto L_c(\lambda)$ achieves its minimum at $\lambda = 0$ and, when $R_0 > 1$, therefore $L_c(\lambda) > 1$ for all λ. By the property (10.19) the set $\{c :$ 'there exists $\lambda < 0$ such that $L_c(\lambda) < 1$'$\}$ is non-empty and, moreover, contains $[\bar{c}, \infty)$ whenever it contains $\bar{c}$. Hence this set is either the whole line or a half-line. Since zero does not belong to the set, it must be a half-line. In line with earlier notation, we call the boundary point c_0.

Exercise 10.10 L_c is a convex function of λ that, for $c = c_0$, 'touches' the level one, but does not dip below this level. Equation (10.12) reads $L_c(\lambda) = 1$ and equation (10.18) says we should have a minimum when varying λ. Together therefore, the two equations state that the minimum should be one. To make more precise assertions, we have to distinguish the case where the minimum is actually attained for a finite value of λ from the case where L_c has, as a function of λ, an infimum at minus infinity. This involves the behaviour of $L_c(\lambda)$ for $\lambda \to -\infty$. This behaviour is, in turn, determined by the competition between the factor $\exp(-\lambda\alpha)$, tending to $+\infty$ for $\alpha > 0$, and the factor $\exp(\lambda c\tau)$, tending to zero. Hence it is determined by the support of the function B (i.e. the interval(s) where $B > 0$). We think that it is helpful to be aware of the possibility that c_0 is characterized by $\lim_{\lambda\to-\infty} L_{c_0}(\lambda) = 1$, but that this is kind of exceptional and that it makes good pragmatic sense to try and determine c_0 and λ_0 from $L_c(\lambda) = 1$ and (10.18).

Exercise 10.11 Ignoring the reduction of the density of susceptibles, the incidence is the product of the force of infection A and the host population density S_0. So, with the right interpretation of the word 'offspring' in the definition of B, we arrive at (10.21). In order for B thus defined to only depend on $|x - \xi|$, it is necessary that, among other things, S_0 is constant as a function of x.

17.5 Elaborations for Chapter 11

Exercise 11.1 An age representation is not applicable to macroparasitic infections since infected individuals can receive additional doses of the infectious agent after the start-up dose. In the microparasite case the fast reproduction after the start-up dose will cause further incoming doses of the same agent to go 'unnoticed' in the large amount already built up inside the host. Consequently, we can neglect this influence. In the macroparasite case the additional doses make the crucial difference between various infected individuals, and both the number and size of the doses received and the timing of these re-infections determines the impact of the parasite on the host and the contribution of the host to the spread of infection in the population.

Exercise 11.2 i) Invasion in the 'virgin' situation is still characterized as a 'best case' (from the point of view of the parasite). In other words, we are in the situation where the transmission process can proceed optimally. For microparasites the key point is that no infectious material enters hosts that are already infected. For macroparasites one should interpret 'adverse conditions' as anything that hinders development of parasites. This means, for example, that parasites do not compete with each other for resources within the host, or interact with the immune system of the host, but can each optimally produce eggs to be shed into the environment. In other words, there are no density-dependent feedback effects acting on the parasite in any of the stages of its life cycle, i.e. there is no nonlinear interaction between parasites, or between parasites and the environment they inhabit (including hosts).

ii) In the invasion phase there are no density-dependent constraints, and parasite densities in all stages of the life cycle will be very low. Also there will be few infected hosts, and the chances of infecting the same host more than once will be rather small in