The SIR model: Kermack and McKendrick (1926)

The assumption is the population has fixed size N consisting of three groups: the Infected I, the Susceptibles S and the Recovered R. Thus

$$I(t) + S(t) + R(t) = N \qquad \text{for all } t \geq 0 \tag{1}$$

The equations are

$$\begin{aligned} \frac{dS}{dt} &= -\beta\, S\, I \\ \frac{dI}{dt} &= \beta\, S\, I - kI \\ \frac{dR}{dt} &= kI \end{aligned} \tag{2}$$

and the assumption that at $t = 0$,

$$R(0) = 0, \quad S(0) = S_0, \quad I(0) = I_0 = N - S_0 \tag{3}$$

Some jargon: the ratio $p = k/\beta$ is the *relative removal rate*, and $c = \beta/k$ is the *contact number*.

Remark: A little care is needed here to include equation (1) in the implementation. One can certainly scale the equations in (2) by dividing by N and obtain $[S, I, R]$ as ratios. In this case the product SI occurring in the first two equations needs clarification. Some references will write these as $\frac{dS}{dt} = -\beta\frac{S\,I}{N}$ and $\frac{dI}{dt} = \beta\,\frac{S\,I}{N} - kI$ for this reason. However, the $\frac{1}{N}$ can be absorbed into the constant β as we have done; it just changes the units of this rate (or probability).

First thing to notice is that if $\vec{x} = \begin{bmatrix} S \\ I \\ R \end{bmatrix}$ then the system (2) can be written as $\frac{d\vec{x}}{dt} = F(\vec{x})$ and is autonomous. Checking for critical points we see that this means $S\,I = 0$, $\beta S\,I - kI = 0$ and $kI = 0$ so that the line $I = 0$ is critical. The Jacobian matrix for the system (2) is

$$J = \begin{bmatrix} -\beta I & -\beta S & 0 \\ \beta I & \beta S - k & 0 \\ 0 & k & 0 \end{bmatrix}$$

If we linearise the second equation in (2) about the initial point $S = 1$ then $I'(t) = (\beta - k)I$ and so if $\beta > k$ then I will grow exponentially originally.

The following are obvious conclusions from equations (2).

Since $\beta > 0$, $S \geq 0$, $I \geq 0$, we have $\frac{dS}{dt} \leq 0$ so that $S(t)$ is monotonically decreasing in t. Also, $R(t)$ is monotonically increasing in t. The sign of $\frac{dI}{dt}$ depends on the sign of $\beta S - k$ and $\frac{dI}{dt}$ is positive only when $S > k/\beta = p = 1/c$.

Thus since $0 \leq R$, $S \leq N$ it follows from the monotonicity that both $R(t)$ and $S(t)$ must tend to a limit as $t \to \infty$; call these R_∞ and S_∞. Since $I(t) = N - R(t) - S(t)$, we must also have I tending to a limit I_∞ as $t \to \infty$.

The number R_∞/N is the proportion that gets the disease and provides a measure of the intensity of the epidemic.

The autonomous nature of the basic model equation allow for several important observations

$$\frac{dS}{dR} = -\frac{\beta\, S\, I}{kI} = -cS \tag{4}$$

provided that I is not yet zero. This is easily solved to give

$$S(R) = S_0 e^{-cR} \tag{5}$$

Since $R(t) \leq R_\infty \leq N$, we have $e^{-cR} \geq e^{-cN}$ so that

$$S(t) \geq S_0 e^{-cN}. \tag{6}$$

This is a crucial prediction of the model: $S_\infty > 0$. There will always be some individuals who escape the disease; the epidemic will die out but not because there are no Susceptibles left.

Now we also have

$$\frac{dI}{dS} = -\frac{(\beta S - k)I}{-\beta S I} = -1 + \frac{p}{S} \tag{7}$$

This is a separable equation with solution

$$I(t) = -S(t) + p\log(S) + C \tag{8}$$

for some constant C. At time $t - 0$ there are $I(0) = I_0$ Infectives and S_0 Susceptibles so that $C = I_0 + S_0 - p\log S_0 = N - p\log S_0$ and so

$$I(t) := g(S) = N - S(t) + p\log\Big(\frac{S(t)}{S_0}\Big) \tag{9}$$

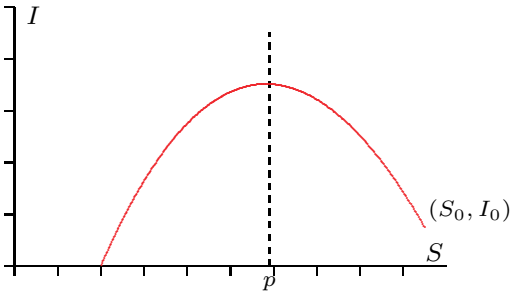
This means the orbits for the autonomous system lie along the curve given by equation (9). Also, since $g(S_0) = I_0 > 0$ and $lim_{S\to 0^+}\, g(S) = -\infty$ the curve crosses the line $I = 0$ at some positive value of S less than S_0. Since all critical points for the system (2) lie on the line $I = 0$ this orbit must approach the point $(S_\infty, 0)$ as $t \to \infty$. Thus $I_\infty = 0$.

We also have from (7) that

$$g''(S) = \frac{d^2 I}{dS^2} = -\frac{p}{S^2} \tag{10}$$

showing the graph of I as a function of S is concave downwards and reaches its maximum where $\frac{dI}{dS} = 0$ which, from equation (7) this occurs when $S = p$

The figure to the left shows this relationship between S and I. The initial state is (S_0, I_0) and if this point falls to the left of the line $S = p$ then $I(t)$ decreases monotonically to zero; there is no epidemic. However, if $S_0 > p$ then the number of Infectives initially increases until $S = p$ after which the number of Infectives again approaches zero.



To locate the value of S_∞, since $I_\infty = 0$ we have $S_\infty = N - R_\infty$ and using equation (5) we have, letting $t \to \infty$

$$S_\infty = S_0 e^{-cR_\infty} \tag{11}$$

and thus $S_\infty = S_0 e^{-c(N - S_\infty)}$ so that S_∞ is determined as a solution of the equation

$$f(x) := S_0 e^{-cN} e^{cx} - x = 0. \tag{12}$$

Note that $f(0) = S_0 e^{-cN} > 0$ and $f(N) = S_0 - N < 0$ so given the fact that f is continuous the *Intermediate Value Theorem* shows there is at least one number x^* between 0 and N for which $f(x^*) = 0$, that is at least one positive root of equation (12). However, an easy calculation shows that $f'(x) = c[f(x) + x] - 1$ and so $f'(x^*) = cx^* - 1$. Since $x^* = S_\infty < p$, $f'(x^*) < 0$. Also, $f''(x) = S_0 c^2 e^{-c(N-x)}$ and this is always positive.

If now there were two roots then *Rolle's Theorem* guarantees there is a point where the derivative is zero between them. But the derivative is negative at both roots and the second derivative is always positive so the derivative f' must always be negative between the roots. This is a contradiction, showing the existence of a unique root x^*. We have shown

Theorem. Threshold Theorem of Epidemiology. *If $S_0 < k/\beta$ then $I(t)$ goes monotonically to zero. If $S_0 > k/\beta$ then then the number of Infectives increases initially and then tends monotonically to zero. The limit $lim_{t\to\infty} S(t)$ exists and is given by the unique root of the equation $f(x) = 0$.*

We now look to derive an equation for $R(t)$. Rewrite $R'(t) = kI$ from equation (2) as $R'(t) = k(N - S - R)$ we obtain

$$\frac{dR}{dt} = k\big[N - R - S_0 e^{-cR}\big] \tag{13}$$

Although this is separable, the integrations are messy and we resort to an approximation that will suffice. We replace the exponential term in (13) by the the first three terms of its Maclaurin series, $e^x = 1 + x + \frac{1}{2}x^2$ with $x = -cR$ so that

$$\frac{dR}{dt} \approx k\big(N - R - S_0(1 - cR + \frac{1}{2}c^2R^2)\big).$$

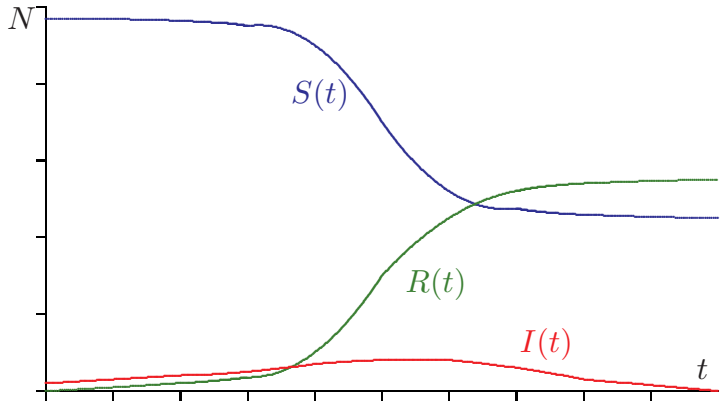
We now assume that the initial number of infectives is small so we can take $S_0 \sim N$. Then $R'(t) = 0$ gives the approximation for R_∞

$$k\big(N - R_\infty - S_0 + cS_0R_\infty - \frac{1}{2}c^2S_0R_\infty^2)\big) \approx 0$$

which occurs when

$$R_\infty \approx 2p\big(1 - \frac{p}{S_0}\big)$$

This gives an approximation to the total number who have contacted the disease.



There are many deficiencies in this model, mostly omissions of common characteristics caused by a quest for sufficient simplicity to allow analysis of the dynamics. In particular, in order to keep the equations autonomous the rate constants β and k cannot depend on time and must hold the same value for all individuals. The three "compartments" S, I and R allow only the transition $S \to I \to R$ so that there is no possibility of someone who is recovered playing any further role – in particular they cannot be re-infected. One reading of the constants β and k in equations (2) is that they are rates by the fact that they couple a rate of change of one state into another. An

alternative viewpoint is they are probabilities in the sense that a certain proportion of one state will transition into the next each time period, typically measured in days so that for example going from $S \to I$ has this simple interpretation. The much-quoted Ro value - the ratio β/k or $contact\ number\ c$ is then interpreted as the total number of individuals an infected person infects. The ultimate strategy of epidemic management is lower this figure; in particular, to ensure it remains less than unity.

Also, there is no possibility of someone exposed to the disease going to any state but the infected. The next extensions to the basic model takes this into account. Many infectious diseases have an incubation period before being infectious during which the host cannot yet spread the disease.

The SIRS model

A standard model to allow for re-infection has the simple modification The equations are

$$\begin{aligned} \frac{dS}{dt} &= -\beta\, S\, I + \eta R \\ \frac{dI}{dt} &= \beta\, S\, I - kI \\ \frac{dR}{dt} &= kI - \eta R \end{aligned} \tag{14}$$

so that there is simply now an interchange from $R \to S$ allowed with probability η.

The SEIRD model

We now assume the existence of an $exposed\ state\ E$ that fits in with $S \to E \to I \to R$ that allows for an incubation period. In short, to avoid adding in a delay to the term β which would destroy the autonomous nature we add another state and with this another coupling constant σ.

The new system is then

$$\begin{aligned} S' &= -\beta\, S\, I \\ E' &= \beta\, S\, I - \sigma E \\ I' &= \sigma\, E - kI \\ R' &= kI \end{aligned} \tag{15}$$

Again it is clear that $(1, 0, 0, 0)$ is an equilibrium point. A linearisation of (15) about this point gives the pair

$$E' = -\sigma E - \beta I, \qquad I' = \sigma E - kI \tag{16}$$

Now the Jacobian matrix of (16) is $J = \begin{bmatrix} -\sigma & \beta \\ \sigma & k \end{bmatrix}$ with eigenvalues

$$\lambda = \frac{1}{2}[-(\sigma + k) \pm \sqrt{(\sigma - k)^2 + 4\sigma\beta}.$$

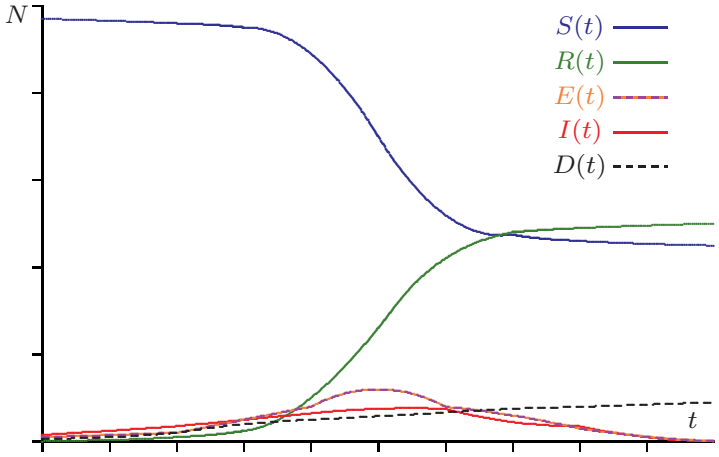
The dominant eigenvalue is λ_1 and this is positive if $\beta > k$ giving an initial exponential growth.

Before proceeding further there is an issue that we might address at the same time. As described, our epidemic lumps together into the single state R all those coming from the state I; *recovered* means actual recovery as well as death. Since this last possibility is one of critical interest a further state D is added but the long chain model is modified by a branch; the infected class can transition

into either a true recovered state R or a state D and the probability of $I \to D$ is α (and so that from I to R is then $1-\alpha$). However, our rate (or the time in days it takes) out of the infected state k might be different for each of the branch transitions; we retain k for entry into R but a new constant ρ for entry to D

$$\begin{aligned} S' &= -\beta\, S\, I \\ E' &= \beta\, S\, I - \sigma E \\ I' &= \sigma\, E - (1-\alpha) k\, I - \alpha\rho\, I \\ R' &= (1-\alpha) k\, I \\ D' &= \alpha\rho\, I \end{aligned} \tag{17}$$

Note that the stable points of this system as well that in equation (15) all lie along the line $I = 0$, indicating that any steady state solution will result in $I_\infty = 0$. The graph of each of these compartment magnitudes $[S, E, I, R, D]$ is shown in the figure below.



Adding more compartments

The ability to add further elements in the above was at the expense of adding further compartments. I done in moderation from a computational perspective this is quite negligible.

Other states that could be added might be to accommodate differences in the constants β and k within subgroups of the population but it is usually hard to make sufficiently accurate measurements to allow for this. An exception for the current virus is the death state which varies with gender and dramatically with age. However, the state D is an end state of the chain and from a pure epidemic standpoint does not alter the I or S categories.

The model based on ODE systems have no space component and cannot take into account in population density variations or in population movement. Dynamical system models that include space and movement within become partial differential equations and the diffusion of particles (or people) is governed by a random walk process such that at the macroscopic level becomes a *reaction-diffusion* equation and the classical form in one space variable is

$$\frac{\partial u}{\partial t} - a\frac{\partial^2 u}{\partial x^2} + d\frac{\partial u}{\partial x} = f(x,t,u) \tag{18}$$

Here a is the diffusion coefficient (and it can depend on x), d is the "drift coefficient" indicating a net movement in one direction. In more than one space variable this becomes

$$\frac{\partial u}{\partial t} - a\triangle = f(x,t,u) \tag{19}$$

The above brings in considerable complexity and computation cost and the "fix" is to assume that we have m regional compartments in each of which one of the above SIR models holds with its own local values of β and k then to allow transitions between these m compartments. [One problem though is the assumption that each regional compartment C_i has a fixed number N_i of total individuals.] If one such regional compartment has a high contact number then prohibiting movement from this region will reduce the number of new infectives being added elsewhere – quarantining.

The state E in addition to providing a transition between the I and S states allows a certain latency into the picture. Most epidemics are short-lived in comparison to additions/subtractions due to births and deaths changes so time-dependence for such factors can often taken to be negligible. However, this is not always the case and in fact the management strategy is to reduce the contact number (typically through β) - which means essentially changing it with time. If one can live with a model using a piecewise constant function $\beta = \beta(t)$, then simply restart the equation system after each jump in β.

However, there is one significant factor that has not been included that can have considerable significance; that of time delays. An ODE $x'(t) = f(t, x; \tau)$ with time delay in its simplest form just replaces the argument in x as appearing in f by $f\big(x(t-\tau)\big)$ for some $\tau > 0$. The inclusion in epidemic models makes a great deal of sense as the lag between actually getting infected and able to subsequently reinfect can be days and furthermore the time taken to be recognised that an individual has moved from the S class to the I can be longer. This can cause consider error in estimating the progression of the epidemic.

In practice a fixed delay time τ affects the standard solution process of solving ordinary differential equations in which a solution value at time t is moved to one at $t + \delta t$ for some suitably chosen/computed time step δt. There is no need to store previous steps within the computation. This changes with delay where enough previous time steps have to be stored in order to recover the value $x(t-\tau)$ for use in the integrator step.

A fixed delay may not be appropriate; for example the ability of an exposed individual to re-infect may vary considerably over the delay period. In this case the delay term is of the form $\int_{t_0}^{t} \phi(t, s)x(s)\, ds$ for some spread function ϕ that has to built in. A simpler case is $\phi(t, s) = \phi(s)$ that is the delay process is the same at all times t throughout the epidemic and the simple delay by a fixed amount can be obtained by taking $\phi(s) = \delta(s - \tau)$ where $\delta(s)$ is the standard Dirac-δ function. When this integral term is included one gets an *integro-differential equation.* The computational costs now rise. There is also the issue of finding the right model for ϕ.