SIR models of epidemics

Contributed by Sebastian Bonhoeffer; adapted for BioSym by Stefan Schafroth

Infectious diseases are a major cause of death worldwide, and have in the past killed many more people than all the wars (think, for instance, of the Spanish flu). Mathematical modelling of infectious diseases was initiated by Bernoulli in 1760. The work of Kermack and McKendrick, published in 1927, had a major influence on the modelling framework. Their SIR model is still used to model epidemics of infectious diseases. We will study this basic model, and some of its extensions.

1. Introduction

The SIR model tracks the numbers of susceptible, infected and recovered individuals during an epidemic with the help of ordinary differential equations (ODE). The model can be coded in a few lines in MATLAB. We will learn how to simulate the model and how to plot and interpret the results. We will use simulation to verify some analytical results. We will plot time courses, phase diagrams and contour plots. The following concepts and methods can be illustrated with the SIR model:

- Basic reproductive ratio (R0)
- Herd immunity
- Numerical simulation of ordinary differential equations
- Graphical tools (phase portrait, contour plot)

2. Mathematical Model

2.1 Description of the model

A major assumption of many mathematical models of epidemics is that the population can be divided into a set of distinct compartments. These compartments are defined with respect to disease status. The simplest model, which was described by Kermack and McKendrick in **1927,** consists of three compartments: susceptible (S), infected (I), recovered (R). We will later discuss some possible modifications and extensions to this simple scheme.

Susceptible Individuals that are susceptible have, in the case of the basic SIR model, never been infected, and they are able to catch the disease. Once they have it, they move into the Infected compartment.

Infected Infected individuals can spread the disease to susceptible individuals. The time they spend in the infected compartment is the infectious period, after which they enter the recovered compartment.

Recovered Individuals in the recovered compartment are assumed to be immune for life.

The SIR model is easily written using ordinary differential equations (ODEs), which implies a deterministic model (no randomness is involved, the same starting conditions give the same output), with continuous time (as opposed to discrete time). Analogous to the principles of reaction kinetics, we assume that encounters between infected and susceptible individuals occur at a rate proportional to their respective numbers in the population. The rate of new infections can thus be defined as βSI, where β is a parameter for infectivity. Infected individuals are assumed to recover with a constant probability at any time, which translates into a constant per capita recovery rate that we denote with r, and thus an overall rate of recovery rI. Based on these assumptions we can draw the scheme of the model:

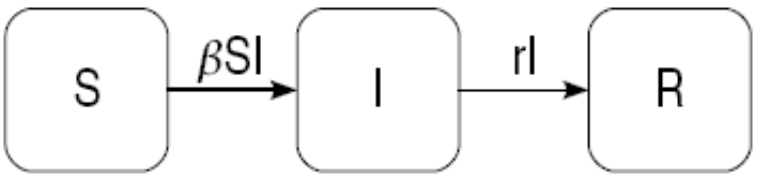

Figure 1: Scheme of the basic SIR model. Boxes represent compartments, and arrows indicate flux between the compartments.

The scheme can also be translated into a set of differential equations:

$$\frac{dS}{dt} = -\beta S I$$

$$\frac{dI}{dt} = \beta S I - r I$$
$$\frac{dR}{dt} = r I \quad (1)$$

Using this model, we will consider a mild, short-lived epidemic, e.g. influenza, in a closed population. Closed means that there is no immigration or emigration. Moreover, given the time scale of influenza epidemics, we will not consider demographic turnover (birth or death), and all infections are assumed to end with recovery. The size of the population (S+I+R) is therefore constant and equal to the initial population size, which we denote with the parameter N.

Let us now consider a population which is naive with respect to the disease we are considering. What happens if a single infected individual is introduced into such a population? Is there going to be an epidemic? How many people will be infected? We will answer these questions by implementing and simulating the model in MATLAB.

2.2 Tips to develop the SIR model

Let us now implement the model in MATLAB, using the `ode45` command to numerically solve differential equations. The script

- `sir.m`

provided on the web page will also help you to plot the results as in Fig. 2. Run the model with the preset parameters. Now take some time to think about the interpretation of the simulation. What is the population size with these settings? What is the time frame of the simulation? Is it possible to tell? What defines the time scale? Would you expect to get such a smooth curve in a real life epidemic?

Questions: Explore the model: play around with initial conditions and parameters and see what happens in the simulations (a more systematic analysis will be performed later). Observe the effect of each parameter and the possible courses of simulated epidemics. Is it possible to simulate a sustained epidemic in this model?

Answers...

3. Questions that can be investigated

- What conditions are necessary for the outbreak of an epidemic?
- What fraction of a population is going to be infected in a transient epidemic?
- Can partial vaccination in a population protect against the outbreak of an epidemic?

Advanced questions: Modify the model to

- allow for the loss of immunity
- model treatment of the disease
- model the emergence of drug resistance and find the optimal rate of treatment
- model longer time scales that allow for the birth and death of individuals

3.1 Basic exercises

A key parameter in epidemiology is the basic reproductive ratio, R_0. It is defined as the average number of secondary cases transmitted by a single infected individual that is placed into a fully susceptible population. In other words, R_0 tells us about the initial rate of spread of the disease. Hence, if $R_0 > 1$, there will be an epidemic, and if $R_0 < 1$, the introduced infecteds will recover (or die) without being able to replace themselves by new infections. In this model, it is pretty easy to derive R_0. The disease-free state corresponds to: S=N, I=0, R=0. If one infected individual appears in the population, there will be an epidemic if and only if $dI/dt > 0$. By replacing S with N in Equation 1, this yields $\beta N/r > 1$. That is:

$$R_0 = \beta N r \quad (2)$$

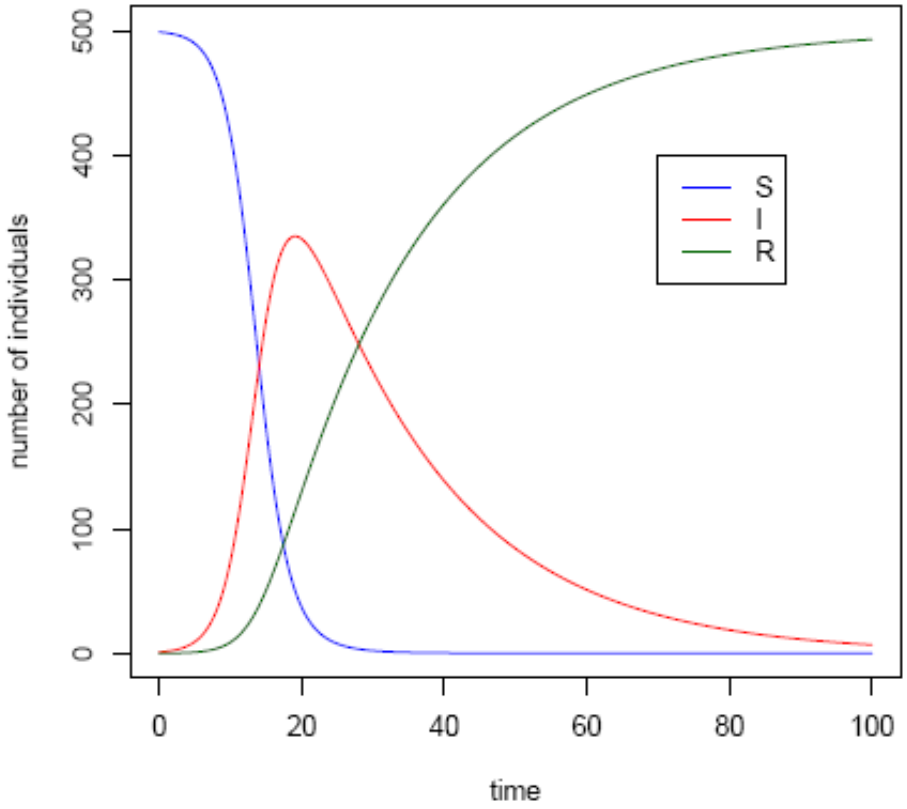

Figure 2: An epidemic simulated in the SIR model.

Eb1★: Check this formula by simulating the model for different sets of parameters. Fix N, and vary β and r. Choose your values such as to have combinations with both $R_0 > 1$ and $R_0 < 1$, as predicted by Equation 2. Because it is the ratio of the two parameters that matters, it is better to use a geometric, rather than an arithmetic sequence for the parameter values. Run the model for each parameter combination (Hint: use a loop structure, e.g. for, iterated over a pre-defined vector of parameter values; you will need two embedded loops for the two parameters), and check and record whether there has been an epidemic or not. For this, you can compare the initial and maximum number of infecteds: was there growth above the initial level? You can perform a simple "automated" test to verify the key role of R_0. For each parameter combination, record a logical value whether the maximum number of infecteds was above the initial, and another logical value whether R_0 was above 1. Store these values in two vectors and then check the identity of the two vectors.

Eb2★: After the simple logical test, now take advantage of MATLAB's graphical capabilities to visualize the result. Repeat the simulations, and now store the maximum number of infecteds for each parameter combination. A simple way to do so is to record the numbers in a matrix with rows and columns corresponding to the appropriate values in the vectors of the parameters that are being varied (i.e. β and r). The number of rows and columns in this matrix will thus equal the number of elements in the vectors that contain the values employed for the two parameters. Use the matrix and the two parameter vectors to plot the number of infections during an epidemic according to β and r with the help of the filled.contour function. Finally, superimpose the graph of R_0 as calculated by the analytical formula in Equation 2 on your plot. If your script is OK, this will fit!

3.2 Advanced/additional exercises

4. Glossary

Compartment models: models where the population is divided between several spatial compartments or classes, which are connected by the flow (migration or transformation) of individuals.

SIR models: models where the population is divided into 3 classes - susceptible individuals are uninfected and susceptible to the disease; infected individuals are infected and can infect susceptibles; recovered individuals have recovered from the infection and are immune to re-infection.

Basic reproductive ratio: the key parameter of epidemiology. It represents the number of secondary cases initiated by the introduction of a single infected individual into a 'naive' population.

Phase portrait: also called a phase diagram, it shows the temporal evolution of two or three variables of a system. It consists of trajectories plotted in a coordinate system that has axes corresponding to the variables of the system.

Herd immunity: immunity of a population to the outbreak of an epidemic, provided by the immunity of only a fraction of the individuals.

5. Literature & Weblinks

- Kermack, W. and McKendrick, A., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. London A 115, 700-721. Proc. R. Soc. London A 115, 700-721.
- Anderson, R. M. and May, R. M. 1991. Infectious Diseases of Humans. Oxford. Oxford University Press
- Matt Keeling's article in Plus: The mathematics of diseases (open access)
- Compartment models of epidemiology in Wikipedia