SIR Model

D. Sulsky

June 21, 2012

The diseases we are discussing have been classified as microparasitic. Examples of microparasitic diseases are chicken pox, measles, HIV/AIDS, influenza and tuberculosis. Diseases due to worms, for example, are called macroparasitic. Other than the size of the infecting agent, the main distinction is that the microparasites reproduce within their hosts and are transmitted directly from one host to another. Most macroparasites, on the other hand, have a somewhat more complicated life cycle, often with a secondary host or carrier. Schistosomiasis is an example of a macroparasitic disease. Schistosomiasis is passed only through contact with water in which live snails that can incubate the disease causing helminths. In reality, the distinction between micro- and macro-parasitic diseases is not always clear and there is a range of infecting agents and transmission paths. In the case of malaria, the disease is not passed from human to human but through the anopheline mosquitoes, but some of the parasite's life cycle occurs in the mosquito and some in humans. When a mosquito bites an infected individual, it sucks the gametocytes, the sexual forms of the parasite, along with blood. These gametocytes continue the sexual phase of the cycle within the mosquito gut and the sporozoites that develop then fill the salivary glands of the infested mosquito. When this female mosquito bites another human for a blood meal, the sporozoites are inoculated into the blood stream of the fresh victim, thus spreading the infection.

Microparasitic diseases (viral or bacterial) can pass through a population like a flame through fuel, and useful analogies have been made between epidemics and chemical reactions. A theory of these epidemics was derived by W.O. Kermack, a chemist, and A. G. McKendrick, a physician, working at the Royal College of Surgeons in Edinburgh between 1900 and 1930. The simple model we considered this morning, is based on their work. One important result of theirs is that an infection determines a threshold size for the susceptible population, above which an epidemic will propagate. Their theoretical epidemic threshold is observed in practice,

and it measures to what extent a real population is vulnerable to the spread of an epidemic.

The basic model due to Kermack and McKendrick is deterministic. It can be extended to include diseases with intermediate hosts, diseases with stratified populations (families, preschools, schools, social groups with mixing probabilities). A simple stochastic model that describes the spread of disease through random sampling, is due to L. Reed and W. H. Frost at Johns Hopkins in the 1920s. They never published their model but it has been described in the literature by others, [5]. There are connections between these two simple models to explore.

We've been examining diseases caused by viruses and bacteria that live and reproduce within their host and are transmitted through contacts between hosts. We've made the distinction between sick individuals who harbor the disease and those who are as yet healthy. Thus, we've divided the population into subclasses of susceptibles, S, and infectives, I. We will add a third class of removed individuals, R, who can no longer contract the disease because they have recovered with immunity, have been placed in isolation, or have died. This model is now called an SIR model, and is attributed to the classic work on the theory of epidemics done by Kermack and McKendrick (1927). Each of the classes of individuals is assumed to consist of identically healthy or sick individuals.

Population Classes in the SIR model:

Susceptible: capable of becoming infected

Infective: capable of causing infection

Recovered: removed from the population: had the disease and recovered, now immune, immune or isolated until recovered, or deceased.

We make the same assumptions as in the discrete model:

1. The population is fixed.
2. After recovery, immunity is conferred.
3. The gain in the infective class is at a rate proportional to the number of infectives and susceptibles, aSI, $a > 0$. Susceptibles are lost at the same rate.
4. The rate of removal of infectives to the removed class is proportional to the number of infectives, bI, $b > 0$.

5. The incubation period is short enough to be negligible. (A susceptible who contracts the disease is infective right away.)

6. The various groups are uniformly mixed.

Now, we allow changes to occur continuously in time, rather than in discrete steps. The set of nonlinear, ordinary differential equations for this disease model is

$$\begin{aligned} \frac{dS}{dt} &= -aSI \\ \frac{dI}{dt} &= aSI - bI \\ \frac{dR}{dt} &= bI \end{aligned} \tag{1}$$

with initial conditions, $S(0) = S_0 > 0$, $I(0) = I_0 > 0$ and $R(0) = 0$. Note that the parameter a has units of one over time per individual; but the parameter b has units of one over time. In this model, these parameters characterize the disease.

There are various approaches one can take to understanding the predictions of this model and the behavior of its solutions. Kermack and McKendrick derived an approximate solution for the removal rate, dR/dt, in terms of a hyperbolic secant function and then fit parameters and found the model to compare well with data for death by plague in Bombay during an epidemic in 1906. We will also look at qualitative methods that tell us properties of the solution (without explicitly calculating the solutions). These methods are standard tools in the analysis of ordinary differential equations. Finally, we can compute approximate solutions by numerical methods.

A primary question is, given, a, b, S_0 and I_0, when will there be an epidemic? First, look for $I(t) > I_0$ for some $t > 0$. This condition says that the infective population at some time t is larger than the initial number, and could indicate an epidemic. Look at the derivative of I initially,

$$\frac{dI}{dt}|_{t=0} = I_0(aS_0 - b). \tag{2}$$

We see that the infective population increases initially if $I_0(aS_0 - b) > 0$ and decreases initially if $I_0(aS_0 - b) < 0$. Since $I_0 > 0$, we can state this observation in terms of S_0. If $S_0 < b/a \equiv \rho$ then the infective population decreases initially, and if $S_0 > b/a \equiv \rho$, the infective population increases initially. Also note that always $dS/dt < 0$. Thus, the susceptible population always decreases. In particular, $S(t) < S_0$. We can

conclude that if $S_0 < \rho$ then $S(t) < \rho$ for all time. Thus, $dI/dt \leq 0$ for all time, in which case $I(t) < I_0$, and the infection dies out (no epidemic).

On the other hand, if $S_0 > \rho$ then $I(t)$ increases initially and we have the chance for an epidemic. These observations describe what is known as a threshold phenomenon. Call $S_c = \rho = b/a$ the critical value of the susceptible population. If the initial susceptible population is larger than this value, $S_0 > S_c$ then there is an epidemic, otherwise, there is not. The susceptible population must be large enough for an epidemic to occur.

Let's try to understand this threshold better. The ratio $\rho = b/a$ is sometimes called the relative removal ratio and $\sigma = a/b$ is called the contact ratio. Note that b has the units of 1/time and gives the removal rate from the infective class. Thus, the average period of infectivity is $1/b$. The ratio a/b is the fraction of the population that comes into contact with an infective individual during the period of infectiousness.

An important quantity that gives a rudimentary understanding of a disease is the basic reproductive rate of the infection, $R_0 = aS_0/b$. The basic reproductive rate is the number per unit time of secondary infections produced by one primary infection in a wholly susceptible population. If $R_0 > 1$ (more than one secondary infection from a primary infection) then an epidemic will occur.

In further analyzing the model, we can take into account the particularly convenient fact that the total population, $N(t) = S(t) + I(t) + R(t)$, is actually constant and equal to $N_0 = S_0 + I_0$. (Recall, we took $R(0) = 0$.) This fact means that one variable, say $R(t)$, can always be eliminated so that the model can be given in terms of two equations and two unknowns, $S(t)$ and $I(t)$. Qualitative analysis for such problems can be examined using a phase plane. We first ask, are there any equilibrium solutions to this model? We see that the line $I = 0$ is a line of equilibrium points. The equilibrium susceptible population $\bar{S}$ can be any value and is determined as a function of the initial conditions. For example, two equilibria are, $\bar{I}_1 = 0$, $\bar{S}_1 = N_0$ ($\bar{R}_1 = 0$) and $\bar{I}_2 = 0$, $\bar{S}_2 = b/a$ ($\bar{R}_2 = N_0 - b/a$). In the first case, the whole population is healthy, but susceptible, and there is no disease. In the second case, part of the population remains susceptible after the disease, the rest has recovered. The disease runs its course by exhausting the infective population ($\bar{I}_2 = 0$), not by exhausting the susceptible population, as one might guess *a priori*.

We can derive a useful analytical result by considering I as a function of S. From equation (1), we obtain

$$\frac{dI}{dS} = \frac{aSI - bI}{-aSI} = -(1 - \frac{b}{aS}) = -1 + \frac{\rho}{S}, \tag{3}$$

provided $I \neq 0$. Integrating this equation gives the (I, S) phase-plane trajectories as

$$I + S - \rho \ln S = \text{constant} = I_0 + S_0 - \rho \ln S_0. \tag{4}$$

Figure 1 is a sketch of the phase trajectories. If an epidemic exists, we would like

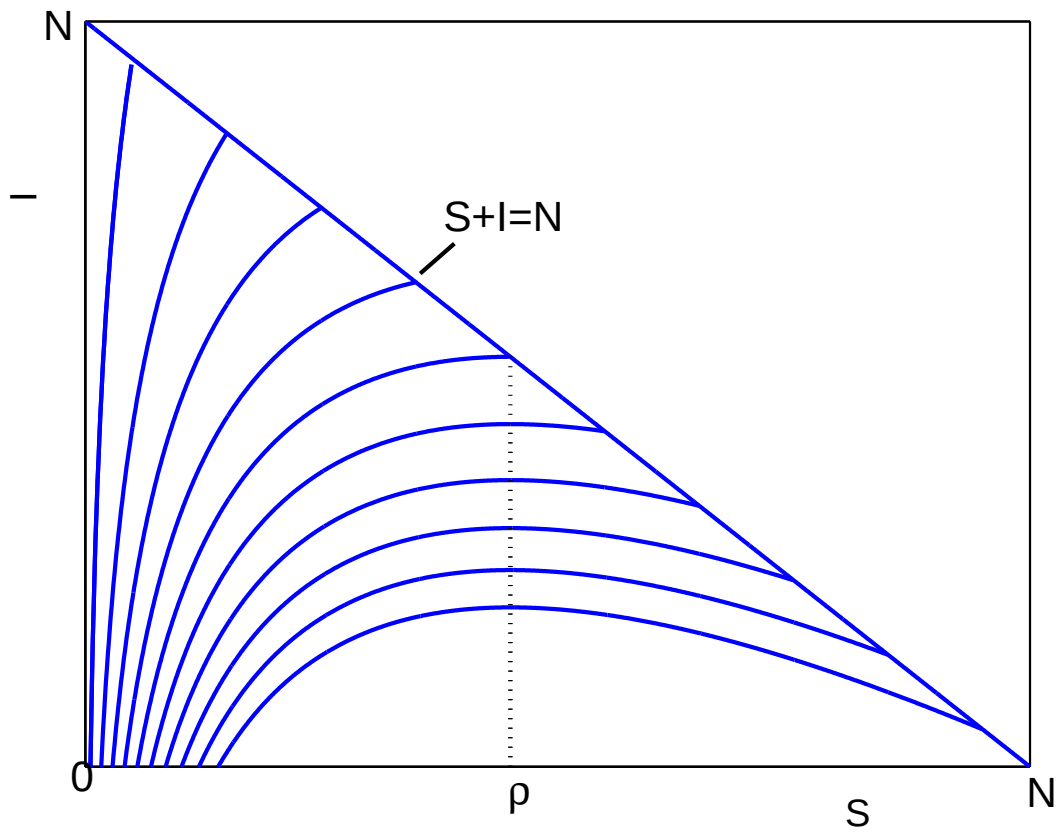

Figure 1: Phase trajectories for the SIR epidemic model. The curves are determined by the initial conditions $I(0) = I_0$ and $S(0) = S_0$. With $R(0) = 0$, all of the trajectories start on the line $S + I = N$ and remain within the triangle since $0 < S + I \leq N_0$ for all time. An epidemic situation exists if $I(t) > I_0$ and this always occurs if $S_0 > \rho$ and $I_0 > 0$.

to know how severe it is. From equation (3), the maximum of I, $I_{\max}$, occurs when

$S = \rho$, where $dI/dt = 0$. Using equation (4) with $S = \rho$,

$$\begin{aligned} I_{\max} &= \rho \ln \rho - \rho + I_0 + S_0 - \rho \ln S_0 \\ &= I_0 + (S_0 - \rho) + \rho \ln \frac{\rho}{S_0} \\ &= N_0 - \rho + \rho \ln \frac{\rho}{S_0}. \end{aligned} \tag{5}$$

For any initial values, I_0 and $S_0 > \rho$, the phase trajectory starts with $S > \rho$ and we see that I increases from I_0 and hence an epidemic ensues. It may not necessarily be a severe epidemic, as is the case if I_0 is close to $I_{\max}$. It is also clear that if $S_0 < \rho$ then I decreases from I_0 and no epidemic occurs.

Since the axis $I = 0$ is a line of singularities, on all trajectories $I \to 0$ as $t \to \infty$. From equation (1), S decreases since $dS/dt < 0$ for $S \neq 0$ and $I \neq 0$. Also from equation (1)

$$\begin{aligned} &\frac{dS}{dR} = -\frac{S}{\rho} \\ &\quad \to S = S_0 \exp[-R/\rho] \geq S_0 \exp[-N_0/\rho] > 0 \\ &\quad \to 0 < S(\infty) \leq N_0. \end{aligned} \tag{6}$$

In fact, from the first equation in (1), $0 < S(\infty) < \rho$. Since $I(\infty) = 0$, and the total population is constant, we have $R(\infty) = N_0 - S(\infty)$. Thus from equation (6) we have

$$S(\infty) = S_0 \exp[-R(\infty)/\rho] = S_0 \exp(-(N_0 - S(\infty))/\rho], \tag{7}$$

and so $S(\infty)$ is the positive root of the transcendental equation

$$S_0 \exp\left[-\frac{N - z}{\rho} \right] = z. \tag{8}$$

We then get the total number of susceptibles who catch the disease in the course of the epidemic as

$$I_{\text{total}} = I_0 + S_0 - S(\infty), \tag{9}$$

where $S(\infty)$ is the positive solution of equation (8). An important implication of this analysis, namely that $I(t) \to 0$ and $S(t) \to S(\infty) > 0$, is that the disease dies out from lack of infectives and not from the lack of susceptibles. Thus, under the assumptions of this model, the disease cannot wipe out the entire population.

The threshold result for an epidemic is directly related to the relative removal ratio ρ. If $S_0 > \rho$ there is an epidemic and if $S_0 < \rho$ there is not. For a given disease, the relative removal ratio varies with the community and hence determines

whether an epidemic may occur in one community and not in another. The number of susceptibles, S_0, also plays a role. For example, if S_0 is high, and the removal rate a of infectives is low (through ignorance, lack of medical care, inadequate isolation, etc.) then an epidemic is likely to occur. The maximum number of infectives is $I_{\max}$, given above, and the total number who get the infection is I_{total}, also given above.

It is worth stopping to reflect on the model we've been studying and examine our choices. What if we modeled a disease like a predator-prey model? Keshet [1] examined this idea and pointed out the flaws. Let x be the population of human hosts, and y be the viral population.

The assumptions are that

1. There is a constant human birth rate α.
2. Viral infection causes an increase in mortality due to disease, so $g(y) > 0$.
3. Reproduction of viral particles depends on human presence.
4. In the absence of human hosts, the virus particles 'die' or become nonviable at rate γ.

The equations would be

$$\begin{aligned} \frac{dx}{dt} &= (\alpha - g(y))x \\ \frac{dy}{dt} &= axy - \gamma y. \end{aligned} \tag{10}$$

This approach leads to a modified Lotka-Volterra predation model. The viruses are predatory organisms searching for human prey to consume. This model is not realistic for several reasons. The total viral population is huge and is not really relevant. What matters is the distribution of these viruses over the hosts since that determines what percentage of people will actually suffer from the disease. Moreover, there is an underlying assumption in this model that viruses roam free in the environment, randomly encountering new hosts. This assumption is rarely true. Diseases spread by contact or close proximity between infected and healthy individuals. So, it is the interaction between classes of individuals that is key.

What criticisms do you have of the simple SIR model? If the disease is not of short duration, then the equation for the susceptible population should include birth and death terms. Mortality due to other causes than the disease should be included in the equation for the infective population, and in the removed class. We have also assumed that everyone in the susceptible population is equally likely to catch the disease. This assumption might not be valid since some may be more susceptible than

others. Also, what if complete immunity or only temporary immunity results from having the disease? Moreover, many diseases have a latent or incubation period in which a susceptible has become infected but is not yet infectious. Measles for example has an 8-13 day latent period. The latent period for AIDS has been found to be anywhere from a few months to years after the individual has been shown to have antibodies for HIV. We can model latency either with a delay equation or by introducing another class. There are even more complications that can be considered in the model - changes due to immigration, susceptibility depending on age or weight, or other factors which might also keep classes of individuals from mixing completely. For global pandemics, we surely must also look at the spatial spread of the disease. We have also not considered macroparasitic diseases that are transmitted through a secondary host. Such diseases are modeled differently since it is not I that is important, but the distribution of parasites within the infective class.

Things to try:

1. Set $I = 0$ in equation (4) and use Newton's method to find S. Plot results as a function of aS_0/b.

2. As noted some diseases have a latent period. Suppose new recruits from the susceptible class go into an exposed class, but not yet infectious class, which we denote by $E(t)$.

   (a) Derive a mathematical model for this disease using either difference equations or differential equations. Consider two cases:

      i. A certain proportion of the exposed group becomes infective at each sampling time ($dE/dt = -rE$).

      ii. Individuals remain in the exposed class for a fixed time, say T ($dE/dt = rI(t)S(t) - rI(t-T)S(t-T)$)).

   (b) Derive a threshold result for the system in either of these cases.

3. Schistosomiasis is a disease caused by a wormlike parasite (a helminth). Male and female helminths must mate in a host (eg. humans, ducks or swine). Thereafter, some of the fertilized eggs leave the hosts in their feces. When an egg comes into contact with fresh water, it hatches and attempts to find a snail that it can penetrate. Once a snail is infected, a large number of larvae are produced. They swim freely in search of a host in which to reproduce. It might penetrate the skin of a host or be ingested with water or food grown in the water. We sample the populations at regular intervals, and we denote by H_n

the mean number of worms infecting each host, I_n is the number of infected snails. Then

$$\begin{aligned} H_{n+1} &= (1-\mu)H_n + cI_n \\ I_{n+1} &= (1-\delta)I_n + b\frac{(S-I_n)H_n^2}{1+H_n} \end{aligned} \tag{11}$$

where μ and δ are death probabilities for hosts and snails, c is the number of helminths becoming established in a host due to each infected snail, S is the number of snails (fixed) and the term $bH_n^2/(1+H_n)$ gives the mean number of paired worms per infected host and so is proportional to the number of eggs produced. The last term describes the interaction between eggs and susceptible snails.

Consider the steady-state distribution of helminths and infected snails: $H_n = H$ and $I_n = I$. The equations for these values are

$$\begin{aligned} H &= (1-\mu)H + cI \\ I &= (1-\delta)I + b\frac{(S-I)H^2}{1+H} \end{aligned} \tag{12}$$

(a) Describe the static states of the schistosomiasis model.

(b) Which of these solutions is stable?

References

[1] Keshet, Leah, "Mathematical Models in Biology." Random House, 1988.

[2] Hoppenstadt, F.C. and C.S. Peskin, "Mathematics in Medicine and the Life Sciences." Springer-Verlag, 1992.

[3] Murray, J.D., "Mathematical Biology." Springer-Verlag, 1989.

[4] Kermack, W.O. and A.G. McKendrick, "A Contribution to the Mathematical Theory of Epidemics." Proceedings of the Royal Society of London. Series A, 115:700-721, 1927.

[5] Abbey, H, " An examination of the Reed Frost theory of epidemics." Human Biology, 24:201-233, 1952.