

CSPeskin 2/25/2021

Epidemics: The SIR Model

Consider a closed population (no births, no deaths, no migration) and a non-fatal disease to which people are immune once they have recovered.

Let

$S(t)$ = number of susceptible people

$I(t)$ = number of infected people

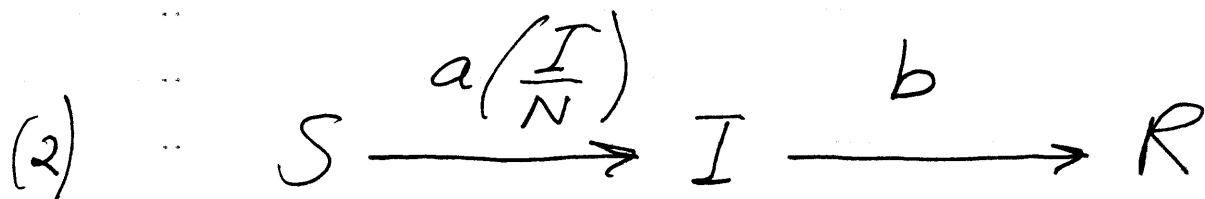
$R(t)$ = number of recovered people

at time t . Then

$$(1) \quad S(t) + I(t) + R(t) = N$$

where N is the constant number of people in the population.

The process of infection and recovery can be diagrammed as follows



The constants a and b have units of $1/\text{time}$. I will call a the infectivity of the disease and b the recovery rate. Their meanings will be discussed more fully below.

The interpretation of the diagram (2) is that the number of new infections per unit time is

(3)
$$a\left(\frac{I}{N}\right)S$$

and the number of people who are recovering per unit time is

(4) bI

Note in each case that the number on the arrow (sometimes called a "rate constant") has to be multiplied by the number of people available to undergo a transition to get the overall rate at which the transition is occurring.

Given the diagram (2) and its interpretation (3-4), it is straightforward to write down differential equations that govern $S(t)$, $I(t)$, and $R(t)$. They are

$$(5) \quad \frac{dS}{dt} = -a\left(\frac{I}{N}\right)S$$

$$(6) \quad \frac{dI}{dt} = a\left(\frac{I}{N}\right)S - bI$$

$$(7) \quad \frac{dR}{dt} = bI$$

Note that by adding these equations we get

$$(8) \quad \frac{dN}{dt} = 0$$

which just confirms that the population size is constant.

Interpretation of parameters:

We start with b , since its interpretation is simpler than that of a

The easiest way to think about b is that

(9) $\frac{1}{b} =$ average time it takes to recover

Another way is to imagine a population in which everyone is infected at $t=0$. Since there are no susceptible people in such a population, we only need equations (6-7), and they become

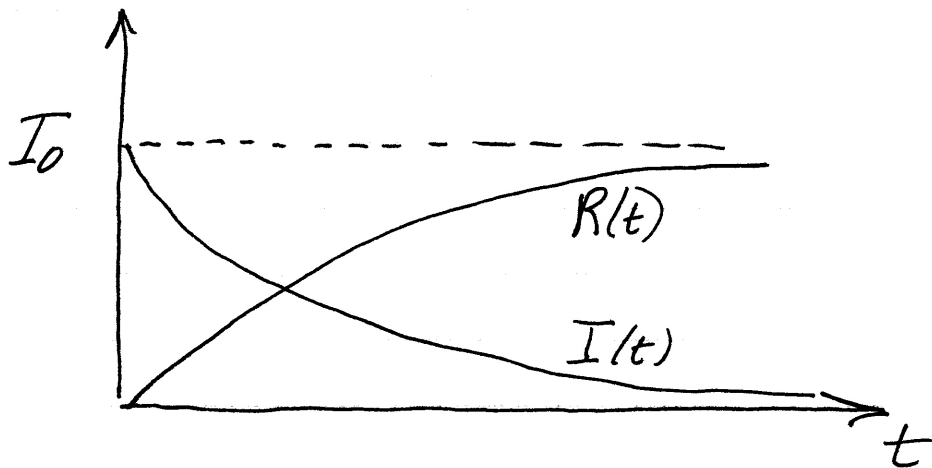
(10) $\frac{dI}{dt} = -bI$, $I(0) = I_0$

(11) $\frac{dR}{dt} = bI$, $R(0) = 0$

Equations (10-11) are solved by

$$(12) \quad I(t) = I_0 e^{-bt}$$

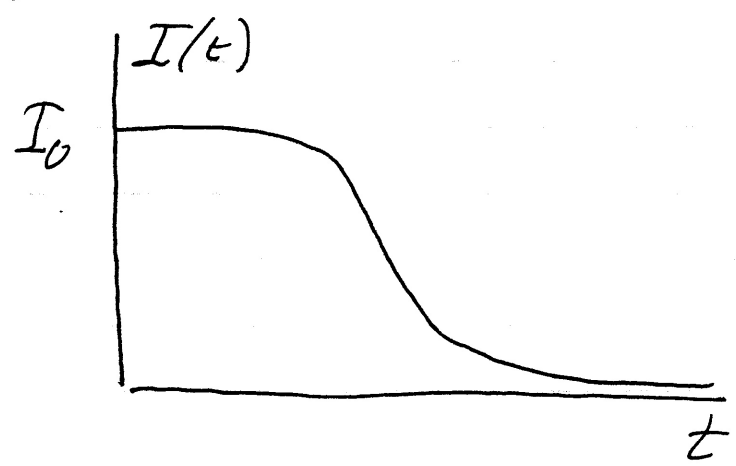
$$(13) \quad R(t) = \int_0^t b I(t') dt'$$
$$= I_0 \int_0^t e^{-bt'} b dt'$$
$$= I_0 (1 - e^{-bt})$$



Thus, b is the rate constant for the exponential decay of the number of infected people in a population where everyone is initially infected.

The equation $\frac{dI}{dt} = -bI$ is based

on the unrealistic assumption that recovery at any time is equally likely regardless of how long a person has been infected. In a population with everyone initially infected,* a more realistic result would look like this



*with all of the infections being initiated at $t=0$

One way to get such a result is to assume that recovery is an m -step process

$$(14) \quad I^{(1)} \xrightarrow{mb} I^{(2)} \xrightarrow{mb} \dots I^{(m)} \xrightarrow{mb} R$$

Here the average time it takes to recover is still $1/b$, but the shape of the plot of

$$(15) \quad I(t) = \sum_{k=1}^m I^{(k)}(t)$$

will be like the more realistic curve sketched on the previous page.

What happens here as m increases with b fixed is interesting, and especially the limit $m \rightarrow \infty$ is very interesting. See if you can figure this out, either analytically or by numerical simulation.

An advantage of (14) is that a different infectivity can be attributed to each of the stages $I^{(k)}$. You might consider this in a project, but here we consider only the case $m=1$.

Next, we consider the infectivity parameter a
It can be written as

(16) $a = a_0 P_{trans}$

where a_0 is the number of encounters that a typical person in the population has per unit time, where an encounter is defined as an interaction between two people that could result in transmission of the pathogen. On any one encounter, transmission will only occur if one person is infected and the other is susceptible, and then it will only occur with probability P_{trans} .

The parameter a is strongly influenced by behavior. Staying home as much as possible reduces a_0 , and wearing a mask reduces P_{trans} (for an airborne pathogen).

Recall that the rate of new infections per susceptible person is given by

$$(17) \quad a \left(\frac{I}{N} \right) = a_0 P_{\text{trans}} \left(\frac{I}{N} \right)$$

In this expression, the factor I/N is the probability that a person chosen at random from the population is infected. Thus, the overall probability of transmission during one encounter is

$$(18) \quad P_{\text{trans}} \left(\frac{I}{N} \right)$$

and this explains why the factor (I/N) appears as it does in our model.

An important remark related to this is that the SIR model is often written with (a/N) combined into a single parameter so that N does not appear explicitly in the equations. This makes no fundamental difference if N is constant (although of course it makes a huge difference in the parameter value if

.. N is large), but if N is varying
 .. (because of births, deaths, or migration),
 .. or even if we just want to compare
 .. two populations with different N , then
 .. it makes a difference whether we think
 .. of a as constant or a/N as constant.

.. When we say that a is constant, which is
 .. the viewpoint adopted here, we are
 .. saying that a_0 , the number of encounters
 .. experienced by one person per unit time,
 .. is independent of the population size. This
 .. seems only reasonable for large populations,
 .. since no one person interacts with the
 .. population as a whole, but rather with
 .. a circle of friends the size of which does
 .. not depend on the size of the whole population.
 .. On the other hand, it is certainly true
 .. that population density does influence
 .. the number of encounters that a given person
 .. has per day. Thus, by keeping a constant
 .. as N varies we are essentially
 .. saying that the population density is
 .. remaining the same.

The most important parameter of the SIR model is neither a nor b separately, but rather their ratio

(19) $\frac{a}{b} = \text{reproduction number}$

This is usually denoted R_0 , but I am avoiding that notation here to avoid confusion with the number of recovered people $R(t)$.

To see why (a/b) is the reproduction number, consider the number of infections that are caused directly by a single infected person in an otherwise susceptible population.

The infected person has an amount of time $1/b$ to spread the disease, and during that time the rate at which new infections occur is

(20) $a \left(\frac{1}{N} \right) (N-1) \approx a$

since $I=1$ and $S=N-1$. Of course,

$I(t)$ is changing as new infections occur, but the fraction $1/N$ remains constant anyway because we are only counting infections caused directly by the one initially infected person. The factor $(N-1)$, however, should really be $N-I(t)$, but this does not matter because

(21)
$$\frac{N-I(t)}{N} \approx 1$$

when $I(t) \ll N$, which is the situation we are considering here.

Thus a is the rate at which new infections are caused directly by a single infected person in an otherwise susceptible population, and since the time interval in which these transmission events can occur is $1/b$, the total number of infections caused is $a(1/b) = \frac{a}{b}$, the reproduction number.

Of course, the actual number of people infected directly by a single infected person in an otherwise susceptible population is an integer, and moreover it is not a fixed number but instead is a random variable.

What a/b actually gives is the expected value, or mean, of this random variable. Thus a/b is not necessarily an integer.

Intuitively, if $(a/b) > 1$ an epidemic will occur, and if $(a/b) < 1$, it will not. As we will see later, this intuition is correct.

Deterministic simulation

We can solve equations (5-7) numerically by Euler's method

$$(22) \quad \frac{S(t+\Delta t) - S(t)}{\Delta t} = -a \frac{I(t)}{N} S(t)$$

$$(23) \quad \frac{I(t+\Delta t) - I(t)}{\Delta t} = a \frac{I(t)}{N} S(t) - b I(t)$$

$$(24) \quad \frac{R(t+\Delta t) - R(t)}{\Delta t} = b I(t)$$

Note that

$$(25) \quad \begin{aligned} S(t+\Delta t) + I(t+\Delta t) + R(t+\Delta t) \\ = S(t) + I(t) + R(t) \end{aligned}$$

so the population size does not change

```

.. clean all; close all;
.. % 7 days to recover
.. % reproduction number 2.3
.. % unit of time is one day

.. b = 1/7 ; a = 2.3 * b ;
.. N = 1000 ; I = 3 ; R = 0 ; S = N - I - R ;
.. tmax = 200 % days
.. dt = 1/24 % one hour
.. clockmax = tmax / dt
.. for clock = 1 : clockmax
..   t = clock * dt ;
..   newI = dt * a * (I/N) * S ;
..   newR = dt * b * I ;
..   S = S - newI ;
..   I = I + newI - newR ;
..   R = R + newR ;
..   tsave(clock) = t ;
..   Ssave(clock) = S ;
..   Isave(clock) = I ;
..   Rsave(clock) = R ;
.. end
.. subplot(2, 1, 1), ...
.. plot(tsave, Ssave, tsave, Isave, tsave, Rsave)
.. subplot(2, 1, 2), plot(tsave, Isave)

```


Notes :

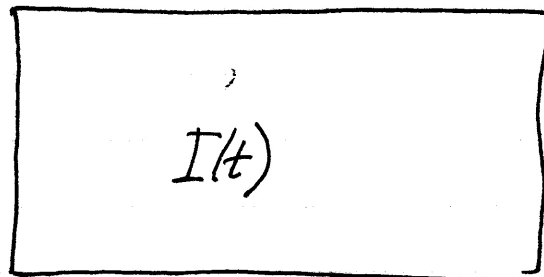
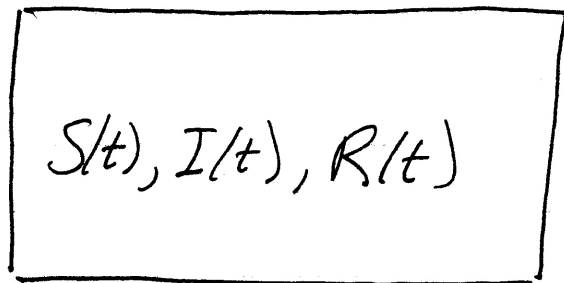
- Although the code will work as written, it would be better to pre-allocate the arrays that save results for future plotting. Therefore, insert the following right after clockmax has been defined :

```

tsave = zeros(1, clockmax);
Ssave = zeros(1, clockmax);
Isave = zeros(1, clockmax);
Rsave = zeros(1, clockmax);

```

- The subplot commands put ^{the} two plots in one figure :



(sometimes needed to make a better scale for $I(t)$.)

Stochastic simulation

Two unrealistic aspects of the foregoing are

- 1) S, I, R are considered real numbers, but they should be integers
- 2) The model is deterministic, so the randomness of the real world is not considered.

To overcome these limitations, we reinterpret the SIR model in terms of individuals and probabilities.

Consider a single susceptible person during a short time interval of duration dt . The average number of encounters that this person has during that time interval is

$$(26) \quad a_0 dt$$

and if the time interval dt is so short that the probability of more than one encounter is negligible, then $a_0 dt$ is also the

probability of one encounter, and $1 - a_0 dt$ is the probability of no encounter.

If an encounter occurs, the probability that the susceptible person becomes infected is $P_{trans} (I/N)$,

and therefore the probability that a particular susceptible person becomes infected during a time interval of duration dt is

$$(27) \quad a_0 dt P_{trans} \frac{I}{N} = (dt) a \frac{I}{N}$$

Similarly, the probability that a particular infected person will recover during a short time interval of duration dt is given by

$$(28) \quad (dt) b$$

Note that the probabilities given by (27) & (28) are just the rate constants from the diagram (2) multiplied by the duration dt of the time interval. Here we do not multiply by the numbers of people

available to undergo the transition, since we are going to make a random decision for each such person individually.

It is important to understand that the probabilities given by (27) & (28) are approximate, and that the approximation is only a good one if the probability is much smaller than 1. The approximation gets better and better as dt is made smaller. It might seem that as $dt \rightarrow 0$ nothing will happen, since the probabilities approach 0, but this is incorrect. As $dt \rightarrow 0$, it does become more and more unlikely that anything will happen in any one interval of duration dt , but there are more and more such intervals, and this compensates for the low probability of anything happening in any one interval.

It is also very important to understand that the rate constants themselves are not probabilities. They have units of $1/\text{time}$, and their numerical value depends on the choice of units. In particular

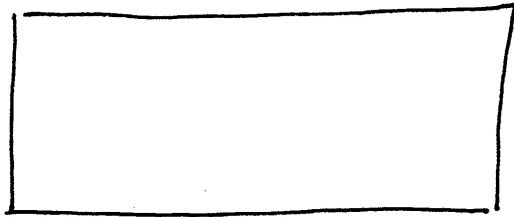
the numerical value of a rate constant can be any positive number and there is nothing wrong with that number being greater than 1. For example, if it takes a week to recover, then

$$(29) \quad b = (1/7)/\text{day} = 1/\text{week} = 52/\text{year} \\ = 5200/\text{century}$$

A very common mistake is to confuse a rate constant with a probability and to assume, therefore, that it has to be smaller than one.

Now that we have the probabilities of infection and recovery in a time interval dt as given by (27-28), the next question is how to make something happen with a specified probability p in a computer program. This is accomplished as follows:

if (rand < p)



end

In Matlab, `rand` generates a random number with the uniform distribution on the interval $(0, 1)$. This means that the probability that `rand` is in any subinterval of $(0, 1)$ is equal to the length of that subinterval. In particular, the probability that `rand` lies in $(0, p)$ is equal to p . Therefore, whatever code is in the box will be executed with probability p , and not executed with probability $1-p$.

Therefore, we can convert our previous SIR program into a stochastic simulation by taking out the two lines that define newI and newR , and instead evaluate these quantities in the following way

```

newI = 0
for s = 1:S
    if rand < dt * a * (I/N)
        newI = newI + 1
    end
end
newR = 0
for i = 1:I
    if rand < dt * b
        newR = newR + 1
    end
end
end

```

Thus, we look at each susceptible person and decide randomly whether that person becomes infected during the current time step, and we look at each infected person and decide randomly whether that person recovers during the current time step.

A vectorized way to get the same result is as follows:

```
if S > 0
    newI = sum(rand(S, 1) < dt * a * I / N);
else
    newI = 0;
end
```

```
if I > 0
    newR = sum(rand(I, 1) < dt * b);
else
    newR = 0;
end
```

This should be much faster!

Since logical expressions evaluate to 0 ("false") or 1 ("true"), sum is counting the number of instances in which the random number was less than the specified probability.

Since newI and newR are integers, S, I, R change in steps that are integers, and hence they remain integers if they are integers at $t=0$.

Also, repeated runs of the same program will not give the same result.

(This assumes, however, that the repeated runs are done within the same Matlab session. If you re-start Matlab, you will get the same result as before, since the random numbers are only pseudo-random, and Matlab is re-generating them again by the same algorithm. You can defeat this by choosing a different seed for the random number generator.)

Since the deterministic and stochastic programs are so similar, it is very straight forward to compare their results, and this could be the focus of a project. When N is large, we

Expect the deterministic and stochastic results to be very similar. For smaller N , the stochastic results should look like a noisy version of the deterministic results. By averaging a large number of stochastic results for the same problem we can remove the noise, and the averaged result of many stochastic simulations can be compared to the deterministic result for the same problem. I think that the averaged stochastic result will show systematic deviations from the deterministic result, but that these deviations will be small.