Understanding R\_0

Chris H. Wilson, University of Florida

March 20, 2020

# Introduction

If you are anything like me, you have been reading a *lot* of articles about the virology and epidemiology of SARS-COV2, the agent behind the new disease Covid-19. Our knowledge about this virus is developing rapidly - thankfully - however the scale of the public health crisis we are facing is daunting. Rather than attempt any kind of comprehensive analysis here, I want to help unpack one of the key pieces of terminology that is being thrown around a lot, “R naught” () or the “basic reproduction number”. This will be a mid-level mathematical exposition, intended to make this epidemiological concept intuitive to those with a decent generic quantitative fluency, but no special training in disease ecology or related theory. It may even be useful to those who have studied some population and disease ecology, but are feeling rusty or never fully understood the logic behind the models.

The basic “toy model” in epidemiology is the so-called “SIR” model, standing for Susceptible/Infected/Recovered. This is the bread and butter of modeling disease outbreaks in a population. In many cases, its assumptions are way too restrictive to provide an accurate quantitative predictive framework, but it is a useful framework to understand in order to educate intuition. In the case of SARS-COV2, I am told that SIR models actually do fit the data surprisingly well, at least within cities, for reasons that will become clear. So the structure in this little tutorial is simple: first we will derive SIR models from first principles, and then second we will see where comes from.

# Multiplicative Growth: From Discrete to Continuous

To begin, recall that the basic process underying biological growth is multiplicative. That is to say, cells divide, organisms reproduce, and so do viruses! This multiplication process tells us that, absent structural constrains from outside, populations of cells, viruses and organisms do not accumulate linearly over time but exponentially (i.e. non-linearly). In addition to multiplying, they spread. Multiplicative growth is captured quite simply by the following difference equation:

$$\tag{1} X[t+1] = X[t] + \tau X[t]$$

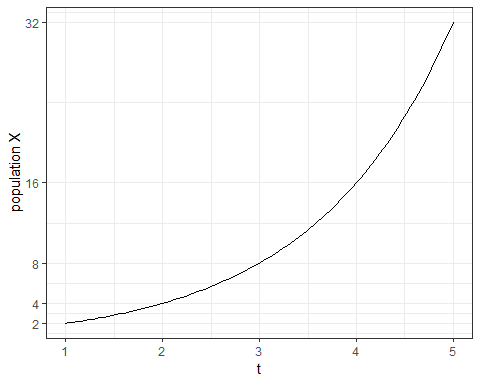
We can rewrite this as:

$$\tag{2} X[t+1] = (1 + \tau) X[t]$$

and after substitution , we get:

$$\tag{3} X[t+1] = \lambda X[t]$$

Here, represents the multiplicative/geometric growth rate of the population over some discrete time interval indexed by . If we take , our time-step is a doubling interval, and the population growth looks like



Given discrete time intervals, and an initial population , it is easily shown that our equation is

$$\tag{4} X[T] = \lambda^T X[0]$$

Now, discrete equations are nice, and in many ways most congenial for biological processes. However, because we want to use the tools of calculus and differential equations, we need to map this into continuous form, which we do by defining an instantaneous growth rate , and setting , hence , and our continuous equation is:

$$\tag{5} X[t] = e^{kt} X[0]$$

. This turns out to be the integral solution of the differential equation:

$$\tag{6} \frac{dX}{dt} = kX$$

and with equation 6 we are finally ready to jump into the SIR models, from which the concept of “basic reproductive rate” is most easily derived.

# Deriving R\_0

Our first goal is to model how the number of infected people in a population grows over time. To do this, we start with equation 6 and then we make additional assumptions. The basic logic is this: you start with an infected individual. That person then comes into contact with a certain number of non-infected individuals at a certain rate. Call this, “rate of contact” (). In each contact, there is a certain probability that the infection will be transmitted, “transmission probability”(). When the number of infected individuals is small relative to the total population (), the growth in rate of infections will be, to a high degree of approximation:

$$\tag{7} \frac{dI}{dt} = \kappa \rho I$$

As always with differential equations, it is helpful to be super clear about the **dimensional analysis** (a hugely under-taught tool, see Sanjoy Mahavan’s excellent book “Street Fighting Math”). On the left hand side, we have something with dimensions of persons/time. On the right hand side we have I (with dimensions ‘persons’), multiplied by dimensionless probability , multiplied by a rate . So must ultimately have dimension of 1/time. Indeed, thinking it through it is #persons-contacted/(person-infected X time). The dimensions of persons cancel, and we have 1/time, just as needed.

What equation 7 assumes so far is that *every person contacted by an infected individual is susceptible to infection*. However, as the infected proportion of the population grows, this assumption breaks down as the probability that an individual that is contacted is already infected, or has recovered and is now resistant increases. We now introduce a pool of susceptible individuals, “S”, and we represent this probability as (a ratio that is also dimensionless). Since this is dimensionless, we can simply stick it onto equation 7 no problem, and we get:

$$\tag{8} \frac{dI}{dt} = \kappa \rho I \frac{S}{N}$$

Over time, declines to 0, and the growth rate of infections goes to zero. At this point, we need to point out that there is a stringent assumption here. That is, this probability is homogeneous across the population and can be represented by the proportion . This is the so-called “law of mass action” imported from equilibrial dynamics in chemistry and physics. In this context, it requires the notion that populations are *well mixed*. The real world can get in the way of this being a good assumption when there is strong spatial structure, strong heterogeneity in the population, and so forth.

Another thing happens though. Infected individuals either die or recover and become resistant. The simplest assumption here is that death and recovery are also both first-order processes, so we subtract them from the growth rate and complete our differential equation. For now, I am going to pretend that no one dies (yay!) and people just become resistant, as the math fundamentally does not change either way:

$$\tag{9} \frac {dI}{dt} = \kappa \rho I \frac{S}{N} - \nu I$$

Now, briefly, we must maintain the RHS with dimensions of persons/time, so is simply another rate with dimension of 1/time, and represents rate of recovery into resistance.

And now we can see exactly what is and where it comes from. An individual does not stay infectious forever. They recover and become resistant “R” instead, at the rate . But before they do this, they are generating new infections at the rate . So if we take the following ratio:

$$\tag{10} \frac {\kappa \rho}{\nu}$$

, we have a nice dimensionless number that represents the average number of new infections per infected individual. This is the “basic reproductive number”!

Return to the dimensionality for a second. It is customary to write as . Remember that this has a dimension of 1/time. Therefore, the “typical time” to a new infection is the reciprocal or . Likewise, the “typical time” to recovery is . “” is clearly then the typical time to recovery over the typical time to infection. A larger could arise from either recovery taking longer, time to infection being shorter, or both.

Large is bad. Small is good.

In fact, we can now easily recover the single most important epidemiological consequence of . Is the number of infected individuals growing or declining? Our goal is to make a set of interventions that turn equation 9 negative, thus guaranteeing the decline of infected individuals, and eventual extinction of the outbreak.

We re-write equation 9 using from equation 10 and we get:

$$\tag {11} \frac {dI}{dt} = R\_0 \nu I \frac{S}{N} - \nu I = \nu I (R\_0 \frac{S}{N} - 1) $$

At the beginning of an outbreak, where , this differential equation is clearly positive when > 1, and negative when < 1.

Here is the punchline: at the beginning of an outbreak, if you severely limit average number of contacts between people, transmission probabilities, or both, you can drive below one and prevent the outbreak from ever materializing. This is the goal of social distancing measures. We can also be more surgical and effective at this if we target infected individuals and their contacts for quarantine. This requires widespread testing and is why the Trump administration’s response to Covid-19 is criminal malpractice.

As the pandemic proceeds, the number of susceptibles presumably declines, which leads to self-limitation of the outbreak through the ratio . This is the so-called “herd immunity” effect that the British seem to be banking on. In better scenarios, we deploy mass scale vaccination campaigns to drive enough people out of the “S” pool so that the equation goes negative that way. “Herd immunity” via infection is a terribly risky proposition since it **requires** a lot of people to get sick, and in the case of SARS-COV2 this is going to mean lots of people dead both directly from the virus and indirectly from a flooded healthcare system that can no longer take care of people with all kinds of ailments. This, too, is criminal malpractice, given all the uncertainties about Covid-19.

# Recovering the full SIR model

The rest of this tutorial will be brief. We complete our model with understanding the disease by considering the three pools separately, in the order S-I-R.

$$\tag {12} \frac {dS}{dt} = - R\_0 \nu I \frac{S}{N} $$

$$\tag {13} \frac {dI}{dt} = R\_0 \nu I \frac{S}{N} - \nu I $$

$$\tag {14} \frac {dR}{dt} = \nu I $$

This represents a system of three coupled ODEs. We have removal of Susceptibles by infection, already derived in detail. The initial condition is . The dynamics of the Susceptible pool have already been derived in detail. Then we simply add an equation for the change in Recovered/Resistant pool which receives everyone dropping out of the Infected pool.

And there you have it! This is the most basic foundation of infectious disease epidemiology. Along the way, I hope the assumptions underlying the epidemiology are more clear, as is the meaning of the ubiquitous term. In parting, I want to emphasize that is NOT an inherent property of a virus. It emerges dialectically from viral biology interacting with human social behavior, which in turn is shaped historically by culture, politics and economics.

We have a lot of leverage over epidemics. **Let us use our collective agency to write a better history for this pandemic!**