Epidemiology: The math of disease outbreaks

BMC Advanced
Chris Overton
Class notes after 2020-03-04

These notes are meant to list topics we covered in person on March 4, and to help students who missed that prepare for our video session on March 18.
For questions at the bottom of a page (starting with "Q"): please think about these before turning to the next page.

The context:
Last time, were were seeing early exponential growth in Covid-19 cases outside of China (at the time the only country where disease outbreak seemed clearly under control.)
We said then that the situation would change significantly until our next meeting - then scheduled for last week, but canceled due to Covid-19 and replaced by an online meeting this week.
Here we try to emphasize analytical content that will help you navigate the unfolding global crisis - even as media coverage varies dramatically.

Last time, we established that most students were comfortable with basic calculus, which provides an entry to differential equations. You mainly just need an intuitive understanding of what a derivative is, and how it tells you how functions grow.

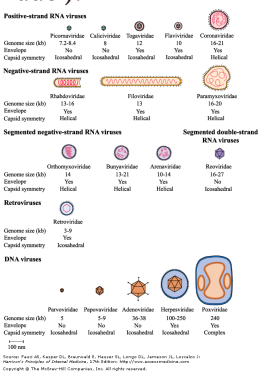
Useful links - please take a look at the first before class.
1) 2008, David Earn: A light Introduction to Modeling Recirrent Epidemics:
https://ms.mcmaster.ca/earn.old/pdfs/Earn2008_LightIntro.pdf
Author's page:https://davidearn.mcmaster.ca/publications/Earn2008
Easier high school version:
https://ms.mcmaster.ca/earn.old/pdfs/Earn2004_PiInTheSky.pdf
3) Counts of global Covid-19 cases by country and status:
http://worldometers.inco/coronavirus
4) A detailed "call to arms" by Tomas Pueyo (as usual, some models debatable):
https://medium.com/@tomaspueyo/coronavirus-act-today-or-people-will-die-f4d3d9cd99ca
5) Useful data, with daily updates:
https://github.com/CSSEGISandData/COVID-19
interactive map: https://coronavirus.jhu.edu/map.html

Topics discussed:

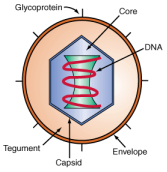
- Current situation of covid 19 epidemic
- How math fits in: Intro to ODE's (ordinary differential equations)
- Derivation and critique of the "SIR" model (susceptible/infected/recovered) - the basic model for epidemics
- How can we modify this model to reflect Covid-19?

Background

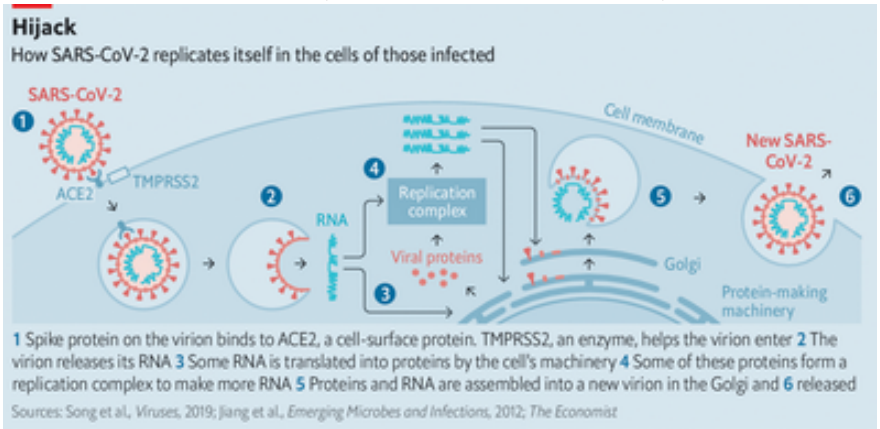
Here, we can't cover all relevant backgound -- much less stay current with the explosion of available information.
But for perspective, here is how coronaviruses (upper right) compare to other viruses that infect humans (from Harrison's Internal Medicine, co-author Fauci):



Note the extra lipid (fatty) outer layer. It includes ways to enter our cells but is washble with soap:



Basic infection mechanism (Economist, issue 2020-03-14):



Q1: What is the probability you will be infected?
Q2: What is the probability you will require hospitalization, or worse?
Q3: How many people in the US will be infected in 2020?
Q4: How many people in the US will die as a result?

Answers about disease incidence:

We assume you are a student under 20. Your probability of infection is highly uncertain, and depends very much actions of yourself, your household, your community, and your government. But whatever the answer to Q1, that for Q2 should be well over two orders of magnitude smaller, because younger people are much less likely than average (over the entire population) to experience severe problems.
Again, the answer to Q3 is highly uncertain. Numbers reported this week are already around an order of magnitude higher than two weeks ago. These severely underestimate totals, because many infections have not yet been verified, and many have not yet caused symptoms. But the ratio of Q4 to Q3 is a subject of active investigation. It has been estimated as around 1%, but would be higher when health systems are overwhelmed by too many cases to provide usual treatment. This is around an order of magnitude higher than for annual flu outbreaks, and so is of great concern.

To get a handle on incidence, we'll need models for how infectious diseases spread.

To get there, three steps:

1) Review of calculus
Given a function y(x) plot and understand how both its derivative and antiderivative behave (not replicated in these notes.)

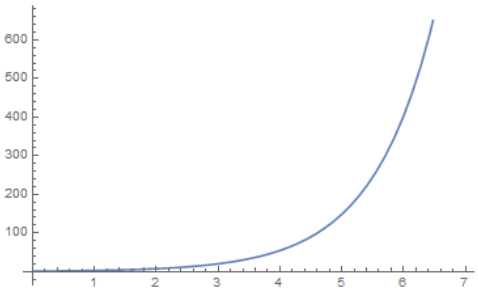
2) Exponential growth and decay:
Q5: if z'(x) = c * z(x) for a constant c, how does z(x) behave for positive c and for negative c?
Can you think of real-life examples of each?

3) Positive and negative feedback loops - qualitative behavior and sample equations (see two pages ahead.)

Gaining a healthy respect for exponential growth

For positive c, z'(x) = c * z(x) is satisfied by $z(x) = d * e^{c*x}$. This includes the new constant d, that might be determined by knowing the value of z(x) for one particular x. For arbitrarily small c > 0 and any positive power of x (e.g. $x^{googol} = x^{10^{100}} = x^{10000000.....0000}$), eventually z(x) will become bigger than this function.
Examples of this include the balance in a bank account earning a fixed interest rate, or the number of bacteria if growing on adequate culture media.

```
Plot[expGrowth[1, x], {x, 0, 7}]
```

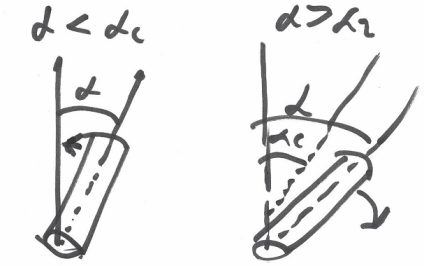


It turns out the same solution works when c < 0, but in this case the function z(x) converges quickly to 0. This is called exponential decay, and occurs in real life with radioactive decay.

3) To contrast qualitatively the convergence to ∞ and 0:
Q6: What is the difference between positive feedback and negative feedback?

Positive feedback and negative feedback

In class, we illustrated the difference with a tall cylindrical container returning to vertical, as long as the angle α of its initial displacement from vertical was less than a "critical" angle α_c.



Such stability is called negative feedback: below α_c, α is accelerated toward 0. Our health, and even our physical balance, is maintained by many such systems.
On the other hand, once $\alpha > \alpha_c$, the greater this difference, the faster α is accelerated - positive feedback resulting in collapse. This is a prototype of how large changes occur, such as exponential growth or when a forest catches fire.
Together, both situations are sort-of captured by the equation
$|\alpha|'' = d * (|\alpha| - \alpha_c)$,
with constant d. A bit of a problem is that absolute value is not differentiable, but we simplify this way to illustrate how (stable) solutions can wiggle back and forth across zero.
One consolation of real life is that its positive feedback loops usually sit inside larger negative feedback loops. For example, an accelerating forest fire eventually runs up against limits on what is flammable.

The SIR model for epidemic spread

With these ideas in mind, we explore how to capture qualitative behavior of epidemics.
Note first that there is a positive feedback loop in the number of people who are infected: the more who are infected, the more likely other (uninfected) are to run into someone who is infected.
On the other hand, there is sort of a negative feedback loop in that the more people are (or have been) infected, the fewer people are still left to infect.
Caution: this assumes that one cannot get infected twice (not always true.)

To capture dynamics, we use a compartmental model, namely one that splits the population into several groups, or "compartments". The simplest has only:

- S for susceptible - those who could be infected
- I for infected / infectious
- R for recovered

Our assumption is that the population remains constant at N people. Then R = N - S - I, so for a complete model, we only need to model S and I. Then people can move between these compartments only as S -> I -> R.

Q7: How can you do this by modeling derivatives S' and I' ?

The SIR equations

The basic approach uses two constants: tr is a rate of transmission, and $recov$ is a rate of recovery:

$S' = -\frac{tr*I*S}{N}$

$I' = +\frac{tr*I*S}{N} - recov * I$

Transmissions: the first terms in these equations track how people transfer compartments from S to I.

Recovery: the second term of the second equation captures transition from compartment I to R. Apart from new infections, this shows a pattern of exponential decay. The larger $recov$, the faster the decay.

To make sense of the transmission terms, imagine a regular susceptible person meeting random others. On average, a proportion $\frac{I}{N}$ of them are infected, and this times tr captures how many new infections occur.

Early stage behavior: initially, I << N, so S is roughly proportional to N. In this case, the rate of transmission becomes $S' \approx -tr * I * 1$. This looks like exponential growth in I, disregarding recoveries. Currently, I << S in each country!

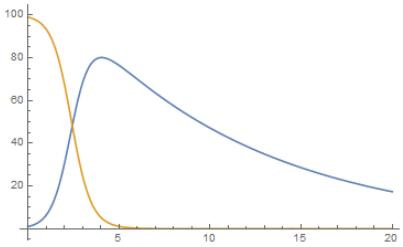
Eventually, when fewer people (S) are left to infect, or whenever transmission is smaller than recovery, exponential decay takes over.

We illustrated several examples of this in class. Here is just one example, when tr=2, recov=.1

The basic setup, here as percent of population

```
n = 100.;
i0 = 1;
tr = 2;
recov = .1;
duration = 20;

Plot[Evaluate[{i[t], s[t]} /. sol], {t, 0, duration}]
```



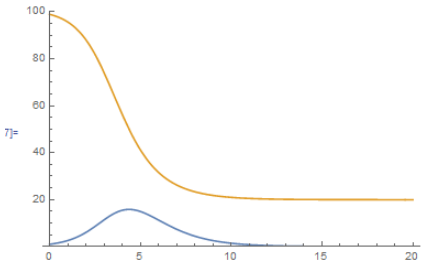
Q8: In this example what fraction of the population eventually has gotten infected?

100% - in other words, this is a lousy way to control disease!
Even so, the rate of infection eventually does taper off, when there are fewer S people left to infect.
Notice that some evade infection (i.e. remain susceptible) if the recovery rate is faster:

Similar, but now with faster recovery

```
n = 100.;
i0 = 1;
tr = 2;
recov = 1;
duration = 20;
```

```
sol; Plot[Evaluate[{i[t], s[t]} /. sol], {t, 0, duration},
 PlotRange → {0, 100}]
```



Controlling infection

The only way to control disease is to arrange things in such a way that the transmission becomes less than the decay even when there are still susceptible people left.
Examples include:

- Increasing the recovery rate through medical treatment
- Reducing the transmission rate by immunizing a proportion of the population or keeping people from infecting each other.

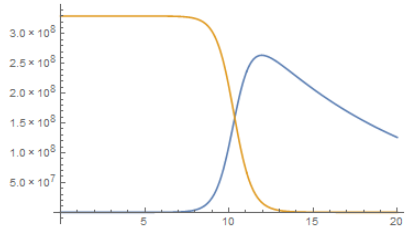
Once decay is faster than transmission, the population is said to have "herd immunity."

Q9: Another example: a single person leads to infection of the entire US. In this case, why is there a "time bomb", in which nothing appearing to happen for the first several days? Does this mean there is a low transmission rate?

Sleeper?

```
n = nUs; (*330 million*)
i0 = 1;
tr = 2;
recov = .1;
duration = 20;
```

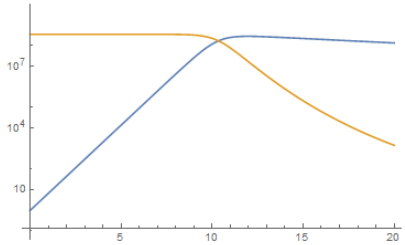
```
sol; Plot[Evaluate[{i[t], s[t]} /. sol], {t, 0, duration}]
```



No - absolutely not! As you can see from the same graph with a logarithmic y-axis, the near-exponential growth is going on all the time until it suddenly becomes visible and then huge on the linear scale we saw on the last page.

... no - it just looks that way because of the early exponential rise

```
LogPlot[Evaluate[{i[t], s[t]} /. sol], {t, 0, duration}]
```



It is very important to understand how easy it is to underestimate ongoing exponential growth when relatively small data values take time to appear on national or global scales.
This explains how almost every country has failed to initiate stronger actions earlier, when the total cost to society would have been much smaller.

See below for how reported cases lagged presumed case starts in China.

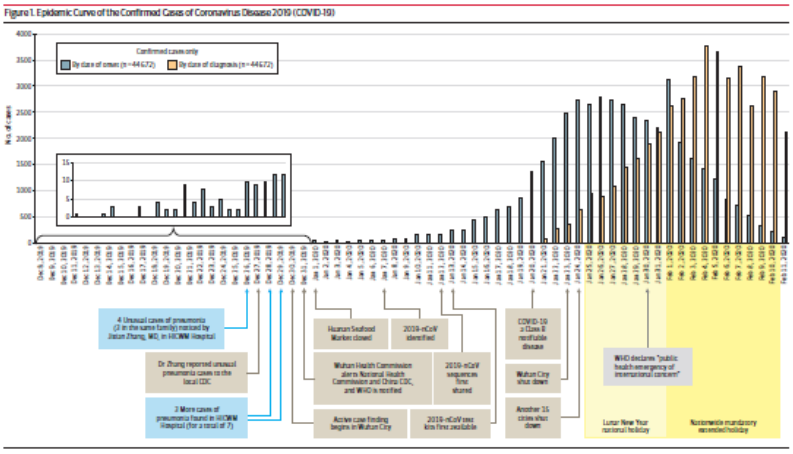
So to recap, the basic SIR model does capture several realistic qualitative features of disease outbreaks.

Q10: what is not realistic about the SIR model as discussed so far?

Possible improvements to the SIR model (I)

We will continue to discuss several possibilities. Common ones include:

- Stochastic behavior: anything differentiable is continuous. But humans occur only in multiples of 1. We could think of S, I , and R as proportions, but even so, when any represent few individuals, behavior of individuals could alter a whole trajectory. For Example, if you start with I(0) = 1, that single person could kill the entire epidemic by removing themselves until the infection passes.
- Quarantines are a way to isolate the infected from transmitting further. Including a separate Q compartment (a subset of infected, so we could call it I_q) This requires a models for moving from I to Q. But assuming no transmission of disease from Q to S, this is a way to alter the trajectory of the epidemic.
  (Problem: there was a real-life "quarantine" of several thousand people on the Grand Princess ship recently docked in Oakland. Unfortunately, this added another new compartments, namely susceptible people S_q stuck in this "quarantine" - and suffering a much higher transmission rate than in the general population. This has led so far to 7 deaths.)
- Time lags: time infected may not be the same as time infectious (i.e. able to infect others.) Also, there may be a systematic lag between time infected (gray) and the time when this was recorded clinically (yellow), as recorded in China:



Possible improvements to the SIR model (II)

- Multiple incidence events: our first SIR model only initiates infections by specifying I(0). In real life, there can be multiple additions of new infections cases coming from other geographical units than the one we are attempting to model.
- Human behavior changes the transmission rate tr: once society becomes sufficiently alarmed, governments mandate different behavior. So far, this has been done successfully in 2003 for the Sars Coronavirus in China, and successfully this year for the Sars 2 Coronavirus (agent causing Covid-19) in Mainland China, and apparently in Singapore, Taiwan, and Hong Kong. Although success is not yet established in most other places, a lowered transmission rate should be visible in retrospect due to all of the travel restrictions and social distancing that now are being mandated internationally.
- New treatments could alter both transmission (via vaccines) and recovery (a faster rate due to curative medicines.)
- Rates of transit into and out of I are not the same across individuals - e.g. they differ based on behaviors and on health. This is a significant problem with the model, but the model is partially robust in that the tr and recov coefficients apply as an average across the whole population. In California, the poor health care and living condition of the homeless and impoverished could create a high rate of transmission within these groups, which could then spread to the rest of the population.
- Unfortunately, in real life, population characteristics differ significantly between susceptible and recovered: an additional transition is from I -> D (death.) For less severe diseases, this has a low proportional effect on the outbreak, and so is ignored in models. For Covid-19, there were already over 7000 deaths reported by March 17. This is a critically important outcome, but one that reduces I(t) just as recovery does.
- Once health facilities are overwhelmed, rate of recovery could decrease and rate of death could increase - with substantive effects on the epidemic.
- Finally, the numbers we are seeing in the data are not all actual cases! As of a couple days ago (March 12), total tests ever conducted in the US lagged behind then current daily tests in South Korea. So in the US (among other places), the number of reported cases is a radical underestimate. When this catches up with reality, it will look temporarily like higher growth than is accurate. More generally, many milder cases never get reported.

Next time, we will consider some of these improvements in further detail, and will take another look at a world that has changed dramatically in two weeks!