

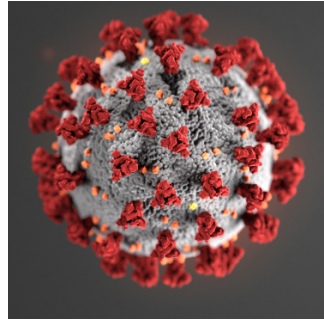
Citizens and cities facing new hazards and threats

30th November to 4th December 2020

Modeling and Battling COVID-19

Dr. George Markokwsky





Modeling and Battling COVID-19

Prof. George Markowsky

Cyber Society Lab

Department of Computer Science

Missouri University of Science & Technology



Goal of this Talk

- The goal of this talk is to introduce you to a basic model used to model epidemics
- We will show how this model applies to the COVID-19 pandemic and how it leads to a strategy for dealing with the pandemic
- There are many important aspects of the COVID-19 pandemic that are not covered in this talk — many more details at DocGM.com
- Our goal to achieve a qualitative understanding of the behavior of epidemics in general and apply this understanding to battling COVID-19
- We hope that this presentation will help you understand why accurate modeling is so challenging, but that even rough models can still be very useful
- Understanding this model will help you to understand more complex and detailed models and highlight the important role that computer modeling plays in dealing with disasters



Compartmental Models

- In a compartmental model, we partition the members of our group into compartments and establish rules on how individuals move from one compartment to another
- We assume that the behavior of all individuals in a given compartment is the same
- Usually, the more compartments we establish, the more accurate the model
- We want to create a short-term model, so we will ignore events such as births and immigration



The SIR Model

- The SIR model is one of the simplest of the compartmental models and gets its name from the three compartments (S, I, R) that it divides the entire population into
- The model was first proposed by William Kermack and Anderson McKendrick in a series of 3 papers written between 1927 and 1933
- The SIR model forms the basis for many of the models in use today
- It assumes that population does not change during the modeling period and that there are three compartments of people – Susceptible, Infectious, and Recovered



The S, I, and R Compartments

- You might think that you understand the nature of the S, I, and R compartments from their names and you would be mostly correct, but there are subtleties associated with these definitions
- One thing that might strike you as strange is that there do not seem to be any dead people in this model
- Actually, the dead people are in the Recovered compartment
- The term recovered is a bit misleading, since what we are really talking about is people who can no longer infect other people
- It might have been better to call this compartment U for uninfectious



Infectious People

- Note also, that we refer to the compartment I as consisting of **infectious** people and not infected people
- For the sake of the model, we care primarily whether a person can infect other people and not whether they are infected or not
- For example, suppose a disease typically takes a person 15 days to recover from, but the person only produces particles that can infect other people on the first day of the disease, then we would put the person into the Recovered group after 1 day and not after 15 days
- Similarly, if a person is infectious for 15 days, but after day 5 the person goes into quarantine and can no longer infect other people they would go into the Recovered compartment as soon as they are quarantined even though they might still be sick
- Also, a person might get sick and end up with kidney failure but stop being infectious
- In this case, we classify that person as recovered but not everyone would apply the term “recovered” to this person



Recovered People

- The language is a bit tricky here
- Usually, when we say that someone recovered from the flu we mean that they are no longer sick and are back to “normal”
- If someone died from the flu, we would not normally say that “they recovered from the flu”
- The important thing to understand here is that recovered means that the people in question can no longer infect other people
- Another tricky instance is if a person had genetic immunity to the disease that person should be considered recovered even though that person was never infected
- Also, if an infectious person is quarantined so that they can no longer infect anyone, we would move them to the Recovered compartment even though they are still sick and could potentially infect people
- Because dead people are placed in the Recovered compartment, we can say that the population does not change during the modeling



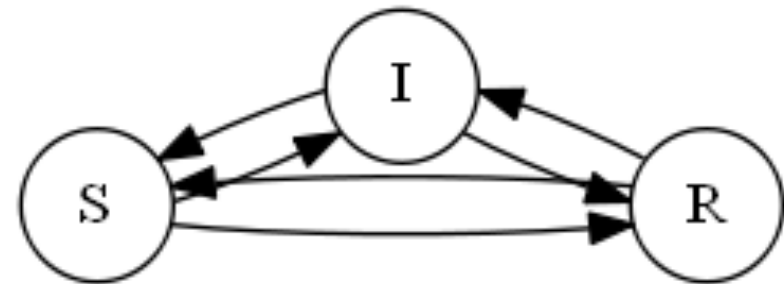
Susceptible People

- A susceptible person is a person who is not infected, but who could become infectious by exposure to infectious people



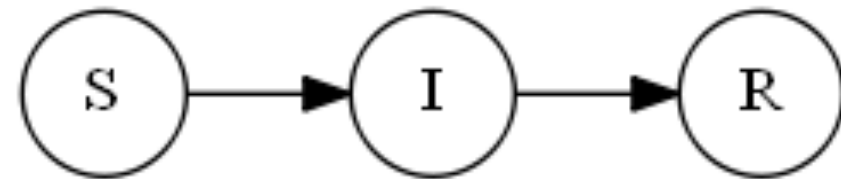
Relationships Between Compartments

- In modeling an epidemic, we want to understand how people move from one compartment to another
- In general, we can envisage a diagram such as the one to the right in which there are circumstances under which people can move directly from one compartment to any other compartment
- We want to simplify this picture to avoid analyzing 6 transitions



Simplifying the Diagram

- At this time, it is still not completely known exactly what type of immunity “recovered” COVID-19 victims maintain
- This is not an idle question since we know there is enough mutation among other Corona viruses (e.g., the common cold) so that people can get reinfected, even in the same cold season
- Needless to say, the more arrows we have in the model linking the compartments the more complex the model
- To simplify our discuss, we will assume that we have the diagram to the right which we will call the $S \rightarrow I \rightarrow R$ diagram
- We are now ready to start writing some equations!



The $S \rightarrow I \rightarrow R$ Diagram



Some Equations

- If the population we are studying has N people, the first equation we can write is $S + I + R = N$
- Notice we do not subtract people who were killed from the population for the purposes of the model – people who died are counted by R
- We are now going to reach into our bag of tricks and announce that we can differentiate the above equation
- You might object on the grounds that S , I , and R are not real numbers, and it makes no sense to talk about 4.5 infected people, etc.
- However, there are two important counter arguments
 1. We want to primarily apply our model to large populations so the difference between 1,000,000 and 1,000,000.6 people is not significant
 2. These models work and we use them in many different situations



A Differential Equation

- You might ask, why do we even want to differentiate the equation?
- The answer is, that we are interested in the movement between compartments which suggests that we are interested in the rate of change of the numbers in all three compartments
- Note that we are dealing with functions $S(t)$, $I(t)$, and $R(t)$ but we will often suppress the t to reduce clutter
- If we differentiate both sides of $S + I + R = N$, we get
- $\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$
- This means that once we figure out how two of the compartments change, we will know how the third one changes



The Fundamental Equations

1. $\frac{dR}{dt} = aI$ ($a = 1/d$, where d is the number of days it takes to recover)
2. $\frac{dS}{dt} = \frac{-bcIS}{N}$ (b is the probability of being infected from a contact and c is the number of contacts/day)
3. $\frac{dI}{dt} = \frac{bcIS}{N} - aI$ (the sum of the derivatives = 0)

- Notice the following: I appears in every term on the right-hand side of every equation
- Notice also that some terms involve N which suggests that we would have different equations for different populations – we would prefer to have a universal set of equations that works for any population
- Fortunately, there is a simple trick that we can use to make our equations “universal” so they will work for any population



The Simple Trick

- Derivatives are linear so if we replace S , I , and R by S/N , I/N , and R/N , we get what we want
- In particular, if we introduce the functions $s(t) = S(t)/N$, $i(t) = I(t)/N$, and $r(t) = R(t)/N$, we get

$$1. \frac{ds(t)}{dt} = -bci(t)s(t)$$

$$2. \frac{di(t)}{dt} = bci(t)s(t) - ai(t)$$

$$3. \frac{dr(t)}{dt} = ai(t)$$

- The beauty of the above system is that it is population independent – it just involves ratios!



“Flatten the Curve”

- You have undoubtedly heard the expression: “we have to flatten the curve”
- There are three curves here, which one do we need to flatten?
- The answer is we need to flatten $i(t)$ as you can see when we write the equations a bit differently

$$1. \frac{ds(t)}{dt} = -bci(t)s(t) = (-bcs(t))i(t)$$

$$2. \frac{di(t)}{dt} = bci(t)s(t) - ai(t) = (bcs(t) - a)i(t)$$

$$3. \frac{dr(t)}{dt} = ai(t)$$

- We noted this earlier when we pointed out that $i(t)$ is a factor in every derivative – if you can keep $i(t)$ small you keep all the derivatives small!



Numerical Approximations

- Taking great liberties, which I will not try to justify here, we end up with the following equations written in computer-friendly style
 1. $s(t+1) = s(t) - b*c*i(t)*s(t)$
 2. $i(t+1) = i(t) + b*c*i(t)*s(t) - a*i(t)$
 3. $r(t+1) = r(t) + a*i(t)$
- The previous 3 equations are easy to implement in Python and in a spreadsheet like Excel
- I have an expanded version of this talk that has both working Python code and a working Excel version that you can get from my website : DocGM.com



Let's Run Some Examples

- When we run examples, we need to pick some values for $s(0)$, $i(0)$, and $r(0)$ along with coefficients a , b , and c
- If we pick $i(0) = 0$, then nothing happens in the model because all derivatives are 0
- It is simple to assume that $r(0) = 0$ since we are assuming that no one has recovered yet since no one got infected – we are ignoring genetic immunity
- Thus we need $i(0) = \epsilon$ and $s(0) = 1 - \epsilon$
- We will investigate the choice of ϵ soon
- We then need to pick a , b , and c



Picking a, b, and c

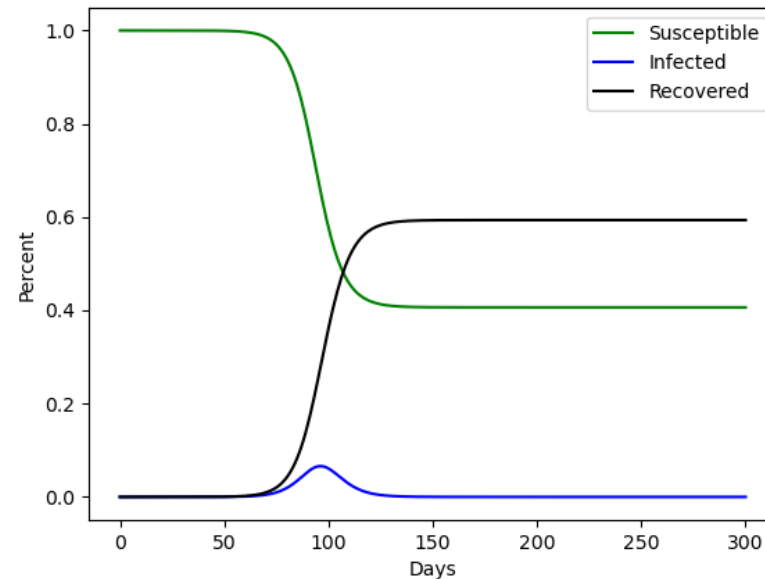
- Let's start off with a disease, something like a cold, during which people are infectious for 3 days
- We will assume that $a = 1/3$
- What about b and c?
- Just to get started, let's suppose that a person has 20 contacts a day and that the chance of getting infected from meeting an infected person is 2.5% = .025, so $bc = 20 \times .025 = .5$
- Let's run the program and see what we get



Our Starting Scenario

- With the parameters set up top, we get the graph shown to the right
- What does $i[0] = .0000001$ mean?
- It means that there is only 1 infection for every 10,000,000 people – this is the equivalent of about 35 infected people in the entire USA
- Why did the graph stabilize? Shouldn't everyone get infected?
- The reason the graph stabilized is that $I(t)$ became 0, so all changes stop
- The peak for $i(t)$ happens about 100 days out while the peak for $r(t)$ is maybe 115 days out with that being around the maximum of deaths if $dead(t)$ is counted in $r(t)$

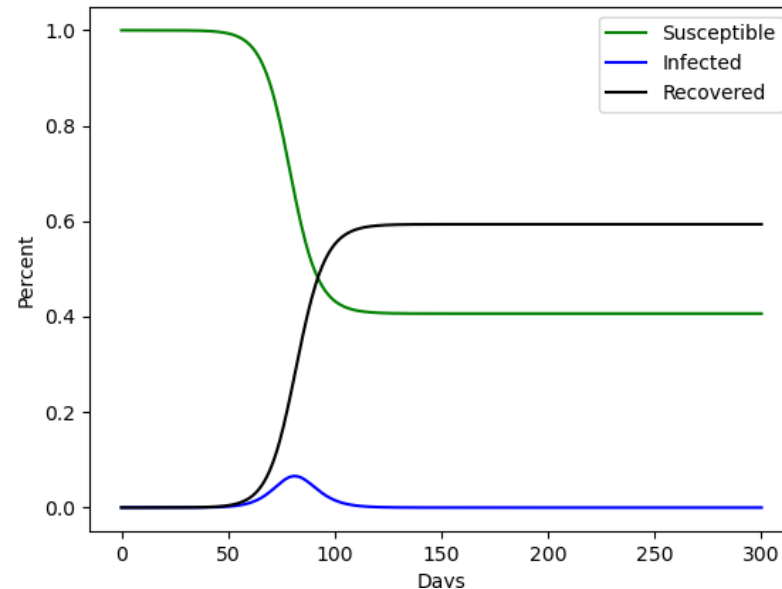
Epidemic with $i(0) = 0.0000001$, $b = 0.025$, $c = 20.000$, and $a = 0.333$



Starting With More Infected People

- Let's change $i[0]$ to 1 in a million instead of 1 in 10 million
- We then get the graph to the right
- We got pretty much the same curve except that the peaks happened earlier which make sense
- About 60% of the population ends up in the recovered category and about 40% never get infectious

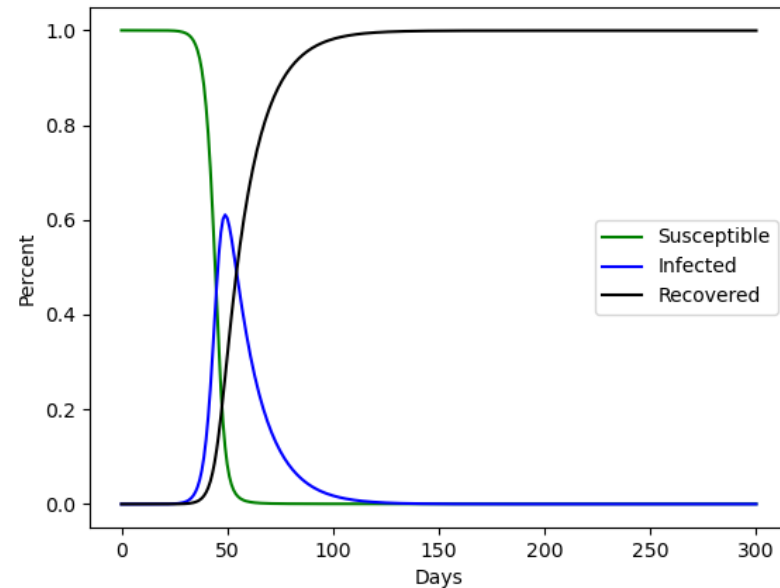
Epidemic with $i(0) = 0.0000010$, $b = 0.025$, $c = 20.000$, and $a = 0.333$



Extending Days of Infectiousness

- Let's change $i[0]$ back to 1 in 10 million
- Let's assume that people are infectious for 14 days instead of 3 so that $a = 1/14$
- With these parameters, we get the graph to the right
- The graph to the right is much scarier than the graph on the previous slides!
- Remember that this graph is population independent
- Clearly, we did not flatten the $I(t)$ curve enough

Epidemic with $i(0) = 0.0000001$, $b = 0.025$, $c = 20.000$, and $a = 0.071$



The abc-Attack on COVID-19

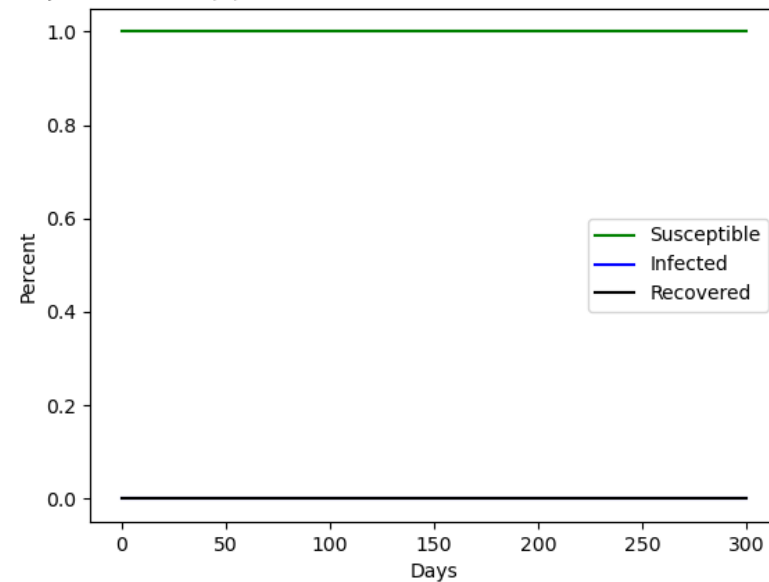
- The model we have come up with depends on 3 parameters: a, b, and c
- To control the pandemic we need to **increase a ($a = 1/d$, so decrease d), decrease b, and decrease c**
- The simplest and **fastest** way to decrease d (increase a) is to **find infectious individuals and quarantine** them until they stop being infectious
- **The role of testing is crucial here!**
- If we test only sick people, we will never find the asymptomatic carriers!
- Ideally, we should test everyone
- **We must quarantine infectious people!**
- We can decrease b by creating greater distances between people, having people wear masks, and washing your hands and not touching your face (**distancing, mask wearing, washing your hands and not touching your face**)
- We can decrease c by people seeing fewer people (**social isolation**)
- An **abc-attack** can also be called a **lockdown**



An Effective abc-Attack

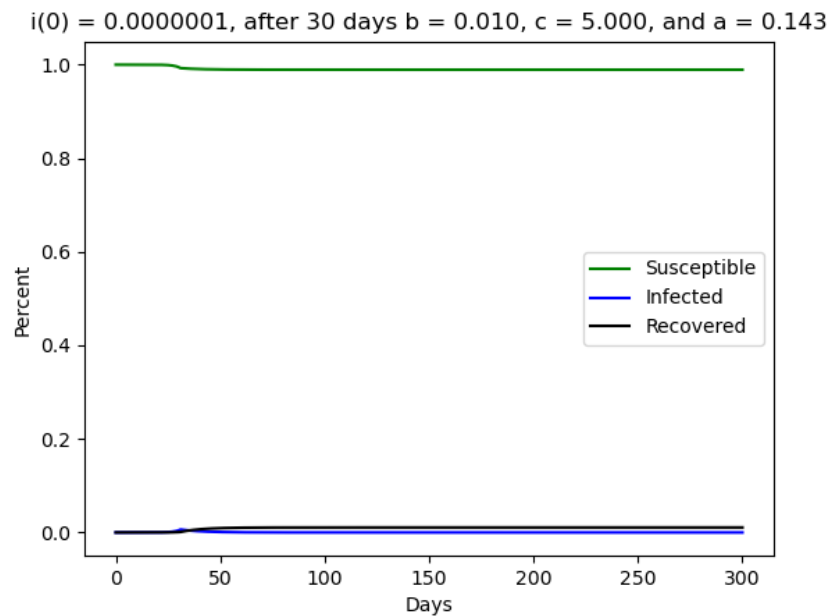
- Let's keep $i[0]$ at 1 in 10 million
- Let's assume that people are infectious for 7 days instead of 14 so that $a = 1/7$
- Let's assume we knock b down from .025 to .010 and c from 20 to 5
- This abc-attack generates the graph to the right
- The graph to the right is much more pleasant to look at than the earlier graphs
- Remember that this graph is population independent
- We have really “flattened the curve”!

Epidemic with $i(0) = 0.0000001$, $b = 0.010$, $c = 5.000$, and $a = 0.143$



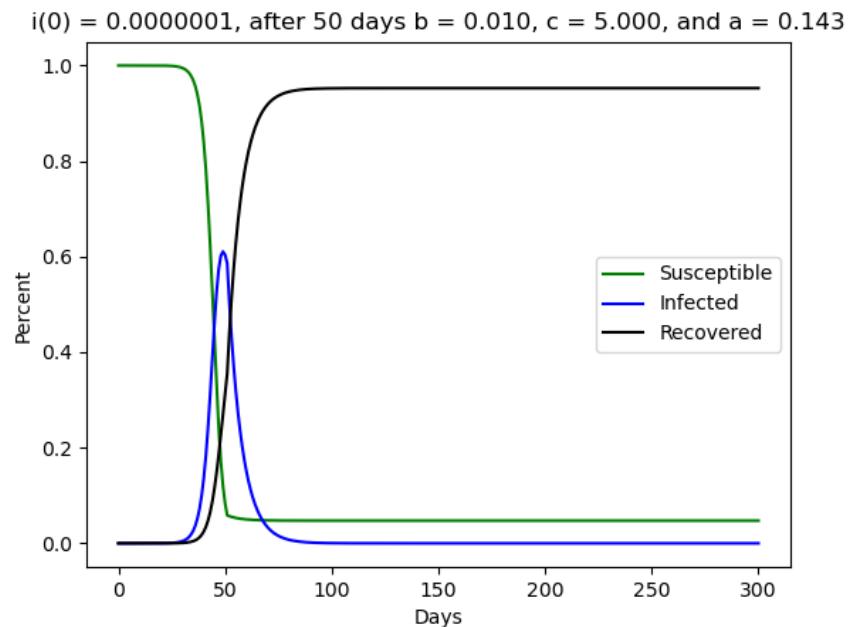
“Normal” and then an abc-Attack

- Suppose that we start out with the “normal” parameters with $a = 1/14$, $b = .025$, and $c = 20$ for 30 days
- After 30 days, we switch to $a = 1/7$, $b = .010$, and $c = 5$
- This scenario generates the graph to the right!
- What a difference!
- What if we wait for 50 days before we switch to the new a , b , and c ?



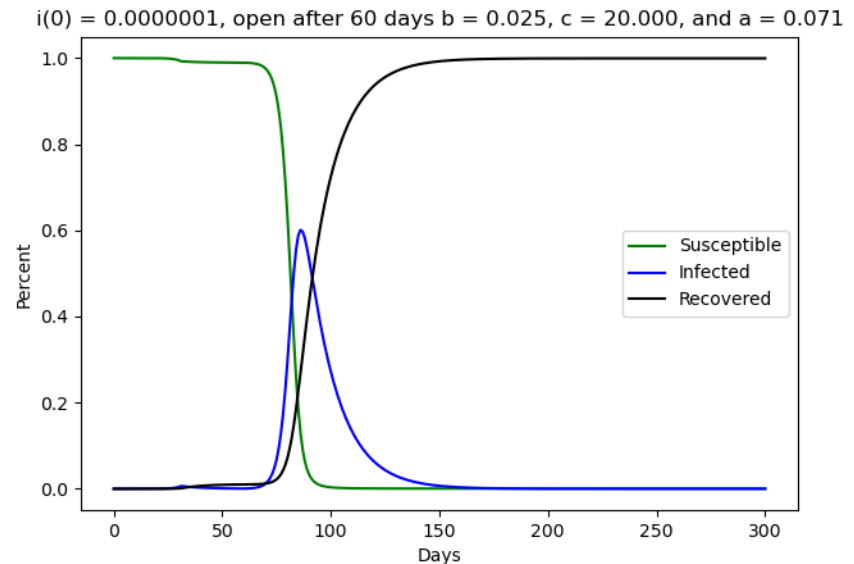
A “Small” Delay

- The graph to the right shows what happens if we just delay 20 days before applying a lockdown
- The graph on the right demonstrates vividly the importance of implementing your abc-attack as quickly as you can
- We’ve gone from a not-so-bad scenario to a very bad scenario with nearly everyone getting infected just by delaying for 20 days
- **LACK OF SPEED, KILLS**



Reopening

- Let's consider the following scenario: we start out with our "normal" scenario with $a = 1/14$, $b = .025$, and $c = 20$ for 30 days
- After 30 days we lockdown to $a = 1/7$, $b = .010$, and $c = 5$
- After 60 days we decide to go back to "normal" with $a = 1/14$, $b = .025$, and $c = 20$
- The graph you get is shown to the right
- It is clear that reopening must be done carefully!



General Conclusions

- This simple model shows very graphically that in the absence of a “cure” or a vaccine, controlling the constants a , b , and c is extremely important
- This model is not detailed enough for precise planning, but it offers strong evidence that controlling a , b , and c can be an effective way to limit the damage done by COVID-19
- The model very clearly illustrates the value of **testing and quarantining, distancing, social isolation, and washing your hands and not touching your face**
- It also illustrates that keeping $i(t)$ as low as possible – “**flattening the curve**”, is extremely important!
- Also, we have shown that **LACK OF SPEED, KILLS!**
- **It makes clear that pandemics like COVID-19 are exponential processes that can explode if we are not very careful**



Looking at Real Data

- Once you have a model, regardless of how crude it might be, there is an overwhelming temptation to apply it to real data
- There is some preliminary work that must be done since the data does not directly measure the constants and functions that appear in our SIR model
- In particular, the number most commonly cited is the number of people who got infected over the course of the epidemic, let's call this number X
- This number never goes down so it clearly cannot be $i(t)$ which goes up and eventually goes down
- Another problem is that when people recover, they are never removed from the count
- There are other issues as well



Looking at Real Data

- As mentioned at the beginning of this talk, there are many problems with creating models with the lack of information that we face
- COVID-19, however, will just do its thing and not care whether we understand it or not
- In short, we need to make our best assumptions and move on
- Note that $X(t)$ is given in terms of individuals, so we need to use the function $I(t)$ rather than $i(t)$ since $I(t)$ counts individuals and $i(t)$ is a ratio
- At this point the simplest thing we can do is to assume that $X(t+1) = X(t) + I(t)$, so if we want to estimate $I(t)$ we should use the equation $I(t) = X(t+1) - X(t)$
- Because of asymptomatic carriers we really don't know $I(t)$, but this is the best that we can do under the circumstances



Looking at Real Data

- Let's return to our differential equation for $I(t)$
- $\frac{dI}{dt} = \frac{bcIS}{N} - aI$
- Let's also use $I(t+1) = I(t) + \frac{dI(t)}{dt} = I(t) + (bcS(t)/N - a)I(t)$ so
- $(I(t+1)/I(t)) = 1 + bcS(t)/N - a$ which is not quite a constant because of the term $S(t)/N$
- For short periods of time, $S(t)/N$ is a constant, let's call it λ
- The population of the USA is about 330 million and so far, there have been 12 million confirmed infections
- Even if we estimate that the number of infections is really 33 million, this means that λ varies between .9 and 1
- So after all this work, we have come up with $\frac{I(t+1)}{I(t)} \approx (1 + \lambda bc - a)$
- Now we are ready to look at some data and see how it matches up with this analysis



Data for the US

- Here's our plan of attack
- We first collect our best data for X
- Then we determine that $I(t) = X(t+1) - X(t)$
- We used a 7-day average of $X(t)$ to smooth out the data
- Then we look at the ratios of $I(t+1)/I(t)$ and see whether we get a constant
- For this test run we used data for the entire USA from the New York Times data repository (<https://github.com/nytimes/covid-19-data>) running from October 1, 2020 until November 20, 2020
- The sequence of values that we calculated is shown to the right and show our estimates of $1 + \lambda bc - a$, which should be a constant
- The results are very close to constant with an average of 1.02 and a standard deviation of .08

1.029	1.035	1.020
1.015	1.030	1.027
1.011	1.027	1.021
1.027	1.010	1.015
0.961	1.026	0.477
1.032	1.033	1.064
1.019	1.030	1.014
1.024	1.042	1.046
1.030	1.048	1.052
1.003	1.049	1.039
1.008	1.059	
1.044	1.039	
1.015	1.047	
1.010	1.058	
1.023	1.041	
1.034	1.047	
1.058	1.052	
1.025	1.034	
1.018	1.031	
1.028	1.034	

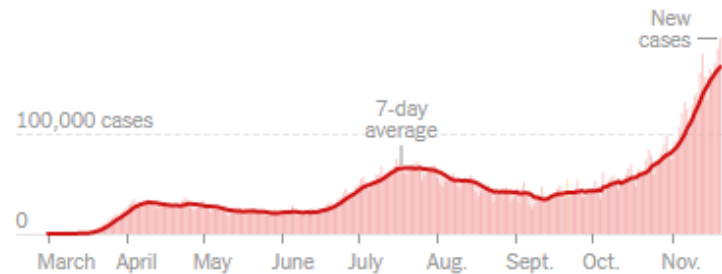
Avg = 1.02, SD = 0.08
with $\lambda = 1$ and $a = 0.07$,
 $bc = 0.09$

Avg = 1.02, SD = 0.08
with $\lambda = .9$ and $a = 0.07$,
 $bc = 0.1$



By The New York Times Updated November 21, 2020, 12:17 A.M. E.T.

[Leer en español](#)



	TOTAL REPORTED	ON NOV. 20	14-DAY CHANGE
Cases	12 million+	198,537	+67% →
Deaths	254,320	1,947	+63% →
Hospitalized		82,178	+50% →

■ Day with data reporting anomaly.
Hospitalization data from the Covid Tracking Project; 14-day change trends use 7-day averages.

We still have a difficult battle ahead of us to control this pandemic!

