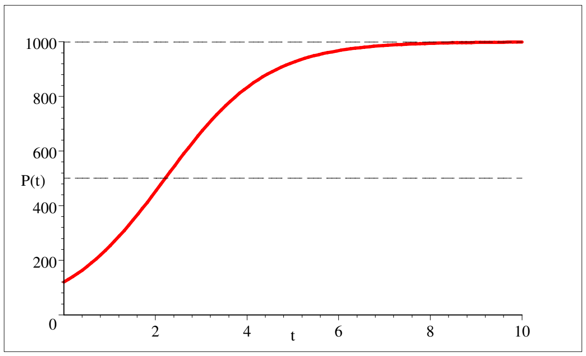
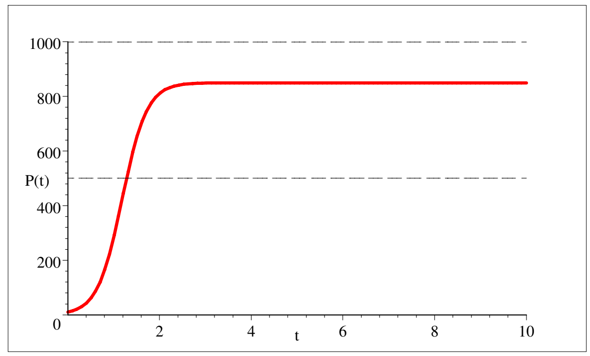
# Epidemics and the Logistic Model

**3/16/2020. Luis F. Moreno, Professor Emeritus, Mathematics, SUNY Broome Community College.**

Certain things have been known since around the 1830's about how organisms reproduce and spread into an available environment as a function of time. I hope you will add this to your knowledge about the current epidemic (and conversely, please add to it).

A **population** is any group of interest to the scientist at a certain time. It can be the number of locusts at time $t$ in an infestation, or the group of people infected at time $t$ by the latest virus, or the number of bacteria at $t$ in a single person. A curve that gives a good model for these processes is the logistic or Verhulst equation (Pierre Verhulst, 1804-1849, et al.). Such curves can look like these, depending on three independent parameters:




Logistic model A

Logistic model B

The X axis is time, measured in whichever units are appropriate: hours for bacteria in a Petri dish or in your body, days for an epidemic or a locust infestation, or years for the spread of deer. In all cases, the vertical axis is $P(t)$, the number of members in the population of interest (bacteria count, number of people infected, deer count). In the case of epidemics, the logistic model counts the total number of infected subjects at any time $t$, assuming that no one is cured and then reinfected.

In all cases of the model, the population eventually levels off, i.e., $P(t)$ becomes practically constant after some time. This constant is called the **carrying capacity** of the environment, $M$ (for "maximum"). In graph A, it looks like $M = 1000$. The other graph shows a carrying capacity of about 850, practically reached after about time $t = 3$. $M$ is one of the parameters of the model, and we may say that $P(t)$ "reaches" $M$, although it can equal $M$ only after infinite time (it is an asymptote).

It is important to note that the popularly hoisted "exponential growth" is a poor model for any living things. This is because it has no carrying capacity. Thus, the population has no bounds! If bacteria could increase exponentially, in a relatively short time they would cover the entire Earth in a layer miles thick, and it still wouldn't stop. Clearly, this model is very limited. Only for times close to zero is exponential growth a tenable model.

The carrying capacity of the environment is affected by many things. In the case of an epidemic, it is reached when everyone in the environment has been infected. The environment may be a nursing home, or a country, or the world. In all scenarios, no new infections occur when $M$ is reached.

Another parameter is the **initial population** $P_0$, i.e., the population at time $t = 0$. In an epidemic caused by a new strain of virus, $P_0$ is the number of subjects infected at the start of record keeping, so $P_0$ is one or more. In model B above, $P_0 = 10$ (maybe the first 10 locusts that flew in). In model A, $P_0 = 120$.

The third parameter, often named $r$, affects the steepness or slope of the logistic curve. We will only use positive $r$, and of the two models above, A has the smaller $r$. For epidemics, we can call it the **virulence**. Model B shows a pathogen that reached its carrying capacity in under three days (or weeks, etc.), starting with just 10 subjects. That's more virulent than model A.

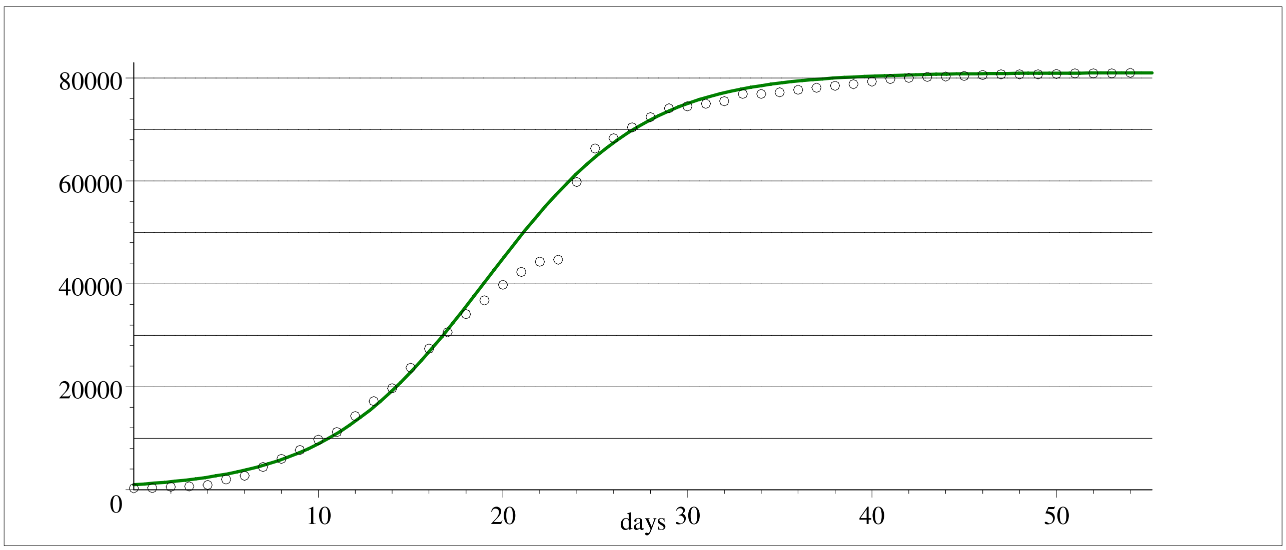
The general formula for the logistic model is

$$P(t) = \frac{P_0 M}{P_0 + (M - P_0)e^{-rt}},$$

where we see the three parameters, and the only variable is time $t$. Logistic model A is $P(t) = \frac{120(1000)}{120 + 880e^{-0.9t}}$. By comparison with the general formula, one can see that $r = 0.9$, $P_0 = 120$, and $M = 1000$. Logistic model B is $P(t) = \frac{10(850)}{10 + 840e^{-3.73t}}$ (with higher virulence).

## Applying the Model

The outbreak in mainland China has been tracked by Johns Hopkins University at this site:
https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6
The data points for the graph below are from that site. The X axis is the number of days since day zero, Jan. 20, 2020. On that day, 278 cases were reported. We can see the carrying capacity at 81000 cases.
The CDC as of March 9 said, "For the majority of people, the immediate risk of being exposed to the virus that causes COVID-19 is thought to be low. There is not widespread circulation in most communities in the United States." So, we are uncertain of the value of $r$.

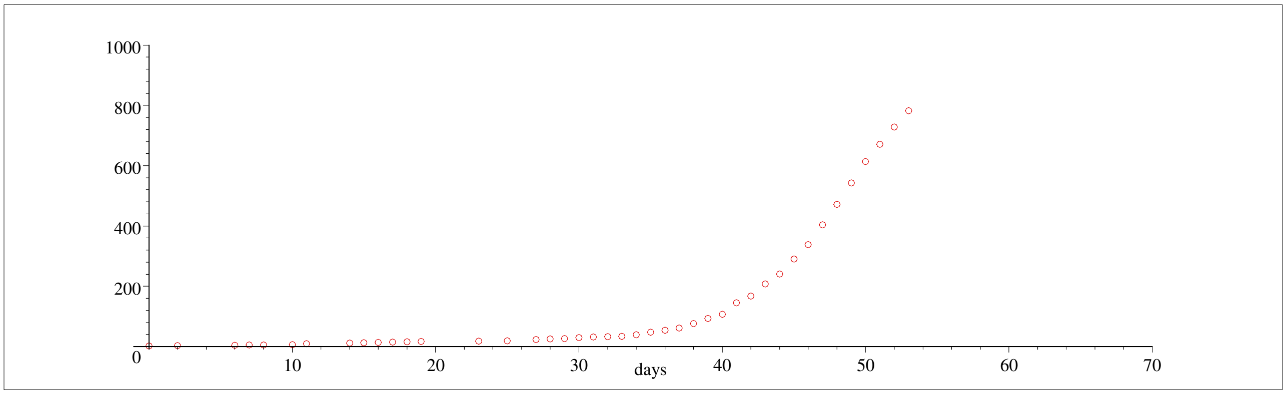



COVID-19 in China. Johns Hopkins data.

Here, the logistic model $P(t) = \frac{P_0 M}{P_0 + (M - P_0)e^{-rt}}$ has $P_0 = 278$, and $M = 81000$. An $r = 0.231$ gives a proper rise to the curve. Finally, a leftward shift produces a good fit to the data points:
$P(t) = \frac{278(81000)}{278 + 80722e^{-0.231(t+5.5)}}$ is the curve you see. One also sees a systematic underreporting error around day 19, which was corrected on day 24. (Another slight shift could also be made so that the curve crosses through the initial 278 value.)

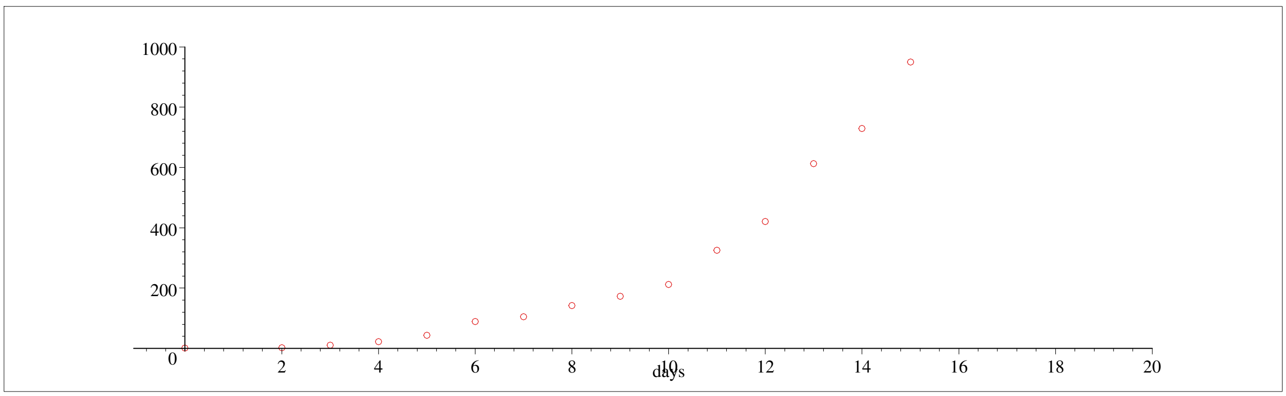
Now let us look at CDC data for cumulative cases in the U.S., as of 3/16. Day zero was Jan. 14, 2020, with two infected subjects. The data is from

https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus




COVID-19 in U.S. CDC 3/16/2020

Below, we see the cumulative cases for New York, beginning at day zero March 1, 2020, with one infected subject. (Source: New York State)




COVID-19 in N.Y.S. 3/16/2020

Neither of these graphs is approaching a carrying capacity $M$, yet. It is too soon to predict its value in either case, but it is surely there. Also, the virulence $r$ is lower for the national data, compared to New York's data.

## Observations and Conclusions

(1) The reason Verhulst's model is so flexible is that a scientist can change its three parameters according to available data. But this is also why it can lead to erroneous conclusions. If the data is incomplete or not statistically representative, then the parameters will be far from true. Applied mathematics interprets reality, but it cannot create truth from falsehood. On the other hand, when the model is judiciously applied, it is just as relevant today as 180 years ago, when Verhulst first published it. Does it apply here? I think so. But we need good values for $M$ and $r$.

In the hands of science, engineering and technology, mathematics creates beautiful generalizations. This one model unifies the phenomena of the spread of real viruses, the spread of malicious codes aptly named "viruses," and an expanding wolf population in Yellowstone Park.

(2) At **no time** should exponential growth be considered. Other terms for it are "geometric progression," "geometric growth," and "the hockey stick curve." These are essentially fake news indicators.

(3) The current epidemic will eventually reach its carrying capacity, whatever that is. Government actions, including vaccination, quarantine, will lower the rate $r$ for the curve in the model. The public perception will be that the virus will inevitably leak out of any containment like sand going through a sieve, but that effective

planning will make the holes in the sieve smaller, buying time. This is very important to hospitals and health care professionals in general.

(4) Mainstream media has little patience for this level of analysis. Reporters should at least point to references. Wikipedia is excellent in these venues. The media should engage epidemiologists who are willing to give a presentation about these functions and models. Substantial discussions shouldn't be relegated to TED talks in times like these.

(5) There are more advanced models, for instance, ones that allow for the entrance of vaccines, so that $M$ may change. One can also go back to the drawing board and not assume that $r$ is a constant, allowing it to vary with time. There are but few "free lunches" in mathematics. As always, the more realistic the model, the harder it becomes to work with.

(6) I have not used statistical fitting for the China data yet, because I wanted to look at the original mathematics upon which later statistical models rely. I am confident that the results will agree well.

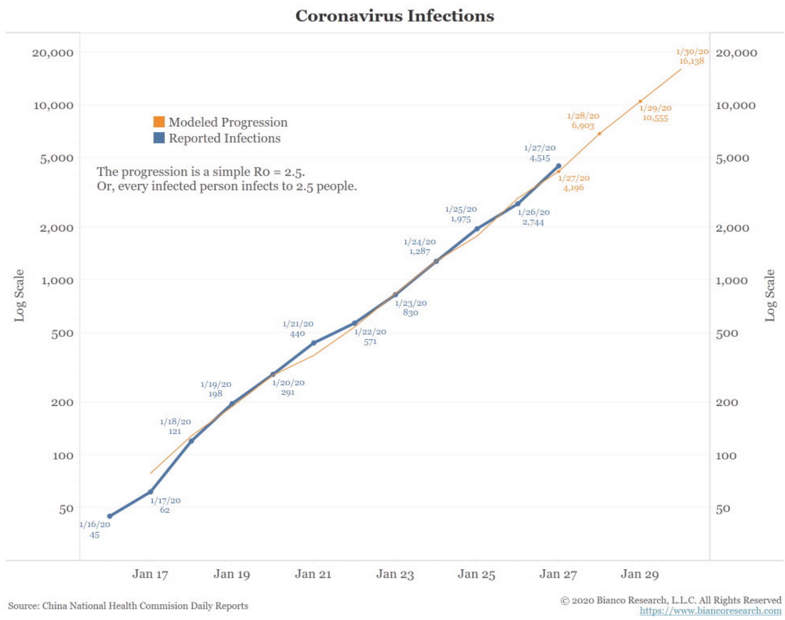
(7) In the example below, *Bianco Research* in late January presented data from China and proceeded to model it by exponential growth. The inevitable conclusion: by February 20, they expected 138 million infected subjects. This is a classic case of good data fed into a miserable model. Exponential growth models should be used only in calculus classes for pedagogical purposes (with appropriate caveats), and never by journalists.

*Bianco Research* 1/28/2020

The growth in coronavirus infections has continued along a geometric progression for the last 12 days [1]. Should it continue along this path, infection cases could approach 100,000 in a week.

The following charts were constructed from the daily update from the National Health Commission of the People's Republic of China. The blue line in the chart below shows the actual number of reported coronavirus cases stands at 4,515 as of January 27. The orange line is a simple progression that assumes a 53% increase in the cases every day. Or, one person infects 2 to 2.5 people. So it is a simple multiplier, nothing more. This is known as R0 (R-Naught), or the infection rate. Note the chart is a log scale. The reported number of infections perfectly track this simple multiplier. This is how viral inflections growth, along a geometric path.[2] The chart below shows that the virus has tracked this growth rate 12 straight days. If this track is not altered, the number of reported cases will top 16,000 by Friday. To many, such a geometric progression is alarming...

As the orange line below shows, this type of growth rate would suggest 80,000+ infections next Monday and 138 million by February 20... To be absolutely clear, this is NOT a prediction that 100 million people will be infected by Feb 20.[3] Rather, this has been its growth rate for the last 12 days. A vaccine, mutation or successful quarantine/isolation could help reduce this growth rate.[4] Is this growth rate possible? Over the near-term yes.




[1] This is the assumption for exponential growth.

[2] They have fallen for the exponential model.

[3] They are now realizing that this is crazy. The exponential model that fits the data is $y = 45e^{0.419t}$, where $t$ is the number of days after 1/16/2020. But it is only good for the time period within the chart. On February 20, 35 days after 1/16, this model predicts 105 million infected subjects! (I don't know how they got 138 million.)

[4] Or, try a better model...