Introduction

P and N functions. The cumulative total number of Covid-19 cases confirmed up to and including day t is denoted by $P(t)$. We allow $t = 0$ to correspond to any given date.

The number of new cases confirmed on N on day t is given by

$$N(t) = P(t) - P(t-1).$$

We sometimes also call $N(t)$ the...

- daily increase in cases on day t.
- daily cases on day t.
- new cases.

Note that with these definitions

$$\begin{aligned} P(t) &= P(t-1) + N(t) \\ &= (\text{Total cases confirmed up to and including day } t) + (\text{New cases confirmed on day } t) \end{aligned}$$

Data sources. All data is sourced from the *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University* accessible at the following link

- https://github.com/CSSEGISandData/COVID-19.

We reference this source as JHU in short.

Getting the data for your own explorations, or for students to explore. Raw data for countries and US states, together with their 7-day smoothing (see below) can be downloaded using a couple of websites that Isaac Flath built for the Consortium.

- https://isaac-flath.shinyapps.io/coronavirus2/
- https://isaac-flath.shinyapps.io/Covid-19/

These two sites also generate plots for any given time range for all the quantities we discuss in the problems we have written (P, N, relative change of P, log plots of P, etc).

Data smoothing. Covid-19 data (specially N-data) is very noisy, so in all the problems we have written we use the 7-day moving average of the data which we compute as follows ($\widetilde{P}(t)$ is the raw time series obtained JHU):

$$\begin{aligned} P(t) &= (\widetilde{P}(t) + \widetilde{P}(t-1) + \ldots + \widetilde{P}(t-6))/7 \\ N(t) &= P(t+1) - P(t) \end{aligned}$$

so, the new cases time series $N(t)$ gets smoothed through the smoothing of $P(t)$ and does not get smoothed separately. This makes the relation $N(t) = P(t+1) - P(t)$ continue to hold after the smoothing. We also make sure that we pick up JHU data well before $t = -7$, so the data is smoothed using the above formula even for $t = 0$ using the previous 7 data points in the JHU time series.

Figures 1 and 2 show the effect of this smoothing for the cumulative cases P and new cases N for Romania, with t in days since March 1, 2020 (thick blue is the smoothed data, thin black is the raw data). A side product of this smoothing is that data gets shifted forward (to the left) by 3.5 days.

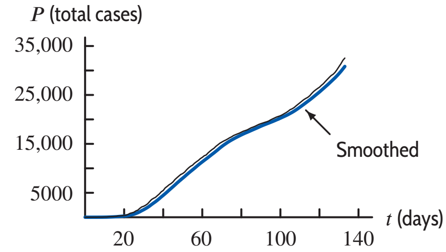

Figure 1

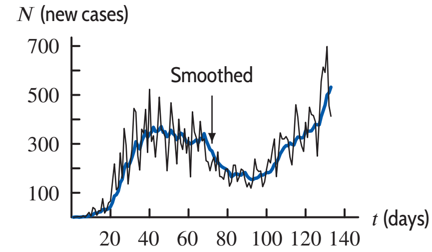

Figure 2

Expected shape of P and N. The Covid-19 outbreaks have occurred in waves in many countries. Each wave has an N-plot which is bell-shaped and that will correspond to a sigmoid shaped P-curve.

Subsequent waves stack on top of one another and result in N-plots that look like a sum of bell-shaped curves (which may or may not drop all the way down to zero between the bells), and this results in a P-plot that will look like a sequence of stacked sigmoid s-curves. Figures 3 and 4 are the P and N plots for Australia between March 1 and August 16, 2020.

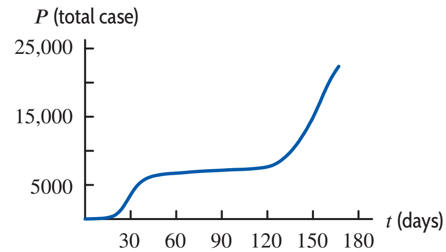

Figure 3

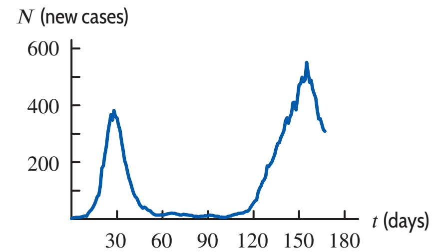

Figure 4

Of course, the outbreaks in many countries do not look so distinctly as a sum of waves. Besides possible data collecting issues or changes in data collecting strategies, drastic measures implemented to control the outbreak have noticeable effects on the way the outbreak developed. Figures 1 and 2 are good examples of expected shapes P and N plots when the waves are not separate.

Log plots of P

Logarithmic scales on the vertical axis have been used extensively in news outlets covering Covid-19 outbreaks due to the initial approximately exponential phases in many outbreaks around the world.

Expected shape of the log plot of P. For a single wave in an outbreak, the log plot of P is expected to look like Figure 5, which shows the first wave of the outbreak in the Netherlands.

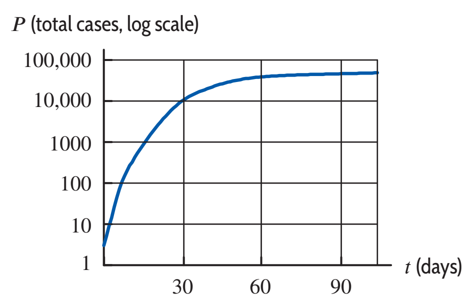

Figure 5

What students need to have seen to do these exercises. Many of the exercises that have been written allow students to think about log plots without using logs at all. They just need to understand that

- in log scales equally spaced marks correspond to multiplication by the same constant.
- exponential growth will look straight on a semi-log plot using a logarithmic scale on the vertical axis.

Several exercises have been written to make students comfortable with these two facts.

Radical changes in the data. Log plots sometimes reveal radical changes in the way the Covid-19 outbreaks developed. For example Figures 9 and 10 show the outbreak South Africa with $t = 0$ corresponding to March 18, 2020. There is a sharp change around $t = 14$, and an internet search for the Covid-19 timeline in South Africa reveals that a strict lockdown was imposed precisely around $t = 14$.

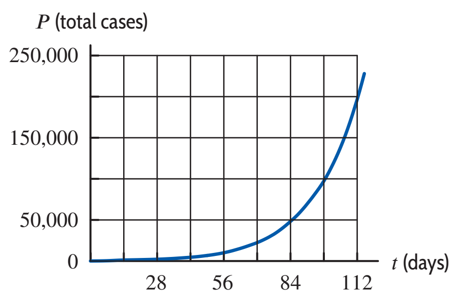

Figure 6

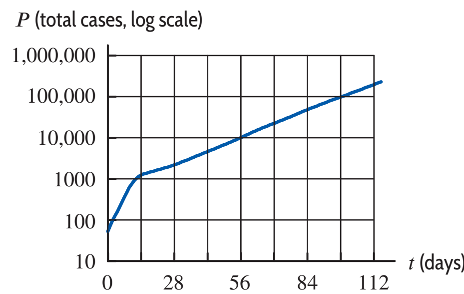

Figure 7

Shortcomings of logs plots in the Covid-19 context. News outlets mostly stopped using them after the first wave. Probable reasons are:

- Eventually many P plots for countries started looking very similar (concave down stabilizing at a point), like Figure 5.
- Slow linear growth at the end of the sigmoid of the first wave looks almost flat on a log scale. Figure 5 itself is a good example of this: it is hiding the fact that outbreak is growing linearly with approximately 1000 daily cases in June. This is discussed in one of the problems we have authored.

Data exploration ideas for students. Use https://isaac-flath.shinyapps.io/coronavirus2/ to

- find countries or states with approximate exponential phases.
- find countries or states with two different approximately exponential phases with different daily growth rates. Try to figure out what led to the change in the rate of spread.
- for countries or states with exponential phases how much variation is seen in the growth rate between them?

Text for students. An expository text on log plots for students which includes the above ideas (among many other interesting things) has been written by Dan Flath from the Consortium accessible at the MCWG Covid resource page.

More problems and problem ideas at the MCWG Covid resource page. Visit `https://mcwg.github.io/covid`.

Sample problems - Log Scales - see the Instructor notes package for more

1. Figure 8 is a log plot of the number P of cases of Covid-19 for four countries A, B, C and D.

(a) How can you tell from the plot that P is growing exponentially in all four countries?

(b) Since growth is exponential, there is a constant doubling time and a constant time for P to be multiplied by 10. Complete the table.

Country	A	B	C	D
doubling time (days)	6.0		12.0	
days to multiply by 10		33.2		99.7

(c) Find the ratio (days to multiply by 10)/(doubling time) for each country. Give a simple explanation for the values of the ratios.

(d) The ratios in part (c) are all the same. Is the time for P to multiply by 10 proportional to the doubling time?

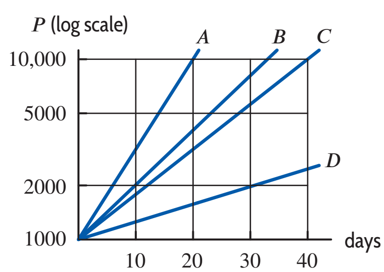

Figure 8

2. In March 2020, the government of South Africa announced a 21 day national lockdown starting on March 27 and ending on April 16 to combat the spread of the Covid-19 outbreak in the country.[1] Figure 9 shows the cumulative total reported number of Covid-19 cases[2] in South Africa, P, during the initial phase of the 2020 pandemic where t is the number of days after 3/18/20. Figure 10 shows a plot of P using a logarithmic scale on the vertical axis, and Table 1 contains the values of P on some dates.

Even though the linear scale plot in Figure 9 looks exponential, one can see in Figure 10 that there were in fact two exponential growth phases with different growth rates.

(a) Estimate the value of t when the total number of confirmed cases was 100,000.

(b) Approximately how many confirmed cases where there when the lockdown began?

(c) Explain why the change between the two exponential phases is not visible in the linear scale plot in Figure 9.

(d) Find the percent growth rate, doubling time and tenfold time during each exponential phase. You may find the data in Table 1 of use.

(e) If lockdown measures had not been implemented and the first exponential phase had continued, approximately on what date would have P reached its actual July 8 value (which corresponds to $t = 112$)? Use Figure 10 to make a rough estimate.

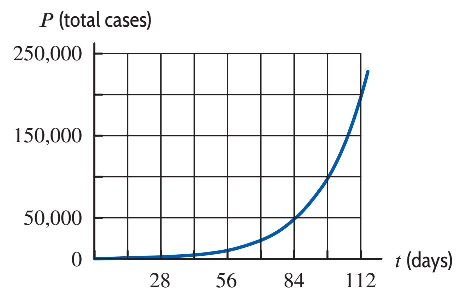

Figure 9

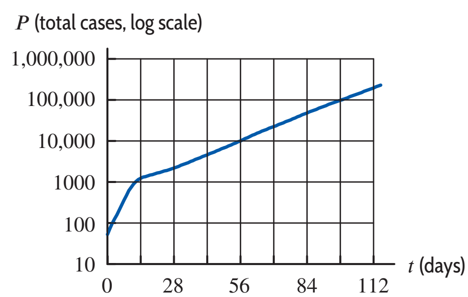

Figure 10

Table 1

Date	t, days since 3/18/20	P, confirmed cases
3/18/20	0	53
3/23/20	5	207
3/28/20	10	746
04/02/20	15	1308
04/17/20	30	2397
05/17/20	60	12,887
06/16/20	90	65,937

[1]https://en.wikipedia.org/wiki/COVID-19_pandemic_in_South_Africa , accessed on 19 January, 2020
[2]Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on July 12, 2020.

3. Figure 11 shows the total number of confirmed Covid-19 cases $P(t)$ in the United Kingdom[3] up to and including day t, where $t = 0$ is March 4, 2020. Figure 12 shows the same data using a logarithmic scale on the vertical axis. Note that P is increasing and concave up in the linear-scale plot (Figure 11). Was P growing exponentially during this period of time? How does the logarithmic scale help you decide?

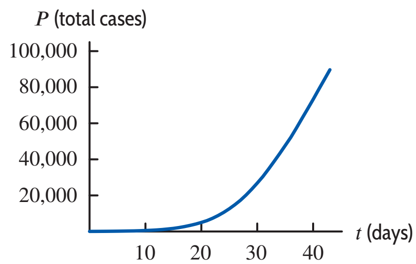

Figure 11

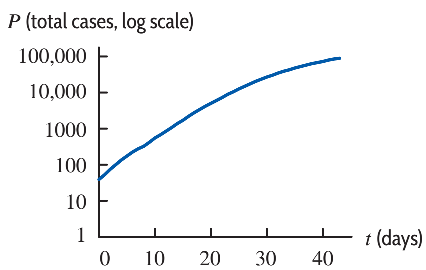

Figure 12

4. Figure 13 is a plot of the number P of Covid-19 cases in a state during 1 month using a logarithmic scale on the vertical axis.

(a) How can you tell from the graph that the number of cases is growing exponentially?

(b) How many cases were there on day 0? What is the doubling time?

(c) Mark 4000 and 8000 on the proper places on the vertical axis.

(d) Draw on the same set of axes an exponential growth function with 2000 cases on day 0 and the same doubling time as in part (b).

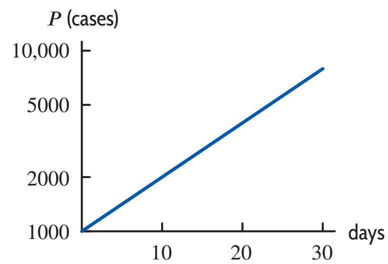

Figure 13

5. Use Figure 14 to determine whether the total number of confirmed Covid-19 cases in India[4] increased more between April 16 and May 1 ($0 \leq t \leq 15$), or between June 15 and June 30 ($75 \leq t \leq 90$). Note the logarithmic scale on the vertical axis.

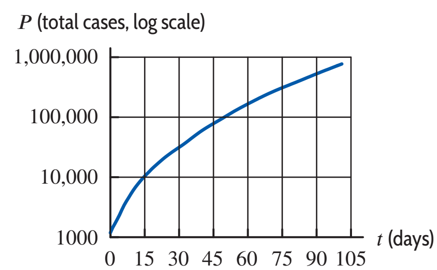

Figure 14

[3] Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on June 26, 2020.

[4] Data from Johns Hopkins University, downloaded from https://github.com/CSSEGISandData/COVID-19 accessed on July 12, 2020.