Modelling the spread of the SARS-CoV-2 in China

George Watkins, MATH310, Simon Fraser University, 2020-03-05

Exponential Growth

Let $Q(t)$ be the total number of reported cases of the virus at a point in time. Assuming that new cases are proportial to the population, with proportionality constant *r*, we can express the realtionship with the differential equation:
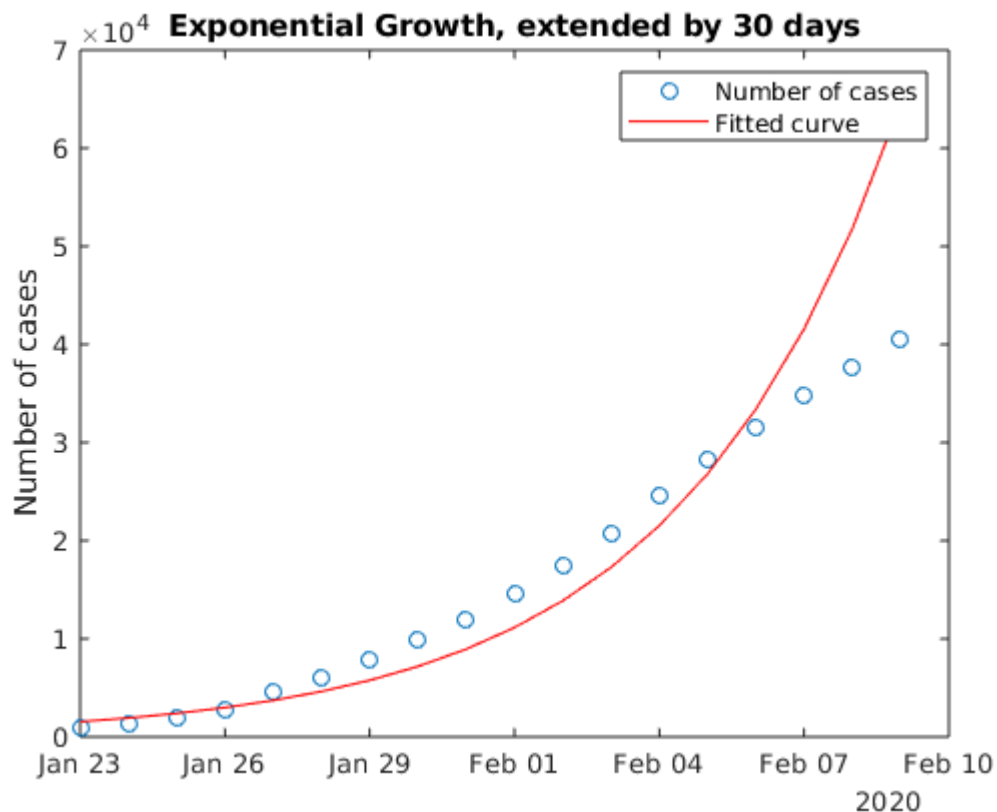
$$\frac{dQ}{dt} = rQ$$

This equation has an analytic solution. For initial condition $Q(0) = Q_0$:

$$Q(t) = Q_0 e^{rt}$$

We have values of $Q(t)$ for many days, those are the number of reported cases. This is data publicly available in [1].

What we want to find are Q_0 and *r*. So we find the best fit exponential curve through the data.

```
a = 1.5438e+03
b = 0.2196
```

Doesn't seem to fit the data too well. This model is good in the early stages of the spread of a disease when the new case proprtionality assumption holds, but as time goes by cases start increasing more slowly.

Logistic Growth

In logistic growth we assume that the proportionality constant r of the exponential growth model decreases over time. So we replace it by a function of Q that decreases as cases increase: $r - \alpha Q$.

The differential equation for this type of growth becomes:
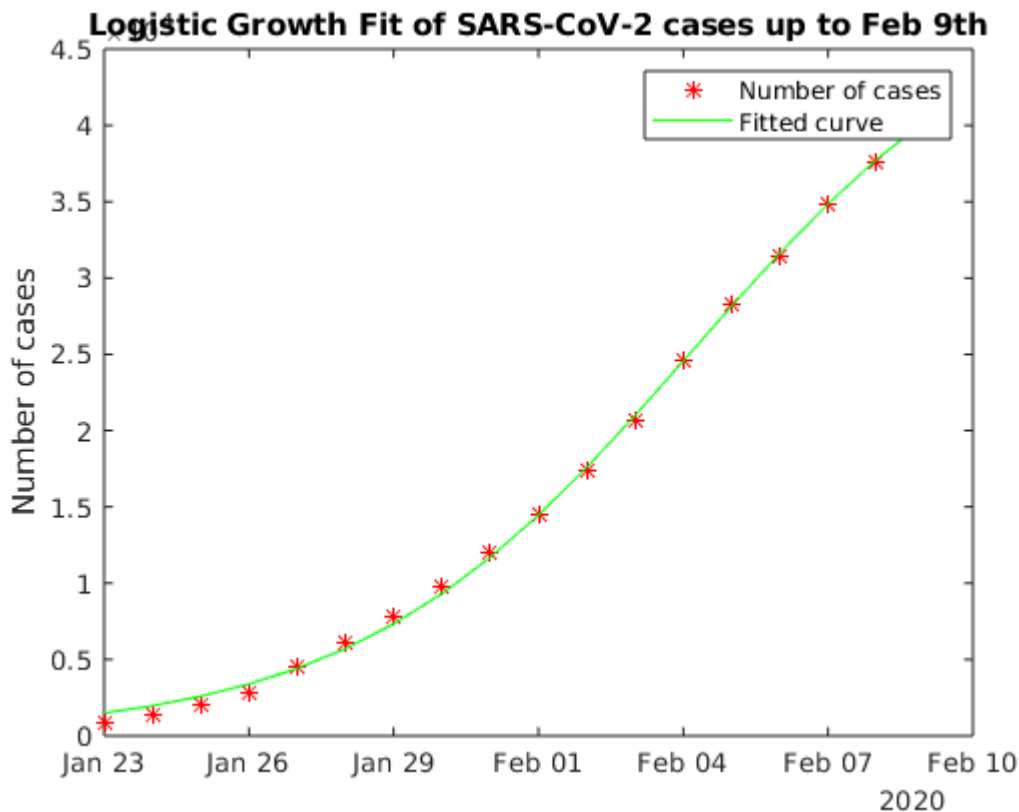
$$\frac{dQ}{dt} = (r - \alpha Q)Q$$

The analytic solution can be expressed in the form:

$$Q(t) = \frac{Q_{\text{inf}}}{(1 + \exp(-\alpha(t - t_{\text{half}})))}$$

Where Q_{inf} is the number of cases as $t \to \infty$, and t_{half} the time at which $Q(t) = \frac{Q_{\text{inf}}}{2}$.

That is the form a helper library[2] uses. So we can use it to fit a curve of such form to the data.

```
thalf = 12.1724
Qinf = 5.0361e+04
alpha = 0.2866
```


Logistic Growth Fit of SARS-CoV-2 cases up to Feb 9th

This model seems to fit the data much better.

Logistic Growth Model Prediction

The curve fitting was done at the beginning of February.

Logistic Growth and cases up to Feb 9th extended by 30 days
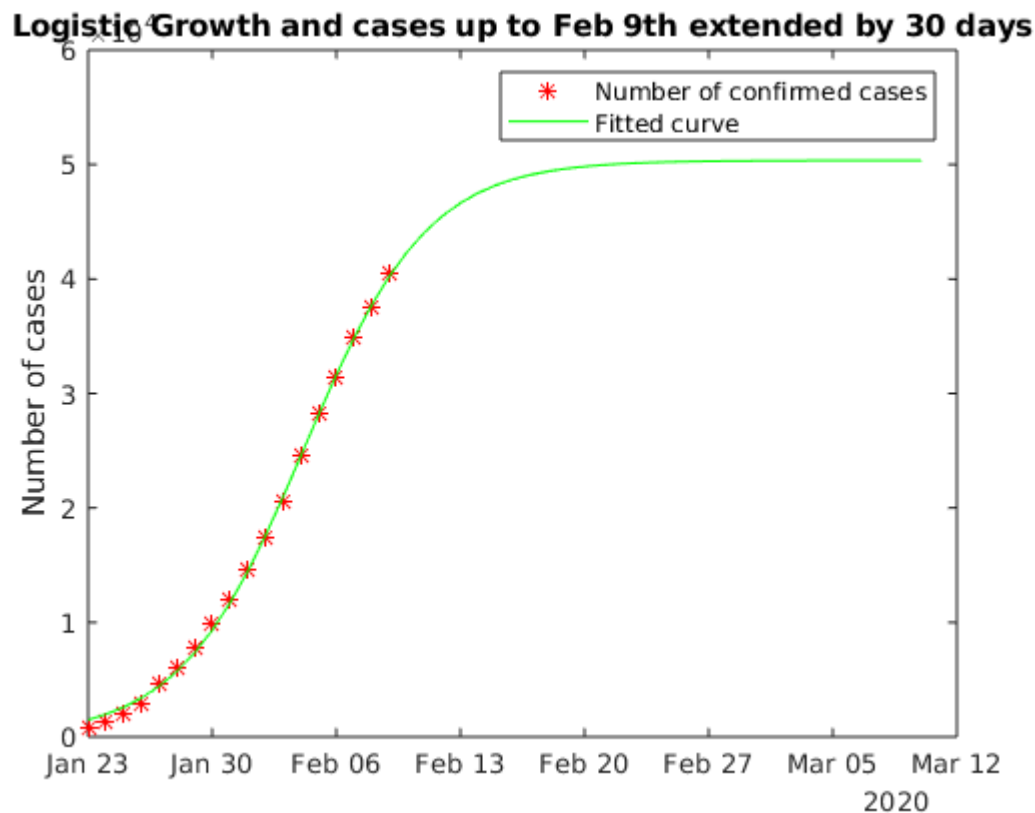
Fits the data better, but common sense might give us a different intuition for the future development of cases of the virus.

Almost 30 lays later

Up to date data was retrieved from the Hong Kong government about cases of the SARS-CoV-2 cases in Mainland China [3]. So we can plot the new data against the old estimate

Logistic Growth and cases up to Feb 9th against new data
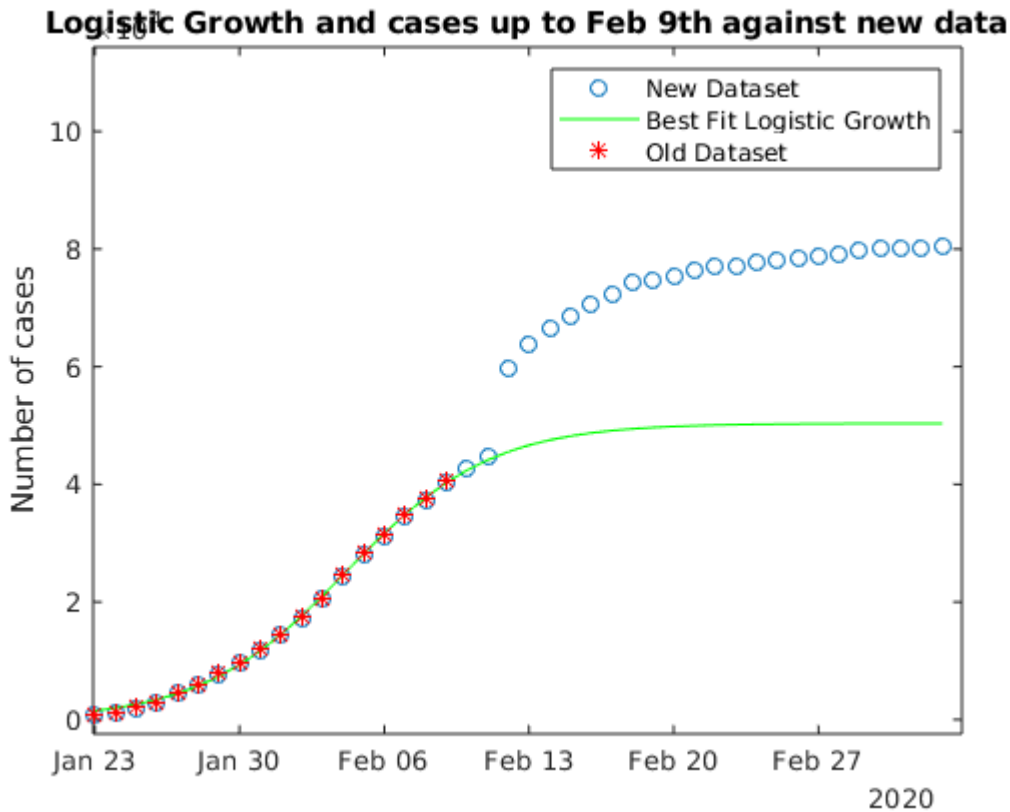
The big "jump" in the number of cases is when the the medical procedure for testing for the virus changes. This resulted in many more cases being recognized as of the SARS-CoV-2 infections.

Assuming that the the new SARS-CoV-2 test detects a number of cases proportional to the number of cases being detected before, we could try to correct for this change in testing procedure as follows: take the ratio between the actual number of cases and the projected number as an adjustment factor. Then scale the curve and the values before the day of the jump by the adjustment factor.

We choose to adjust the preceding data so that our prediction is using current values.

Logistic Growth and adjusted number of cases
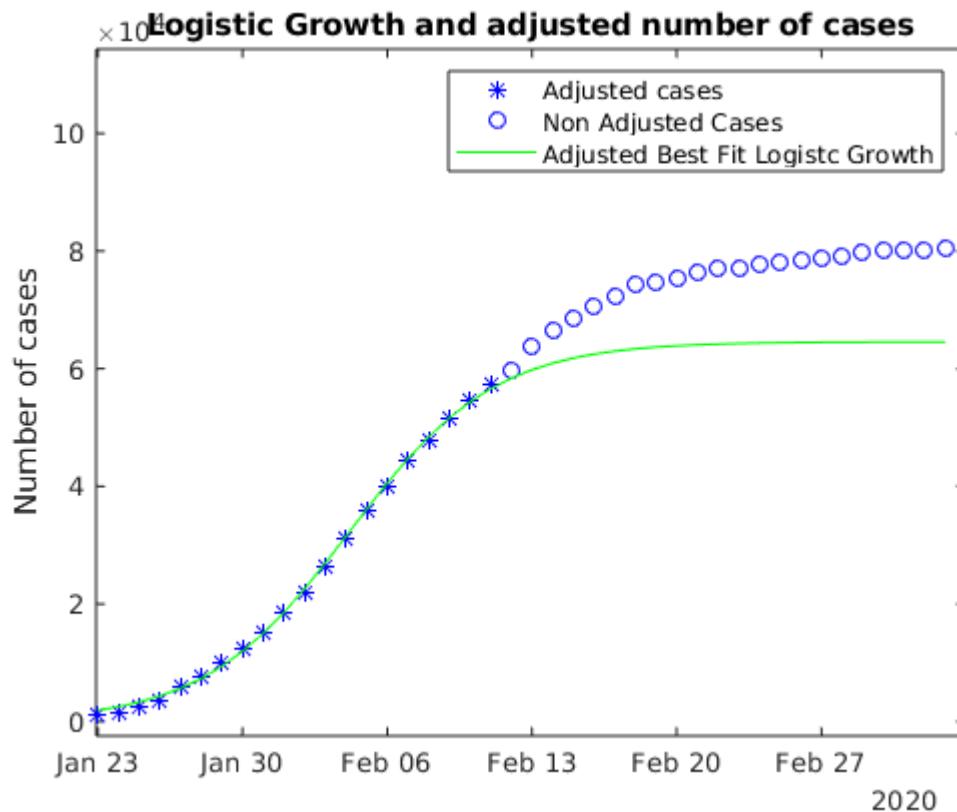
Over all it seems like the model catpured the general pattern of the data, but it diddn't quite give us an acurate prediction. Even after introducing the ajustment factor ajustment factor.

Towards a better model

During the EpiCoronaHck event, organized by SFU, I learned more about modelling epidemics.

A frequently used model in epidemology is the SIR model. Its name comes from the fact that it describes the evolution over time of:

- *S*, the number of people succeptible to the disease. For new diseases like this one it is often assumed to be the entire population.
- *I*, the number of infected people.
- *R*, the number of people *removed* from the population, either by healing (assuming that the disease can't be contracted again) or otherwise

The relationships between the above quantities are given by the following system of differential equations:

$$\frac{dS}{dt} = -\frac{S}{N} \cdot \frac{R_0}{D_I} \cdot I$$

$$\frac{dI}{dt} = +\frac{S}{N} \cdot \frac{R_0}{D_I} \cdot I - \frac{I}{D_I}$$

$$\frac{dR}{dt} = \frac{I}{D_I}$$

Where

- R_0 is the Basic Reproduction Number. Is the expected number of new cases generated by a single case. The news like to talk about this number.
- D_I is the he infectuos period.
- *N* is the total population.

This model is not linear (because of the product of *S* and I), and there is no known analytic solution, so numerical methods must be used.

In this model we assume that an individual shows symptoms and is contagious immediately after it coming in contact with the virus. There exist models such as the SEIR model that add an *Exposed* part of the population that has been exposed to the virus but might not show any symptoms yet, while already being contagious. Models of the sort are probably more accurate for the SARS-Cov-2 virus.

References

[1] Worldometers Dataset: https://www.worldometers.info/coronavirus/coronavirus-cases/

[2] James Conder (2020). fit_logistic(t,Q) (https://www.mathworks.com/matlabcentral/fileexchange/41781-fit_logistic-t-q), MATLAB Central File Exchange. Retrieved February 8, 2020.

[3] Hong Kong Dataset: https://data.gov.hk/en-data/dataset/hk-dh-chpsebcddr-novel-infectious-agent

The EpiCoronaHack event's Github page: https://github.com/EpiCoronaHack/Hackathon2020