A Covid-19 Puzzle: When is Exponential Growth not Exponential Growth?

Benjamin Lengerich*

Carnegie Mellon University

March 30, 2020

1 Introduction

Suppose a population of N individuals is infected by a novel virus. Because this virus is novel, there is no testing available at the start of the infection. After some delay, testing becomes incrementally available. Each day, more and more test kits identify more and more infected members of the community. Because these growth factors occur simultaneously, it is very difficult to separate the growth in the size of the infected population from the growth in testing availability. Here, we propose a simple probabilistic model to separate these growth factors.

For simplicity, let's assume that (1) the diagnostic test is perfectly accurate, and (2) each individual is tested only once.

We have N total individuals in the population. At each timepoint t, we observe the following random variables:

- $n_t(t)$, the number of tested individuals
- $n_p(t)$, the number of positive tests
- $n_n(t)$, the number of negative tests

In addition, we have the unobserved latent variables:

- $z^i(t)$, a Boolean variable indicating whether individual i is truly infected
- $t^i(t)$, a Boolean variable indicating whether individual i is tested
- $T(t)$, overall testing availability (this will end up being expressed in terms of the previous variables)

Now let $\mathbb{P}(t^i(t) = 1 | z^i(t)) = T(t)\big((1 - z^i(t)) + cz^i(t)\big)$, i.e. an infected individual is a factor of c more likely to be tested than an uninfected individual, with overall testing availability $T(t)$. We would like to estimate the total size of the infected population at time t:

$$Z(t) = \sum_{i=1}^{N} z^i(t) \tag{1}$$

*blengeri@cs.cmu.edu

To do so, let us first express the total number of tests as a sum over individual tests:

$$\mathbb{E}[n_t(t)|z(t)] = \sum_{i=1}^{N} \mathbb{P}(t^i(t) = 1|z^i(t)) \tag{2}$$

$$= T(t) \sum_{i=1}^{N} 1 - z^i(t) + c(z^i(t)) \tag{3}$$

$$= T(t)\big(N - \sum_{i=1}^{N} z^i(t)) + c \sum_{i=1}^{N} z^i(t)\big) \tag{4}$$

$$= T(t)\big(N - Z(t) + cZ(t)\big) \tag{5}$$

The expected numbers of positive tests are:

$$\mathbb{E}[n_p|z(t)] = \sum_{i=1}^{N} \mathbb{P}(t^i(t) = 1|z^i(t))z^i(t) \tag{6}$$

$$= T(t) \sum_{i=1}^{N} \big(1 - z^i(t) + c(z^i(t))\big)z^i(t) \tag{7}$$

$$= T(t) \sum_{i=1}^{N} c(z^i(t)) \tag{8}$$

$$= T(t)cZ(t) \tag{9}$$

i.e., $T(t) = \frac{\mathbb{E}[n_p(t)]}{cZ(t)}$. Plugging this back into Eq. 5, we have

$$\mathbb{E}[n_t(t)|z(t)] = \frac{\mathbb{E}[n_p(t)]}{cZ(t)}\big(N - Z(t) + cZ(t)\big) \tag{10}$$

$$= \frac{\mathbb{E}[n_p(t)]}{c}\big(\frac{N}{Z(t)} - 1 + c\big), \tag{11}$$

which allows us to solve for $\mathbb{E}[Z(t)]$:

$$\mathbb{E}[Z(t)|n_t(t), n_p(t), n_n(t)] = \frac{n_p(t)N}{cn_t(t) - (c-1)n_p(t)} \tag{12}$$

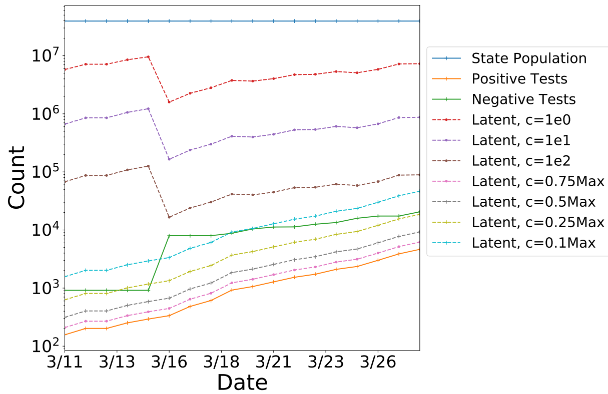
$$= \frac{Nn_p(t)}{cn_n(t) + n_p(t)} \tag{13}$$

2 Applying to Real Covid-19 Counts

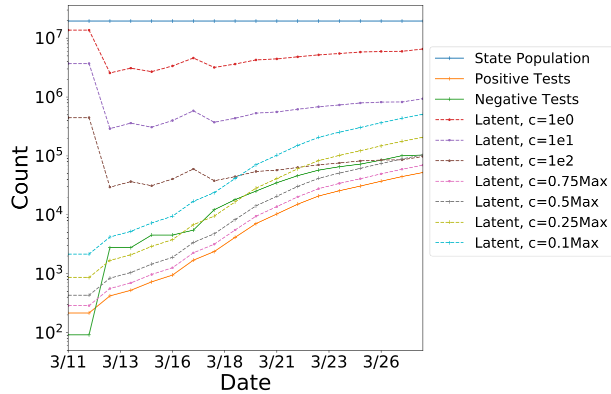
Not let's see what happens when we apply this model to estimate the number of latent cases in each state as of March 29, 2020. Results for a few large states are visualized in Fig. 1. At each timepoint, there is a maximum feasible c, defined by $c \le \frac{N - n_p(t)}{n_n(t)}$, which is bounded by the observed number of cases. At this maximum c, the implied number of latent cases exactly matches the observed number of positive tests.

These plots show a striking flatness (little response to the growing number of positive cases) for each value of the c parameter. This represents the counter-intuitive finding that *if we observe only these total counts, an exponentially-growing infection is statistically indistinguishable from exponentially-growing testing*. Intuitively, we can consider an infection which is fixed in some proportion of the full population. If testing grows exponentially, then the counts of positive tests would grow exponentially without the infection spreading at all.

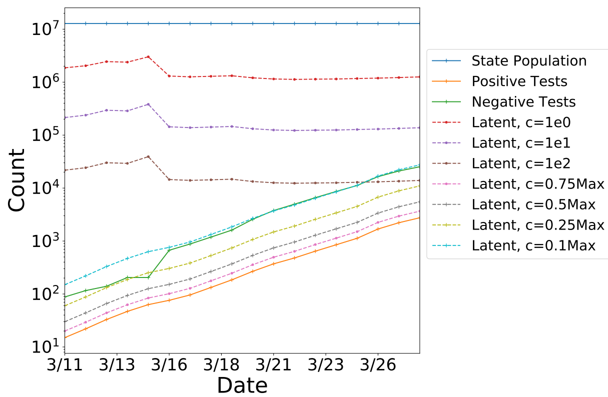
In the present case of COVID-19, testing is expanding. In addition, testing protocols are changing and the value of c is changing over time. This means that the true curve of the size of the infected population is likely to be transitioning between curves. However, we should not necessarily view exponentially-growing case counts as complete proof that the true numbers of cases are exponentially-growing. As the testing availability continues to increase, this picture will come into clearer resolution.



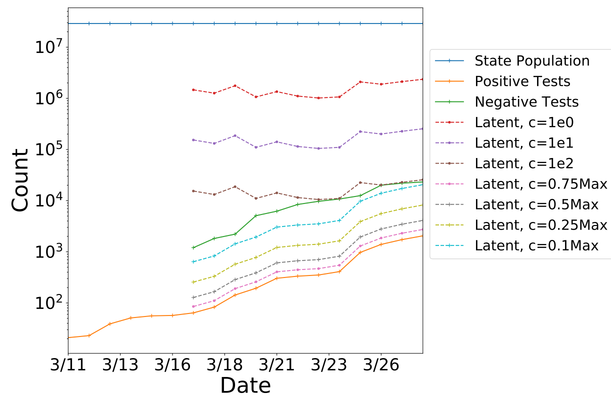
(a) California



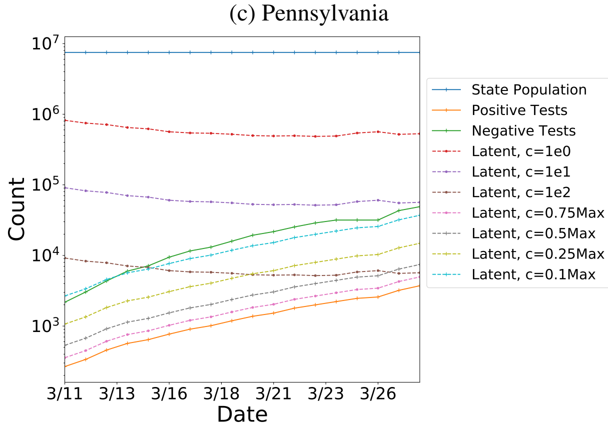
(b) New York



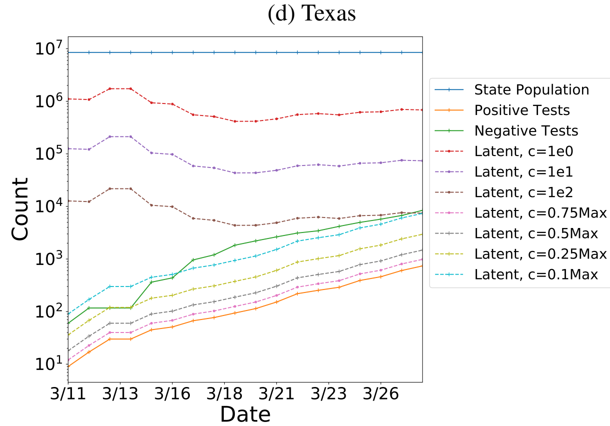
(c) Pennsylvania



(d) Texas



(e) Washington



(f) Virginia

Figure 1: Applying the model to the counts in different states can produce very different estimates of latent cases by changing the value of c. However, for constant c, each curve is relatively flat and mostly unchanged by the growing number of positive tests. When the true positive curve crosses above the latent curve for a given value of c, we can consider that value of c to have been too high.