About Statistical Methods

LiJin Joo lijin.joo at gmail.com

2020-11-03

Smoothing Data and a Decision Rule

To determine a trajectory of data, we smoothed the data by 21-day moving average (21-DMA). The raw counts include substantial day-to-day variability for the different data collection mechanisms, reviewing processes of hospitals and wards (i.e. administrative units in India), test capacity, and occasional checkups and regulation measures followed a sudden increase of cases. A 21-day window, chosen by a convention in the field, includes the data of current and past two weeks from today. A statistic of today reflects the previous 20 days of information and smooths out stochastic fluctuations during this period.

Most statistics and metrics, provided by our dashboard, are smoothed by 21-DMA. Any decision based on the data must consider the fact that the observation of today is not yet complete. We determine if data is in a downward trajectory testing if a mean of this week (7D) is smaller than a mean of the previous two weeks (14D) and vice versa.

Disease Dynamics

Reproduction number (R) and epidemic growth rate (r) are two statistics used to describe disease dynamics, or transmission of COVID-19. Each of them measures a different aspect of disease dynamics, thus, neither of them is interchangeable with the other. Reproduction number R tells about a direction of growth, whether an epidemic is growing or shrinking, but does not tell how fast a number of cases is growing at a given time point. On the other hand, the growth rate r provides the information about size and speed of change at one time point, but does not predict how likely it is that the number of cases will grow or decline over the next time period. The goal of presenting these two statistics is to assess the effectiveness of public health interventions, such as community lockdowns, and to assist decision making around relaxing or strengthening interventions based on the most recent data (Anderson et al. 2020) (Science and Scientific Advisory Group for Emergencies 2020).

Reproduction Number (R)

The reproduction number (R) is the average number of secondary infections produced by a single infected person and shows the potential for transmission. If R is greater than 1, every infected patient produces more than one new case on average, thus, the disease continues to spread. If R is smaller than 1, new cases from each existing patient are less than one on average, and the incidence of the disease will decline and eventually stop.

Initially, the basic reproduction number, R0, measures potential for disease transmission at the beginning of an epidemic, assuming a population never previously exposed to the pathogen (Delamater et al. 2019). However, in reality, it takes some time for an epidemic to be manifested. To estimate the number of secondary infections from an extremely small number of initial cases, complex mathematical models under various sets of assumptions are necessary. One well-known model for separating secondary infections from observed cases is “susceptible, incubation, infectious and recovered (SEIR) model”, which models each of the four populations using a series of ordinary differential equations (Anderson et al. 2020).

Among the variations, we used a version of R called "time-varying" reproduction number, which concerns how disease transmissibility evolves over time, rather than the inherent biological characteristics of a pathogen. As infectious cases accumulate over time, cases grow faster with more cases from the susceptible, exposed and recovered populations, yet the speed of growth slows down over time within a population of finite size. Therefore, Rt is always smaller than R0.

For Rt, we used the method developed by (Cori et al. 2013), which takes a series of incidence values only and estimates the values of the necessary parameters from their empirical density and Bayesian model assumptions. The formulas are provided by R package {EpiEstim} and {EpiNow2}.

- The formula for Rt from {EpiEstim} is as follows.

$$R_t = \frac{I_t}{\sum_{s=1}^{t} I_{t-s}\omega_s}$$

  where I_t is incidence at time t and ω is a discounting weight for an incidence at time s at the present time t ($s < t$). This weight is estimated from an empirical density called "serial intervals", the average time between the first symptom of a transmitter and that of newly infected. To be compatible with the Poisson model for case series, a conjugate Gamma model is used. The Poisson model predicts the number of new cases at time t, I_t by approximation with a series of estimates of incidence as below.

$$P(I_t|I_0, I_1, \ldots, I_{t-1}, \omega, R_t) = \frac{(R_t\Lambda_t)^{I_t}}{I_t!}$$

with $\Lambda_t = \sum_{s=1}^{t} I_{t-s}\omega_s$.

- As I_t is a stochastic process, R_t is as well. At time t, R_t is right censored due to unreported cases because of 1) delay in reporting, 2) incubation period of new cases, 3) generation time, the time between infection events. These three parameters are incorporated in calculations by {EpiNow2} by (Abbott et al. 2020), which contribute to the better uncertainty estimates for Rt.
- We used the estimate_R() function with a "parametric_si" option assuming a gamma prior with a mean of 4 days, a standard deviation of 4.75 days, and a 7-day smoothing window suggested by (Abbott et al. 2020) , (Basu et al. 2020). However, Bayesian estimates give unacceptably small variations (sd $<$ 0.01) for some period where Rt remains near 1 with a large number of new cases. In this case, extra permutations are pursued to obtain a greater uncertainty estimate.

Growth Rate (r)

The growth rate r is calculated by the second order derivative of an incidence (generating) function and indicates how quickly infections are changing at time point t. A growth rate also shows how many new cases are likely to occur for the next time period, usually under the exponential growth assumption for an infectious disease, and can serve as an instantaneous indicator for the impact of a public health intervention. If the growth rate is greater than zero (or positive), the epidemic is growing. If the growth rate is less than zero (or negative) then the epidemic is declining (Anderson et al. 2020).

- Calculation of r is simple and does not require a model as in Rt. However, care must be taken to interpret the size of r on an epidemic growth curve. For instance, even if r is same, $, the growing phase of an epidemic with a small number of new cases and a larger Rt is more critical than the stable phase of an epidemic with a large number of new cases but a smaller Rt. To enhance the interpretability of r, we provide two more statistics.
- Doubling Time (d_t): we used the formula from (Basu et al. 2020) as follows:

$$d_t = \tau \frac{log(2)}{log(1+r)}$$

  where the growth rate r is given by

$$r = \frac{I_t - I_{t-1}}{I_{t-1}}$$

for $t = 0, 1, \ldots, T$ and τ is the smoothing window size. We used $\tau = 7$ days.

Doubling time estimates the time it takes for the number of future cases to be twice what it is today, based on the current growth rate r as estimated by new cases within a window size τ, in our case over the past week. Given that the growth rate of today reflects the data of previous 7 days, doubling time relates past observations to a prediction for the future.

- Levitt Statistics: The Levitt statistic is an alternative way to represent change in cases (Levitt, Scaiewicz, and Zonta 2020).

$$H(t) = log(\frac{X_t}{X_{t-1}})$$

where X_t is cumulative cases up to day t. While the growth rate r represents the instantaneous rate of change in cases for each day, Levitt's statistic calculates the change in cases relative to cumulative cases, which shows how the overall number of cases is changing, or the first derivative of change. If $H(t) < 0$, cases are decreasing. If $H(t) > 0$, cases are increasing.

We compute the Levitt statistic for deaths only, following the suggestion of (Raman 2020). When the number of recorded incident cases is capped by testing capacity, an incidence series does not capture the growth of the epidemic accurately. Deaths are less prone to this type of confounding effect when testing capacity causes censoring of cases.

Editted by Emily Harville (mailto: harvilleconsult@gmail.com)

Reference

Abbott, Sam, Joel Hellewell, Robin N Thompson, Katharine Sherratt, Hamish P Gibbs, Nikos I Bosse, James D Munday, et al. 2020. "Estimating the Time-Varying Reproduction Number of Sars-Cov-2 Using National and Subnational Case Counts." *Wellcome Open Research* 5 (112). F1000 Research Limited: 112.

Anderson, Roy, Christl Donnelly, Deirdre Hollingsworth, Matt Keeling, Vegvari Carolin, Rebecca Baggaley, and Rosie Maddren. 2020. "Reproduction Number (R) and Growth Rate (R) of the Covid-19 Epidemic in the Uk: Methods of Estimation, Data Sources, Causes of Heterogeneity, and Use as a Guide in Policy Formulation." *Royal Society SET-C (Science in Emergencies Tasking – COVID)*. https://royalsociety.org/-/media/policy/projects/set-c/set-covid-19-R-estimates.pdf; The Royal Society.

Basu, Deepankar, Maxwell Salvatore, Debashree Ray, Michael Kleinsasser, Soumik Purkayastha, Rupam Bhattacharyya, and Bhramar Mukherjee. 2020. "A Comprehensive Public Health Evaluation of Lockdown as a Non-Pharmaceutical Intervention on Covid-19 Spread in India: National Trends Masking State Level Variations." *medRxiv*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/2020.05.25.20113043.

Cori, Anne, Neil M Ferguson, Christophe Fraser, and Simon Cauchemez. 2013. "A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics." *American Journal of Epidemiology* 178 (9). Oxford University Press: 1505–12.

Delamater, Paul L, Erica J Street, Timothy F Leslie, Y Tony Yang, and Kathryn H Jacobsen. 2019. "Complexity of the Basic Reproduction Number (R0)." *Emerging Infectious Diseases* 25 (1). https://wwwnc.cdc.gov/eid/article/25/1/17-1901_article; Centers for Disease Control; Prevention: 1.

Gostic, Katelyn M, Lauren McGough, Edward Baskerville, Sam Abbott, Keya Joshi, Christine Tedijanto, Rebecca Kahn, et al. 2020. "Practical Considerations for Measuring the Effective Reproductive Number, Rt." *medRxiv*. Cold Spring Harbor Laboratory Press. https://doi.org/10.1101/2020.06.18.20134858.

Levitt, Michael, Andrea Scaiewicz, and Francesco Zonta. 2020. "Predicting the Trajectory of Any Covid19 Epidemic from the Best Straight Line." *medRxiv*. Cold Spring Harbor Laboratory Press.

Obadia, Thomas, Romana Haneef, and Pierre-Yves Boëlle. 2012. "The R0 Package: A Toolbox to Estimate Reproduction Numbers for Epidemic Outbreaks." *BMC Medical Informatics and Decision Making* 12 (1). BioMed Central: 1–9.

Raman, Bhaskaran. 2020. "When Will the Covid-19 Pandemic End? Levitt's Metric on Indian Data."

Science, Government Office for, and UK Scientific Advisory Group for Emergencies. 2020. "The R Number and Growth Rate in the Uk." https://www.gov.uk/guidance/the-r-number-in-the-uk. 2020.

Wallinga, Jacco, and Peter Teunis. 2004. "Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures." *American Journal of Epidemiology* 160 (6). Oxford University Press: 509–16.