Oriane Kopp

Computational epidemiology: Project – Question 3

Description of the data

I choose the dataset SARS2003. This dataset contains information about the outbreak of the severe acute respiratory syndrome (SARS). This is an infectious disease that mainly affected Asia between 2002 and 2004. The disease also spread outside of Asia, and 29 countries/regions on five continents were impacted (Lam W et *al.*, 2003).
In the dataset from the EpiEstim package, there is two things: a vector containing the observation of the daily incidence of onset of symptoms in Hong Kong during the pandemic and another vector containing 25 discrete daily distribution of the serial interval of SARS, assuming it follows a Gamma distribution, with a mean of 8.4 day, a standard deviation of 3.8 days and a shift of 1 day (Anne Cori et *al.*, 2021).

Effective reproduction number estimation

The effective reproduction number can be seen as the "fitness" measurement of a disease. It represents the average number of individuals infected by one individual (secondary cases). Unlike the basic reproduction number which assumes all individuals are susceptible, the effective reproduction number considers more factors, like the fact that the susceptibility of the population varies.
To estimate R_t (effective reproduction number), I used the package EpiEstim (Anne Cori et *al.*, 2013). Note that R_t is a random variable, so the plot gives the posterior mean (best estimation of the effective reproduction number) and a 95% confidence interval (which means there is 95% chance that the true value is in the range).

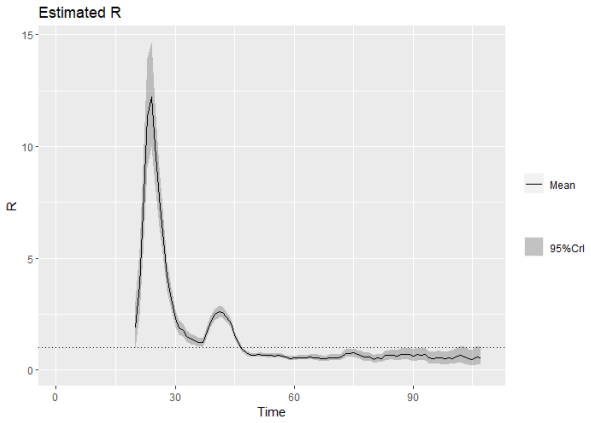

*Figure 1: Estimation of R for SARS 2003 using the **non-parametric** method*

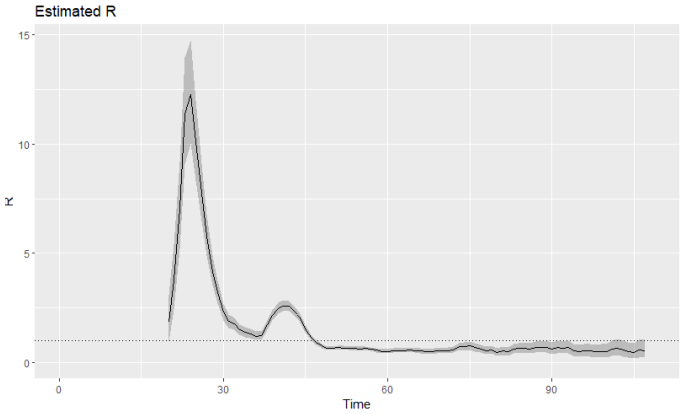

*Figure 2: Estimation of R for SARS 2003, using the **parametric** method with mean 8.4 and standard deviation 3.8.*

Description of the findings

To generate the plot (Figure 1), I choose the non-parametric method and used the window suggested in the EpiEstim documentation (7 days window, starting at week two).

On the plot, we can see that R_t increases fast, which is normal because in the beginning, everybody is susceptible. Then, it decreases and approaches 1 (because there is less and less susceptible individuals). However, there is another small peak, which suggested a big spreading event, like a cluster in a hospital or something like that. After that, R_t decreases again and stabilizes under one. This means the pandemic is going to an end, because an effective reproduction number lower than 1 means that one infected people will transmit it to less than one person.

Variation of assumptions

When I did the first plot, I made two assumptions: the window and the serial interval. I decide to not change the window but play with the serial interval. In the dataset, the distribution of the serial interval is provided in one of the vectors.

Estimation of the serial interval

The serial interval is the time it takes for a person to show symptoms after getting infected by someone else. In other words, it is the time between a person gets sick and the people that they infect show symptoms (Vink M. et *al.*, 2014).
As said in the data description, the distribution of the serial interval is calculated with the assumption that it follows a Gamma distribution with a mean of 8.4 days, a standard deviation of 3.8 days and a shift of 1 day. The mean is the number of days that are needed for an infected person to show symptoms on average. In the package EpiEstim, there is a function to estimate it called "discr_si". If we use this function, we can reproduce the distribution of the serial interval provided in the second vector. We can check that in R by doing `round(discr_si(0:24, mu = 8.4, sigma = 3.8), digits = 3) == SARS2003$si_distr`.

We can also directly input it by estimating R_t in a parametric way. In Figure 2, I used the parametric method, using the same mean and standard deviation that were used to calculate the distribution of the serial interval in the non-parametric method, that's why we get the same plot.
Then I decided to change the mean when estimating the effective reproduction number with the parametric method. I choose a value lower (2 days, Figure 3) and a value bigger (15 days, Figure 4) than the mean provided (8.4 days). I didn't change the standard deviation. We can see that with different serial interval, the behaviour of the outbreak could have been different. With a smaller mean (Figure 3), R_t would have reached a smaller peak than with the actual mean, and then varies between 2 and 0.5, that means that the disease could have not disappeared.
With a bigger mean (Figure 4), Rt would have reached a bigger peak, but then, the behaviour is the same as with the actual mean: a smaller peak and then a stabilisation under 1.

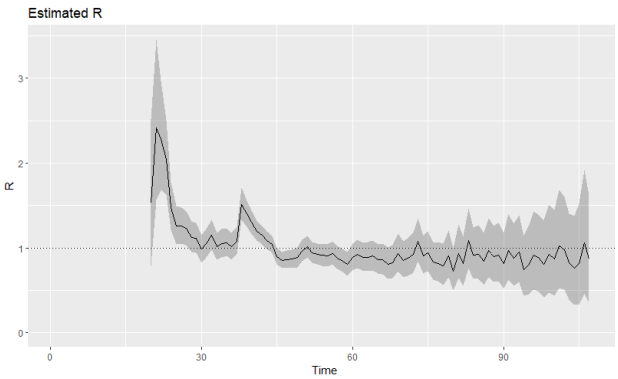

Figure 3: Estimation of R for SARS 2003, using the parametric method with mean 2 and standard deviation 3.8.

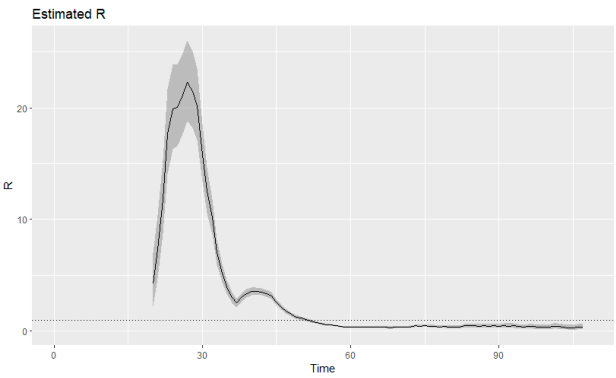

Figure 4: Estimation of R for SARS 2003, using the parametric method with mean 15 and standard deviation 3.8.

We can see that the serial interval as an impact on the estimation of R_t and so on the interpretation of the pandemic behaviour. If it takes less time for people to show symptoms (as in Figure 3), the diseases will infect fewer people in a short period of time (compared to Figure 1), but this also means that the pandemic will last longer because the herd immunity will take longer to develop (and people can lose their immunity and get sick again if it's too long). In contrast, with a longer time to show symptoms (as in Figure 4), more people will get infected in a short period of time (which is a challenge for the hospitals), but then, most of the population will be immunised and the diseases will (hopefully) disappear.

So, a wrong estimation of the serial interval can change the estimation of the effective reproduction number and maybe the measures taken during the pandemic.

References

Cori, Anne, Neil M. Ferguson, Christophe Fraser, and Simon Cauchemez. “A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics.” *American Journal of Epidemiology* 178, no. 9 (November 1, 2013): 1505–12. https://doi.org/10.1093/aje/kwt133.

Cori, Anne. 2021. “EpiEstim: a demonstration.” Accessed December 28, 2023. https://cran.r-project.org/web/packages/EpiEstim/vignettes/demo.html.

Cori, Anne. 2021. “EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves” (Version 2.2-4). R package. Maintained by Anne Cori. https://github.com/mrc-ide/EpiEstim.

Lam, Wk, Ns Zhong, and Wc Tan. “Overview on SARS in Asia and the World.” *Respirology* 8, no. s1 (2003): S2–5. https://doi.org/10.1046/j.1440-1843.2003.00516.x.

Vink, Margaretha Annelie, Martinus Christoffel Jozef Bootsma, and Jacco Wallinga. “Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis.” *American Journal of Epidemiology* 180, no. 9 (November 1, 2014): 865–75. https://doi.org/10.1093/aje/kwu209.