COVID Data Science Project

Group

Sep 25th 2020

We model the probability of exposure to a COVID-19 infected patient by point of interest.

We focus on the event of "Exposure" or E for brevity, which we characterize as the probability that 1 or more people that visit your point of interest have active COVID-19 cases. Importantly, we do not address the epidemiological risk of infection and we limit the model to exposure as described above. We note, on the other hand, that the model could be augmented to consider the user specific epidemiological risk but it requires substantial epidemiological knowledge and may be dependent on the visitor specific demographic and health history information.

In this simplified model we assume that the user visit the point of interest and interacts with other visitors at the same point of interest. Each visitor possibly comes from a different district. We stress the importance of a model that, while being local to the point of interest, can also capture mobility from neighboring districts. In particular each district $d \in D$ has a d-specific infection rate. The probability of the event E is computed as follows:

$$P(E|p,t) = 1 - \Pi_{d \in D} \left(1 - \frac{I_{dt}}{N_{dt}}\right)^{n_{pdt}}$$

where I_{dt} is the number of people with active COVID cases at time t in district. N_{dt} is the total population in district d. Finally n_{pdt} are the number of district d residents that are visiting the point of interest at time t. One way to look at the model is to think about the $P(E|p,d,t)$ as $1 - P(E^c|p,d,t)$ where $P(E^c|p,d,t)$ captures the probability that none of the other visitors we come in contact have active COVID-19 cases. Here we take the number of visitors coming from each district as given so n_{pdt} is known.
Consider the following example, avoiding the t subscript for simplicity. For a given point of interest p, if there are only two districts A and B and $n_{pA} = 3$ and $n_{pB} = 5$ visitors respectively and the infection rates in the two districts are $\frac{I_A}{N_A} = 0.08$ and $\frac{I_B}{N_B} = 0.1$ respectively the risk is given by:

$$P(E|p) = 1 - (1 - 0.8)^3 \cdot (1 - 0.10)^5 = 1 - 0.7787 \cdot 0.59049 = 0.5598$$

The first factor is the probability that none of the $n_{pA} = 3$ visitors from district A have COVID-19. It assumes that visitors to the point of interest are sampled independently from the same population that has infection rate $\frac{I_A}{N_A} = 0.08$. The second factor is the probability that none of the $n_{pB} = 5$ visitors from district B, which has an infection rate of $\frac{I_B}{N_B} = 0.10$ has COVID-19. Hence the probability of being exposed to COVID-19 is given by 0.5598.