TSSL Lab 3: Predicting Covid-19

Description of the SEIR model

Fredrik Lindsten

September 28, 2020

1 Problem formulation

In this lab we will use a state space model to describe the dynamics of SARS-CoV-2, the virus that causes covid-19. We will use a common epidemiological model referred to as a Susceptible–Exposed–Infectious–Recovered (SEIR) model. Similar models are regularly used to model disease dynamics by health authorities and epidemiologists around the world, not least for modeling covid-19.

1.1 A stochastic SEIR model

Let $\mathcal{P}$ denote the number of individuals in a closed population. We will use a (stochastic) nonlinear state space model to describe the disease dynamics within this population. Define the state vector as,

$$\alpha_t = \begin{bmatrix} s_t \\ e_t \\ i_t \\ r_t \end{bmatrix}. \tag{1}$$

That is, the state consists of the number of susceptible, exposed, infectious, and recovered individuals in the population at time t. We can then write the state dynamics as

$$s_{t+1} = s_t - \Delta e_t \tag{2}$$
$$e_{t+1} = e_t + \Delta e_t - \Delta i_t \tag{3}$$
$$i_{t+1} = i_t + \Delta i_t - \Delta r_t \tag{4}$$
$$r_{t+1} = r_t + \Delta r_t \tag{5}$$

where the "Δ-variables" denote the number of individuals that enter a certain state at time t. Specifically, Δe_t denotes the number of individuals that go from being susceptible to exposed, Δi_t the number of individuals that go from being exposed to infectious, and Δr_t the number of individuals that go from being infectious to recovered. Note that there is a "mass balance" in the model,

in the sense that the individuals that leave one state will enter the next state. The population is assumed to be closed and the total size of the population is constant, i.e., $\mathcal{P} = s_t + e_t + i_t + r_t$ for any t.

To complete the dynamical model, it remains to define suitable models for the Δ-variables. We start with the $e \to i$ and $i \to r$ transitions. For each *exposed* individual at time t, we assume that there is a probability $p_{e\to i}$ to transition to the *infectious* state. Similarly, for each *infectious* individual, the probability to transition to the *recovered* state is $p_{i\to r}$. Aggregating over the entire population we get that the number of transitions are binomially distributed,

$$\Delta i_t \sim \text{Bin}(e_t, p_{e\to i}), \tag{6}$$

$$\Delta r_t \sim \text{Bin}(i_t, p_{i\to r}). \tag{7}$$

Following FHM[1], we assume that $1/p_{e\to i} = 5.1$ and $1/p_{i\to r} = 5$ are known.

Next, we consider the transition $s \to e$, which is a bit more complicated due to the fact that the number of *exposed* individuals will depend both on the number of *susceptible* and the number of *infectious* individuals. Indeed, any encounter between a susceptible and an infectious person may lead to a newly exposed individual. To model this, we introduce the variable b_t as the average number of social interactions that a single person has during time step t (i.e., one day). This variable is allowed to depend on t to account for the fact that the expected number of interactions may vary with time due to, e.g., recommendations regarding social distancing. We also introduce the variable ρ which corresponds to the probability that a single encounter between a susceptible and an infectious individual results in a newly exposed case[2]. In can then be shown (see appendix A for a derivation) that a reasonable model for the number of $s \to e$ transitions at time t is given by,

$$\Delta e_t \sim \text{Bin}\left(s_t, 1 - e^{-\rho b_t \frac{i_t}{\mathcal{P}}}\right). \tag{8}$$

1.2 Stochastic transmission rate

The FHM model uses a deterministic transmission rate b_t. To get a more flexible and data-driven model we can instead view it as a stochastic process and introduce it in the state vector. Since $b_t > 0$ we let $b_t = \exp(z_t)$ where z_t is an unconstrained state variable. Different models for z_t could be considered (e.g., a generic AR(p) model), but for simplicity we assume that z_t follows a random walk (similarly to the local level model),

$$z_{t+1} = z_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2).$$

[1] https://www.folkhalsomyndigheten.se/publicerat-material/publikationsarkiv/e/estimates-of-the-peak-day-and-the-number-of-infected-individuals-during-the-covid-19-outbreak-in-the-stockholm-region-sweden-february–april-2020/

[2] In everyday language we would typically say that the healthy subject was infected as a result of the encounter, but in the SEIR framework this is referred to as being "exposed" in order to more clearly distinguish this from the "infectious" state. The person will later transition to the "infectious" state when s/he starts developing symptoms and therefore can carry the disease on.

We let the initial state be $z_1 \sim N(0, \sigma_z^2)$.

1.3 Observation model

The disease dynamics is observed indirectly through, e.g., PCR and antibody tests as well as the number of deaths, hospitalizations and intensive care unit (ICU) registrations per day. All of these time series are sources of information that could be used to estimate the underlying disease dynamics. Indeed, in principle we could have included multiple observation models in our SEIR framework to aggregate all these sources of information in a principled way. However, for simplicity we will only consider a single data source in this assignment, namely the number of newly registered patients at ICU each day. Furthermore, we will consider a simple model for this observations[3], namely that the number of ICU cases is directly proportional to the number of infectious individuals each day. This leads to the model

$$y_t \sim \text{Bin}(i_t, p_{i\to c}), \tag{9}$$

where $p_{i\to c}$ is the probability that an infectious individual enters ICU at any given day.

A Derivation of $p_{e\to i}$

Consider a single susceptible individual. We will derive the probability that this person becomes exposed at time t, and then obtain Δe_t by aggregating over the entire susceptible population.

The expected number of social interactions that this person has during time step t is b_t, and more specifically we assume that the number of interactions K' is Poisson distributed with rate b_t. Assuming that these interactions are uniform over the population, there is a probability of $i_t/\mathcal{P}$ that each interaction is with an infectious individual.

We note that the number of interactions that our susceptible subject has *with infectious individuals*, denoted by K, is obtained by first drawing the number of candidates K' from a Poisson distribution, but then only keeping each candidate with probability $i_t/\mathcal{P}$. This is a well-studied procedure referred to as *thinning* and it is well known that the resulting number of such interactions is given by

$$K \sim \text{Po}\Big(\underbrace{b_t \frac{i_t}{\mathcal{P}}}_{=:\lambda}\Big).$$

Let E denote the event that our subject is exposed, i.e., $E = 1$ means exposure and $E = 0$ means that the subject avoids exposure. We have assumed

[3]You should be aware of the fact that this model is a gross simplification of reality. More sophisticated observation models could have been used with the proposed SEIR model, but not to make it too complicated we will stick with this simple observation model as a proxy for the number of ICU cases.

that there is a probability ρ of exposure for each encounter with an infectious individual. Consequently, the probability of *avoiding exposure* is

$$\mathbb{P}(E = 0|K) = (1 - \rho)^K$$

The marginal probability of avoiding exposure is thus,

$$\begin{aligned}
\mathbb{P}(E = 0) &= \sum_{k=0}^{\infty} \mathbb{P}(E = 0|K)\mathbb{P}(K = k) \\
&= \sum_{k=0}^{\infty} (1 - \rho)^k \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \frac{e^{-\lambda}}{e^{-\lambda(1-\rho)}} \underbrace{\sum_{k=0}^{\infty} \frac{[(1 - \rho)\lambda]^k e^{-\lambda(1-\rho)}}{k!}}_{=1} \\
&= e^{-\lambda\rho},
\end{aligned}$$

where the sum equals one since the summand is the probability mass function of a Poisson distribution with rate $(1 - \rho)\lambda$ (and it thus sums to 1 when summing over the entire domain of the distribution). Consequently, the probability that our subject is exposed is,

$$\mathbb{P}(E = 1) = 1 - e^{-\lambda\rho}.$$

Finally, this holds for *any* susceptible individual, so if we now aggregate the total number of susceptible persons that become exposed during time step t, this gives our final model,

$$\Delta e_t \sim \text{Bin}(s_t, 1 - e^{-\lambda\rho}) = \text{Bin}(s_t, 1 - e^{-\rho b_t \frac{i_t}{P}}).$$