Math 519 Stochastic Process - Final Project

Damien Guého, Minyan Zhang

April 2020

When in presence of an unknown dynamical system, the very first step is generally to gather real data and perform some type of system identification to understand the situation and study the evolution of what is observable. In the case of COVID-19, as many models are available to describe a pandemic, a simple parameters fitting is enough to initiate the identification phase. We considered two models (very close in their overall structure) based on the famous SIR model to describe the evolution of the coronavirus: SIRD and SEIRD. In those models, **S** is the stock of susceptible people, **E** is the stock of exposed people, **I** is the stock of the infected people, **R** is the recovery count and **D** is the death count. Considering that the infected people will be quarantined immediately and that only exposed people (they carry the virus but haven't shown any sign of infection) will infect others, we build the SEIRD in this way,

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta ES}{N} \\
\frac{dE}{dt} &= \frac{\beta ES}{N} - \alpha E \\
\frac{dI}{dt} &= \alpha E - (\gamma + \mu) I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \mu I
\end{aligned} \tag{1}$$

with N the total population, β the infected rate per person per time, α the reciprocal of the expectation of the incubation days, γ the recovery rate and μ the death rate. The SIRD model is built with the same assumptions wit the only difference that we do not separate exposed people from the infected. Therefore, the SIRD model is

$$
\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta I S}{N} \\
\frac{dI}{dt} &= -(\gamma + \mu) I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \mu I
\end{aligned}
\tag{2}
$$

Parameters fitting is performed using an adapted gradient descent algorithm for the SEIRD model whereas the BFGS algorithm is used for the SIRD model. There is no specific reason for choosing one or the other.

Once a suitable model is found, we study the effect of error introduced in new-coming data in the propagation. As a matter of fact, counting and reporting the exact number of people in each category is subject to errors, approximations or even deliberate falsification. Considering the last available data as a random variable with a defined mean and covariance (we only consider the first two moments even if the method works also for higher order moments), the idea is to propagate that pdf through the non-linear dynamical model found before.

Finally, (if time allows us to do so), we want to study a similar case: introducing error in the model parameters. This can be relevant for several reasons: the parameters are very likely to be time-varying, more data points always help the analyst to define a more accurate model, changing in quarantine/travel/political strategies will undoubtedly affect the dynamics of the pandemic...etc. Using an extended Kalman Filter with a quadrature rule based on an unscented transformation, we are capable to update the mean and covariance of our model parameters following a maximum likelihood principle everytime a new data point is coming.