Question on Final Project

Group 1

5/7/2020

We are interested in modelling the excess death due to COVID-19

Until now we have been modelling

$$(1) \quad O_i \sim Poisson(\lambda_i)$$

where O_i is the number of deaths in week i. The i's are weeks in February, March and April.

We have obtained the weekly deaths for several years for the Netherlands. So we have a sample of the number of deaths O_i in week i. We estimated each λ_i, built a predictive posterior distribution for O_i, and a credible interval for O_i. Then we decide whether there is excess death or not by seeing if $O_{i,2020}$ is in the credible interval or not.

We can next think of building a hierarchical model, a multilevel model on this, adding countries and explanatory variables for λ_i

Then it came to my mind we could do something similar to the spatial epidemiological model we have seen and propose:

$$(2) \quad O_i \sim Poisson(E_i\theta_i)$$

where E_i would be the known expected number of deaths in week i that we could estimate as the historical average of death for week i and θ_i the risk in week i. In a way it is a decomposition of λ_i between its known expected value E_i and a perturbation θ_i. $\theta_{i,2020}$ larger than 1 would mean excess death, we can build a credible interval for $\theta_{i,2020}$ to see if it includes 1. Then we can build on a hierarchical model adding countries and explanatory variables for the risk θ_i (% of elderly people, time after lockdown). Doing it in similar way it is done in spatial epidemiology, if I remember well, for example:

$log(\theta_i) \sim Normal(\beta_o + \beta_1 X_1 + \beta_2 X_2, \sigma^2)$

It seems to me that the parameters in (2) are more interpretable. We can say from week i_1 to week i_2 there has been an increased mortality in the country ($\theta_i >$1, credible interval not containing 1). Large percentage of person at risk for covid and time to lockdown are relevant predictors of this increased mortality. I think that if we work directly with λ_i is going to be harder to compare them across weeks and countries because they are not scaled. Also we can think of modelling the dependence between weeks i as spatial epidemiology models dependence between contiguous spatial units.

However, I have doubts I might be trying to force model I am familiar with on a different setting. I see (2) as richer but that might just be an excess of complications.