SCoVMod – movement simulation

June 5, 2020

The purpose of `moveSim` is to provide population and movement input to the SCoVMod (Scotland's CoronaVirus Model). Input is in the form of a national population, drawn from census data, and a set of movements—where each individual travels in each time period—drawn from census flows data and modified by mobility data.

1 Data sources

Population statistics are drawn from National Records Scotland[1](NRS). From the current population estimates we draw the number of individuals whose primary residence is in each geographic unit (location)[2], and their age category. The smallest geographic unit provided publicly by NRS is the Census Output Area, of which there are approx. 46k and each with a household population of 100–500.

The data for assignment of individuals to work locations is drawn from NRS Census Flows data[3], Table WF01BUK, which provides origin/destination workplace data for the population from the 2011 census.

The data for adjusting daily movements for the period after lack-down is taken from Google's Community Mobility Reports[4], which reports the proportion of baseline mobility daily.

The population and census data was retrieved on 1st April 2020, and the mobility data is updated every few days—last updated 25th May 2020.

2 Population

The first output of `moveSim` is a national population of individuals, each with a unique ID, a starting location, and an age category. At this point we also decide if each individual goes to work and assign them a work location (if it is outside their home location).

2.1 Algorithm

The algorithm for assignment of individuals to home locations and age categories is straightforward. A unique ID for each individual in the national population is generated and home locations and age categories are assigned according to the distribution defined by the population estimate data.

An individual's workplace is assigned by distributing a proportion of the population of each location to each work location, weighted by the proportion of individuals from each home location in the census flows data who work in another location. For each origin o and destination d we assign a weight $w_{o,d}$ from the census flows data: $w_{o,d} = \frac{n_{o,d}}{t_o}$ where $n_{o,d}$ is the total number of

[1]NRS statistics: https://statistics.gov.scot/

[2]Geographic units defined by the ONS: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

[3]ONS WICID tool for exporting census flows data: https://wicid.ukdataservice.ac.uk/

[4]Google Community Mobility Reports: https://www.google.com/covid19/mobility/

people who move from o to d to work, and t_o is the total number who move from origin o to any location for work.

We then take the individuals of each home location if they are eligible to work (total n_o); in this case we assume all individuals of adult age 16–65. Each destination is assigned to $n_o \times w_{o,d}$ of these individuals. The individuals who remain have no assigned workplace—either they do not work, or they work within their home location.

2.2 Output format

The population generator of `moveSim` outputs two tables: the first with *PersonID, Origin, Age* as input to SCoVMod, and the second *PersonID, Origin, Age, Worker, Destination* as input to the `moveSim` movement simulator. *PersonID* is a unique integer identifier, *Origin* and *Destination* are string location identifiers, *Age* is a categorical identifier in {*Young, Adult, Elderly*}, and *Worker* a boolean indicating an eligible worker.

3 Movement

The movement simulator of `moveSim` produces input for SCoVMod, in the form of the set of individuals who move from each location, in each time step of the simulation. In this case, we use two time steps per day. In the first time step of the day, worker move to work, and in the second time step they return home. The output movements are a stochastic simulation with a Poisson distributed number of workers moving from each origin to each destination per day, distributed according to the census flows weighted population data, as described above. The volume of movement is reduced uniformly across the population according to the proportional decrease given by the mobility data.

We also introduce an optimisation to reduce the number of movements that need to be handled by SCoVMod. The number of movements are trimmed to one in five and the transmission rates are adjusted accordingly (indirectly by the fitting process).

3.1 Algorithm

For each day of the simulation we consider two time steps: a *day* step where individuals can move to their place of work, and a *night* step where those individuals move back to their home location.

In each day step, we take each destination location d. Let λ_d be the number of eligible workers who are expected to move to the destination location. For each day the sampled number who move s is drawn from a Poisson distribution:

$$s \sim Poisson(\lambda_d)$$

The sampled number of moves s is then scaled according to the per cent change in mobility m for the given day:

$$\lfloor s_m = s(1 + (m/100)) \rfloor$$

The number of moves is then trimmed 4 in 5 by drawing from a Binomial distribution:

$$s_{mt} \sim B(s_m, \frac{1}{5})$$

If the sampled number of workers s_{mt} is less than or equal to the number of workers who may normally move to destination d, then those who move are sampled randomly from those who may normally move. However, if s_{mt} is greater than the number of workers who may normally move to d, then the additional workers are drawn randomly from workers who have no assigned destination location.

The sampled PersonIDs for each destination location are then collected for output, for each day. For each night of the simulation, the worker who moved in the day step are moved back to their origin location.

3.2 Output format

Each time step of the simulation is output in JSON format for input to SCoVMod. Each time step contains the set of destination location IDs, each containing the set of PersonIDs who move to them.

4 Seeds

Infected seed individuals are distributed according to observed case data. As case data represents chiefly severely symptomatic or hospitalised individuals, then we assume those individuals are present in the population 5 days previous. The number of seeded individuals is a fitted paramteer in SCoVMod.