COVID-Predictor

Derivation and Extension of the Population Model

Robin Kramer, Tom Kuchler, Lorin Mühlebach,
Cedric Münger, María Ruiz Martnez, Yannick Schaffner,
Dominik Schulte, Kamil Sklodowski

April 5, 2020

1 Generating Households, Implementation Versus-Virus

Goal: get the age distribution of the household members $\Pr[\mathsf{A}_1, ..., \mathsf{A}_\mathsf{S}, \mathsf{T}]$ which largely determines the infection rate within household. Assumptions about the natural interaction [1] and occupation [2] of the household members has an important impact on how the disease spreads. The method to create natural households by Ajelli et Al. can be summarized as follows:

$$\Pr[\mathsf{T}] : \text{Prb of Houshold type} \tag{1}$$

$$\Pr[\mathsf{S}|\ \mathsf{T}] : \text{Prb of Houshold size given Type} \tag{2}$$

$$\Pr[\mathsf{C}|\ \mathsf{T}, \mathsf{S}] : \text{Prb of Houshold Head Age class,} \tag{3}$$

Is an intermediate step from data-availablility on age in houshold

$$\Pr[\mathsf{A}_1|\ \mathsf{C}] : \text{Prb of Houshold head age} \tag{4}$$

$$\Pr[\mathsf{A}_2, ..., \mathsf{A}_s|\ s = S] : \text{Prb of age of other houshold members} \tag{5}$$

The available data for Italy does however not have a one-to-one correspondence to Switzerland. Thanks to the swiss Federal Statistical Office, a wide variety of relevant data is available:

- Number of persons of age x per municipality
  `https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/stand-entwicklung/bevoelkerung.assetdetail.9635941.html`
- Number of households with n members $n \in \{1, 2, ..., 5, 6+\}$ per municipality
  `https://www.bfs.admin.ch/bfs/de/home/statistiken/bevoelkerung/stand-entwicklung/haushalte.assetdetail.9787080.html`

[1] e.g. family vs. roommates
[2] Influence of infections imported from outside

- Prevalence of different forms of households (Familienbericht G2.1)
- Household type per person and per household (Familienbericht G2.2)
- Parent-childern households, per age of youngest child (over/under 25) and number of parents (Familienbericht G2.3)
- Proportion of parents in same household according to age categories 0-3, 4-12,13-17,18-14 (Familienbericht G2.4)
- Distribution of number of children below 25 in household (Familienbericht G2.8)
- Distribution of age difference in couples living in same household (Familienbericht G3.2)

The "Familienbericht" is available under `https://www.bfs.admin.ch/bfs/de/home/statistiken/kataloge-datenbanken/publikationen.assetdetail.2347880.html` and the relevant datasets can be found in the "Anhang". From this, we propose the following algorithm for generating realistic population and household data:

1.1 Used Approximation

From the available data for Switzerland we decided to use for version 0.1

- the distribution of person/household per municipality,
- the age distribution per municipality.

Then we divided the citizens in the three age-classes 0-18, 18-65, 65+ and proceeded as follows:

1. generate per municipality the correct number of households w. occupants (1,2,3,4,5,6+) members)
2. assign first member according to local age distribution for 18+
3. assign rest of members according to the resulting marginal distribution after excluding the first member (6+ households get only 6 members assigned)

2 Proposed Algorithmic Refinements

In order to refine the approach developed and implemented during VersusVirus, we can propose some improvements, starting with increasing the variety of age categories. A reasonable granularity of the age categories, considering the different natural behaviour of each age group, different effects by quarantine measures, and available data is shown in table 1. Some possible categories are not proposed due to the low prevalence to simplify the household generation. The constraints within each household are as follows:

- Adults are in age categories C-F

Cat	Age	Likeli Occupation
A	0-5	Preschool
B	6-17	School/Apprenticeship
C	18-24	Working[3]/Student
D	25-49	Working
E	50-65	Working
F	65+	retired, high risk

TABLE 1: Proposed classification based on age.

cat	#Pers	Constraints
I	1	adult
II-C	2	Couple
II-P	2	Parent-Child
III-C	3	Couple. 1 Child
III-P	3	Single, 2 Children
IV-C	4	Couple, 2 Children
IV-A	4	All adults
V-C	5	Couple, 3 children
V-A	5	All adults
VI	6+	All adults

TABLE 2: Proposed simplified household categories

- Children are in age categories A-C
- Couples constitute of adults, differing at most by 12 years
- Parents and children in the same household have an age-difference between 18 and 40 years
- thus adults in household-types[4] x-C, x-P are in age categories C-E

2.1 Generating the Households

1. generate municipality households with correct distribution of member numbers categories
2. calculate the ratio $\frac{\text{II-P}}{\text{II-C}}$ and assign households w. 2 members accordingly.

   II-P single parents living with children: 4.3% of all households, around 61% of all single parents live with one child.

   II-C around 25% of all households constitute couples w/o kids
3. in the same fashion, calculate the split for the household types III-X, IV-X, V-X

[4]except II-C

2.2 Populating the Households

1. generate population according to municipal age distribution
2. assign[5] for each household of categories II-P, III-P, III-C, IV-P, V-P one adult in age group C-E
3. assign for III-C, IV-C households spouse meeting the age-constraints
4. assign correct number of children meeting the age constraints
5. Form valid couples for category $II - P$
6. Assign I to remaining adults
7. Assign households of type VI one adult each
8. Fill up the households of type VI as good as possible

2.3 Constraints Violations

The proposed algorithm does not guarantee that all constraints can be met. In particular, the population data is per municipality, but the data on different household forms are country-wide averages. Possible solutions for violations are:

- Try again with different RNG-seed
- Implement optimization algorithm
- take agent of closest fit instead
- Generate new random agent according to population distribution

[5] assignments happen uniform at random from the unassigned population meeting the criteria. This corresponds to the empirical conditional distribution