Weekly Report (4.25-5.1)*

Yanpei Cai†

May 1, 2025

During the past week, I have managed to do the sanity check for my codes for our smoking contagion model and to figure out how Adarsh performed the parameter estimation according to the reference of Christakis et al.. I also had some discussions with Michael, Adarsh, and Valerio for possible extensions.

Coding

Sanity Check for the Simulator

After testing a number of different sets of parameter input values and checking whether the simulated data make sense, I believe that the simulator is plausible.

BOLFI

Since our smoking contagion is characterised by a stochastic mechanism with multiple probability parameters, it is feasible to choose a Beta prior for all spontaneous parameters and interaction parameters. Choosing hyperparameter in a more reasonable way leaves as a task of doing more literature review, for now we start with testing different combinations of hyperparameter.

The most important step in BOLFI is to trace the likelihood function with the discrepancy between the observed and simulated data. Since our observed data is a table of fractions of the population in different smoking states over certain years, the most straightforward way to obtain the discrepancy is to compute the l^2-norm (although it might be too costly).

In fact, Michael suggested that at the beginning we choose some common summary statistics (mean, variance, median, MAD, etc.) to reduce the dimensionality of the data, and we ignore the temporal structure of the data. Later, we can introduce an inspection window for a length of just a few years and we keep rotating it to achieve the sort of temporal effects.

Maths

Parameter estimation from Christakis et al.

Following discussion with Adarsh, (from Christakis et al.) **g** denotes the relative increase in the probability of quitting smoking due to social ties over t years, relative to the spontaneous

†Artificial Intelligence and its Applications Institute, School of Informatics, The University of Edinburgh
Email Address: yanpei.cai@ed.ac.uk

*This is a summary of the works was done in the past week.

baseline. I am still very confused about the whole setting of doing parameter estimation in this way, so I suppose it would be very helpful to leave this as a question for this week and have a discussion about this during this week's meeting.

For Adarsh's derivation, please check the last two pages of this report.

New Ideas

Multiple Social Ties, Multiple Types of Contagion, Subgraphs and Multilayer Networks

Obviously, there can be different types of social ties among one's connections, such as parents, siblings, spouse, friends, colleagues, neighbours, etc. Instead of taking the average effects among social ties (doing this may cause the initiation of certain types of contagion that should not occur in general), we can split the contagion according to different types of social ties, in both the contagion dynamics and the network structure.

For network structure, one way can be that we first construct multiple subgraphs (subnetworks) for different types of social ties and then compress them into a whole network structure for the population. Another way can be that we simply split the whole network into multiple layers by different types of social ties. Both ways allow us to set up more specific contagion dynamics for different types of social ties.

For contagion dynamics, there has been some immature thinking in my mind, but I would like to discuss them during this week's meeting and then construct them in a more rigorous way later this week.

Derivation of Sup. section Eq. 18

Adarsh Prabhakaran

April 2025

Multi-year spontaneous quitting (Eqs 13–17)

Let $\delta_{S\to Q}$ denote the annual spontaneous cessation rate and t the interval length in years. After t years, the fraction of original smokers who remain is $(1-\delta_{S\to Q})^t$, so the multi-year spontaneous quit probability is

$$P_{\text{multi}}^{\text{spont}} = 1 - (1 - \delta_{S\to Q})^t.$$

To recover the equivalent annual cessation probability from any observed multi-year probability g, one uses the inversion

$$p = 1 - (1 - g)^{1/t},$$

as in Eq 17.

Multi-year interaction probability

From Christakis et al., g denotes the relative increase in the quitting probability over t years due to social ties, relative to the spontaneous baseline. Hence

$$g = \frac{P_{\text{multi}}^{\text{int}}}{P_{\text{multi}}^{\text{spont}}}$$

$$\implies \quad P_{\text{multi}}^{\text{int}} = g\, P_{\text{multi}}^{\text{spont}}$$

$$\implies \quad P_{\text{multi}}^{\text{int}} = g\big[1 - (1 - \delta_{S\to Q})^t\big] \quad \text{(using Eq. 17)}$$

This intermediate step was omitted in the original write-up.

Back to an annual interaction rate (Eq 18)

We an now invert $P_{\text{multi}}^{\text{int}}$ to obtain the annual interaction-driven quitting rate $\beta_{S,Q\to Q,Q}$:

$$\beta_{S,Q\to Q,Q} = 1 - \big[1 - P_{\text{multi}}^{\text{int}}\big]^{1/t} = 1 - \Big[1 - g\big(1 - (1 - \delta_{S\to Q})^t\big)\Big]^{1/t}.$$

Similarly, for initiation (non-smoker to smoker), with annual spontaneous initiation rate $\delta_{N \to S}$ and relative increase b,

$$\beta_{N,S \to S,S} = 1 - \Big[1 - b\big(1 - (1 - \delta_{N \to S})^t\big)\Big]^{1/t}.$$