Weekly Report - Presentation

Yanpei Cai

Artificial Intelligence and its Applications Institute
School of Informatics, The University of Edinburgh

30 May 2025

Table of Contents

1. A Brief Review
2. Results
3. Discussions
4. Possible Direction for Further Improvements

THE UNIVERSITY of EDINBURGH
informatics

Smoking Contagion Model

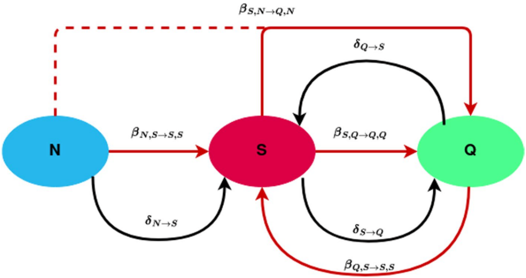


Fig. 1 The figure shows the schematic representation of the state change processes involved in the ABM. The interaction parameters are represented by the red arrows, while the black arrows show the spontaneous terms in the schematic. All three state-change processes are shown in the figure. First, an N-agent can initiate smoking spontaneously ($\delta_{N\to S}$) or due to the interaction with an S-agent ($\beta_{N,S\to S,S}$). Similarly, an S-agent can quit spontaneously ($\delta_{S\to Q}$) or due to interaction with other non-smoker agents (Q-agent: $\beta_{S,Q\to Q,Q}$ or N-agent: $\beta_{S,N\to Q,N}$). Like the other processes, Q-agents relapse into smoking spontaneously ($\delta_{Q\to S}$) or due to interaction ($\beta_{Q,S\to S,S}$) with an S-agent

THE UNIVERSITY of EDINBURGH
informatics

UK Smoking Trend Data (1974-2023)

Index	% NEVERSMOKER	% SMOKER	% QUITTER
1974	37.4	45.6	17
1976	39.1	41.8	19.1
1978	39.9	40.2	19.9
1980	40.1	39.4	20.5
1982	42.2	35.3	22.5
1984	43	34	23
1986	42.9	32.7	24.4
1988	43.7	31.6	24.7
1990	45.2	30	24.8
1992	45.6	28.4	26
1994	47.7	26.8	25.5
1996	46.9	28	25.1
1998	47.7	27.1	25.2
2000	49.6	27	23.4
2001	49.5	26.9	23.6
2002	50.2	25.9	23.9
2003	50.5	26	23.5
2004	51.8	24.6	23.6
2005	52.5	23.9	23.6
2006	54.3	22	23.7
2007	54.8	20.9	24.3
2008	53.4	21.1	25.5
2009	53.7	21	25.3
2010	54.9	20.3	24.8
2011	55.8	19.8	24.4
2012	59	20.4	20.6
2013	58.1	19.2	22.7
2014	58.6	18.8	22.6
2015	58.8	17.8	23.4
2016	59.2	16.1	24.7
2017	58.6	16.8	24.6
2018	57.2	16.6	26.2
2019	58	15.8	26.2
2020	59.8	14.5	25.7
2021	61.7	12.7	25.6
2022	63.4	11.2	25.4
2023	63.8	10.5	25.7

Bayesian Optimization for Likelihood-Free Inference (BOLFI)

- Bayesian Optimisation (BO):
  - BOLFI uses Bayesian optimisation to efficiently explore the parameter space, focusing on regions where the discrepancy between observed and simulated data is minimised.
  - It builds a probabilistic surrogate model (often a Gaussian Process) to approximate the relationship between parameters and the discrepancy.
- Likelihood-Free Inference (LFI):
  - Instead of computing the likelihood, BOLFI relies on forward simulations from the model to compare simulated and observed data using a distance metric.

Bayesian Optimization for Likelihood-Free Inference (BOLFI)

- Active Learning:
  - BOLFI sequentially selects new parameter points to evaluate, balancing exploration (uncertain regions) and exploitation (promising regions).
- Posterior Estimation:
  - After optimisation, BOLFI provides an approximation of the posterior distribution over parameters, conditioned on observed data.

Simulations

Simulations were conducted in the following settings:

- 1000 agents in total, with the initial configuration of agents in each state as same as the data in 1974.
- Erdös-Rényi graph: $n = 1000$, $p = 0.003$.
- All simulated data were generated in the same format as the observed data.

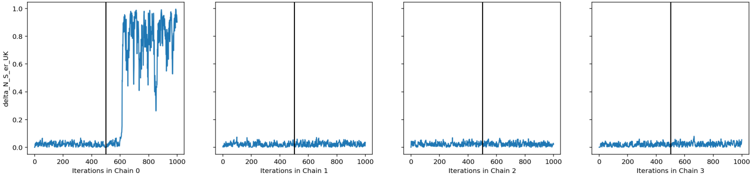
BOLFI Settings

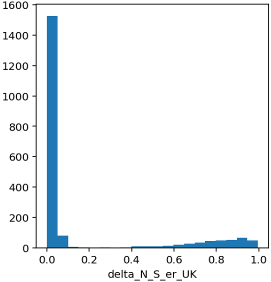
BOLFIs were conducted in the following settings:

- $U(0, 1)$ priors for all parameters need to be calibrated.
- All other parameters stuck to Adarsh's best-fitted values.
- Summary statistics were chosen to focus solely on % SMOKER and % QUITTER.
- Discrepancies were measured by Euclidean distance.

Single Parameter Calibration: $\delta_{N\to S}$

```
Progress [==================================================] 100.0% Complete
INFO:elfi.methods.posteriors:Using optimized minimum value (4.4263) of the GP discrepancy mean
function as a threshold
Parameter order: ['delta_N_S_er_UK']
Optimal parameters: [0.]
Lowest discrepancy: 4.076217291830845
```

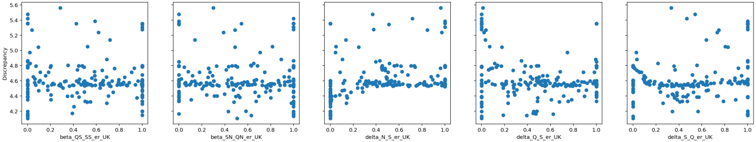




Multiple Parameter Calibration: All 5 Parameters

```
Progress [==================================================] 100.0% Complete
INFO:elfi.methods.posteriors:Using optimized minimum value (4.0339) of the GP discrepancy mean
function as a threshold
Parameter order: ['beta_QS_SS_er_UK', 'beta_SN_QN_er_UK',
'delta_N_S_er_UK', 'delta_Q_S_er_UK', 'delta_S_Q_er_UK']
Optimal parameters: [0.         0.50680321 0.         0.         0.        ]
Lowest discrepancy: 4.104976519451192

ValueError: NUTS: Cannot find acceptable stepsize starting from point
[0.         0.50680321 0.         0.         0.        ]. All trials ended in region with 0 probability.
```



Discussions

It can be observed that,

- the minimum discrepancy can still be very large;
- the point estimate of the best-fitted parameter is not unique (instead of the optimal point estimate, one should focus on is the posterior distribution;
- the posterior distribution sometimes makes no sense at all;
- BOLFI sometimes does not work well.

Possible Direction for Further Improvements

- Perhaps it would benefit from running BOLFI multiple times to refine the knowledge of the posterior.
- Using rolling summary statistics to improve the capture of how the smoking trend evolves over time.
- Introducing vital dynamics to ensure the model naturally corresponds to the smoking trend data.

The End

Thank you very much for your time!