PHD PROJECT IN COMPUTATIONAL SCIENCE: SIMULATION-BASED INFERENCE IN EPIDEMICS

STEFAN ENGBLOM

Effective summary

Public health agencies can be expected to be increasingly dependent on sophisticated sets of data in order to accurately plan for disease scenarios, to design mitigation- and suppression programs, and to assess risks. The quality of such strategies ultimately depend on the development of a range of tools to analyze this data.

In this project we are particularly interested in designing flexible data-driven computational models which are effective in producing not only forecasts, but also accurate "nowcasts" in order to provide situation awareness during decision making. Given the specifics of the application, there is a strong component relating to *trustworthiness* and *robustness*.

The project builds on a common foundation in computational epidemiology and Bayesian inference. On this basis, there are different possible research directions, offering different ideas for a PhD project:

Bayesian Epidemics and State Estimation: The foundational part concerns the design and implementation of flexible stochastic epidemiological models, with support for Bayesian inference and approximate likelihoods. An important objective is to combine computational models with state estimation techniques, allowing different data sources and modeling resolutions to be integrated within a common framework.

Diagnostics for Simulation-Based Inference: An follow-up opening towards the statistical foundations of trustworthy model-based analysis. The focus is on developing methods to assess the consistency of posterior models, in particular under model misspecification and approximate likelihoods. Possible approaches include parametric bootstrapping and data consistency checks, comparing model simulations with observations to detect discrepancies and assess the reliability of resulting inferences.

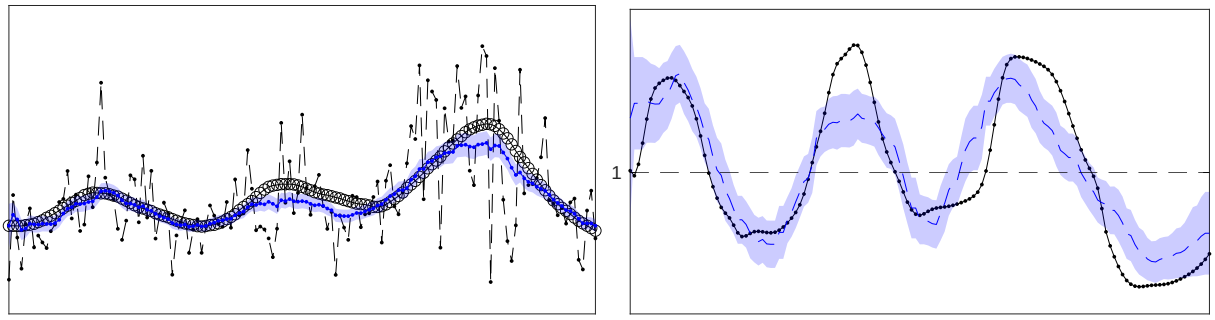
Neural Network Priors: A distinct opening towards machine learning and Bayesian methodology. The goal is to develop a framework in which neural networks generate effective Bayesian priors for epidemiological models. The calibration will take place under maximum entropy constraints, with additional terms accounting for the stability and identifiability of the underlying inference problem.

Epidemic Games and Value of Information: Another research direction towards computational game theory and decision analysis. Here the focus is on modeling

interactions between public health authorities and individuals or other decision-making agents. We are particularly interested in how information affects intervention strategies, and how its value can be quantified through the decisions it enables and the outcomes that follow. This includes investigating adaptive intervention systems in which decisions also influence the information available for subsequent decisions.

The foundational modeling and inference framework in the first part is the starting point for the project. The remaining directions represent different research openings rather than a fixed set of tasks that a candidate is expected to pursue. Depending on candidate's background and interests, a PhD project can focus on one or combine aspects of these directions. The overall aim is to advance computational and statistical methods for reliable model-based analysis and decision support in epidemiology.

Perspectives



Left: noisy measurements of virus concentration in wastewater together with the underlying smooth synthetic truth (circles). *Right:* reproduction number estimates with synthetic truth in black. Blue shade is the estimated uncertainty.

Mathematical modeling of the spread of disease is a fairly mature field. However, the increasing volumes and the improved granularity of data puts different perspectives on modeling in epidemics. Computational models have a tendency to become more complex, more detailed, and contain more information. But without sufficient data to drive such models, the model precision may well be imaginary.

In this project we are interested to partially counter these issues by designing epidemiological models driven by data. Given the constraints of most Public Health related applications, black-box solutions are not acceptable. For any estimates obtained using computational models, there is a substantial need for independent statistical validation. Beyond model construction and inference, there is also a need to understand how information from such models can support decision-making, and how interventions in turn affect the information available for subsequent decisions.

Expectations

The PhD project will involve research on specific setups and applications, using both real and synthetic data. The work will include implementation and evaluation of computational methods, as well as collaboration in the preparation of scientific publications.

A suitable background includes one or more of

- Computational Science or Data Science
- Compute-intensive Statistics or Applied Mathematics

Proficiency in a relevant programming language, e.g., Matlab, Python, R, C/C++ is expected and practical or research experience with machine learning techniques in general will be highly beneficial.

The working language of the research environment is English, and applicants from all countries are welcome. Good written and spoken English is therefore expected.

Welcome to contact me for informal enquiries and discussions.

(S. Engblom) Division of Scientific Computing, Department of Information Technology, Uppsala University, Sweden.

Email address: stefane@it.uu.se, https://stefanengblom.github.io