Simulation-Based Inference for Global Health Decisions

Christian Schroeder de Witt[1*] Bradley Gram-Hansen[1] Nantas Nardelli[1] Andrew Gambardella[1] Rob Zinkov[1]
Puneet Dokania[1] N. Siddharth[1] Ana Belen Espinosa-Gonzalez[2] Ara Darzi[2] Philip Torr[1] Atılım Güneş Baydin[1]
[1]Department of Engineering Science, University of Oxford [2]Department of Surgery and Cancer, Imperial College London *cs@robots.ox.ac.uk

Simulations in Global Health

- Machine learning (ML) has a growing role global health, enabling improvements in health service access and efficiency.
- Design and implementation of public health interventions are assisted by epidemiological modeling through computer simulation.
- Multi-modal intervention portfolios for endemic diseases have been studied, for example in malaria.
- *COVID-19* poses similar challenges in the context of a pandemic, for example modeling social distancing measures.

Can we use novel ML methods from simulation-based inference and control to push existing boundaries?

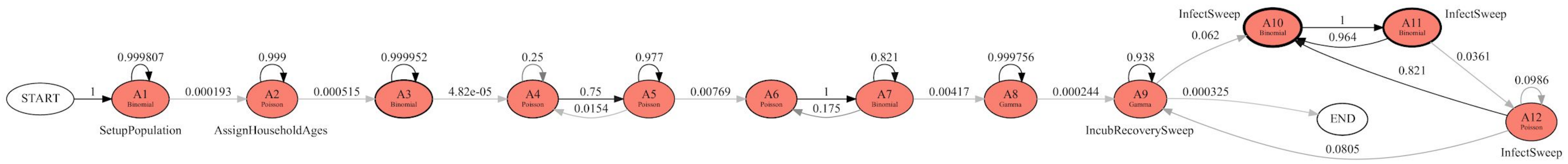
Modelling Challenges

Stochastic individual-based models are significantly more complex than traditional **compartmental models**. This leads to new challenges with respect to:

- **Model calibration**
  The extent to which a simulator can reliably inform real-world prediction and planning is bounded by both model discrepancy and how well the model has been calibrated to empirical data.

- **Optimising decision-making**
  Identifying optimal multimodal intervention strategies and corresponding risks and uncertainties requires searching through potentially vast parameter spaces, which, due to the computational cost of running large simulators (e.g., in some epidemiological studies), usually cannot be exhaustively evaluated.

CovidSim: A Case Study

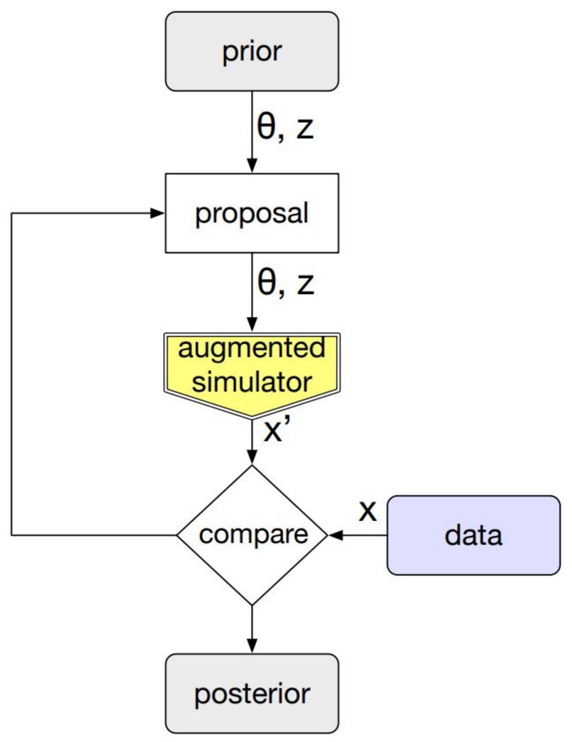


Latent probabilistic structure uncovered using PyProb* probabilistic programming library from an Imperial College London CovidSim simulator† run on Malta, demonstrating the first step in working with this simulator as a probabilistic program. Uniform distributions are omitted for simplicity. *https://github.com/pyprob/pyprob †https://github.com/mrc-ide/covid-sim/

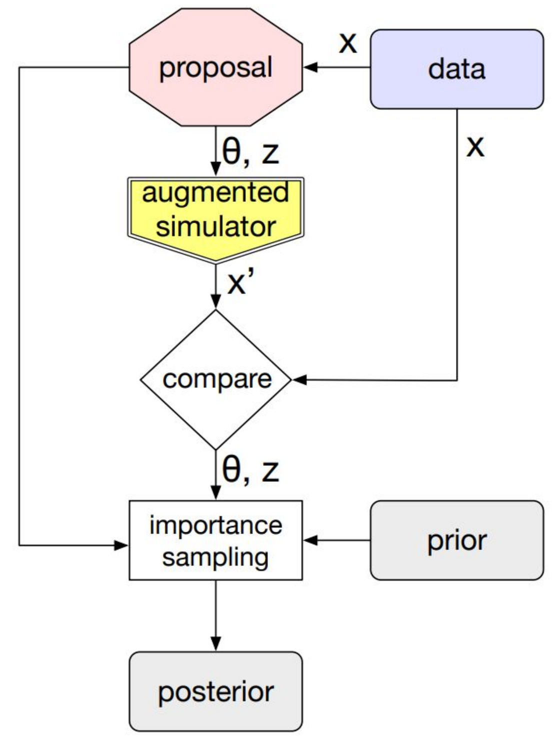
Simulation-Based Inference and Control

- Recent advances in machine learning have led to a new family of promising approaches to **simulation-based inference** (SBI).
- Probabilistic programming allows one to express probability models using computer code and perform statistical inference over the inputs and latent variables of the program, conditioned on data observations (or constraints).
- This is achieved by using special-purpose probabilistic programming languages (PPLs), which augment a host language with features to express probabilities and Bayesian conditioning.
- Recent work made it possible to use pre-existing stochastic simulators as probabilistic programs, with minimal code modification to capture and redirect random number draws scaling up to very large simulators, and particularly relevant to this work, with application to individual-based epidemiology simulators.

Probabilistic Programming with Monte Carlo sampling



Probabilistic Programming with Inference Compilation



Figures by Cranmer et al. 2020

Traditional Approaches

- Inference in individual-based simulators is usually **doubly intractable,** as both simulator likelihood and evidence cannot be evaluated efficiently.
- **Approximate Bayesian computation** (ABC) is likelihood-free but suffers from exponential scaling of inference with data dimension
  → requires domain experts to define low-dimensional summary statistics

Approximate Bayesian Computation with learned summary statistics

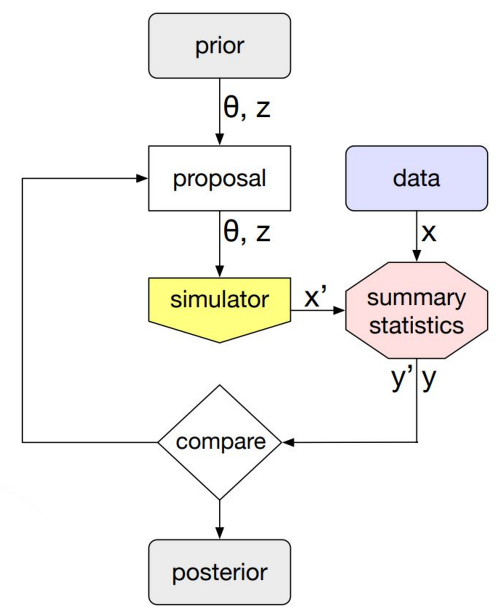


Figure by Cranmer et al. 2020

Conclusions and Outlook

- SBI can benefit from a new set of tools: **automated amortization** by surrogate methods; **source-to-source transformations** to make existing simulators differentiable, enabling efficient gradient-based optimisation and inference (e.g., HMC); **standardized inference interfaces** to facilitate use by non-SBI experts.
- We expect the mentioned techniques to play a role in dealing with communicable and non-communicable diseases, which pose a significant burden for health systems worldwide.

Key References

- Baydin, Heinrich, Bhimji, Shao, Naderiparizi, Munk, Liu, Gram-Hansen, Louppe, Meadows, Torr, Lee, Prabhat, Cranmer, and Wood. 2020. Efficient probabilistic inference in the quest for physics beyond the standard model. In Advances in Neural Information Processing Systems 33 (NeurIPS)
- Cranmer, Brehmer and Louppe. 2020. The frontier of simulation-based inference. Proceedings of the National Academy of Sciences.
- Gram-Hansen, Schroeder de Witt, Torr, Teh,Rainforth, and Baydin. 2019. “Hijacking Malaria Simulators with Probabilistic Programming.” In ICML Workshop on AI for Social Good, Thirty-Sixth International Conference on Machine Learning (ICML 2019), Long Beach, CA, US.
- Gram-Hansen, Schroeder de Witt, Zinkov, Naderiparizi, Munk, Wood, Torr, Teh, Baydin, and Rainforth. 2019. “Efficient Bayesian Inference in Nested Simulators.” In Advances in Approximate Bayesian Inference Symposium (AABI 2019), Vancouver, BC, CAN.

ICML 2020 Workshop on Machine Learning for Global Health

Acknowledgments: CS is funded by Armasuisse, BGH is funded by EPSRC CDT in Autonomous Intelligent Machines and Systems. Phil Torr is supported by the ERC grant ERC-2012-AdG 321162-HELIOS, EPSRC grant Seebibyte EP/M013774/1 and EPSRC/MURI grant EP/N019474/1. We would also like to acknowledge the Royal Academy of Engineering and FiveAI.