Bayesian Inference for Intractable Infectious Disease Models

Michael Gutmann

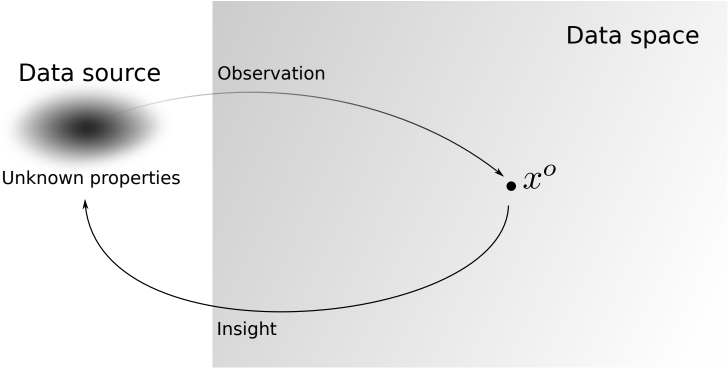
http://homepages.inf.ed.ac.uk/mgutmann

Institute for Adaptive and Neural Computation
School of Informatics, University of Edinburgh

29 November 2017

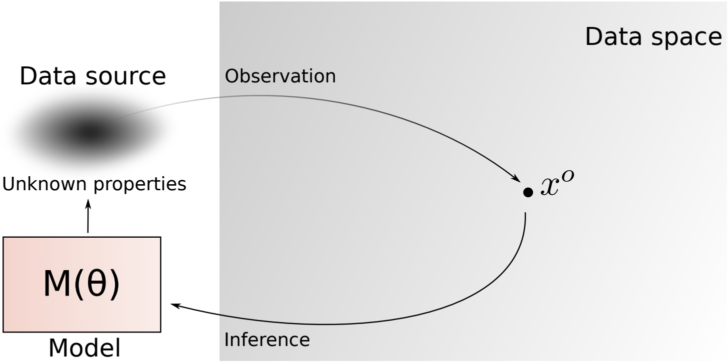
Learning from data

- Goal: Using observed data $\boldsymbol{x}^o$, learn about their source
- Enables decision making, predictions, ...



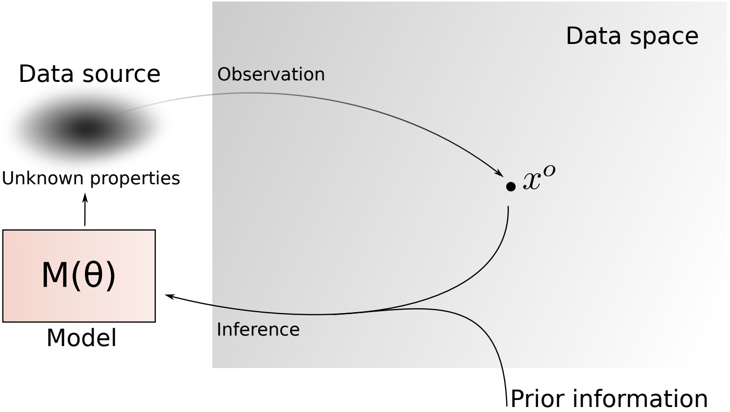
General approach

- Set up a model with potential properties $\boldsymbol{\theta}$ (parameters)
- See which $\boldsymbol{\theta}$ are in line with the observed data $\boldsymbol{x}^o$



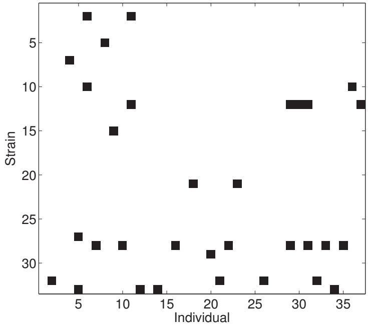
General approach

- Set up a model with potential properties $\boldsymbol{\theta}$ (parameters)
- See which $\boldsymbol{\theta}$ are in line with the observed data $\boldsymbol{x}^o$



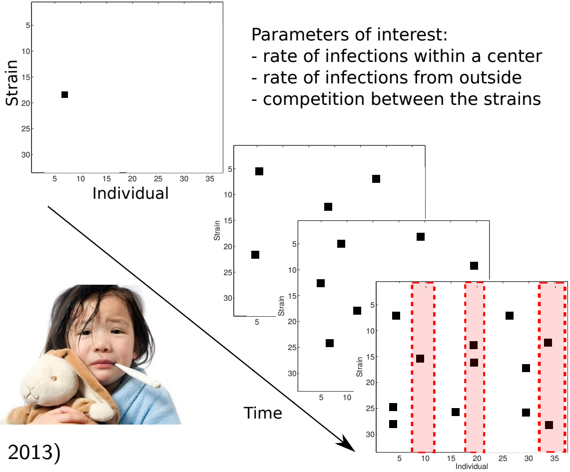
Example: Bacterial infections in child care centres

- Data: Colonization states of sampled attendees in 29 child day care centres (DCCs).
- Each square indicates a child colonized with a strain of the bacterium *Streptococcus pneumoniae*.



Example: Bacterial infections in child care centres

- Model: latent continuous-time Markov chain for the transmission dynamics in a DCC and an observation model
- What can we say about the parameters of interest?

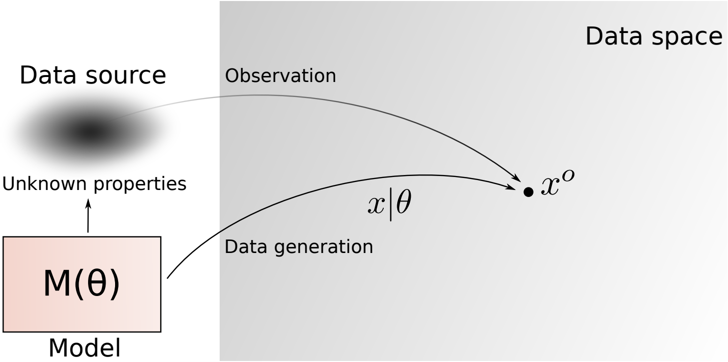


Parameters of interest:
- rate of infections within a center
- rate of infections from outside
- competition between the strains

(Numminen et al, 2013)

The likelihood function

- Measures agreement between $\boldsymbol{\theta}$ and the observed data $\boldsymbol{x}^o$
- Probability to generate data like $\boldsymbol{x}^o$ if hypothesis $\boldsymbol{\theta}$ holds



Difficulty

- For the child care centre model and other (individual-based) models: likelihood function is too expensive to compute.
- General computer science/statistics research question:

  How to efficiently perform (Bayesian) inference when
  - the likelihood function cannot be evaluated
  - but sampling from the model is possible

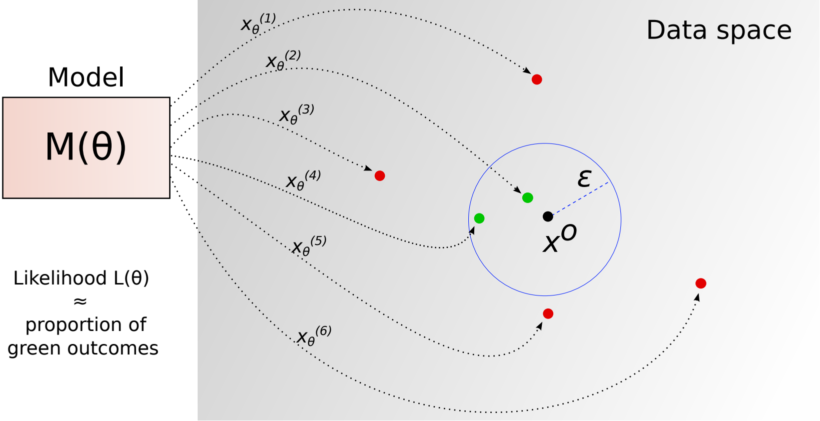
Difficulty

- For the child care centre model and other (individual-based) models: likelihood function is too expensive to compute.
- General computer science/statistics research question:

  How to efficiently perform (Bayesian) inference when
  - the likelihood function cannot be evaluated
  - but sampling from the model is possible
- Research area called "likelihood-free inference" or "approximate Bayesian computation"
  (recent review article: Lintusaari et al, Systematic Biology, 2017)

Simple approach: approximate by counting

Likelihood: Probability to generate data like $\boldsymbol{x}^o$ for parameter value $\boldsymbol{\theta}$

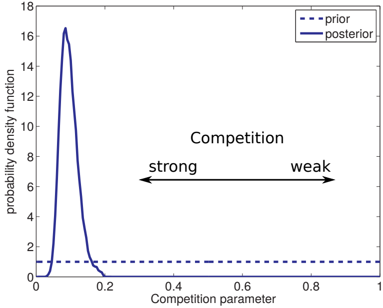


Example: Bacterial infections in child care centers

- Data: *Streptococcus pneumoniae* colonization for 29 centers
- Inference with a smarter version of the counting-based approach (population Monte Carlo ABC)
- Reveals strong competition between different bacterial strains

Expensive:

- 4.5 days on a cluster with 200 cores
- More than one million simulated data sets



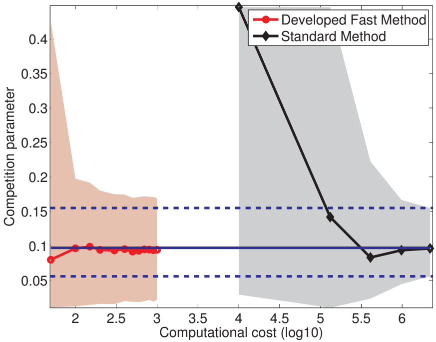
Fast Bayesian inference using machine learning

- We developed a fast inference algorithm using machine learning (Bayesian optimisation).
- Roughly equal results using 1000 times fewer simulations.

4.5 days with 200 cores
↓
90 minutes with seven cores

Posterior means: solid lines,

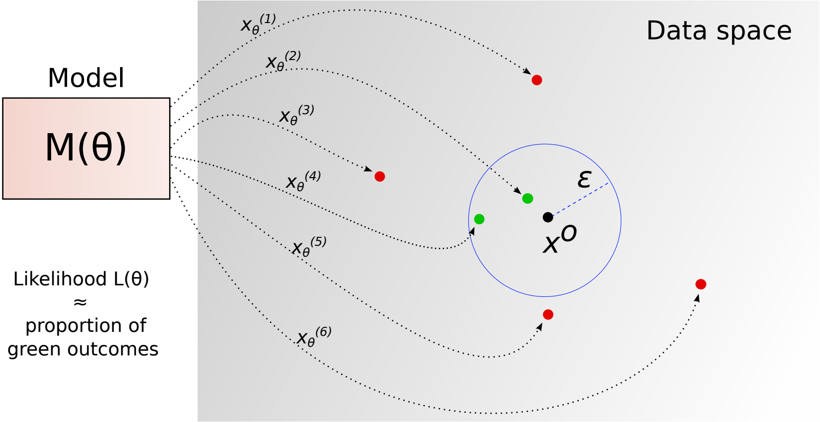
credibility intervals: shaded areas or dashed lines.



(Gutmann and Corander, JMLR, 2016)

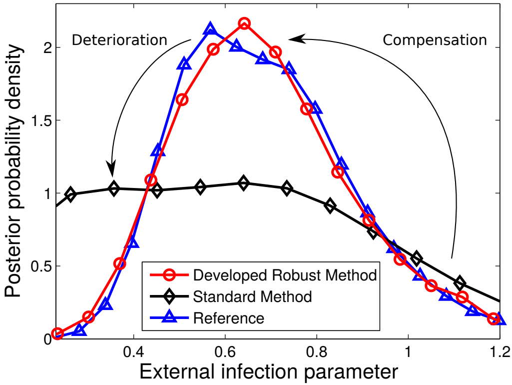
Standard approach: approximate by counting

Likelihood: Probability to generate data like $\boldsymbol{x}^o$ for parameter value $\boldsymbol{\theta}$



Robust Bayesian inference using machine learning

- Traditionally, expert knowledge is used to judge whether the simulated and observed data are close
- But experts make mistakes too
- Robustify using machine learning (Gutmann et al, 2014, 2017)



Conclusions

- Inference for models where the likelihood is intractable but sampling is possible (likelihood-free inference)
- Relevant for complex infectious disease models with many unobserved variables
- Machine learning to accelerate and robustify the inference

Conclusions

- Inference for models where the likelihood is intractable but sampling is possible (likelihood-free inference)
- Relevant for complex infectious disease models with many unobserved variables
- Machine learning to accelerate and robustify the inference

Further information:

- My homepage: `http://homepages.inf.ed.ac.uk/mgutmann`
- Review paper: Lintusaari et al, Systematic Biology, 2017
- Software: ELFI – Engine for Likelihood-Free Inference `http://elfi.readthedocs.io`