Likelihood-Free Inference

— An Introduction to My Research —

Michael Gutmann

https://sites.google.com/site/michaelgutmann

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

April 2017

Program

Overview of my research interests

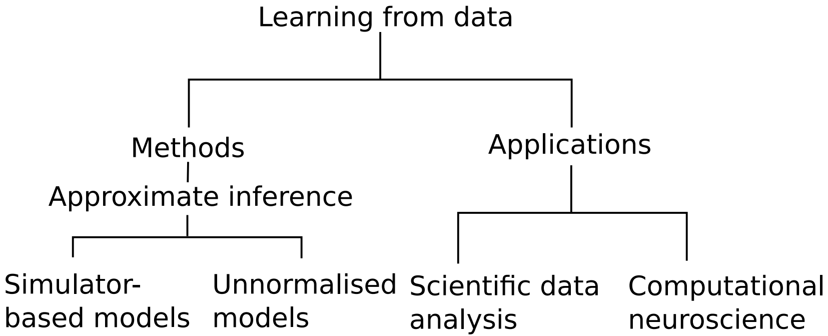
Efficient inference for simulator-based models

Program

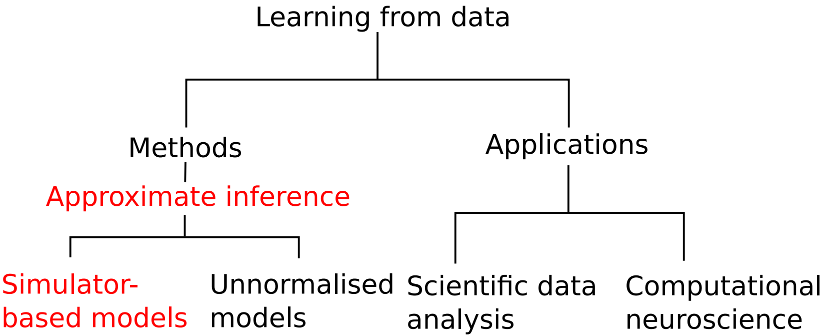
Overview of my research interests

Efficient inference for simulator-based models

My research interests in brief

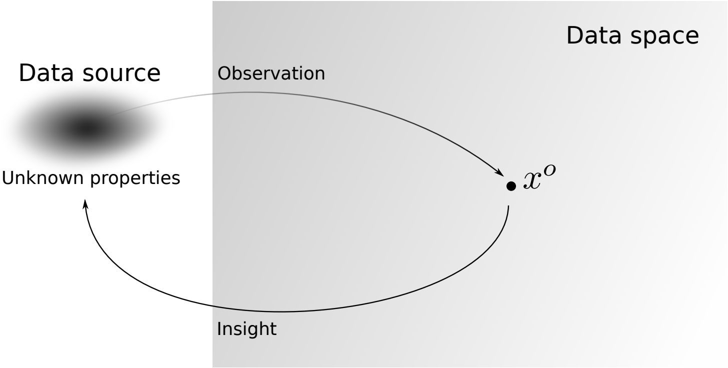


My research interests in brief



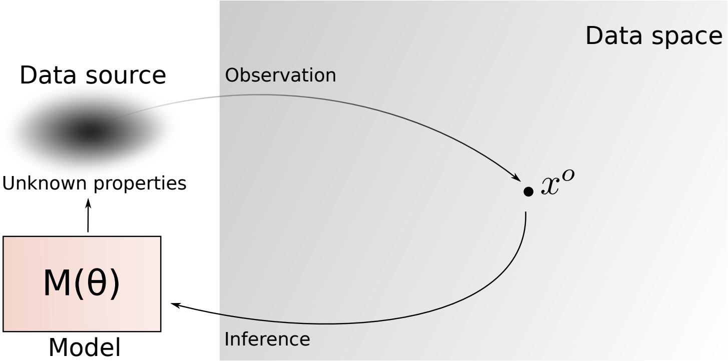
Learning from data

- Goal: Using observed data x^o, learn about their source
- Enables decision making, predictions, ...



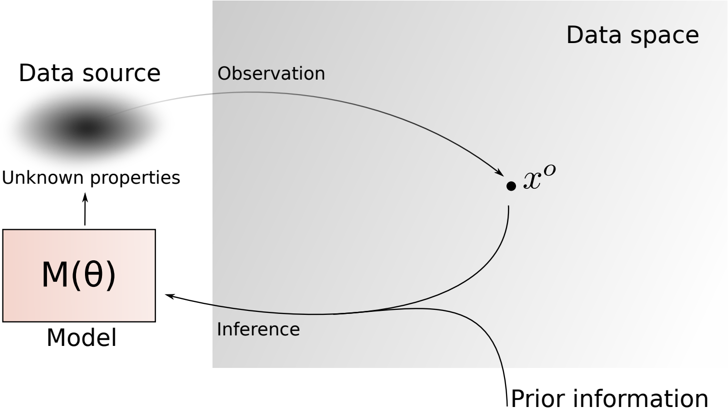
General approach

- Set up a model with potential properties θ (parameters)
- See which θ are in line with the observed data x^o



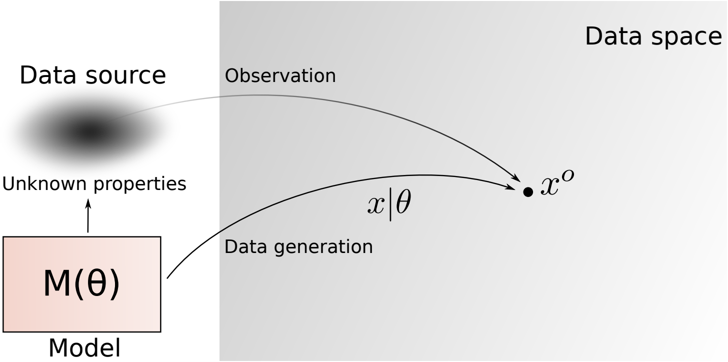
General approach

- Set up a model with potential properties θ (parameters)
- See which θ are in line with the observed data x^o



The likelihood function

- Measures agreement between θ and the observed data x^o
- Probability to generate data like x^o if hypothesis θ holds



Performing statistical inference

- If $L(\theta)$ is known, inference is straightforward
- Maximum likelihood estimation

$$\hat{\theta} = \mathrm{argmax}_\theta \, L(\theta)$$

- Bayesian inference

$$p(\theta|x^o) \propto p(\theta) \times L(\theta)$$

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

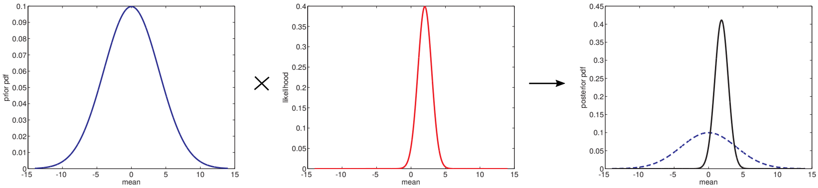
Allows us to learn from data by updating probabilities

Model specification

- Textbook: model $\equiv$ family of probability density functions
- Probability density functions (pdfs) $p(x|\theta)$ satisfy

$$\underbrace{p(x|\theta) \geq 0}_{\text{non-negativity}} \qquad \underbrace{\int p(x|\theta)\mathrm{d}x = 1}_{\text{normalization}}$$

- Likelihood function $L(\theta) = p(x^o|\theta)$
- Closed form solutions are possible



prior $\times$ likelihood $\rightarrow$ posterior

Models with intractable likelihoods

- Not all models are specified as family of pdfs $p(x|\theta)$.
- I worked on
  1. Simulator-based models
  2. Unnormalised models
- The models are rather different, common point:

  Multiple integrals needed to be computed to represent the models in terms of pdfs $p(x|\theta)$.
- Solving the integrals exactly is computationally impossible. (curse of dimensionality)

  $\Rightarrow$ No model pdfs $p(x|\theta)$
  $\Rightarrow$ No likelihood function $L(\theta) = p(x^o|\theta)$
  $\Rightarrow$ No exact inference

General research question

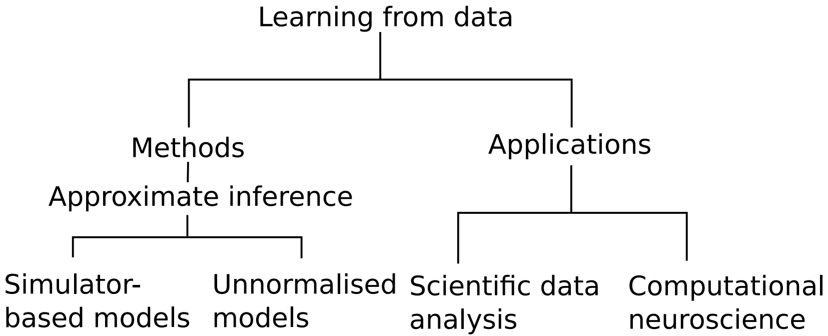
How to efficiently perform (Bayesian) inference when the likelihood function is too expensive too compute?

General research question

How to efficiently perform (Bayesian) inference when the likelihood function is too expensive too compute?

Efficiency $\equiv$ good trade-off between speed and accuracy.

My research interests in brief



Unnormalised models

- Used for modelling
  - images (Markov random fields)
  - social networks (exponential random graphs)
  - ferro-magnetism (Ising model)
  - ...
- Specified via a non-negative function $q(x|\theta) \propto p(x|\theta)$,

$$\int \cdots \int q(x|\theta) dx = Z(\theta) \neq 1 \qquad p(x|\theta) = \frac{q(x|\theta)}{Z(\theta)}$$

- Advantage: Specifying unnormalised models is often easier than specifying normalized models
- Disadvantage: Integral defining $Z(\theta)$ can generally not be computed. Likelihood function is intractable.

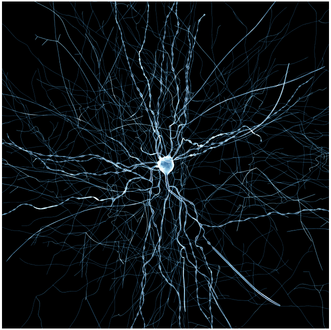
Simulator-based models

- Models which specify a mechanism for generating data
  - e.g. stochastic dynamical systems
  - computer models / simulators of some complex physical process
- Occur in multiple and diverse scientific fields.
- Different communities use different names:
  - Generative (latent-variable) models
  - Implicit models
  - Stochastic simulation models
  - Probabilistic programs

Simulator-based models are widely used

Examples:

- Evolutionary biology:
  Simulating evolution
- Neuroscience:
  Simulating neural circuits
- Ecology:
  Simulating species migration
- Health science:
  Simulating the spread of an infectious disease
- Physics:
  Simulating quantum systems (?)



Simulated neural activity in rat somatosensory cortex
(Figure from `https://bbp.epfl.ch/nmc-portal`)

Pros and Cons

- Advantage: detailed and realistic modelling
- Disadvantage: likelihood function is generally intractable due to unobserved variables.
- To compute $p(x|\theta)$ one has to take into account all possible states of the unobserved variables

$$p(x|\theta) = \int \cdots \int p(x, z|\theta) dz$$

- This is generally computationally impossible.

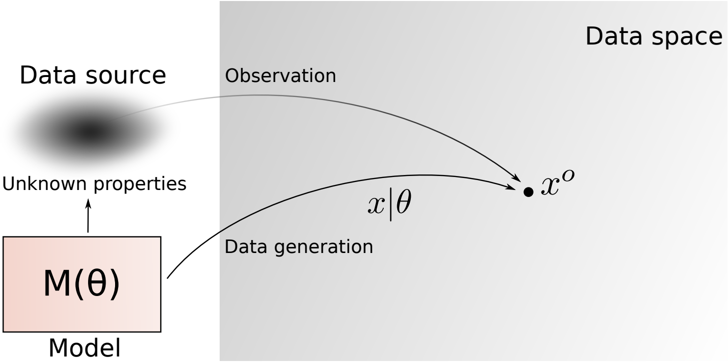
Program

Overview of my research interests

Efficient inference for simulator-based models

The likelihood function

- Measures agreement between θ and the observed data x^o
- Probability to generate data like x^o if hypothesis θ holds

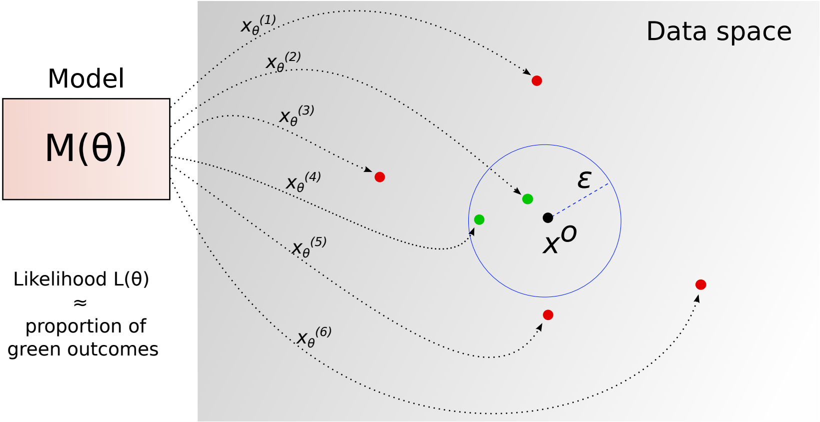


Three foundational issues

1. How should we assess whether $x_\theta \equiv x^o$?
2. How should we compute the probability of the event $x_\theta \equiv x^o$?
3. For which values of θ should we compute it?

Traditional approach

Likelihood: Probability to generate data like x^o if hypothesis θ holds



$$L(\theta) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbb{1} \left(d(x_\theta^{(i)}, x^o) \leq \epsilon \right)$$

Traditional approach

1. How should we assess whether $x_\theta \equiv x^o$?
   $\Rightarrow$ Check whether $||T(x_\theta) - T(x^o)|| \leq \epsilon$
2. How should we compute the probability of the event $x_\theta \equiv x^o$?
   $\Rightarrow$ By counting
3. For which values of θ should we compute it?
   $\Rightarrow$ Sample from the prior

Traditional approach

1. How should we assess whether $x_\theta \equiv x^o$?
   $\Rightarrow$ Check whether $||T(x_\theta) - T(x^o)|| \leq \epsilon$
2. How should we compute the probability of the event $x_\theta \equiv x^o$?
   $\Rightarrow$ By counting
3. For which values of θ should we compute it?
   $\Rightarrow$ Sample from the prior

- Corresponds to a traditional version of a method called approximate Bayesian computation.
- Trade-off between computational cost and accuracy of the inference may be poor.

Some of my work

1. How should we assess whether $x_\theta \equiv x^o$?
   $\Rightarrow$ Use classification (Gutmann et al, 2014, 2017)

1. How should we assess whether $x_\theta \equiv x^o$?
2. How should we compute the probability of the event $x_\theta \equiv x^o$?
   $\Rightarrow$ Use density ratio estimation (Dutta et al, 2016, arXiv:1611.10242)

2. How should we compute the probability of the event $x_\theta \equiv x^o$?
3. For which values of θ should we compute it?
   $\Rightarrow$ Use Bayesian optimisation (Gutmann and Corander, 2013-2016)

Example: Bacterial infections in child care centres

- Comparison of the Bayesian optimisation approach with a standard population Monte Carlo ABC approach.
- Roughly equal results using 1000 times fewer simulations.

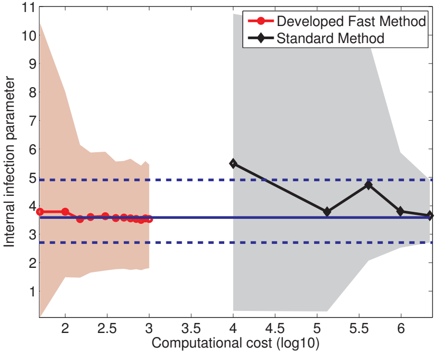
4.5 days with 200 cores

↓

90 minutes with seven cores

Posterior means: solid lines,

credibility intervals: shaded areas or dashed lines.



(Gutmann and Corander, 2016)

Summary

- Learning from data
  - Methods
    - Approximate inference
      - Simulator-based models
        - Importance
        - Difficulty of performing inference
        - Solutions using classification, density ratio estimation, and Bayesian optimimisation
      - Unnormalised models
  - Applications
    - Scientific data analysis
    - Computational neuroscience