# Bayesian Inference via Classification

Michael Gutmann

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

michael.gutmann@ed.ac.uk

4th October 2016

# Reference

M.U. Gutmann, R. Dutta, S. Kaski, and J. Corander
Statistical Inference of Intractable Generative Models via Classification, *arXiv:1407.4981*.

J. Lintusaari, M.U. Gutmann, R. Dutta, S. Kaski, and J. Corander
Fundamentals and Recent Developments in Approximate Bayesian Computation, *Systematic Biology*, in press, 2016.

# Take-home message

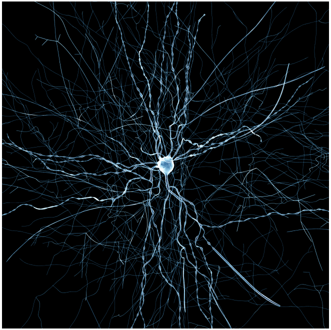
By using classification, we can measure the distance between two data sets and infer the parameters of intractable generative models.

# Problem statement

- Given:
  - Observed data $\mathbf{y}^o$
  - Intractable generative model, parametrised by $\boldsymbol{\theta}$
  - (Possibly) A prior probability density function (pdf) for $\boldsymbol{\theta}$, $p_{\boldsymbol{\theta}}$
- Task: Perform inference about $\boldsymbol{\theta}$
  - Maximum likelihood estimate
  - Posterior distribution of $\boldsymbol{\theta}$

# Generative model

- Model that specifies a mechanism for generating data **y**
  - e.g. stochastic dynamical system
  - computer model / simulator of some complex biological process
- Such models are widely used
  - Evolutionary biology: Simulating evolution
  - Neuroscience: Simulating neural circuits
  - Ecology: Simulating species migration
  - Health science: Simulating the spread of an infectious disease



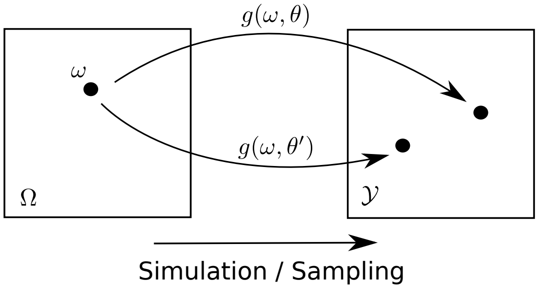
Simulated neural activity in rat somatosensory cortex
(Figure from https://bbp.epfl.ch/nmc-portal)

# Generative model

- Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space.
- A generative model is a collection of (measurable) functions $g(., \boldsymbol{\theta})$ parametrized by $\boldsymbol{\theta}$,

$$\boldsymbol{\omega} \in \Omega \mapsto \mathbf{y} = g(\boldsymbol{\omega}, \boldsymbol{\theta}) \in \mathcal{Y} \tag{1}$$

- For any fixed $\boldsymbol{\theta}$, $\mathbf{y} = g(., \boldsymbol{\theta})$ is a random variable.



# Intractable generative model

- Generally impossible to write down the pdf $p(\mathbf{y}|\boldsymbol{\theta})$ of the generated data.
- Likelihood function $L(\boldsymbol{\theta}) = p(\mathbf{y}^o|\boldsymbol{\theta})$ not known
  well defined, but computation not tractable

  supplemental
- Exact inference not possible

# Approximate inference for generative models

- Inference is approximate due to approximation of $p(\mathbf{y}^o|\boldsymbol{\theta})$
- Rough classification:
  - Parametric approximation: e.g. synthetic likelihood
    (Wood, Nature, 2010)
  - Nonparametric approximation: e.g approximate Bayesian computation
    (Recent review: Gutmann et al, Systematic Biology, 2016)
- General idea: Identify the values of the parameters $\boldsymbol{\theta}$ for which simulated data resemble the observed data

# Meta algorithm for approximate Bayesian computation

- Let $\mathbf{y}^o$ be the observed data.
- Iterate many times:
  1. Sample $\boldsymbol{\theta}$ from a proposal distribution $q(\boldsymbol{\theta})$
  2. Sample $\mathbf{y}|\boldsymbol{\theta}$ according to the model
  3. Compute the distance $d(\mathbf{y}, \mathbf{y}^o)$ between simulated and observed data
  4. Retain $\boldsymbol{\theta}$ if $d(\mathbf{y}, \mathbf{y}^o) \leq \epsilon$
- Different choices for $q(\boldsymbol{\theta})$ give different algorithms
- Produces samples from the (approximate) posterior when the bandwidth $\epsilon$ is small.

# Two major difficulties

1. How to handle the computational cost?
2. How to measure the distance $d$ between simulated and observed data?

# Two major difficulties

1. How to handle the computational cost?

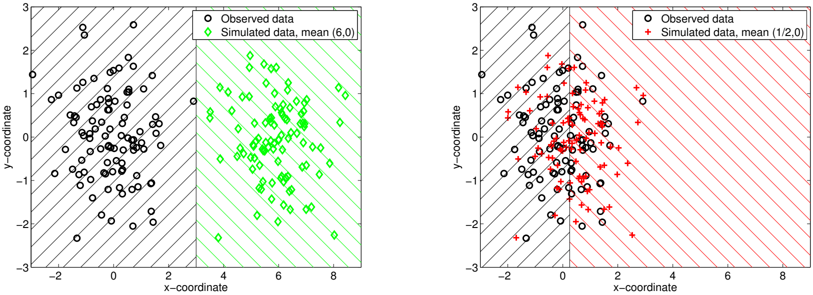
   → Use Bayesian optimization
   M.U. Gutmann and J. Corander
   Bayesian optimization for likelihood-free inference of simulator-based statistical models
   *Journal of Machine Learning Research*, 17(125): 1–47, 2016

2. How to measure the distance $d$ between simulated and observed data?

   → Use classification
   M.U. Gutmann, R. Dutta, S. Kaski, and J. Corander
   Statistical Inference of Intractable Generative Models via Classification, *arXiv:1407.4981*

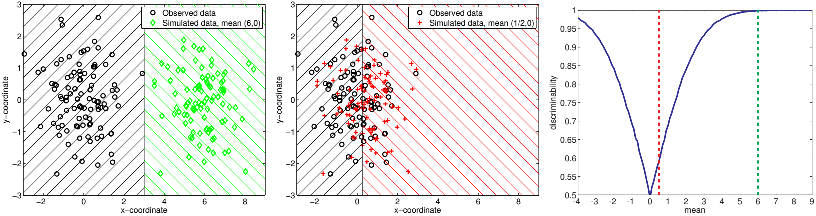
# Using classification to measure the distance

Correctly classifying data into two categories is usually easier if the two data sets were generated with very different values of $\boldsymbol{\theta}$ (left) than with similar values (right).



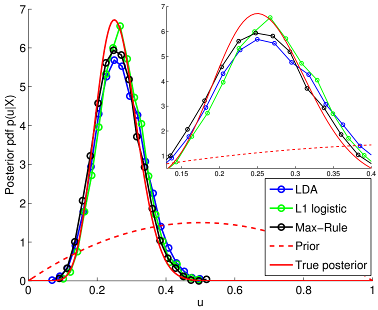
# Using classification to measure the distance

- Classification accuracy (discriminability) serves as distance measure.
- Value of $1/2$: close; Value of 1: far
- Complete arsenal of classification methods becomes available to inference.

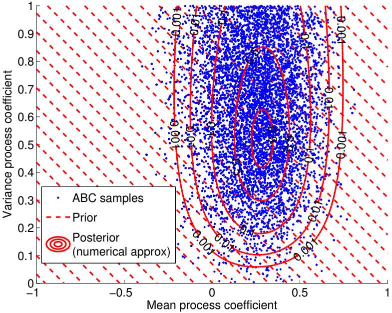


# Using the distance for posterior inference

- Approximate Bayesian computation (ABC) with classification accuracy as distance measure
- Results for toy models:



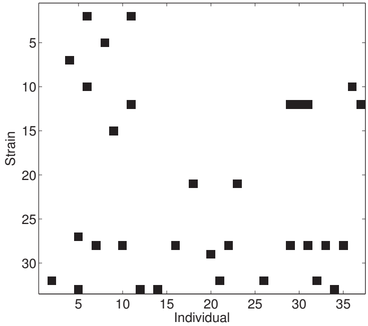

(a) binary data (Bernoulli)




(b) time series (ARCH)

# Application to epidemiology of infectious diseases

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

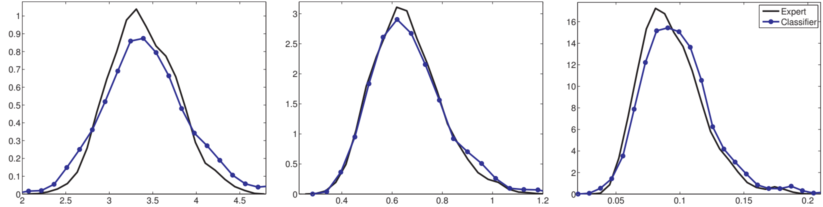


# Application to epidemiology of infectious diseases

- Generative model: latent continuous-time Markov chain for the transmission dynamics in a DCC and an observation model (Numminen et, Biometrics, 2013).
- The model has three parameters:
  - $\beta$: rate of infections within a DCC
  - $\Lambda$: rate of infections outside a DCC
  - $\theta$: possibility to be infected with multiple strains
- Likelihood is intractable (data at a single time point are available only).

# Application to epidemiology of infectious diseases

- Our classification-based distance measure does not use domain/expert knowledge.
- Performs as well as a distance measure based on domain knowledge (Numminen et, Biometrics, 2013).



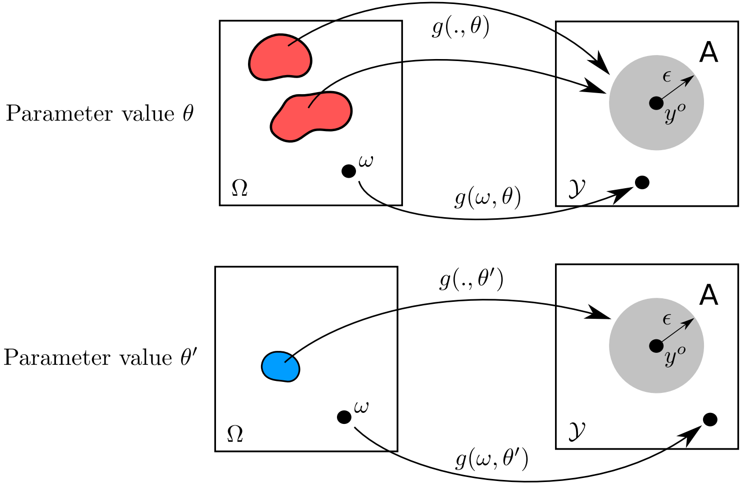

(c) Posterior pdf for $\beta$ (d) Posterior pdf for $\Lambda$ (e) Posterior pdf for $\theta$

# Summary

- Topic: Inference for generative models where the likelihood function is intractable
- Inference principle: Find parameter values for which the distance between simulated and observed data is small
- Showed that, by using classification, we can
  - measure the distance between two data sets, and
  - infer the parameters of intractable generative models.
- More (results, math, related work, ...) in the reference paper!

# Likelihood function is implicitly defined

$$\Pr(y \in A \mid \theta) = \mathcal{P}(\{\omega : g(\omega, \theta) \in A\})$$




back