Likelihood-free inference

GAMBIT X

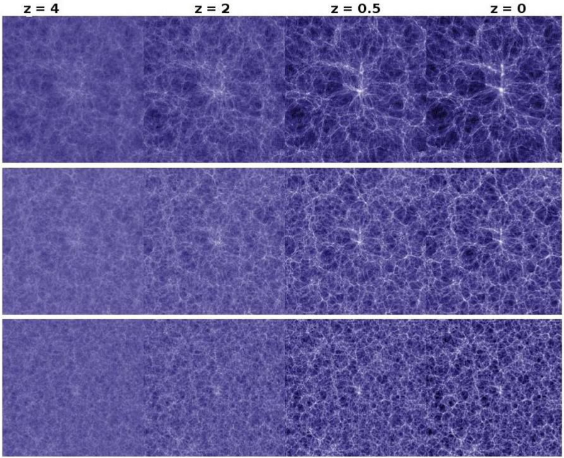
Will Handley
wh260@cam.ac.uk

Kavli Institute for Cosmology
Cavendish Laboratory (Astrophysics Group)
University of Cambridge

June 6^{th} 2019

- The term "Likelihood-free" is a misnomer – there is still very much a likelihood involved in the analysis
- LFI is a framework for situations where you don't know what the likelihood is, but can still simulate your system
  - i.e. you are freed from having to write down the likelihood explicitly
- Fundamental idea: construct a flexible proxy likelihood, and fit this to simulations.
- Fitted likelihood can then be used in both Frequentist and Bayesian analyses.
- Related to Approximate Bayesian Computation (ABC) but better.
- Key references from cosmology:
  - arXiv:1801.01497
  - arXiv:1903.00007
  - arXiv:1903.01473
  - arXiv:1904.05364
  - Another paper coming soon: "Compromise-free Likelihood-free inference"

Motivating example: Cosmological large-scale structure



- What is the likelihood $P(D|\theta)$ for large scale structure formation?
- Can simulate data $\hat{D}$ given set of cosmological parameters θ.
- Image shows effect of varying the amount of dark matter in simulation
- Would like to compare simulation to actual data: $\chi^2(\theta) \sim |\hat{D}(\theta) - D|^2$

Two key problems

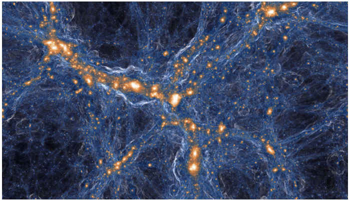
$$\chi^2(\theta) \sim |\hat{D}(\theta) - D|^2$$

1. Datasets are in general exponentially large
   - The whole point of science is to compress data into models with a small number of parameters.
2. What is the correct metric to measure the difference between $\hat{D}$ and D?
   - ABC works by rejection sampling up to some ϵ difference above to select the correct θ.

The recent advance in LFI is a framework that solves both of these problems.

Solution 1: Massive compression

- All we are interested in is "What do the data D tell me about the n parameters of my model θ?"
- In theory, the dataset D should be compressible into n numbers without losing information about θ.
- For example, when inferring the underlying mean μ and variance σ^2 of some numbers $\{x_1, \cdots, x_d\}$, all you need is the sample mean $\bar{x}$ and sample variance S^2 (c.f. sufficient statistics).

$\theta = [\Omega_m, \Omega_b, \sigma_8]$  $\longrightarrow D = [0.3, 0.4, 0.9]$

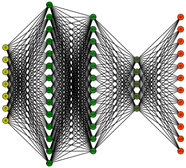
- The advance in cosmology which made the recent work possible (arXiv:1712.00012) was to generalise Karhunen-Loéve and MOPED to *score compression*:
  - If you have an approximate likelihood $L(\theta)$, then $\nabla_\theta L$ is in some sense an optimal compression.
- More compression schemes are possible: watch this space.

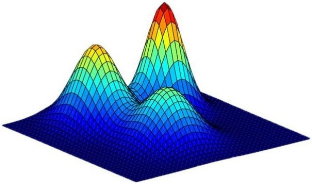
- One can generate a set of $K \sim 1000$ training simulations $\{(D_1, \theta_1), \cdots, (D_K, \theta_K)\}$.
- Apply compression, and these are samples in a $2n$-dimensional space.
- Build a proxy joint distribution $P(\theta, D) = f(\theta, D; \alpha)$ with some free parameters α:
- With this proxy joint, one has a likelihood of α defined by:

$$L(\alpha) = \prod_{i=1}^{K} f(\theta_i, D_i; \alpha)$$

- Use $L(\alpha)$ as a misfit function for maximising/marginalising α.

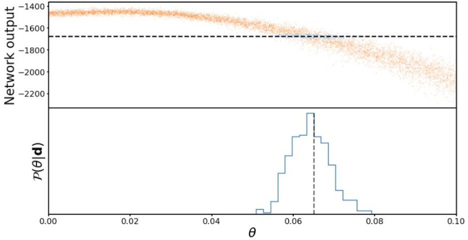
- Gaussian mixture models
- Neural density estimators
- ...

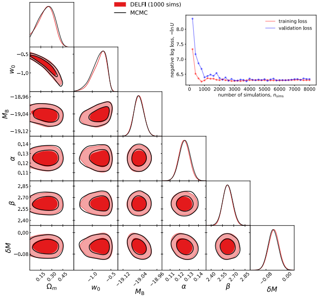




The full framework

1. In lieu of a likelihood, write data simulator $\hat{D}(\theta)$
2. Generate training data $\{(D_1, \theta_1), \cdots, (D_K, \theta_K)\}$
3. Choose massive compression scheme (e.g. score) and compress training data
4. Choose proxy distribution $f(D, \theta; \alpha)$ (e.g. neural density estimator) and fit to training data to give e.g. $P(D, \theta) = f(D, \theta; \alpha_{\max})$
5. Use trained joint for all your usual inference:
   - Can calculate likelihood by inputting the actual data (in compressed form)
   - For some proxy distributions (e.g. mixture models) you can get the evidence for free.





My research:

- Instead of maximising wrt α, in a Bayesian framework one should marginalise
- Can use Bayesian evidences to select the best proxy, and to pick/marginalise over the number of mixture components/neural network nodes.

Summary

- Recent advances have brought LFI into the realm of "possible" with current technology
- In ten years time, with advances in both theory and computing power everyone will be doing this (think MCMC twenty years ago).
- GAMBIT should be thinking about incorporating these techniques over the next few years.