unimpeded (& other cosmological inference tools)

Will Handley
<wh260@cam.ac.uk>

Royal Society University Research Fellow
Astrophysics Group, Cavendish Laboratory, University of Cambridge
Kavli Institute for Cosmology, Cambridge
Gonville & Caius College
willhandley.co.uk/talks

19th October 2023

Highlighted tools for this audience

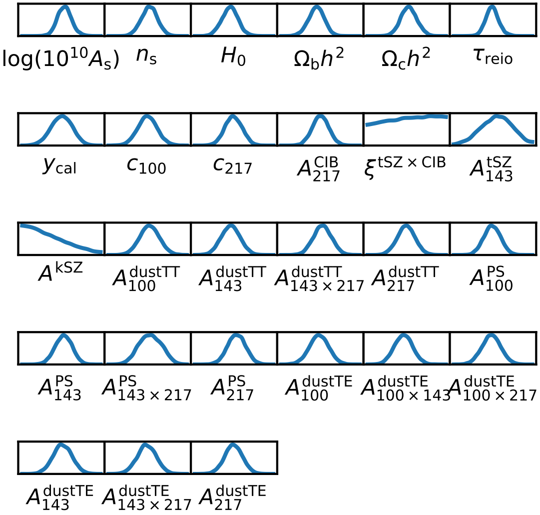
1. `anesthetic` [1905.04768]
    - how to get the most out of your nested sampling runs
    - `github.com/handley-lab/anesthetic`
2. `margarine` [2207.11457]
    - easy-to-use (neural) density estimation for marginal statistics
    - `github.com/htjb/margarine`
3. `prescience` [2309.06942]
    - Fully Bayesian forecasts (no more need for Fisher)
    - `github.com/ThomasGesseyJones/FullyBayesianForecastsExample`
4. `unimpeded`
    - up/downloading tool for transfering inference products (beyond chains)
    - `github.com/handley-lab/unimpeded`

Marginal inference

- Many cosmological likelihoods come with nuisance parameters that have limited relevance for onward inference.
- Notation: $\mathcal{L} = P(D|\theta, \alpha, M)$

$\mathcal{L}$	Likelihood	(e.g. `plik`),
D	Data	(e.g. CMB),
θ	Cosmological parameters	(e.g. Ω_m, $H_0 \ldots$),
α	Nuisance parameters	(e.g. $A_{\mathrm{planck}} \ldots$),
M	Model	(e.g. ΛCDM).

- Some marginal statistics (e.g. marginal means, posteriors…) are easy to compute.
- More machinery is needed for e.g. nuisance marginalised likelihoods and marginal KL divergences $\mathcal{D}_{\mathrm{KL}}$.



Relevant examples for this audience

1. CMB
   - Galatic foreground parameters: A_i, n_i,...
   - Calibration: c, y,...
2. Gravitational wave events
   - Nuisance event parameters: M_1, M_2, θ, ϕ, r, z, ...
3. (time-delay) supernovae
   - Nuisance lens parameters: Φ_{lens}
   - Supernova systematics

- Many modern cosmological problems present themselves in hierarchical form
- These (often well-studied) objects have a set of event-specific parameters α alongside parameters of global interest θ (e.g. H_0).
- Ideally we would be able to have a "nuisance marginalised" likelihood for combining hierarchically.

Nuisance marginalised likelihoods: Theory [2207.11457]

- Bayes theorem

$$\mathcal{L}(\theta,\alpha)\times\pi(\theta,\alpha)=\mathcal{P}(\theta,\alpha)\times\mathcal{Z} \tag{1}$$

Likelihood × Prior = Posterior × Evidence

α: nuisance parameters, θ: cosmo parameters.

- Marginal Bayes theorem

$$\mathcal{L}(\theta)\times\pi(\theta)=\mathcal{P}(\theta)\times\mathcal{Z} \tag{2}$$

- Non-trivially gives **nuisance-free likelihood**

$$\boxed{\mathcal{L}(\theta)=\frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}=\frac{\int\mathcal{L}(\theta,\alpha)\pi(\theta,\alpha)d\alpha}{\int\pi(\theta,\alpha)d\alpha} \tag{3}$$

Key properties

- Given datasets A and B, each with own nuisance parameters α_A and α_B:
- If you use $\mathcal{L}_A(\theta)$, you get the same (marginal) posterior and evidence if you had run with nuisance parameters α_A (ditto B).
- If you run inference on $\mathcal{L}_A(\theta)\times\mathcal{L}_B(\theta)$, you get the same (marginal) posterior and evidence if you had run with all nuisance parameters α_A, α_B on.

(weak marginal consistency requirements on joint $\pi(\theta,\alpha_A,\alpha_B)$ and marginal priors)

Nuisance marginalised likelihoods: Practice

Harry Bevins

PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

$\boxed{\mathcal{L}(\theta,\alpha)}$ $\boxed{\pi(\theta,\alpha)}$

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

Harry Bevins

PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

$\boxed{\mathcal{L}(\theta,\alpha)}$ $\boxed{\pi(\theta,\alpha)}$

$\downarrow$ $\downarrow$

$\boxed{\text{Nested Sampling}}$

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

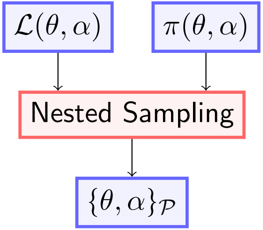
Harry Bevins

PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.



- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

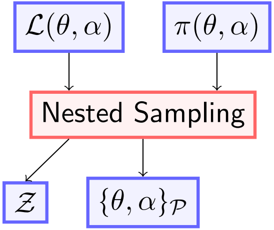
Harry Bevins

PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_\mathcal{P}$ and $\{\theta, \alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.



- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

Harry Bevins

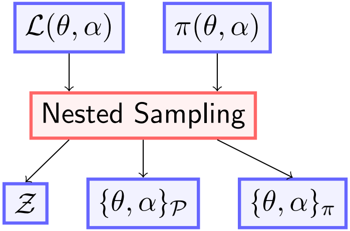
PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_\mathcal{P}$ and $\{\theta, \alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

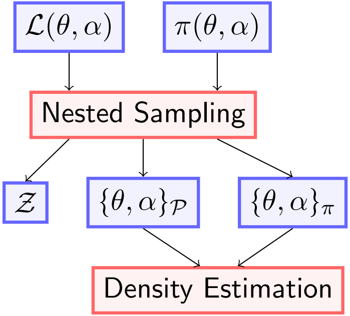
Harry Bevins

PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_\mathcal{P}$ and $\{\theta, \alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.



- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.

Nuisance marginalised likelihoods: Practice

Harry Bevins

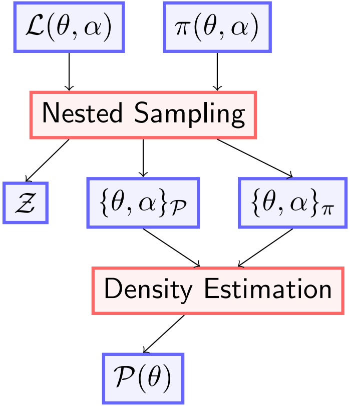
PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_\mathcal{P}$ and $\{\theta, \alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

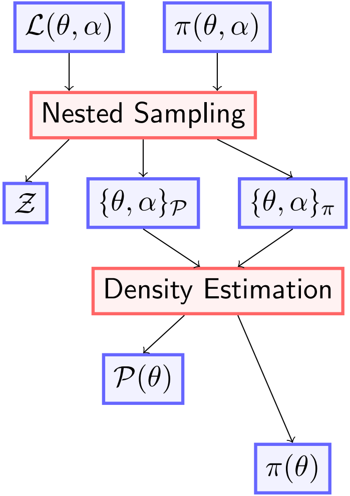
Harry Bevins
PhD→JRF

$$\boxed{\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

Harry Bevins

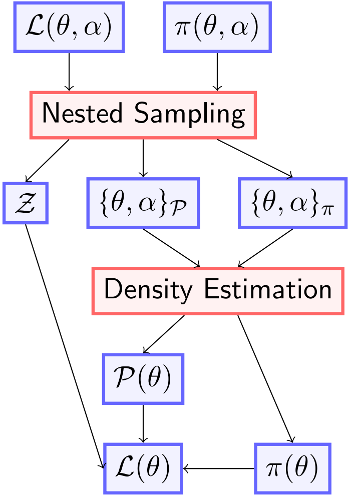
PhD→JRF

$$\mathcal{L}(\theta) = \frac{\mathcal{P}(\theta)\mathcal{Z}}{\pi(\theta)}$$

- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta, \alpha\}_\mathcal{P}$ and $\{\theta, \alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Nuisance marginalised likelihoods: Practice

Harry Bevins

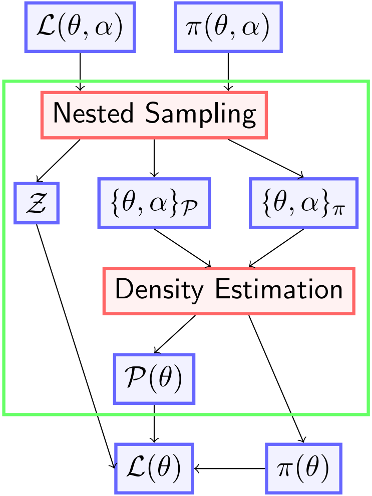
PhD→JRF

$$\boxed{\color{green}{\mathcal{L}}(\theta) = \frac{\color{blue}{\mathcal{P}}(\theta)\color{red}{\mathcal{Z}}}{\color{orange}{\pi}(\theta)}}$$

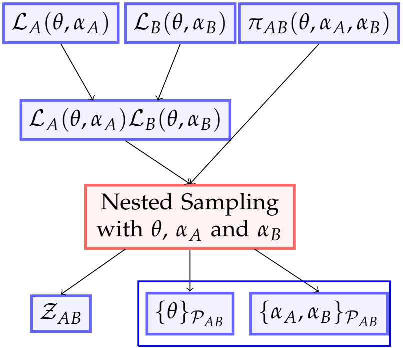
- To compute the nuisance marginalised likelihood, need:
  1. Bayesian evidence $\mathcal{Z}$
  2. Marginal prior and posterior densities

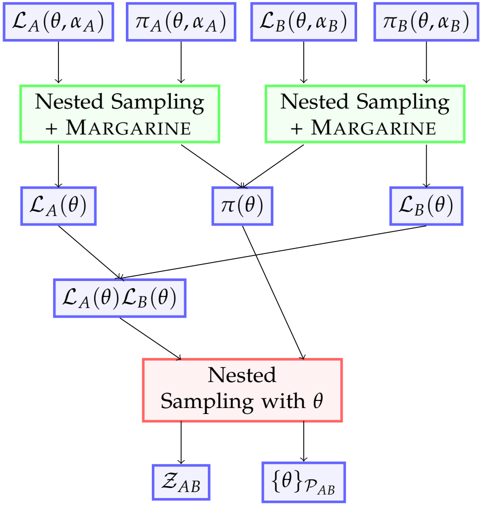
1. Use nested sampling to compute evidence $\mathcal{Z}$ and marginal samples $\{\theta,\alpha\}_\mathcal{P}$ and $\{\theta,\alpha\}_\pi$.
2. Use normalising flows to compute density estimators $\mathcal{P}(\theta)$, $\pi(\theta)$ from marginal samples.

- Emulators usually much faster than original likelihoods
- Library of pre-trained bijectors to be used as priors/emulators/nuisance marginalised likelihoods
- e.g. easy to apply a *Planck*/DES/HERA/JWST prior or likelihood to your existing MCMC chains without needing to install the whole cosmology machinery.



Combination





History of margarine

- Papamakarios et al [1912.02762] (normalising flows)
- Alsing et al [1903.00007] (Delfi)
- Nested sampling with any prior you like (Alsing & Handley) [2102.12478]
- `margarine` (theory) Bevins et al [2207.11457]
- `margarine` (practice) Bevins et al [2205.12841]

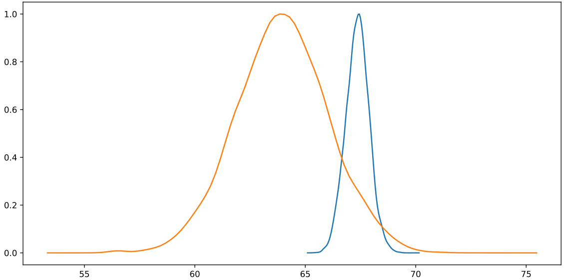
unimpeded

Universal Model comparison and Parameter Estimation Distributed over Every Dataset

Dily Ong
PhD

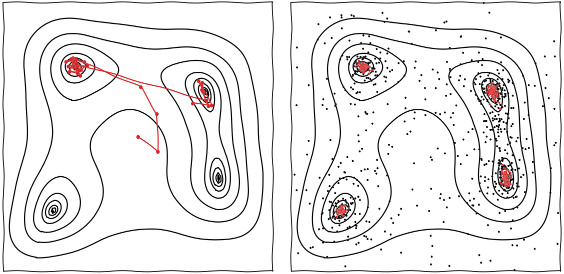
- Python tool for seamlessly downloading, uploading and cacheing of chains
- Data stored on `zenodo`
- hdf5 storage for fast & reliable storage
- `anesthetic` compatible for processing of chains [1905.04768]
- α-testers wanted! (email wh260@cam.ac.uk)
- End goal – community library which everyone contributes to so expensive inference products are reusable and reused.

```python
from unimpeded import Unimpeded
store = Unimpeded(cache='data.hdf5')
samps = store('planck')
samps.H0.plot.kde_1d()
samps = store('planck', model='klcdm')
samps.H0.plot.kde_1d()
```



Inference Legacy Archive

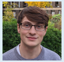
- DiRAC 2020 RAC allocation of 30MCPUh
- Main goal: Planck Legacy Archive equivalent
- Parameter estimation $\rightarrow$ Model comparison
- MCMC $\rightarrow$ Nested sampling
- Planck $\rightarrow$ $\{\text{Planck}, \text{DESY1}, \text{BAO}, \ldots\}$
- Pairwise combinations
- Suite of tools for processing these
  - `anesthetic` 2.0
  - `unimpeded` 1.0
  - `zenodo` archive
  - `margarine`
- MCMC chains also available.
- Library of bijectors emulators for fast re-use



Cosmological forecasting

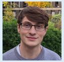
Thomas Gessey-Jones

PhD



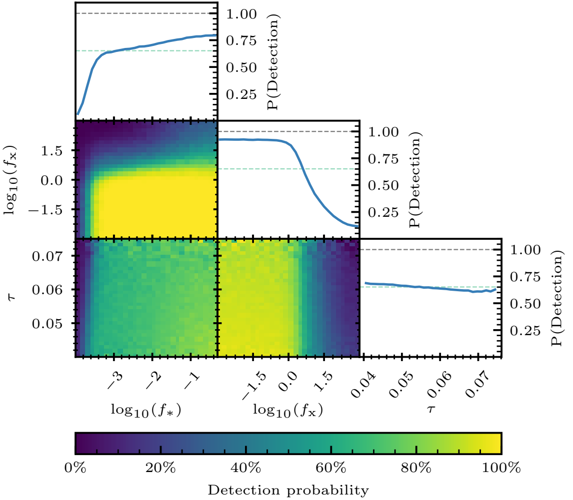
Have you ever done a Fisher forecast, and then felt Bayesian guilt?

- Cosmologists are interested in forecasting what a Bayesian analysis of future data might produce.
- Useful for:
  - white papers/grants,
  - optimising existing instruments/strategies,
  - picking theory/observation to explore next.
- To do this properly:
  1. start from current knowledge $\pi(\theta)$, derived from current data
  2. Pick potential dataset D that might be collected from $P(D)$ $(= \mathcal{Z})$
  3. Derive posterior $P(\theta|D)$
  4. Summarise science (e.g. constraint on θ, ability to perform model comparison)
- This procedure should be marginalised over:
  1. All possible parameters θ (consistent with prior knowledge)
  2. All possible data D
- i.e. marginalised over the joint $P(\theta, D) = P(D|\theta)P(\theta)$.
- Historically this has proven very challenging.
- Most analyses assume a fiducial cosmology θ_*, and/or a Gaussian likelihood/posterior (c.f. Fisher forecasting).
- This runs the risk of biasing forecasts by baking in a given theory/data realisation.



Fully Bayesian Forecasting [2309.06942]

- Simulation based inference gives us the language to marginalise over parameters θ and possible future data D.
- Evidence networks [2305.11241] give us the ability to do this at scale for forecasting.
- Demonstrated in 21cm global experiments, marginalising over:
  - theoretical uncertainty
  - foreground uncertainty
  - systematic uncertainty
- Able to say "at 67mK radiometer noise", have a 50% chance of 5σ Bayes factor detection.
- Can use to optimise instrument design
- Re-usable package: `prescience`



Conclusions

github.com/handley-lab

1. `anesthetic` [1905.04768]
   - how to get the most out of your nested sampling runs
   - `github.com/handley-lab/anesthetic`
2. `margarine` [2207.11457]
   - easy-to-use (neural) density estimation for marginal statistics
   - `github.com/htjb/margarine`
3. `prescience` [2309.06942]
   - Fully Bayesian forecasts (no more need for Fisher)
   - `github.com/ThomasGesseyJones/FullyBayesianForecastsExample`
4. `unimpeded`
   - up/downloading tool for transfering inference products (beyond chains)
   - `github.com/handley-lab/unimpeded`