Sampling methods for high energy physics & particle astrophysics

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

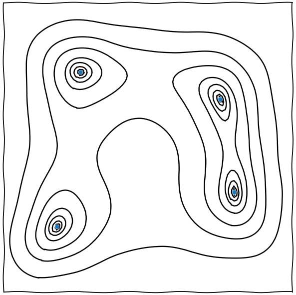
19th August 2024

Numerical inference tasks

- Given a (scalar) function f with a vector of parameters θ, one might want to:

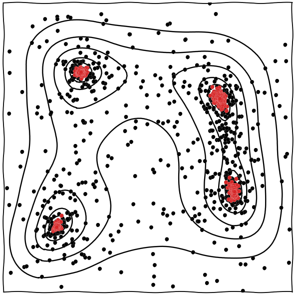
Optimise

$$\theta_{\mathrm{Max}} = \max_{\theta} f(\theta)$$



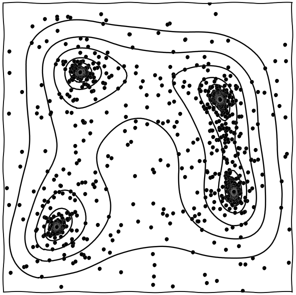
Explore

$$\text{draw/sample} \quad \theta \sim f$$



Integrate

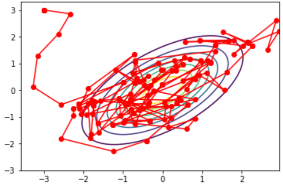
$$\int f(\theta) dV$$

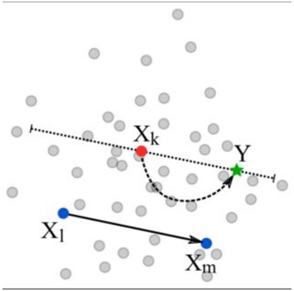


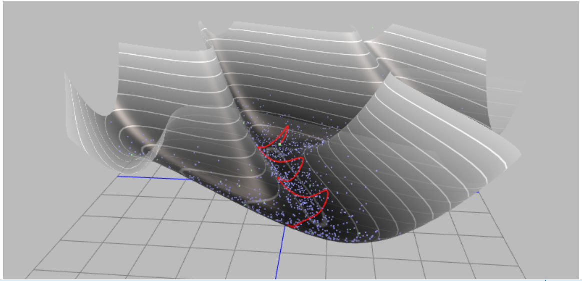
(incomplete) list of techniques

- Optimisers
  - Gradient descent (ADAM, BFGS)
  - simplex method (Nelder-Mead)
  - Genetic algorithms (Diver)
- Samplers
  - **Metropolis-Hastings** (PyMC, MontePython)
  - Hamiltonian Monte Carlo (Stan, blackjax)
  - Ensemble sampling (emcee, zeus).
  - Variational Inference (Pyro, NIFTY)
- Integrators
  - **Nested sampling** (MultiNest, dynesty)
  - Thermodynamic integration
  - Sequential Monte Carlo (pocomc)

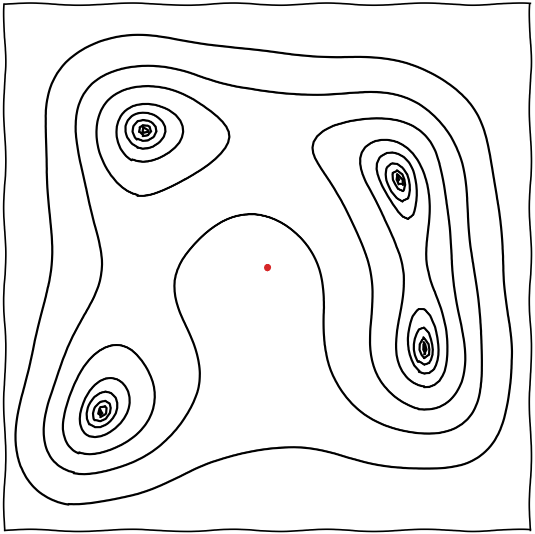
emcee
the mcmc hammer



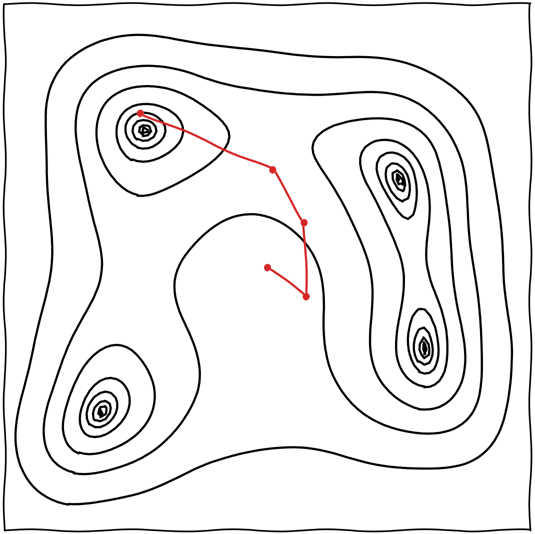




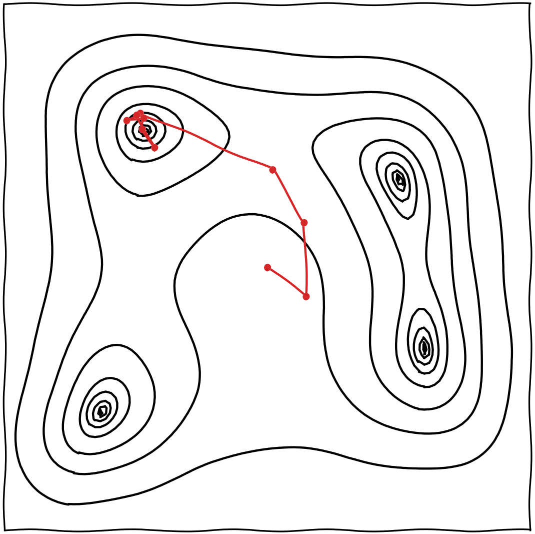
MCMC



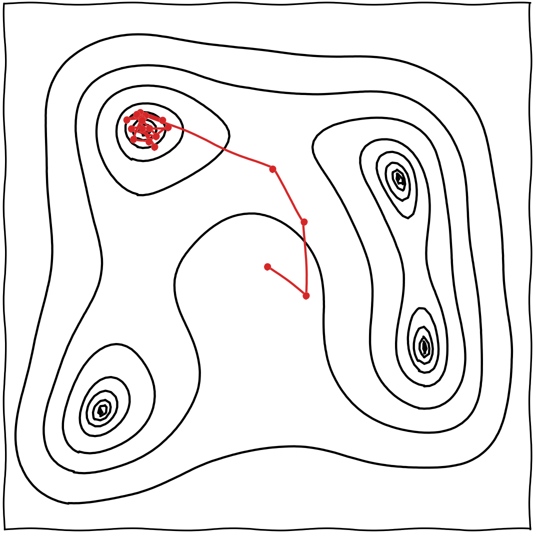
MCMC



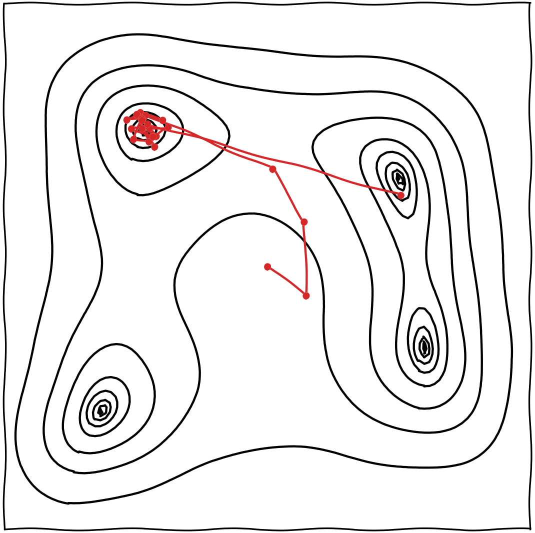
MCMC



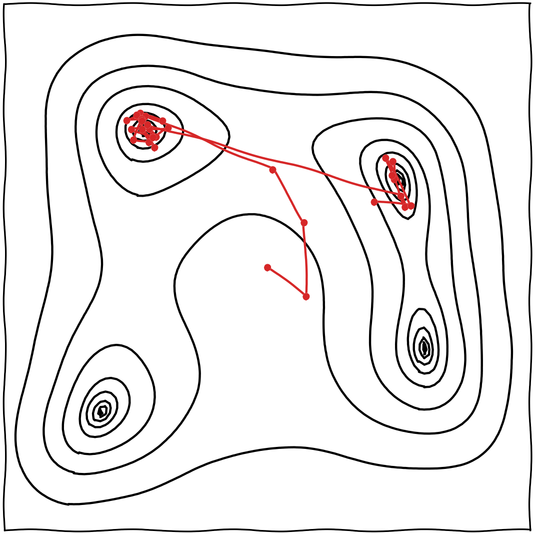
MCMC



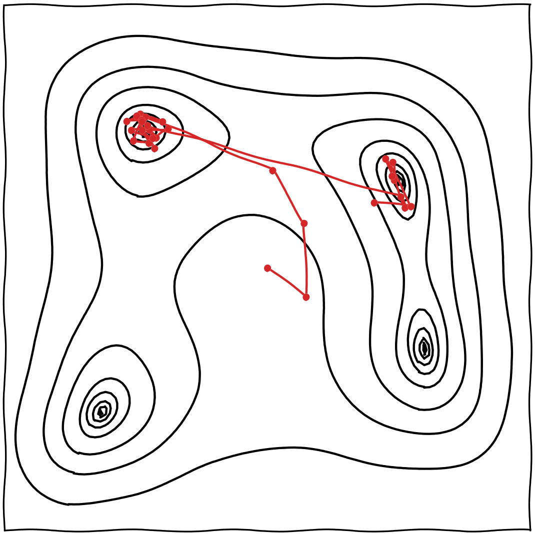
MCMC



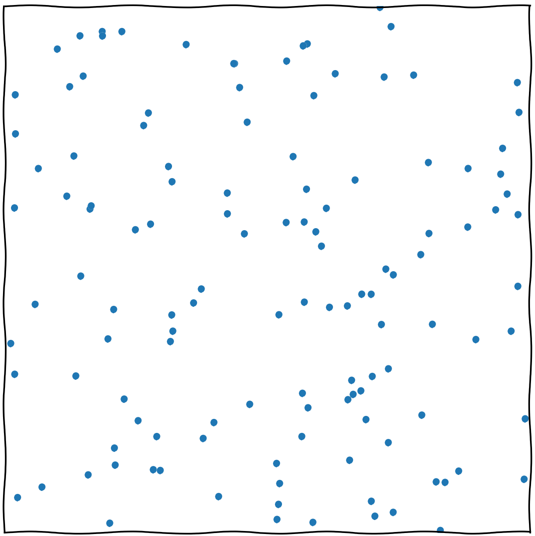
MCMC



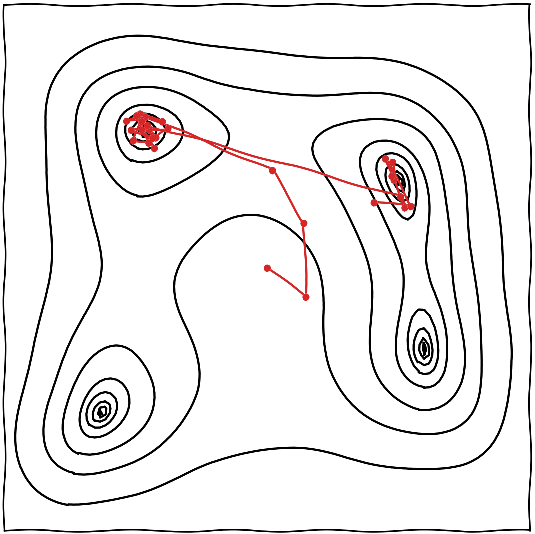
MCMC



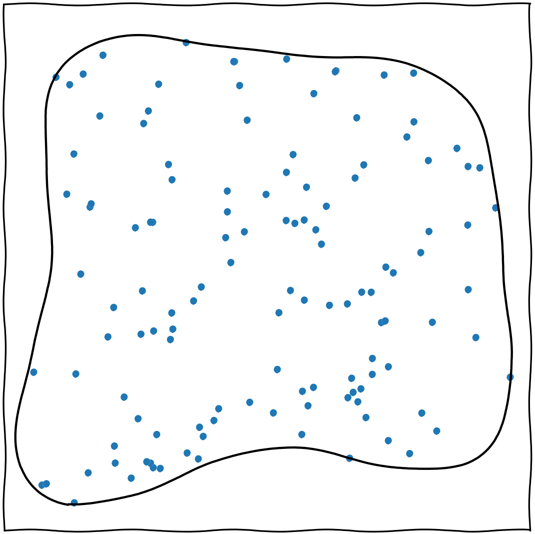
Nested sampling



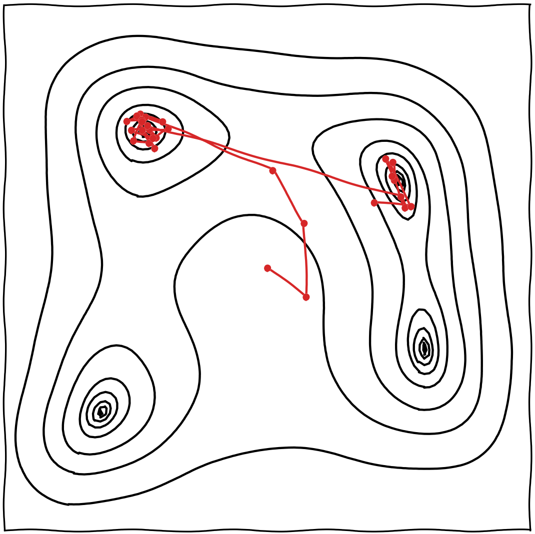
MCMC



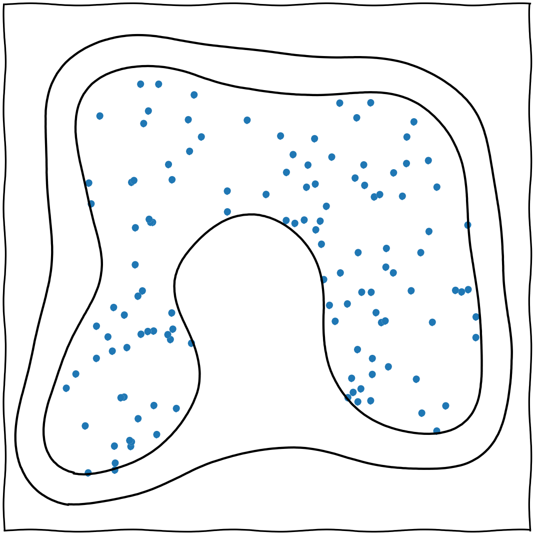
Nested sampling



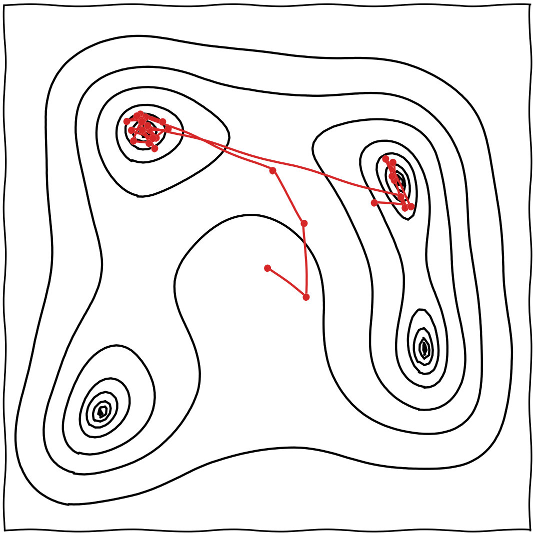
MCMC



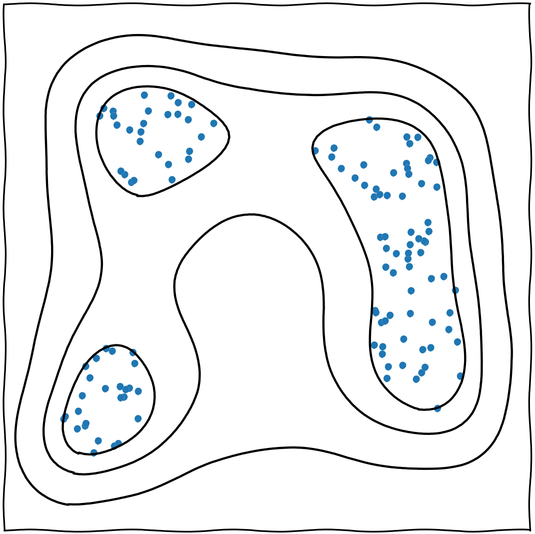
Nested sampling



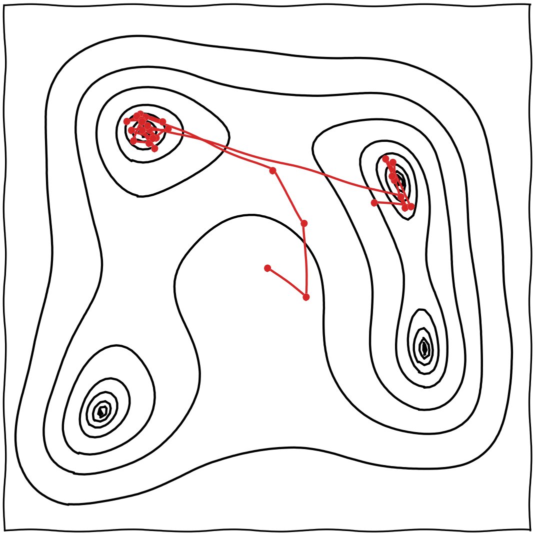
MCMC



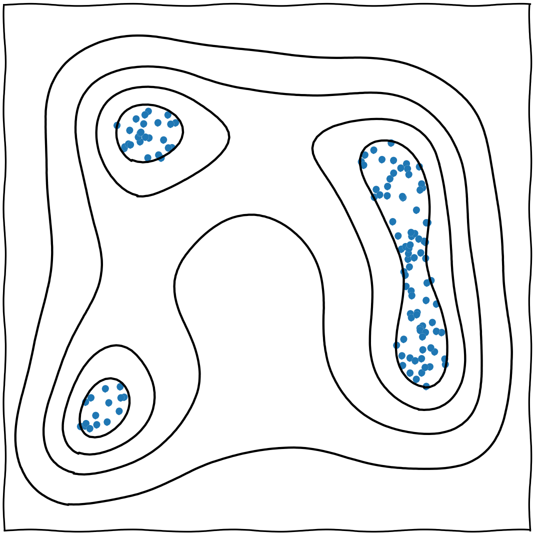
Nested sampling



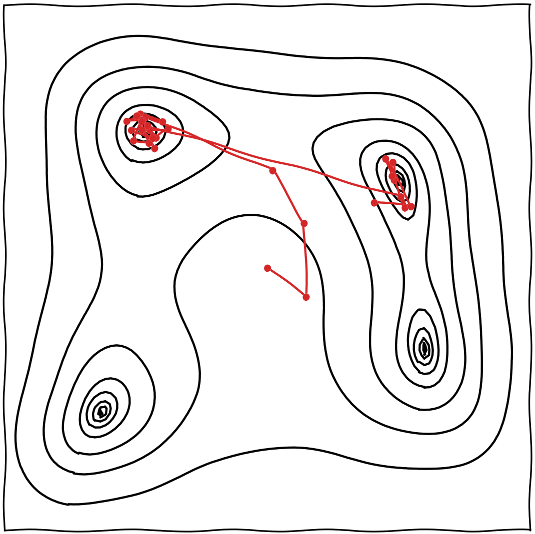
MCMC



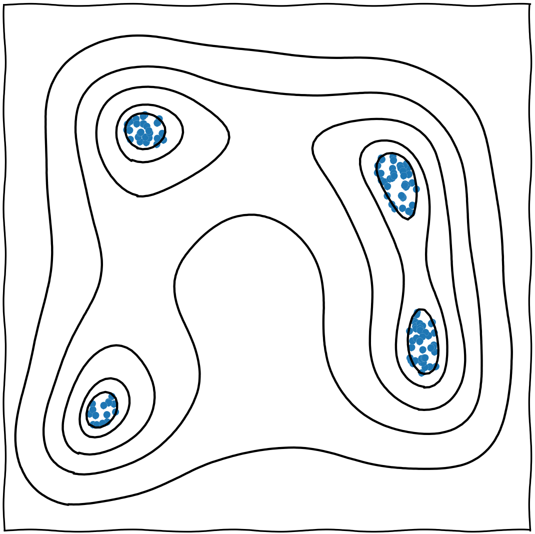
Nested sampling



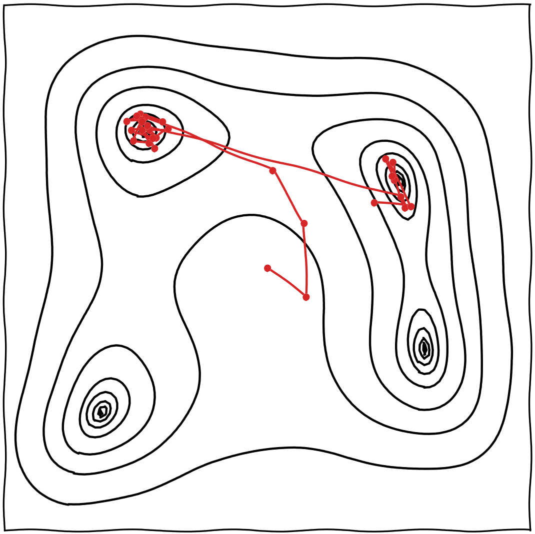
MCMC



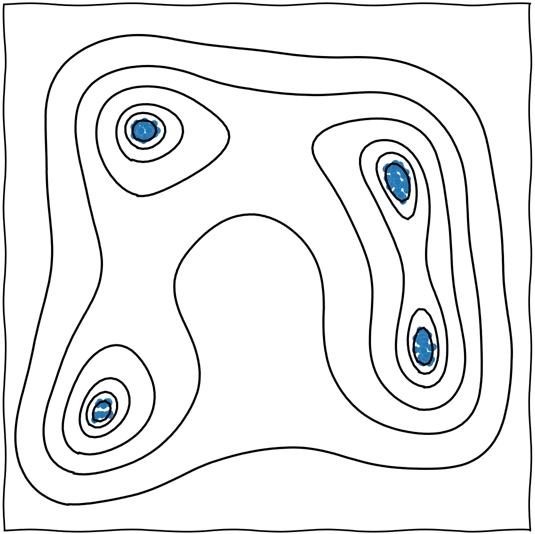
Nested sampling



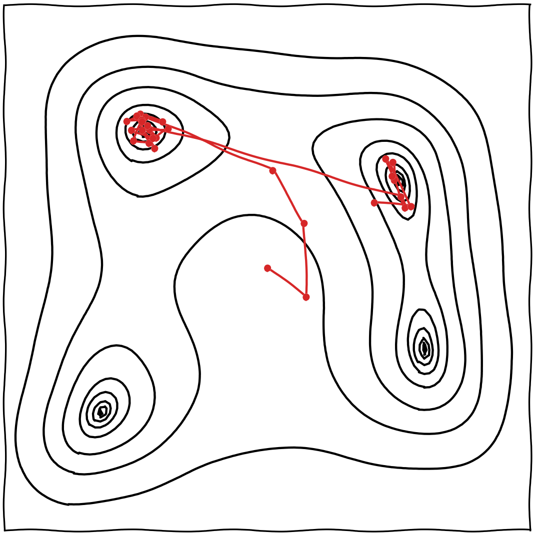
MCMC



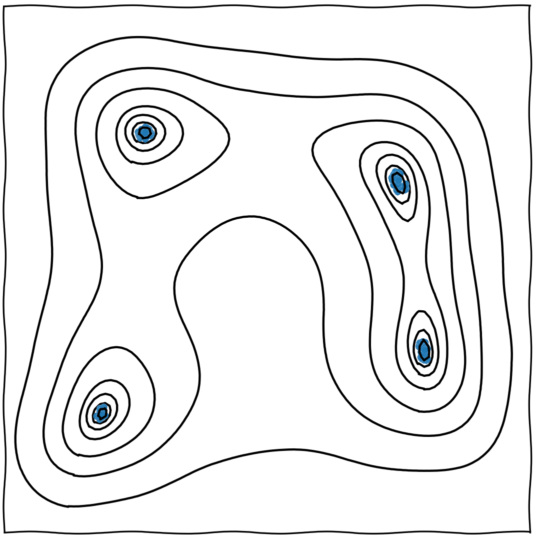
Nested sampling



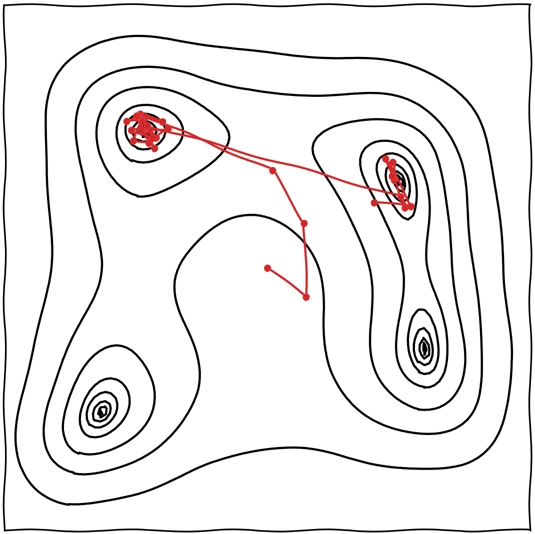
MCMC



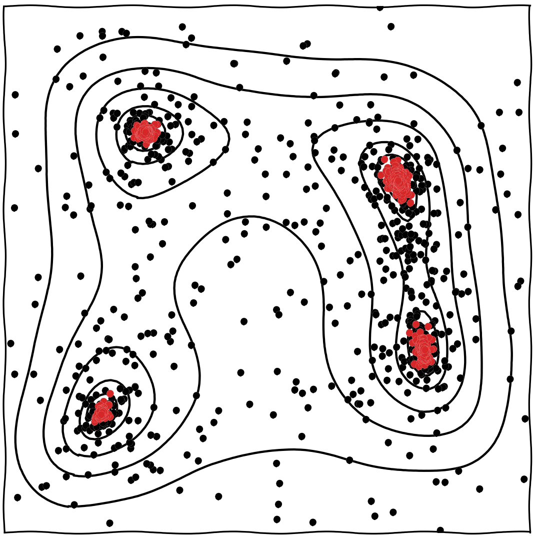
Nested sampling



MCMC

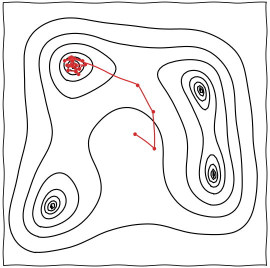


Nested sampling



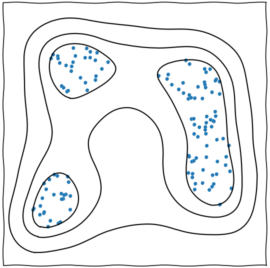
MCMC

- Single "walker"
- Explores posterior
- Fast, if proposal matrix is tuned
- Parameter estimation, suspiciousness calculation
- Channel capacity optimised for generating posterior samples



Nested sampling

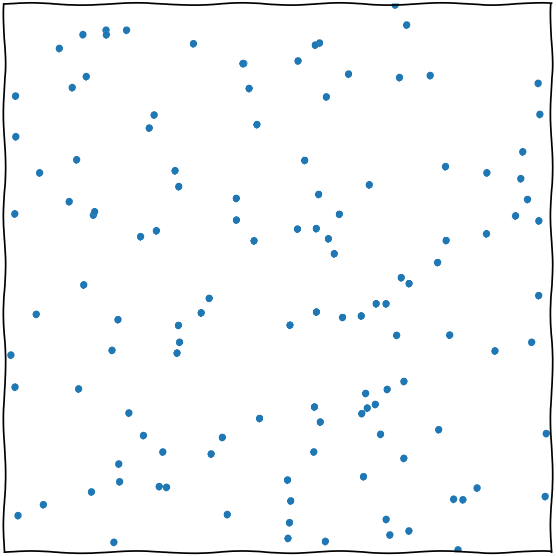
- Ensemble of "live points"
- Scans from prior to peak of likelihood
- Slower, no tuning required
- Parameter estimation, model comparison, tension quantification
- Channel capacity optimised for computing partition function



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

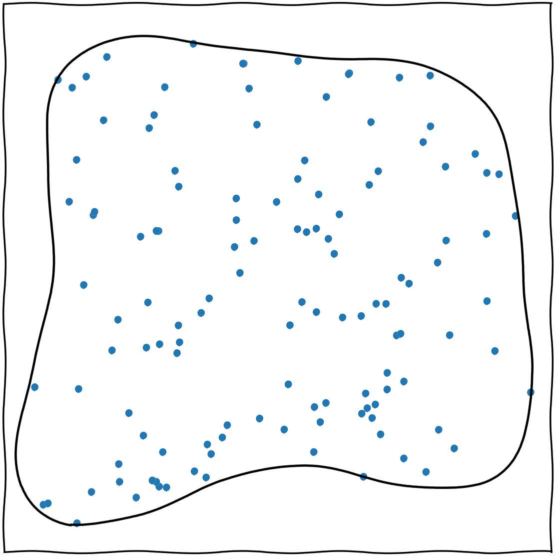
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

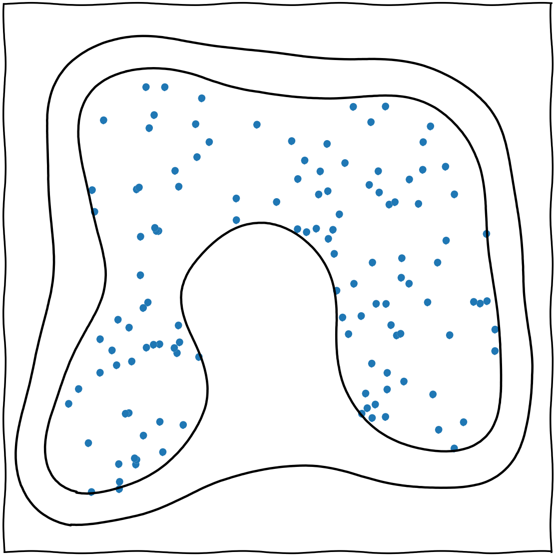
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

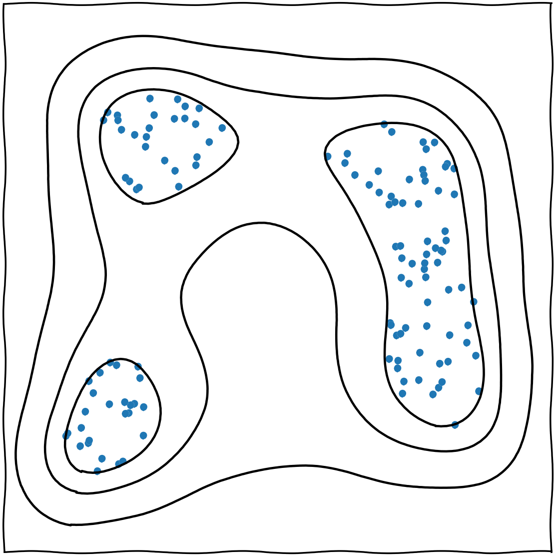
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

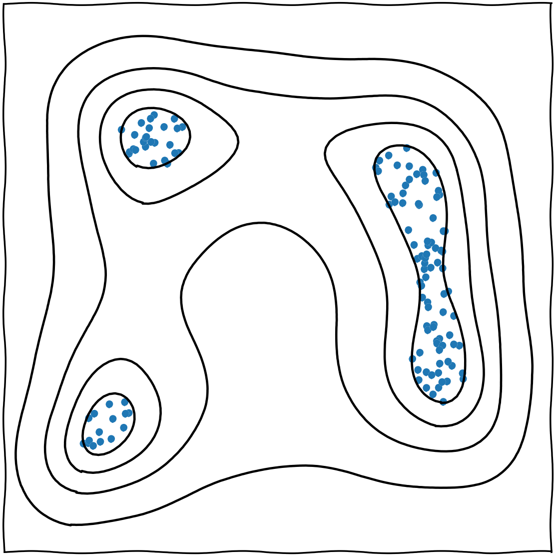
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

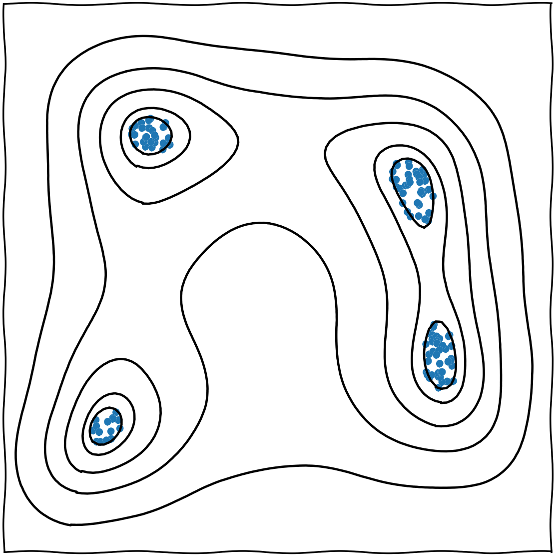
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

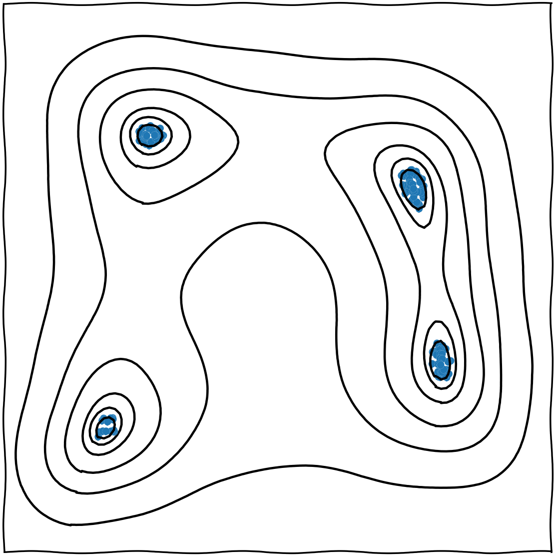
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

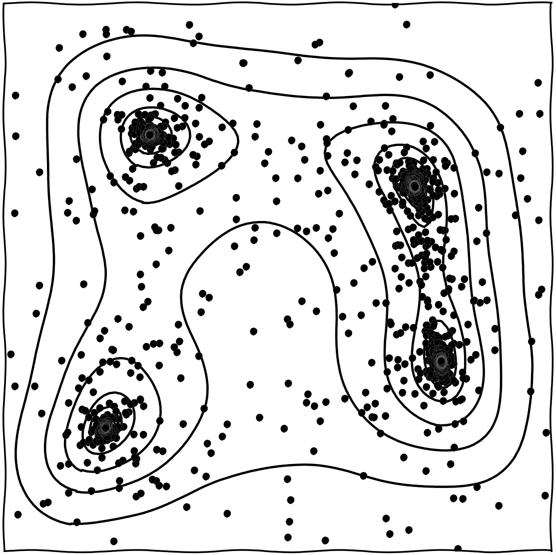
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



Nested sampling: numerical Lebesgue integration

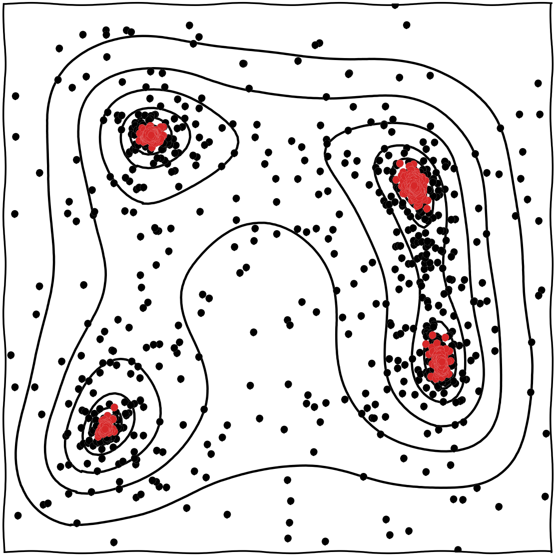
0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log \mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.

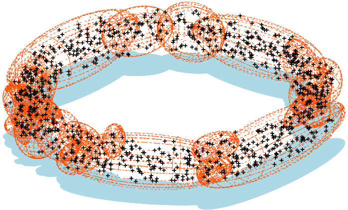


0. Start with N random samples over the space.
i. Delete outermost sample, and replace with a new random one at higher integrand value.

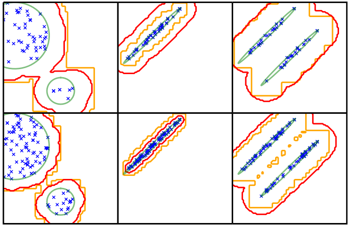
- The "live points" steadily contract around the peak(s) of the function.
- Discarded "dead points" can be weighted to form posterior, prior, or anything in between.
- Estimates the **density of states** and **partition function** $\log\mathcal{Z}(\beta)$.
- The evolving ensemble of live points allows:
  - implementations to self-tune,
  - exploration of multimodal functions,
  - global and local optimisation.



MultiNest [0809.3437]

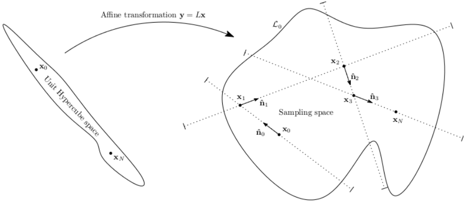


UltraNest [2101.09604]

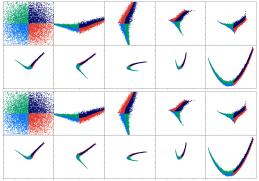


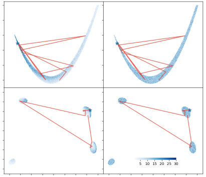
nautilus [2306.16923]

PolyChord [1506.00171]



NeuralNest [1903.10860]



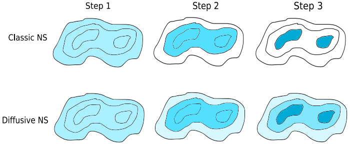


nessai [2102.11056]

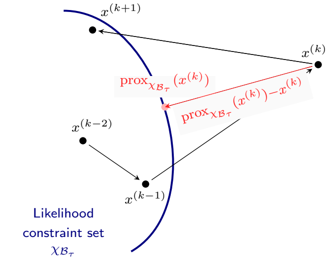
nora [2305.19267]

jaxnest [2012.15286]

DNest [1606.03757]



ProxNest [2106.03646]

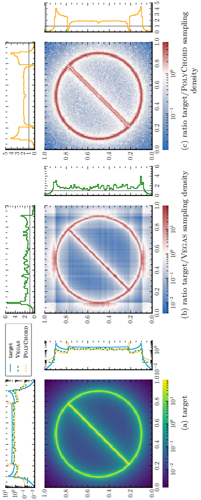


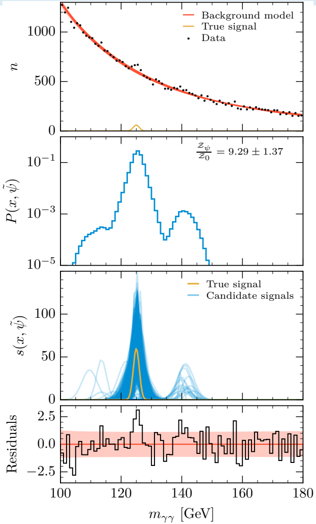
dynesty [1904.02180]

David Yallup

PDRA

- Nested sampling for cross section computation/event generation
$$\sigma = \int_\Omega d\Phi |\mathcal{M}|^2.$$
- Nested sampling can explore the phase space Ω and compute integral blind with comparable efficiency to HAAG/RAMBO [2205.02030].
- Bayesian sparse reconstruction [1809.04598] applied to bump hunting allows evidence-based detection of signals in phenomenological backgrounds [2211.10391].





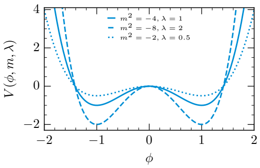
Lattice field theory

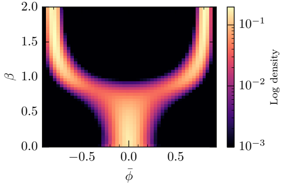
Applications of nested sampling

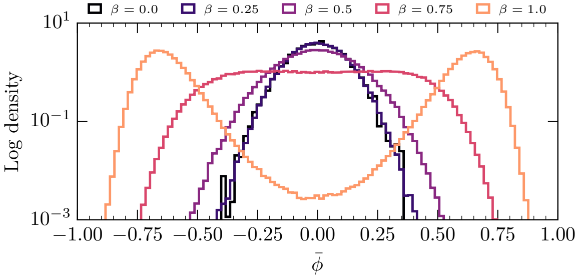
- Consider standard field theory Lagrangian:

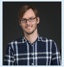
$$Z(\beta) = \int D\phi e^{-\beta S(\phi)}, \quad S(\phi) = \int dx^\mu \mathcal{L}(\phi)$$

- Discretize onto spacetime grid.
- Compute partition function
- NS unique traits:
  - Get full partition function for free
  - allows for critical tuning
  - avoids critical slowing down
- Applications in lattice gravity, QCD, condensed matter physics





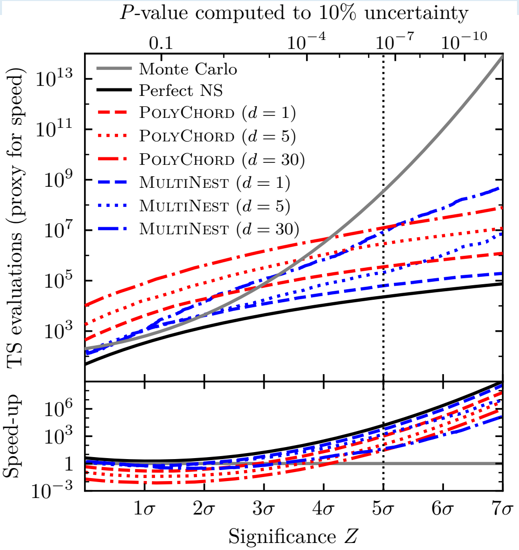




Fast estimation of small p-values [2106.02056](PRL)

Applications of nested sampling

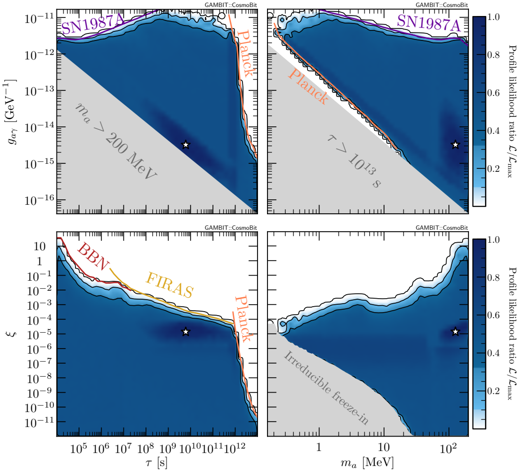
- p-value: $P(\lambda > \lambda^* | H_0)$ – probability that test statistic λ is at least as great as observed λ^*.
- Computation of a tail probability from sampling distribution of λ under H_0.
- For gold-standard 5σ, this is very expensive to simulate directly ($\sim 10^9$ by definition).
- Need insight/approximation to make efficient.
- Nested sampling is tailor-made for this, just make switch: $X \leftrightarrow p$, $\mathcal{L} \leftrightarrow \lambda$, $\theta \leftrightarrow x$.
- The only real conceptual shift is switching the integrator from parameter- to data-space.



Quantification of fine tuning [2101.00428] [2205.13549]

Applications of nested sampling

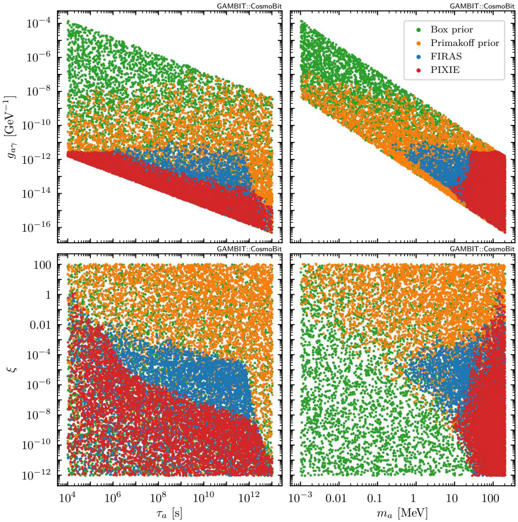
- Example: Cosmological constraints on decaying axion-like particles [2205.13549]. (Also vary cosmology, τ_n and nuisance params)
- Data: CMB, BBN, FIRAS, SMM, BAO.
- Standard profile likelihood fit shows ruled out regions and best-fit point.
- Nested sampling scan:
  - Quantifies amount of parameter space ruled out with Kullback-Liebler divergence $\mathcal{D}_{\mathrm{KL}}$.
  - Identifies best fit region as statistically irrelevant from information theory/Bayesian.
  - No evidence for decaying ALPs. Fit the data equally well: but more constrained parameters create Occam penalty.



Quantification of fine tuning [2101.00428] [2205.13549]

Applications of nested sampling

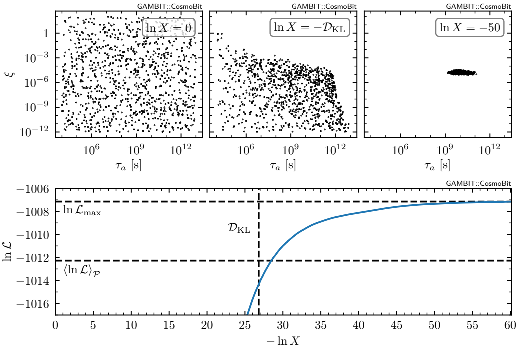
- Example: Cosmological constraints on decaying axion-like particles [2205.13549]. (Also vary cosmology, τ_n and nuisance params)
- Data: CMB, BBN, FIRAS, SMM, BAO.
- Standard profile likelihood fit shows ruled out regions and best-fit point.
- Nested sampling scan:
  - Quantifies amount of parameter space ruled out with Kullback-Liebler divergence $\mathcal{D}_{\mathrm{KL}}$.
  - Identifies best fit region as statistically irrelevant from information theory/Bayesian.
  - No evidence for decaying ALPs. Fit the data equally well: but more constrained parameters create Occam penalty.



Quantification of fine tuning [2101.00428] [2205.13549]

Applications of nested sampling

- Example: Cosmological constraints on decaying axion-like particles [2205.13549]. (Also vary cosmology, τ_n and nuisance params)
- Data: CMB, BBN, FIRAS, SMM, BAO.
- Standard profile likelihood fit shows ruled out regions and best-fit point.
- Nested sampling scan:
  - Quantifies amount of parameter space ruled out with Kullback-Liebler divergence $\mathcal{D}_{\mathrm{KL}}$.
  - Identifies best fit region as statistically irrelevant from information theory/Bayesian.
  - No evidence for decaying ALPs. Fit the data equally well: but more constrained parameters create Occam penalty.



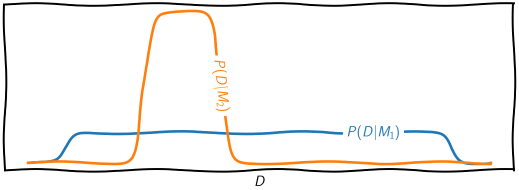
Model comparison $\mathcal{Z} = P(D|M)$

- Bayesian model comparison allows mathematical derivation of key philosophical principles.

Viewed from data-space D:

Popper's falsificationism

- Prefer models that make bold predictions.
- if proven true, model more likely correct.



- Falsificationism comes from normalisation

Viewed from parameter-space θ:

Occam's razor

- Models should be as simple as possible
- . . . but no simpler

- Occam's razor equation:

$$\log \mathcal{Z} = \langle \log \mathcal{L} \rangle_{\mathcal{P}} - \mathcal{D}_{\mathrm{KL}}$$

- "Occam penalty": KL divergence between prior π and posterior $\mathcal{P}$.

$$\mathcal{D}_{\mathrm{KL}} \sim \log \frac{\text{Prior volume}}{\text{Posterior volume}}$$

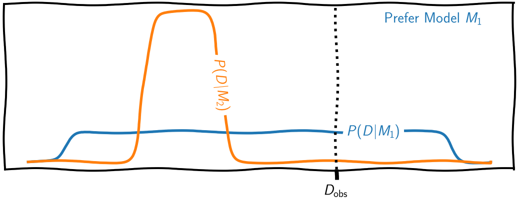
Model comparison $\mathcal{Z} = P(D|M)$

- Bayesian model comparison allows mathematical derivation of key philosophical principles.

Viewed from data-space D:

Popper's falsificationism

- Prefer models that make bold predictions.
- if proven true, model more likely correct.



- Falsificationism comes from normalisation

Viewed from parameter-space θ:

Occam's razor

- Models should be as simple as possible
- . . . but no simpler

- Occam's razor equation:

$$\log \mathcal{Z} = \langle \log \mathcal{L} \rangle_{\mathcal{P}} - \mathcal{D}_{\mathrm{KL}}$$

- "Occam penalty": KL divergence between prior π and posterior $\mathcal{P}$.

$$\mathcal{D}_{\mathrm{KL}} \sim \log \frac{\text{Prior volume}}{\text{Posterior volume}}$$

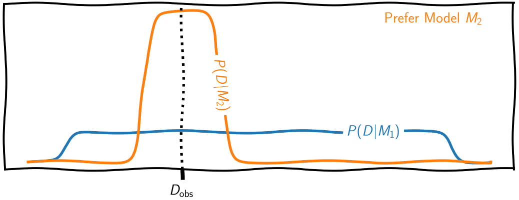
Model comparison $\mathcal{Z} = P(D|M)$

- Bayesian model comparison allows mathematical derivation of key philosophical principles.

Viewed from data-space D:

Popper's falsificationism

- Prefer models that make bold predictions.
- if proven true, model more likely correct.



- Falsificationism comes from normalisation

Viewed from parameter-space θ:

Occam's razor

- Models should be as simple as possible
- ... but no simpler

- Occam's razor equation:

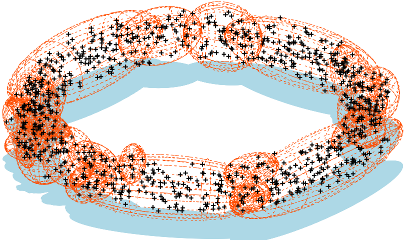
$$\log \mathcal{Z} = \langle \log \mathcal{L} \rangle_{\mathcal{P}} - \mathcal{D}_{\mathrm{KL}}$$

- "Occam penalty": KL divergence between prior π and posterior $\mathcal{P}$.

$$\mathcal{D}_{\mathrm{KL}} \sim \log \frac{\text{Prior volume}}{\text{Posterior volume}}$$

Conclusions

github.com/handley-lab



- Nested sampling is a multi-purpose numerical tool for:
  - Numerical integration $\int f(x)dV$,
  - Exploring/scanning/optimising *a priori* unknown functions,
  - Quantifying fine-tuning with Bayesian theory
- It is applied widely across cosmology & particle physics.
- It's unique traits as the only numerical Lebesgue integrator mean with compute it will continue to grow in importance.