Nested sampling: powering next-generation Bayesian inference tools

for cosmology, particle physics and beyond

Will Handley
<wh260@cam.ac.uk>

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

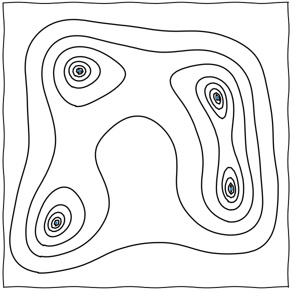
14th December 2022

What is Nested Sampling?

- Nested sampling is a radical, multi-purpose numerical tool.
- Given a (scalar) function f with a vector of parameters θ, it can be used for:

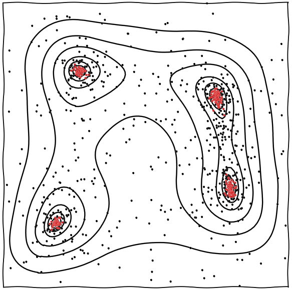
Optimisation

$$\theta_{\max} = \max_{\theta} f(\theta)$$



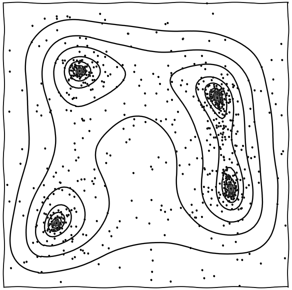
Exploration

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



Integration

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



General setting

- Integration is a fundamental concept in physics, statistics and data science:

Partition functions

$$Z(\beta) = \int e^{-\beta H(q,p)} dqdp$$

Path integrals

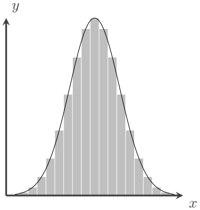
$$\Psi = \int e^{iS} \mathcal{D}x$$

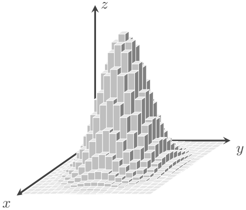
Bayesian marginals

$$\mathcal{Z}(D) = \int \mathcal{L}(D|\theta)\pi(\theta)d\theta$$

- Need numerical tools if analytic solution unavailable.
- High-dimensional numerical integration is hard.
- Riemannian strategy estimates volumes geometrically:

$$\int f(x)d^nx \approx \sum_i f(x_i)\Delta V_i \sim \mathcal{O}(e^n)$$



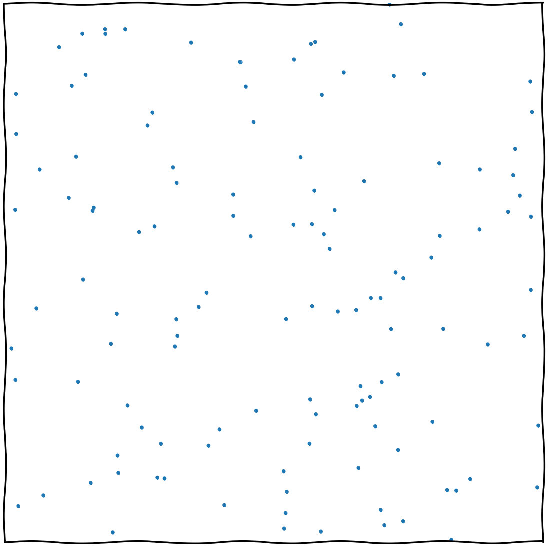


- Curse of dimensionality $\Rightarrow$ exponential scaling.
- Nested sampling integrates **probabilistically**.

The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

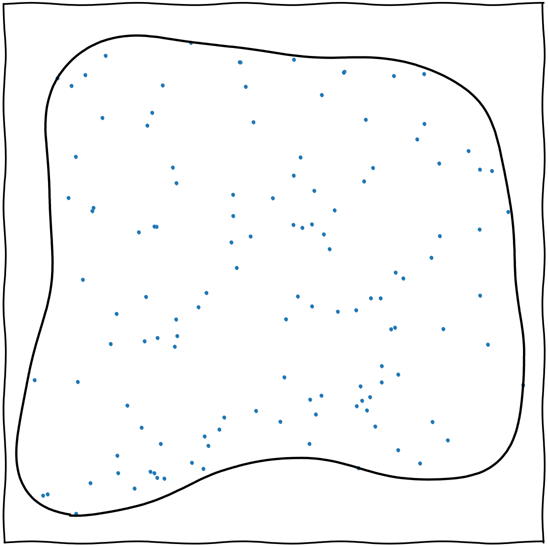
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

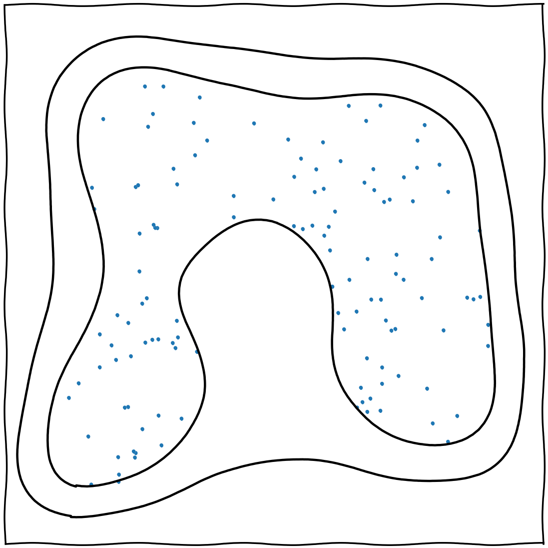
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

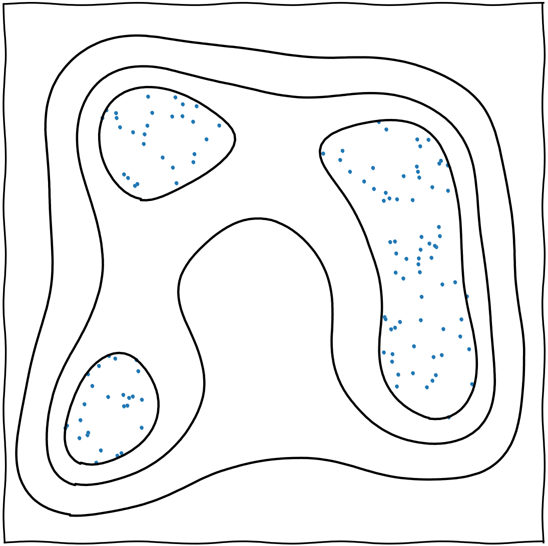
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

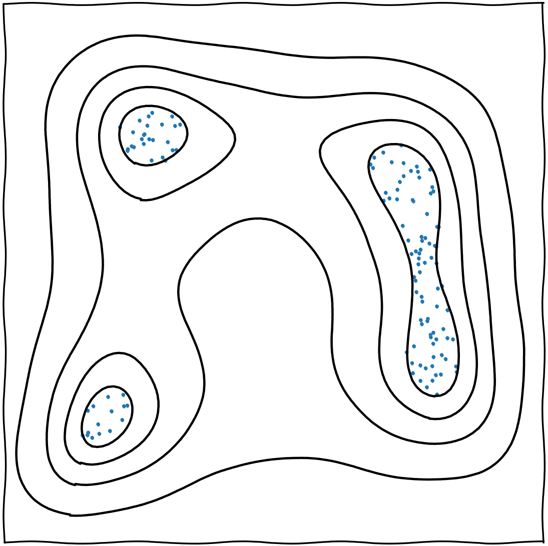
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

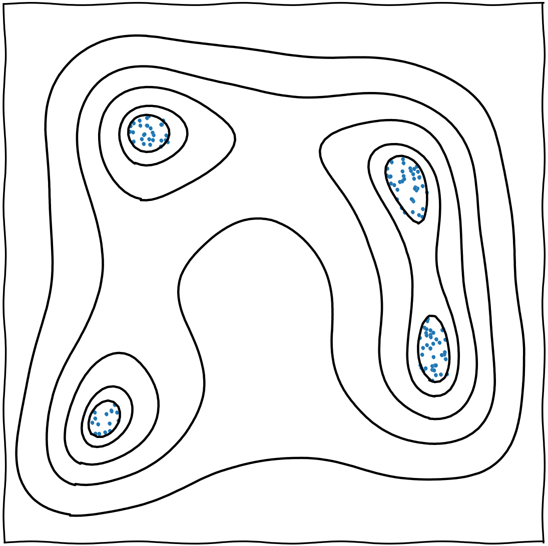
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

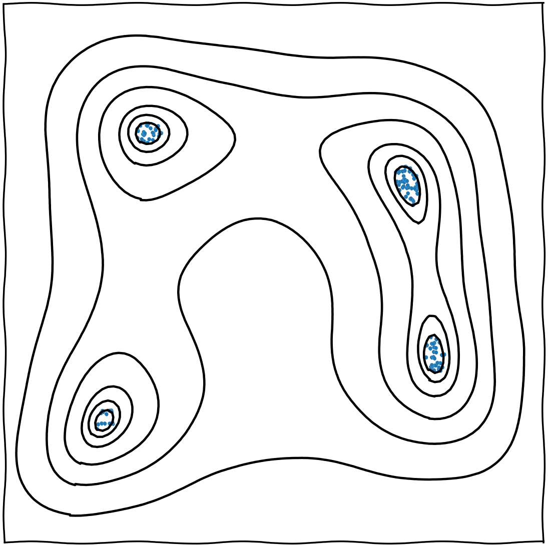
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

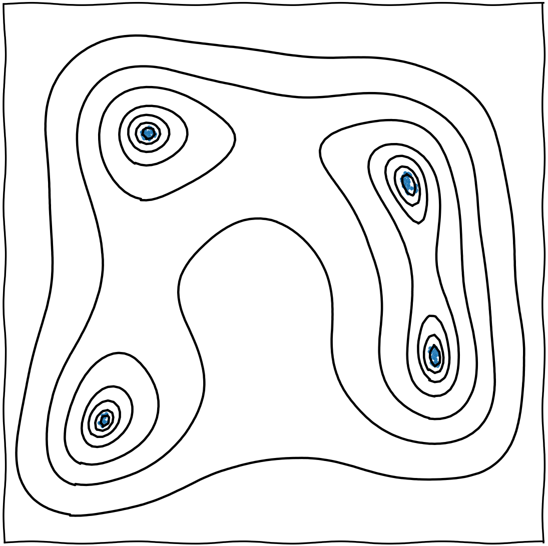
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

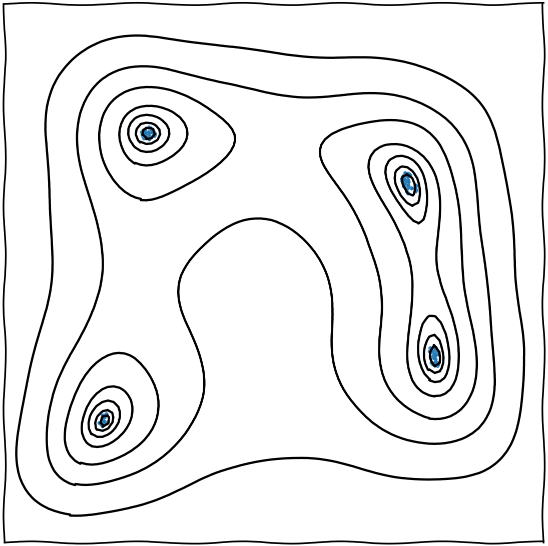
$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-i/n}$$



The nested sampling meta-algorithm

- Start with n random samples over the space.
- 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.
- We can use this evolution to estimate volume *probabilistically*.
- At each iteration, the contours contract by $\sim \frac{1}{n} \pm \frac{1}{n}$ of their volume.
- This is an exponential contraction, so

$$\sum_i f(x_i)\Delta V_i, \qquad V_i = V_0 e^{-(i\pm\sqrt{i})/n}$$

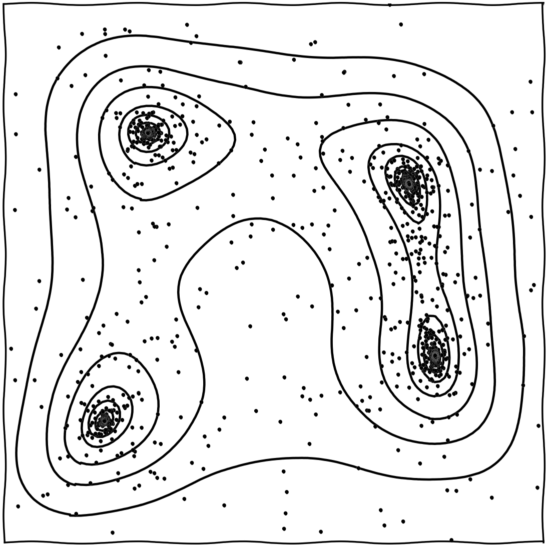


The nested sampling meta-algorithm

- At the end, one is left with a set of discarded "dead" points.
- Nested sampling estimates the **density of states** and calculates partition functions

$$Z(\beta) = \sum_i f(x_i)^\beta \Delta V_i$$

- The evolving ensemble of live points allows:
  - implementations to self-tune
  - exploration of multimodal functions
  - global and local optimisation
- For this kind of numerical, generic, high-dimensional integration, it is the only game in town.



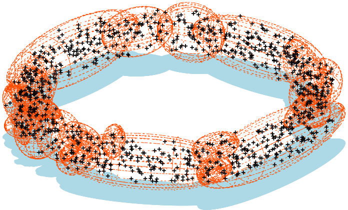
Sampling from a hard likelihood constraint

> *"It is not the purpose of this introductory paper to develop the technology of navigation within such a volume. We merely note that exploring a hard-edged likelihood-constrained domain should prove to be neither more nor less demanding than exploring a likelihood-weighted space."*
>
> — *John Skilling*

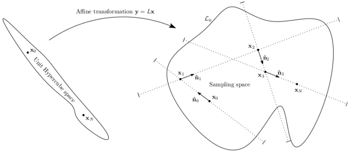
- A large fraction of the work in nested sampling to date has been in attempting to implement a hard-edged sampler in the nested sampling meta-algorithm.
- `https://projecteuclid.org/euclid.ba/1340370944`.
- There has also been much work beyond this (focus of this talk).

Implementations of Nested Sampling [arxiv:2205.15570](NatReview)

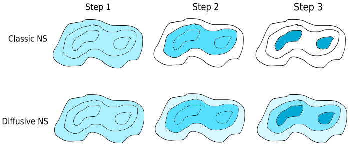
MultiNest [arxiv:0809.3437]



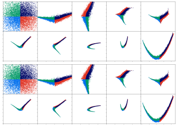
PolyChord [arxiv:1506.00171]

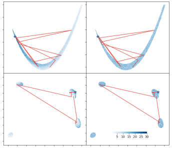


DNest [arxiv:1606.03757]



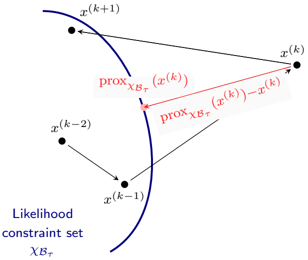
NeuralNest [arxiv:1903.10860]



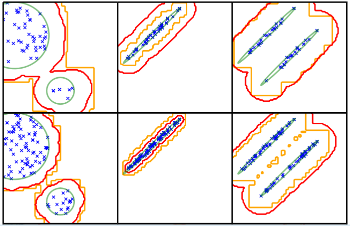


dynesty [arxiv:1904.02180]

ProxNest [arxiv:2106.03646]



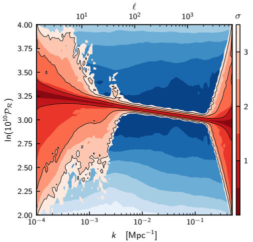
UltraNest [arxiv:2101.09604]

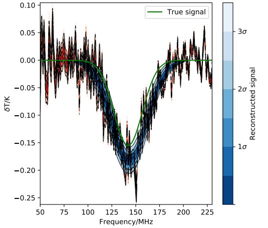


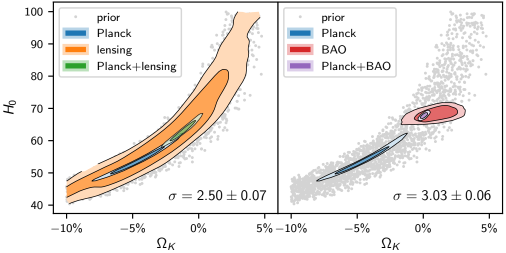
Applications of nested sampling

Cosmology

- Battle-tested in Bayesian cosmology on
  - Parameter estimation: multimodal alternative to MCMC samplers.
  - Model comparison: using integration to compute the Bayesian evidence
  - Tension quantification: using deep tail sampling and suspiciousness computations.
- Plays a critical role in major cosmology pipelines: Planck, DES, KiDS, BAO, SNe.
- The default ΛCDM cosmology is well-tuned to have Gaussian-like posteriors for CMB data.
- Less true for alternative cosmologies/models and orthogonal datasets, so nested sampling crucial.



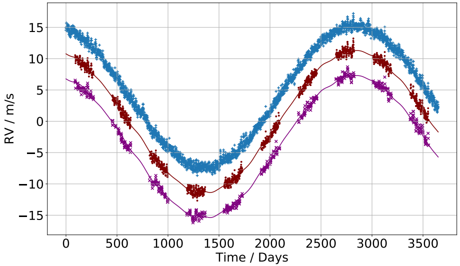


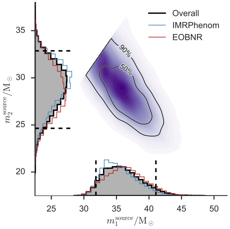


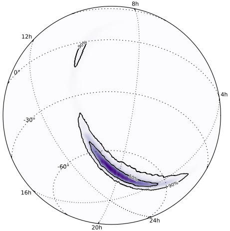
Applications of nested sampling

Astrophysics

- In exoplanets [arxiv:1806.00518]
  - Parameter estimation: determining properties of planets.
  - Model comparison: how many planets? Stellar modelling [arxiv:2007.07278].
  - exoplanet problems regularly have posterior phase transitions [arxiv:2102.03387]
- In gravitational waves
  - Parameter estimation: Binary merger properties
  - Model comparison: Modified theories of gravity, selecting phenomenological parameterisations [arxiv:1803.10210]
  - Likelihood reweighting: fast slow properties



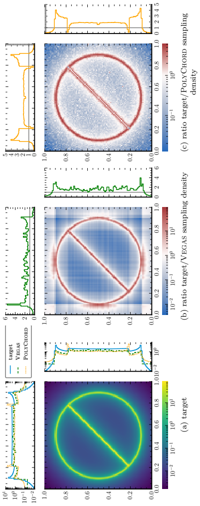


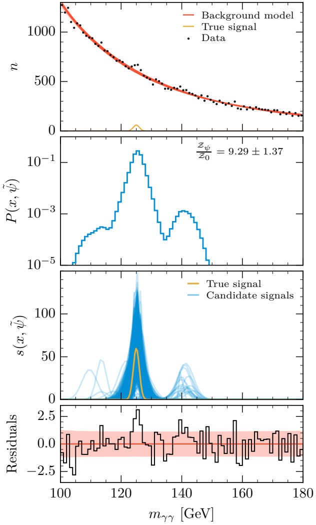


Applications of nested sampling

Particle physics

- 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 [arxiv:2205.02030].
- Bayesian sparse reconstruction [arxiv:1809.04598] applied to bump hunting allows evidence-based detection of signals in phenomenological backgrounds [arxiv:2211.10391].
- Now applying to lattice field theory, and lattice gravity Lagrangians.
- Particle statistics: fast estimation of small p-values [arxiv:2106.02056](PRL).

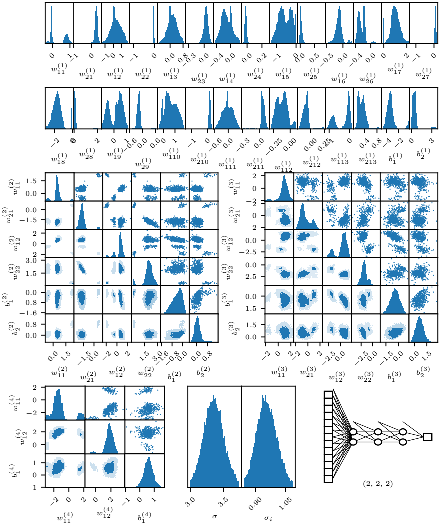




Applications of nested sampling

Machine learning

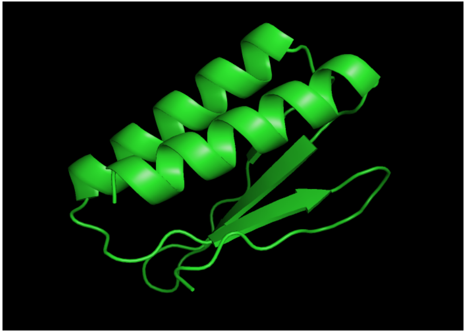
- Machine learning requires:
  - Training to find weights
  - Choice of architecture/topology/hyperparameters
- Bayesian NNs treat training as a model fitting problem
- Compute posterior of weights (parameter estimation), rather than optimisation (gradient descent)
- Use evidence to determine best architecture (model comparison), correlates with out-of-sample performance!
- Solving the full "shallow learning" problem without compromise [arxiv:2004.12211][arxiv:2211.10391].
- Promising work ongoing to extend this to transfer learning and deep nets.



Applications of nested sampling

and beyond...

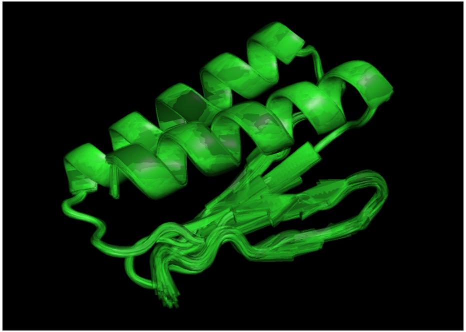
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

and beyond...

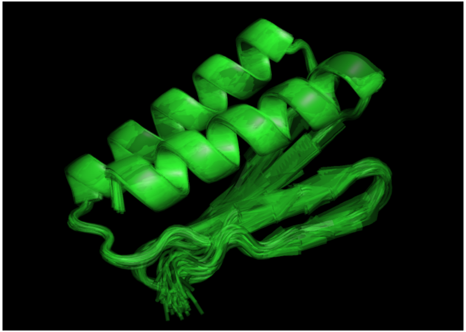
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

and beyond…

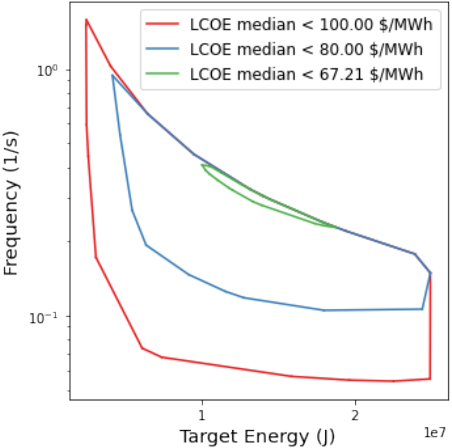
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

and beyond. . .

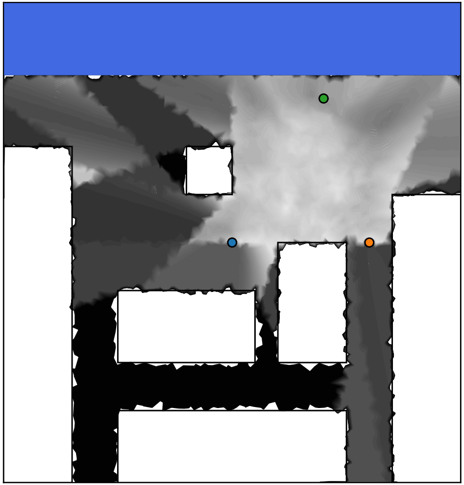
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

and beyond...

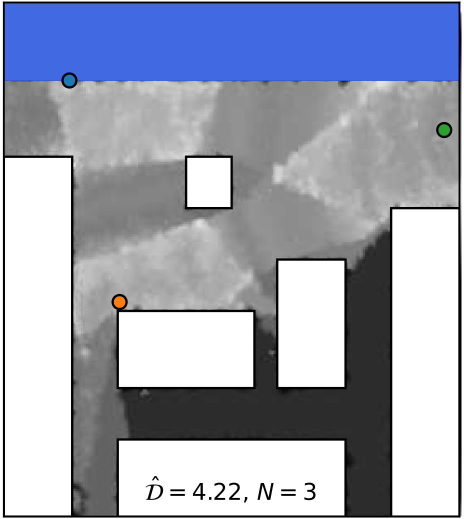
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

and beyond...

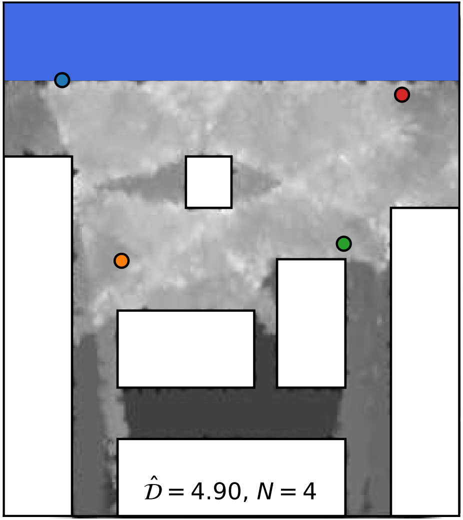
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

and beyond...

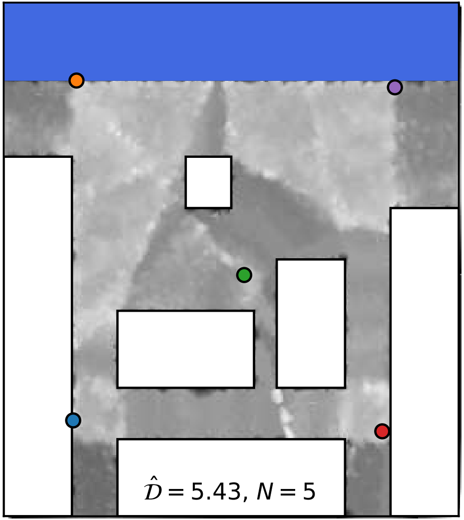
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.



Applications of nested sampling

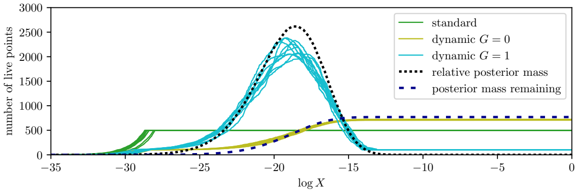
and beyond. . .

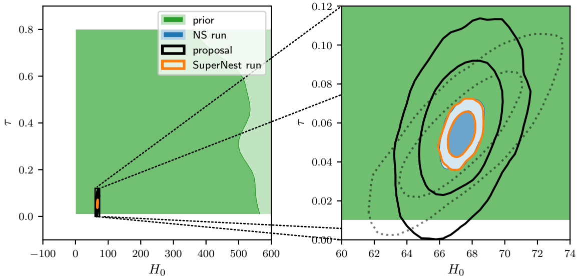
- Techniques have been spun-out (PolyChord Ltd) to:
- Protein folding
  - Navigating free energy surface.
  - Computing misfolds.
  - Thermal motion.
- Nuclear fusion reactor optimisation
  - multi-objective.
  - uncertainty propagation.
- Telecoms & DSTL research (MIDAS)
  - Optimising placement of transmitters/sensors.
  - Maximum information data acquisition strategies.

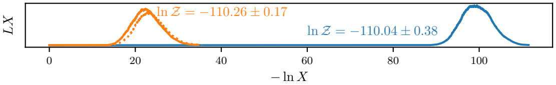


Beyond the meta-algorithm

- Dynamic nested sampling [arxiv:1704.03459]
- Unwoven nested sampling [arxiv:1703.09701]
- Accelerated nested sampling [arxiv:2212.01760]
- Precision nested sampling [arxiv:2006.03371]
- Multiobjective nested sampling
- Nested sampling with gradients?
- Reversible nested sampling?
- Transdimensional nested sampling?
- postprocessing: `anesthetic` [arxiv:1905.04768]
- crosschecking: `nestcheck` [arxiv:1804.06406]

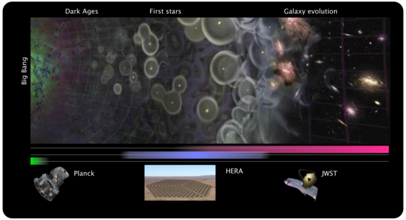


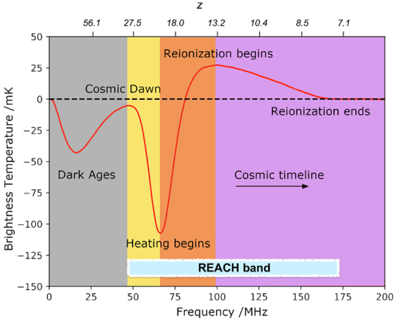


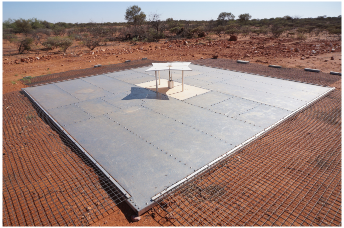


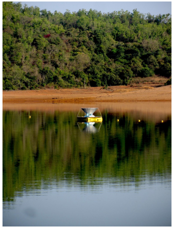
REACH: Global 21cm cosmology [arxiv:2210.07409](NatAstro)

- Imaging the universal dark ages using CMB backlight.
- 21cm hyperfine line emission from neutral hydrogen.
- Global experiments measure monopole across frequency.
- Challenge: science hidden in foregrounds $\sim 10^4\times$signal.
- Lead data analysis team (REACH first light in January)
- Nested sampling woven in from the ground up (calibrator, beam modelling, signal fitting, likelihood selection).
- All treated as parameterised model comparison problems.



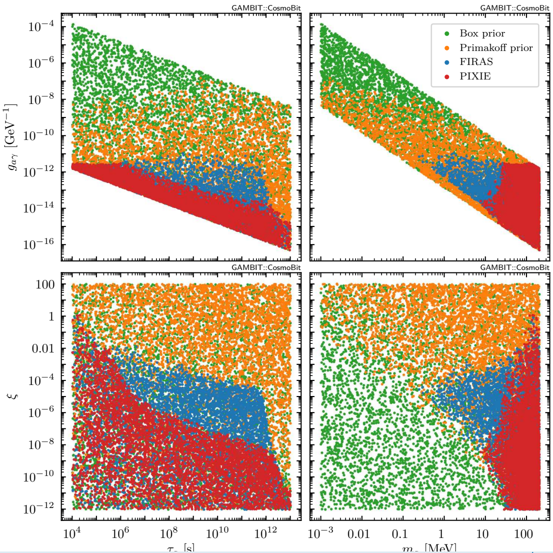






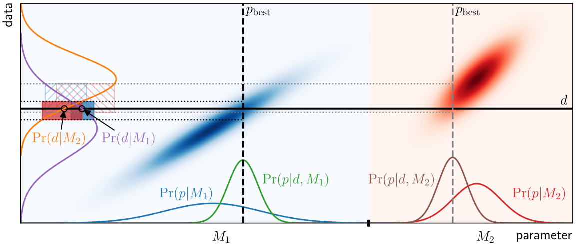
GAMBIT: combining particle physics & cosmological data

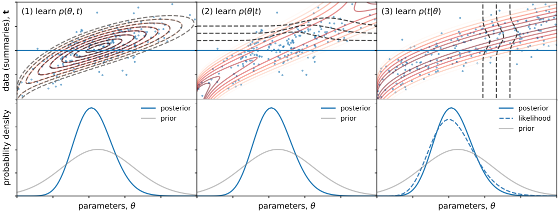
- Multinational team of particle physicists, cosmologists and statisticians.
- Combine cosmological data, particle colliders, direct detection, & neutrino detectors in a statistically principled manner [arxiv:2205.13549].
- Lead Cosmo/Dark Matter working group [arxiv:2009.03286].
- Nested sampling used for global fitting, and fine-tuning quantification [arxiv:2101.00428]



Likelihood-free inference (& nested sampling)

- How do you do inference if you don't know the likelihood $P(D|\theta)$?
  - e.g. if you can simulate a disease outbreak, how can you infer a posterior on R_0, or select the most predictive model?
- If you can forward simulate/model $\theta \rightarrow D$, then you have an implicit likelihood.
- LFI aims to (machine-)*learn* the likelihood from carefully chosen training data $\{(\theta, D)\}$.
- Nested sampling has much to offer
  - truncation strategies
  - evidence driven compression
  - marginalised machine learning
- In my view, LFI represents the future of inference – in twenty years time this will be as well-used as MCMC techniques are today.





CosmoTension

Resolving cosmological tensions with diverse data, novel theories and Bayesian machine learning

- ERC grant $\Rightarrow$ UKRI Frontier, commencing 2023.
- Funds 3 PDRAs and 4 PhDs over 5 years.
- Research programme centered around combining novel theories of gravity, Boltzmann solvers [arxiv:1906.01421], reconstruction [arxiv:1908.00906], nested sampling & likelihood free inference.
- Aims to disentangle cosmological tensions H_0, σ_8, Ω_K with next-generation data analysis techniques.