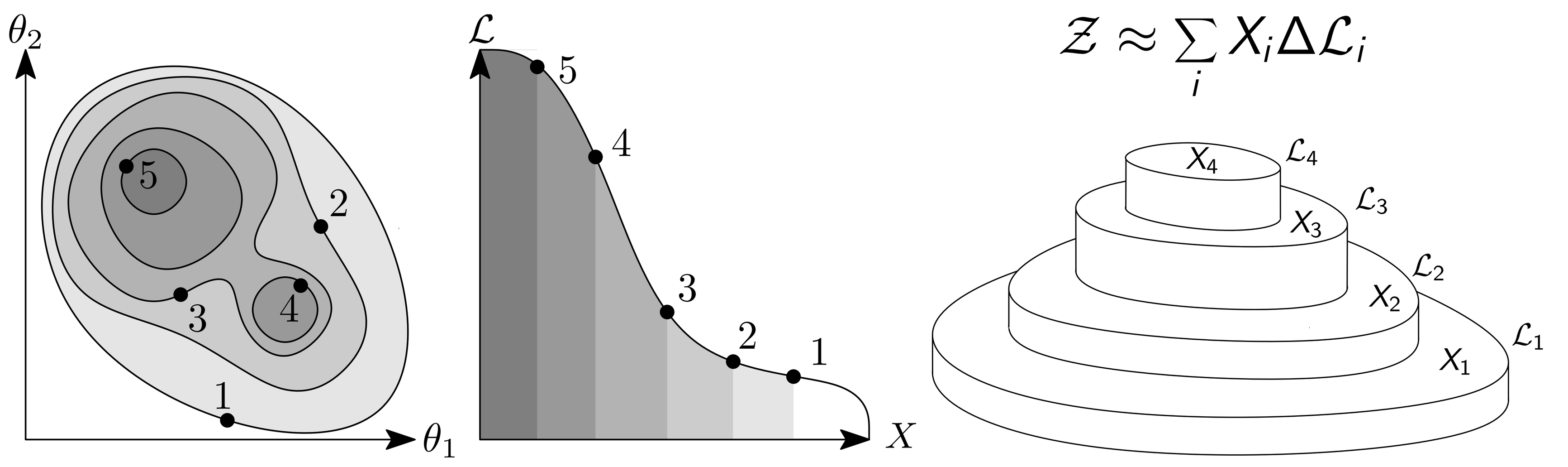
PolyChord: next generation nested sampling

Will Handley <wh260@cam.ac.uk>

Kavli Institute for Cosmology · Cavendish Laboratory · Gonville & Caius College, Cambridge

Nested sampling is an efficient and robust Bayesian inference tool for data science and machine learning. In contrast to techniques such as Metropolis-Hastings or Hamiltonian Monte Carlo, it is capable of computing the Bayesian evidence, does not require gradients and has little-to-no problem-specific tuning. Given an objective function, nested sampling navigates and optimises athermally, simultaneously sampling the posterior and computing the evidence, the density of states and the full partition function. With access to evidences one can perform Bayesian model comparison, which is applicable to an extremely wide variety of scientific and machine learning problems. `PolyChord` represents the state-of-the-art in high-performance nested sampling.

The nested sampling meta algorithm



John Skilling's [1] nested sampling meta algorithm, given a likelihood $\mathcal{L}(\theta)$ and prior $\pi(\theta)$:

S_0:	Generate n_{live} samples θ drawn from the prior $\theta \sim \pi$.
S_{n+1}:	Delete lowest likelihood $\mathcal{L}_*$ sample in S_n, and replace with $\theta \sim \pi : \mathcal{L}(\theta) > \mathcal{L}_*$.

With the set of deleted points one can compute the evidence $\mathcal{Z}$ and posterior weights $\mathcal{P}$

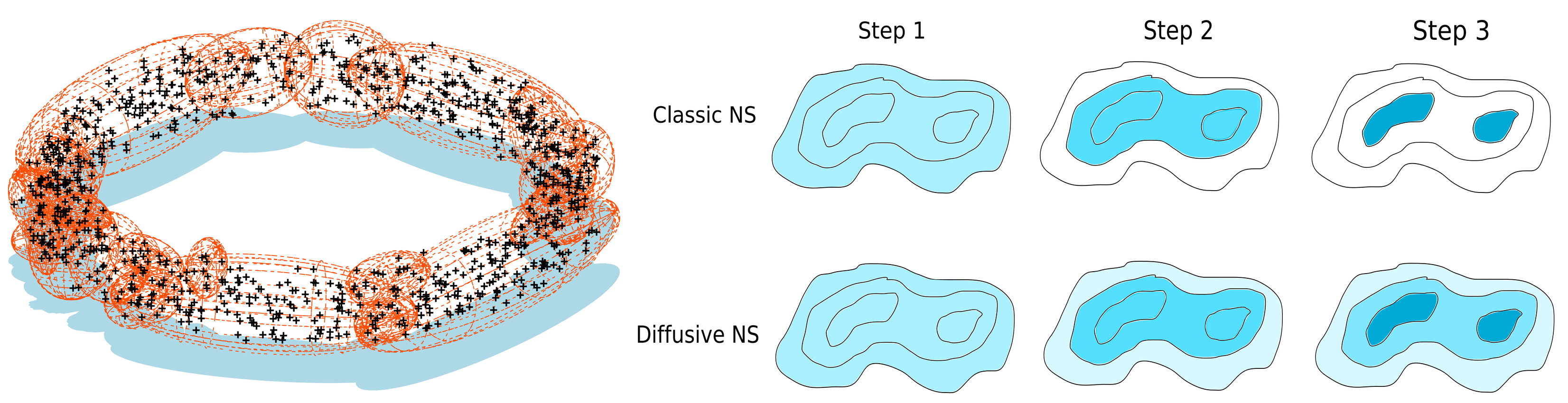
$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)\mathrm{d}\theta \approx \sum_i \mathcal{L}_i \times (X_{i-1} - X_i), \qquad \mathcal{P}_i = \mathcal{L}_i \times (X_{i-1} - X_i)$$

where X_i is fraction of prior volume enclosed by contour $\mathcal{L}_i$. We don't know these volumes exactly, but statistically estimate them given that the contours contract by $\sim \frac{1}{n}$ each time

$$X_0 = 1, \qquad X_{i+1} = t \times X_i, \qquad P(t) = nt^{n-1} \quad \Rightarrow \quad t \approx \frac{n}{n+1}$$

Nested sampling is terminated when the live points reach the posterior bulk $\mathcal{Z}_{\mathrm{live}} \ll \mathcal{Z}_{\mathrm{dead}}$.

Historical implementations



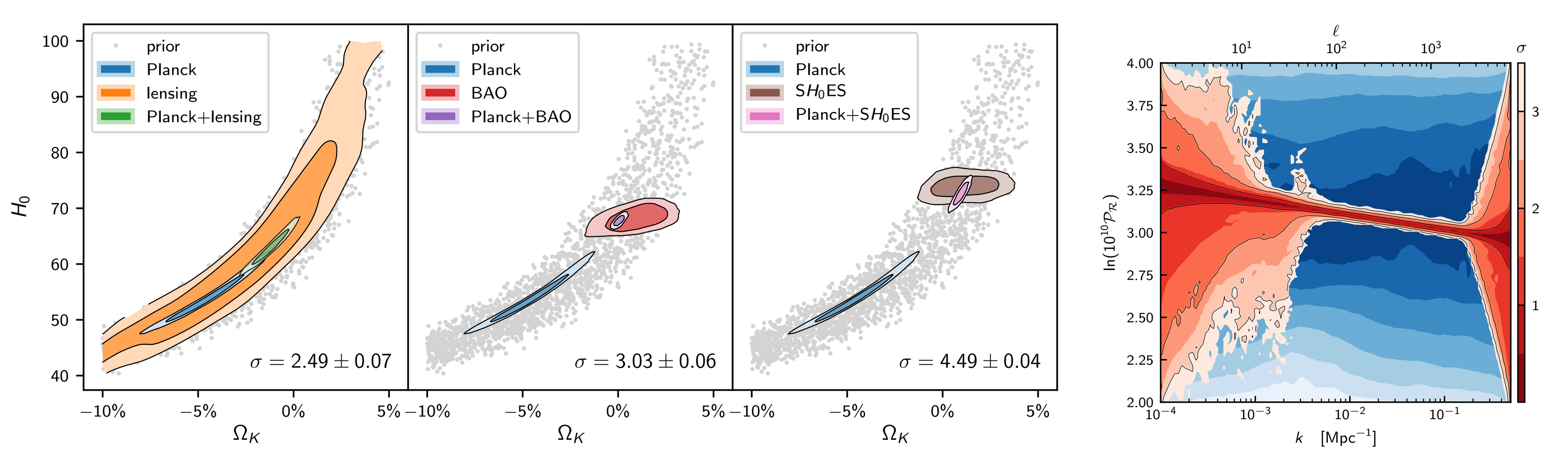
The meta-algorithm does not instruct on how to draw $\theta \sim \pi : \mathcal{L}(\theta) > \mathcal{L}_*$. The procedure to perform this is the distinguishing factor between implementations in the literature:

Simple MCMC	[2]	Metropolis-Hastings sampling of tabletop likelihood
`MultiNest`	[3]	Ellipsoidal rejection sampling
Galilean NS	[4]	Gradient-based reflection
Hamiltonian NS	[5]	Gradient-based navigation
Neural Nest	[6]	Neural network latent space whitening
`dynesty`	[7]	python implementation of dynamic nested sampling

The live points at each iteration are used to self-tune a given algorithm, for example by whitening the space and performing clustering analysis. Nested sampling is limited by:

1. The convergence time it takes to compress from prior to the posterior bulk
2. The efficiency with which a new live point can be generated at each iteration

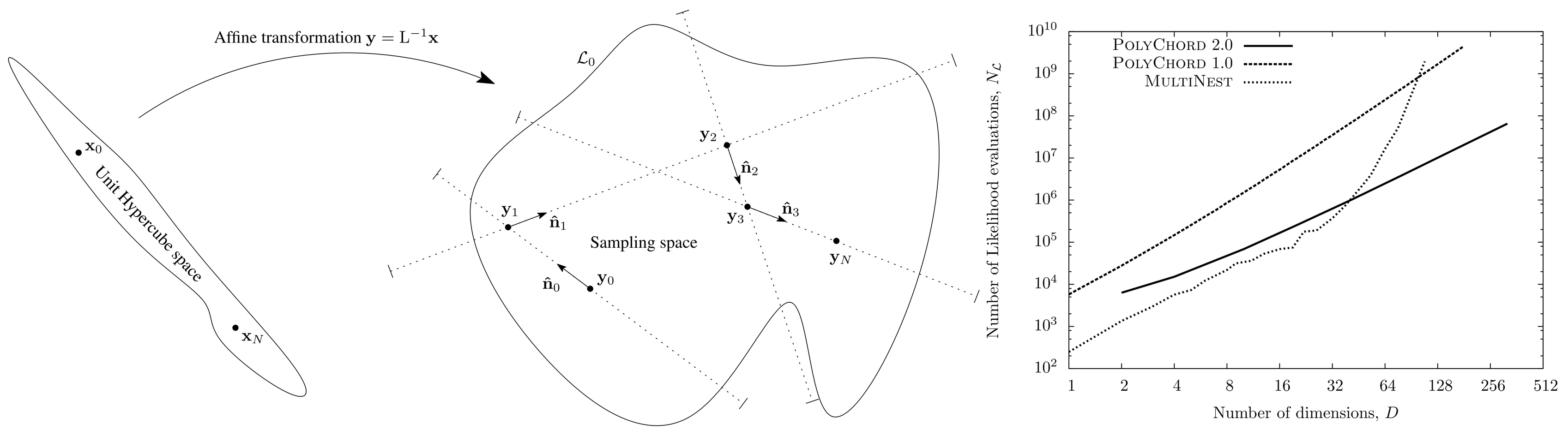
Scientific applications



Nested sampling has been applied to a wide variety of scientific applications, including Cosmology [8, 9], exoplanets [10], particle physics [11], geophysics [12] and Likelihood free inference [13]. The anesthetic python package [14] provides a unified interface for processing the outputs of nested sampling algorithms.

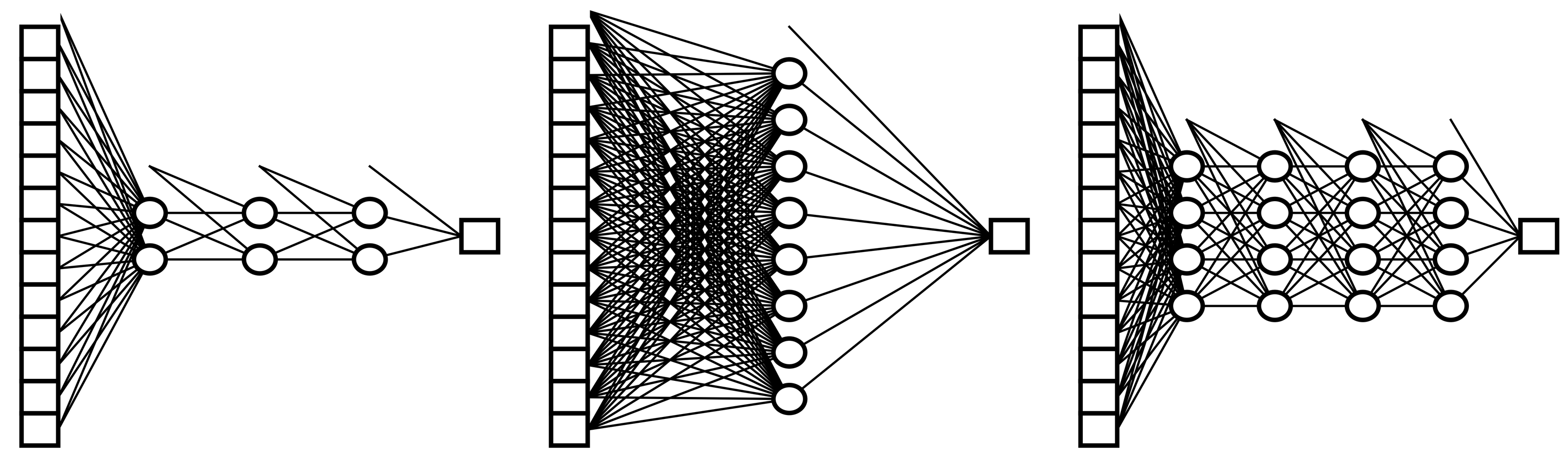
Advances in the field since the original paper [1] include dynamic NS [15], posterior repartitioning [12], diffusive NS [16], superposition enhanced NS [17], diffusive NS [16] and `BAMBI` [18].

PolyChord



`PolyChord` [19, 20] represents the state-of-the-art in high-performance nested sampling. It utilises slice sampling to enable scaling to hundreds of dimensions, employs a sophisticated mode separation and tracking algorithm and is written in `FORTRAN` for maximum speed with Python & C++ interfaces. Version 2.0 represents a further step-change, capable of scaling to thousands of dimensions by removing a key inefficiency in traditional approaches.

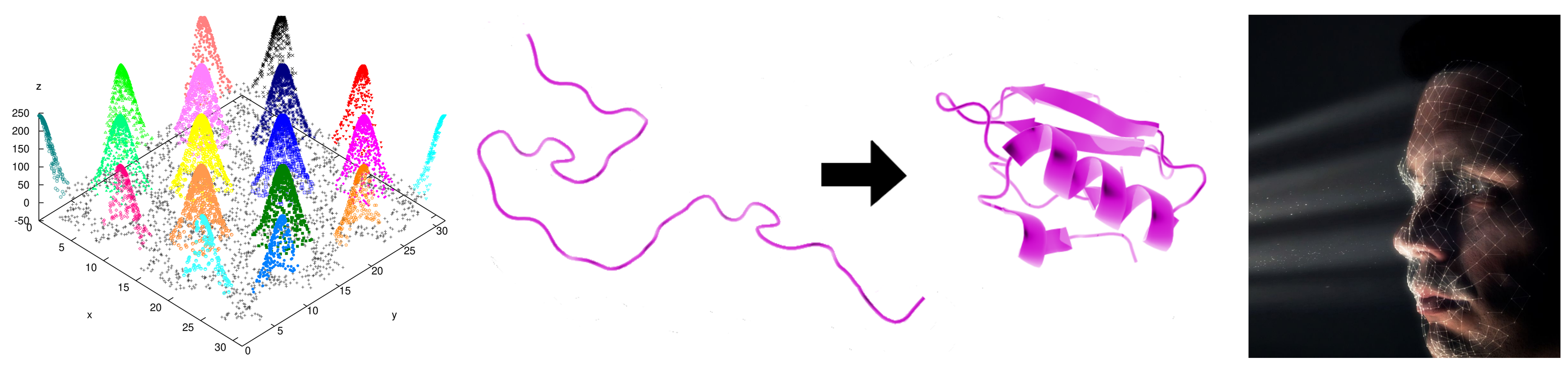
Bayesian neural networks



`PolyChord` can be used to train small Bayesian neural networks with hundreds of parameters [21]. Instead of training network parameters θ by minimising a misfit function $\chi^2_{\mathrm{train}}(\theta)$, `PolyNet` treats it as a Bayesian inference problem with $\log\mathcal{L} \propto -\chi^2_{\mathrm{train}}(\theta)$. This generates a posterior over the network weights, which can be used to give error bars on predictions. `PolyNet` also returns evidences, which can be used to measure the quality of a network architecture, and sample over an ensemble of networks. Key results:

1. Bayesian neural networks outperform traditional neural networks in their predictions.
2. The Bayesian evidence correlates with out-of-sample performance on testing data
3. Marginalising over networks using the Bayesian evidence produces robust predictors.

PolyChord Ltd



PolyChord Ltd is a spinout company working with industrial partners applying nested sampling to their data science problems (www.polychord.co.uk). It has helped win three STFC grants and Innovate UK funding to fund further postdoctoral research. Current projects range from facial recognition (AnyVision), through geophysics (Shell) to food safety (Agri-Neo). Future projects include protein folding and neural network training and design.

References

[1] John Skilling. Nested sampling for general bayesian computation. *Bayesian analysis*, 1(4):833–859, 2006.
[2] D. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford science publications. OUP Oxford, 2006.
[3] Feroz, Hobson, and Bridges. MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398(4):1601–1614, Oct 2009.
[4] F. Feroz and J. Skilling. Exploring multi-modal distributions with nested sampling. In *American IoP Conference Series*, volume 1553, pages 106–113, Aug 2013.
[5] Michael Betancourt. Nested Sampling with Constrained Hamiltonian Monte Carlo. In *American IofP Conference Series*, volume 1305, pages 165–172, Mar 2011.
[6] Adam Moss. Accelerated Bayesian inference using deep learning. *arXiv e-prints*, page arXiv:1903.10860, Mar 2019.
[7] Joshua S Speagle. dynesty: A Dynamic Nested Sampling Package for Estimating Bayesian Posteriors and Evidences. *arXiv e-prints*, page arXiv:1904.02180, Apr 2019.
[8] **W. J. Handley**, A. N. Lasenby, H. V. Peiris, and M. P. Hobson. Bayesian inflationary reconstructions from Planck 2018 data. *PRD*, 100(10):103511, Nov 2019.
[9] **Will Handley**. Curvature tension: evidence for a closed universe. *arXiv*, 1908.09139, Aug 2019.
[10] Hall, Thompson, **Handley**, and Queloz. On the Feasibility of Intense Radial Velocity Surveys for Earth-Twin Discoveries. *MNRAS*, 479(3):2968–2987, Sep 2018.
[11] Gregory D. Martinez, James McKay, Ben Farmer, Pat Scott, Elinore Roebber, Antje Putze, and Jan Conrad. Comparison of statistical sampling methods with ScannerBit, the GAMBIT scanning module. *European Physical Journal C*, 77(11):761, Nov 2017.
[12] Xi Chen, Farhan Feroz, and Michael Hobson. Bayesian automated posterior repartitioning for nested sampling. *arXiv e-prints*, page arXiv:1908.04655, Aug 2019.
[13] **W. Handley** and J. Alsing. Compromise-free Likelihood free inference. *Bayesian analysis (In preparation)*, 2020.
[14] **Will Handley**. anesthetic: nested sampling visualisation. *JOSS*, 4:1414, May 2019.
[15] E. Higson, **W. Handley**, L Hobson, and A Lasenby. Dynamic nested sampling. *Statistics and Computation*, 29(5):891–913, Sep 2019.
[16] Brendon J. Brewer and Daniel Foreman-Mackey. DNest4: Diffusive Nested Sampling in C++ and Python. *arXiv e-prints*, page arXiv:1606.03757, Jun 2016.
[17] Stefano Martiniani, Jacob D Stevenson, David J Wales, and Daan Frenkel. Superposition enhanced nested sampling. *Physical Review X*, 4(3):031034, 2014.
[18] Philip Graff, Farhan Feroz, Michael P. Hobson, and Anthony Lasenby. BAMBI: blind accelerated multimodal Bayesian inference. *MNRAS*, 421(1):169–180, Mar 2012.
[19] **W. J. Handley**, M. P. Hobson, and A. N. Lasenby. polychord: nested sampling for cosmology. *MNRAS*, 450:L61–L65, Jun 2015.
[20] **W. J. Handley**, M. P. Hobson, and A. N. Lasenby. POLYCHORD: next-generation nested sampling. *MNRAS*, 453(4):4384–4398, Nov 2015.
[21] K. Javid, **W. J. Handley**, M. P. Hobson, and L. Lasenby. Compromise-free Bayesian neural networks. *Bayesian analysis (In preparation)*, 2020.