The scaling frontier of nested sampling

How fast in nested sampling?

$$\boxed{T = T_\mathcal{L} \times n_\mathrm{live} \times f_\mathrm{sampler} \times \mathcal{D}_\mathrm{KL}}$$

How accurate is nested sampling?

$$\boxed{\sigma \approx \sqrt{\mathcal{D}_\mathrm{KL}/n_\mathrm{live}}}$$

in d dimensional parameter space:

- $T_\mathcal{L}$: likelihood eval time $\sim \mathcal{O}(d)$
- n_live: number of live points $\sim \mathcal{O}(d)$
- f_sampler: efficiency of point generation region$\sim \mathcal{O}(e^{d/d_0})$ or path$\sim \mathcal{O}(d)$
- $\mathcal{D}_\mathrm{KL}$: KL between prior and posterior $\approx \log V_\pi/V_\mathcal{P}$ $\sim \mathcal{O}(d)$

- Most attention on algorithmically improving f_sampler, but only a fraction of the story!
- In my summary, will highlight where others fit into this picture
- $\mathcal{D}_\mathrm{KL}$ appears twice, so improvements here are quadratically important.
- Gradients give you d more information.

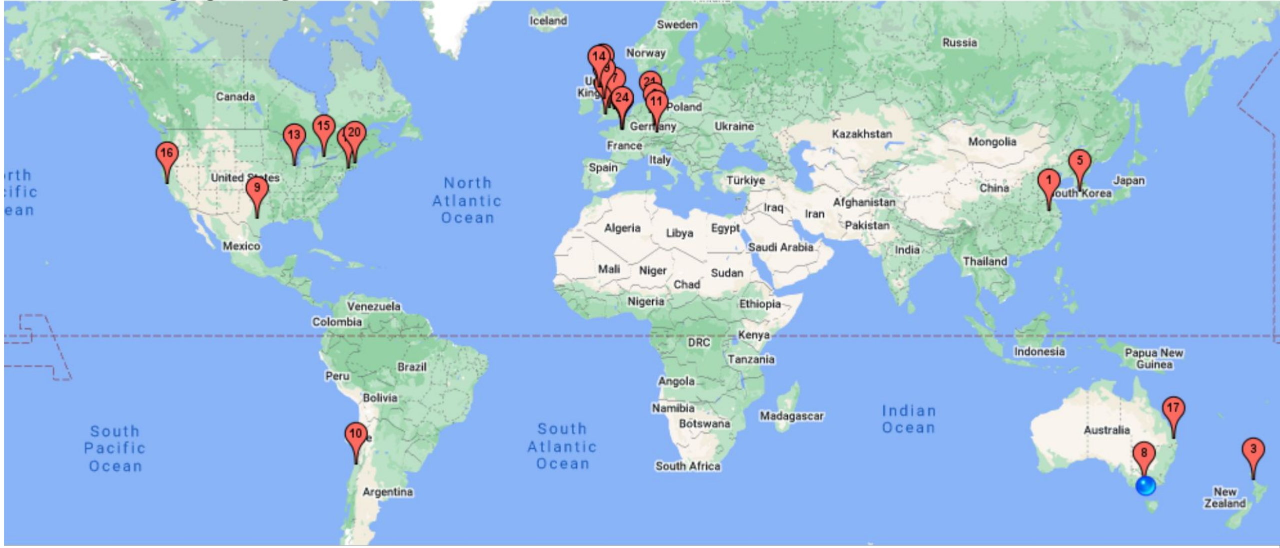
Introduction
Johannes Buchner

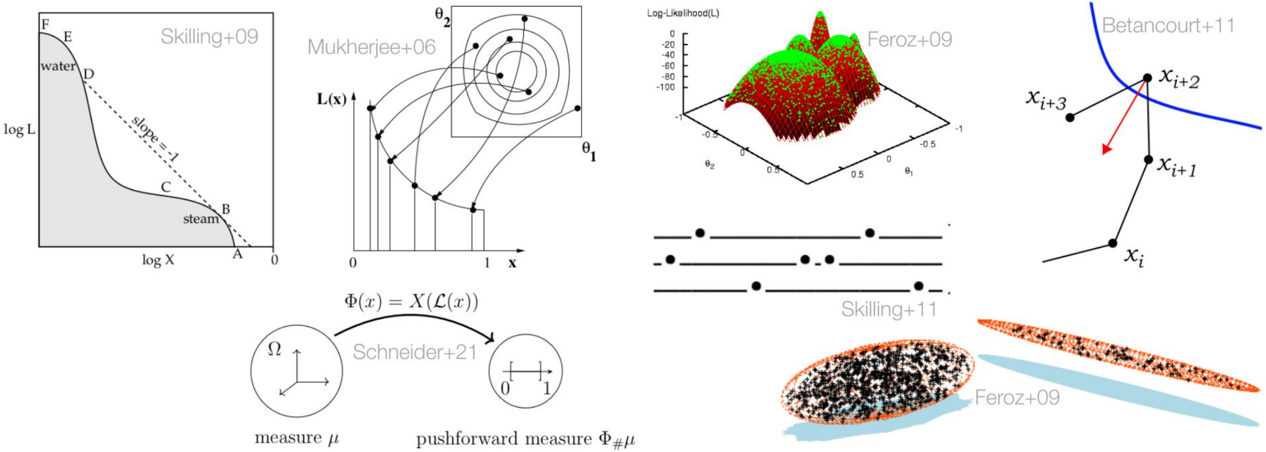
Global nested sampling!

Technical Buchner
Review: [2101.09675](https://arxiv.org/abs/2101.09675)

Pedagogical Fowlie
Review: [2205.15570](https://arxiv.org/abs/2205.15570)

Hope to have NS2024UK
as a week-long
conference in the UK





[Will Handley <wh260@cam.ac.uk>](mailto:wh260@cam.ac.uk)

Nested Sampling

(A rose by any other name)

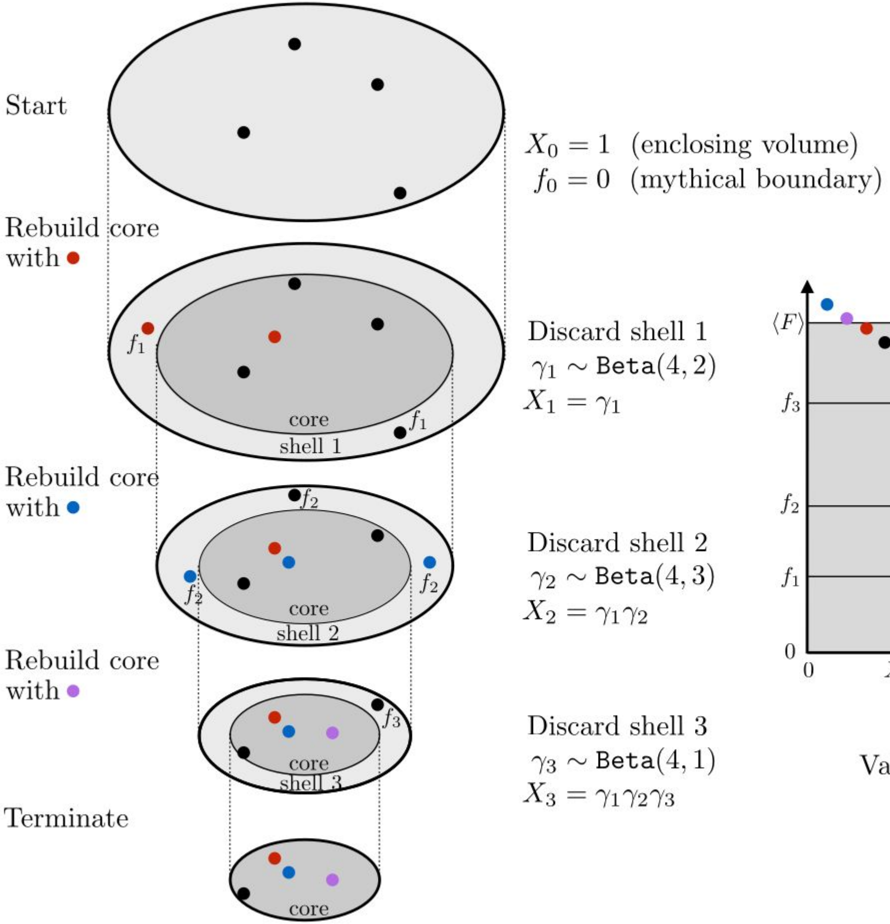
John Skilling

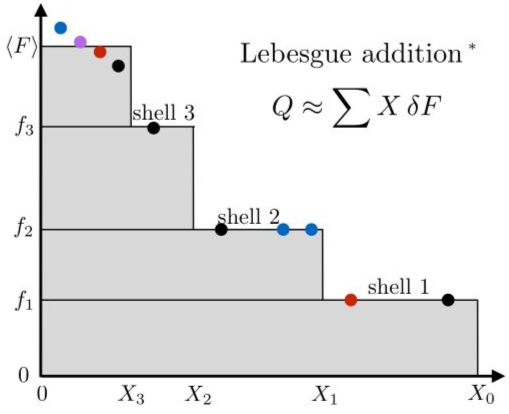
Estimating volumes* using counting (no geometry, topology, or dimensionality)

Alternative names:

- Box sampling
- Onion sampling
- Lesbesgue sampling

These meetings are very good at bringing wide-ranging people together and swapping ideas.





Value of Q is statistical. (✓)

* ref: Ning Xiang

"You will do it my way – unless your way is better, then we'll do it your way" – Steve Gull

* volumes or measures

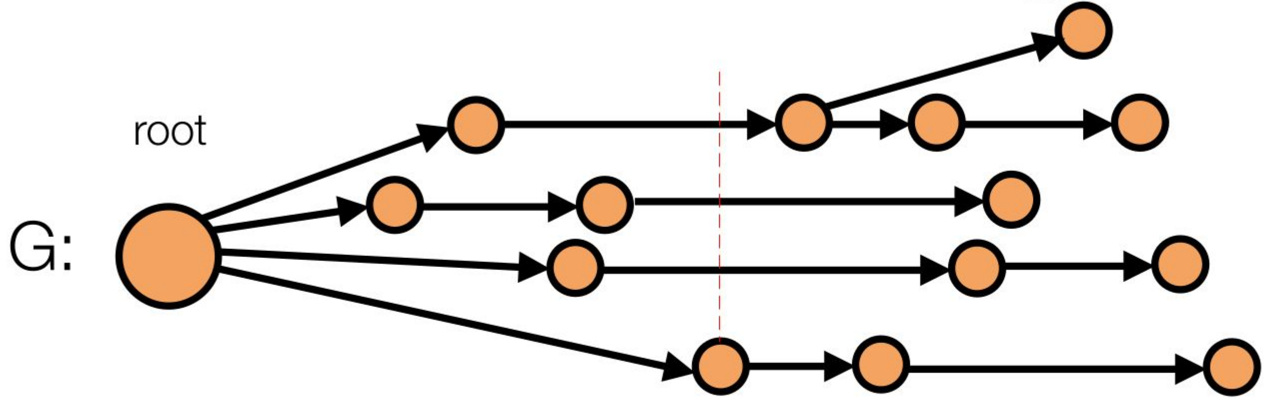
Will Handley <wh260@cam.ac.uk>

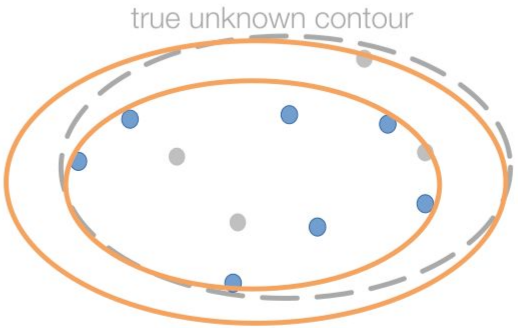
Sampling strategies
Johannes Buchner

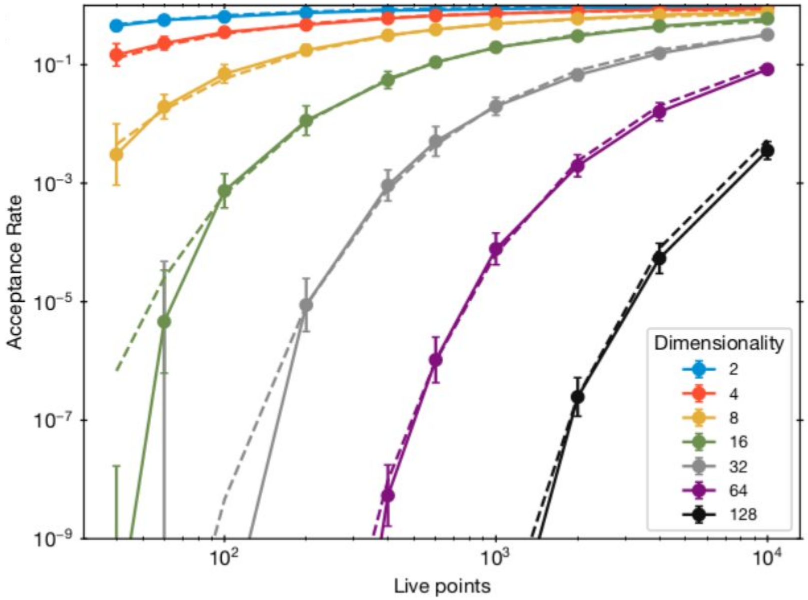
Nested sampling as a tree

Autotuning with UltraNest

"Snowballing Nested Sampling"



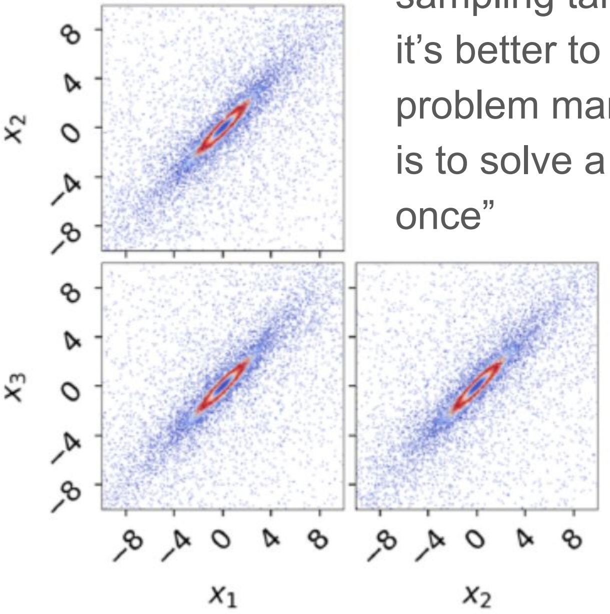




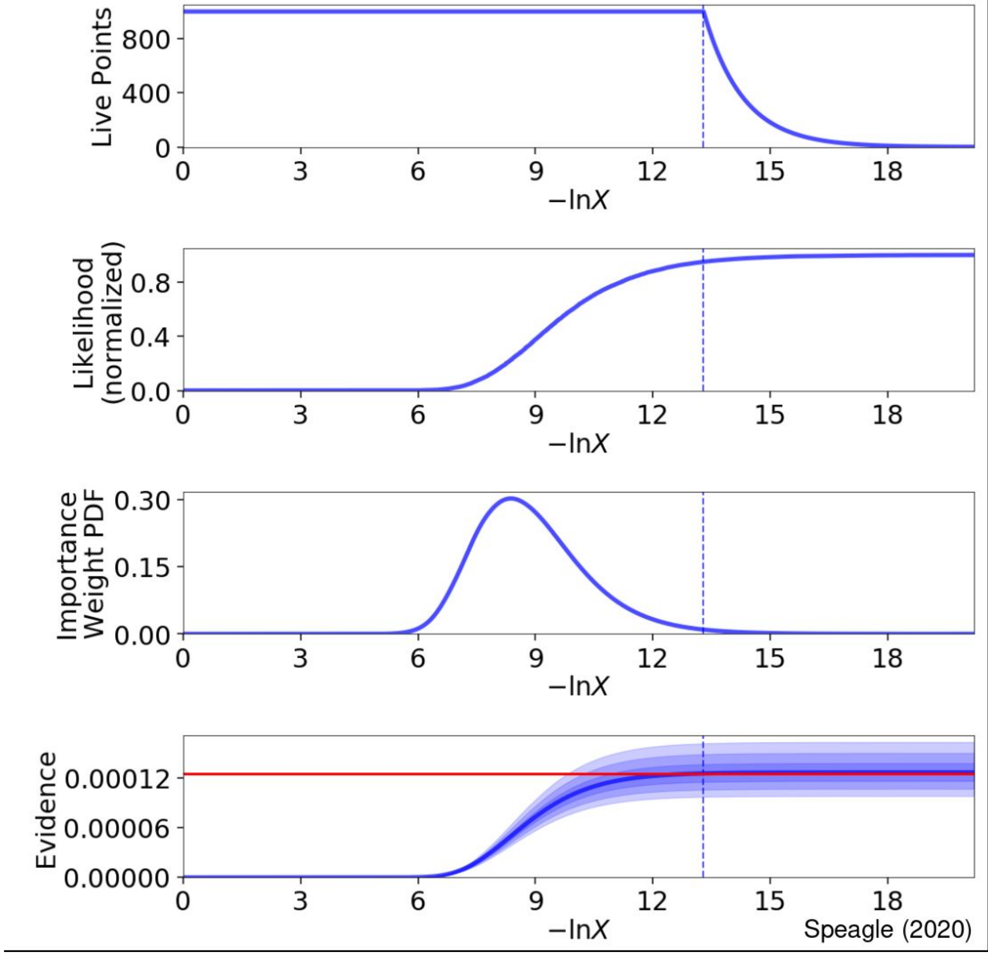
Will Handley <wh260@cam.ac.uk>

Musings on user interaction
Josh Speagle

Nested sampling codes as Bayesian ~~indoctrination~~ teaching tools.



"The bet that nested sampling takes:
it's better to solve an easy problem many times than it is to solve a hard problem once"



Will Handley <wh260@cam.ac.uk>

NS & SMC

Leah South & Nicolas Chopin

SMC is qualitatively & quantitatively similar to nested sampling

Much larger mathematical literature to

Possible to combine the two! (ANS-SMC)

Gives mathematical guarantees and unbiased estimators of Z

Waste-free SMC doesn't move all particles (more like nested sampling)

Sequential Monte Carlo (SMC)

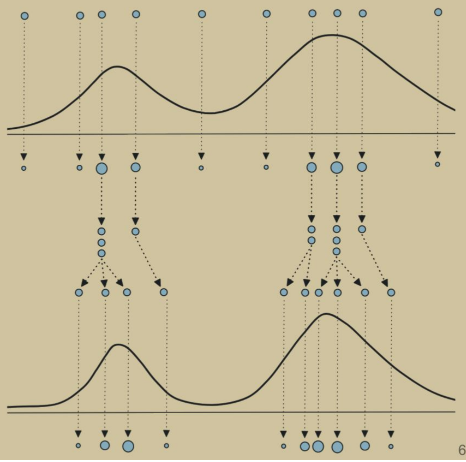
Evolve particles θ^i from prior to posterior

$$p_t(\theta) \propto \mathcal{L}^{\beta_t}(\theta)\pi(\theta)$$

$$0 = \beta_1 < \cdots < \beta_T = 1$$

Steps:

1. Compute weights $w_i = \mathcal{L}^{\beta_t - \beta_{t-1}}(\theta^i)$ and $\mathcal{Z}_t = \mathcal{Z}_{t-1} \times N^{-1} \sum_{i=1}^{N} w_i$
1. Resample particles
2. Move particles using MCMC



6

The following choices/adjustments for NS give ANS-SMC

- Resample and move *all* samples at each iteration
- Use weights for all samples (including above last threshold) based on MC rather than quadrature

This is equivalent to using $\alpha = (N-1)/N$ in ANS-SMC, which would be very inefficient.

Using ANS-SMC with $\alpha \approx e^{-1}$ gives closer computational effort.

Will Handley <wh260@cam.ac.uk>

Adaptive Ensemble Annealing
Michael Habeck

Nested sampling without NS?

Introduce a (θ-tunable) ensemble annealing function q(x|θ).

Construct an ensemble annealing algorithm which obeys NS characteristic of "constant thermodynamic speed"

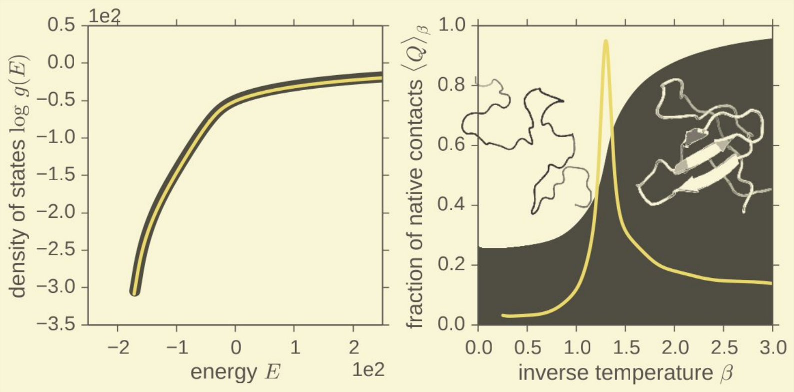
Similar performance to NS.

What assumptions/axioms turn this into NS?

ADAPTIVE ENSEMBLE ANNEALING

Require: θ_0 (initial control parameter), δ (entropic distance), N (#particles)
Initialize particles $\{\mathbf{x}_{0n}\}_{n=1}^{N}$ where $\mathbf{x}_{0n} \sim q(\mathbf{x} \mid \theta_0)$
while not converged **do**
Estimate importance weights w_m based on all samples
Pick next parameter θ_m such that $D_{\mathrm{KL}}(q(\cdot \mid \theta_m) \| q(\cdot \mid \theta_{m-1})) \overset{!}{=} \delta$.
Draw new particle(s) $\mathbf{x}_{mn} \sim q(\mathbf{x} \mid \theta_m)$ (by starting a Markov chain at a resampled particle)
end while
Approximate prior: $\pi \approx \sum_{mn} w_{mn} \delta_{\mathbf{x}_{mn}}$

APPLICATION TO PROTEIN FOLDING



Will Handley <wh260@cam.ac.uk>

Discussion Panel 1:

Open issues on statistical foundations, synergies with work on related algorithms (bridge/path sampling, SMC, rare event simulation)

1. Are we satisfied with proofs of convergence of integral and posterior
2. What is the link between SMC and NS? Is NS a subset of SMC, or is it distinct?
3. What can be the two communities learn from each other?
   - Why is NS more widely used in astro/particles -- is it just sociolology?
4. Are statisticians re-inventing nested sampling under "rare event sampling"?
5. Where is this field going of these various algorithms?
   - What is driving it?
   - What can it do/can't it do?
6. Does measure theory give us any useful insights?

Gradients

Pablo Lemos

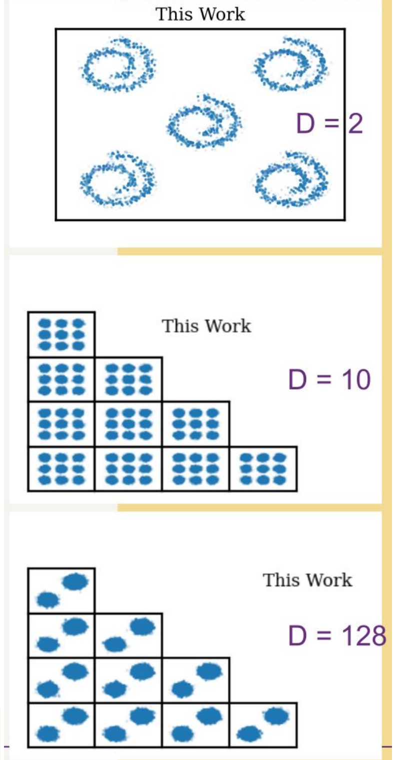
Updating Galilean/CHMC for the differentiable programming era

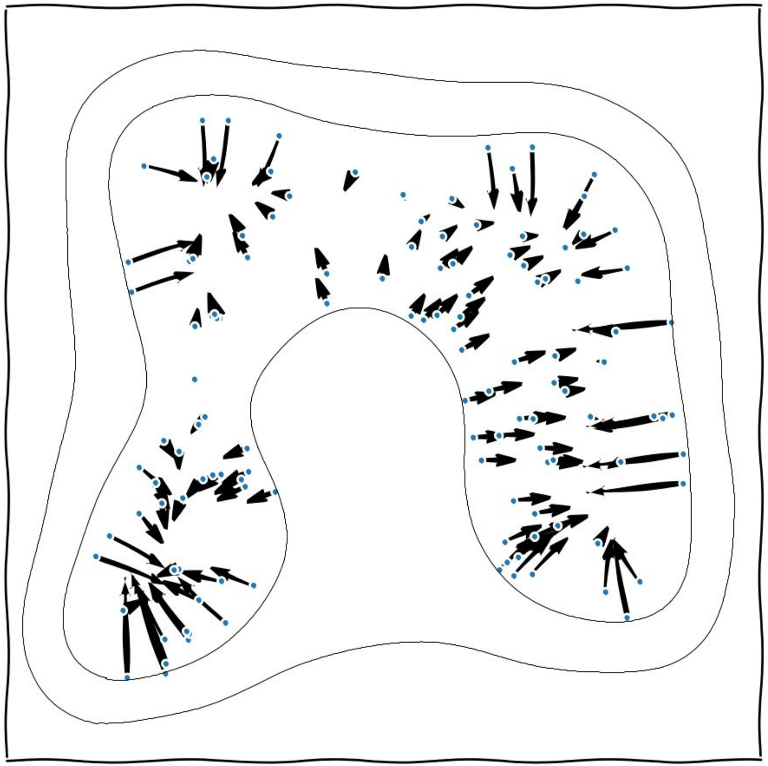
Hamiltonian slice sampling + clustering in jax

Other musings on using gradients:

- We only use gradients when we "step out" of the iso-likelihood contour.
- We have access to information about the gradient of the likelihood everywhere

- Current techniques don't make use of:
- We have "cloud" of gradients at every point
- We have an estimate of the prior volume X
- $\nabla \log X \propto \nabla \log L$





"Using gradients must be hard, because I haven't been able to do it" – *John Skilling*

Will Handley <wh260@cam.ac.uk>

Posterior Repartitioning
Xi Chen & Aleksandr Petrosyan

"Sucking some of the likelihood into the prior"

Separation of L & π is unique to NS: *"sample from the **prior**, subject to a hard **likelihood** constraint"*

Can correct user mis-specified priors (Chen)

Reshuffling can be made dynamic

Can accelerate nested sampling (SuperNest)

"Proposal/hint" distributions for NS (distinct from prior)

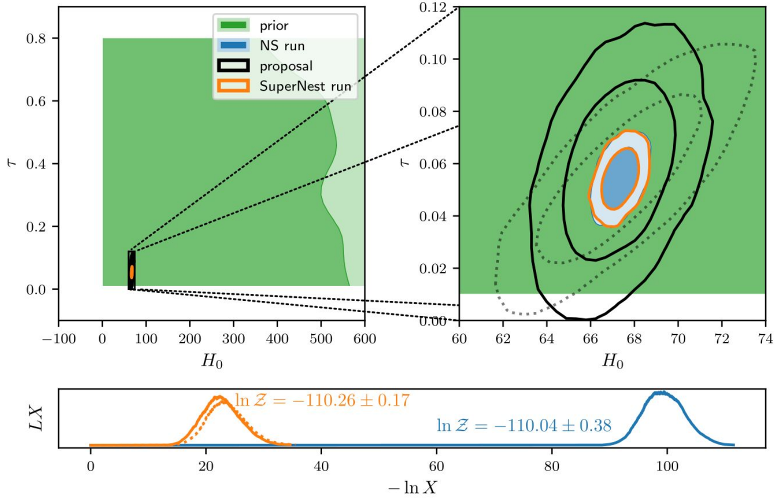
Non-trivial connection to importance sampling

Maintain posterior through detailed balance equation:

$$\mathcal{L}(\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \tilde{\mathcal{L}}(\boldsymbol{\theta})\tilde{\pi}(\boldsymbol{\theta})$$

- $\tilde{\mathcal{L}}(\boldsymbol{\theta})$ and $\tilde{\pi}(\boldsymbol{\theta})$ are the newly-defined effective (or modified) likelihood and prior, respectively
- $\tilde{\pi}(\boldsymbol{\theta})$ can be any tractable distributions that can be appropriately normalised to unit volume

$$\tilde{\mathcal{P}}(\boldsymbol{\theta},\beta) \propto \tilde{\mathcal{L}}(\boldsymbol{\theta},\beta)\tilde{\pi}(\boldsymbol{\theta},\beta) = \tilde{\mathcal{L}}(\boldsymbol{\theta},\beta)\tilde{\pi}(\boldsymbol{\theta}|\beta)\pi(\beta),$$



Will Handley <wh260@cam.ac.uk>

Importance nested sampling
Michael Williams & Johannes Lange

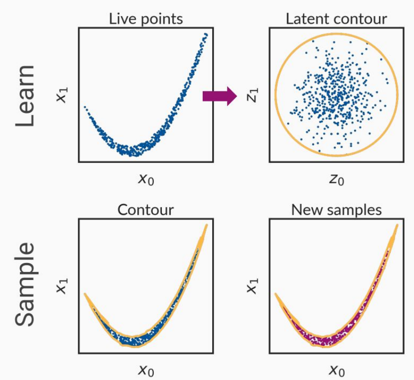
Region samplers for the deep learning era

i-nessai: normalising flows

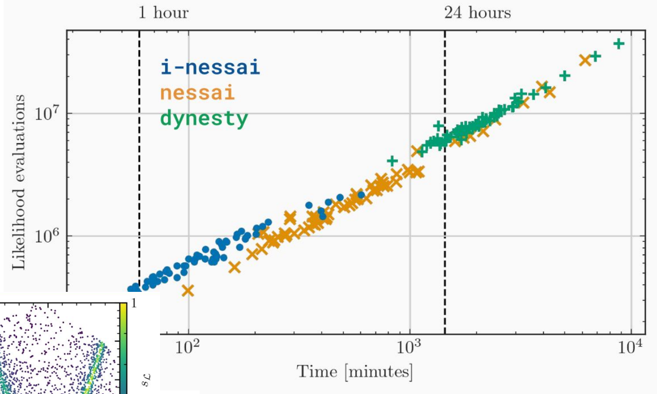
nautilus: classifiers

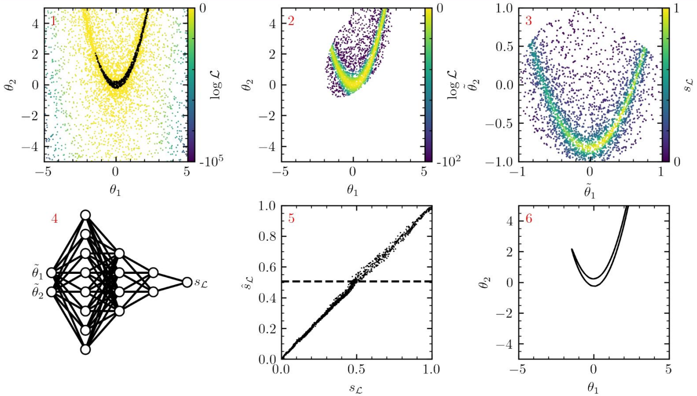
Use NNs to define regions and importance sample to improve efficiency.

If MultiNest had been written today, would likely used these rather than ellipsoids.



$$\hat{Z} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathcal{L}(\theta_i)\pi(\theta_i)}{Q(\theta_i)}$$





$$Q(\theta) = \sum_{i=1}^{N_{\mathrm{Flows}}}\alpha_i q_i(\theta)$$

Anesthetic

Lukas Hergt

Open-source community built nested sampling software suite

More than just another corner plotter!

Handles weighted statistics & feels like numpy/scipy/matplotlib/pandas

Give it a try:

pip install anesthetic

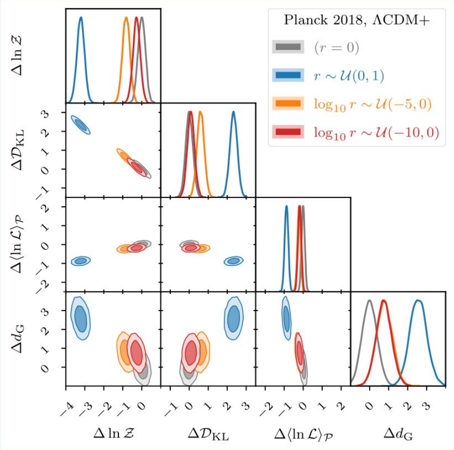
anesthetic: nested sampling post-processing

Authors:	Will Handley and Lukas Hergt
Version:	2.0.0
Homepage:	https://github.com/handley-lab/anesthetic
Documentation:	http://anesthetic.readthedocs.io/

CI passing | codecov 100% | docs passing | pypi package 2.0.0 | DOI 10.5281/zenodo.8097934 | JOSS 10.21105/joss.01414 | license MIT

`anesthetic` brings together tools for processing nested sampling chains by leveraging standard scientific python libraries.

Bayesian statistics: logarithmic vs uniform prior on r



weighted `pandas.DataFrame`

```
[4]: ns = read_chains(gaussian3d_path + "gaussian3d_ns_100nlive_polychord_raw/gaussian3d_ns_100nlive")
[5]: ns
```

[5]:

		x1	x2	x3	logprior__0	loglike__gaussian3d	logL	logL_birth	nlive
	labels	x_1	x_2	x_3	$\log \pi_0$	$\log \mathcal{L}_{\mathrm{gaussian3d}}$	$\ln \mathcal{L}$	$\ln \mathcal{L}_{\mathrm{birth}}$	n_{live}
	weights								
0	**3.120715e-17**	-5.492322	5.443531	5.329458	-7.45472	-46.857193	-46.857193	-inf	114
1	**1.455935e-16**	-5.407397	-5.153382	-5.415083	-7.45472	-45.317022	-45.317022	-inf	113
2	**1.704418e-14**	-3.600835	5.347057	-5.834200	-7.45472	-40.554276	-40.554276	-inf	112
3	**1.778227e-14**	4.457432	4.539513	-5.918974	-7.45472	-40.511883	-40.511883	-inf	111
4	**2.113157e-14**	5.046166	4.235301	-5.635906	-7.45472	-40.339317	-40.339317	-inf	110
...	**...**	...	...	...	...	...	...	...	...
1377	**3.269885e-03**	-0.009055	-0.031159	0.002386	-7.45472	-2.757345	-2.757345	-2.781492	5
1378	**3.269975e-03**	-0.028005	-0.014035	0.004734	-7.45472	-2.757317	-2.757317	-2.767012	4
1379	**3.270127e-03**	-0.028708	0.003180	-0.008758	-7.45472	-2.757271	-2.757271	-2.770868	3
1380	**3.270155e-03**	-0.021166	-0.003618	0.020802	-7.45472	-2.757263	-2.757263	-2.789061	2
1381	**3.271606e-03**	0.001819	0.001527	-0.000824	-7.45472	-2.756819	-2.756819	-2.770868	1

1382 rows × 8 columns

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

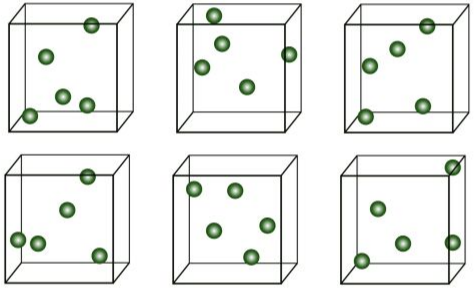
Will Handley <wh260@cam.ac.uk>

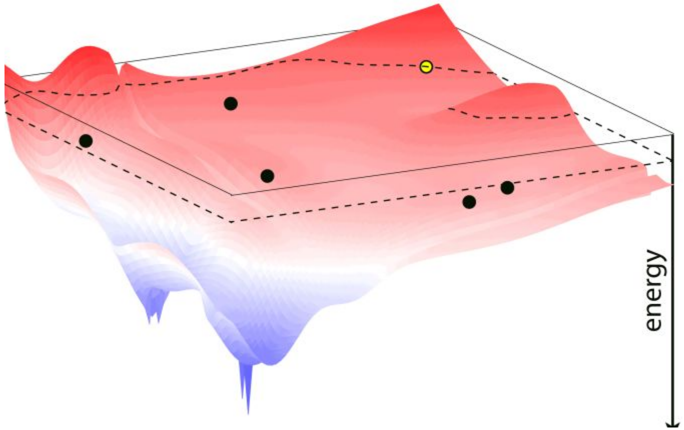
Discussion Panel 2:

The relation between machine learning and NS

1. When do we use machine learning/when do we use nested sampling?
2. What are the prospects for hybrid methods? (Normalising flows, model emulators/BAMBI)
3. What is the future?
4. Gradients/differentiable programming. Why is it hard to use them?
5. Hierarchical Bayesian models with 1000s of parameters
    - is this an area that nested sampling is ever going to be useful?
6. Simulation based inference and nested sampling?
    - Evidence Networks
7. Can NS benefit from hardware accelerators similar to ML?
    - GPUs were key to ML developments, could there be an equivalent for NS?
    - Quantum computing?
8. Connections between NS and variational inference?
    - May be parallels to Importance NS

NS for atomic systems
Livia Bartok-Partay



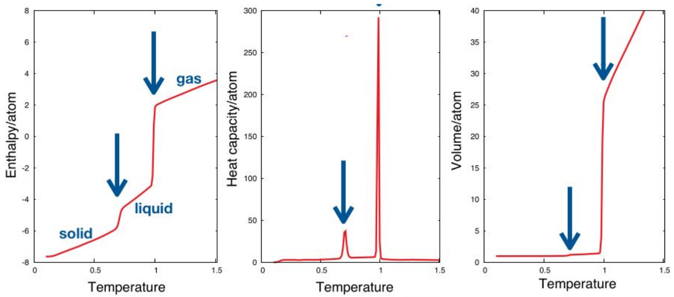


Earliest adopters of NS (2007)

Community that actually uses the partition function

Cannot use rejection samplers

Beauty of nested sampling is that it gives you the thermodynamically important basins



Review paper:

[10.1140/epjb/s10051-021-00172-1](https://doi.org/10.1140/epjb/s10051-021-00172-1)

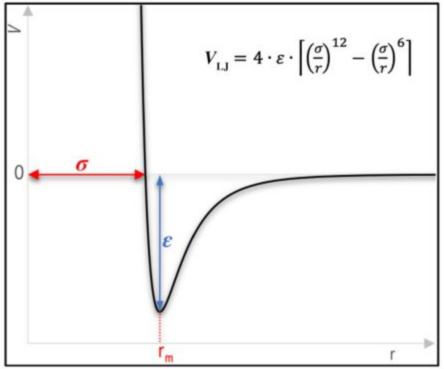
Will Handley <wh260@cam.ac.uk>

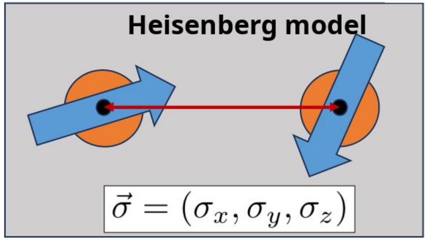
Magnets
George Marchant

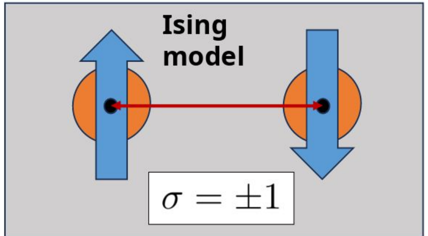
NiFe as gap magnets:

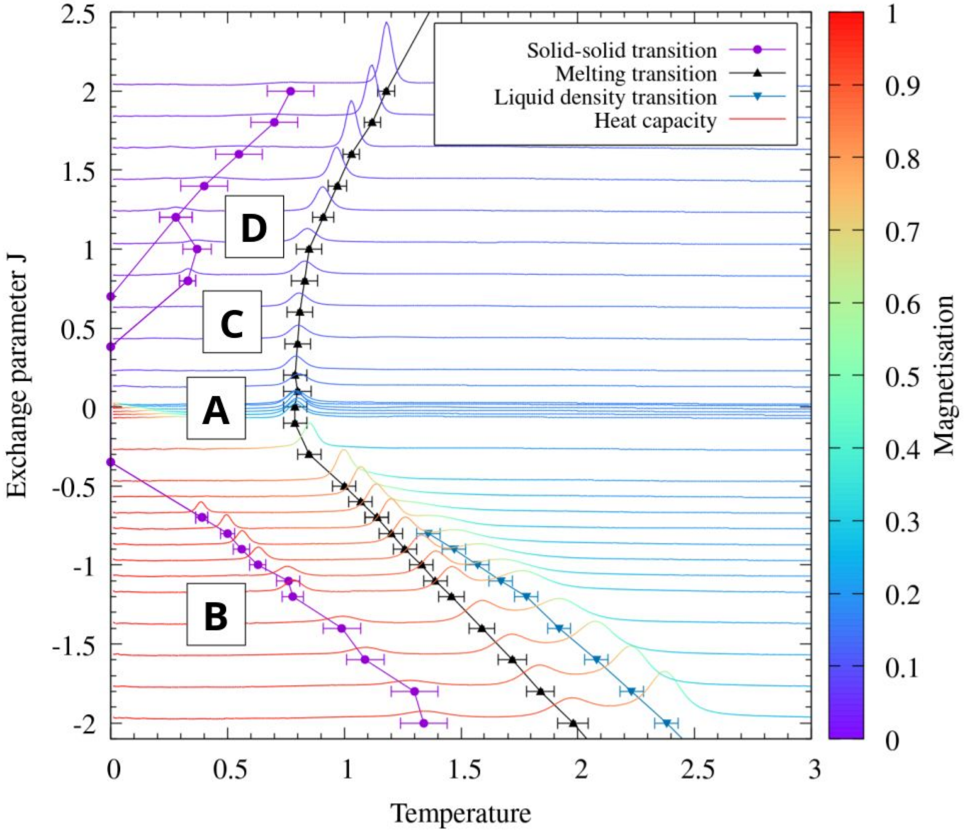
Using atomistic models + NS can recover rich phenomenology (ordered, frustrated & compensated phases)

Get all temperatures in one go!





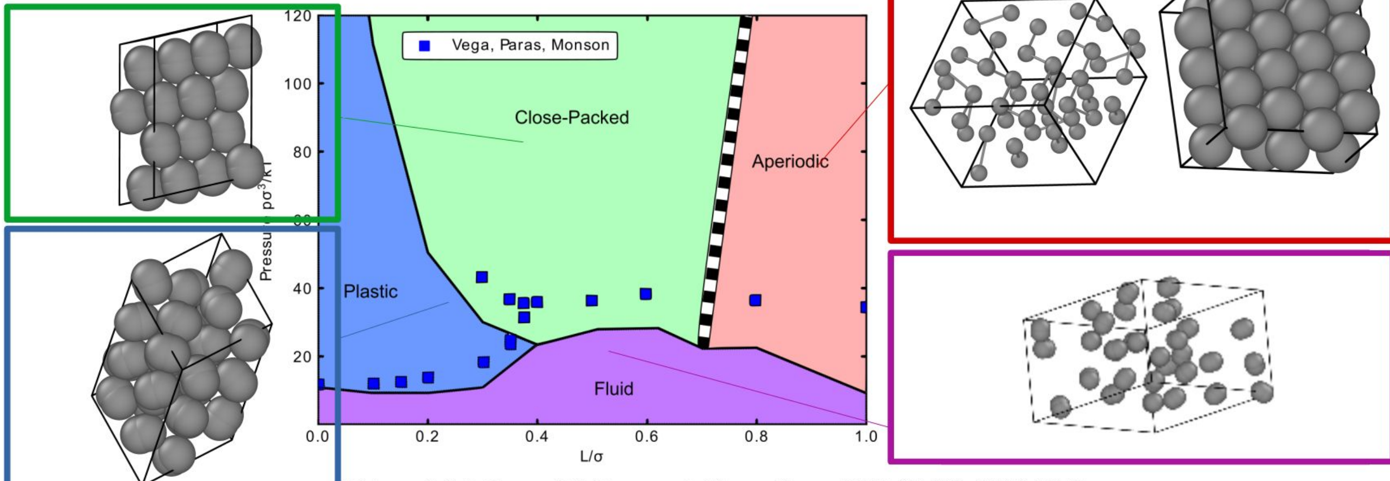




Will Handley <wh260@cam.ac.uk>

Hard spheres
Omar Adesida

Phase Diagram



C Vega, E P A Paras, P A Monson, J. Chem. Phys. 1992; 97 (11): 8543–8548.

Dimers (two spheres stuck together) as a model organism exhibiting interesting & complex phase behaviour

Looking at the physics rather than the chemistry (clean a-priori knowledge)

Challenging for traditional approaches because of ‘forbidden regions’

NS solves this as a matter of course

NS discovers aperiodic phase in compressibility graph

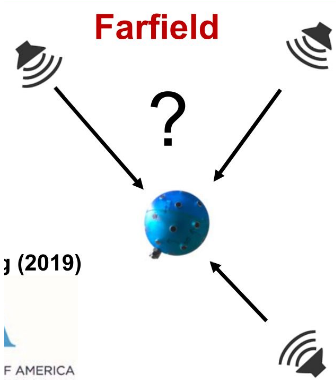
Will Handley <wh260@cam.ac.uk>

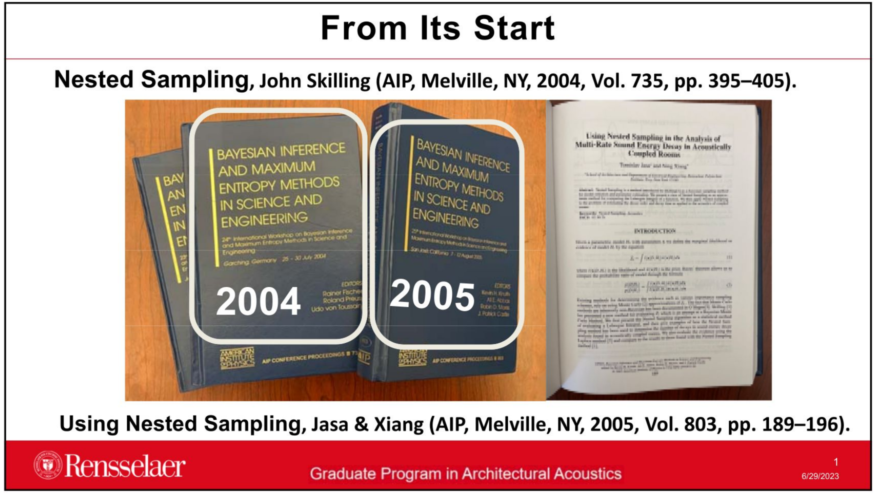
Spherical microphone array
Ning Xiang

Whole 19 year research programme arising from John's original 2019 paper

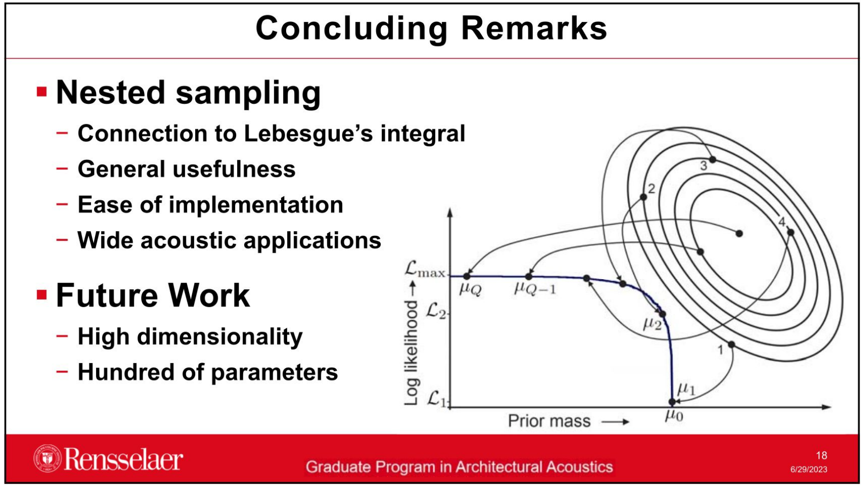
Uses nested sampling to explore parameter space and compare models for beamforming spherical microphone arrays

Nested sampling as a Lebesgue integrator





1



18

Also!

Margret Westerkamp (tomorrow – precision nested sampling)

Cesar Godhino NS + QED

Louis Duval: Atomic physics & spectroscopy

Nico Unglert Silicon-based phase diagrams + ML

Jack O’Brian (Supernovae)

Mingrui Yang: Surface science + Nested sampling phase diagrams

Lune Maillard: Nested Sampling with materials and quantum effect

Martino Trassinelli: Atomic spectroscopy

Images
Jason McEwan

Drawing ideas from other fields (signal processing)

NS on images!

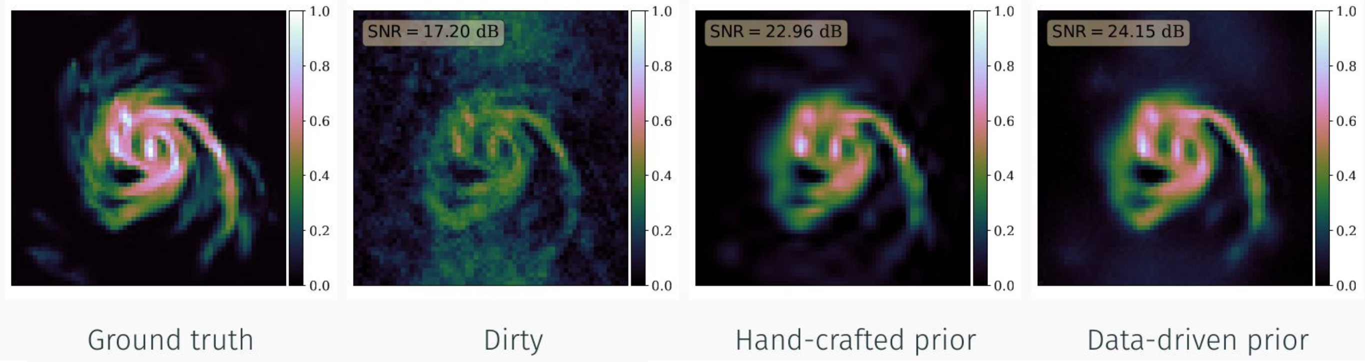
d~1e6 (256x256)

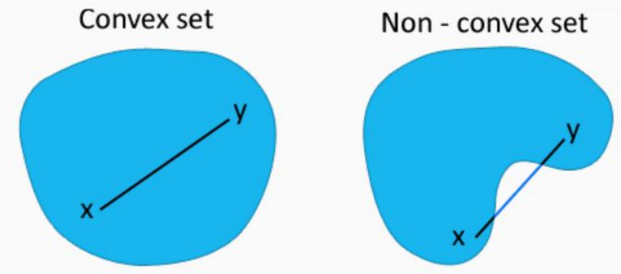
Relies on convexity of likelihood-prior combination

Suggests DKL !~O(d)

[2307.00056](2307.00056)

NS separation of prior means one can use deep data-driven priors

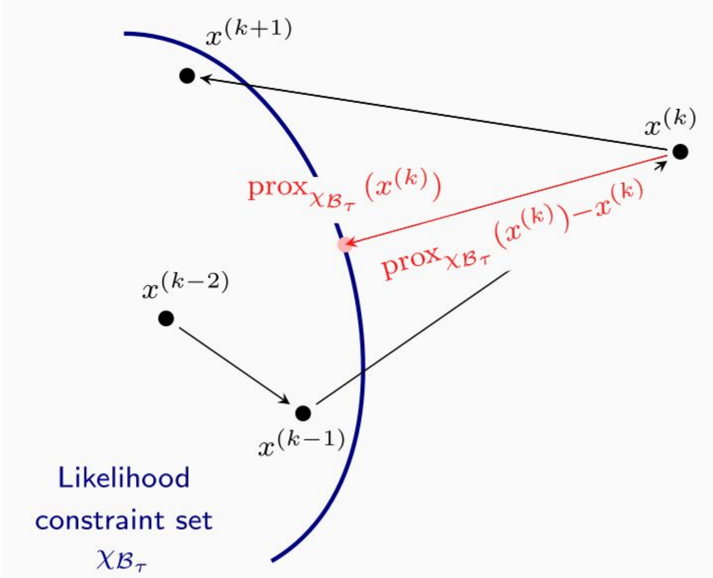




pip install ProxNest

also shout out to Harmonic – Alicja Polanska (see poster)

An alternative evidence calculator (minor heresy)



Opening up Nested Sampling
Andrew Fowlie

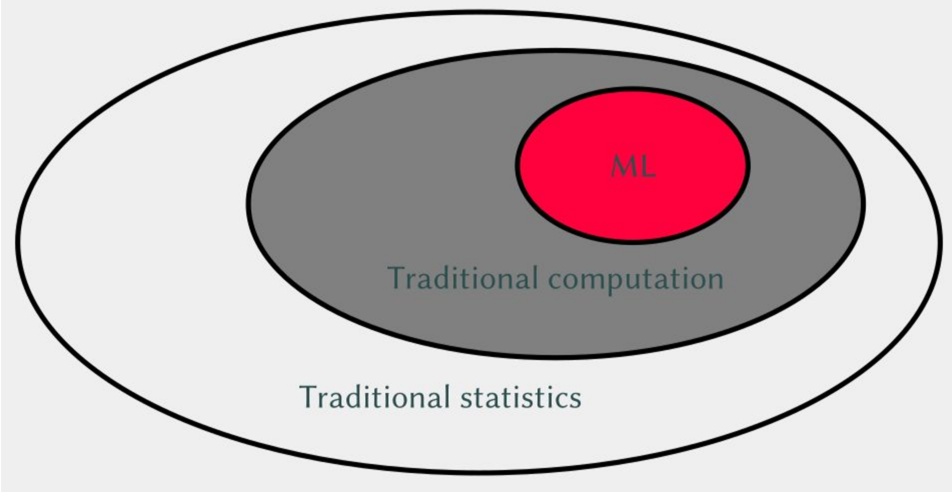
Applications: NS for p values (major heresy!), NS for ABC (approximate Bayesian computation)

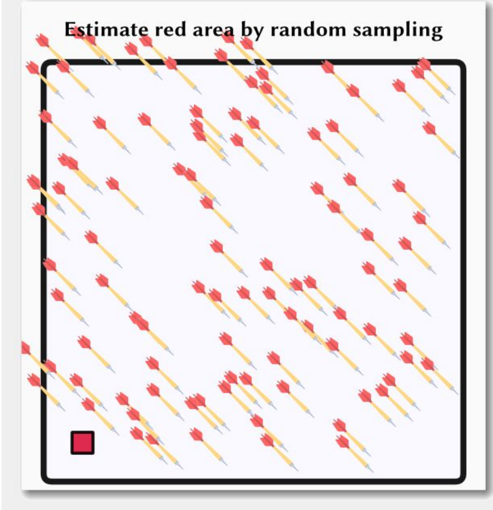
Errors: Heuristic √(H/n) valid when

$\langle \log X \rangle \gg \sigma(\log X)$

Live points: insertion index checking [2006.03371](2006.03371)

Constrained sampling: room to accommodate ML even for skeptics





Will Handley <wh260@cam.ac.uk>

(Information) Entropy with NS
Brendon Brewer

Looking for applications!

standard deviation quantifies 'horizontal' uncertainty, entropy does it 'vertically'.

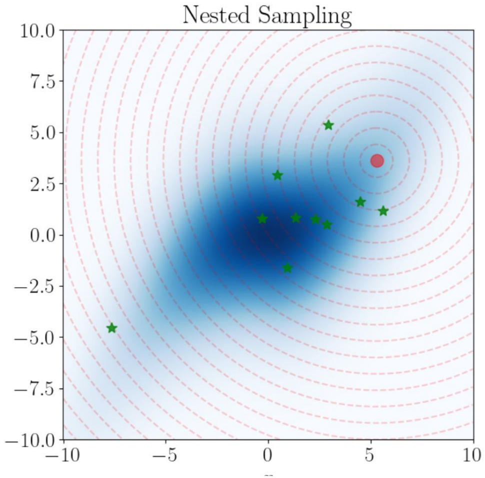
Can estimate unknown log probability at a point using number of NS iterations it takes to get to that point

Works for marginal entropies (e.g. if there are nuisance parameters).

InfoNest

To compute $H(\theta) = -\int f(\theta) \log f(\theta)\, d\theta$ when f can be sampled but not evaluated:

1. Generate a 'reference point' θ_{ref} from f
2. Do a Nested Sampling run with f as "prior" and minus the distance to θ_{ref} as "likelihood".
3. Measure how many NS iterations were needed to make the distance to θ_{ref} really small, and divide by N. That gives an unbiased estimate of the log-prob near θ_{ref}.
4. Repeat steps 1–3 many times.
5. Average the estimated log-probs, then apply corrections to convert to density.



Thank you

To you all

To Udo

To Johannes

To John

See you next year for NS2024 – the 20th anniversary