

Nested sampling for small p -values

Andrew Fowlie, Sebastian Hoof, and Will Handley (May 2021). In: arXiv: 2105.13923 [physics.data-an]

github.com/andrewfowlie/ns_for_p_values

Andrew Fowlie

22 November 2021

Nanjing Normal University



1. Why small p -values?
2. Why is it difficult to compute small p -values?
3. Nested sampling
4. Example & Tools

Why small p -values?

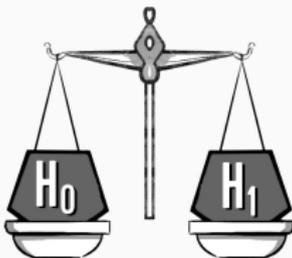
***P*-value**

The *p*-value, *p*, is the probability of observing data as or more extreme than that observed, given the null hypothesis, H_0 , i.e.,

$$p = P(\lambda \geq \lambda_{\text{Observed}} \mid H_0)$$

where λ is a test-statistic that summarises the data and defines extremeness.

P-values in high-energy physics



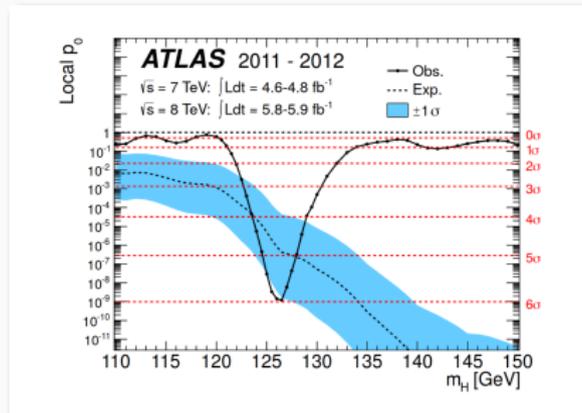
In high-energy physics, we want to discover new phenomena and new particles. Perform null hypothesis test:

- H_0 — Standard Model (SM) backgrounds only
- H_1 — SM + new physics, e.g. Higgs boson or supersymmetric particles

We conventionally require a tiny global p -value less than about 10^{-7} , corresponding to 5σ

Higgs discovery

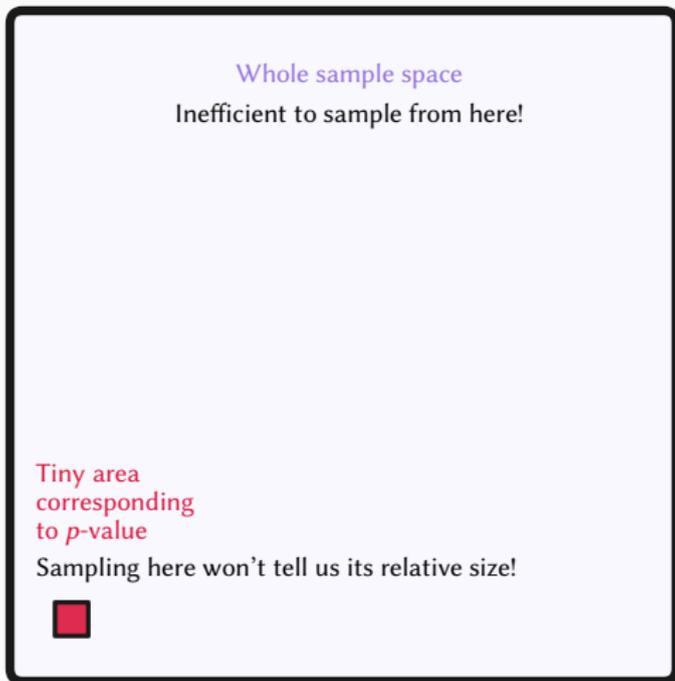
Classic example. Higgs discovery in 2012 (Aad et al. 2012).



Wait until reach 5σ global. We need to compute tiny p -values.

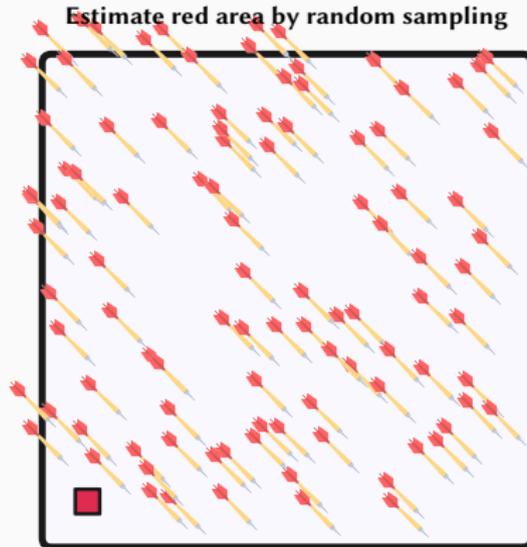
Why is it difficult to compute small p -values?

Illustrate problem with two-dimensional data, \mathbf{x} .



In reality, red region exponentially tiny. Illustrate with squares but assume nothing about geometry/topology in problem or solution.

Sample from whole sample space — Monte Carlo



Draw n samples from whole sampling distribution. We really need at least one sample to fall in red region.

Sample from whole sample space — Monte Carlo

Estimate p by fraction of them that fall in red region

$$\hat{p} = \frac{m}{n}$$

Error of order Wald (Brown, Cai, and DasGupta 2001)

$$\frac{\Delta p}{p} = \sqrt{\frac{1/p}{n}}$$

Usual $1/\sqrt{n}$ statistical error scaling. **For fixed fractional uncertainty, number of samples scales as $1/p$**

Need $n \gtrsim 1/p$ at very least for reasonable estimate.

If we have some more information, we can compute p analytically.

- E.g., I know the red area is a box of side 0.05. $p = 0.05^2$.
- Sometimes, we know (or hope!) our problem satisfies certain regularity conditions and large sample limit. We can apply asymptotics (Cowan et al. 2011).

But the conditions aren't always satisfied.

We want generality. Generality is power to tackle any problem we want.

Nested sampling

Nested sampling

Nested sampling (Skilling 2004; Skilling 2006) originally algorithm for Bayesian computation. We reinterpreted it in the context of p -values

- Evolves collection of n_{live} live points to greater and greater **test-statistics**
- Evolution controlled by single user parameter — n_{live}
- Replaces one point at a time
- **Meta-algorithm** — different solutions to finding replacement points
- Many existing public implementations

General idea

1. Sample n points from the sampling distribution
2. Rank them by test-statistic
3. Delete the least extreme half

You just compressed by factor 1/2! Repeat it i times and you'll achieve exponential compression $1/2^i$ at a constant rate!

If it took n_{iter} iterations to reach the area corresponding to p -value, then

$$p = \frac{1}{2^{n_{\text{iter}}}}$$

Complete algorithm

1. Draw n_{live} sets of pseudo-data from the sampling distribution – the live points
2. Initialize $n_{\text{iter}} = 0$
3. repeat
4. $n_{\text{iter}} = n_{\text{iter}} + 1$
5. Find the minimum TS λ^* amongst the live points
6. Replace live point corresponding to λ^* by one drawn from the sampling distribution subject to $\lambda > \lambda^*$
7. until $\lambda^* \geq \lambda_{\text{Observed}}$
8. return *Estimate of $p = e^{-n_{\text{iter}}/n_{\text{live}}}$*

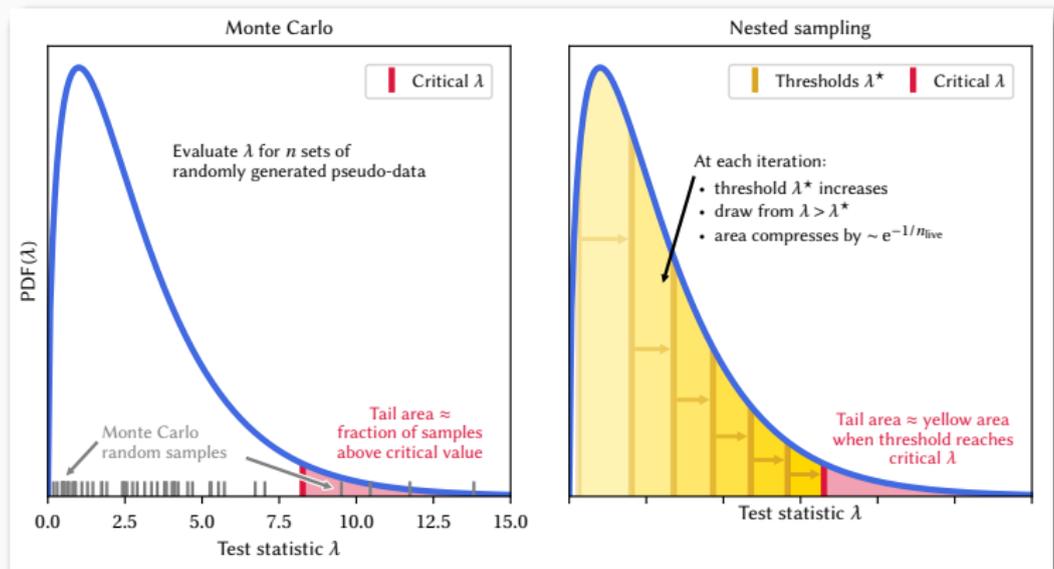
Why does that work?

- By replacing **half** the points, we compress by 1/2 per iteration
- That isn't optimal — better to **replace one point at a time**
- Compress by about $e^{-1/n_{\text{live}}} \simeq 1 - 1/n_{\text{live}}$ each iteration

This breaks the computation of a tiny p -value into the product of n_{iter} moderate factors

$$p = \prod_{i=1}^{n_{\text{iter}}} e^{-1/n_{\text{live}}} = e^{-n_{\text{iter}}/n_{\text{live}}}$$

Nested sampling



Compress into tail of TS distribution. Stop once we get to p -value.

Theoretical speedup

For fixed fractional uncertainty on p , we expect to obtain a speed-up versus MC

$$\frac{\text{Evaluations for NS}}{\text{Evaluations for MC}} = \frac{(\log^2 1/p)/\epsilon}{1/p}$$

Massive gains for small p ! Provided that the efficiency factor ϵ doesn't spoil things.

Exploration

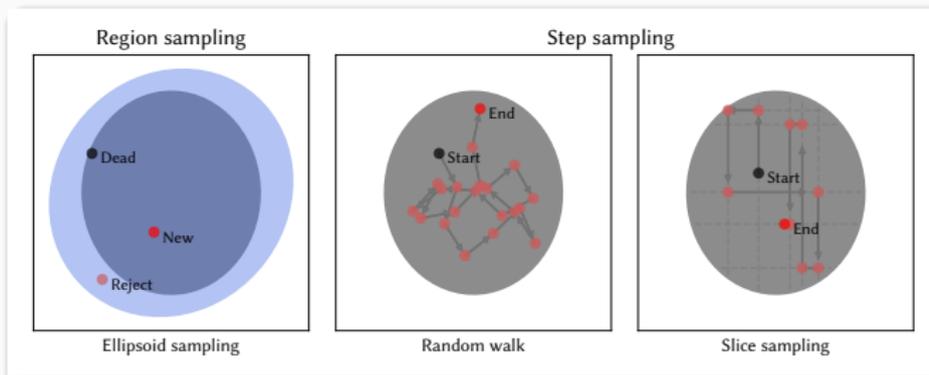
So far nothing depended on the problem at hand or the dimension of the sampling space. I glossed over a detail though

- How to draw a replacement live point from the constrained sampling distribution, $\lambda > \lambda^*$?
- Involves trial and error, and **some inefficiency (the factor ϵ)**
- Re-introduces dependence on dimensionality of the sampling space (though it needn't be exponential)

Strategies

MultiNest (Feroz and Hobson 2008; Feroz, Hobson, and Bridges 2009; Feroz et al. 2013) — bound live points by ellipsoids that approximate the λ^* contour and sample from them

PolyChord (Handley, Hobson, and Lasenby 2015a; Handley, Hobson, and Lasenby 2015b). Take walk starting from a randomly chosen existing live point — in PolyChord a slice sampling walk.



Example & Tools

A simple problem

The p -value associated with d independent Gaussian measurements

- d dimensional sampling space, $\mathbf{x}_i \sim \mathcal{N}(\mu, \sigma^2)$.
- Test-statistic

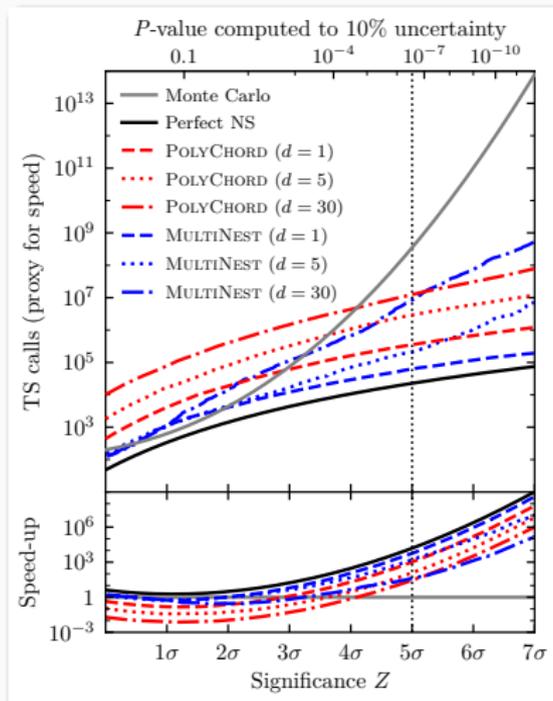
$$\lambda = \sum_{i=1}^d \left(\frac{\mathbf{x}_i - \mu}{\sigma} \right)^2$$

- We know analytically

$$\lambda \sim \chi_d^2 \quad \text{such that} \quad p = 1 - F_{\chi_d^2}(\lambda_{\text{Observed}})$$

- Toy example that allows us to easily control dimension, size of p and check correctness

Number of evaluations for fixed fractional uncertainty



Perfect NS means if 100% efficiency, $\epsilon = 1$, was possible

Performance on simple problem

- For large p , nested sampling could only narrowly beat MC in the best-case scenario, and typically worse. MC just fine when p is moderate
- For small $p \lesssim 4\sigma$, the scaling kicks in. Nested sampling wins by orders of magnitude
- Nested sampling performance depends on dimensionality but even for $30d$ sampling space, winning by 10^6 at 7σ

Implementing the example in Python

Import our PolyChord wrapper from our code

 github.com/andrewfowlie/ns_for_p_values

```
>>> from p_value import pc
```

This wraps PolyChord to enable the required stopping conditions and return the p -value from NS.

We also wrapped dynesty and MultiNest. Healthy ecosystem of publicly available NS implementations and NS analysis software.

Implementing the example in Python

Make a function that transforms $U(0, 1)$ draws to chi-squared draws. These chi-squared draws are the pseudo-data

```
>>> from scipy.stats import chi2

>>> def transform(unit_hyper_cube):
...     return chi2.ppf(unit_hyper_cube, df=1)
```

This is the inverse transform method using the chi-squared distribution's percent point function — `chi2.ppf`

Implementing the example in Python

Define the test-statistic — here it's just the sum of the chi-squared draws

```
>>> def test_statistic(data):  
...     return data.sum()
```

Implementing the example in Python

Run PolyChord on this problem for 5 draws and an observed $\lambda = 50$, and using 100 live points.

```
>>> n_dim = 5
>>> observed = 50.
>>> pc(test_statistic, transform, n_dim, observed,
      n_live=100)
```

Same signature for our MultiNest and dynesty wrappers, just replace `pc` with `mn` or `dynesty`.

Simple easy-to-use signature. Wraps established, fast libraries for NS.

Implementing the example in Python

The results are displayed as

```
p-value = (1.5910 +/- 0.7161)e-09.  
log10(p-value) = -8.7983 +/- 0.19548.  
Significance = 5.922 sigma.  
Function calls = 309765
```

We computed a p -value of 1 in a billion in only 300,000 calls!
Agrees with analytic result within uncertainty:

```
>>> chi2.sf(50, df=5)  
1.3857973367009573e-09
```

Summary

- Nested sampling particularly suitable for p -value computation, as it naturally builds path to the p -value
- Orders of magnitude faster than Monte Carlo for small p , as scaling $\log^2 1/p$ rather than $1/p$ for fixed relative error
- Performance understood theoretically and demonstrated numerically
- With small modifications, standard nested sampling software can be used to compute p -values
-  github.com/andrewfowlie/ns_for_p_values

References

- Aad, Georges et al. (2012). “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC.” In: *Phys. Lett. B* 716, pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].
- Brown, Lawrence D., T. Tony Cai, and Anirban DasGupta (2001). “Interval Estimation for a Binomial Proportion.” In: *Statistical Science* 16.2, pp. 101–117. DOI: 10.1214/ss/1009213286.
- Cowan, Glen et al. (2011). “Asymptotic formulae for likelihood-based tests of new physics.” In: *Eur. Phys. J. C* 71. [Erratum: *Eur. Phys. J. C* 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

References ii

- Feroz, F., M. P. Hobson, and M. Bridges (2009). “MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics.” In: *Mon. Not. Roy. Astron. Soc.* 398, pp. 1601–1614. DOI: 10.1111/j.1365-2966.2009.14548.x. arXiv: 0809.3437 [astro-ph].
- Feroz, F. et al. (2013). “Importance Nested Sampling and the MultiNest Algorithm.” In: *The Open Journal of Astrophysics*. DOI: 10.21105/astro.1306.2144. arXiv: 1306.2144 [astro-ph.IM].
- Feroz, Farhan and M. P. Hobson (2008). “Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis.” In: *Mon. Not. Roy. Astron. Soc.* 384, p. 449. DOI: 10.1111/j.1365-2966.2007.12353.x. arXiv: 0704.3704 [astro-ph].

Fowlie, Andrew, Sebastian Hoof, and Will Handley (May 2021).

“Nested sampling for frequentist computation: fast estimation of small p -values.” In: arXiv: 2105.13923 [physics.data-an].

Handley, W. J., M. P. Hobson, and A. N. Lasenby (2015a).

“PolyChord: nested sampling for cosmology.” In: *Mon. Not. Roy. Astron. Soc.* 450.1, pp. L61–L65. DOI: 10.1093/mnras/slv047. arXiv: 1502.01856 [astro-ph.CO].

– (Nov. 2015b). “PolyChord: next-generation nested sampling.” In: *Mon. Not. Roy. Astron. Soc.* 453.4, pp. 4384–4398. DOI: 10.1093/mnras/stv1911. arXiv: 1506.00171 [astro-ph.IM].

- Skilling, John (Nov. 2004). “Nested Sampling.” In: *American Institute of Physics Conference Series*. Ed. by Rainer Fischer, Roland Preuss, and Udo Von Toussaint. Vol. 735. American Institute of Physics Conference Series, pp. 395–405. DOI: 10.1063/1.1835238.
- (2006). “Nested sampling for general Bayesian computation.” In: *Bayesian Analysis* 1.4, pp. 833–859. DOI: 10.1214/06-BA127.