Title: An Introduction to (Dynamic) Nested Sampling and Model Selection

Speakers: John Speagle

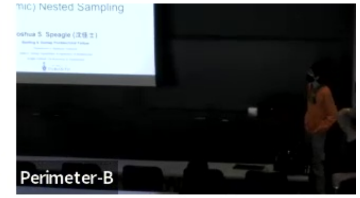
Series: Cosmology & Gravitation

Date: April 26, 2022 - 11:00 AM

URL: https://pirsa.org/22040129

Abstract: I will present a brief introduction to Nested Sampling, a complementary framework to Markov Chain Monte Carlo approaches that is designed to estimate marginal likelihoods (i.e. Bayesian evidences) and posterior distributions. This will include some discussion on the philosophical distinctions and motivations of Nested Sampling, a few ways of understanding why/how it works, some of its pros and cons, and more recent extensions such as Dynamic Nested Sampling. If time/interest permits, I can either (a) highlight how this can work in practice using the public Python package dynesty or (b) discuss the more general problem of model selection and why Bayesian evidences may (or may not) be helpful.

Zoom Link: https://pitp.zoom.us/j/95990705337?pwd=VzB4cjhzSDhoM0RCYTNwZHUzUVlzdz09



An Introduction to (Dynamic) Nested Sampling

Joshua S. Speagle (沈佳士)

Banting & Dunlap Postdoctoral Fellow

Department of Statistical Sciences

David A. Dunlap Department of Astronomy & Astrophysics

Dunlap Institute for Astronomy & Astrophysics

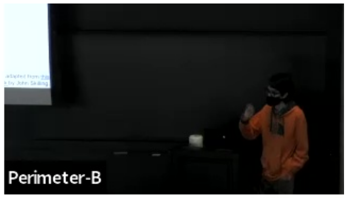
UNIVERSITY OF TORONTO

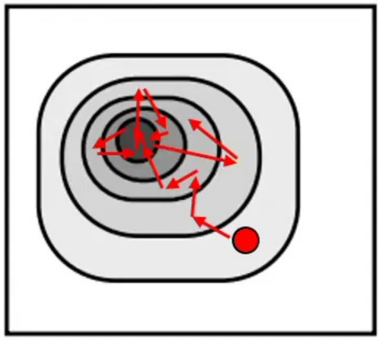
Background

$$\underbrace{\Pr(\boldsymbol{\Theta}|\mathbf{D},\mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{\Pr(\mathbf{D}|\boldsymbol{\Theta},\mathrm{M})}^{\text{Likelihood}}\ \overbrace{\Pr(\boldsymbol{\Theta}|\mathrm{M})}^{\text{Prior}}}{\underbrace{\Pr(\mathbf{D}|\mathrm{M})}_{\text{Evidence}}}$$

Bayes' Theorem

Motivation: Sampling the Posterior



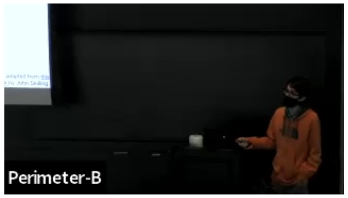


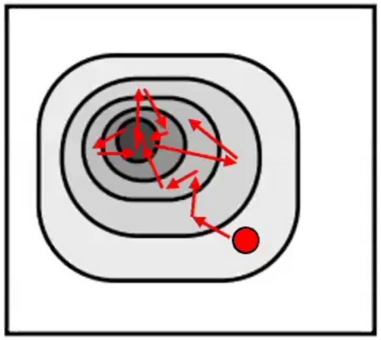
Sampling directly from the likelihood $\mathcal{L}(\Theta)$ is **hard**.

MCMC: Solving a Hard Problem once.
(Markov Chain Monte Carlo)

Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior



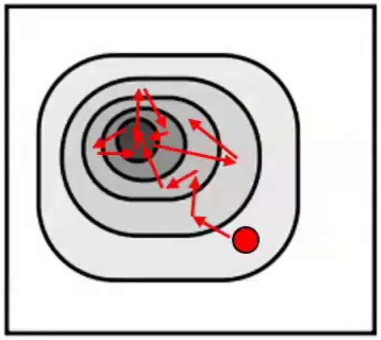


Sampling uniformly within $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ bound is **easier**.

MCMC: Solving a Hard Problem once.

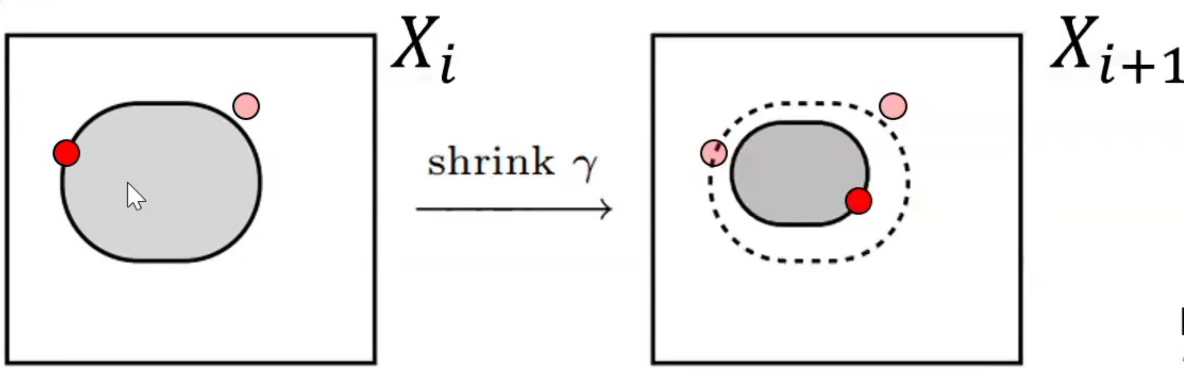
Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior



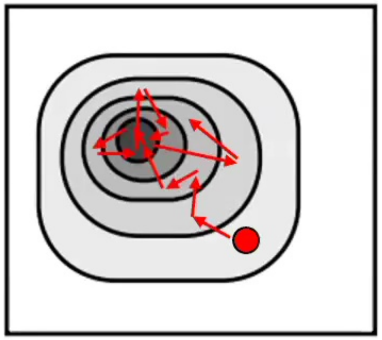
MCMC: Solving a Hard Problem once.
vs
Nested Sampling: Solving an Easier Problem many times.

Sampling uniformly within $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ bound is **easier**.



Pictures adapted from this 2010 talk by John Skilling.

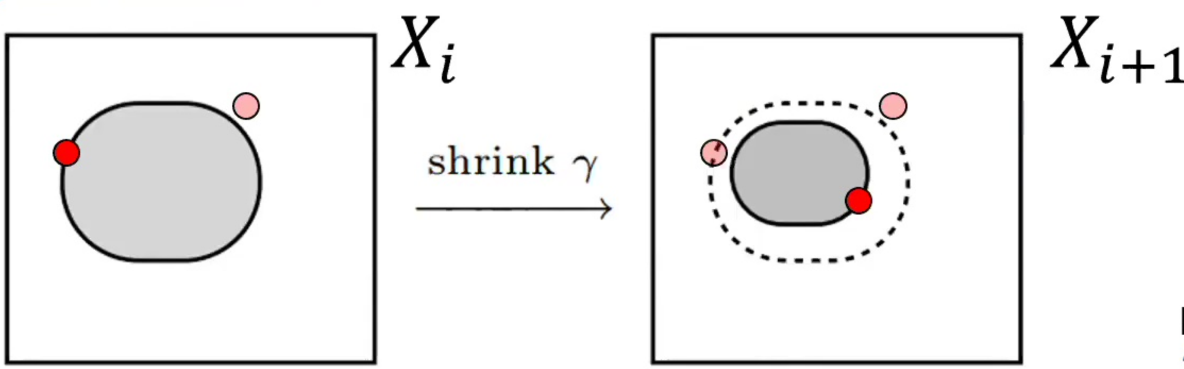
Motivation: Sampling the Posterior



MCMC: Solving a Hard Problem once.
vs
Nested Sampling: Solving an Easier Problem many times.

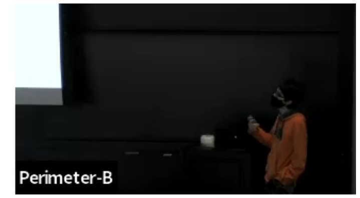
Sampling uniformly within $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ bound is **easier**.

If you have a **prior transform** that converts your priors to look uniform, then this case is equivalent.



Pictures adapted from this 2010 talk by John Skilling.

Motivation: Integrating the Posterior

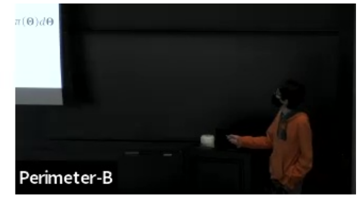


Posterior — Likelihood — Prior — Evidence

$$\underbrace{\Pr(\boldsymbol{\Theta}|\mathbf{D},\mathrm{M})}_{\text{Posterior}} = \frac{\overbrace{\Pr(\mathbf{D}|\boldsymbol{\Theta},\mathrm{M})}^{\text{Likelihood}}\ \overbrace{\Pr(\boldsymbol{\Theta}|\mathrm{M})}^{\text{Prior}}}{\underbrace{\Pr(\mathbf{D}|\mathrm{M})}_{\text{Evidence}}}$$

Bayes' Theorem

Motivation: Integrating the Posterior



Posterior, Likelihood, Prior

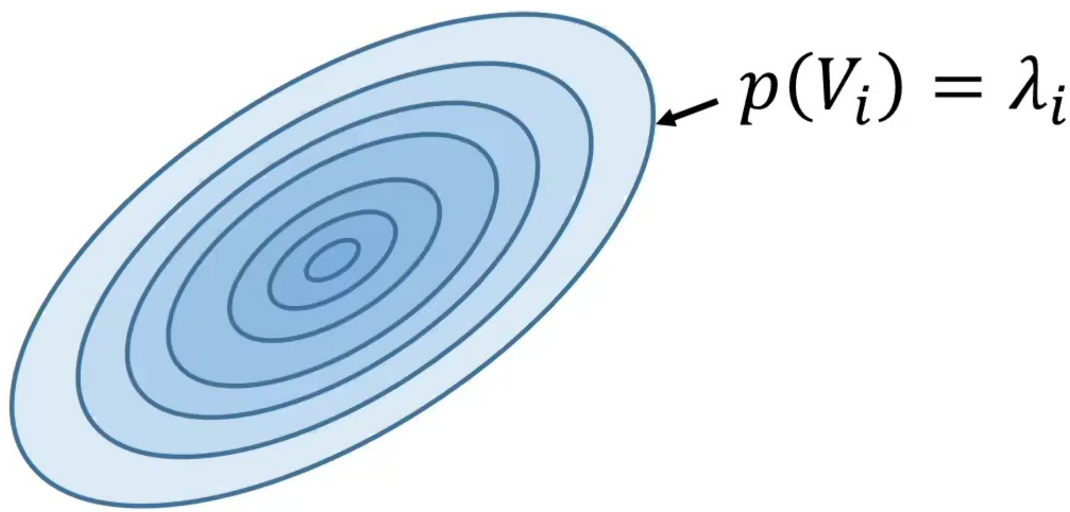
$$p(\boldsymbol{\Theta}) = \frac{\mathcal{L}(\boldsymbol{\Theta})\,\pi(\boldsymbol{\Theta})}{\mathcal{Z}}$$

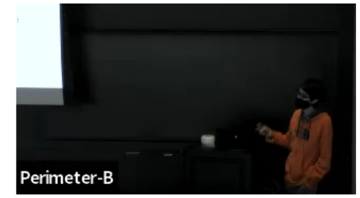
$$\text{Evidence} \equiv \int_{\Omega_{\boldsymbol{\Theta}}} \mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$$

Bayes' Theorem

Motivation: Integrating the Posterior

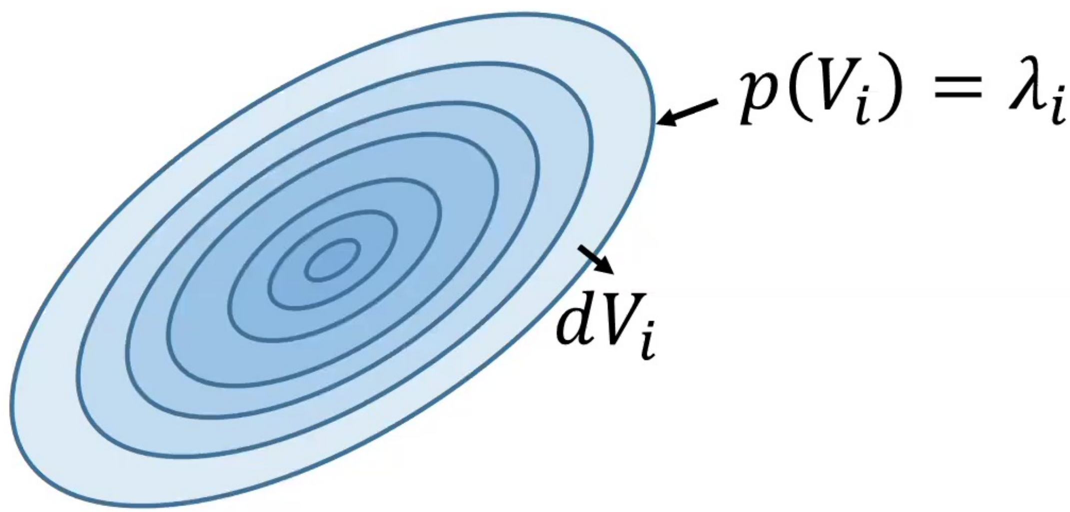
$$\int_{\Omega_{\boldsymbol{\Theta}}} p(\boldsymbol{\Theta})\, d\boldsymbol{\Theta}$$





Motivation: Integrating the Posterior

$$\int_{\{\boldsymbol{\Theta}:\, p(\boldsymbol{\Theta})=\lambda\}} \lambda\, dV(\lambda)$$

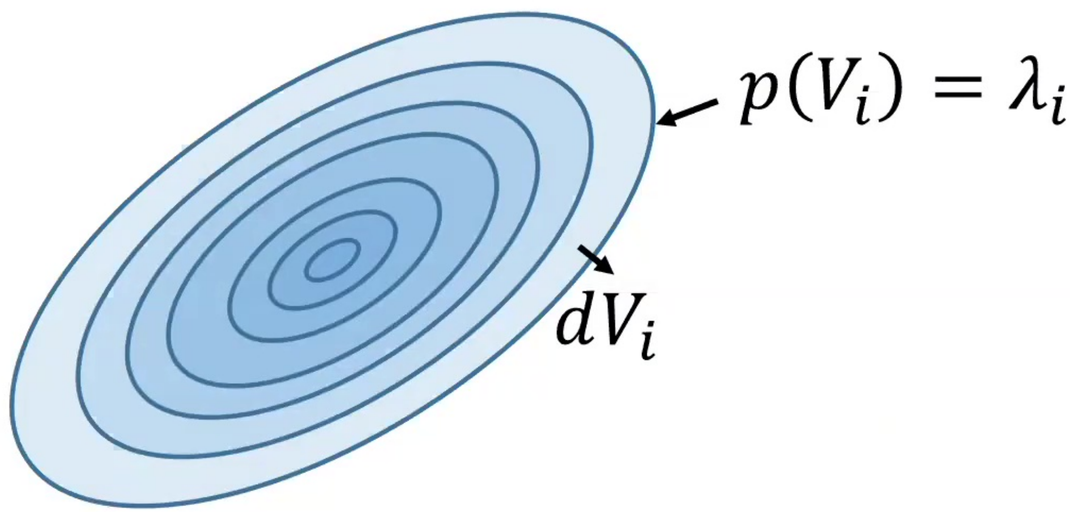


Motivation: Integrating the Posterior

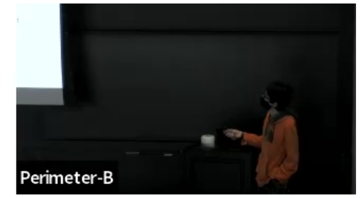
"Amplitude"

$$\int_0^\infty p(V)\,dV(\lambda)$$

"Volume"

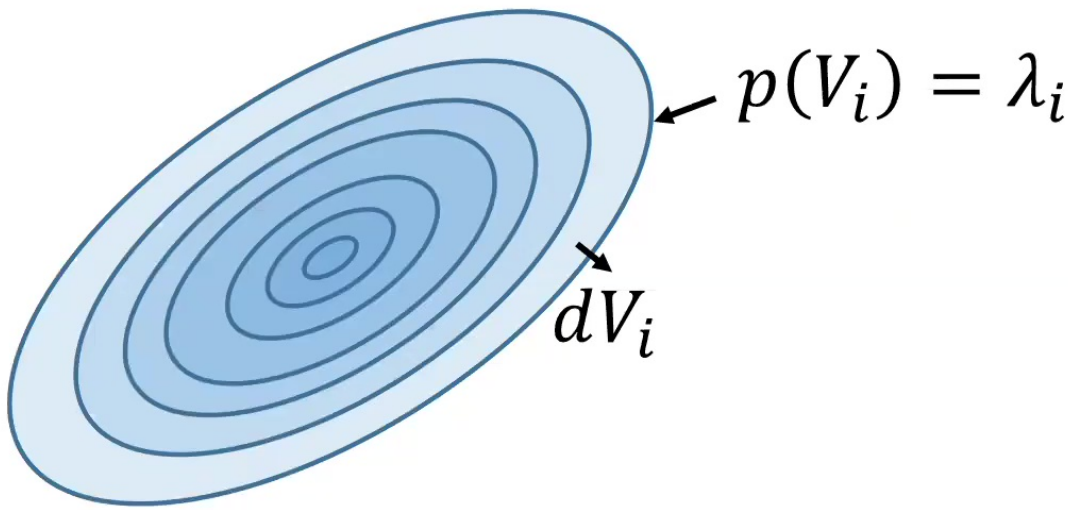


Motivation: Integrating the Posterior



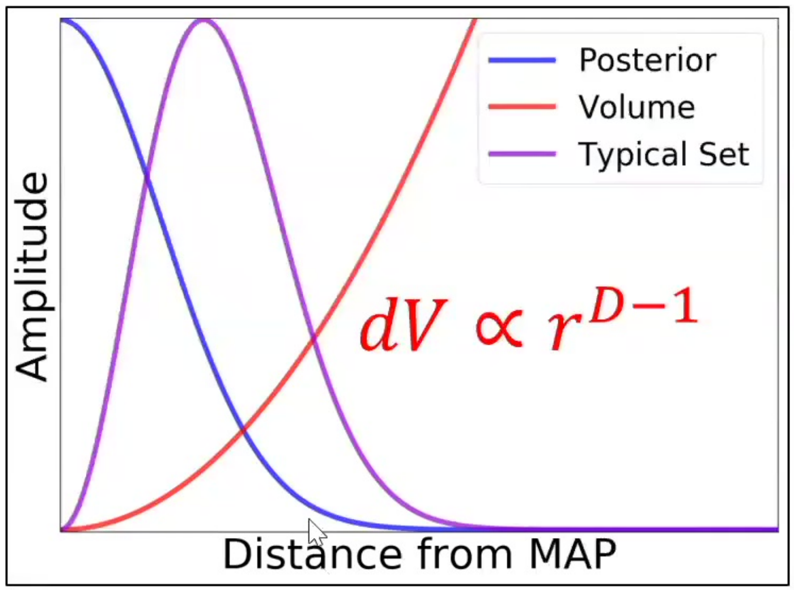
"Typical Set"

$$\int_0^\infty \overbrace{p(V)\,dV(\lambda)}^{\text{"Typical Set"}}$$



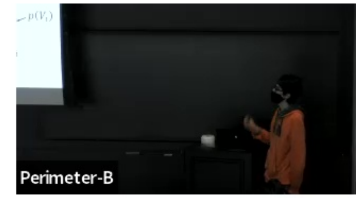
Motivation: Integrating the Posterior

$$\int_0^\infty \overbrace{p(V)\,dV(\lambda)}^{\text{“Typical Set”}}$$

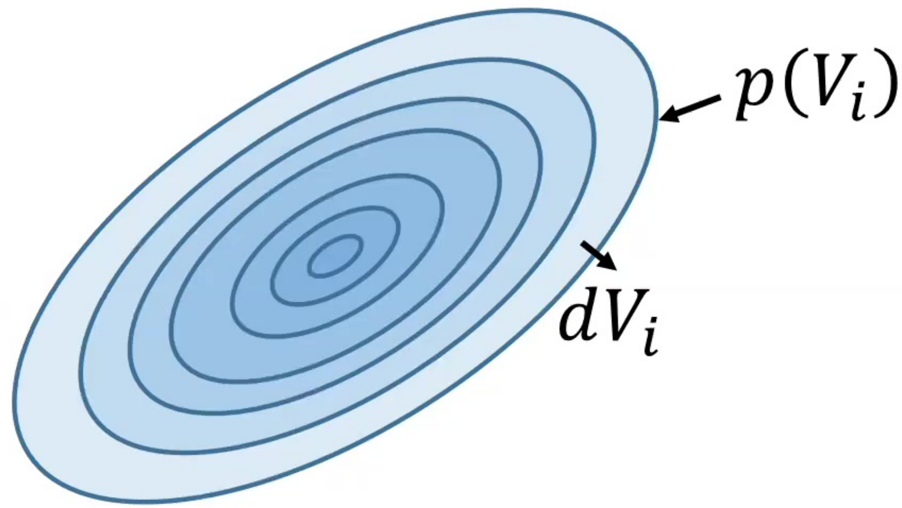


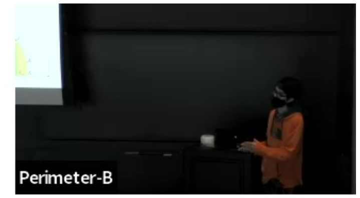
See also Speagle (2019)

Motivation: Integrating the Posterior



$$\mathcal{Z} \equiv \int_{\Omega_{\boldsymbol{\Theta}}} \mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$$



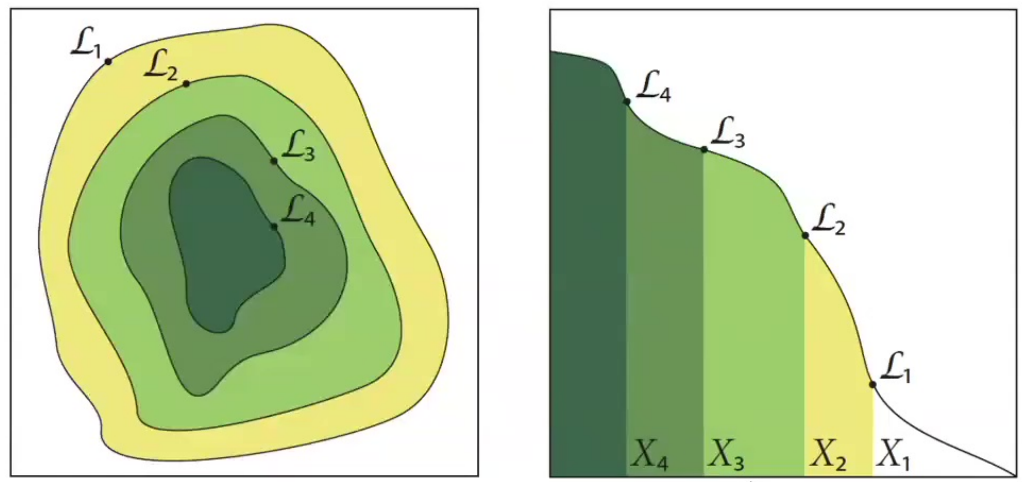


Motivation: Integrating the Posterior

$$\mathcal{Z} \equiv \int_{\Omega_{\boldsymbol{\Theta}}} \mathcal{L}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$$

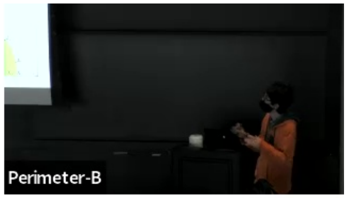
$$X(\lambda) \equiv \int_{\{\boldsymbol{\Theta}:\mathcal{L}(\boldsymbol{\Theta})>\lambda\}} \pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$$

“Prior Volume”



Feroz et al. (2013)

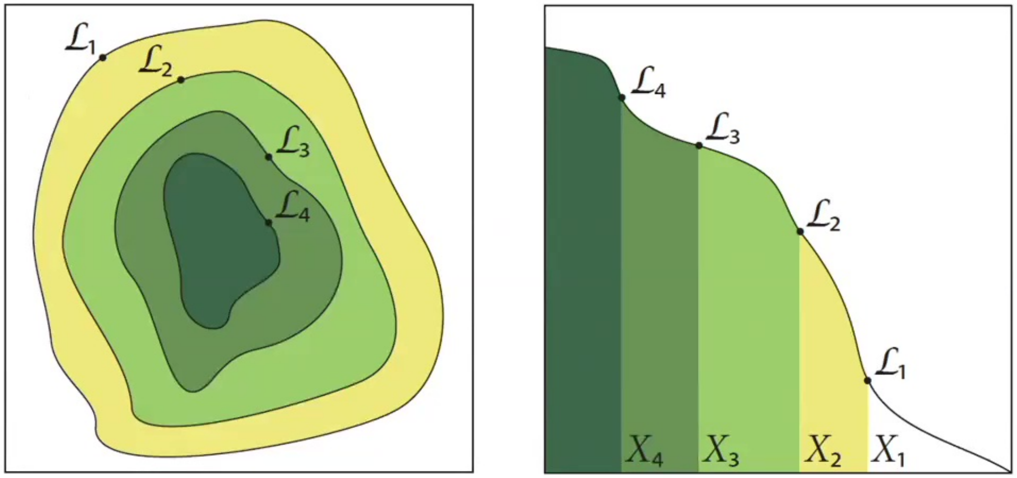
Motivation: Integrating the Posterior



$$\mathcal{Z} = \int_0^1 \mathcal{L}(X)\,dX$$

$$X(\lambda) \equiv \int_{\{\boldsymbol{\Theta}:\mathcal{L}(\boldsymbol{\Theta})>\lambda\}} \pi(\boldsymbol{\Theta})\,d\boldsymbol{\Theta}$$

“Prior Volume”



Feroz et al. (2013)

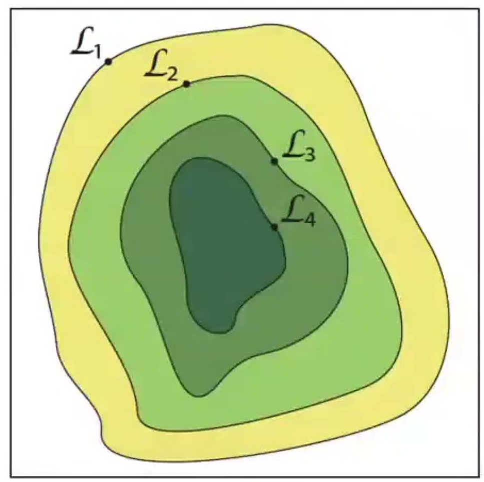
Motivation: Integrating the Posterior

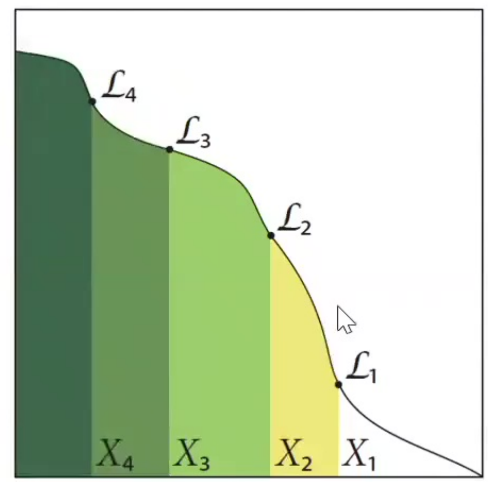
Assumes bijective relationship!
(i.e. no likelihood “plateaus”)

$$\mathcal{Z} = \int_0^1 \boxed{\mathcal{L}(X)} dX$$

$$X(\lambda) \equiv \int_{\{\mathbf{\Theta}:\mathcal{L}(\mathbf{\Theta})>\lambda\}} \pi(\mathbf{\Theta}) d\mathbf{\Theta}$$

“Prior Volume”





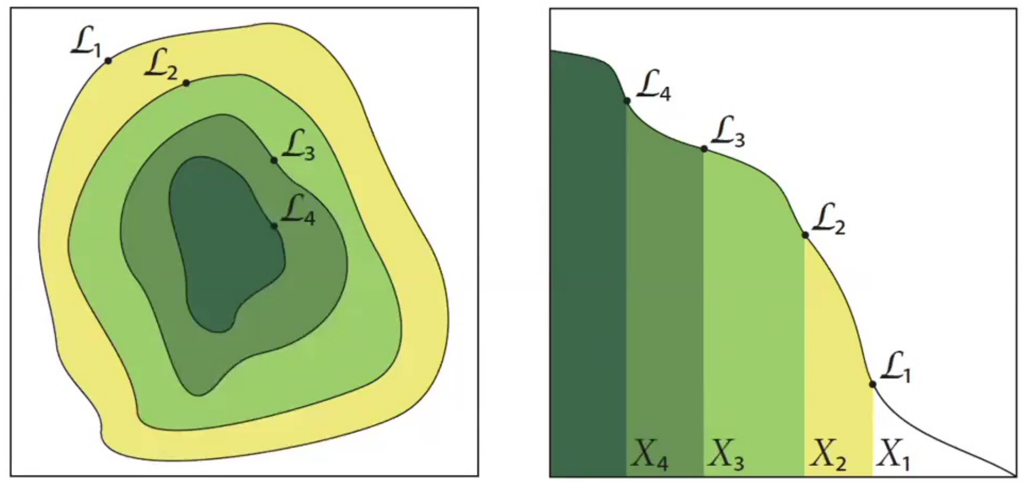
Feroz et al. (2013)

Motivation: Integrating the Posterior

$$\mathcal{Z} \approx \sum_{i=1}^{n} \mathcal{L}_i \times \Delta X_i$$

$$X(\lambda) \equiv \int_{\{\boldsymbol{\Theta}:\mathcal{L}(\boldsymbol{\Theta})>\lambda\}} \pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$$

"Prior Volume"



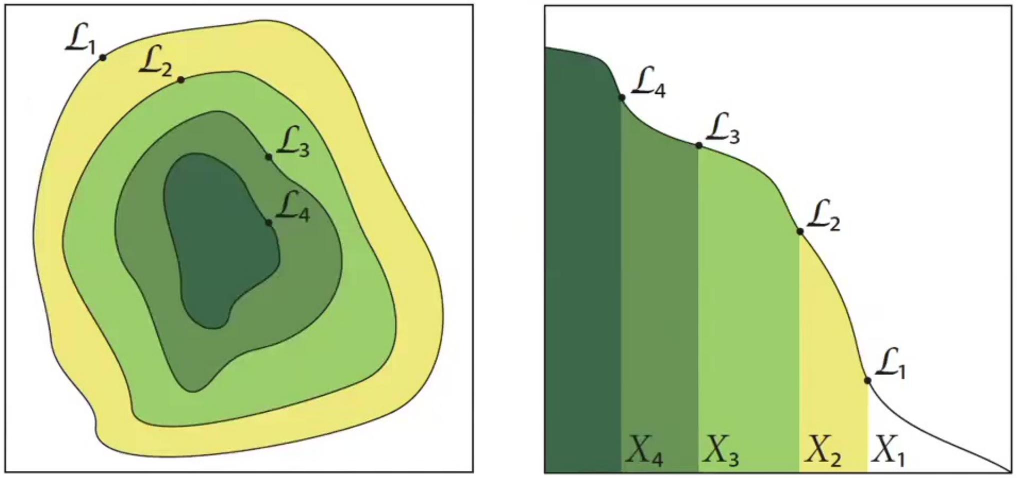
Feroz et al. (2013)

Motivation: Integrating the Posterior

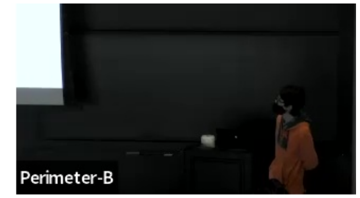
$$\hat{Z} \approx \sum_{i=1}^{n} \mathcal{L}_i \times \boxed{\Delta \widehat{X}_i}$$

$$X(\lambda) \equiv \int_{\{\boldsymbol{\Theta}:\mathcal{L}(\boldsymbol{\Theta})>\lambda\}} \pi(\boldsymbol{\Theta}) d\boldsymbol{\Theta}$$

"Prior Volume"



Feroz et al. (2013)



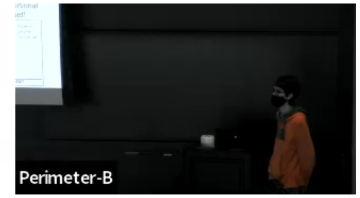
Motivation: Integrating the Posterior

$$\hat{Z} \approx \sum_{i=1}^{n} \widehat{w}_i$$

We get posteriors “for free”

Importance Weight : $\hat{p}_i = \dfrac{\widehat{w}_i}{\hat{Z}}$

Motivation: Integrating the Posterior

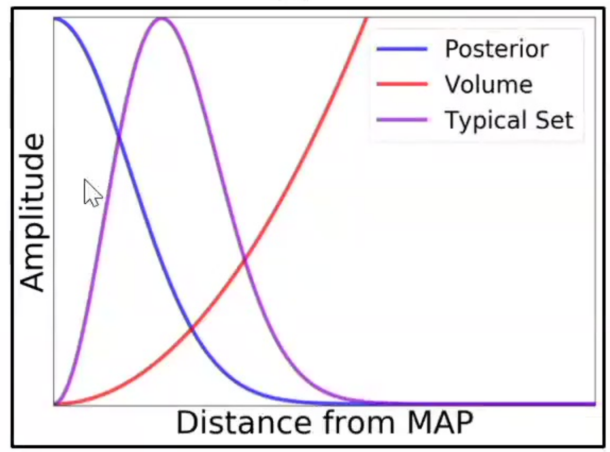


$$\hat{Z} \approx \sum_{i=1}^{n} \hat{w}_i$$

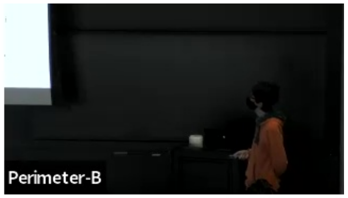
We get posteriors "for free"

Importance Weight : $\hat{p}_i = \dfrac{\hat{w}_i}{\hat{Z}}$

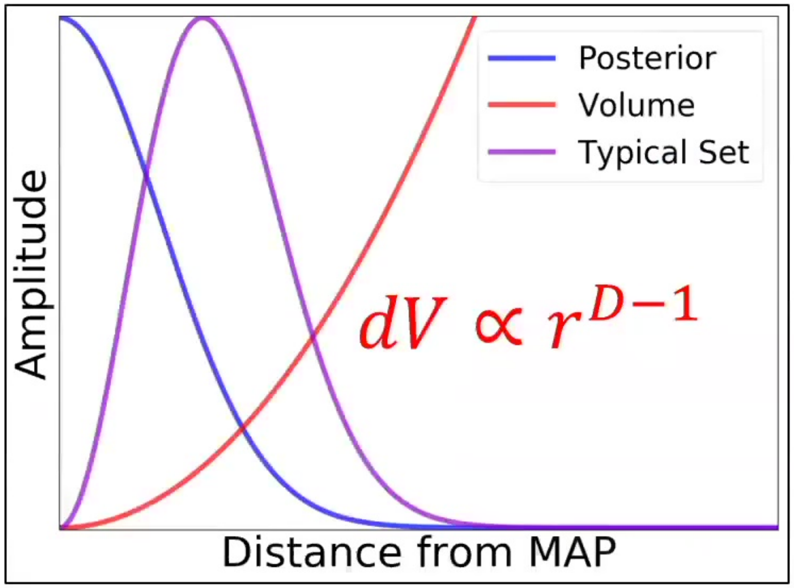
Weights are proportional to the typical set!



Motivation: Integrating the Posterior

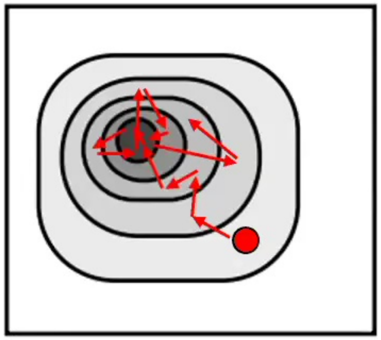


$$\int_0^\infty \overbrace{p(V)\,dV(\lambda)}^{\text{“Typical Set”}}$$



See also Speagle (2019)

Motivation: Sampling the Posterior



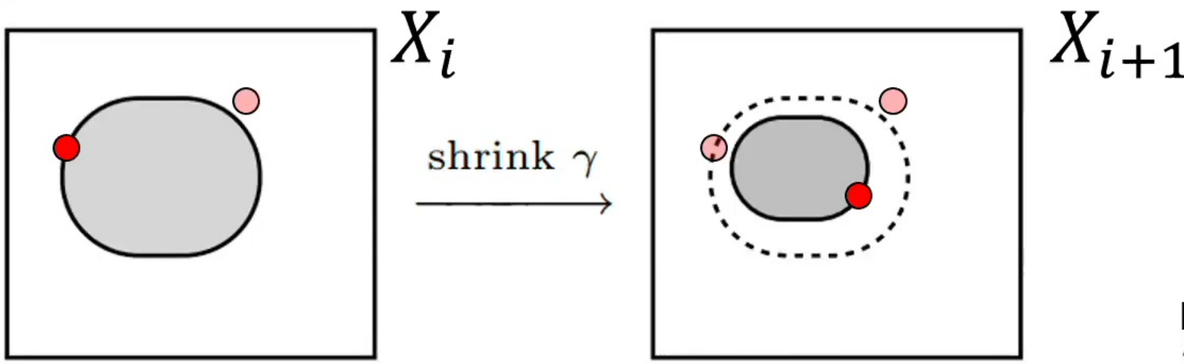
MCMC: Solving a Hard Problem once.

vs

Nested Sampling: Solving an Easier Problem many times.

Sampling uniformly within $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ bound is **easier**.

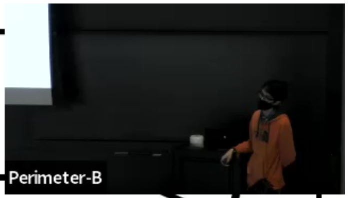
If you have a **prior transform** that converts your priors to look uniform, then this case is equivalent.



Pictures adapted from this 2010 talk by John Skilling.

Stopping Criteria

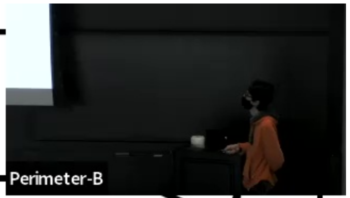
$$\frac{\hat{\mathcal{Z}}_i}{\hat{\mathcal{Z}}_i + \mathcal{Z}_{\text{in}}}$$



$$\hat{\mathcal{Z}} = \hat{\mathcal{Z}}_i + \mathcal{Z}_{\text{in}}$$

Stopping Criteria

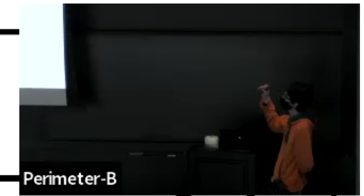
$$\frac{\hat{Z}_i}{\hat{Z}_i + Z_{\text{in}}}$$



$$\hat{Z} = \sum_{j=1}^{i} \widehat{w}_j + Z_{\text{in}}$$

Stopping Criteria

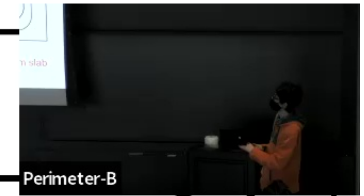
$$\frac{\hat{Z}_i}{\hat{Z}_i + Z_{\text{in}}} < f$$



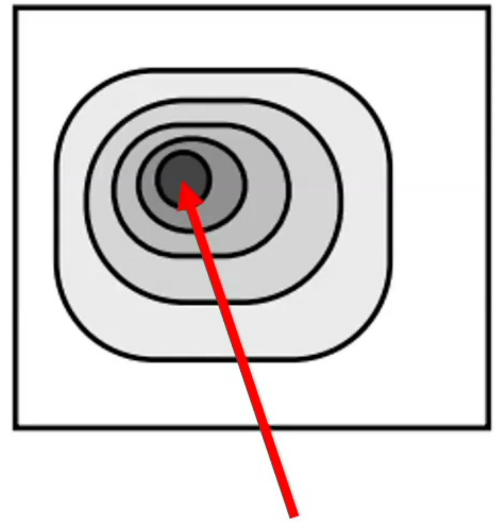
$$\hat{Z} \leq \sum_{j=1}^{i} \hat{w}_j + \mathcal{L}_{\text{max}} X_n$$

Stopping Criteria

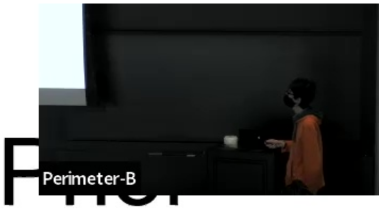
$$\frac{\hat{\mathcal{Z}}_i}{\hat{\mathcal{Z}}_i + \mathcal{Z}_{\text{in}}} <$$

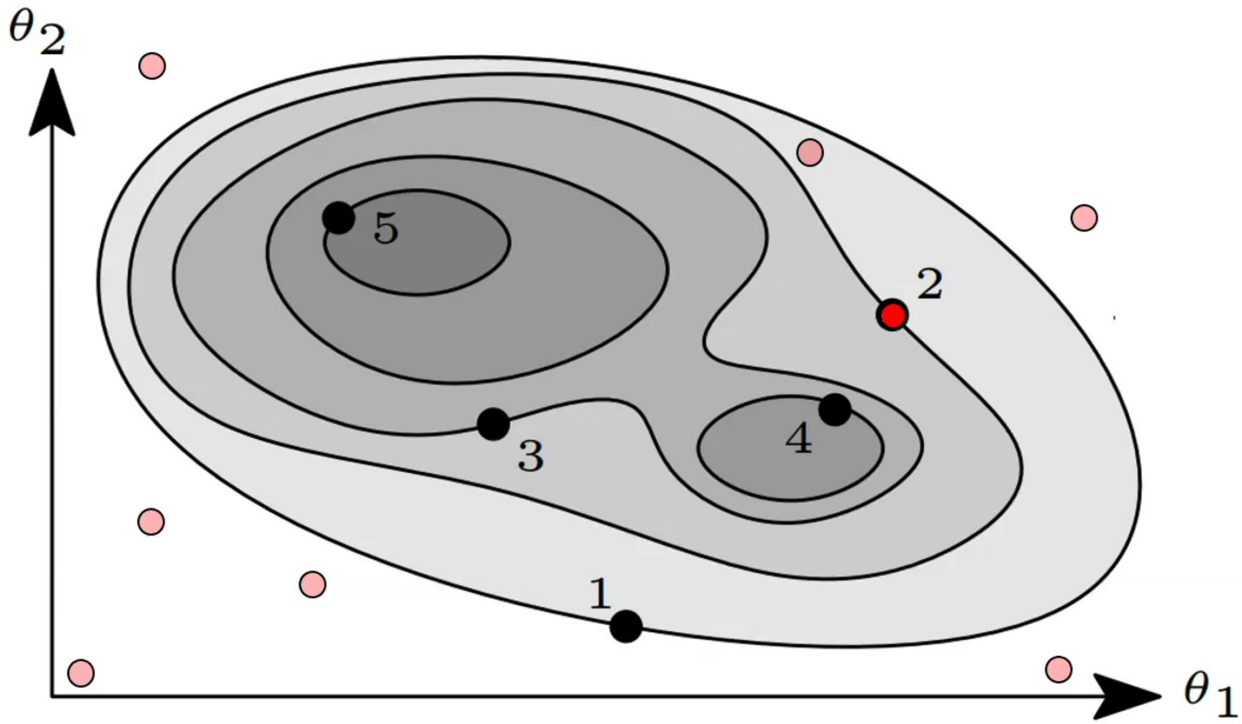


$$\hat{\mathcal{Z}} \lesssim \sum_{j=1}^{i} \widehat{w}_j + \hat{\mathcal{L}}_{\max}\hat{X}_n$$



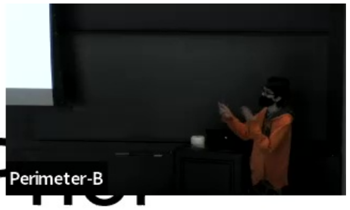
Naïve Approach: Sampling from the Prior

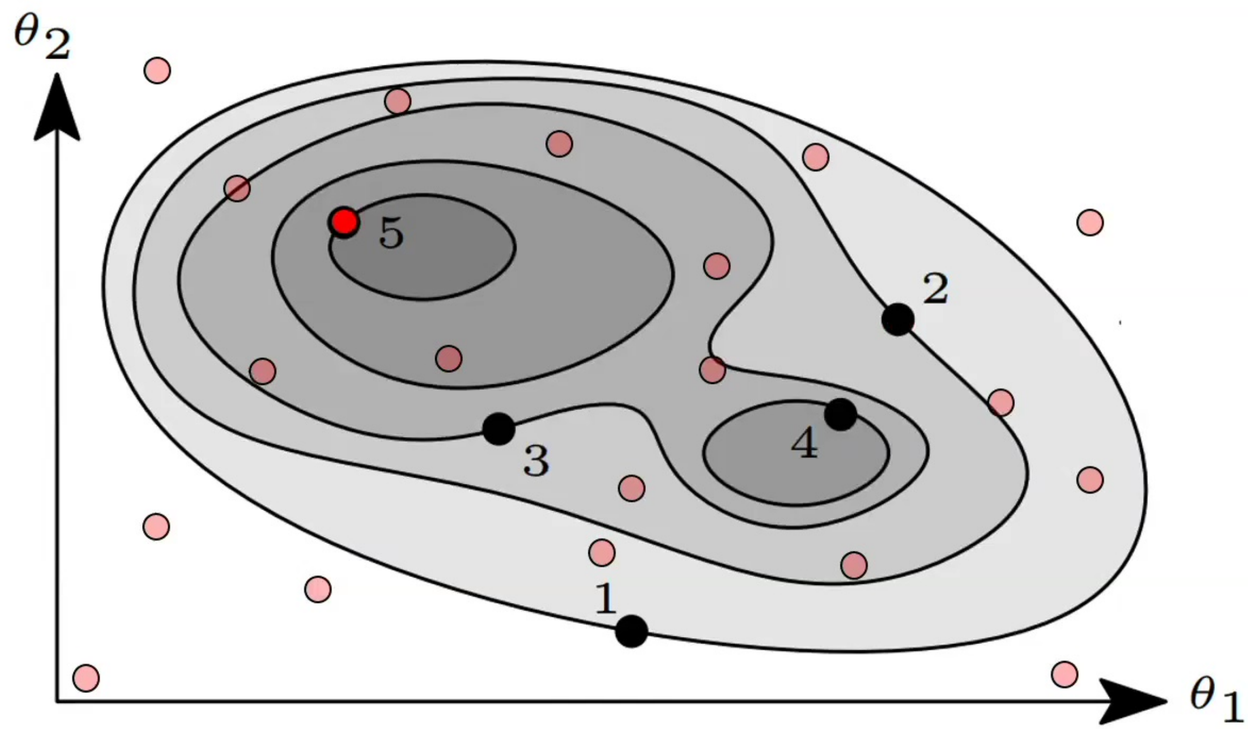




Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

Naïve Approach: Sampling from the P

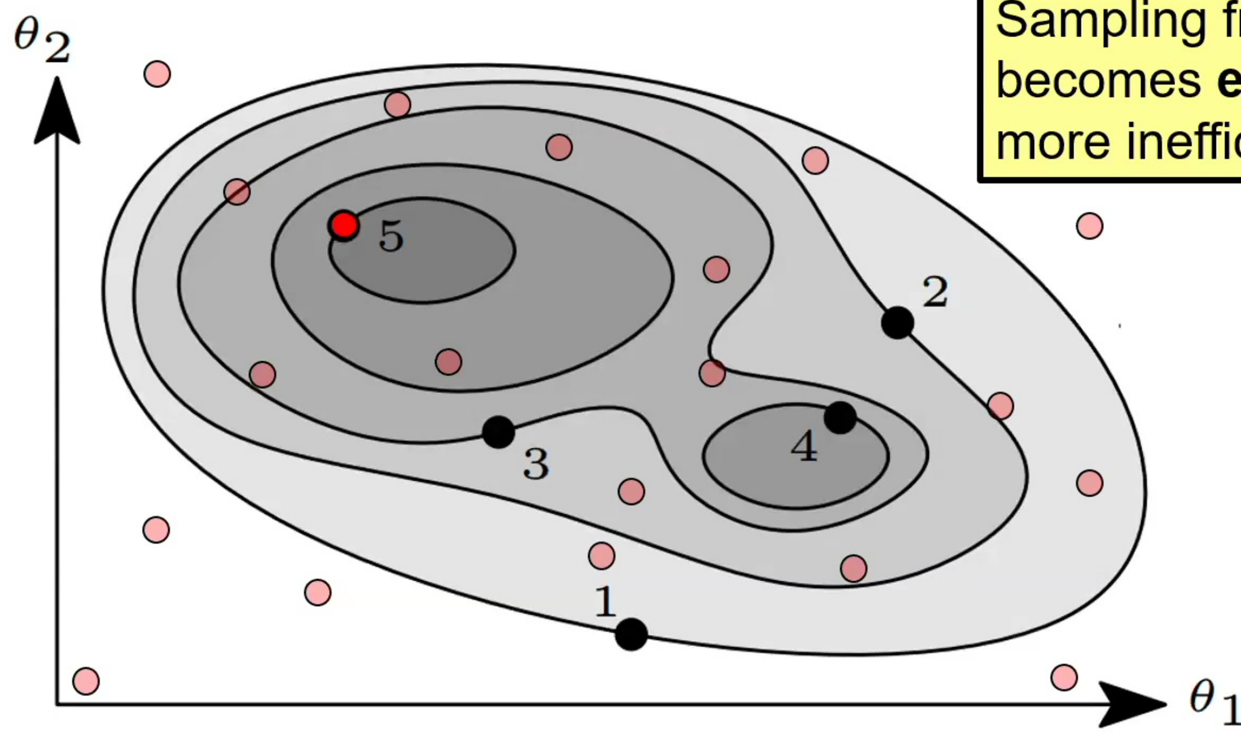




Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

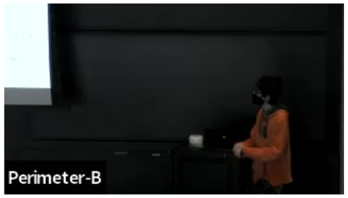
Naïve Approach: Sampling from the Prior

Sampling from the prior becomes **exponentially** more inefficient over time.

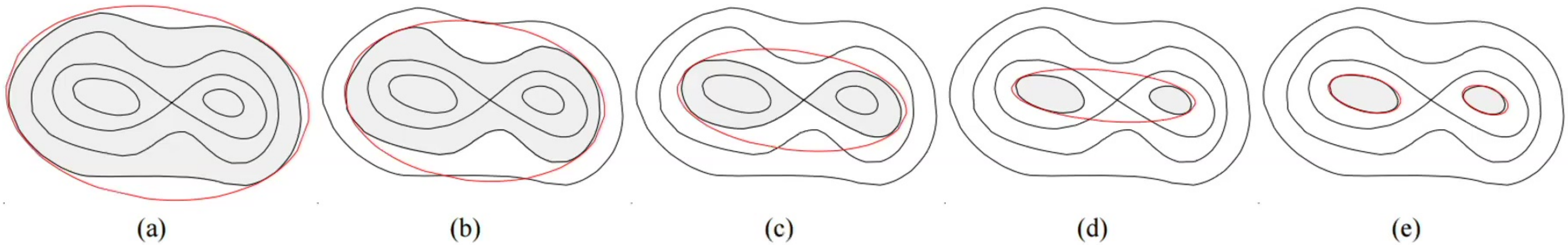


Higson et al. (2017)
arxiv:1704.03459

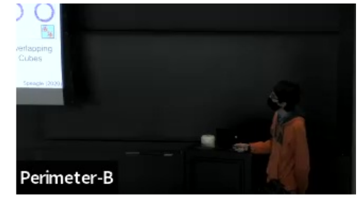
Sampling from the Constrained Prior



Proposal:
Try to bound the iso-likelihood contours in real time.

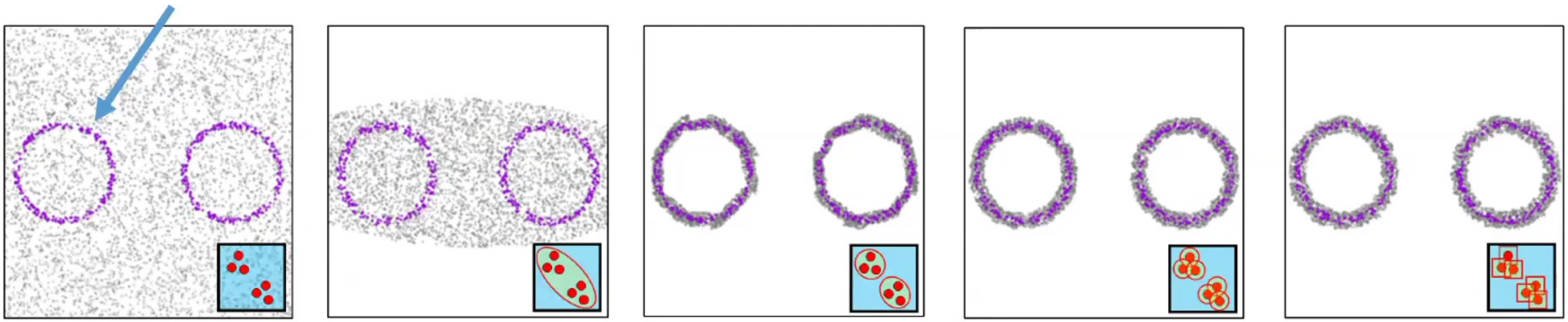


Feroz et al. (2009)



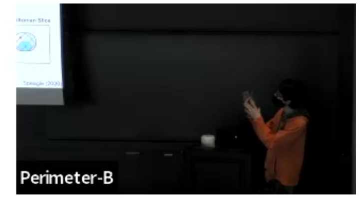
Examples of Bounding Strategies

“Live points” (i.e. “chains”)

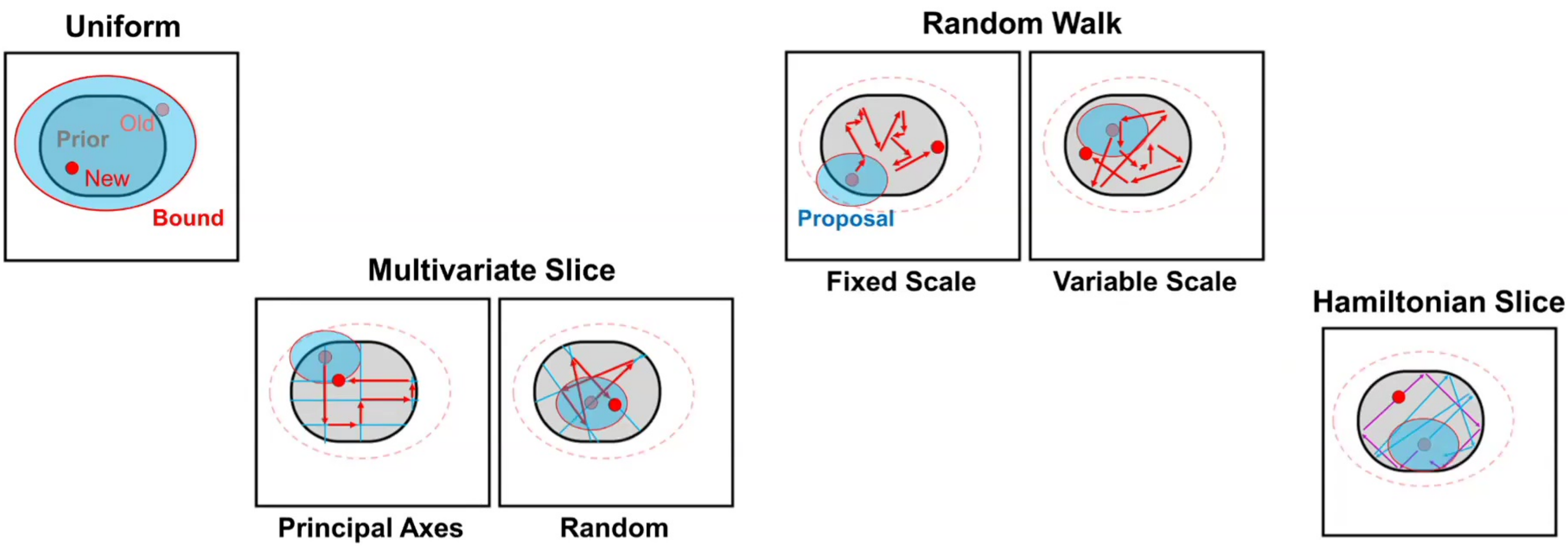


Unit Cube (no bound) | Single Ellipsoid | Multiple Ellipsoids | Overlapping Balls | Overlapping Cubes

Speagle (2020)

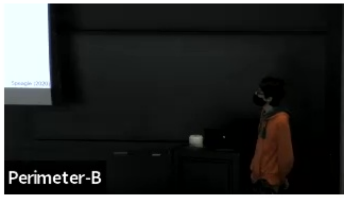


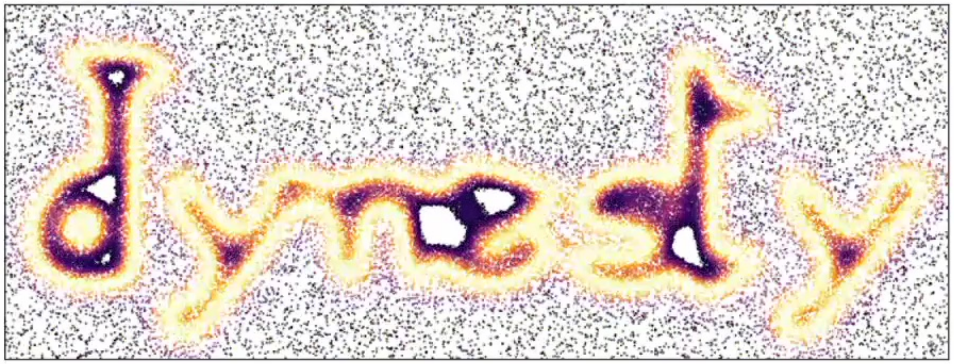
Examples of Sampling Strategies



Speagle (2020)

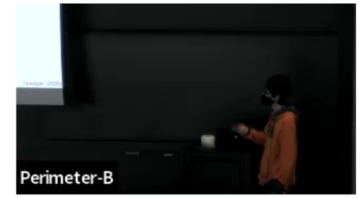
dynesty





https://dynesty.readthedocs.io

Speagle (2020)



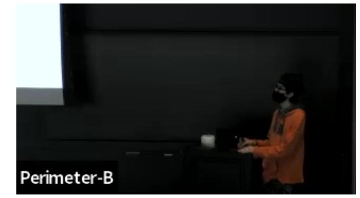
dynesty

- **Open-source Python package** designed to make (Dynamic) Nested Sampling easy to use but also easy to customize.
- Designed to be **highly modular** and can mix-and-match methods.
- Includes **built-in plotting utilities** and post-processing tools.

https://dynesty.readthedocs.io

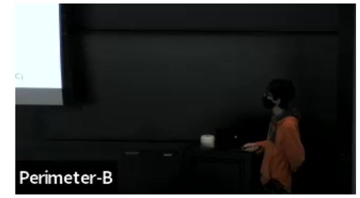
Speagle (2020)

Advantages and Disadvantages



Advantages to Nested Sampling:

Disadvantages to Nested Sampling:



Advantages and Disadvantages

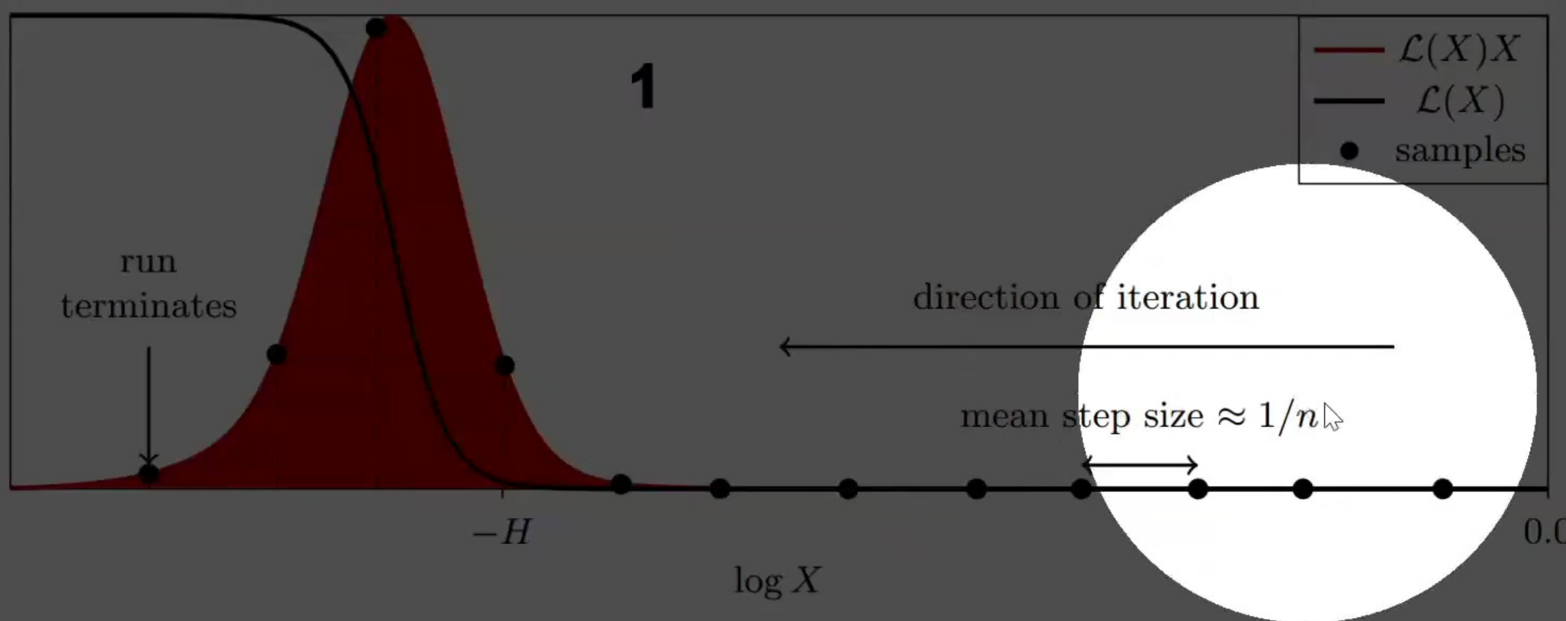
Advantages to Nested Sampling:

1. Can characterize complex uncertainties in real-time.
2. Can allocate samples much more efficiently in some cases.
3. Possesses well-motivated stopping criteria (Skilling 2006; Speagle 2020).
4. Can help perform model selection.

Disadvantages to Nested Sampling:

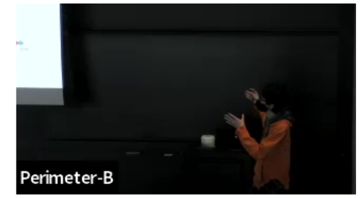
1. Implementations require a prior transform.
2. Runtime sensitive to size of prior.
3. Overall approach can sometimes miss certain types of solutions.
4. Sampling is more involved.
5. Can’t use gradients as “naturally” as Hamiltonian Monte Carlo (HMC).

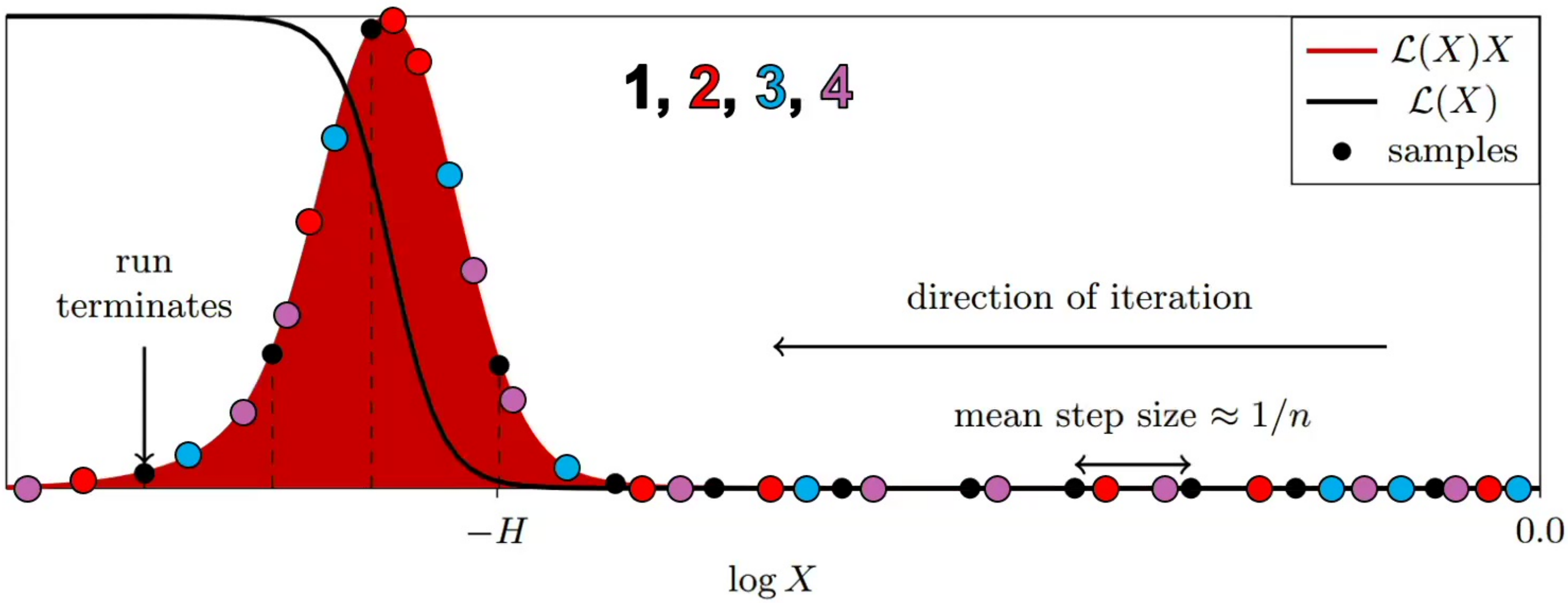
Dynamic Nested Sampling



Higson et al. (2017)
arxiv:1704.03459

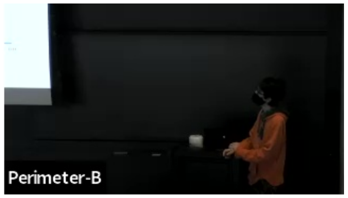
Dynamic Nested Sampling

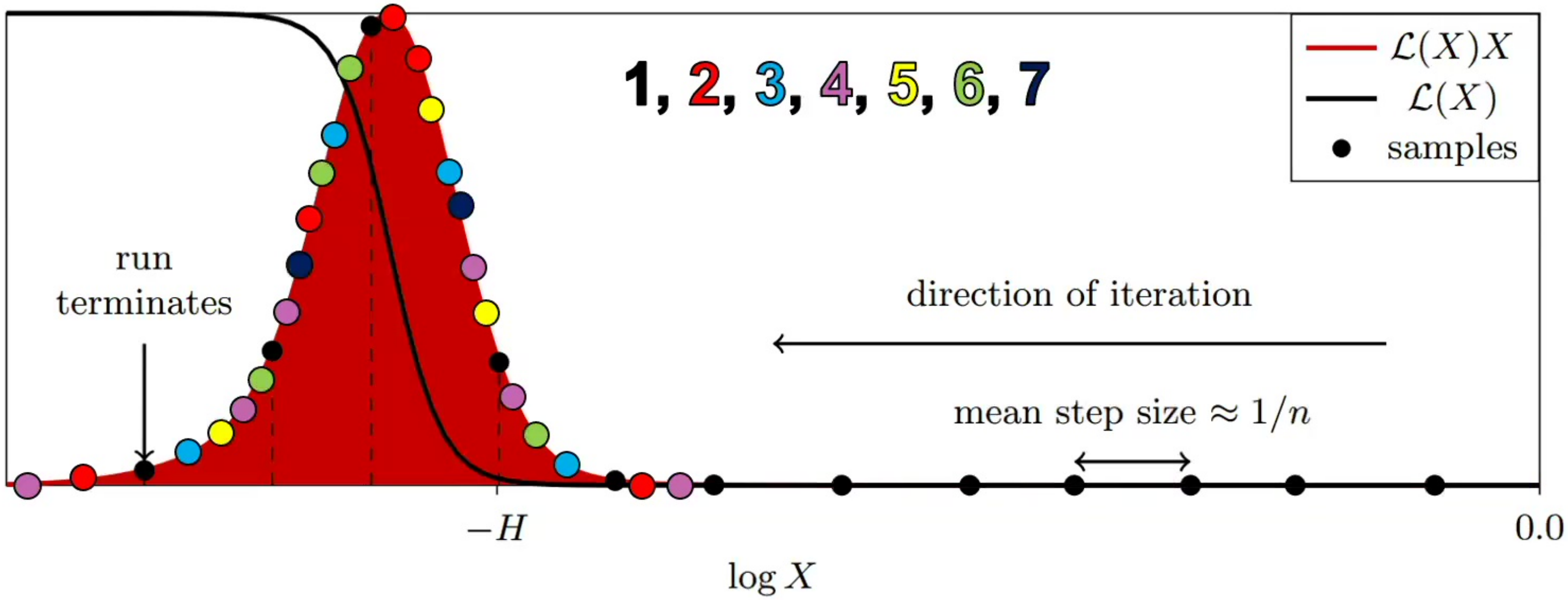




Higson et al. (2017)
arxiv:1704.03459

Dynamic Nested Sampling

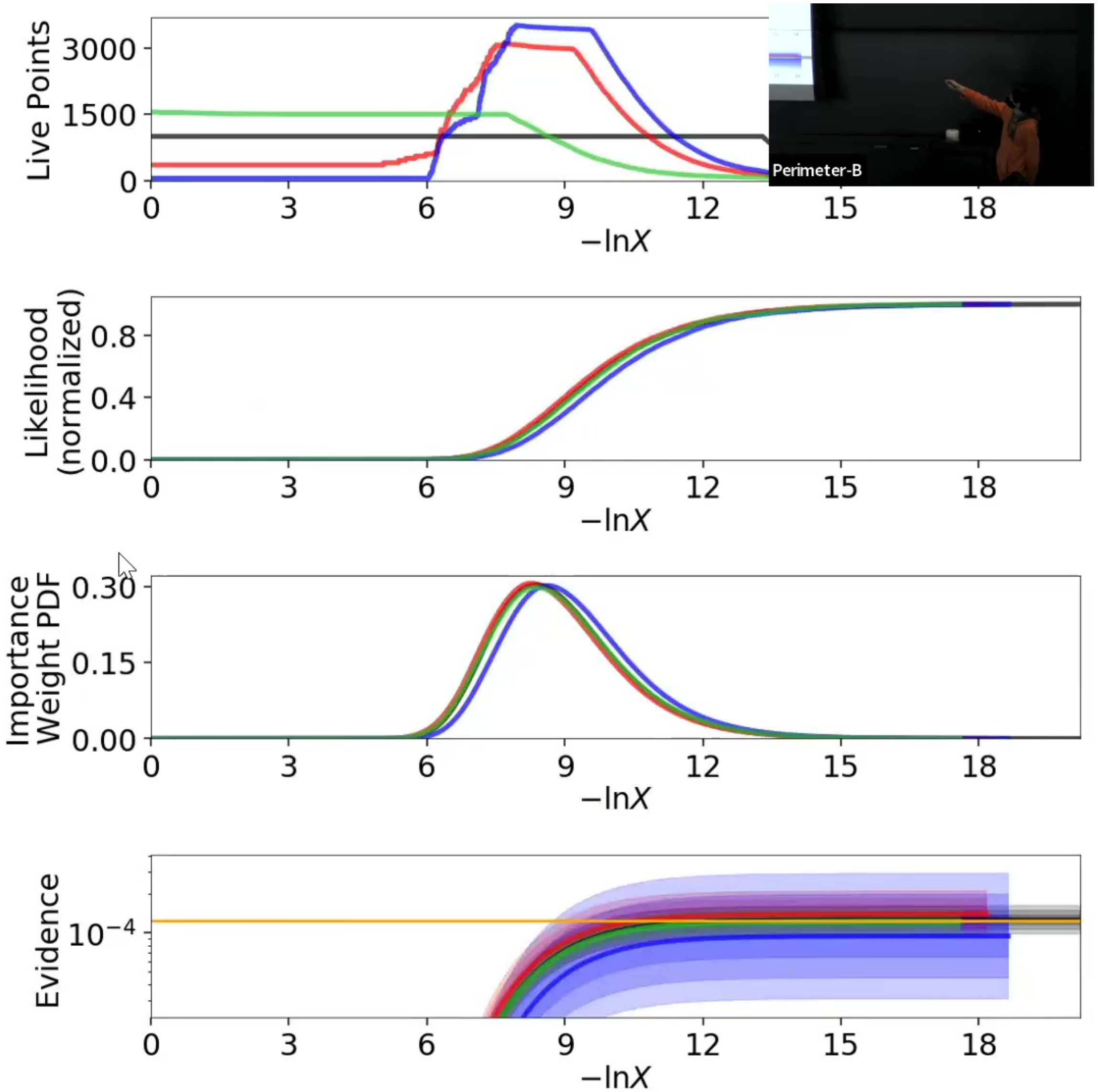




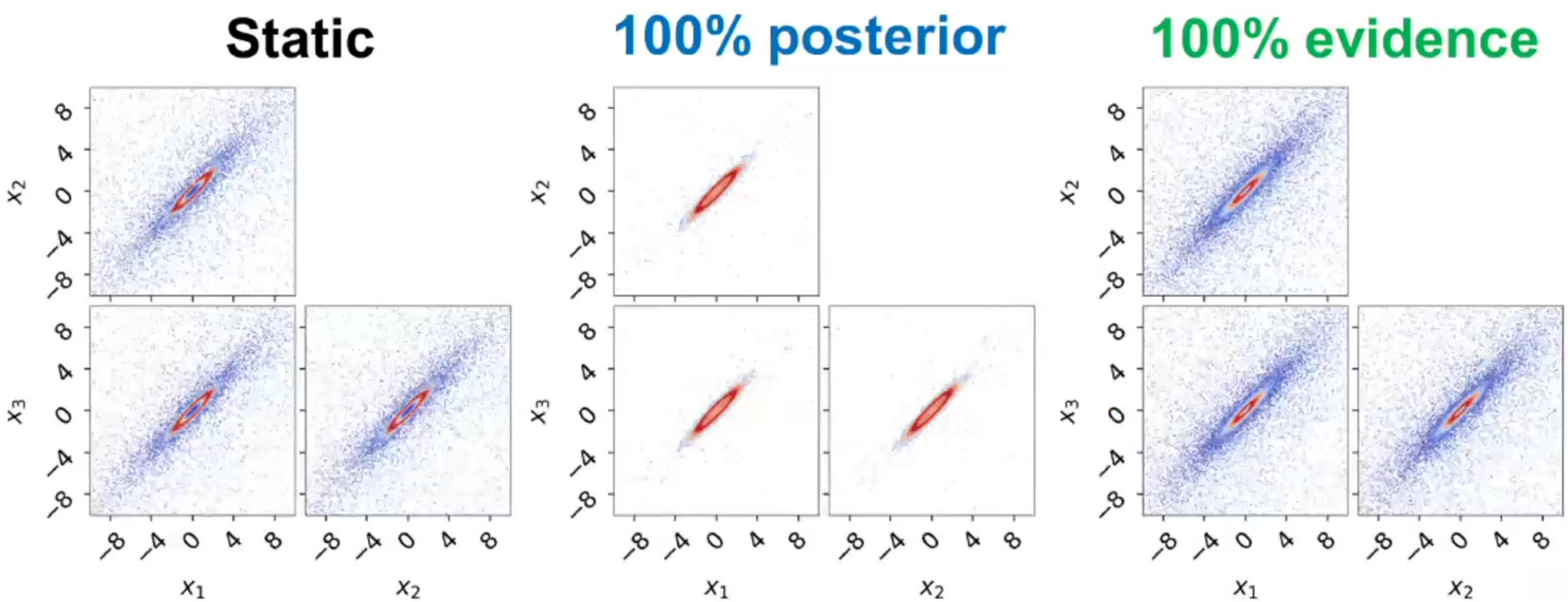
Higson et al. (2017)
arxiv:1704.03459

Comparisons

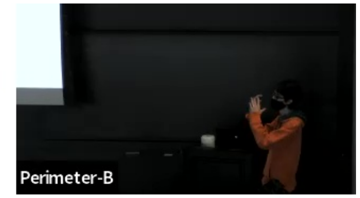
- Fixed number of samples.
- Only change is in overall Dynamic Nested Sampling **strategy**.



Comparisons

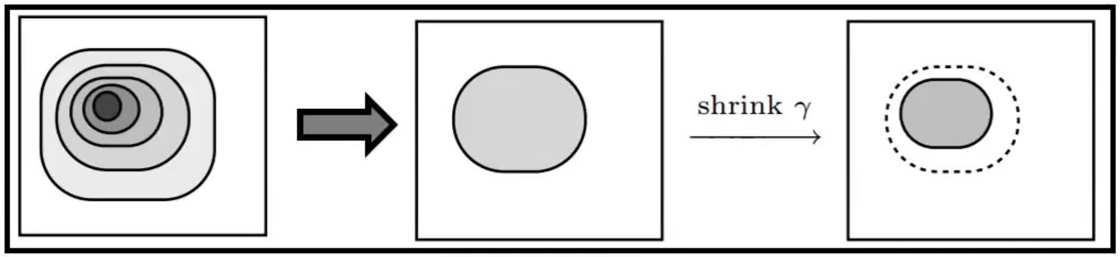


Speagle (2020)

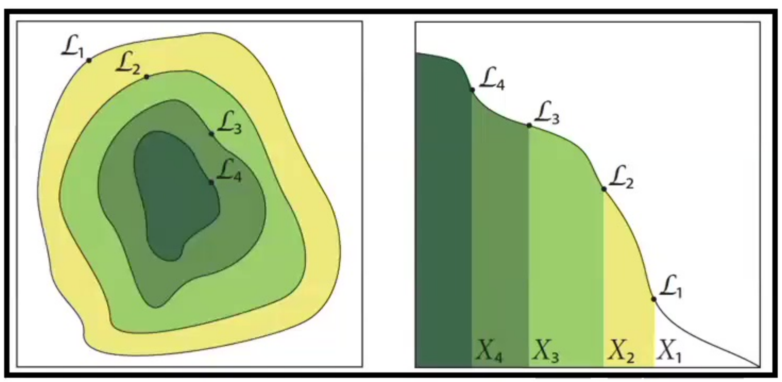


Summary

Core Idea Behind Nested Sampling

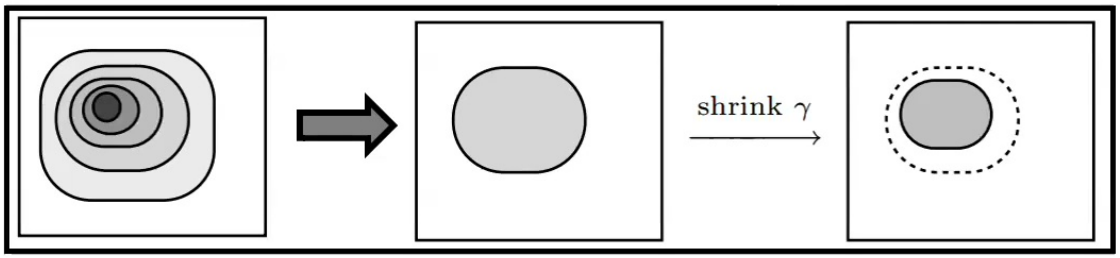


How Nested Sampling Works

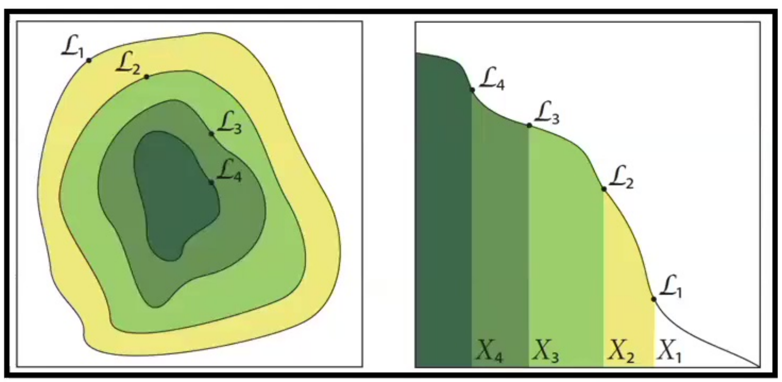


Summary

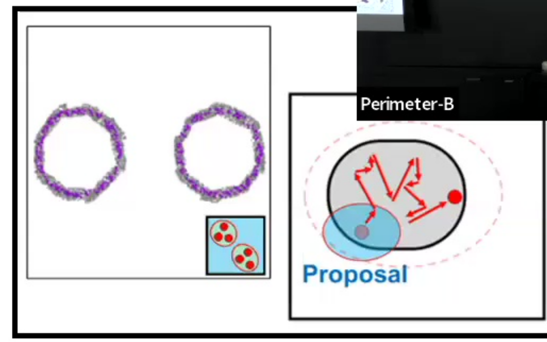
Core Idea Behind Nested Sampling

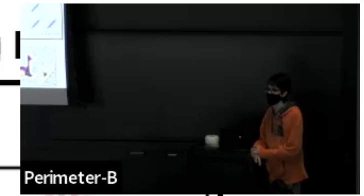


How Nested Sampling Works



Nested Sampling I





Dynamic Nested Sampling