An Introduction to (Dynamic) Nested Sampling

Joshua S. Speagle

Banting & Dunlap Postdoctoral Fellow

Department of Statistical Sciences

David A. Dunlap Department of Astronomy & Astrophysics

Dunlap Institute for Astronomy & Astrophysics

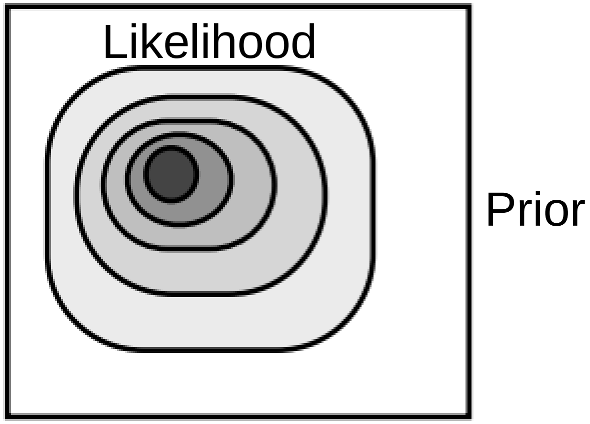
UNIVERSITY OF
TORONTO

Background

$$\underset{\text{Posterior}}{\Pr(\boldsymbol{\Theta}|\mathbf{D},\mathrm{M})} = \frac{\overset{\text{Likelihood}}{\Pr(\mathbf{D}|\boldsymbol{\Theta},\mathrm{M})}\ \overset{\text{Prior}}{\Pr(\boldsymbol{\Theta}|\mathrm{M})}}{\underset{\text{Evidence}}{\Pr(\mathbf{D}|\mathrm{M})}}$$

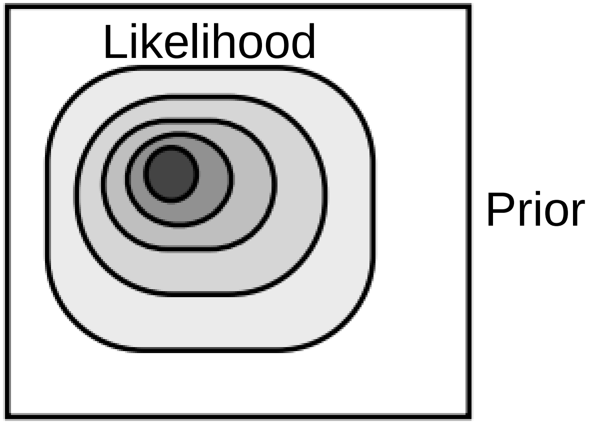
Bayes' Theorem

Motivation: Sampling the Posterior



Pictures adapted from this 2010 talk by John Skilling.

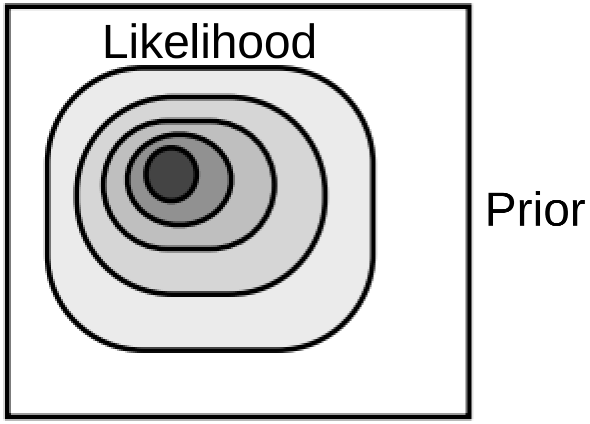
Motivation: Sampling the Posterior



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

Pictures adapted from [this 2010 talk](#)by John Skilling.

Motivation: Sampling the Posterior



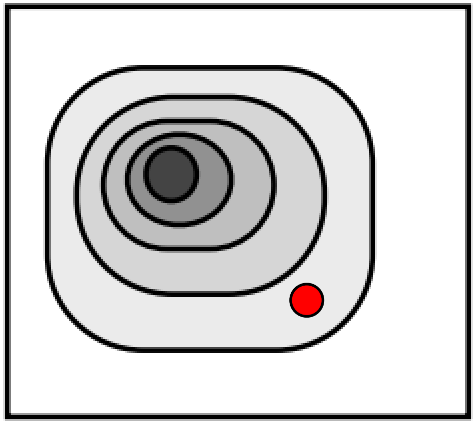
Sampling directly from the likelihood $\mathcal{L}(\boldsymbol{\Theta})$ is **hard**.

MCMC: Solving a Hard Problem once.

(Markov Chain Monte Carlo)

Pictures adapted from [this 2010 talk](#) by John Skilling.

Motivation: Sampling the Posterior



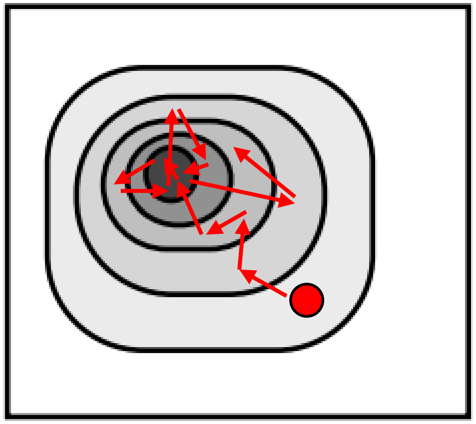
Sampling directly from the likelihood $\mathcal{L}(\boldsymbol{\Theta})$ is **hard**.

MCMC: Solving a Hard Problem once.

(Markov Chain Monte Carlo)

Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior



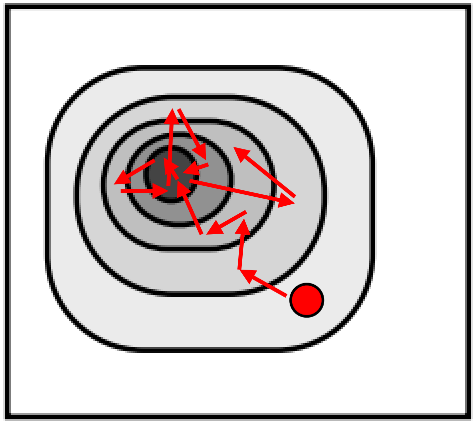
Sampling directly from the likelihood $\mathcal{L}(\mathbf{\Theta})$ is **hard**.

MCMC: Solving a Hard Problem once.

(Markov Chain Monte Carlo)

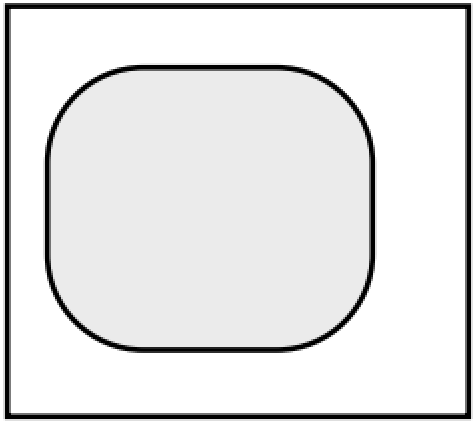
Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior



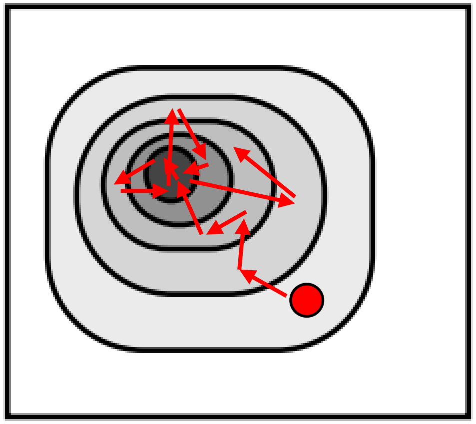
MCMC: Solving a Hard Problem once.
vs
Nested Sampling: Solving an Easier Problem many times.

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



Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior



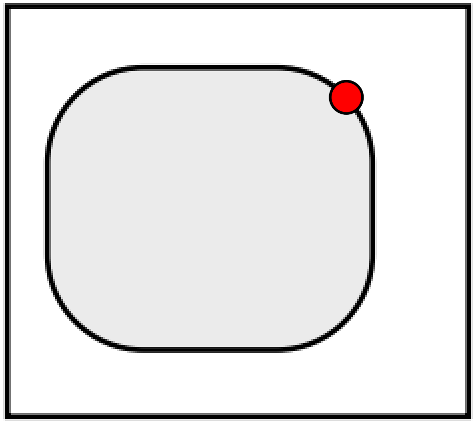
MCMC: Solving a Hard Problem once.

vs

Nested Sampling: Solving an Easier Problem many times.

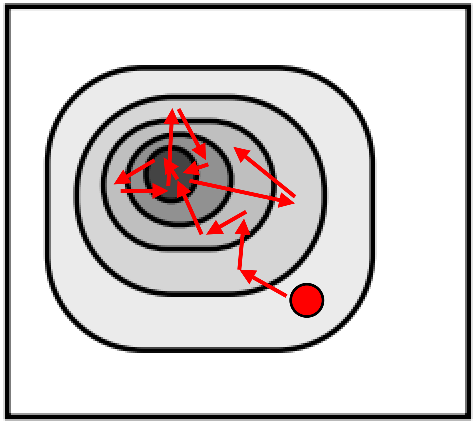
Sampling uniformly within bound $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ is **easier**.

X_{i-1}



Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior

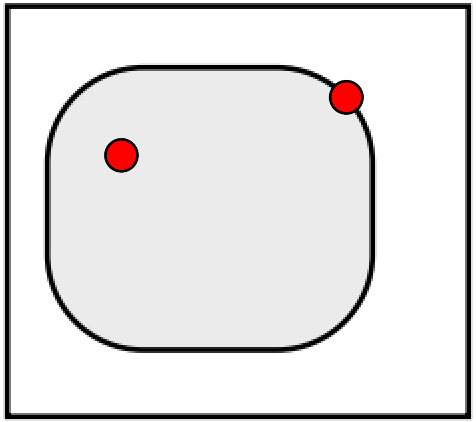


MCMC: Solving a Hard Problem once.

vs

Nested Sampling: Solving an Easier Problem many times.

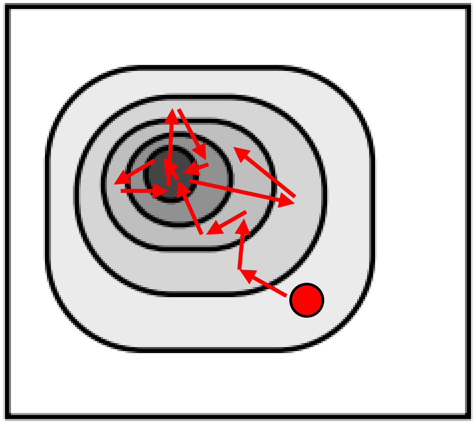
Sampling uniformly within bound $\mathcal{L}(\mathbf{\Theta}) > \lambda$ is **easier**.



X_{i-1}

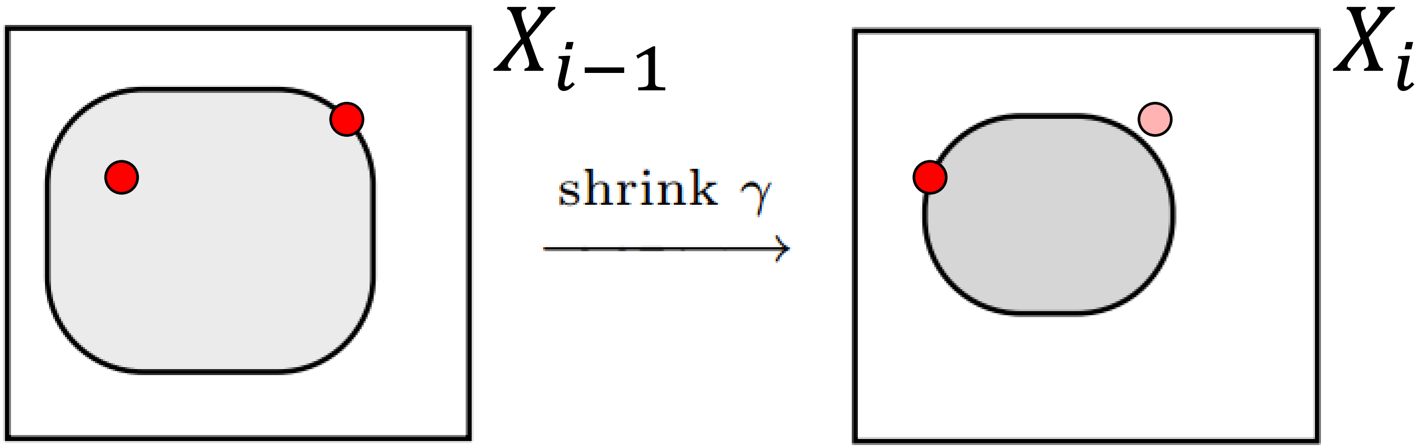
Pictures adapted from [this 2010 talk](#) by John Skilling.

Motivation: Sampling the Posterior



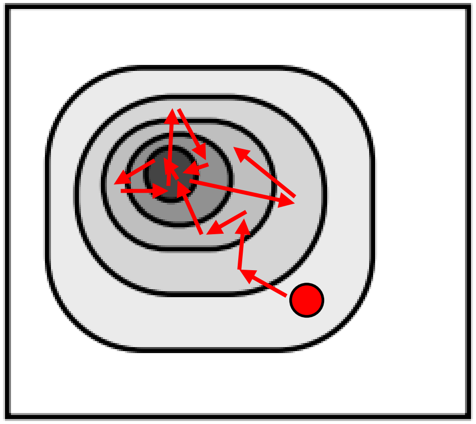
MCMC: Solving a Hard Problem once.
vs
Nested Sampling: Solving an Easier Problem many times.

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



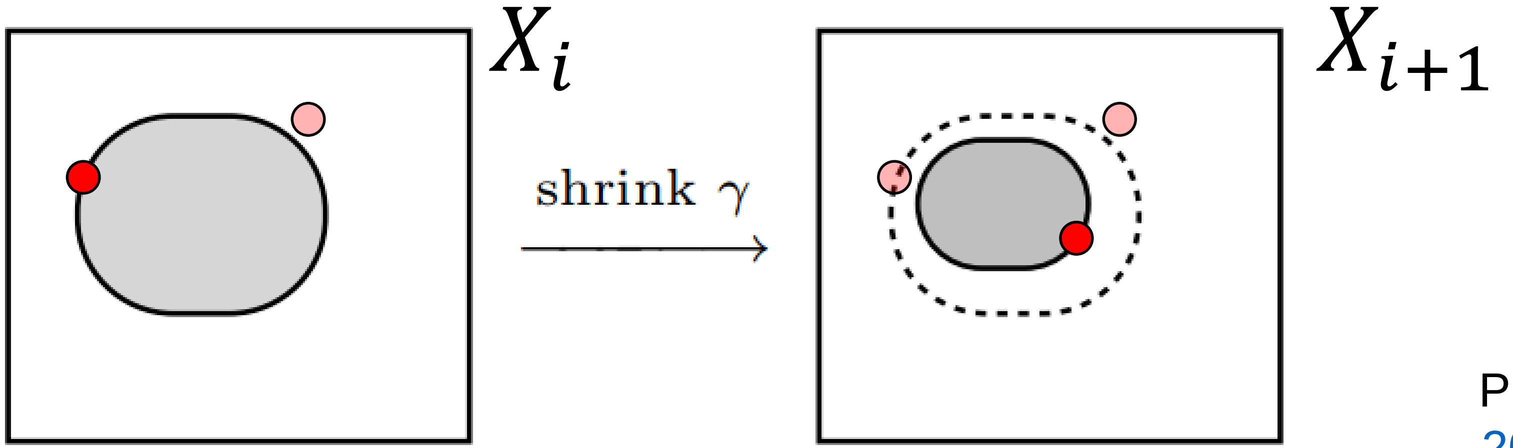
Pictures adapted from this 2010 talk by John Skilling.

Motivation: Sampling the Posterior



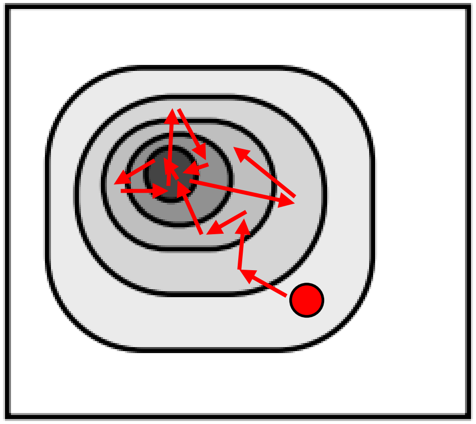
MCMC: Solving a Hard Problem once.
vs
Nested Sampling: Solving an Easier Problem many times.

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



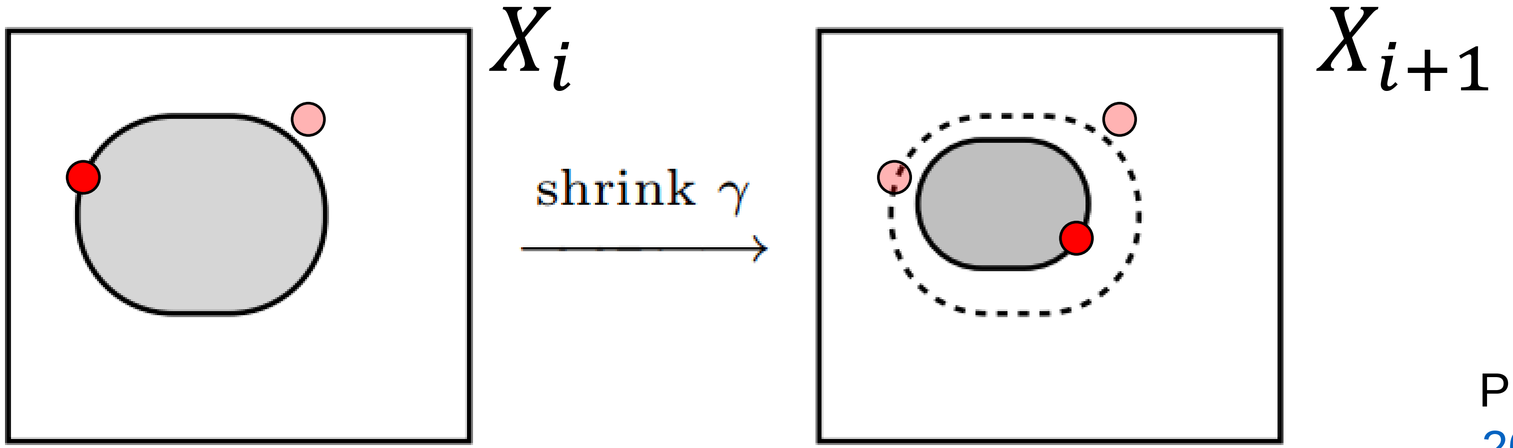
Pictures adapted from [this 2010 talk](#) by John Skilling.

Motivation: Sampling the Posterior



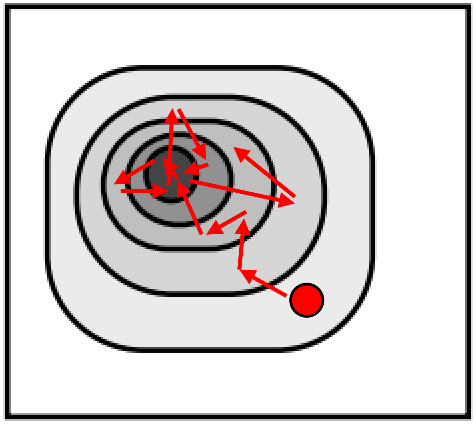
MCMC: Solving a Hard Problem once.
vs
Nested Sampling: Solving an Easier Problem many times.

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



Pictures adapted from [this 2010 talk] by John Skilling.

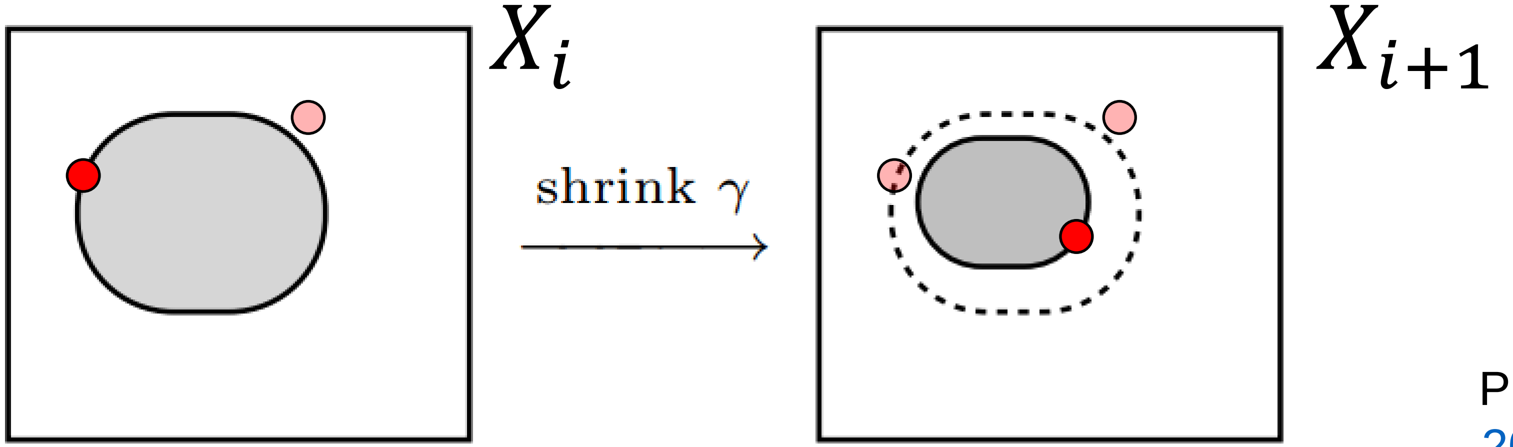
Motivation: Sampling the Posterior



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

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

Motivation: Integrating the Posterior

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

Bayes' Theorem

Motivation: Integrating the Posterior

$$\underbrace{p(\boldsymbol{\Theta})}^{\text{Posterior}} = \frac{\overbrace{\mathcal{L}(\boldsymbol{\Theta})}^{\text{Likelihood}}\,\overbrace{\pi(\boldsymbol{\Theta})}^{\text{Prior}}}{\mathcal{Z}}$$

Bayes' Theorem

Motivation: Integrating the Posterior

$$\overset{\text{Posterior}}{p(\boldsymbol{\Theta})} = \frac{\overset{\text{Likelihood}}{\mathcal{L}(\boldsymbol{\Theta})}\,\overset{\text{Prior}}{\pi(\boldsymbol{\Theta})}}{\mathcal{Z}}$$

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

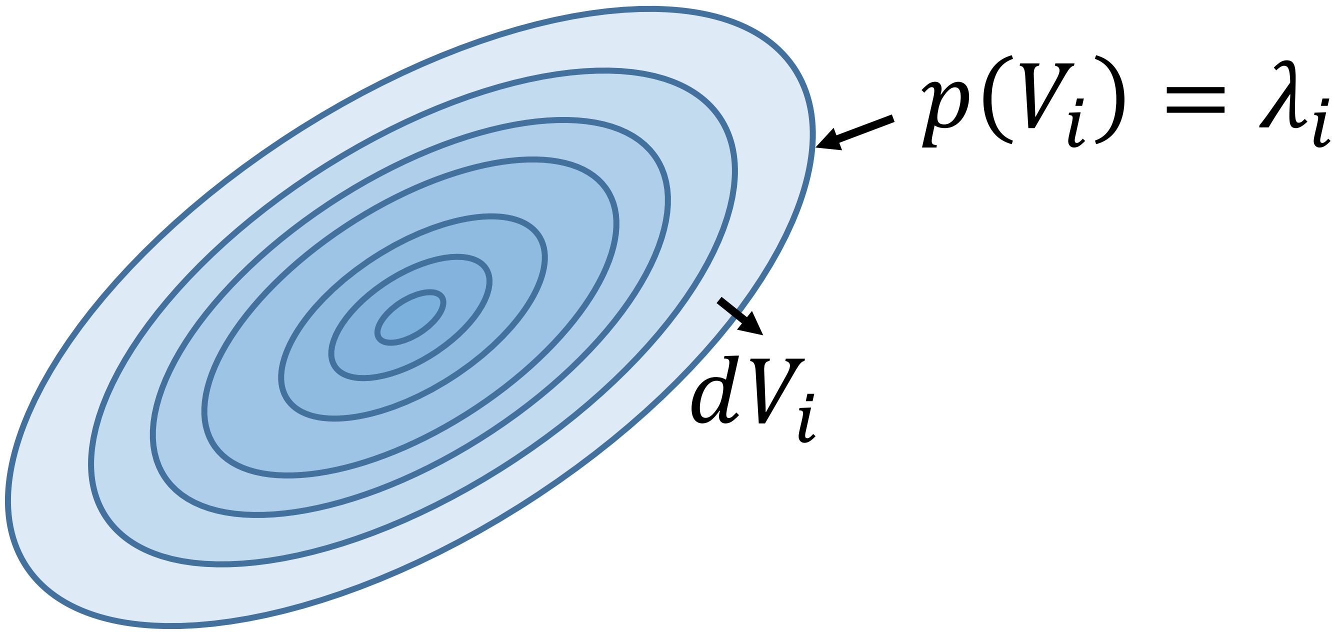
Bayes' Theorem

Motivation: Integrating the Posterior

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

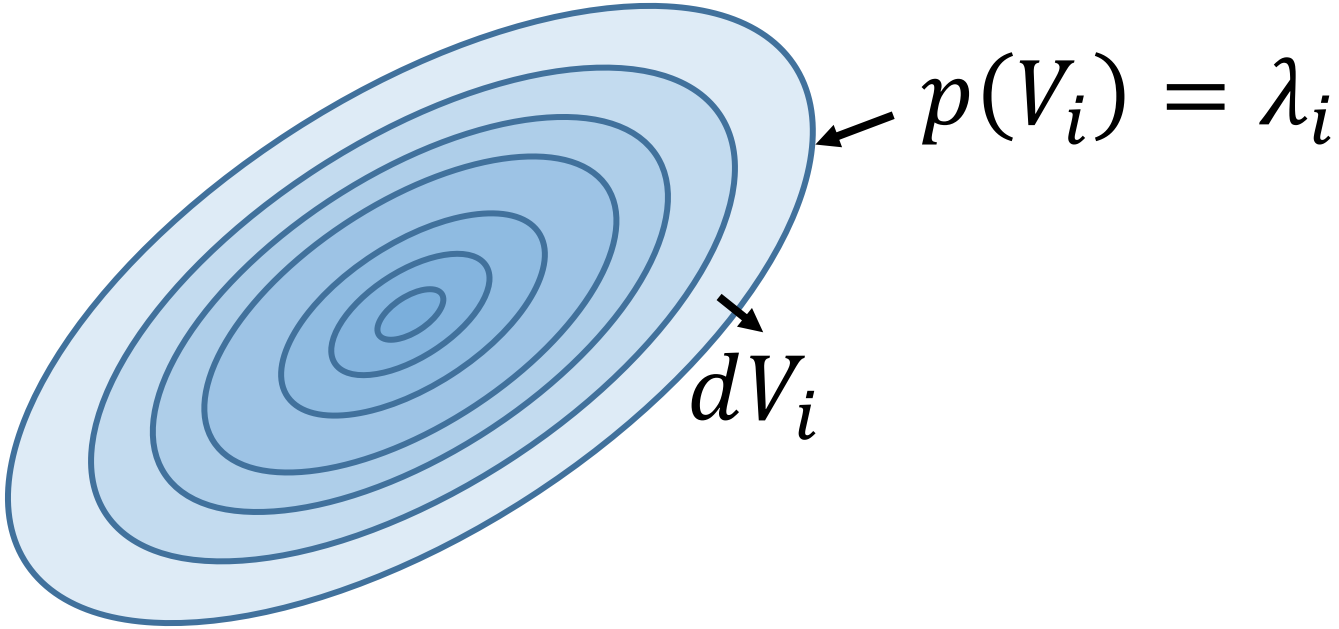
Motivation: Integrating the Posterior

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



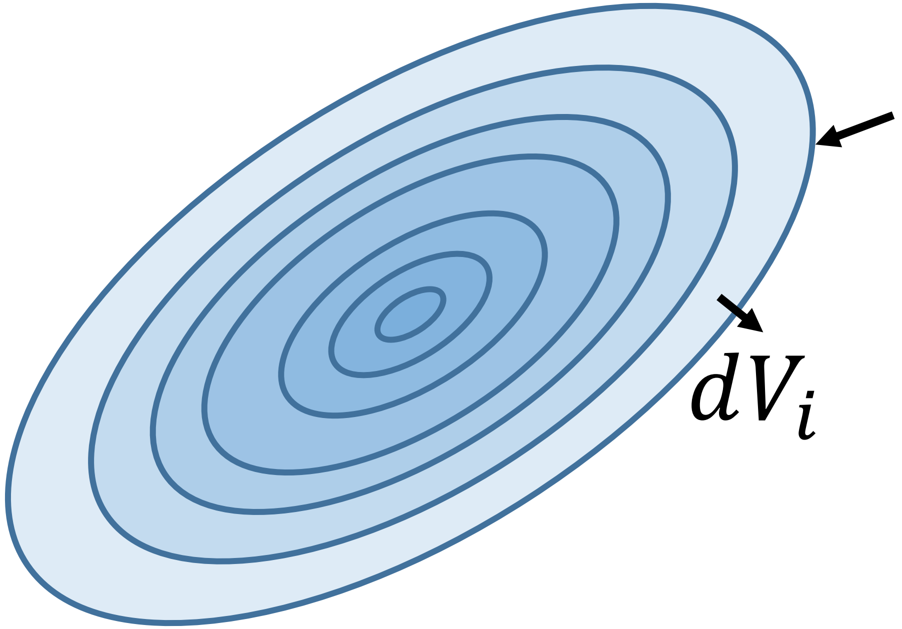
Motivation: Integrating the Posterior

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



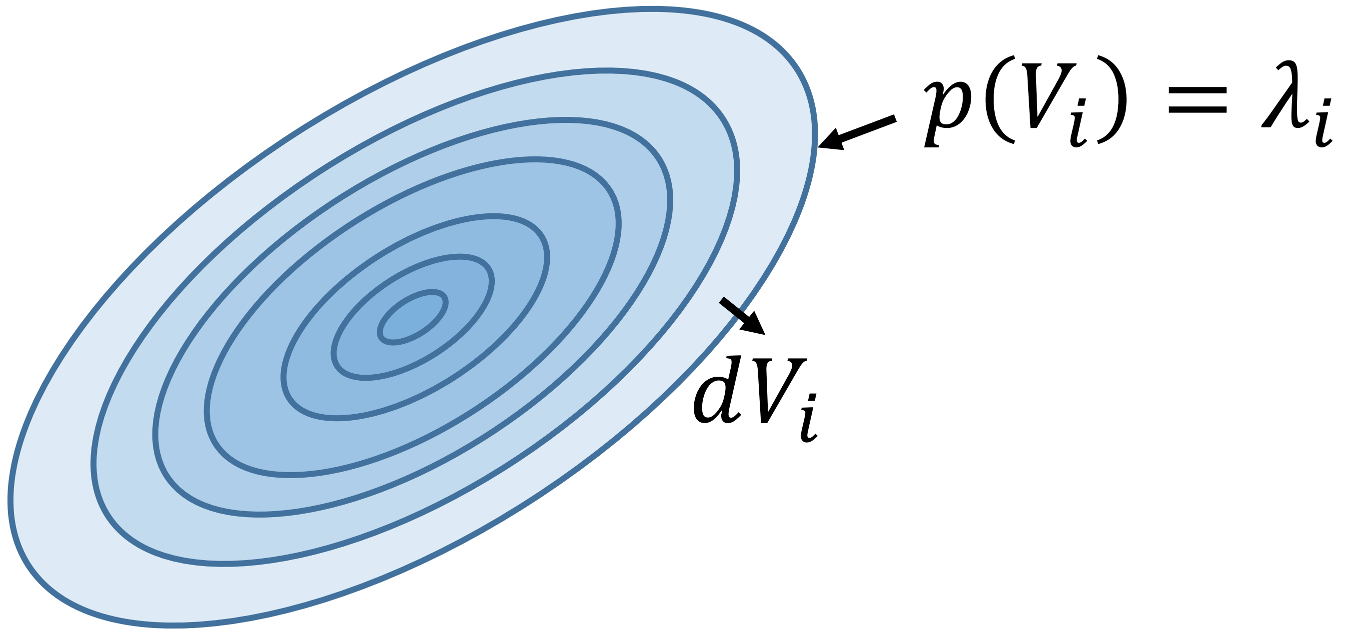
Motivation: Integrating the Posterior

$$\int_0^\infty \underbrace{p(V)}^{\text{“Amplitude”}} \underbrace{dV(\lambda)}_{\text{“Volume”}}$$



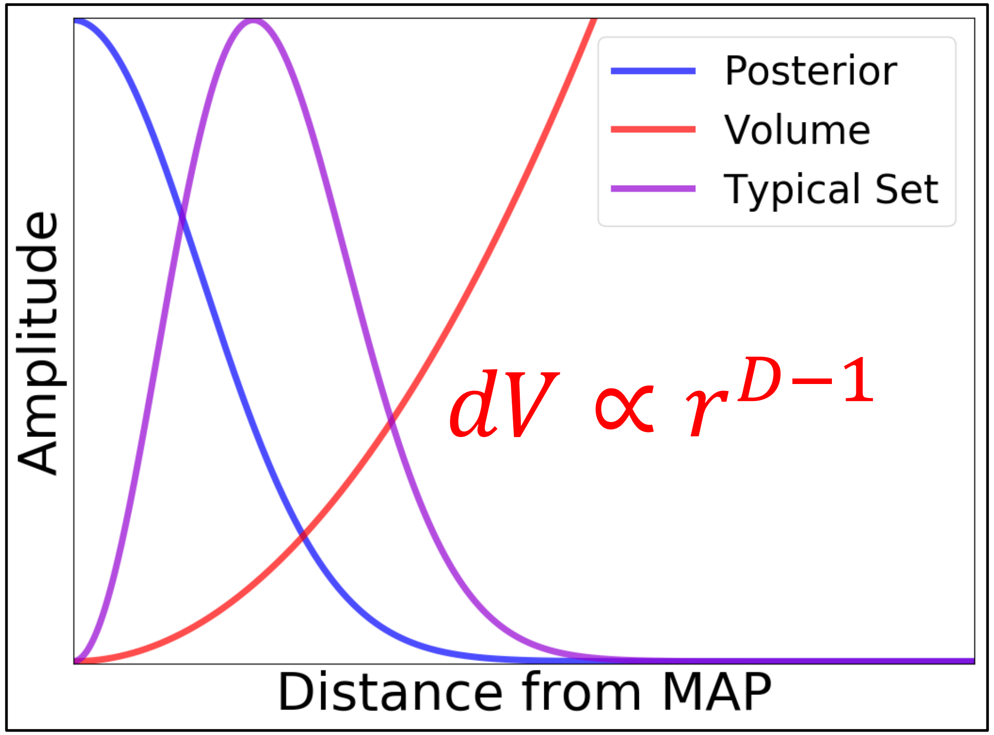
Motivation: Integrating the Posterior

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



Motivation: Integrating the Posterior

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



See also Speagle (2019)

Motivation: Integrating the Posterior

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

Motivation: Integrating the Posterior

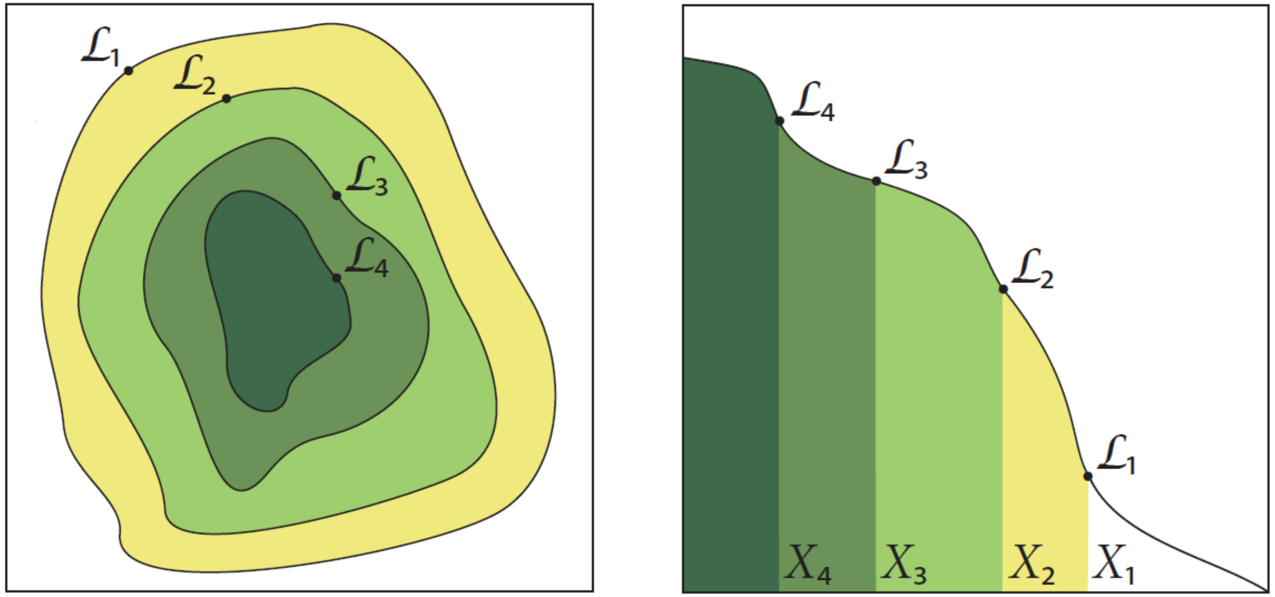
$$\mathcal{Z} \equiv \int_{\Omega_{\mathbf{\Theta}}} \mathcal{L}(\mathbf{\Theta})\pi(\mathbf{\Theta})d\mathbf{\Theta}$$

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"



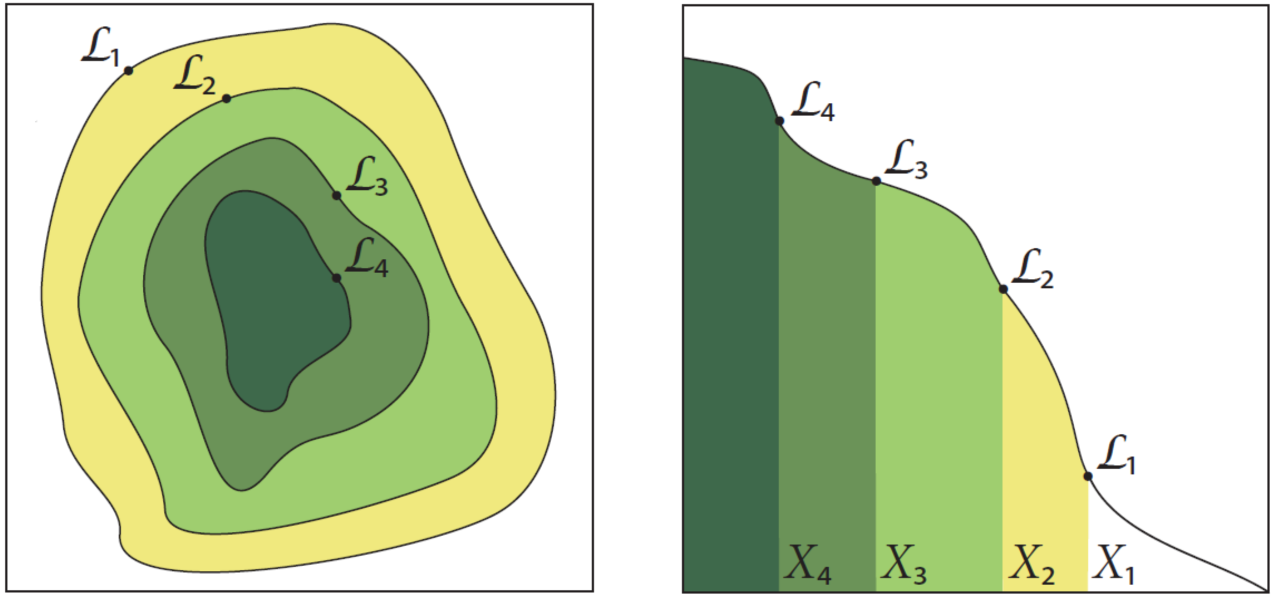
Feroz et al. (2013)

Motivation: Integrating the Posterior

$$\mathcal{Z} = \int_0^\infty X(\lambda)d\lambda$$

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

"Prior Volume"



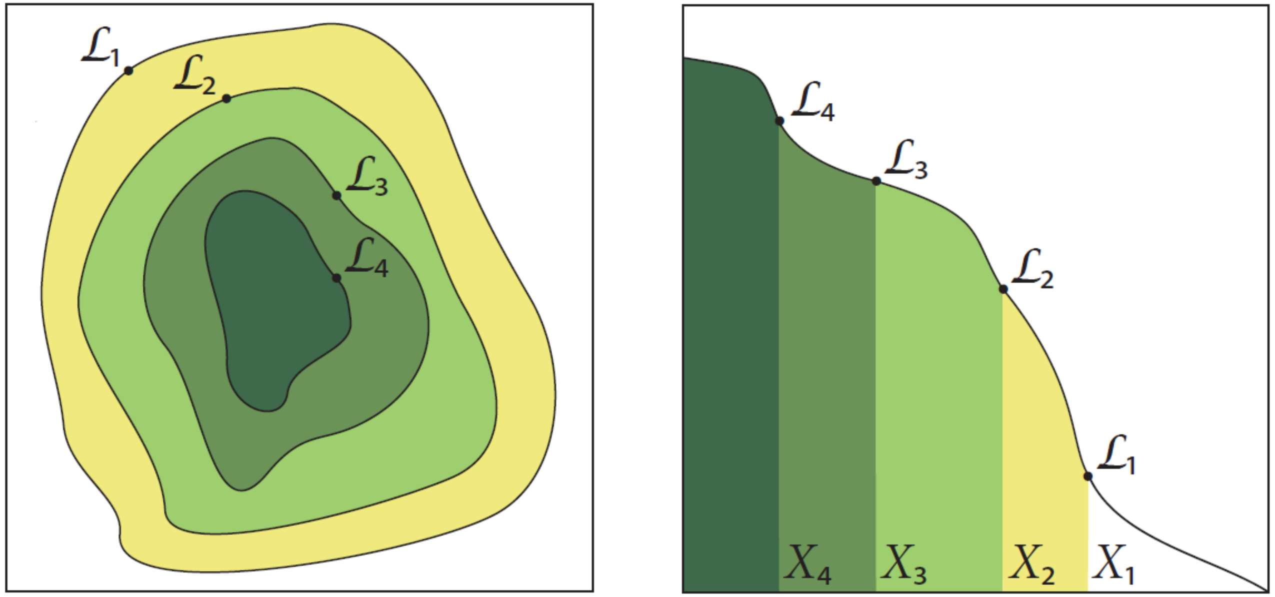
Feroz et al. (2013)

Motivation: Integrating the Posterior

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

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

"Prior Volume"



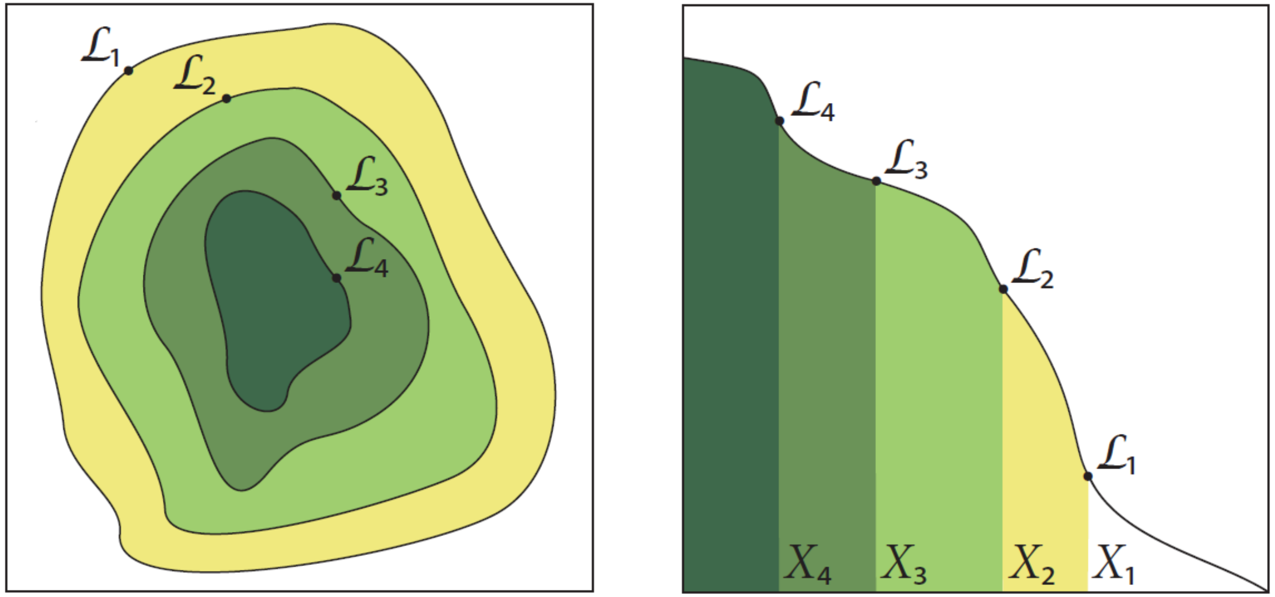
Feroz et al. (2013)

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"



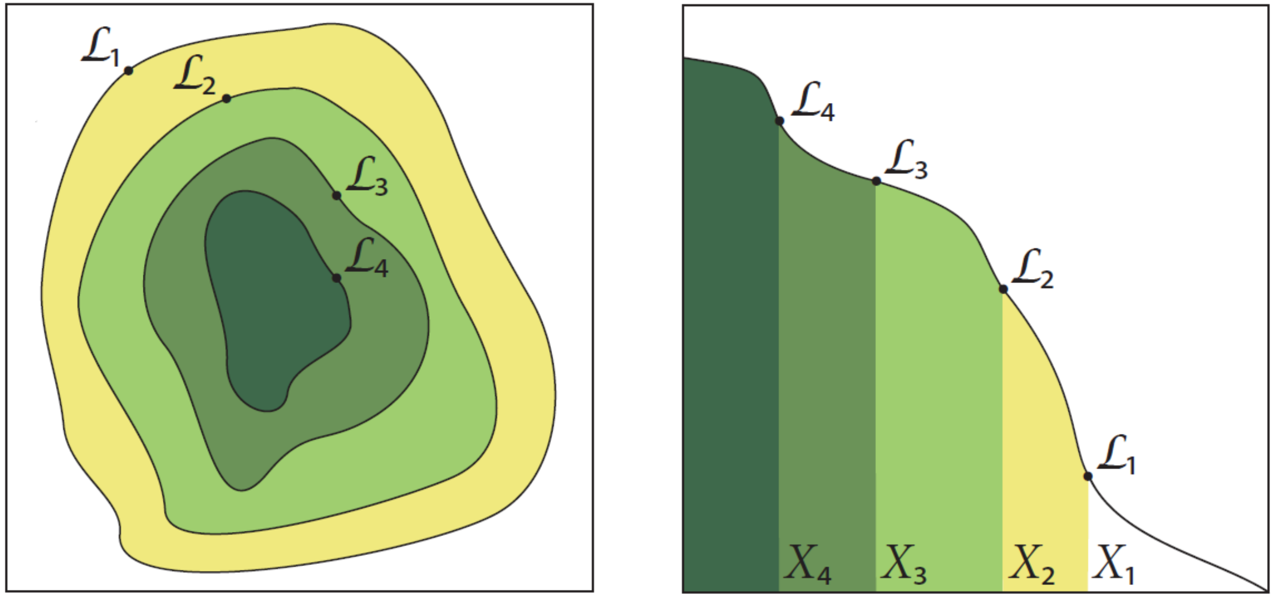
Feroz et al. (2013)

Motivation: Integrating the Posterior

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

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

"Prior Volume"



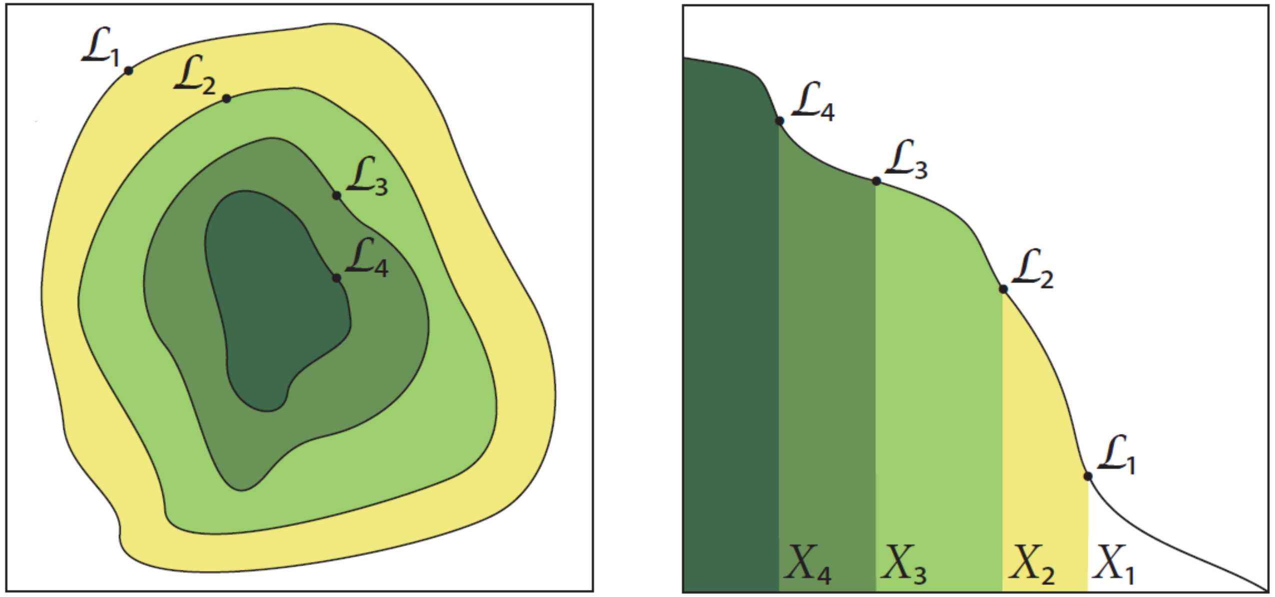
Feroz et al. (2013)

Motivation: Integrating the Posterior

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

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

"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”

$$\hat{p}_i = \frac{\widehat{w}_i}{\hat{Z}}$$

Importance Weight

Motivation: Integrating the Posterior

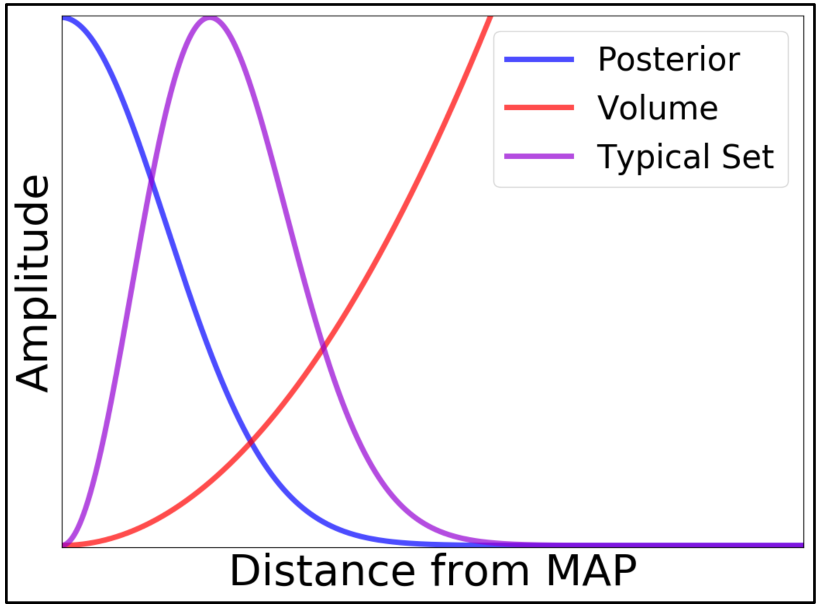
$$\hat{Z} \approx \sum_{i=1}^{n} \widehat{w}_i$$

We get posteriors “for free”

$$\hat{p}_i = \frac{\widehat{w}_i}{\hat{Z}}$$

Importance Weight

Directly proportional to the typical set!



Stopping Criteria

$$\hat{z} = \hat{z}_i + z_{\text{in}}$$

Stopping Criteria

$$\boxed{\frac{\hat{z}_i}{\hat{z}_i + z_{\text{in}}} < f}$$

$$\hat{z} = \hat{z}_i + z_{\text{in}}$$

Stopping Criteria

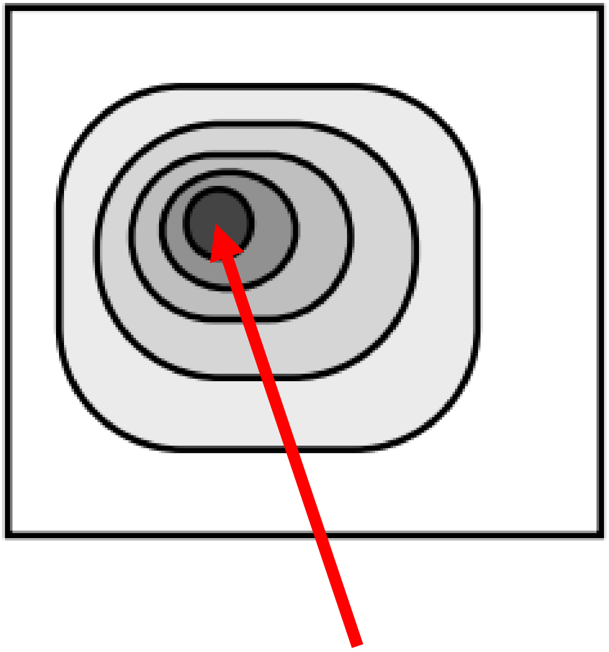
$$\boxed{\frac{\hat{z}_i}{\hat{z}_i + z_{\text{in}}} < f}$$

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

Stopping Criteria

$$\boxed{\frac{\hat{z}_i}{\hat{z}_i + z_{\text{in}}} < f}$$

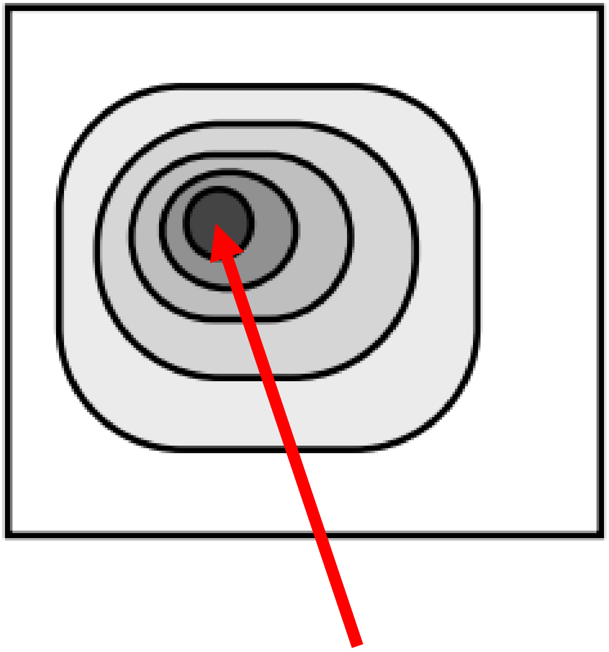
$$\hat{z} \leq \sum_{j=1}^{i} \widehat{w}_j + \mathcal{L}_{\max} X_n$$



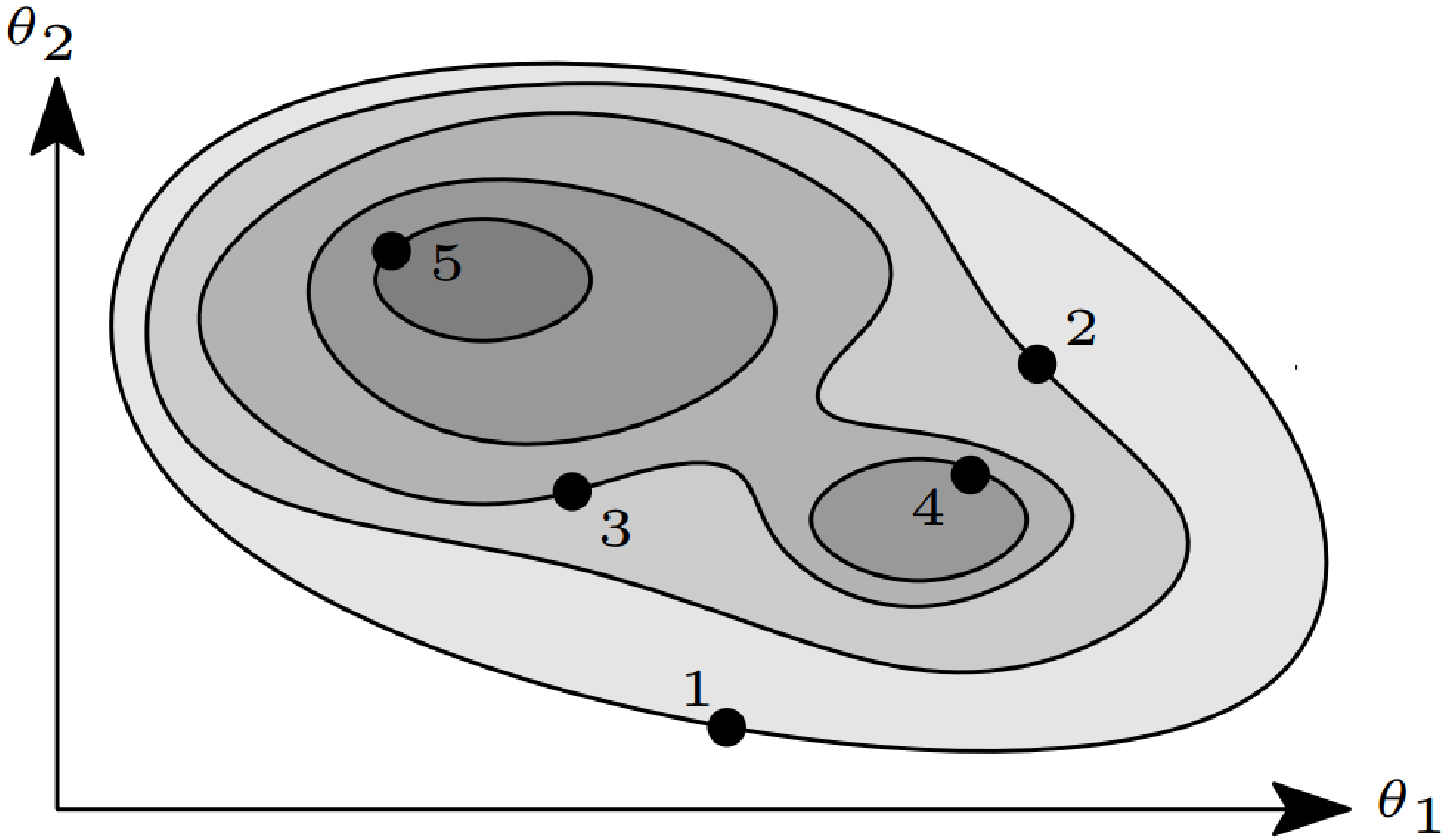
Stopping Criteria

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

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

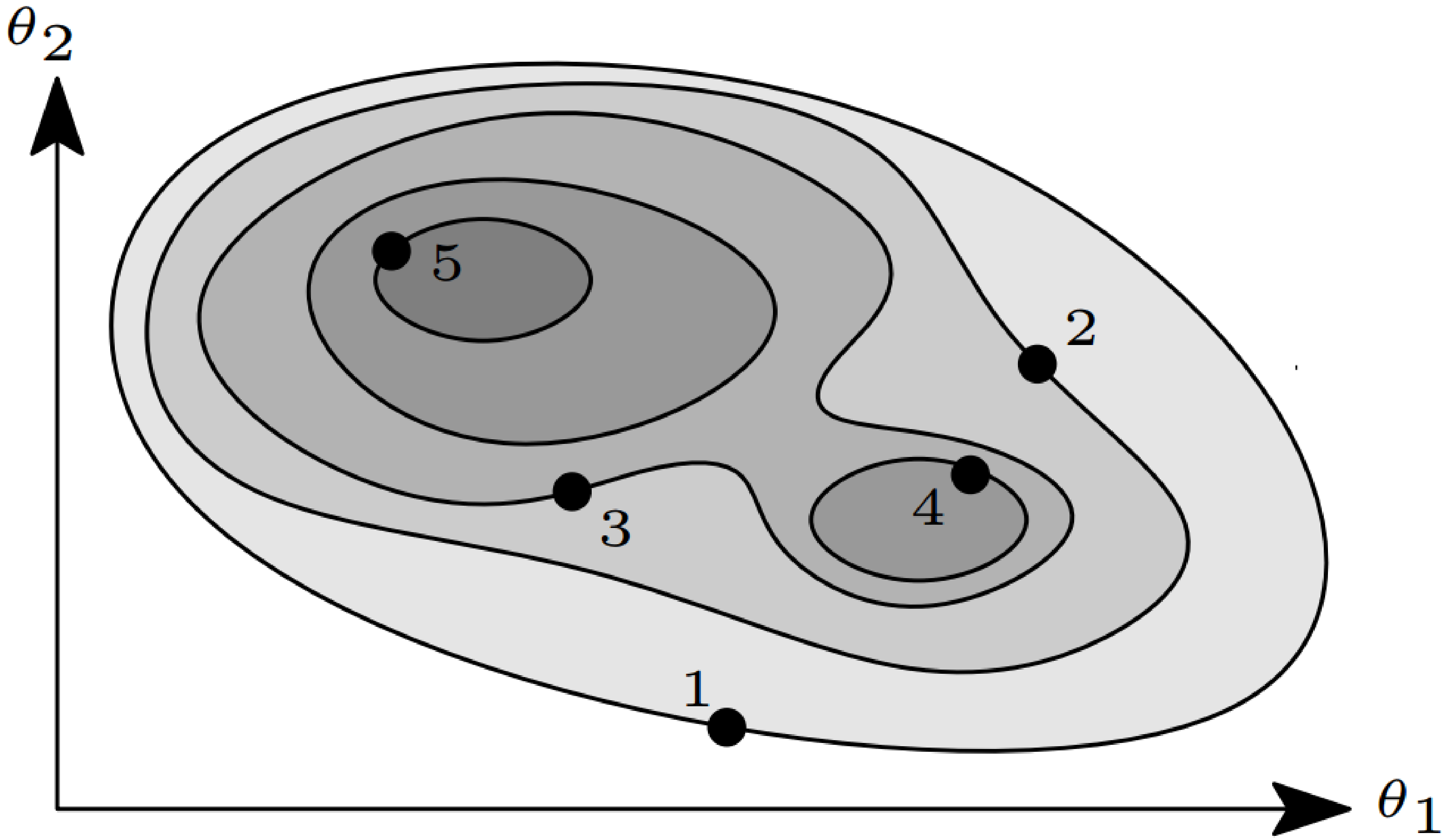


Nested Sampling In Practice



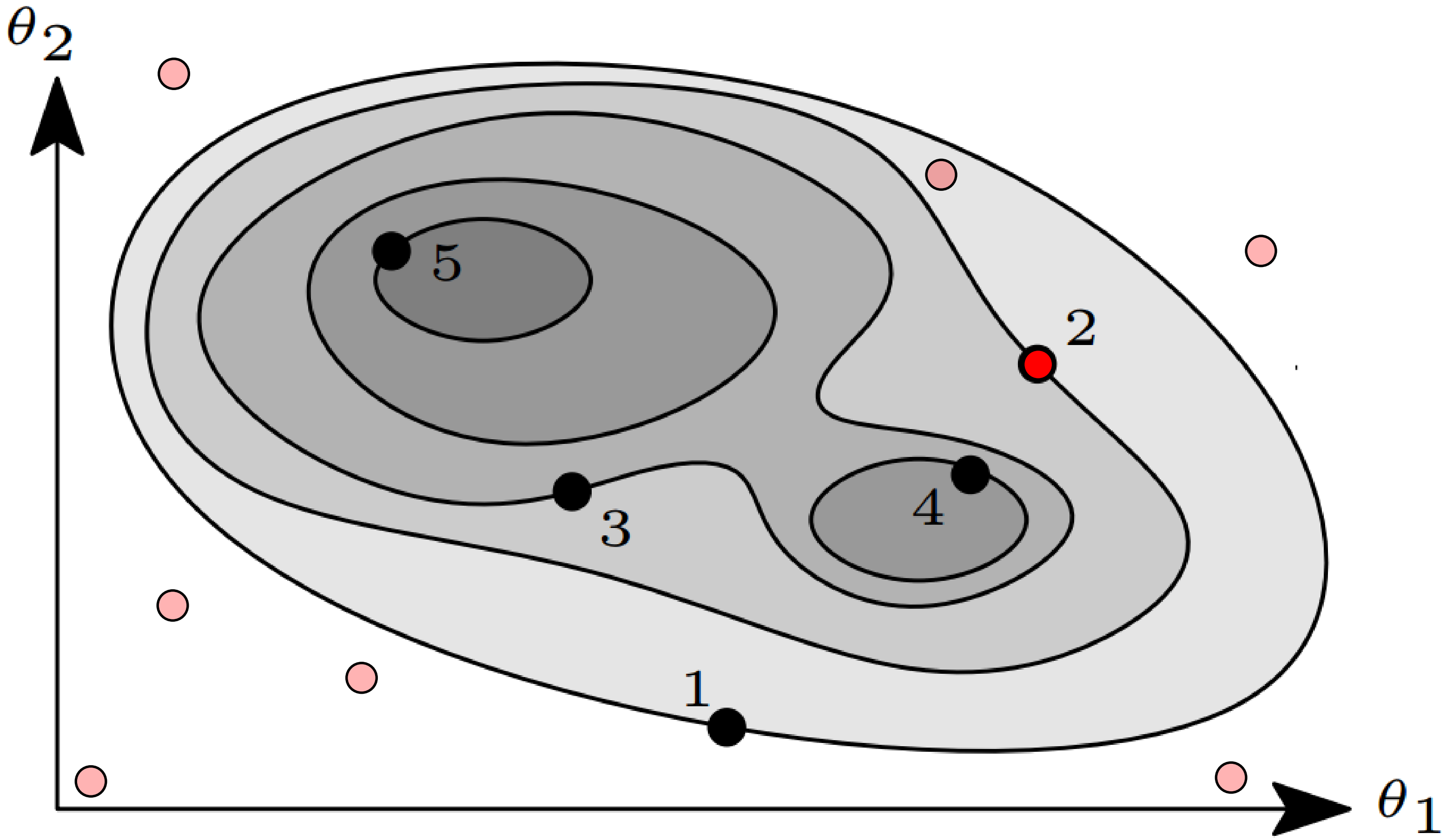
Higson et al. (2017)
arxiv:1704.03459

Naïve Approach: Sampling from the Prior



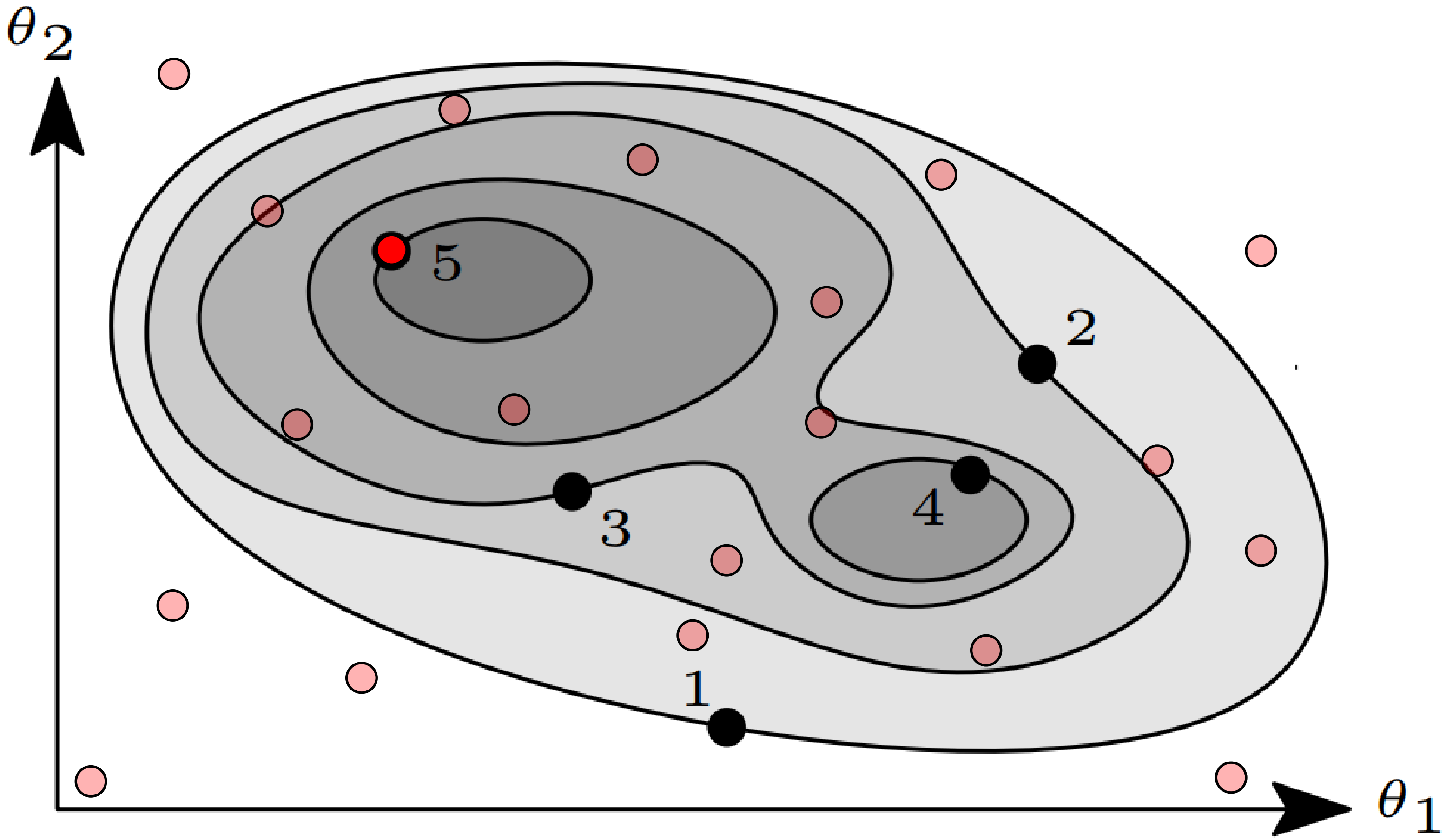
Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

Naïve Approach: Sampling from the Prior



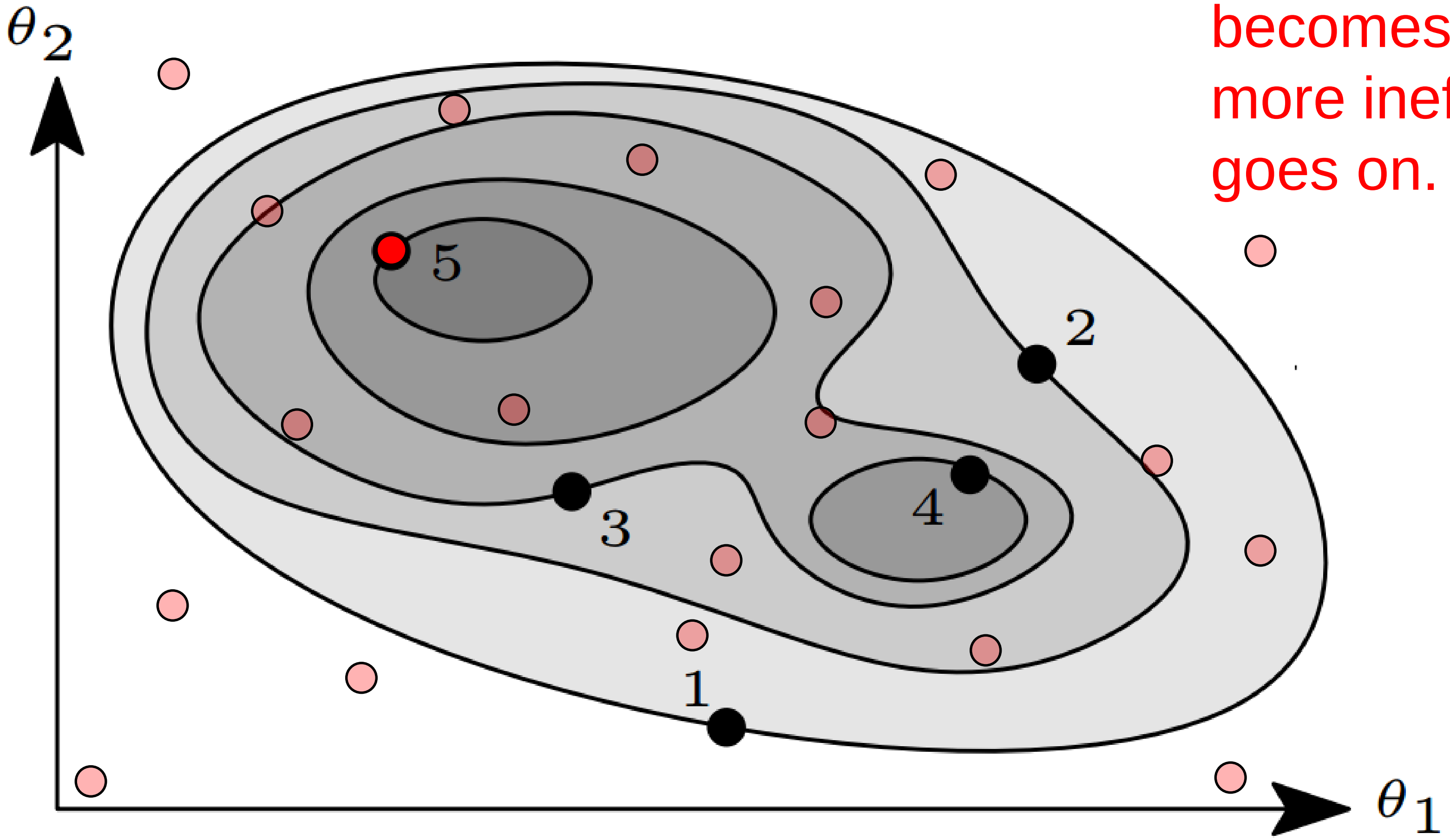
Higson et al. (2017)
arxiv:1704.03459

Naïve Approach: Sampling from the Prior



Higson et al. (2017)
arxiv:1704.03459

Naïve Approach: Sampling from the Prior

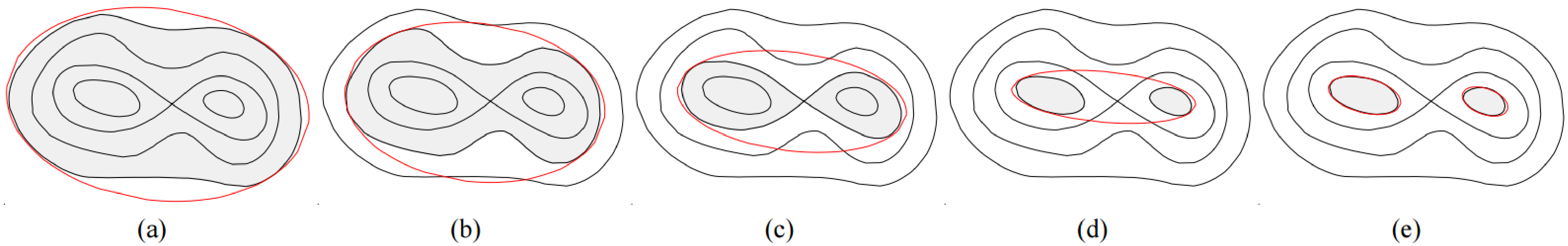


Sampling from the prior becomes **exponentially** more inefficient as time goes on.

Higson et al. (2017)
arxiv:1704.03459

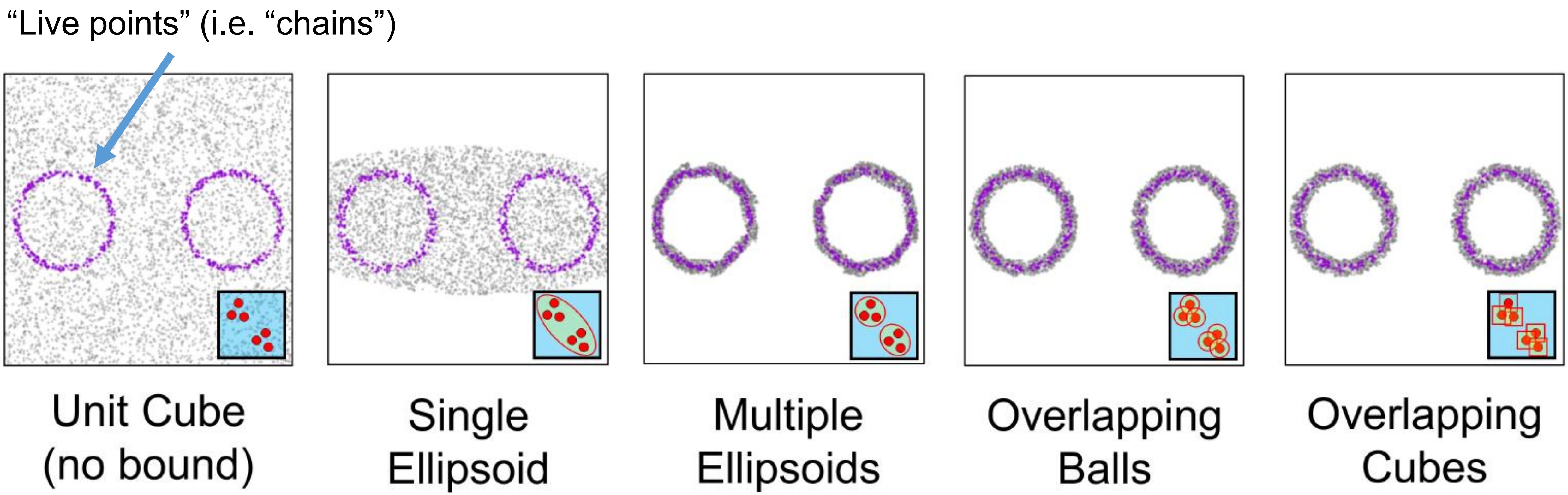
Sampling from the Constrained Prior

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



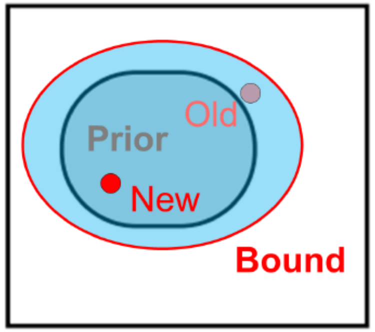
Feroz et al. (2009)

Examples of Bounding Strategies

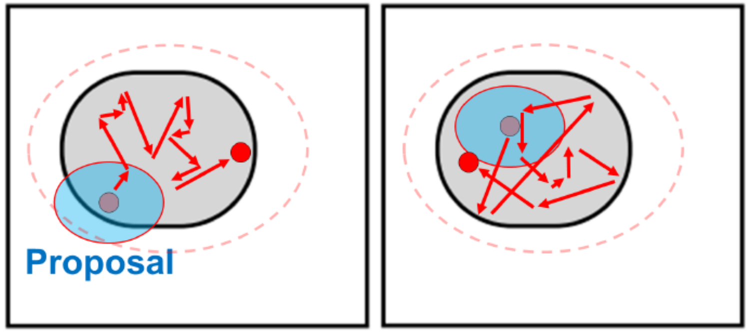


Examples of Sampling Strategies

Uniform

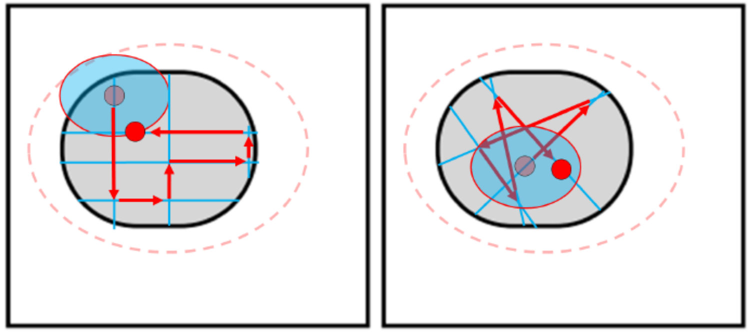


Random Walk



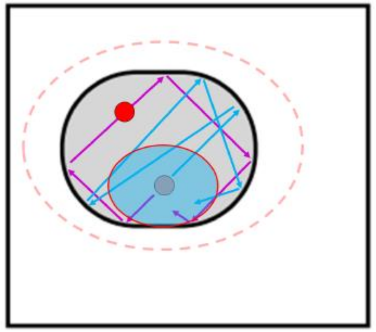
Fixed Scale **Variable Scale**

Multivariate Slice



Principal Axes **Random**

Hamiltonian Slice



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:

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 HMC.

Advantages and Disadvantages

Advantages to Nested Sampling:

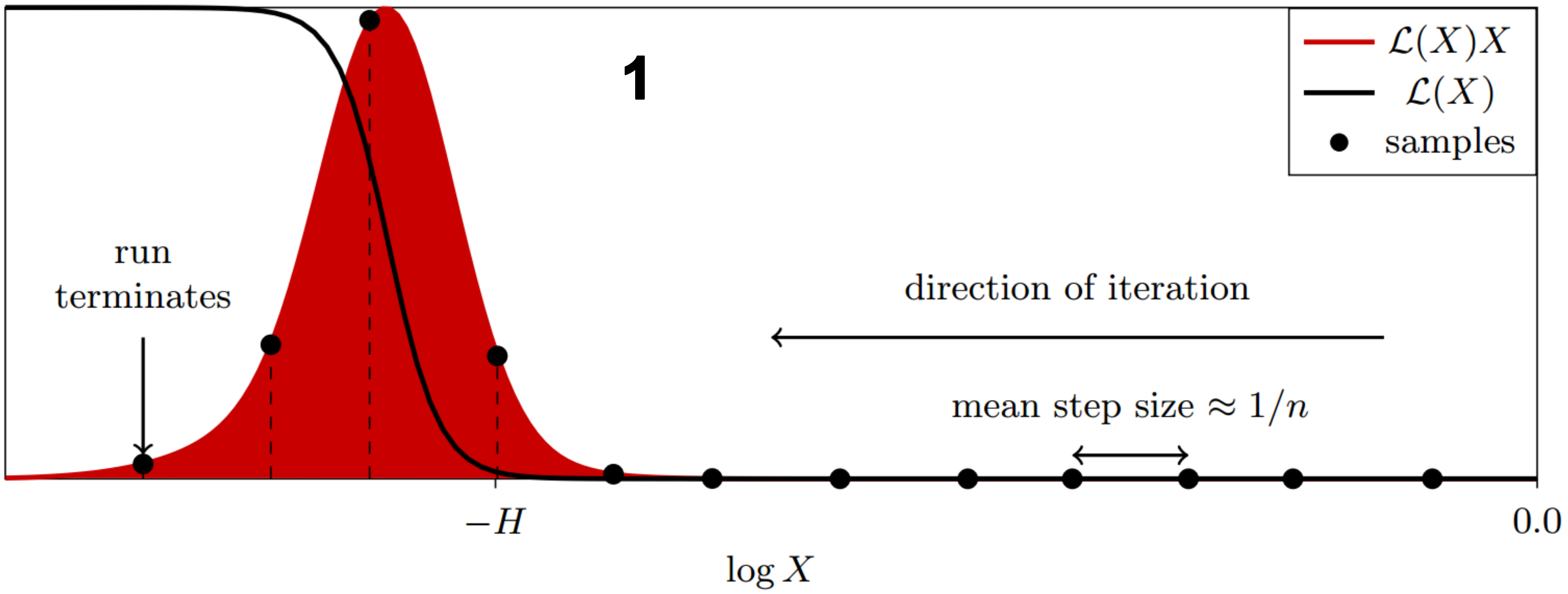
1. Can characterize complex uncertainties in real-time.
2. Can allocate samples much more efficiently in some cases.
3. Possesses well-motivated stopping criteria (Skilling 2006; Speagle 2020).
4. Can help perform model selection.

Disadvantages to Nested Sampling:

1. Implementations require a prior transform.
2. Runtime
3. Overall a … utions.
4. Sampling
5. Can't use gradients as "naturally" as HMC.

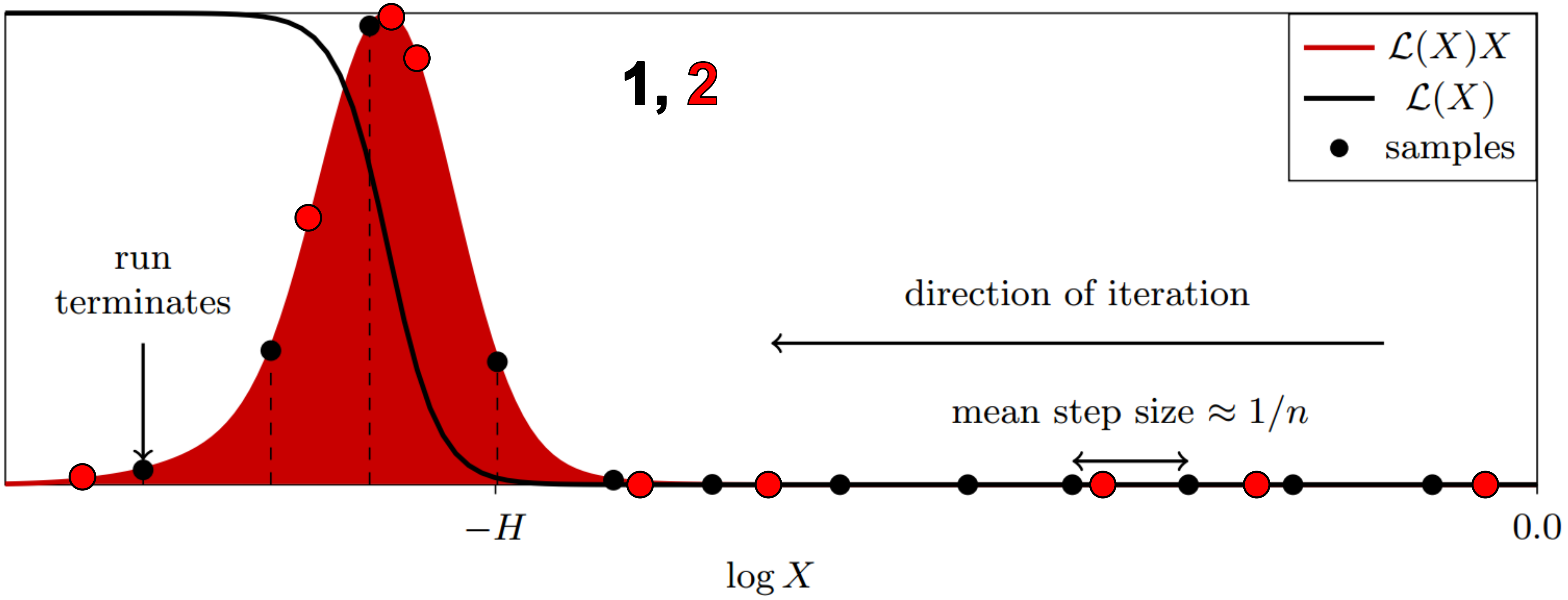
Fraction of "wasted" samples doesn't adapt to the shape of the posterior over time

Dynamic Nested Sampling



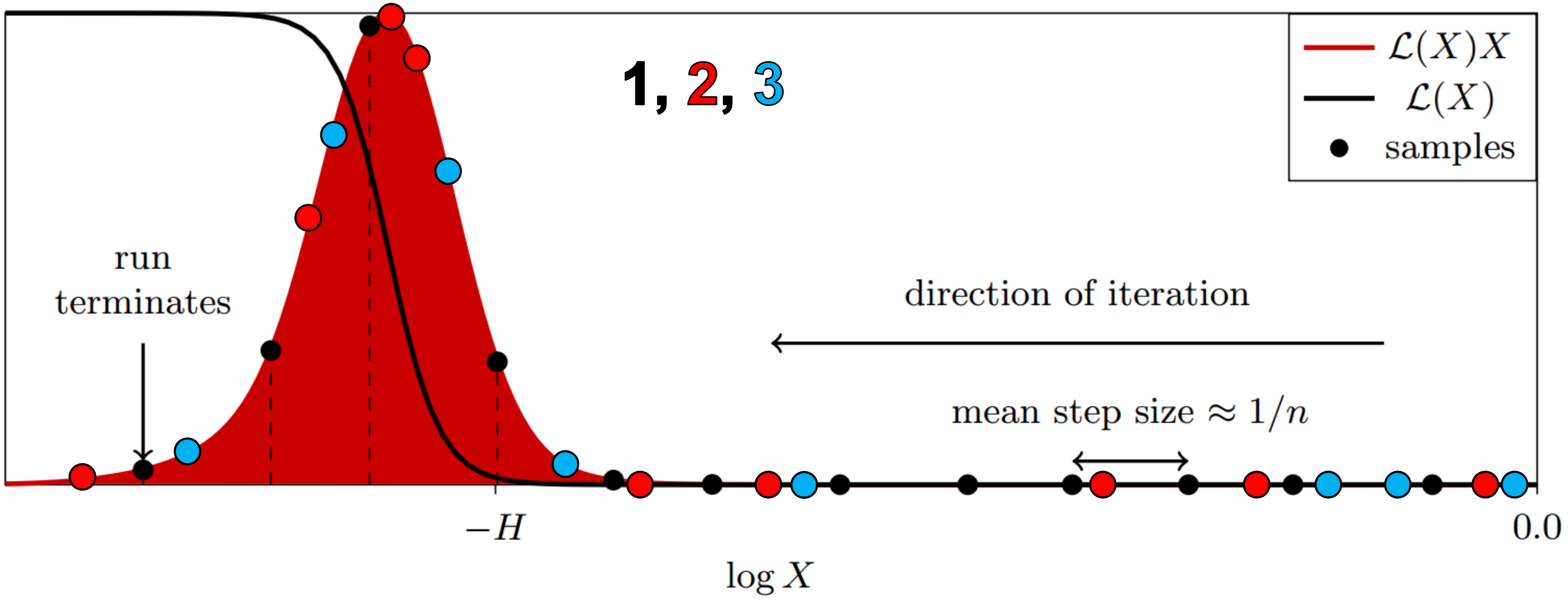
Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

Dynamic Nested Sampling



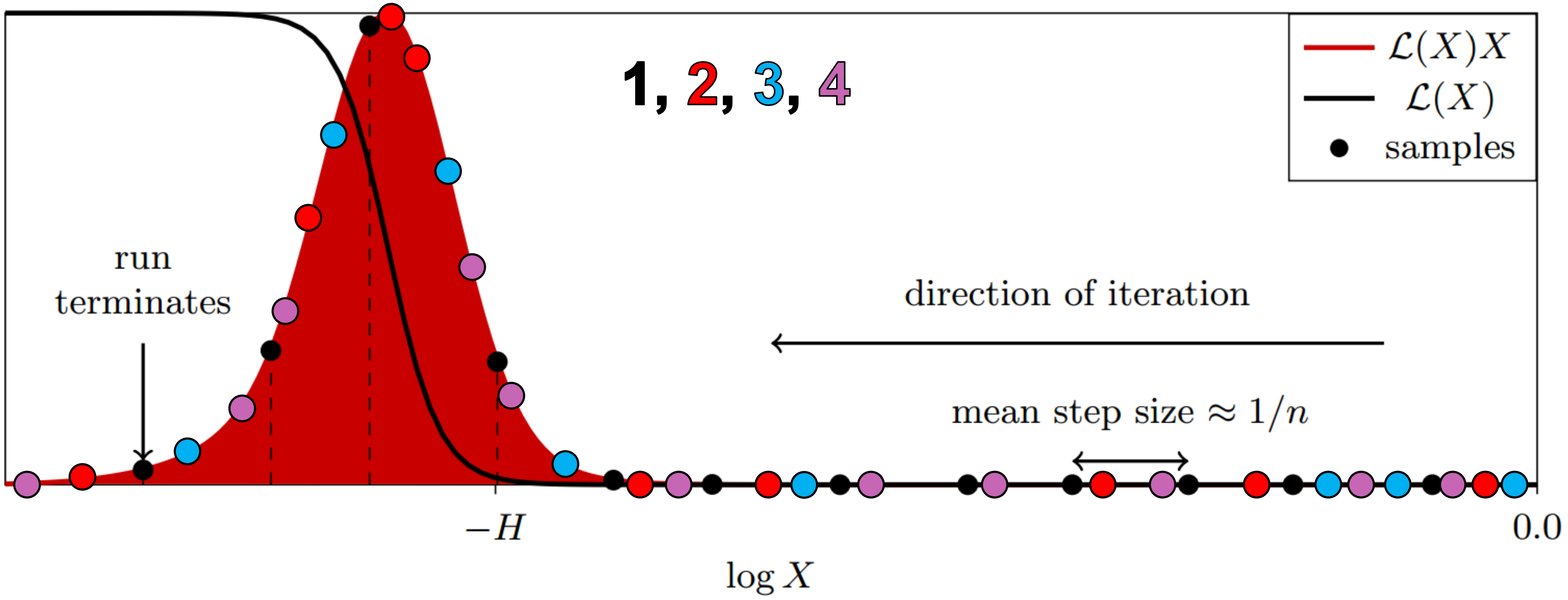
Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

Dynamic Nested Sampling



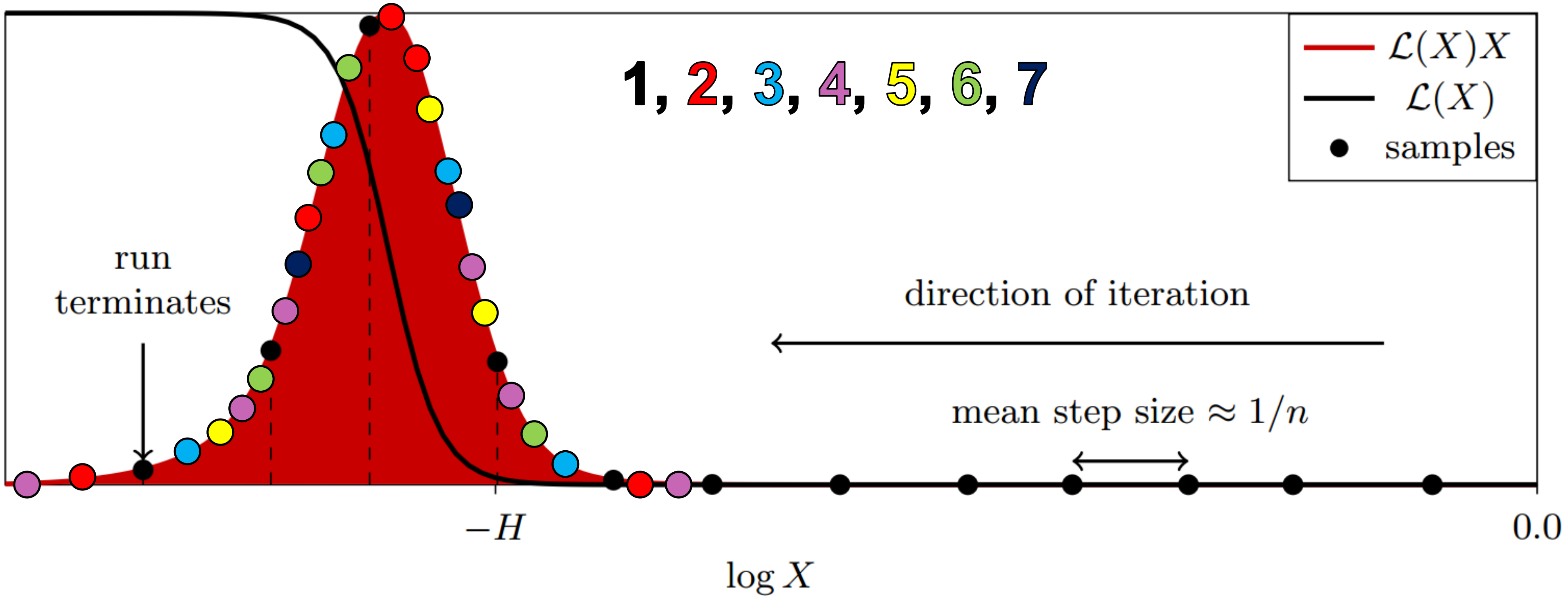
Higson et al. (2017)
arxiv:1704.03459

Dynamic Nested Sampling



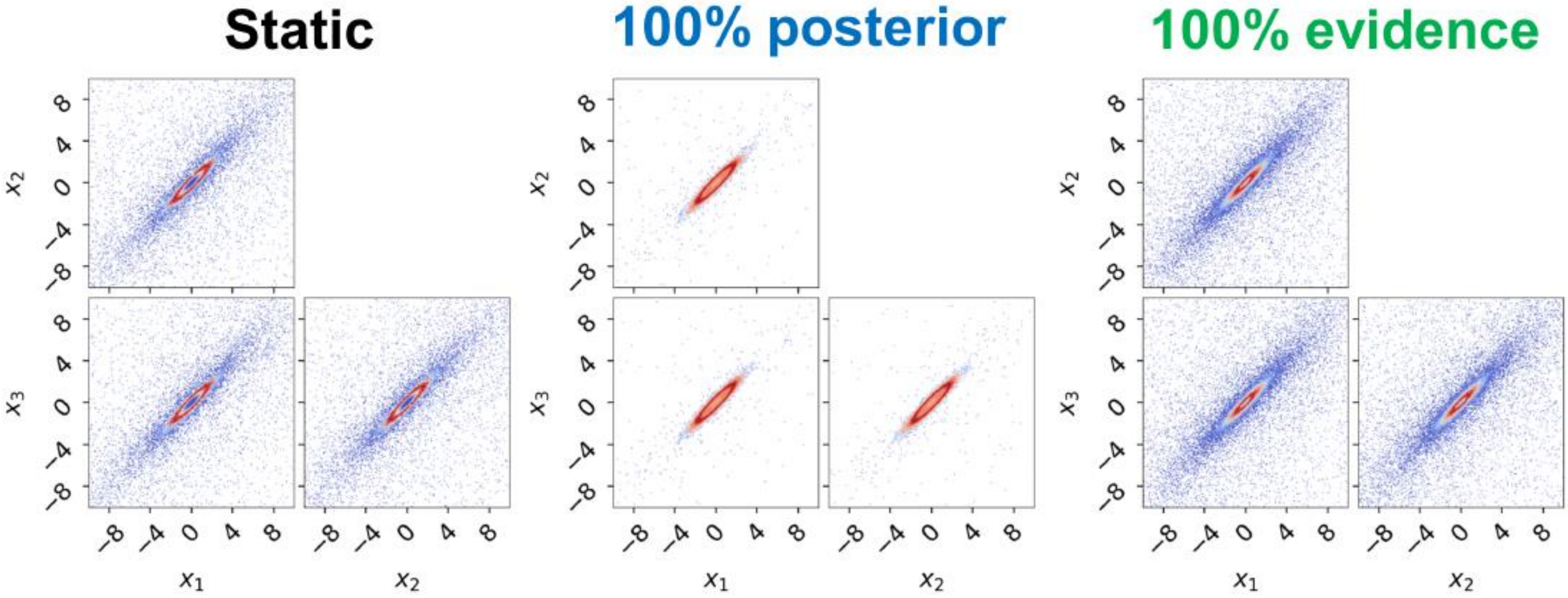
Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

Dynamic Nested Sampling



Higson et al. (2017)
arxiv:1704.03459

Illustrative Example



Speagle (2020)

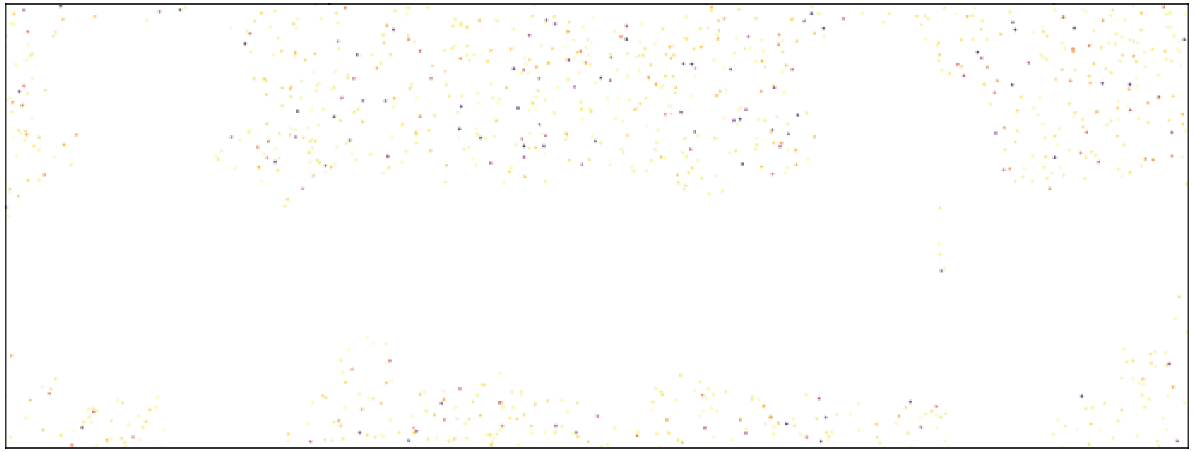
dynesty

Inspired by *emcee*! (Foreman-Mackey et al. 2013)

- **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.



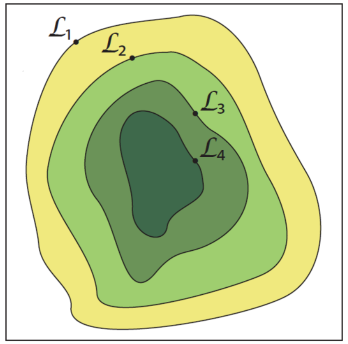
Speagle (2020)

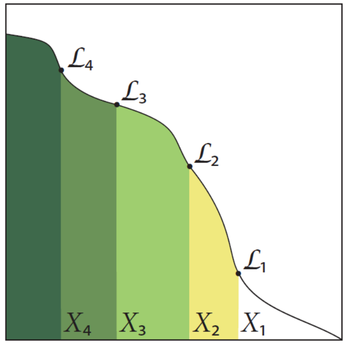
https://dynesty.readthedocs.io

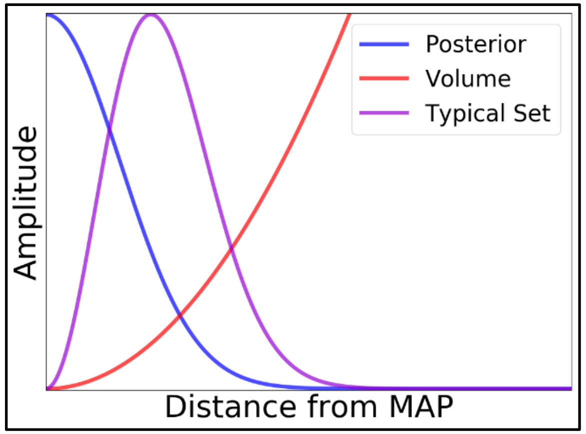
Summary

Summary

Motivation and Concepts

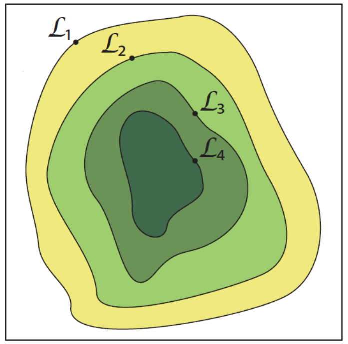


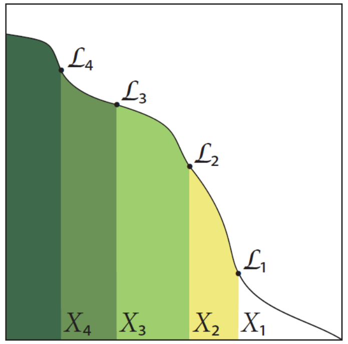


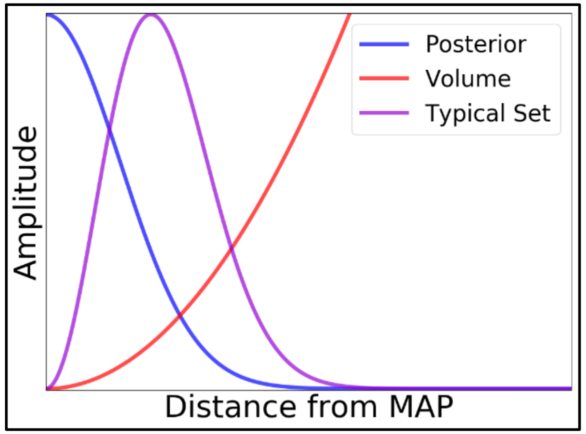


Summary

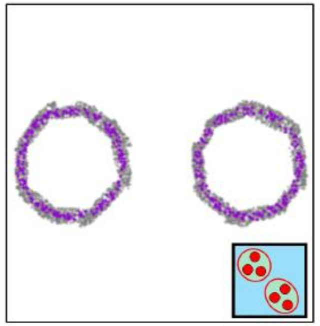
Motivation and Concepts



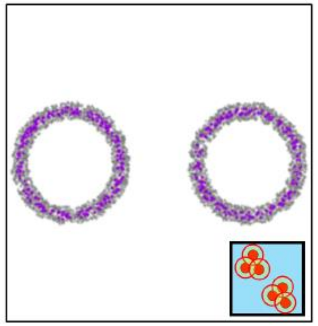




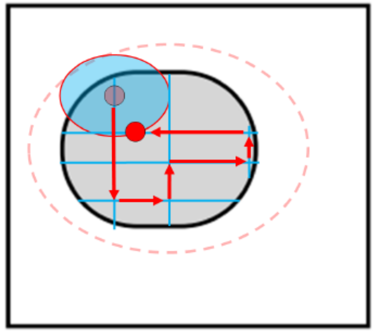
Practical Issues



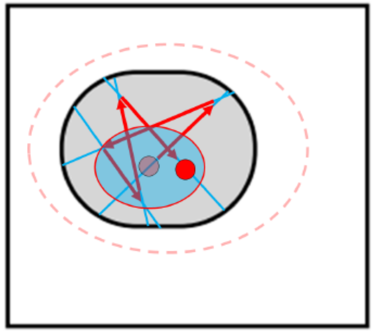
Multiple Ellipsoids



Overlapping Balls



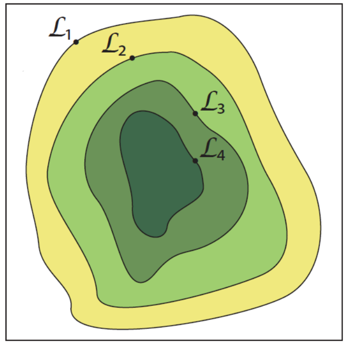
Principal Axes

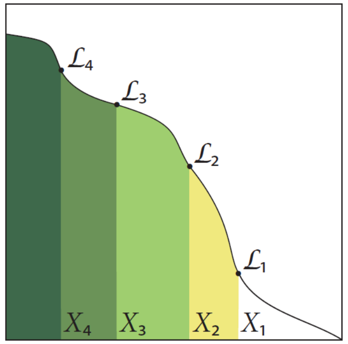


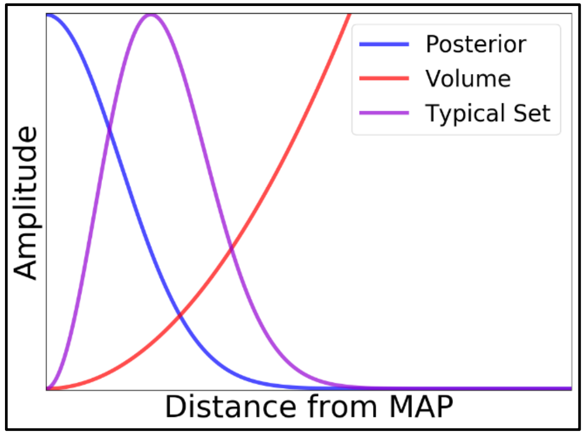
Random

Summary

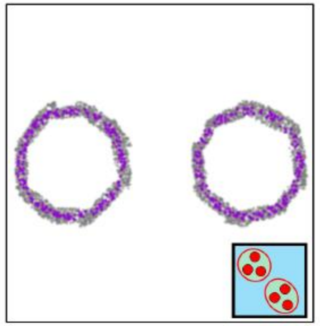
Motivation and Concepts



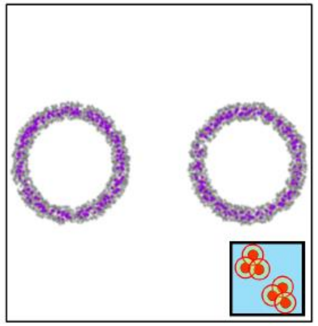




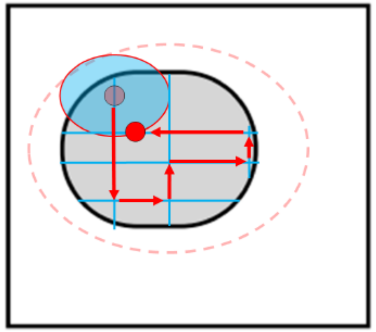
Practical Issues



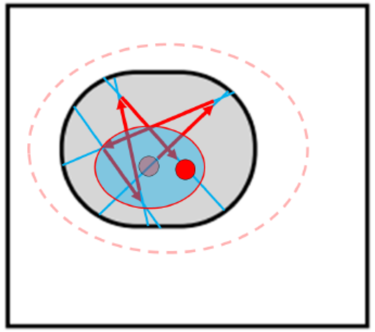
Multiple Ellipsoids



Overlapping Balls

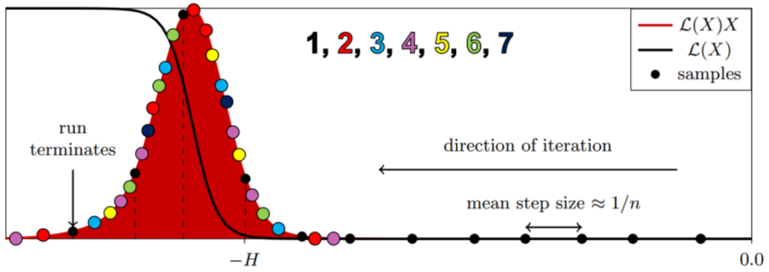


Principal Axes



Random

Modifications and Applications



Static **100% posterior**