An Introduction to (Dynamic) Nested Sampling

Josh Speagle

jspeagle@cfa.harvard.edu

Introduction

Background

$$\Pr(\boldsymbol{\Theta}|\mathbf{D},\mathrm{M}) = \frac{\Pr(\mathbf{D}|\boldsymbol{\Theta},\mathrm{M})\Pr(\boldsymbol{\Theta}|\mathrm{M})}{\Pr(\mathbf{D}|\mathrm{M})}$$

Bayes' Theorem

Background

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

Bayes' Theorem

Background

Posterior

Likelihood Prior

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

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

Bayes' Theorem

Background

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

Posterior Estimation via Sampling

Samples $\boldsymbol{\Theta}_N, \boldsymbol{\Theta}_{N-1}, \ldots, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_1 \in \Omega_{\boldsymbol{\Theta}}$

Posterior Estimation via Sampling

Samples $\mathbf{\Theta}_N, \mathbf{\Theta}_{N-1}, \ldots, \mathbf{\Theta}_2, \mathbf{\Theta}_1 \in \Omega_{\mathbf{\Theta}}$

Weights $\hat{p}_N, \hat{p}_{N-1}, \ldots, \hat{p}_2, \hat{p}_1 \Rightarrow \hat{p}(\mathbf{\Theta})$

$$= \sum_{i=1}^{N} \hat{p}_i \delta_D(\mathbf{\Theta} - \mathbf{\Theta}_{\mathrm{i}})$$

Posterior Estimation via Sampling

Samples $\boldsymbol{\Theta}_N, \boldsymbol{\Theta}_{N-1}, \dots, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_1 \in \Omega_{\boldsymbol{\Theta}}$

↓

Weights $\hat{p}_N, \hat{p}_{N-1}, \dots, \hat{p}_2, \hat{p}_1 \Rightarrow \hat{p}(\boldsymbol{\Theta})$

MCMC 1, 1, ... , 1, 1

$$= \sum_{i=1}^{N} \hat{p}_i \delta_D(\boldsymbol{\Theta} - \boldsymbol{\Theta}_\mathrm{i})$$

Posterior Estimation via Sampling

Samples $\boldsymbol{\Theta}_N, \boldsymbol{\Theta}_{N-1}, \ldots, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_1 \in \Omega_{\boldsymbol{\Theta}}$

Weights $\hat{p}_N, \hat{p}_{N-1}, \ldots, \hat{p}_2, \hat{p}_1 \Rightarrow \hat{p}(\boldsymbol{\Theta})$

MCMC $1, \quad 1, \quad \ldots, \quad 1, \quad 1$

Importance Sampling $\frac{p_N}{q_N}, \frac{p_{N-1}}{q_{N-1}}, \ldots, \frac{p_2}{q_2}, \frac{p_1}{q_1}$

$$= \sum_{i=1}^{N} \hat{p}_i \delta_D(\boldsymbol{\Theta} - \boldsymbol{\Theta}_\mathrm{i})$$

Posterior Estimation via Sampling

Samples $\mathbf{\Theta}_N, \mathbf{\Theta}_{N-1}, \dots, \mathbf{\Theta}_2, \mathbf{\Theta}_1 \in \Omega_{\mathbf{\Theta}}$

Weights $\hat{p}_N, \hat{p}_{N-1}, \dots, \hat{p}_2, \hat{p}_1 \Rightarrow \hat{p}(\mathbf{\Theta})$

$$= \sum_{i=1}^{N} \hat{p}_i \delta_D(\mathbf{\Theta} - \mathbf{\Theta}_{\mathrm{i}})$$

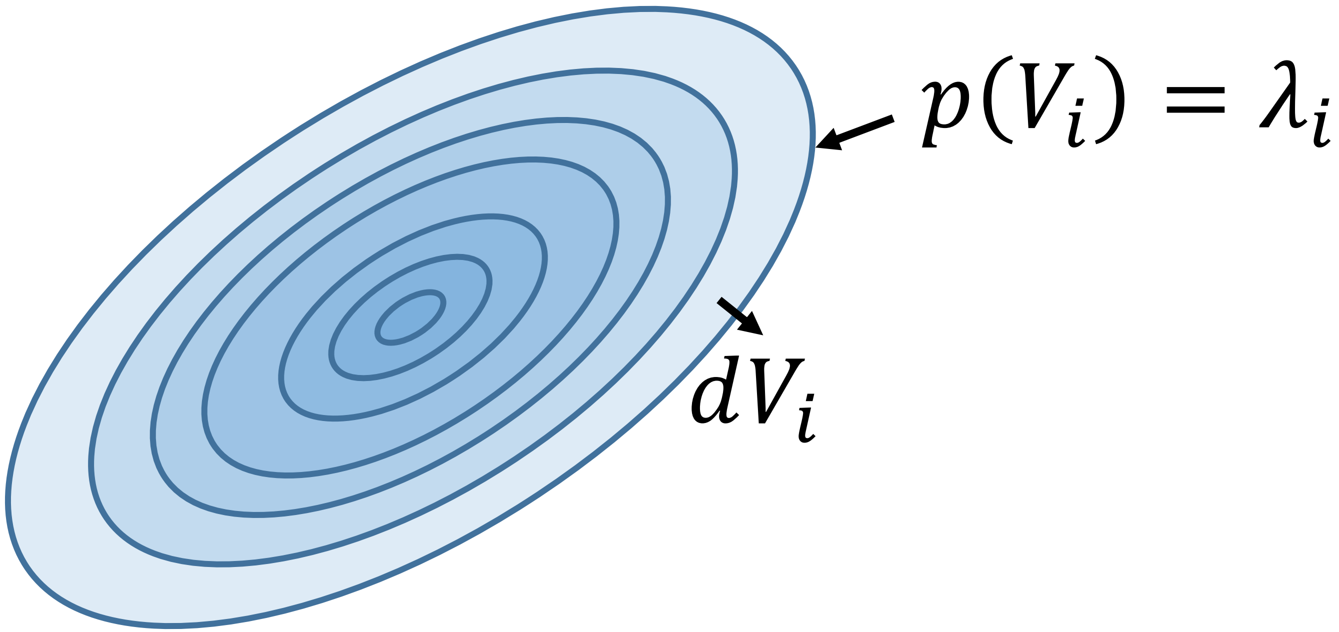
Nested Sampling?

Motivation: Integrating the Posterior

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

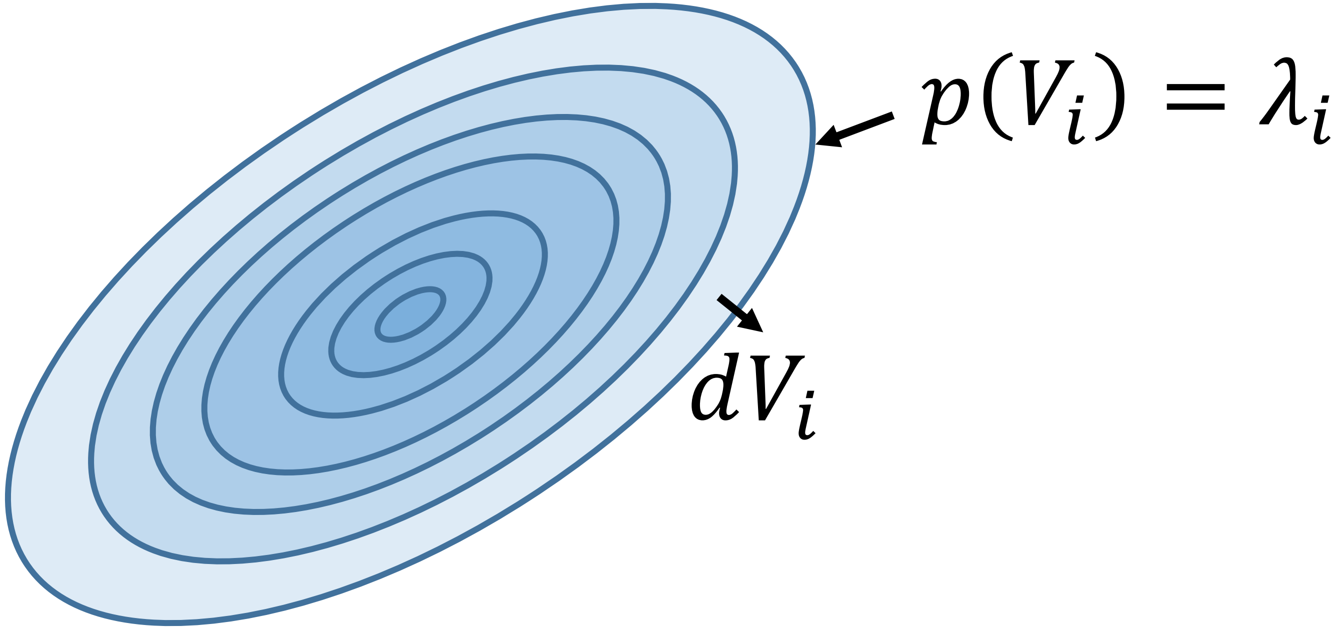
Motivation: Integrating the Posterior

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



Motivation: Integrating the Posterior

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

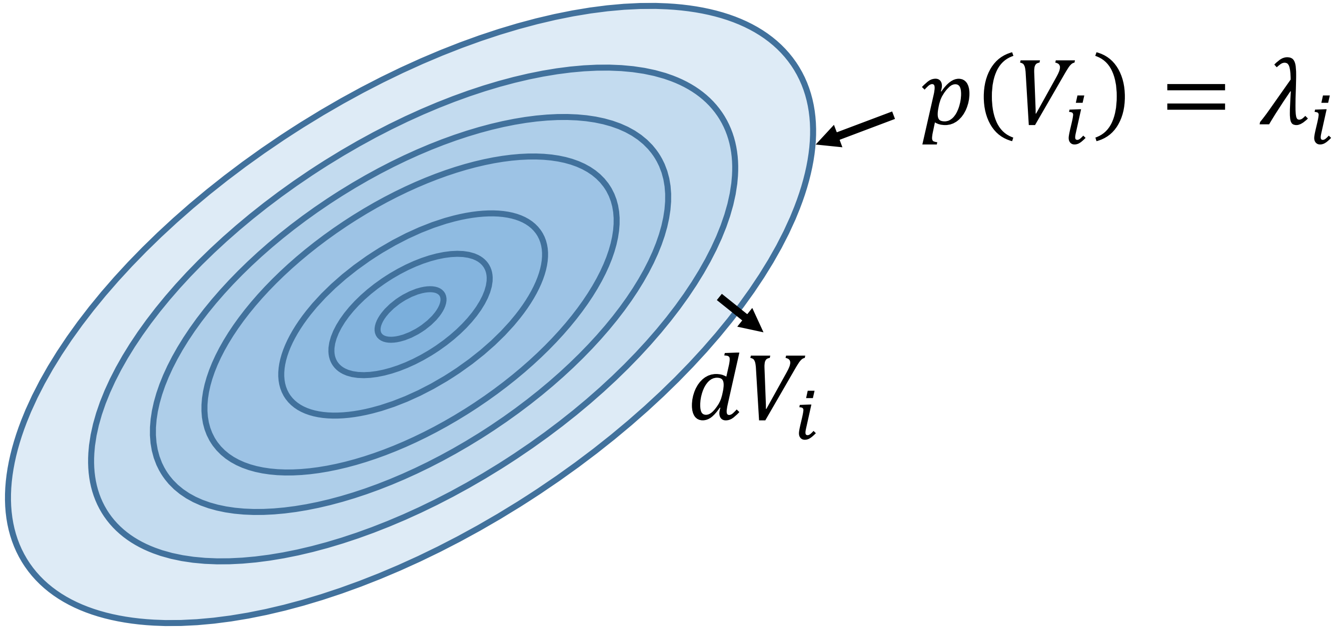


Motivation: Integrating the Posterior

"Amplitude"

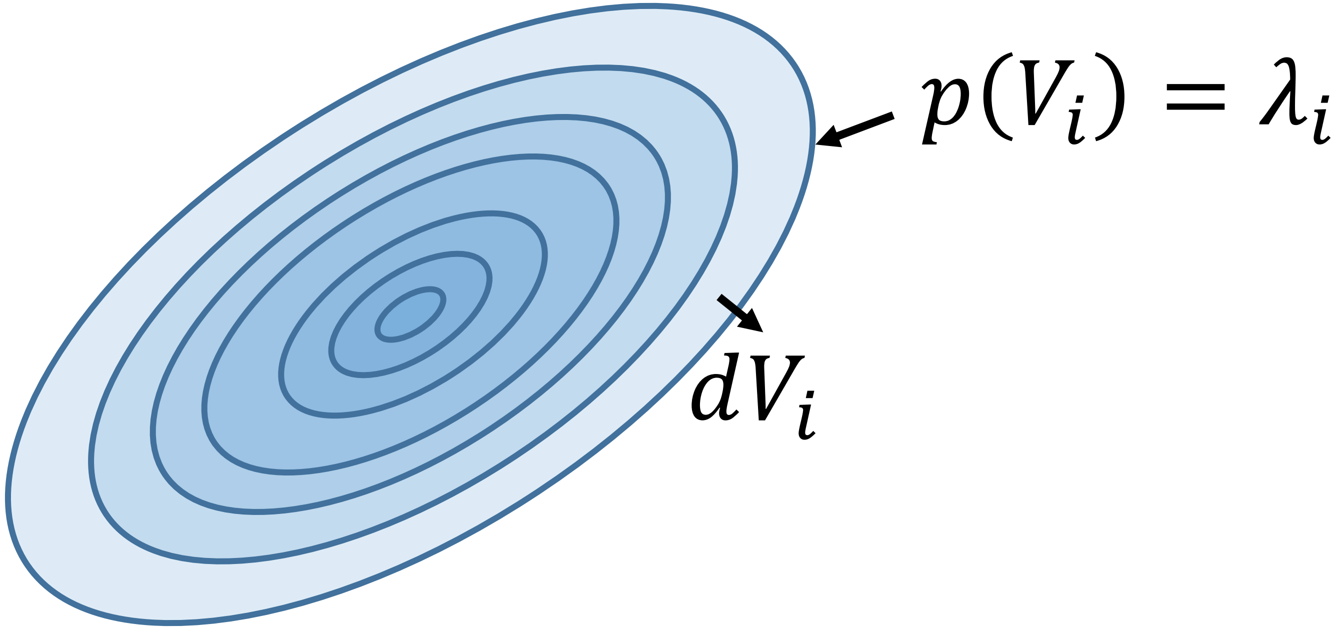
$$\int_0^\infty p(V)\, dV(\lambda)$$

"Volume"



Motivation: Integrating the Posterior

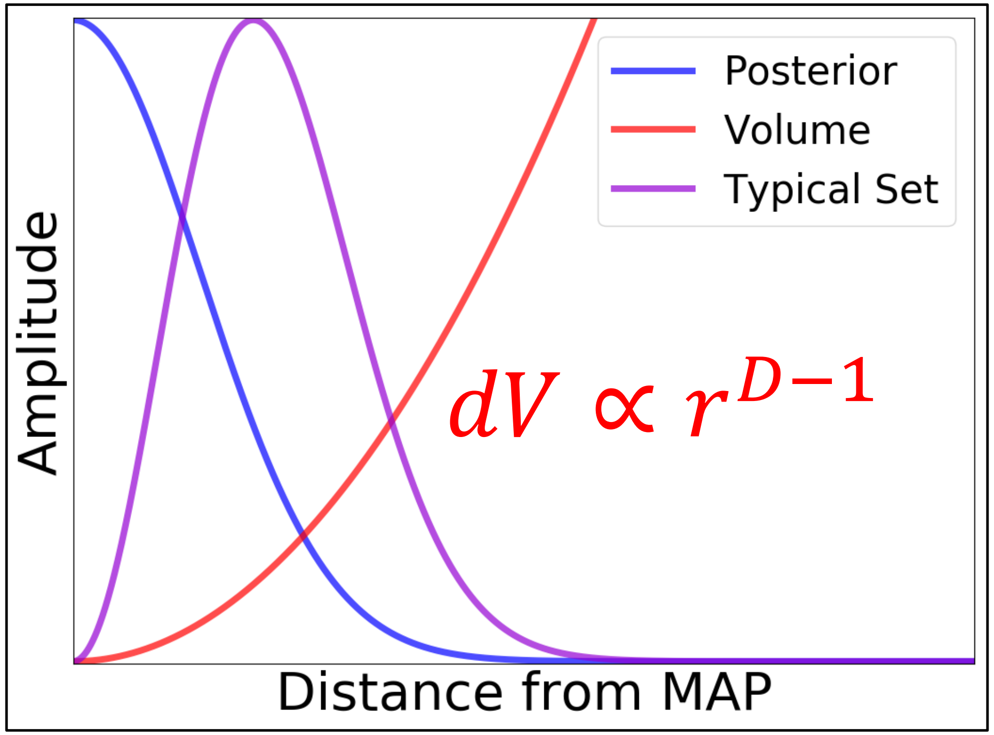
$$\int_0^\infty \overbrace{p(V)\,dV(\lambda)}^{\text{“Typical Set”}}$$



Motivation: Integrating the Posterior

“Typical Set”

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



Motivation: Integrating the Posterior

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

Motivation: Integrating the Posterior

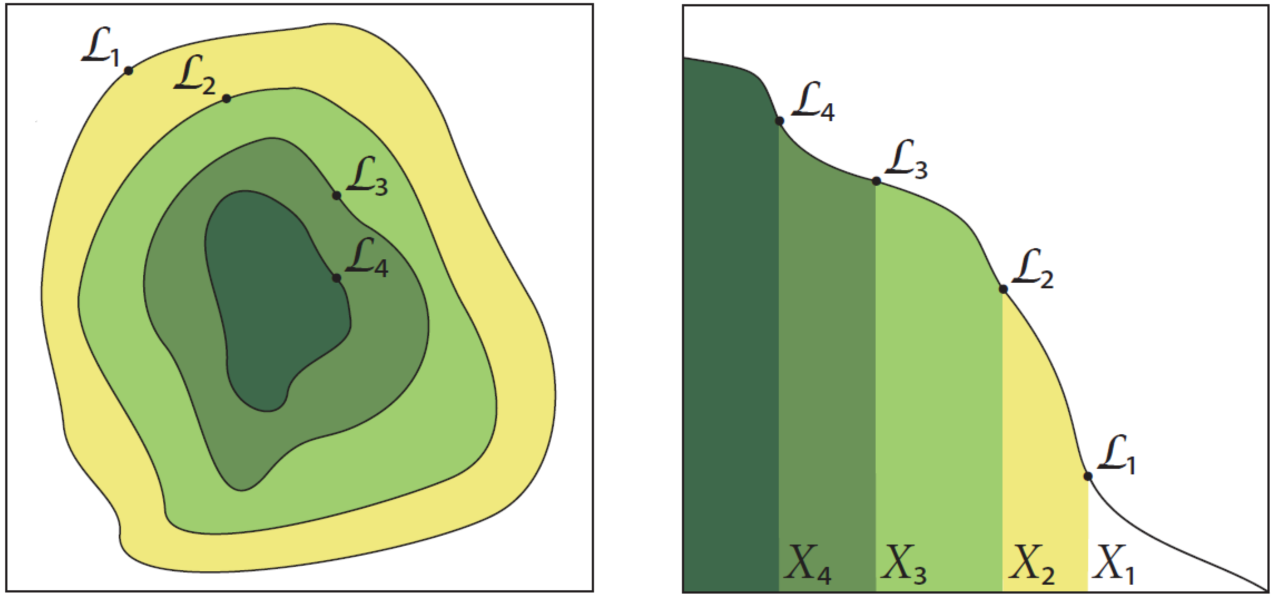
$$\mathcal{Z} \equiv \int_{\Omega_{\boldsymbol{\Theta}}} \mathcal{L}(\boldsymbol{\Theta}) \pi(\boldsymbol{\Theta}) d\boldsymbol{\Theta}$$

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"



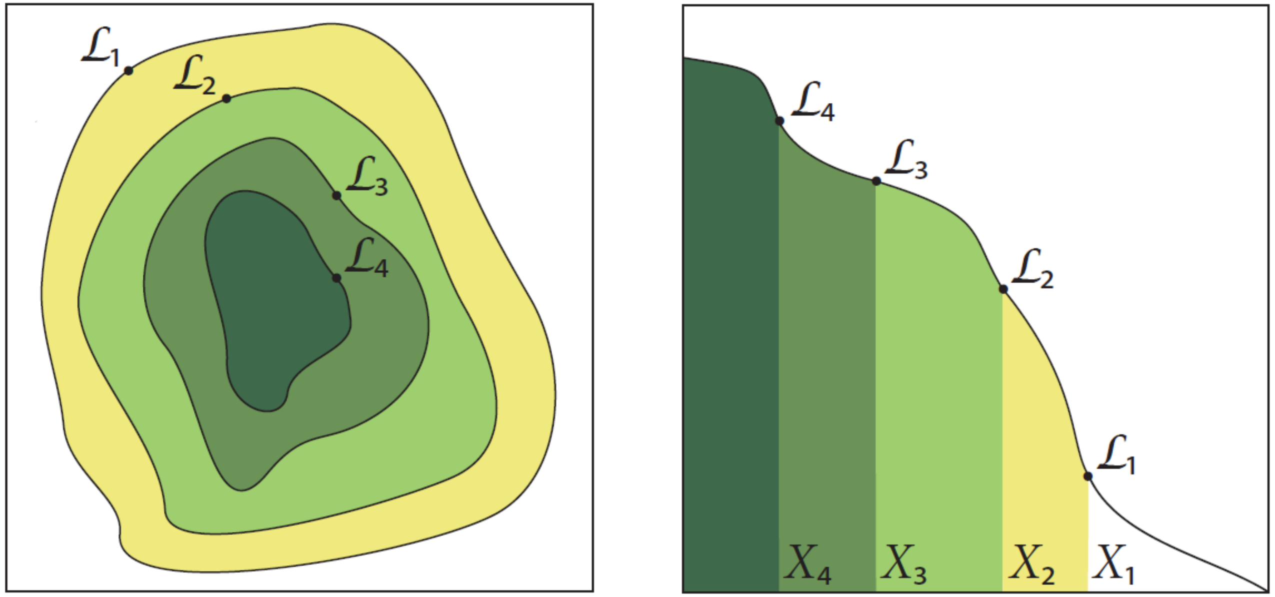
Feroz et al. (2013)

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"



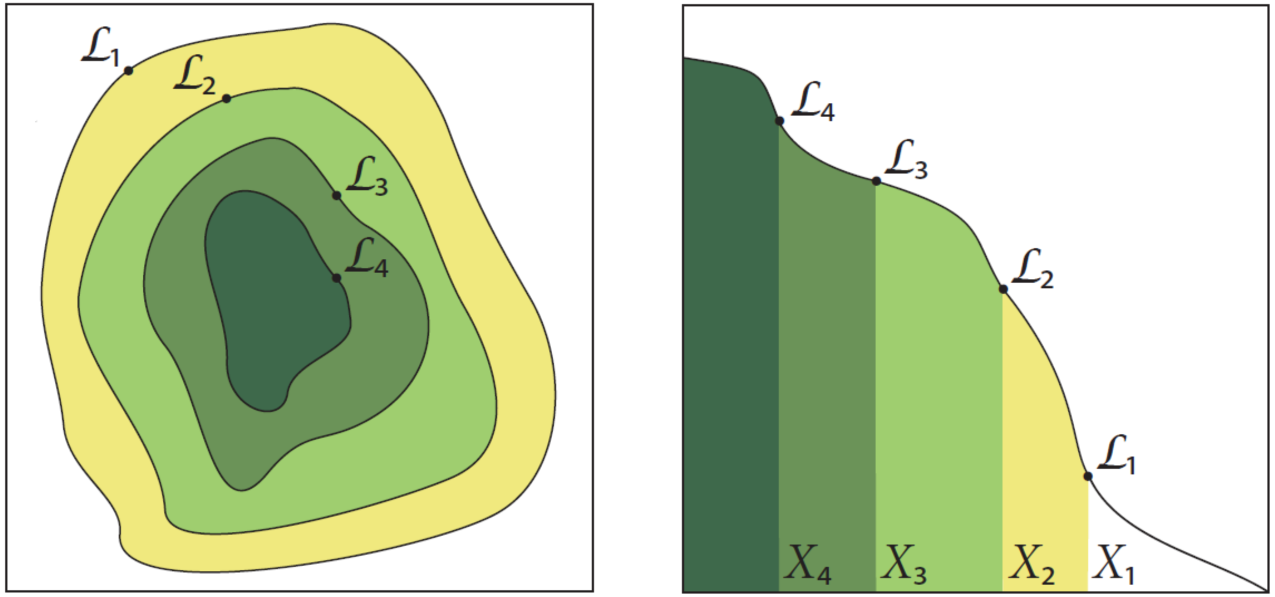
Feroz et al. (2013)

Motivation: Integrating the Posterior

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

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

"Prior Volume"



Feroz et al. (2013)

$f(\mathcal{L}_i, \dots)g(X_i, \dots)$ can be rectangles, trapezoids, etc.

Motivation: Integrating the Posterior

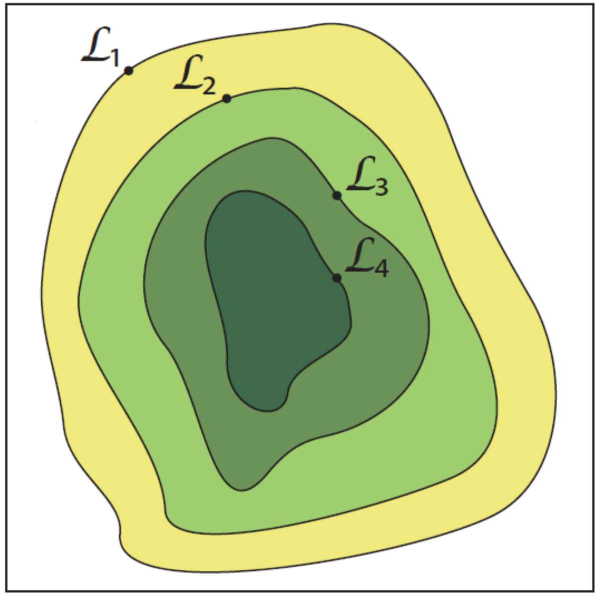
Differential Volume ΔX_i

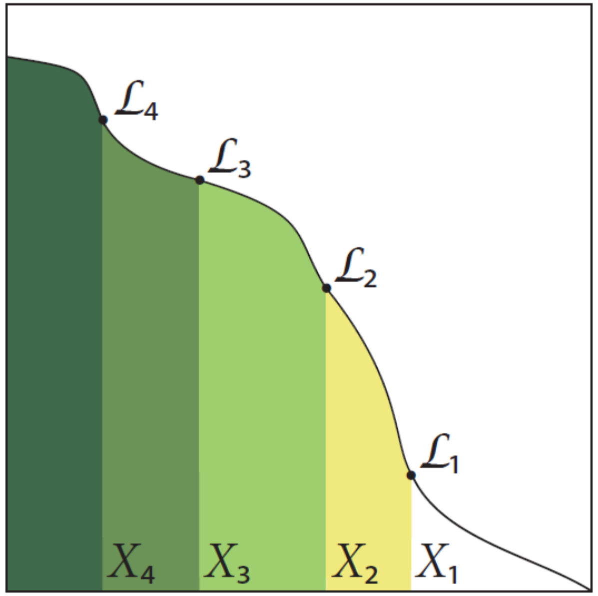
$$\hat{Z} \approx \sum_{i=1}^{n} f(\mathcal{L}_i, \dots)g(X_i, \dots)$$

Amplitude

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

"Prior Volume"





Feroz et al. (2013)

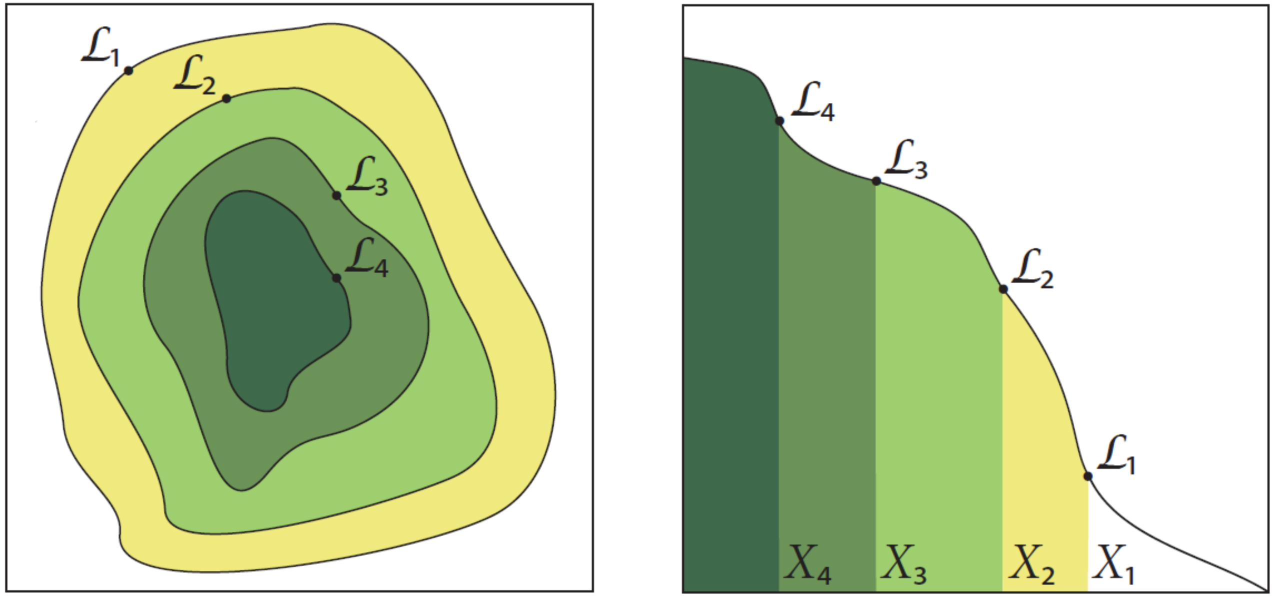
$$\widehat{w}_i = f(\mathcal{L}_i, \ldots)\, g(X_i, \ldots)$$

Motivation: Integrating the Posterior

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

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

"Prior Volume"



Feroz et al. (2013)

$$\hat{w}_i = f(\mathcal{L}_i, \ldots)g(X_i, \ldots)$$

Motivation: Integrating the Posterior

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

We get posteriors "for free"

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

Importance Weight

$$\widehat{w}_i = f(\mathcal{L}_i, \dots)g(X_i, \dots)$$
$$\sim f(p_i, \dots)g(V_i, \dots)$$

Motivation: Integrating the Posterior

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

We get posteriors "for free"

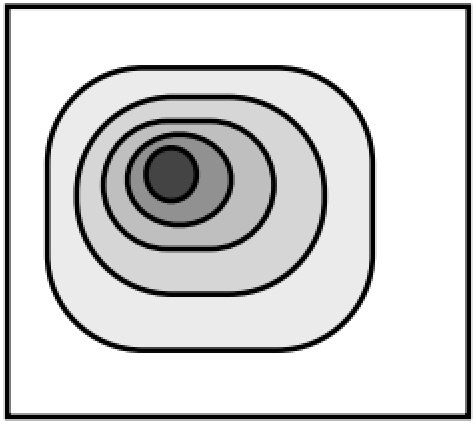
$$\hat{p}_i = \frac{\widehat{w}_i}{\hat{Z}}$$

Importance Weight

Directly proportional to typical set.

Pictures from this 2010 talk by Skilling.

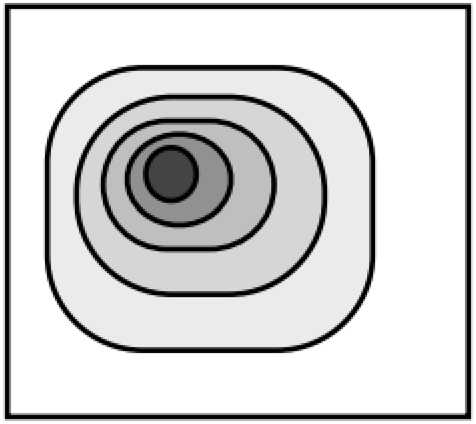
Motivation: Sampling the Posterior



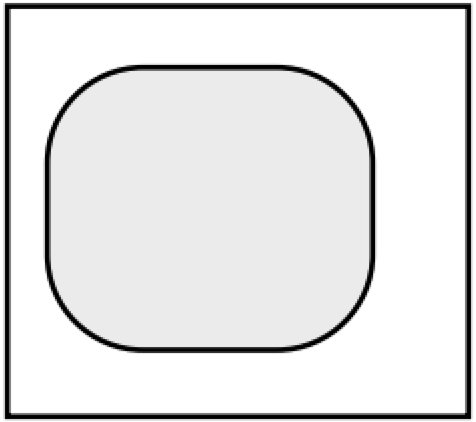
Sampling directly from the likelihood $\mathcal{L}(\boldsymbol{\Theta})$ is **hard**.

Pictures from this 2010 talk by Skilling.

Motivation: Sampling the Posterior

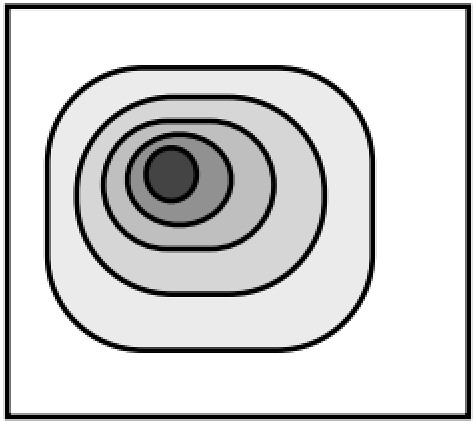


Sampling uniformly within bound $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ is **easier**.

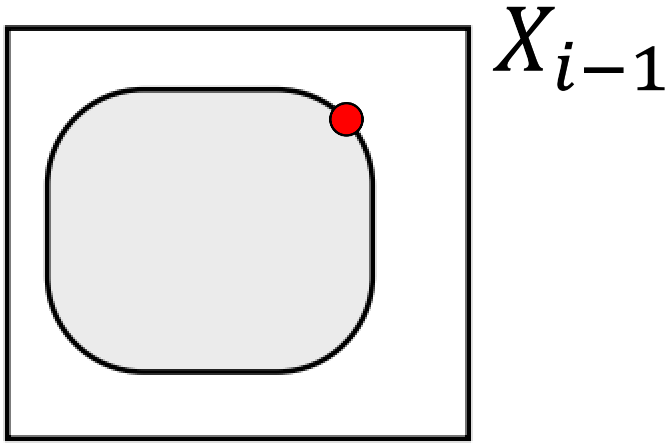


Pictures from this 2010 talk by Skilling.

Motivation: Sampling the Posterior

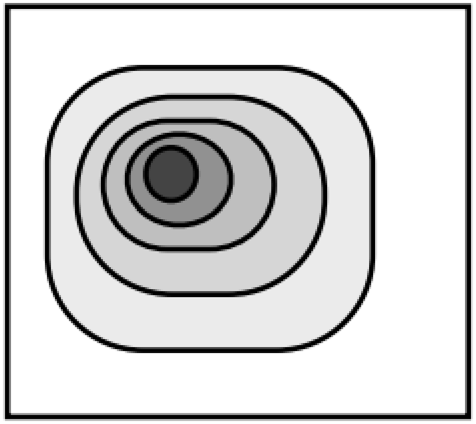


Sampling uniformly within bound $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ is **easier**.

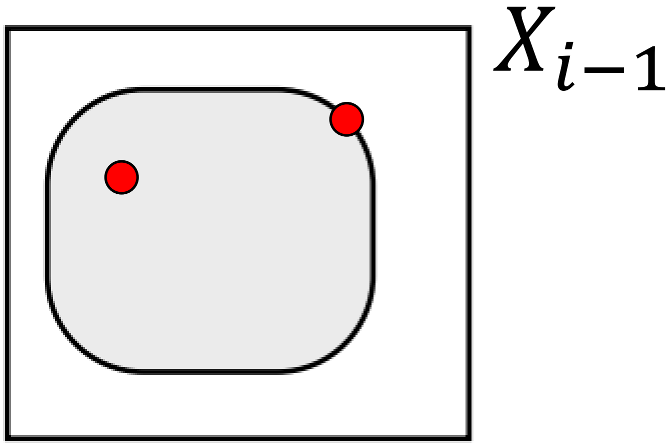


Pictures from this 2010 talk by Skilling.

Motivation: Sampling the Posterior

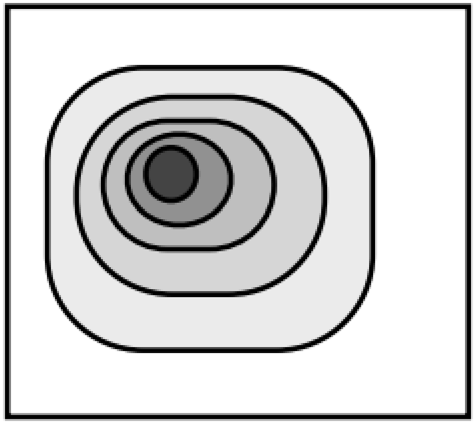


Sampling uniformly within bound $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ is **easier**.

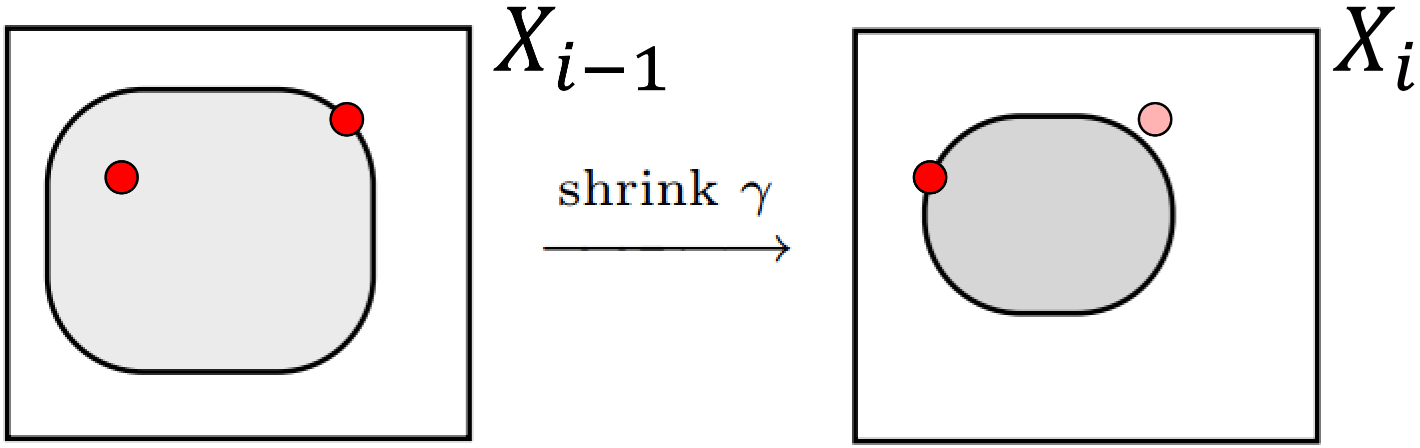


Pictures from [this 2010 talk](#)by Skilling.

Motivation: Sampling the Posterior

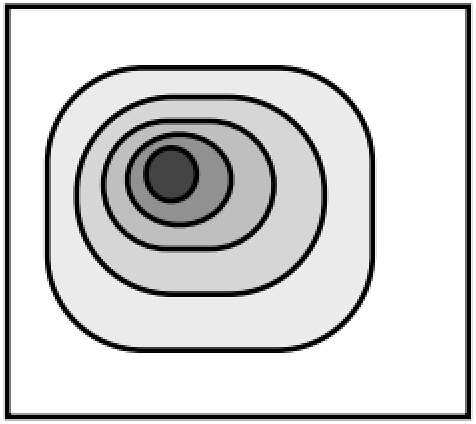


Sampling uniformly within bound $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ is **easier**.

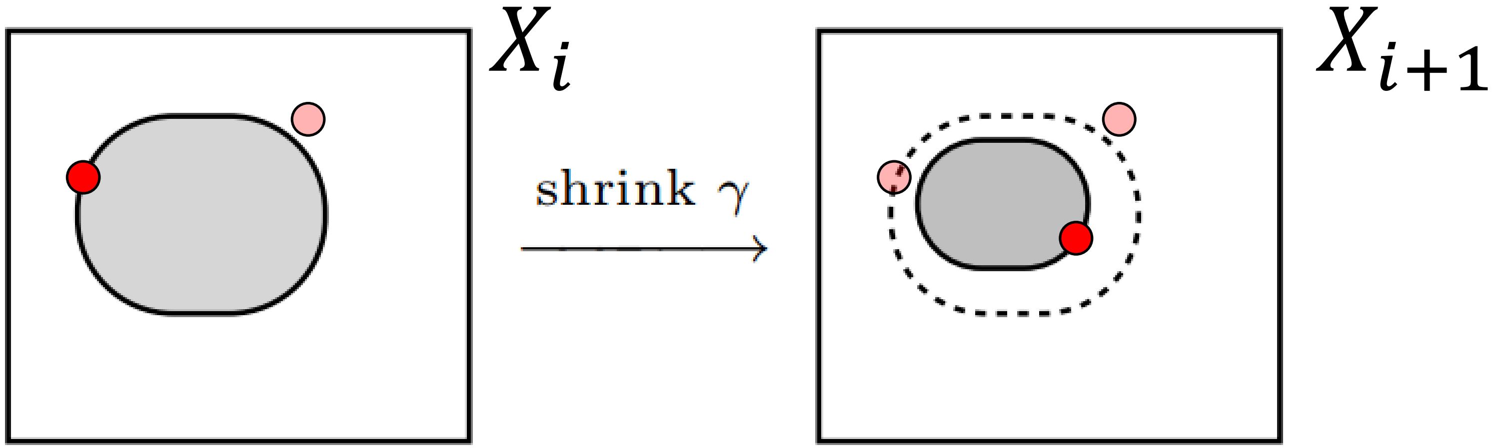


Pictures from [this 2010 talk](#) by Skilling.

Motivation: Sampling the Posterior

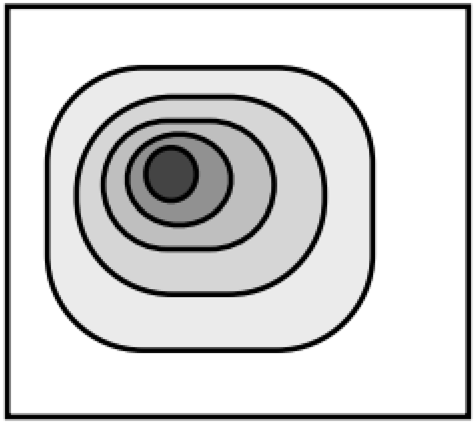


Sampling uniformly within bound $\mathcal{L}(\boldsymbol{\Theta}) > \lambda$ is **easier**.



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

Motivation: Sampling the Posterior

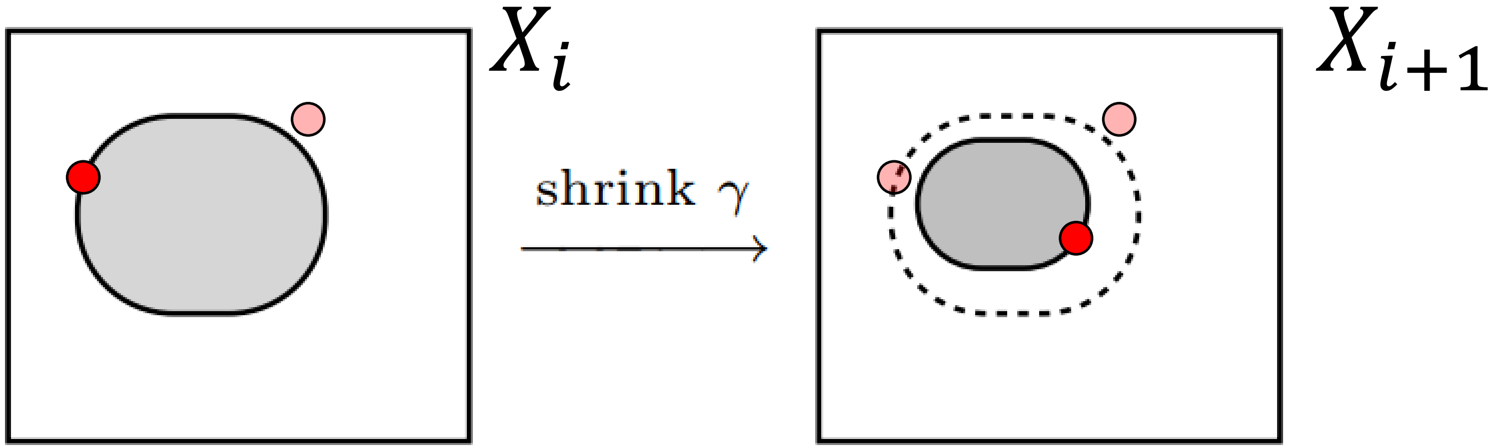


MCMC: Solving a Hard Problem once.

vs

Nested Sampling: Solving an Easier Problem many times.

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



How Nested Sampling Works

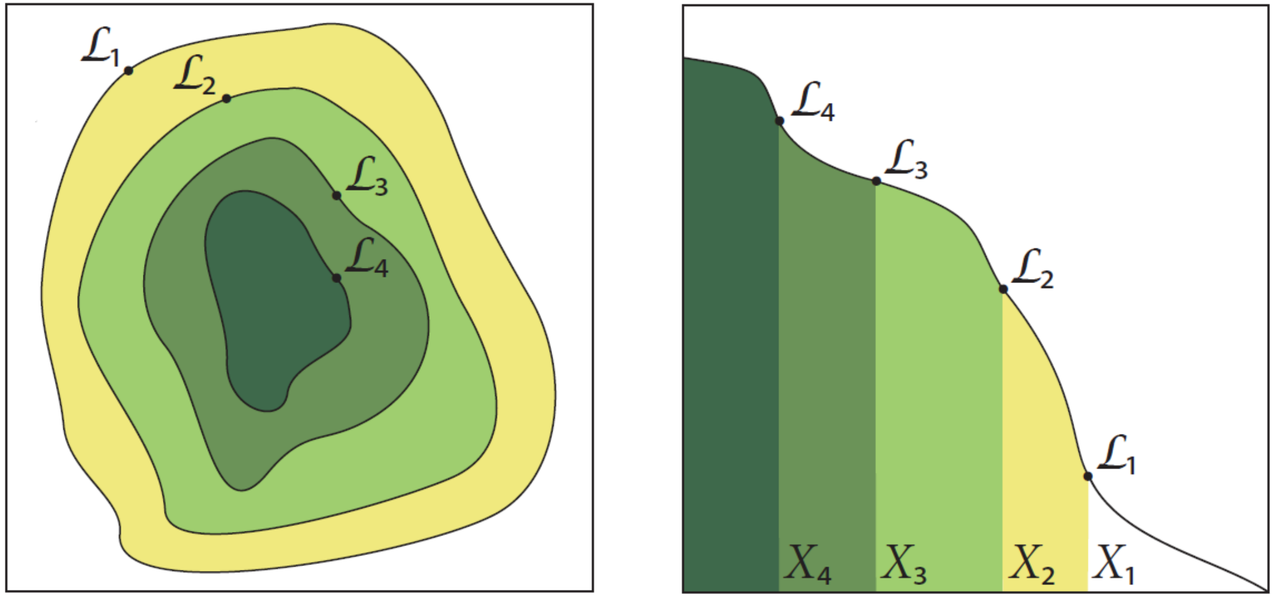
Estimating the Prior Volume

$$\hat{w}_i = f(\mathcal{L}_i, \ldots)\, g(X_i, \ldots)$$

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

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

"Prior Volume"



Feroz et al. (2013)

Estimating the Prior Volume

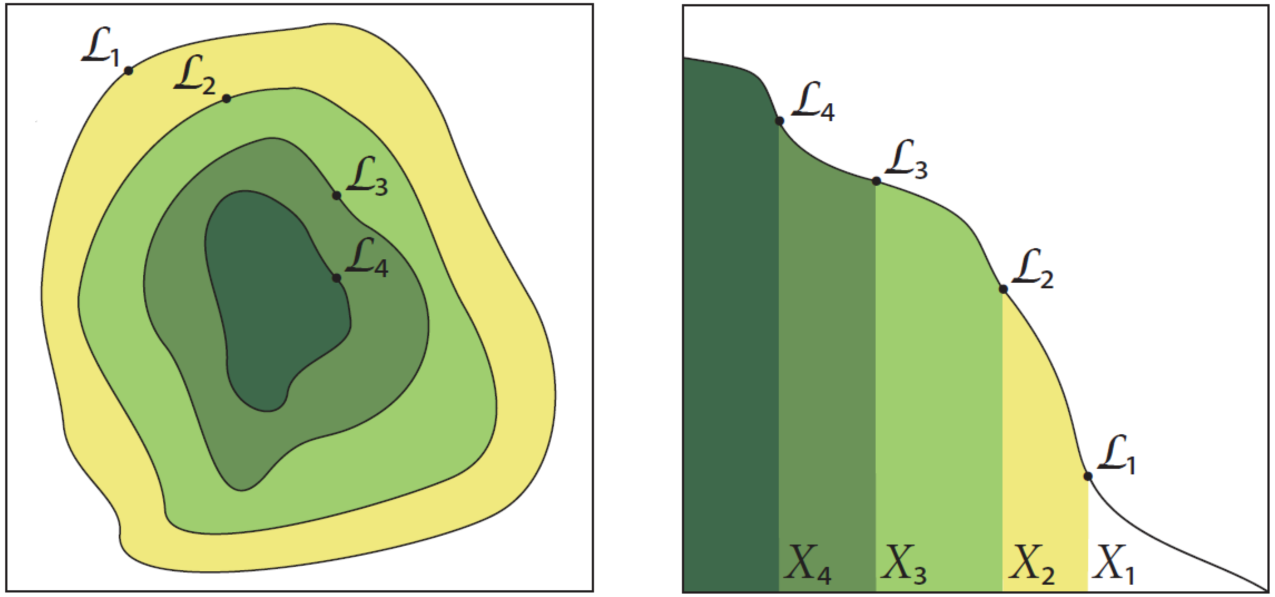
$$\widehat{w}_i = f(\mathcal{L}_i, \ldots)\, g(X_i, \ldots)$$

✓

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

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

"Prior Volume"



Feroz et al. (2013)

Estimating the Prior Volume

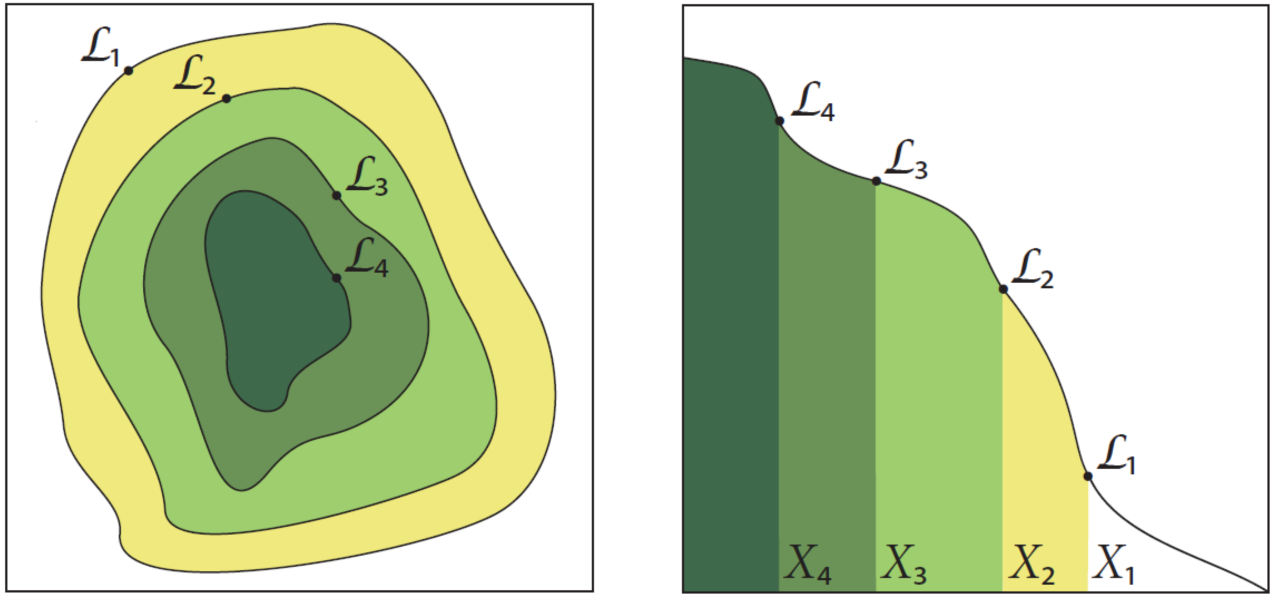
$$\hat{w}_i = f(\mathcal{L}_i, \dots)\, g(X_i, \dots)$$

✓ ???

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

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

"Prior Volume"



Feroz et al. (2013)

Estimating the Prior Volume

$$\hat{w}_i = f(\mathcal{L}_i, \ldots)\, g(X_i, \ldots)$$

???

Probability Integral Transform

PDF

$$X \sim f(x)$$

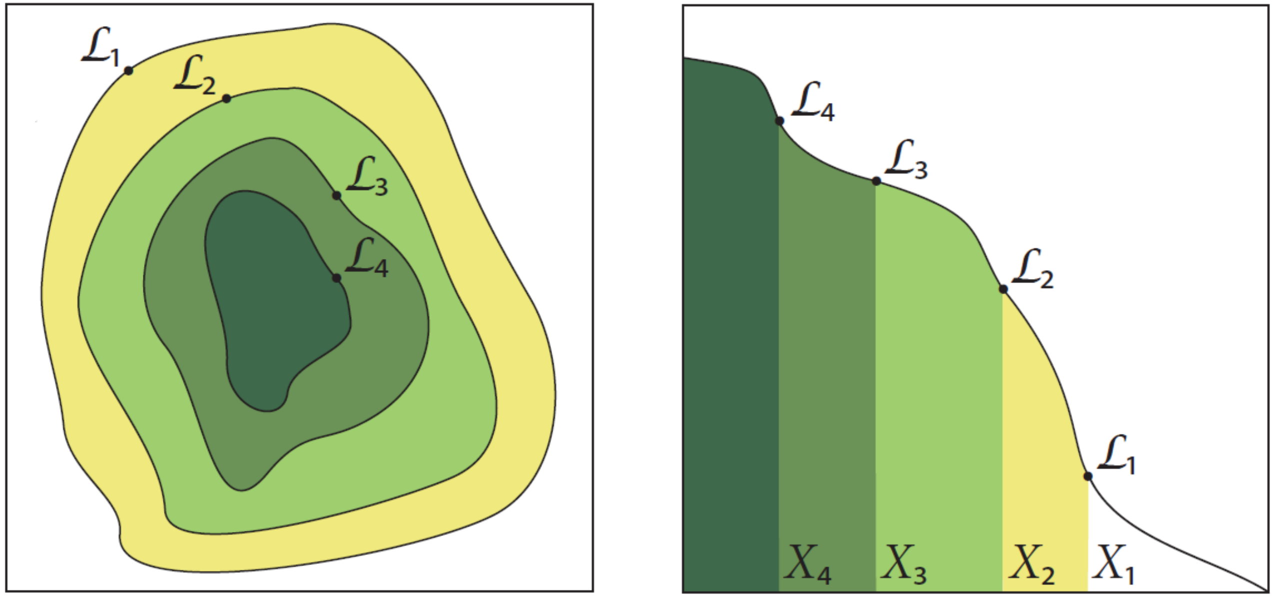
$$F(X) \sim \text{Unif}$$

CDF

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

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

"Prior Volume"



Feroz et al. (2013)

Estimating the Prior Volume

$$\hat{w}_i = f(\mathcal{L}_i, \ldots)\, g(X_i, \ldots)$$

✓ ???

Probability Integral Transform

PDF

$$X \sim f(x)$$

$$F(X) \sim \text{Unif}$$

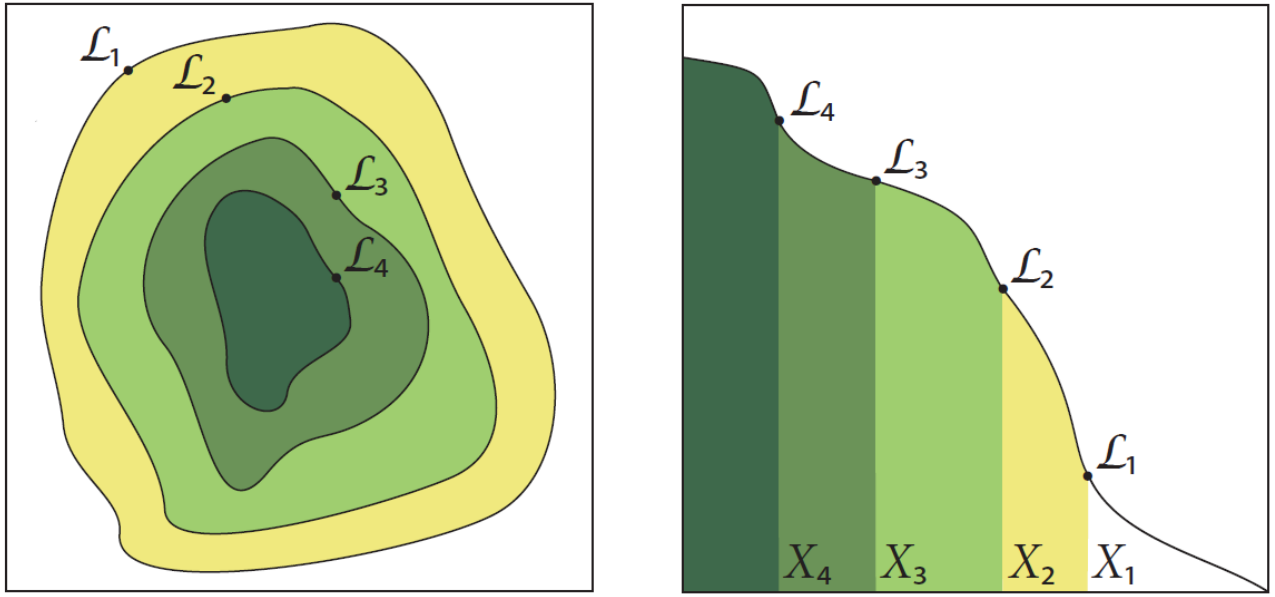
CDF

n

Need to sample from the constrained prior.

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

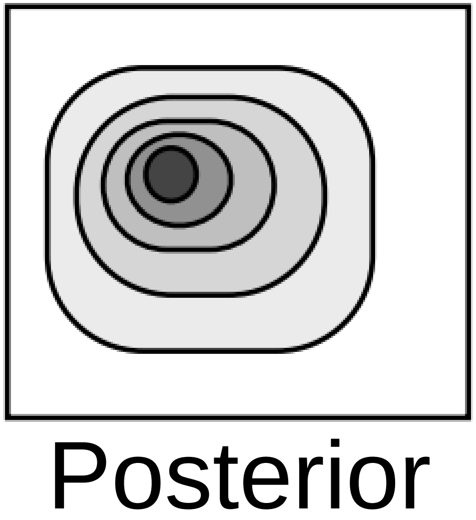
"Prior Volume"

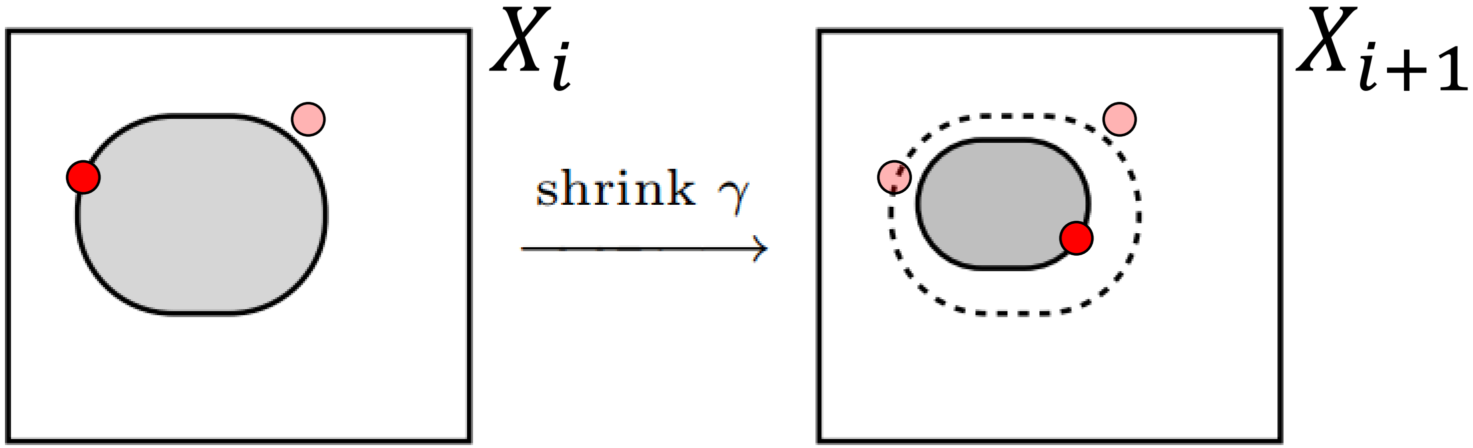


Feroz et al. (2013)

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

Estimating the Prior Volume

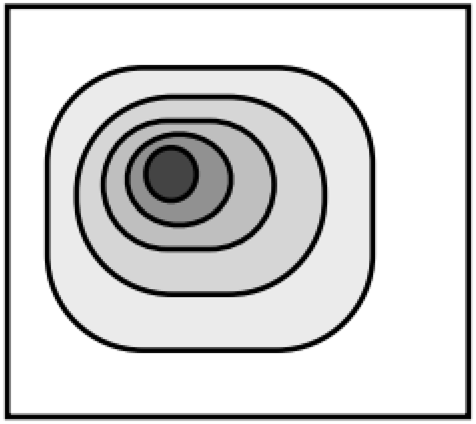




Pictures from [this 2010 talk](#) by Skilling.

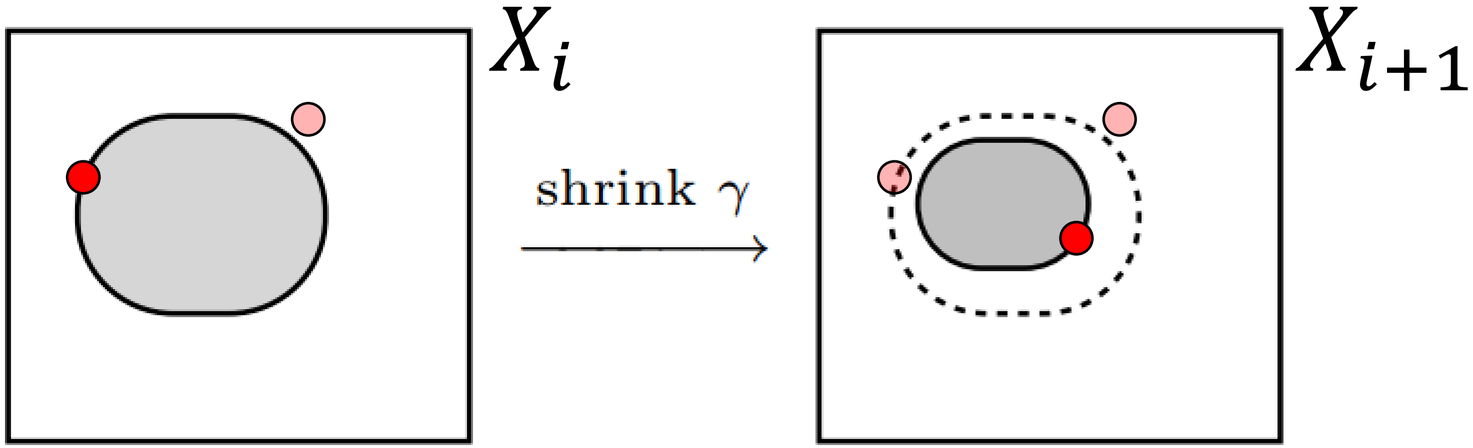
Estimating the Prior Volume

$U_i \sim \mathrm{Unif}$



Posterior

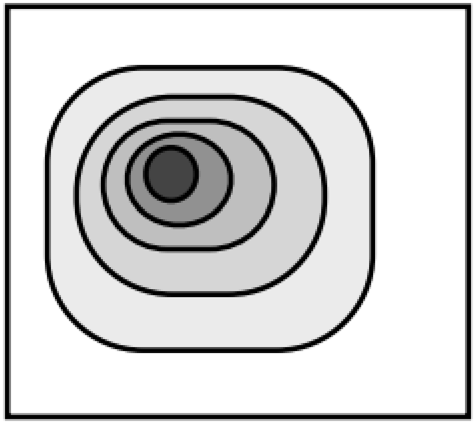
$X_{i+1} = U_i X_i$



Pictures from this 2010 talk by Skilling.

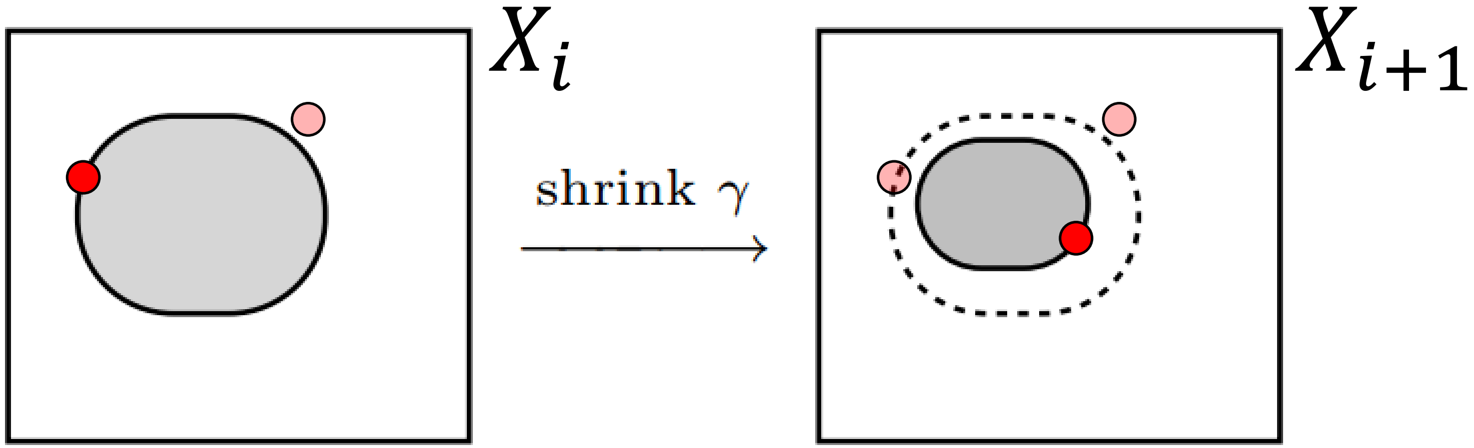
Estimating the Prior Volume

$U_0, \ldots, U_i \overset{\text{i.i.d.}}{\sim} \text{Unif}$



Posterior

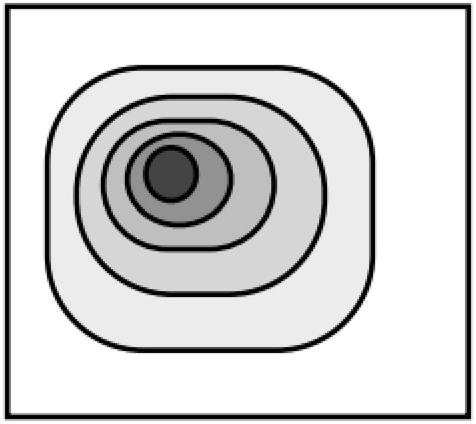
$$X_{i+1} = \prod_{j=0}^{i} U_j \, X_0$$



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

Estimating the Prior Volume

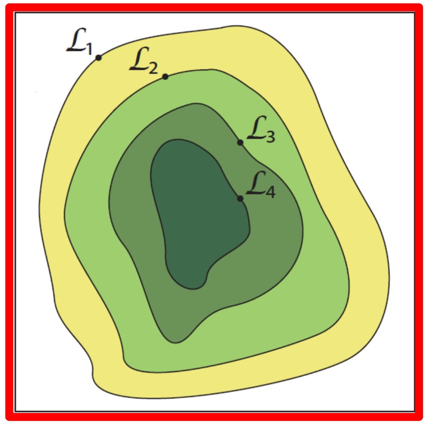
$$U_0, \ldots, U_i \overset{\text{i.i.d.}}{\sim} \text{Unif}$$

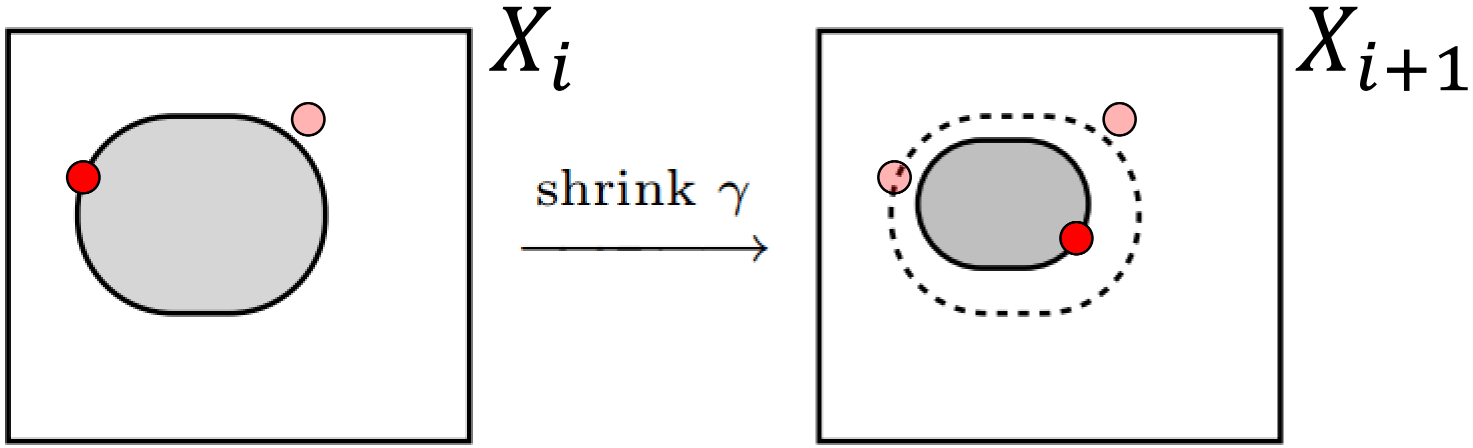


Posterior

$$X_{i+1} = \prod_{j=0}^{i} U_j \, X_0$$

$$X_0 \equiv 1$$

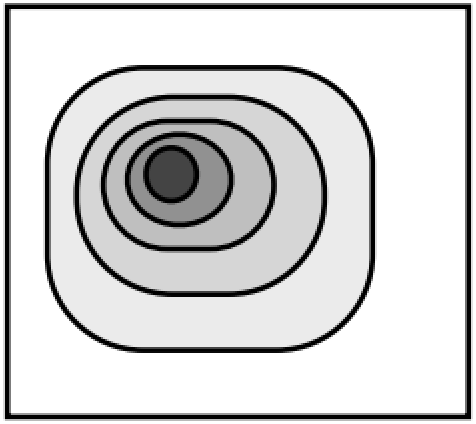




Pictures from this 2010 talk by Skilling.

Estimating the Prior Volume

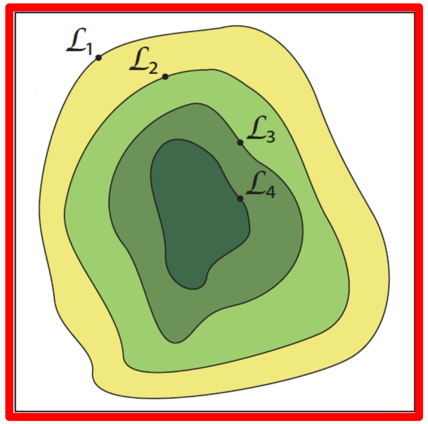
$$U_0, \ldots, U_i \overset{\text{i.i.d.}}{\sim} \text{Unif}$$

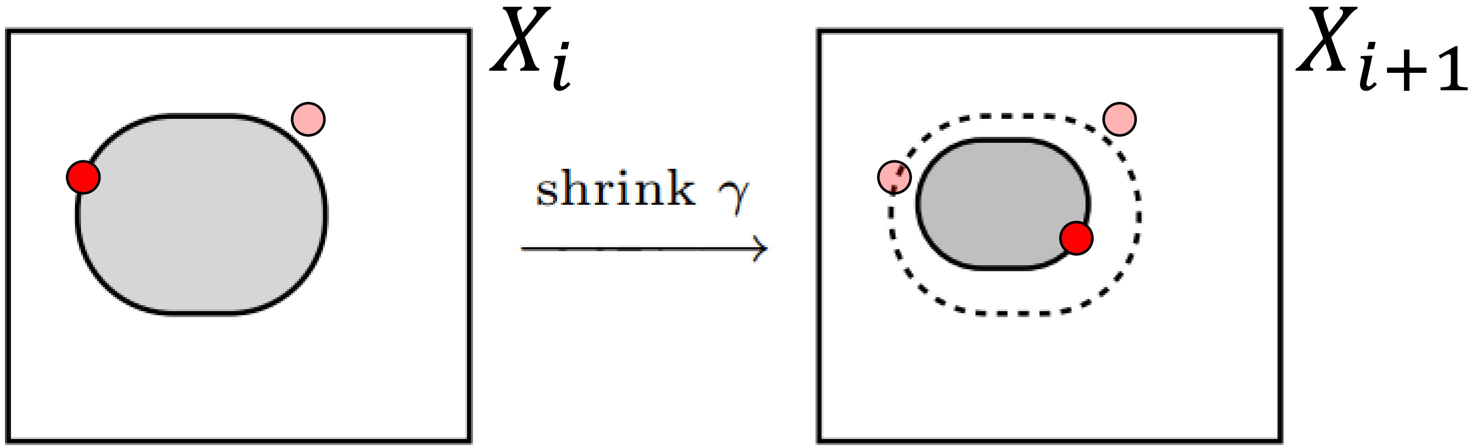


Posterior

$$X_{i+1} = \prod_{j=0}^{i} U_j$$

$$X_0 \equiv 1$$





Estimating the Prior Volume

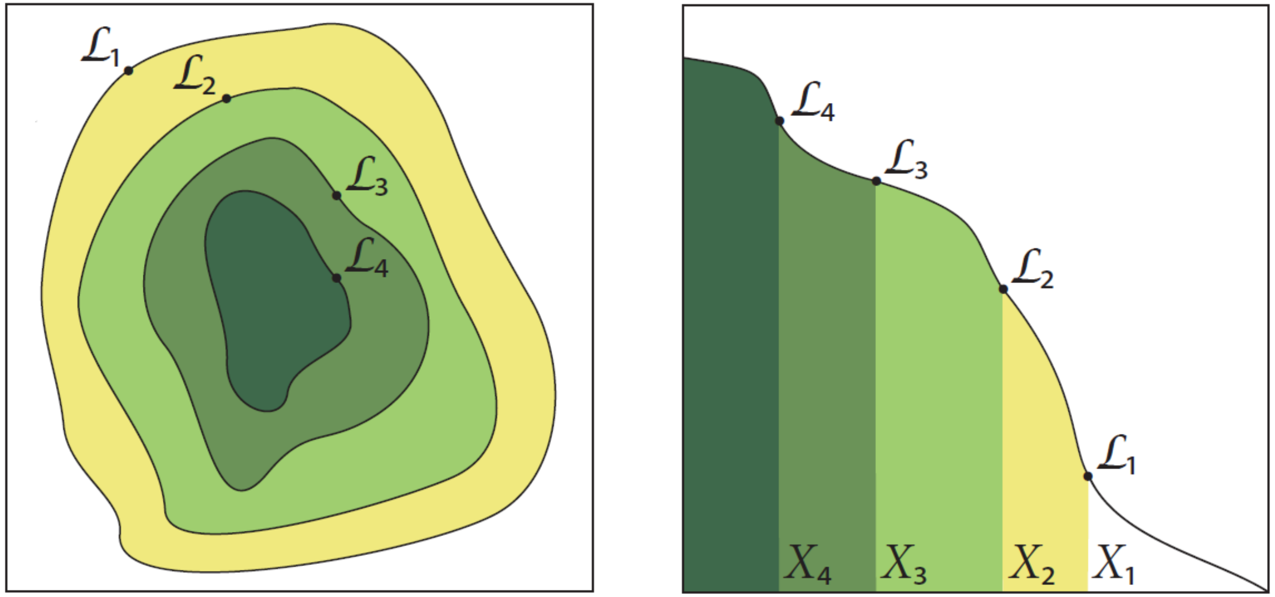
$$\hat{w}_i = f(\mathcal{L}_i, \dots)\, g(X_i, \dots)$$

✓ ???

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

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

"Prior Volume"



Feroz et al. (2013)

Estimating the Prior Volume

$$\hat{w}_i = f(\mathcal{L}_i, \ldots)\, g(X_i, \ldots)$$

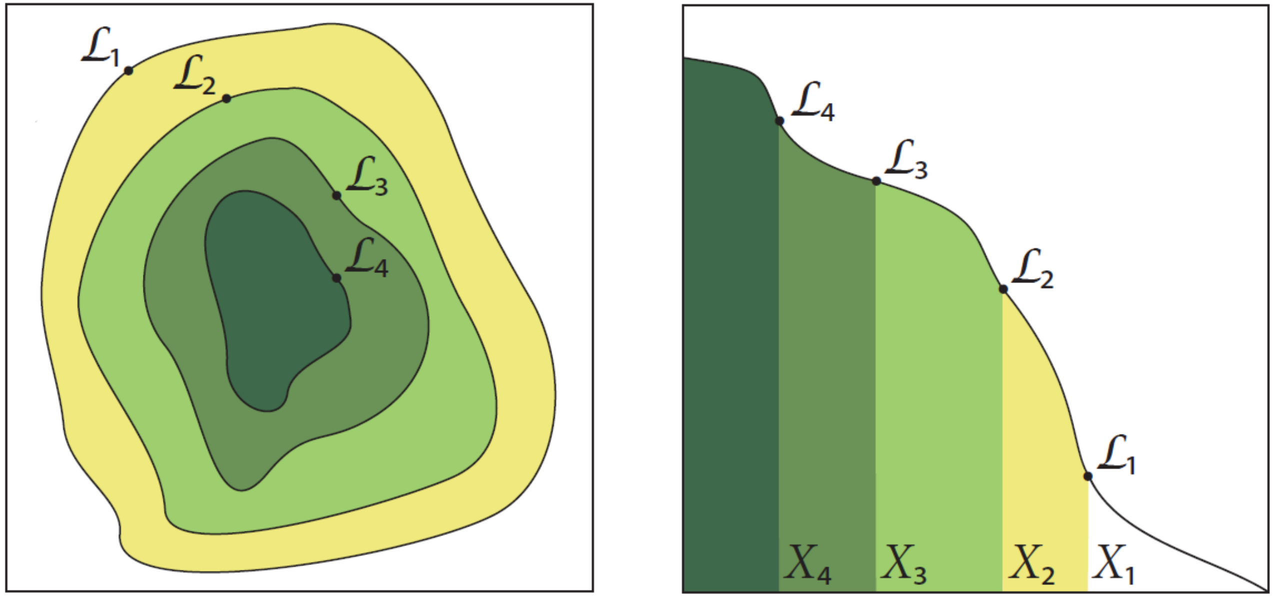
$$\hat{Z} \approx \sum_{i=1}^{n} \hat{w}_i$$

$$\Pr(X_1, X_2, \ldots, X_i)$$

where 1st & 2nd moments can be computed

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

"Prior Volume"



Feroz et al. (2013)

Nested Sampling Algorithm

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$\boldsymbol{\Theta}_N \quad \boldsymbol{\Theta}_{N-1} \quad \cdots \quad \boldsymbol{\Theta}_2 \quad \boldsymbol{\Theta}_1$$

Nested Sampling Algorithm (Ideal)

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$\boldsymbol{\Theta}_N \quad \boldsymbol{\Theta}_{N-1} \quad \cdots \quad \boldsymbol{\Theta}_2 \quad \boldsymbol{\Theta}_1$$

Samples sequentially drawn
from constrained prior $\pi_{\mathcal{L}>\mathcal{L}_i}(\boldsymbol{\Theta})$.

Nested Sampling Algorithm (Naïve)

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$\boldsymbol{\Theta}_1 \sim \pi(\boldsymbol{\Theta})$$

1. Samples sequentially drawn from prior $\pi(\boldsymbol{\Theta})$.
2. New point only accepted if $\mathcal{L}_{i+1} > \mathcal{L}_i$.

Nested Sampling Algorithm (Naïve)

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$\boldsymbol{\Theta}_2 \quad \boldsymbol{\Theta}_1 \sim \pi(\boldsymbol{\Theta})$$

1. Samples sequentially drawn from prior $\pi(\boldsymbol{\Theta})$.
2. New point only accepted if $\mathcal{L}_{i+1} > \mathcal{L}_i$.

Nested Sampling Algorithm (Naïve)

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$\mathbf{\Theta}_N \quad \mathbf{\Theta}_{N-1} \quad \cdots \quad \mathbf{\Theta}_2 \quad \mathbf{\Theta}_1 \sim \pi(\mathbf{\Theta})$$

1. Samples sequentially drawn from prior $\pi(\mathbf{\Theta})$.
2. New point only accepted if $\mathcal{L}_{i+1} > \mathcal{L}_i$.

Building a Nested Sampling Algorithm

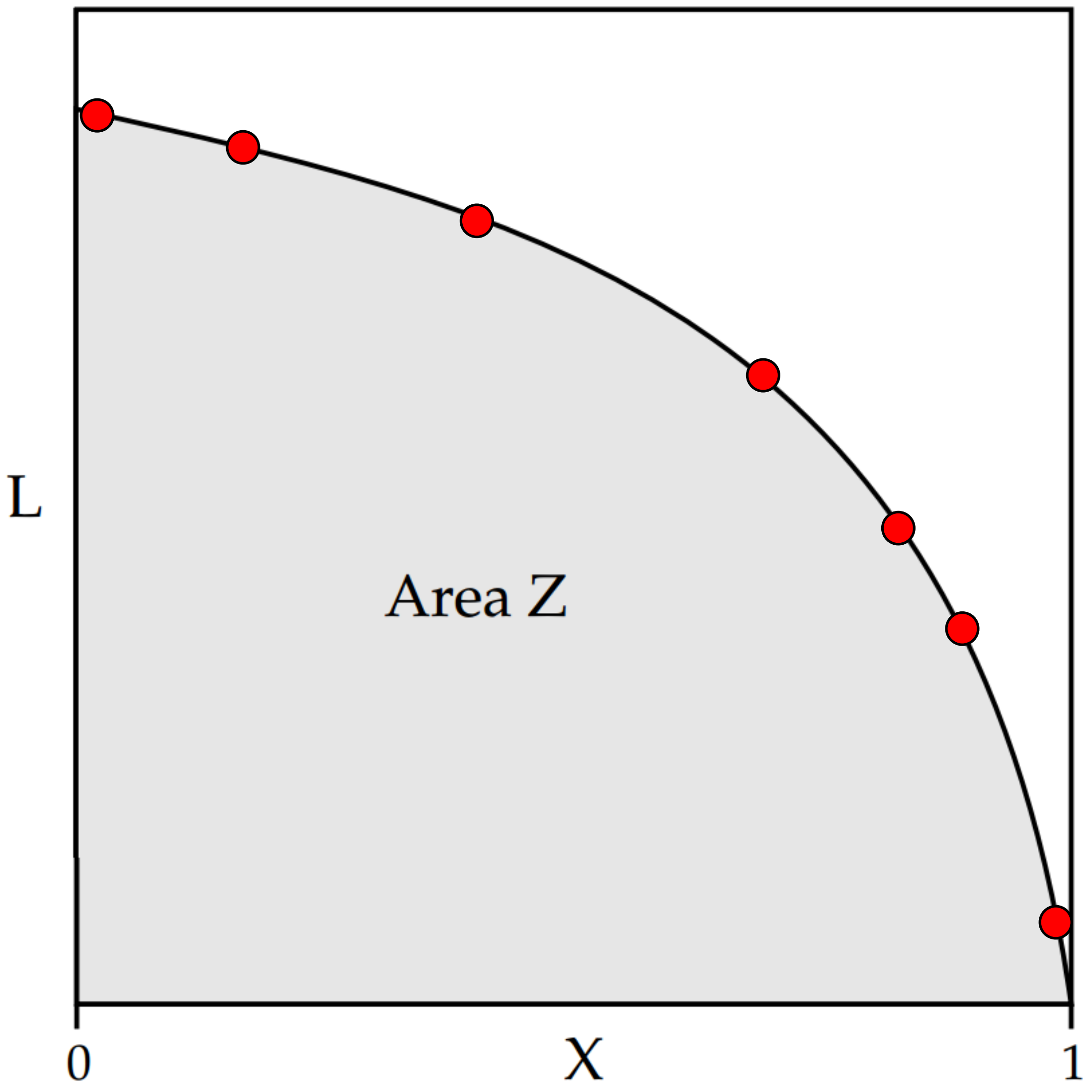
Components of the Algorithm

1. Adding more particles.
2. Knowing when to stop.
3. What to do after stopping.

Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \cdots > \mathcal{L}_2^{(1)} > \mathcal{L}_1^{(1)} > 0$$

Skilling (2006)

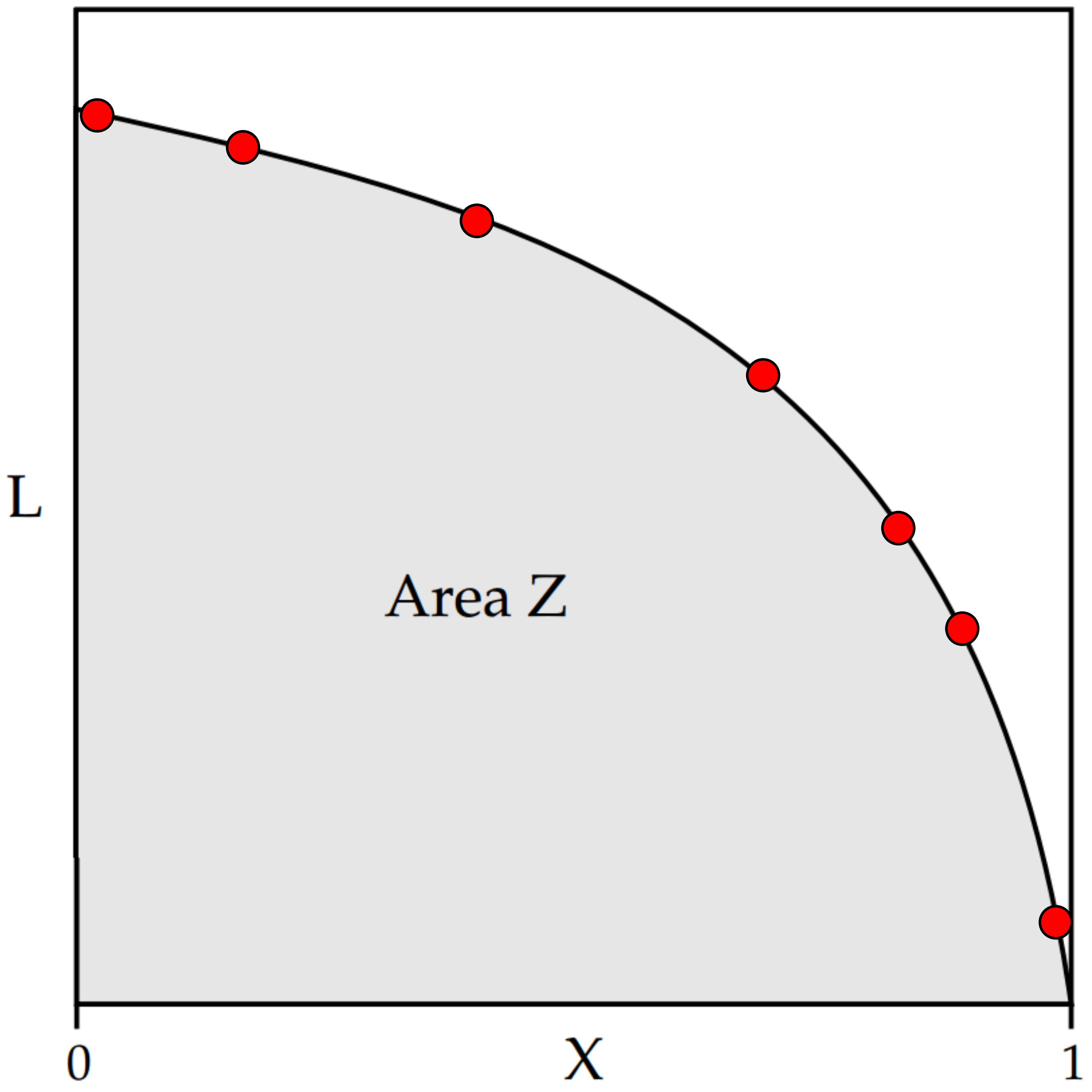


Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \cdots > \mathcal{L}_2^{(1)} > \mathcal{L}_1^{(1)} > 0$$

$$X_{i+1}^{(1)} = \prod_{j=0}^{i} U_j$$

Skilling (2006)

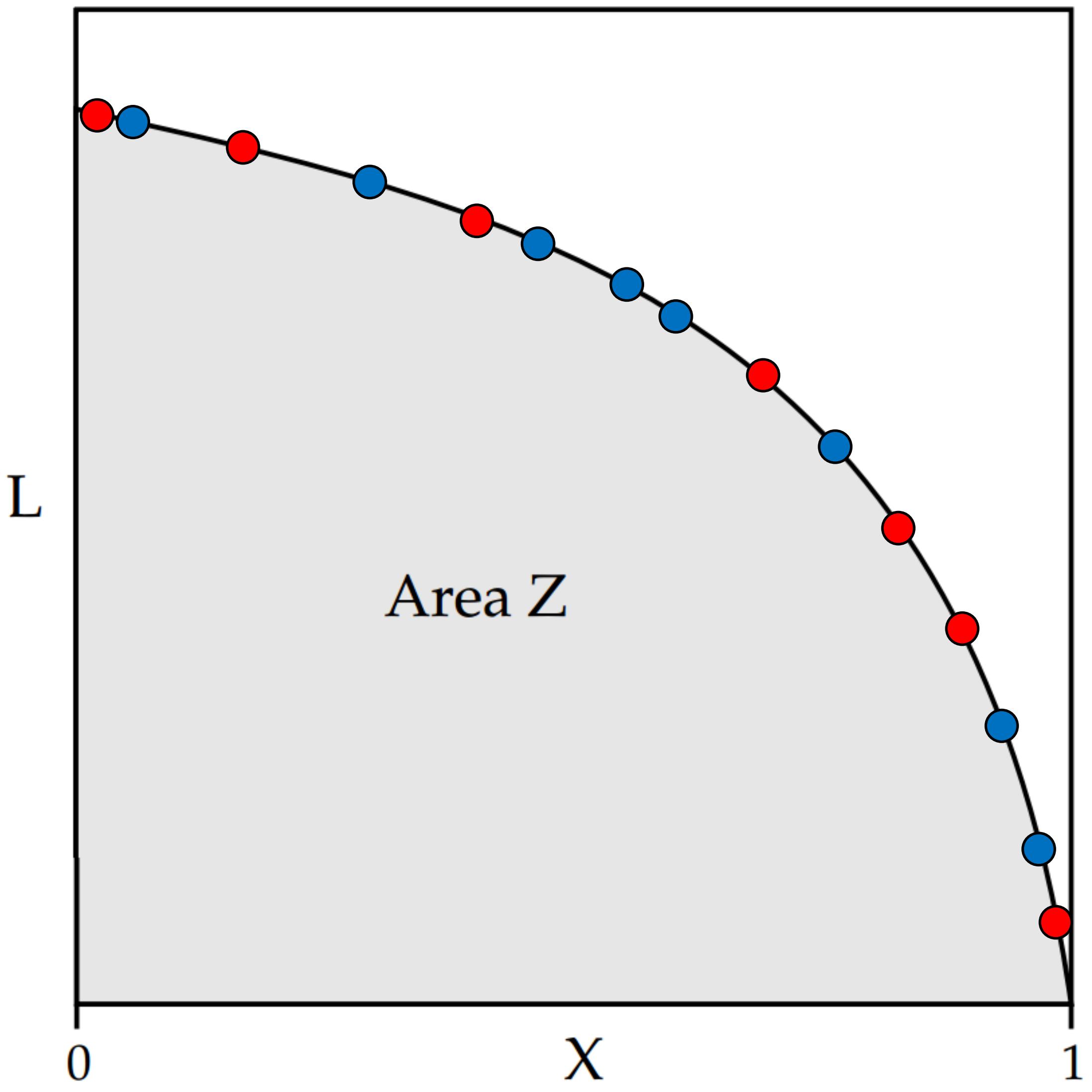


Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \cdots > \mathcal{L}_2^{(1)} > \mathcal{L}_1^{(1)} > 0$$

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > 0$$

Skilling (2006)



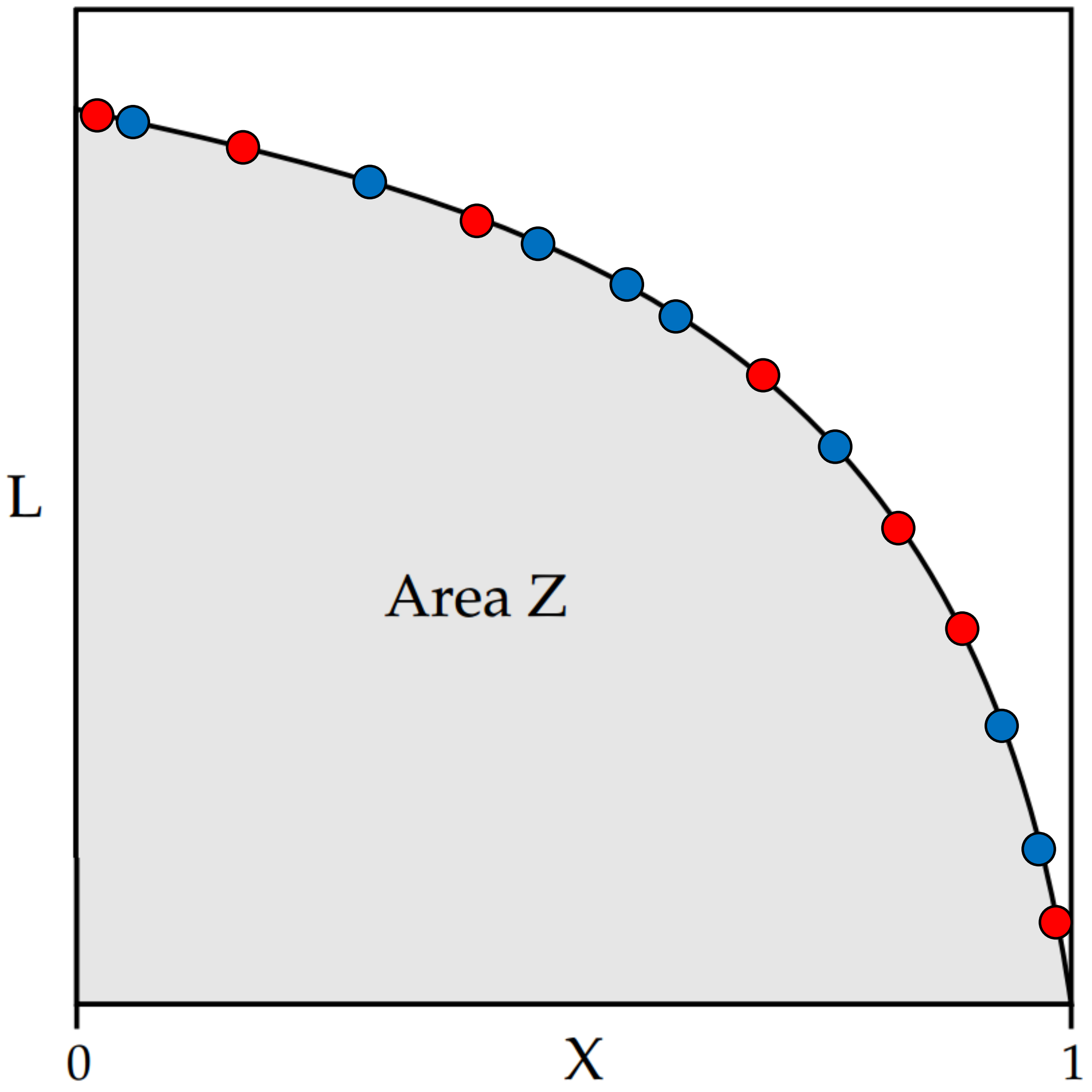
Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \cdots > \mathcal{L}_2^{(1)} > \mathcal{L}_1^{(1)} > 0$$

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > 0$$

$$X_{i+1}^{(1)} = \prod_{j=0}^{i} U_j \qquad X_{j+1}^{(2)} = \prod_{k=0}^{j} U_k$$

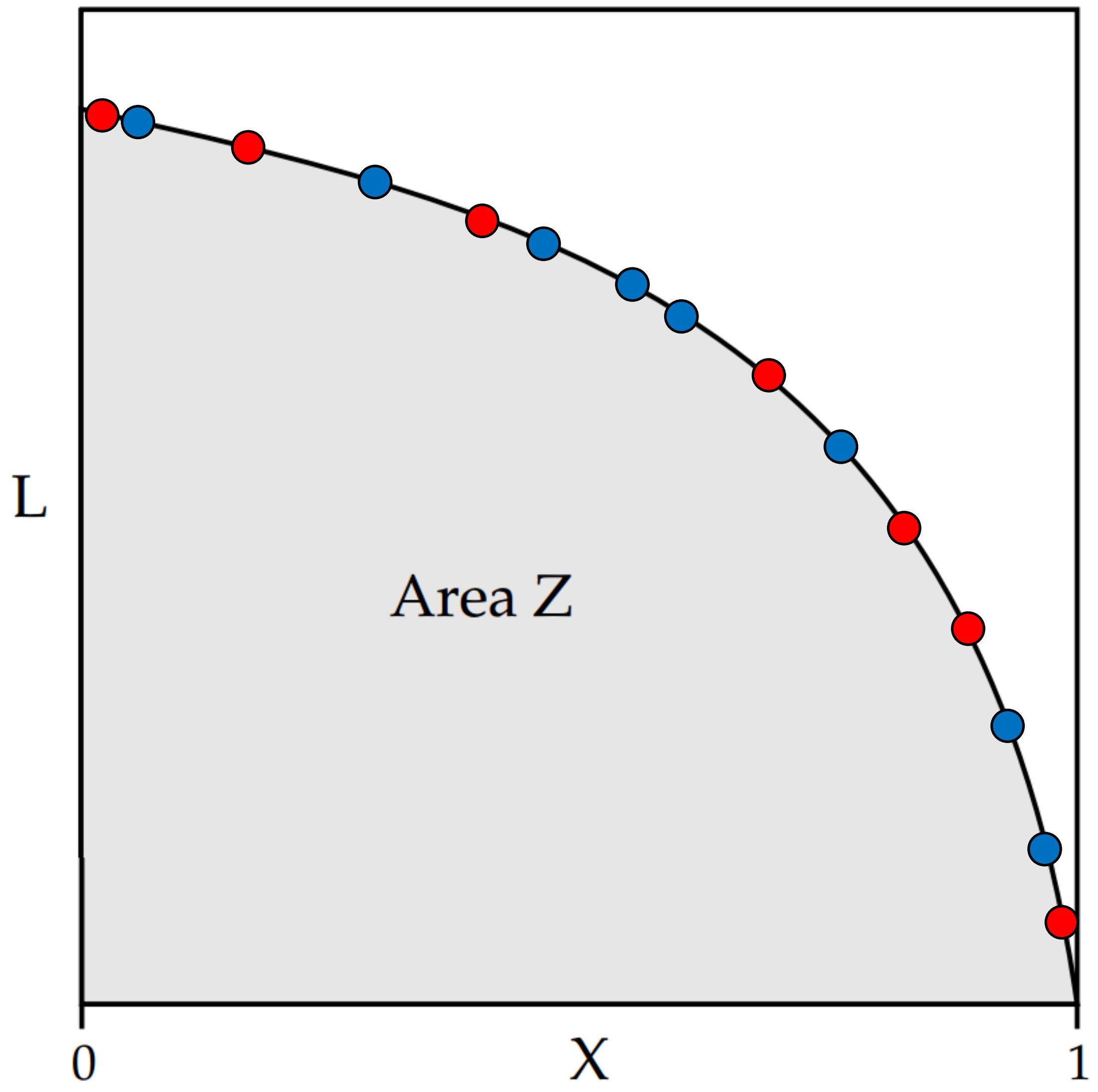
Skilling (2006)



Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

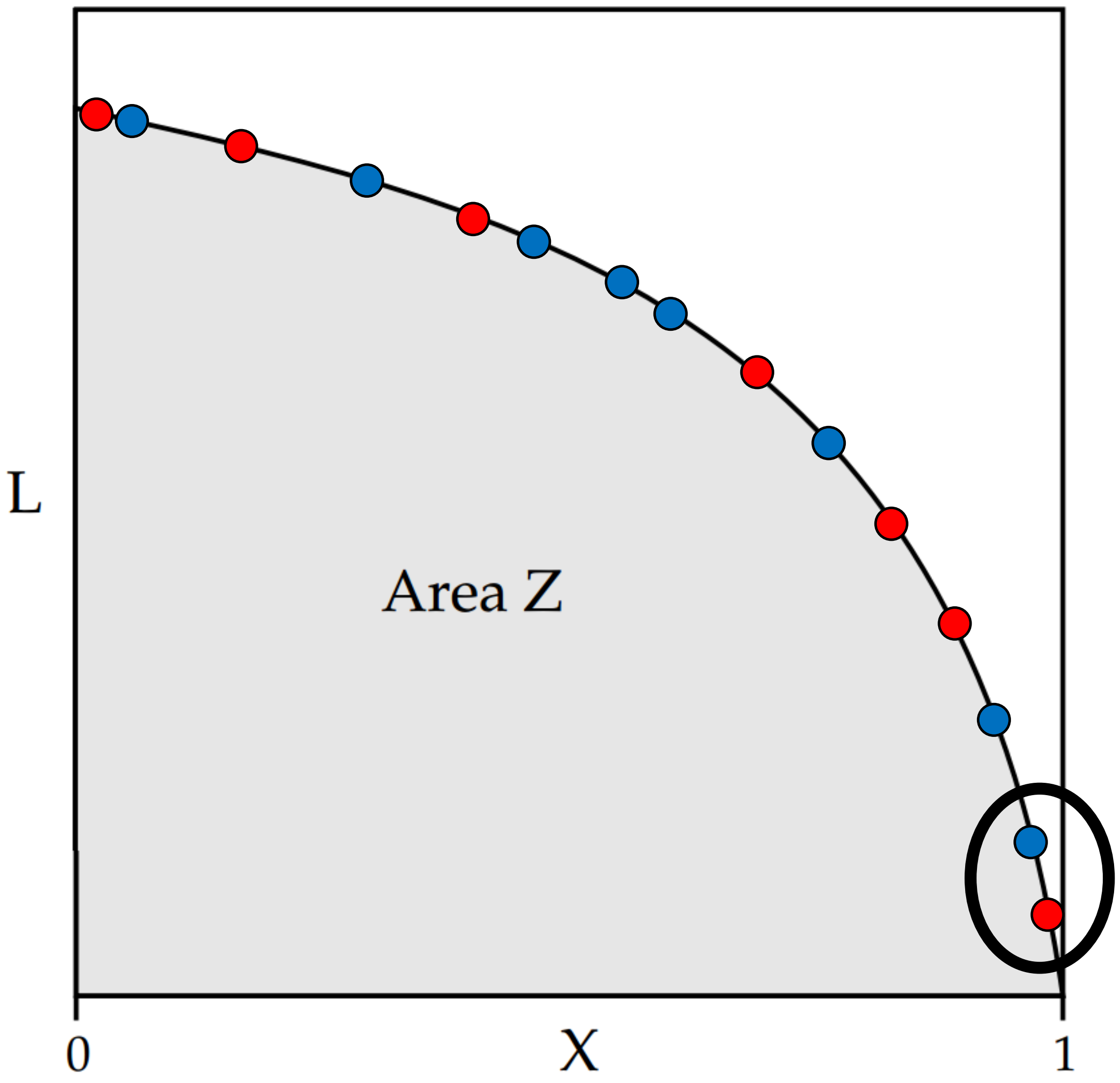
Skilling (2006)



Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

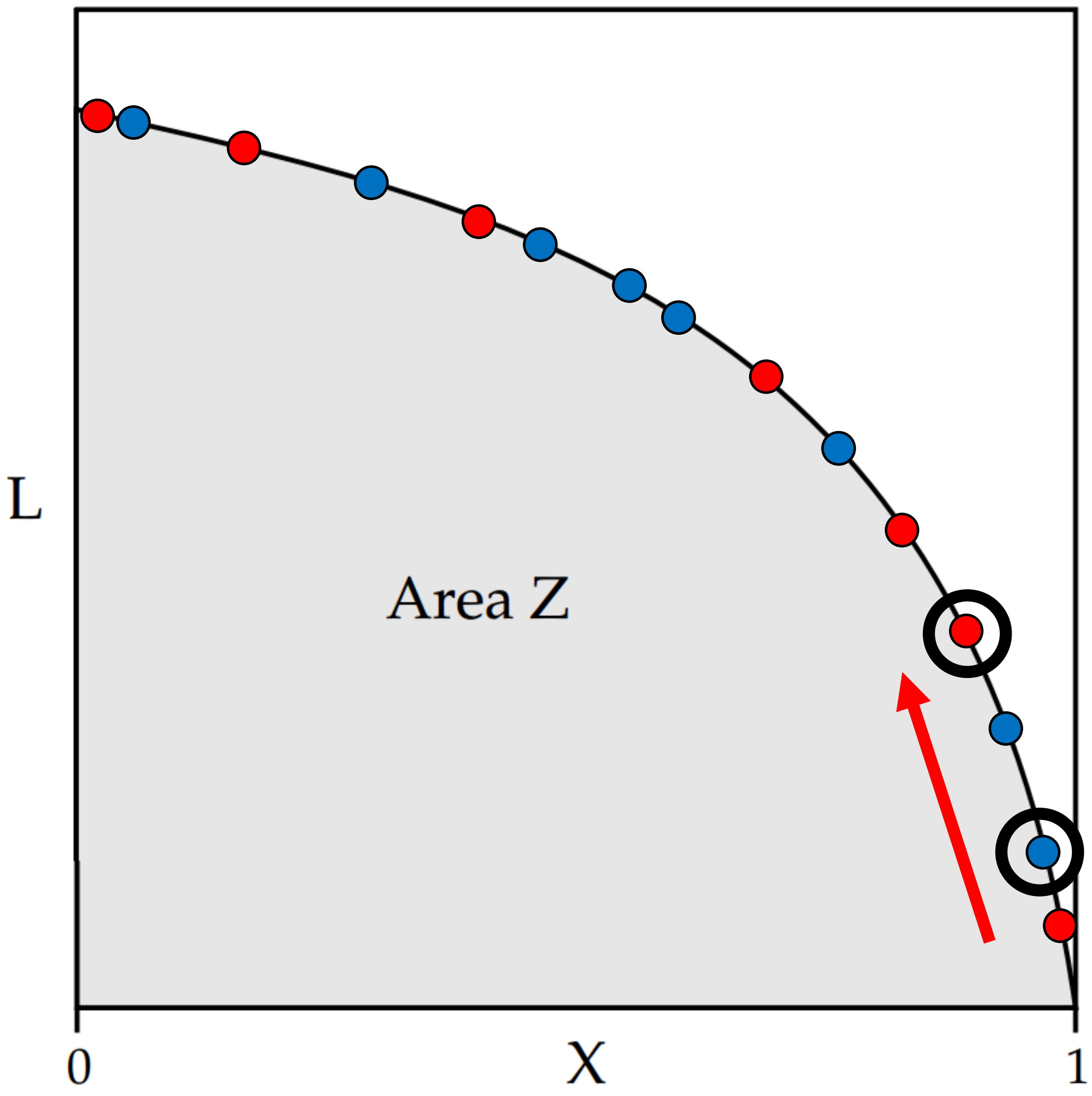
Skilling (2006)



Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

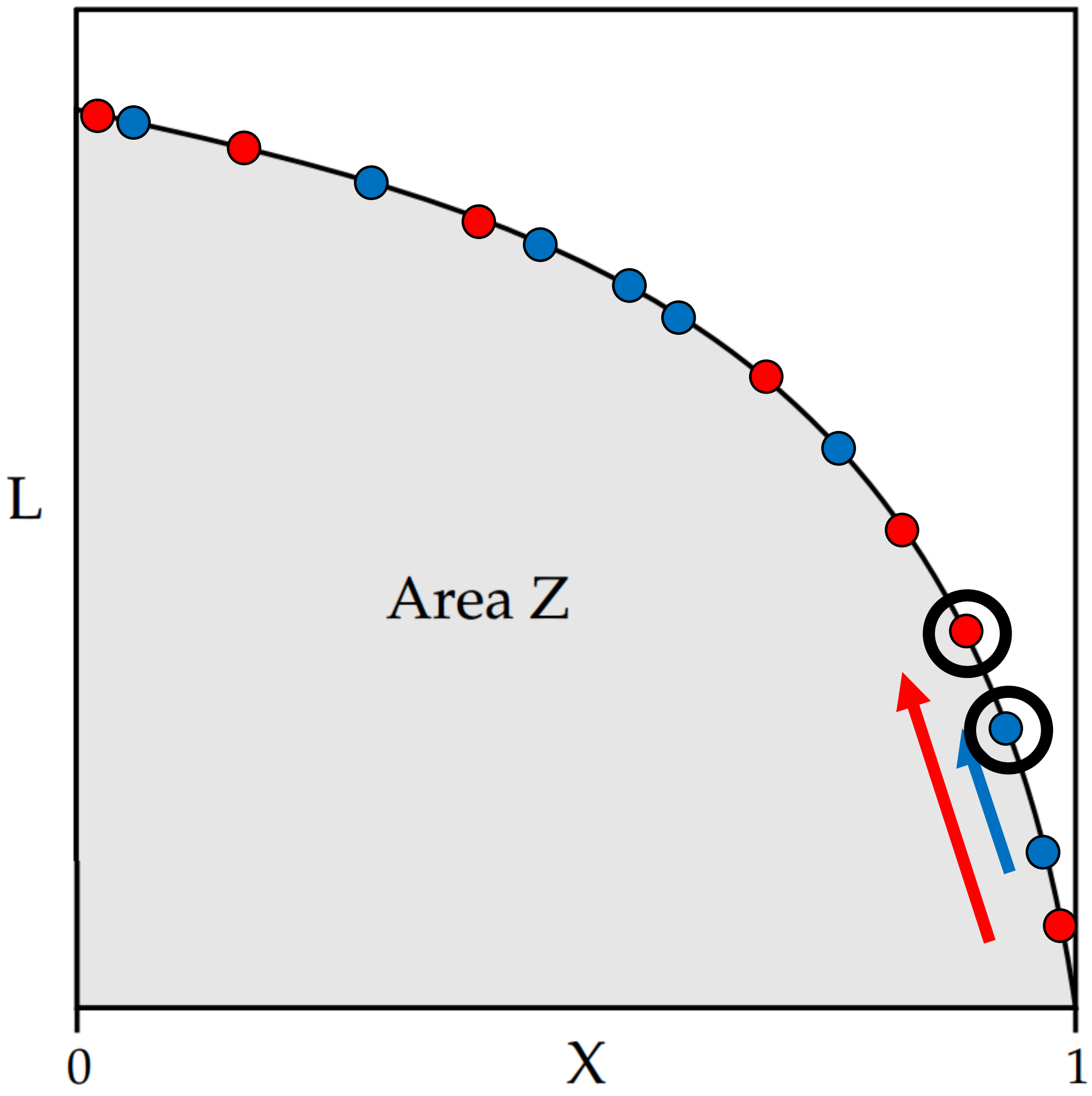
Skilling (2006)



Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

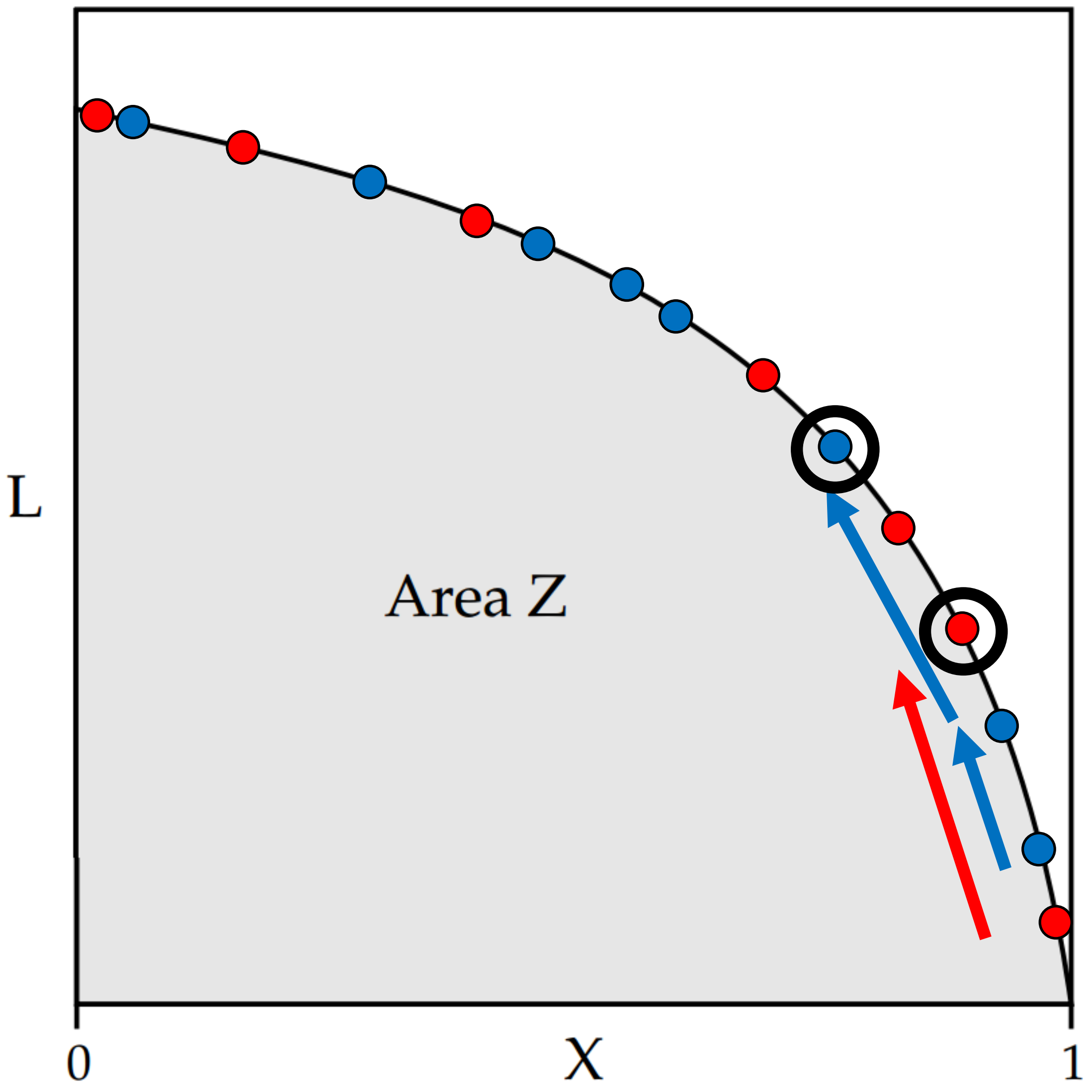
Skilling (2006)



Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

Skilling (2006)

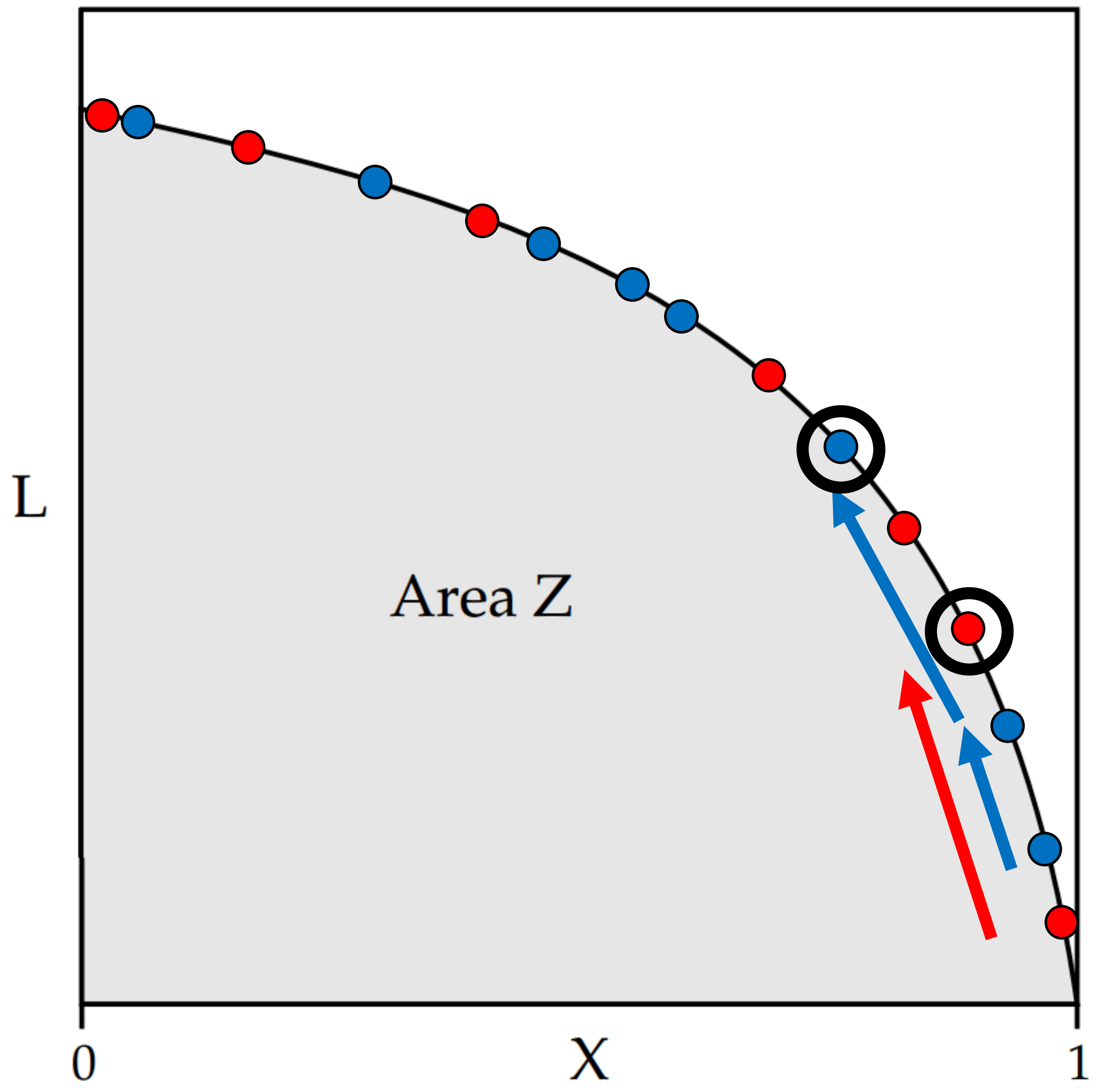


Adding More Particles

$$U_0, \ldots, U_i \sim \mathrm{Unif}$$

$$\ln X_{i+1} = \sum_{j=0}^{i} \ln U_j$$

Skilling (2006)



Adding More Particles

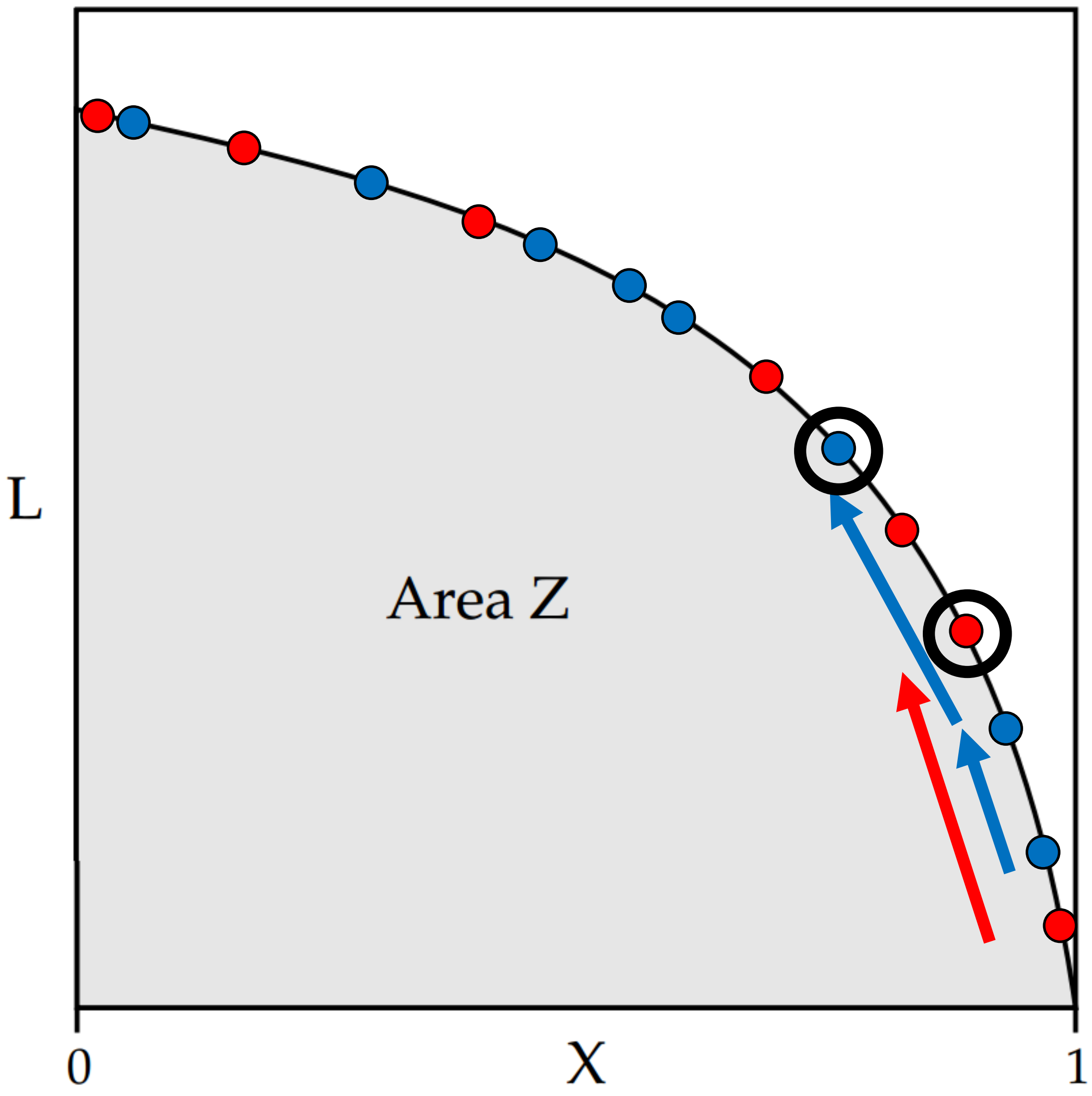
$$T_0, \ldots, T_i \sim U_{(2)}$$

Order Statistics

$$U_1, U_2 \sim \text{Unif} \Rightarrow U_{(1)}, U_{(2)}$$

$$\ln X_{i+1} = \sum_{j=0}^{i} \ln T_j$$

Skilling (2006)



Adding More Particles

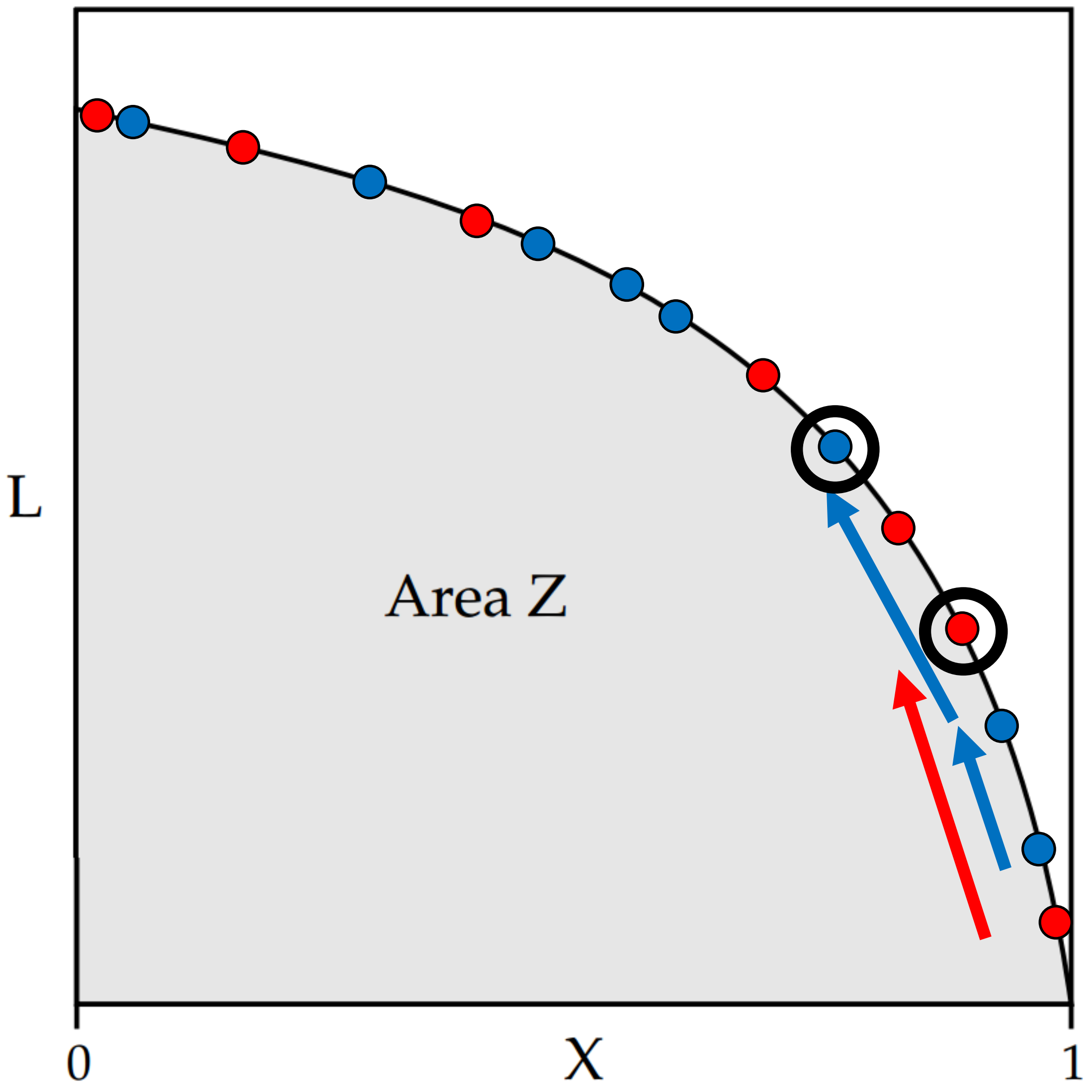
$$T_0, \ldots, T_i \sim \mathrm{Beta}(2,1)$$

Order Statistics

$$U_1, U_2 \sim \mathrm{Unif} \Rightarrow U_{(1)}, U_{(2)}$$

$$\ln X_{i+1} = \sum_{j=0}^{i} \ln T_j$$

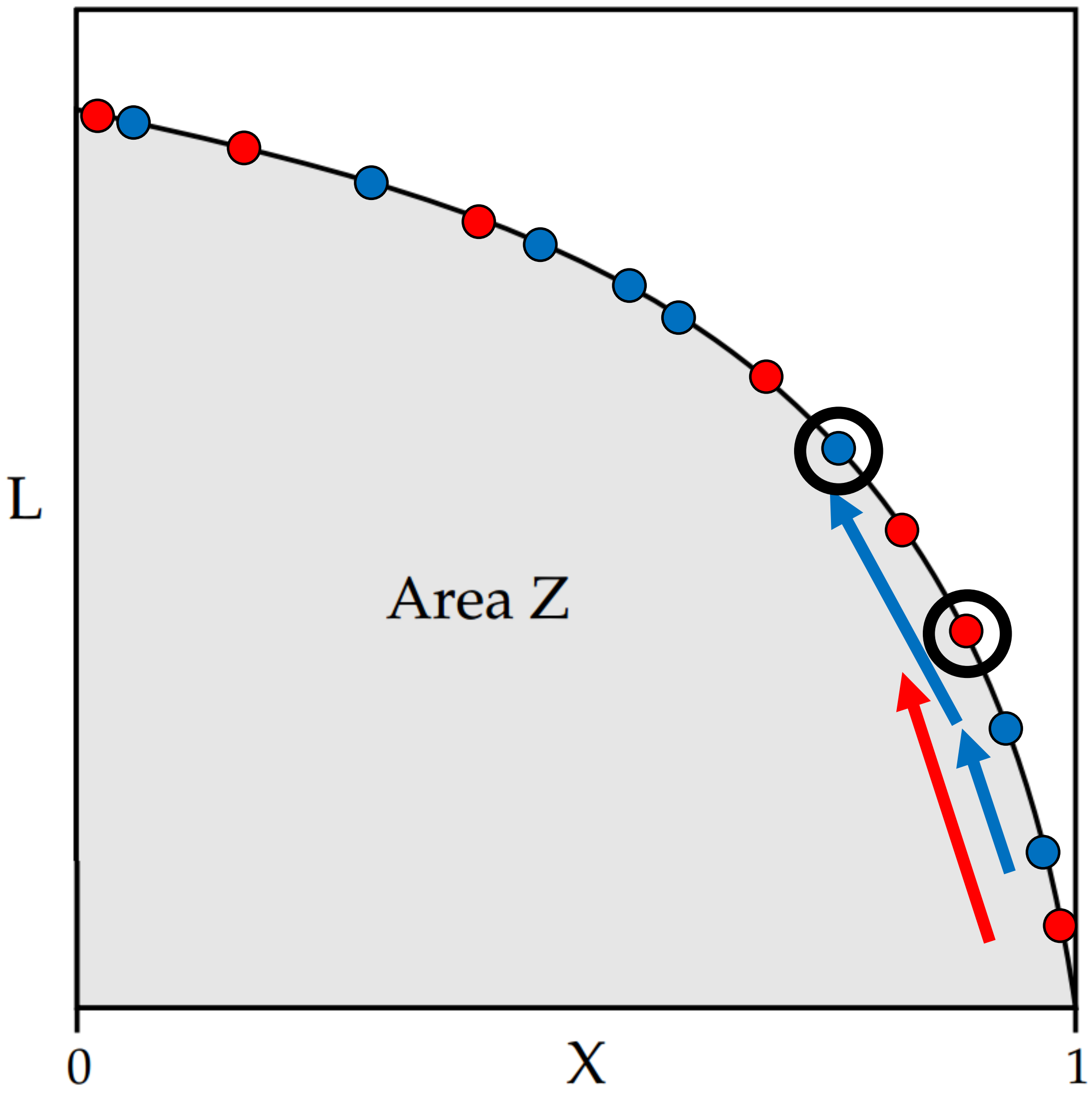
Skilling (2006)



Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

Skilling (2006)

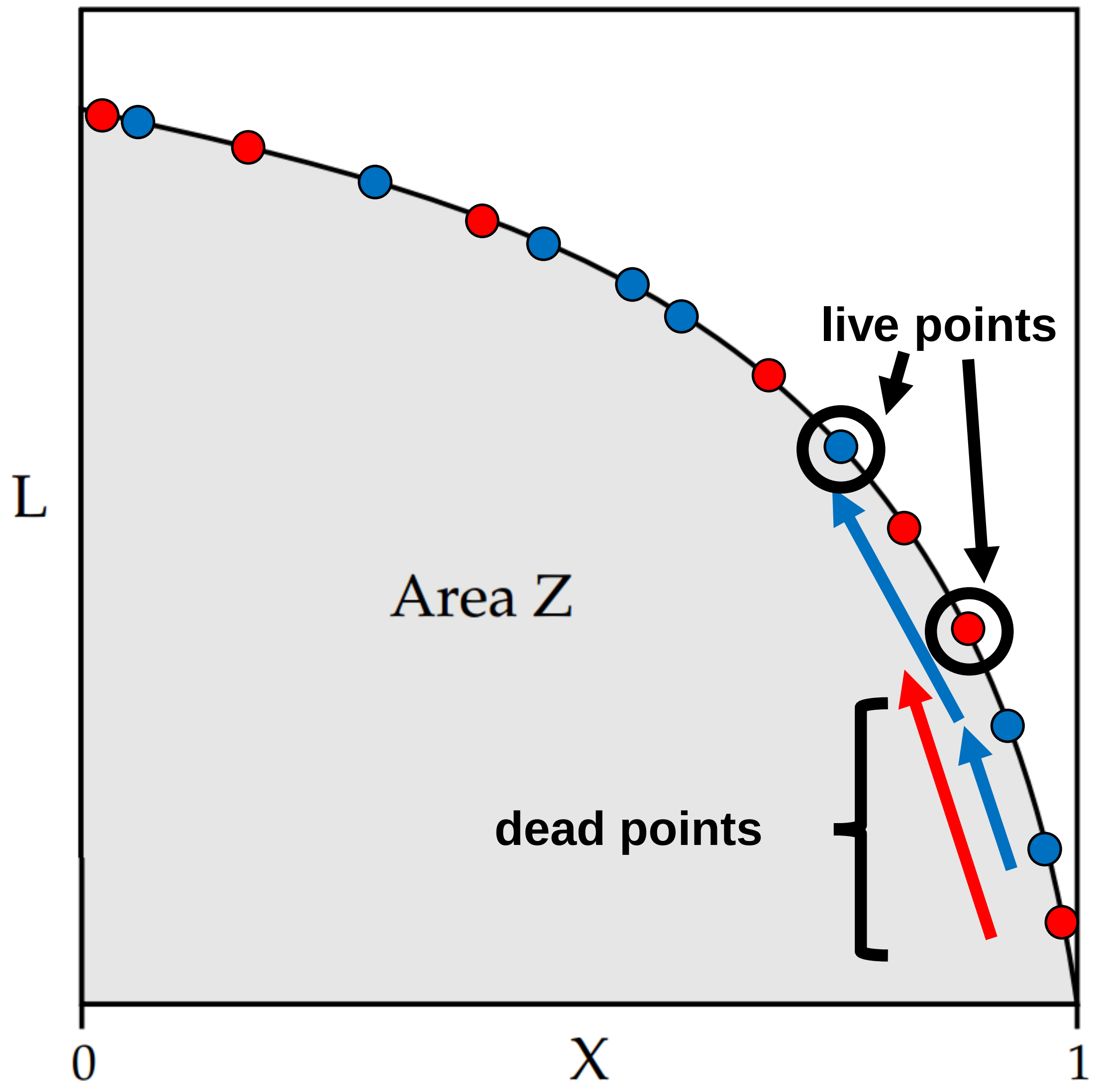


Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(1)}$$
$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > \mathcal{L}_1^{(1)} > 0$$

One run with 2 “live points”
= 2 runs with 1 live point

Skilling (2006)

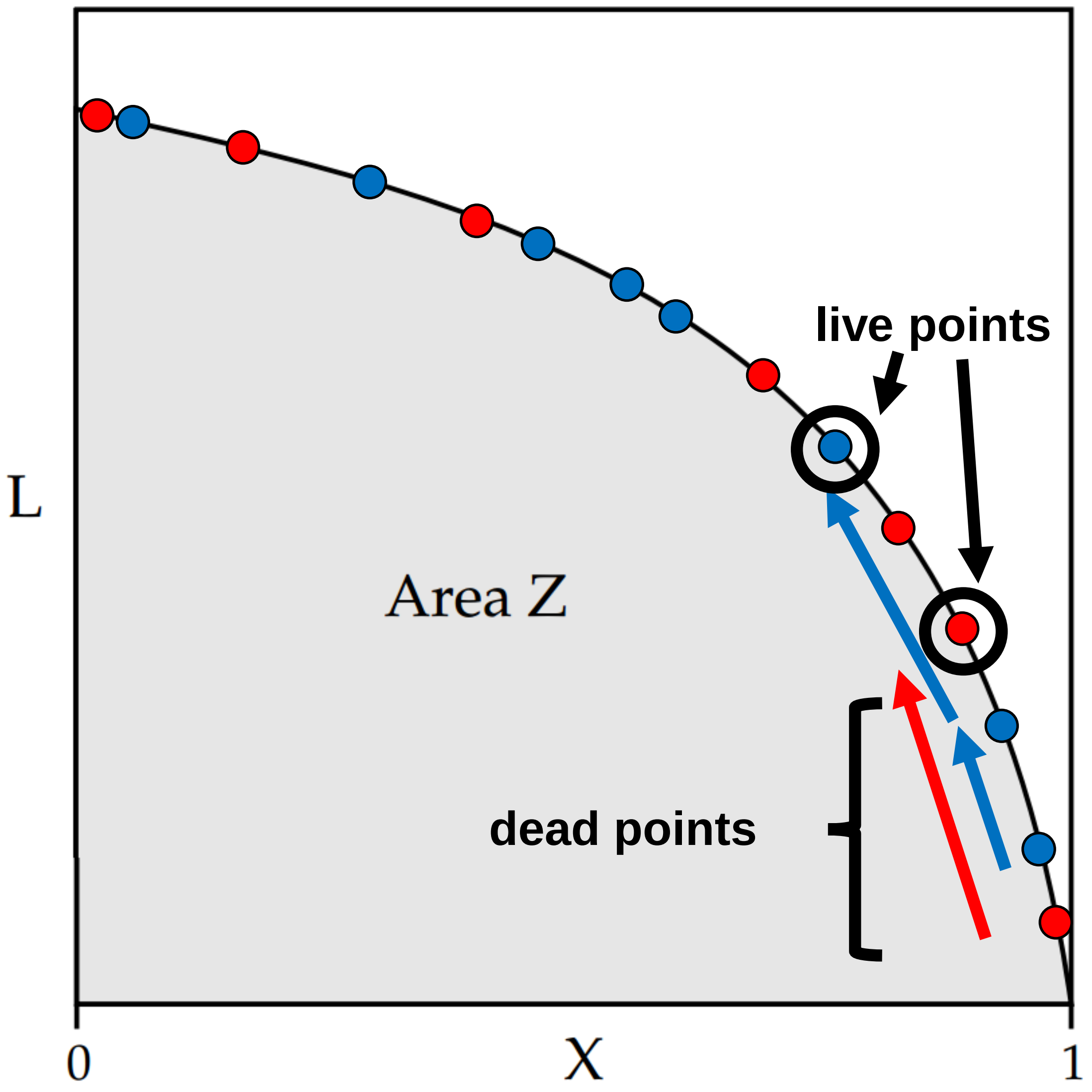


Adding More Particles

$$\mathcal{L}_{N_1}^{(1)} > \mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_{2}^{(1)}$$
$$> \mathcal{L}_{2}^{(2)} > \mathcal{L}_{1}^{(2)} > \mathcal{L}_{1}^{(1)} > 0$$

One run with K "live points"
= K runs with 1 live point

Skilling (2006)



Adding More Particles

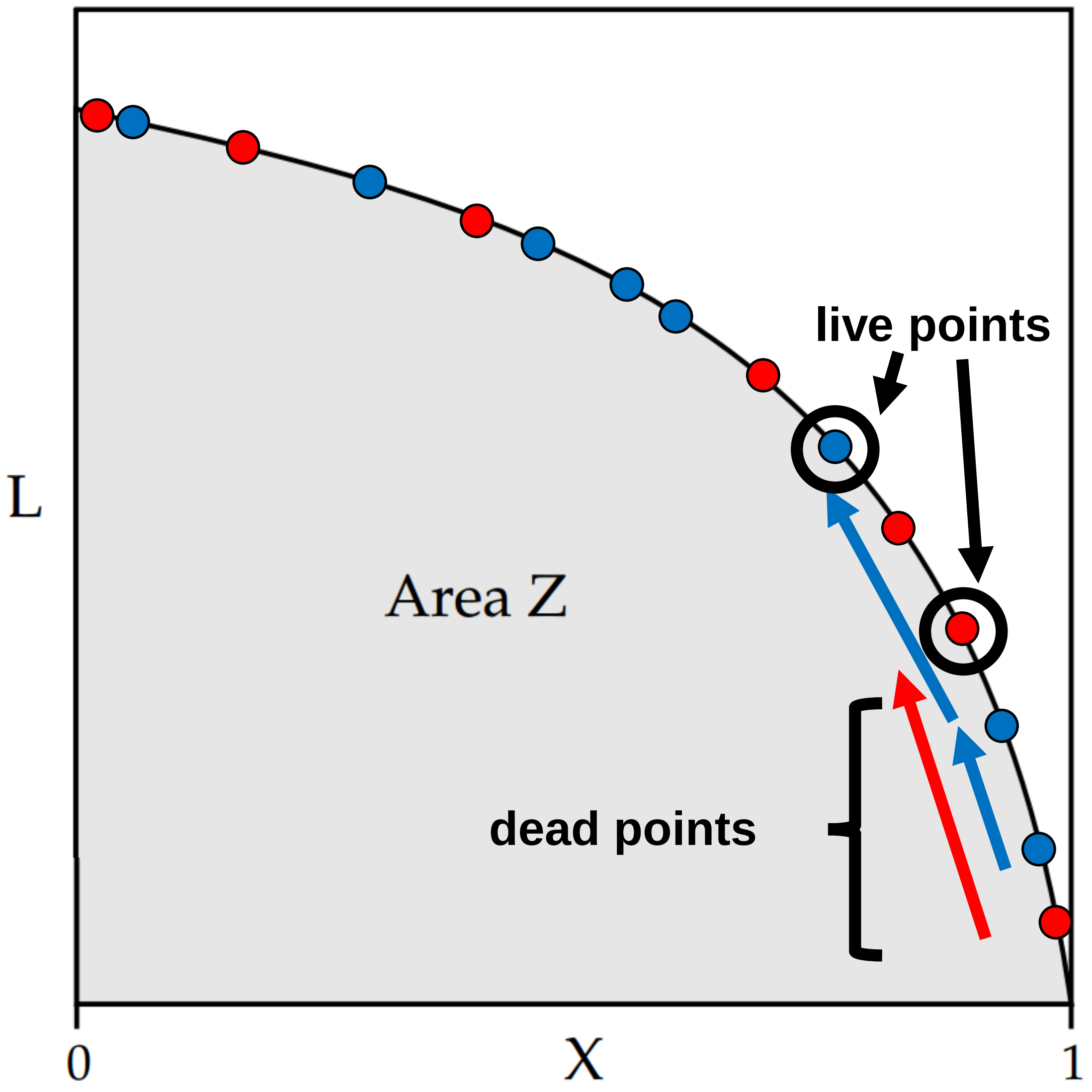
$$T_0, \ldots, T_i \sim \text{Beta}(2,1)$$

Order Statistics

$$T_1, T_2 \sim \text{Unif} \Rightarrow U_{(1)}, U_{(2)}$$

$$\ln \mathbb{E}[X_{i+1}] = \sum_{j=0}^{i} \ln \mathbb{E}[T_j]$$

Skilling (2006)



Adding More Particles

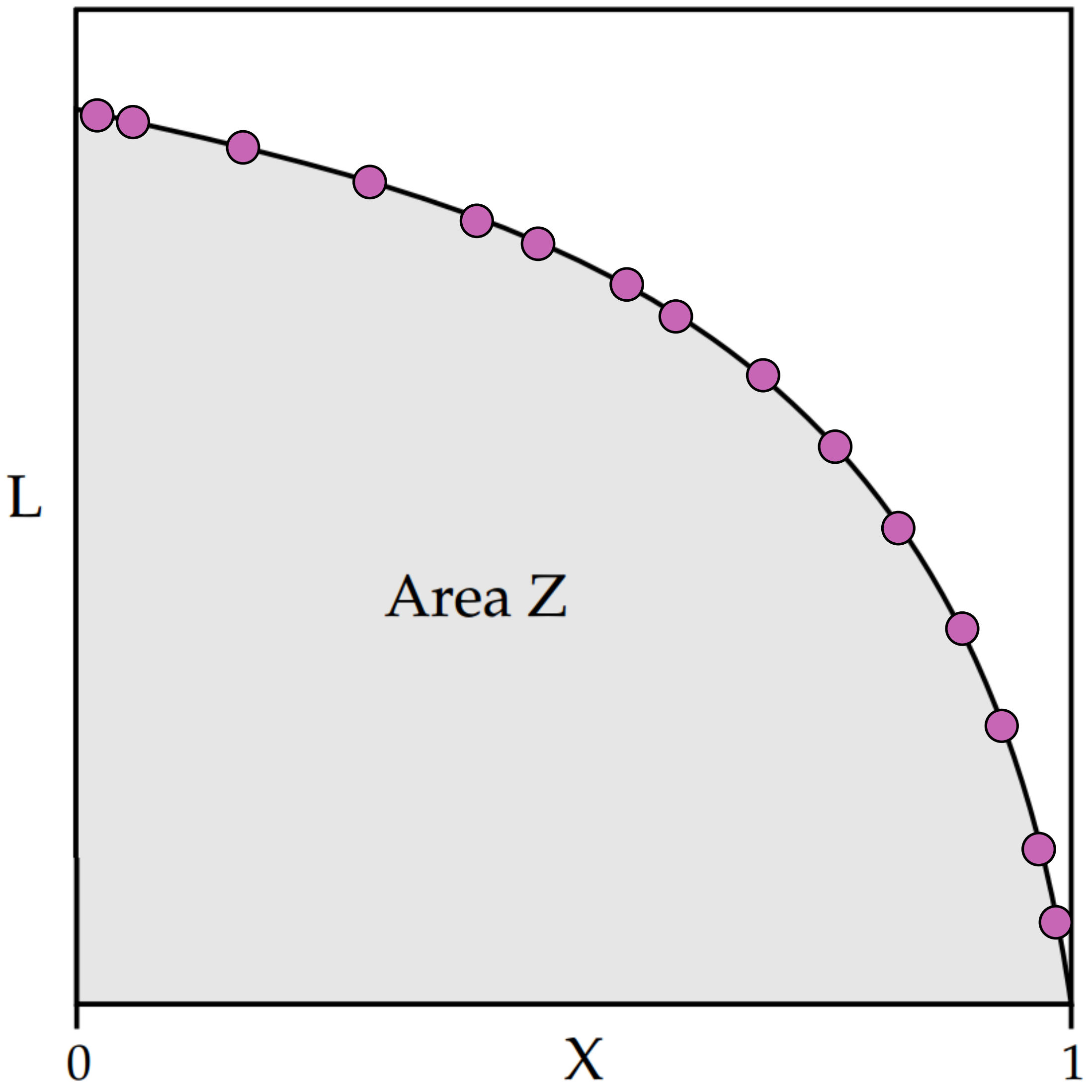
$$T_0, \dots, T_i \sim \text{Beta}(K, 1)$$

Order Statistics

$$U_1, \dots, U_K \sim \text{Unif} \Rightarrow U_{(1)}, \dots, U_{(K)}$$

$$\ln \mathbb{E}[X_{i+1}] = \sum_{j=0}^{i} \ln \mathbb{E}[T_j]$$

Skilling (2006)



Adding More Particles

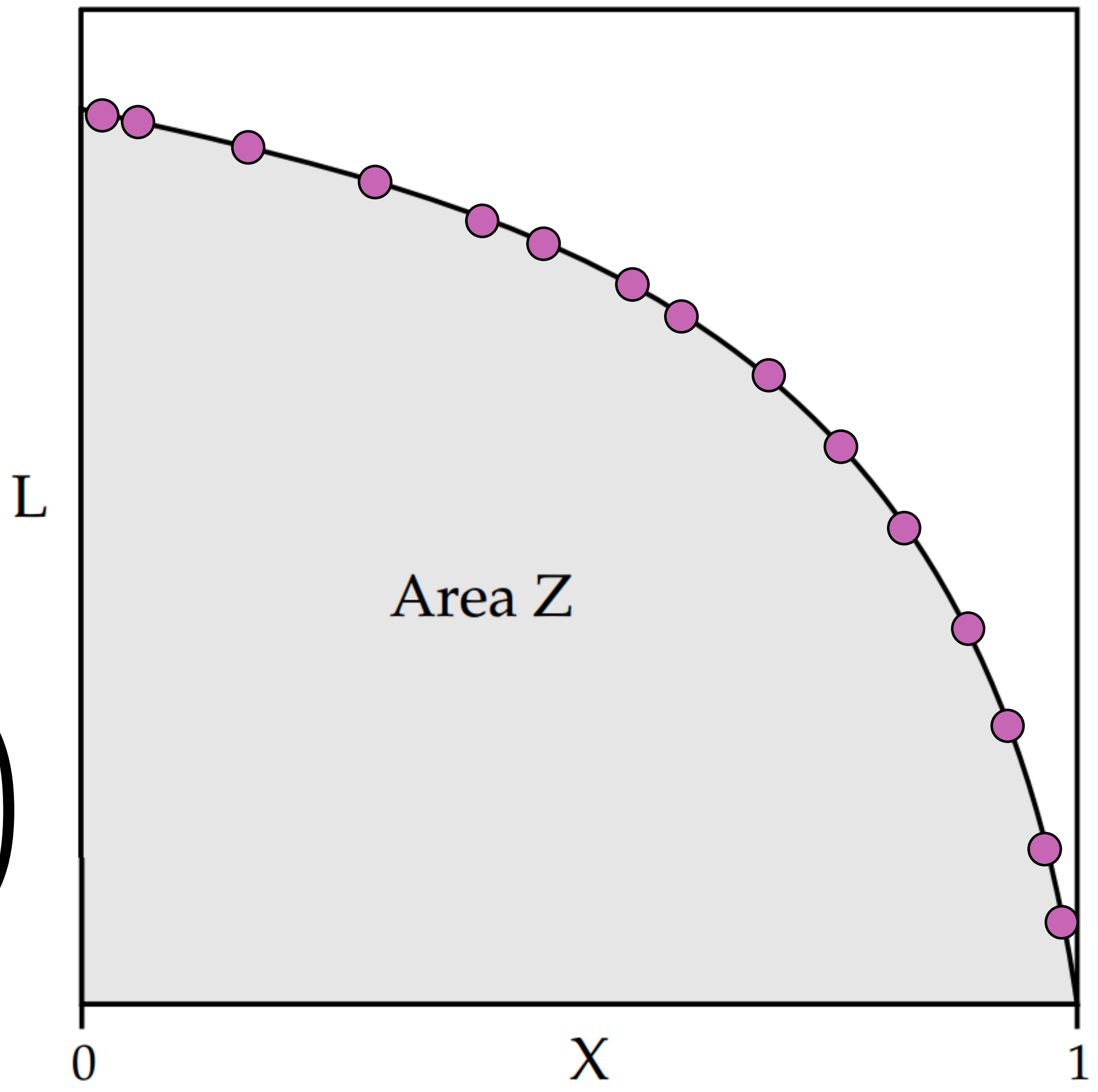
$$T_0, \ldots, T_i \sim \mathrm{Beta}(K, 1)$$

Order Statistics

$$U_1, \ldots, U_K \sim \mathrm{Unif} \Rightarrow U_{(1)}, \ldots, U_{(K)}$$

$$\ln \mathbb{E}[X_{i+1}] = \sum_{j=0}^{i} \ln\left(\frac{K}{K+1}\right)$$

Skilling (2006)



Stopping Criteria

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

Stopping Criteria

$$\frac{\hat{z}_i}{\hat{z}_i + z_{\mathrm{in}}} > f$$

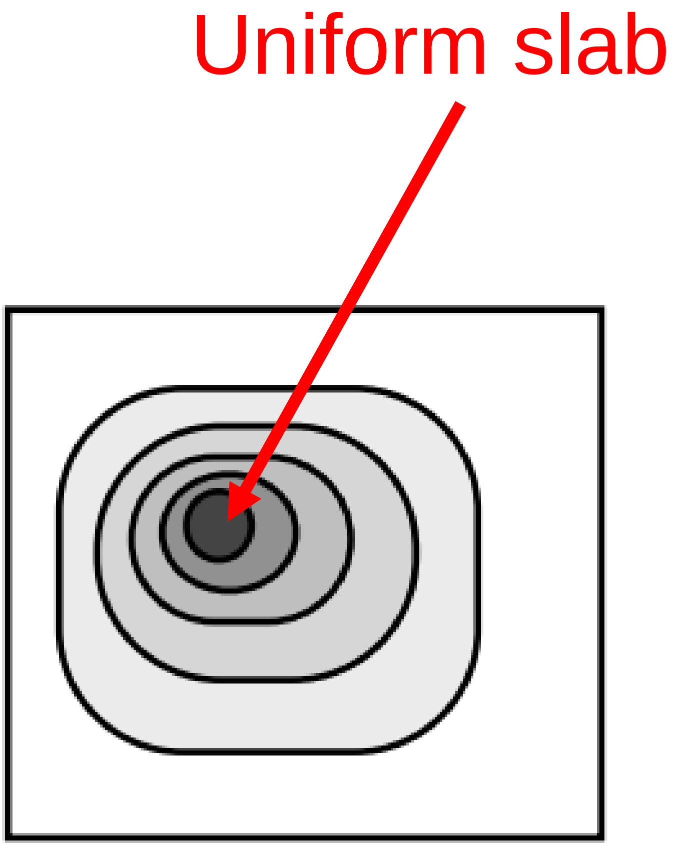
Stopping Criteria

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

Stopping Criteria

Maximum Likelihood

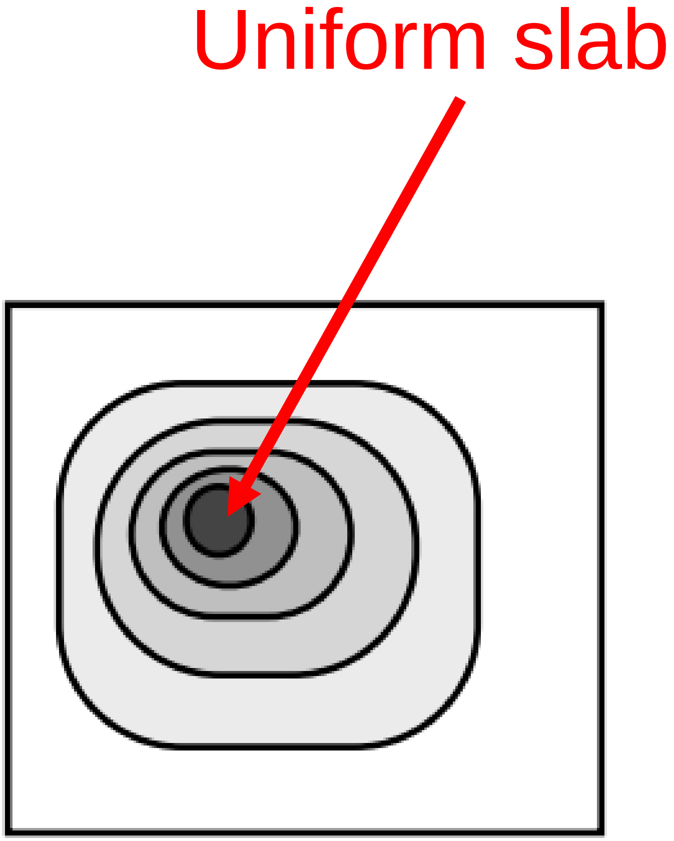
$$\hat{Z} \leq \sum_{j=1}^{i} \widehat{w}_j + \mathcal{L}_{(\mathrm{K})} X_n$$



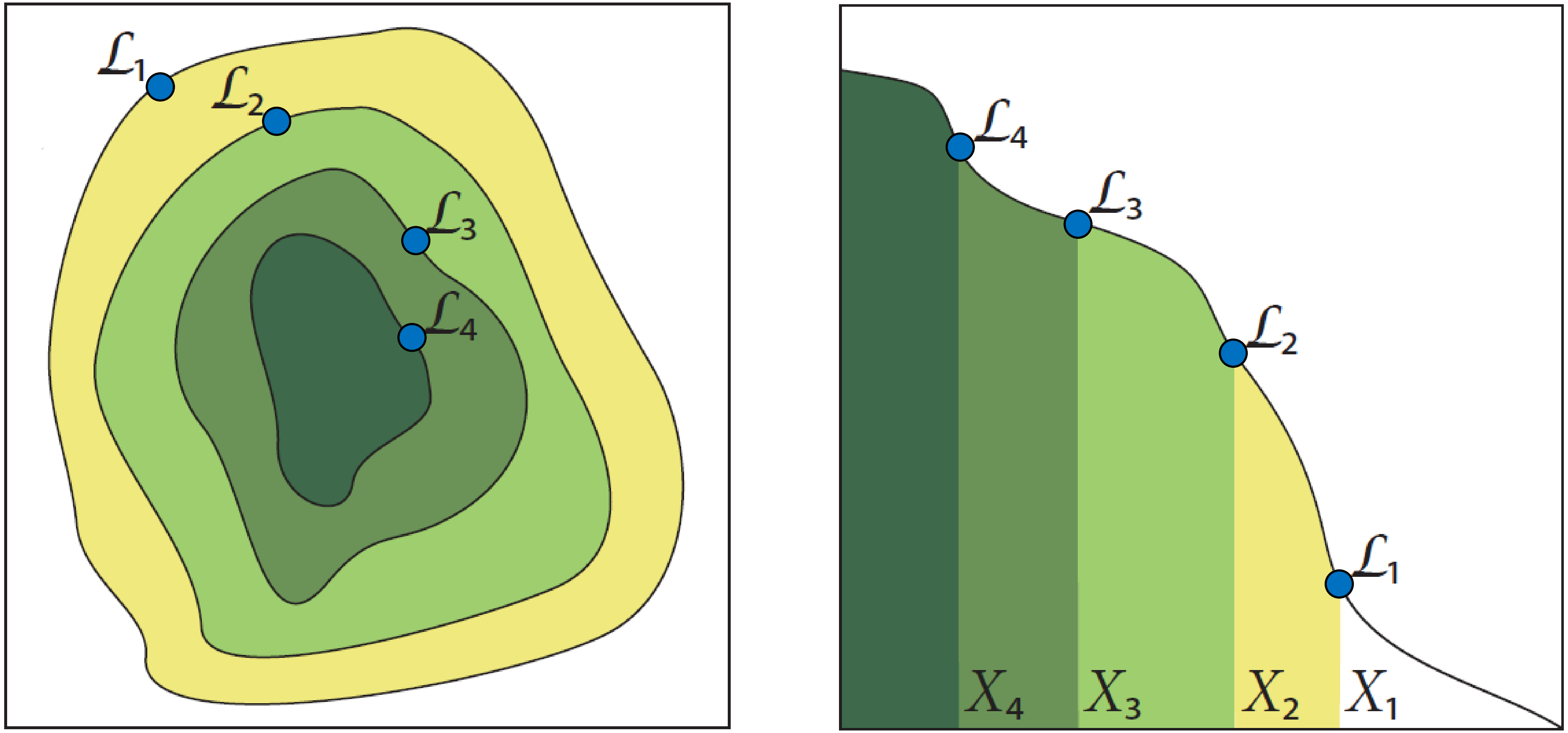
Stopping Criteria

Maximum Likelihood

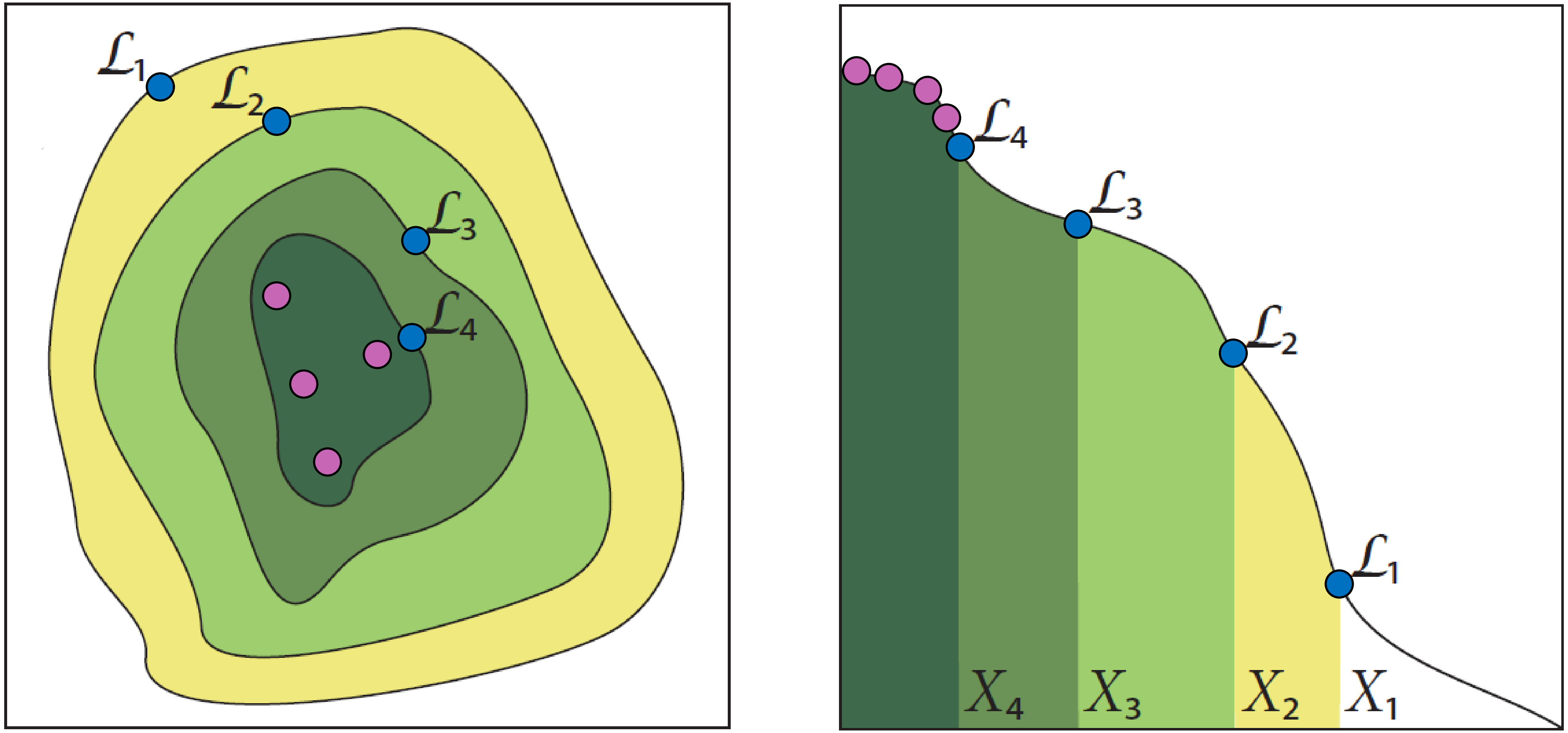
$$\hat{Z} \lesssim \sum_{j=1}^{i} \widehat{w}_j + \mathcal{L}_{(K)} \widehat{X}_n$$



"Recycling" the Final Set of Particles



“Recycling” the Final Set of Particles



“Recycling” the Final Set of Particles

$U_1, \ldots, U_K \sim \text{Unif}$

$$\ldots, \chi_{N+1}^{[k]}, \ldots \sim U_k X_N$$

“Recycling” the Final Set of Particles

$$U_1, \dots, U_K \sim \text{Unif}$$

$$\dots, \chi_{N+1}^{[k]}, \dots \sim U_k X_N$$

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

“Recycling” the Final Set of Particles

$U_1, \ldots, U_K \sim \mathrm{Unif}$

$$\ldots, \chi_{N+1}^{[k]}, \ldots \sim U_k X_N$$

$$\mathcal{L}_{N+K} > \cdots > \mathcal{L}_{N+1} > \mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

"Recycling" the Final Set of Particles

$$U_1, \ldots, U_K \sim \text{Unif}$$

$$\ldots, \chi_{N+1}^{[k]}, \ldots \sim U_k X_N$$

$$\mathcal{L}_{N+K} > \cdots > \mathcal{L}_{N+1} > \mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$X_{N+k} = \chi_{N+1}^{(K-k+1)}$$

"Recycling" the Final Set of Particles

$$U_1, \ldots, U_K \sim \text{Unif}$$

$$\ldots, X_{N+k}, \ldots \sim U_{(K-k+1)} X_N$$

$$\mathcal{L}_{N+K} > \cdots > \mathcal{L}_{N+1} > \mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$X_{N+k} = \chi_{N+1}^{(K-k+1)}$$

"Recycling" the Final Set of Particles

$$U_1, \ldots, U_K \sim \text{Unif}$$

$$\ldots, X_{N+k}, \ldots \sim U_{(K-k+1)} X_N$$

$$\mathcal{L}_{N+K} > \cdots > \mathcal{L}_{N+1} > \mathcal{L}_N > \mathcal{L}_{N-1} > \cdots > \mathcal{L}_2 > \mathcal{L}_1 > 0$$

$$X_{N+k} = \chi_{N+1}^{(K-k+1)}$$

Rényi Representation

“Recycling” the Final Set of Particles

$$\ln \mathbb{E}[X_N] = \sum_{i=1}^{N} \ln\left(\frac{K}{K+1}\right)$$

“Recycling” the Final Set of Particles

$$\ln \mathbb{E}[X_{N+k}] = \sum_{i=1}^{N} \ln\left(\frac{K}{K+1}\right) + \ln\left(\frac{K-k+1}{K+1}\right)$$

“Recycling” the Final Set of Particles

$$\ln \mathbb{E}[X_{N+k}] = \sum_{i=1}^{N} \ln\left(\frac{K}{K+1}\right) + \sum_{j=1}^{k} \ln\left(\frac{K-j+1}{K-j+2}\right)$$

"Recycling" the Final Set of Particles

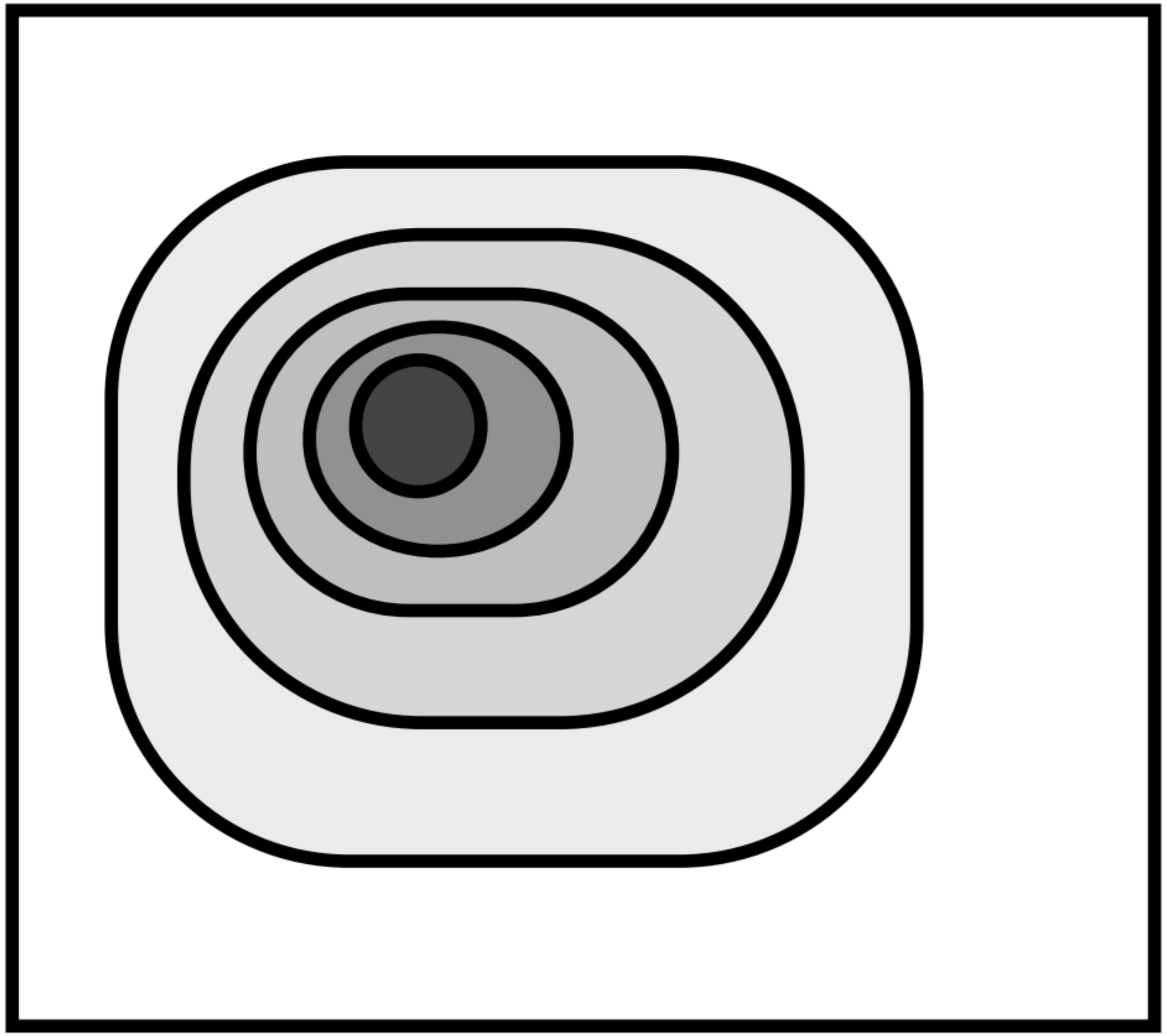
Exponential Shrinkage Uniform Shrinkage

$$\ln \mathbb{E}[X_{N+k}] = \sum_{i=1}^{N} \ln\left(\frac{K}{K+1}\right) + \sum_{j=1}^{k} \ln\left(\frac{K-j+1}{K-j+2}\right)$$

Nested Sampling Errors

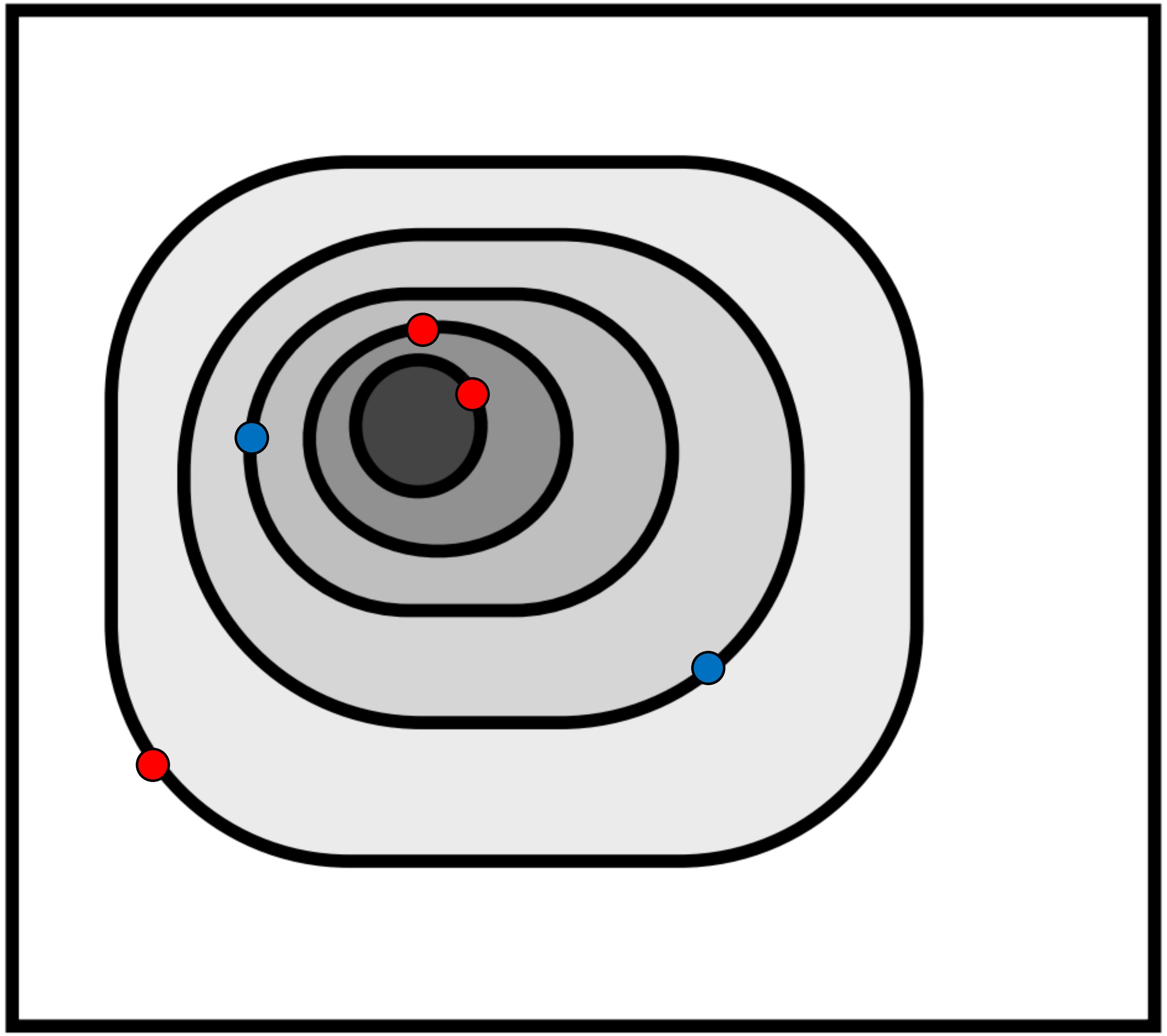
Pictures from this 2010 talk by Skilling.

Nested Sampling Uncertainties



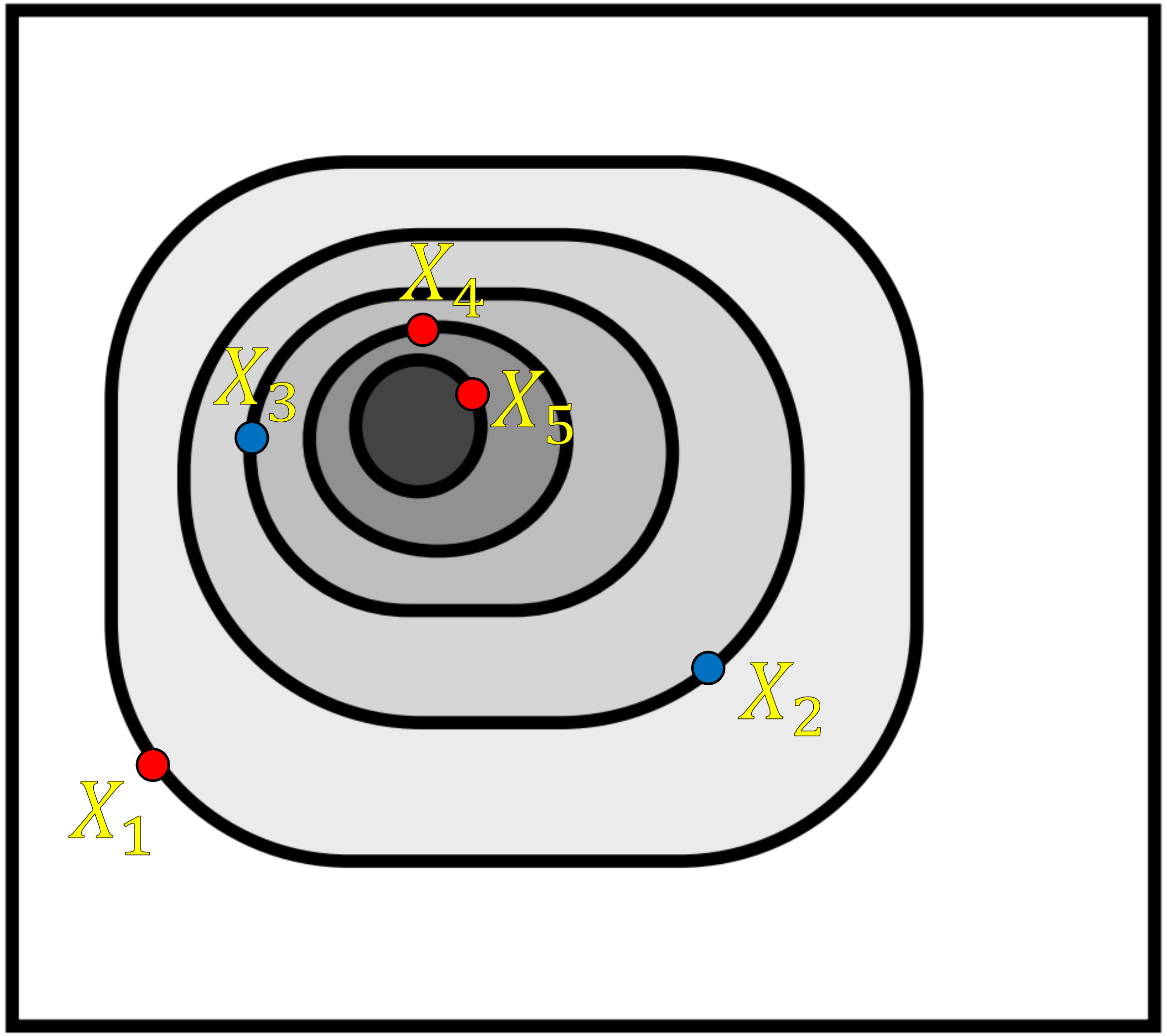
Pictures from [this 2010 talk](#) by Skilling.

Nested Sampling Uncertainties



Pictures from this 2010 talk by Skilling.

Nested Sampling Uncertainties



Statistical uncertainties

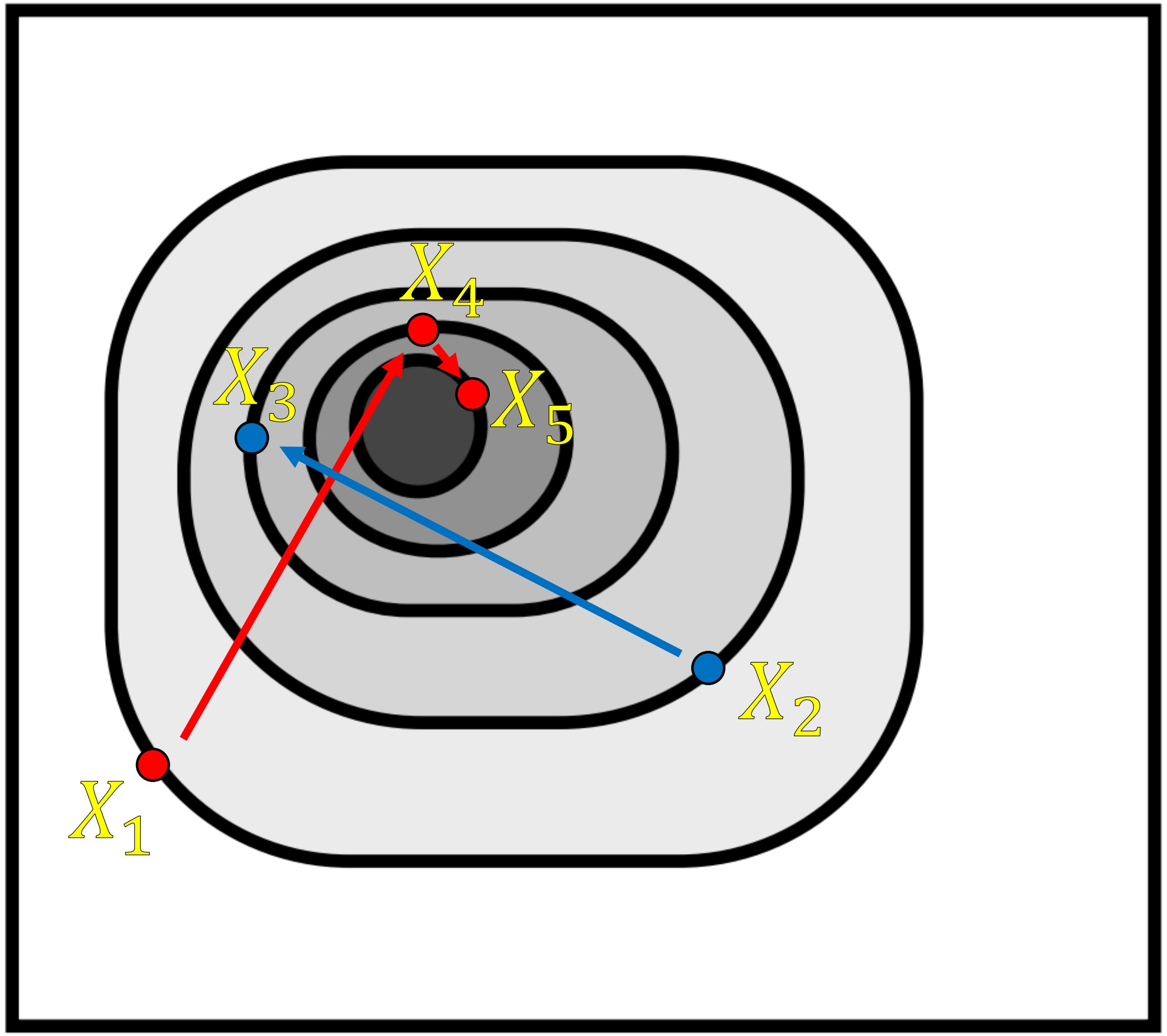
- Unknown prior volumes

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

Nested Sampling Uncertainties

Sampling uncertainties

- Number of samples (counting)
- Discrete point estimates for contours
- Particle path dependencies



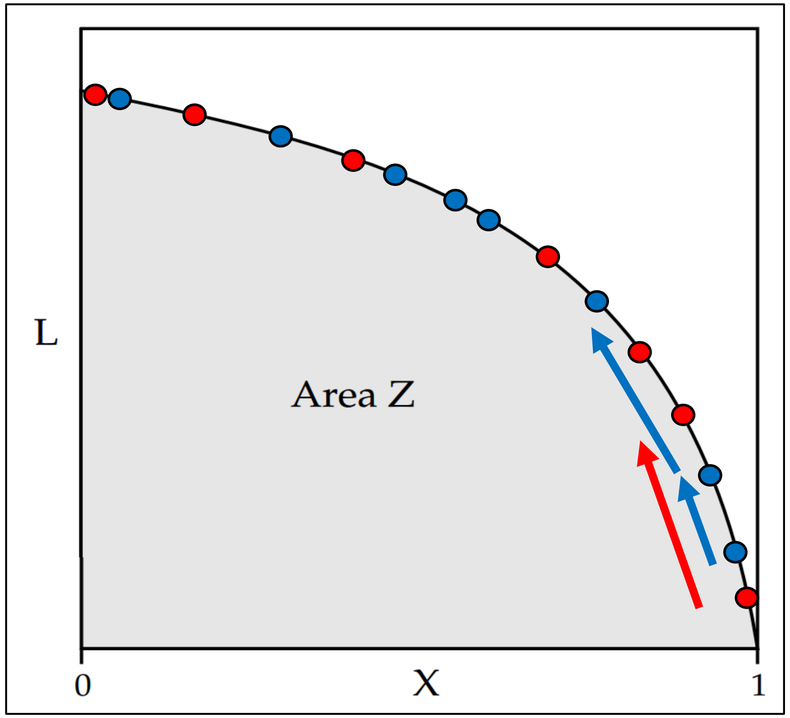
Statistical uncertainties

- Unknown prior volumes

Sampling Error: Poisson Uncertainties

"Distance" from prior to posterior

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{???}{\mathbb{E}[\Delta \ln X]}$$



Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

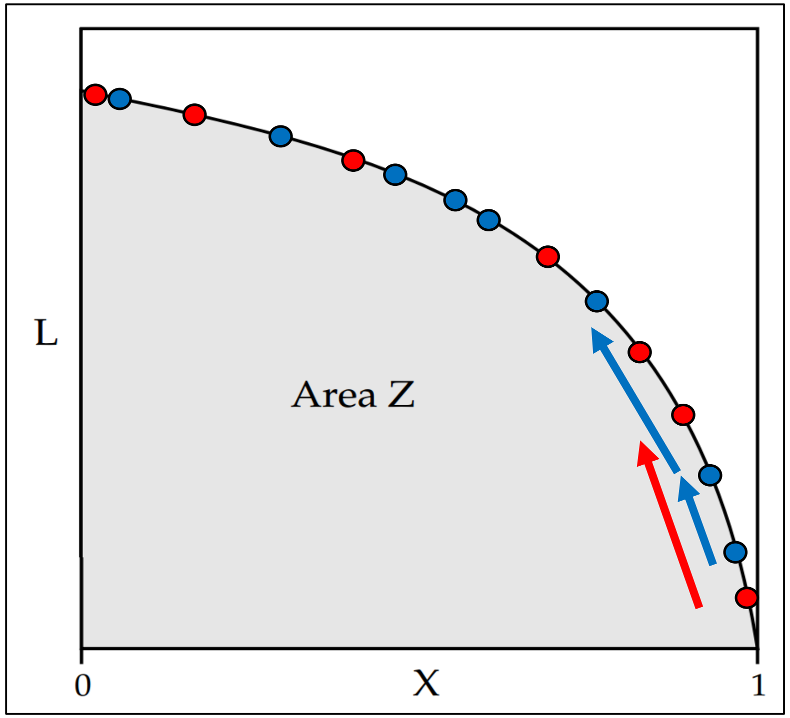
Kullback-Leibler divergence from π to p → "information gained".

$$H(\pi|p) \equiv \int_{\Omega_{\mathbf{\Theta}}} p(\mathbf{\Theta}) \ln \frac{p(\mathbf{\Theta})}{\pi(\mathbf{\Theta})} d\mathbf{\Theta}$$

Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{H(\pi|p)}{\mathbb{E}[\Delta \ln X]}$$



Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{H(\pi|p)}{\mathbb{E}[\Delta \ln X]}$$

$$\mathbb{V}\left[\ln \hat{\mathcal{Z}}\right]$$

Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{H(\pi|p)}{\mathbb{E}[\Delta \ln X]}$$

$$\mathbb{V}\left[\ln \hat{\mathcal{Z}}\right] \sim \mathbb{V}\left[\ln \hat{X}_H\right]$$

Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{H(\pi|p)}{\mathbb{E}[\Delta \ln X]}$$

$$\mathbb{V}\left[\ln \hat{\mathcal{Z}}\right] \sim \mathbb{V}\left[\ln \hat{X}_H\right] \sim (\sigma[N]\mathbb{E}[\Delta \ln X])^2$$

Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{H(\pi|p)}{\mathbb{E}[\Delta \ln X]}$$

$$\mathbb{V}\left[\ln \hat{\mathcal{Z}}\right] \sim \mathbb{V}\left[\ln \hat{X}_H\right] \sim (\sigma[N]\mathbb{E}[\Delta \ln X])^2$$
$$\sim H(\pi|p)\mathbb{E}[\Delta \ln X]$$

Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Poisson Uncertainties

$$\mathbb{V}[N] = \mathbb{E}[N] \sim \frac{H(\pi|p)}{\mathbb{E}[\Delta \ln X]}$$

$$\mathbb{V}\left[\ln \hat{\mathcal{Z}}\right] \sim \mathbb{V}\left[\ln \hat{X}_H\right] \sim (\sigma[N]\mathbb{E}[\Delta \ln X])^2$$

$$\sim \sum_{i=1}^{N+K} \Delta H_i \mathbb{E}[\Delta \ln X_i]$$

$$= 1/K$$

Based on Skilling (2006)
and Keeton (2011)

Sampling Error: Monte Carlo Noise

$$\mathbb{E}_p[f(\boldsymbol{\Theta})] = \int_{\Omega_{\boldsymbol{\Theta}}} f(\boldsymbol{\Theta}) p(\boldsymbol{\Theta}) d\boldsymbol{\Theta}$$

Formalism following Higson et al. (2017) and Chopin and Robert (2010)

Sampling Error: Monte Carlo Noise

$$\mathbb{E}_p[f(\boldsymbol{\Theta})] = \int_{\Omega_{\boldsymbol{\Theta}}} f(\boldsymbol{\Theta}) p(\boldsymbol{\Theta}) d\boldsymbol{\Theta} = \frac{1}{\mathcal{Z}} \int_0^1 \tilde{f}(\mathrm{X}) \mathcal{L}(X) dX$$

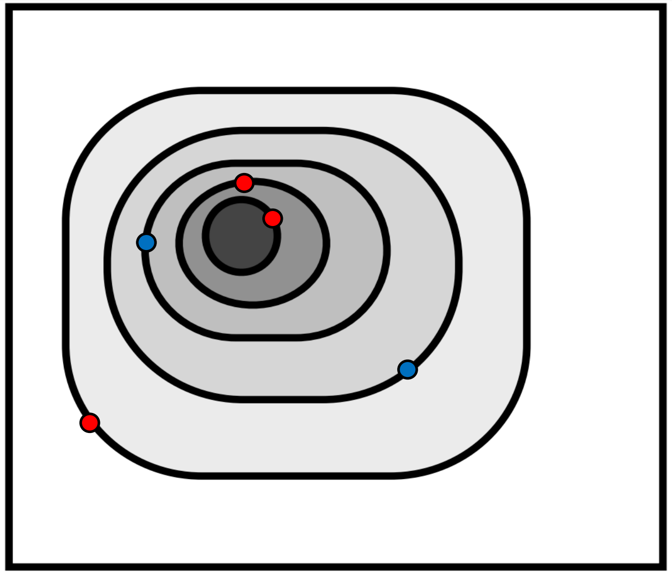
$$\tilde{f}(\mathrm{X}) = \mathbb{E}_\pi[f(\boldsymbol{\Theta}) | \mathcal{L}(\boldsymbol{\Theta}) = \mathcal{L}(X)]$$

Formalism following Higson et al. (2017) and Chopin and Robert (2010)

Sampling Error: Monte Carlo Noise

$$\mathbb{E}_p[f(\boldsymbol{\Theta})] = \int_{\Omega_{\boldsymbol{\Theta}}} f(\boldsymbol{\Theta}) p(\boldsymbol{\Theta}) d\boldsymbol{\Theta} = \frac{1}{\mathcal{Z}} \int_0^1 \tilde{f}(\mathrm{X}) \mathcal{L}(X) dX$$

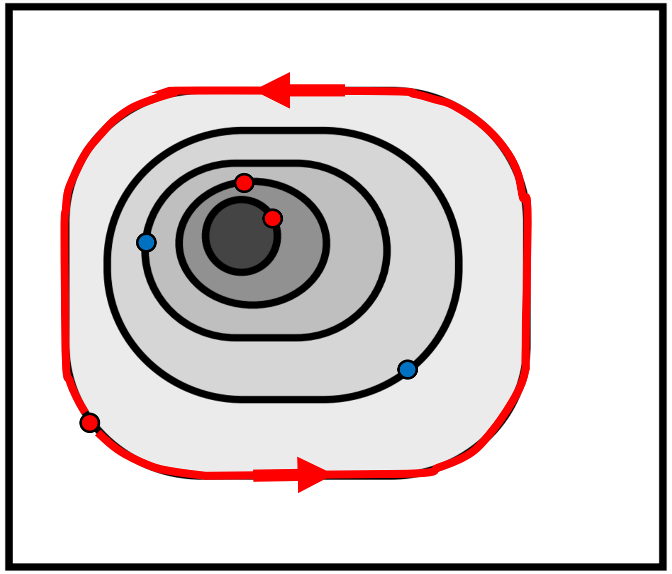
$$\tilde{f}(\mathrm{X}) = \mathbb{E}_\pi[f(\boldsymbol{\Theta}) | \mathcal{L}(\boldsymbol{\Theta}) = \mathcal{L}(X)]$$



Formalism following Higson et al. (2017) and Chopin and Robert (2010)

Sampling Error: Monte Carlo Noise

$$\mathbb{E}_p[f(\boldsymbol{\Theta})] = \int_{\Omega_{\boldsymbol{\Theta}}} f(\boldsymbol{\Theta}) p(\boldsymbol{\Theta}) d\boldsymbol{\Theta} = \frac{1}{Z} \int_0^1 \tilde{f}(\mathrm{X}) \mathcal{L}(X) dX$$



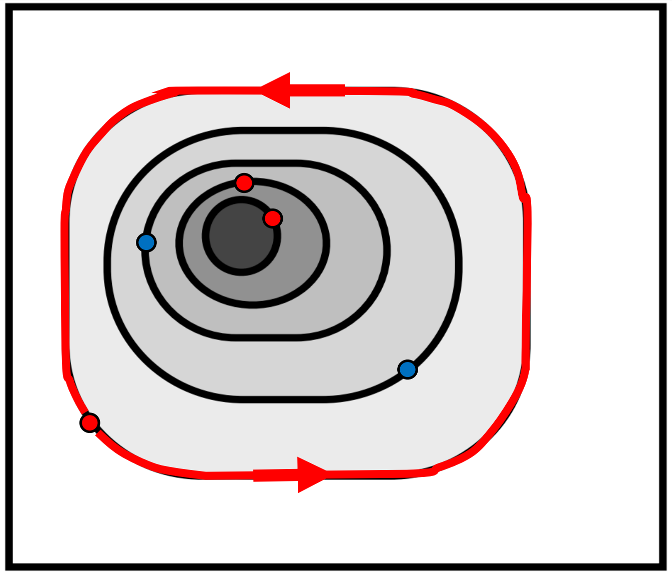
$$\tilde{f}(\mathrm{X}) = \mathbb{E}_\pi[f(\boldsymbol{\Theta}) | \mathcal{L}(\boldsymbol{\Theta}) = \mathcal{L}(X)]$$

Formalism following Higson et al. (2017) and Chopin and Robert (2010)

Sampling Error: Monte Carlo Noise

$$\mathbb{E}_p[f(\boldsymbol{\Theta})] = \int_{\Omega_{\boldsymbol{\Theta}}} f(\boldsymbol{\Theta})p(\boldsymbol{\Theta})d\boldsymbol{\Theta} = \frac{1}{\mathcal{Z}}\int_0^1 \tilde{f}(\mathrm{X})\mathcal{L}(X)dX$$

$$\approx \sum_{i=1}^{N+K} \hat{p}_i\, f(\boldsymbol{\Theta}_i)$$

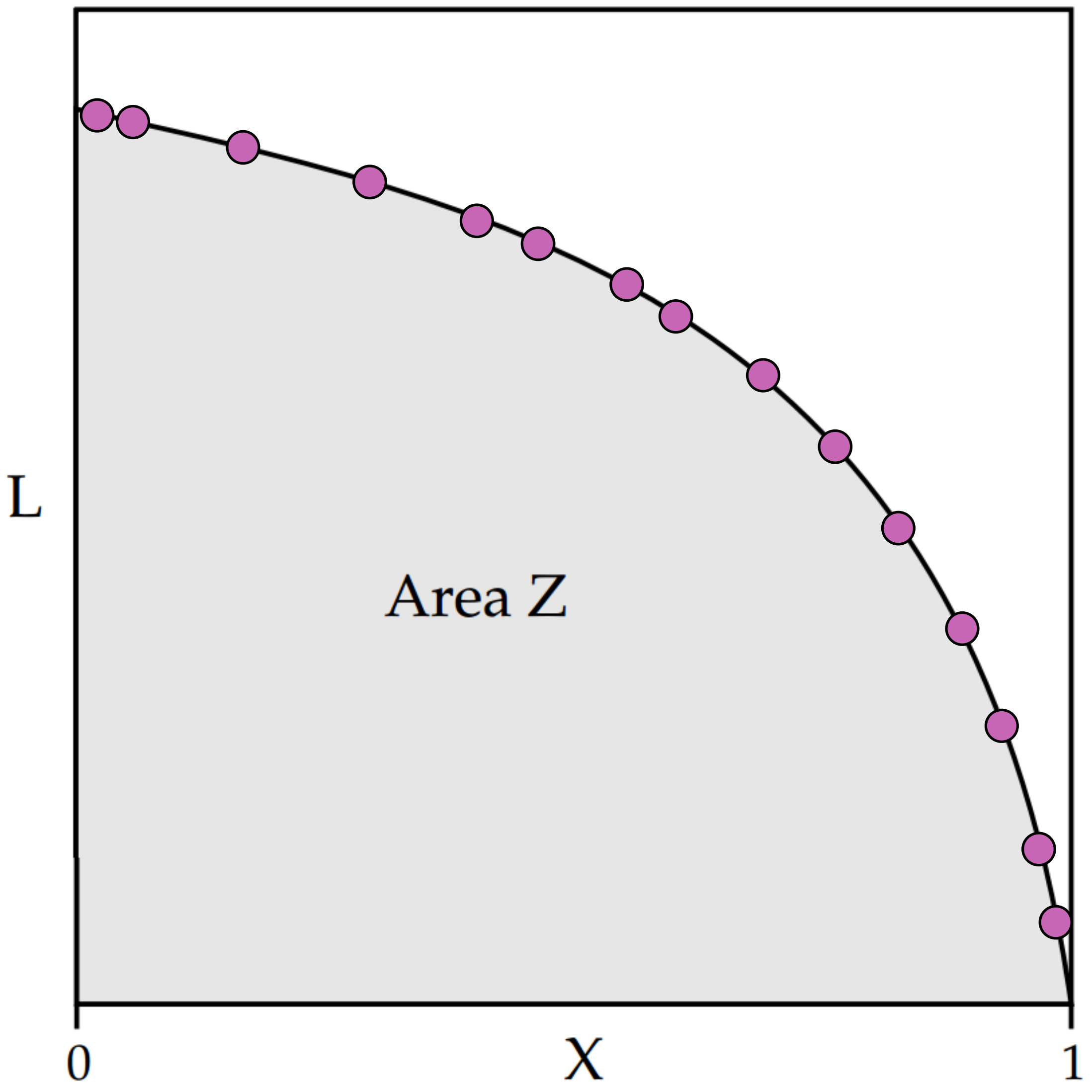


Formalism following Higson et al. (2017) and Chopin and Robert (2010)

Exploring Sampling Uncertainties

$$\mathcal{L}_N > \mathcal{L}_{N-1} > \cdots$$
$$> \mathcal{L}_2 > \mathcal{L}_1 > 0$$

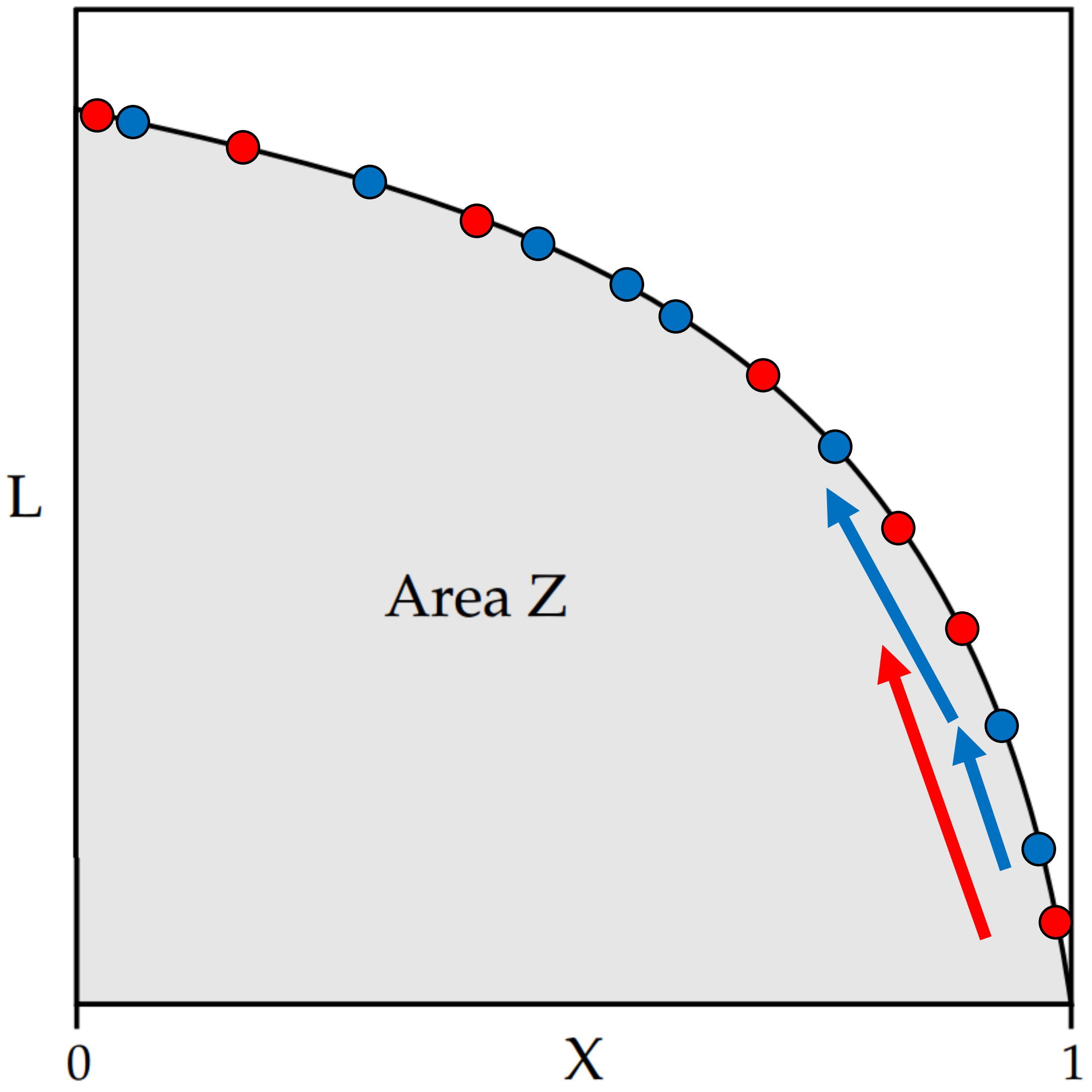
One run with **K** "live points"
= **K** runs with 1 live point!



Exploring Sampling Uncertainties

$$\mathcal{L}_{N_1}^{(1)} > \cdots > \mathcal{L}_2^{(1)} > \mathcal{L}_1^{(1)} > 0$$

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > 0$$



One run with K "live points"
= K runs with 1 live point!

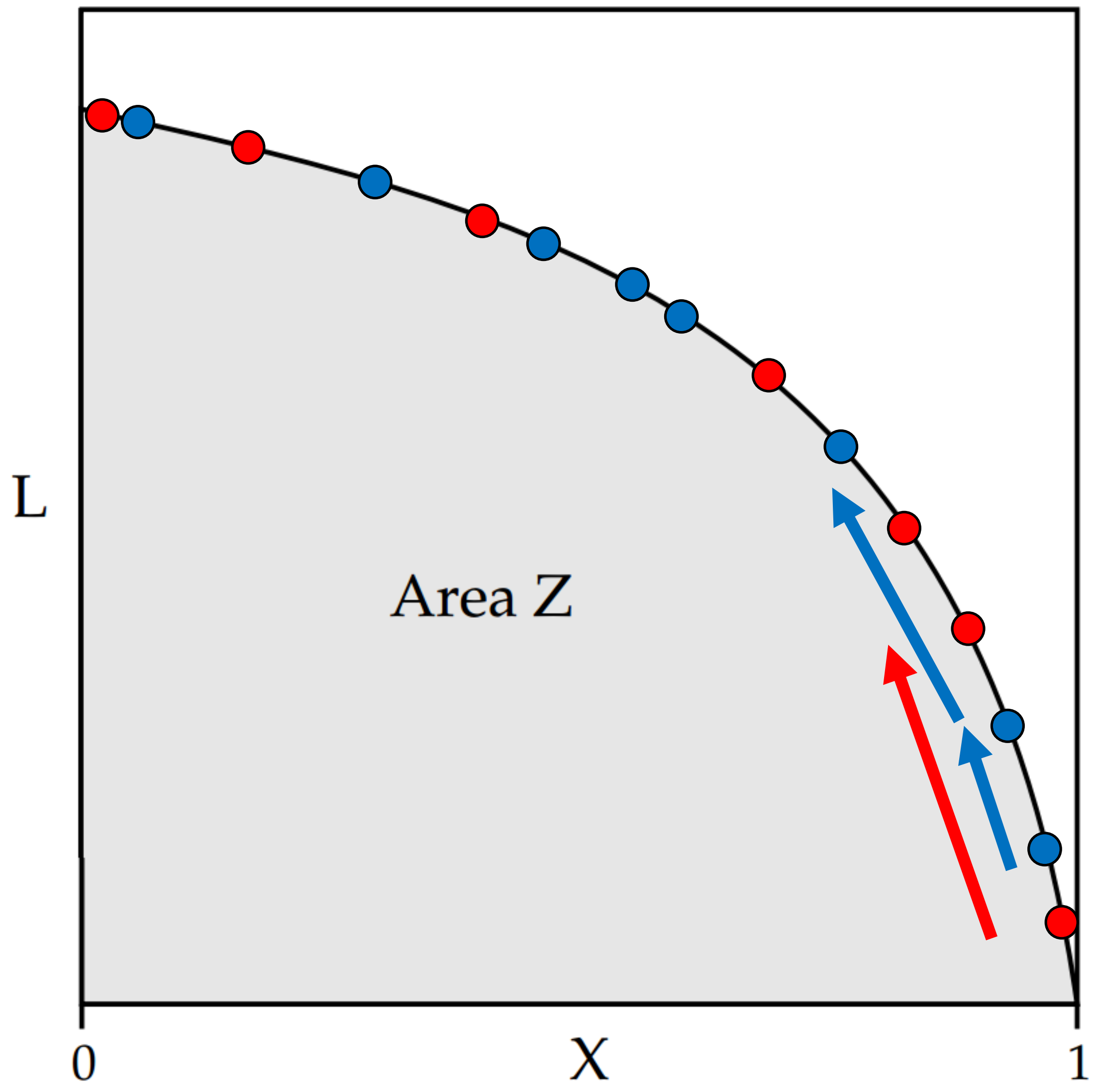
Exploring Sampling Uncertainties

Original run

"strand"

$$\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\} = \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(2)}\right\}, \ldots$$

"strand"



One run with K "live points"

= K runs with 1 live point!

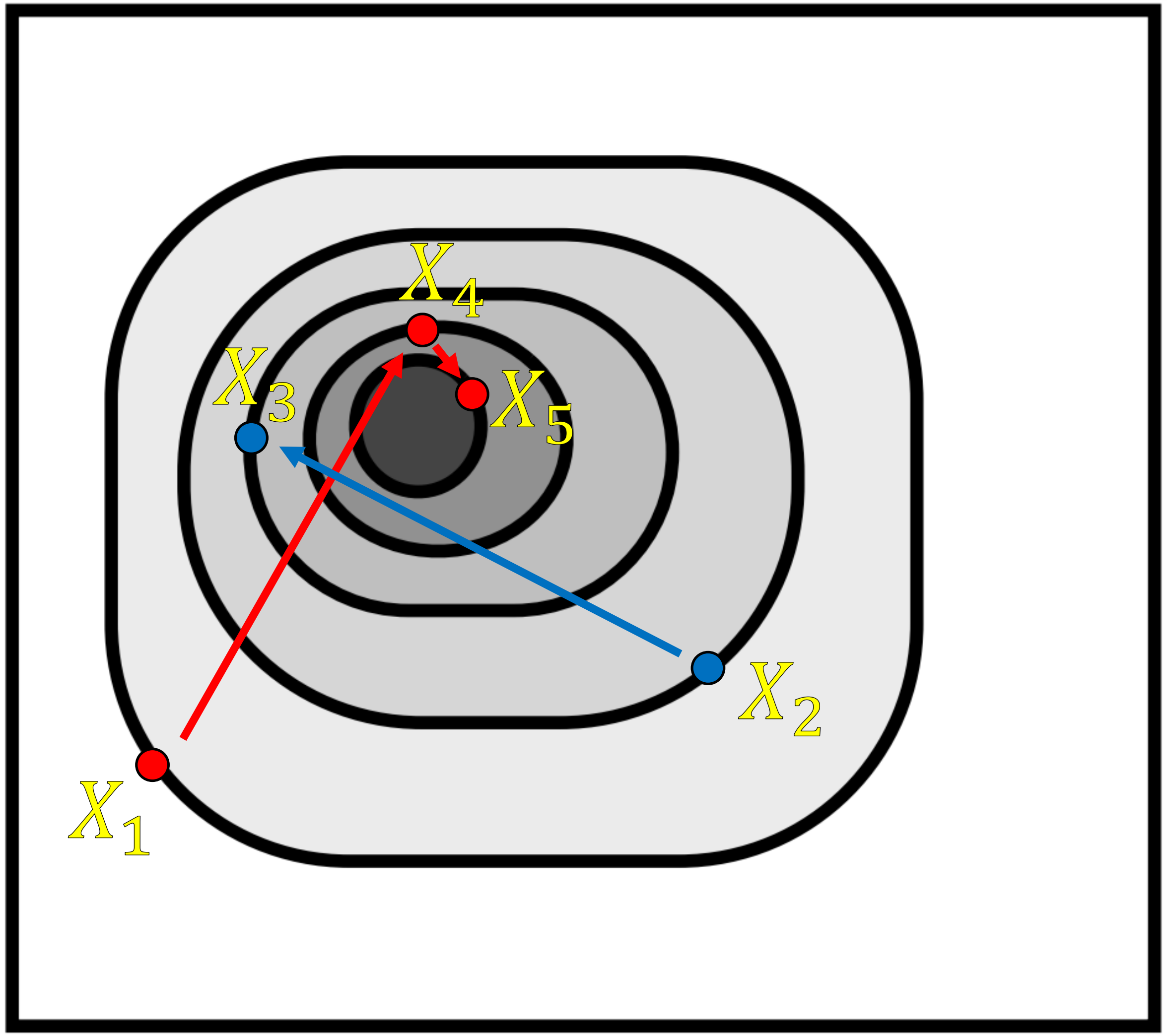
Exploring Sampling Uncertainties

Original run

"strand"

$$\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\} = \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(2)}\right\}, \ldots$$

"strand"



Exploring Sampling Uncertainties

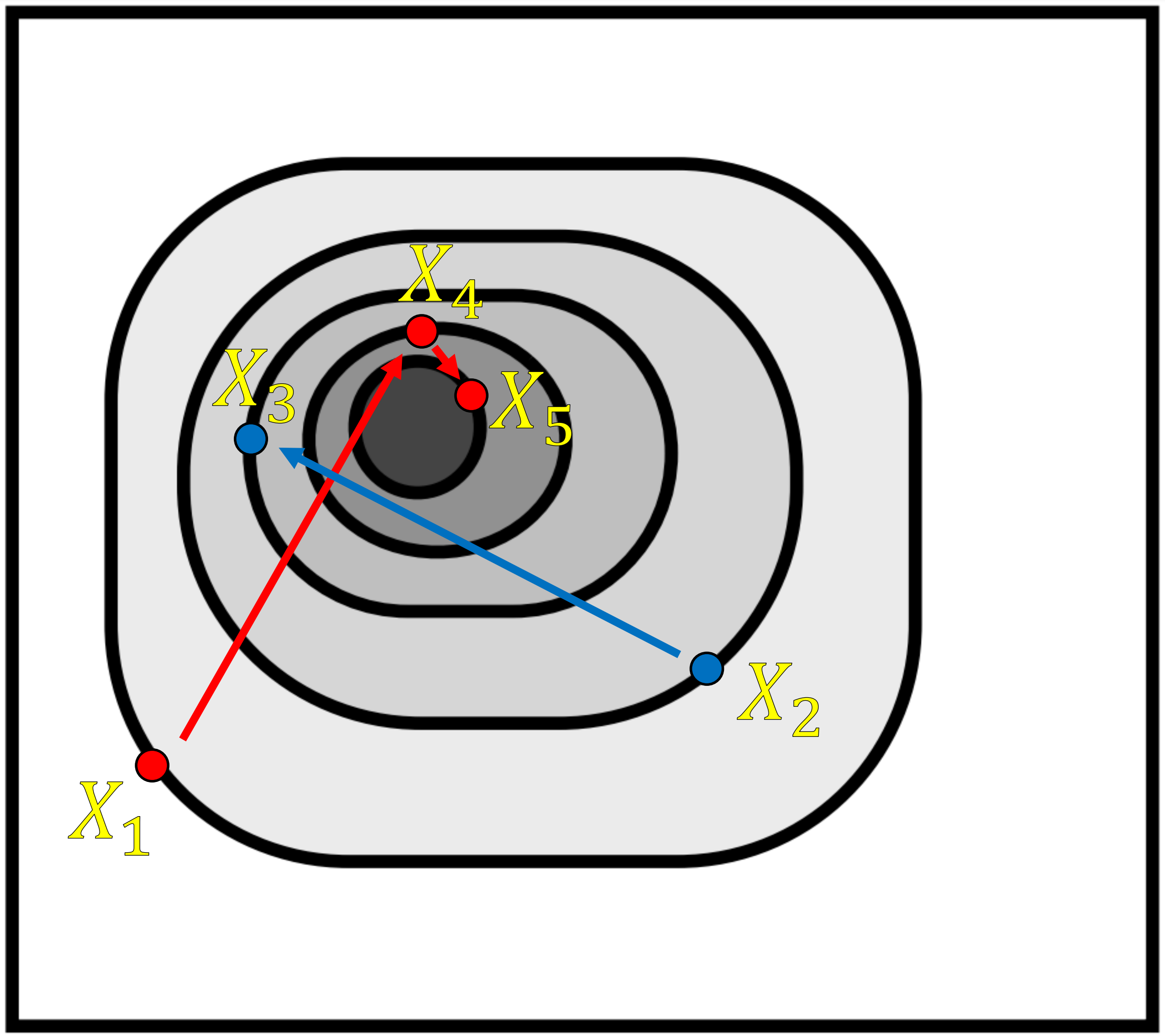
Original run

$$\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\} = \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(2)}\right\}, \ldots$$

"strand" "strand"

We would like to sample K paths from the set of all possible paths $\mathrm{P}(\{\mathcal{L}_{(i)}\}, \ldots)$. However, we don't have access to it.

$$\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\}' = \left\{\mathcal{L}_{(i)}^{(1)}\right\}', \left\{\mathcal{L}_{(i)}^{(2)}\right\}', \ldots$$



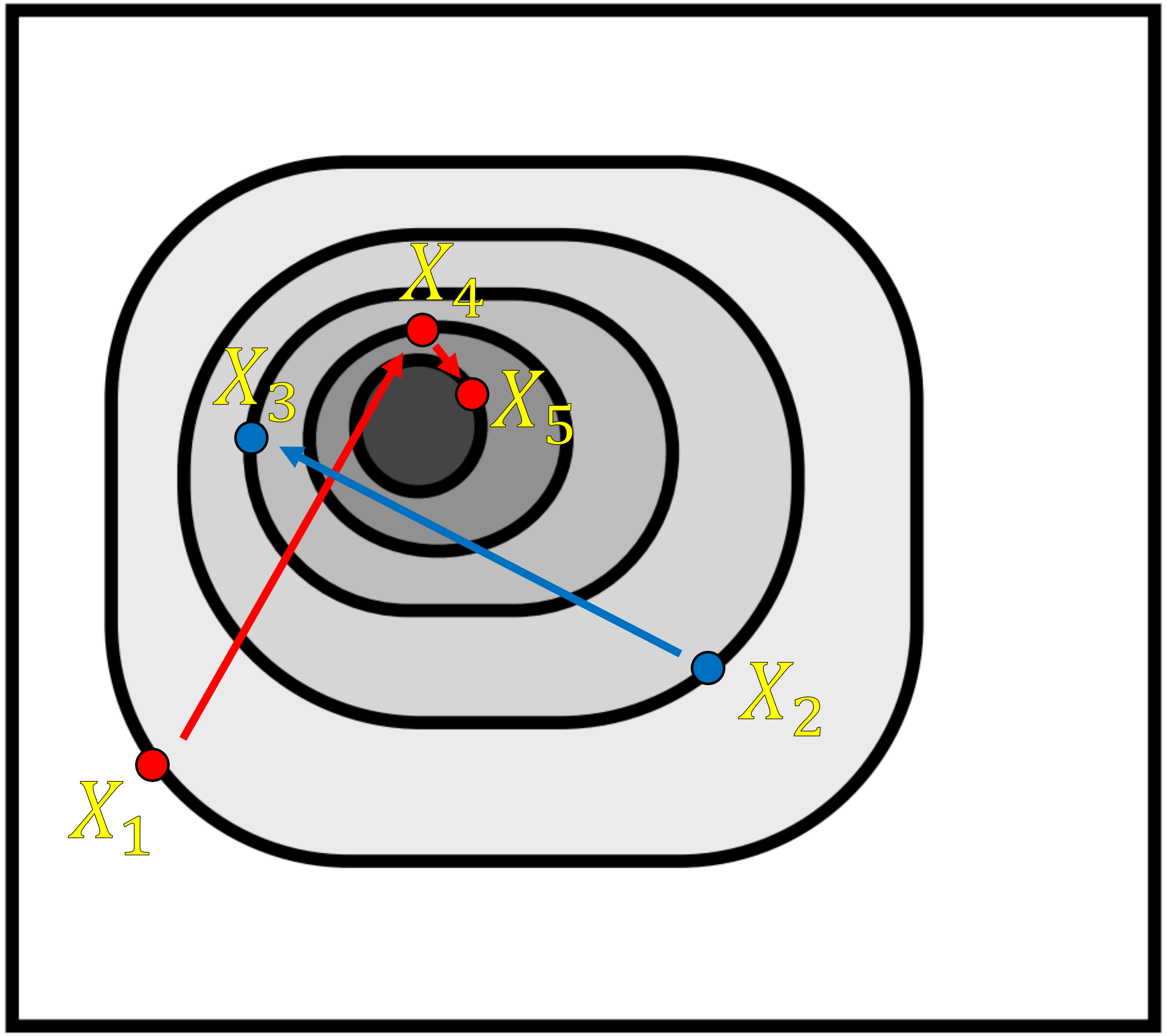
Exploring Sampling Uncertainties

Original run

"strand"

$$\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\} = \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(2)}\right\}, \ldots$$

"strand"



We would like to sample K paths from the set of all possible paths $\mathrm{P}\left(\left\{\mathcal{L}_{(i)}\right\}, \ldots\right)$. However, we don't have access to it.
Use bootstrap estimator.

$$\widetilde{\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\}}' = \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(2)}\right\}, \ldots$$

Exploring Sampling Uncertainties

Original run

"strand"

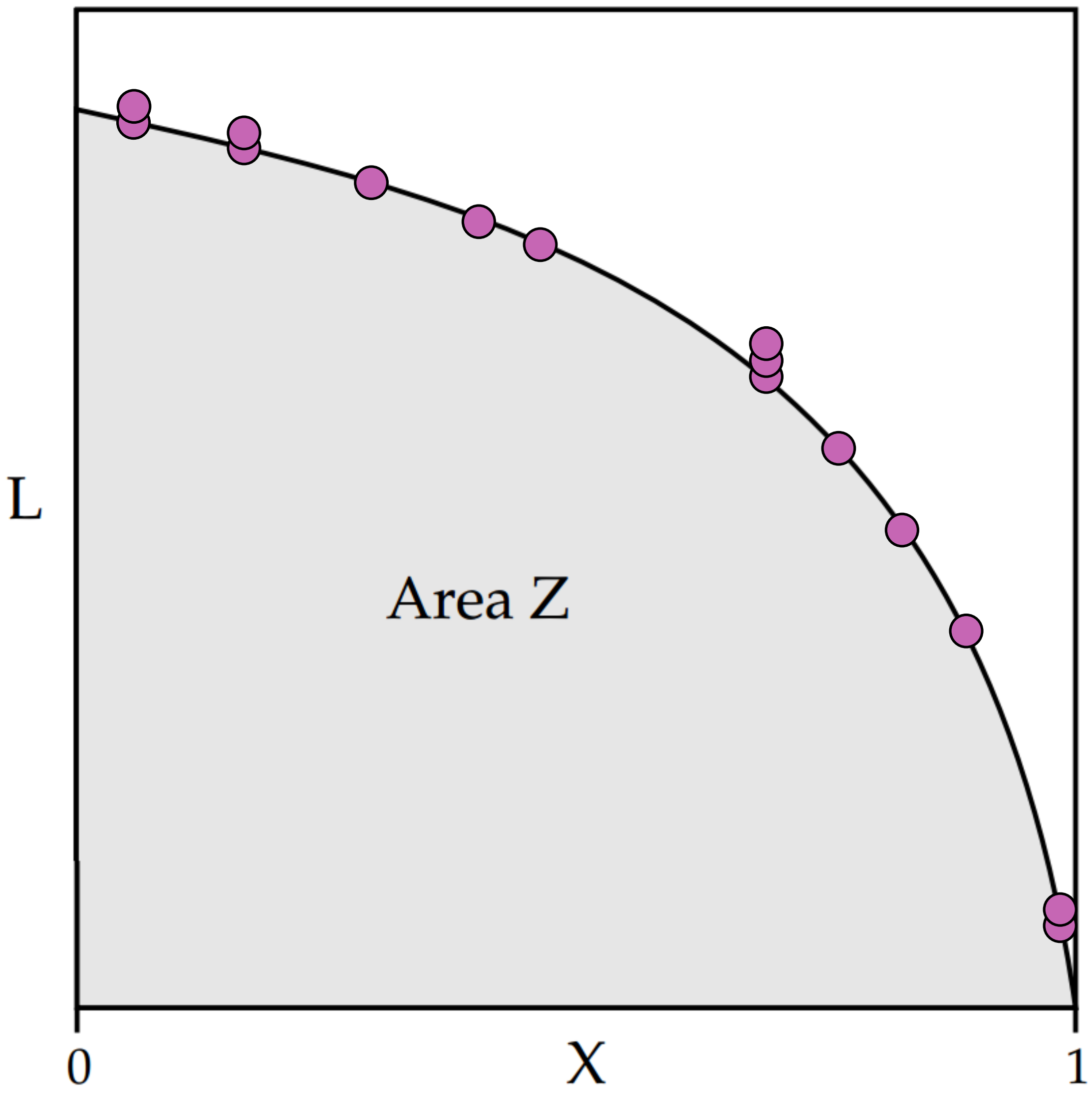
$$\left\{\mathcal{L}_{(i)}^{(\cdot)}\right\} = \left\{\mathcal{L}_{(i)}^{(1)}\right\}, \left\{\mathcal{L}_{(i)}^{(2)}\right\}, \ldots$$

"strand"

We would like to sample K paths from the set of all possible paths $\mathrm{P}\left(\left\{\mathcal{L}_{(i)}\right\}, \ldots\right)$. However, we don't have access to it.

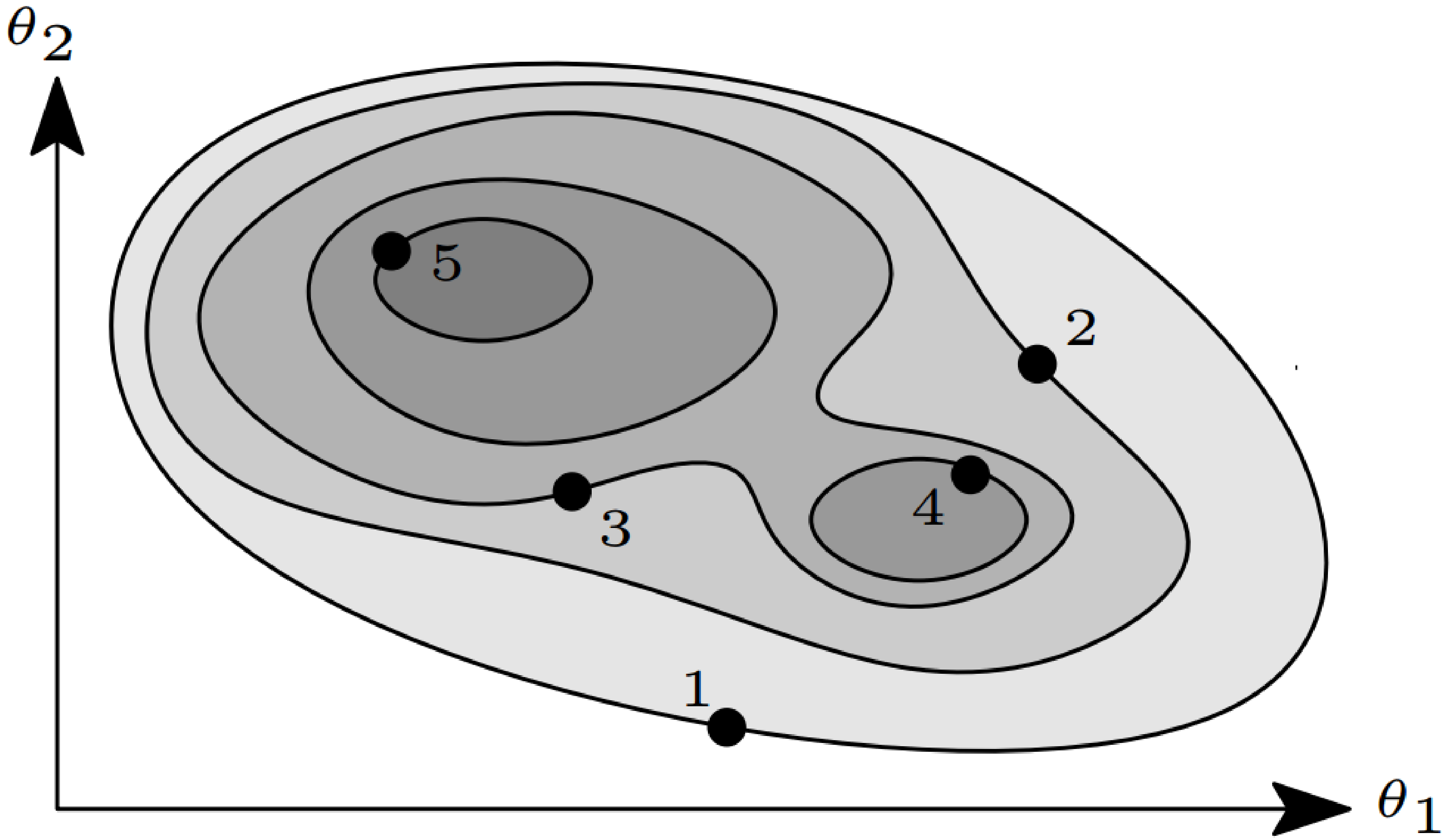
Use bootstrap estimator.

$$\widetilde{\hat{p}}_i' = \frac{\widetilde{\hat{w}}_i'}{\widetilde{\hat{Z}}'} \qquad \widetilde{\hat{Z}}' \approx \sum_{i=1}^{N'} \widetilde{\hat{w}}_i'$$



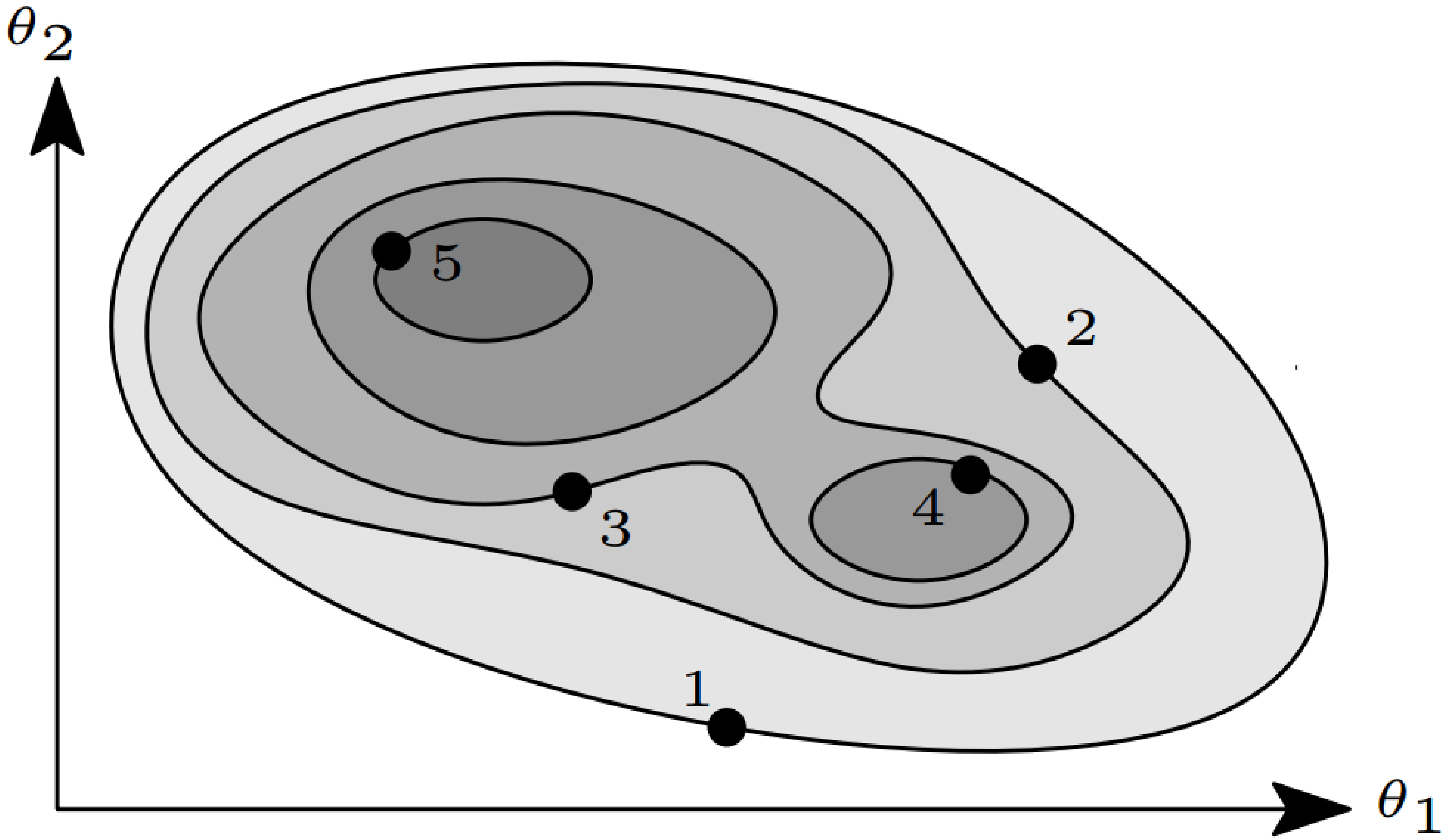
Nested Sampling
In Practice

Nested Sampling In Practice



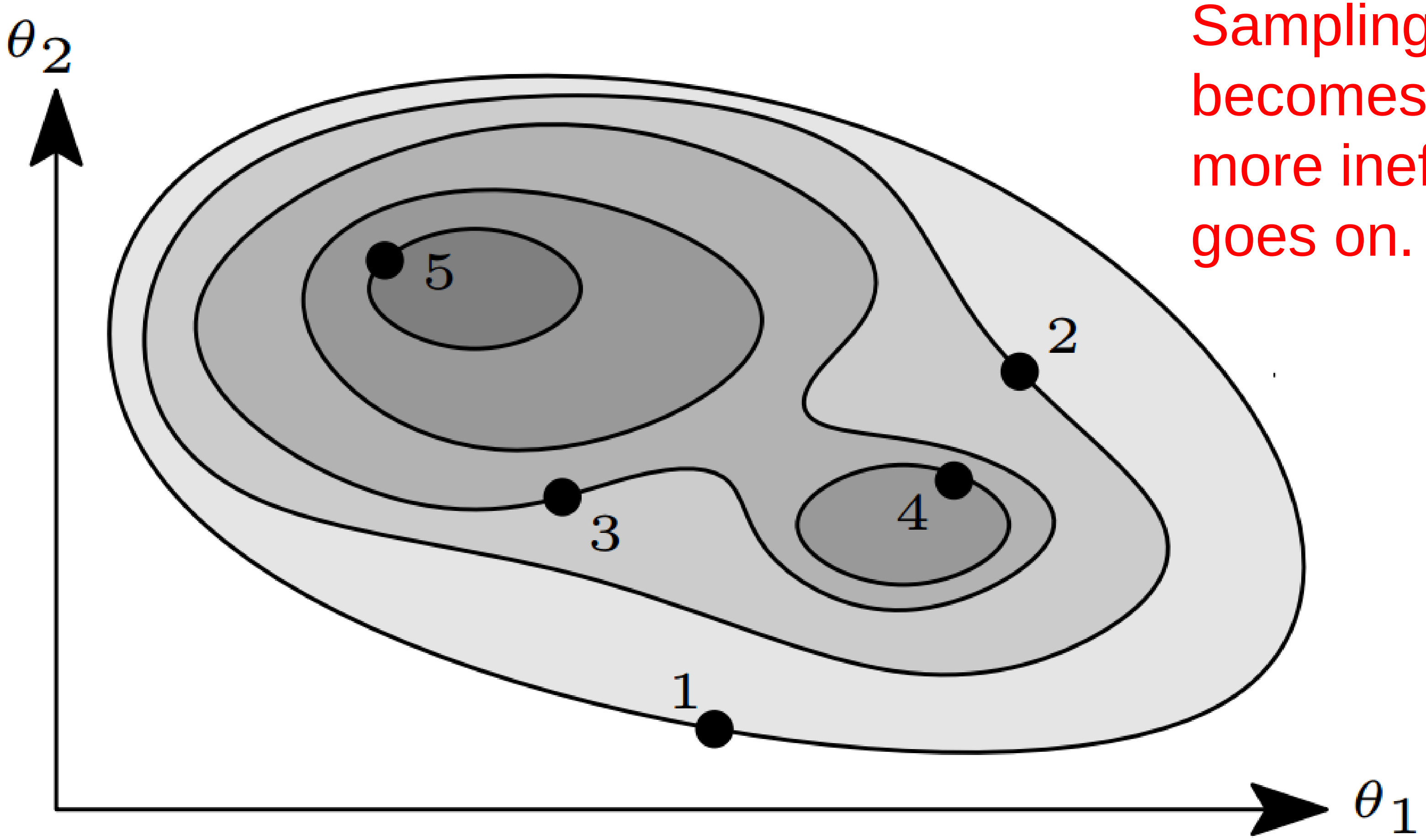
Higson et al. (2017)
arxiv:1704.03459

Method 0: Sampling from the Prior



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

Method 0: Sampling from the Prior

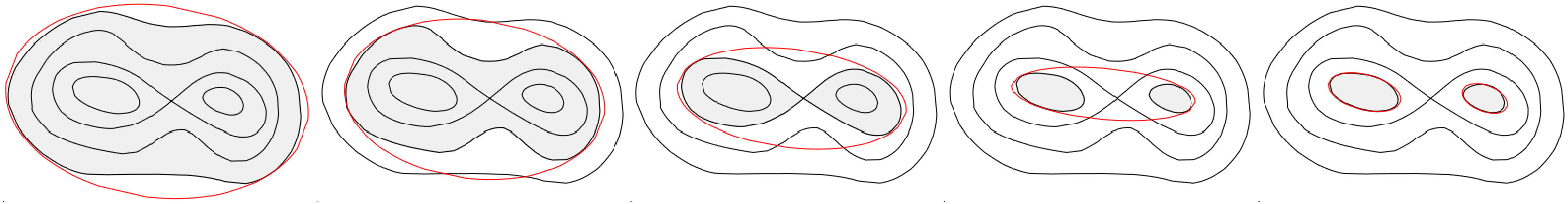


Sampling from the prior becomes exponentially more inefficient as time goes on.

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

Method 1: Constrained Uniform Sampling

Proposal:
Bound the iso-likelihood contours in real time and sample from the newly **constrained prior**.



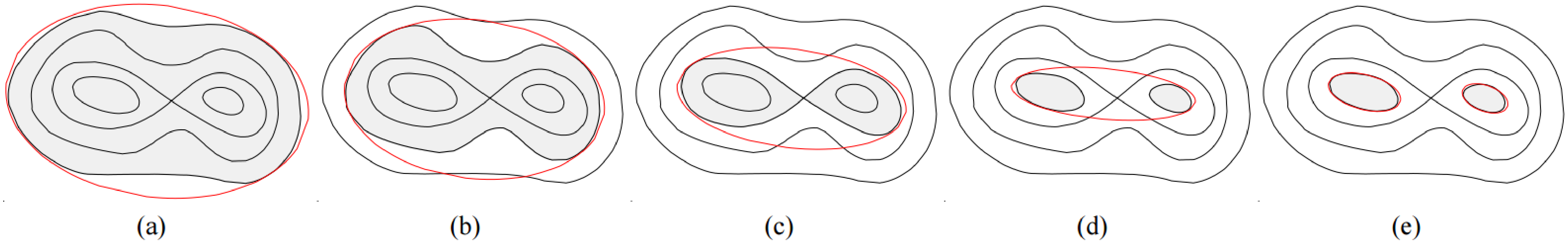
(a) (b) (c) (d) (e)

Feroz et al. (2009)

Method 1: Constrained Uniform Sampling

Issues:

- How to ensure bounds always encompass iso-likelihood contours?
- How to generate flexible bounds?

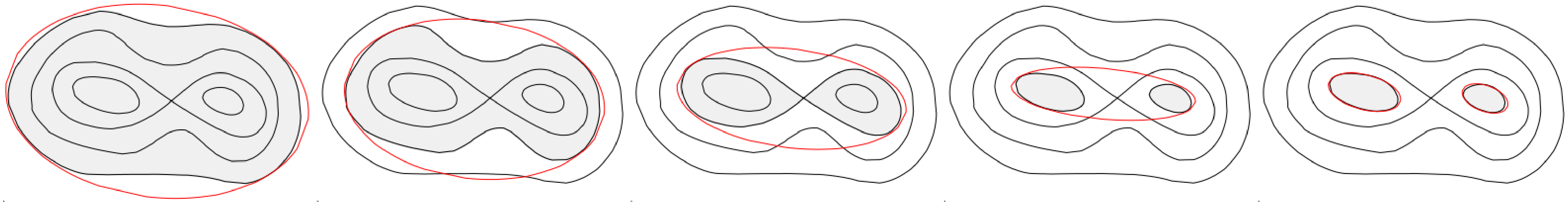


Feroz et al. (2009)

Method 1: Constrained Uniform Sampling

Issues:

- How to ensure bounds always encompass iso-likelihood contours? Bootstrapping.
- How to generate flexible bounds? Easier with uniform (transformed) prior.



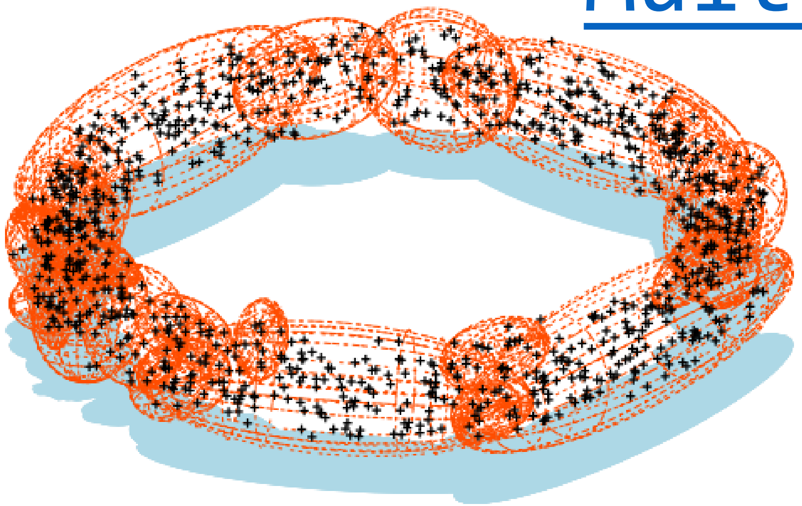
(a) (b) (c) (d) (e)

Feroz et al. (2009)

Method 1: Constrained Uniform Sampling

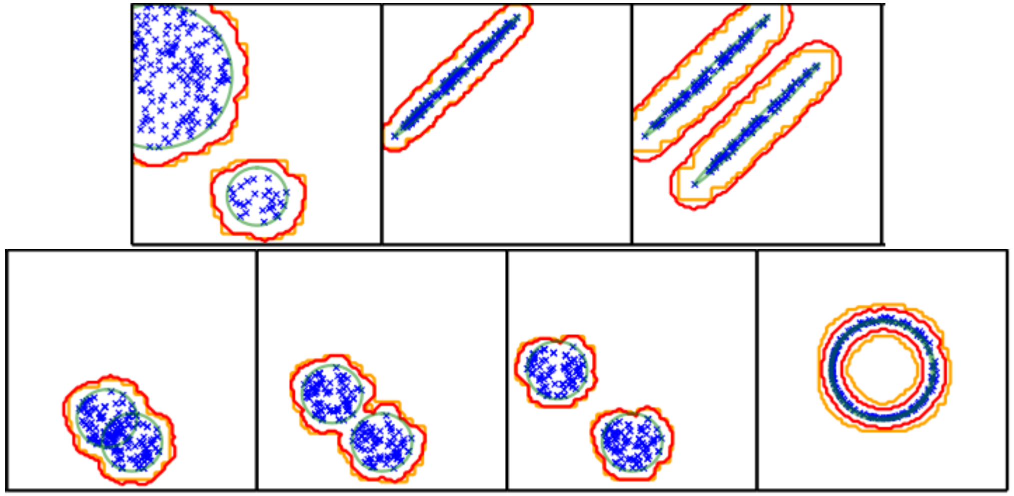
Ellipsoids

[MultiNest]



Feroz et al. (2009)

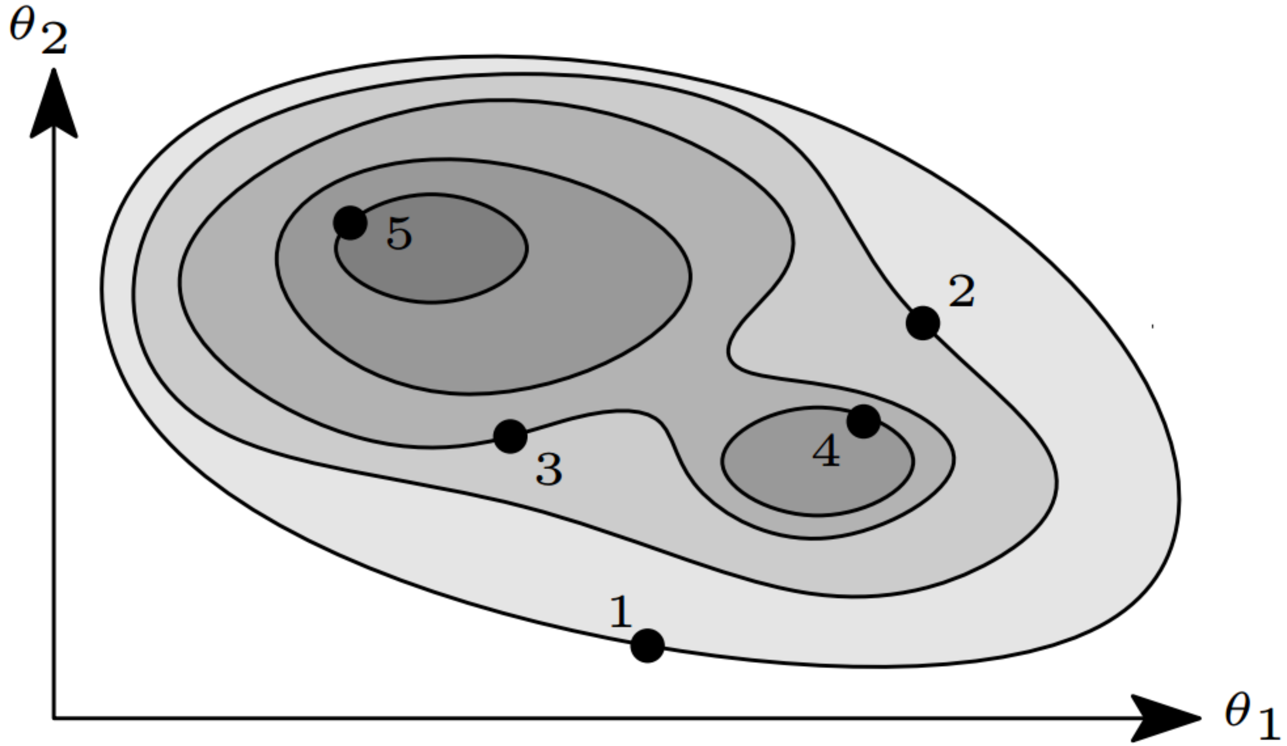
Balls/Cubes



Buchner (2014)

Method 2: “Evolving” Previous Samples

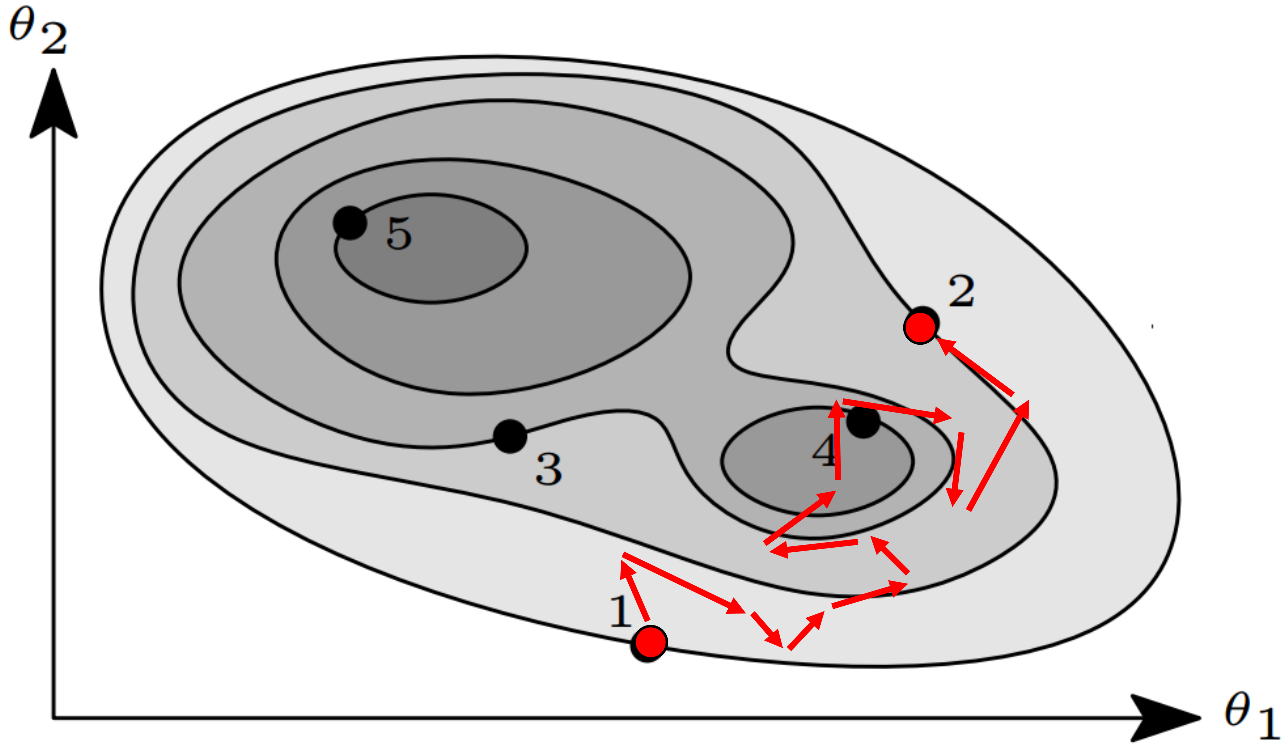
Proposal:
Generate independent samples subject to the likelihood constraint by “evolving” copies of current live points.



Method 2: “Evolving” Previous Samples

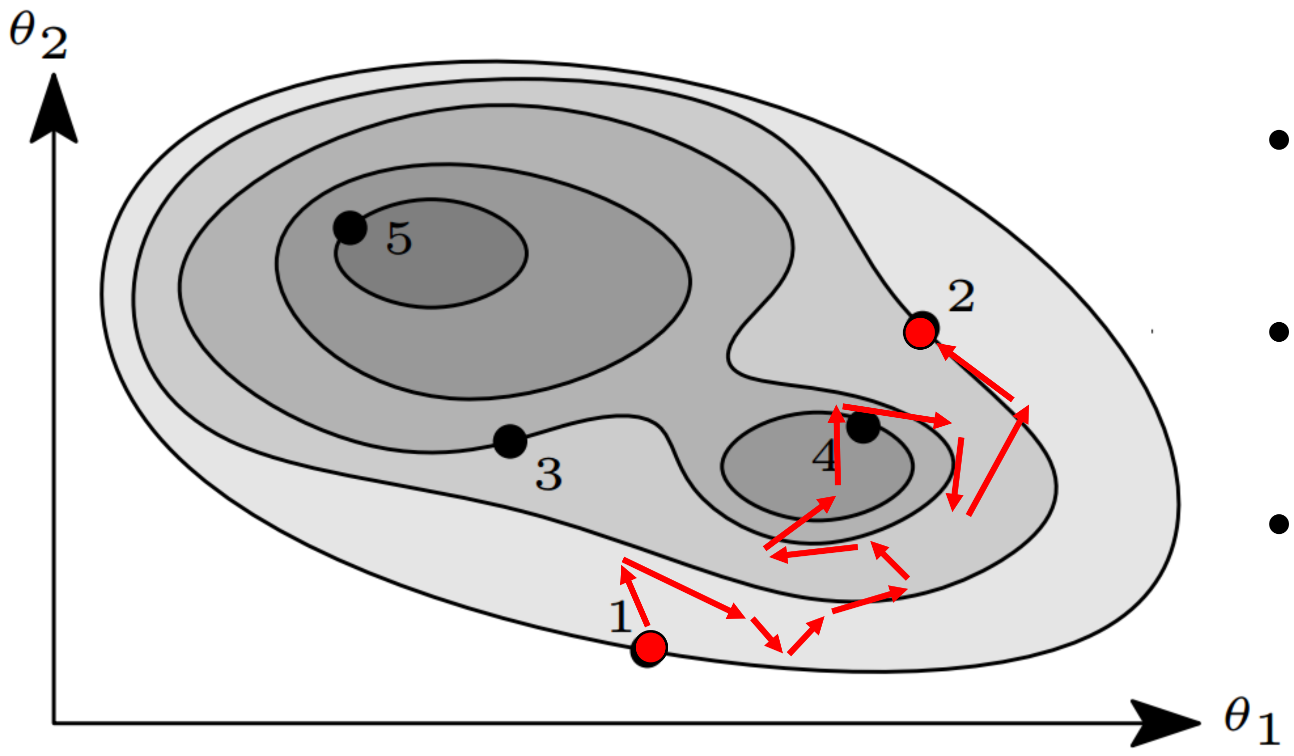
Proposal:

Generate independent samples subject to the likelihood constraint by “evolving” copies of current live points.



Method 2: “Evolving” Previous Samples

Proposal:
Generate independent samples subject to the likelihood constraint by “evolving” copies of current live points.

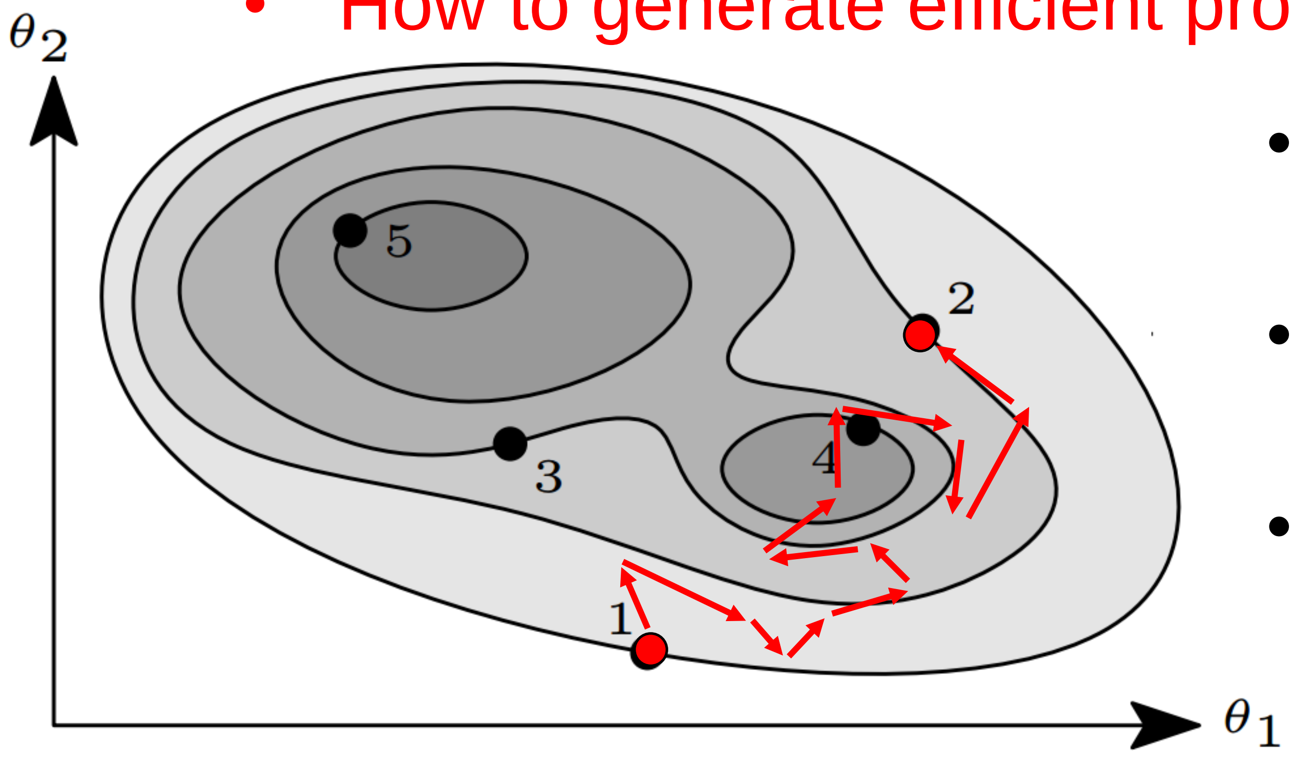


- Random walks (i.e. MCMC)
- Slice sampling [PolyChord]
- Random trajectories (i.e. HMC)

Method 2: “Evolving” Previous Samples

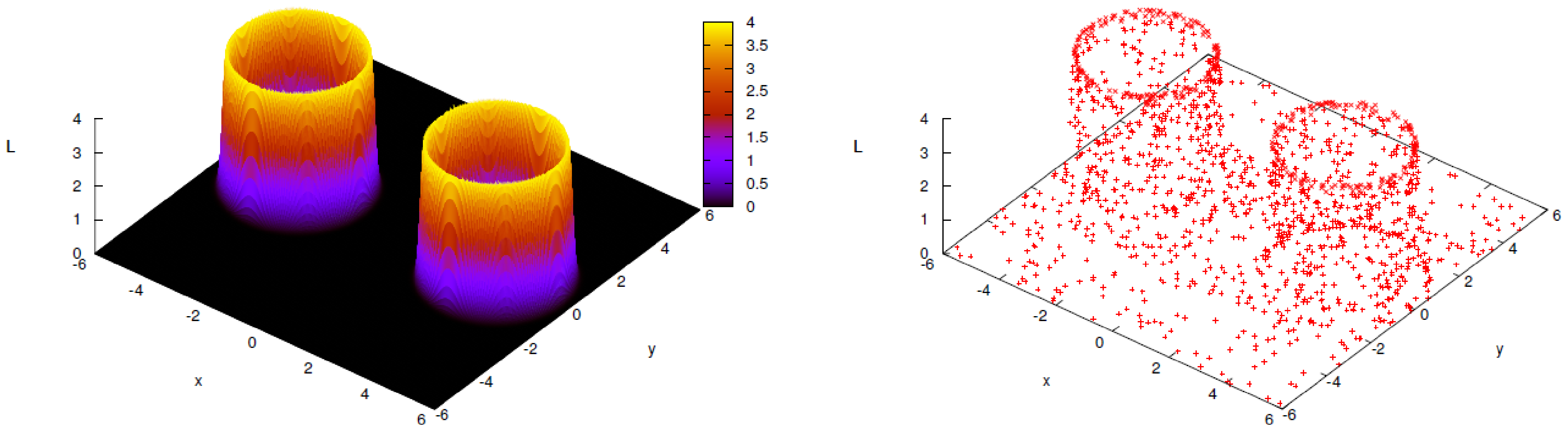
Issues:

- How to ensure samples are independent (thinning) and properly distributed within likelihood constraint?
- How to generate efficient proposals?



- Random walks (i.e. MCMC)
- Slice sampling PolyChord
- Random trajectories (i.e. HMC)

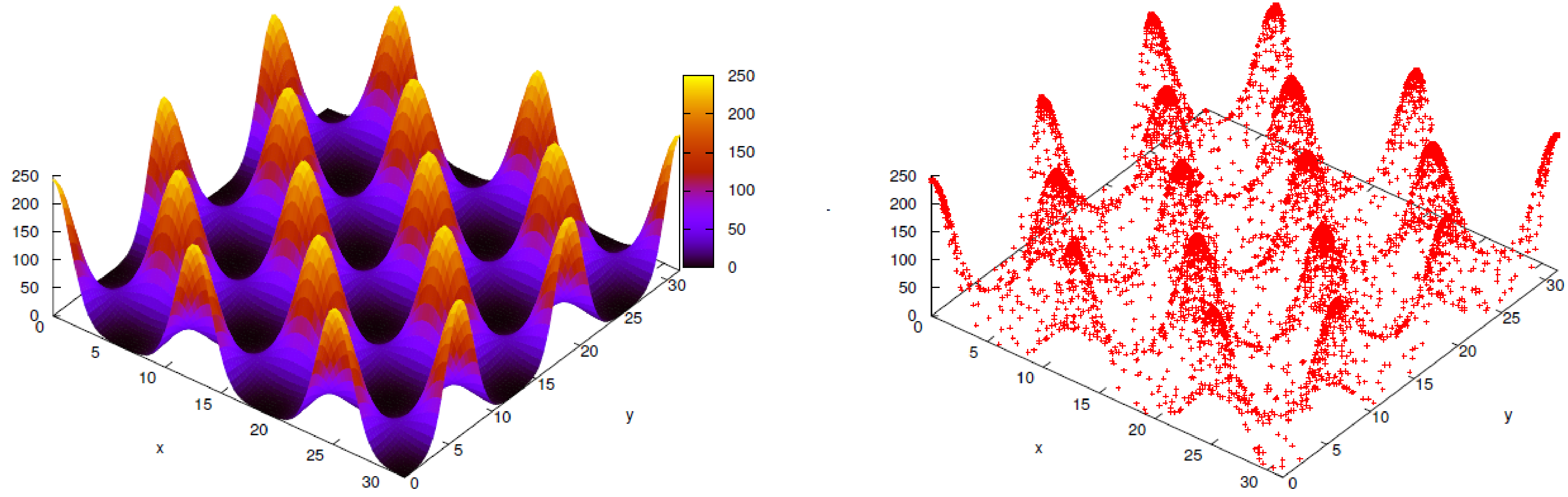
Example: Gaussian Shells



(a) (b)

Feroz et al. (2013)

Example: Eggbox



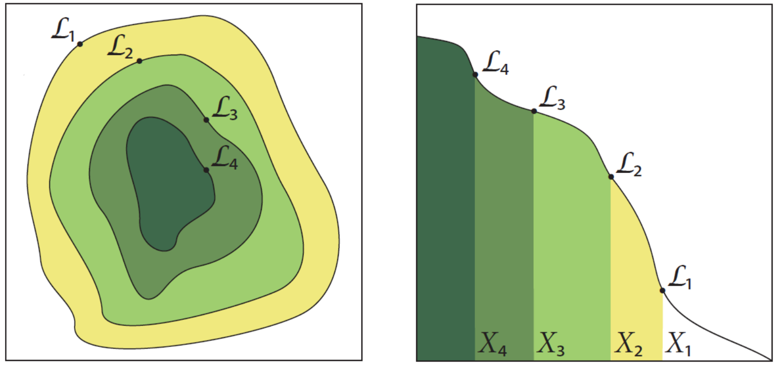
(a) (b)

Feroz et al. (2013)

Summary: (Static) Nested Sampling

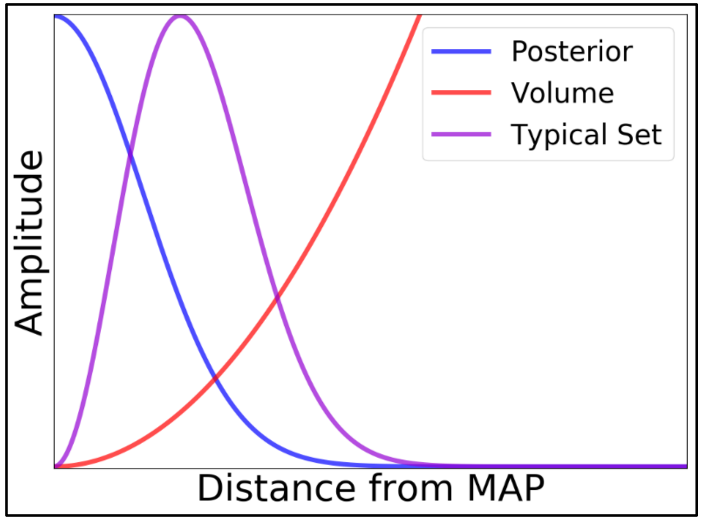
Summary: (Static) Nested Sampling

1. Estimates the evidence $\hat{\mathcal{Z}}$.



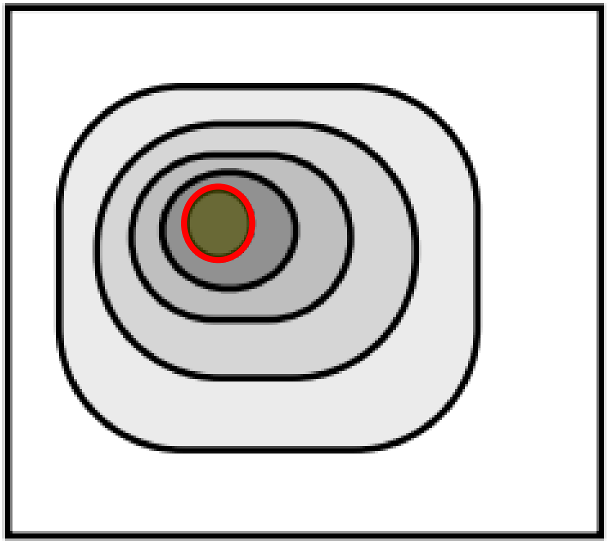
Summary: (Static) Nested Sampling

1. Estimates the evidence $\hat{\mathcal{Z}}$.
2. Estimates the posterior $\hat{p}(\mathbf{\Theta})$.



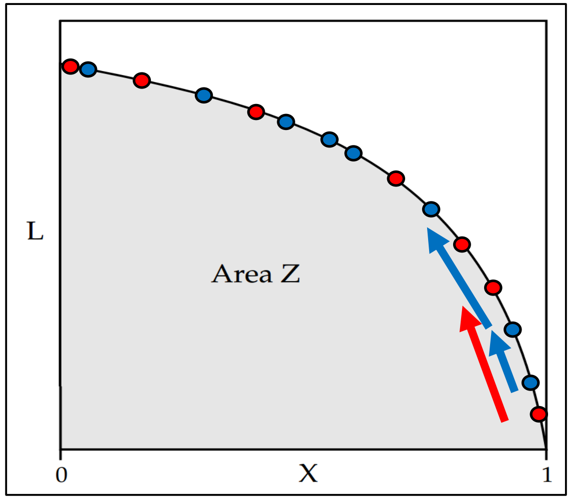
Summary: (Static) Nested Sampling

1. Estimates the evidence $\hat{\mathcal{Z}}$.
2. Estimates the posterior $\hat{p}(\mathbf{\Theta})$.
3. Possesses well-defined stopping criteria.



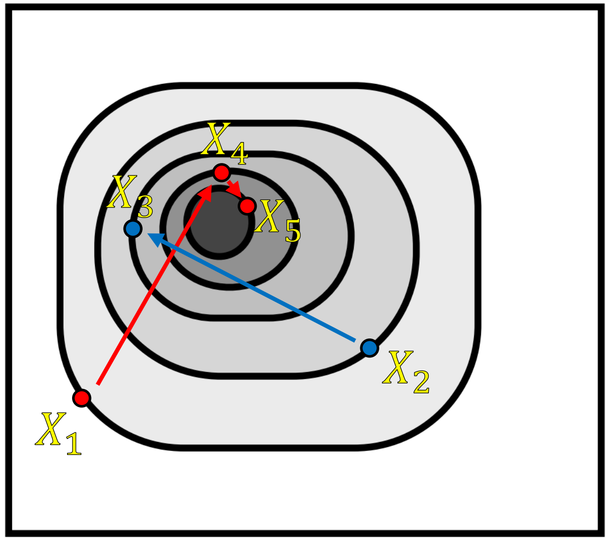
Summary: (Static) Nested Sampling

1. Estimates the evidence $\hat{\mathcal{Z}}$.
2. Estimates the posterior $\hat{p}(\mathbf{\Theta})$.
3. Possesses well-defined stopping criteria.
4. Combining runs improves inference.



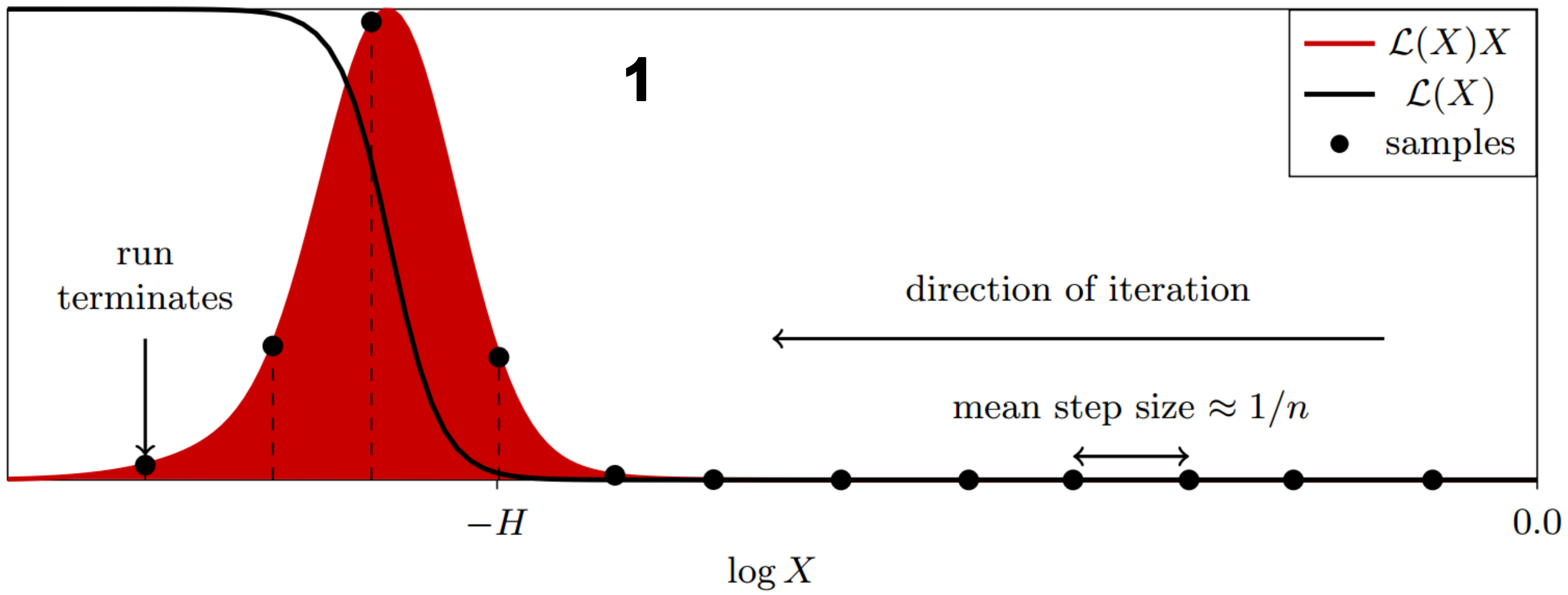
Summary: (Static) Nested Sampling

1. Estimates the evidence $\hat{\mathcal{Z}}$.
2. Estimates the posterior $\hat{p}(\boldsymbol{\Theta})$.
3. Possesses well-defined stopping criteria.
4. Combining runs improves inference.
5. Sampling and statistical uncertainties can be simulated from a single run.



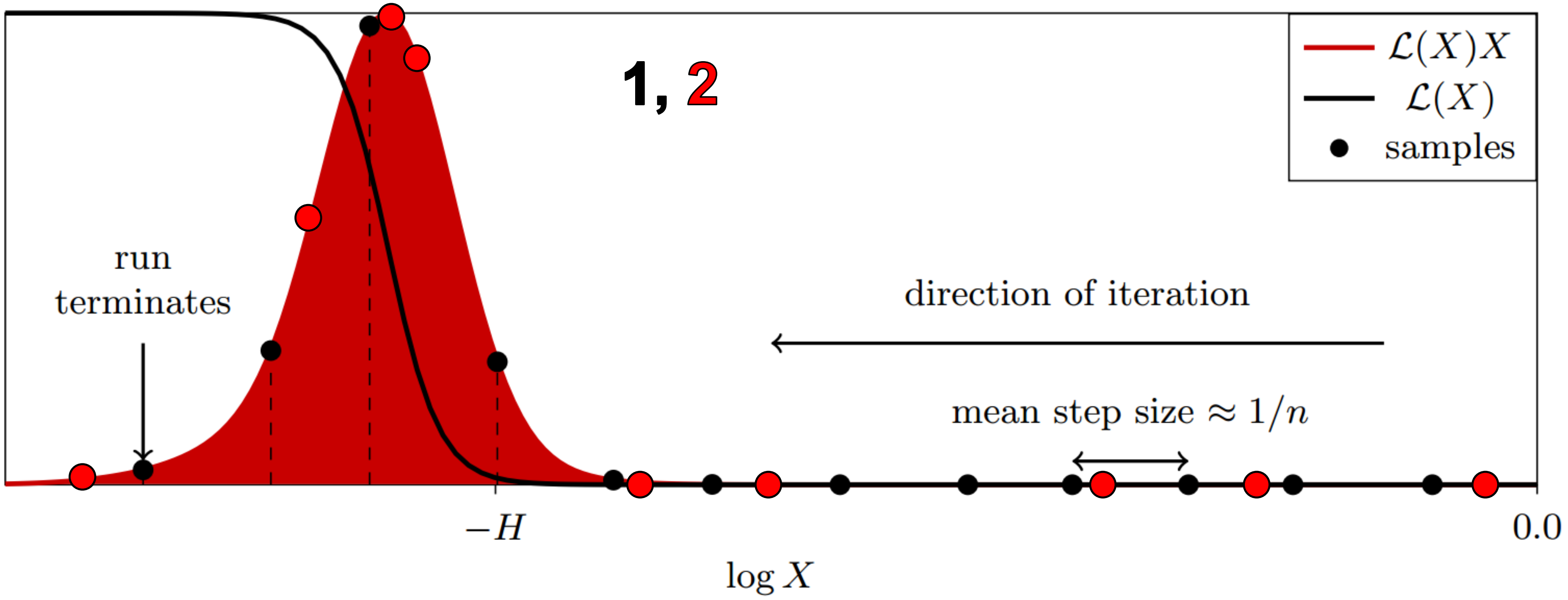
Dynamic Nested Sampling

Dynamic Nested Sampling



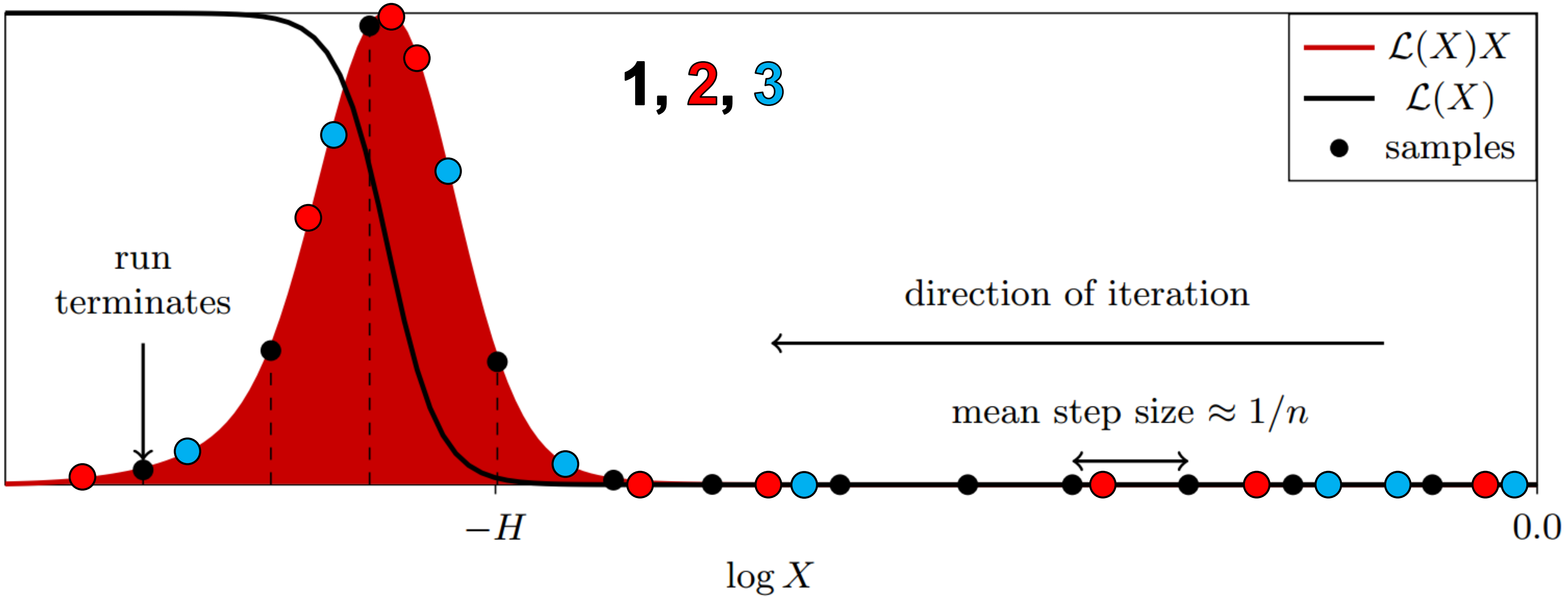
Higson et al. (2017)
arxiv:1704.03459

Dynamic Nested Sampling



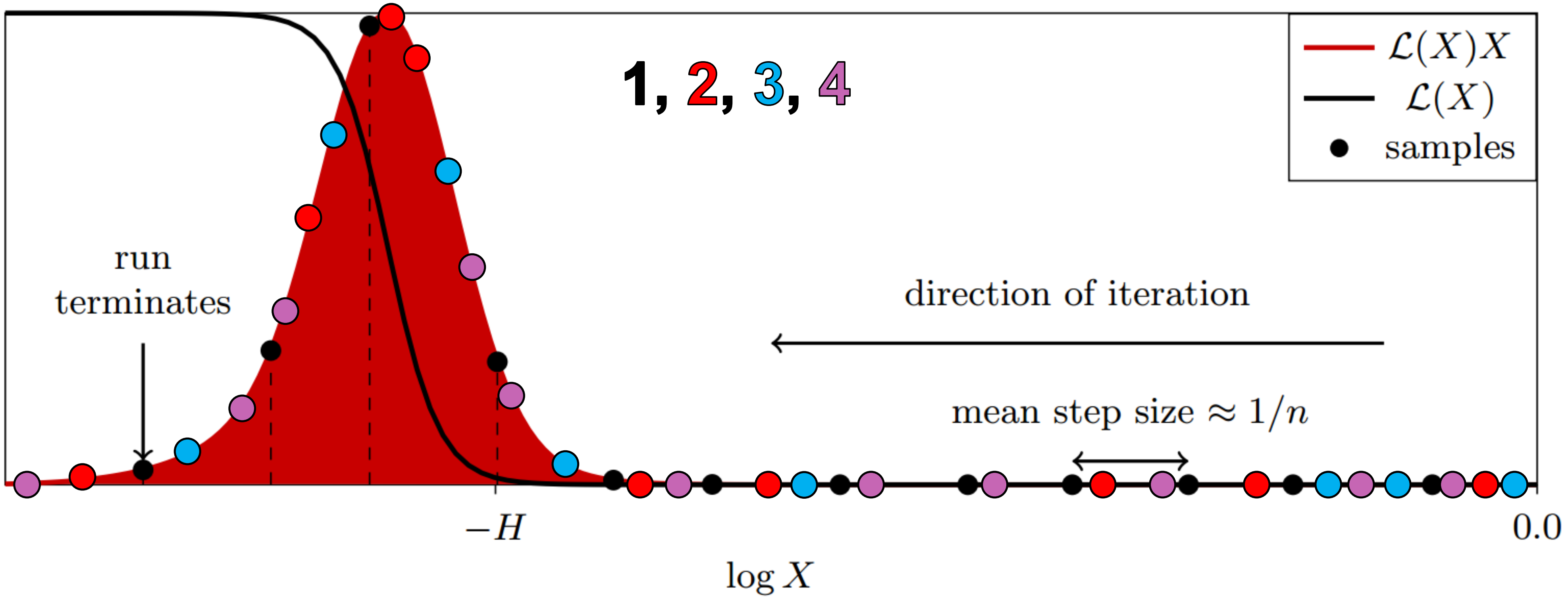
Higson et al. (2017)
[arxiv:1704.03459](https://arxiv.org/abs/1704.03459)

Dynamic Nested Sampling



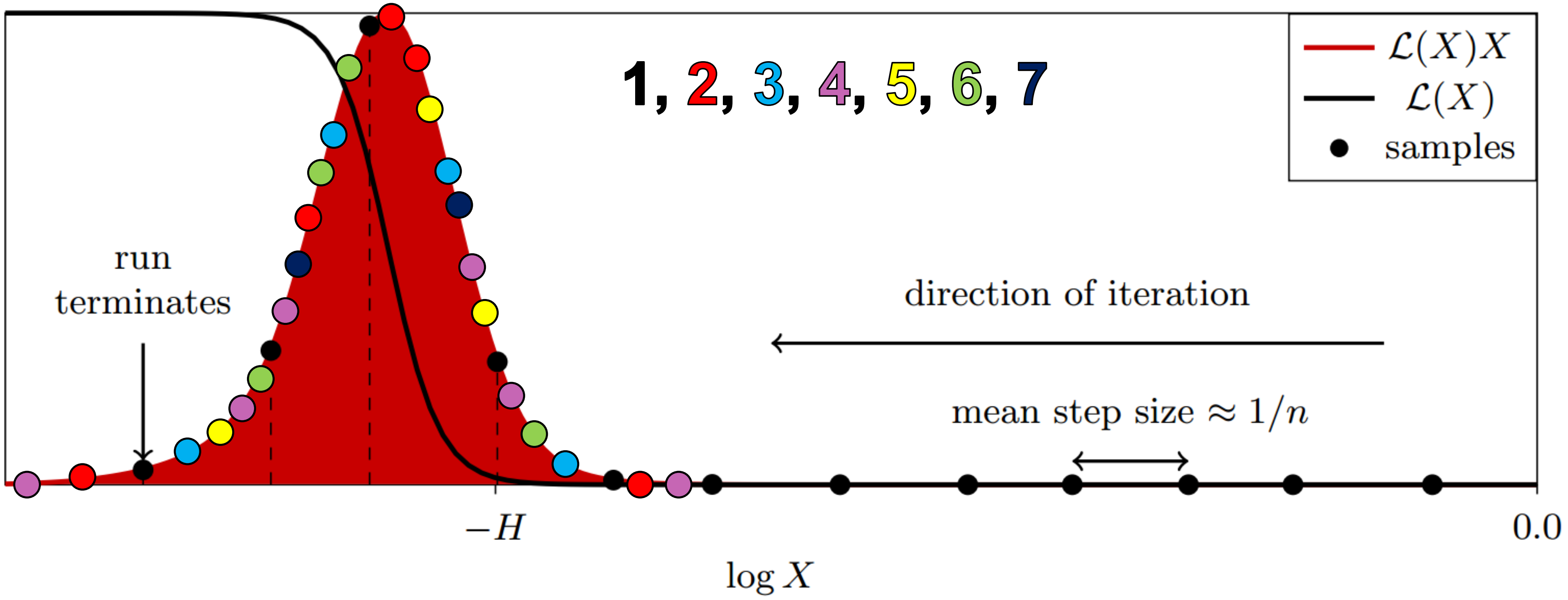
Higson et al. (2017)
arxiv:1704.03459

Dynamic Nested Sampling



Higson et al. (2017)
arxiv:1704.03459

Dynamic Nested Sampling

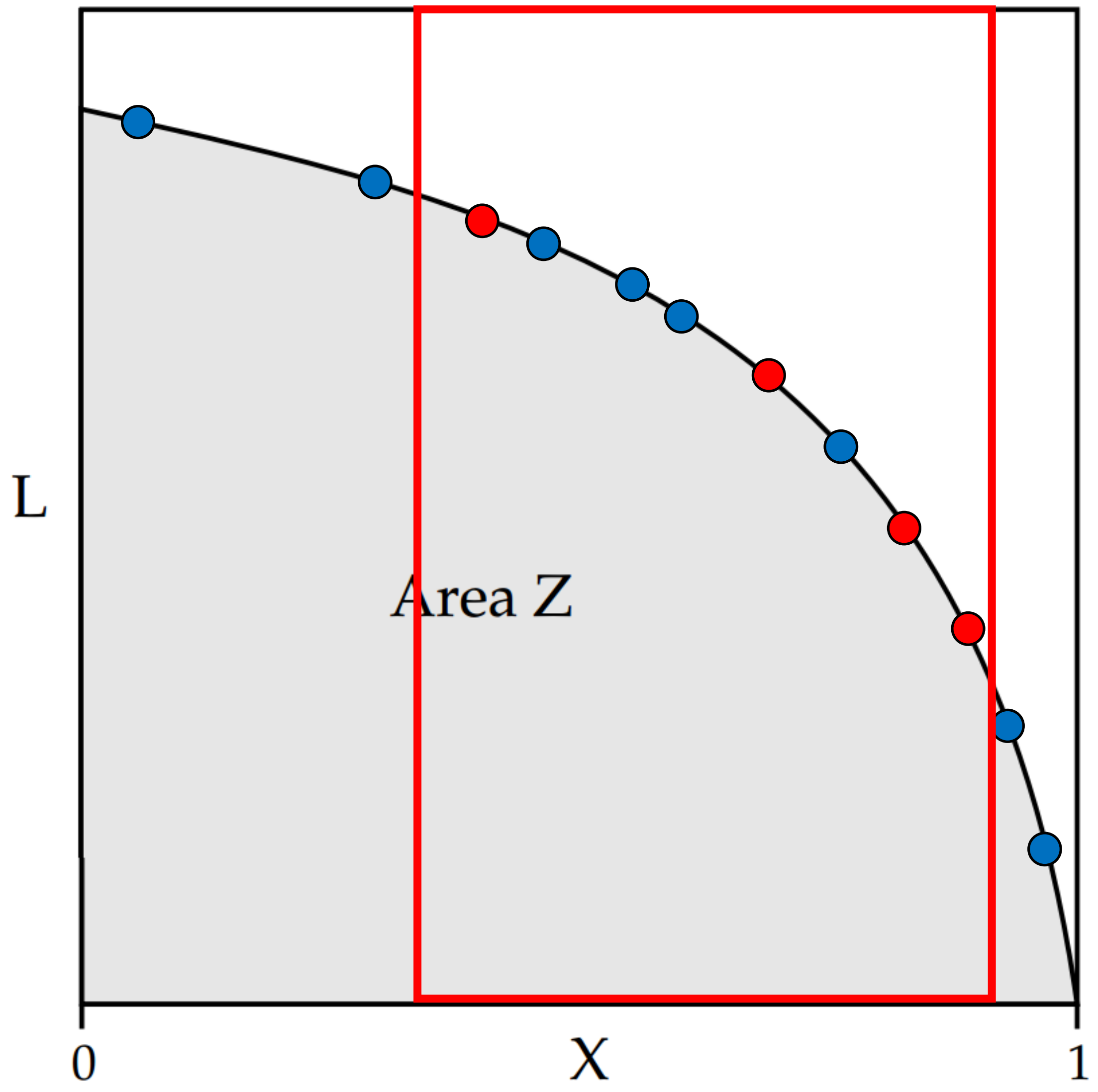


Higson et al. (2017)
arxiv:1704.03459

Dynamic Nested Sampling

$$\mathcal{L}_{\max}^{(1)} > \mathcal{L}_{N_1}^{(1)} > \cdots > \mathcal{L}_2^{(1)} > \mathcal{L}_1^{(1)} > \mathcal{L}_{\min}^{(1)}$$

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > 0$$



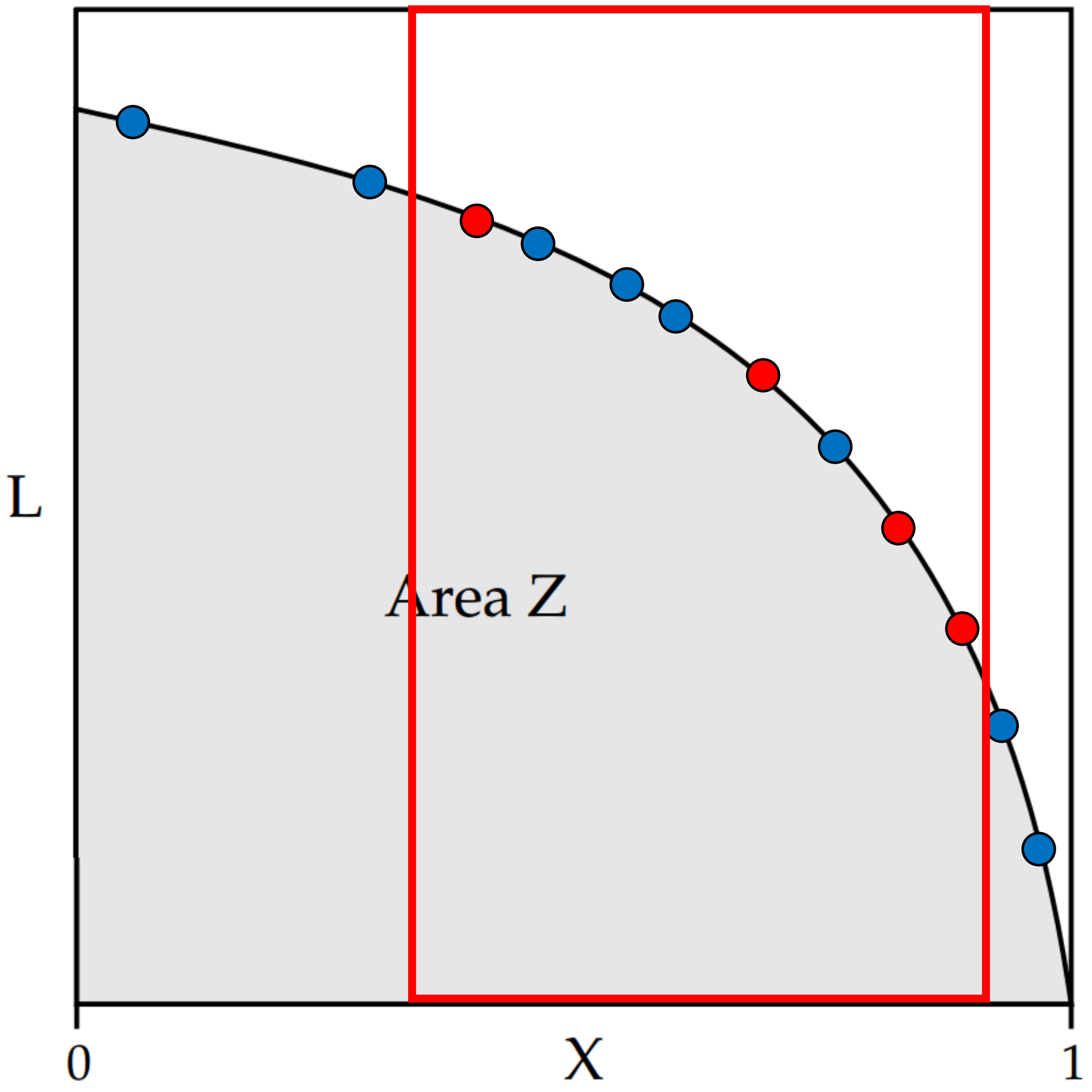
Dynamic Nested Sampling

N_2 live points $N_1 + N_2$ live points

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_{N_1}^{(1)} \ldots > \mathcal{L}_1^{(1)}$$

$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > 0$$

N_2 live points



Dynamic Nested Sampling

N_2 live points $\qquad$ $N_1 + N_2$ live points

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_{N_1}^{(1)} \ldots > \mathcal{L}_{1}^{(1)}$$

$$> \mathcal{L}_{2}^{(2)} > \mathcal{L}_{1}^{(2)} > 0$$

N_2 live points

$$\ln X_{i+1} = \sum_{j=1}^{i} \ln T_j$$

Dynamic Nested Sampling

N_2 live points $N_1 + N_2$ live points

$$\mathcal{L}_{N_2}^{(2)} > \cdots > \mathcal{L}_{N_1}^{(1)} \ldots > \mathcal{L}_1^{(1)}$$

$$> \mathcal{L}_2^{(2)} > \mathcal{L}_1^{(2)} > 0$$

N_2 live points

$$\ln X_{i+1} = \sum_{j=1}^{i} \ln T_j$$

Constant/Increasing

$K_{i+1} \geq K_i$:

$$T_{i+1} \sim \mathrm{Beta}(K_{i+1}, 1)$$

Dynamic Nested Sampling

N_2 live points $\quad$ $N_1 + N_2$ live points

$$\mathcal{L}^{(2)}_{N_2} > \cdots > \mathcal{L}^{(1)}_{N_1} \ldots > \mathcal{L}^{(1)}_{1}$$

$$> \mathcal{L}^{(2)}_{2} > \mathcal{L}^{(2)}_{1} > 0$$

N_2 live points

$$\ln X_{i+1} = \sum_{j=1}^{i} \ln T_j$$

Constant/Increasing

$$K_{i+1} \geq K_i:$$

$$T_{i+1} \sim \mathrm{Beta}(K_{i+1}, 1)$$

Decreasing sequence

$$K_{i+j>i} < K_i:$$

$$T_{i+1}, \ldots T_{i+j} \sim$$

$$U_{(K_i)}, \ldots, U_{(K_i - j + 1)}$$

Static Nested Sampling

Exponential shrinkage

Uniform shrinkage

$$\ln \mathbb{E}[X_{N+k}] = \sum_{i=1}^{N} \ln\left(\frac{K}{K+1}\right) + \sum_{j=1}^{k} \ln\left(\frac{K-j+1}{K-j+2}\right)$$

Dynamic Nested Sampling

Exponential shrinkage

Uniform shrinkage

$$\ln \mathbb{E}[X_N] = \sum_{i=1}^{n_1} \ln\left(\frac{K_i}{K_i + 1}\right) + \sum_{j=1}^{n_2} \ln\left(\frac{K_{n_1} - j + 1}{K_{n_1} - j + 2}\right)$$

$$+ \cdots + \sum_{j=1}^{K_{\mathrm{final}}} \ln\left(\frac{K_{\mathrm{final}} - j + 1}{K_{\mathrm{final}} - j + 2}\right)$$

Benefits of Dynamic Nested Sampling

- Can accommodate new “strands” within a particular range of prior volumes without changing overall statistical framework.

- Particles can be adaptively added until stopping criteria are reached, allowing targeted estimation.

Benefits of Dynamic Nested Sampling

- Can accommodate new “strands” within a particular range of prior volumes without changing overall statistical framework.

- Particles can be adaptively added until stopping criteria are reached, allowing targeted estimation.

Allocating Samples

Posterior weight

$$I_P(i) \sim \hat{p}_i$$

Evidence weight

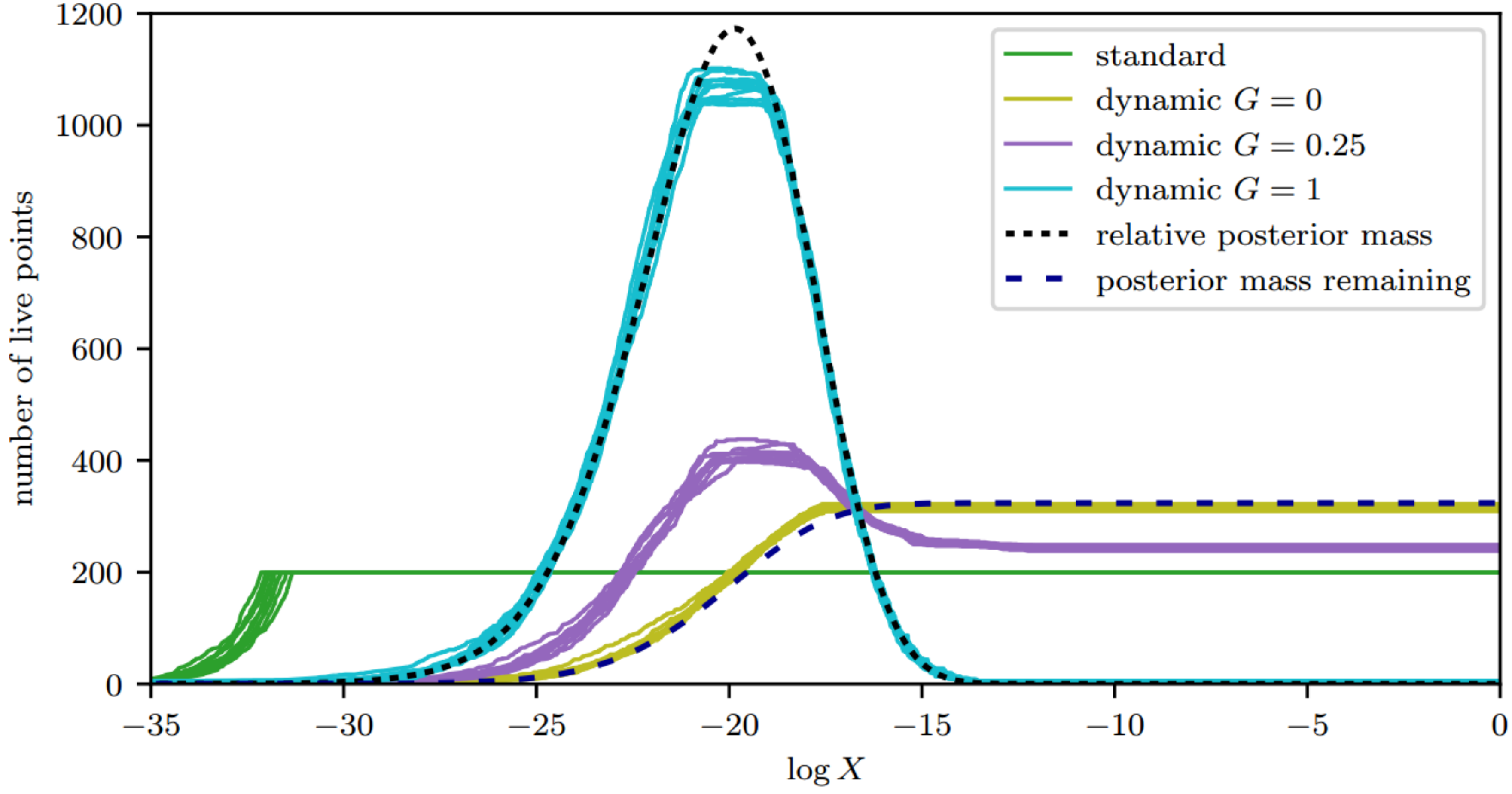
$$I_Z(i) \sim \frac{\mathbb{E}[\mathcal{Z}_{>i}]}{K_i}$$

Weight Function

$$I(G,i) = G\,\frac{I_P(i)}{\sum_j I_P(j)} + (1-G)\,\frac{I_Z(i)}{\sum_j I_Z(j)}$$

Higson et al. (2017)
arxiv:1704.03459

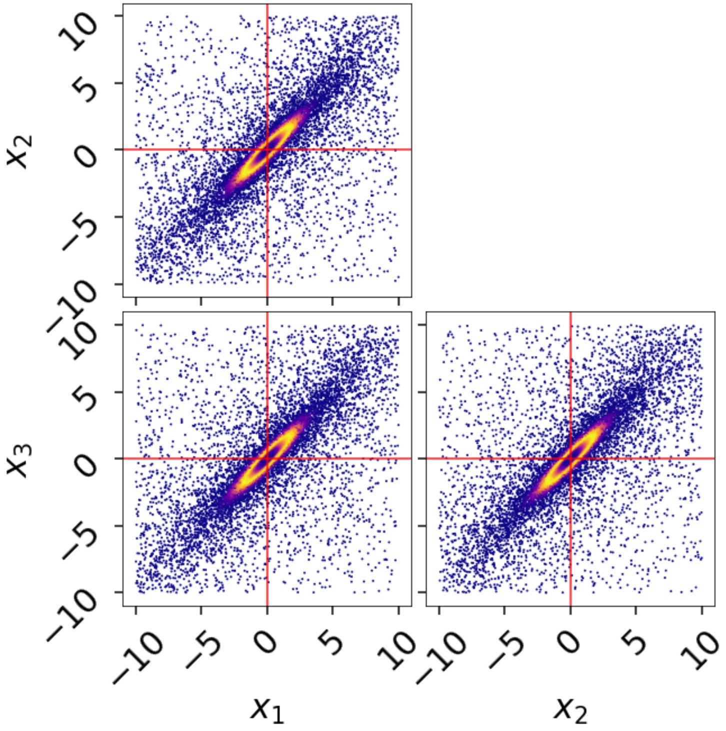
Allocating Samples



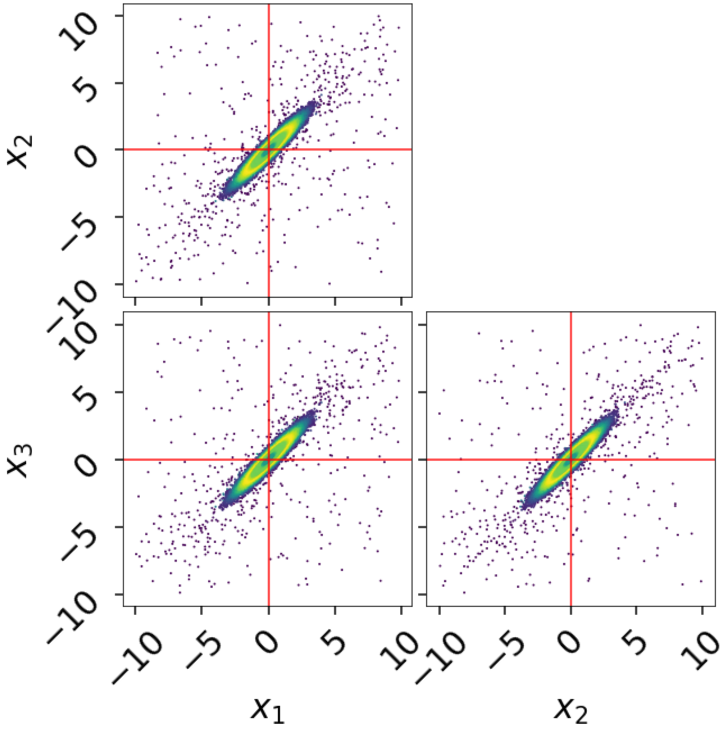
Higson et al. (2017)
[arxiv:1704.03459]

Allocating Samples

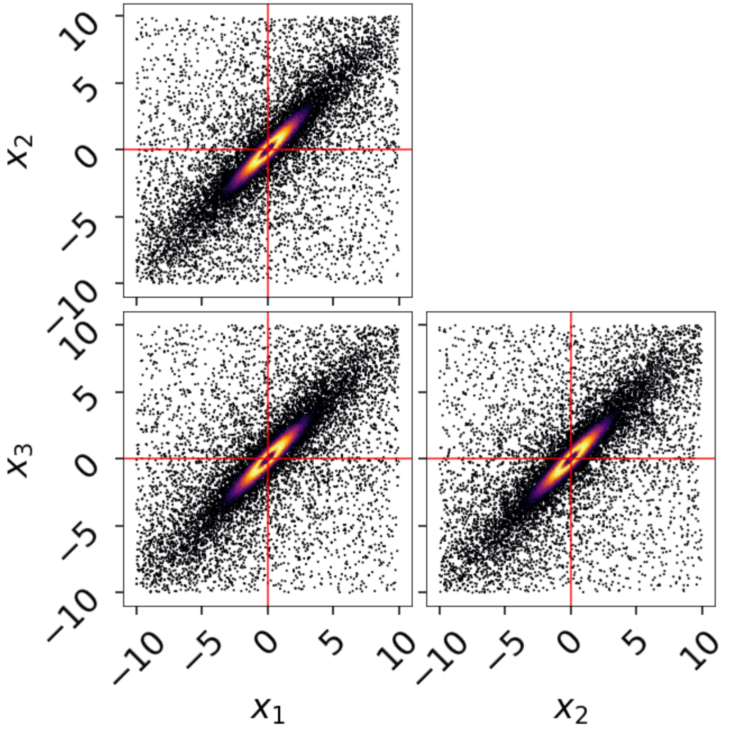
3-D correlated multivariate Normal



Static



Dynamic (posterior)



Dynamic (evidence)

How Many Samples is Enough?

How Many Samples is Enough?

- **Ill-posed question: depends on application!**

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\boldsymbol{\Theta})$ with $\hat{p}(\boldsymbol{\Theta})$, how many samples are necessary?

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\mathbf{\Theta})$ with $\hat{p}(\mathbf{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\mathbf{\Theta}_i\})$ and $\tilde{p}(\{\mathbf{\Theta}_\mathrm{i}\})$ densities to be "close".

"True" posterior constructed over same "domain".

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\mathbf{\Theta})$ with $\hat{p}(\mathbf{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\mathbf{\Theta}_i\})$ and $\tilde{p}(\{\mathbf{\Theta}_\mathrm{i}\})$ densities to be “close”.

$$H(\tilde{p}|\hat{p}) \equiv \int_{\Omega_{\mathbf{\Theta}}} \hat{p}(\mathbf{\Theta}) \ln \frac{\hat{p}(\mathbf{\Theta})}{\tilde{p}(\mathbf{\Theta})} d\mathbf{\Theta}$$

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\boldsymbol{\Theta})$ with $\hat{p}(\boldsymbol{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\boldsymbol{\Theta}_i\})$ and $\tilde{p}(\{\boldsymbol{\Theta}_\mathrm{i}\})$ densities to be “close”.

$$H(\tilde{p}|\hat{p}) = \sum_{i=1}^{N} \hat{p}_i \ln(\hat{p}_i/\tilde{p}_i)$$

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\mathbf{\Theta})$ with $\hat{p}(\mathbf{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\mathbf{\Theta}_i\})$ and $\tilde{p}(\{\mathbf{\Theta}_\mathrm{i}\})$ densities to be “close”.

We **want** access to $\mathrm{P}(\{\hat{p}_i\}|\tilde{p})$, but we don’t know $\tilde{p}(\mathbf{\Theta})$.

$$H(\tilde{p}|\hat{p}) = \sum_{i=1}^{N} \hat{p}_i \ln(\hat{p}_i/\tilde{p}_i)$$

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\boldsymbol{\Theta})$ with $\hat{p}(\boldsymbol{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\boldsymbol{\Theta}_i\})$ and $\tilde{p}(\{\boldsymbol{\Theta}_\mathrm{i}\})$ densities to be "close".

We **want** access to $\Pr(\{\hat{p}_i\}|p(\boldsymbol{\Theta}))$, but we don't know $p(\boldsymbol{\Theta})$.

Use bootstrap estimator.

$$H\left(\hat{p}\middle|\tilde{\hat{p}}'\right) = \sum_{i=1}^{N} \tilde{\hat{p}}'_i \ln\left(\tilde{\hat{p}}'_i / \hat{p}_i\right)$$

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\mathbf{\Theta})$ with $\hat{p}(\mathbf{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\mathbf{\Theta}_i\})$ and $\tilde{p}(\{\mathbf{\Theta}_\mathrm{i}\})$ densities to be "close".

We **want** access to $\Pr(\{\hat{p}_i\}|p(\mathbf{\Theta}))$, but we don't know $p(\mathbf{\Theta})$.

Use bootstrap estimator.

$$H(\hat{p}|\tilde{\hat{p}}') = \sum_{i=1}^{N} \tilde{\hat{p}}'_i \ln(\tilde{\hat{p}}'_i/\hat{p}_i)$$

Random variable

How Many Samples is Enough?

- In any sampling-based approach to estimating $p(\boldsymbol{\Theta})$ with $\hat{p}(\boldsymbol{\Theta})$, how many samples are necessary?

Assume general case: we want D-dimensional $\hat{p}(\{\boldsymbol{\Theta}_i\})$ and $\tilde{p}(\{\boldsymbol{\Theta}_\mathrm{i}\})$ densities to be “close”.

We **want** access to $\mathrm{Pr}(\{\hat{p}_i\}|p(\boldsymbol{\Theta}))$, but we don’t know $p(\boldsymbol{\Theta})$.

Use bootstrap estimator.

Possible stopping criterion: fractional (%) variation in H.

Dynamic Nested Sampling Summary

1. Can sample from multi-modal distributions.
2. Can simultaneously estimate the evidence $\hat{\mathcal{Z}}$ and posterior $\hat{p}(\boldsymbol{\Theta})$.
3. Combining independent runs improves inference ("trivially parallelizable").
4. Can simulate uncertainties (sampling and statistical) from a single run.
5. Enables adaptive sample allocation during runtime using arbitrary weight functions.
6. Possesses evidence/posterior-based stopping criteria.

Examples and Applications

Dynamic Nested Sampling with **dynesty**

Beta release

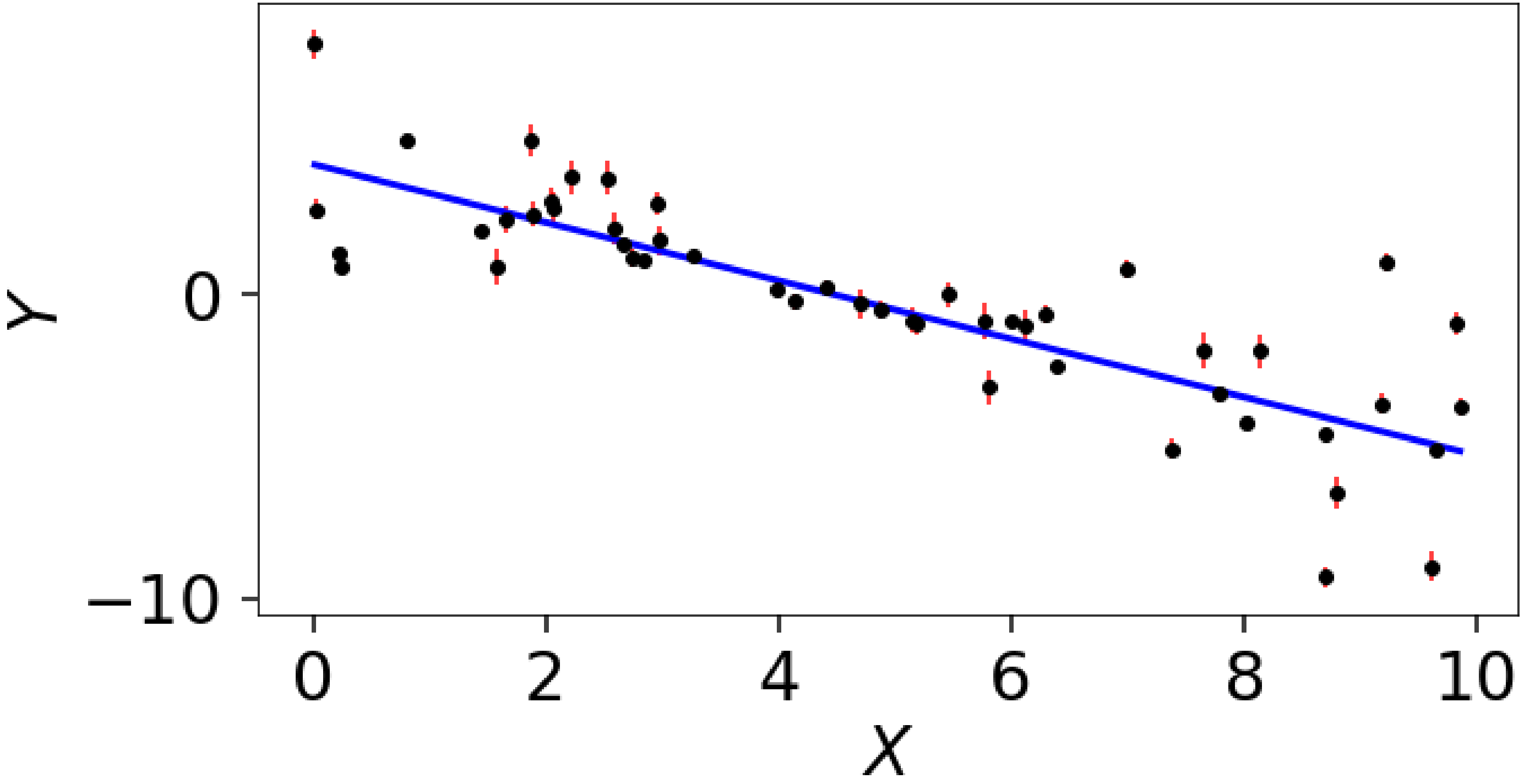
- Pure Python.
- Easy to use.
- Modular.
- Open source.
- Parallelizable.
- Flexible bounding/sampling methods.
- Thorough documentation!

dynesty.readthedocs.io

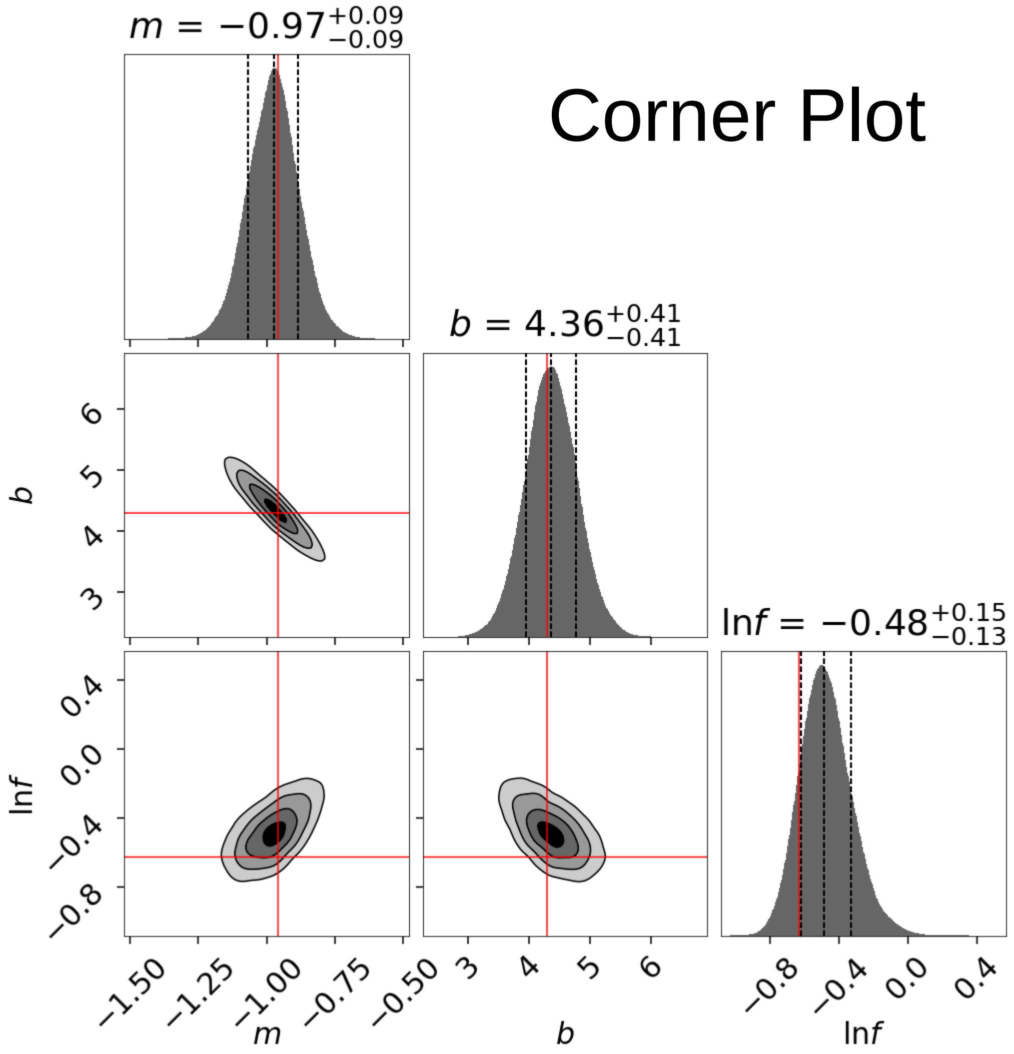
Example:

Example: Linear Regression (Posterior)

$$\mathbf{\Theta} = \{a, b, \ln f\}$$

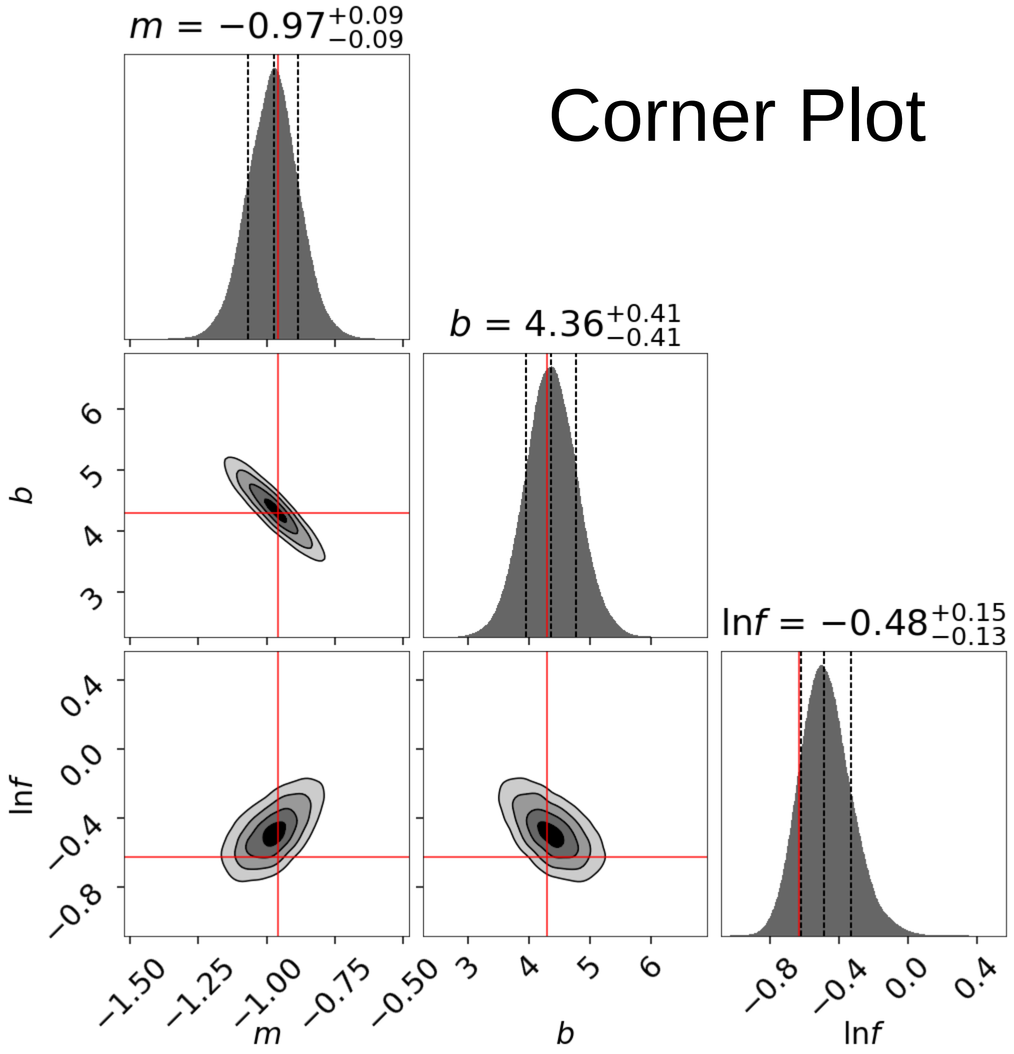


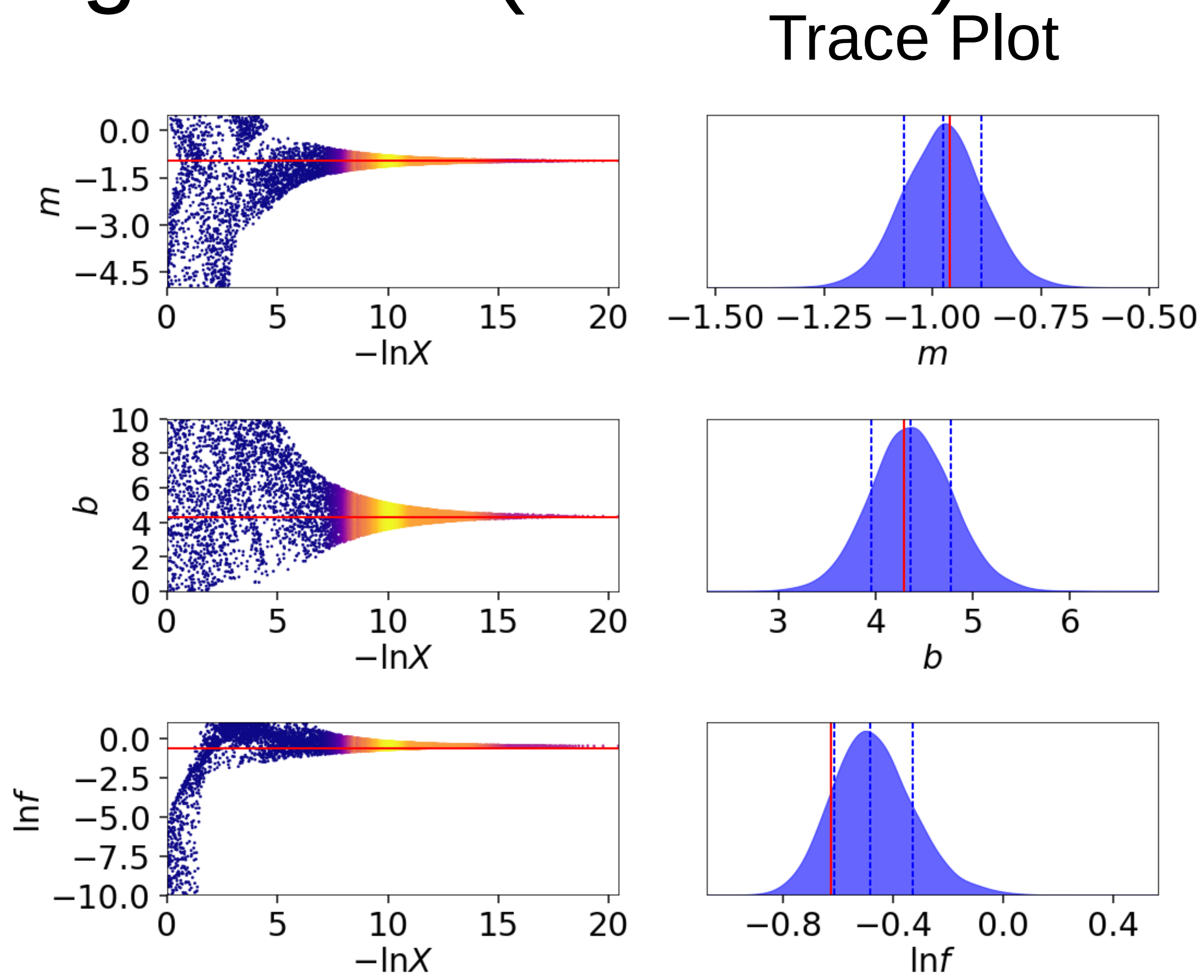
Example: Linear Regression (Posterior)



Corner Plot

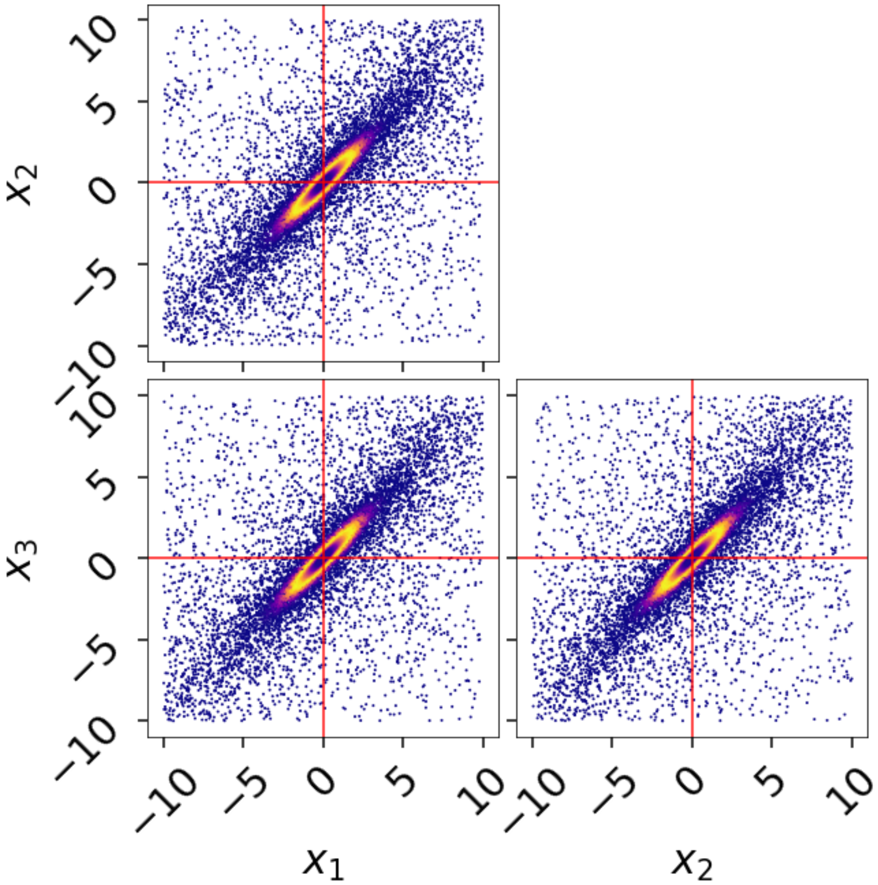
Example: Linear Regression (Posterior)



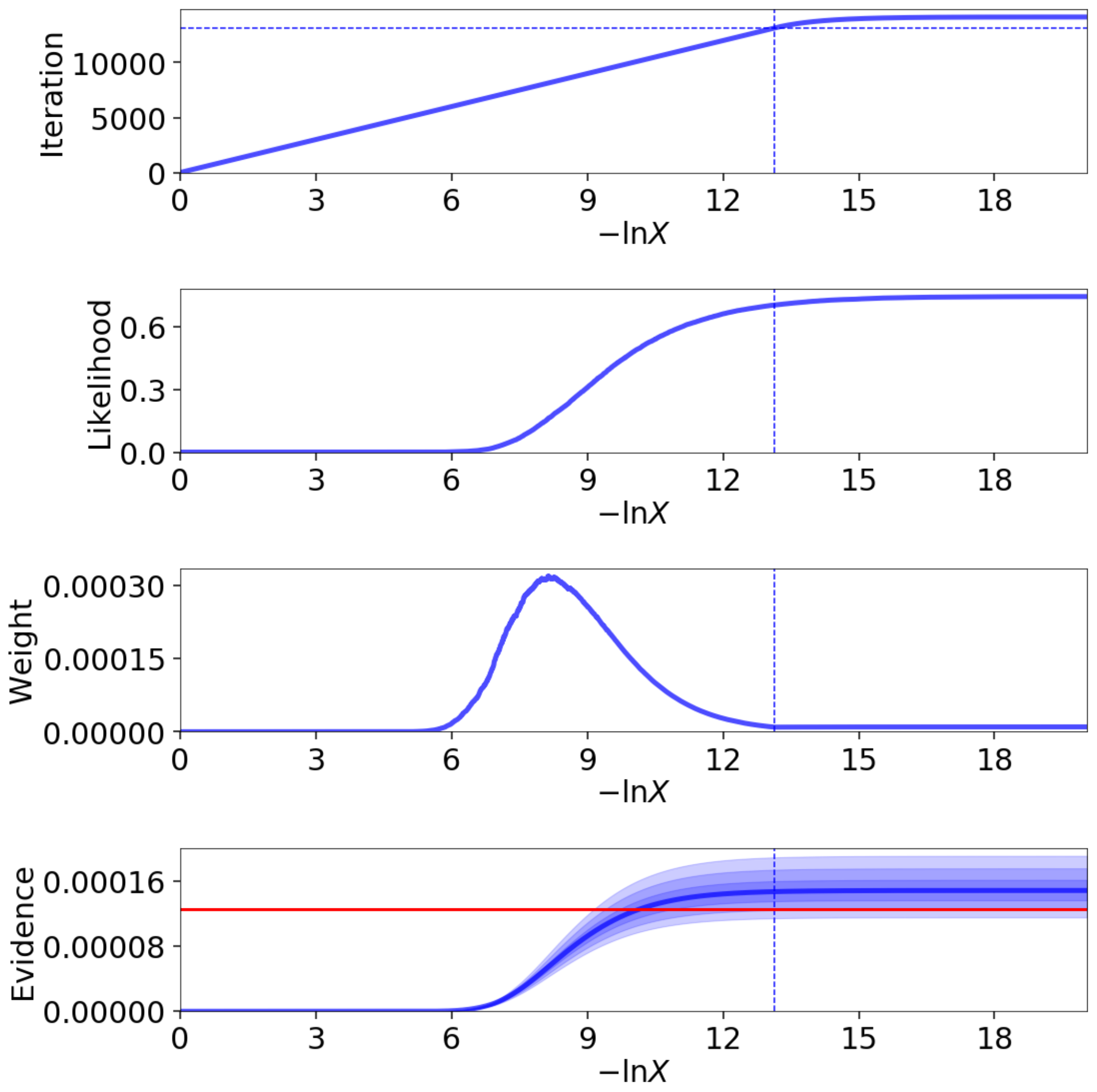


Example: Multivariate Normal (Evidence)

$$\mathbf{\Theta} = \{x_1, x_2, x_3\}$$

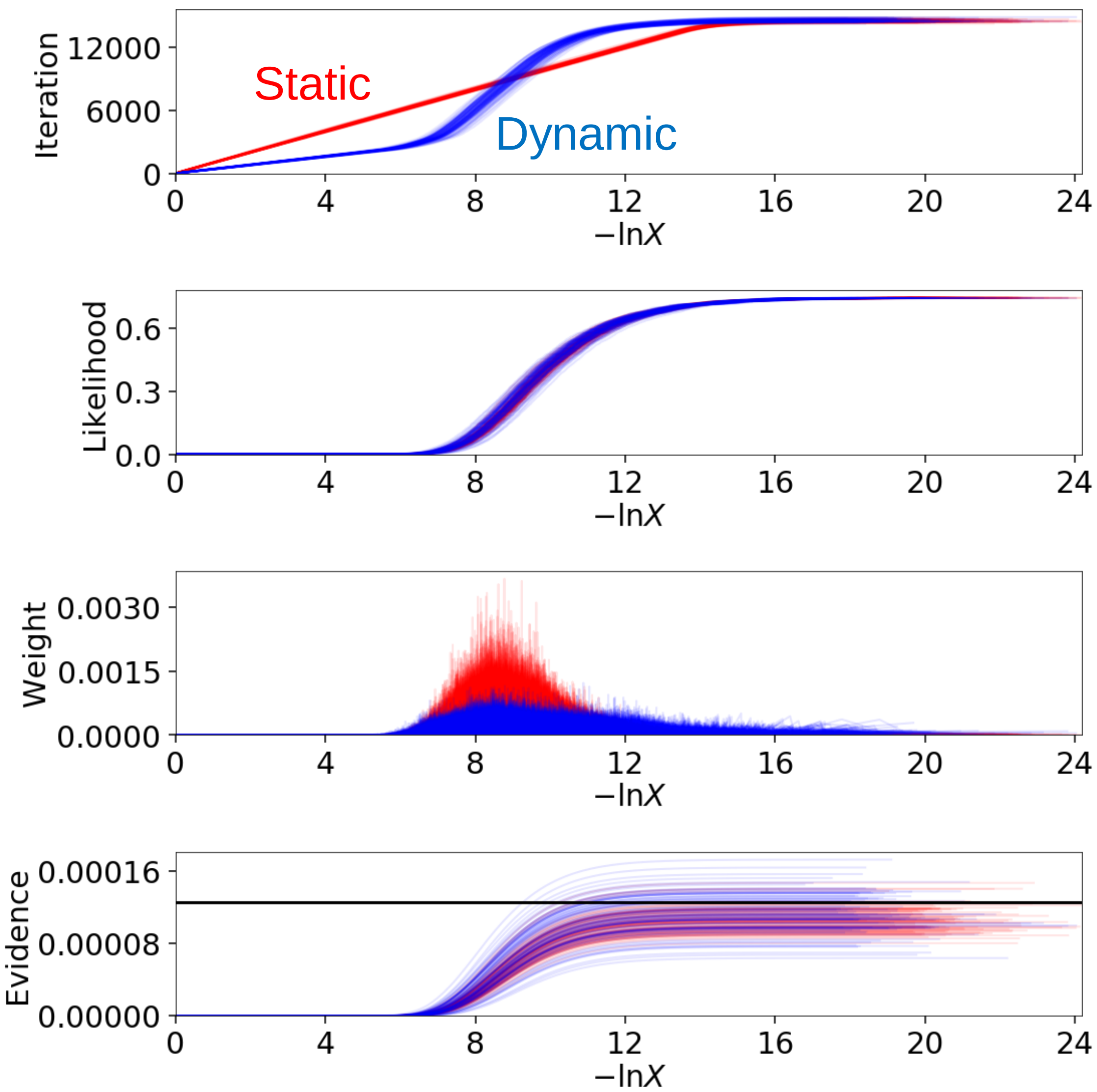


Example: Multivariate Normal (Evidence)



“Summary” Plot

Example: Multivariate Normal (Errors)



"Summary" Plot

Application:

Application:

- All results are preliminary but agree with results from MCMC methods (derived using `emcee`).

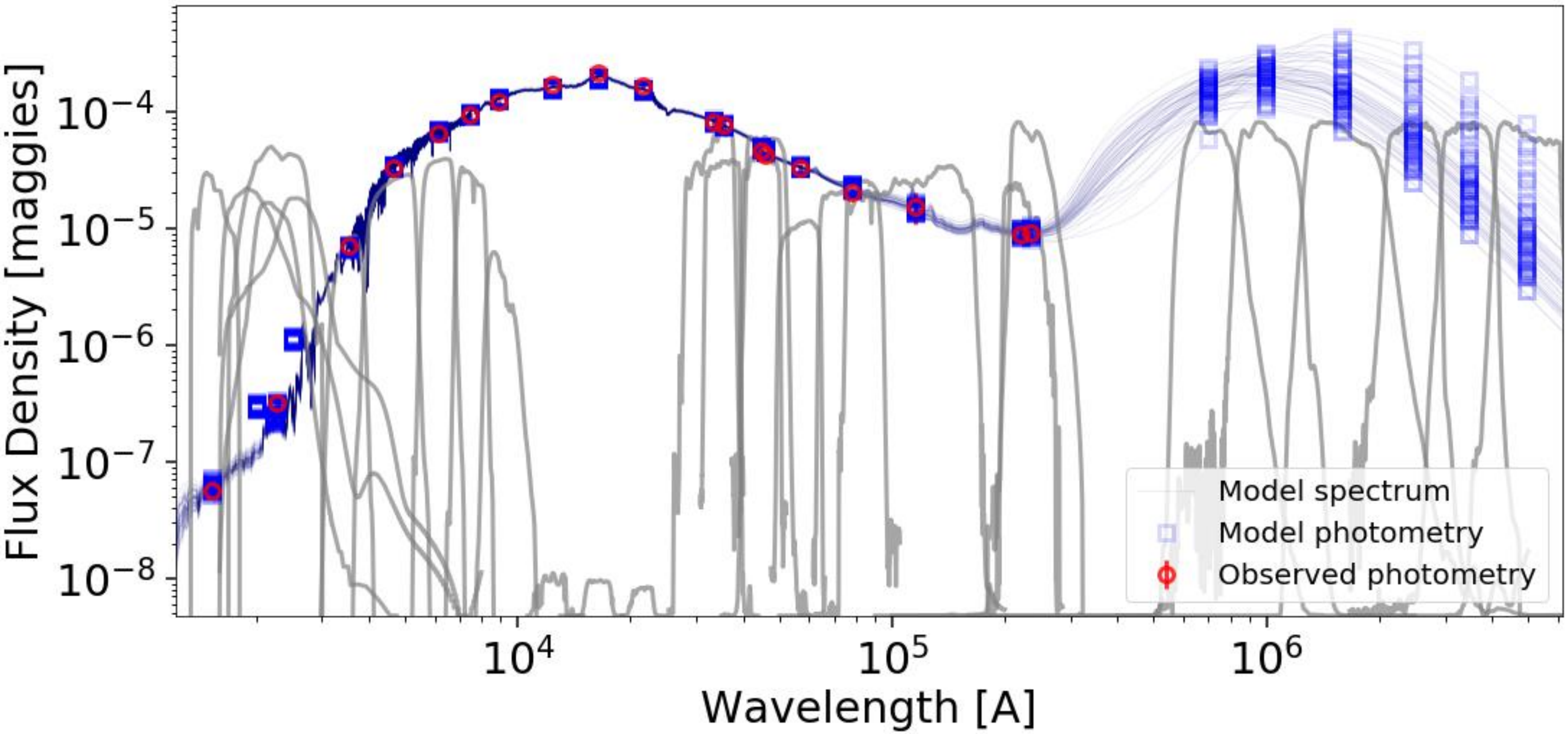
- Samples allocated with 100% posterior weight, automated stopping criterion (2% fractional error in simulated KLD).

- dynesty was substantially (~3-6x) more efficient at generating good samples than emcee, **before thinning**.

With: Joel Leja, Ben Johnson, Charlie Conroy

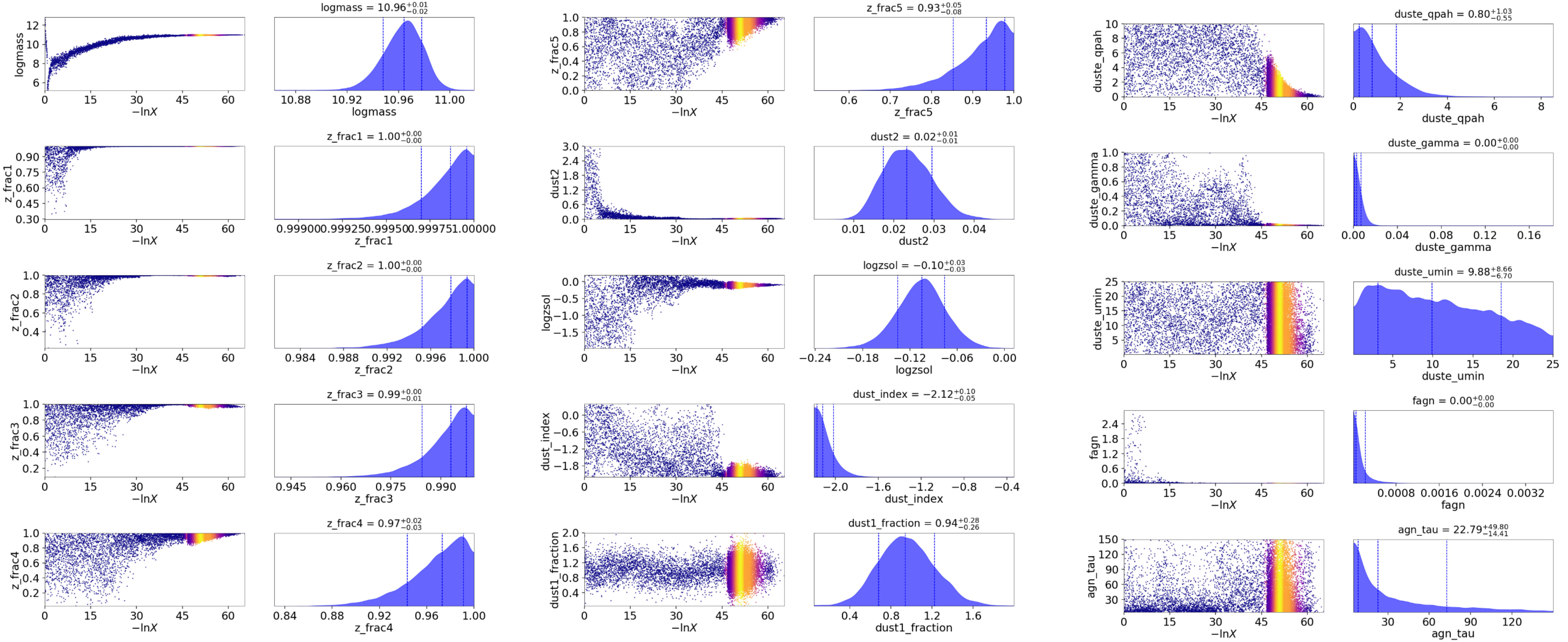
Application: Modeling Galaxy SEDs

$$\boldsymbol{\Theta} = \{\ln M_*, \ln Z, \boldsymbol{\sigma}_5, \boldsymbol{\delta}_6, \boldsymbol{\alpha}_2\}$$ D=15



With: Joel Leja, Ben Johnson, Charlie Conroy

Application: Modeling Galaxy SEDs



With: Joel Leja, Ben Johnson, Charlie Conroy

Application: Modeling Galaxy SEDs

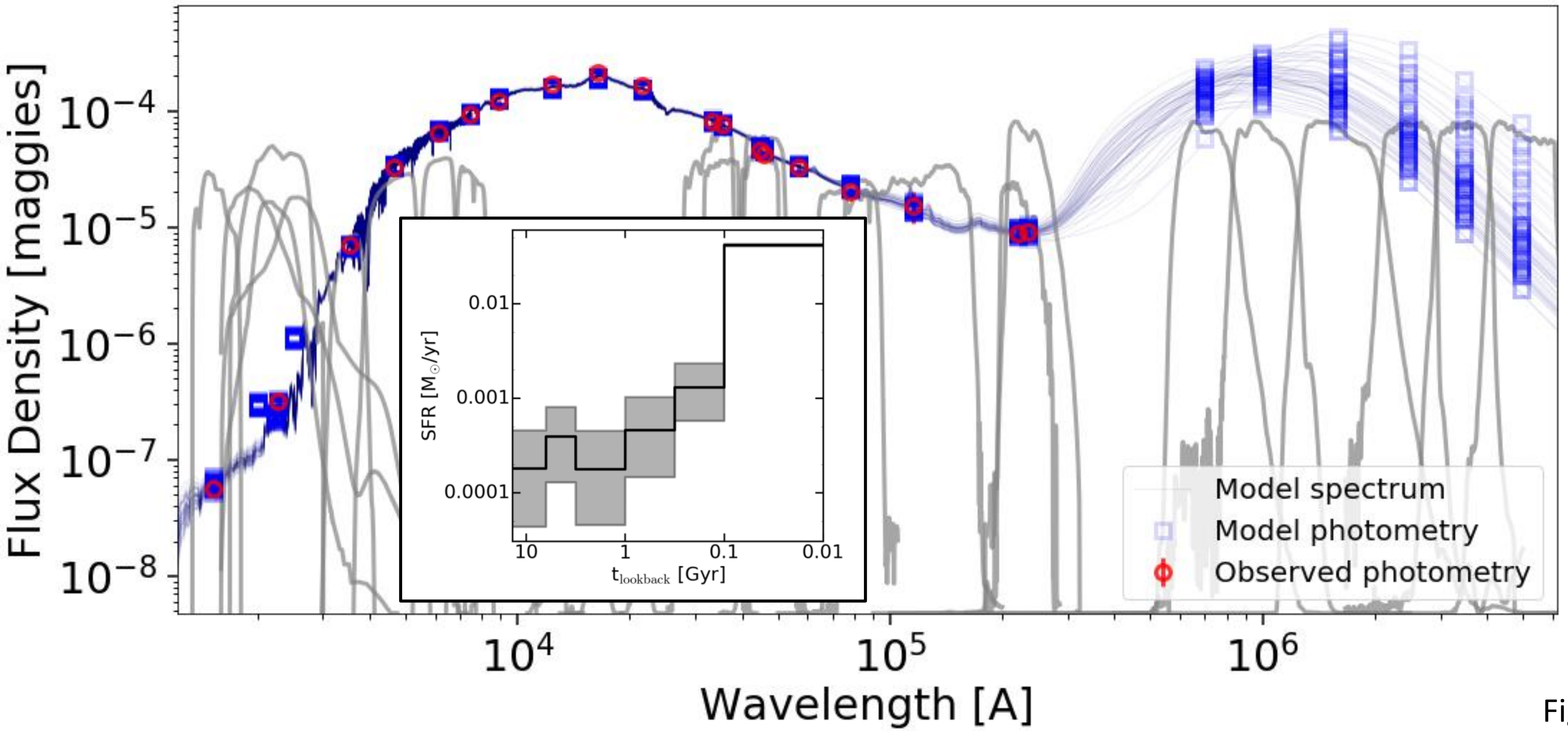
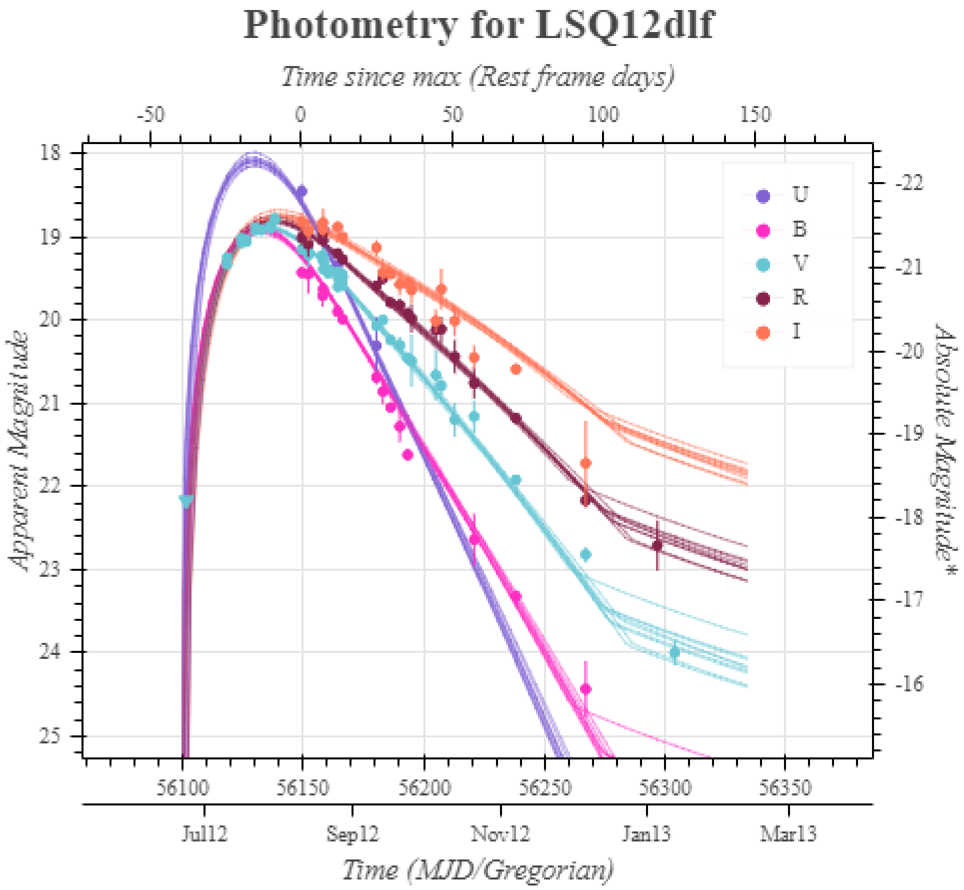


Fig: Joel Leja

With: James Guillochon, Kaisey Mandel

Application: Supernovae Light Curves

$$\boldsymbol{\Theta} = \{t_e, \boldsymbol{\rho}_4, \boldsymbol{\epsilon}_3, \boldsymbol{\delta}_3, \sigma^2\}$$ D=12



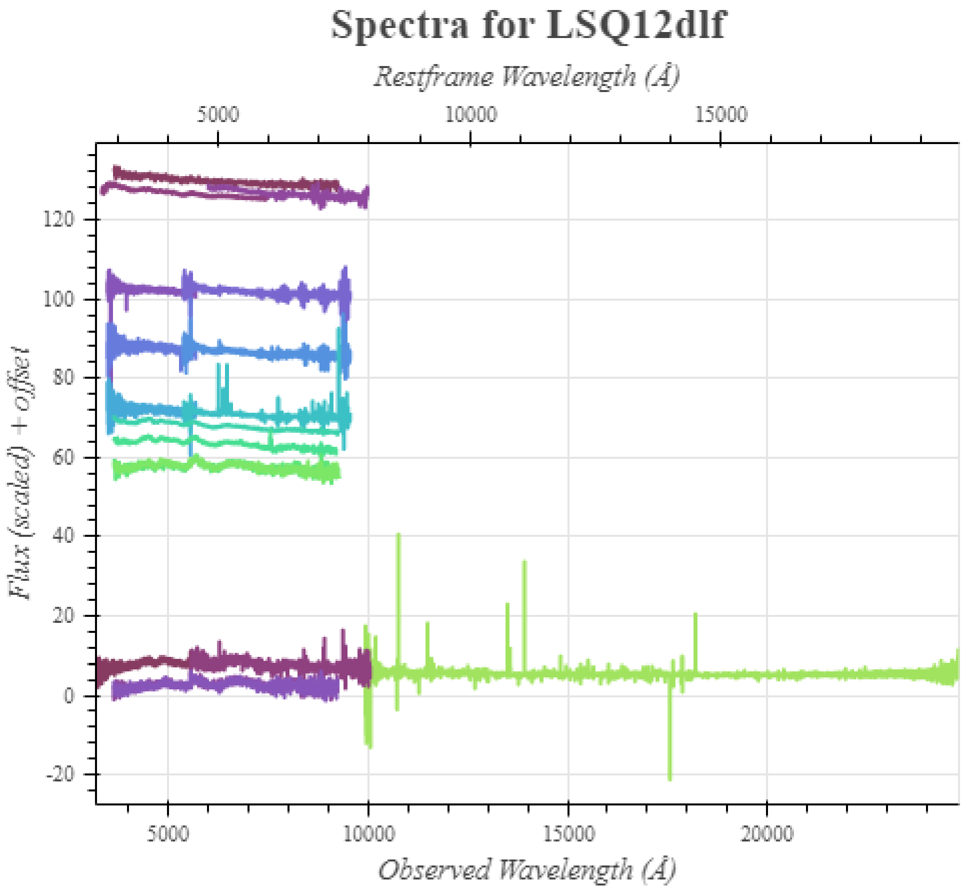
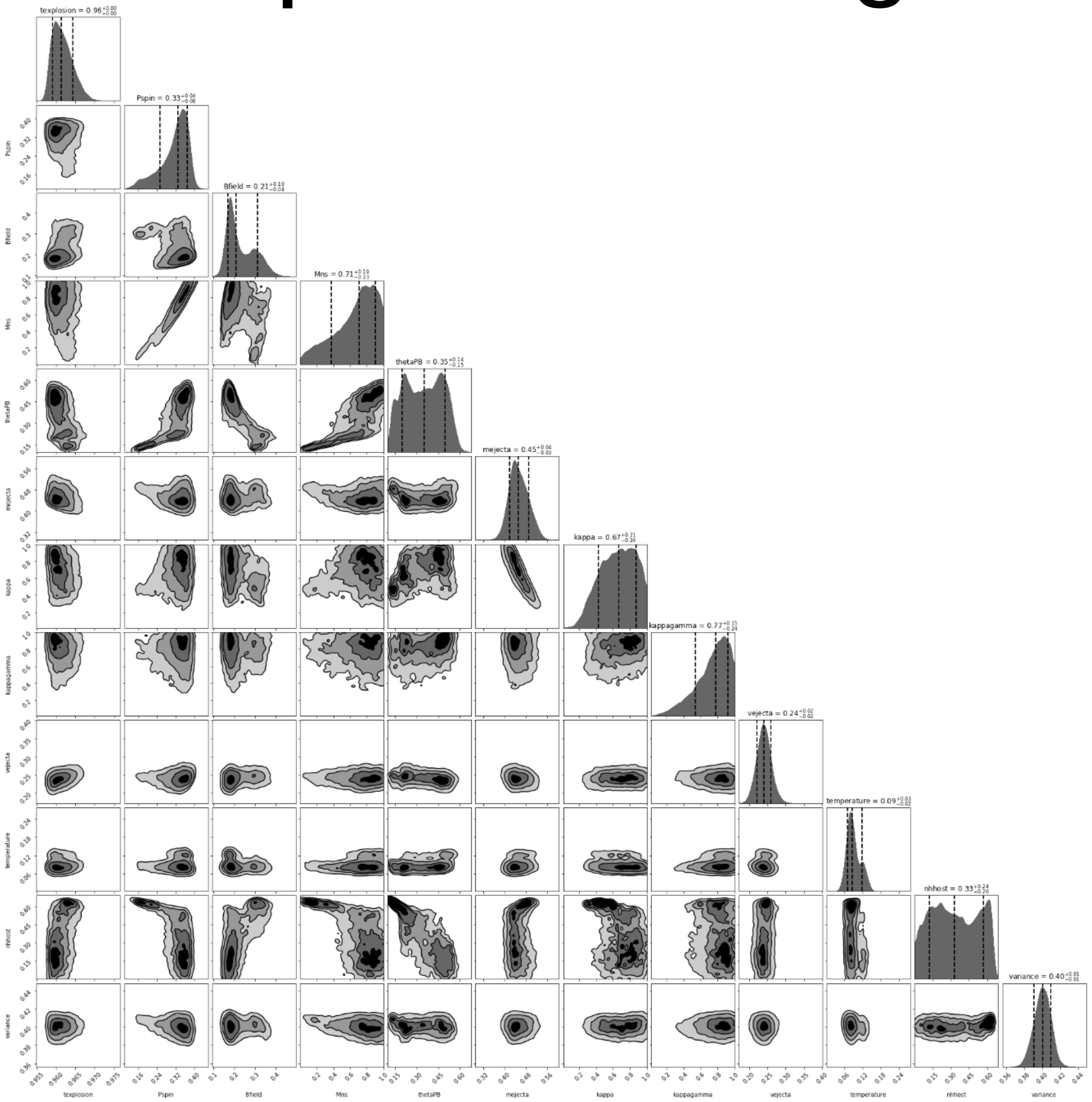


Fig: Open Supernova Catalog (LSQ12dlf), James Guillochon

With: James Guillochon, Kaisey Mandel

Application: Supernovae Light Curves

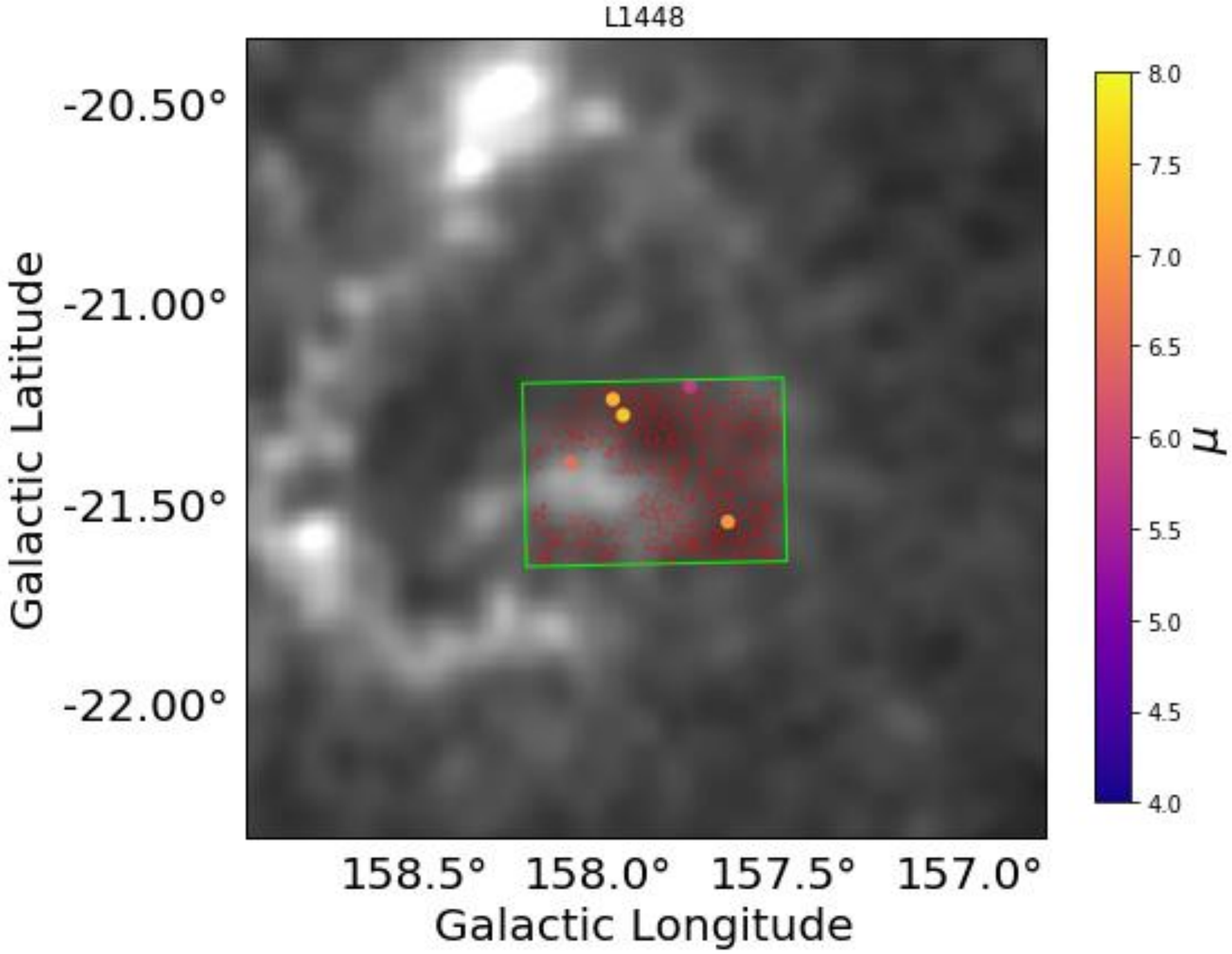


With: Catherine Zucker, Doug Finkbeiner, Alyssa Goodman

Application: Molecular Cloud Distances

$$\boldsymbol{\Theta} = \{\boldsymbol{f}_2, \boldsymbol{d}_5, \boldsymbol{c}_5, p_o\}$$

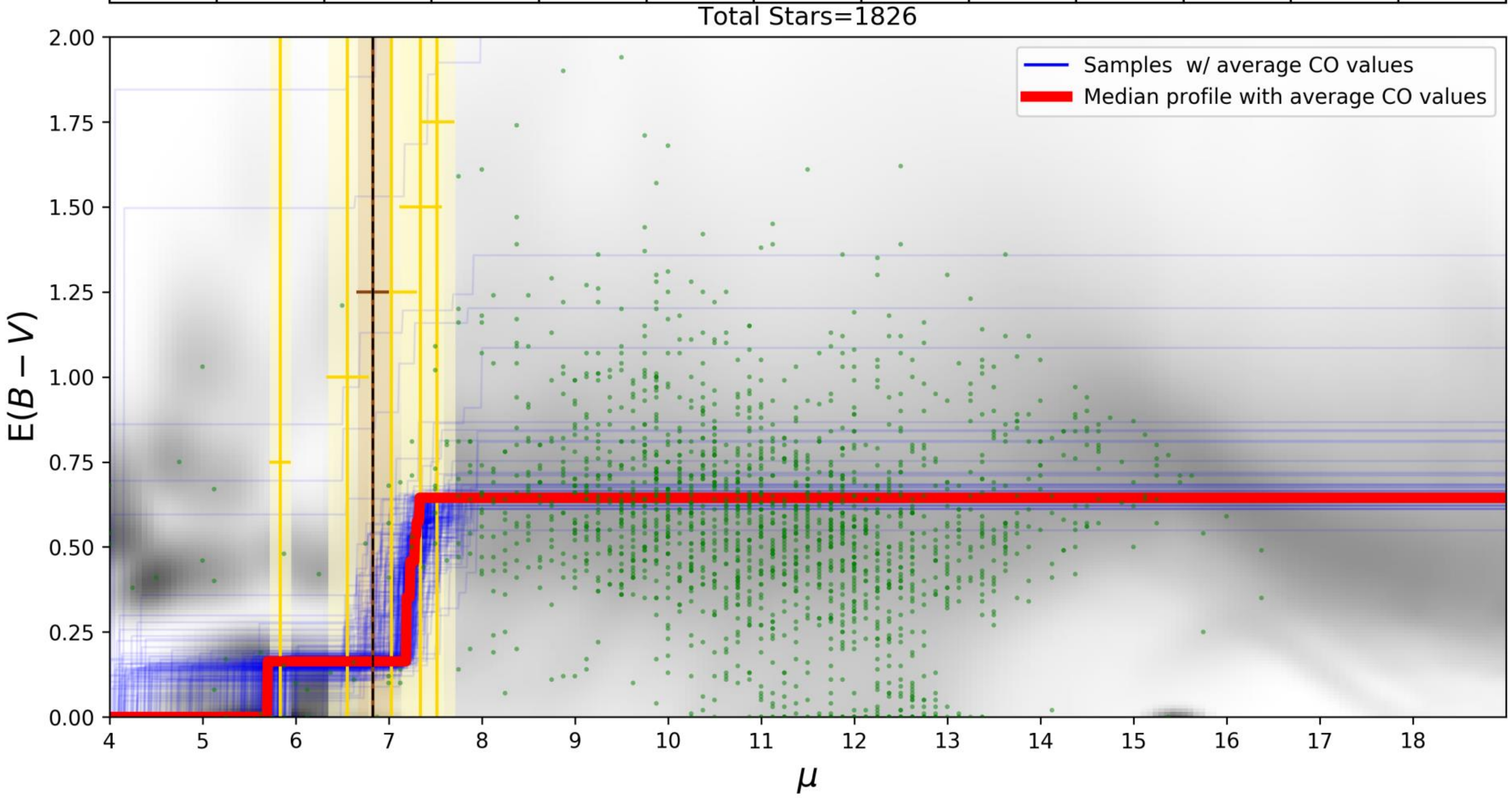
D=13



With: Catherine Zucker, Doug Finkbeiner, Alyssa Goodman

Application: Molecular Cloud Distances

dfore	rfore	P_b	d1	d2	d3	d4	d5	c1	c2	c3	c4	c5
$5.70^{+0.15}_{-0.17}$	$0.16^{+0.01}_{-0.01}$	$0.04^{+0.00}_{-0.00}$	$7.31^{+0.10}_{-0.11}$	$7.29^{+0.13}_{-0.12}$	$7.24^{+0.10}_{-0.08}$	$7.20^{+0.06}_{-0.07}$	$7.34^{+0.13}_{-0.13}$	$0.79^{+0.07}_{-0.07}$	$0.57^{+0.04}_{-0.05}$	$0.76^{+0.05}_{-0.06}$	$1.16^{+0.09}_{-0.08}$	$1.66^{+0.11}_{-0.10}$
			-1.2 km/s	0.8 km/s	2.8 km/s	4.9 km/s	6.9 km/s					



Dynamic Nested Sampling with **dynesty**

Beta release

- Pure Python.
- Easy to use.
- Modular.
- Open source.
- Parallelizable.
- Flexible bounding/sampling methods.
- Thorough documentation!

dynesty.readthedocs.io