PolyChord: Next Generation Nested Sampling

Sampling, Parameter Estimation and Bayesian Model Comparison

Will Handley
wh260@cam.ac.uk

Supervisors: Anthony Lasenby & Mike Hobson
Astrophysics Department
Cavendish Laboratory
University of Cambridge

December 11, 2015

Parameter estimation & model comparison

Metropolis Hastings

Nested Sampling

PolyChord

Applications

Notation

Notation

- Data: D

Notation

- Data: D
- Model: M

Notation

- Data: D
- Model: M
- Parameters: Θ

Notation

- Data: D
- Model: M
- Parameters: Θ
- Likelihood: $\mathrm{P}(D|\Theta, M) = \mathcal{L}(\Theta)$

Notation

- Data: D
- Model: M
- Parameters: Θ
- Likelihood: $\mathrm{P}(D|\Theta, M) = \mathcal{L}(\Theta)$
- Posterior: $\mathrm{P}(\Theta|D, M) = \mathcal{P}(\Theta)$

Notation

- Data: D
- Model: M
- Parameters: Θ
- Likelihood: $\mathrm{P}(D|\Theta, M) = \mathcal{L}(\Theta)$
- Posterior: $\mathrm{P}(\Theta|D, M) = \mathcal{P}(\Theta)$
- Prior: $\mathrm{P}(\Theta|M) = \pi(\Theta)$

Notation

- Data: D
- Model: M
- Parameters: Θ
- Likelihood: $\mathrm{P}(D|\Theta, M) = \mathcal{L}(\Theta)$
- Posterior: $\mathrm{P}(\Theta|D, M) = \mathcal{P}(\Theta)$
- Prior: $\mathrm{P}(\Theta|M) = \pi(\Theta)$
- Evidence: $\mathrm{P}(D|M) = \mathcal{Z}$

Bayes' theorem

Parameter estimation

Bayes' theorem

Parameter estimation

What does the data tell us about the params Θ of our model M?

Bayes' theorem

Parameter estimation

What does the data tell us about the params Θ of our model M?

Objective: Update our prior information $\pi(\Theta)$ in light of data D.

Bayes' theorem

Parameter estimation

What does the data tell us about the params Θ of our model M?

Objective: Update our prior information $\pi(\Theta)$ in light of data D.

$$\pi(\Theta) = \mathrm{P}(\Theta|M) \xrightarrow{D} \mathrm{P}(\Theta|D, M) = \mathcal{P}(\Theta)$$

Bayes' theorem
Parameter estimation

What does the data tell us about the params Θ of our model M?

Objective: Update our prior information $\pi(\Theta)$ in light of data D.

$$\pi(\Theta) = \mathrm{P}(\Theta|M) \xrightarrow{D} \mathrm{P}(\Theta|D, M) = \mathcal{P}(\Theta)$$

Solution: Use the likelihood $\mathcal{L}$ via Bayes' theorem:

Bayes' theorem
Parameter estimation

What does the data tell us about the params Θ of our model M?

Objective: Update our prior information $\pi(\Theta)$ in light of data D.

$$\pi(\Theta) = \mathrm{P}(\Theta|M) \xrightarrow{D} \mathrm{P}(\Theta|D,M) = \mathcal{P}(\Theta)$$

Solution: Use the likelihood $\mathcal{L}$ via Bayes' theorem:

$$\mathrm{P}(\Theta|D,M) = \frac{\mathrm{P}(D|\Theta,M)\mathrm{P}(\Theta|M)}{\mathrm{P}(D|M)}$$

Bayes' theorem

Parameter estimation

What does the data tell us about the params Θ of our model M?

Objective: Update our prior information $\pi(\Theta)$ in light of data D.

$$\pi(\Theta) = \mathrm{P}(\Theta|M) \xrightarrow{D} \mathrm{P}(\Theta|D, M) = \mathcal{P}(\Theta)$$

Solution: Use the likelihood $\mathcal{L}$ via Bayes' theorem:

$$\mathrm{P}(\Theta|D, M) = \frac{\mathrm{P}(D|\Theta, M)\mathrm{P}(\Theta|M)}{\mathrm{P}(D|M)}$$

$$\text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Evidence}}$$

Bayes' theorem

Model comparison

Bayes' theorem

Model comparison

What does the data tell us about our model M_i in relation to other models $\{M_1, M_2, \cdots\}$?

Bayes' theorem

Model comparison

What does the data tell us about our model M_i in relation to other models $\{M_1, M_2, \cdots\}$?

$$\mathrm{P}(M_i) \xrightarrow{D} \mathrm{P}(M_i|D)$$

Bayes' theorem

Model comparison

What does the data tell us about our model M_i in relation to other models $\{M_1, M_2, \cdots\}$?

$$\mathrm{P}(M_i) \xrightarrow{D} \mathrm{P}(M_i|D)$$

$$\mathrm{P}(M_i|D) = \frac{\mathrm{P}(D|M_i)\mathrm{P}(M_i)}{\mathrm{P}(D)}$$

Bayes' theorem

Model comparison

What does the data tell us about our model M_i in relation to other models $\{M_1, M_2, \cdots\}$?

$$\mathrm{P}(M_i) \xrightarrow{D} \mathrm{P}(M_i|D)$$

$$\mathrm{P}(M_i|D) = \frac{\mathrm{P}(D|M_i)\mathrm{P}(M_i)}{\mathrm{P}(D)}$$

$$\mathrm{P}(D|M_i) = \mathcal{Z}_i = \text{Evidence of } M_i$$

Parameter estimation & model comparison

The challenge

Parameter estimation & model comparison

The challenge

Parameter estimation: what does the data tell us about a model?
(Computing posteriors)

Parameter estimation & model comparison

The challenge

Parameter estimation: what does the data tell us about a model?
(Computing posteriors)

Model comparison: what does the data tell us about all models?
(Computing evidences)

Parameter estimation & model comparison

The challenge

Parameter estimation: what does the data tell us about a model?
(Computing posteriors)

Model comparison: what does the data tell us about all models?
(Computing evidences)

Both of these are challenging things to compute.

Parameter estimation & model comparison

The challenge

Parameter estimation: what does the data tell us about a model?
(Computing posteriors)

Model comparison: what does the data tell us about all models?
(Computing evidences)

Both of these are challenging things to compute.

- Markov-Chain Monte-Carlo (MCMC) can solve the first of these (kind of)

Parameter estimation & model comparison

The challenge

Parameter estimation: what does the data tell us about a model?
(Computing posteriors)

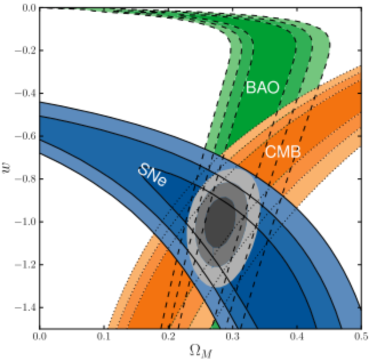
Model comparison: what does the data tell us about all models?
(Computing evidences)

Both of these are challenging things to compute.

- Markov-Chain Monte-Carlo (MCMC) can solve the first of these (kind of)
- Nested sampling (NS) promises to solve both simultaneously.

Parameter estimation & model comparison

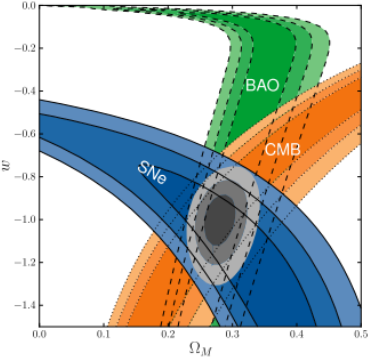
Why is it difficult?



Parameter estimation & model comparison

Why is it difficult?

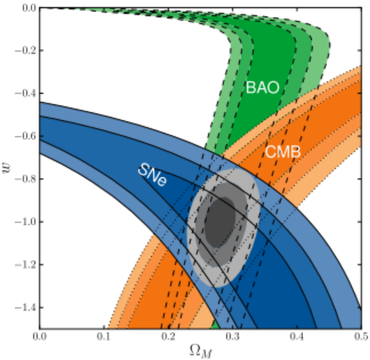
1. In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.



Parameter estimation & model comparison

Why is it difficult?

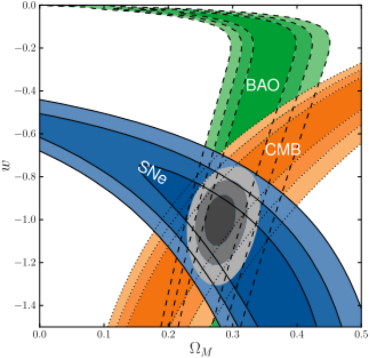
1. In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.
2. Worse, you don't know where this region is.



Parameter estimation & model comparison

Why is it difficult?

1. In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.
2. Worse, you don't know where this region is.

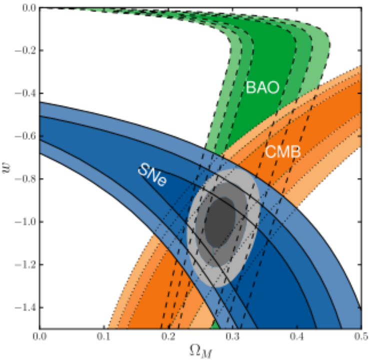


- Describing an N-dimensional posterior fully is impossible.

Parameter estimation & model comparison

Why is it difficult?

1. In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.
2. Worse, you don't know where this region is.

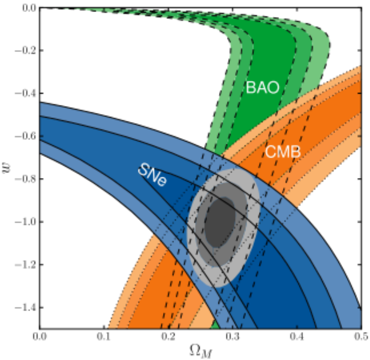


- Describing an N-dimensional posterior fully is impossible.
- Project/marginalise into 2- or 3-dimensions at best

Parameter estimation & model comparison

Why is it difficult?

1. In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.
2. Worse, you don't know where this region is.



- Describing an *N*-dimensional posterior fully is impossible.
- Project/marginalise into 2- or 3-dimensions at best
- *Sampling* the posterior is an excellent compression scheme.

Markov-Chain Monte-Carlo (MCMC)

Metropolis-Hastings, Gibbs, Hamiltonian...

Markov-Chain Monte-Carlo (MCMC)

Metropolis-Hastings, Gibbs, Hamiltonian...

- Turn the N-dimensional problem into a one-dimensional one.

Markov-Chain Monte-Carlo (MCMC)

Metropolis-Hastings, Gibbs, Hamiltonian...

- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.

Markov-Chain Monte-Carlo (MCMC)

Metropolis-Hastings, Gibbs, Hamiltonian...

- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.
    1. Pick random direction

Markov-Chain Monte-Carlo (MCMC)

Metropolis-Hastings, Gibbs, Hamiltonian...

- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.
  1. Pick random direction
  2. Choose step length

Markov-Chain Monte-Carlo (MCMC)

Metropolis-Hastings, Gibbs, Hamiltonian...

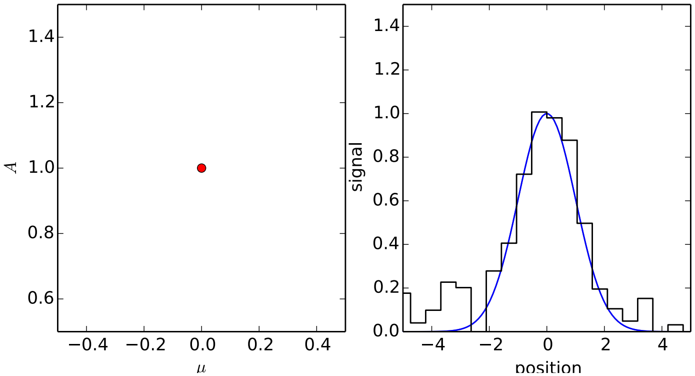
- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.
    1. Pick random direction
    2. Choose step length
    3. If uphill, make step...

Markov-Chain Monte-Carlo (MCMC)

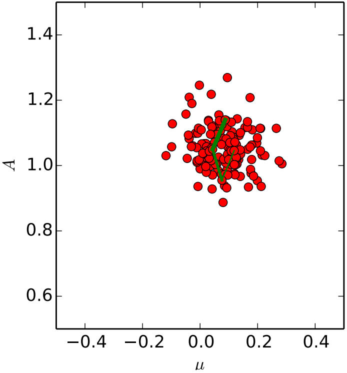
Metropolis-Hastings, Gibbs, Hamiltonian...

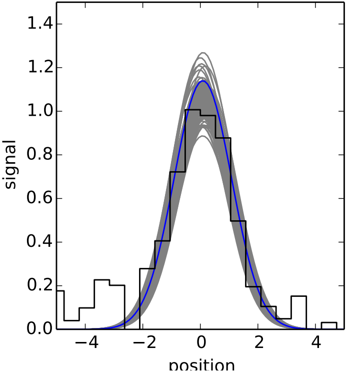
- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.
  1. Pick random direction
  2. Choose step length
  3. If uphill, make step...
  4. ...otherwise sometimes make step.

MCMC in action



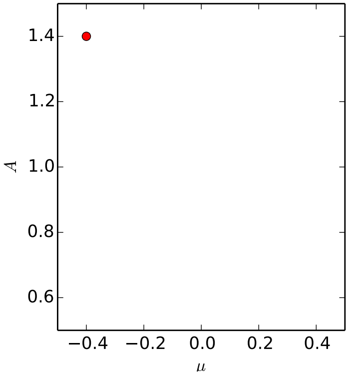
MCMC in action

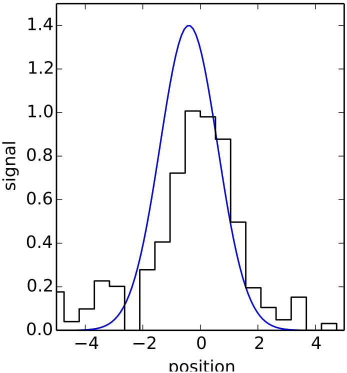




When MCMC fails

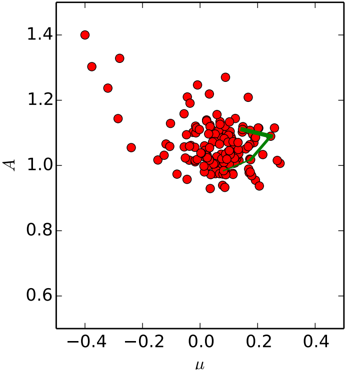
Burn in

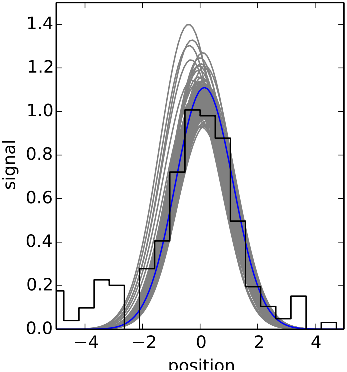




When MCMC fails

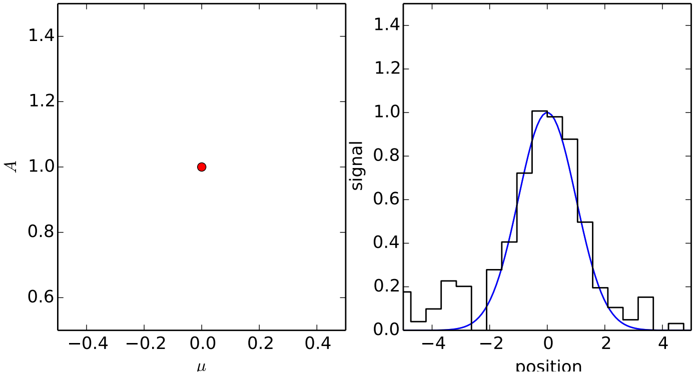
Burn in





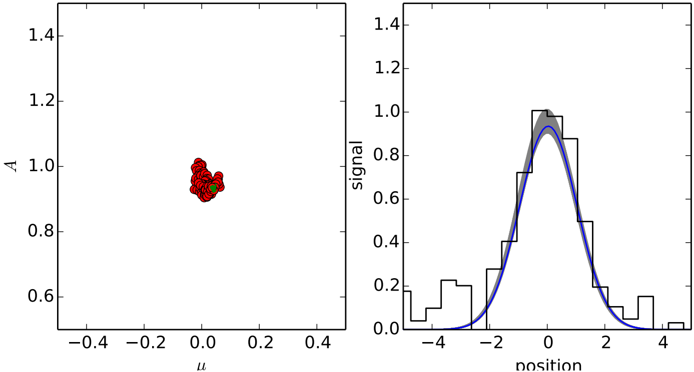
When MCMC fails

Tuning the proposal distribution



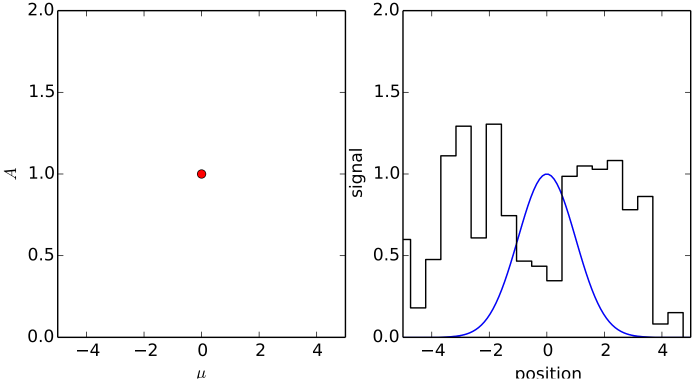
When MCMC fails

Tuning the proposal distribution



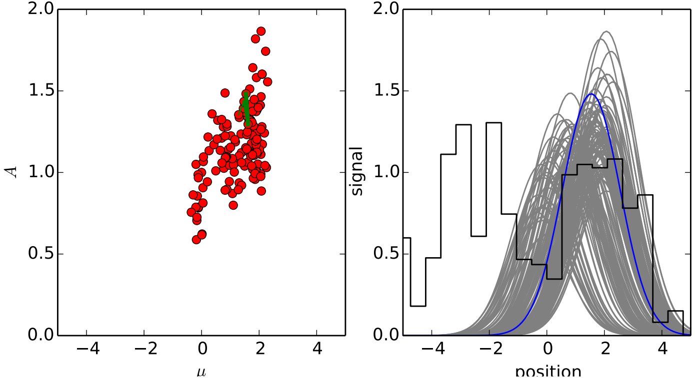
When MCMC fails

Multimodality



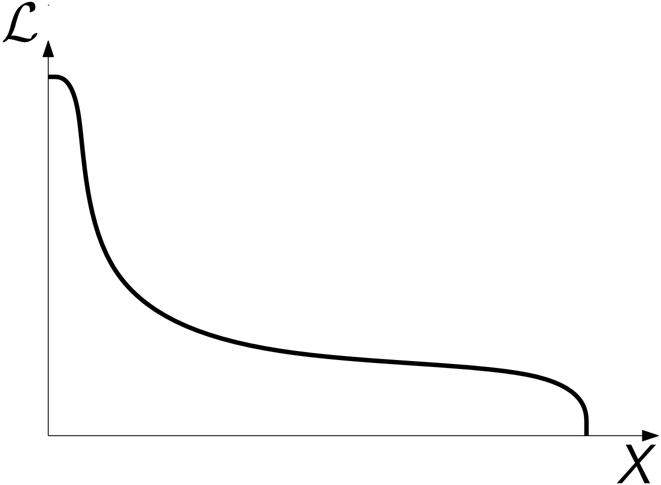
When MCMC fails

Multimodality



When MCMC fails

Phase transitions



When MCMC fails

The real reason. . .

When MCMC fails

The real reason. . .

- MCMC does not give you evidences!

When MCMC fails

The real reason...

- MCMC does not give you evidences!

$$\mathcal{Z} = \mathrm{P}(D|M)$$

When MCMC fails

The real reason...

- MCMC does not give you evidences!

$$\begin{aligned}\mathcal{Z} &= \mathrm{P}(D|M) \\ &= \int \mathrm{P}(D|\Theta, M)\mathrm{P}(\Theta|M) d\Theta\end{aligned}$$

When MCMC fails

The real reason...

- MCMC does not give you evidences!

$$
\begin{aligned}
\mathcal{Z} &= \mathrm{P}(D|M) \\
&= \int \mathrm{P}(D|\Theta, M)\mathrm{P}(\Theta|M) d\Theta \\
&= \int \mathcal{L}(\Theta)\pi(\Theta) d\Theta
\end{aligned}
$$

When MCMC fails

The real reason...

- MCMC does not give you evidences!

$$
\begin{aligned}
\mathcal{Z} &= \mathrm{P}(D|M) \\
&= \int \mathrm{P}(D|\Theta, M)\mathrm{P}(\Theta|M) d\Theta \\
&= \int \mathcal{L}(\Theta)\pi(\Theta) d\Theta \\
&= \langle \mathcal{L} \rangle_\pi
\end{aligned}
$$

When MCMC fails

The real reason...

- MCMC does not give you evidences!

$$
\begin{aligned}
\mathcal{Z} &= \mathrm{P}(D|M) \\
&= \int \mathrm{P}(D|\Theta, M)\mathrm{P}(\Theta|M) d\Theta \\
&= \int \mathcal{L}(\Theta)\pi(\Theta) d\Theta \\
&= \langle \mathcal{L} \rangle_\pi
\end{aligned}
$$

- MCMC fundamentally explores the posterior, and cannot average over the prior.

Nested Sampling

John Skilling's alternative to MCMC!

Nested Sampling

John Skilling's alternative to MCMC!

New procedure:

Nested Sampling

John Skilling's alternative to MCMC!

New procedure:
Maintain a set S of n samples, which are sequentially updated:

Nested Sampling

John Skilling's alternative to MCMC!

New procedure:

Maintain a set S of n samples, which are sequentially updated:

S_0: Generate n samples from the prior π.

Nested Sampling

John Skilling's alternative to MCMC!

New procedure:

Maintain a set S of n samples, which are sequentially updated:

S_0: Generate n samples from the prior π.

S_{n+1}: Delete the lowest likelihood sample in S_n, and replace it with a new sample with higher likelihood

Nested Sampling

John Skilling's alternative to MCMC!

New procedure:

Maintain a set S of n samples, which are sequentially updated:

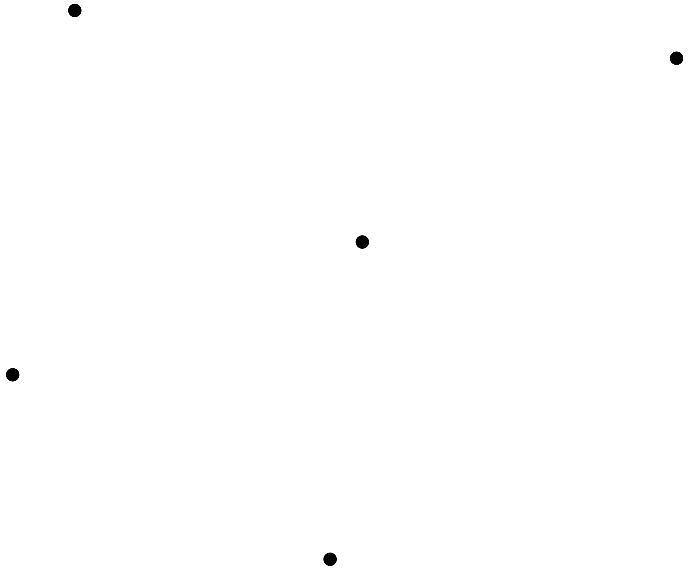
S_0: Generate n samples from the prior π.

S_{n+1}: Delete the lowest likelihood sample in S_n, and replace it with a new sample with higher likelihood

Requires one to be able to sample from the prior, subject to a *hard likelihood constraint*.

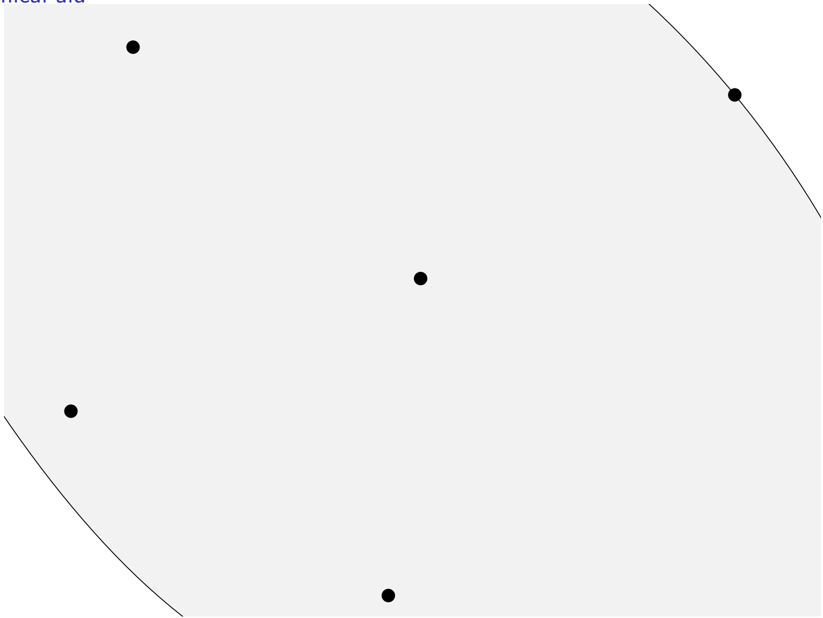
Nested Sampling

Graphical aid



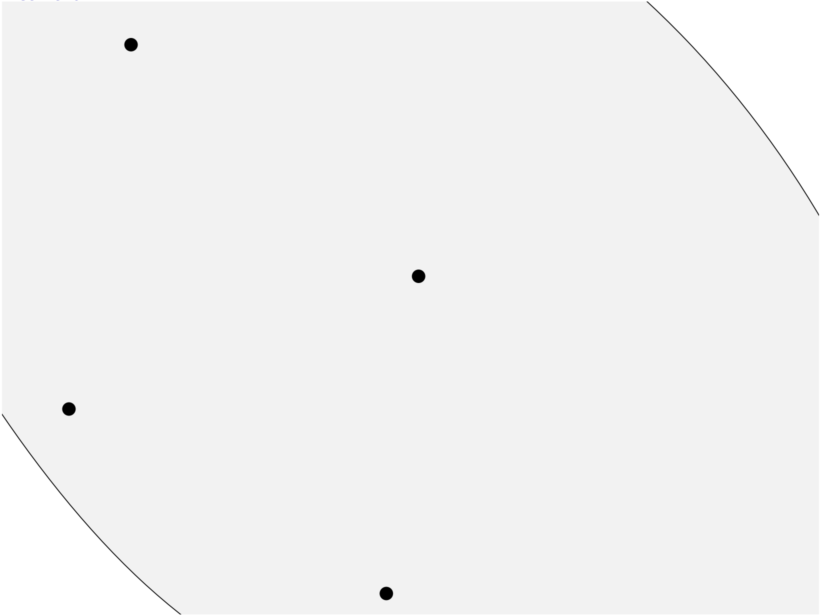
Nested Sampling

Graphical aid



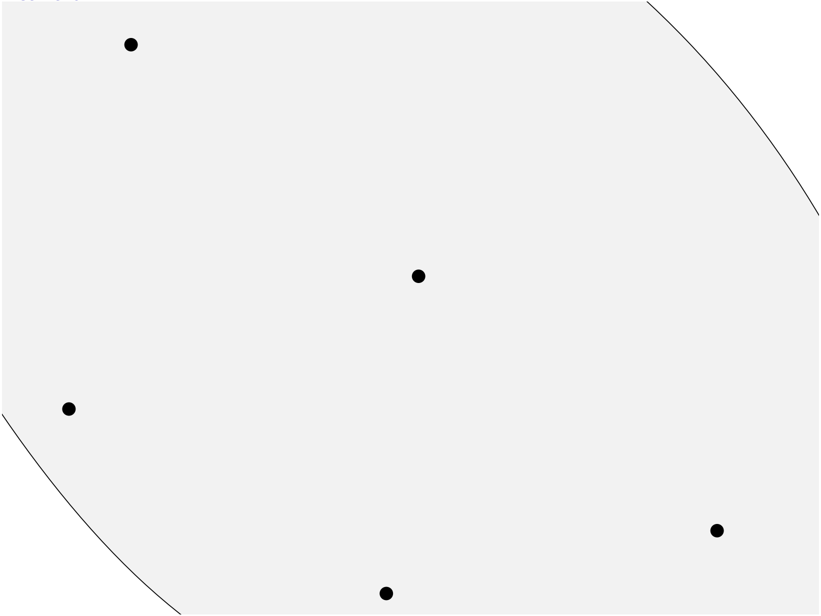
Nested Sampling

Graphical aid



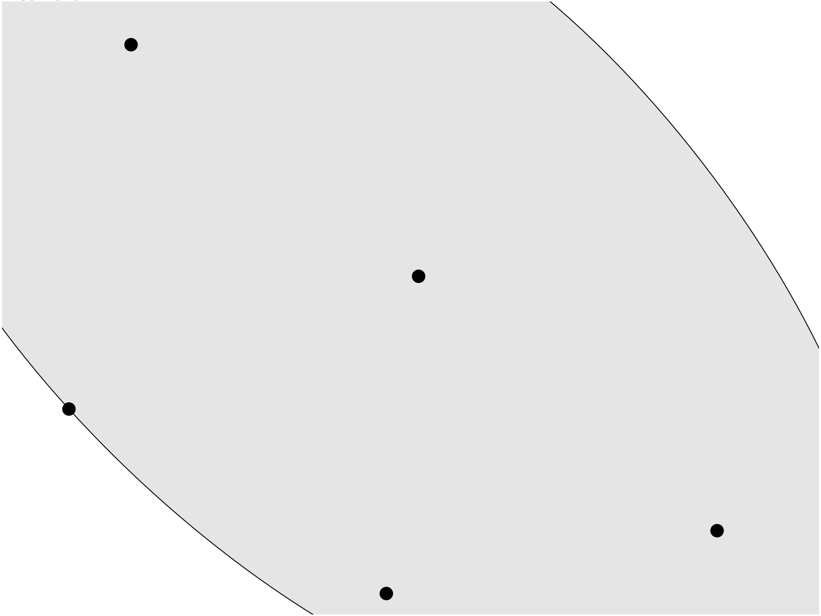
Nested Sampling

Graphical aid



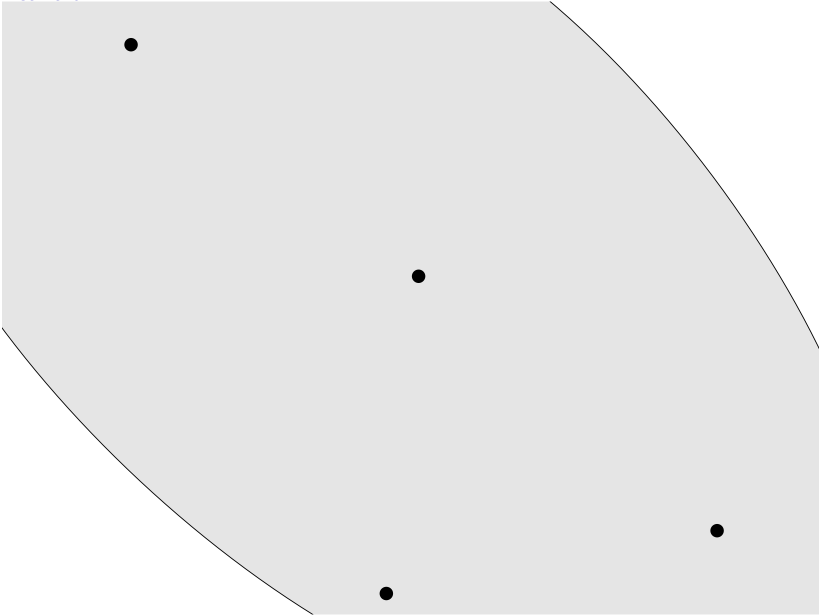
Nested Sampling

Graphical aid



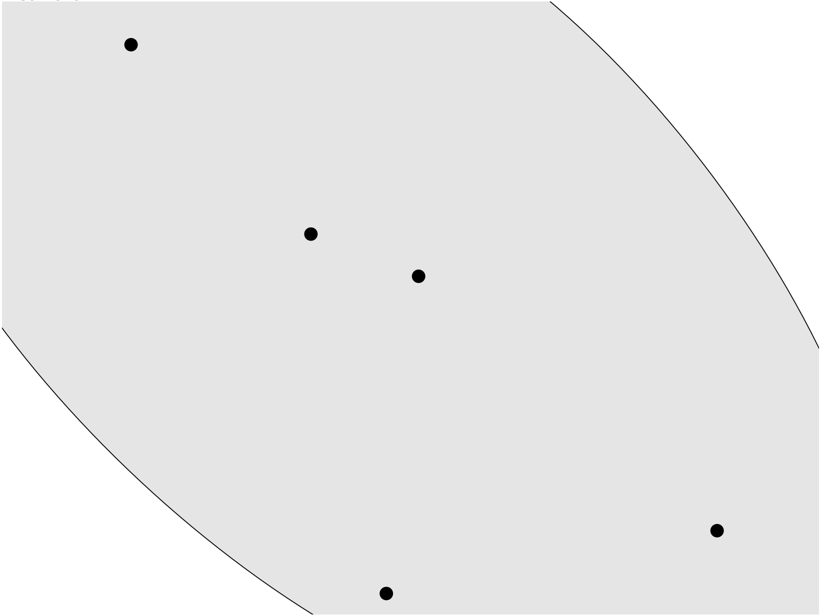
Nested Sampling

Graphical aid



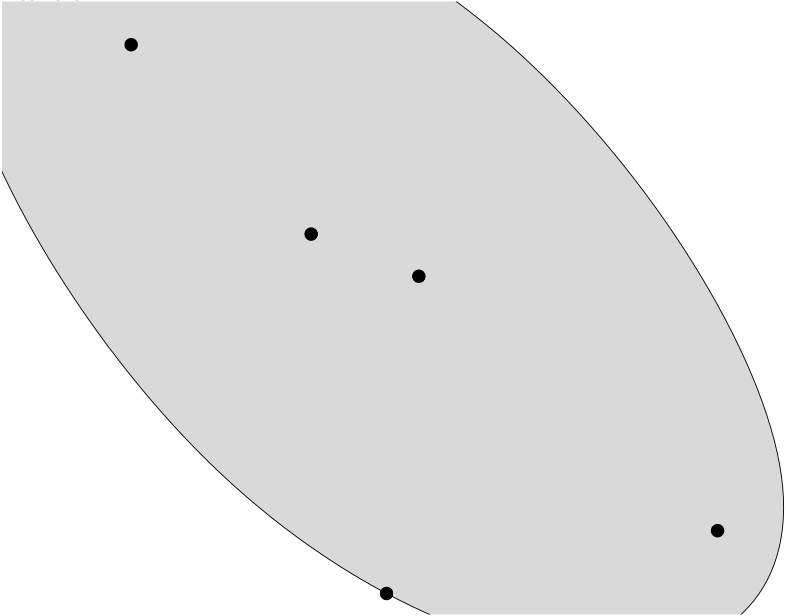
Nested Sampling

Graphical aid



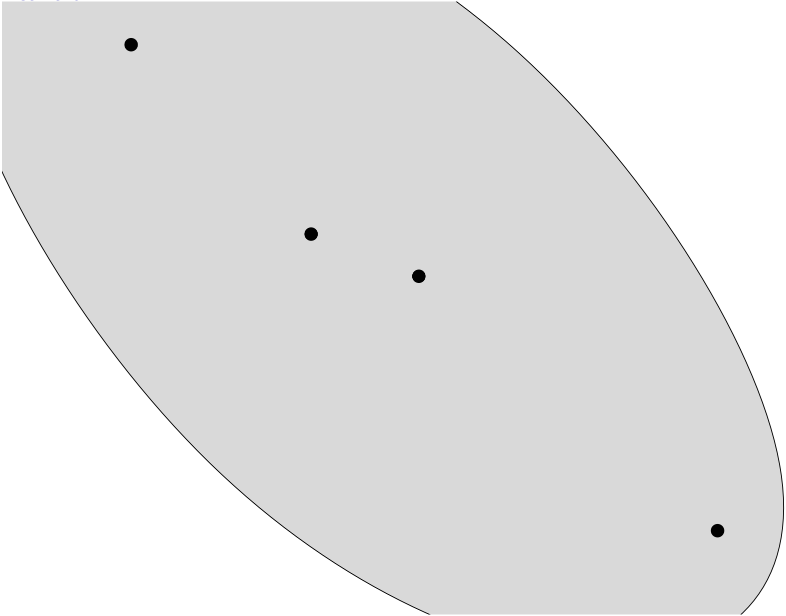
Nested Sampling

Graphical aid



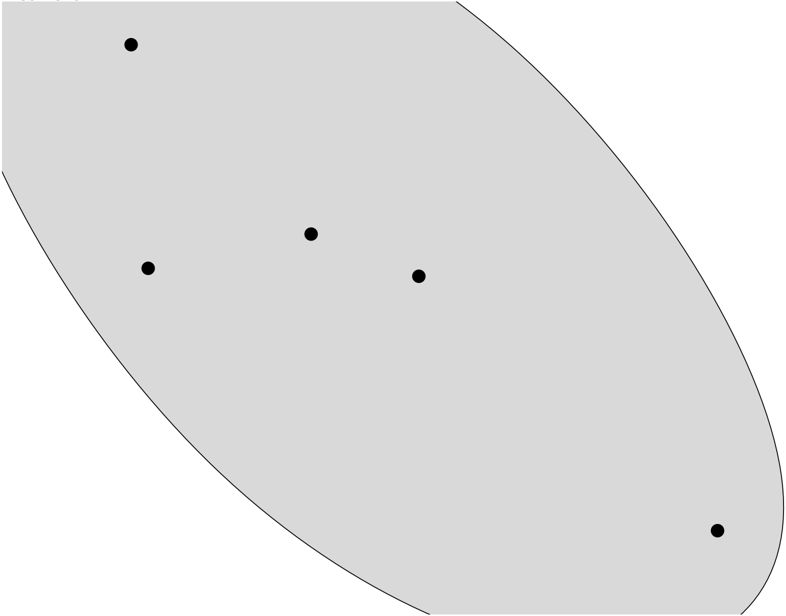
Nested Sampling

Graphical aid



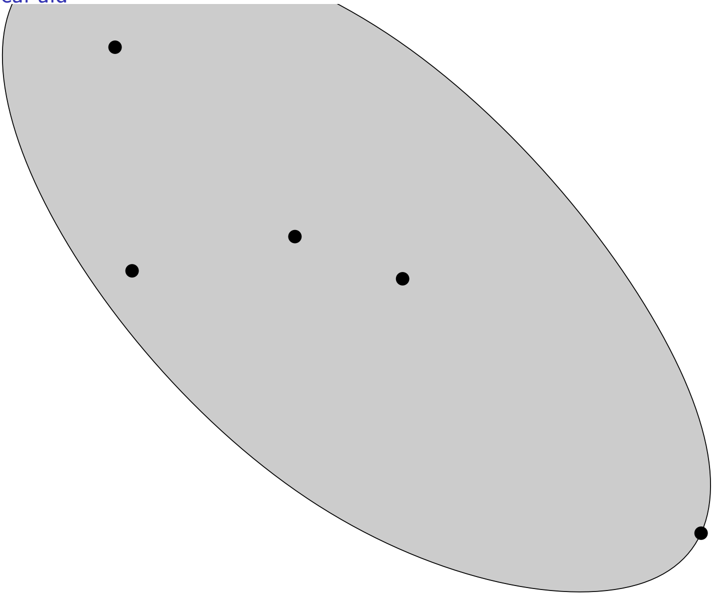
Nested Sampling

Graphical aid



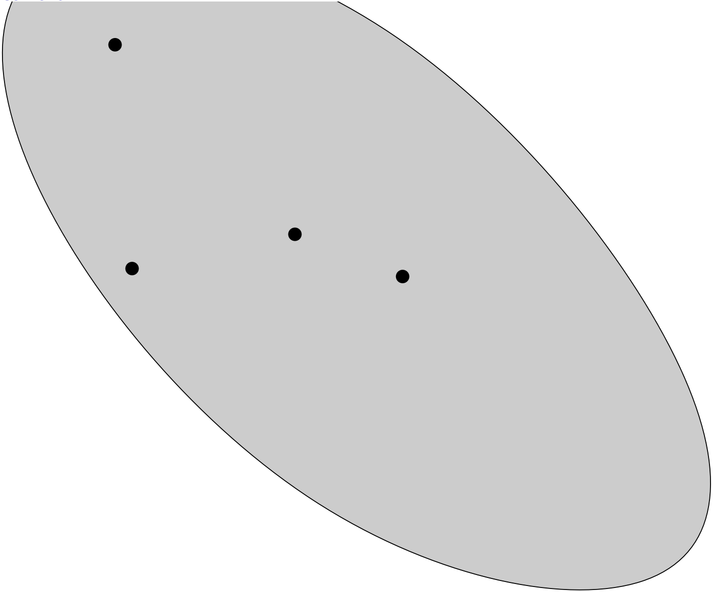
Nested Sampling

Graphical aid



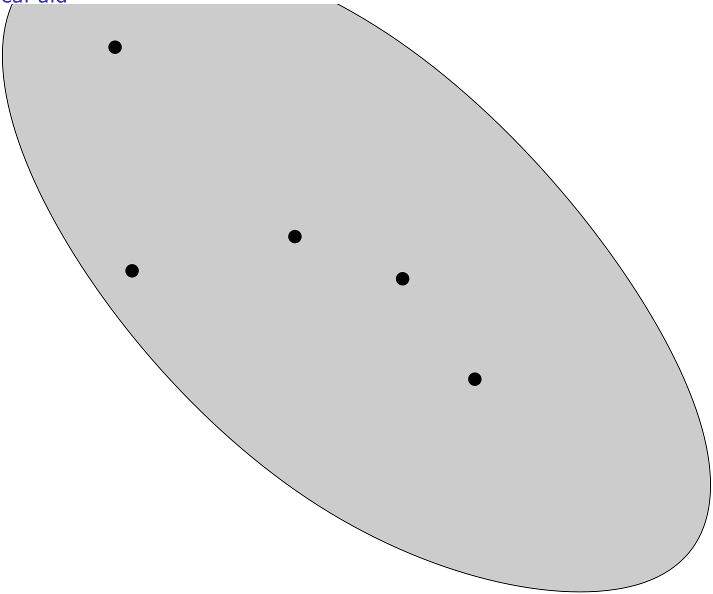
Nested Sampling

Graphical aid



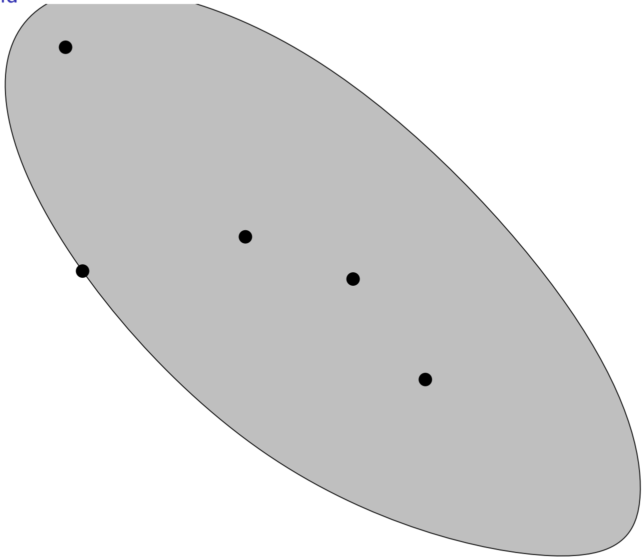
Nested Sampling

Graphical aid



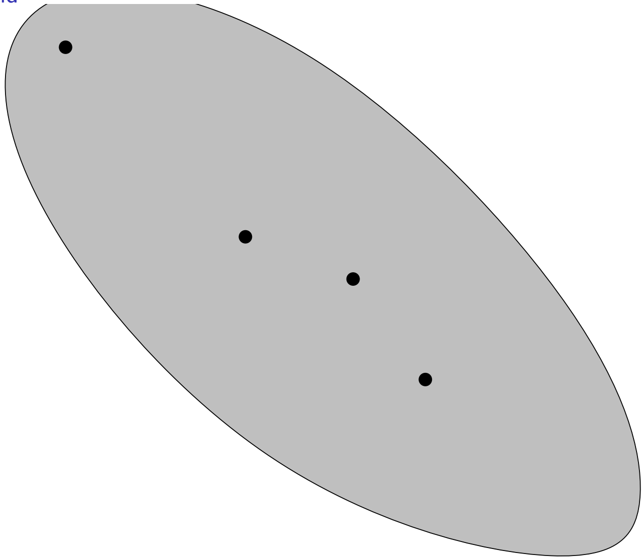
Nested Sampling

Graphical aid



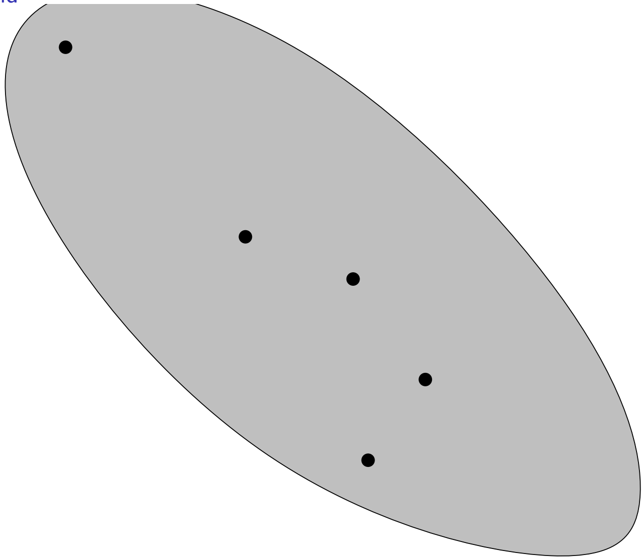
Nested Sampling

Graphical aid



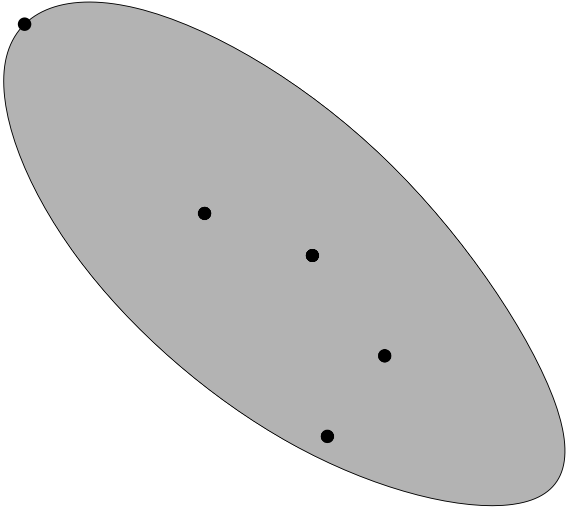
Nested Sampling

Graphical aid



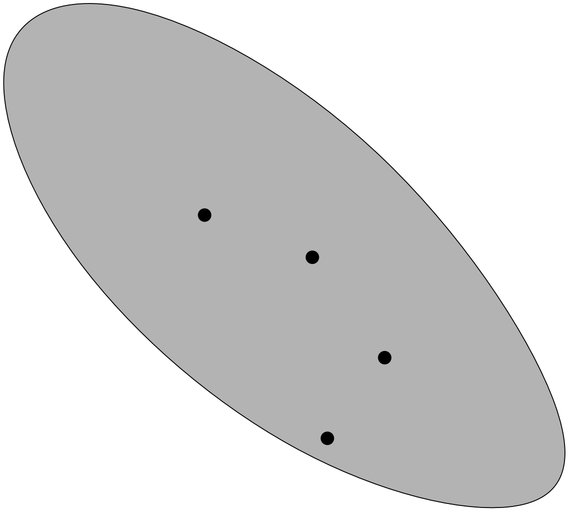
Nested Sampling

Graphical aid



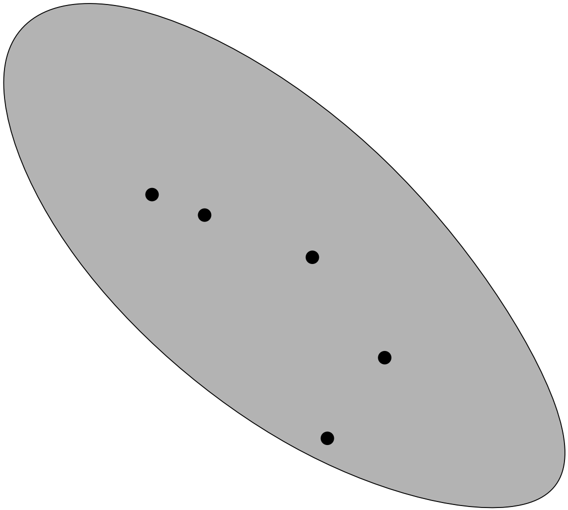
Nested Sampling

Graphical aid



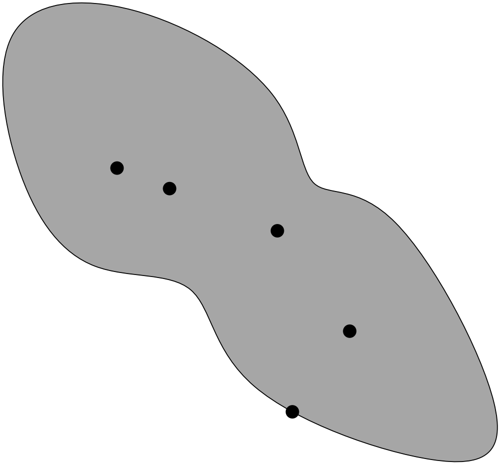
Nested Sampling

Graphical aid



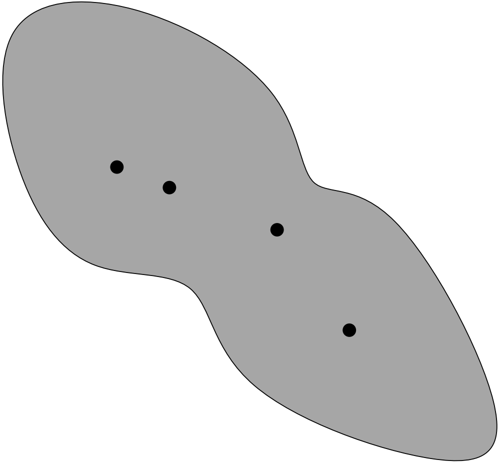
Nested Sampling

Graphical aid



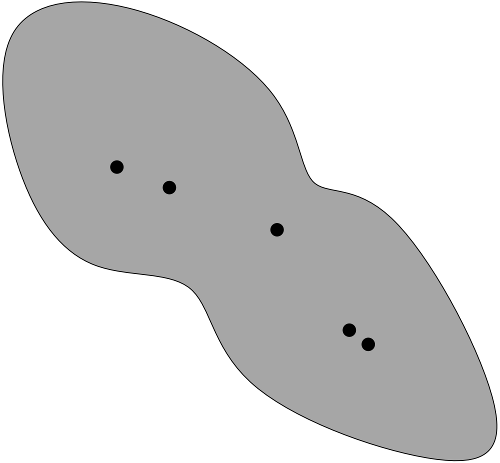
Nested Sampling

Graphical aid



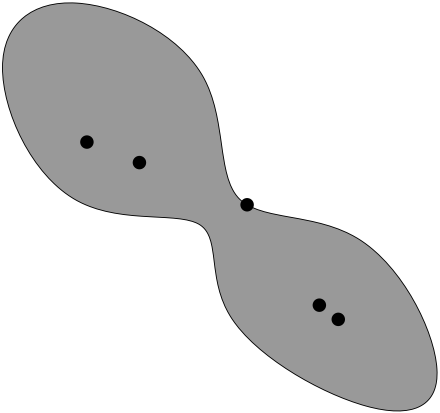
Nested Sampling

Graphical aid



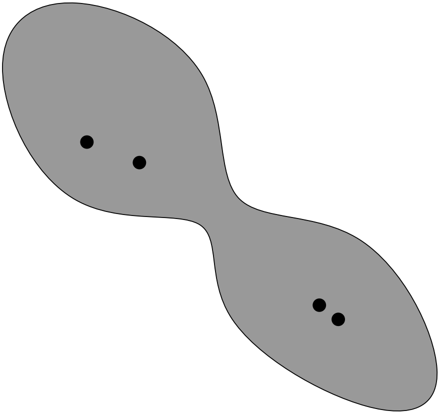
Nested Sampling

Graphical aid



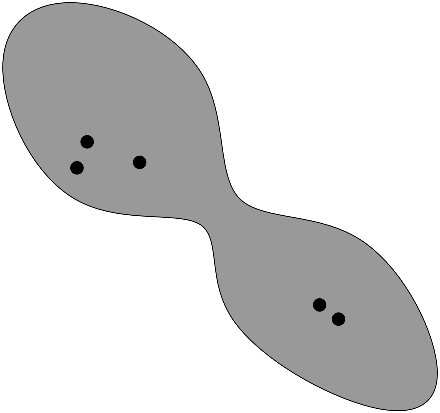
Nested Sampling

Graphical aid



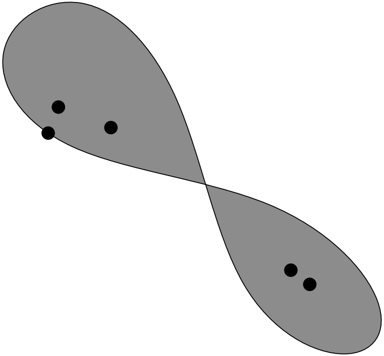
Nested Sampling

Graphical aid



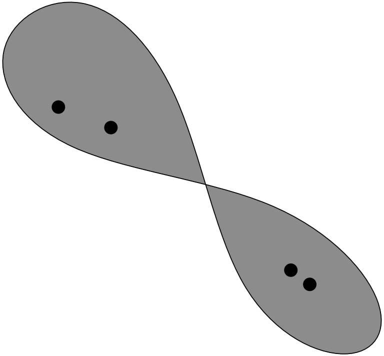
Nested Sampling

Graphical aid



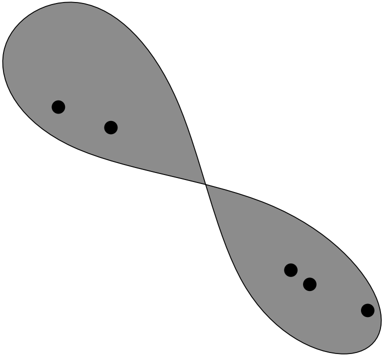
Nested Sampling

Graphical aid



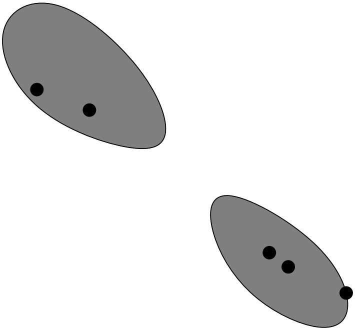
Nested Sampling

Graphical aid



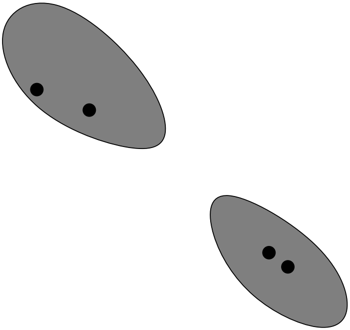
Nested Sampling

Graphical aid



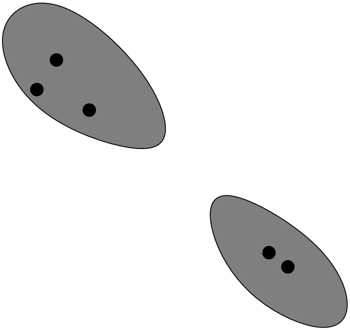
Nested Sampling

Graphical aid



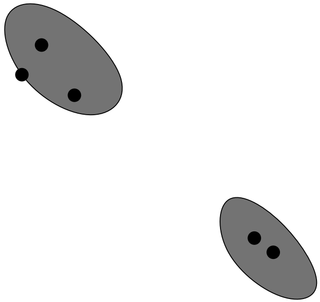
Nested Sampling

Graphical aid



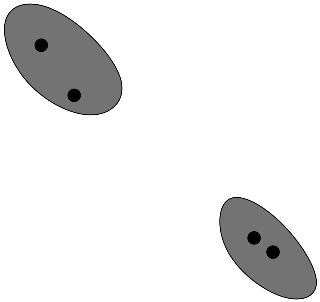
Nested Sampling

Graphical aid



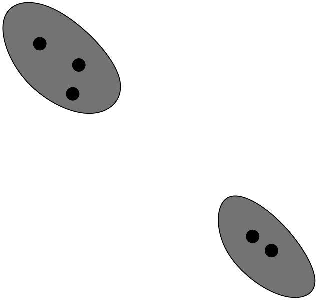
Nested Sampling

Graphical aid



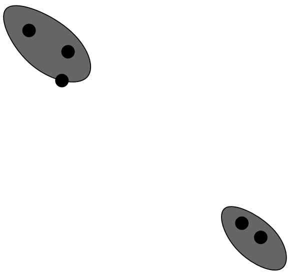
Nested Sampling

Graphical aid



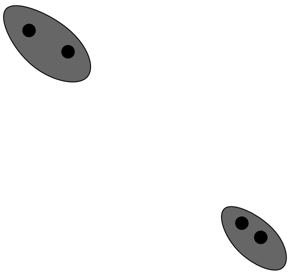
Nested Sampling

Graphical aid



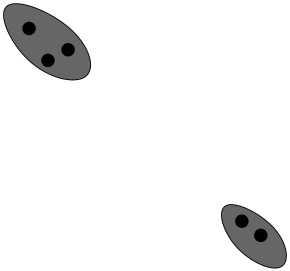
Nested Sampling

Graphical aid



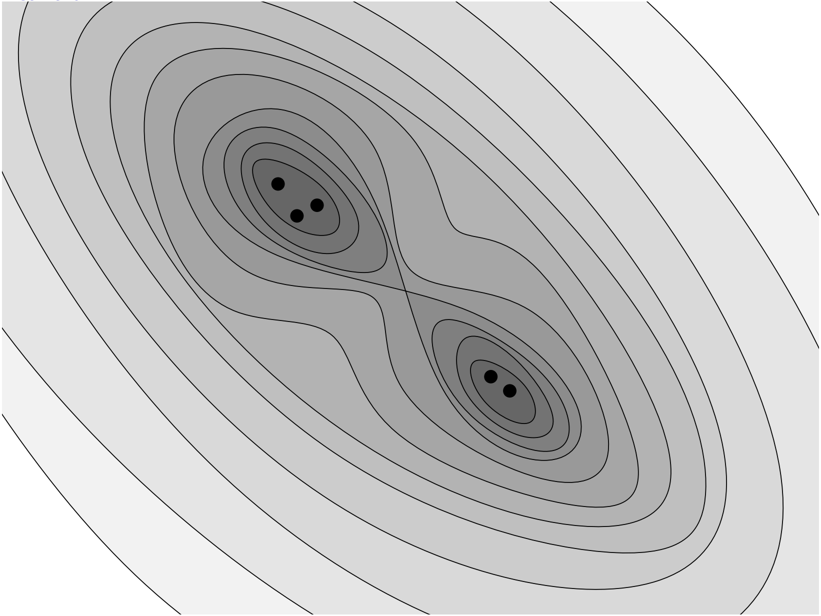
Nested Sampling

Graphical aid



Nested Sampling

Graphical aid



Nested Sampling

Why bother?

Nested Sampling

Why bother?

- At each iteration, the likelihood contour will shrink in volume by a factor of $\approx 1/n$.

Nested Sampling
Why bother?

- At each iteration, the likelihood contour will shrink in volume by a factor of $\approx 1/n$.
- Nested sampling zooms in to the peak of the posterior *exponentially*.

Nested Sampling

Why bother?

- At each iteration, the likelihood contour will shrink in volume by a factor of $\approx 1/n$.
- Nested sampling zooms in to the peak of the posterior *exponentially*.
- Nested sampling can be used to get evidences!

Calculating evidences

Calculating evidences

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta$$

Calculating evidences

- Transform to 1 dimensional integral

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta$$

Calculating evidences

- Transform to 1 dimensional integral $\pi(\theta)d\theta = dX$

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta$$

Calculating evidences

- Transform to 1 dimensional integral $\pi(\theta)d\theta = dX$

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta = \int \mathcal{L}(X)dX$$

Calculating evidences

- Transform to 1 dimensional integral $\pi(\theta)d\theta = dX$

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta = \int \mathcal{L}(X)dX$$

- X is the *prior volume*

Calculating evidences

- Transform to 1 dimensional integral $\pi(\theta)d\theta = dX$

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta = \int \mathcal{L}(X)dX$$

- X is the *prior volume*

$$X(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)d\theta$$

Calculating evidences

- Transform to 1 dimensional integral $\pi(\theta)d\theta = dX$

$$\mathcal{Z} = \int \mathcal{L}(\theta)\pi(\theta)d\theta = \int \mathcal{L}(X)dX$$

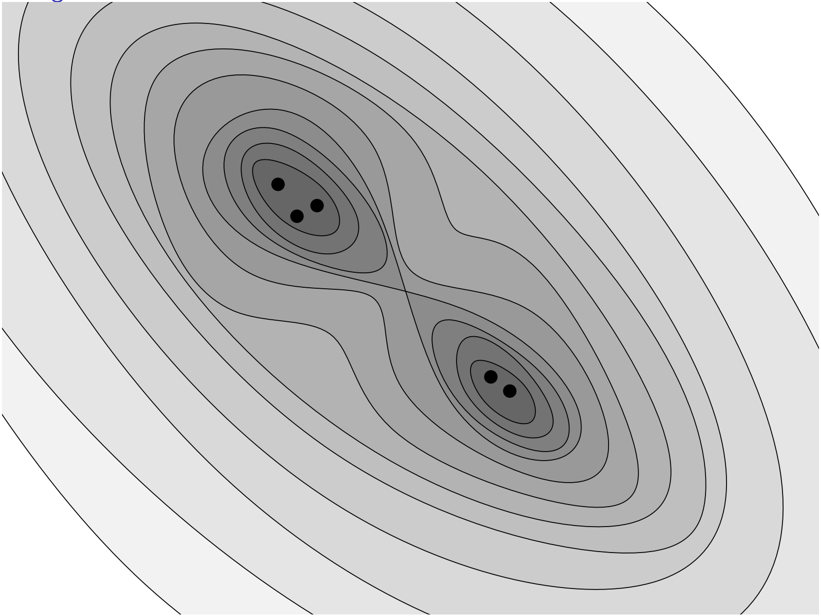
- X is the *prior volume*

$$X(\mathcal{L}) = \int_{\mathcal{L}(\theta)>\mathcal{L}} \pi(\theta)d\theta$$

- i.e. the fraction of the prior which the iso-likelihood contour $\mathcal{L}$ encloses.

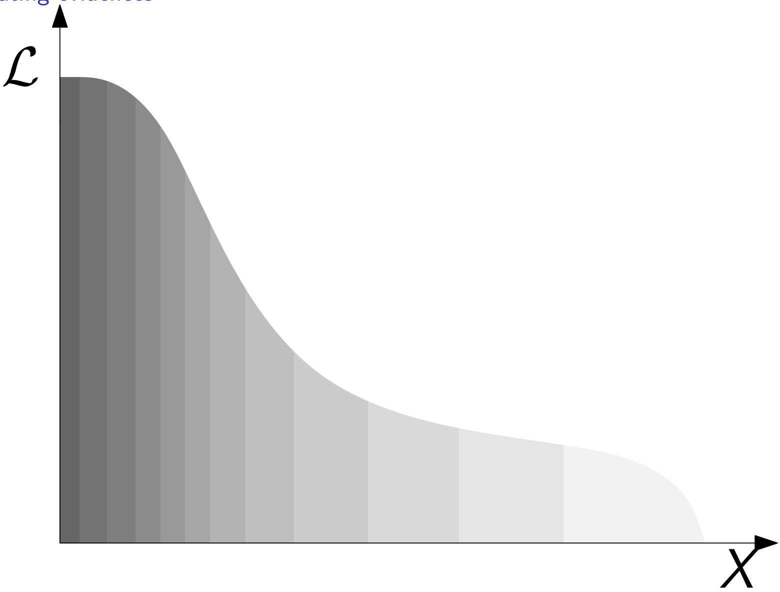
Nested Sampling

Calculating evidences



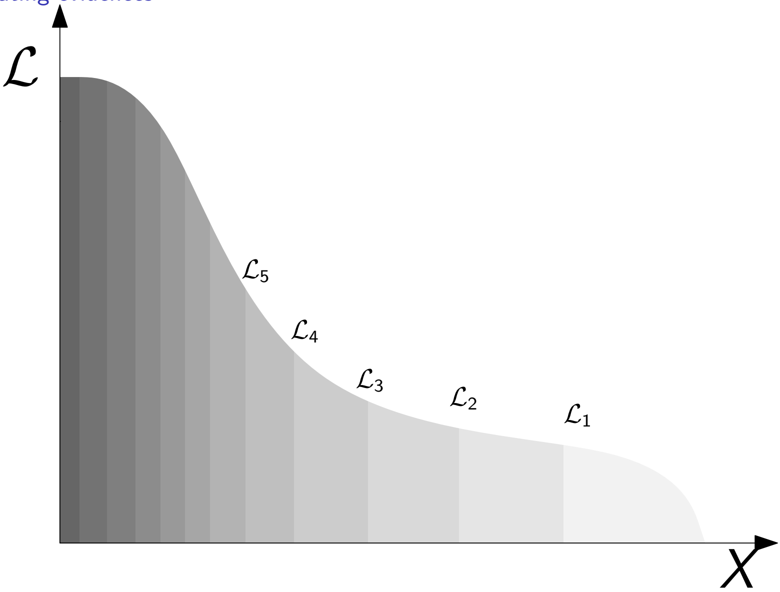
Nested Sampling

Calculating evidences



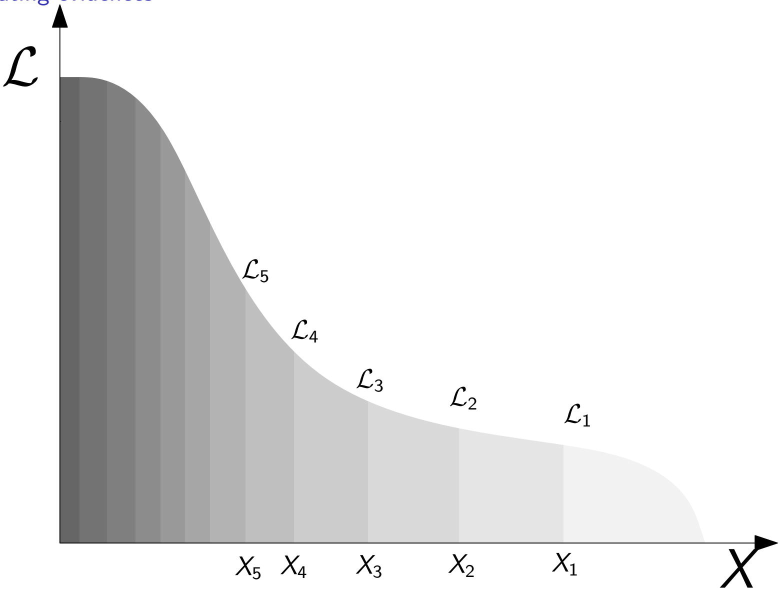
Nested Sampling

Calculating evidences



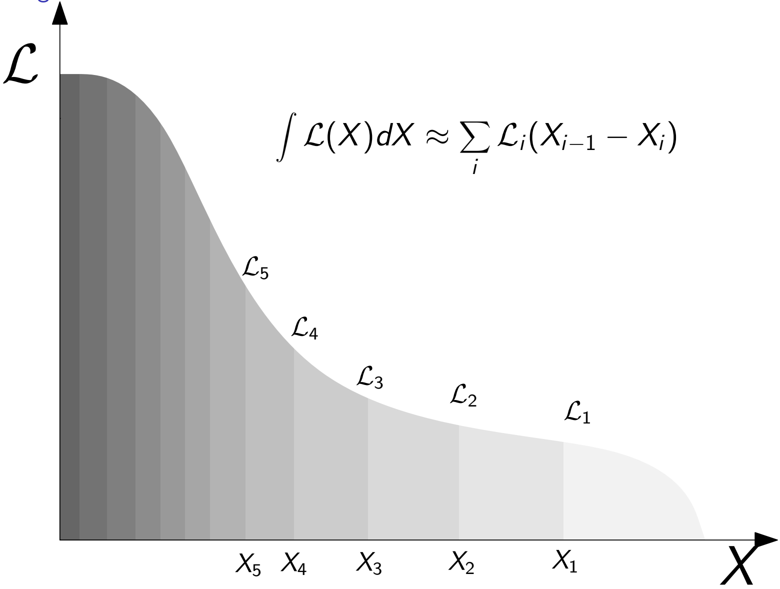
Nested Sampling

Calculating evidences



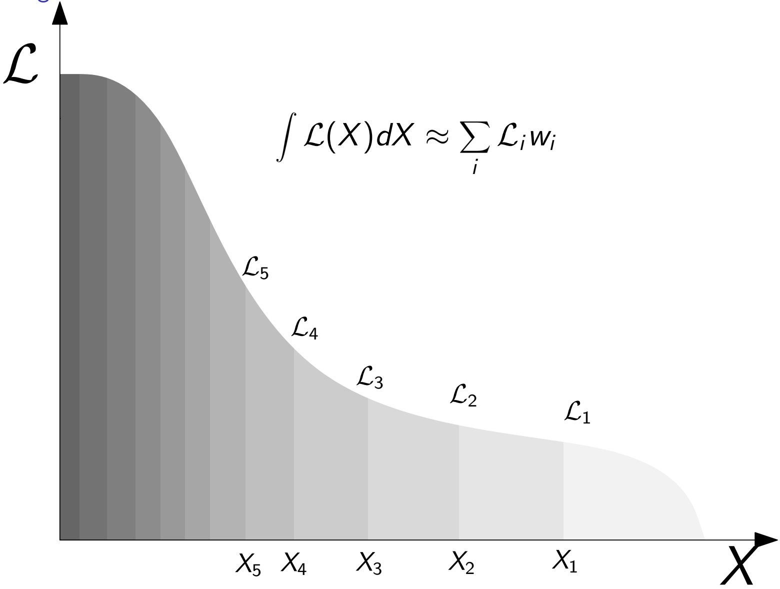
Nested Sampling

Calculating evidences



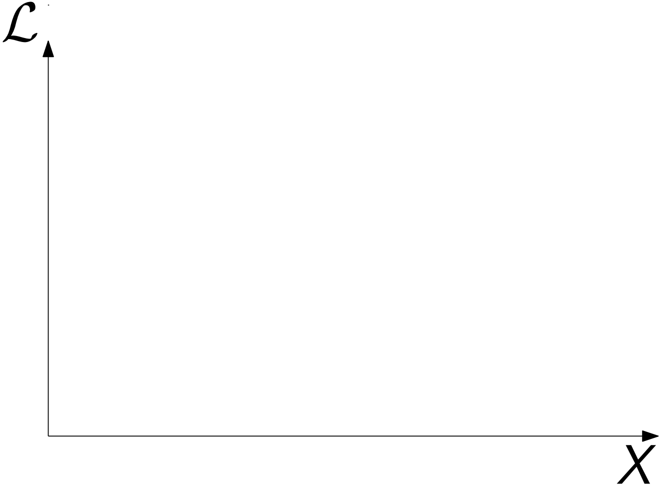
Nested Sampling

Calculating evidences



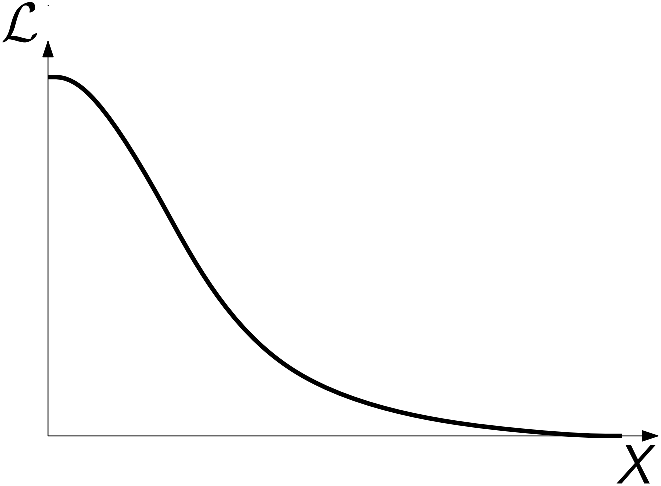
Estimating evidences

Evidence error



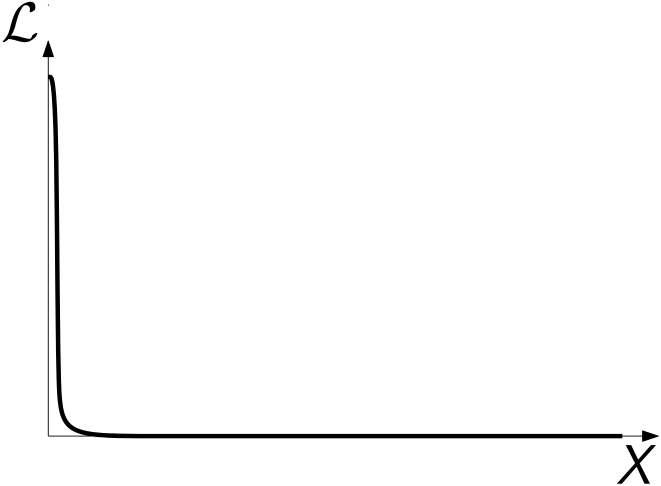
Estimating evidences

Evidence error



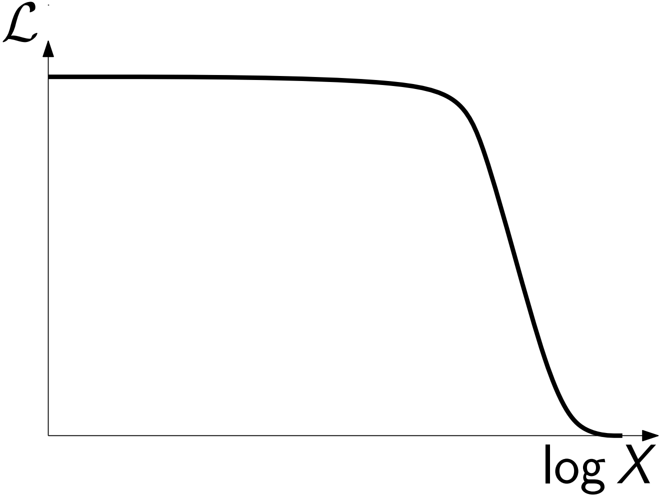
Estimating evidences

Evidence error



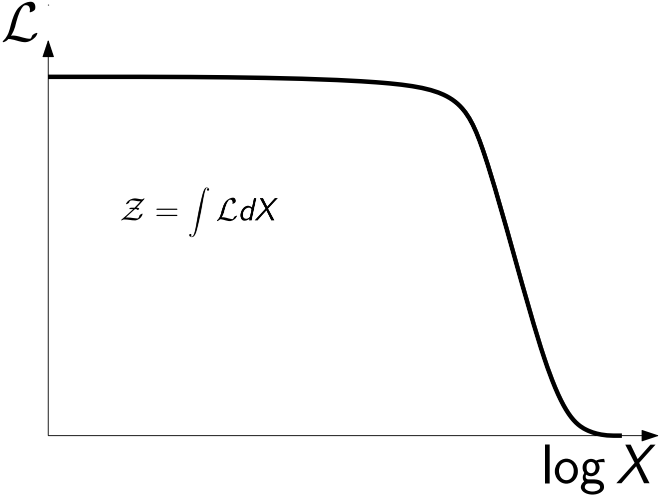
Estimating evidences

Evidence error



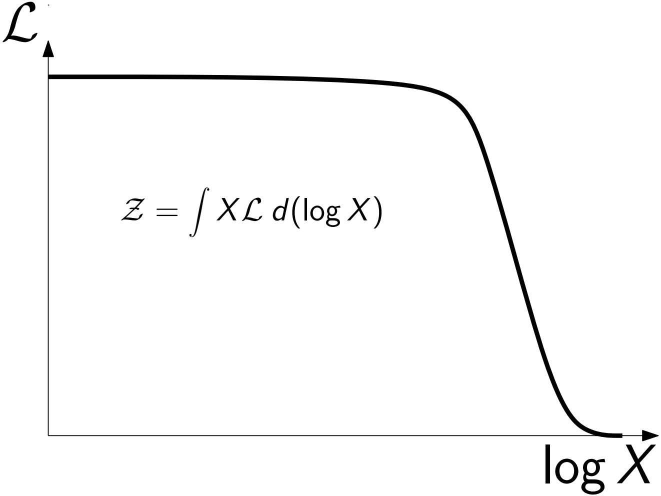
Estimating evidences

Evidence error



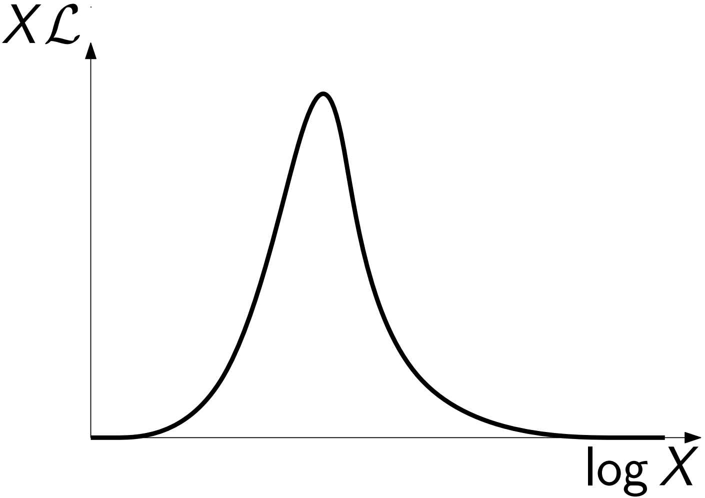
Estimating evidences

Evidence error



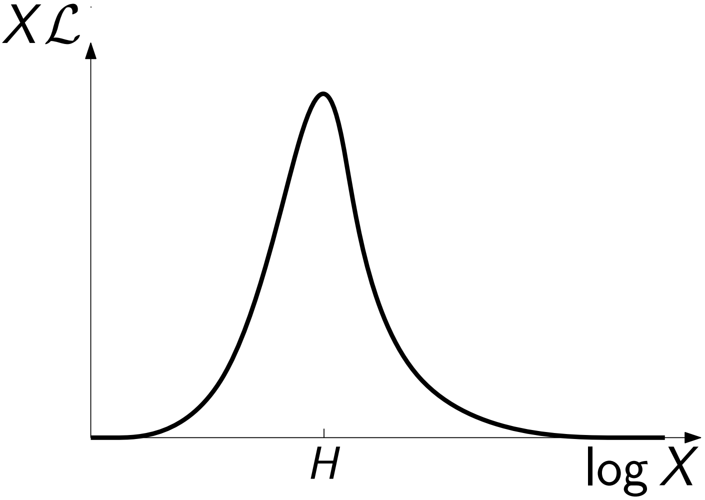
Estimating evidences

Evidence error



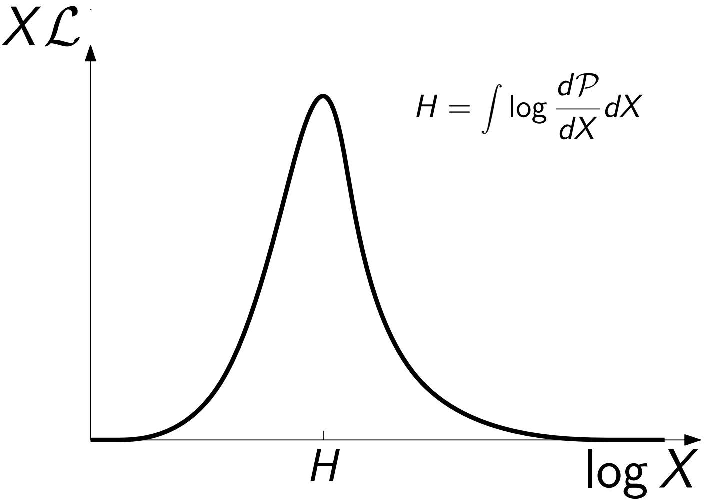
Estimating evidences

Evidence error



Estimating evidences

Evidence error



Estimating evidences

Evidence error

Estimating evidences

Evidence error

- approximate compression:

$$\Delta \log X \sim -\frac{1}{n}$$

Estimating evidences

Evidence error

- approximate compression:

$$\Delta \log X \sim -\frac{1}{n} \pm \frac{1}{n}$$

Estimating evidences

Evidence error

- approximate compression:

$$\Delta \log X \sim -\frac{1}{n} \pm \frac{1}{n}$$

$$\log X_i \sim -\frac{i}{n}$$

Estimating evidences

Evidence error

- approximate compression:

$$\Delta \log X \sim -\frac{1}{n} \pm \frac{1}{n}$$

$$\log X_i \sim -\frac{i}{n} \pm \frac{\sqrt{i}}{n}$$

Estimating evidences

Evidence error

- approximate compression:

$$\Delta \log X \sim -\frac{1}{n} \pm \frac{1}{n}$$

$$\log X_i \sim -\frac{i}{n} \pm \frac{\sqrt{i}}{n}$$

- # of steps to get to H:

$$i_H \sim nH$$

Estimating evidences

Evidence error

- approximate compression:
$$\Delta \log X \sim -\frac{1}{n} \pm \frac{1}{n}$$
$$\log X_i \sim -\frac{i}{n} \pm \frac{\sqrt{i}}{n}$$

- # of steps to get to H:
$$i_H \sim nH$$

- estimate of volume at H:
$$\log X_H \approx -H \pm \sqrt{\frac{H}{n}}$$

Estimating evidences

Evidence error

- approximate compression:

$$\Delta \log X \sim -\frac{1}{n} \pm \frac{1}{n}$$

$$\log X_i \sim -\frac{i}{n} \pm \frac{\sqrt{i}}{n}$$

- $\#$ of steps to get to H:

$$i_H \sim nH$$

- estimate of volume at H:

$$\log X_H \approx -H \pm \sqrt{\frac{H}{n}}$$

- estimate of evidence error:

$$\log \mathcal{Z} \approx \sum w_i \mathcal{L}_i \pm \sqrt{\frac{H}{n}}$$

Nested sampling

Parameter estimation

Nested sampling

Parameter estimation

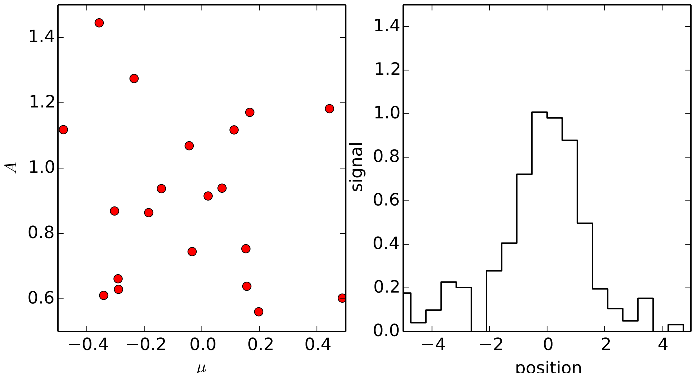
- NS can also be used to sample the posterior

Nested sampling

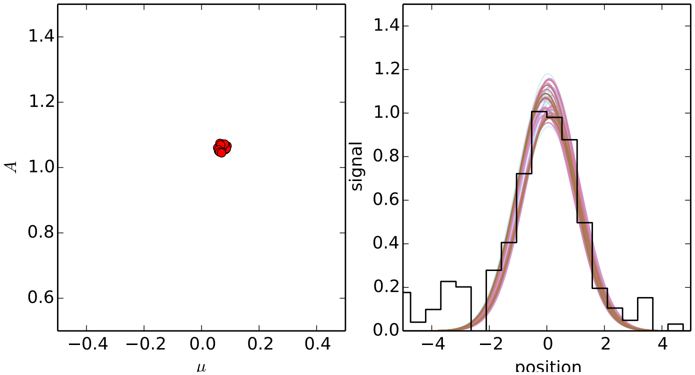
Parameter estimation

- NS can also be used to sample the posterior
- The set of dead points are posterior samples with an appropriate weighting factor

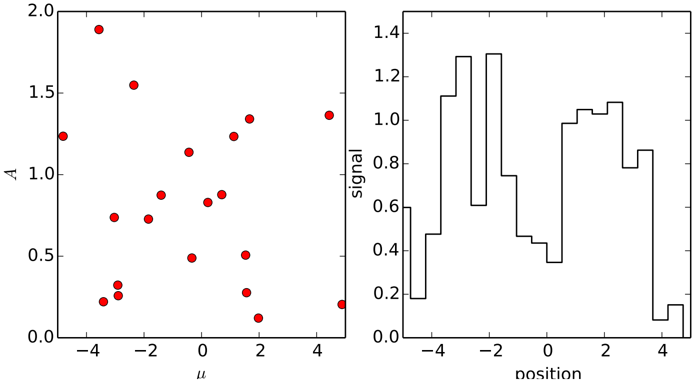
When NS succeeds



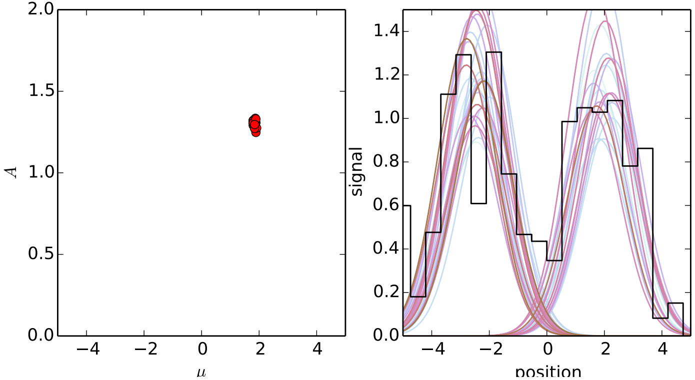
When NS suceeds



When NS succeeds



When NS suceeds



Sampling from a hard likelihood constraint

Sampling from a hard likelihood constraint

> *"It is not the purpose of this introductory paper to develop the technology of navigation within such a volume. We merely note that exploring a hard-edged likelihood-constrained domain should prove to be neither more nor less demanding than exploring a likelihood-weighted space."*
>
> — *John Skilling*

Sampling within an iso-likelihood contour

Previous attempts

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

- Suffers in high dimensions

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

- Suffers in high dimensions

Hamiltonian sampling F. Feroz & J. Skilling (2013).

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

- Suffers in high dimensions

Hamiltonian sampling F. Feroz & J. Skilling (2013).

- Requires gradients and tuning

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

- Suffers in high dimensions

Hamiltonian sampling F. Feroz & J. Skilling (2013).

- Requires gradients and tuning

Diffusion Nested Sampling B. Brewer et al. (2009).

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

- Suffers in high dimensions

Hamiltonian sampling F. Feroz & J. Skilling (2013).

- Requires gradients and tuning

Diffusion Nested Sampling B. Brewer et al. (2009).

- Very promising

Sampling within an iso-likelihood contour

Previous attempts

Rejection Sampling MultiNest; F. Feroz & M. Hobson (2009).

- Suffers in high dimensions

Hamiltonian sampling F. Feroz & J. Skilling (2013).

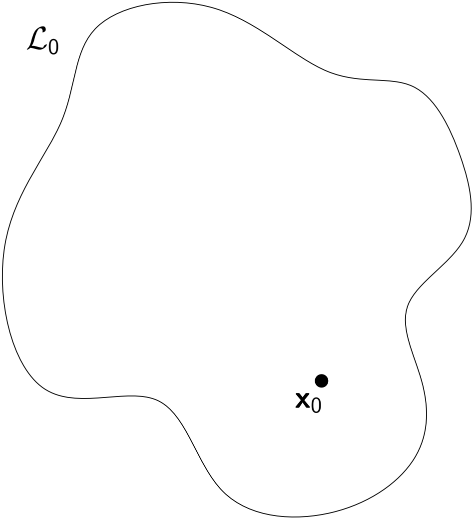
- Requires gradients and tuning

Diffusion Nested Sampling B. Brewer et al. (2009).

- Very promising
- Too many tuning parameters

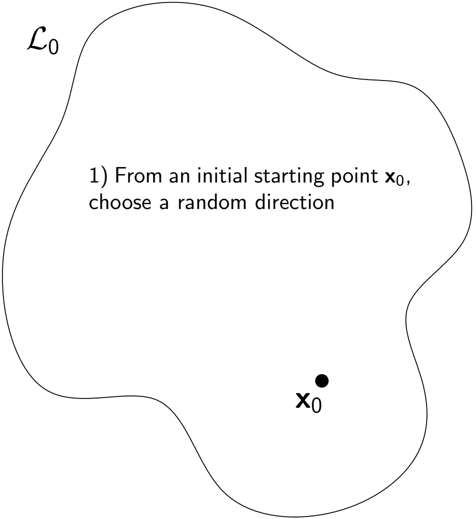
PolyChord

"Hit and run" slice sampling



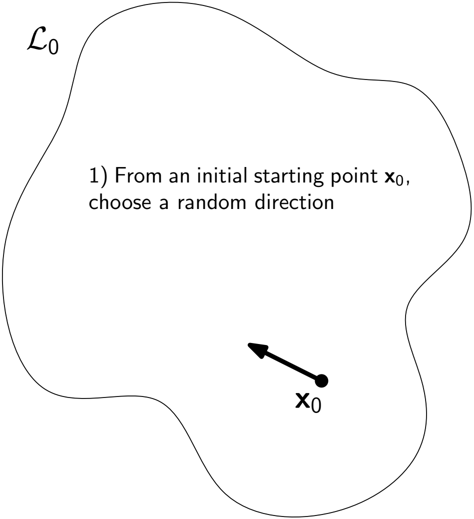
PolyChord

"Hit and run" slice sampling



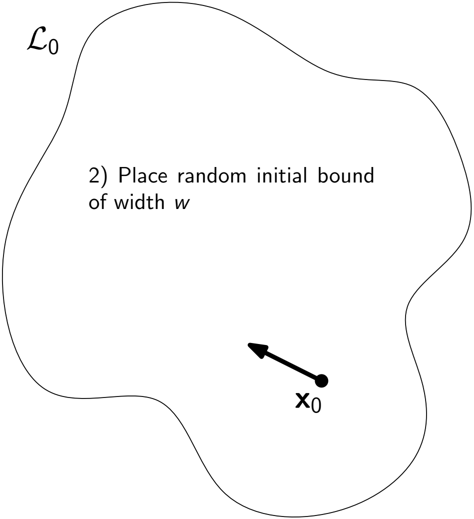
PolyChord

"Hit and run" slice sampling



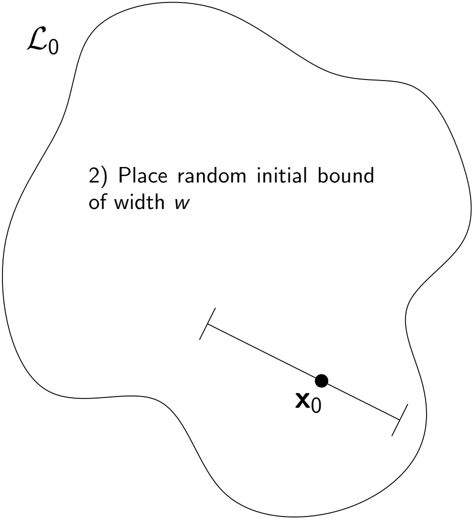
PolyChord

"Hit and run" slice sampling



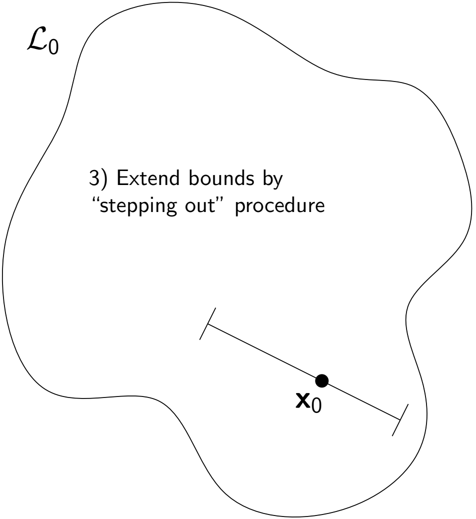
PolyChord

"Hit and run" slice sampling



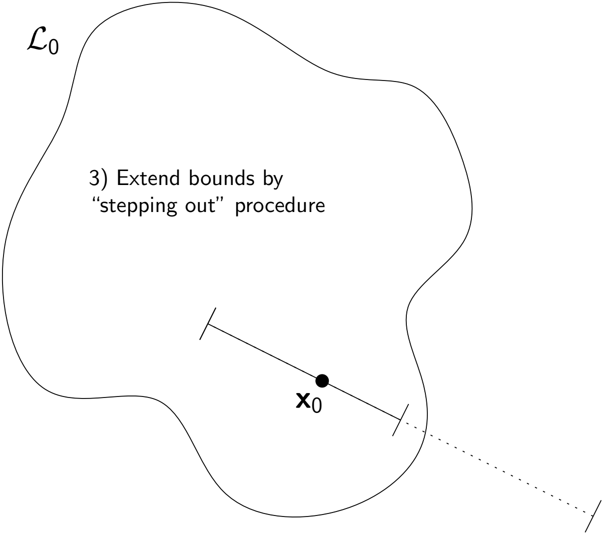
PolyChord

"Hit and run" slice sampling



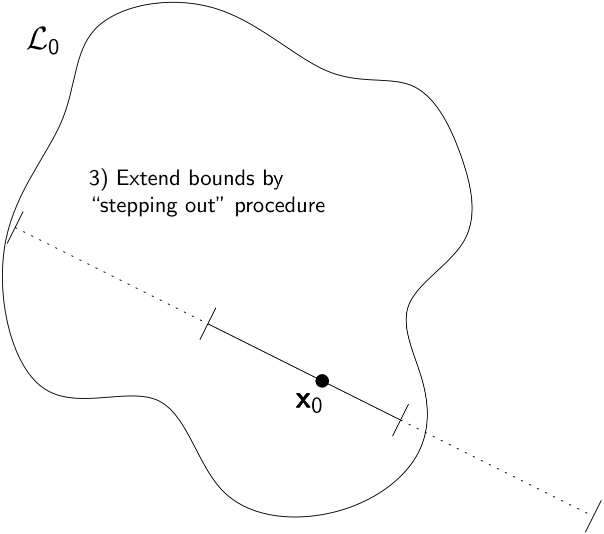
PolyChord

"Hit and run" slice sampling



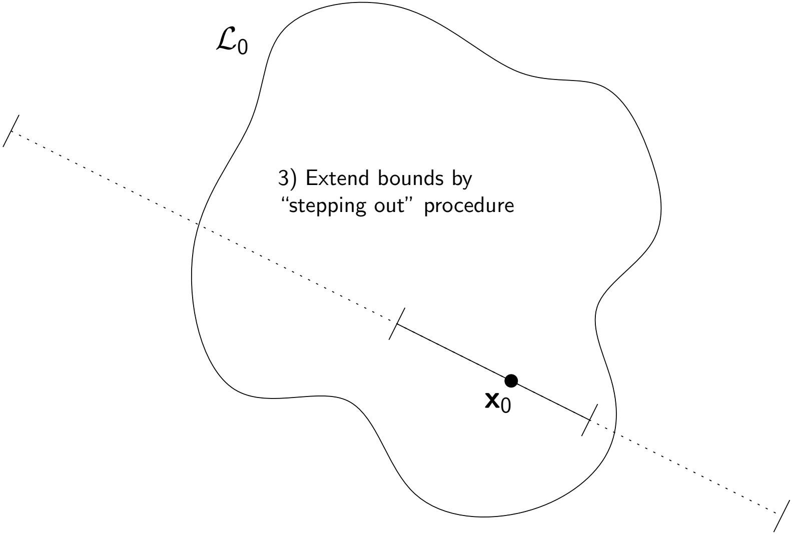
PolyChord

"Hit and run" slice sampling



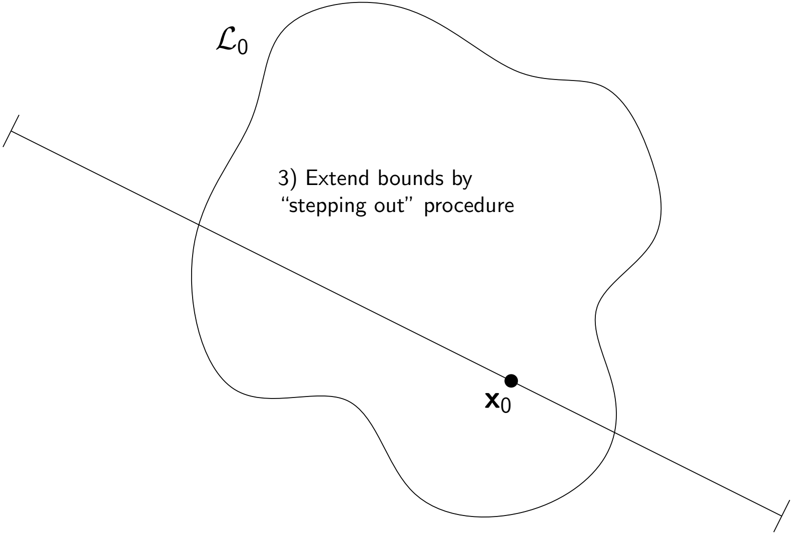
PolyChord

"Hit and run" slice sampling



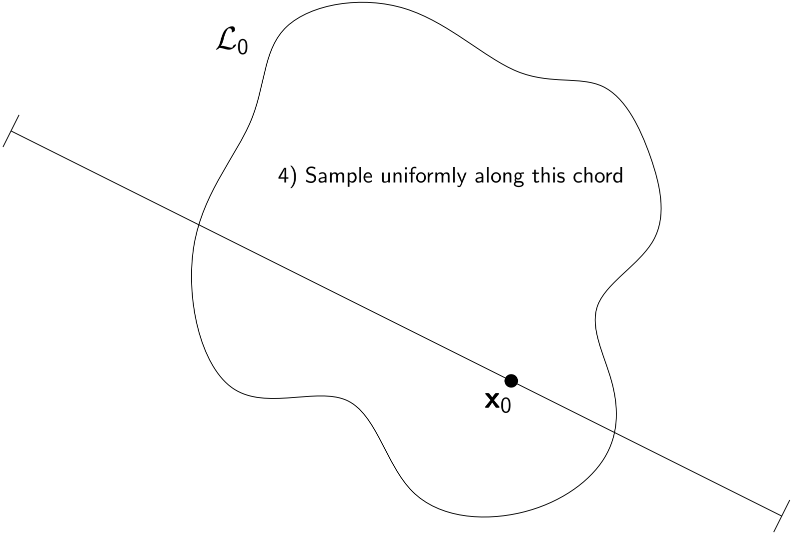
PolyChord

"Hit and run" slice sampling



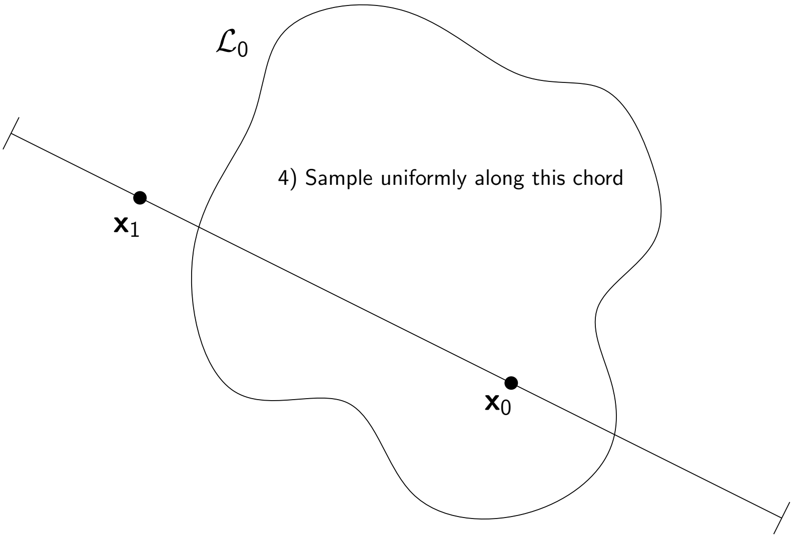
PolyChord

"Hit and run" slice sampling



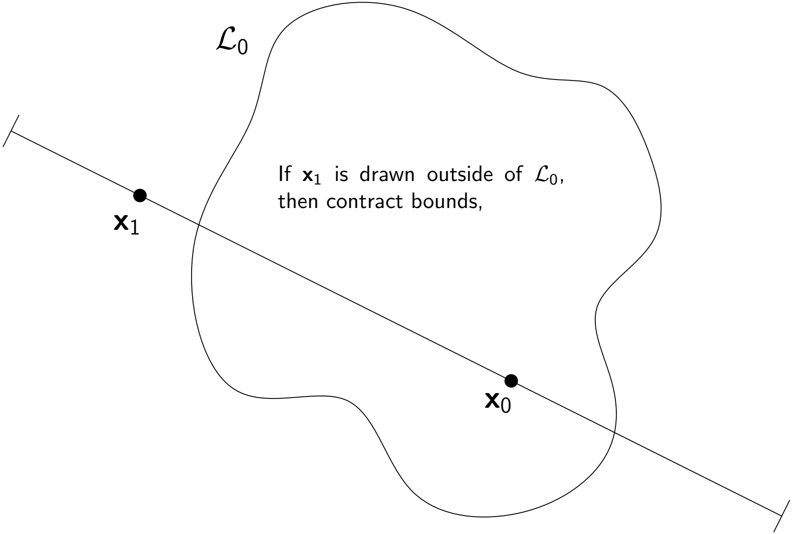
PolyChord

"Hit and run" slice sampling



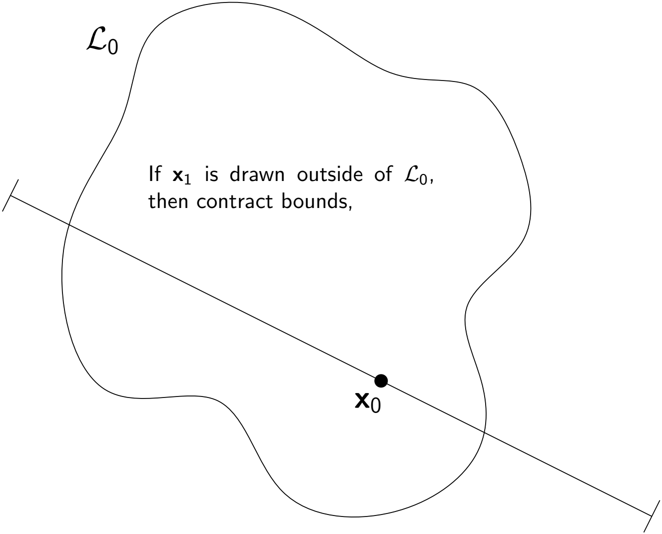
PolyChord

"Hit and run" slice sampling



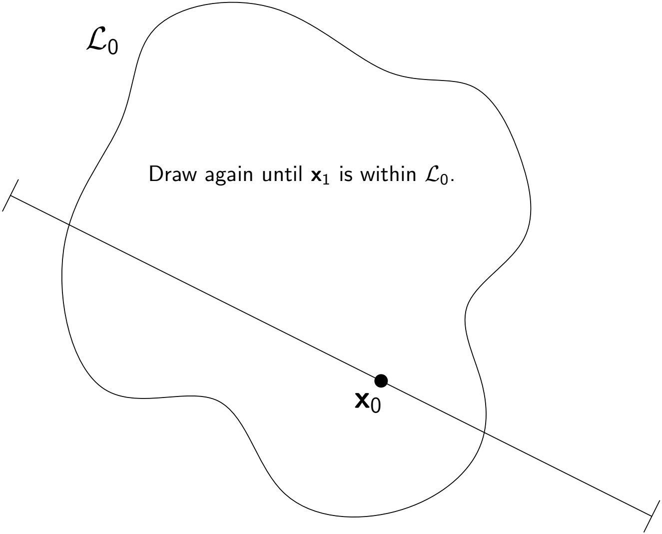
PolyChord

"Hit and run" slice sampling



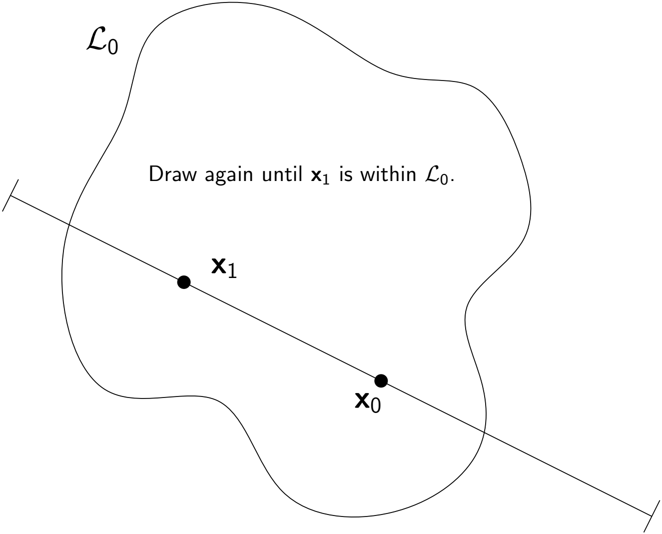
PolyChord

"Hit and run" slice sampling



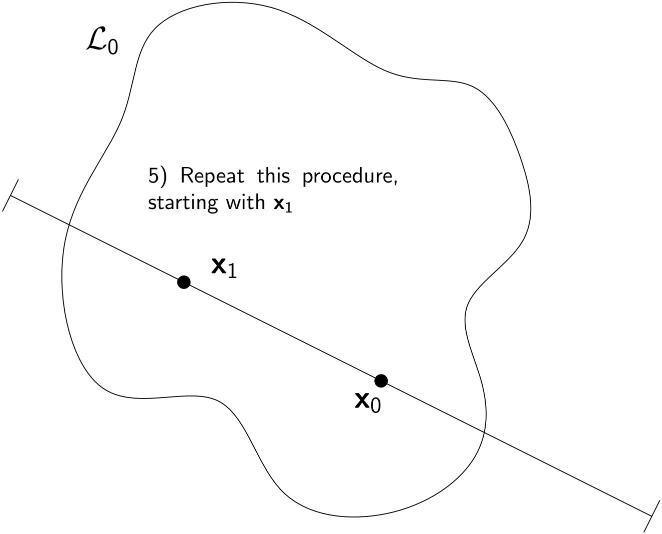
PolyChord

"Hit and run" slice sampling



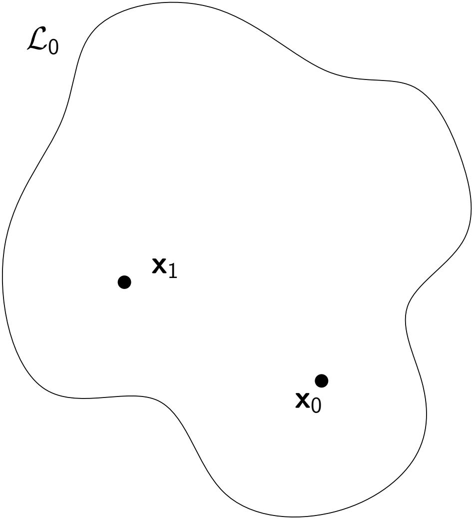
PolyChord

"Hit and run" slice sampling



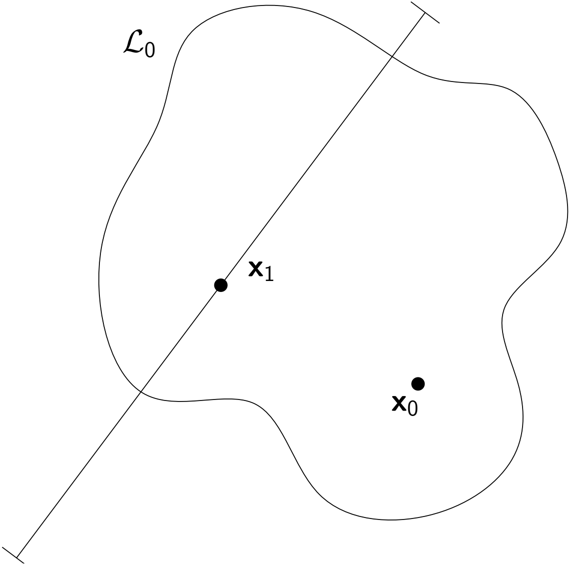
PolyChord

"Hit and run" slice sampling



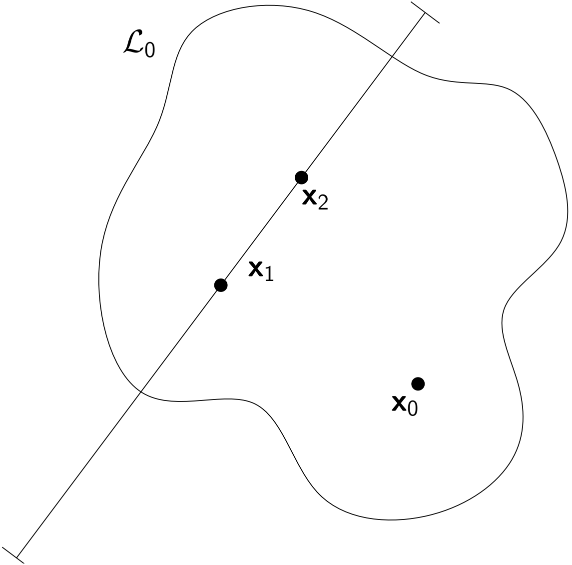
PolyChord

"Hit and run" slice sampling



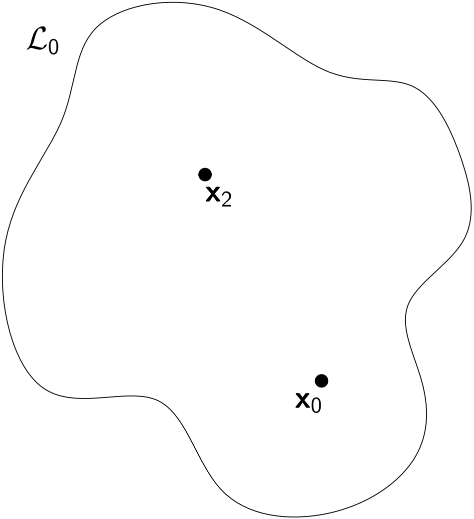
PolyChord

"Hit and run" slice sampling



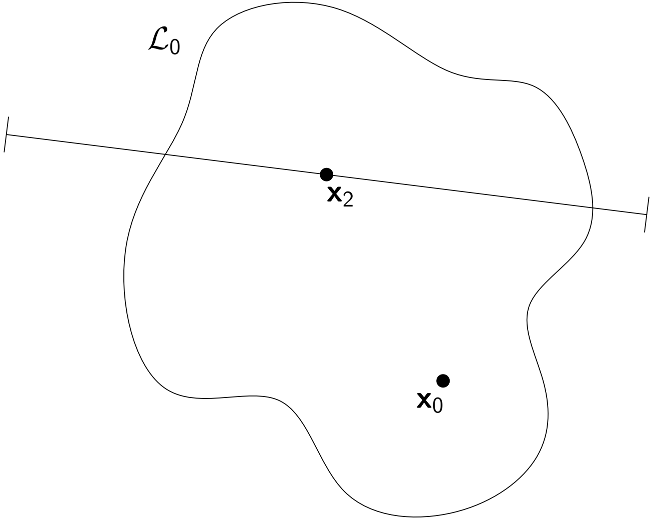
PolyChord

"Hit and run" slice sampling



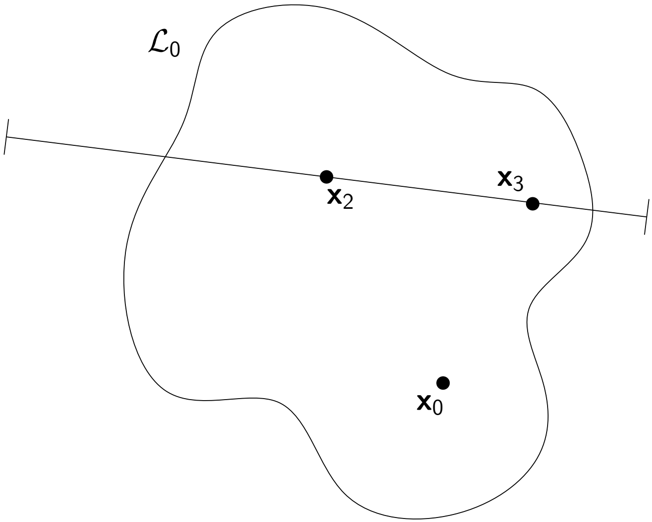
PolyChord

"Hit and run" slice sampling



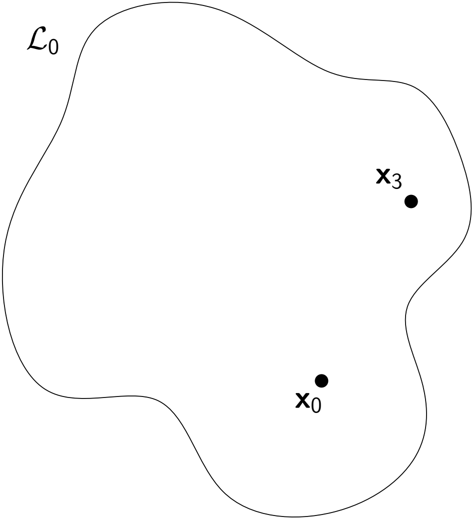
PolyChord

"Hit and run" slice sampling



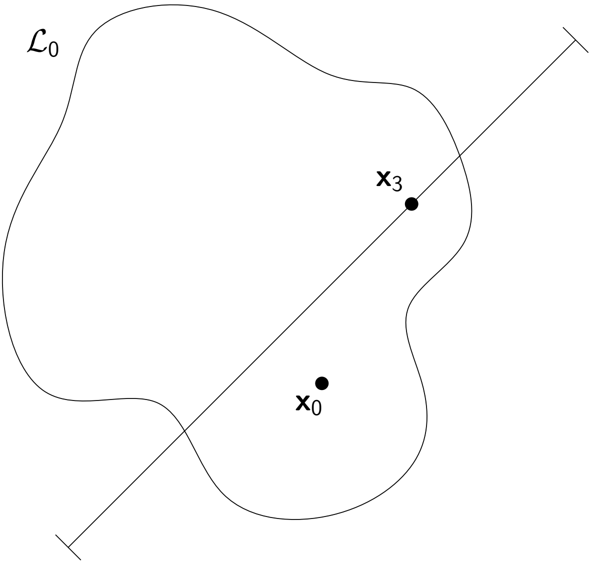
PolyChord

"Hit and run" slice sampling



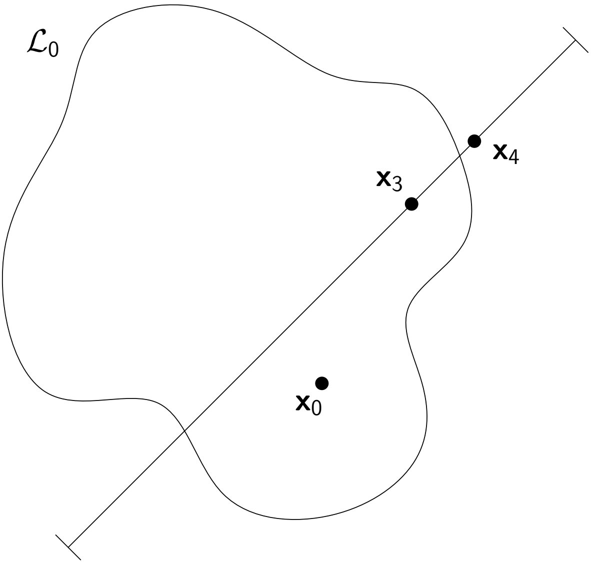
PolyChord

"Hit and run" slice sampling



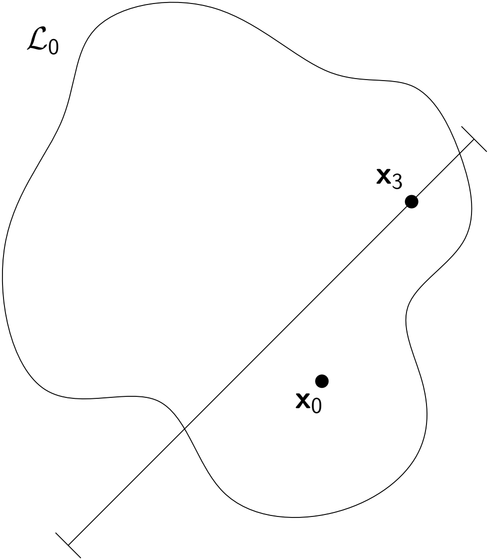
PolyChord

"Hit and run" slice sampling



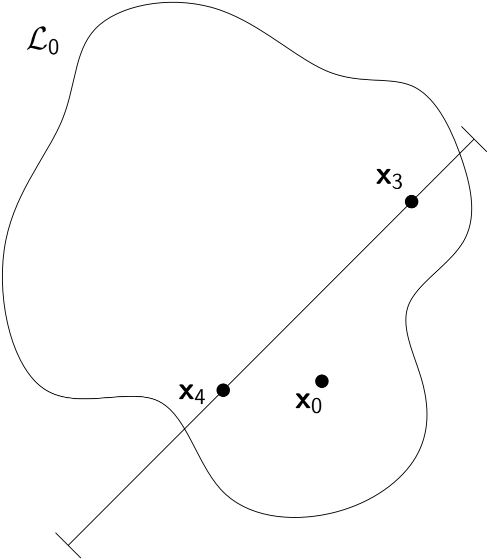
PolyChord

"Hit and run" slice sampling



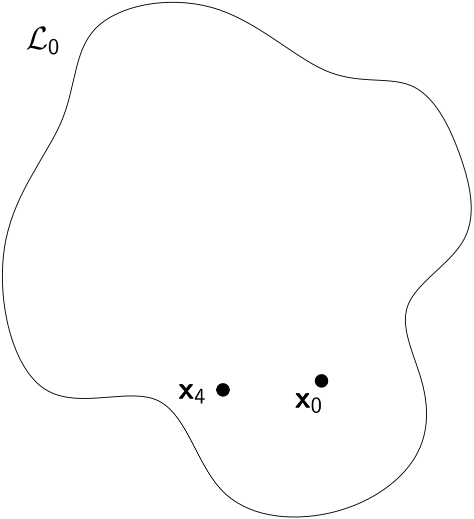
PolyChord

"Hit and run" slice sampling



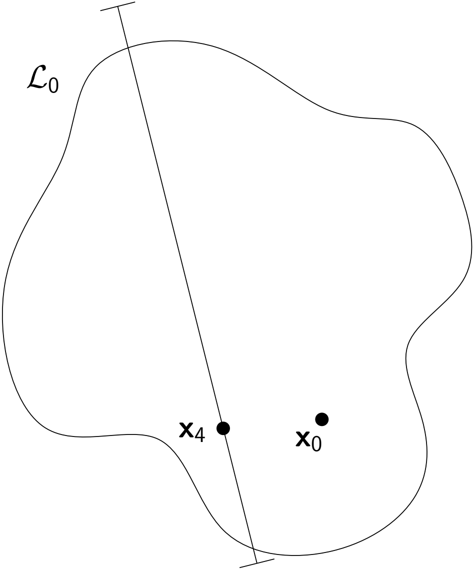
PolyChord

"Hit and run" slice sampling



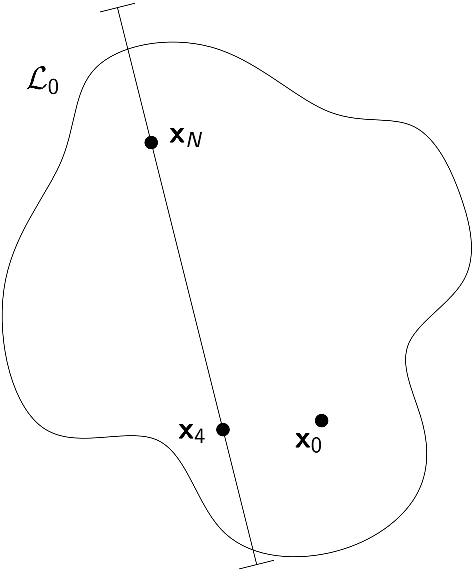
PolyChord

"Hit and run" slice sampling



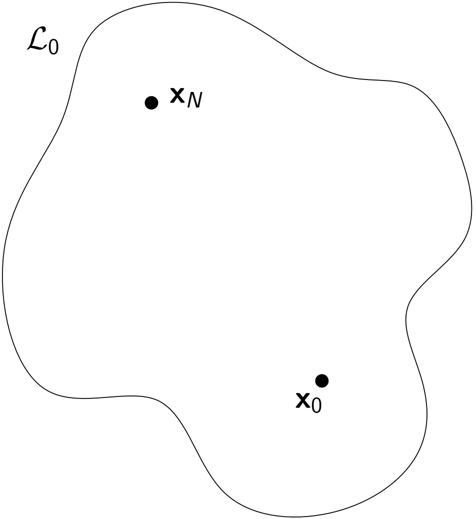
PolyChord

"Hit and run" slice sampling



PolyChord

"Hit and run" slice sampling



PolyChord

Key points

PolyChord

Key points

- This procedure satisfies detailed balance.

PolyChord

Key points

- This procedure satisfies detailed balance.
- Works even if $\mathcal{L}_0$ contour is disjoint.

PolyChord

Key points

- This procedure satisfies detailed balance.
- Works even if $\mathcal{L}_0$ contour is disjoint.

PolyChord

Key points

- This procedure satisfies detailed balance.
- Works even if $\mathcal{L}_0$ contour is disjoint.
- Need N reasonably large $\sim \mathcal{O}(n_{\mathrm{dims}})$ so that x_N is de-correlated from x_1.

Issues with Slice Sampling

Issues with Slice Sampling

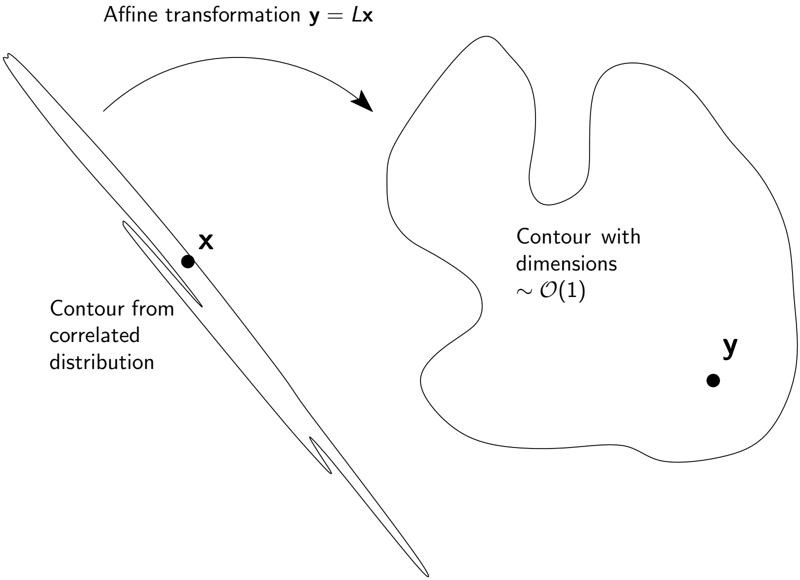
1. Does not deal well with correlated distributions.

Issues with Slice Sampling

1. Does not deal well with correlated distributions.
2. Need to “tune” w parameter.

PolyChord's solutions

Correlated distributions



PolyChord's solutions

Correlated distributions

PolyChord's solutions

Correlated distributions

- We make an affine transformation to remove degeneracies, and "whiten" the space.

PolyChord's solutions

Correlated distributions

- We make an affine transformation to remove degeneracies, and "whiten" the space.
- Samples remain uniformly sampled

PolyChord's solutions

Correlated distributions

- We make an affine transformation to remove degeneracies, and "whiten" the space.
- Samples remain uniformly sampled
- We use the covariance matrix of the live points and all inter-chain points

PolyChord's solutions

Correlated distributions

- We make an affine transformation to remove degeneracies, and "whiten" the space.
- Samples remain uniformly sampled
- We use the covariance matrix of the live points and all inter-chain points
- Cholesky decomposition is the required skew transformation

PolyChord's solutions

Correlated distributions

- We make an affine transformation to remove degeneracies, and "whiten" the space.
- Samples remain uniformly sampled
- We use the covariance matrix of the live points and all inter-chain points
- Cholesky decomposition is the required skew transformation
- $w = 1$ in this transformed space

PolyChord's Additions

PolyChord's Additions

- Parallelised up to number of live points with openMPI.

PolyChord's Additions

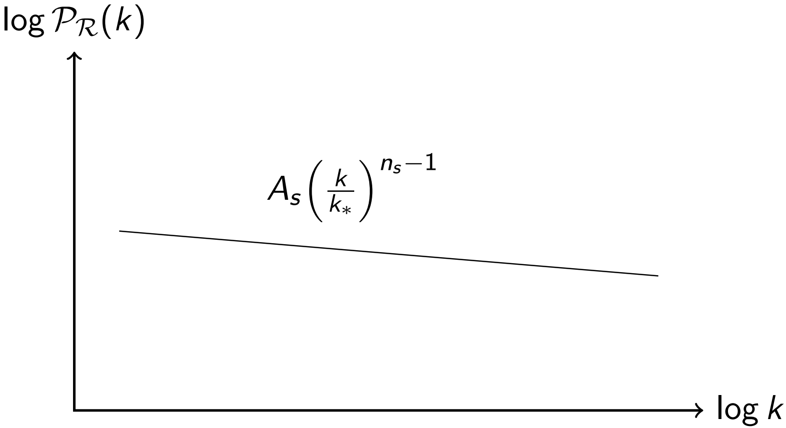
- Parallelised up to number of live points with openMPI.
- Novel method for identifying and evolving modes separately.

PolyChord's Additions

- Parallelised up to number of live points with openMPI.
- Novel method for identifying and evolving modes separately.
- Implemented in CosmoMC, as "CosmoChord", with fast-slow parameters.

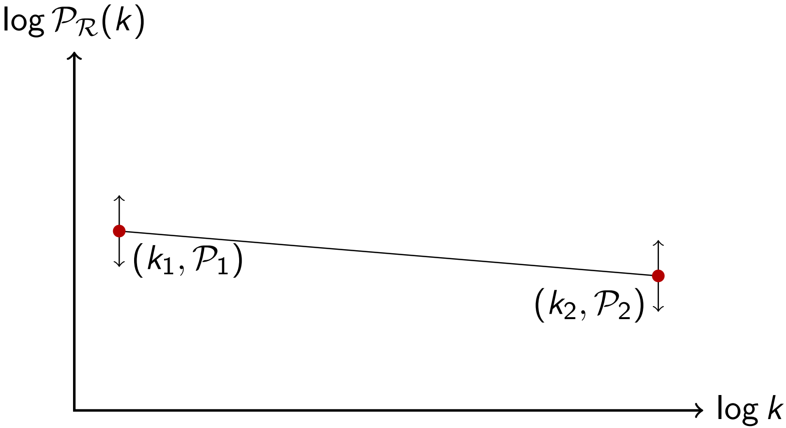
PolyChord in action

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



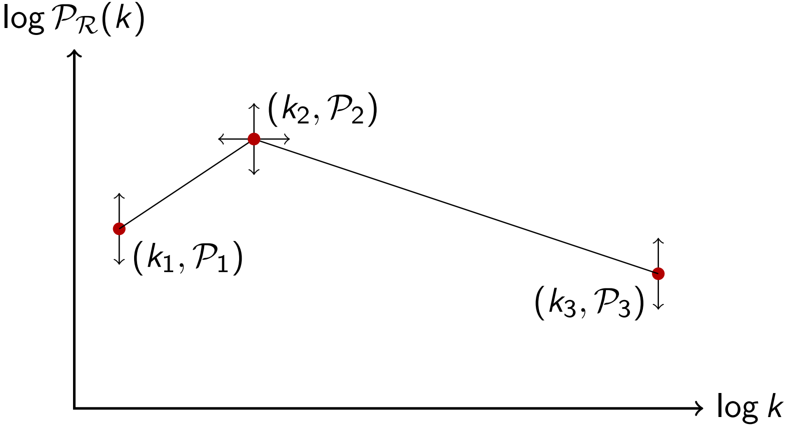
PolyChord in action

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



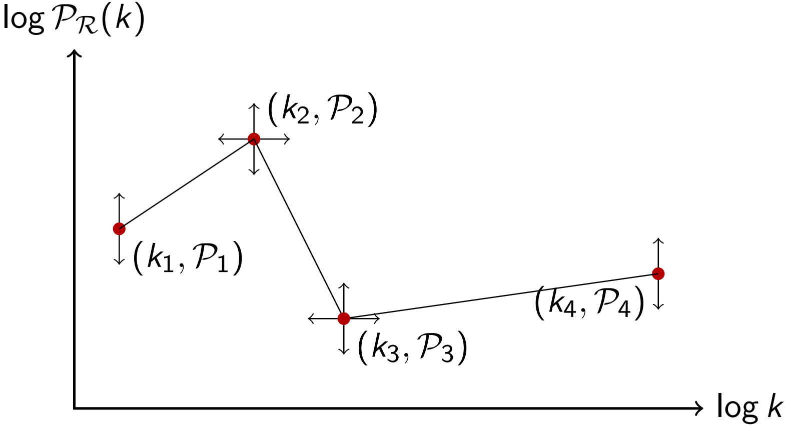
PolyChord in action

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



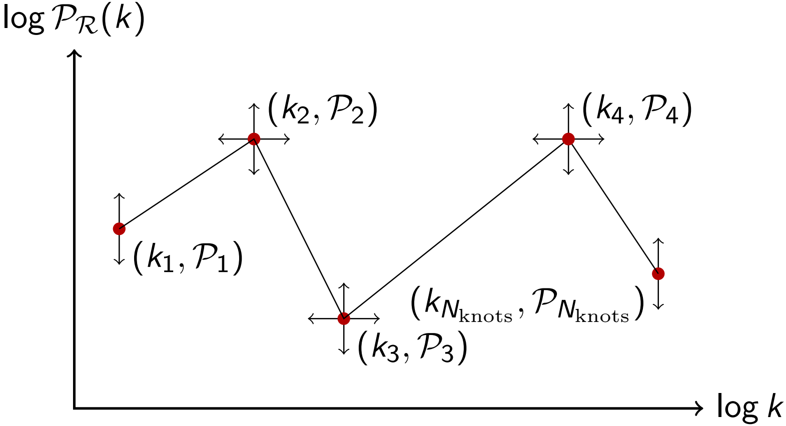
PolyChord in action

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



PolyChord in action

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



Planck data

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction

Planck data

Primordial power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ reconstruction

- Temperature data TT+lowP

Planck data

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction

- Temperature data TT+lowP
- Foreground (14) & cosmological $(4 + 2 * N_{\mathrm{knots}} - 2)$ parameters

Planck data

Primordial power spectrum $\mathcal{P}_{\mathcal{R}}(k)$ reconstruction

- Temperature data TT+lowP
- Foreground (14) & cosmological $(4 + 2 * N_{\mathrm{knots}} - 2)$ parameters
- Marginalised plots of $\mathcal{P}_{\mathcal{R}}(k)$

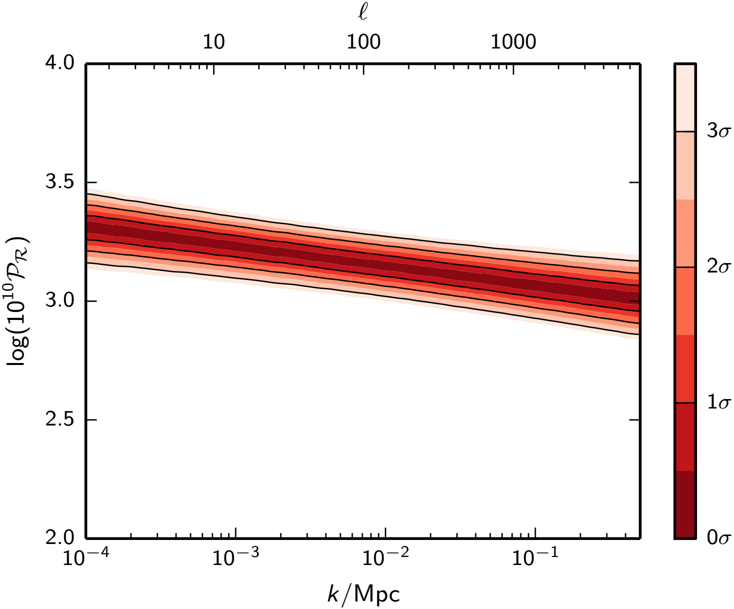
Planck data

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction

- Temperature data TT+lowP
- Foreground (14) & cosmological $(4 + 2 * N_\mathrm{knots} - 2)$ parameters
- Marginalised plots of $\mathcal{P}_\mathcal{R}(k)$
- $$\mathrm{P}(\mathcal{P}_\mathcal{R}|k, N_\mathrm{knots}) = \int \delta(\mathcal{P}_\mathcal{R} - f(k;\theta))\mathcal{P}(\theta)d\theta$$

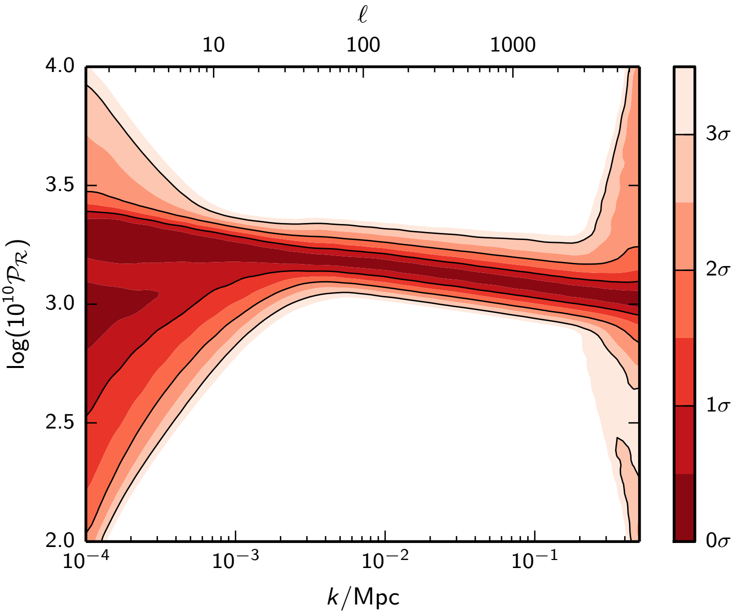
0 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



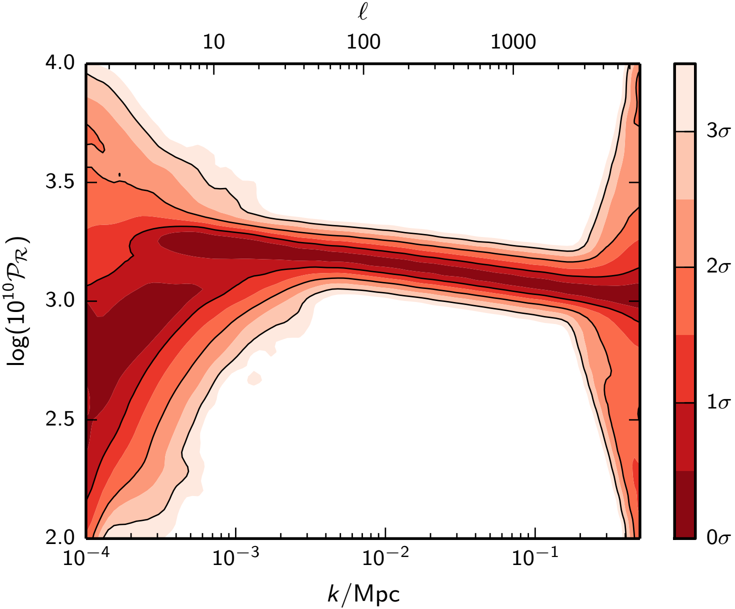
1 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



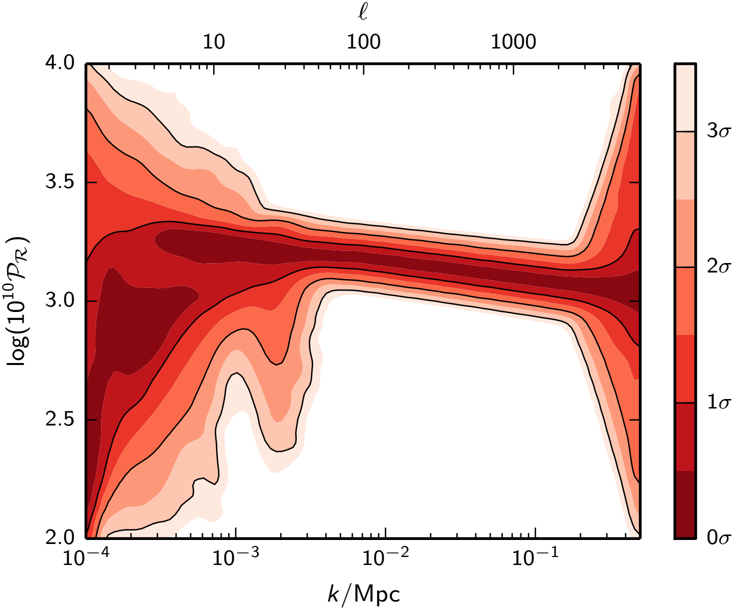
2 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



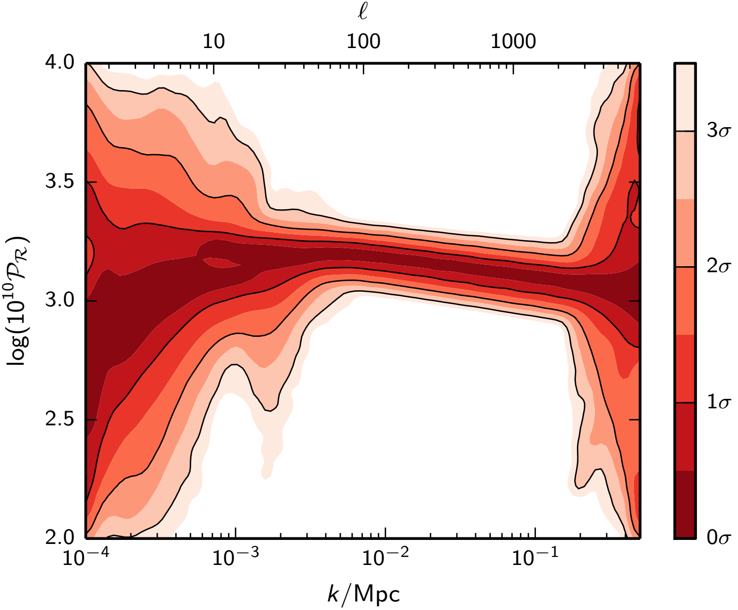
3 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



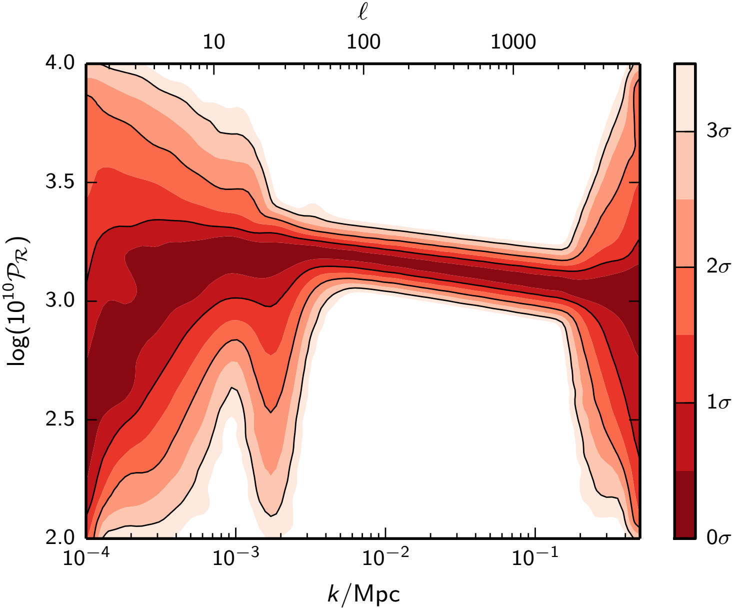
4 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



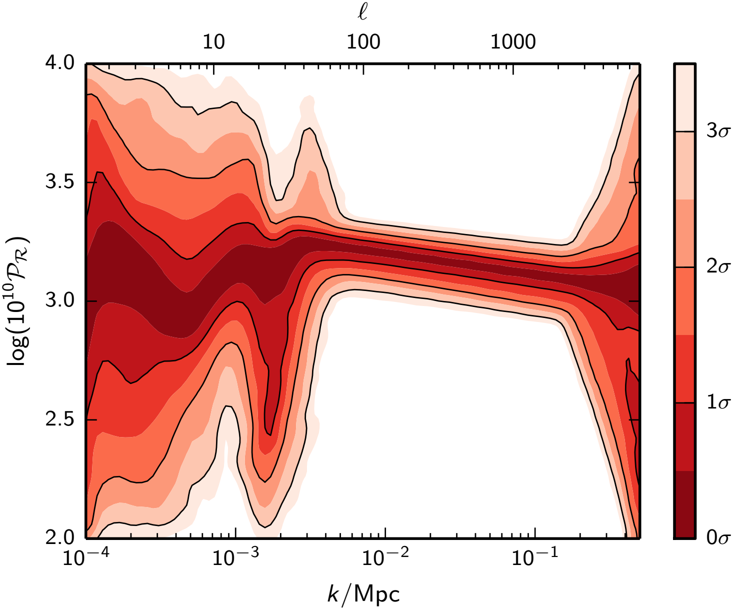
5 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



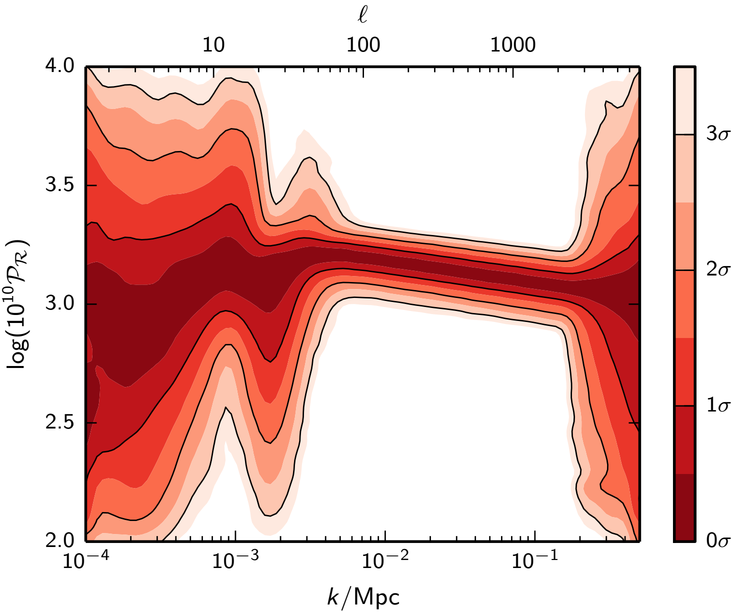
6 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



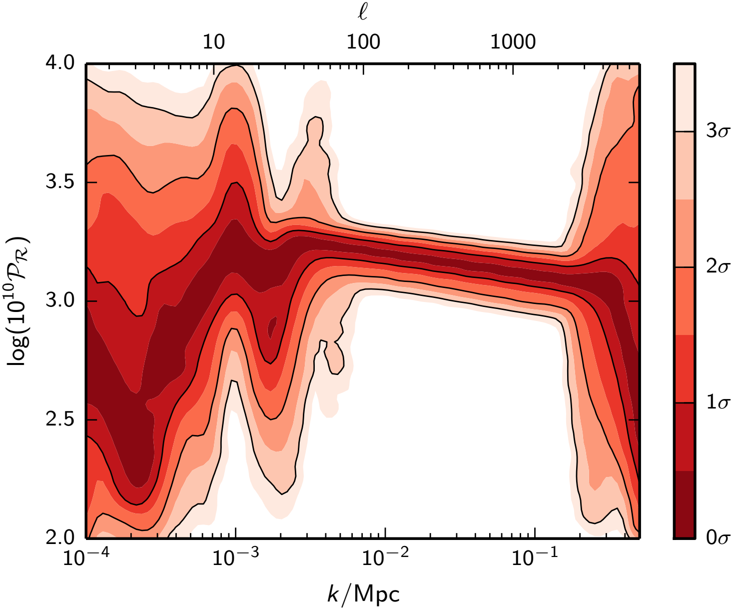
7 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



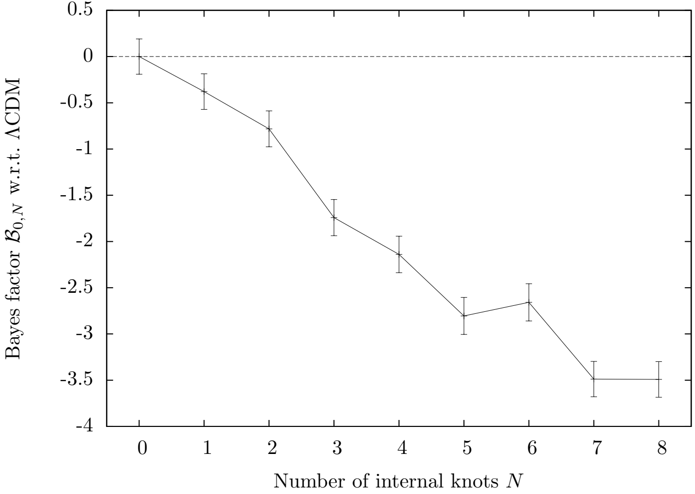
8 internal knots

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



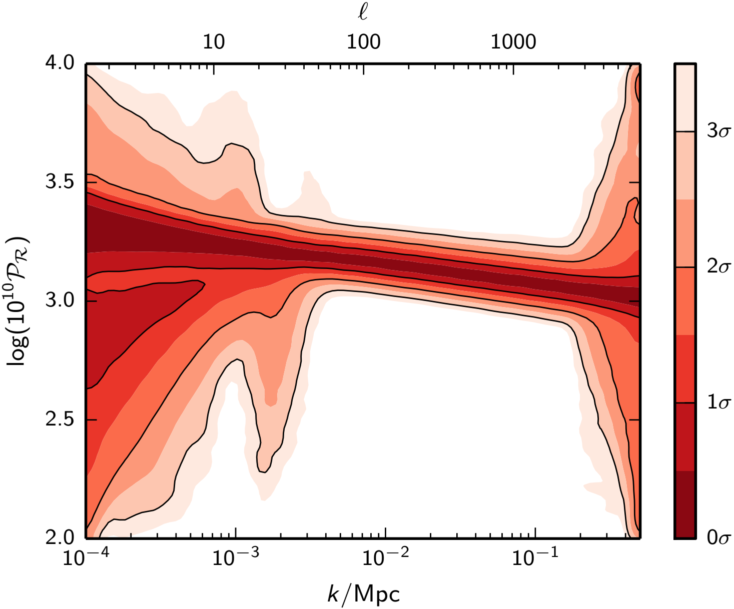
Bayes Factors

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



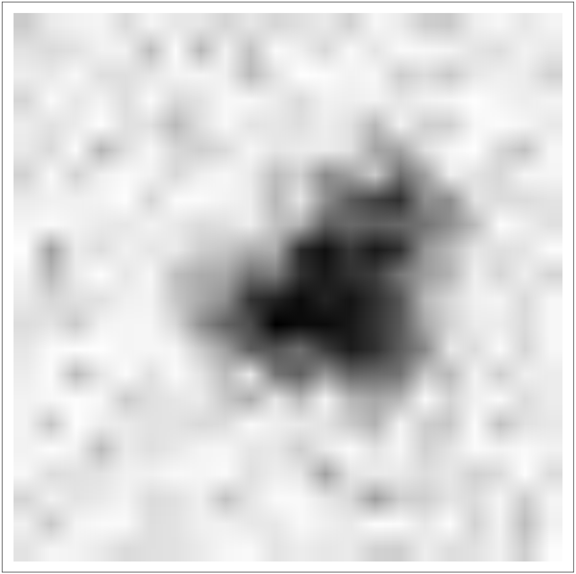
Marginalised plot

Primordial power spectrum $\mathcal{P}_\mathcal{R}(k)$ reconstruction



Object detection

Toy problem



Object detection

Evidences

Object detection

Evidences

- $\log \mathcal{Z}$ ratio: $-251 : -156 : -114 : -117 : -136$

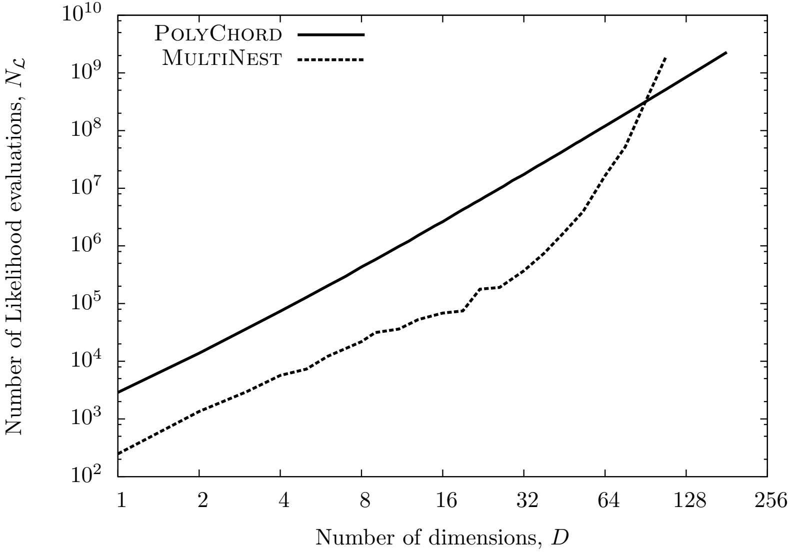
Object detection

Evidences

- $\log \mathcal{Z}$ ratio: $-251 : -156 : -114 : -117 : -136$
- odds ratio: $10^{-60} : 10^{-19} : 1 : 0.04 : 10^{-10}$

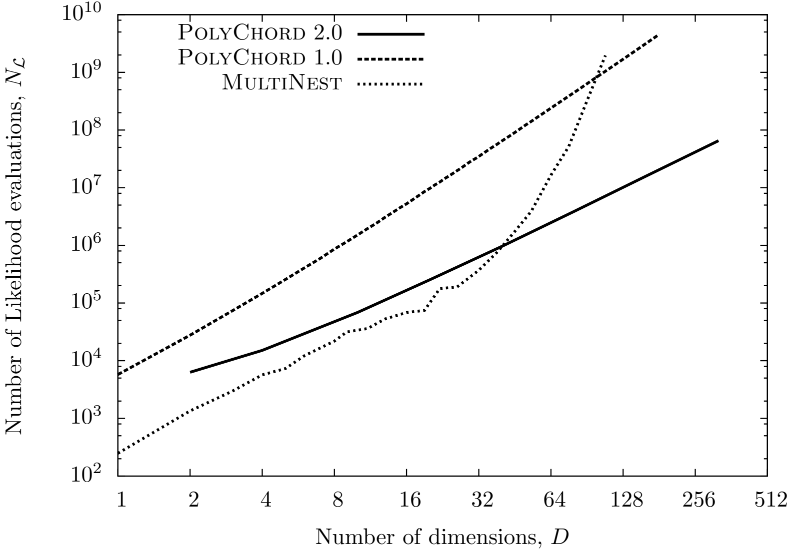
PolyChord vs. MultiNest

Gaussian likelihood



PolyChord vs. MultiNest

Gaussian likelihood



Conclusions

The future of nested sampling

Conclusions

The future of nested sampling

- We are at the beginning of a new era of sampling algorithms

Conclusions

The future of nested sampling

- We are at the beginning of a new era of sampling algorithms
- Plenty of more work in to be done in exploring new versions of nested sampling

Conclusions

The future of nested sampling

- We are at the beginning of a new era of sampling algorithms
- Plenty of more work in to be done in exploring new versions of nested sampling
- Nested sampling is just the beginning

Conclusions

The future of nested sampling

- We are at the beginning of a new era of sampling algorithms
- Plenty of more work in to be done in exploring new versions of nested sampling
- Nested sampling is just the beginning
- arXiv:1506.00171

Conclusions

The future of nested sampling

- We are at the beginning of a new era of sampling algorithms
- Plenty of more work in to be done in exploring new versions of nested sampling
- Nested sampling is just the beginning
- arXiv:1506.00171
- http://ccpforge.cse.rl.ac.uk/gf/project/polychord/