PolyChord & the future of nested sampling

Tools for sampling, Parameter Estimation and Bayesian Model Comparison

Will Handley

wh260@cam.ac.uk

Supervisors: Anthony Lasenby & Mike Hobson

Astrophysics Group

Cavendish Laboratory

University of Cambridge

Monday 9^{th} May, 2016

Parameter estimation & model comparison

Metropolis Hastings

Nested Sampling

PolyChord

PolyChord 2.0

Examples

What is nested sampling?

What is nested sampling?

- Nested sampling is an alternative sampling approach.

What is nested sampling?

- Nested sampling is an alternative sampling approach.
- Unlike traditional methods, it does not sample from the posterior directly.

What is nested sampling?

- Nested sampling is an alternative sampling approach.
- Unlike traditional methods, it does not sample from the posterior directly.
- Instead it gradually compresses the prior onto the posterior.

What is nested sampling?

- Nested sampling is an alternative sampling approach.
- Unlike traditional methods, it does not sample from the posterior directly.
- Instead it gradually compresses the prior onto the posterior.
- In doing so, it circumvents many issues (dimensionality, topology, geometry) that beset normal approaches.

What is nested sampling?

- Nested sampling is an alternative sampling approach.
- Unlike traditional methods, it does not sample from the posterior directly.
- Instead it gradually compresses the prior onto the posterior.
- In doing so, it circumvents many issues (dimensionality, topology, geometry) that beset normal approaches.
- Similar to simulated annealing, but automatically picks the "correct" annealing schedule.

Bayes' theorem

Parameter estimation

Bayes' theorem

Parameter estimation

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

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

Bayes' theorem

Parameter estimation

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

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

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

Bayes' theorem

Parameter estimation

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

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

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

$$\mathcal{P}(\Theta) = \frac{\mathcal{L}(\Theta)\pi(\Theta)}{\mathcal{Z}}$$

Bayes' theorem

Model comparison

Bayes' theorem

Model comparison

What does data tell us about our model M_i?

Bayes' theorem

Model comparison

What does data tell us about our model M_i?

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

Bayes' theorem

Model comparison

What does data tell us about our model M_i?

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

$$\mathrm{P}(M_i|D) = \frac{\mathcal{Z}_i \; \mu_i}{\sum_k \mathcal{Z}_k \; \mu_k}$$

Bayes' theorem

Model comparison

What does data tell us about our model M_i?

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

$$\mathrm{P}(M_i|D) = \frac{\mathcal{Z}_i \, \mu_i}{\sum_k \mathcal{Z}_k \, \mu_k}$$

Model averaging:

- Multiple models with posterior on the same parameter: $\mathrm{P}(y|M_i, D)$

$$\mathrm{P}(y|D) = \sum_i \mathrm{P}(y|M_i, D)\mathrm{P}(M_i|D)$$

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...

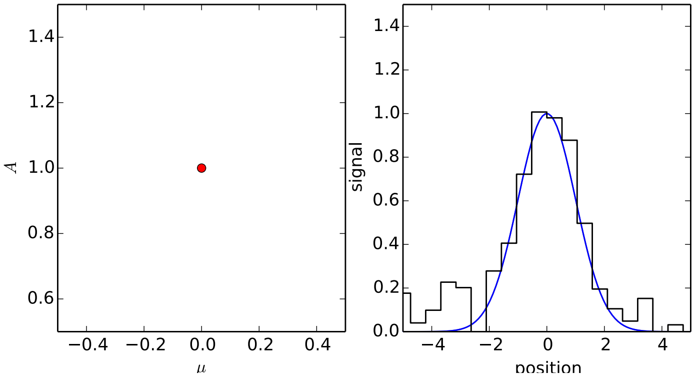
- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.
    1. Pick random direction
    2. Choose step length
    3. If uphill, make step...

Markov-Chain Monte-Carlo (MCMC)

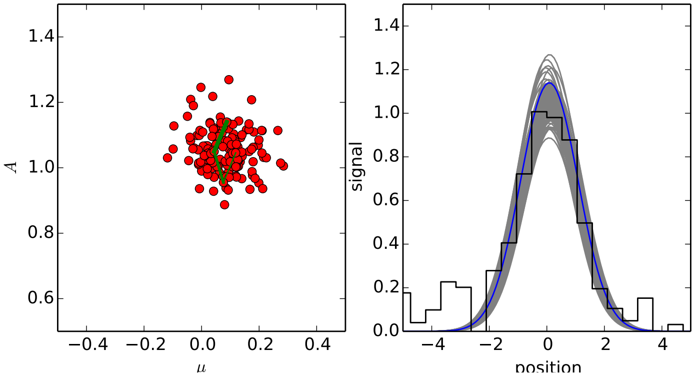
Metropolis-Hastings, Gibbs, Hamiltonian...

- 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

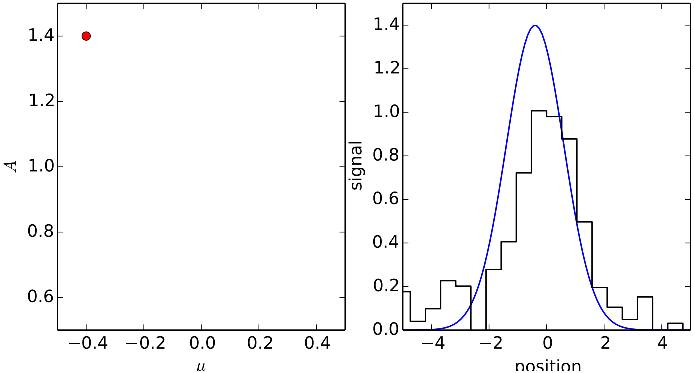


MCMC in action



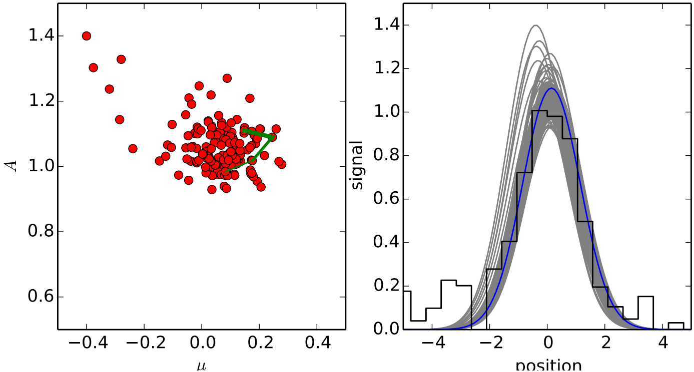
When MCMC fails

Burn in



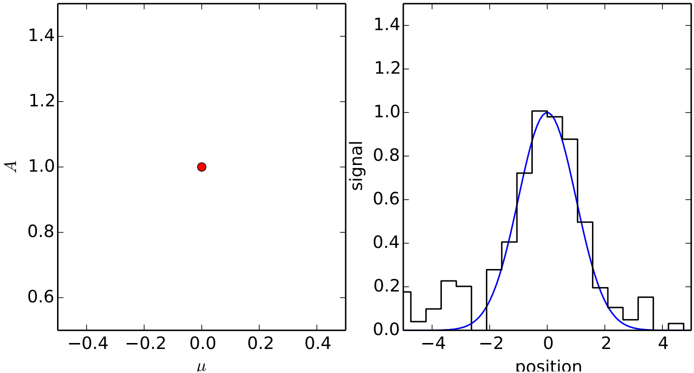
When MCMC fails

Burn in



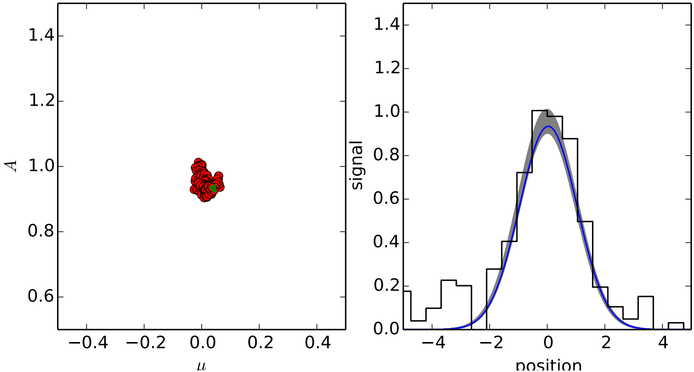
When MCMC fails

Tuning the proposal distribution



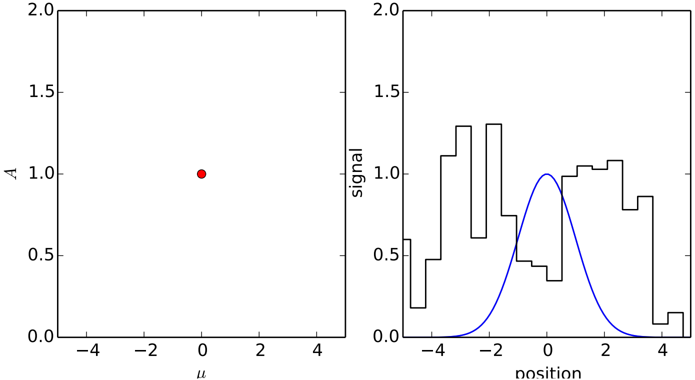
When MCMC fails

Tuning the proposal distribution



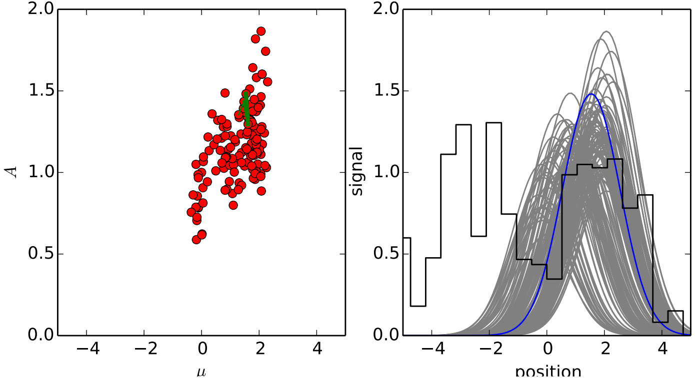
When MCMC fails

Multimodality



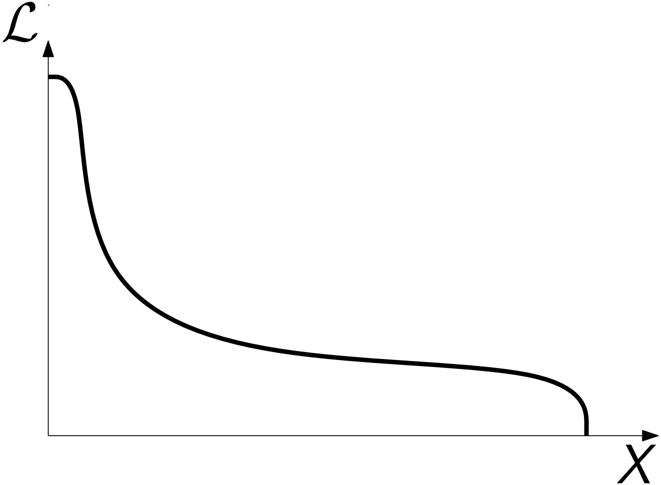
When MCMC fails

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!

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

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

When MCMC fails

The real reason...

- MCMC does not give you evidences!

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

- MCMC fundamentally explores the posterior, and cannot average over the prior.
- Simulated annealing gives one possibility for computing evidences.

When MCMC fails

The real reason...

- MCMC does not give you evidences!

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

- MCMC fundamentally explores the posterior, and cannot average over the prior.
- Simulated annealing gives one possibility for computing evidences.
  - Suffers from similar issues to MCMC, especially phase transitions

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.
- Simulated annealing gives one possibility for computing evidences.
  - Suffers from similar issues to MCMC, especially phase transitions

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 uniformly over the space (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 uniformly over the space (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:

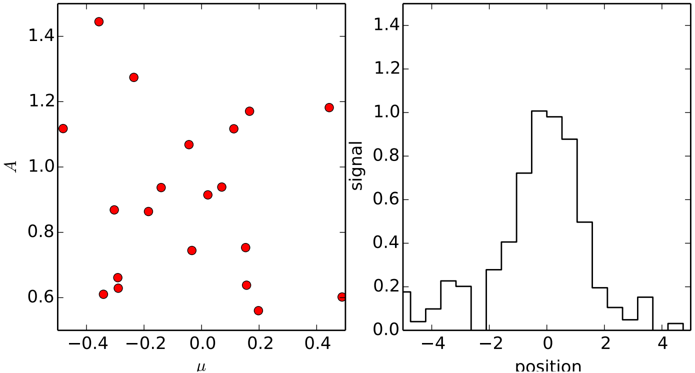
S_0: Generate n samples uniformly over the space (from the prior π).

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 uniformly within a region, subject to a *hard likelihood constraint*.

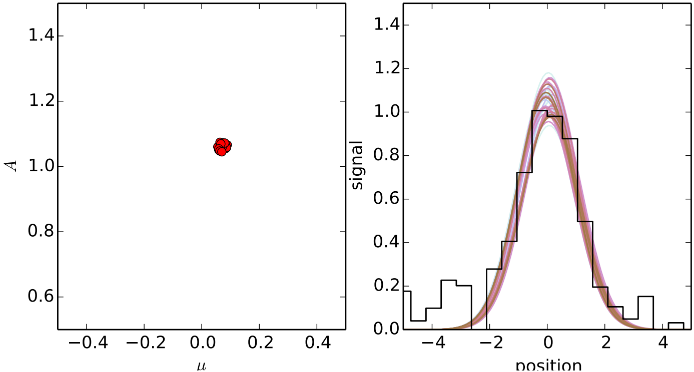
Nested sampling

Unimodal



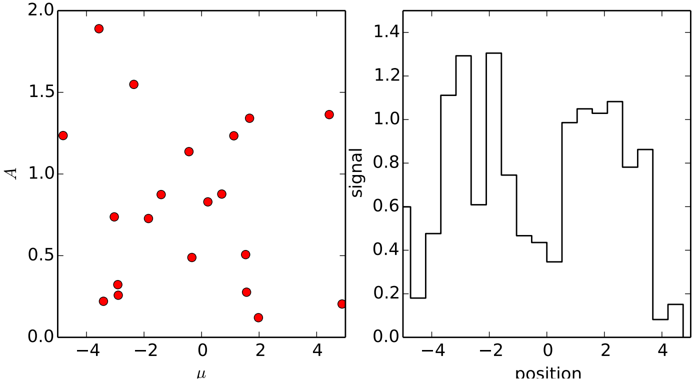
Nested sampling

Unimodal



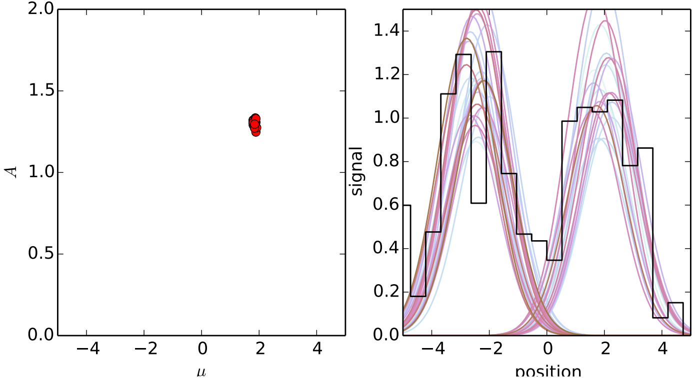
Nested sampling

Multimodal



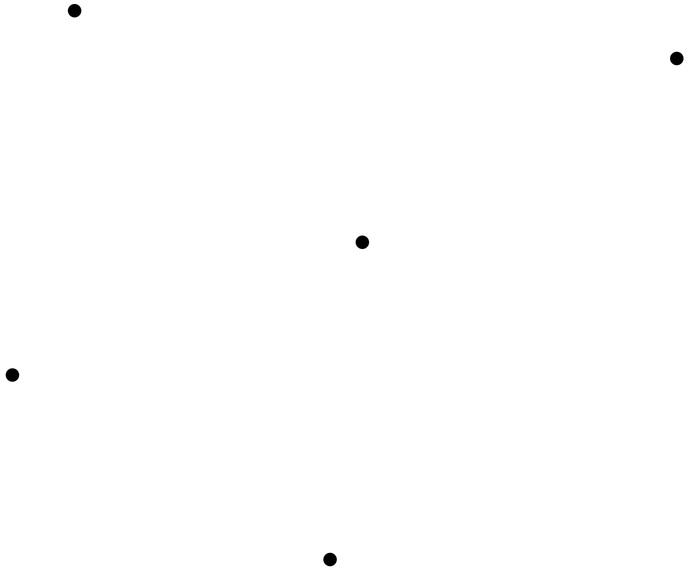
When NS suceeds

Multimodal



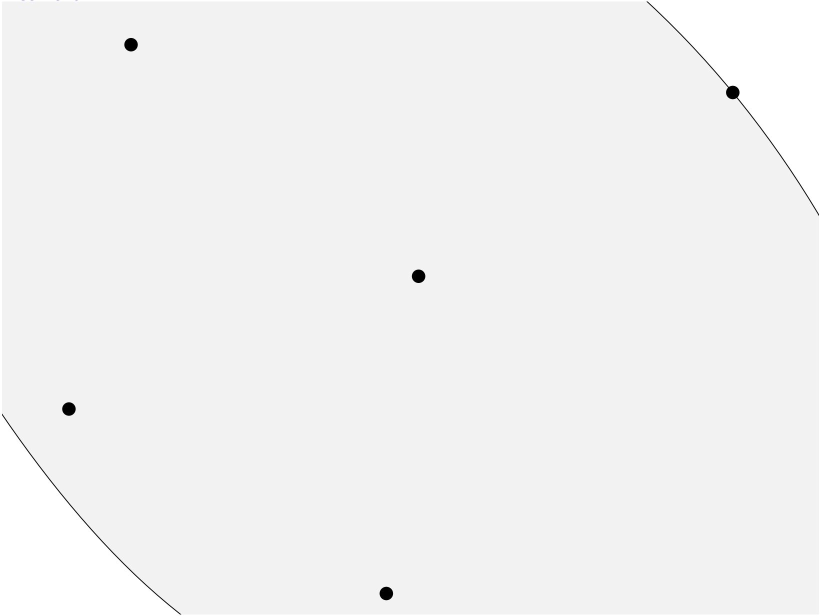
Nested Sampling

Graphical aid



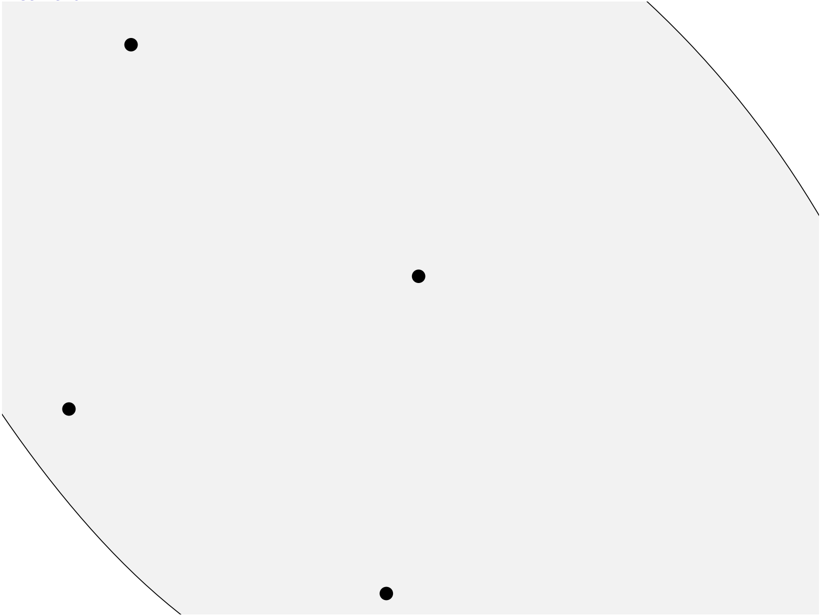
Nested Sampling

Graphical aid



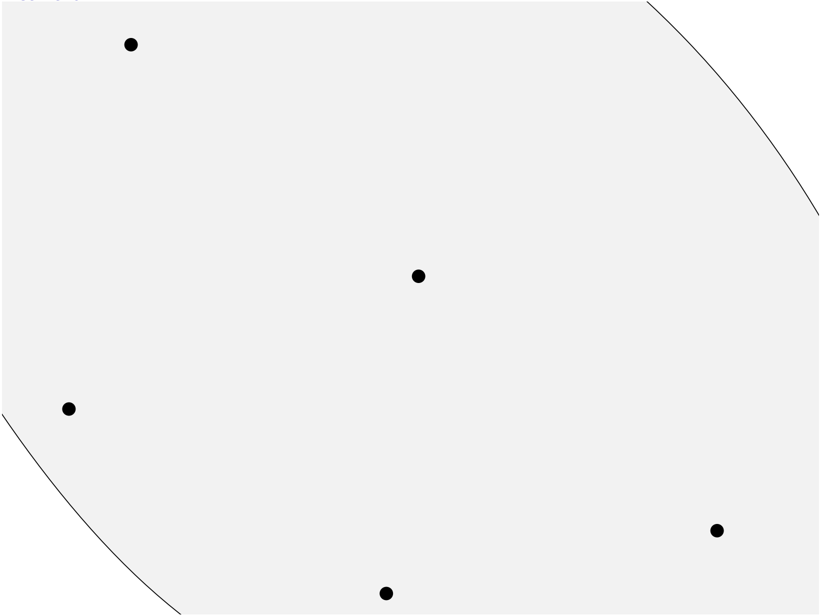
Nested Sampling

Graphical aid



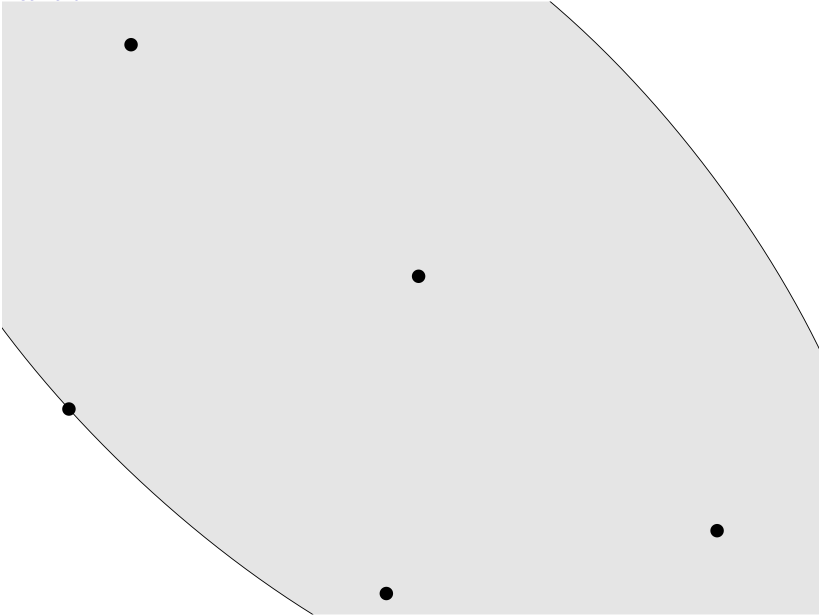
Nested Sampling

Graphical aid



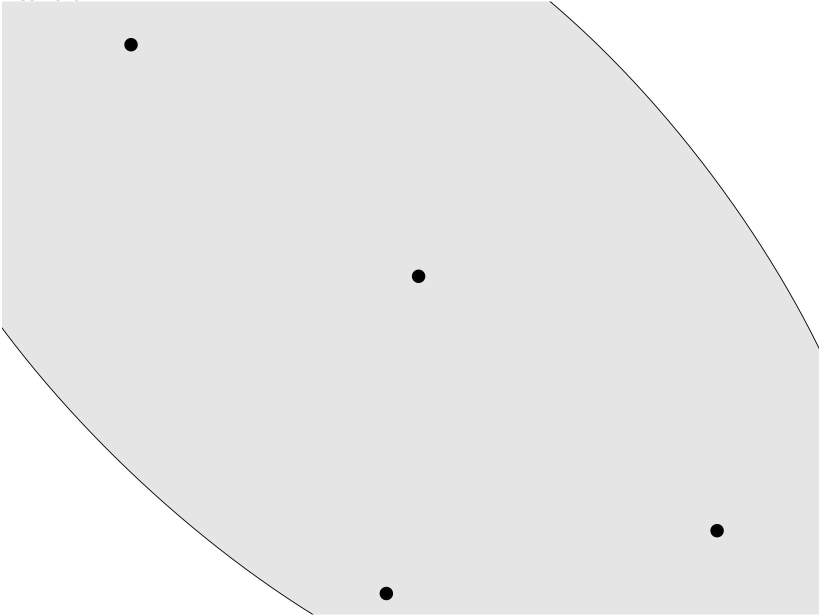
Nested Sampling

Graphical aid



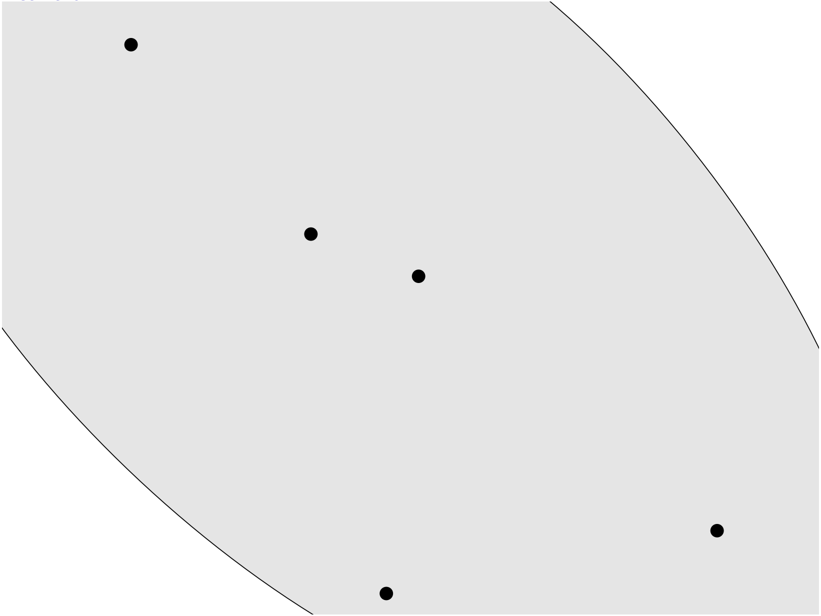
Nested Sampling

Graphical aid



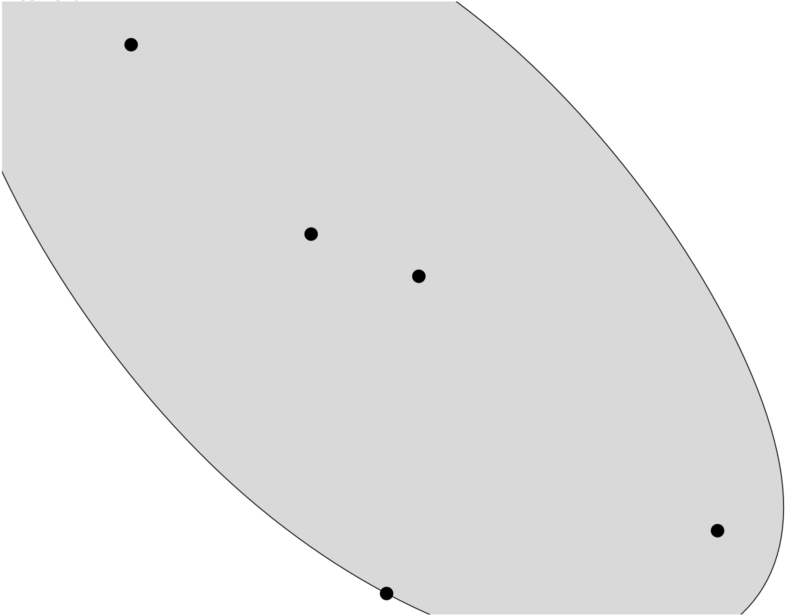
Nested Sampling

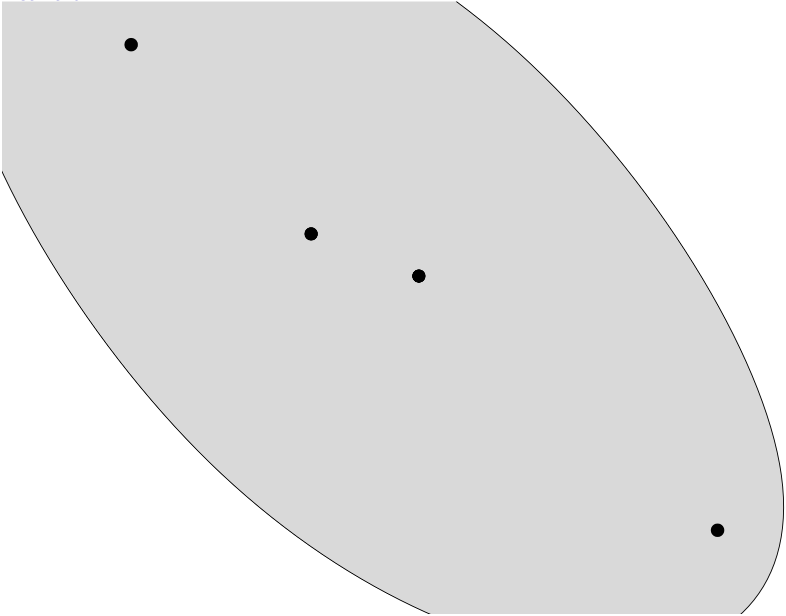
Graphical aid



Nested Sampling

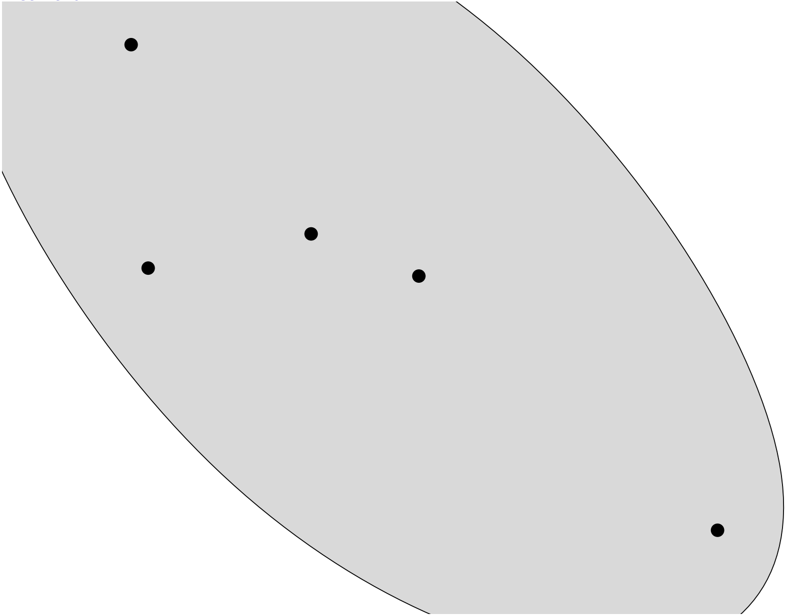
Graphical aid





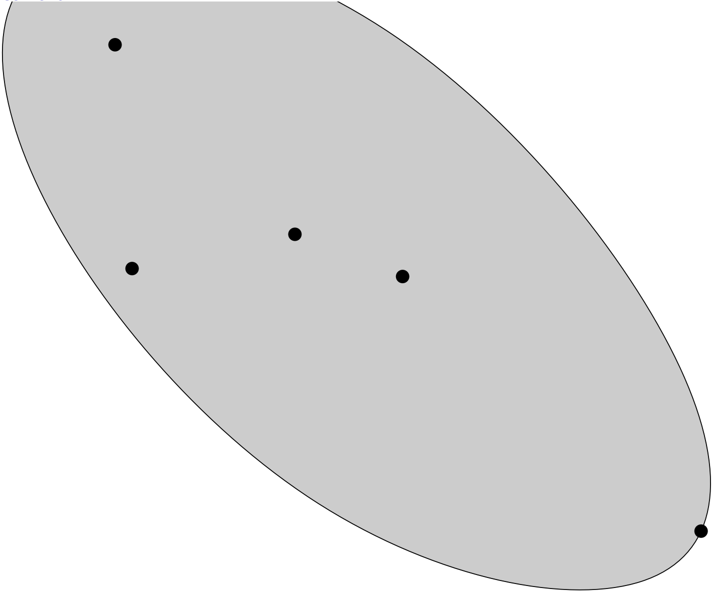
Nested Sampling

Graphical aid



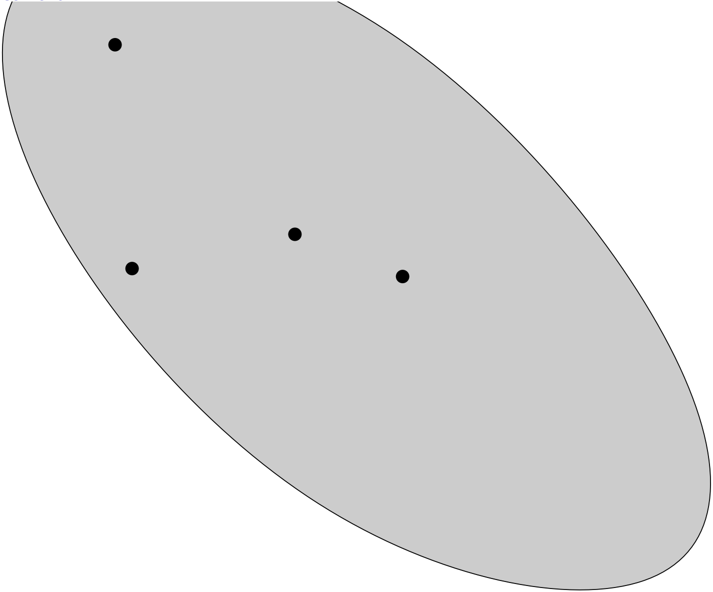
Nested Sampling

Graphical aid



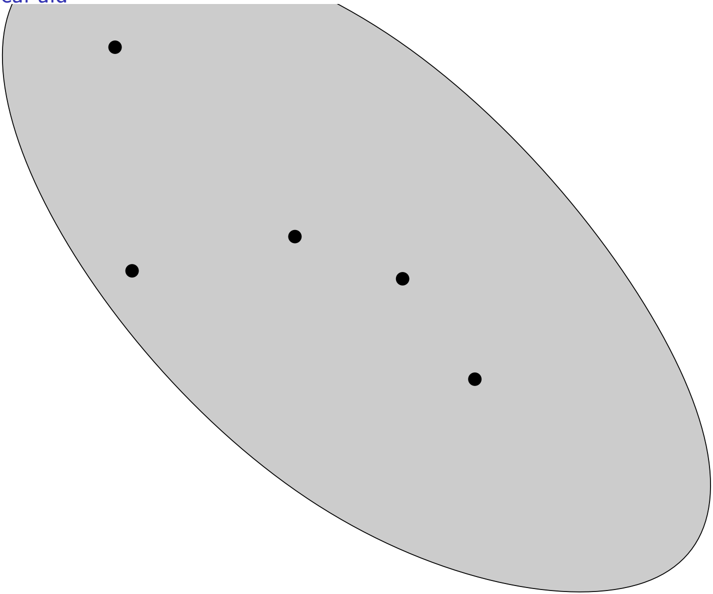
Nested Sampling

Graphical aid



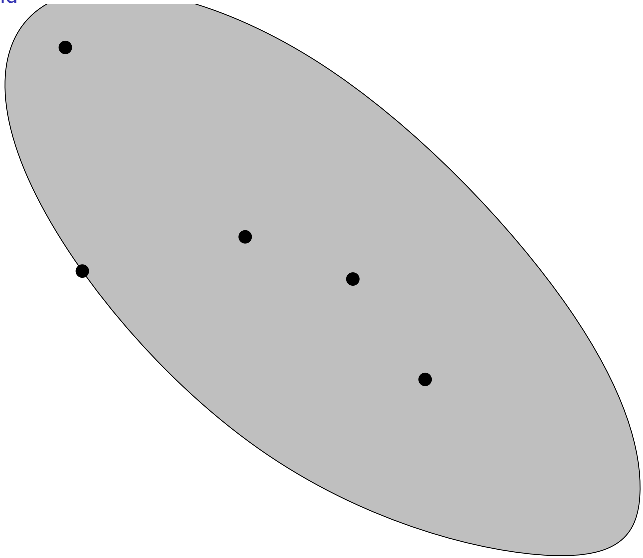
Nested Sampling

Graphical aid



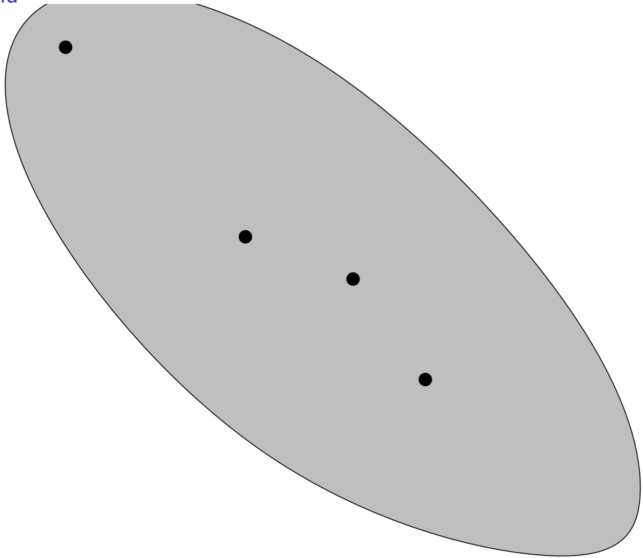
Nested Sampling

Graphical aid



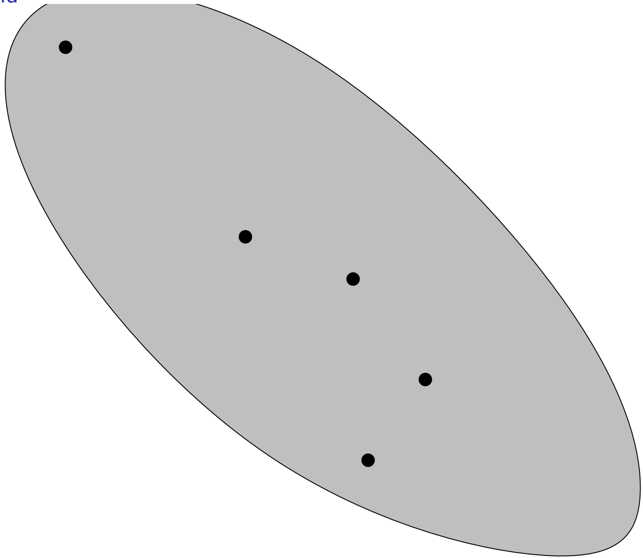
Nested Sampling

Graphical aid



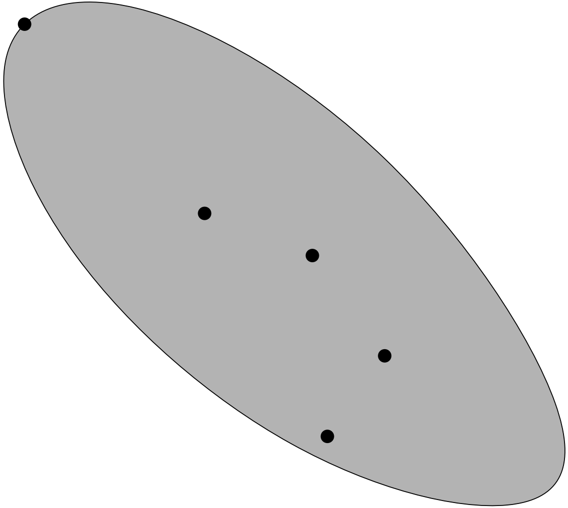
Nested Sampling

Graphical aid



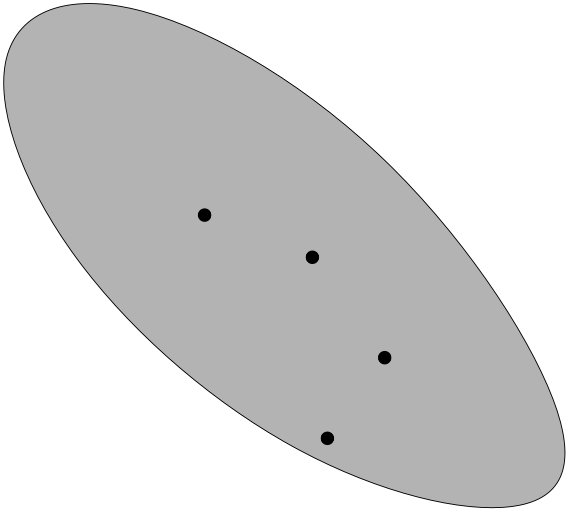
Nested Sampling

Graphical aid



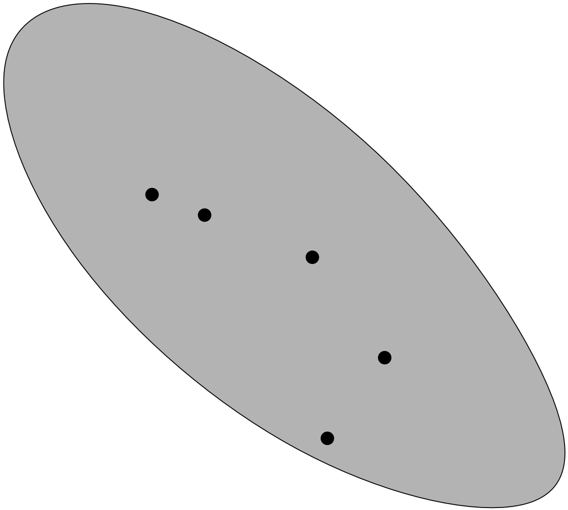
Nested Sampling

Graphical aid



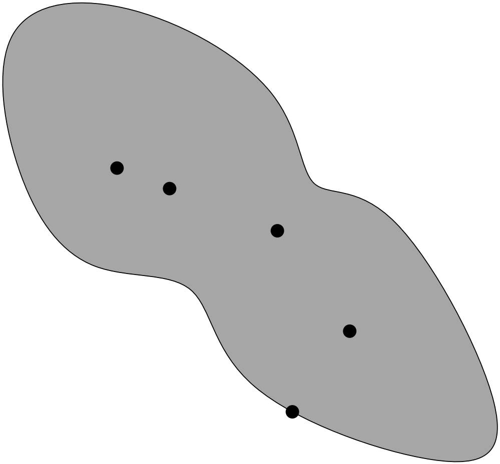
Nested Sampling

Graphical aid



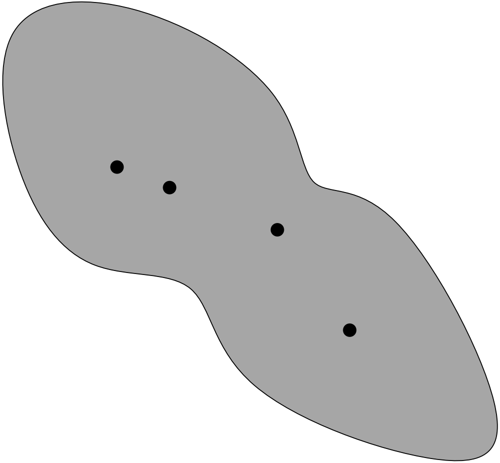
Nested Sampling

Graphical aid



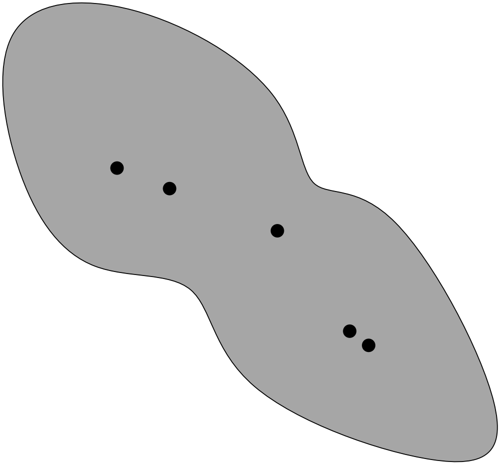
Nested Sampling

Graphical aid



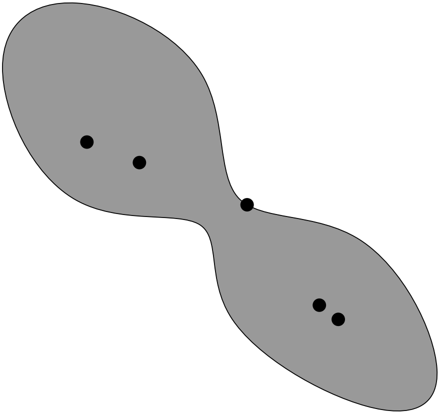
Nested Sampling

Graphical aid



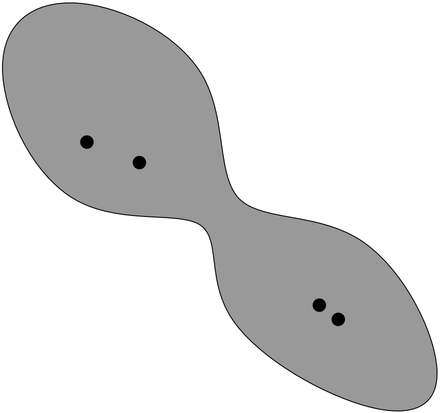
Nested Sampling

Graphical aid



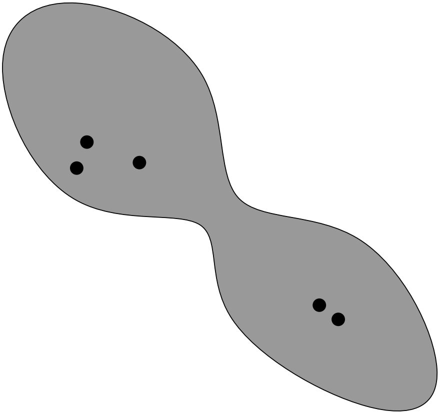
Nested Sampling

Graphical aid



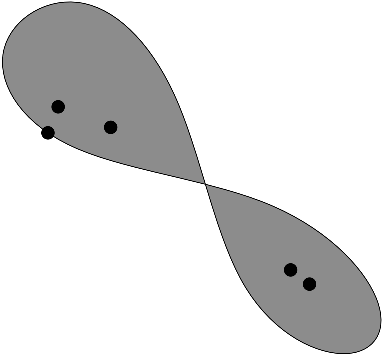
Nested Sampling

Graphical aid



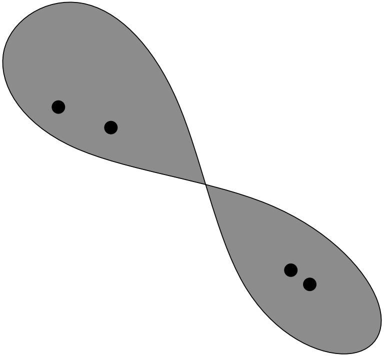
Nested Sampling

Graphical aid



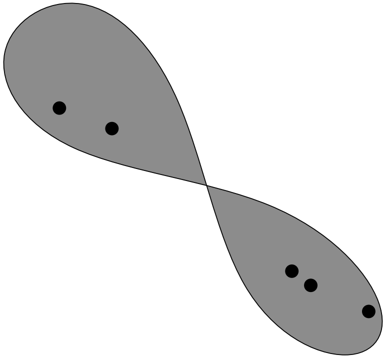
Nested Sampling

Graphical aid



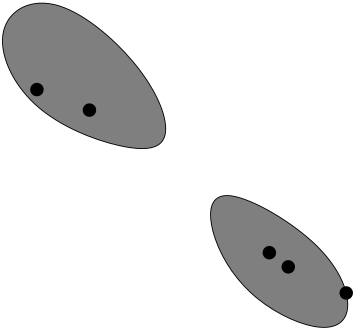
Nested Sampling

Graphical aid



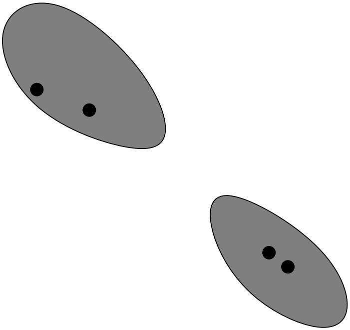
Nested Sampling

Graphical aid



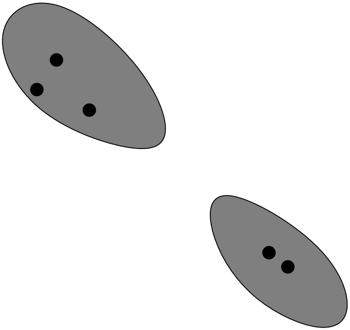
Nested Sampling

Graphical aid



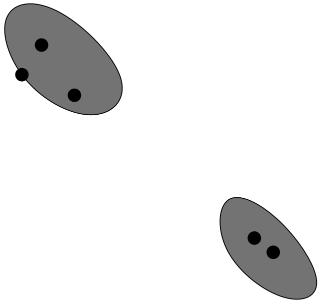
Nested Sampling

Graphical aid



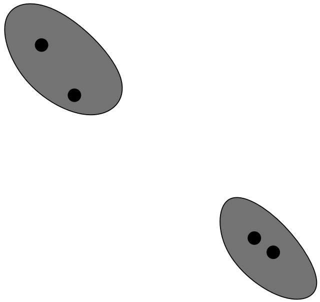
Nested Sampling

Graphical aid



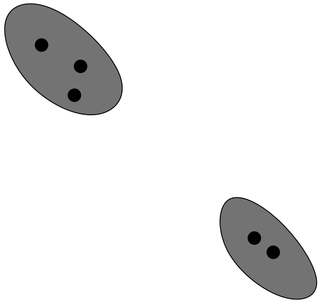
Nested Sampling

Graphical aid



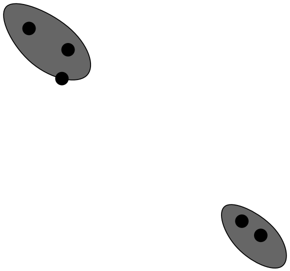
Nested Sampling

Graphical aid



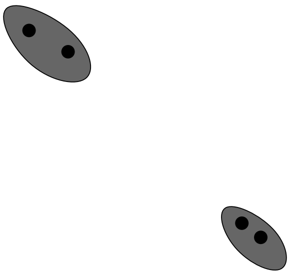
Nested Sampling

Graphical aid



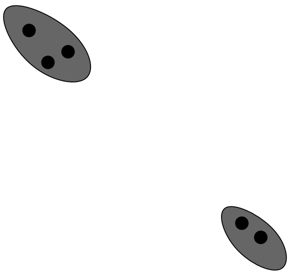
Nested Sampling

Graphical aid



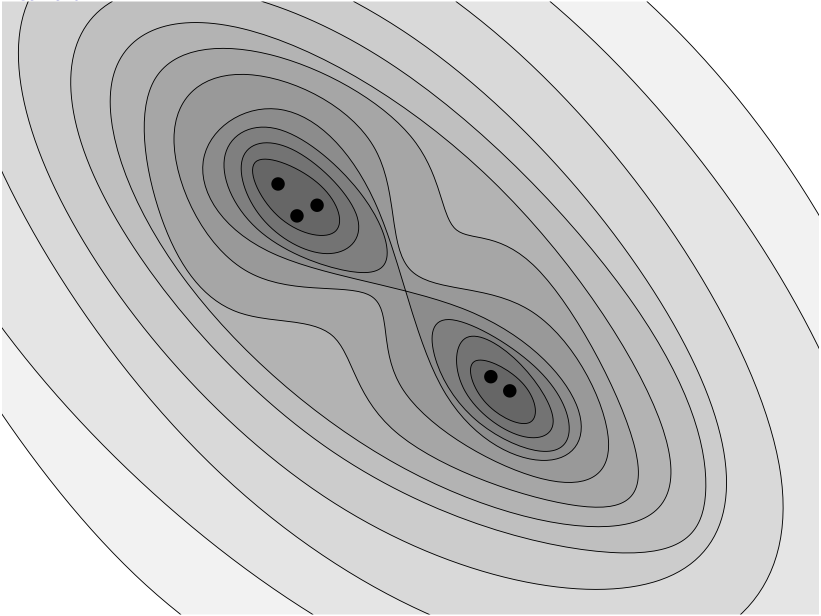
Nested Sampling

Graphical aid



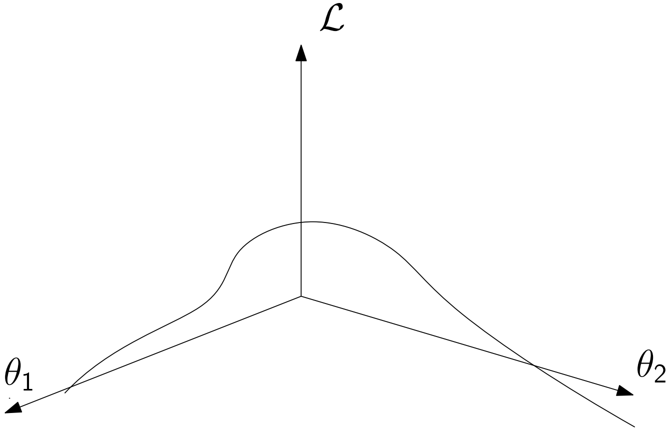
Nested Sampling

Graphical aid



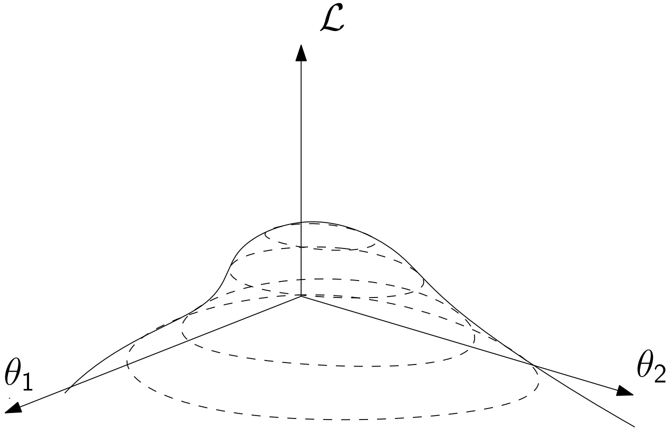
Nested Sampling

Calculating evidences



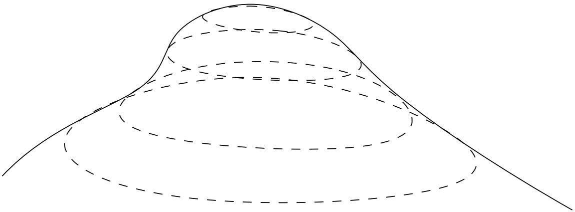
Nested Sampling

Calculating evidences



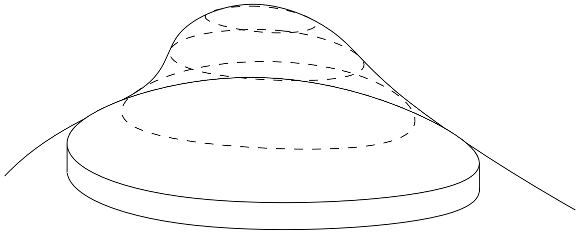
Nested Sampling

Calculating evidences



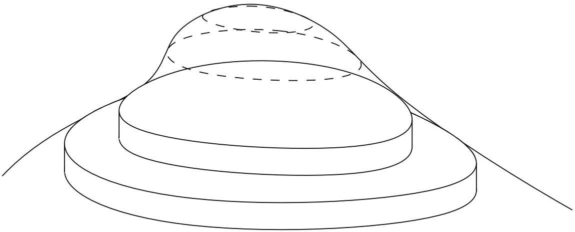
Nested Sampling

Calculating evidences



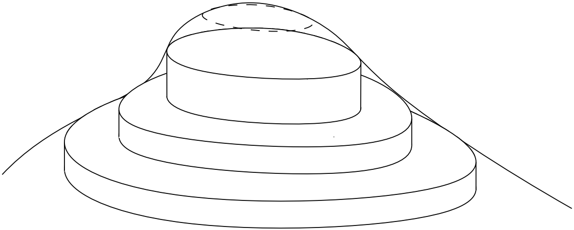
Nested Sampling

Calculating evidences



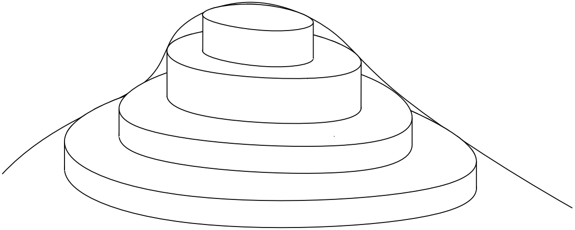
Nested Sampling

Calculating evidences



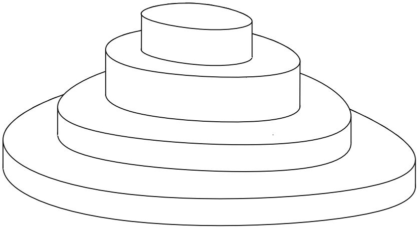
Nested Sampling

Calculating evidences



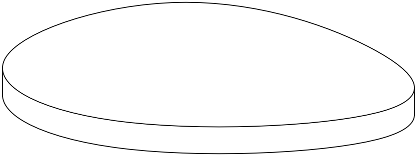
Nested Sampling

Calculating evidences



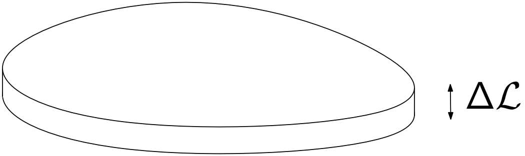
Nested Sampling

Calculating evidences



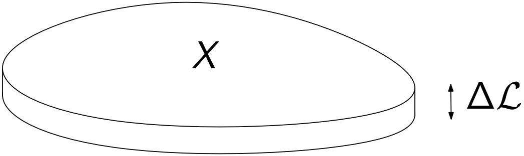
Nested Sampling

Calculating evidences



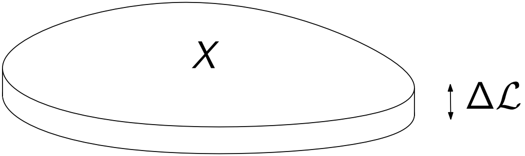
Nested Sampling

Calculating evidences



Nested Sampling

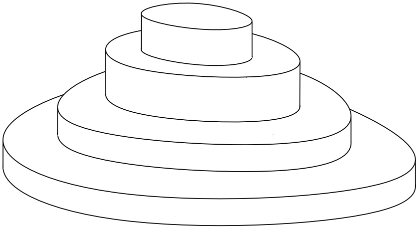
Calculating evidences



Volume $= X\Delta\mathcal{L}$

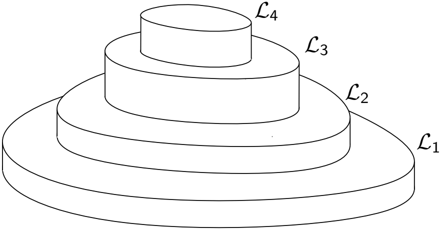
Nested Sampling

Calculating evidences



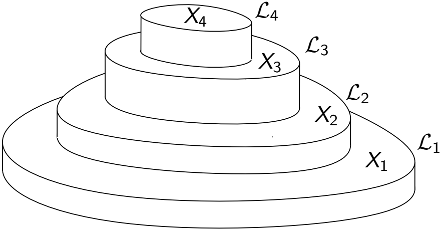
Nested Sampling

Calculating evidences



Nested Sampling

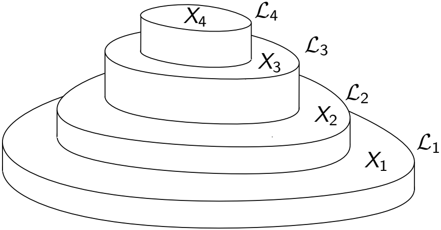
Calculating evidences



Nested Sampling

Calculating evidences

$$\mathcal{Z} \approx \sum_i X_i \Delta\mathcal{L}_i$$



Nested Sampling

Exponential volume contraction

Nested Sampling

Exponential volume contraction

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

Nested Sampling

Exponential volume contraction

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

Nested Sampling

Exponential volume contraction

- At each iteration, the likelihood contour will shrink in volume by $\approx 1/n$.
- Nested sampling zooms in to the peak of the posterior *exponentially*.
- $$\mathcal{Z} \approx \sum_i \Delta\mathcal{L}_i X_i \tag{1}$$

Nested Sampling

Exponential volume contraction

- At each iteration, the likelihood contour will shrink in volume by $\approx 1/n$.
- Nested sampling zooms in to the peak of the posterior *exponentially*.
- $$\mathcal{Z} \approx \sum_i \Delta\mathcal{L}_i X_i \tag{1}$$
- $$X_{i+1} \approx \frac{n}{n+1} X_i, \qquad X_0 = 1 \tag{2}$$

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

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 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*

- Most of the work in NS to date has been in attempting to implement a hard-edged sampler in the NS meta-algorithm.

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

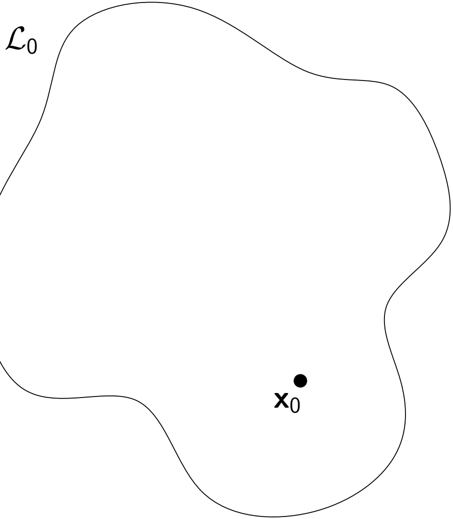
Hamiltonian sampling F. Feroz & J. Skilling (2013).

- Requires gradients and tuning

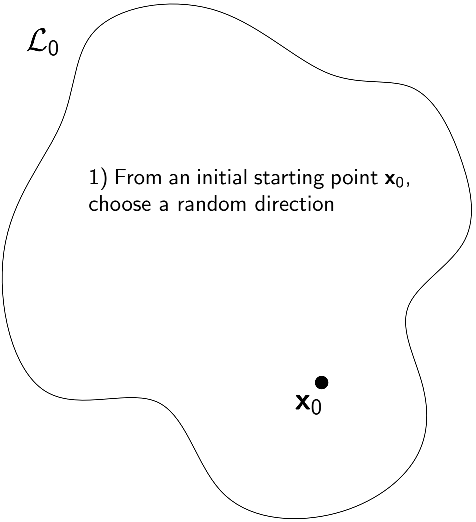
Diffusion Nested Sampling B. Brewer et al. (2009).

- Very promising
- Too many tuning parameters

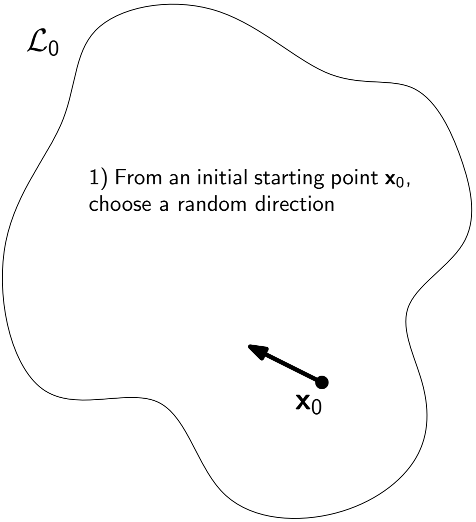
"Hit and run" slice sampling



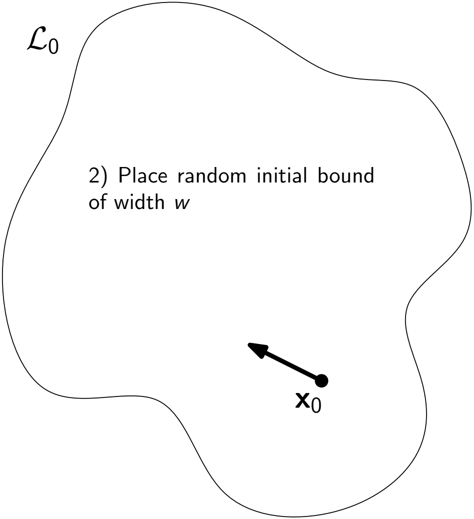
“Hit and run” slice sampling



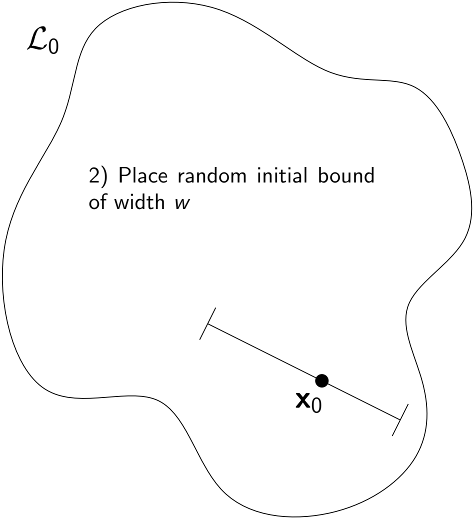
"Hit and run" slice sampling



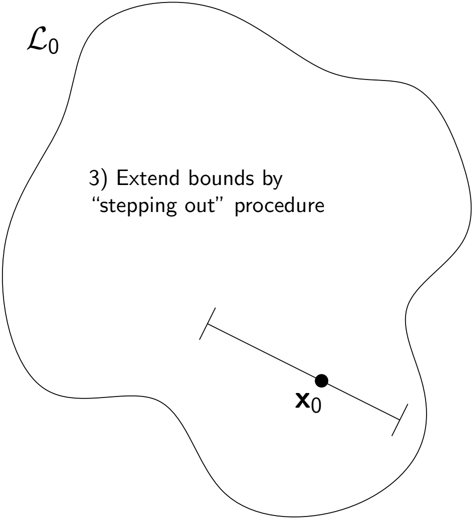
"Hit and run" slice sampling



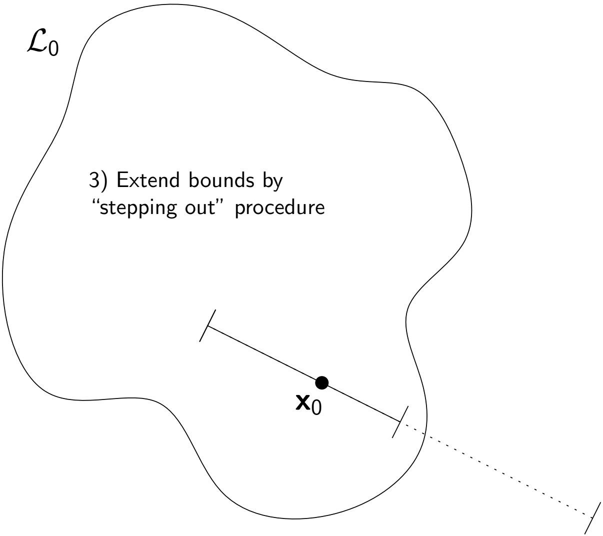
“Hit and run” slice sampling



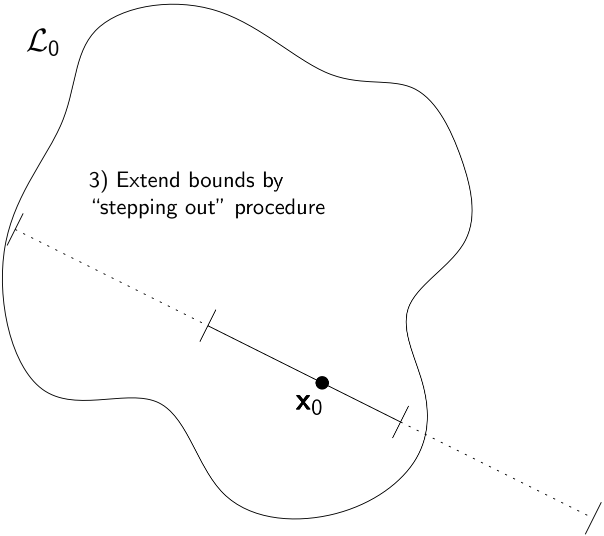
"Hit and run" slice sampling



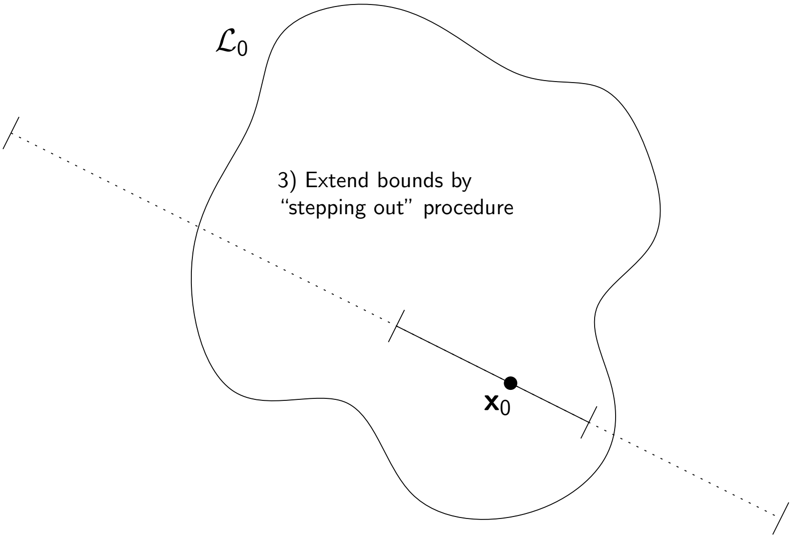
“Hit and run” slice sampling



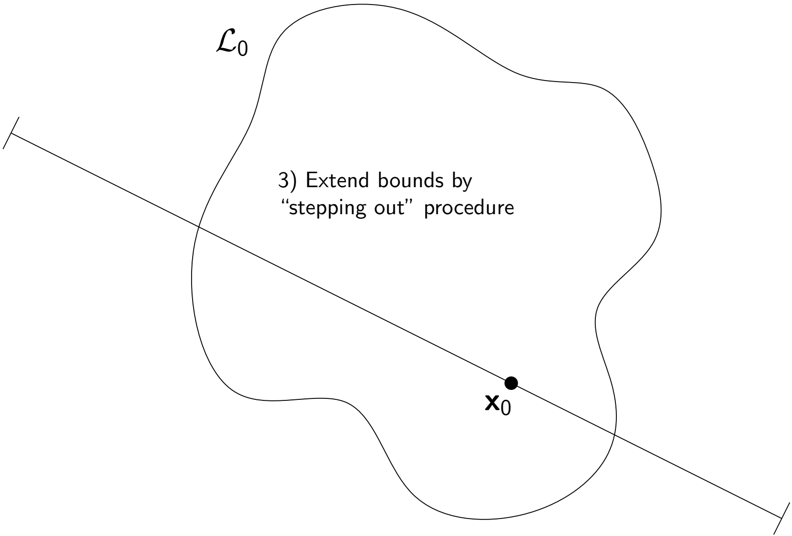
"Hit and run" slice sampling



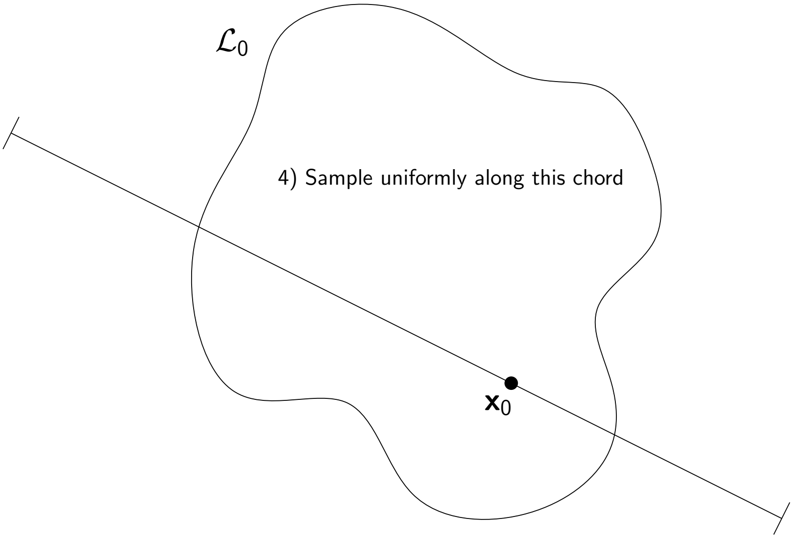
"Hit and run" slice sampling



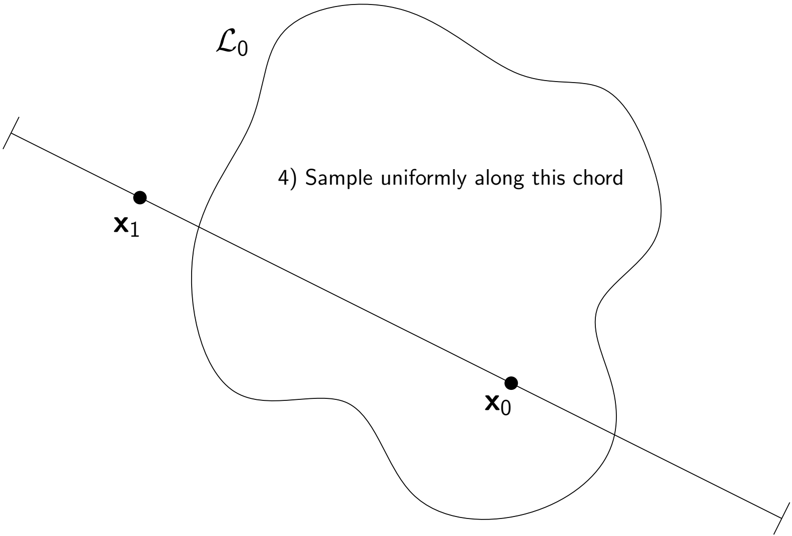
“Hit and run” slice sampling



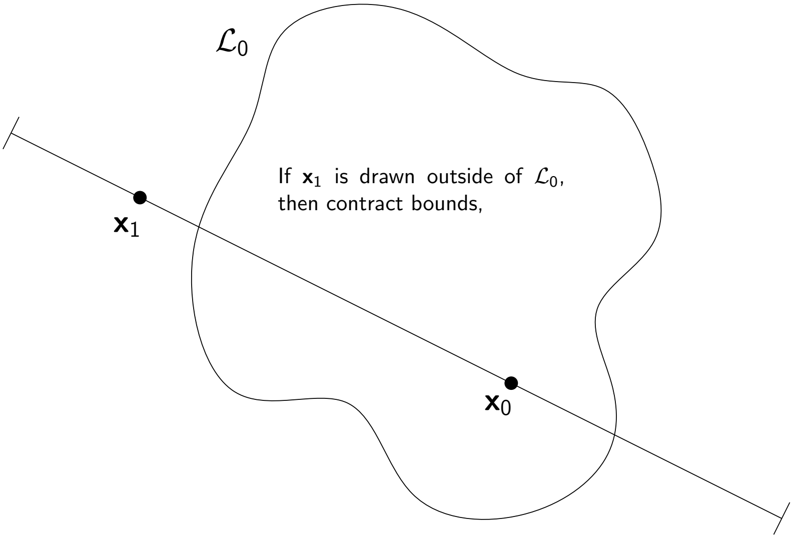
"Hit and run" slice sampling



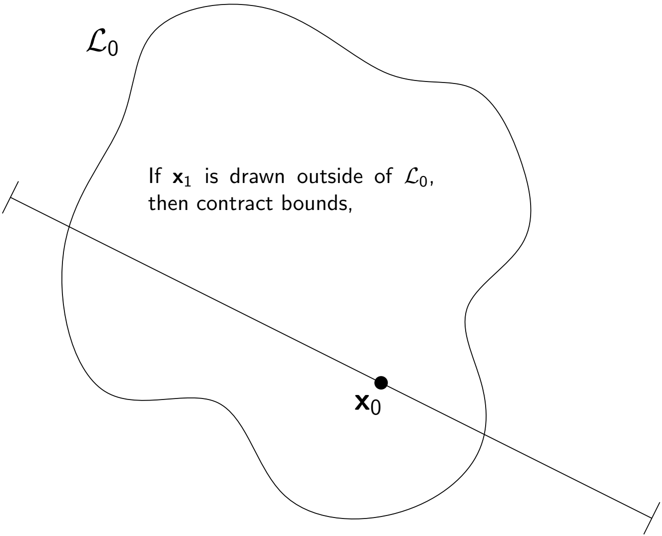
"Hit and run" slice sampling



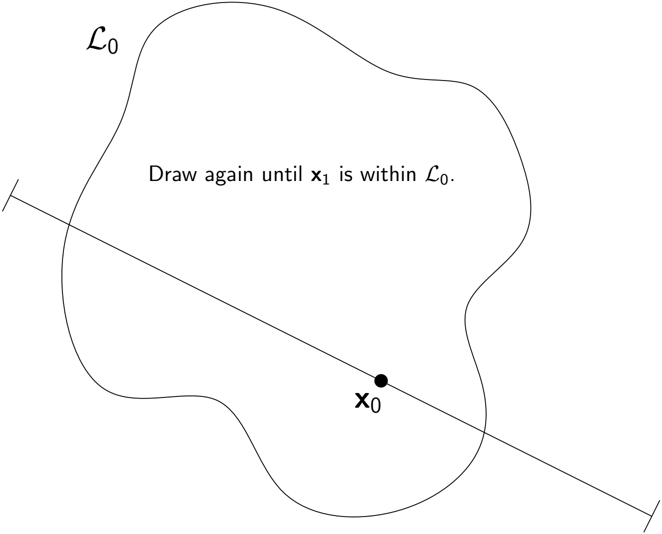
"Hit and run" slice sampling



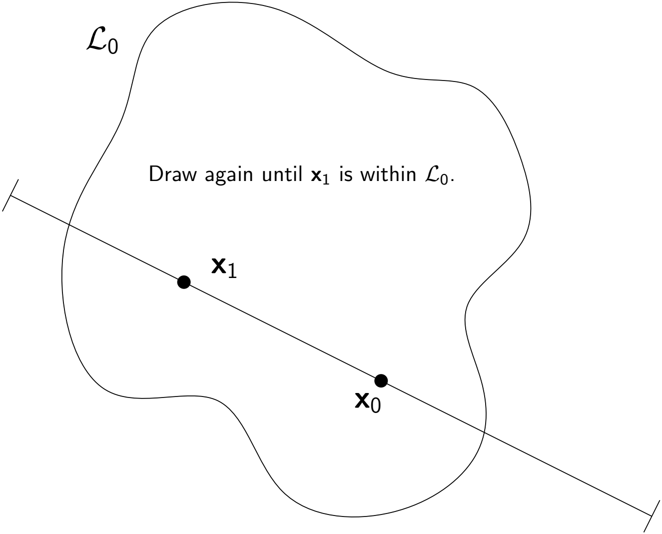
"Hit and run" slice sampling



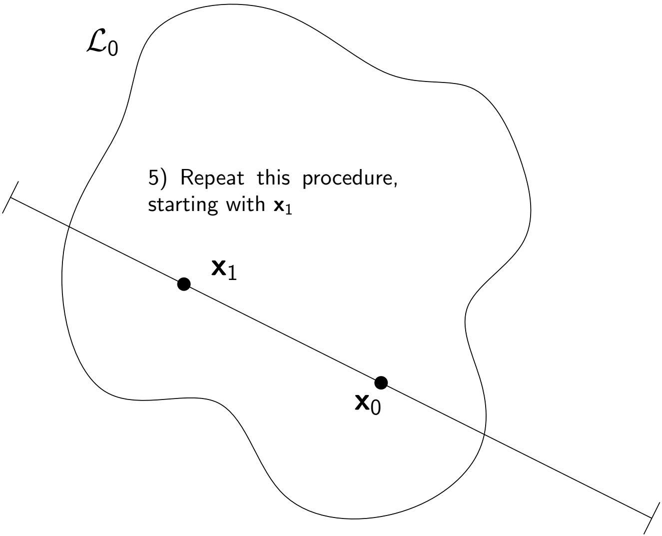
“Hit and run” slice sampling



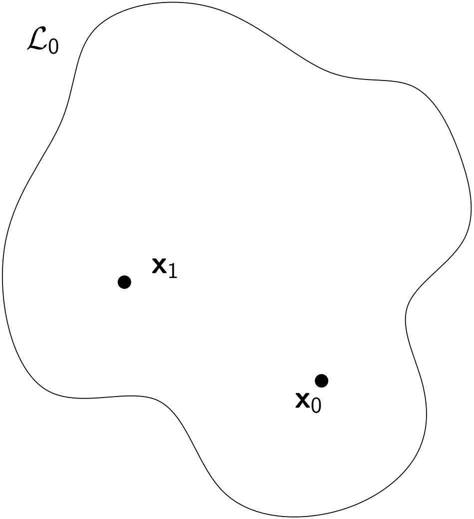
“Hit and run” slice sampling



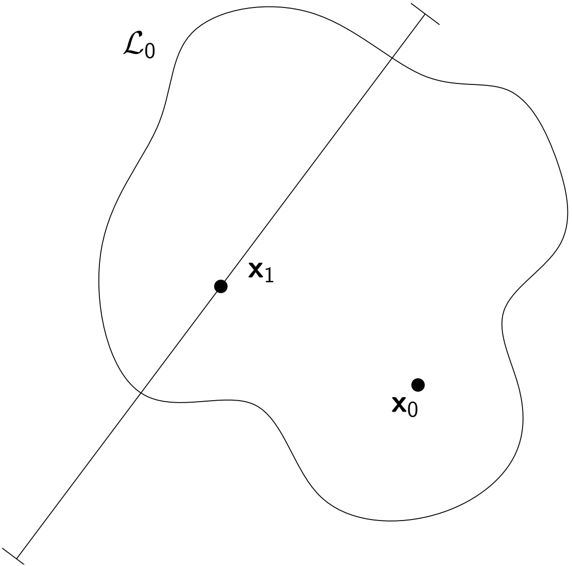
"Hit and run" slice sampling



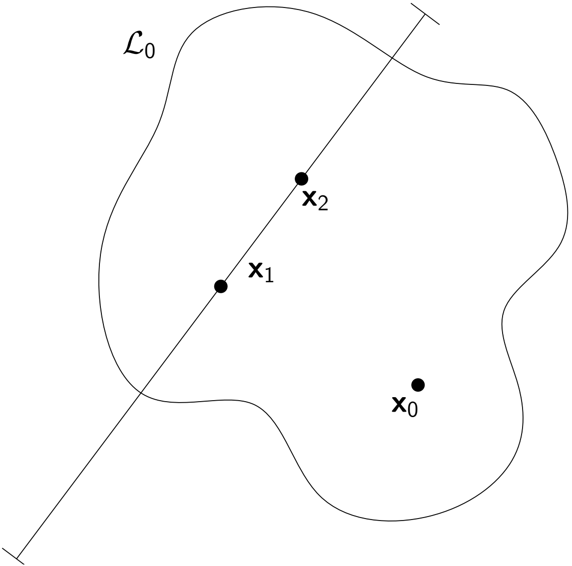
"Hit and run" slice sampling



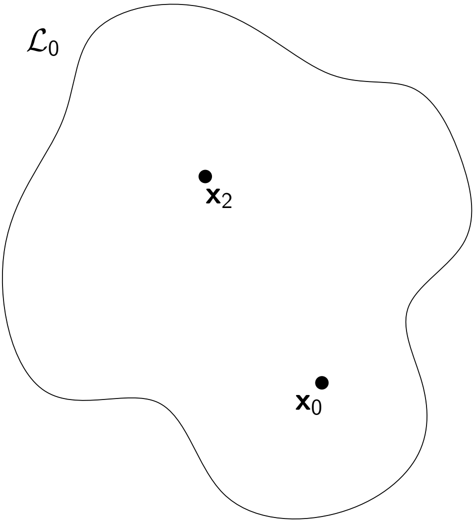
"Hit and run" slice sampling



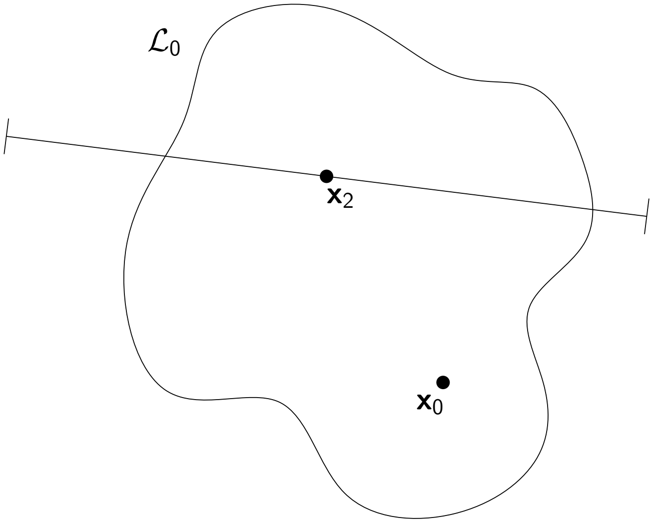
"Hit and run" slice sampling



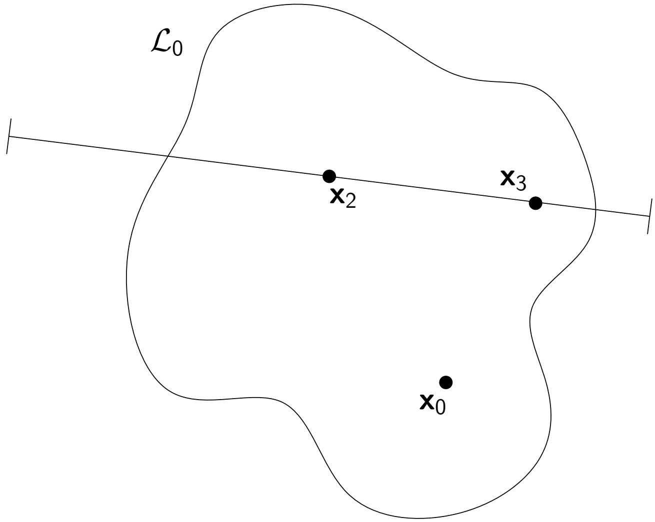
"Hit and run" slice sampling



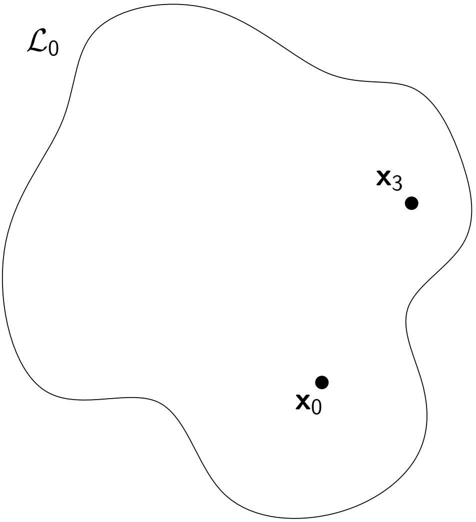
“Hit and run” slice sampling



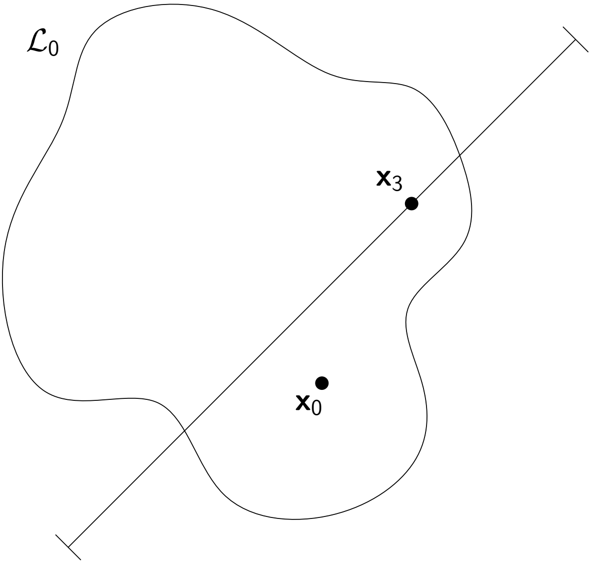
"Hit and run" slice sampling



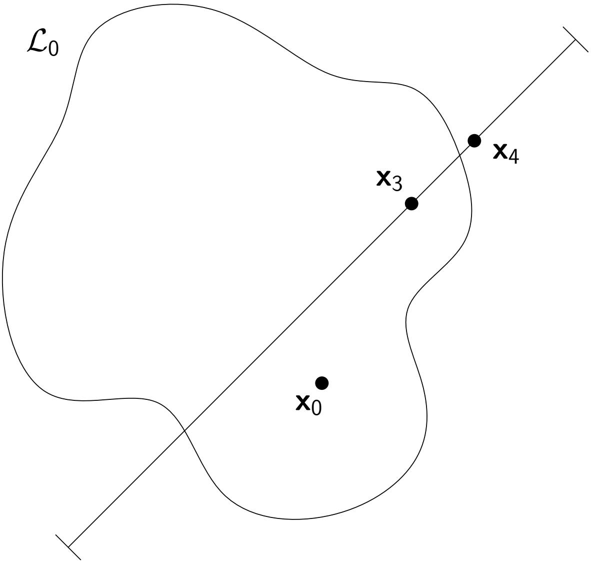
"Hit and run" slice sampling



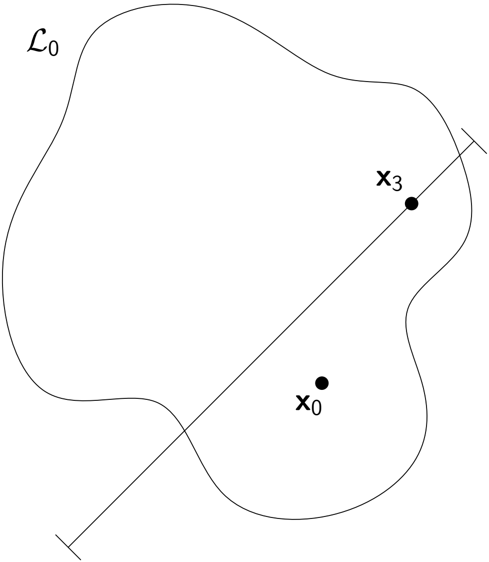
"Hit and run" slice sampling



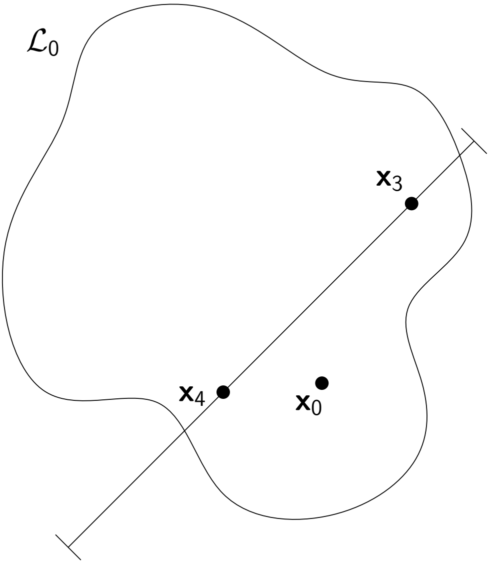
"Hit and run" slice sampling



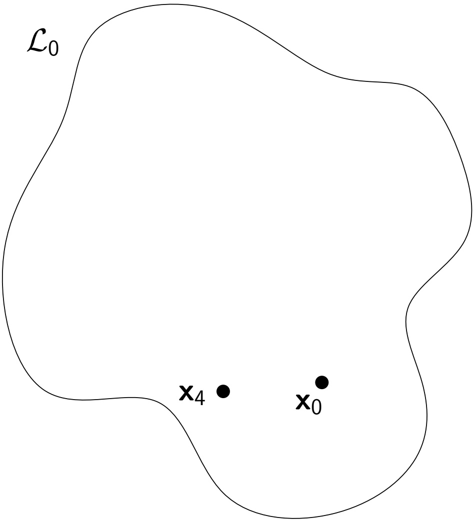
“Hit and run” slice sampling



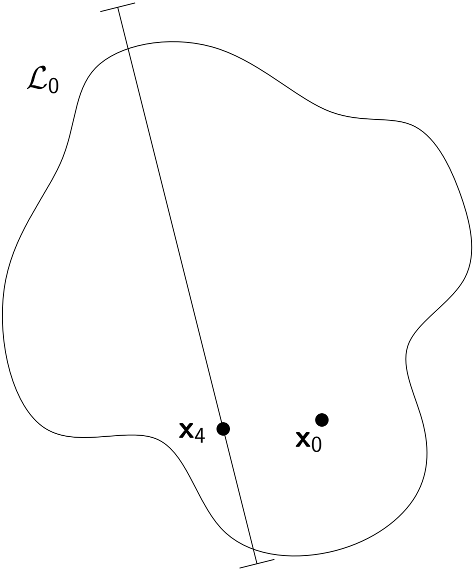
"Hit and run" slice sampling



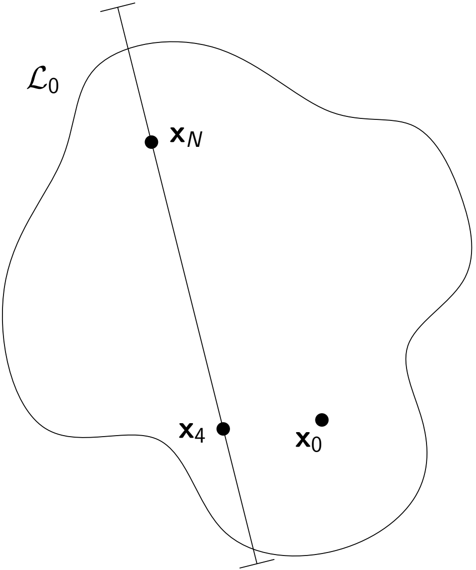
"Hit and run" slice sampling



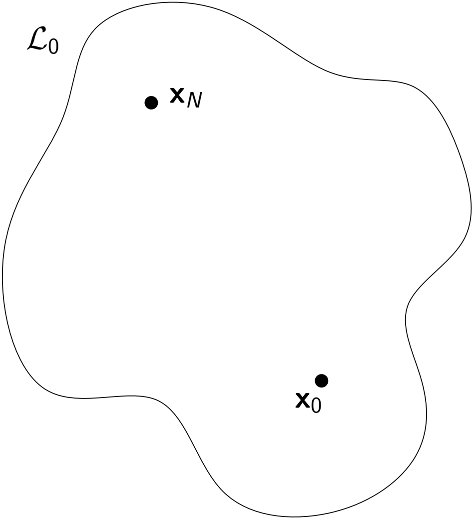
“Hit and run” slice sampling



“Hit and run” slice sampling



"Hit and run" slice sampling



“Hit and run” slice sampling

Key points

“Hit and run” slice sampling

Key points

- This procedure satisfies detailed balance.

"Hit and run" slice sampling

Key points

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

Issues with Slice Sampling

Correlated distributions

Issues with Slice Sampling

Correlated distributions

1. Does not deal well with correlated distributions.

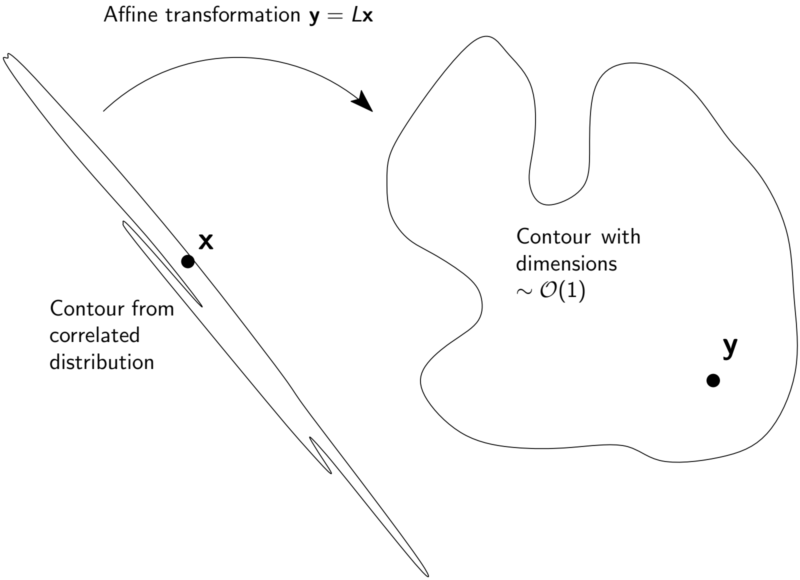
Issues with Slice Sampling

Correlated distributions

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

Issues with Slice Sampling

Multimodality

Issues with Slice Sampling

Multimodality

1. Although it satisfies detailed balance practically this isn't good enough.

Issues with Slice Sampling

Multimodality

1. Although it satisfies detailed balance practically this isn't good enough.
2. Affine transformation is useless.

PolyChord's solutions

Multimodality

PolyChord's solutions

Multimodality

1. Identifies separate modes via clustering algorithm on live points.

PolyChord's solutions

Multimodality

1. Identifies separate modes via clustering algorithm on live points.
2. Evolves these modes "semi-independently"

PolyChord's Additions

PolyChord's Additions

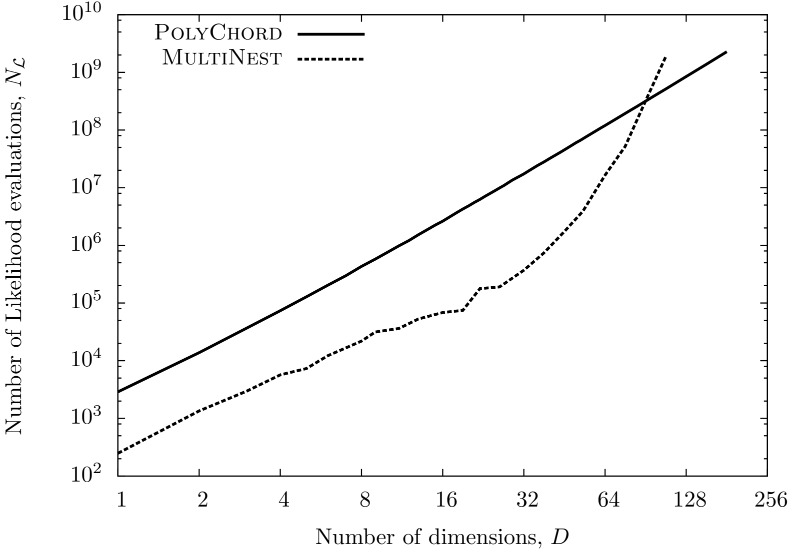
- Parallelised up to number of live points with openMPI.

PolyChord's Additions

- Parallelised up to number of live points with openMPI.
- Implemented in CosmoMC, as "CosmoChord", with fast-slow parameters.

PolyChord vs. MultiNest

Gaussian likelihood



PolyChord 1.0

PolyChord 1.0

- Well tested.

PolyChord 1.0

- Well tested.
- arXiv:1502.01856

PolyChord 1.0

- Well tested.
- arXiv:1502.01856
- arXiv:1506.00171

PolyChord 1.0

- Well tested.
- arXiv:1502.01856
- arXiv:1506.00171
- `ccpforge.cse.rl.ac.uk/gf/project/polychord/`

PolyChord

Scaling with dimensionality

PolyChord

Scaling with dimensionality

- Perfect nested sampling has $N_{\mathcal{L}} \sim \mathcal{O}(D^2)$

PolyChord

Scaling with dimensionality

- Perfect nested sampling has $N_{\mathcal{L}} \sim \mathcal{O}(D^2)$
- PolyChord 1.0 has $N_{\mathcal{L}} \sim \mathcal{O}(D^3)$

PolyChord

Scaling with dimensionality

- Perfect nested sampling has $N_{\mathcal{L}} \sim \mathcal{O}(D^2)$
- PolyChord 1.0 has $N_{\mathcal{L}} \sim \mathcal{O}(D^3)$
  - Need $\sim \mathcal{O}(D)$ to de-correlate at each step

PolyChord

Scaling with dimensionality

- Perfect nested sampling has $N_{\mathcal{L}} \sim \mathcal{O}(D^2)$
- PolyChord 1.0 has $N_{\mathcal{L}} \sim \mathcal{O}(D^3)$
  - Need $\sim \mathcal{O}(D)$ to de-correlate at each step
  - Forced to throw $\sim \mathcal{O}(D)$ inter-chain points away.

PolyChord 2.0

Inter-chain evaluations

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.
- MCMC-like approaches generate many (correlated) samples

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.
- MCMC-like approaches generate many (correlated) samples
- At each iteration:

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.
- MCMC-like approaches generate many (correlated) samples
- At each iteration:
  - Throw away n_{chain} samples.

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.
- MCMC-like approaches generate many (correlated) samples
- At each iteration:
  - Throw away n_{chain} samples.
  - Generate n_{chain} new (correlated) samples to replace them.

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.
- MCMC-like approaches generate many (correlated) samples
- At each iteration:
  - Throw away n_{chain} samples.
  - Generate n_{chain} new (correlated) samples to replace them.
- If $n_{\mathrm{chain}} \sim \mathcal{O}(D)$ (as required), then overall $\sim \mathcal{O}(D^3) \to \sim \mathcal{O}(D^2)$.

PolyChord 2.0

Inter-chain evaluations

- Traditional nested sampling requires the ability to easily generate a single sample at each iteration.
- MCMC-like approaches generate many (correlated) samples
- At each iteration:
  - Throw away n_{chain} samples.
  - Generate n_{chain} new (correlated) samples to replace them.
- If $n_{\mathrm{chain}} \sim \mathcal{O}(D)$ (as required), then overall $\sim \mathcal{O}(D^3) \to \sim \mathcal{O}(D^2)$.
- Need to be able to quantify degree of correlation for correct inference.

Aside: Merging nested sampling runs

Aside: Merging nested sampling runs

- In his original paper, John Skilling noted that nested sampling runs can be merged.

Aside: Merging nested sampling runs

- In his original paper, John Skilling noted that nested sampling runs can be merged.
- Take two complete nested sampling runs generated by $n_{\mathrm{live}}^{(1)}$ and $n_{\mathrm{live}}^{(2)}$ live points.

Aside: Merging nested sampling runs

- In his original paper, John Skilling noted that nested sampling runs can be merged.
- Take two complete nested sampling runs generated by $n_{\mathrm{live}}^{(1)}$ and $n_{\mathrm{live}}^{(2)}$ live points.
- Combining the two runs in likelihood order gives a new run generated by $n_{\mathrm{live}}^{(1)} + n_{\mathrm{live}}^{(2)}$ live points.

Aside: Unweaving nested sampling runs

Aside: Unweaving nested sampling runs

- The reverse is also true.

Aside: Unweaving nested sampling runs

- The reverse is also true.
- Given a nested sampling run with n_{live} points, there is a unique way of separating it into n_{live} single-point runs (threads).

PolyChord 2.0

Handling correlations

PolyChord 2.0

Handling correlations

- Unweave the run into n_{live} threads.

PolyChord 2.0
Handling correlations

- Unweave the run into n_{live} threads.
- Each thread is a "true" nested sampling run, although threads are correlated.

PolyChord 2.0
Handling correlations

- Unweave the run into n_{live} threads.
- Each thread is a "true" nested sampling run, although threads are correlated.
- Can use traditional techniques on threads to quantify correlation

PolyChord 2.0

Handling correlations

- Unweave the run into $n_{\rm live}$ threads.
- Each thread is a "true" nested sampling run, although threads are correlated.
- Can use traditional techniques on threads to quantify correlation
  - Batch means
  - Jacknifing
  - Bootstrapping

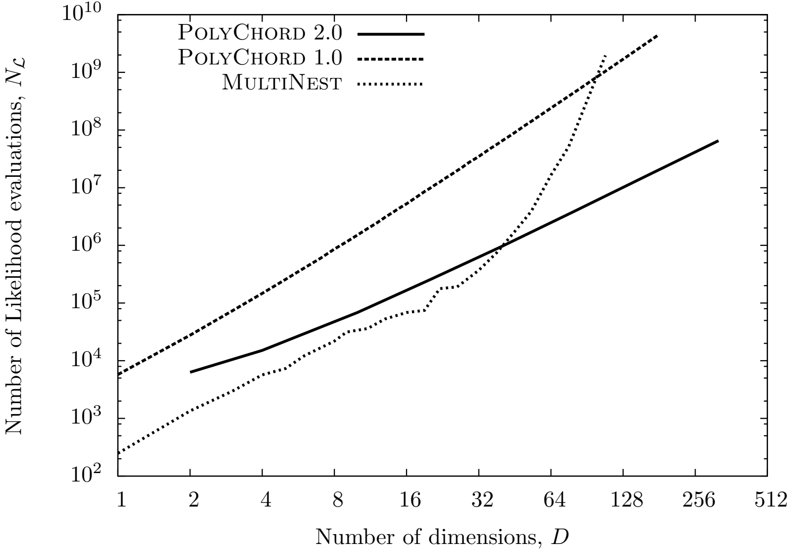
PolyChord 2.0

Handling correlations

- Unweave the run into n_{live} threads.
- Each thread is a “true” nested sampling run, although threads are correlated.
- Can use traditional techniques on threads to quantify correlation
  - Batch means
  - Jacknifing
  - Bootstrapping
- With this in hand, can produce correct inferences from correlated runs.

PolyChord 2.0 vs. MultiNest

Gaussian likelihood

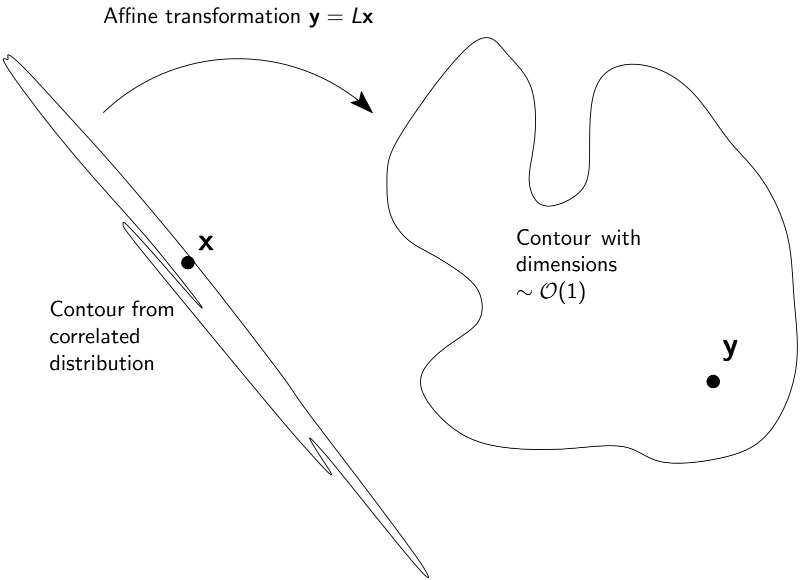


Correlated distributions

Correlated distributions

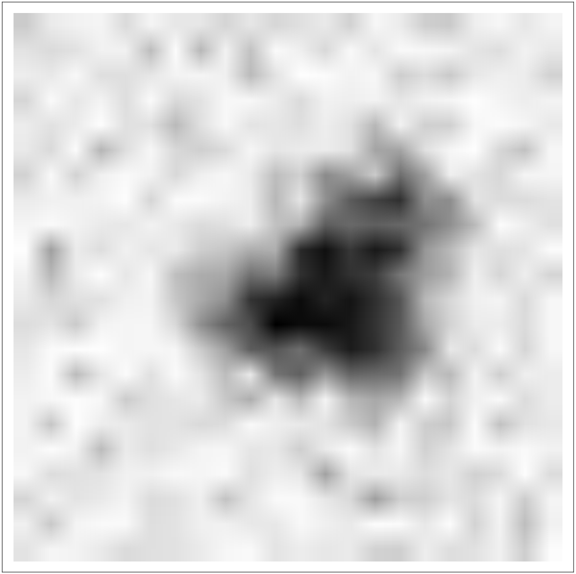
- Correlated distributions are hard

Correlated distributions



Object detection

Toy problem



Object detection

Evidences

Object detection

Evidences

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

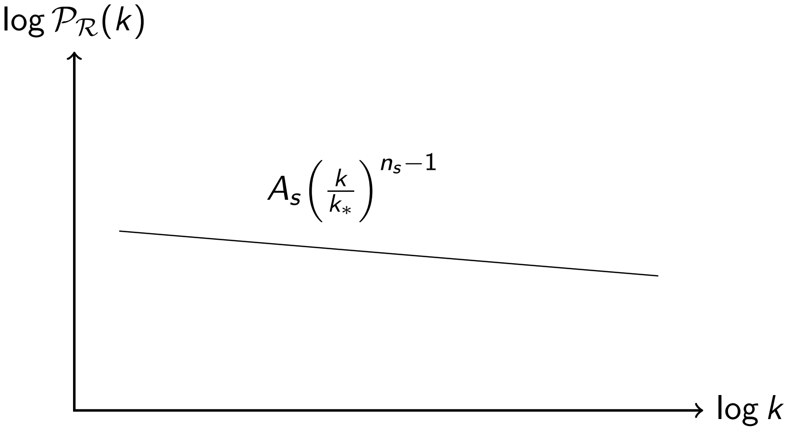
Object detection

Evidences

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

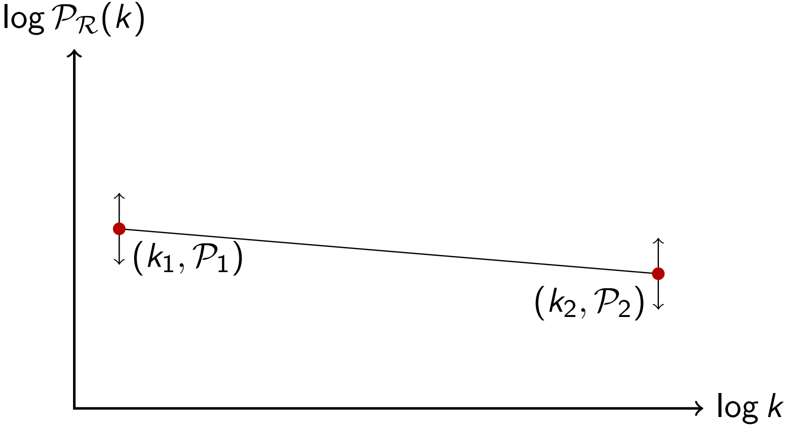
PolyChord in action

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



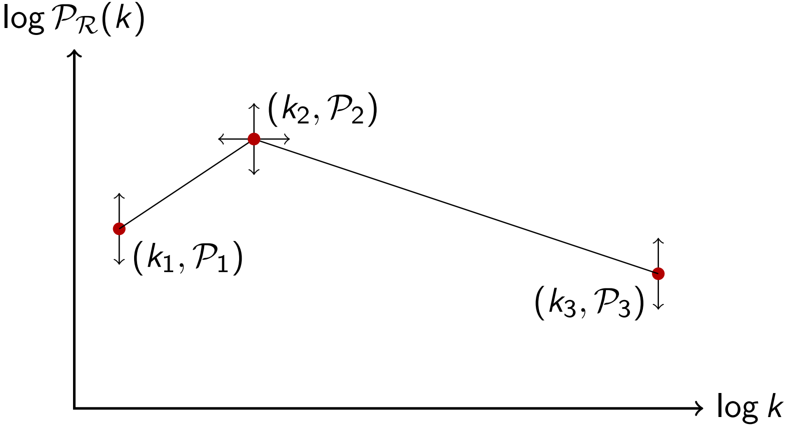
PolyChord in action

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



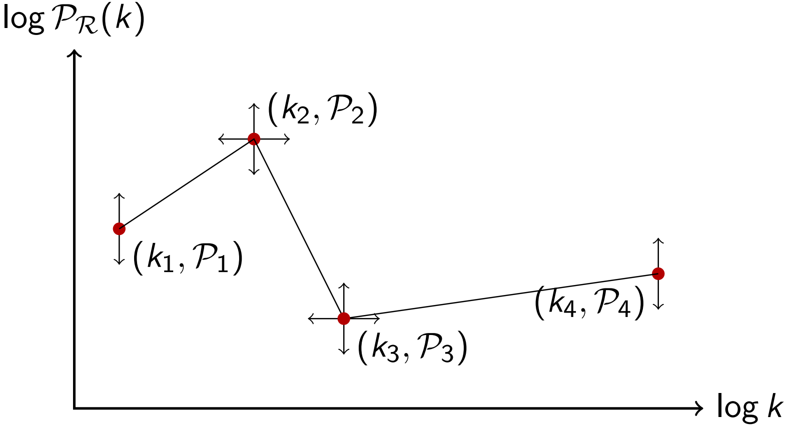
PolyChord in action

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



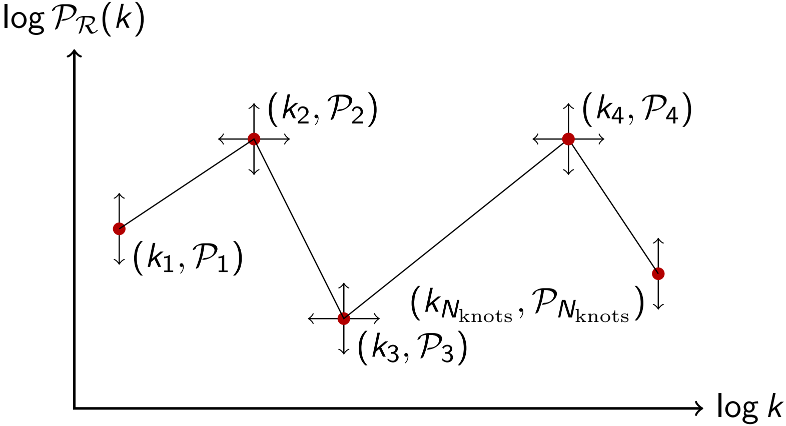
PolyChord in action

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



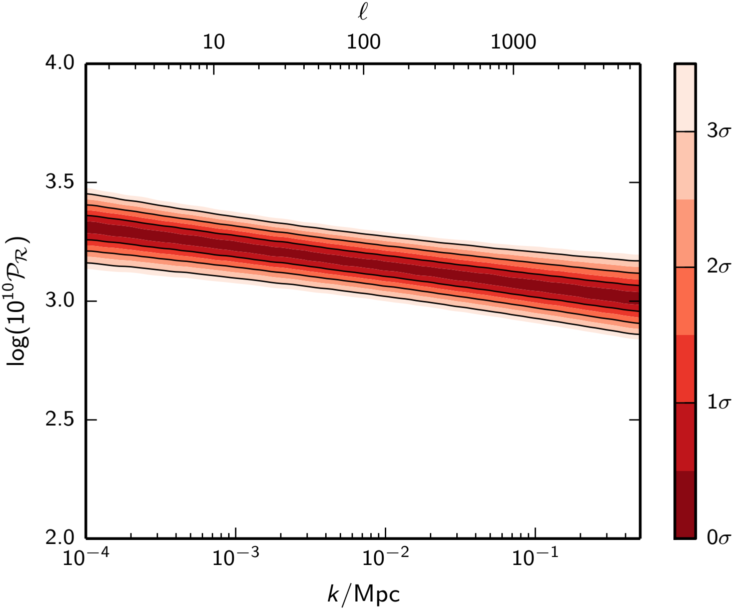
PolyChord in action

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



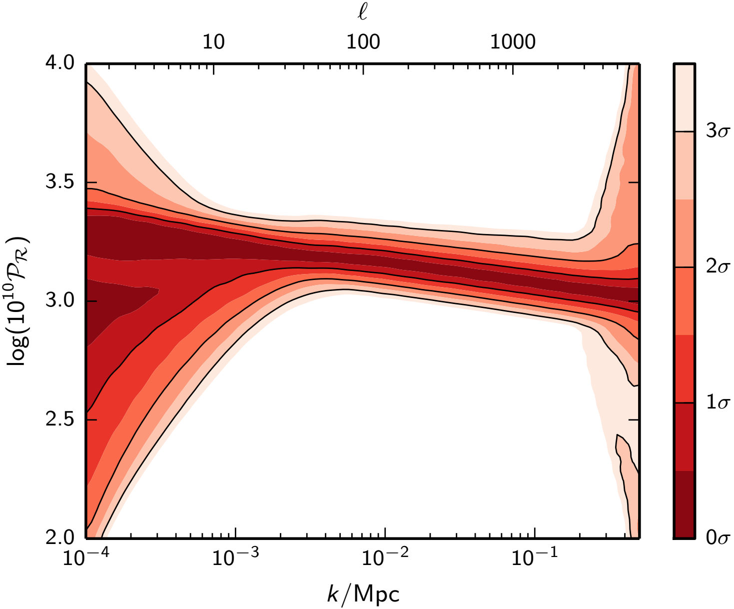
0 internal knots

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



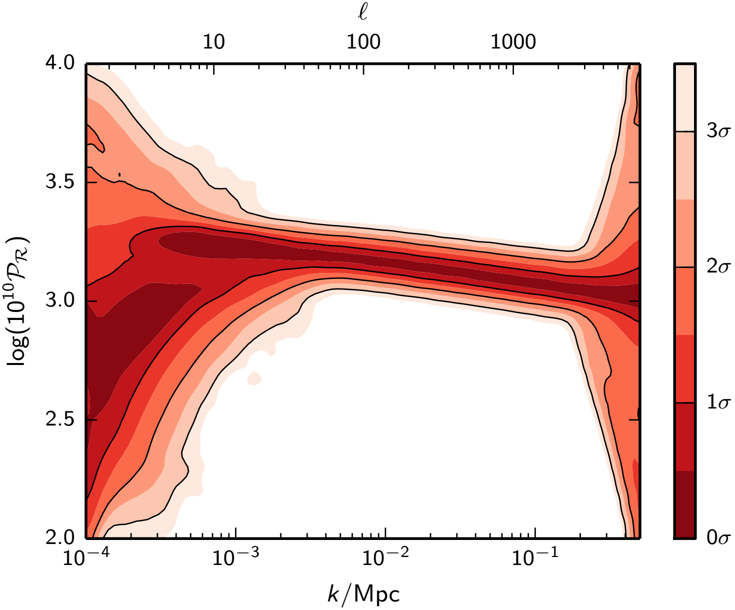
1 internal knots

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



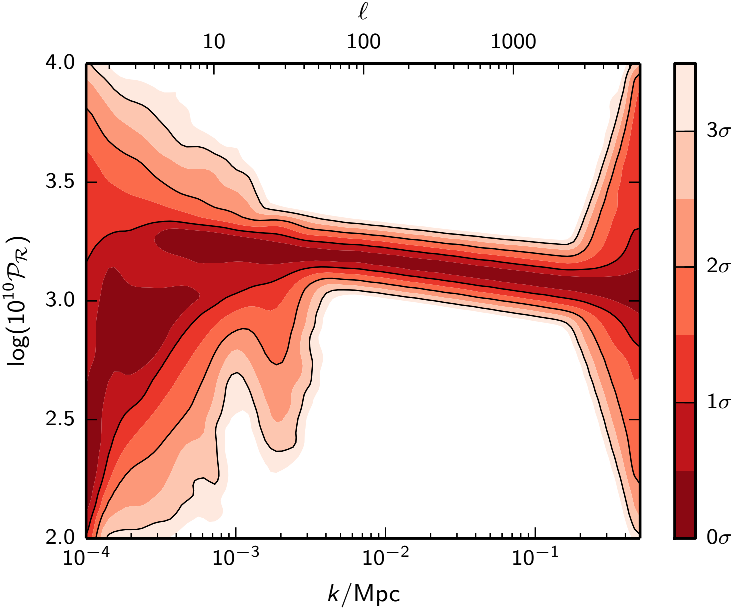
2 internal knots

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



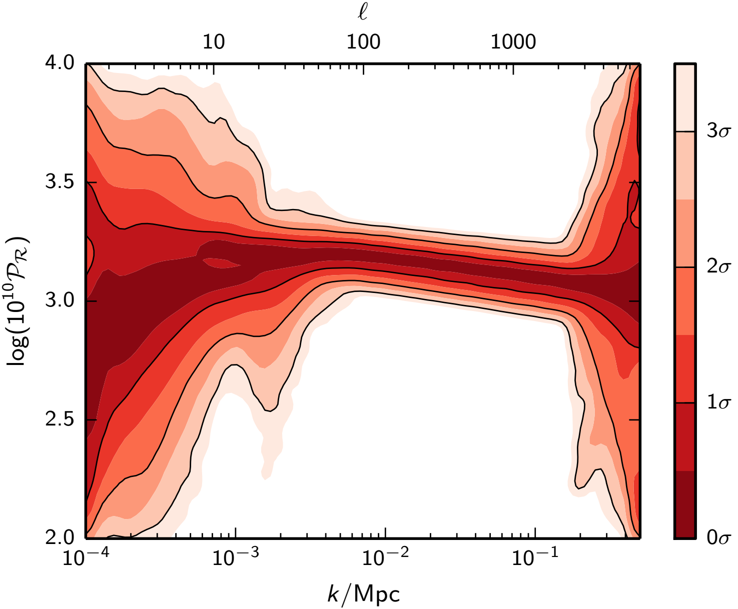
3 internal knots

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



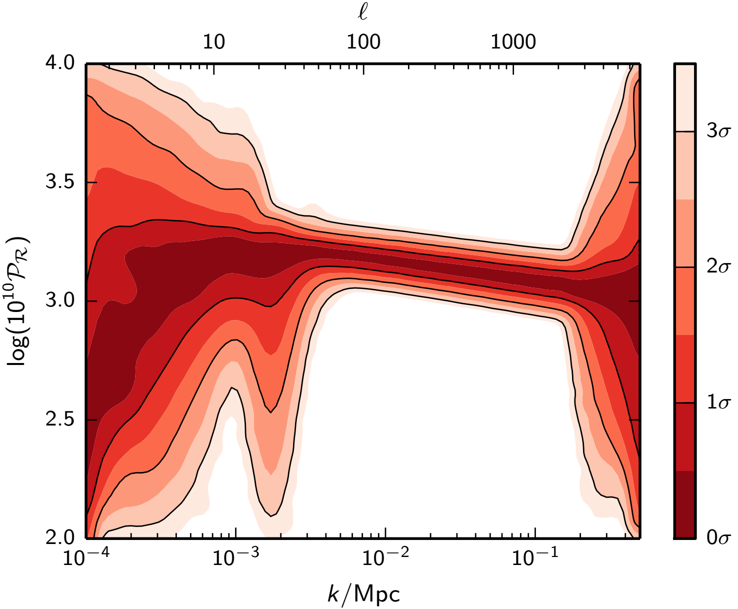
4 internal knots

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



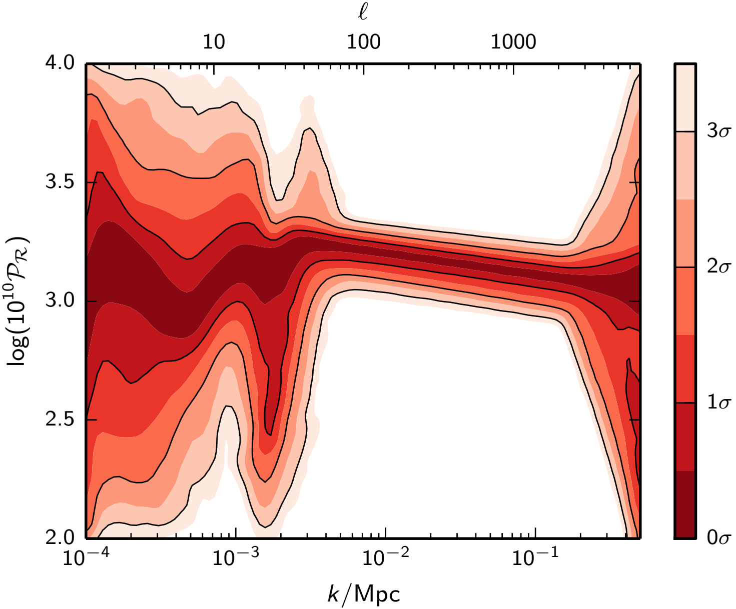
5 internal knots

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



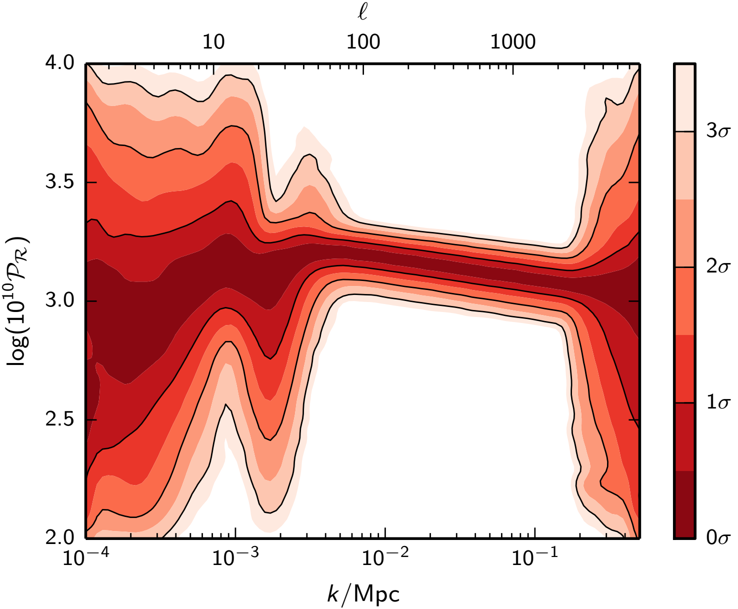
6 internal knots

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



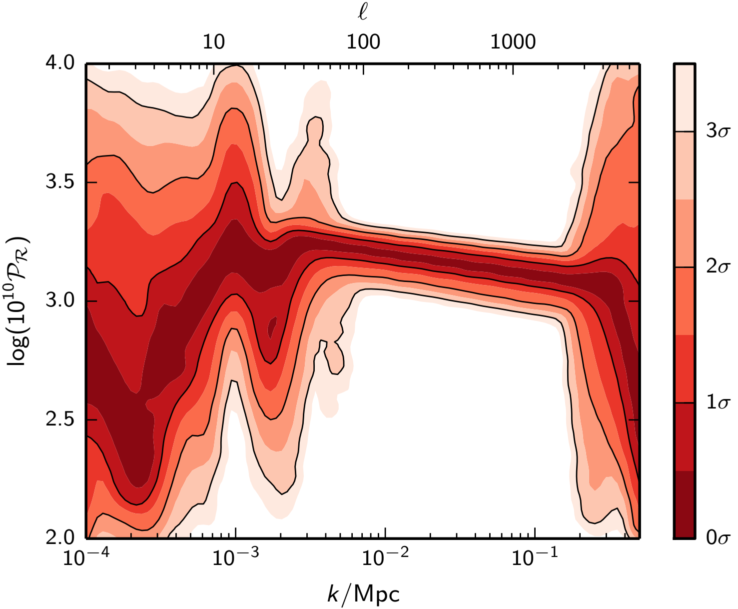
7 internal knots

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



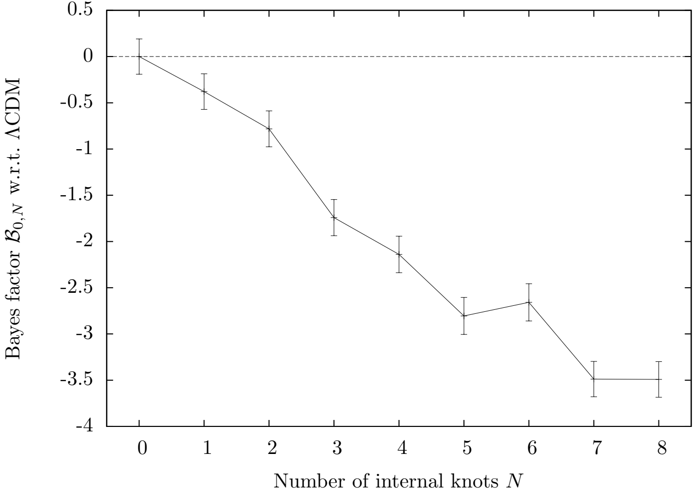
8 internal knots

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



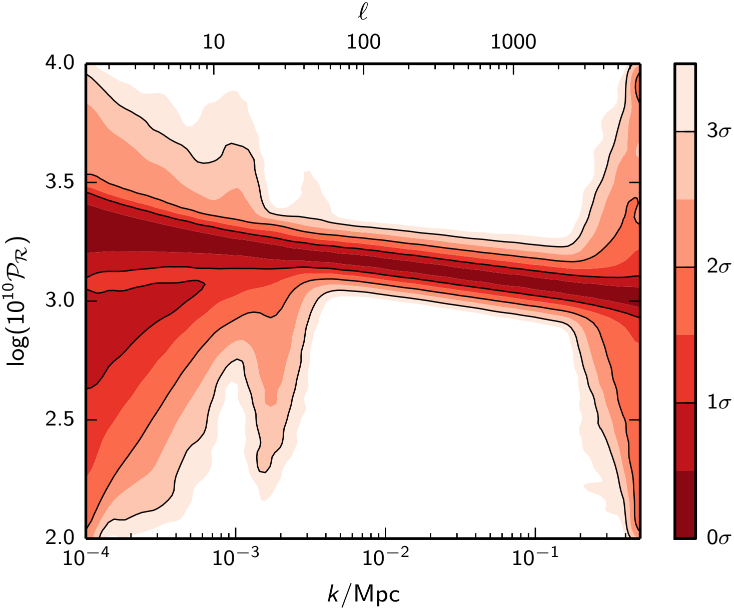
Bayes Factors

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



Marginalised plot

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



Affine invariance

Affine invariance

- The optimal exploration technique is be affine invariant.

Affine invariance

- The optimal exploration technique is be affine invariant.
- Treat distribution $\mathrm{P}(\mathrm{x})$ and $\mathrm{P}(R\mathrm{x})$ the same.

Affine invariance

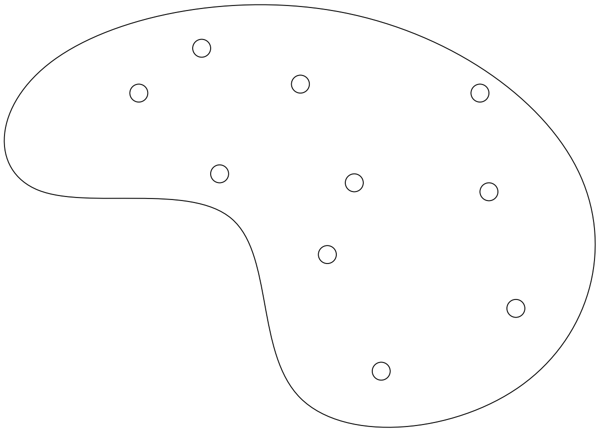
- The optimal exploration technique is be affine invariant.
- Treat distribution $\mathrm{P}(\mathrm{x})$ and $\mathrm{P}(R\mathrm{x})$ the same.
- No need to worry about correlations.

Affine invariance

- The optimal exploration technique is be affine invariant.
- Treat distribution $\mathrm{P}(\mathrm{x})$ and $\mathrm{P}(R\mathrm{x})$ the same.
- No need to worry about correlations.
- Good example: Now highly successful emcee (MCMC hammer).
  - Important: emcee is not unique (or necessarily best)

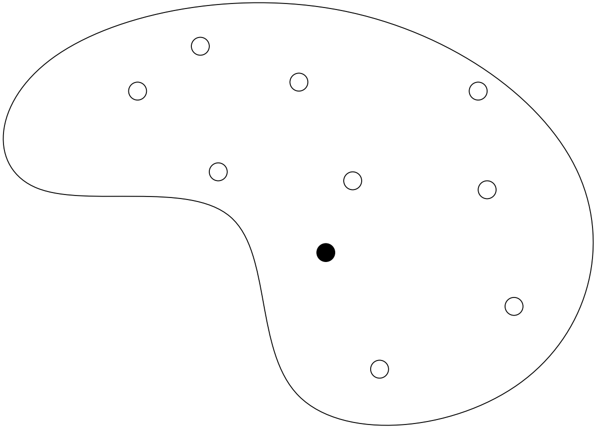
Skillings affine invariant ideas

Leapfrog



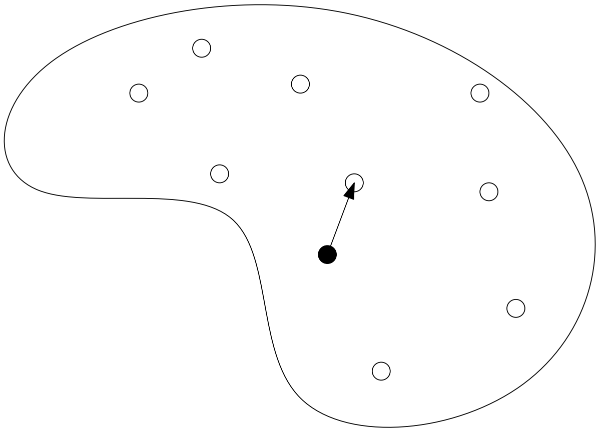
Skillings affine invariant ideas

Leapfrog



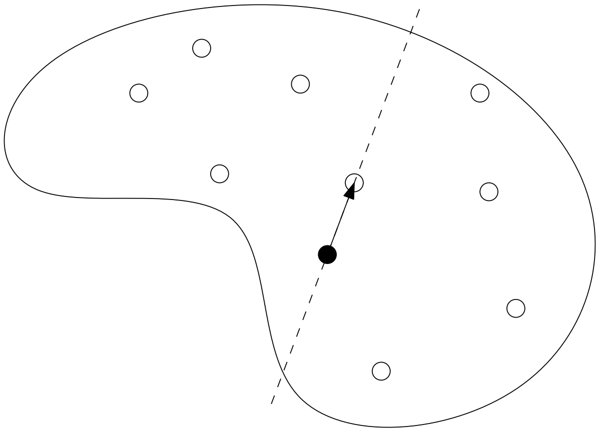
Skillings affine invariant ideas

Leapfrog



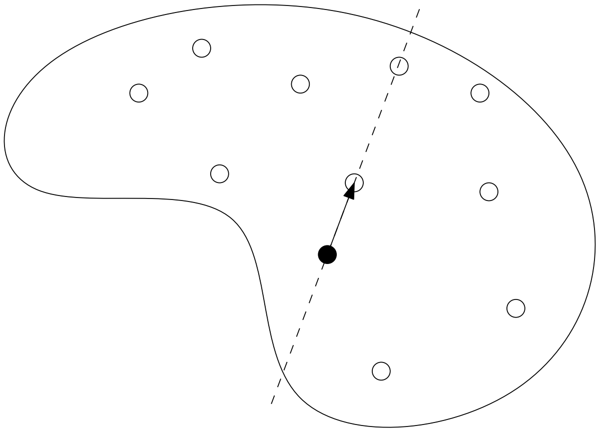
Skillings affine invariant ideas

Leapfrog



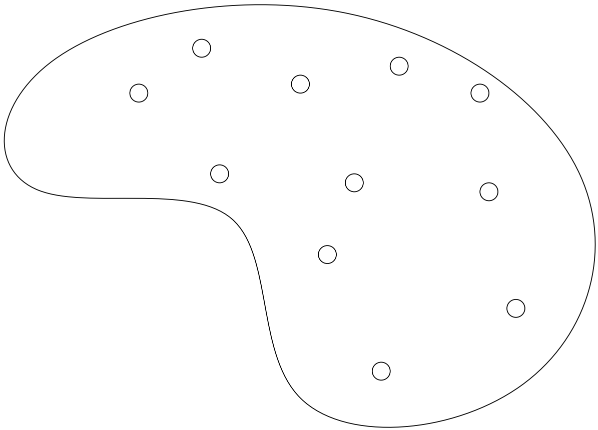
Skillings affine invariant ideas

Leapfrog



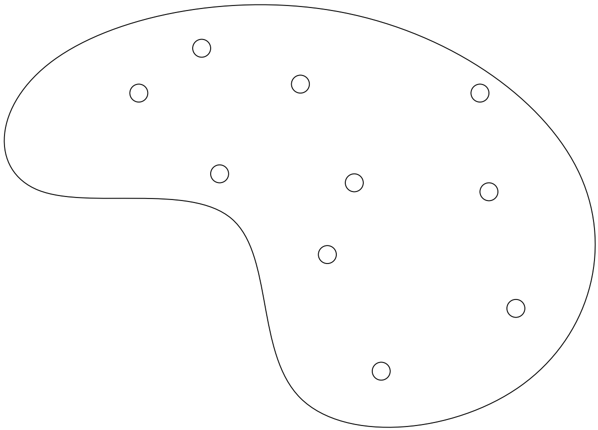
Skillings affine invariant ideas

Leapfrog



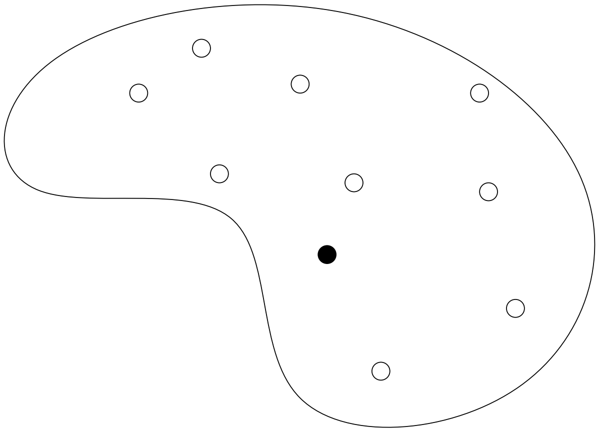
Skillings affine invariant ideas

Parallel walk



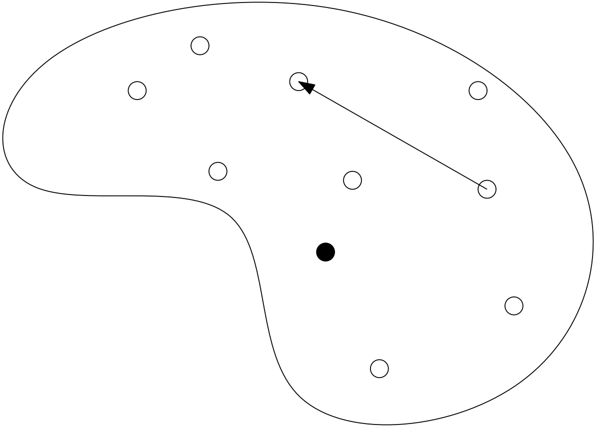
Skillings affine invariant ideas

Parallel walk



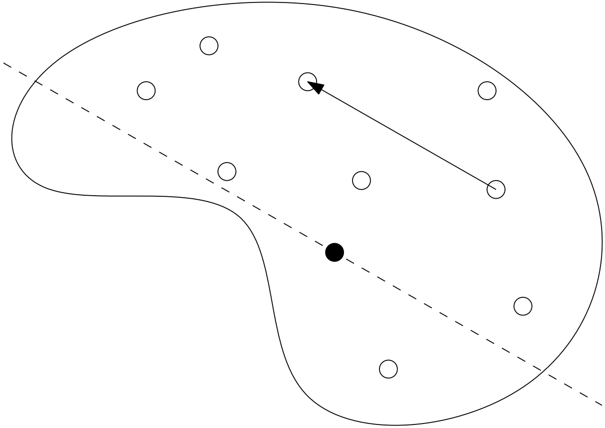
Skillings affine invariant ideas

Parallel walk



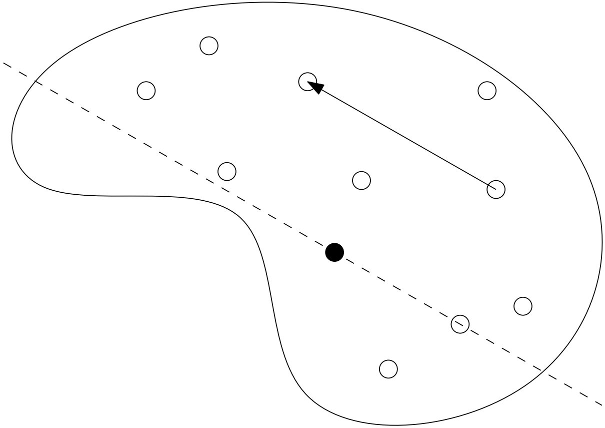
Skillings affine invariant ideas

Parallel walk



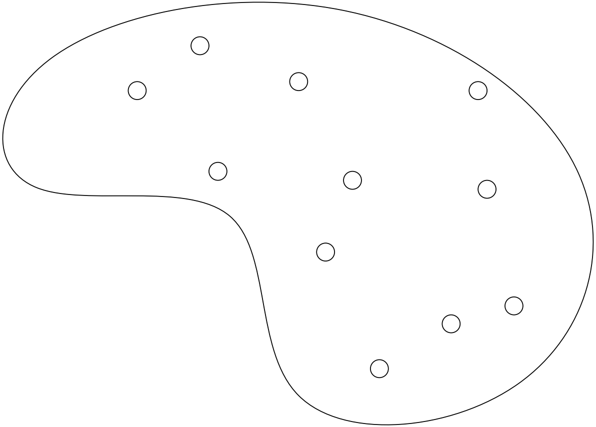
Skillings affine invariant ideas

Parallel walk



Skillings affine invariant ideas

Parallel walk



Affine invariance

Subspace collapse

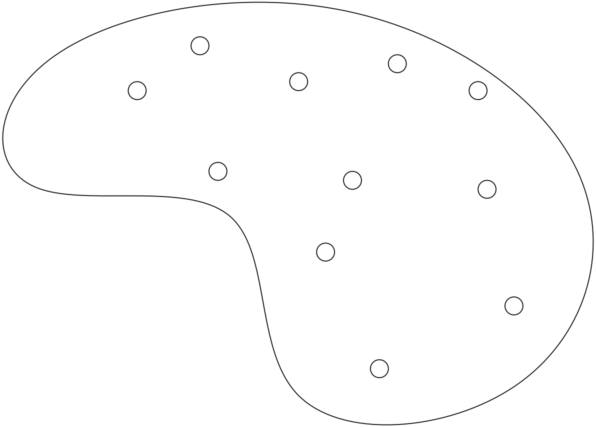
Affine invariance

Subspace collapse

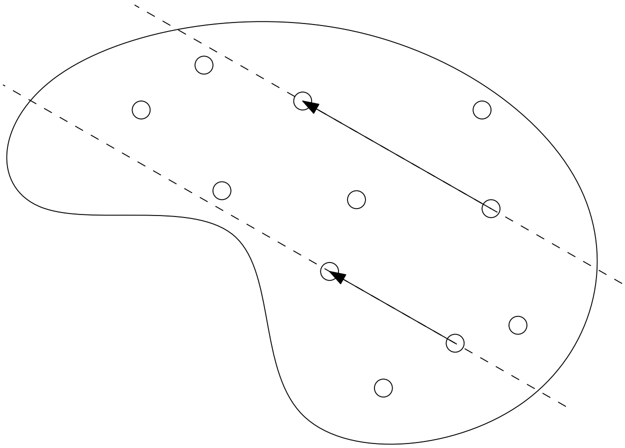
- The main problem that besets these techniques is "subspace collapse".

Subspace collapse

Leapfrog



Subspace collapse



Subspace collapse

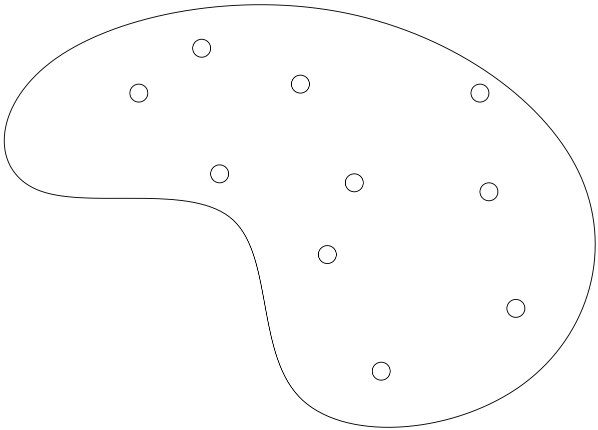
Solution

Subspace collapse
Solution

- Need to use $\sim \mathcal{O}(D)$ points to avoid this.

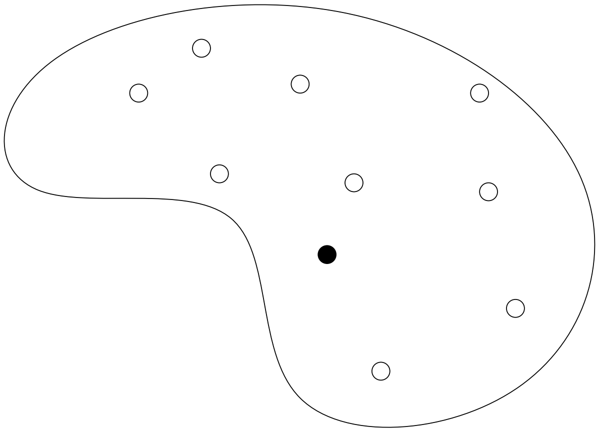
Skillings affine invariant ideas

Guided walk



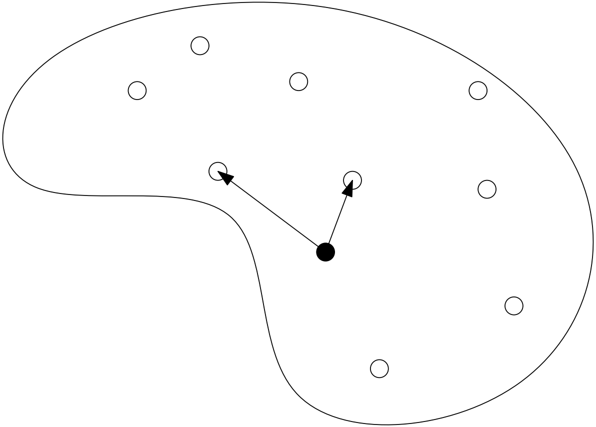
Skillings affine invariant ideas

Guided walk



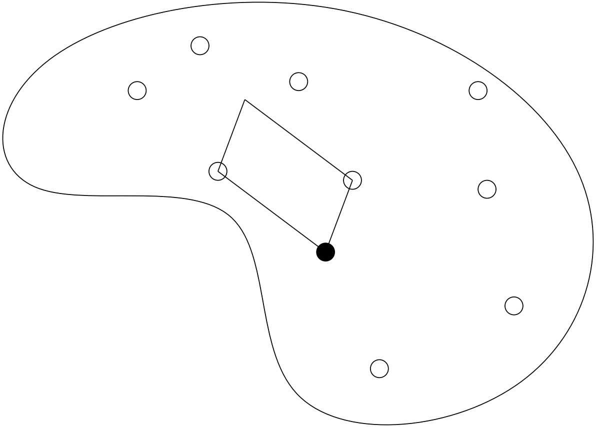
Skillings affine invariant ideas

Guided walk



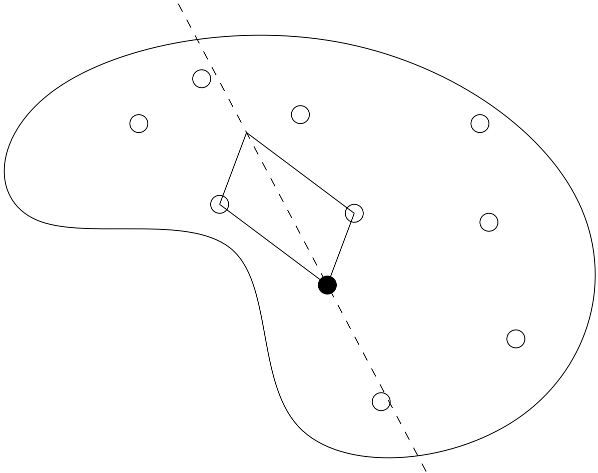
Skillings affine invariant ideas

Guided walk



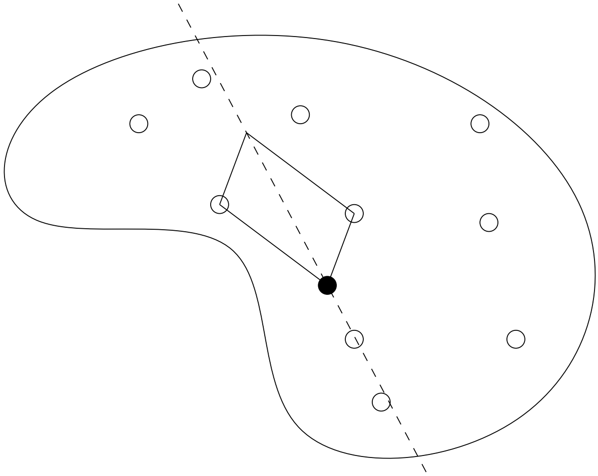
Skillings affine invariant ideas

Guided walk



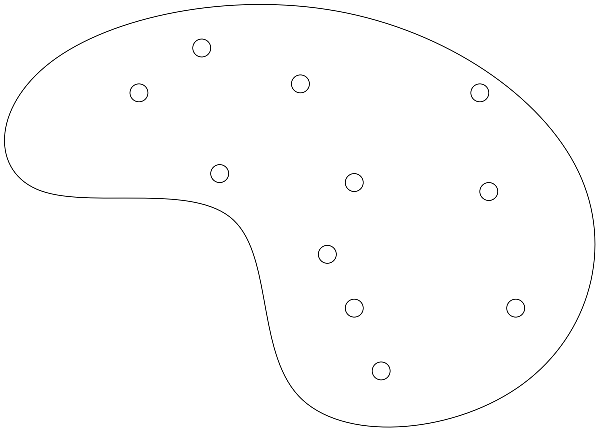
Skillings affine invariant ideas

Guided walk



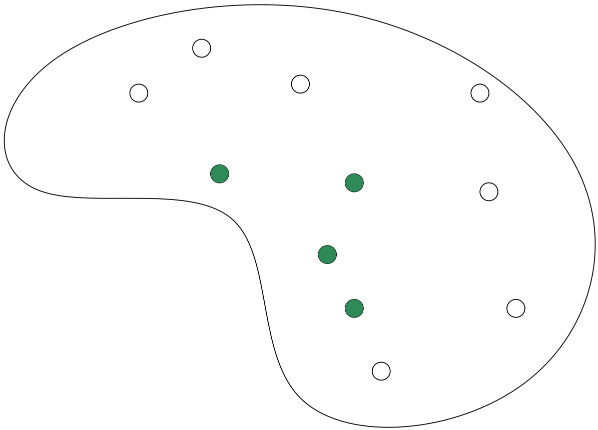
Skillings affine invariant ideas

Guided walk



Skillings affine invariant ideas

Guided walk



Affine invariant

Other variations

Affine invariant

Other variations

- Generalise guided walk to D dimensions (slice through the mean of D other points).

Affine invariant

Other variations

- Generalise guided walk to D dimensions (slice through the mean of D other points).
- Slice through a "random" linear combination of D points.

Affine invariant

Other variations

- Generalise guided walk to D dimensions (slice through the mean of D other points).
- Slice through a "random" linear combination of D points.
- Slice through a "random" linear combination of all points

Affine invariant

Other variations

- Generalise guided walk to D dimensions (slice through the mean of D other points).
- Slice through a "random" linear combination of D points.
- Slice through a "random" linear combination of all points
- There are lots of variations: This is an underused area of the field.

PolyChord 2.0

PolyChord 2.0

- Using intermediate points so $\sim \mathcal{O}(D^3) \to \sim \mathcal{O}(D^2)$.

PolyChord 2.0

- Using intermediate points so $\sim \mathcal{O}(D^3) \rightarrow \sim \mathcal{O}(D^2)$.
- Unweaving runs to quantify correlations

PolyChord 2.0

- Using intermediate points so $\sim \mathcal{O}(D^3) \rightarrow \sim \mathcal{O}(D^2)$.
- Unweaving runs to quantify correlations
- Affine invariant sampling