PolyChord 2.0

Advances in nested sampling with astrophysical applications

Will Handley
wh260@cam.ac.uk

Astrophysics Group
Cavendish Laboratory
University of Cambridge

January 11, 2021

What is nested sampling?

What is nested sampling?

- Nested sampling is an alternative way of sampling posteriors.

What is nested sampling?

- Nested sampling is an alternative way of sampling posteriors.
- Uses ensemble sampling to compress prior to posterior.

What is nested sampling?

- Nested sampling is an alternative way of sampling posteriors.
- Uses ensemble sampling to compress prior to posterior.
- In doing so, it circumvents many issues (dimensionality, topology, geometry) that beset standard approaches.

Outline

Outline

1. Background theory

Outline

1. Background theory
2. Review existing sampling approaches

Outline

1. Background theory
2. Review existing sampling approaches
3. Nested Sampling & Historical implementations.

Outline

1. Background theory
2. Review existing sampling approaches
3. Nested Sampling & Historical implementations.
4. PolyChord

Outline

1. Background theory
2. Review existing sampling approaches
3. Nested Sampling & Historical implementations.
4. PolyChord
5. Applications

Bayes' theorem

Parameter estimation

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

Bayes' theorem

Parameter estimation

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

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

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

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

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$D = \{C_\ell\}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned} D =& \{C_\ell\} \\ M =& \Lambda\mathrm{CDM} \end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned} D =& \{C_\ell\} \\ M =& \Lambda\mathrm{CDM} \\ \Theta =& \Theta_{\Lambda\mathrm{CDM}} \end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned} D =& \{C_\ell\} \\ M =& \Lambda\mathrm{CDM} \\ \Theta =& \Theta_{\Lambda\mathrm{CDM}} \\ \Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned} D =& \{ C_\ell^{(\mathrm{Planck})} \} \\ M =& \Lambda\mathrm{CDM} \\ \Theta =& \Theta_{\Lambda\mathrm{CDM}} \\ \Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned} D =& \{C_\ell^{(\mathrm{Planck})}\} \\ M =& \Lambda\mathrm{CDM} \\ \Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} \\ \Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{C_\ell^{(\mathrm{Planck})}\} \\
M =& \Lambda\mathrm{CDM} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{ C_\ell^{(\text{Planck})} \} \\
M =& \Lambda\text{CDM} + \text{extensions} \\
\Theta =& \Theta_{\Lambda\text{CDM}} + \Theta_{\text{Planck}} \\
\Theta_{\Lambda\text{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\text{Planck}} =& (y_{\text{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times 217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\text{dust}\,TT}, A_{143}^{\text{dust}\,TT}, A_{143\times 217}^{\text{dust}\,TT}, A_{217}^{\text{dust}\,TT}, c_{100}, c_{217})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{C_\ell^{(\mathrm{Planck})}\} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10}A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times 217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times 217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{C_\ell^{(\mathrm{Planck})}\} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10}A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times 217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times 217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{ C_\ell^{(\mathrm{Planck})} \} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{ C_\ell^{(\mathrm{Planck})} \} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times 217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times 217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}}, w)
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{ C_\ell^{(\mathrm{Planck})} \} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}}, w, \Sigma m_\nu, m_{\nu,\mathrm{sterile}}^{\mathrm{eff}})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{C_\ell^{(\mathrm{Planck})}\} + \{\mathrm{LSS}\} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times 217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times 217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}}, w, \Sigma m_\nu, m_{\nu,\mathrm{sterile}}^{\mathrm{eff}})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{C_\ell^{(\mathrm{Planck})}\} + \{\mathrm{LSS}\} + \{\text{“Big Data”}\} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times 217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times 217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}}, w, \Sigma m_\nu, m_{\nu,\mathrm{sterile}}^{\mathrm{eff}})
\end{aligned}$$

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{ C_\ell^{(\mathrm{Planck})} \} + \{\mathrm{LSS}\} + \{\text{“Big Data”}\} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}}, w, \Sigma m_\nu, m_{\nu,\mathrm{sterile}}^{\mathrm{eff}})
\end{aligned}$$

- Likelihoods can be quite complicated!

Parameter estimation

A concrete example.

$$\mathcal{L}(\Theta) = P(D|\Theta, M)$$

$$\begin{aligned}
D =& \{ C_\ell^{(\mathrm{Planck})} \} + \{\mathrm{LSS}\} + \{\text{“Big Data”}\} \\
M =& \Lambda\mathrm{CDM} + \mathrm{extensions} \\
\Theta =& \Theta_{\Lambda\mathrm{CDM}} + \Theta_{\mathrm{Planck}} + \Theta_{\mathrm{extensions}} \\
\Theta_{\Lambda\mathrm{CDM}} =& (\Omega_b h^2, \Omega_c h^2, 100\theta_{MC}, \tau, \ln(10^{10} A_s), n_s) \\
\Theta_{\mathrm{Planck}} =& (y_{\mathrm{cal}}, A_{217}^{CIB}, \xi^{tSZ-CIB}, A_{143}^{tSZ}, A_{100}^{PS}, A_{143}^{PS}, A_{143\times217}^{PS}, A_{217}^{PS}, \\
& A^{kSZ}, A_{100}^{\mathrm{dust}\,TT}, A_{143}^{\mathrm{dust}\,TT}, A_{143\times217}^{\mathrm{dust}\,TT}, A_{217}^{\mathrm{dust}\,TT}, c_{100}, c_{217}) \\
\Theta_{\mathrm{extensions}} =& (n_{\mathrm{run}}, n_{\mathrm{run,run}}, w, \Sigma m_\nu, m_{\nu,\mathrm{sterile}}^{\mathrm{eff}})
\end{aligned}$$

- Likelihoods can be quite complicated!
- We need advanced sampling approaches.

Bayes' theorem

Parameter estimation

What does data D 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

Parameter estimation

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

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

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

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

Prior: Usually uniform, but can be source of controversy.

Bayes' theorem

Parameter estimation

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

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

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

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

Prior: Usually uniform, but can be source of controversy.

Evidence: Just a normalising constant

Bayes' theorem

Parameter estimation

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

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

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

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

Prior: Usually uniform, but can be source of controversy.

Evidence: Just a normalising constant?

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?

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

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

e.g. Should we include running?

Bayes' theorem

Model comparison

What does data tell us about our model M_i?

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

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

e.g. Should we include running? Neutrinos?

Bayes' theorem
Model comparison

What does data tell us about our model M_i?

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

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

e.g. Should we include running? Neutrinos? Dark energy?

Bayes' theorem

Model comparison

What does data tell us about our model M_i?

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

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

e.g. Should we include running? Neutrinos? Dark energy?

Model averaging:

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

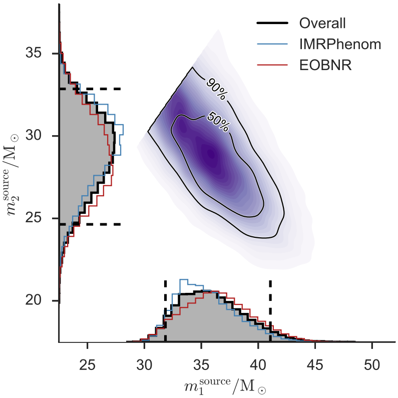
$$\mathrm{P}(y|D) = \sum_i \mathrm{P}(y|M_i, D)\mathrm{P}(M_i|D)$$

Parameter estimation

Why do sampling?

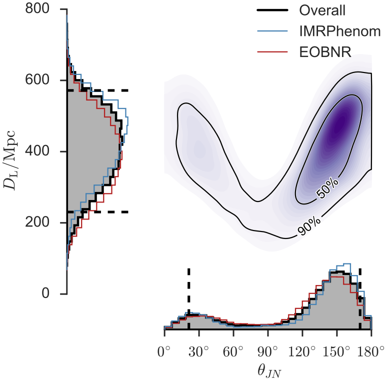
Parameter estimation

Why do sampling?



Parameter estimation

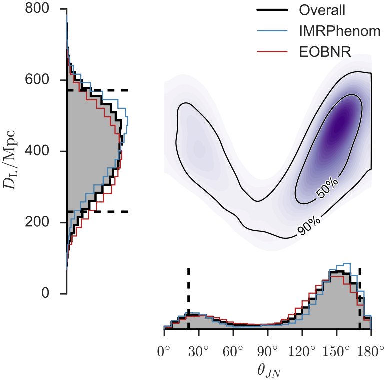
Why do sampling?



Parameter estimation

Why do sampling?

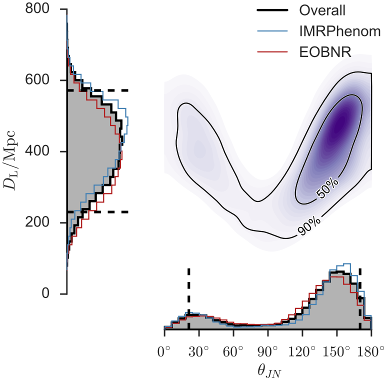
- In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.



Parameter estimation

Why do sampling?

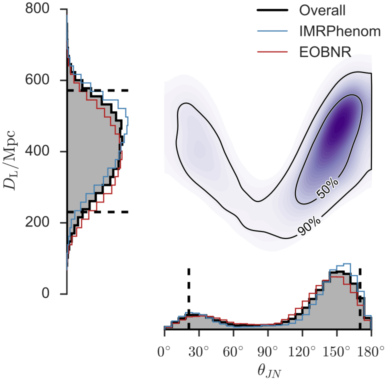
- In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.
- Describing an N-dimensional posterior fully is impossible.



Parameter estimation

Why do sampling?

- In high dimensions, posterior $\mathcal{P}$ occupies a vanishingly small region of the prior π.
- Describing an *N*-dimensional posterior fully is impossible.
- *Sampling* the posterior is an excellent compression scheme.



Current sampling approaches

Current sampling approaches

1. Metropolis Hastings.

Current sampling approaches

1. Metropolis Hastings.
2. Hamiltonian Monte-Carlo (HMC).

Current sampling approaches

1. Metropolis Hastings.
2. Hamiltonian Monte-Carlo (HMC).
3. Ensemble sampling (e.g. emcee).

Metropolis Hastings

Metropolis Hastings

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

Metropolis Hastings

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

Metropolis Hastings

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

Metropolis Hastings

- 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

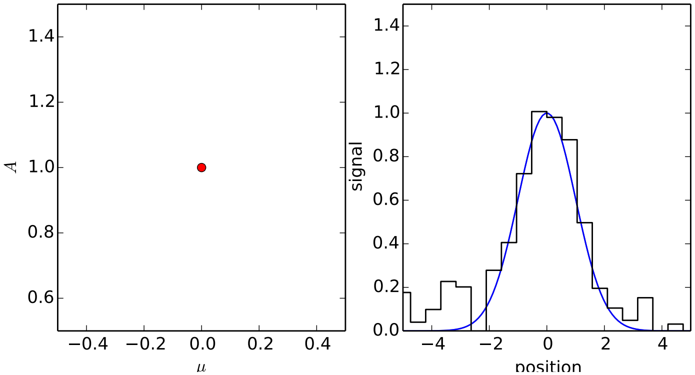
Metropolis Hastings

- 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. . .

Metropolis Hastings

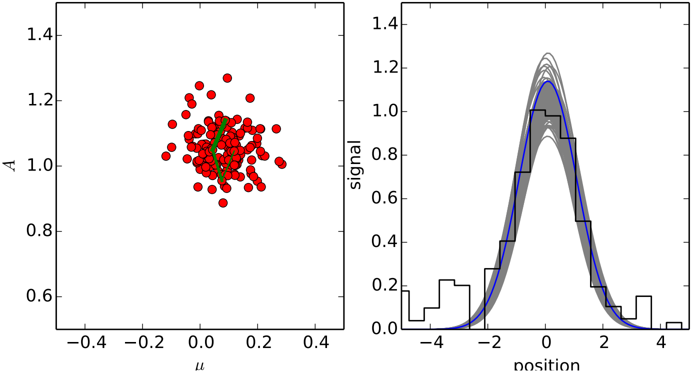
- Turn the N-dimensional problem into a one-dimensional one.
- Explore the space via a biased random walk.
  1. Pick random direction
  2. Choose step length
  3. If uphill, make step...
  4. ...otherwise sometimes make step.

Metropolis Hastings



Metropolis Hastings

Metropolis Hastings



Metropolis Hastings

Struggles with. . .

Metropolis Hastings

Struggles with...

1. Burn in

Metropolis Hastings

Struggles with. . .

1. Burn in
2. Multimodality

Metropolis Hastings

Struggles with...

1. Burn in
2. Multimodality
3. Correlated Peaks

Metropolis Hastings

Struggles with...

1. Burn in
2. Multimodality
3. Correlated Peaks
4. Phase transitions

Hamiltonian Monte-Carlo

Hamiltonian Monte-Carlo

- Key idea: Treat $\log L(\Theta)$ as a potential energy

Hamiltonian Monte-Carlo

- Key idea: Treat $\log L(\Theta)$ as a potential energy
- Guide walker under "force":

$$F(\Theta) = \nabla \log L(\Theta)$$

Hamiltonian Monte-Carlo

- Key idea: Treat $\log L(\Theta)$ as a potential energy
- Guide walker under "force":

$$F(\Theta) = \nabla \log L(\Theta)$$

- Walker is naturally "guided" uphill

Hamiltonian Monte-Carlo

- Key idea: Treat $\log L(\Theta)$ as a potential energy
- Guide walker under "force":

$$F(\Theta) = \nabla \log L(\Theta)$$

- Walker is naturally "guided" uphill
- Conserved quantities mean efficient acceptance ratios.

Hamiltonian Monte-Carlo

Problems

Hamiltonian Monte-Carlo

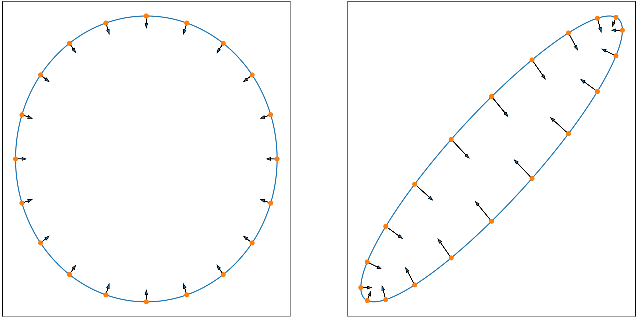
Problems

- "Uphill" is not covariant.

Hamiltonian Monte-Carlo

Problems

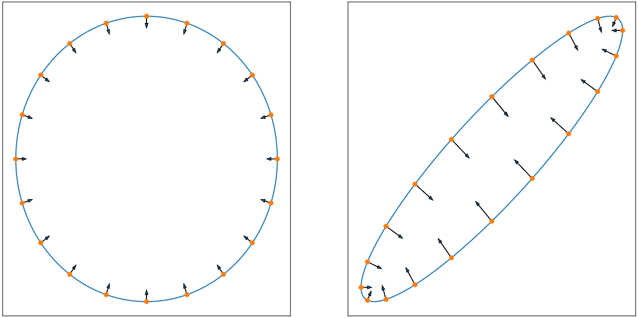
- “Uphill” is not covariant.



Hamiltonian Monte-Carlo

Problems

- "Uphill" is not covariant.



- Requires gradients (autograd – python)

Ensemble sampling

Ensemble sampling

- Instead of one walker, evolve a set of n walkers.

Ensemble sampling

- Instead of one walker, evolve a set of n walkers.
- Can use information present in ensemble to guide proposals.

Ensemble sampling

- Instead of one walker, evolve a set of n walkers.
- Can use information present in ensemble to guide proposals.
- emcee: affine invariant proposals.

Ensemble sampling

- Instead of one walker, evolve a set of n walkers.
- Can use information present in ensemble to guide proposals.
- emcee: affine invariant proposals.
- emcee is not the only (or even best) affine invariant approach.

The real reason these all fail

The real reason these all fail

- MCMC does not give you evidences!

The real reason these all fail

- MCMC does not give you evidences!

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

The real reason these all fail

- MCMC does not give you evidences!

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

The real reason these all fail

- MCMC does not give you evidences!

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

The real reason these all fail

- MCMC does not give you evidences!

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

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

The real reason these all fail

- MCMC does not give you evidences!

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

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

The real reason these all fail

- MCMC does not give you evidences!

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

- MCMC fundamentally explores the posterior, and cannot average over the prior.
- Simulated annealing gives one possibility for computing evidences.
  - Inspired by thermodynamics.
  - Suffers from similar issues to MCMC.
  - Unclear how to choose correct annealing schedule

Nested Sampling

John Skilling's alternative to traditional MCMC!

Nested Sampling

John Skilling's alternative to traditional MCMC!

New procedure:

Nested Sampling

John Skilling's alternative to traditional MCMC!

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

Nested Sampling

John Skilling's alternative to traditional 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 traditional 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 uniform sample with higher likelihood

Nested Sampling

John Skilling's alternative to traditional MCMC!

New procedure:

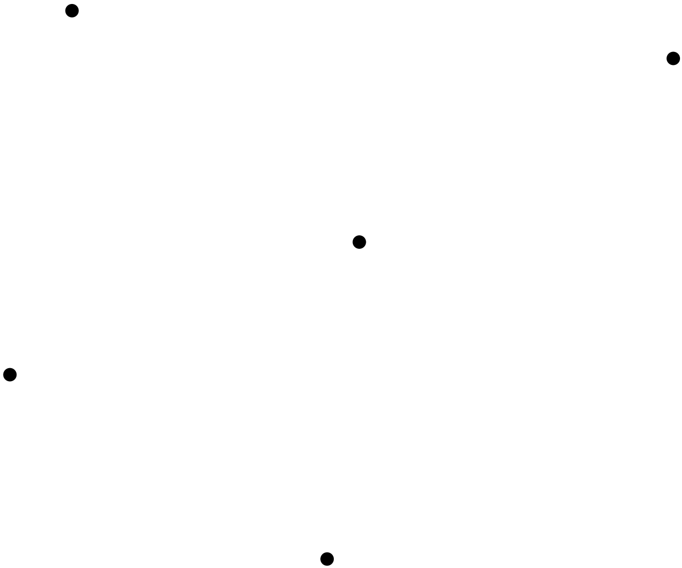
Maintain a set S of n samples, which are sequentially updated:

- 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 uniform sample with higher likelihood

Requires one to be able to uniformly within a region, subject to a *hard likelihood constraint*.

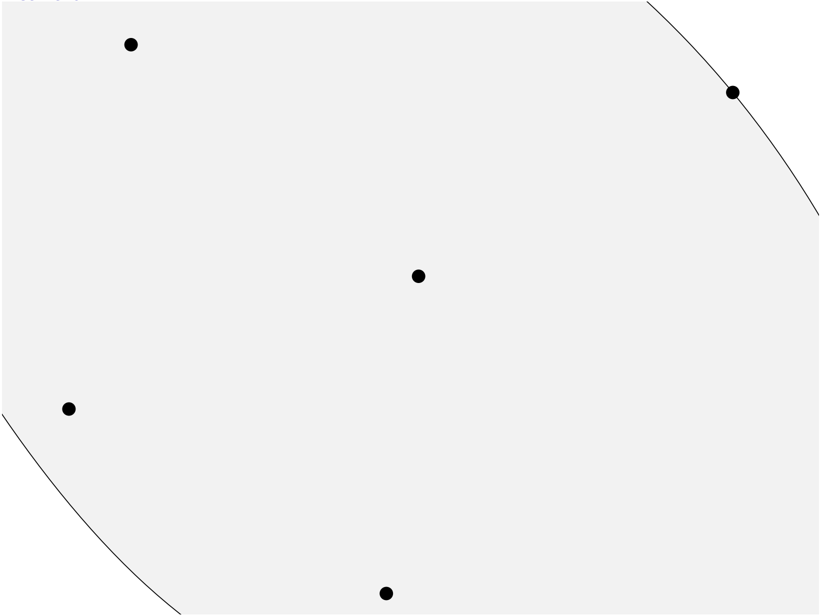
Nested Sampling

Graphical aid



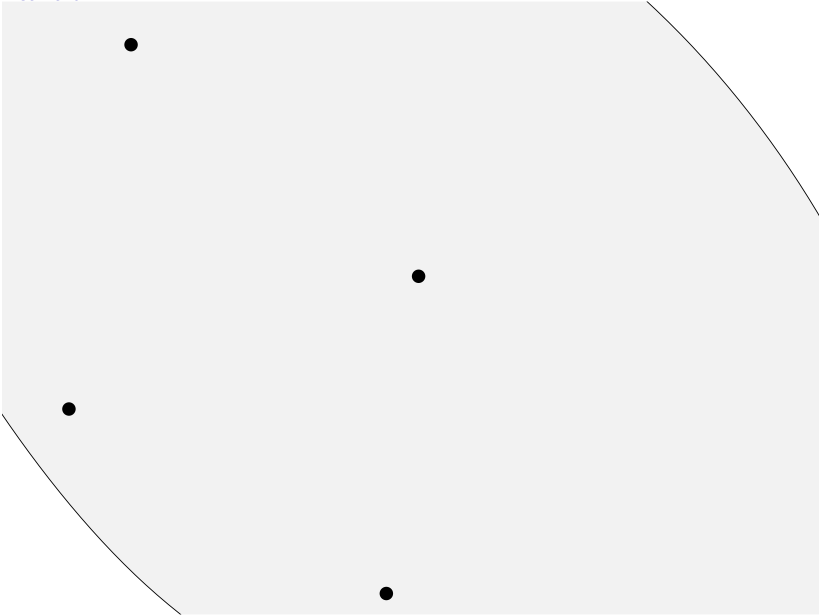
Nested Sampling

Graphical aid



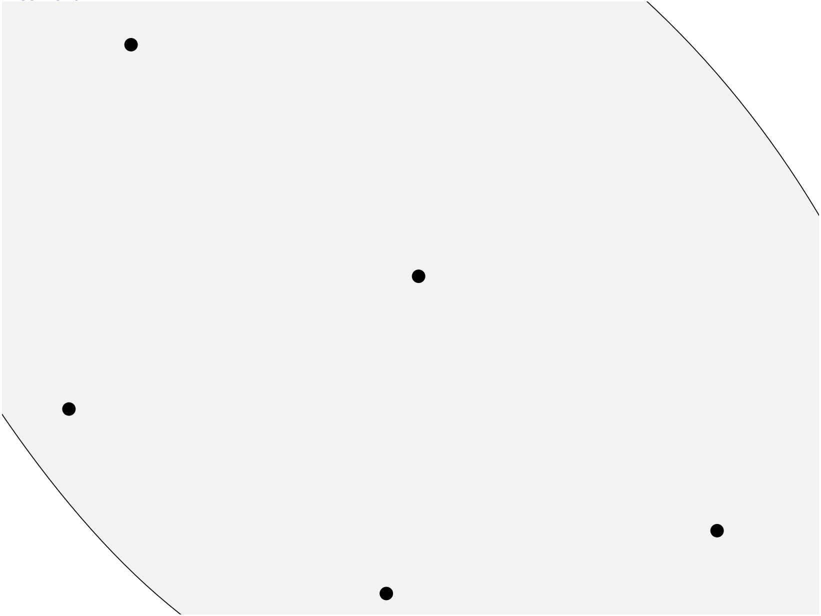
Nested Sampling

Graphical aid



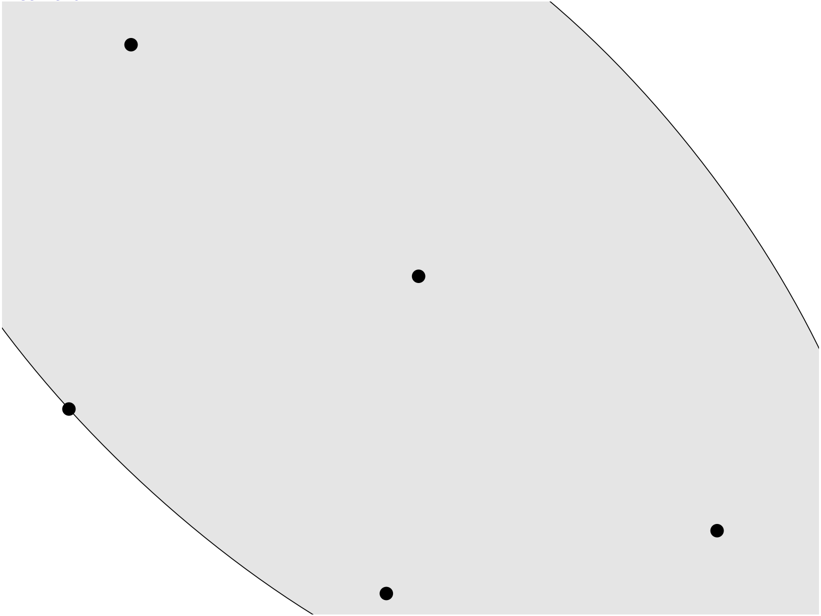
Nested Sampling

Graphical aid



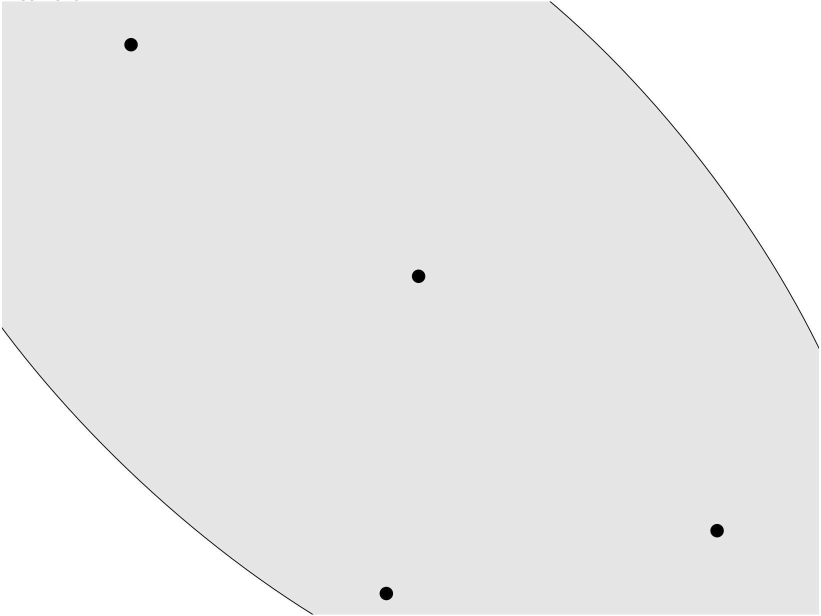
Nested Sampling

Graphical aid



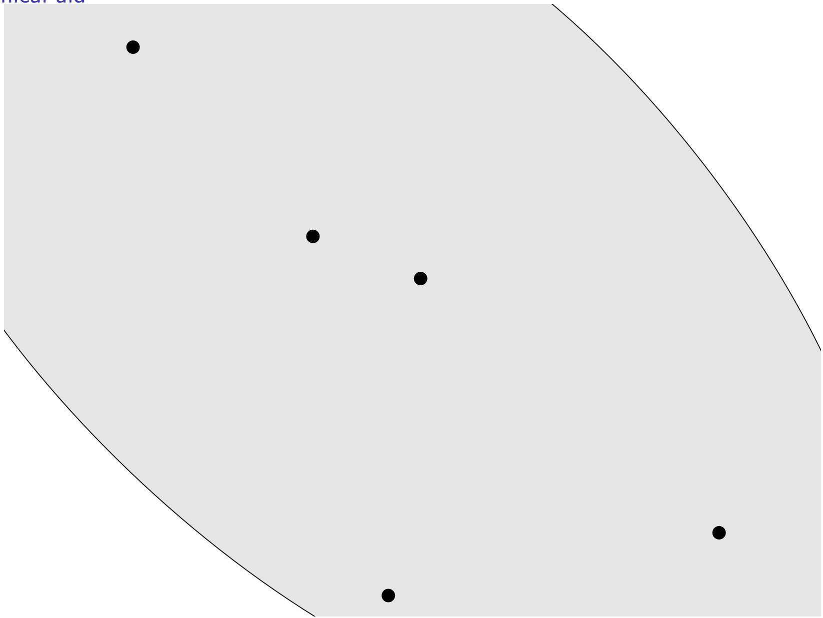
Nested Sampling

Graphical aid



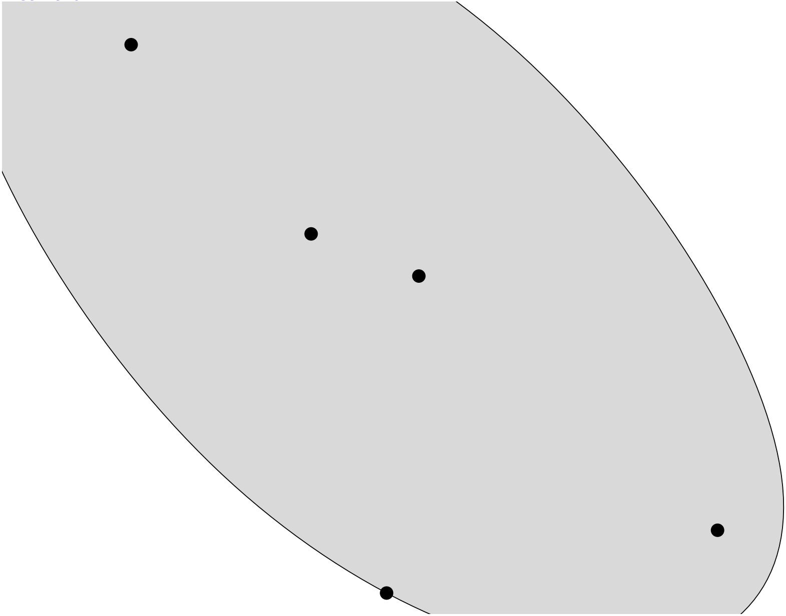
Nested Sampling

Graphical aid



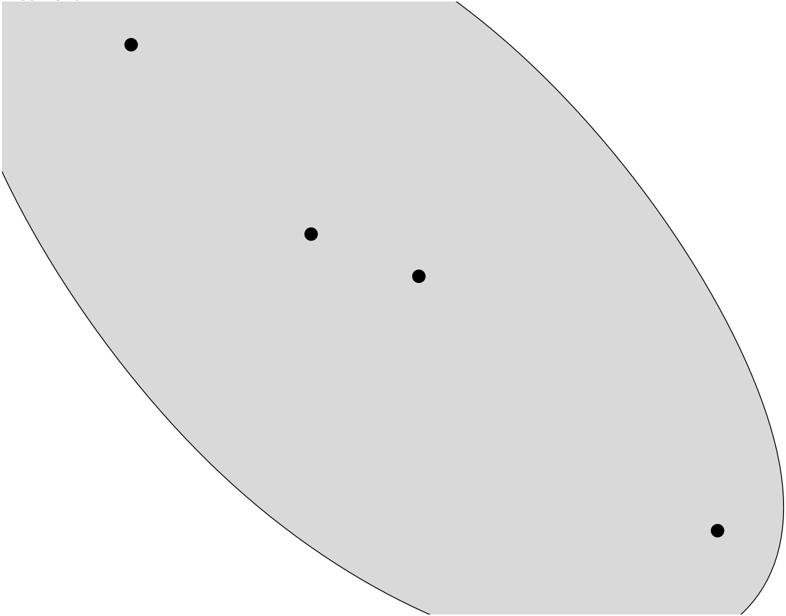
Nested Sampling

Graphical aid



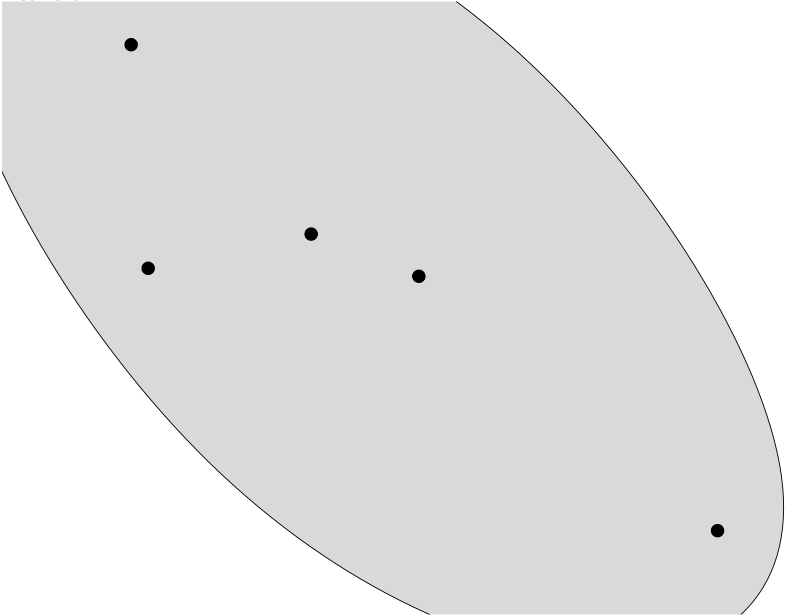
Nested Sampling

Graphical aid



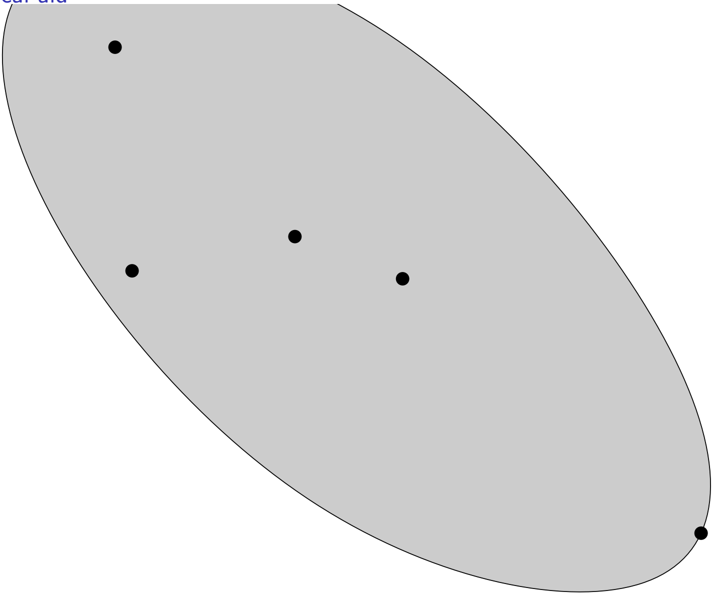
Nested Sampling

Graphical aid



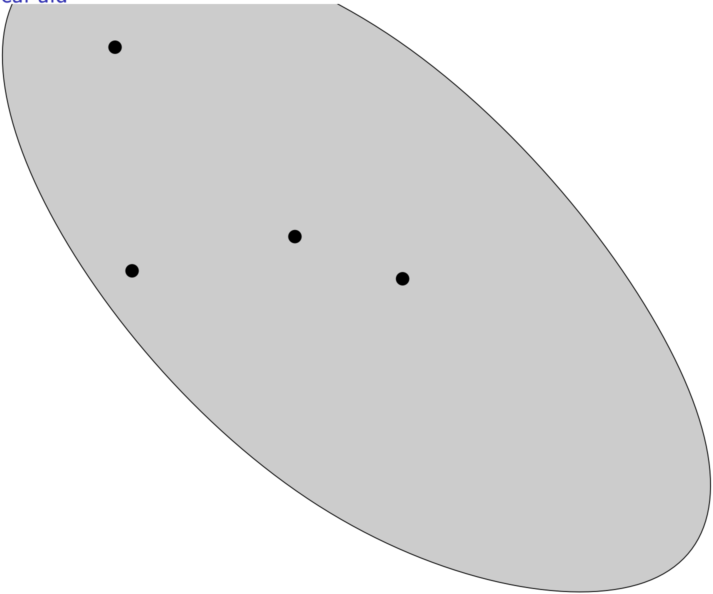
Nested Sampling

Graphical aid



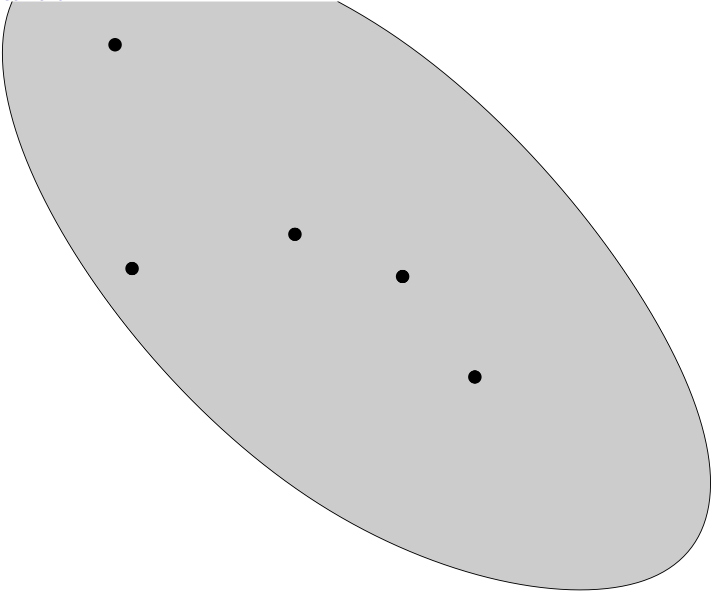
Nested Sampling

Graphical aid



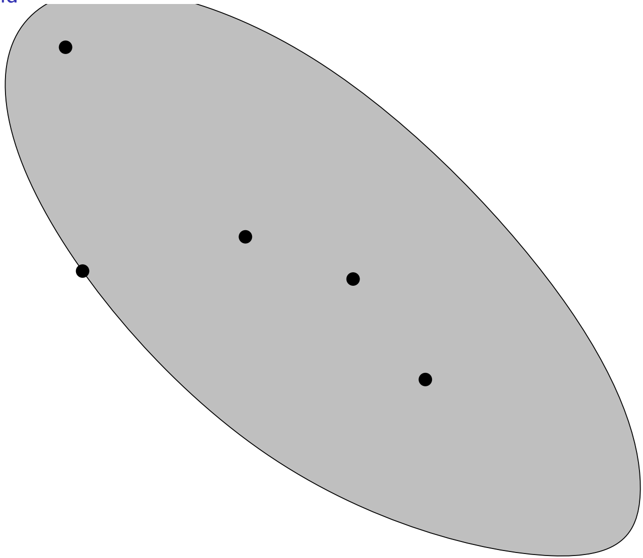
Nested Sampling

Graphical aid



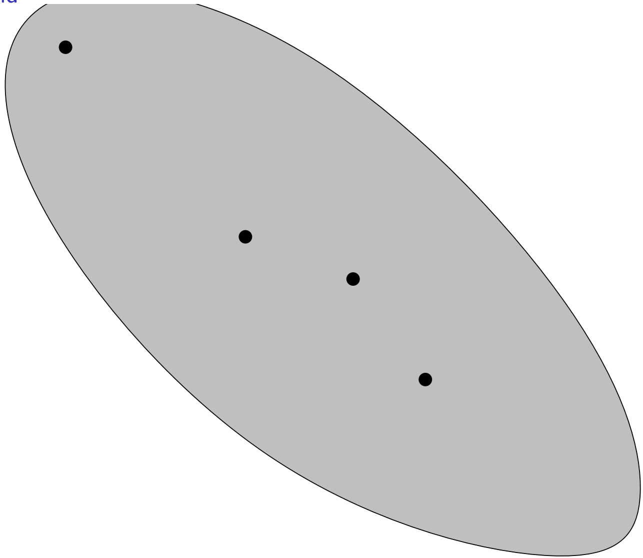
Nested Sampling

Graphical aid



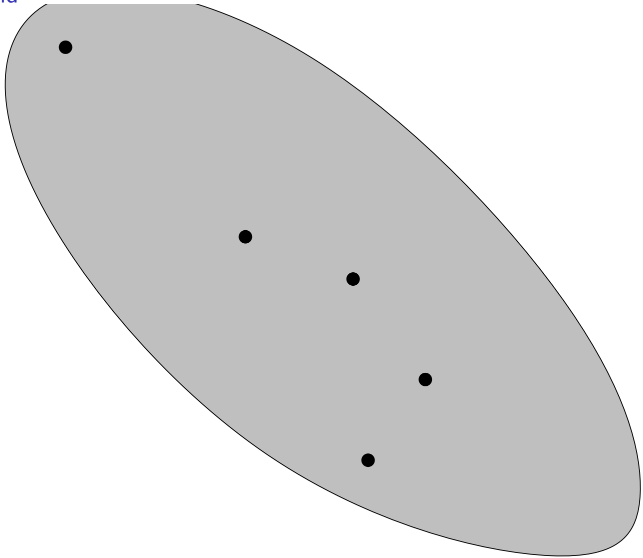
Nested Sampling

Graphical aid



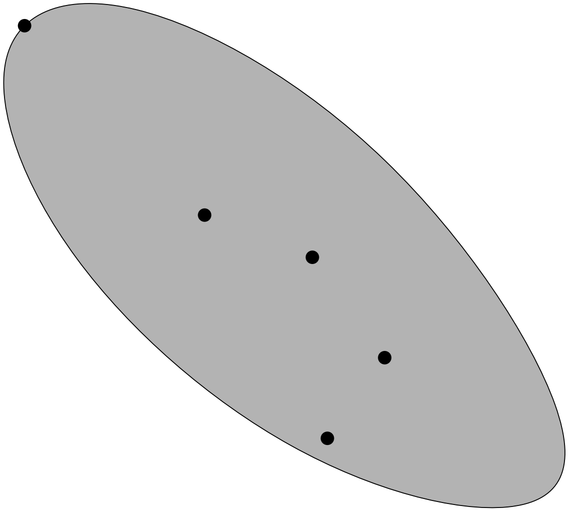
Nested Sampling

Graphical aid



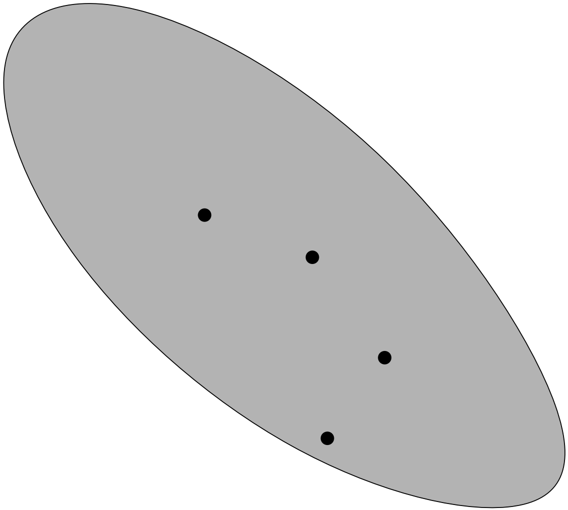
Nested Sampling

Graphical aid



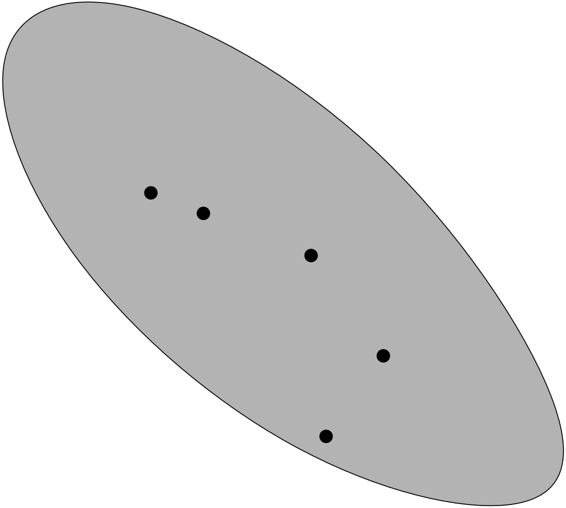
Nested Sampling

Graphical aid



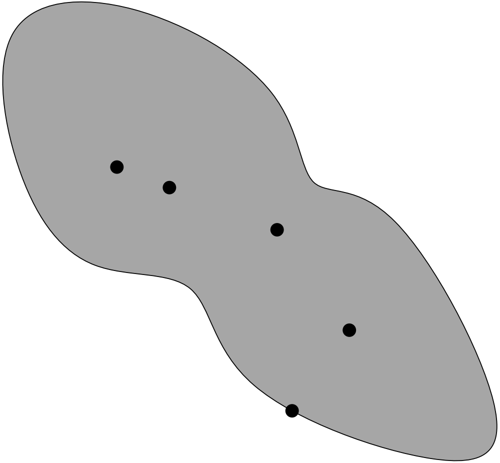
Nested Sampling

Graphical aid



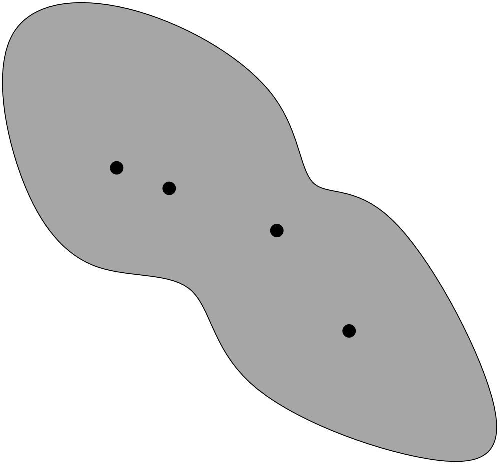
Nested Sampling

Graphical aid



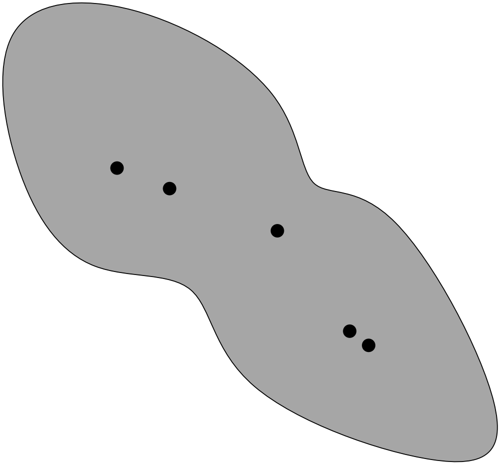
Nested Sampling

Graphical aid



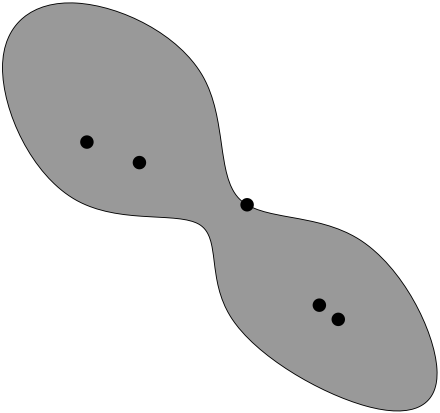
Nested Sampling

Graphical aid



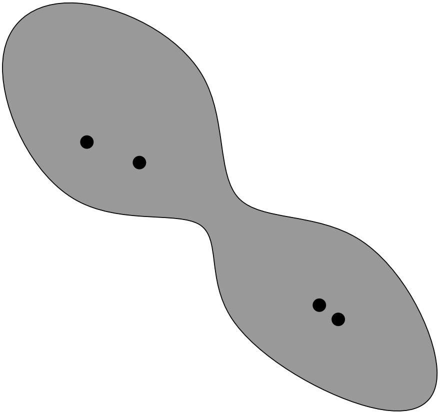
Nested Sampling

Graphical aid



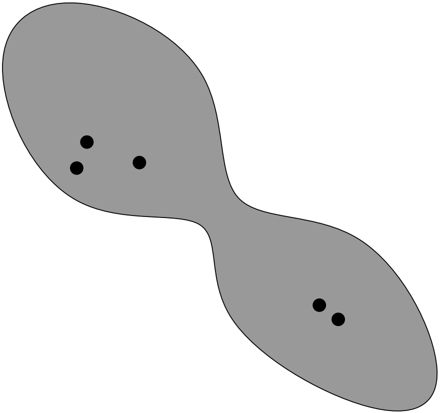
Nested Sampling

Graphical aid



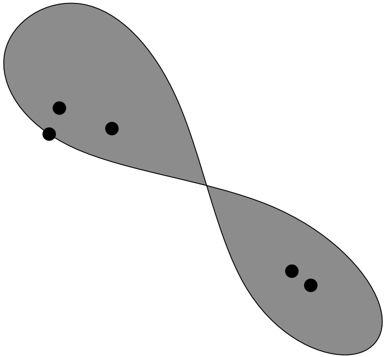
Nested Sampling

Graphical aid



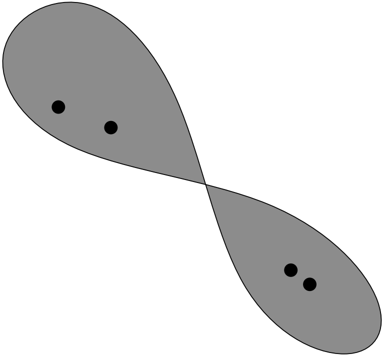
Nested Sampling

Graphical aid



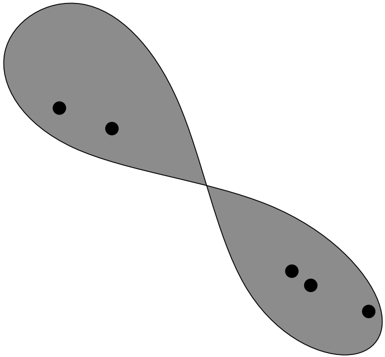
Nested Sampling

Graphical aid



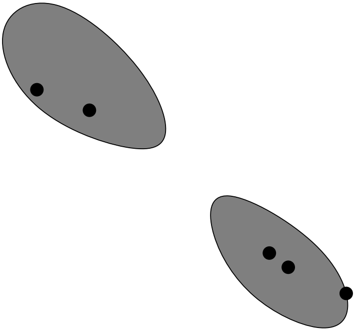
Nested Sampling

Graphical aid



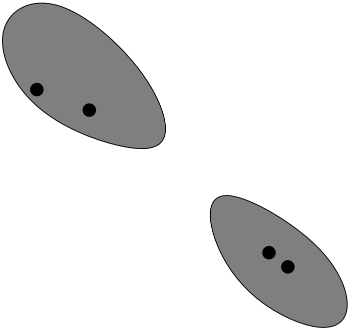
Nested Sampling

Graphical aid



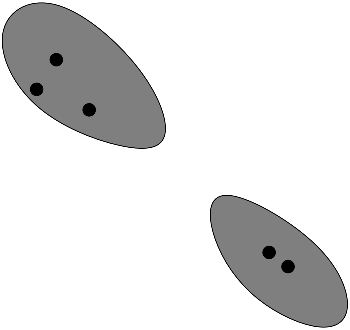
Nested Sampling

Graphical aid



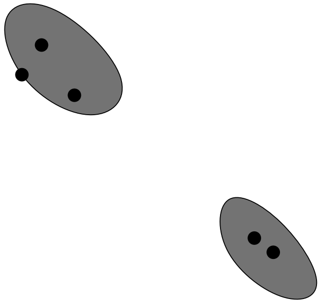
Nested Sampling

Graphical aid



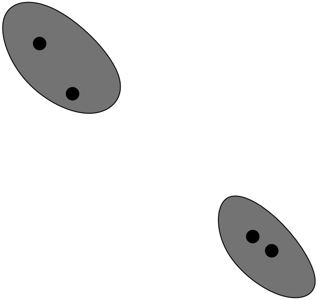
Nested Sampling

Graphical aid



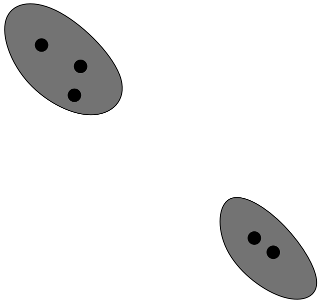
Nested Sampling

Graphical aid



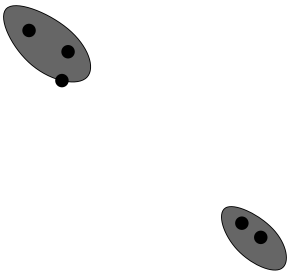
Nested Sampling

Graphical aid



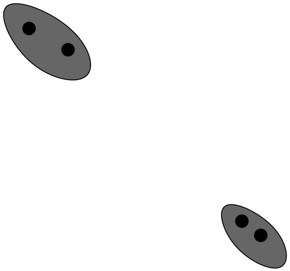
Nested Sampling

Graphical aid



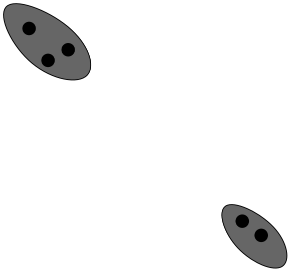
Nested Sampling

Graphical aid



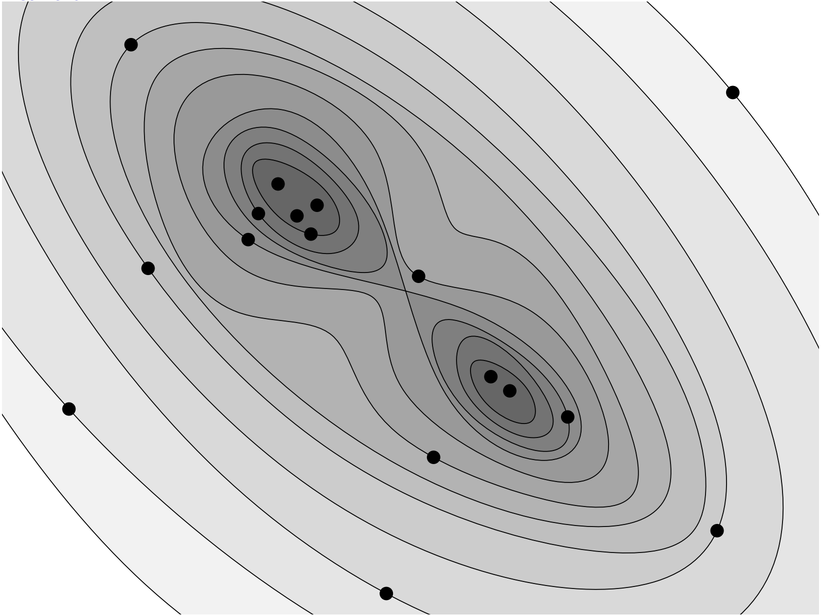
Nested Sampling

Graphical aid



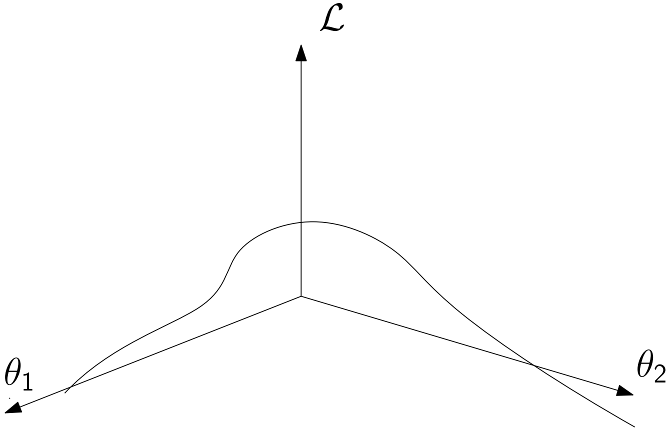
Nested Sampling

Graphical aid



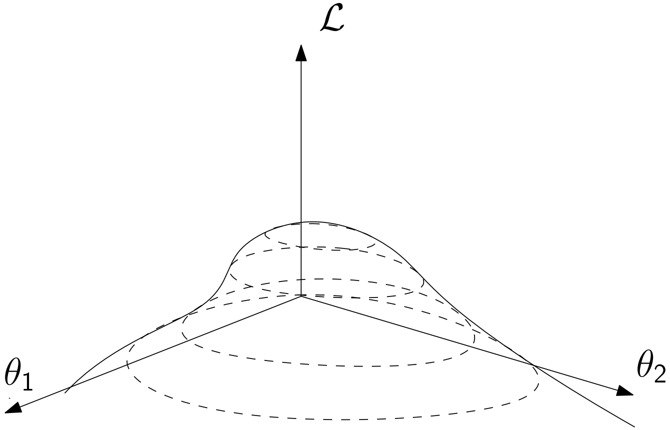
Nested Sampling

Calculating evidences



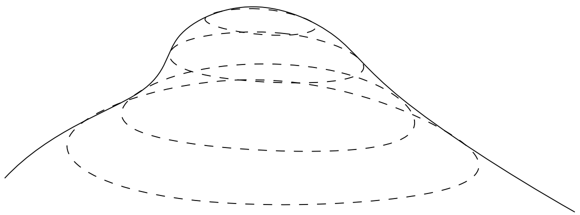
Nested Sampling

Calculating evidences



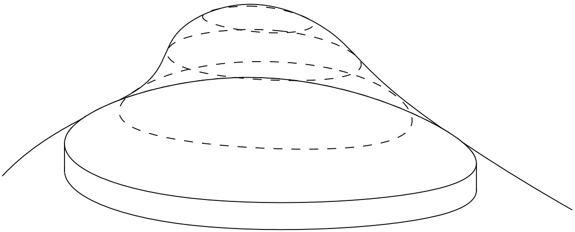
Nested Sampling

Calculating evidences



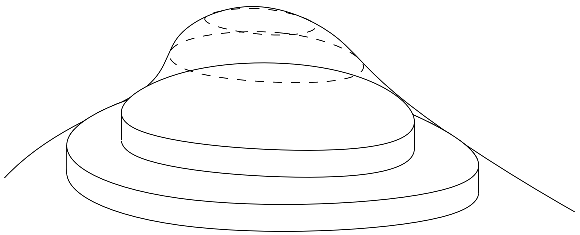
Nested Sampling

Calculating evidences



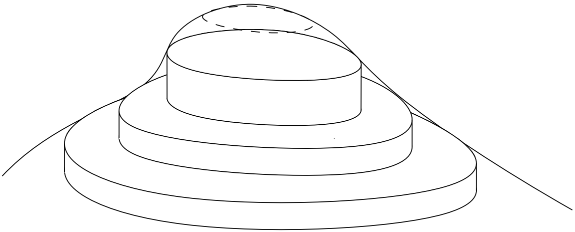
Nested Sampling

Calculating evidences



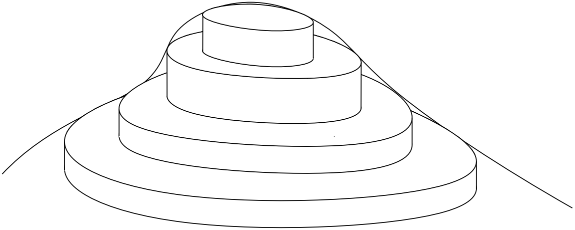
Nested Sampling

Calculating evidences



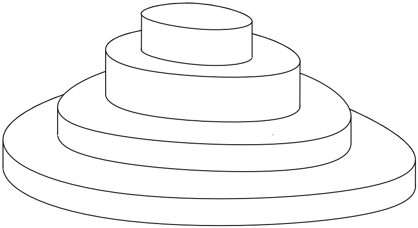
Nested Sampling

Calculating evidences



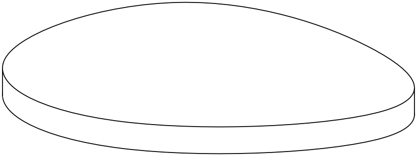
Nested Sampling

Calculating evidences



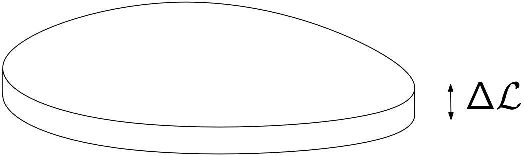
Nested Sampling

Calculating evidences



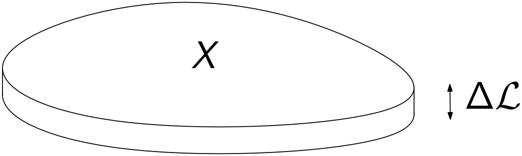
Nested Sampling

Calculating evidences



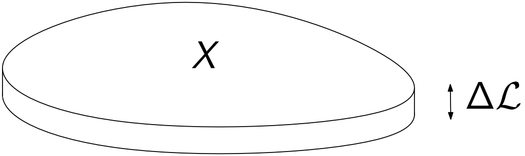
Nested Sampling

Calculating evidences



Nested Sampling

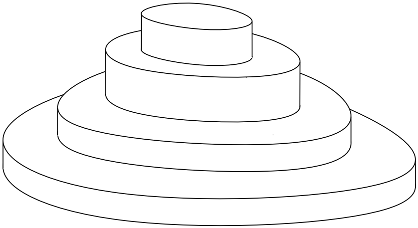
Calculating evidences



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

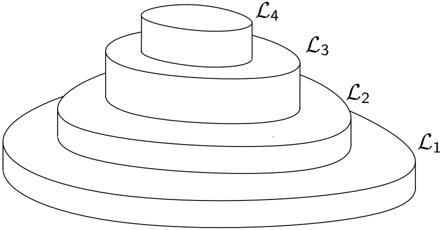
Nested Sampling

Calculating evidences



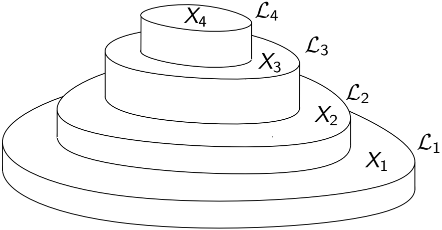
Nested Sampling

Calculating evidences



Nested Sampling

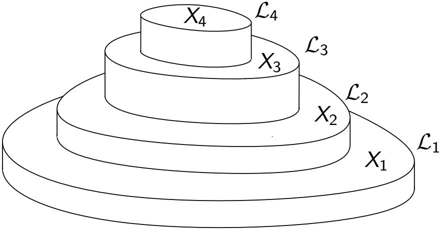
Calculating evidences



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 M.J. Betancourt (2010)

Sampling within an iso-likelihood contour

Previous attempts

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

- Suffers in high dimensions

Hamiltonian: M.J. Betancourt (2010)

Galilean: 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 M.J. Betancourt (2010)

Galilean 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 M.J. Betancourt (2010)

Galilean F. Feroz & J. Skilling (2013)

- Requires gradients and tuning

Diffusive Nested Sampling B. Brewer et al. (2009,2016).

Sampling within an iso-likelihood contour

Previous attempts

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

- Suffers in high dimensions

Hamiltonian M.J. Betancourt (2010)

Galilean F. Feroz & J. Skilling (2013)

- Requires gradients and tuning

Diffusive Nested Sampling B. Brewer et al. (2009,2016).

- Very promising

Sampling within an iso-likelihood contour

Previous attempts

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

- Suffers in high dimensions

Hamiltonian M.J. Betancourt (2010)

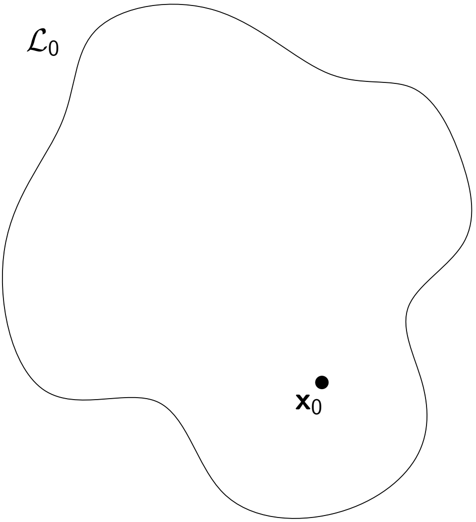
Galilean F. Feroz & J. Skilling (2013)

- Requires gradients and tuning

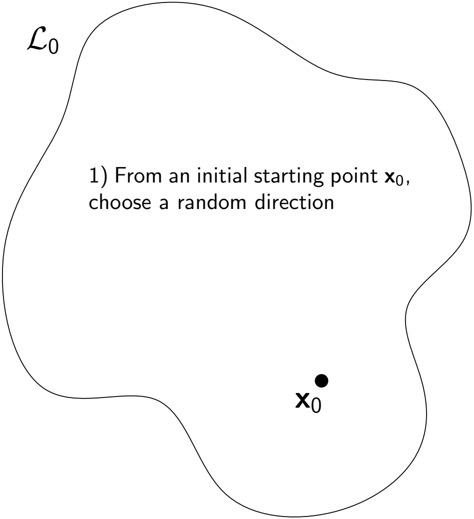
Diffusive Nested Sampling B. Brewer et al. (2009,2016).

- Very promising
- Still needs tuning.

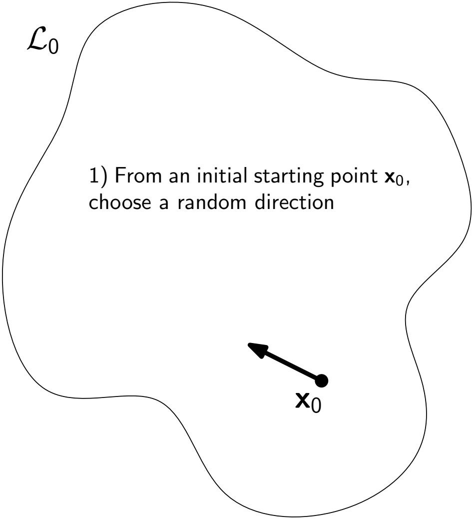
"Hit and run" slice sampling



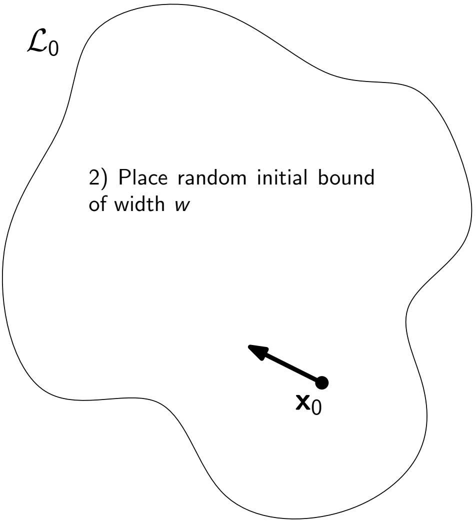
"Hit and run" slice sampling



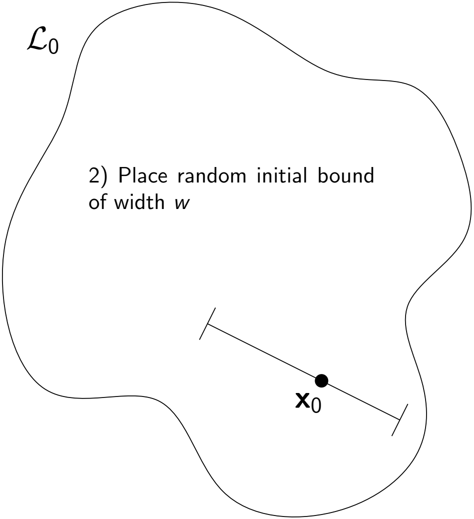
"Hit and run" slice sampling



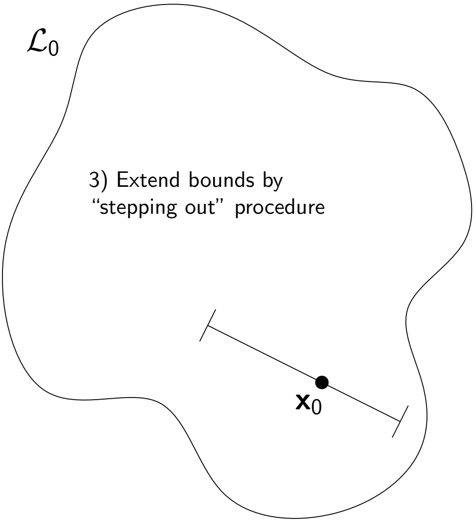
"Hit and run" slice sampling



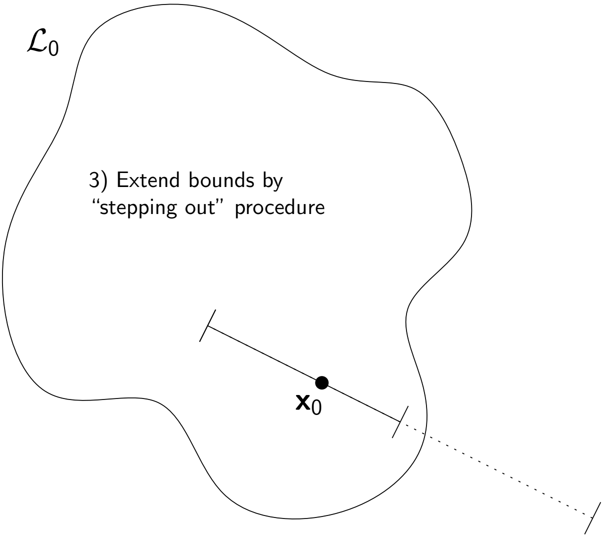
“Hit and run” slice sampling



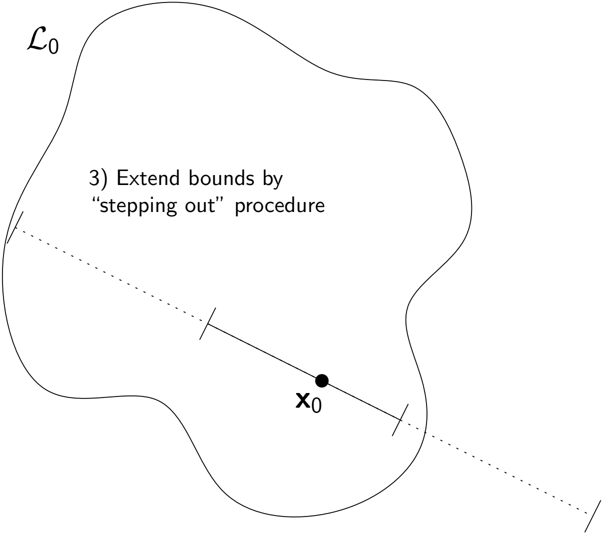
"Hit and run" slice sampling



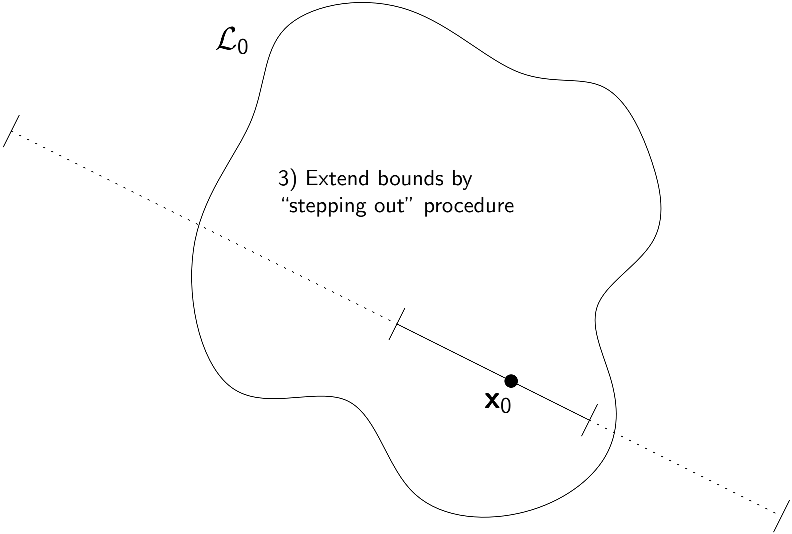
“Hit and run” slice sampling



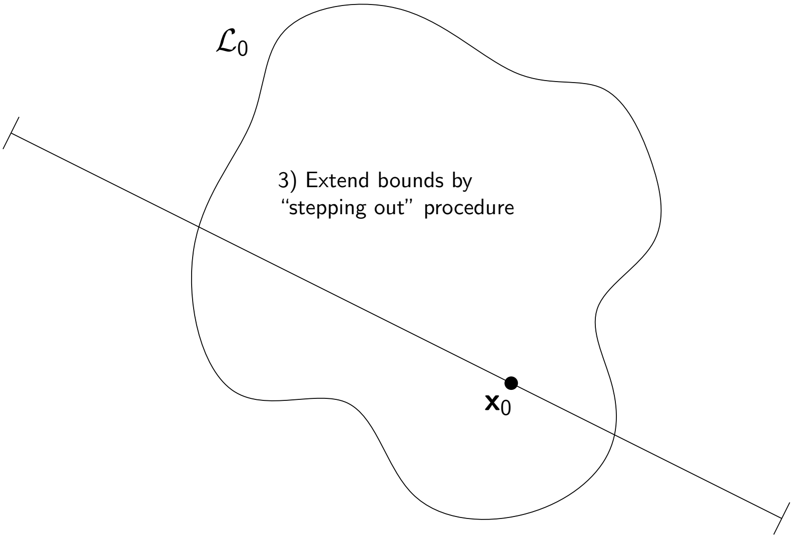
“Hit and run” slice sampling



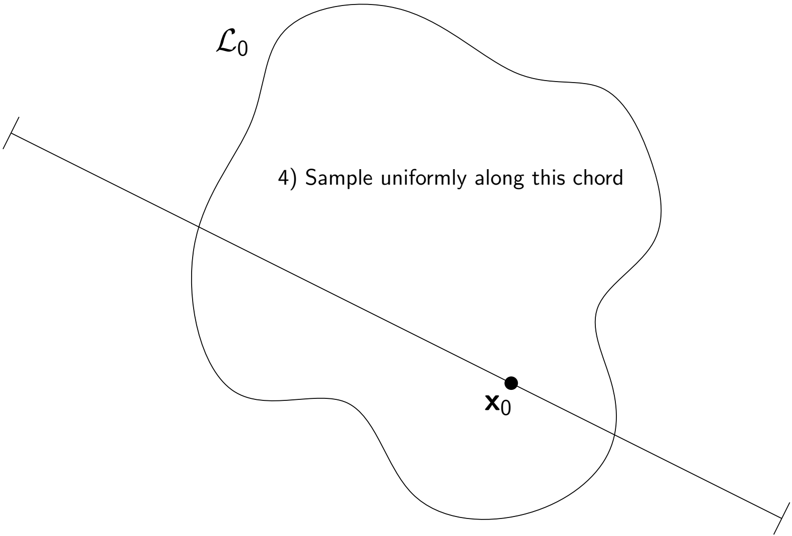
"Hit and run" slice sampling



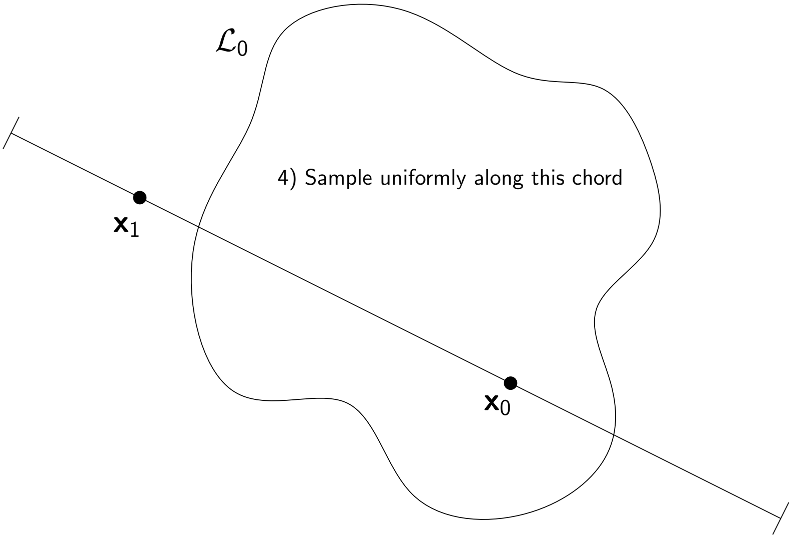
“Hit and run” slice sampling



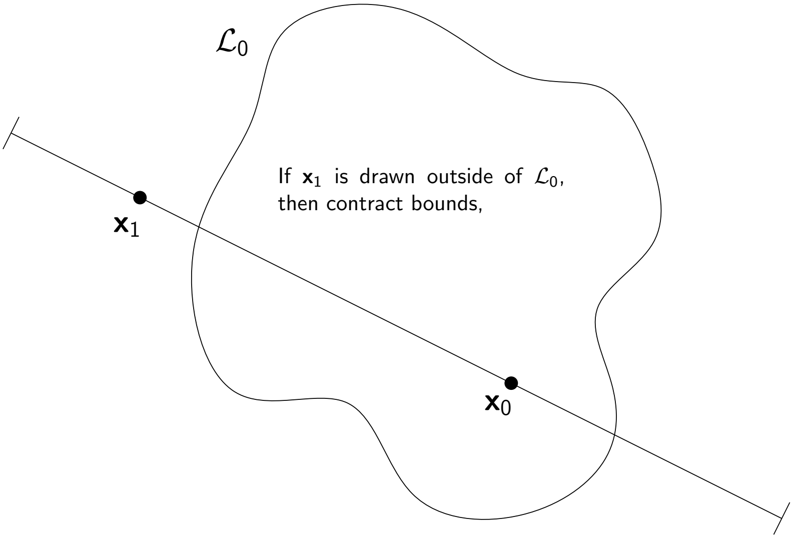
“Hit and run” slice sampling



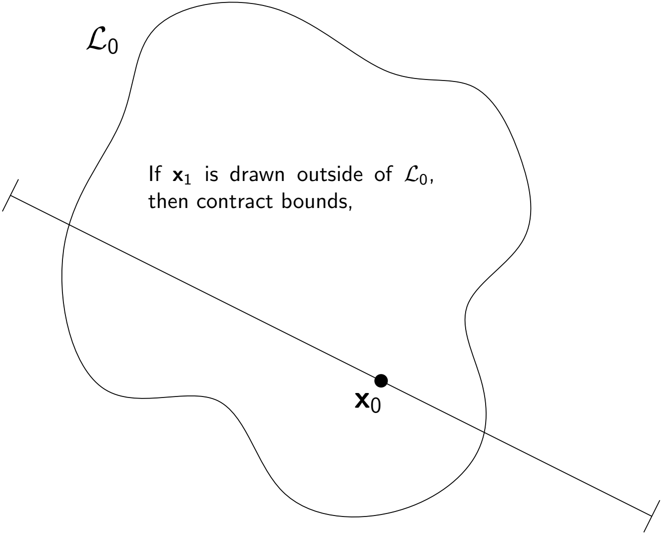
"Hit and run" slice sampling



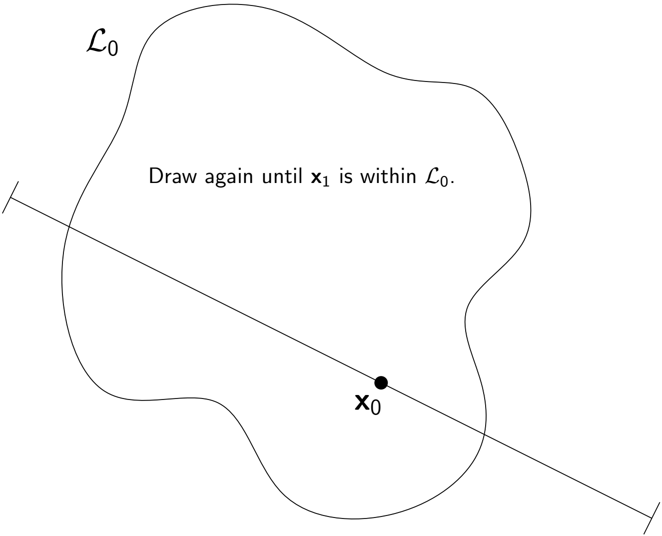
“Hit and run” slice sampling



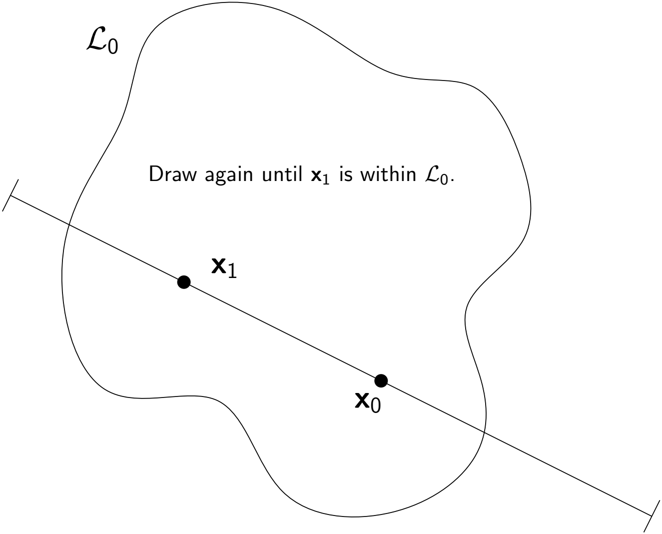
“Hit and run” slice sampling



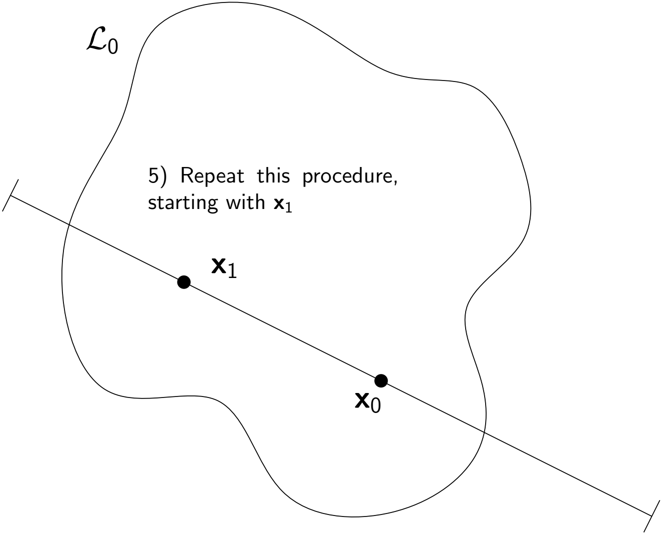
“Hit and run” slice sampling



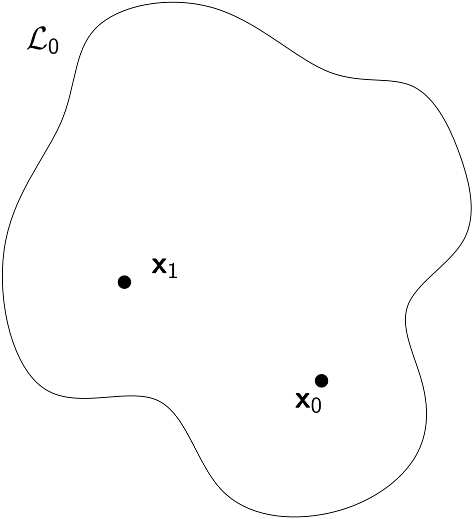
"Hit and run" slice sampling



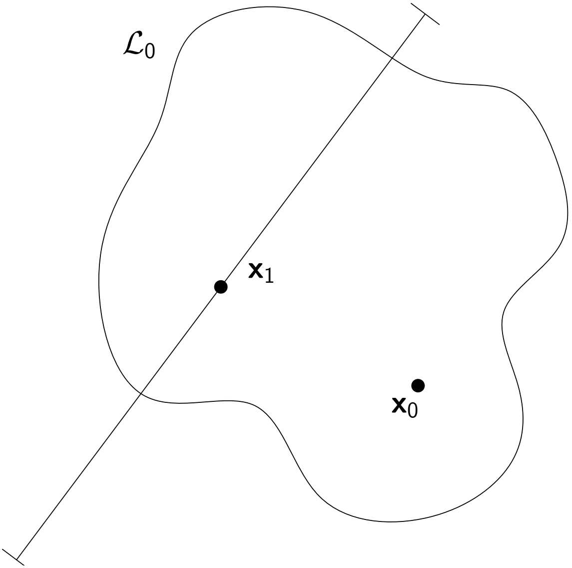
"Hit and run" slice sampling



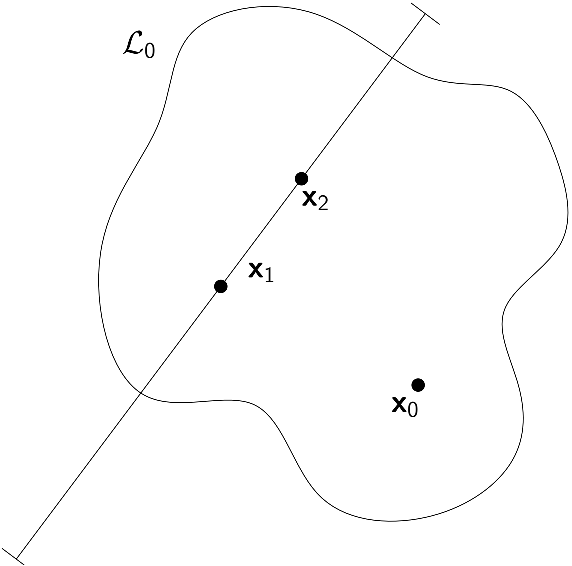
“Hit and run” slice sampling



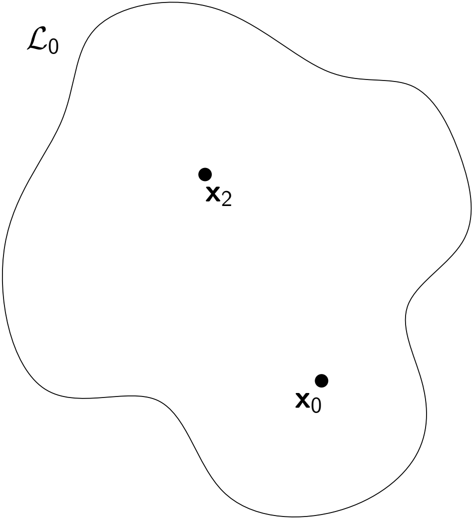
“Hit and run” slice sampling



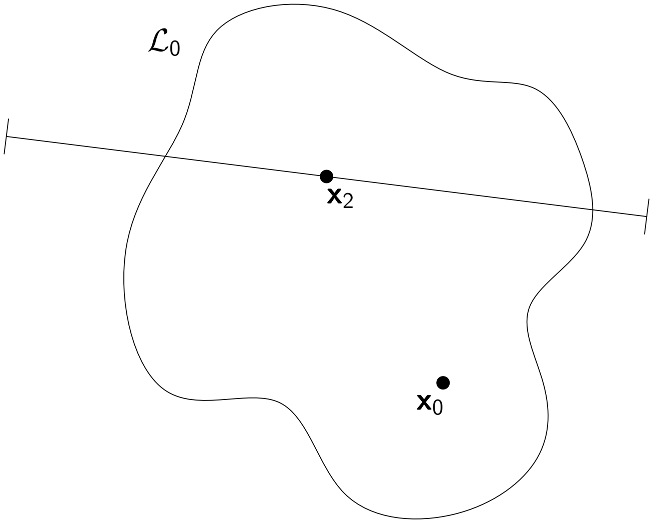
"Hit and run" slice sampling



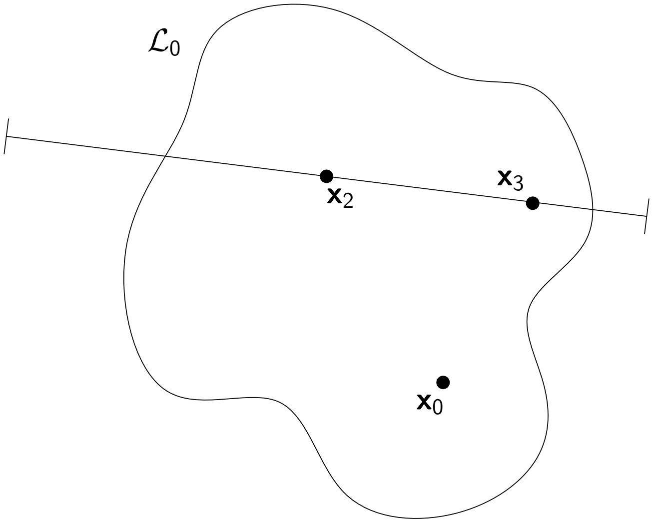
"Hit and run" slice sampling



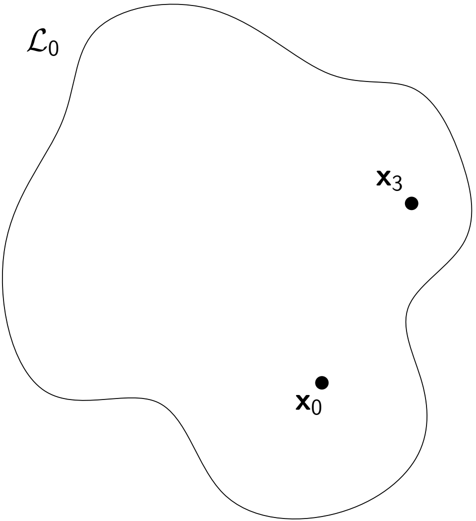
"Hit and run" slice sampling



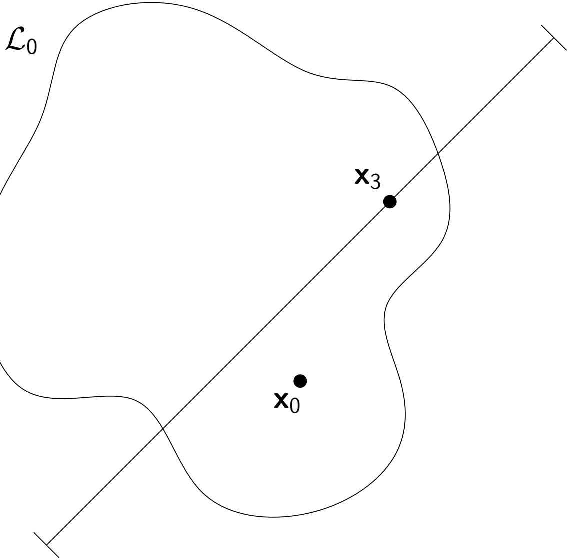
"Hit and run" slice sampling



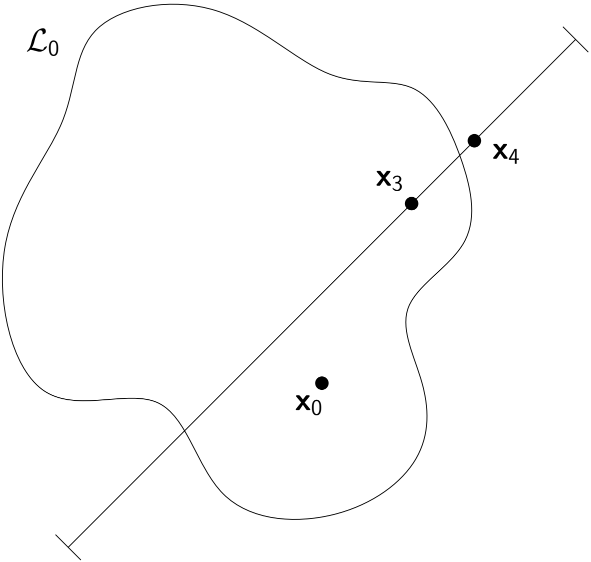
"Hit and run" slice sampling



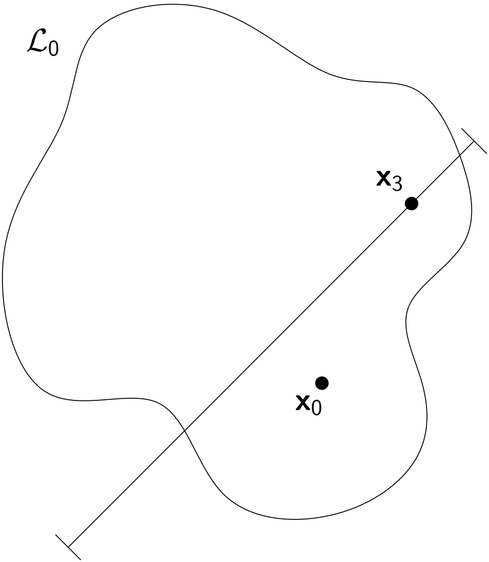
"Hit and run" slice sampling



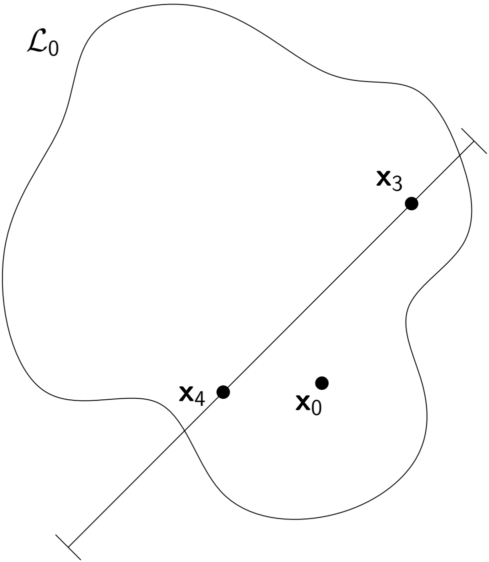
"Hit and run" slice sampling



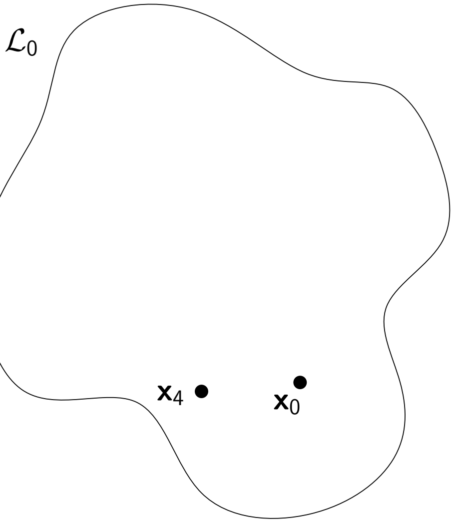
“Hit and run” slice sampling



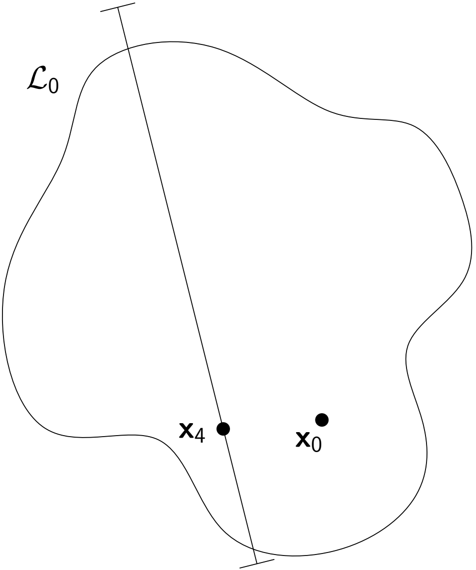
"Hit and run" slice sampling



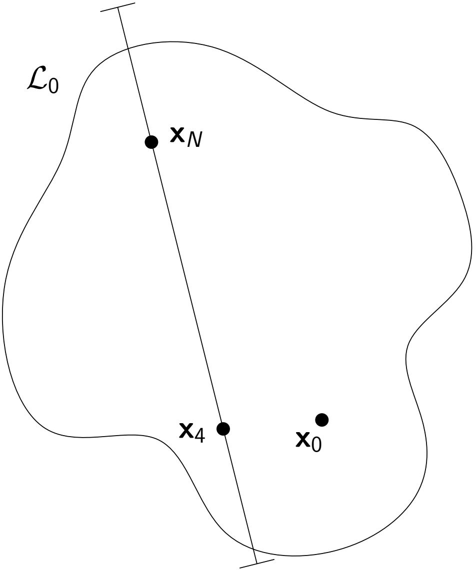
"Hit and run" slice sampling



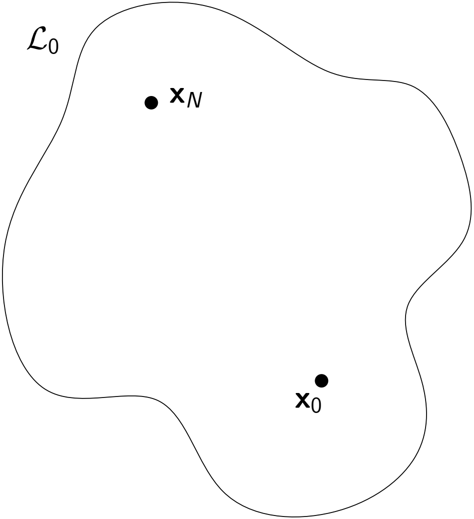
"Hit and run" slice sampling



"Hit and run" slice sampling



“Hit and run” slice sampling



Issues with Slice Sampling

Correlated distributions

Issues with Slice Sampling

Correlated distributions

- 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

- Need N reasonably large $\sim \mathcal{O}(n_{\mathrm{dims}})$ so that x_N is de-correlated from x_1.
- Does not deal well with correlated distributions.

Issues with Slice Sampling

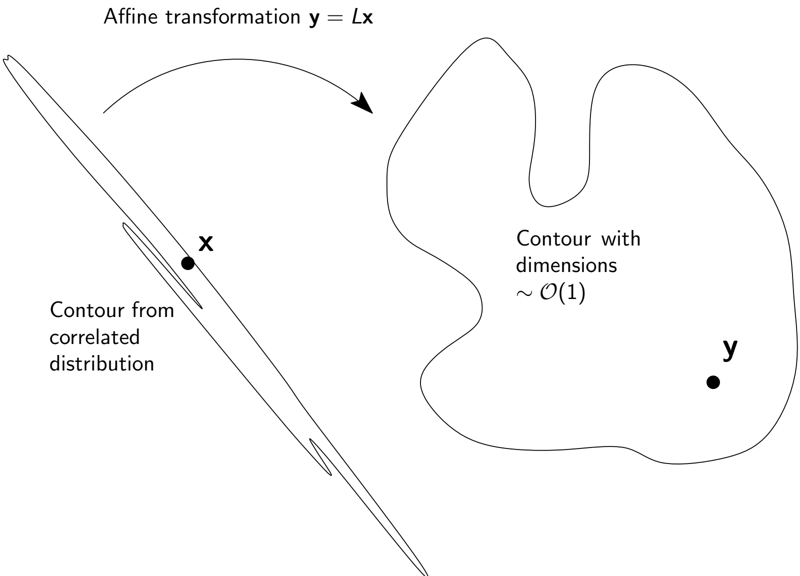
Correlated distributions

- Need N reasonably large $\sim \mathcal{O}(n_{\mathrm{dims}})$ so that x_N is de-correlated from x_1.
- Does not deal well with correlated distributions.
- Need to "tune" w parameter.

PolyChord 1.0's solution

Correlated distributions

,



PolyChord 1.0's solutions

Correlated distributions

PolyChord 1.0's solutions

Correlated distributions

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

PolyChord 1.0's solutions

Correlated distributions

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

PolyChord 1.0'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 1.0'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 1.0'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 1.0's solutions

Multimodality

PolyChord 1.0's solutions
Multimodality

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

PolyChord 1.0's solutions

Multimodality

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

PolyChord 1.0's Additions

PolyChord 1.0's Additions

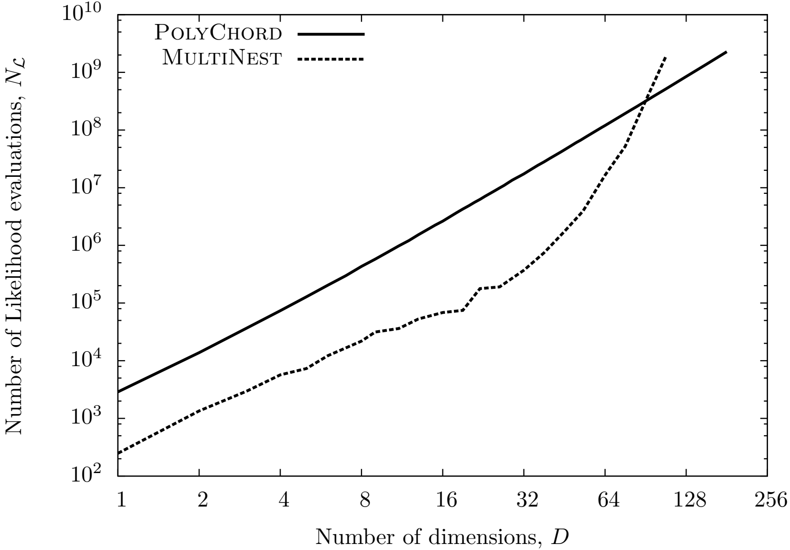
- Parallelised up to number of live points with openMPI.

PolyChord 1.0'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 $n_{\rm live}$ 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 $n_{\rm live}$ 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_{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

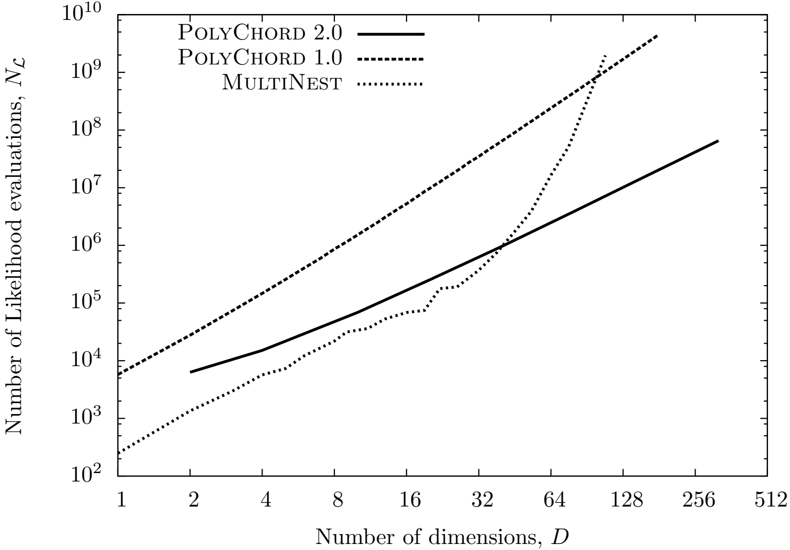
PolyChord 2.0

Handling correlations

- Unweave the run into $n_{\rm live}$ 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
- With this in hand, can produce correct inferences from correlated runs.

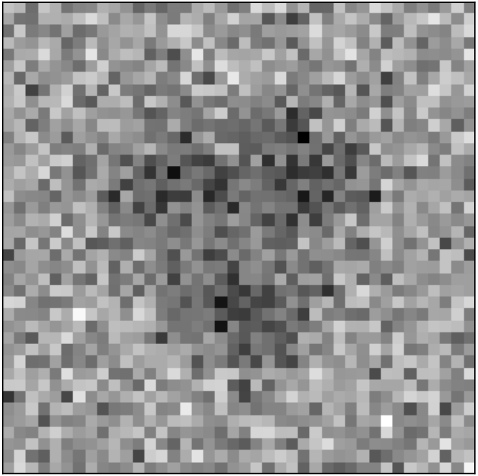
PolyChord 2.0 vs. MultiNest

Gaussian likelihood



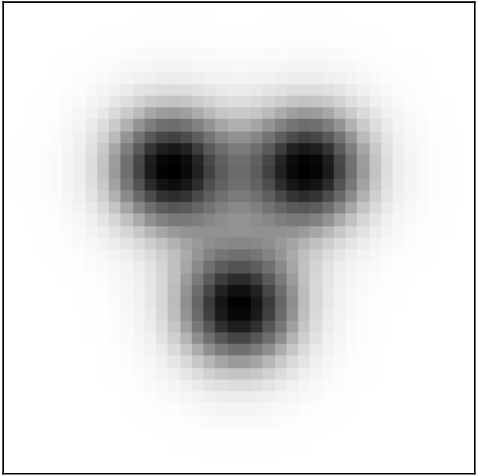
Object detection

Toy problem



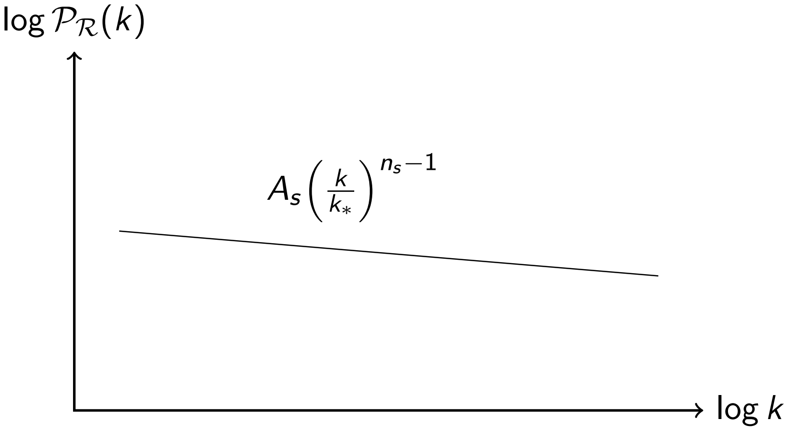
Object detection

Toy problem



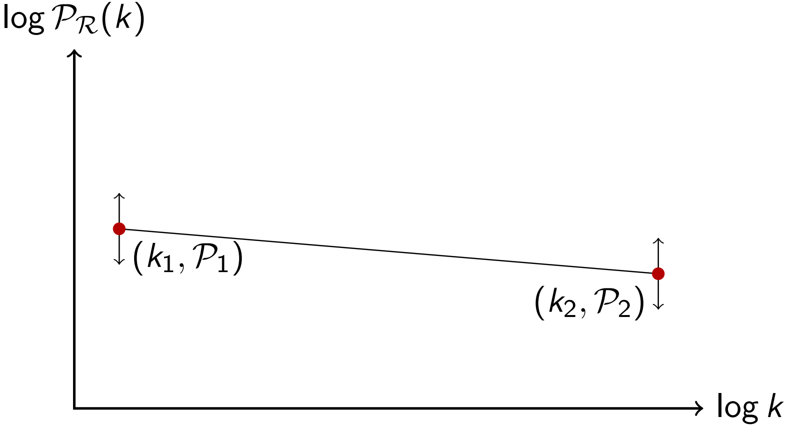
PolyChord in action

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



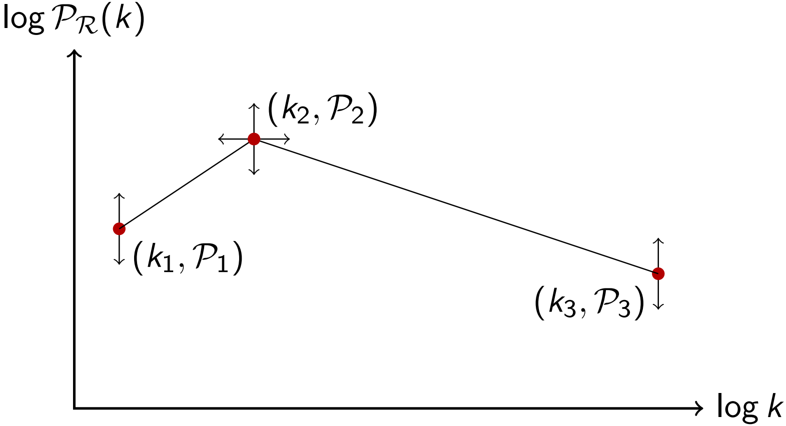
PolyChord in action

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



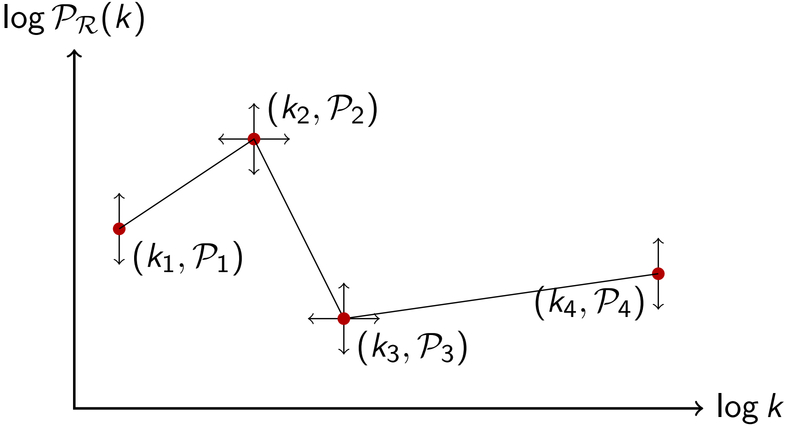
PolyChord in action

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



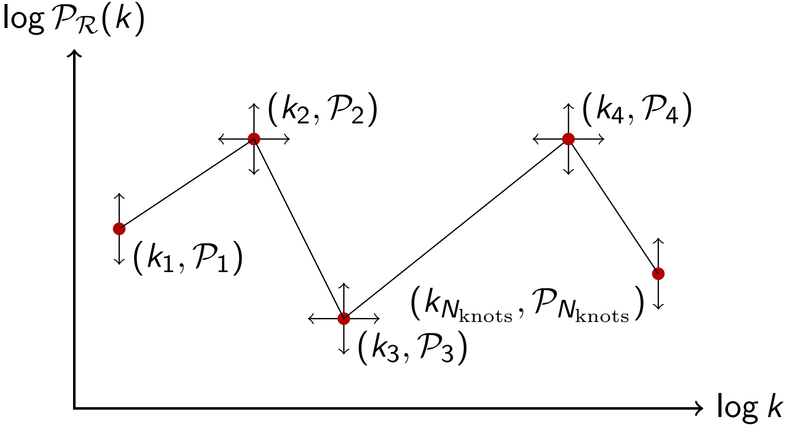
PolyChord in action

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



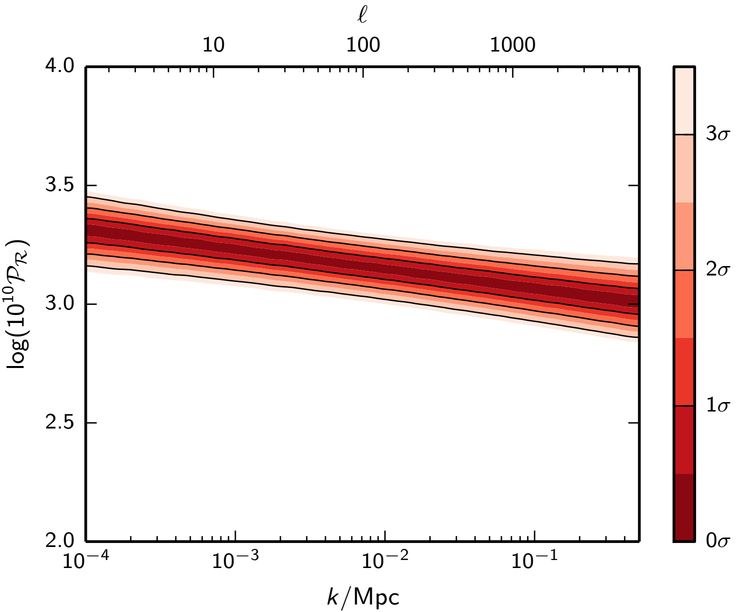
PolyChord in action

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



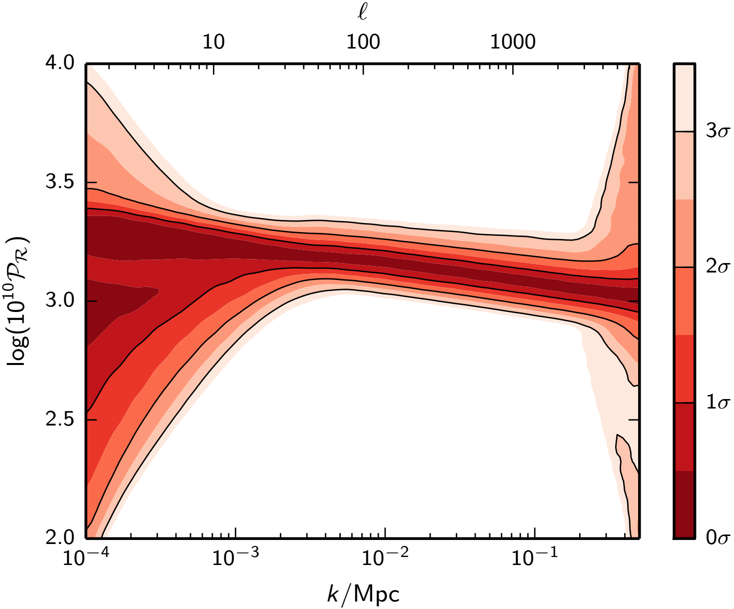
0 internal knots

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



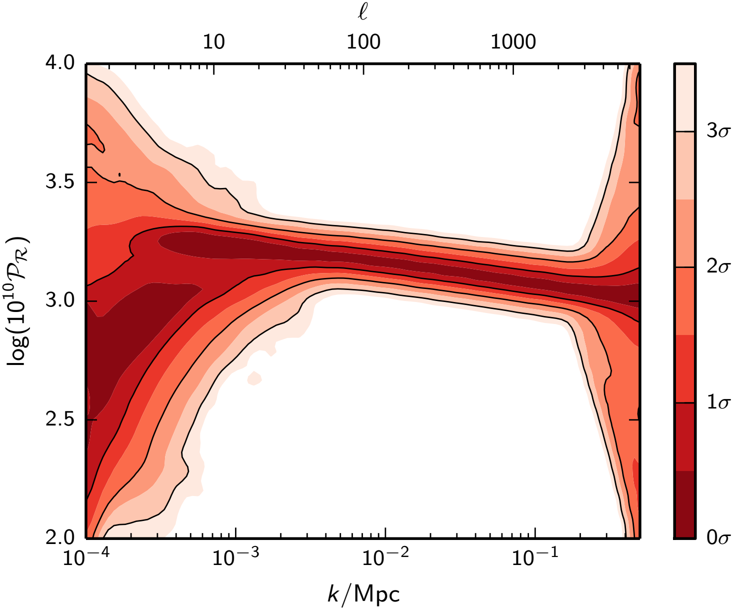
1 internal knots

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



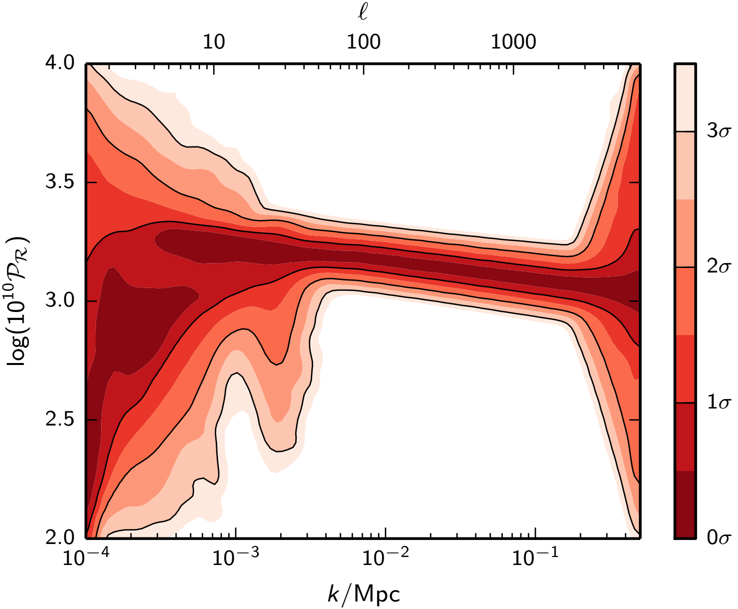
2 internal knots

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



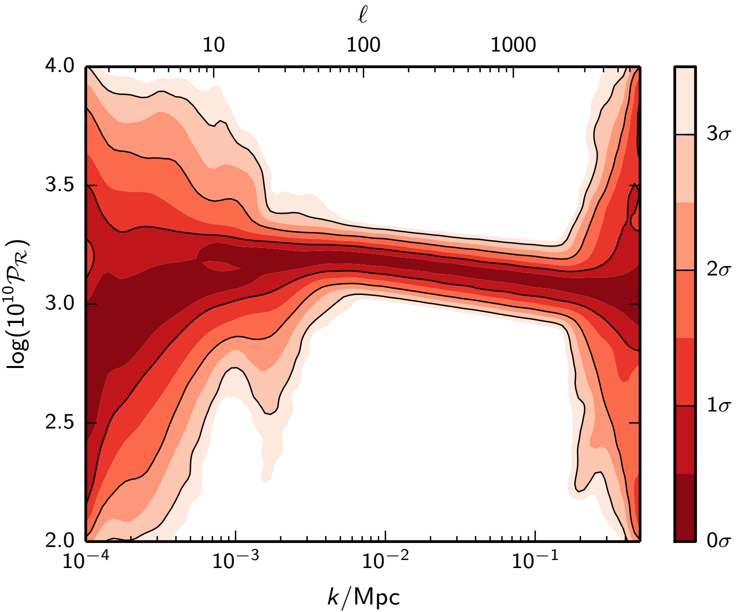
3 internal knots

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



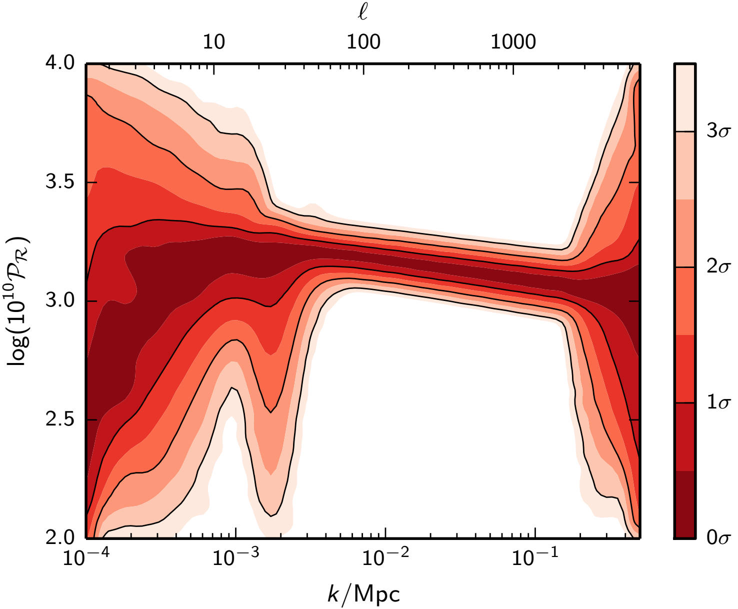
4 internal knots

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



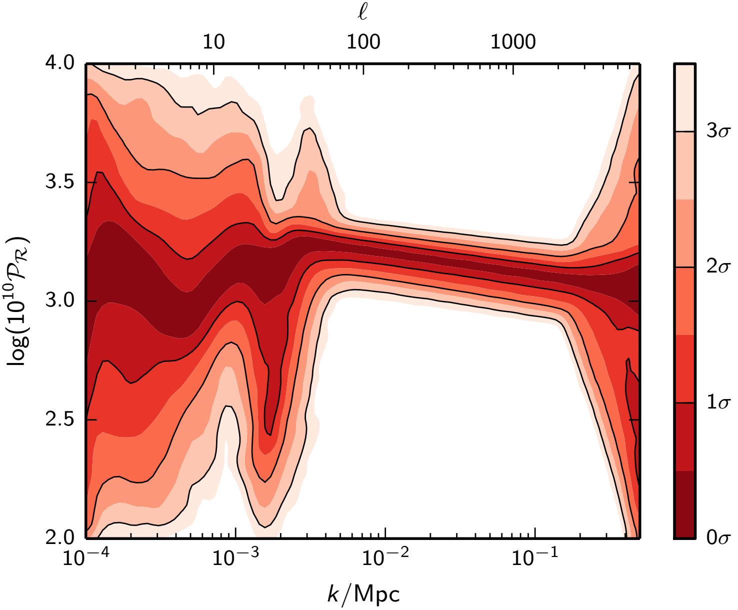
5 internal knots

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



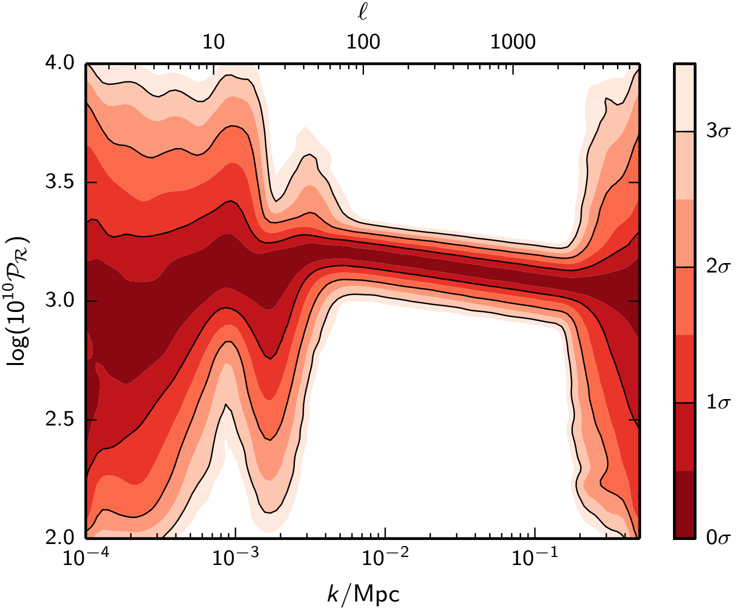
6 internal knots

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



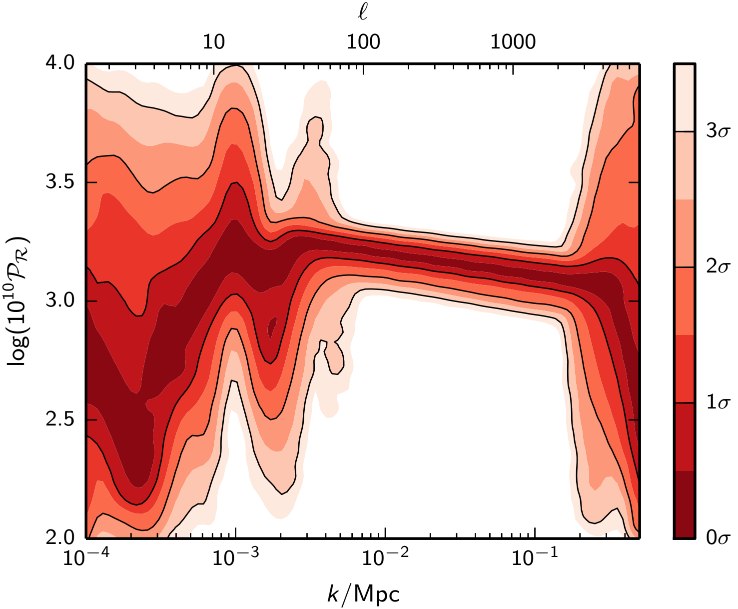
7 internal knots

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



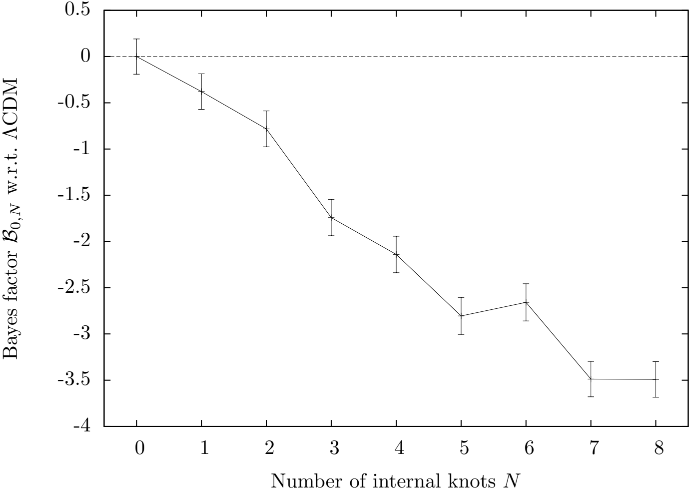
8 internal knots

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



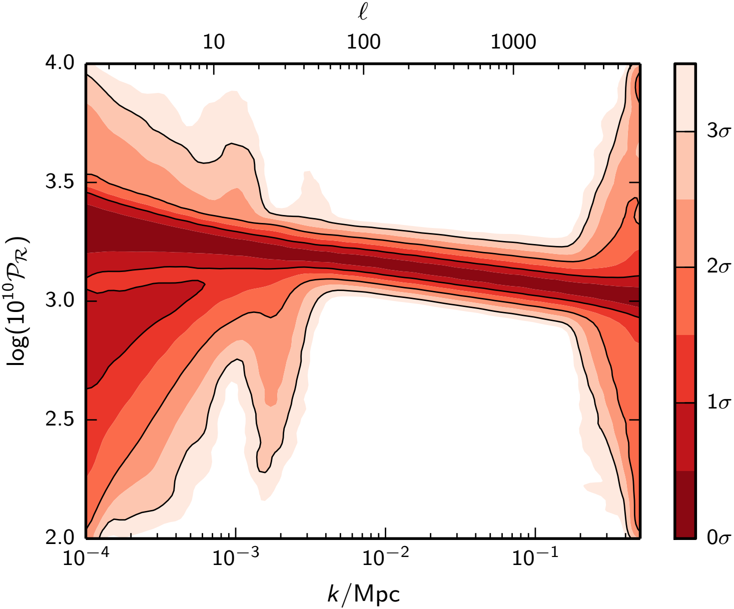
Bayes Factors

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



Marginalised plot

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



Dark energy equation of state reconstruction

Dark energy equation of state reconstruction

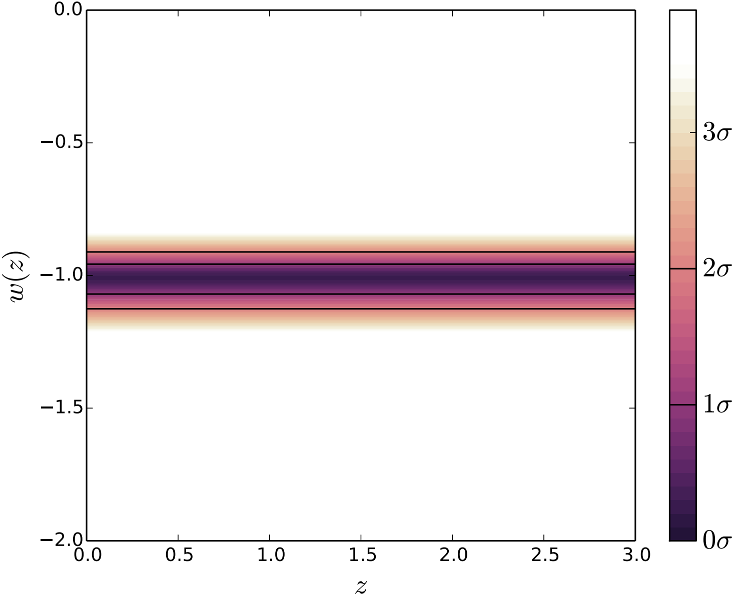
- Same thing, but for Dark energy equation of state $w(z)$ (quintessence).

Dark energy equation of state reconstruction

- Same thing, but for Dark energy equation of state $w(z)$ (quintessence).
- Data used is Planck 2015, BOSS DR 11, JLA supernovae and BOSS Lyα data

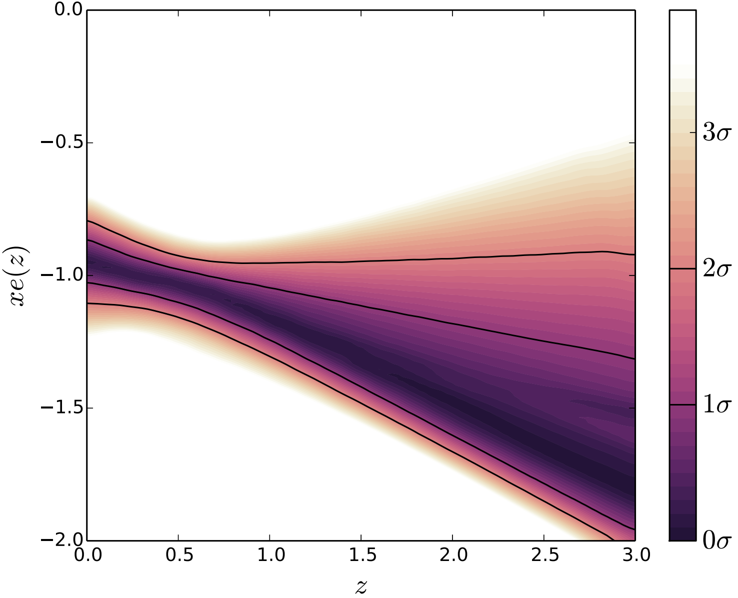
Flat, variable *w*

Dark energy equation of state reconstruction



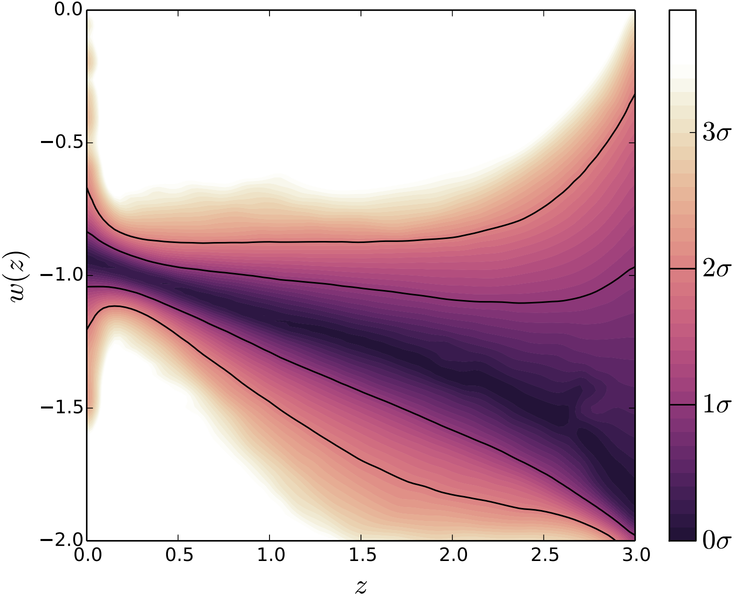
Tilted

Dark energy equation of state reconstruction



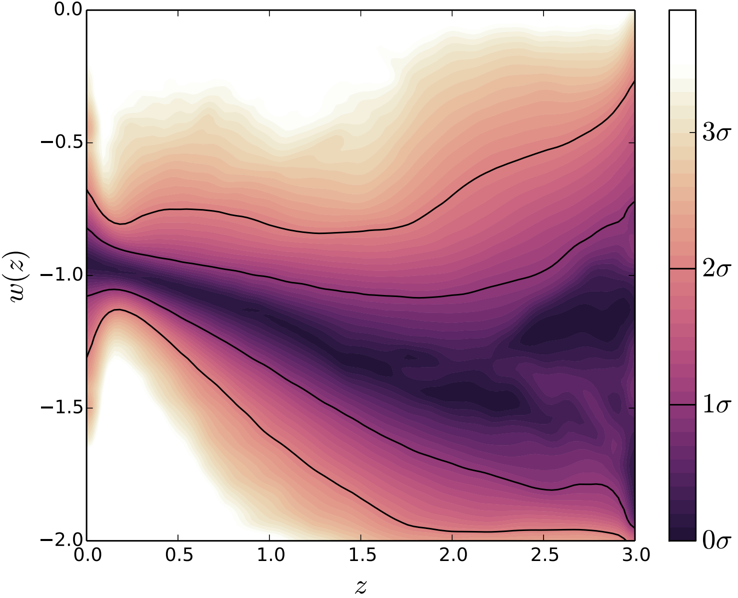
1 internal node

Dark energy equation of state reconstruction



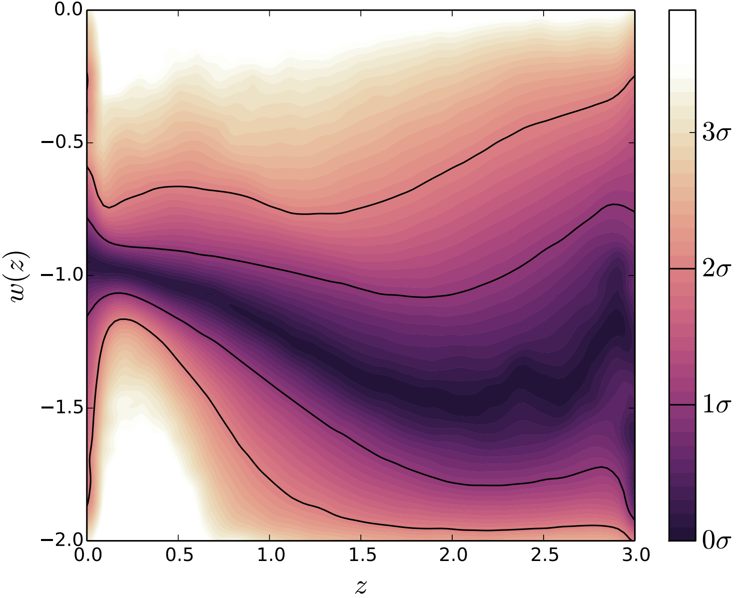
2 internal nodes

Dark energy equation of state reconstruction



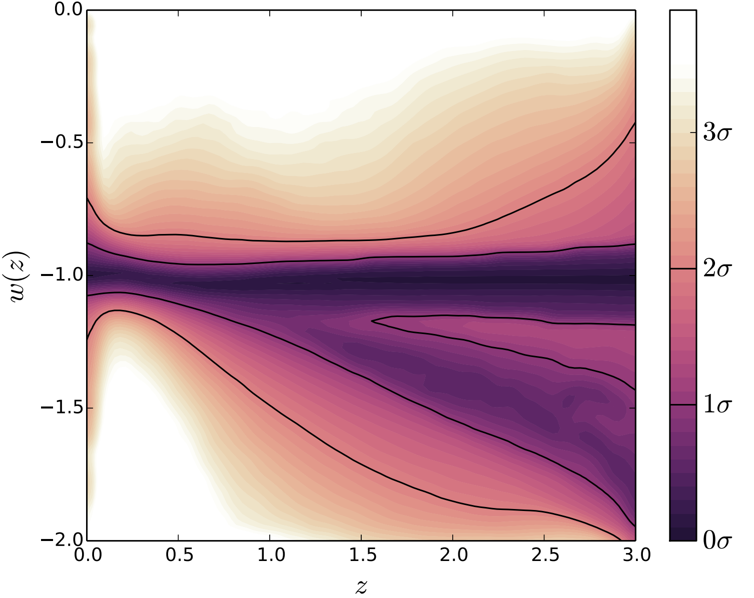
3 internal nodes

Dark energy equation of state reconstruction



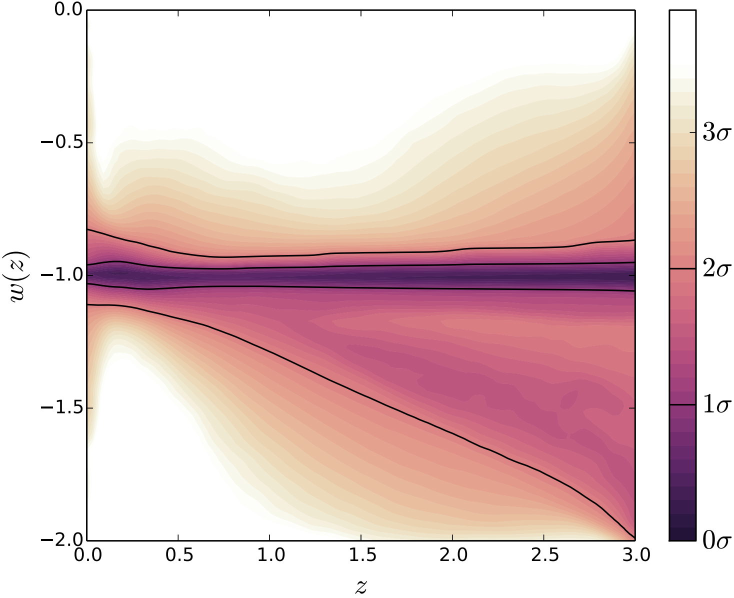
Marginalised plot - just extension models

Dark energy equation of state reconstruction



Marginalised plot - including LCDM

Dark energy equation of state reconstruction



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

Future work

Future work

1. Parallelisation

Future work

1. Parallelisation
2. Affine invariant mode detection.

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

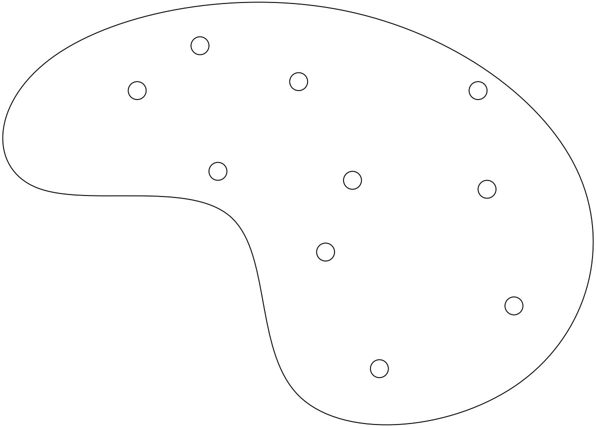
- The optimal exploration technique is be affine invariant.
- Treat distribution $\mathrm{P}(\mathrm{x})$ and $\mathrm{P}(R\mathrm{x})$ the same.
- No need to worry about correlations.

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)

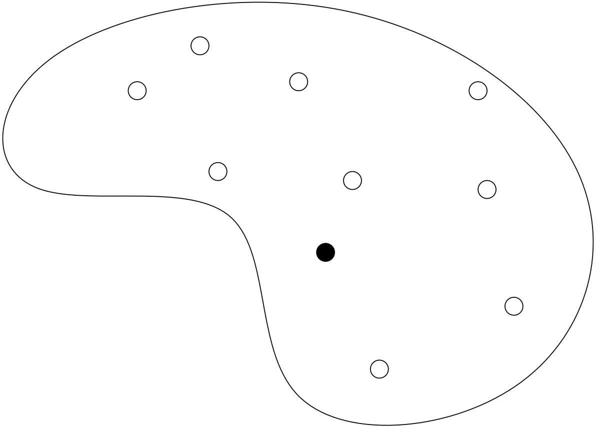
Skillings affine invariant ideas

Leapfrog



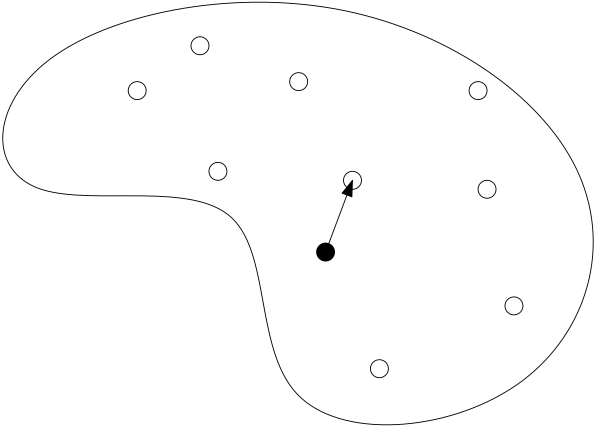
Skillings affine invariant ideas

Leapfrog



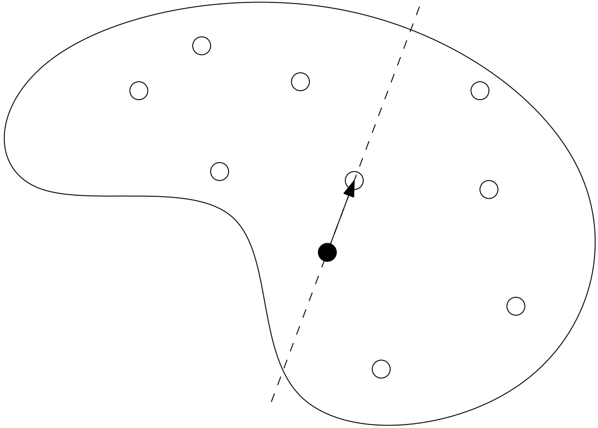
Skillings affine invariant ideas

Leapfrog



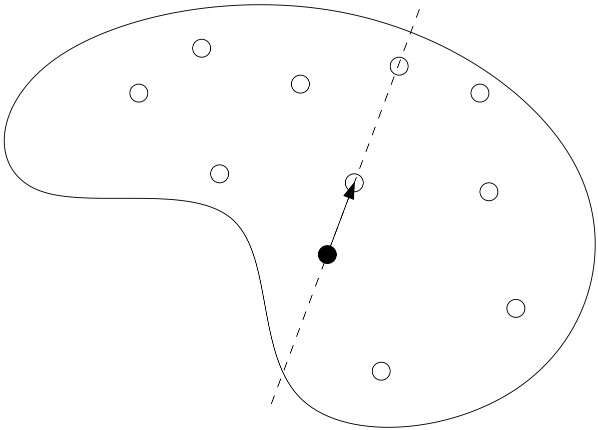
Skillings affine invariant ideas

Leapfrog



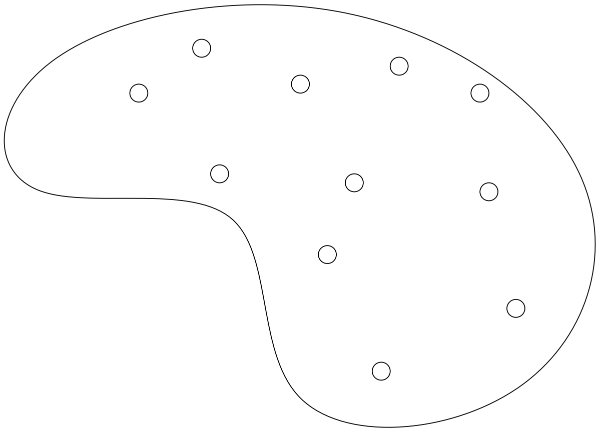
Skillings affine invariant ideas

Leapfrog



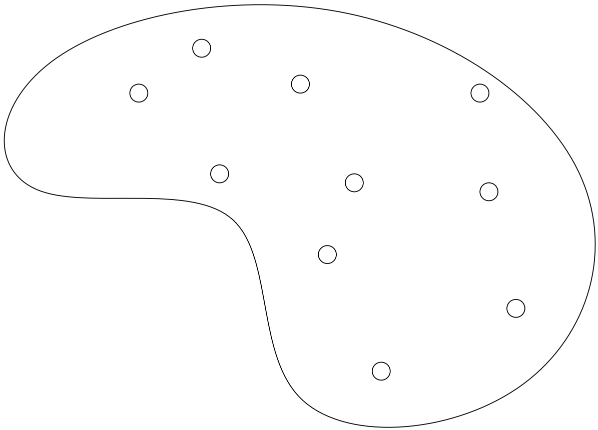
Skillings affine invariant ideas

Leapfrog



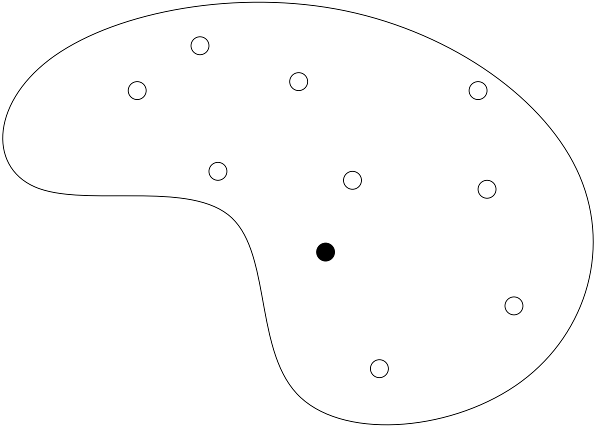
Skillings affine invariant ideas

Parallel walk



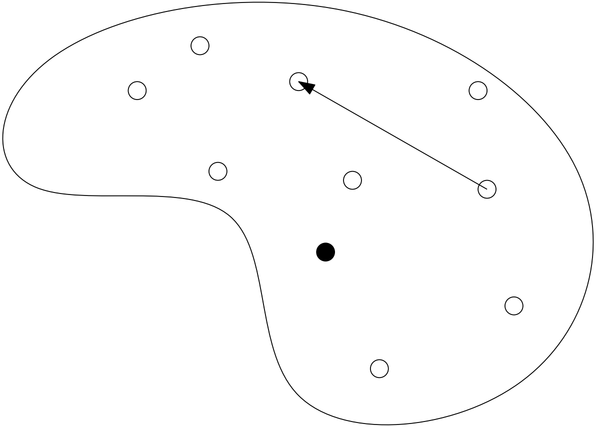
Skillings affine invariant ideas

Parallel walk



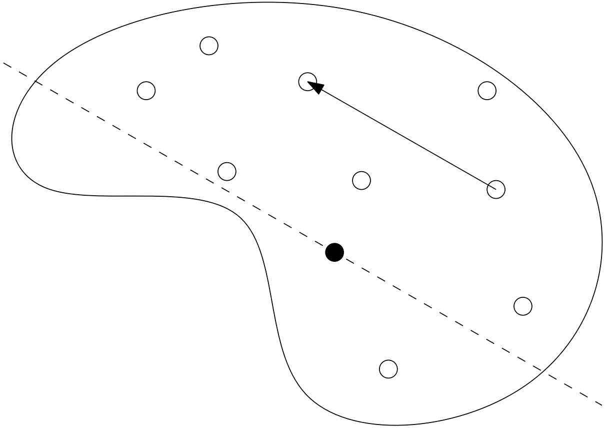
Skillings affine invariant ideas

Parallel walk



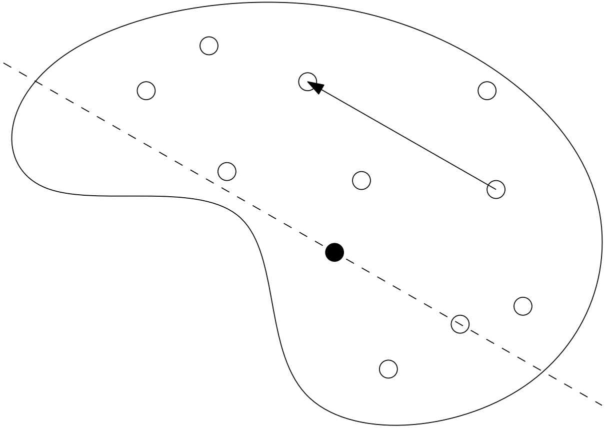
Skillings affine invariant ideas

Parallel walk



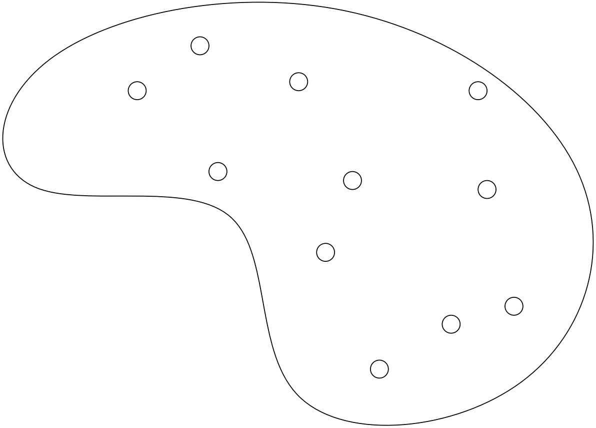
Skillings affine invariant ideas

Parallel walk



Skillings affine invariant ideas

Parallel walk



Affine invariance

Subspace collapse

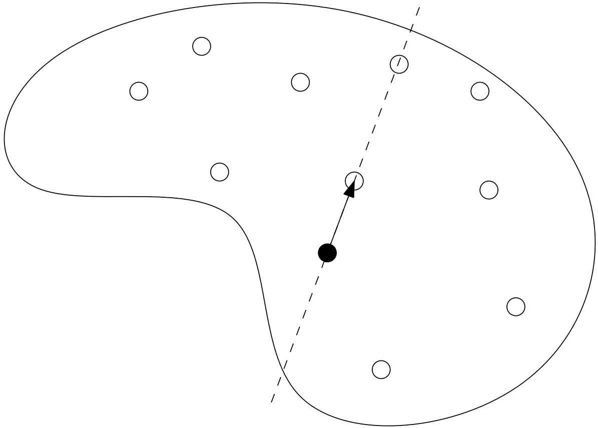
Affine invariance

Subspace collapse

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

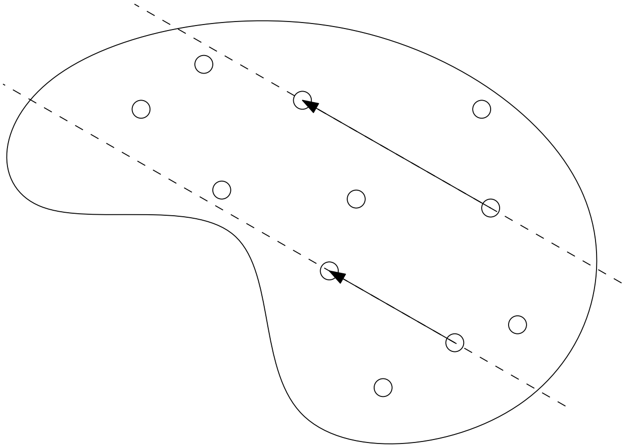
Subspace collapse

Leapfrog



Subspace collapse

Parallel walk



Subspace collapse

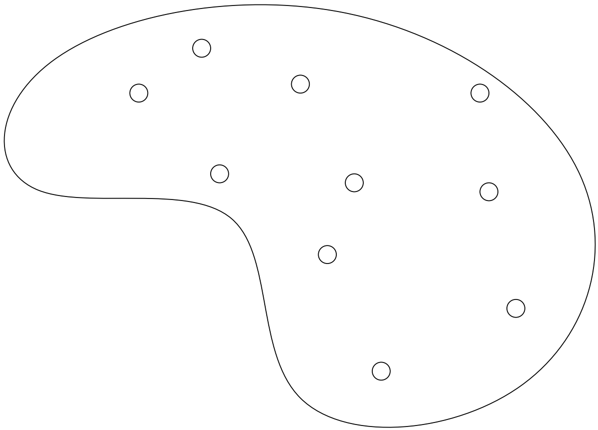
Solution

Subspace collapse
Solution

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

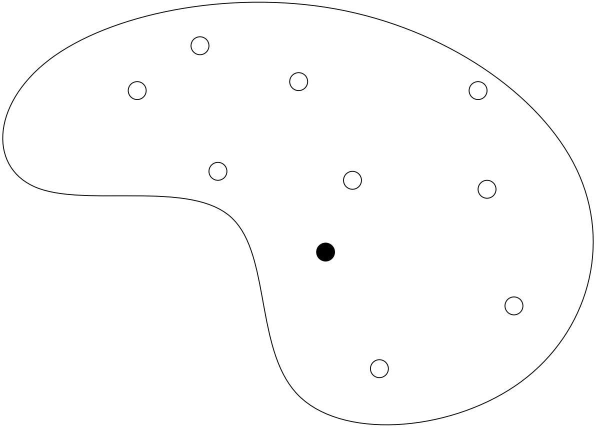
Skillings affine invariant ideas

Guided walk



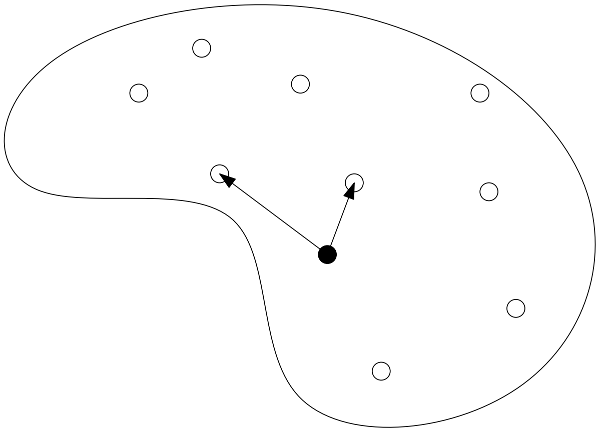
Skillings affine invariant ideas

Guided walk



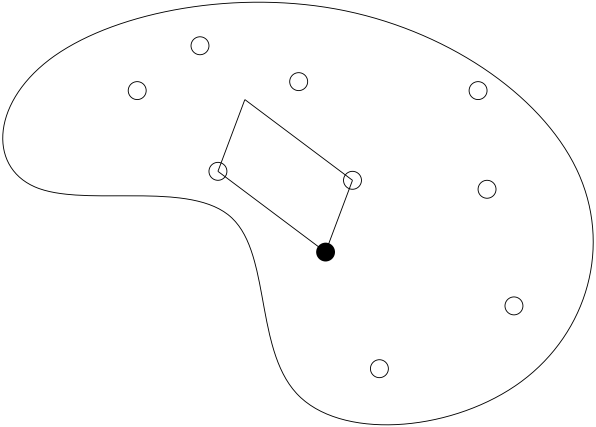
Skillings affine invariant ideas

Guided walk



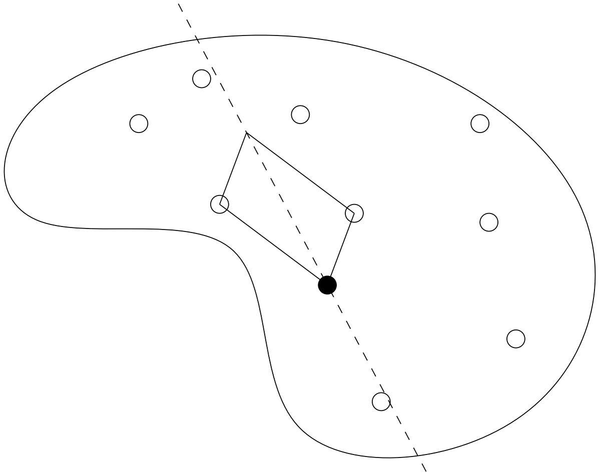
Skillings affine invariant ideas

Guided walk



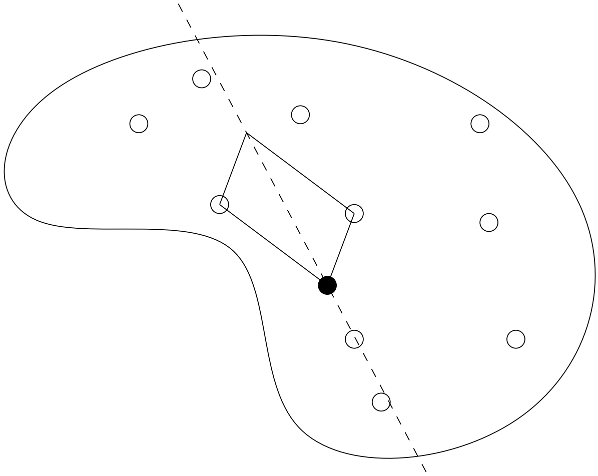
Skillings affine invariant ideas

Guided walk



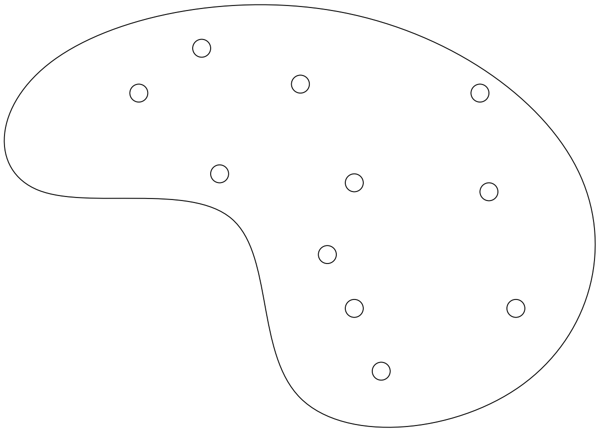
Skillings affine invariant ideas

Guided walk



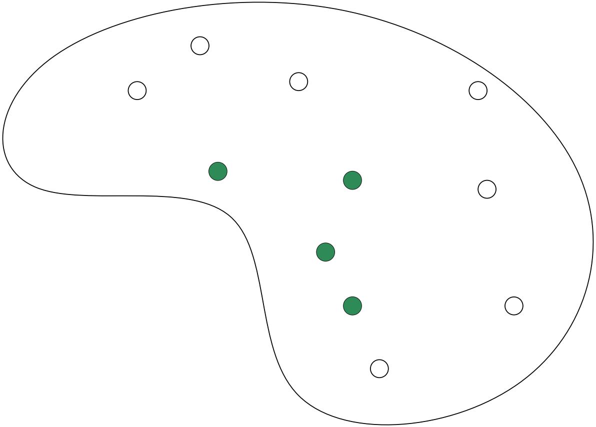
Skillings affine invariant ideas

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.