Nested Sampling

An efficient and robust Bayesian inference tool for cosmology and particle physics

Will Handley
wh260@cam.ac.uk

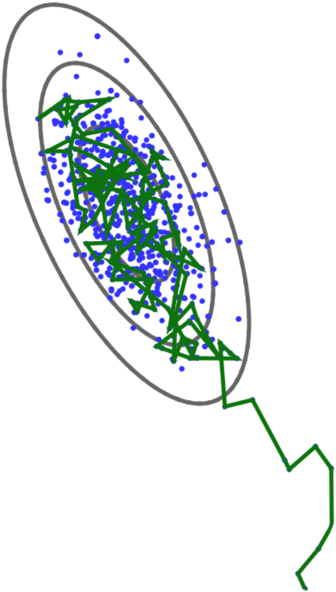
Astrophysics Group
Cavendish Laboratory
University of Cambridge

Nov 6, 2018

Outline

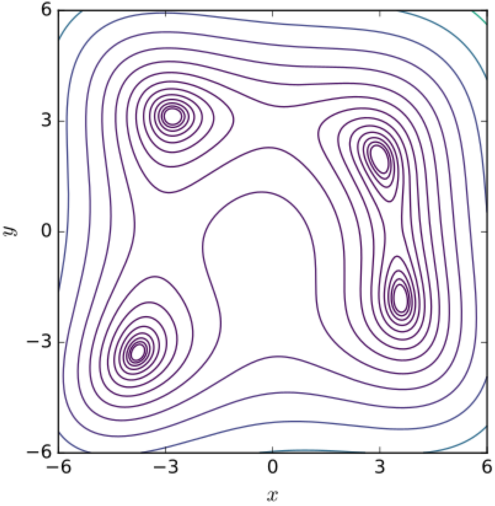
What do I mean by sampling?

- Sampling is the process of generating D-dimensional points $\theta = (\theta_1, \ldots, \theta_D)$ drawn from probability distribution $P(\theta)$.
- $P(\theta)$ is a-priori unknown, and may be expensive to evaluate.
- The name of the game is use as few calls to P as possible.
- Points need not be independent, and indeed normally only need $\sim \mathcal{O}(12)$ for most inference purposes.



Challenges in sampling techniques

- Multimodality
- Burn-in
- Convergence diagnosis
- Correlation/Degeneracy
- Parallelisation
- Phase-transitions
- High dimensions



Nested Sampling

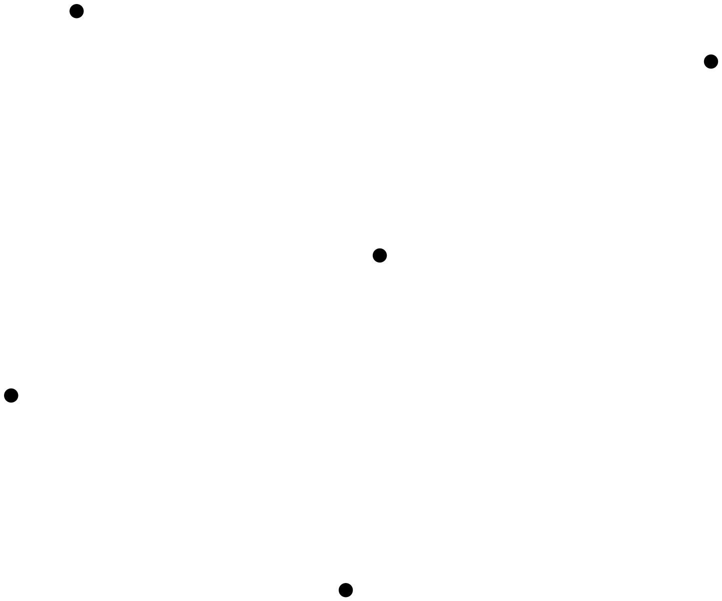
- Completely new approach to sampling:

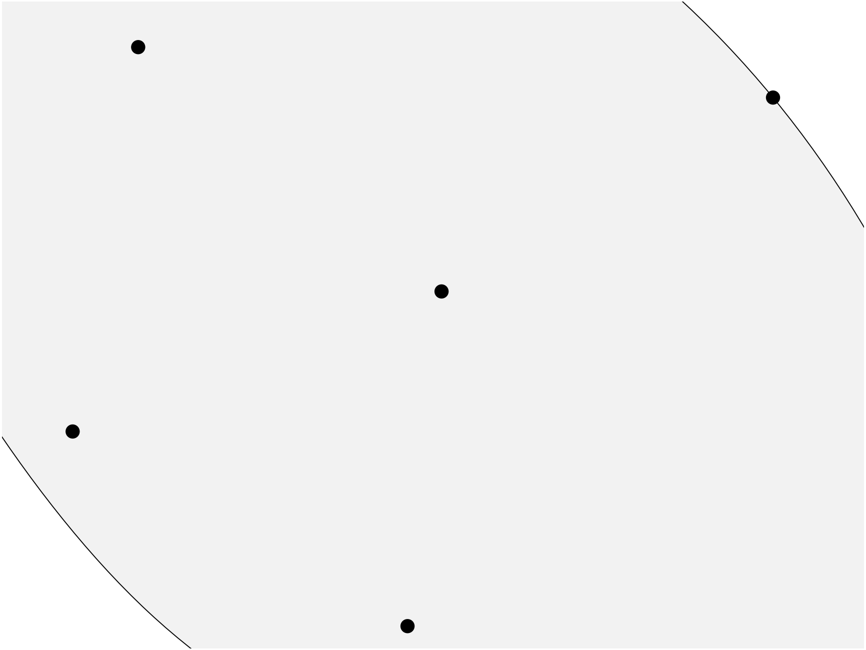
  Maintain a set S of n samples, which are sequentially updated:

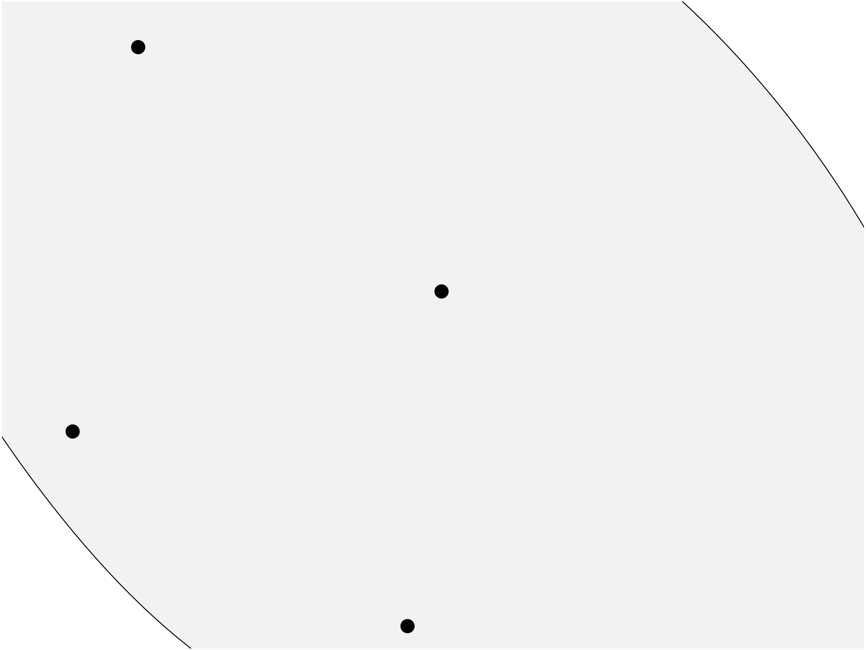
  S_0: Generate n samples uniformly over the space.

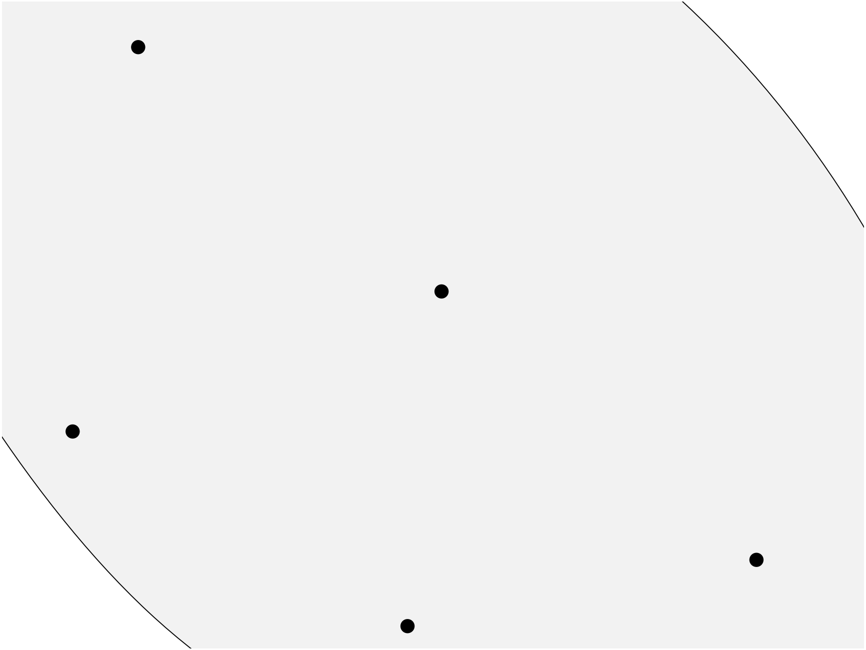
  S_{n+1}: Delete the lowest probability sample in S_n, and replace it with a new sample with higher probability

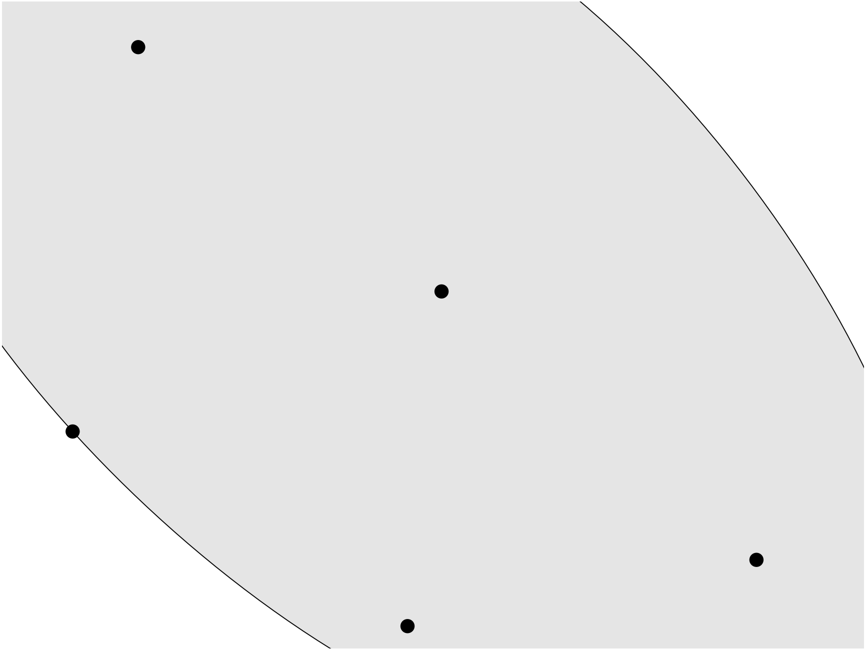
- This generates a *run* of discarded points.
- $n \sim \mathcal{O}(10s - 1000s)$
- Requires one to be able to uniformly within a region, subject to a *hard probability constraint*.
- John Skilling's original paper: euclid.ba/1340370944

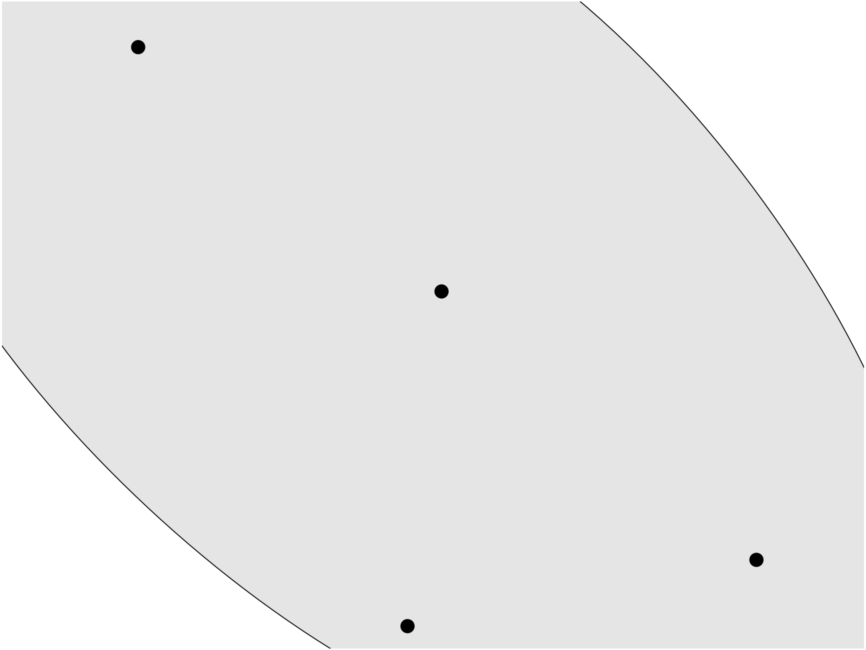


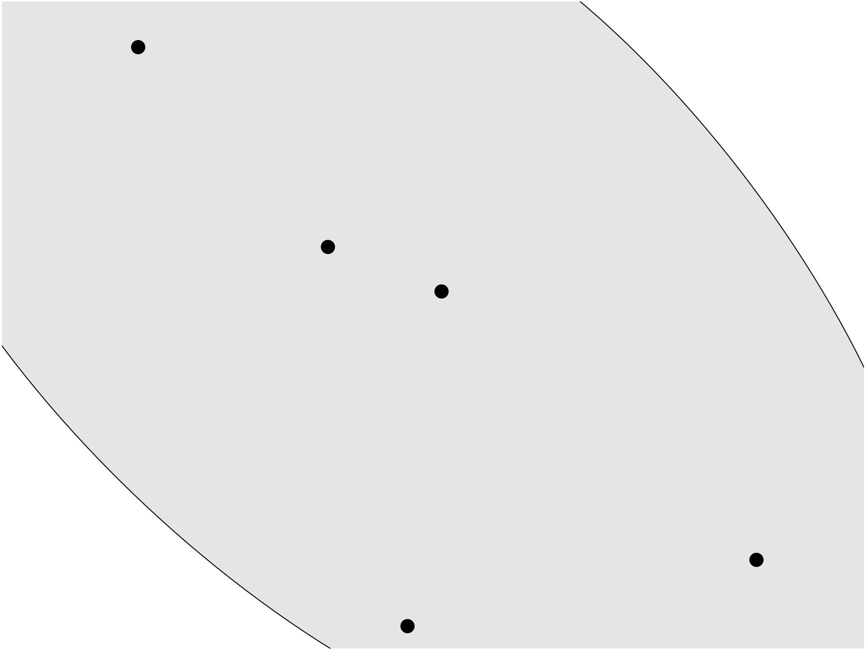


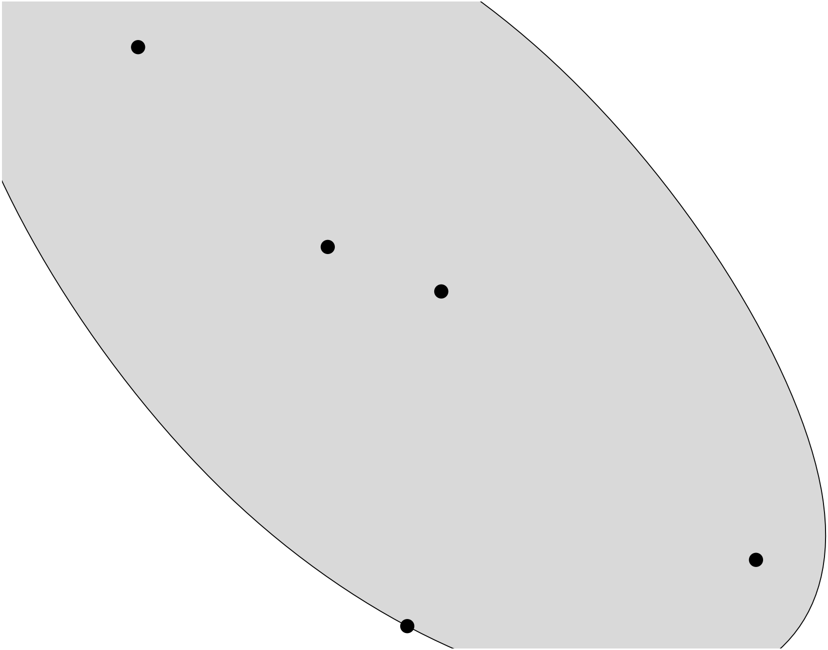


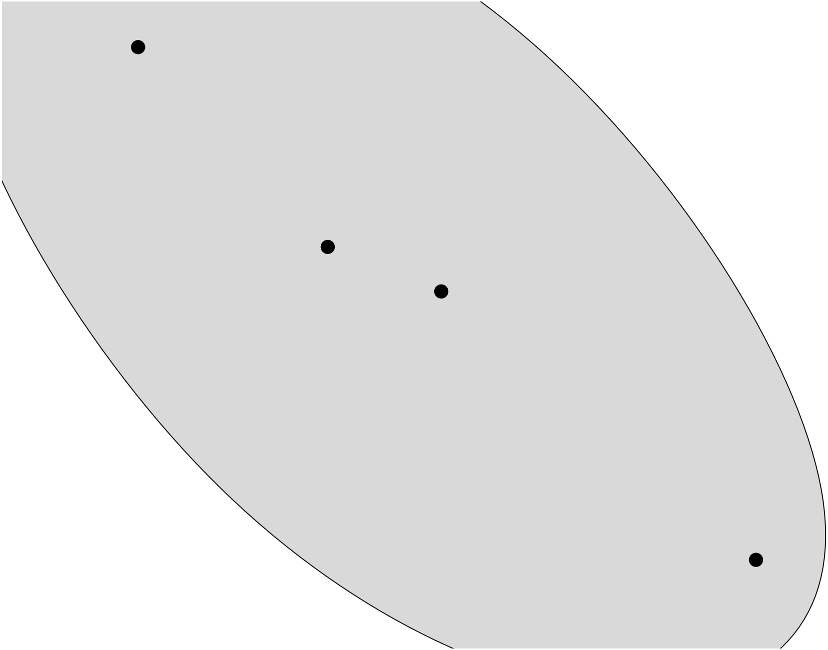


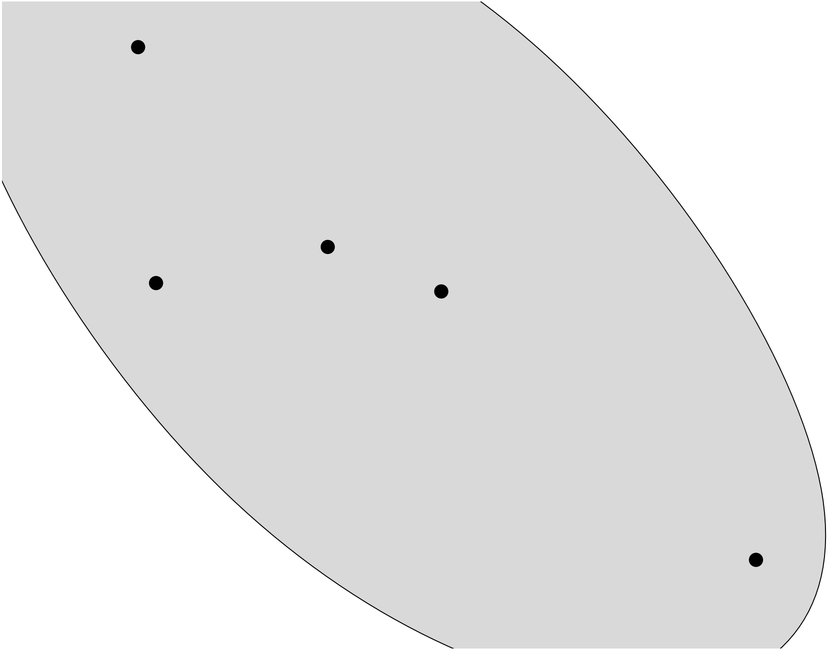


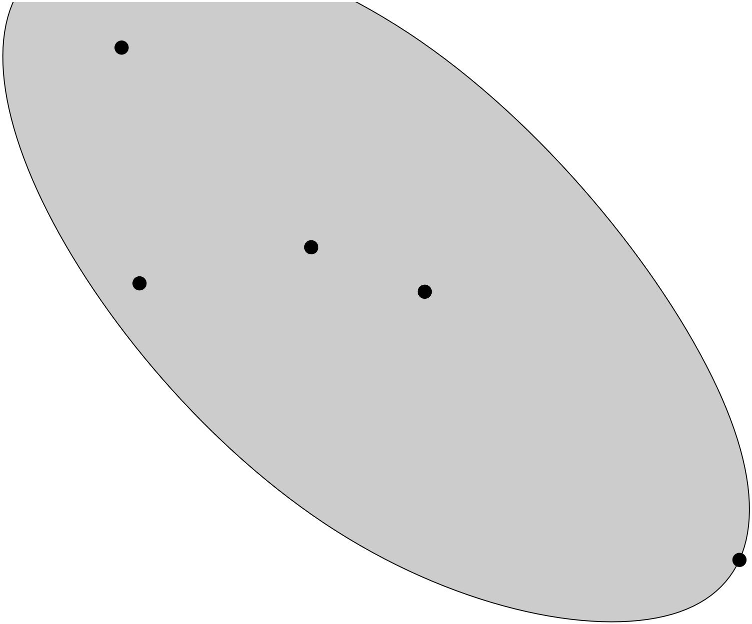


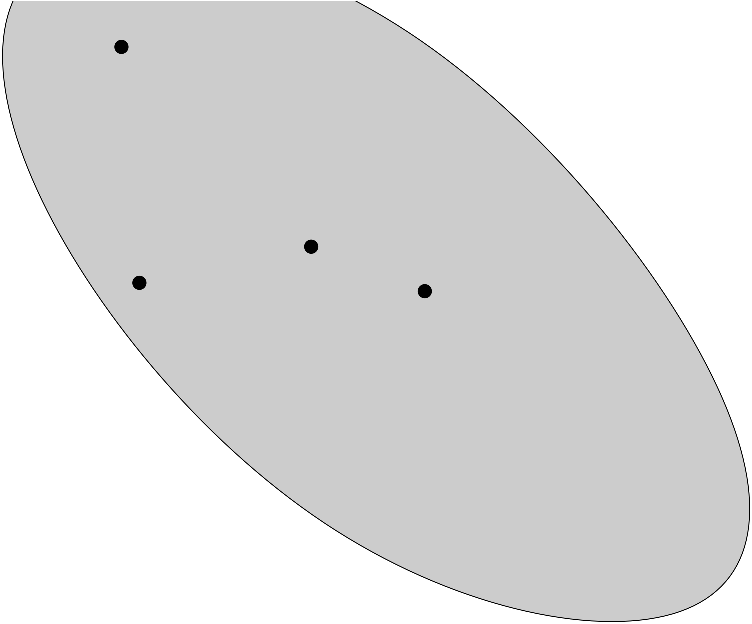


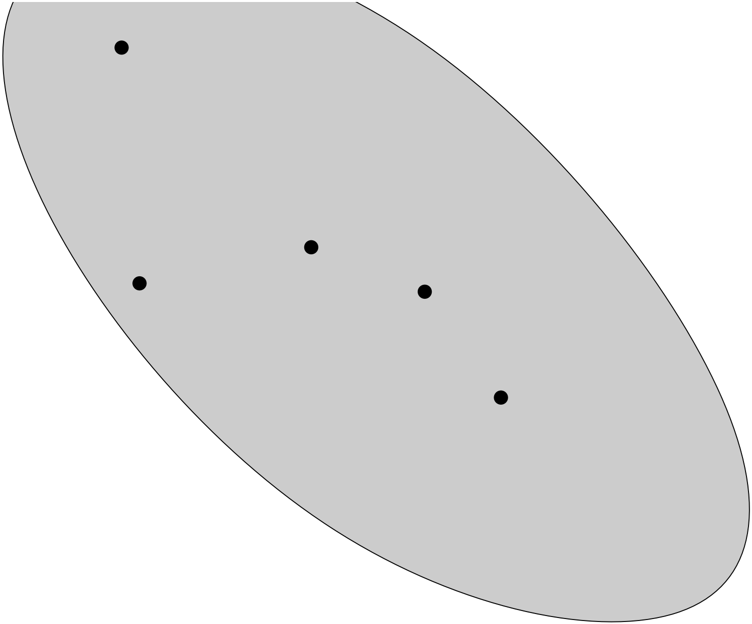


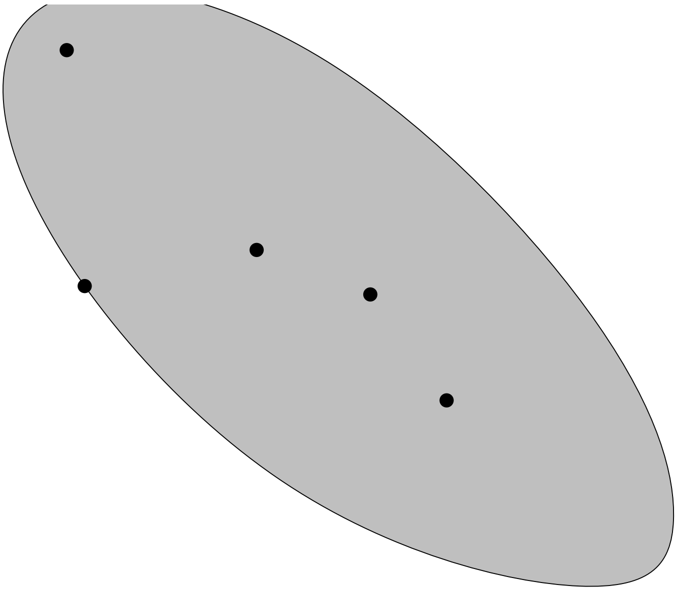


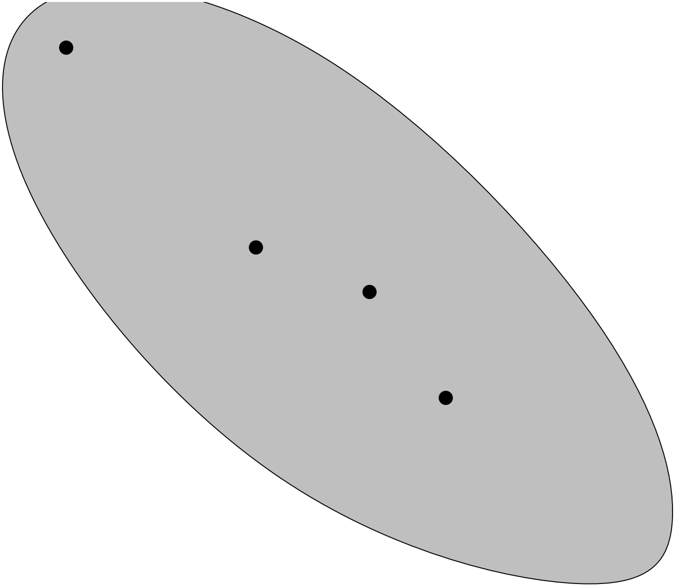


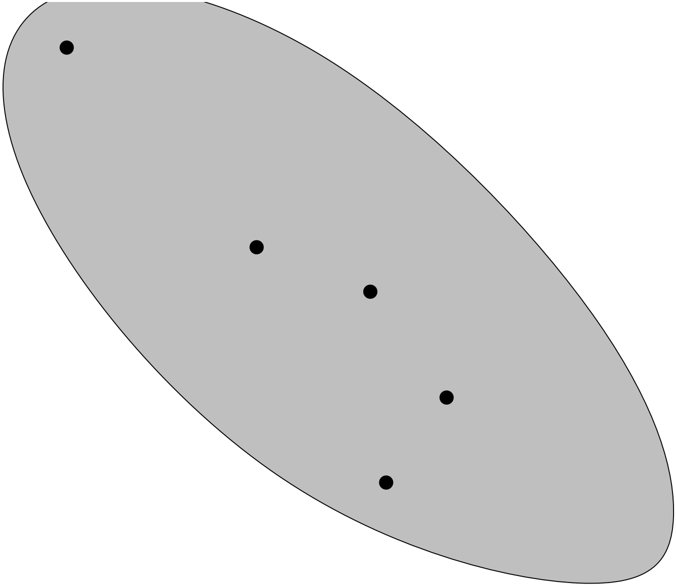


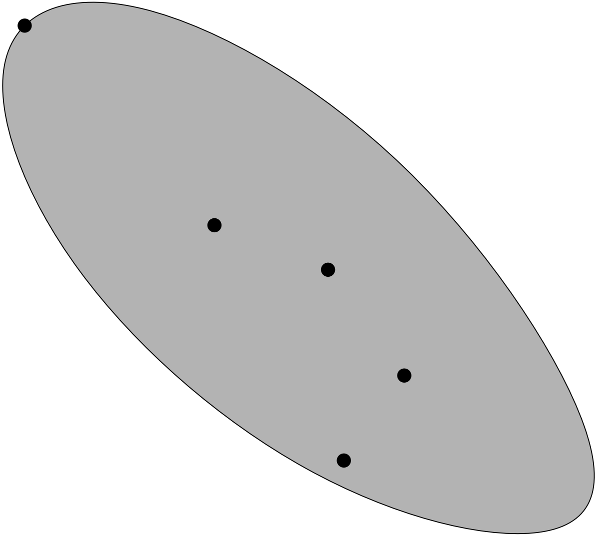


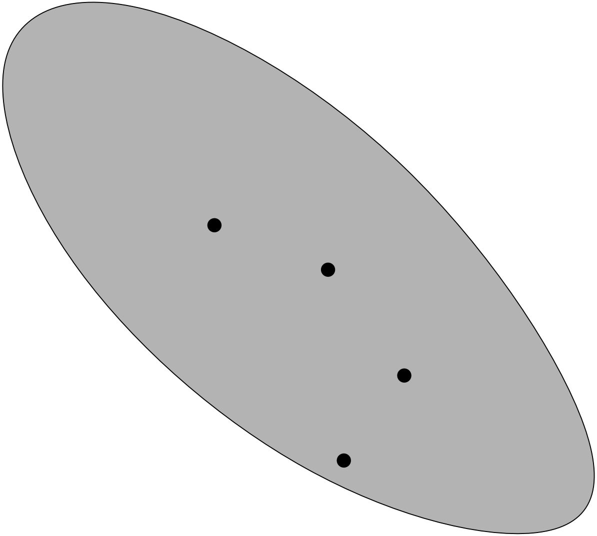


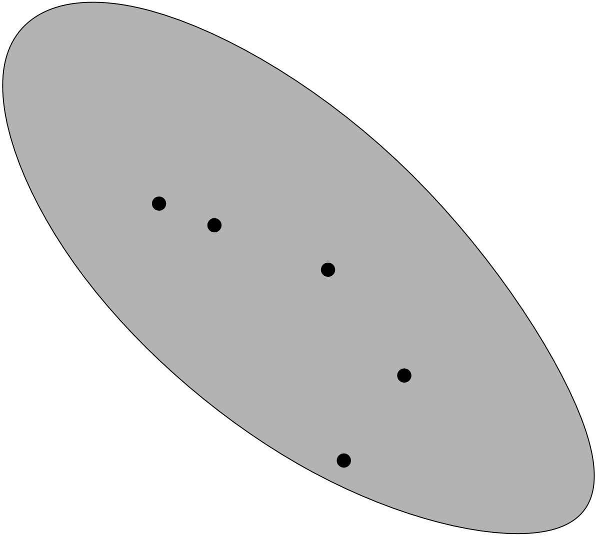


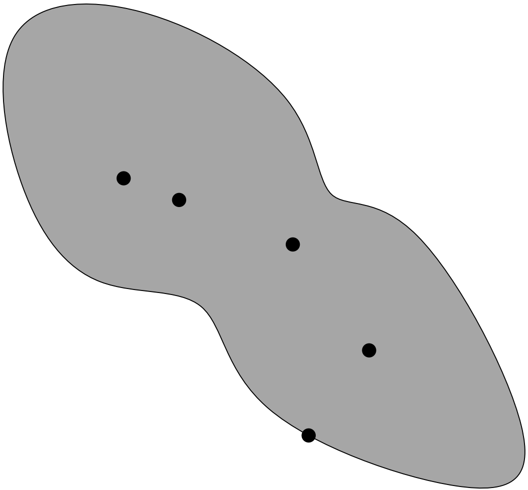


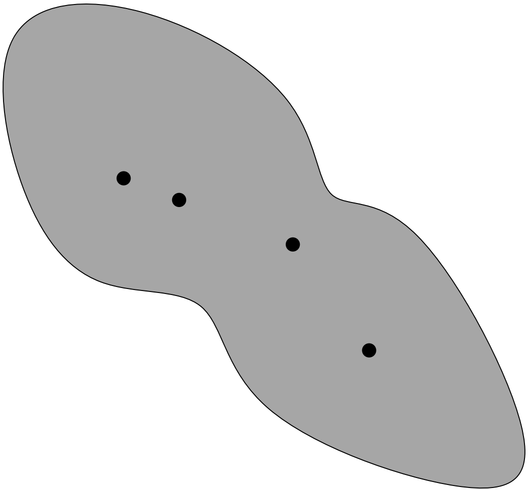


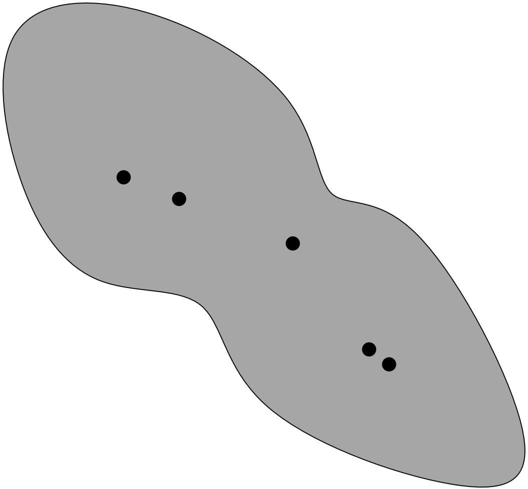


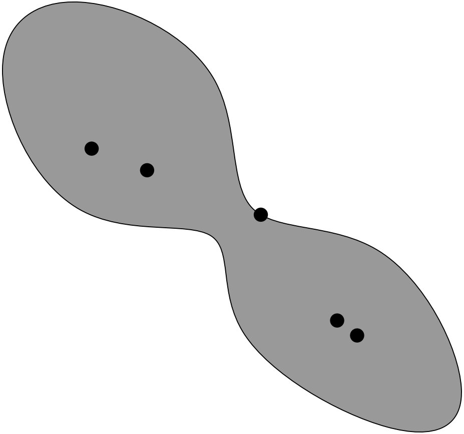


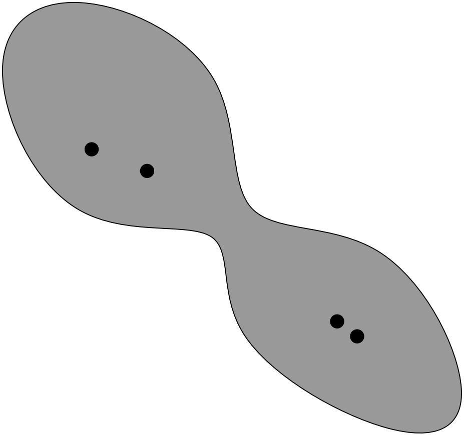


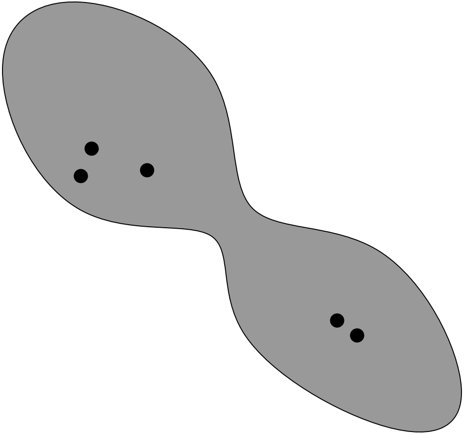


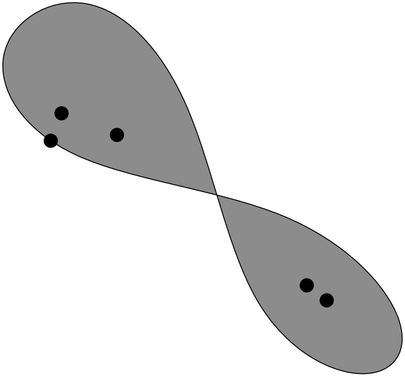


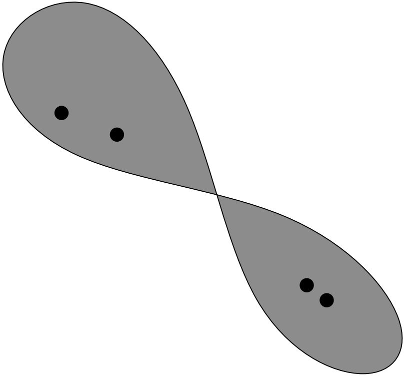


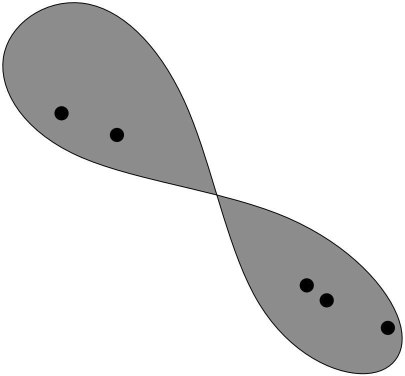


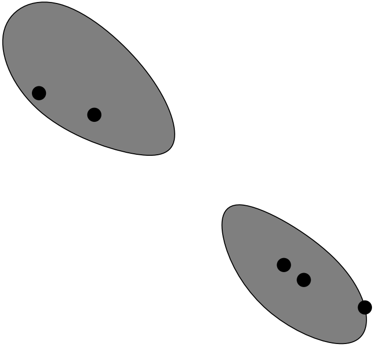


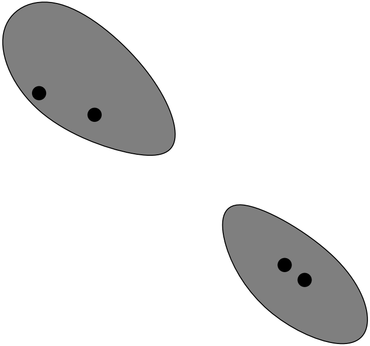


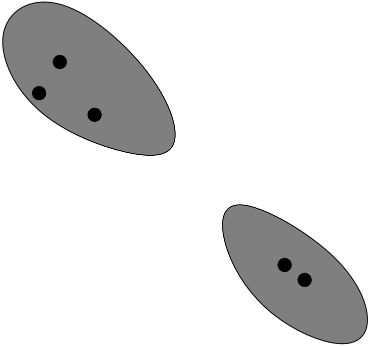


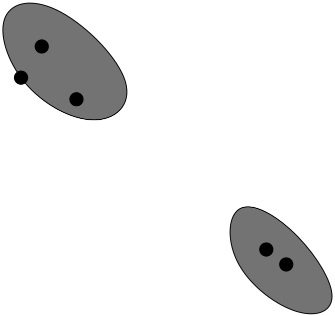


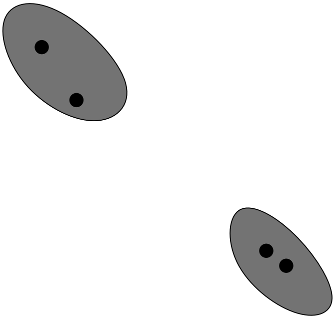


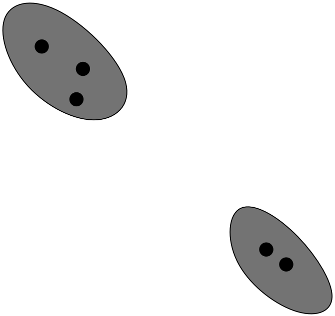


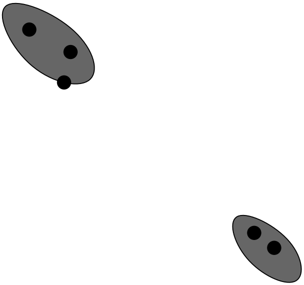


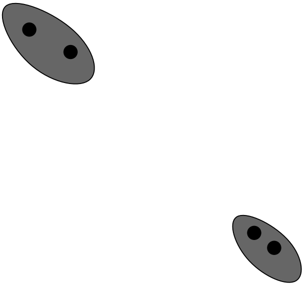


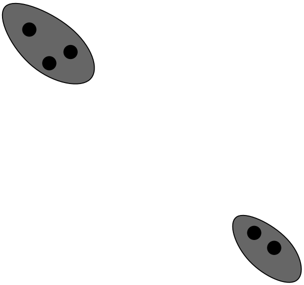


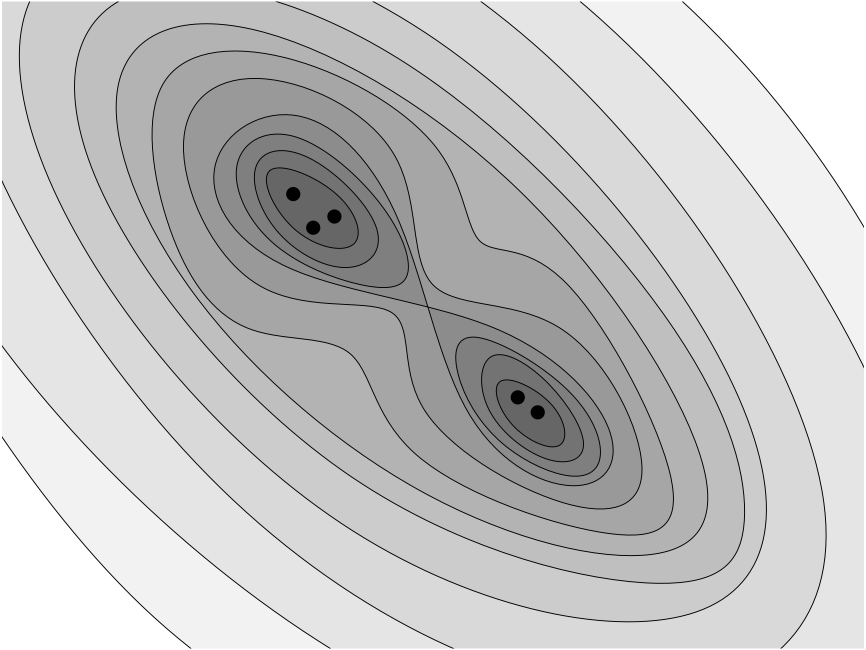








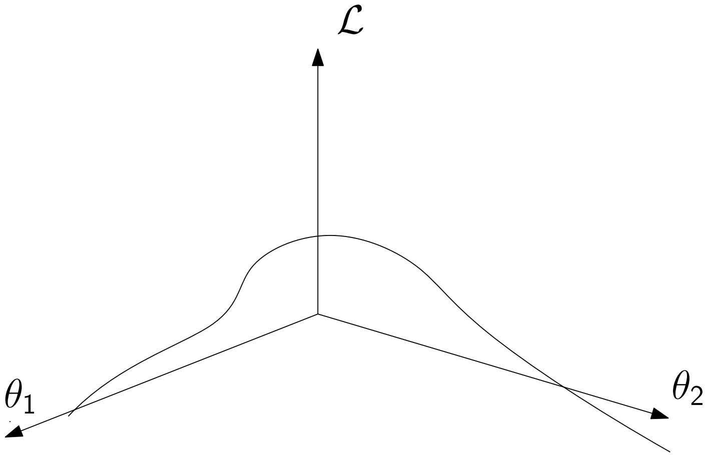




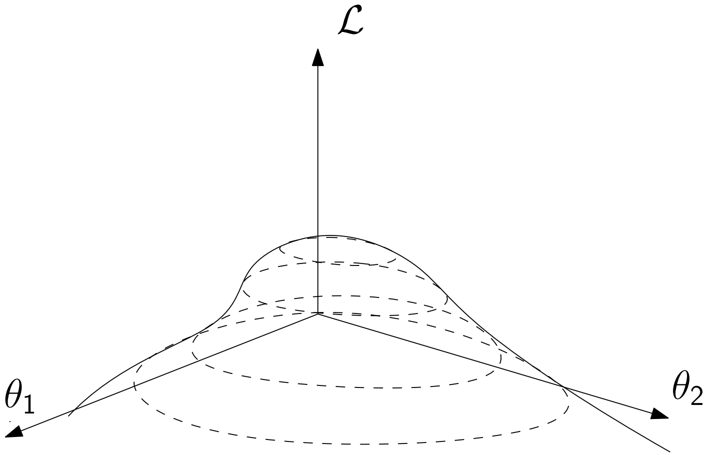
How is Nested Sampling used?

- Nested sampling generates a *run* of discarded points
- These points can be weighted in post-processing to give:
  - Posterior samples
  - Bayesian Evidence (marginal likelihoods)
  - Kullback Liebler divergence
  - Partition function
- This is possible because the nested sampling scheme is a probabilistic integrator, allowing one to estimate the *density of states*.

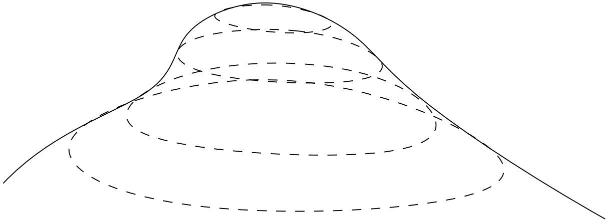
Probabilistic integration



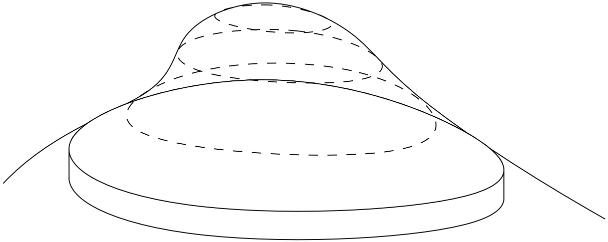
Probabilistic integration



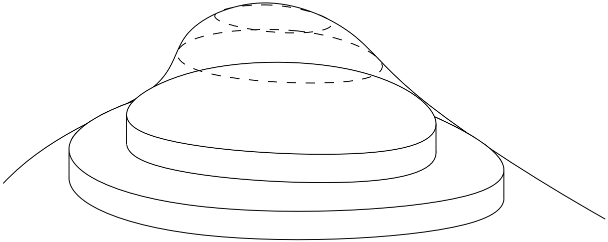
Probabilistic integration



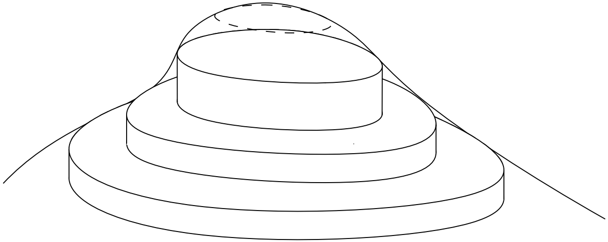
Probabilistic integration



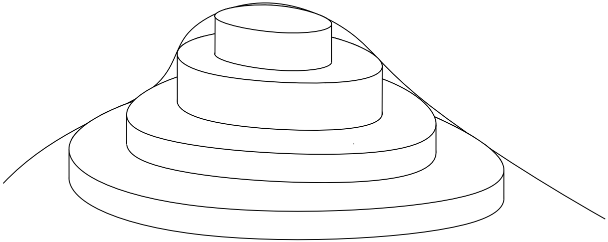
Probabilistic integration



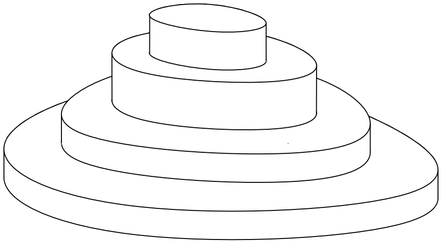
Probabilistic integration



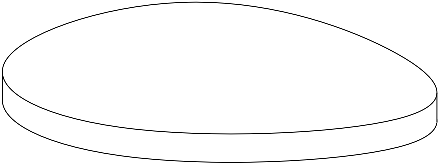
Probabilistic integration

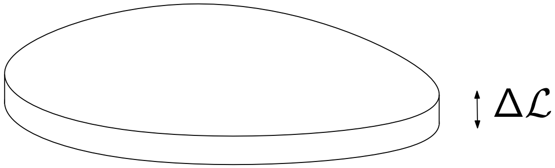


Probabilistic integration

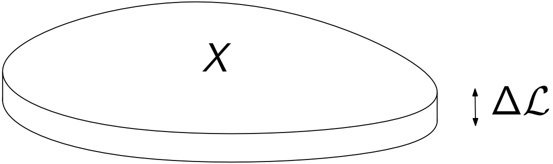


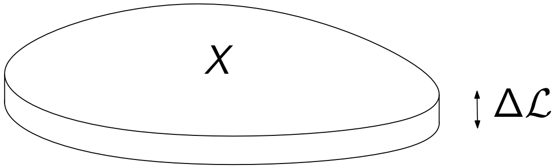
Probabilistic integration




$\Delta\mathcal{L}$

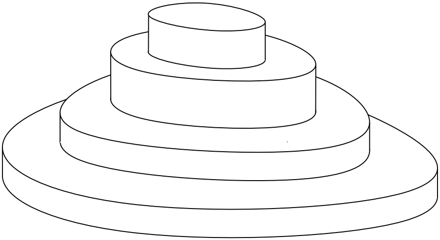
Probabilistic integration



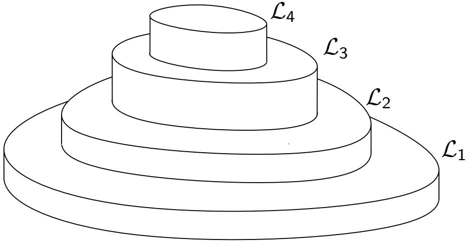


$$\text{Volume} = X\Delta\mathcal{L}$$

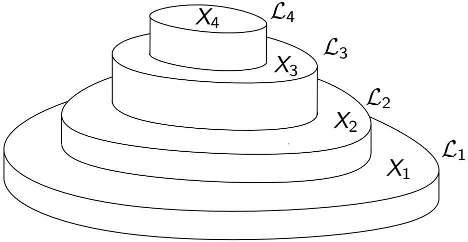
Probabilistic integration



Probabilistic integration

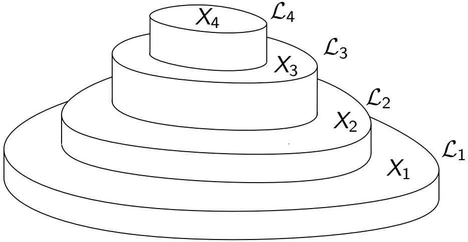


Probabilistic integration



Probabilistic integration

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



Estimating the density of states

- If number of live points $n = 100$, each uniformly sampled point sits in a shell $\approx 1\%$ of volume of outer most contour
- At each iteration, contour shrinks in volume by $\approx 1/n$.

$$\mathcal{Z} \approx \sum_i \Delta\mathcal{L}_i X_i, \qquad X_{i+1} \approx \frac{n}{n+1} X_i, \qquad X_0 = 1$$

- Nested sampling zooms in to the peak of the posterior *exponentially*.
- In fact, we perform precise inference on the volumes:

$$P(X_{i+1}|X_i) = n[X_{i+1}/X_i]^{n-1}$$

- Posterior weights are $\mathcal{P}_i = \mathcal{L}_i \times (X_i - X_{i-1})$

Key advantages of nested sampling

- The density of states (prior volume estimation) is the missing piece in inference, normally avoided/cancelled in traditional methods.
- Allows numerical computation of Bayesian Evidence & KL divergence.
- At each iteration, the set of live points enables self-tuning:
  - Clustering (handling multi-modality)
  - Correlation estimation (constructing proposal distributions)
- The sampling process is athermal, and invariant under monotonic tranformations of the sampled distribution $L(\theta) \to f(L(\theta))$:

$$E(\theta) = -\log L(\theta) \qquad P(\theta) = \frac{1}{Z(\beta)} e^{-\beta E(\theta)} \tag{1}$$

- By appropriate re-weighting, we can post-process the posterior samples to be at any temperature.

How do we perform nested sampling?

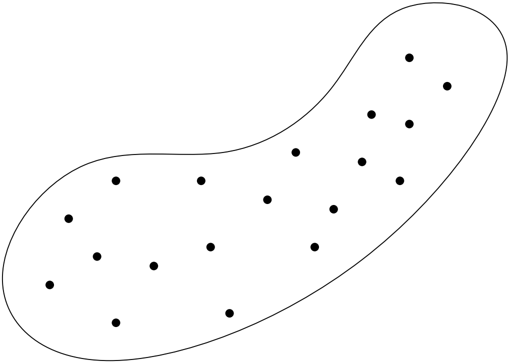
- Requires one to be able to uniformly within a region, subject to a *hard probability constraint*.
- Two main codes that implement traditional NS:

MultiNest	arXiv:0809.3437
PolyChord	arXiv:1506.00171

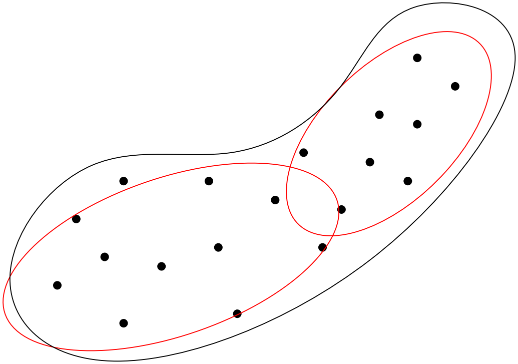
- Alternative frameworks:

Diffusive NS	arXiv:0912.2380
SE NS	arXiv:1402.6306
Dynamic NS	arXiv:1704.03459

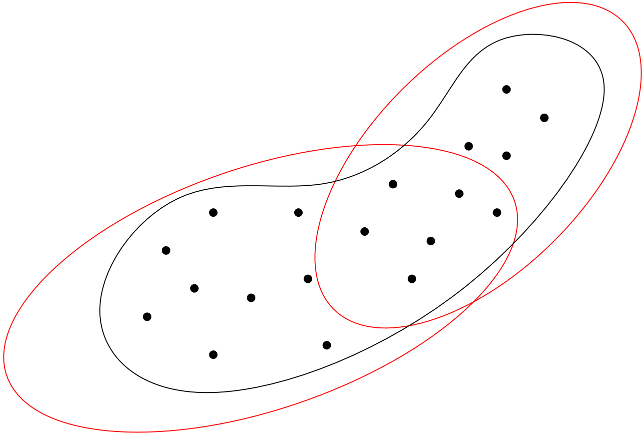
MultiNest



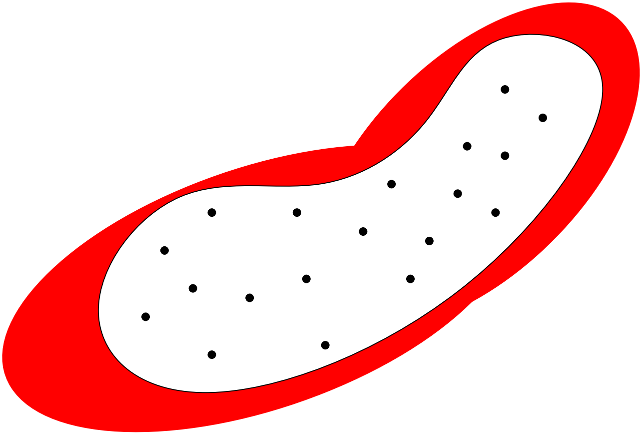
MultiNest



MultiNest



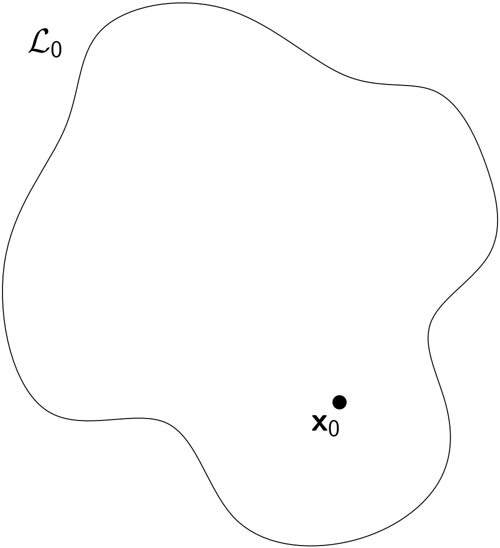
MultiNest

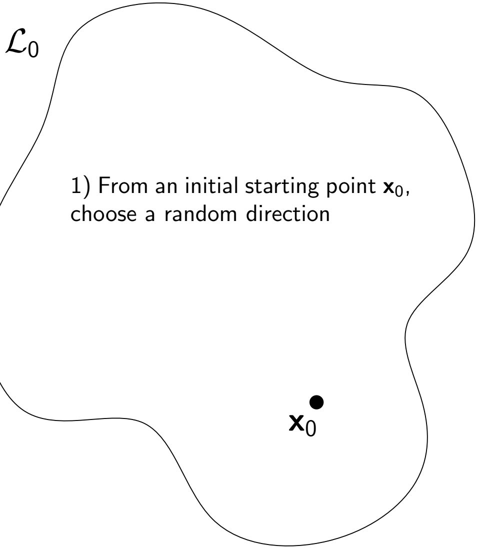


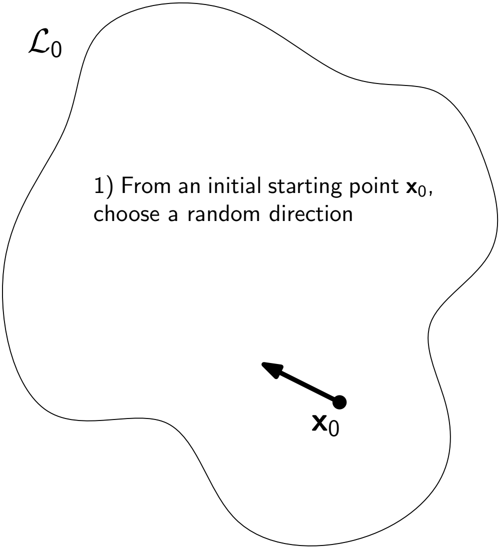
MultiNest

- arXiv:0809.3437
- Uses a set of overlapping ellipsoids to approximate the shape of the contours
- Rejection samples from this representation.
- Maximally efficient in low dimensions…
- …exponentially bad in high dimensions.
- Transition is distribution-dependent (as low as 5 or as high as 60).
- John Skilling originally anticipated/advocated using MCMC-style chain-based approach.

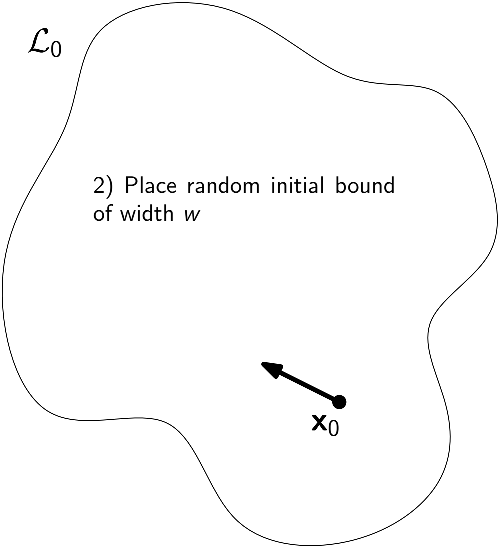
PolyChord



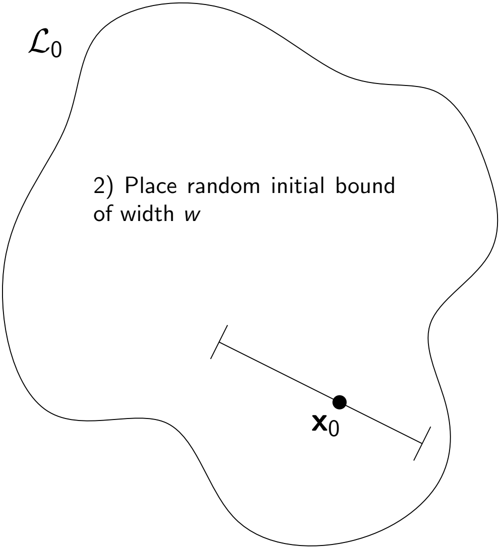

$\mathcal{L}_0$
1) From an initial starting point $\mathbf{x}_0$,
choose a random direction
$\mathbf{x}_0$


$\mathcal{L}_0$
1) From an initial starting point $\mathbf{x}_0$,
choose a random direction
$\mathbf{x}_0$

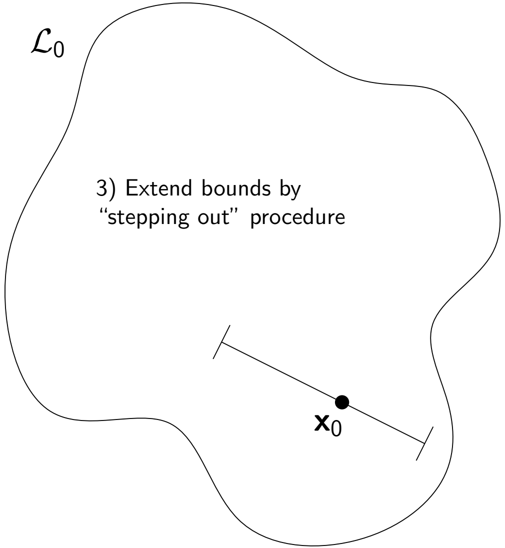
PolyChord

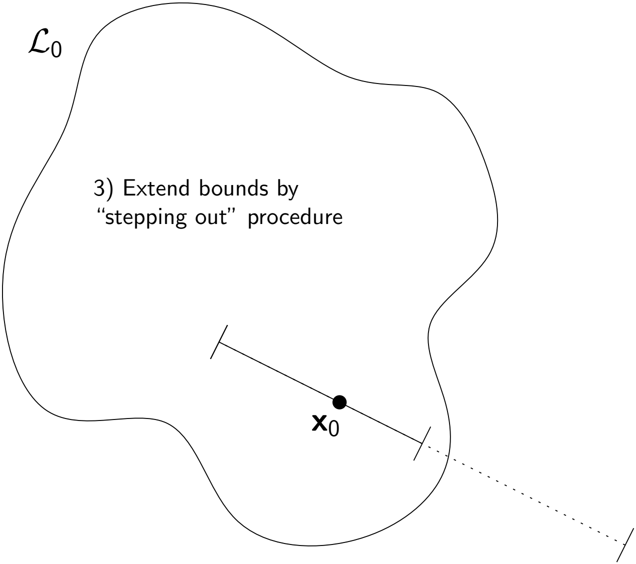


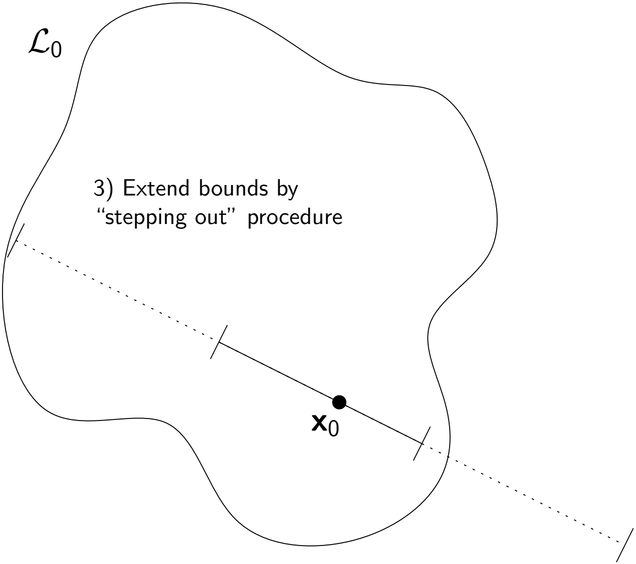
PolyChord



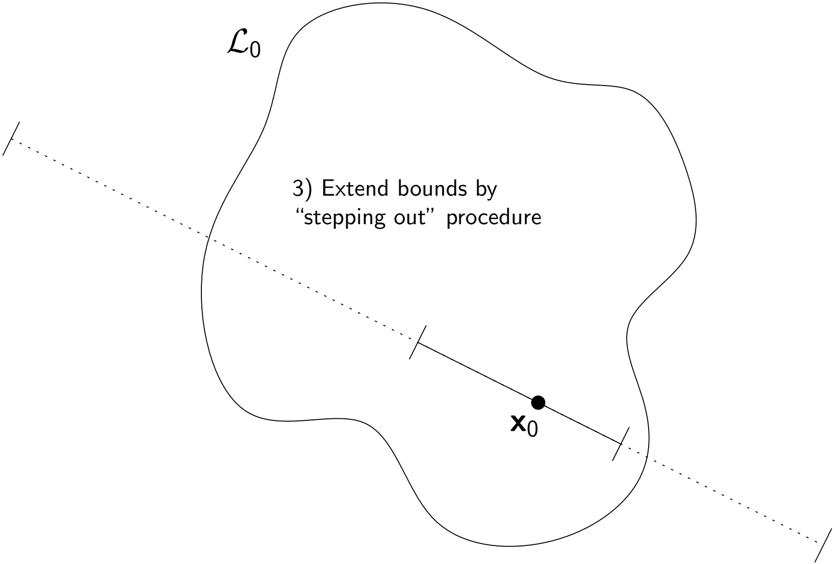
PolyChord


$\mathcal{L}_0$
3) Extend bounds by
"stepping out" procedure
$\mathbf{x}_0$


$\mathcal{L}_0$
3) Extend bounds by
"stepping out" procedure
$\mathbf{x}_0$

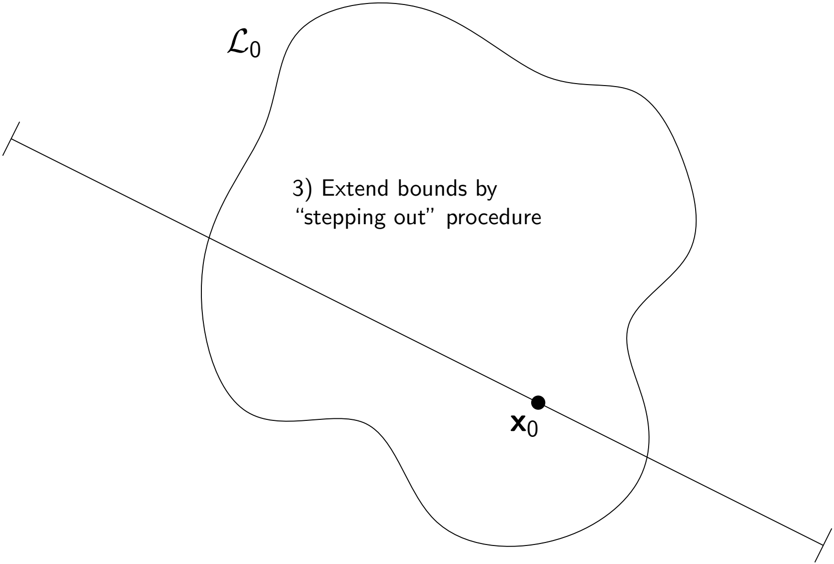

$\mathcal{L}_0$
3) Extend bounds by
"stepping out" procedure
$\mathbf{x}_0$

PolyChord

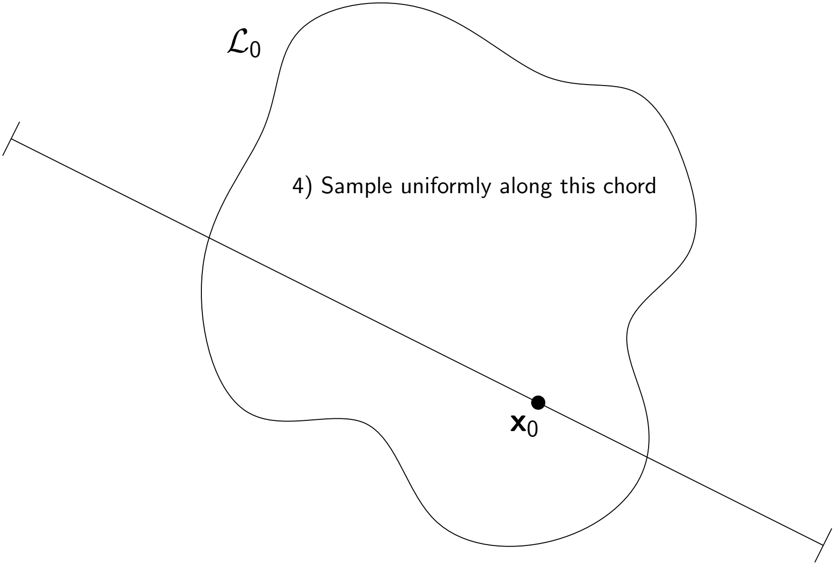


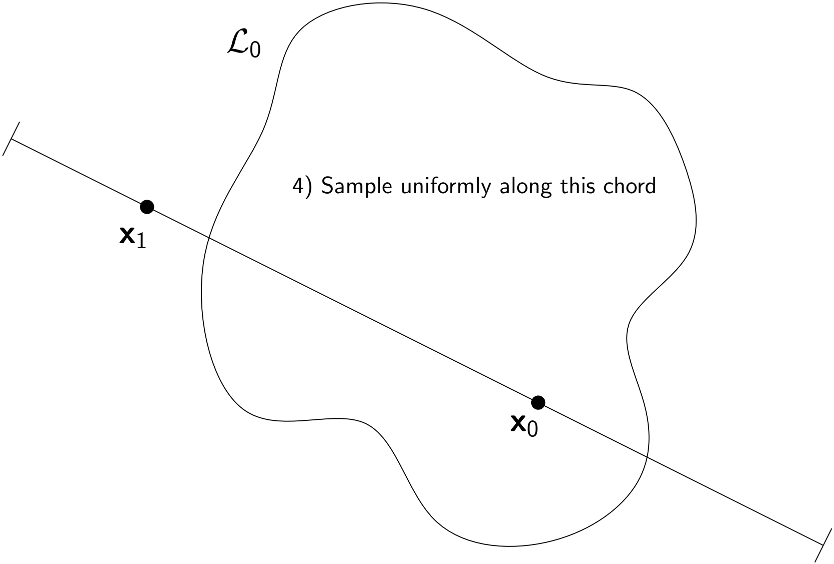
$\mathcal{L}_0$
3) Extend bounds by
"stepping out" procedure
$\mathbf{x}_0$

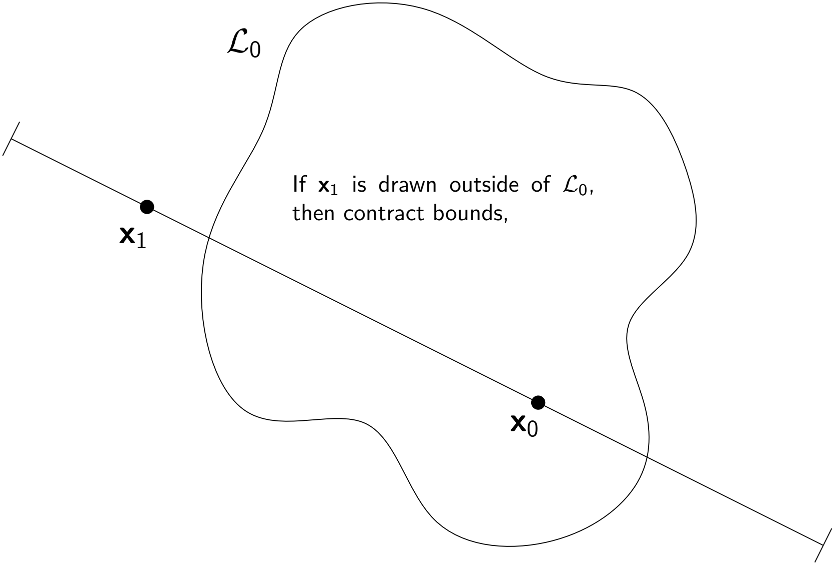
PolyChord


$\mathcal{L}_0$
3) Extend bounds by
"stepping out" procedure
$\mathbf{x}_0$

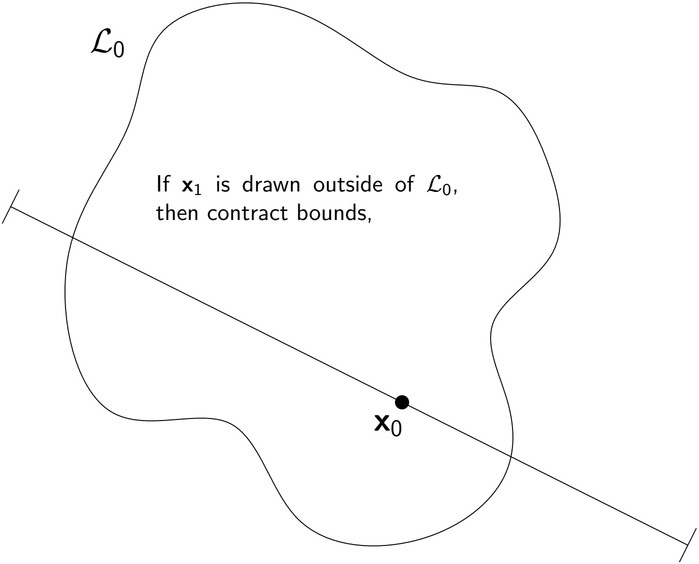
PolyChord



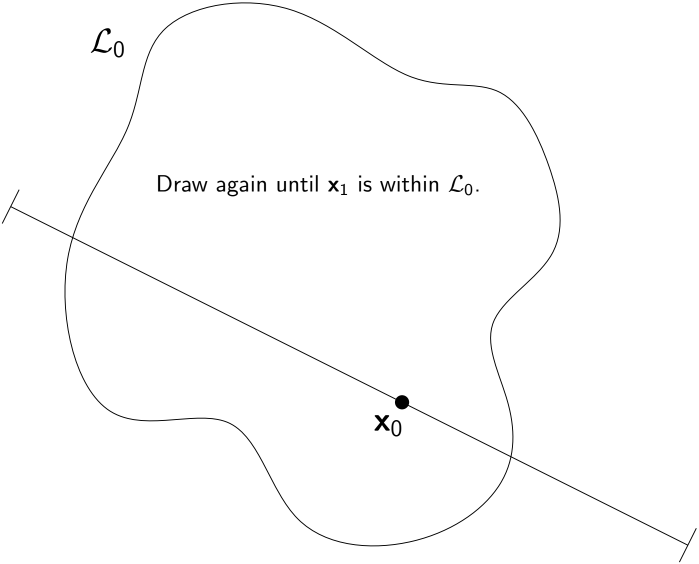

$\mathcal{L}_0$
4) Sample uniformly along this chord
$\mathbf{x}_1$
$\mathbf{x}_0$


$\mathcal{L}_0$
If $\mathbf{x}_1$ is drawn outside of $\mathcal{L}_0$,
then contract bounds,
$\mathbf{x}_1$
$\mathbf{x}_0$

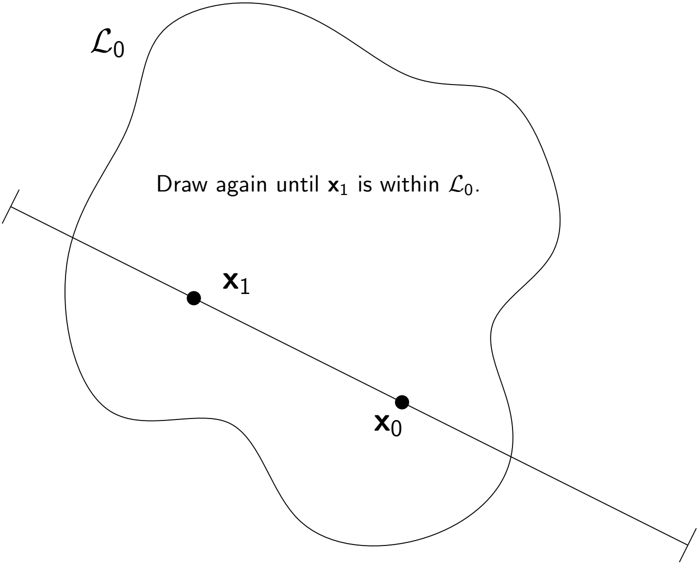
PolyChord



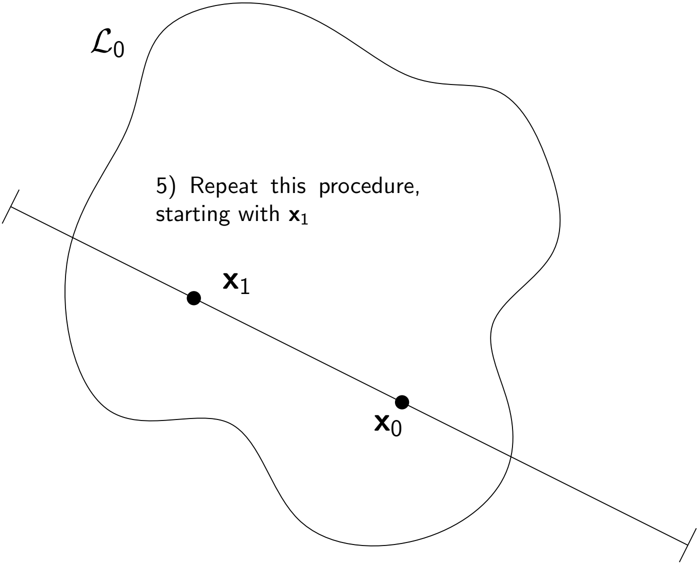
If $\mathbf{x}_1$ is drawn outside of $\mathcal{L}_0$, then contract bounds,


$\mathcal{L}_0$
Draw again until $\mathbf{x}_1$ is within $\mathcal{L}_0$.
$\mathbf{x}_0$

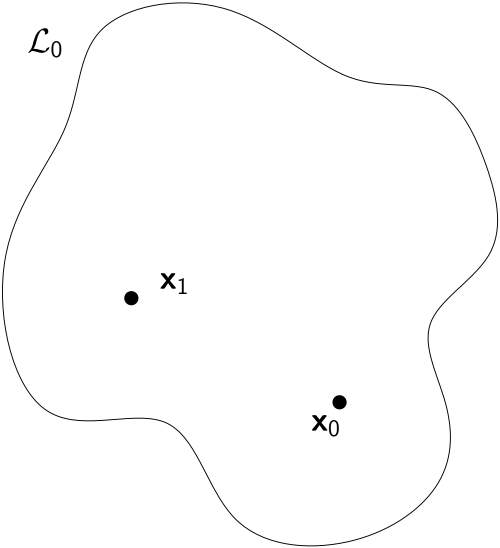
PolyChord

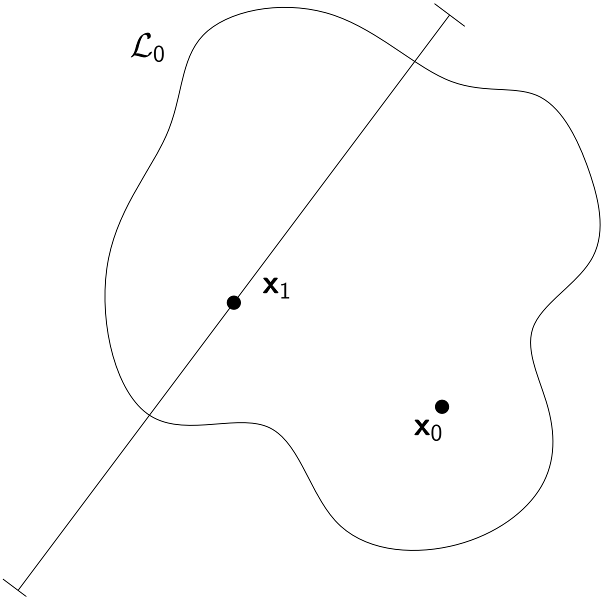


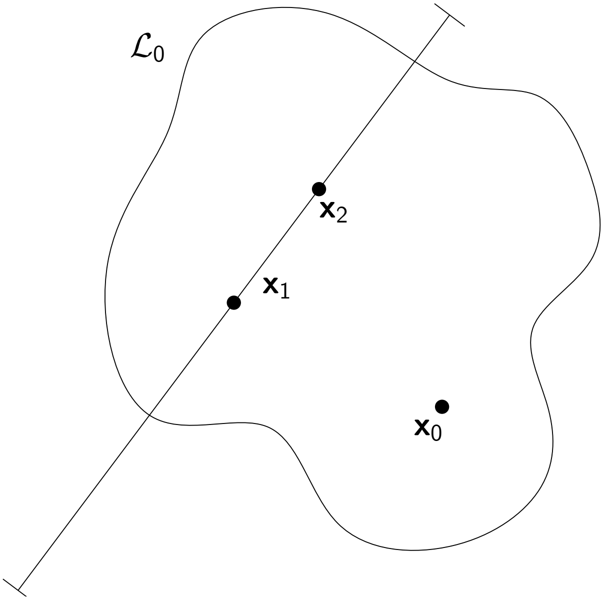
$\mathcal{L}_0$
Draw again until $\mathbf{x}_1$ is within $\mathcal{L}_0$.
$\mathbf{x}_1$
$\mathbf{x}_0$

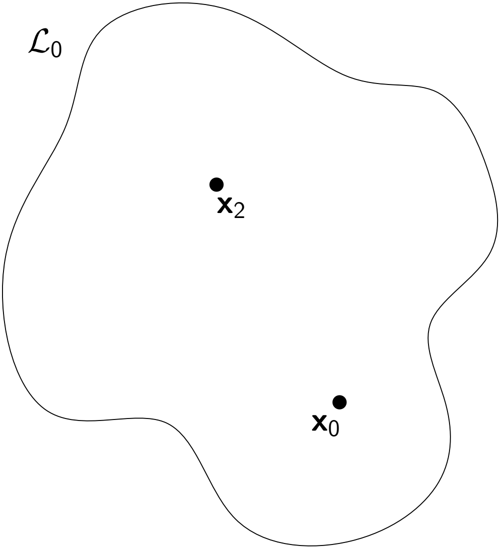

$\mathcal{L}_0$
5) Repeat this procedure, starting with $\mathbf{x}_1$
$\mathbf{x}_1$
$\mathbf{x}_0$

PolyChord

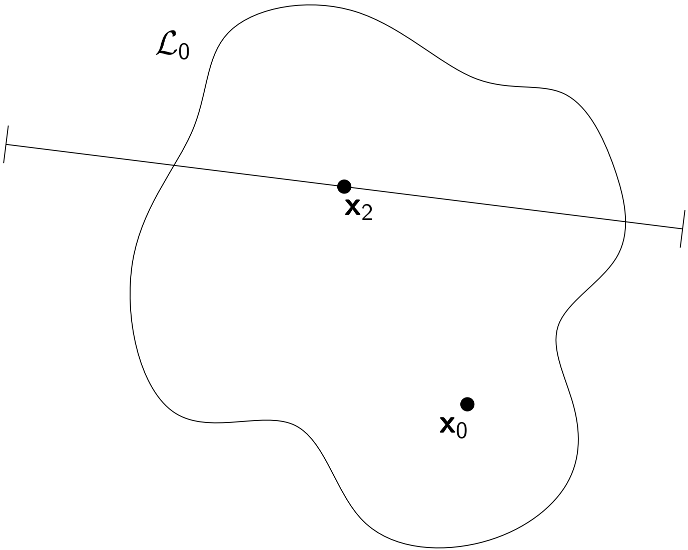


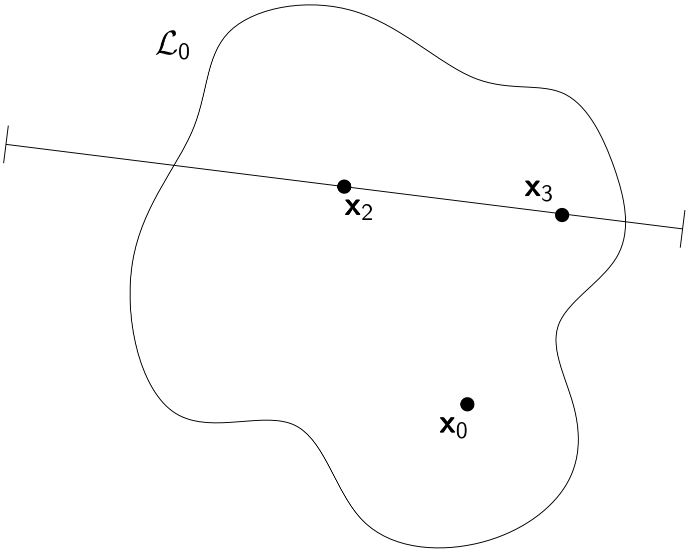

$\mathcal{L}_0$
$\mathbf{x}_1$
$\mathbf{x}_0$


$\mathcal{L}_0$
$\mathbf{x}_2$
$\mathbf{x}_1$
$\mathbf{x}_0$

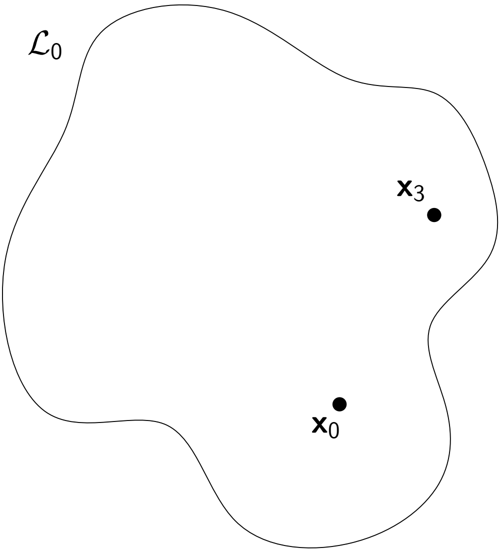

$\mathcal{L}_0$
$\mathbf{x}_2$
$\mathbf{x}_0$

PolyChord

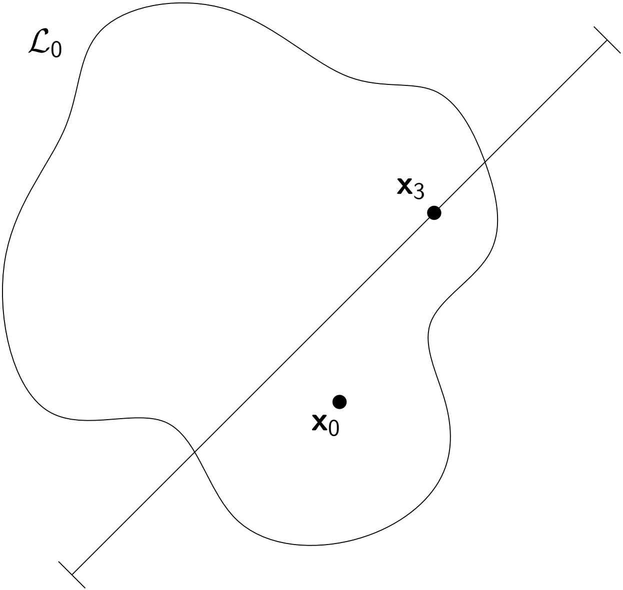


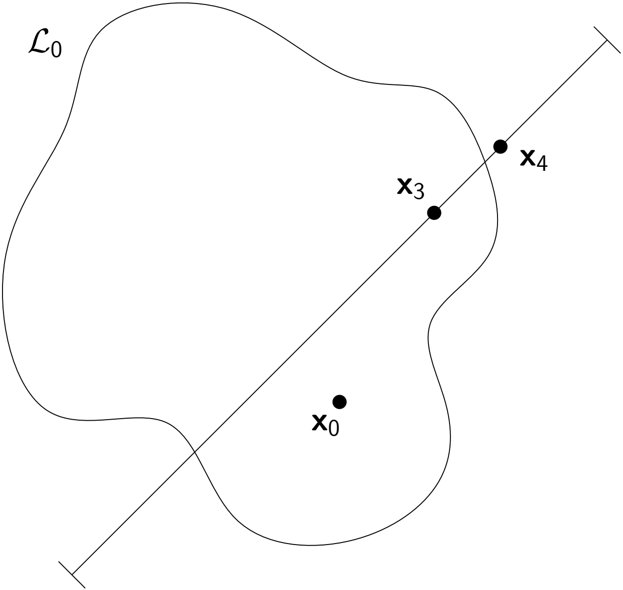

$\mathcal{L}_0$
$\mathbf{x}_2$
$\mathbf{x}_3$
$\mathbf{x}_0$

PolyChord

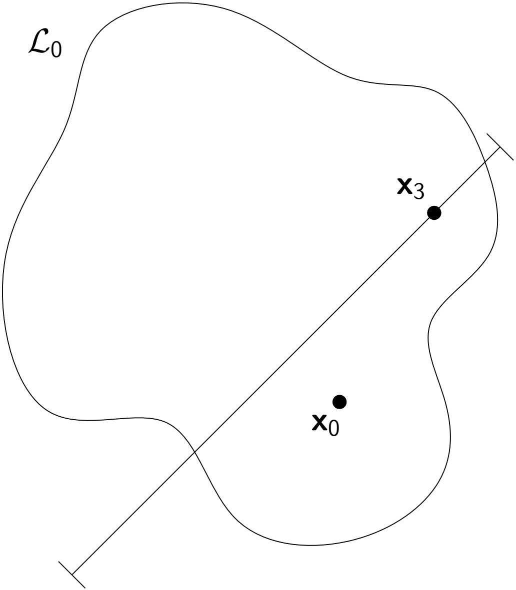


PolyChord

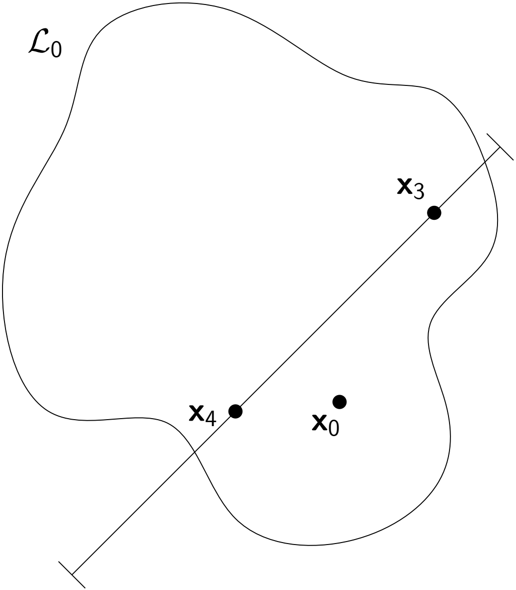


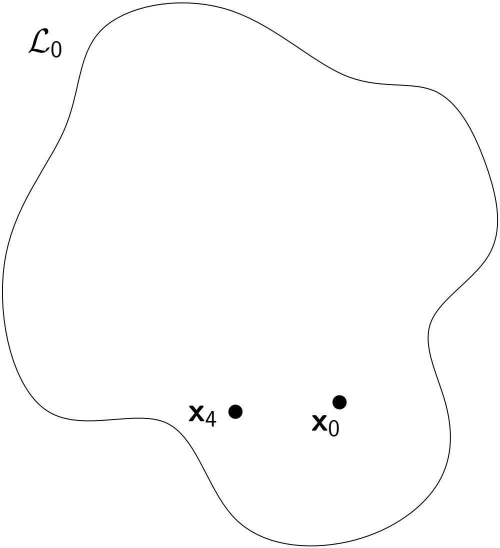

$\mathcal{L}_0$
$\mathbf{x}_4$
$\mathbf{x}_3$
$\mathbf{x}_0$

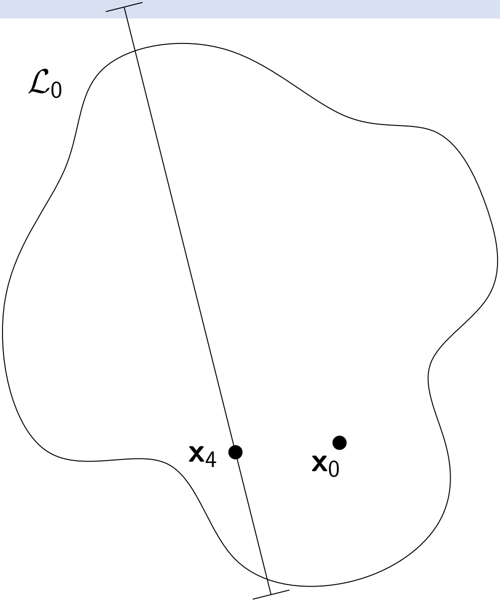
PolyChord

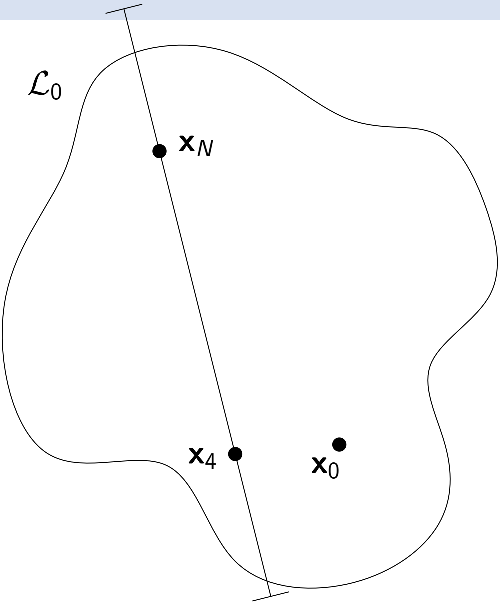


PolyChord

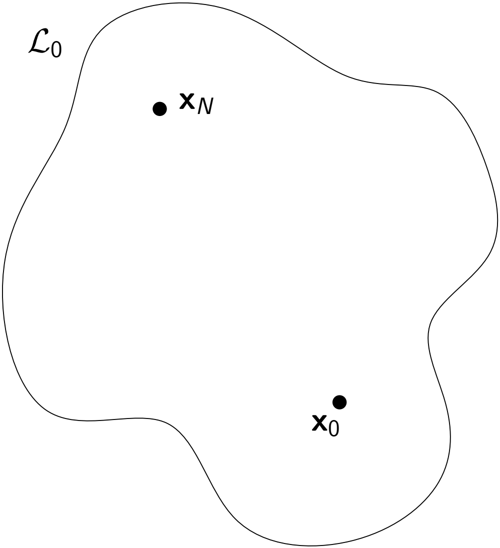



$\mathcal{L}_0$
$\mathbf{x}_4$
$\mathbf{x}_0$


$\mathcal{L}_0$
$\mathbf{x}_4$
$\mathbf{x}_0$


$\mathcal{L}_0$
$\mathbf{x}_N$
$\mathbf{x}_4$
$\mathbf{x}_0$

PolyChord



PolyChord

- arXiv:1506.00171
- Uses slice sampling to generate new points: euclid.aos/1056562461
- Each step requires $\sim \mathcal{O}(3-5)$ evaluations, need $N \sim \mathcal{O}(D)$ steps per chain to decorrelate from start point.
- Worse than MultiNest in low dimensions, exponentially more efficient in high dimensions.
- Chain-based $\Rightarrow$ allows exploitation of fast-slow hierarchy.
- Under active maintenance/development

Further advances in nested sampling

- Dynamic nested sampling
  - arXiv:1704.03459
  - Allows one to refine a run and generate more points, e.g. in the posterior bulk, or prior tails.
- Consistency checking
  - arXiv:1804.06406
  - Unweaving runs allows for cross-checking and testing for imperfect contour sampling.
- Diffusive nested sampling
  - arXiv:0912.2380
  - Fuzzy contours represent an alternative approach to nested sampling
- PolyChord 2.0
  - arXiv:0912.2380
  - Under active develoment, promises $\sim \mathcal{O}(D)$ speed-up

Takeaway points

- Nested sampling is far more than a posterior sampler.
- To do nested sampling in high dimensions, you cannot use MultiNest.
- Fast-slow hierarchies have proven extremely useful in speeding up Planck, AMI and DES analyses.
- PolyChord is available on GitHub: github.com/PolyChord/PolyChordLite
- PolyChord interfaces are available for CosmoMC, cosmosis, MontePython and GAMBIT.