# Cosmological Parameter Extraction from Data

Benjamin Audren

Institute of Theoretical Physics
École Polytechnique Fédérale de Lausanne

12/05/2014

# Outline

1. Introduction
   - Need for a statistical tool
   - Bayesian approach
2. Comparison of existing methods
   - Metropolis-Hastings
   - Importance Sampling
   - EEMCE
   - Nested Sampling

# Outline

# Bibliography

## Book

- Bayesian methods in cosmology (Hobson, Jaffe, Liddle, Mukherjee and Parkinson)

## Review Articles

- Comparison of sampling techniques for Bayesian parameter estimation (*Rupert Allison, Joanna Dunkley*, arXiv:1308.2675)
- Bayes in the sky: Bayesian inference and model selection in cosmology (*Roberto Trotta*, arXiv:0803.4089)

# Precision Cosmology

## Precise measurement of random realisations

- Beginning of **precision** experiments with **WMAP**

# Precision Cosmology

## Precise measurement of random realisations

- Beginning of **precision** experiments with **WMAP**
- Cosmology is a **stochastic theory**, requiring observation of **statistical quantities**

# Precision Cosmology

## Precise measurement of random realisations

- Beginning of **precision** experiments with **WMAP**
- Cosmology is a **stochastic theory**, requiring observation of **statistical quantities**
- A **set of parameters** produce prediction for these **statistical quantities**

# Precision Cosmology

## Precise measurement of random realisations

- Beginning of **precision** experiments with **WMAP**
- Cosmology is a **stochastic theory**, requiring observation of **statistical quantities**
- A **set of parameters** produce prediction for these **statistical quantities**
- How to **infer the value of the parameters from the data** ?

# Precision Cosmology

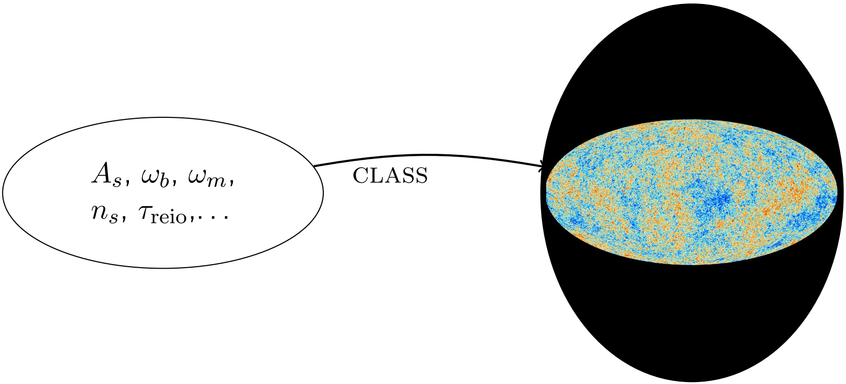
## Precise measurement of random realisations

- Beginning of **precision** experiments with **WMAP**
- Cosmology is a **stochastic theory**, requiring observation of **statistical quantities**
- A **set of parameters** produce prediction for these **statistical quantities**
- How to **infer the value of the parameters from the data** ?

## Similar to Particle Physics

Prediction only for, *e.g.* the rate of decay of a particle. Information acquired when statistically observing this decay channel (how many times did it decay to this particular product ?)

# The big picture



# Bayesian approach

## from Bayesian Methods in Cosmology

### Bayesian methods

- All quantities are considered **statistical**

# Bayesian approach

## from Bayesian Methods in Cosmology

### Bayesian methods

- All quantities are considered **statistical**
- The question we can answer is: **is this model better than another ?**

# Bayesian approach

## from Bayesian Methods in Cosmology

### Bayesian methods

- All quantities are considered **statistical**
- The question we can answer is: **is this model better than another ?**
- We infer **credible regions** in which, given a model, the parameters live.

# Bayesian approach

## from Bayesian Methods in Cosmology

### Bayesian methods

- All quantities are considered **statistical**
- The question we can answer is: **is this model better than another ?**
- We infer **credible regions** in which, given a model, the parameters live.
- Our knowledge depends on the **data measured**.

# Bayesian approach

## Definitions

Given a **context** $\mathcal{I}$ and a set of **data** $D$:

$$\left.\begin{array}{l} \bullet\ \theta\text{: \textbf{continuous} values for a parameter set} \\ \bullet\ \mathrm{pr}(\theta) \geq 0\text{: \textbf{probability} function is positive} \\ \bullet\ \int \mathrm{pr}(\theta)\mathrm{d}\theta = 1\text{: \textbf{Sum rule}} \\ \bullet\ \mathrm{pr}(\phi, \theta) = \mathrm{pr}(\phi|\theta)\mathrm{pr}(\theta)\text{: \textbf{Product rule}} \end{array}\right\} \quad || \; \mathcal{I}$$

# Bayes Theorem

## Bayes Theorem

*All quantities are given in the context $\mathcal{I}$*

$$\begin{array}{ccccc} \mathrm{pr}(\theta)\mathrm{pr}(D|\theta) & = & \mathrm{pr}(\theta, D) & = & \mathrm{pr}(D)\mathrm{pr}(\theta|D) \\ \textbf{Prior} \times \textbf{Likelihood} & = & \textbf{Joint} & = & \textbf{Evidence} \times \textbf{Posterior} \\ \pi(\theta)\mathcal{L}(\theta) & = & \dots & = & E\mathcal{P}(\theta) \\ \text{Input} & & \longrightarrow & & \text{Output} \end{array}$$

# Bayes Theorem

## Bayes Theorem

*All quantities are given in the context $\mathcal{I}$*

$$\begin{array}{ccccc} \mathrm{pr}(\theta)\mathrm{pr}(D|\theta) & = & \mathrm{pr}(\theta, D) & = & \mathrm{pr}(D)\mathrm{pr}(\theta|D) \\ \textbf{Prior} \times \text{ Likelihood} & = & \text{Joint} & = & \text{Evidence} \times \text{ Posterior} \\ \pi(\theta)\mathcal{L}(\theta) & = & \ldots & = & E\mathcal{P}(\theta) \\ \text{Input} & & \longrightarrow & & \text{Output} \end{array}$$

## Prior

*A priori* information on the parameters. Most of the time, it is **flat** (uniform chance to be inside a given volume).

# Bayes Theorem

## Bayes Theorem

*All quantities are given in the context $\mathcal{I}$*

$$\begin{array}{ccccc}
\mathrm{pr}(\theta)\mathrm{pr}(\mathbf{D}|\theta) & = & \mathrm{pr}(\theta, D) & = & \mathrm{pr}(D)\mathrm{pr}(\theta|D) \\
\text{Prior} \times \ \textbf{Likelihood} & = & \text{Joint} & = & \text{Evidence} \times \text{Posterior} \\
\pi(\theta)\mathcal{L}(\theta) & = & \ldots & = & E\mathcal{P}(\theta) \\
\text{Input} & & \longrightarrow & & \text{Output}
\end{array}$$

## Likelihood

Given by the instrument operated at *known input*. If uncontrolled unknow: **nuisance parameters**.

# Bayes Theorem

## Bayes Theorem

*All quantities are given in the context $\mathcal{I}$*

$$\begin{array}{ccccc} \mathrm{pr}(\theta)\mathrm{pr}(D|\theta) & = & \mathrm{pr}(\theta, D) & = & \mathrm{pr}(\mathbf{D})\mathrm{pr}(\theta|D) \\ \text{Prior} \times \text{ Likelihood} & = & \text{Joint} & = & \textbf{Evidence} \times \text{Posterior} \\ \pi(\theta)\mathcal{L}(\theta) & = & \dots & = & \mathbf{E}\mathcal{P}(\theta) \\ \text{Input} & & \longrightarrow & & \text{Output} \end{array}$$

## Evidence

Only recovered with Nested Sampling, gives an information on how well the given context suits the data.

# Bayes Theorem

## Bayes Theorem

*All quantities are given in the context $\mathcal{I}$*

$$
\begin{array}{ccccc}
\mathrm{pr}(\theta)\mathrm{pr}(D|\theta) & = & \mathrm{pr}(\theta, D) & = & \mathrm{pr}(D)\mathrm{pr}(\theta|\mathbf{D}) \\
\text{Prior} \times \text{ Likelihood} & = & \text{Joint} & = & \text{Evidence} \times \textbf{Posterior} \\
\pi(\theta)\mathcal{L}(\theta) & = & \dots & = & E\mathcal{P}(\theta) \\
\text{Input} & & \longrightarrow & & \text{Output}
\end{array}
$$

## Posterior Distribution

The name of the game. Inferred distribution of probability, after using the data.

# Bayes Theorem

## Complete Rules

$$\int \pi(\theta)\mathrm{d}\theta = \int \mathcal{P}(\theta)\mathrm{d}\theta = 1$$

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

$$\mathcal{P}(\theta) = \frac{\pi(\theta)\mathcal{L}(\theta)}{E}$$

# Bayesian approach: issues

## What are the problems ?

- **Evidence** is hard to compute because...
- we don't usually know the **Likelihood** function (analytically).
- so we don't know where to **sample** it (many dimensions)...
- and it might be **computationally expensive**.

# Bayesian approach: issues

### What are the problems ?

- **Evidence** is hard to compute because...
- we don't usually know the **Likelihood** function (analytically).
- so we don't know where to **sample** it (many dimensions)...
- and it might be **computationally expensive**.

### How to deal with it ?

- we have to sample **randomly** the volume (*all methods*)
- we can avoid computing this integral by **computing ratios** (*mcmc*)

# Methods that give the Evidence

## Comparison between models

Evidence is **how well your context explains the data**. Nested Sampling gives this. Others dont. . . but best-fit likelihood gives some indication.

# Methods that give the Evidence

## Comparison between models

Evidence is **how well your context explains the data**. Nested Sampling gives this. Others dont... but best-fit likelihood gives some indication.

## Caution

These methods will only give you a **probable** answer, not a definite one. It is the price you pay for not doing your integral exactly.

# But. . .

### Living without the evidence

**Theoretically motivated model**: we want to know the values of this parameter to explain the data. Then, the **posterior** is an interesting quantity, and the **evidence** can be left aside temporarily.

# But. . .

## Living without the evidence

**Theoretically motivated model**: we want to know the values of this parameter to explain the data. Then, the **posterior** is an interesting quantity, and the **evidence** can be left aside temporarily.

## Nonetheless

Beware of the best-fit likelihood value !

# Outline

# Algorithm and Sampler

## Foreword

### Algorithm

Tells you how to **choose** which point you move, and if you **accept** a new point or not

### Sampler

Tells you how to **move**, **select a** new point.

Sometimes used **interchangeably**

# Algorithm and Sampler

## Foreword

**available in Monte Python**

As of v2.0.0, you can use **MultiNest** (`nested sampling`, by Farhan Feroz & Mike Hobson), and the **CosmoHammer** (`emcee`, by Joel Akeret & Sebastian Seehars)

# Algorithm and Sampler

## Foreword

### available in Monte Python

As of v2.0.0, you can use **MultiNest** (`nested sampling`, by Farhan Feroz & Mike Hobson), and the **CosmoHammer** (`emcee`, by Joel Akeret & Sebastian Seehars)

### MultiNest

**MultiNest** was easily implemented in Monte Python thanks to **PyMultiNest**, done by Johannes Buchner, and the help of Jesus Torrado

# Algorithm and Sampler

## Foreword

### available in Monte Python

As of v2.0.0, you can use **MultiNest** (`nested sampling`, by Farhan Feroz & Mike Hobson), and the **CosmoHammer** (`emcee`, by Joel Akeret & Sebastian Seehars)

### MultiNest

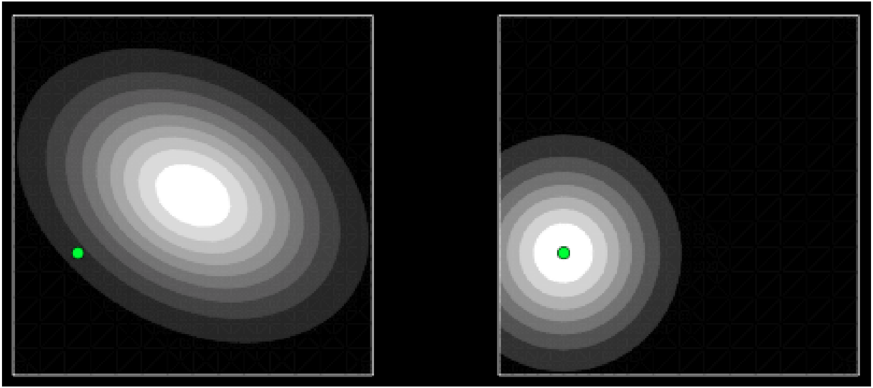
**MultiNest** was easily implemented in Monte Python thanks to **PyMultiNest**, done by Johannes Buchner, and the help of Jesus Torrado

### Cosmo Hammer

Thanks to Joel and Sebastian for helping setting this up

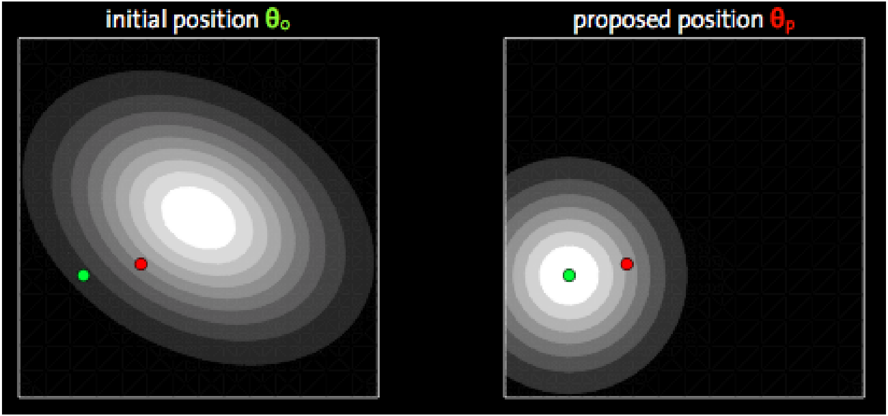
# Metropolis-Hastings

## Courtesy of Sebastian Seehars



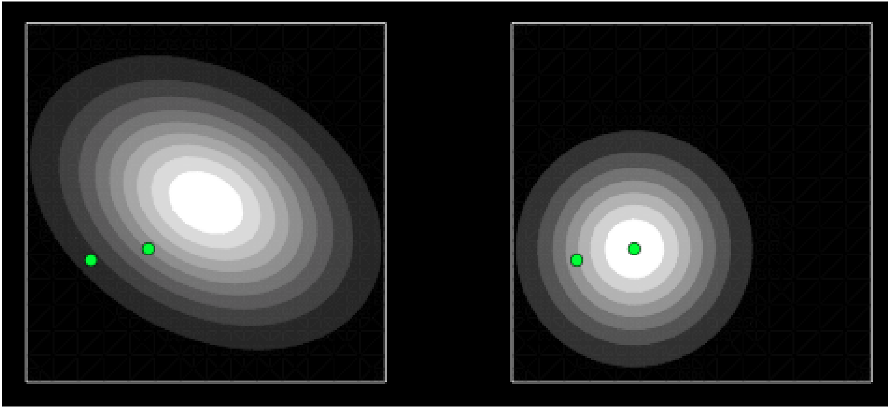
# Metropolis-Hastings

## Courtesy of Sebastian Seehars



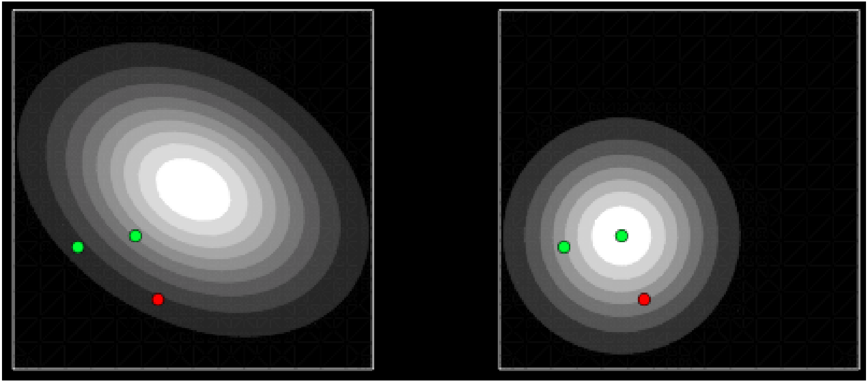
# Metropolis-Hastings

## Courtesy of Sebastian Seehars



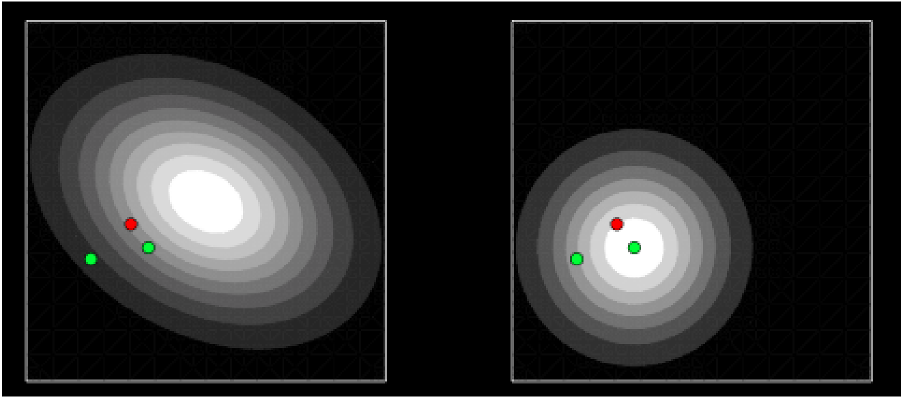
# Metropolis-Hastings

## Courtesy of Sebastian Seehars



# Metropolis-Hastings

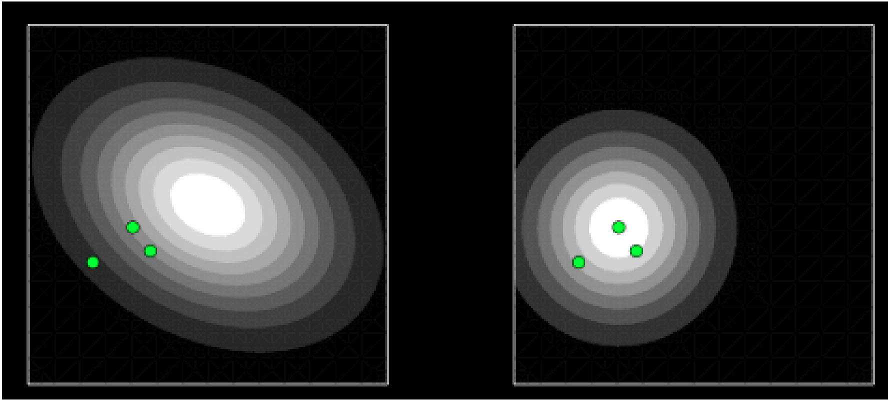
## Courtesy of Sebastian Seehars



multiplicity + 1

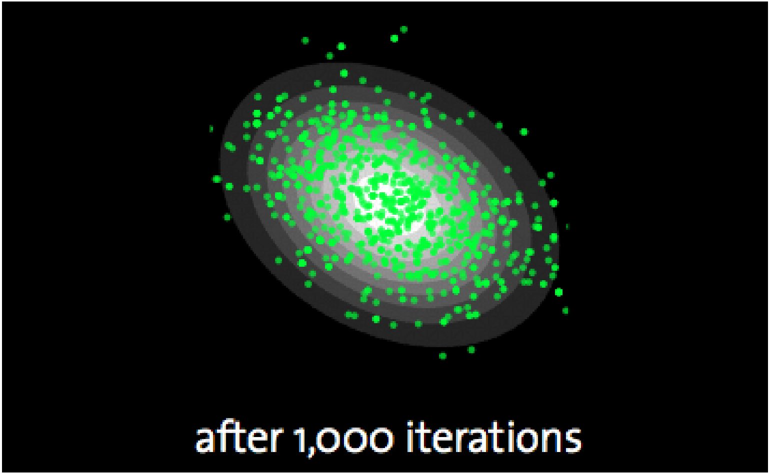
# Metropolis-Hastings

## Courtesy of Sebastian Seehars



# Metropolis-Hastings

## Courtesy of Sebastian Seehars

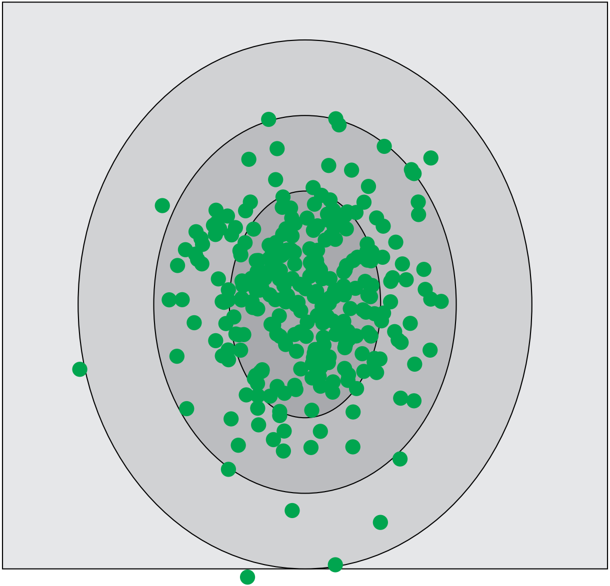


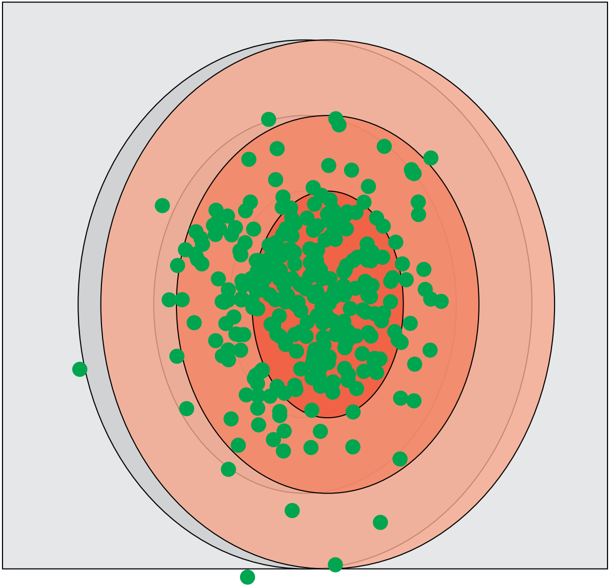
# Importance Sampling

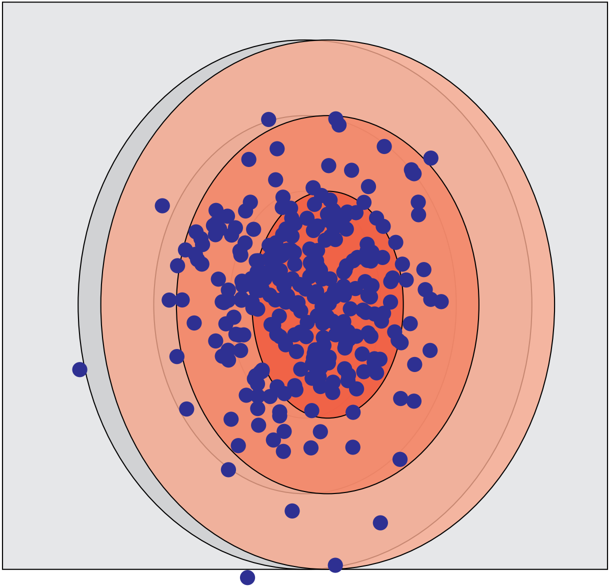
## Not a real sampling

- start from an existing, **similar** distribution
- **add the new likelihood** at every already sampled point
- **write a new chain**, with the multiplicity changed:

$$\tilde{N} = N\frac{\tilde{\mathcal{L}}}{\mathcal{L}}$$







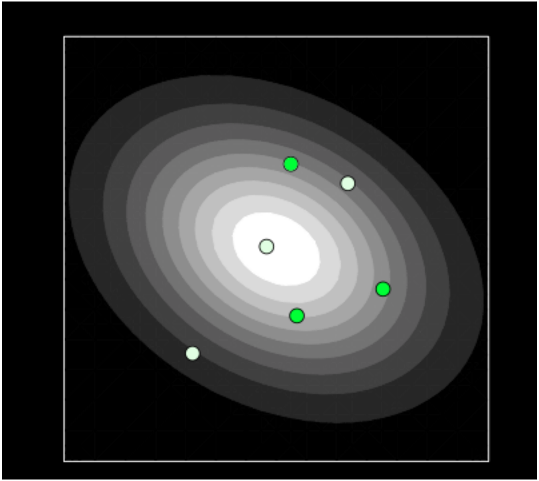
# Benefits of Importance Sampling

**When to use it?**

- **Existing** run with a set of **slow** experiments
- Want to add one more experiment (**prior on** $H_0$)
- This should be very fast!

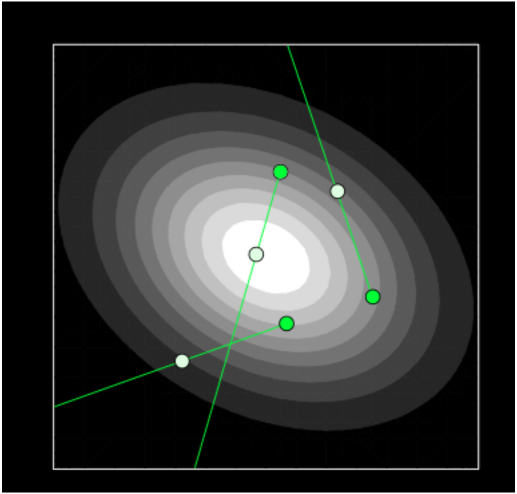
# emcee

## Courtesy of Sebastian Seehars



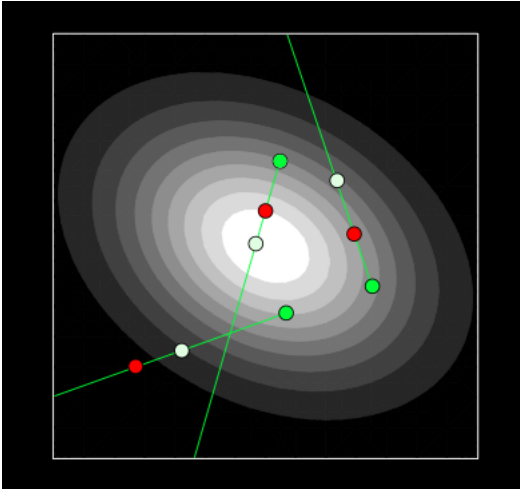
# emcee

## Courtesy of Sebastian Seehars



# emcee

## Courtesy of Sebastian Seehars



# Nested Sampling

## Goals

- Trying to answer a different question: **is this model better?**

# Nested Sampling

## Goals

- Trying to answer a different question: **is this model better?**
- Computes the evidence $E$ (Occam's razor)

# Nested Sampling

## Goals

- Trying to answer a different question: **is this model better?**
- Computes the evidence $E$ (Occam's razor)
- As a by-product: **posterior distribution**!

# Nested Sampling

## Goals

- Trying to answer a different question: **is this model better?**
- Computes the evidence $E$ (Occam's razor)
- As a by-product: **posterior distribution**!

## Pros/Cons

**Good to explore weirdly shaped distributions**,
**but scales not so well with many parameters**

# Nested Sampling

## Procedure

- Active set ($\simeq 200$ points) chosen randomnly in prior range

# Nested Sampling

## Procedure

- Active set ($\simeq$ 200 points) chosen randomnly in prior range
- Iterate over the points: **discard least likely**, pick a new point until a **more likely** is found

# Nested Sampling

## Procedure

- Active set ($\simeq$ 200 points) chosen randomnly in prior range
- Iterate over the points: **discard least likely**, pick a new point until a **more likely** is found
- To propose a new point: **approximate the posterior distribution by encompassing all other points within an ellipse**.

# Nested Sampling

## Procedure

- Active set ($\simeq$ 200 points) chosen randomnly in prior range
- Iterate over the points: **discard least likely**, pick a new point until a **more likely** is found
- To propose a new point: **approximate the posterior distribution by encompassing all other points within an ellipse**.
- The sorting of these points by **increasing likelihood** allows to compute the evidence ($\simeq$ trapezoidal integration)

# Nested Sampling

## Mathematical Procedure

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

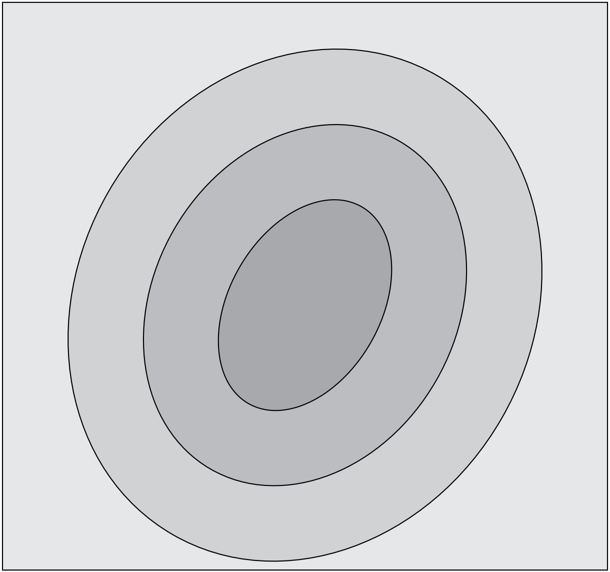
$$dX = \pi(\theta)\mathrm{d}\theta\text{: prior volume}$$

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

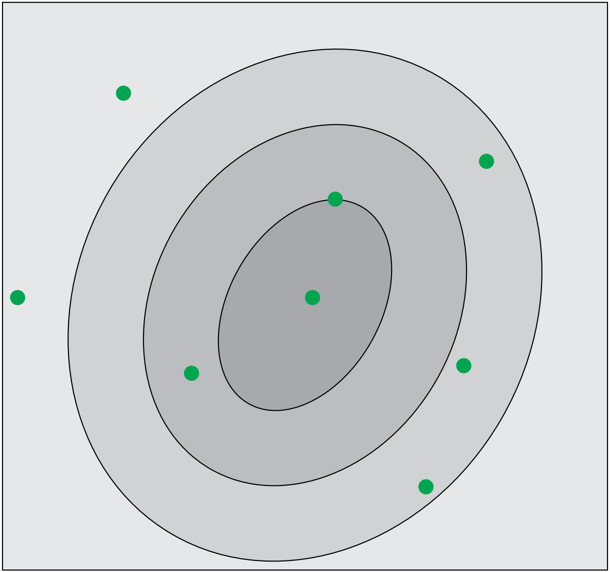
$$E = \int_0^1 \mathcal{L}(X)dX$$

$\mathcal{L}(X)$ is a monotonically decreasing function of $X$

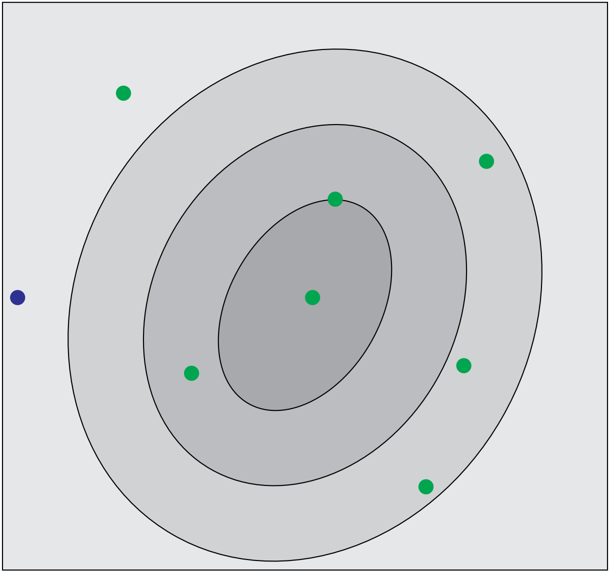
# Nested Sampling



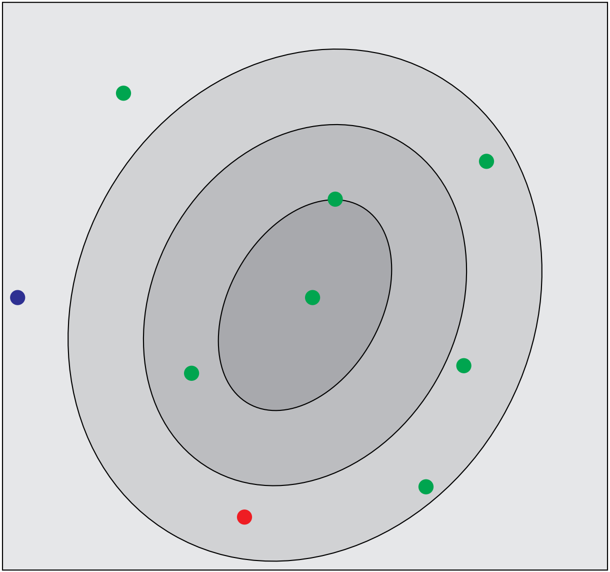
# Nested Sampling



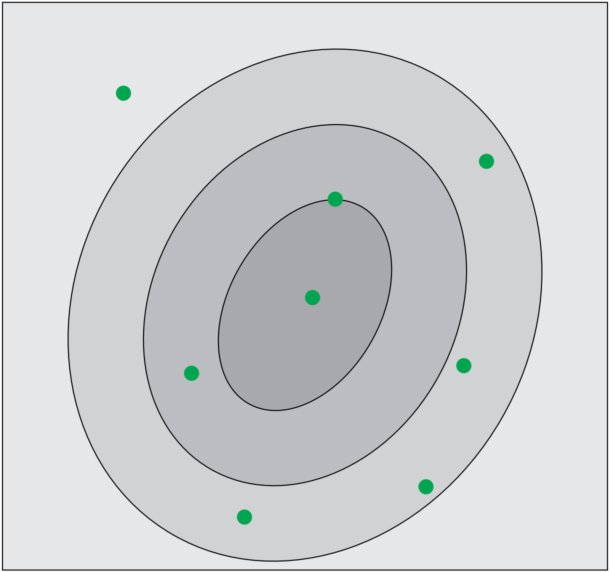
# Nested Sampling



# Nested Sampling



# Nested Sampling



# Nested Sampling

**At each time step**

- The integral $E$ is computed with the **active set**
- The **prior volume** shrinks on the most likely points
- Stops when $\Delta E < 0.5$