Bayesian Parameter Extraction

Benjamin Audren

Institute of Theoretical Physics
École Polytechnique Fédérale de Lausanne

27/10/2014

Outline

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

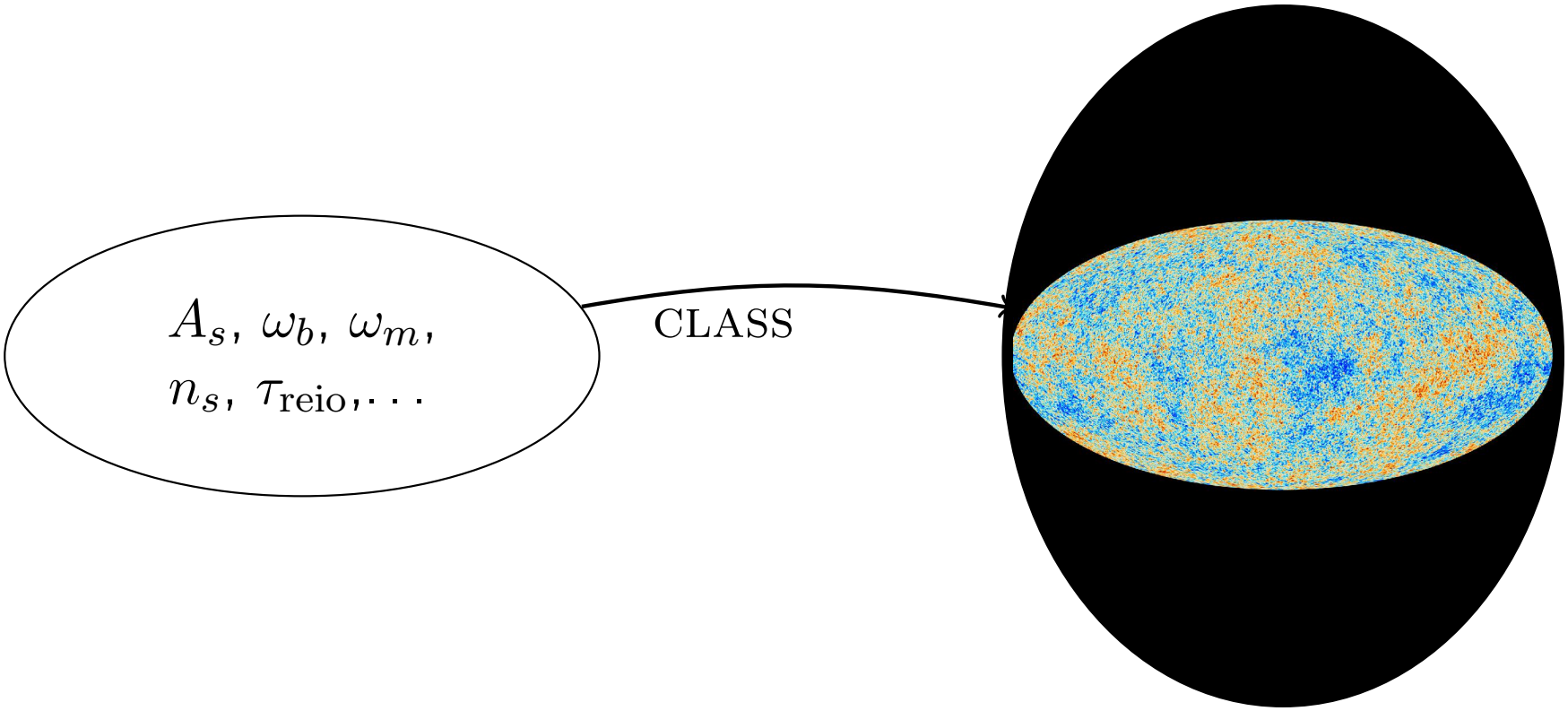
Precise measurement of random realisations

- 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:

- θ: **continuous** values for a parameter set
- $\mathrm{pr}(\theta) \geq 0$: **probability** function is positive
- $\int \mathrm{pr}(\theta)\mathrm{d}\theta = 1$: **Sum rule**
- $\mathrm{pr}(\phi, \theta) = \mathrm{pr}(\phi|\theta)\mathrm{pr}(\theta)$: **Product rule**

$\Big\} \; || \; \mathcal{I}$

Bayes Theorem

Bayes Theorem

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

$\mathrm{pr}(\theta)\mathrm{pr}(D\|\theta)$	$=$	$\mathrm{pr}(\theta, D)$	$=$	$\mathrm{pr}(D)\mathrm{pr}(\theta\|D)$
Prior $\times$ **Likelihood**	$=$	**Joint**	$=$	**Evidence** $\times$ **Posterior**
$\pi(\theta)\mathcal{L}(\theta)$	$=$	$\ldots$	$=$	$E\mathcal{P}(\theta)$
Input		$\longrightarrow$		Output

Bayes Theorem

Bayes Theorem

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

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

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 $\mathcal{I}$

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

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 $\mathcal{I}$

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

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

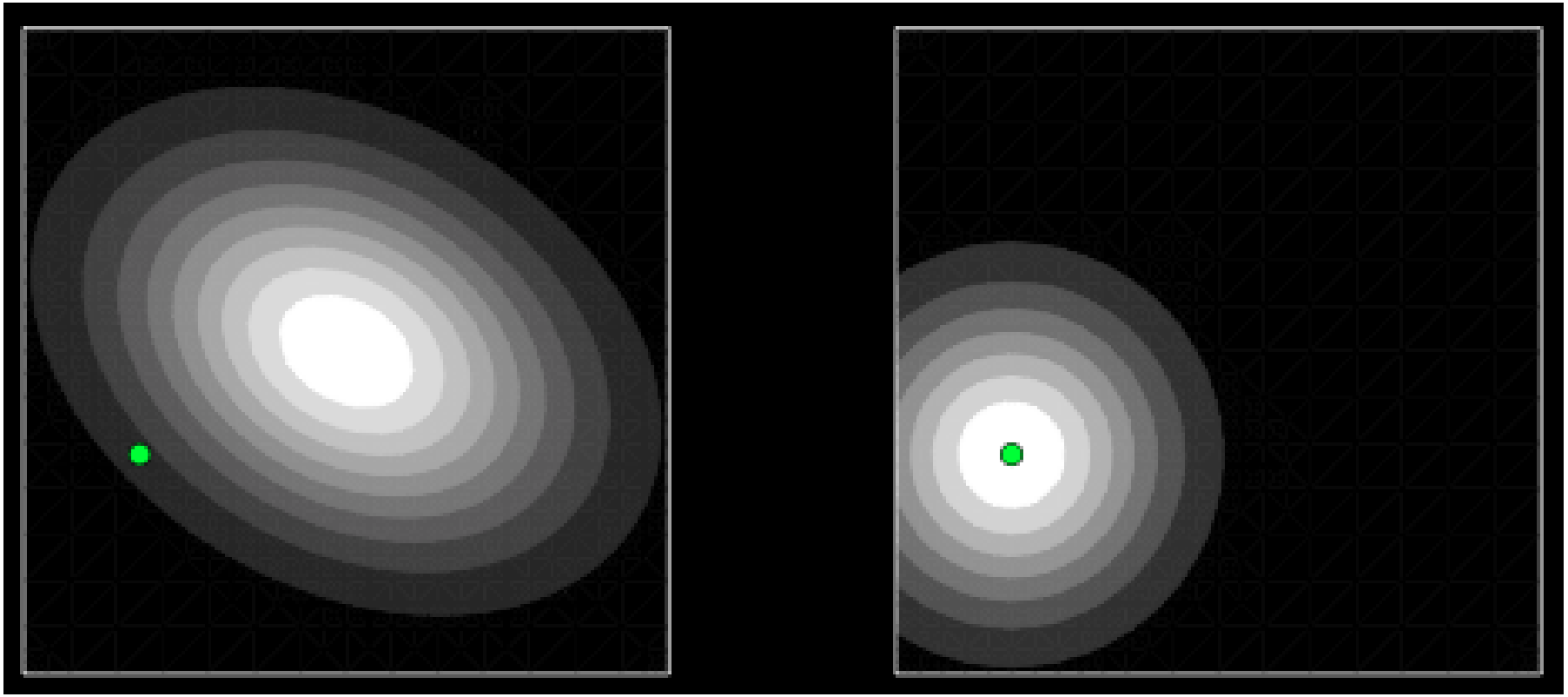
MultiNest was easily implemented in Monte Python thanks to **PyMultiNest**, done by Johannes Buchner, and the help of Jesus Torrado

Cosmo Hammer

Thanks to Joel and Sebastian for helping setting this up

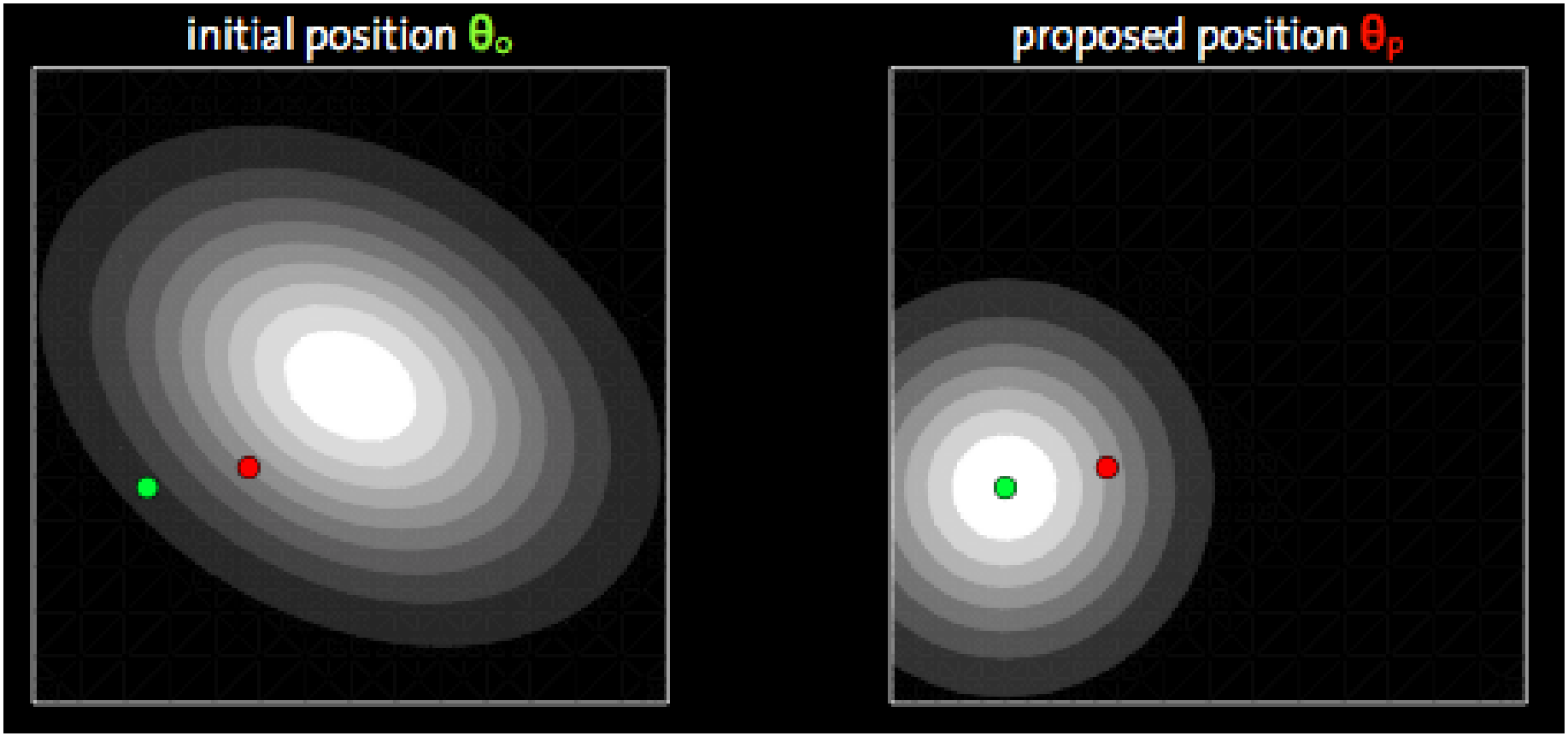
Metropolis-Hastings

Courtesy of Sebastian Seehars



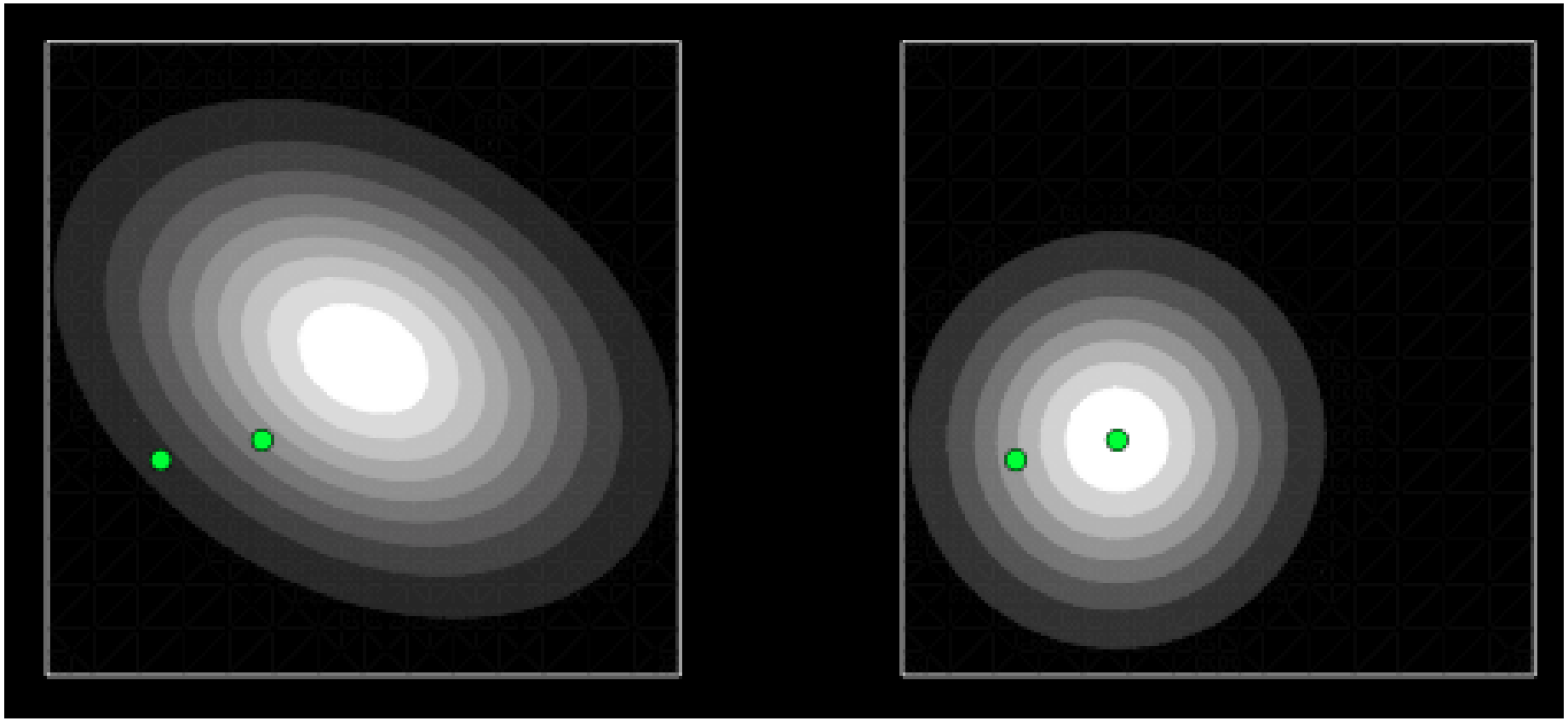
Metropolis-Hastings

Courtesy of Sebastian Seehars



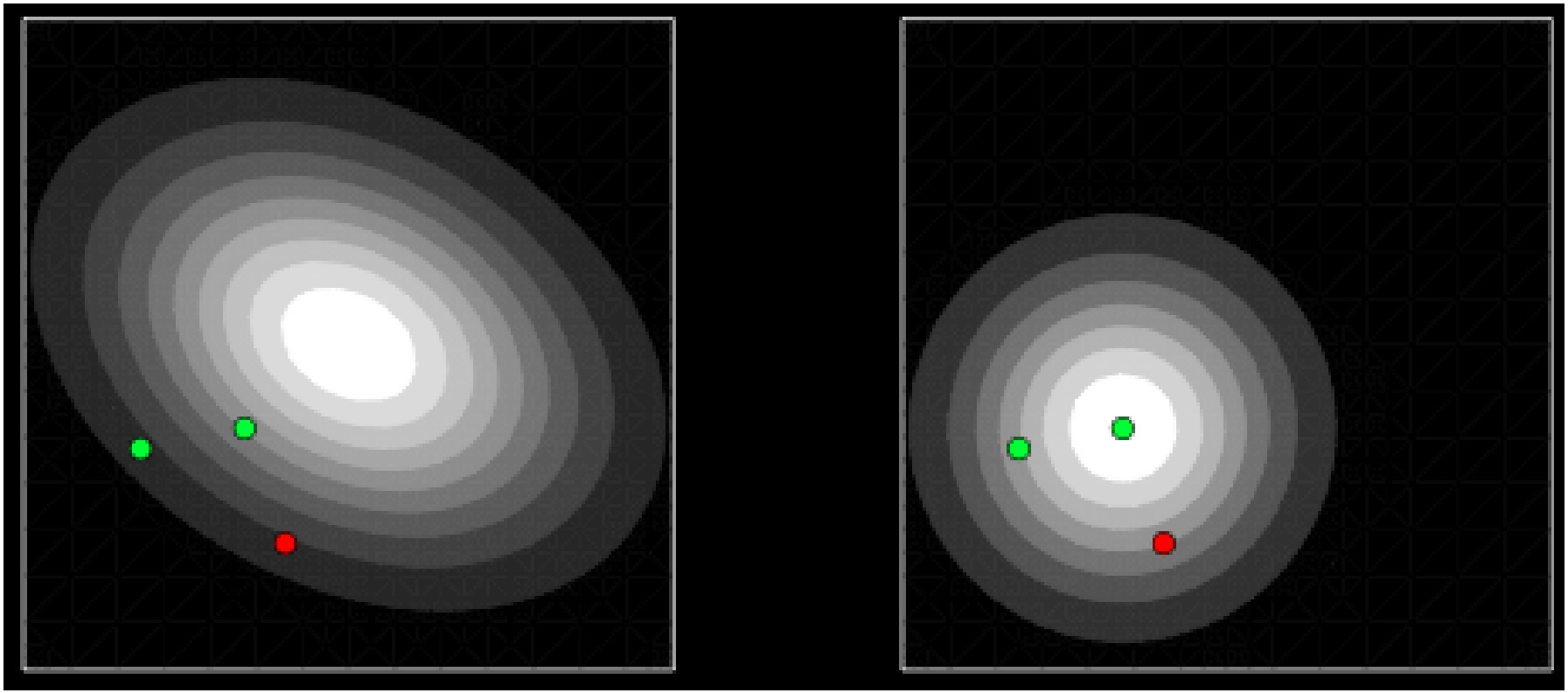
Metropolis-Hastings

Courtesy of Sebastian Seehars



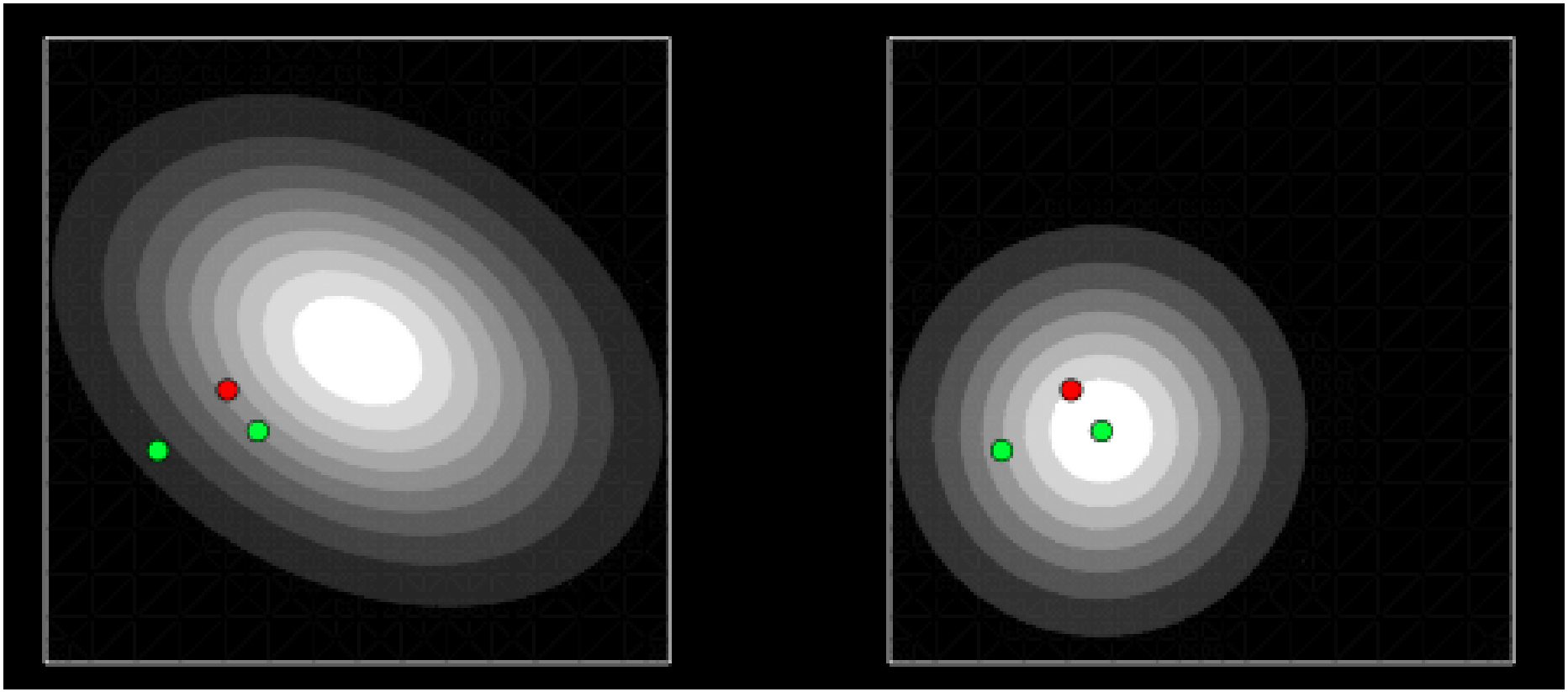
Metropolis-Hastings

Courtesy of Sebastian Seehars



Metropolis-Hastings

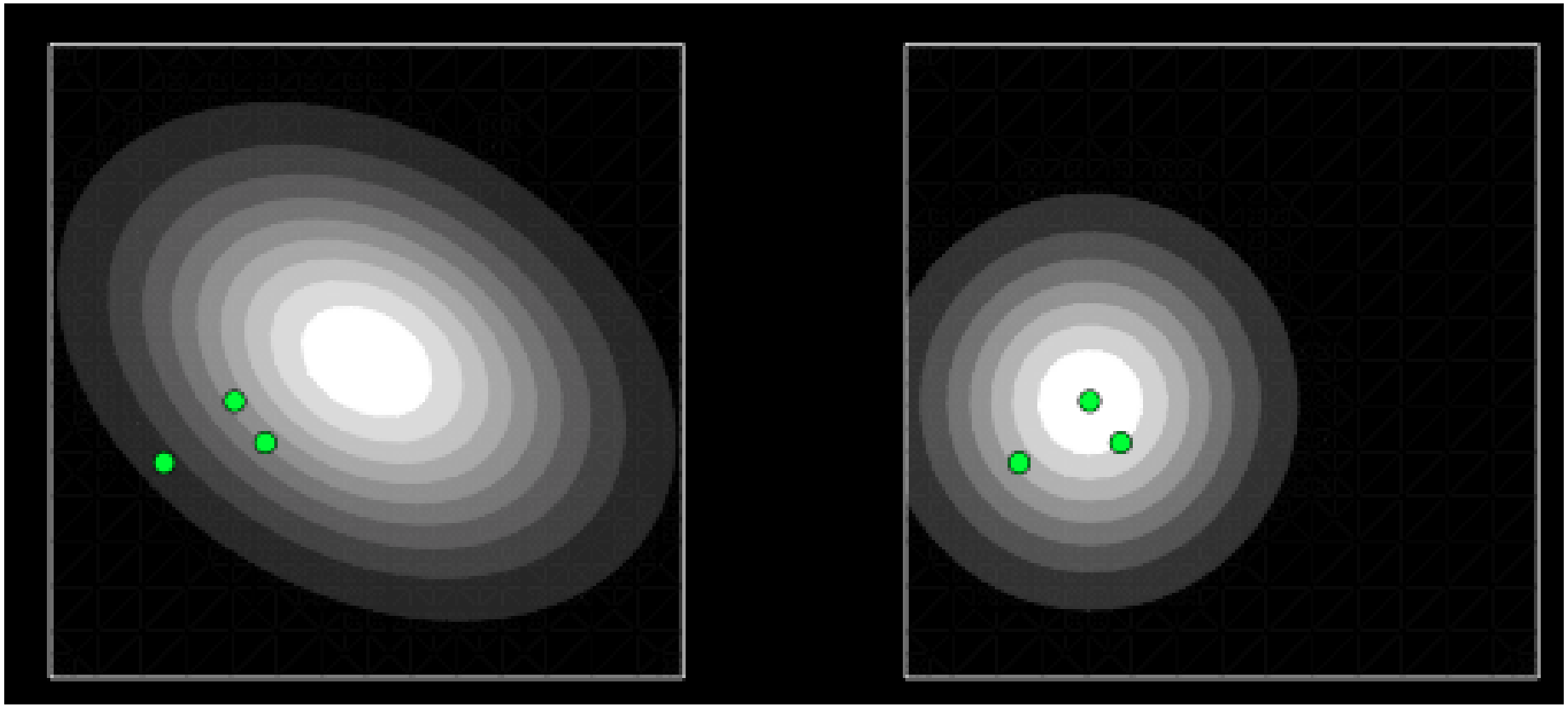
Courtesy of Sebastian Seehars



multiplicity + 1

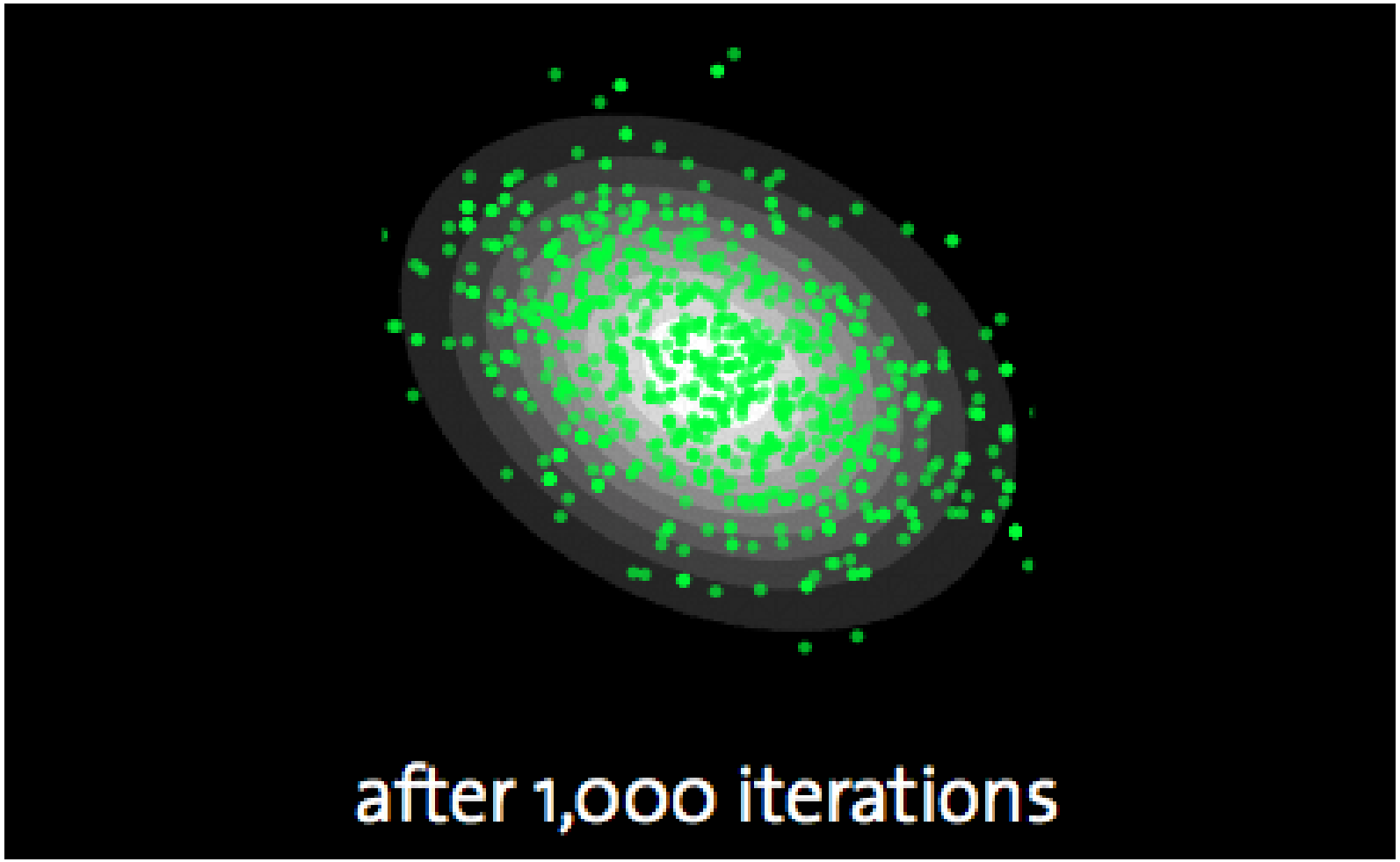
Metropolis-Hastings

Courtesy of Sebastian Seehars



Metropolis-Hastings

Courtesy of Sebastian Seehars

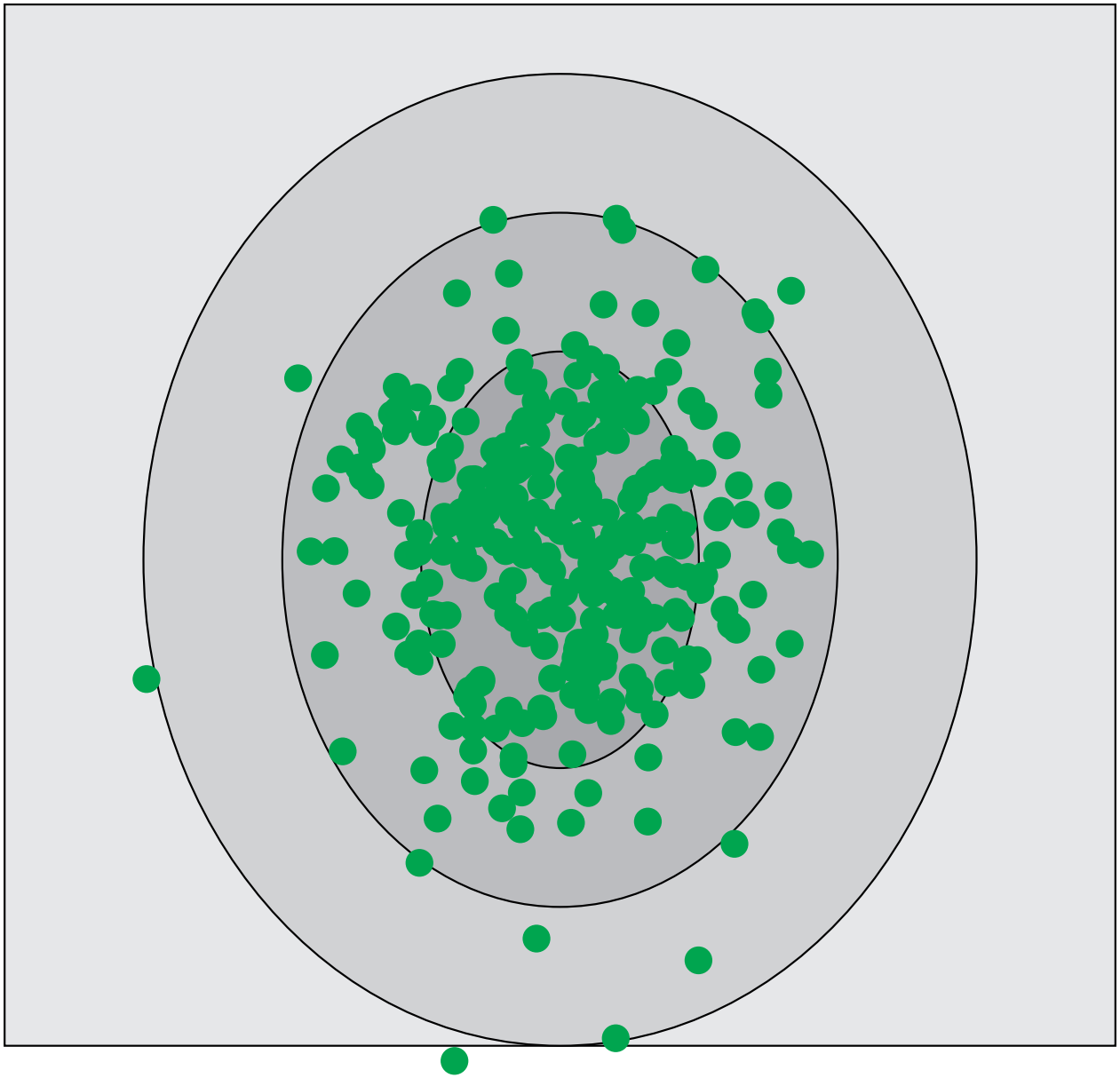


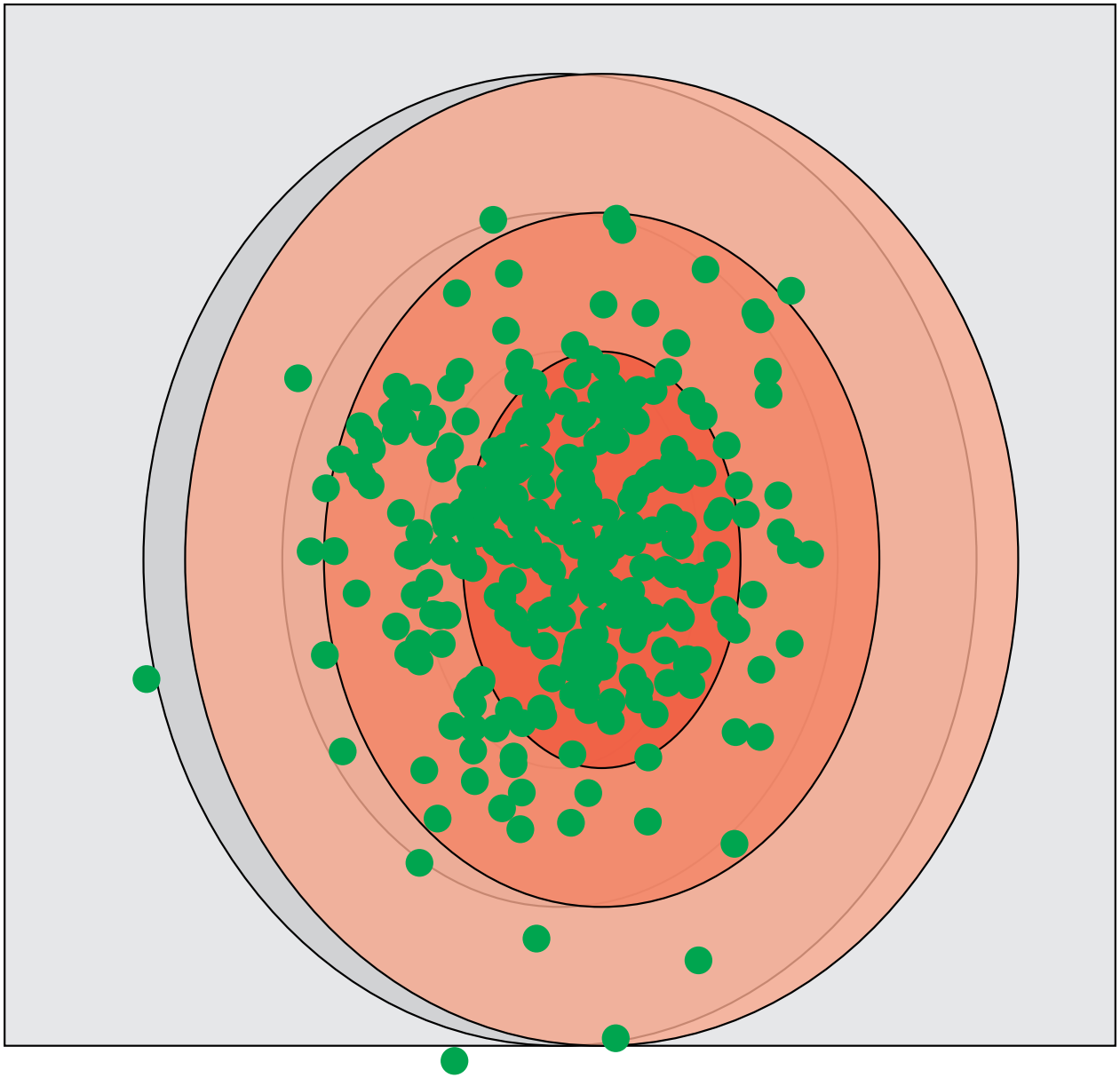
Importance Sampling

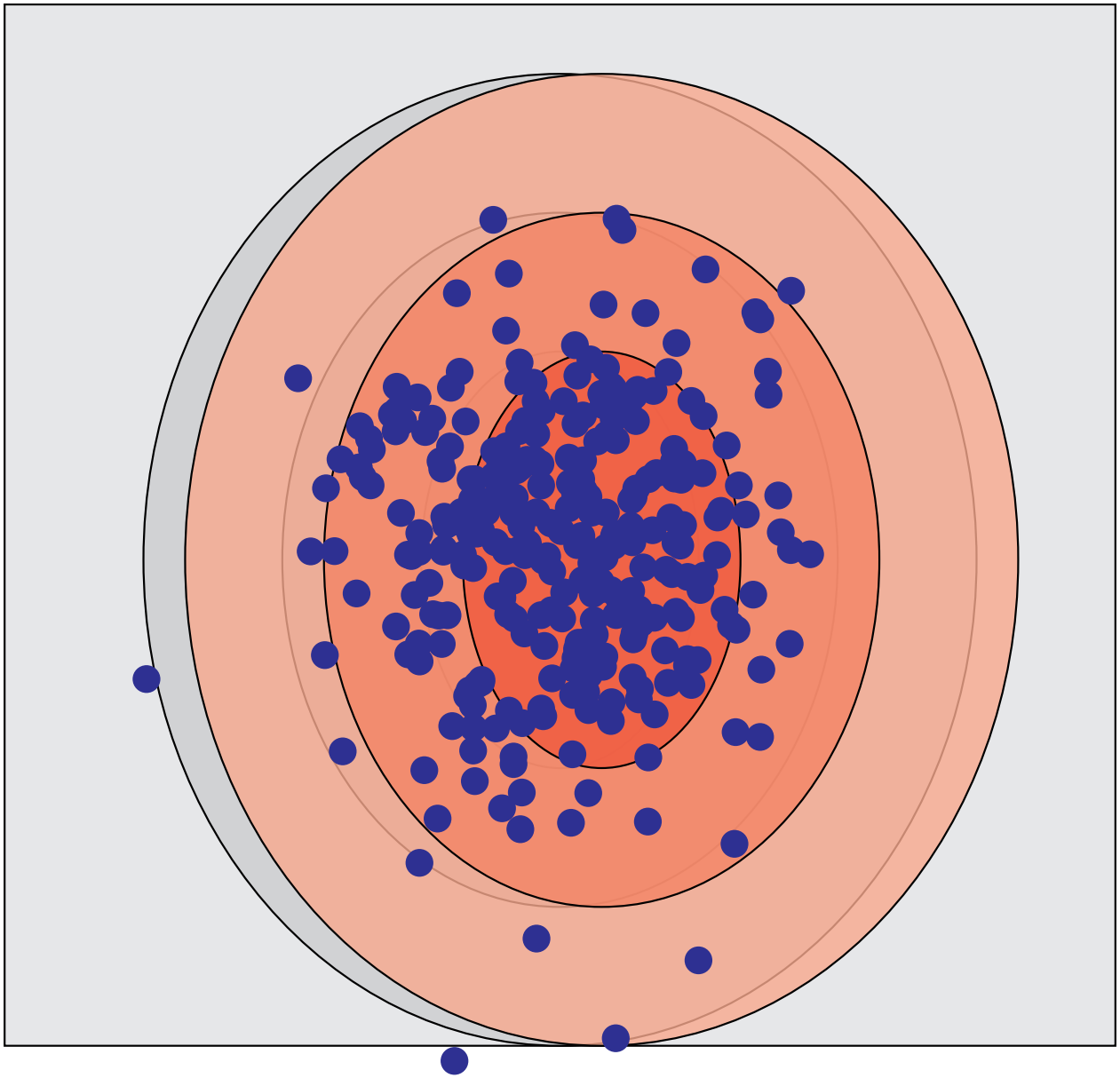
Not a real sampling

- 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!