A brief introduction to Bayesian Statistics through Astronomical Applications (Lecture 3)

Cecilia Mateu J.
Centro de Investigaciones de Astronomía (CIDA)

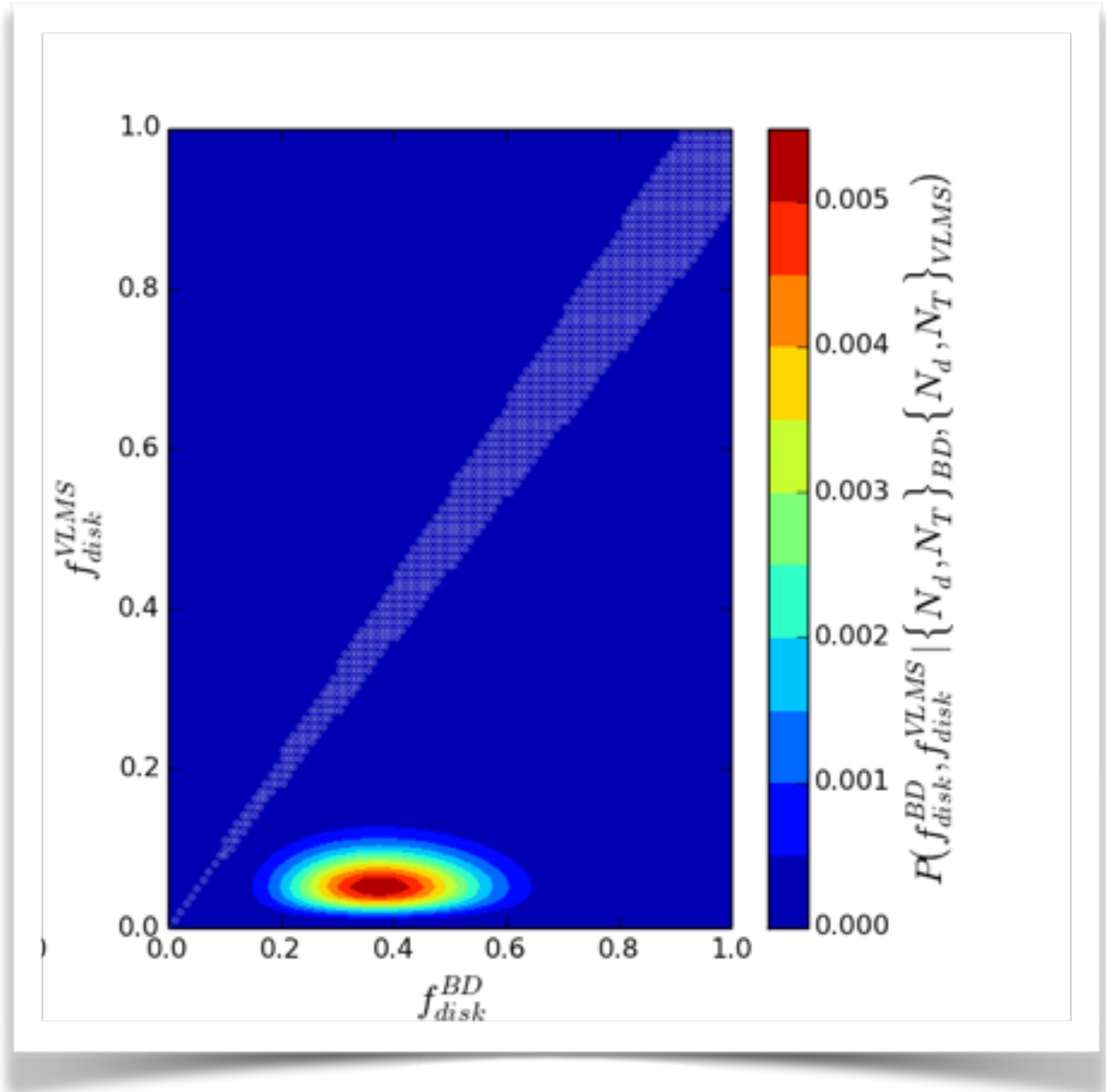
Universidad de Barcelona
27 de octubre de 2016

Exploring the parameter space in high dimensionality problems: Markov Chain Monte Carlo

Computing the Posterior in many-parameter problems

- As we have already discussed, the Posterior can only be found by direct evaluation in problems with very few parameters (<~ 6?)

- We would like to have a way of exploring the parameter space efficiently, spending more computation time around high-probability areas than around low probability ones



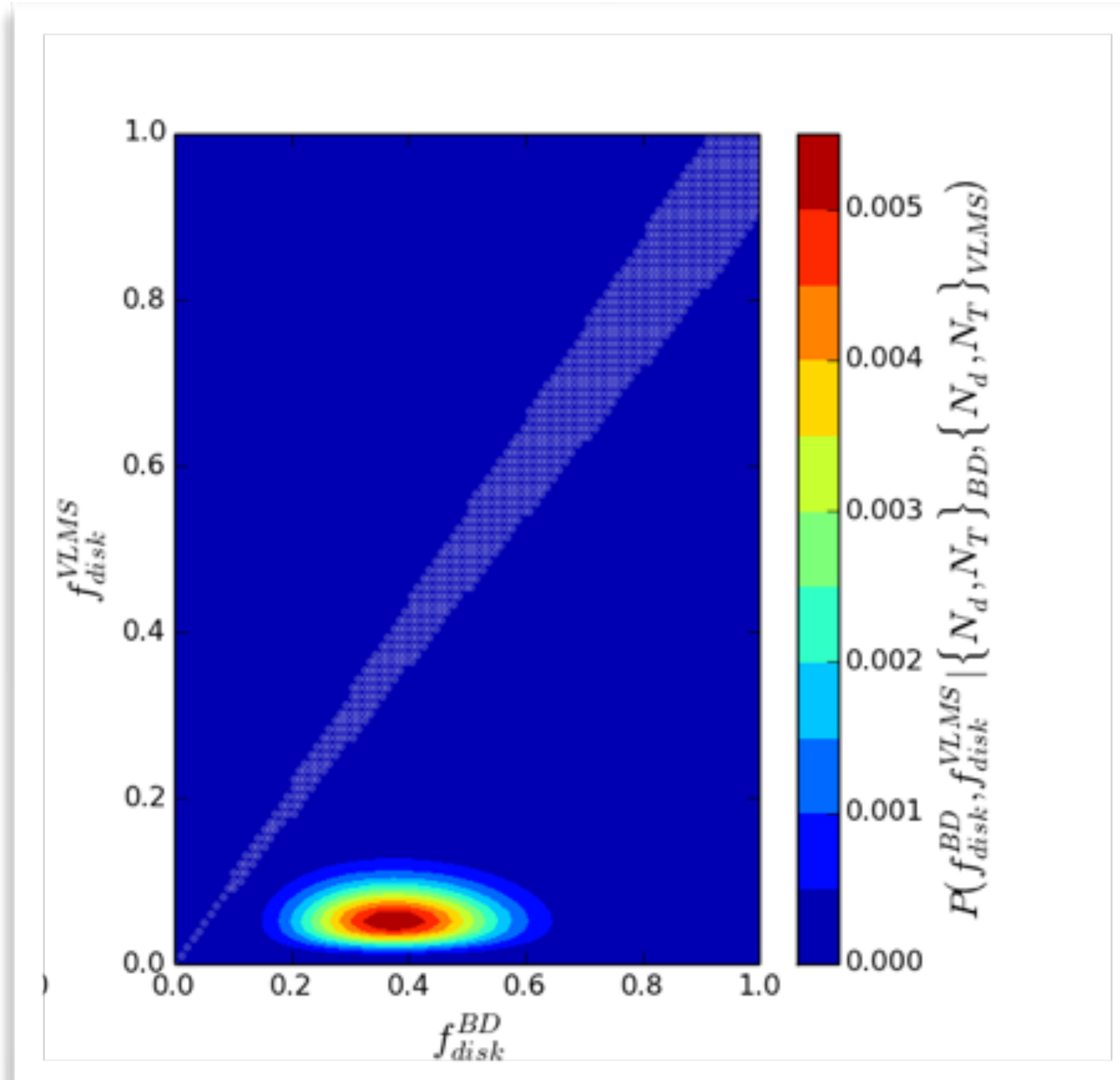
Obtaining samples from the Posterior

- One way is to try to obtain Posterior samples, i.e. a random realisation of the Posterior PDF

- If one has a random realisation of the Posterior with N samples, the Posterior is simply the N-dimensional histogram of this samples

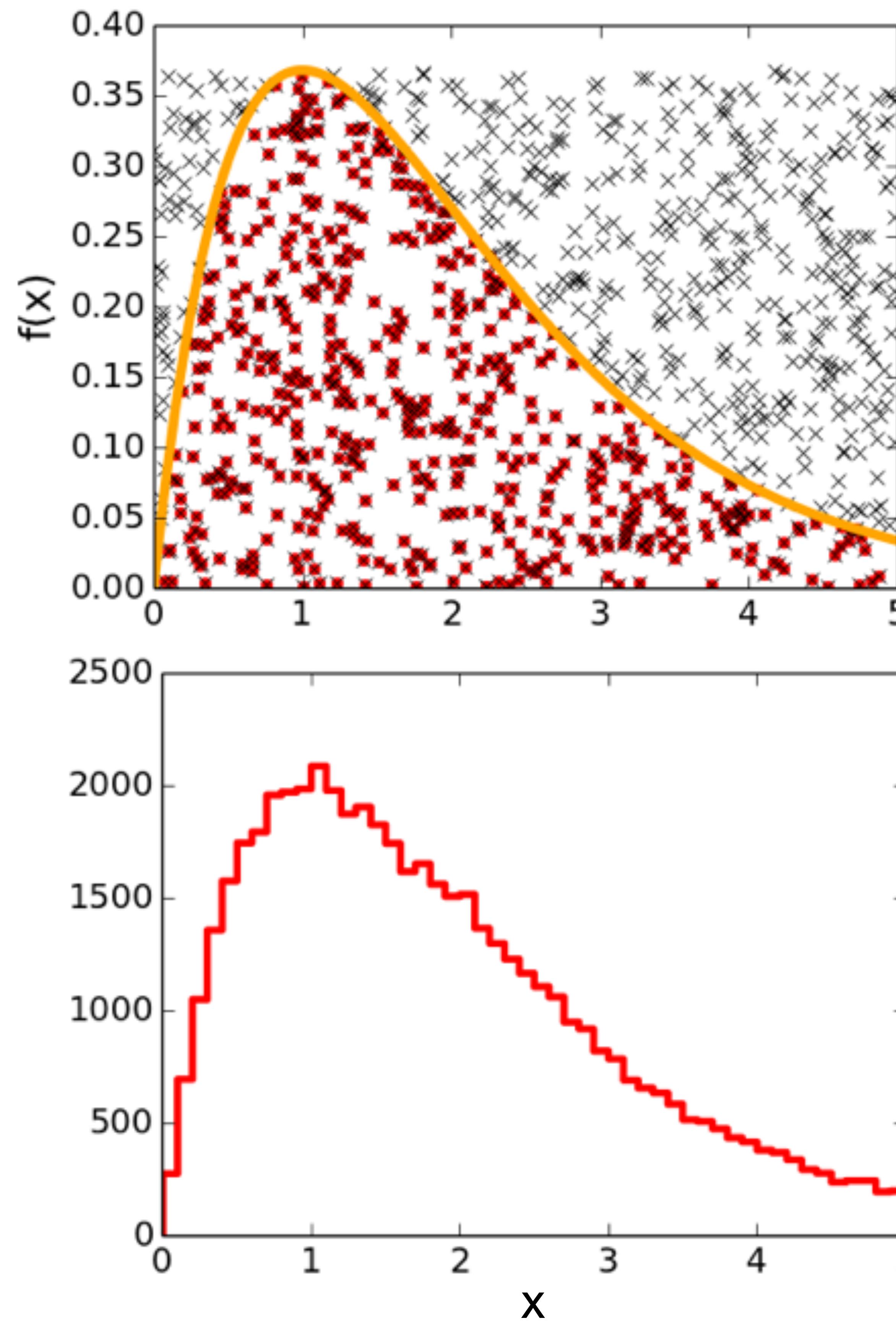
- Having posterior samples, Marginalization is trivial, just the histogram in any lower number of dimensions is the marginal posterior!

- Uncertainties can be easily computed as the standard deviation or percentiles in the resulting histograms



Posterior Samples: Von Neumann Rejection

- A simple way of doing this is with the Von Neumann Rejection Technique, also known as Accept-Reject:
    - Generate random uniformly-distributed samples (x,y) with $x_o<x<x_f$ and $0<y<max(Posterior)$
    - Accept only the samples for which $y<Posterior(x)$
    - … that's it, the accepted points are distributed as the posterior
- This is quite simple and works in any number of dimensions!
- **However…**



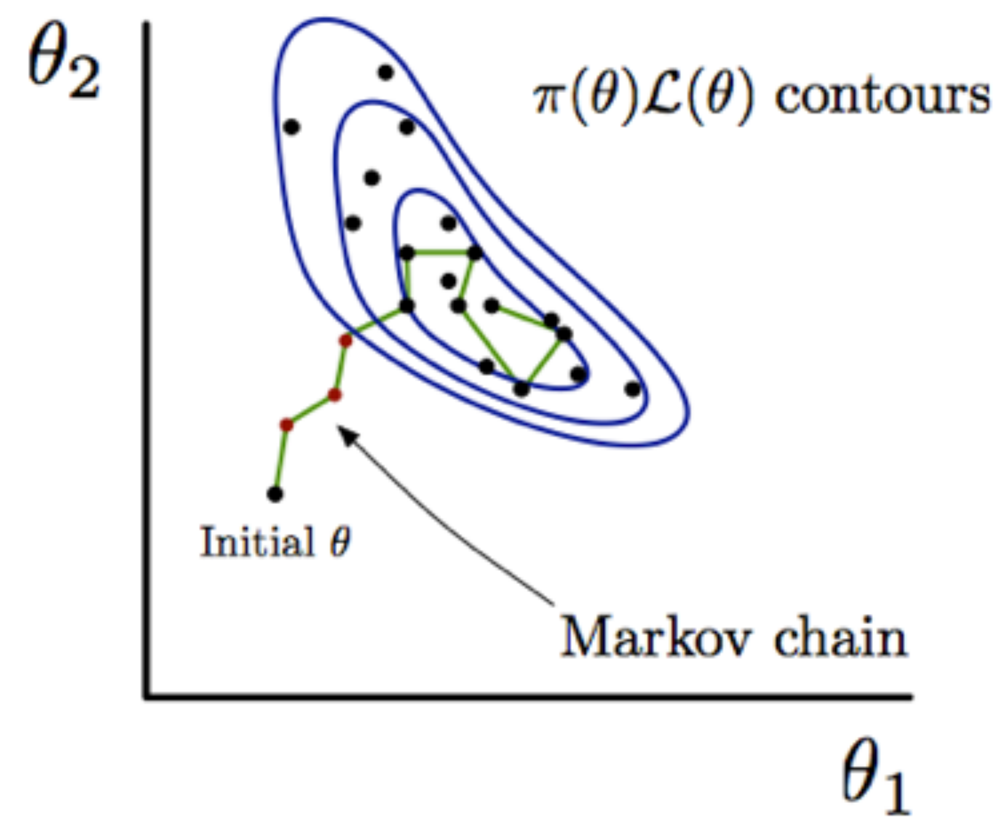
Posterior samples: Markov Chain Monte Carlo

- Von Neumann rejection can still be very inefficient for most problems, so Markov Chain Monte Carlo (MCMC) is preferred

- The idea of MCMC is to start from a point and explore the parameter space by taking steps that may be accepted or rejected, such that the Markov chain:

    - Tends to walk towards higher probability areas

    - Tends to avoid low probability areas



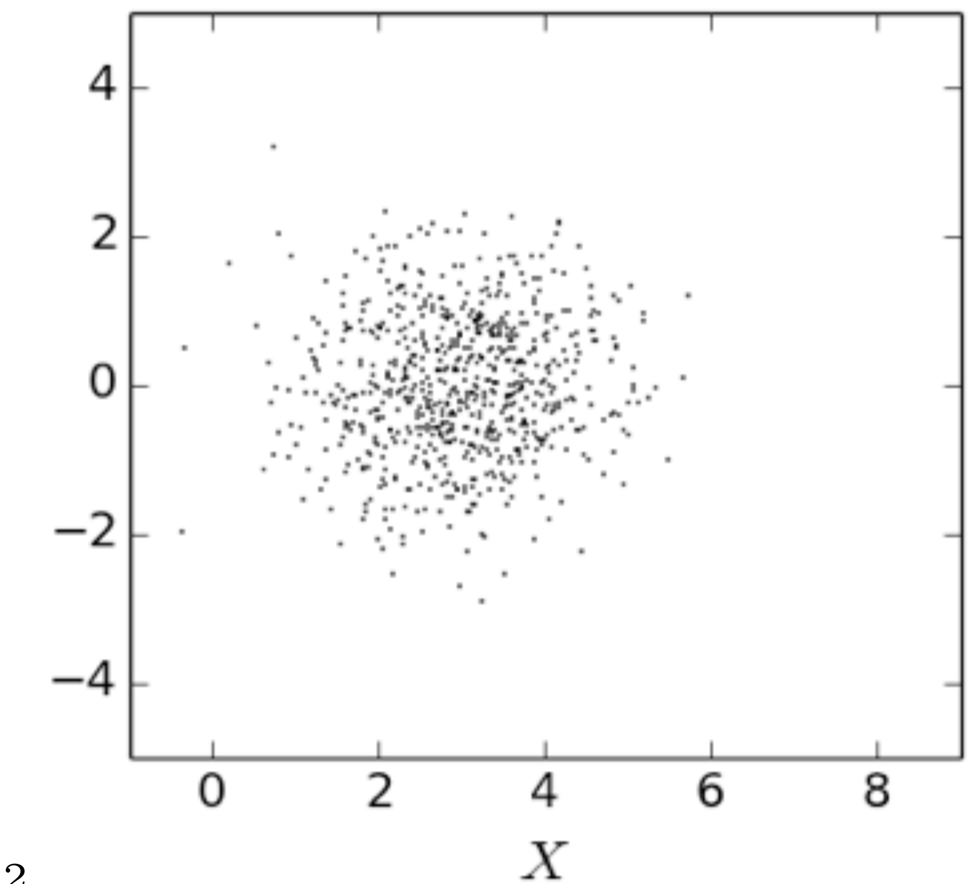
(from Tom Loredo's Lecture Notes)

- Note that the samples are not completely independent, there is some correlation

- After a while, the chain 'forgets' the initial conditions and the accepted (independent) samples have a PDF that is proportional to the Posterior

A two-parameter problem

- We observe the following distribution of N pairs (xi,yi)



- It seems reasonable to assume they were drawn from a random distribution, so lets use a gaussian model with known σx=σy=1 and μX,μY the unknown means in the X and Y directions

- The likelihood is expressed as

$$P(\{x_i, y_i\}|\mu_X, \mu_Y) = \prod_{i=1}^{N} e^{\frac{1}{2}[(x_i - \mu_X)^2 + (y_i - \mu_Y)^2]}$$

- Assuming a uniform prior probability for μX,μY, the posterior is therefore given by

$$P(\mu_X, \mu_Y|\{x_i, y_i\}) = \prod_{i=1}^{N} e^{\frac{1}{2}[(x_i - \mu_X)^2 + (y_i - \mu_Y)^2]}$$

MCMC: The Metropolis-Hastings Recipe

from Hogg et al. 2010

- 1-Choose an initial position for the model params $\{\mu_i\}$
- 2-Advance a step (in one parameter) randomly -> μ_{i+1}
- 3-Evaluate the posterior at current position $P(\mu_{i+1})$
- 4-Draw a random number R with uniform probability in the range $0<R<1$
  - If $R<P(\mu_{i+1})/P(\mu_i)$, keep the point and add it to the chain
  - if not, go back to the previous step and re-add it to the chain
  - repeat …
  - **the set of $\{\mu_i\}$ obtained is a random realization of the Posterior !**

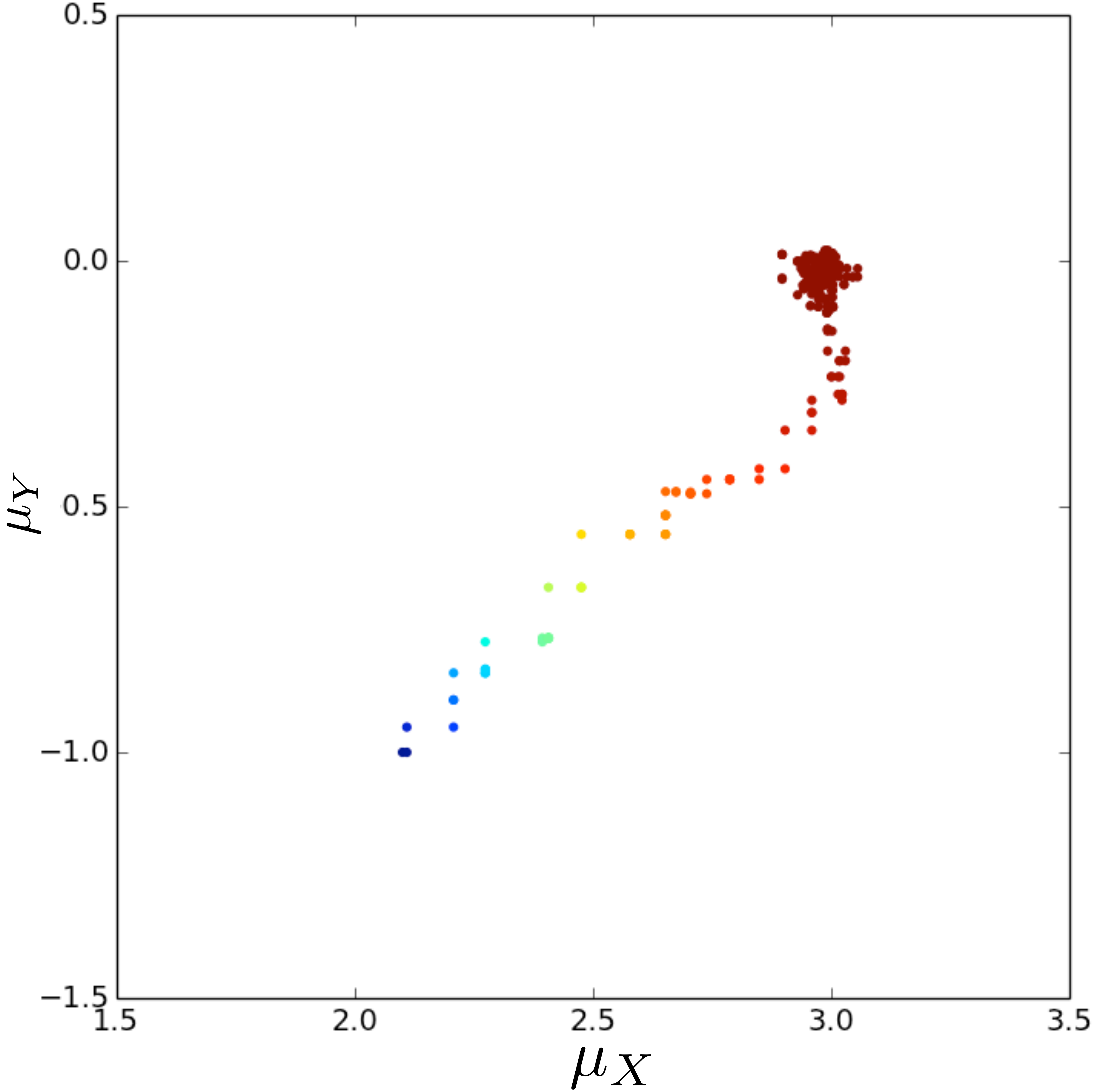
MCMC: The Metropolis-Hastings Recipe

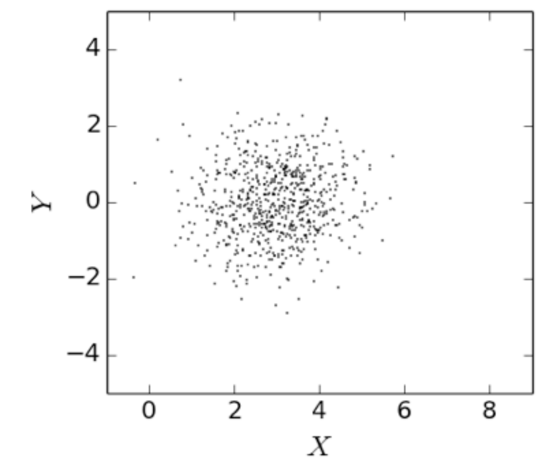
- The step size must be chosen so that the acceptance fraction (fraction of points accepted in the chain) lies between ~0.2 and ~0.5 (see Hogg et. al. 2010 and Foreman-Mackey et al. 2013)

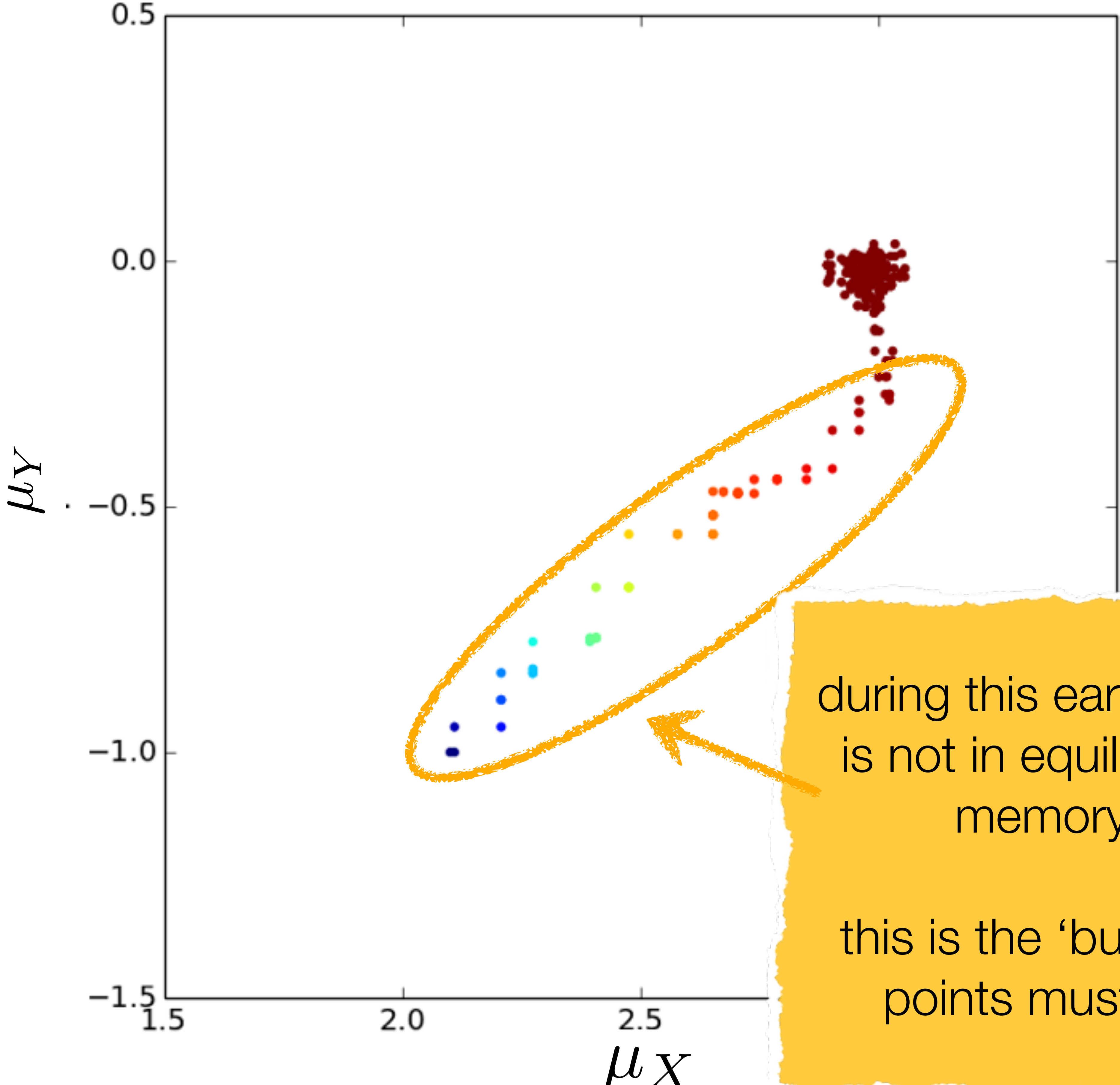
- This algorithm is a piece of cake to write, excellent for playing around to develop some intuition as to how the MCMC works

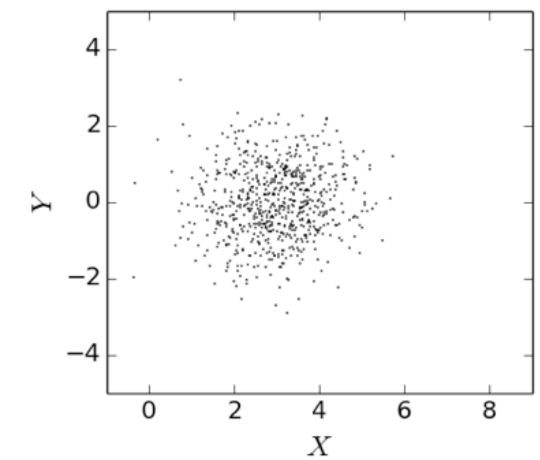
- The problem is that fine-tunning the chain when the number of parameters is large is highly non-trivial! (there's no way of guessing it a priori)

- This is solved by MCMC implementations like ***emcee*** (in Python, Foreman-Mackey et al. 2013) that use algorithms more sophisticated than Metropolis-Hastings, with very few free parameters


0.5
0.0
0.5
−1.0
−1.5
1.5
2.0
2.5
3.0
3.5
μ_Y
μ_X


4
2
0
−2
−4
0
2
4
6
8
Y
X



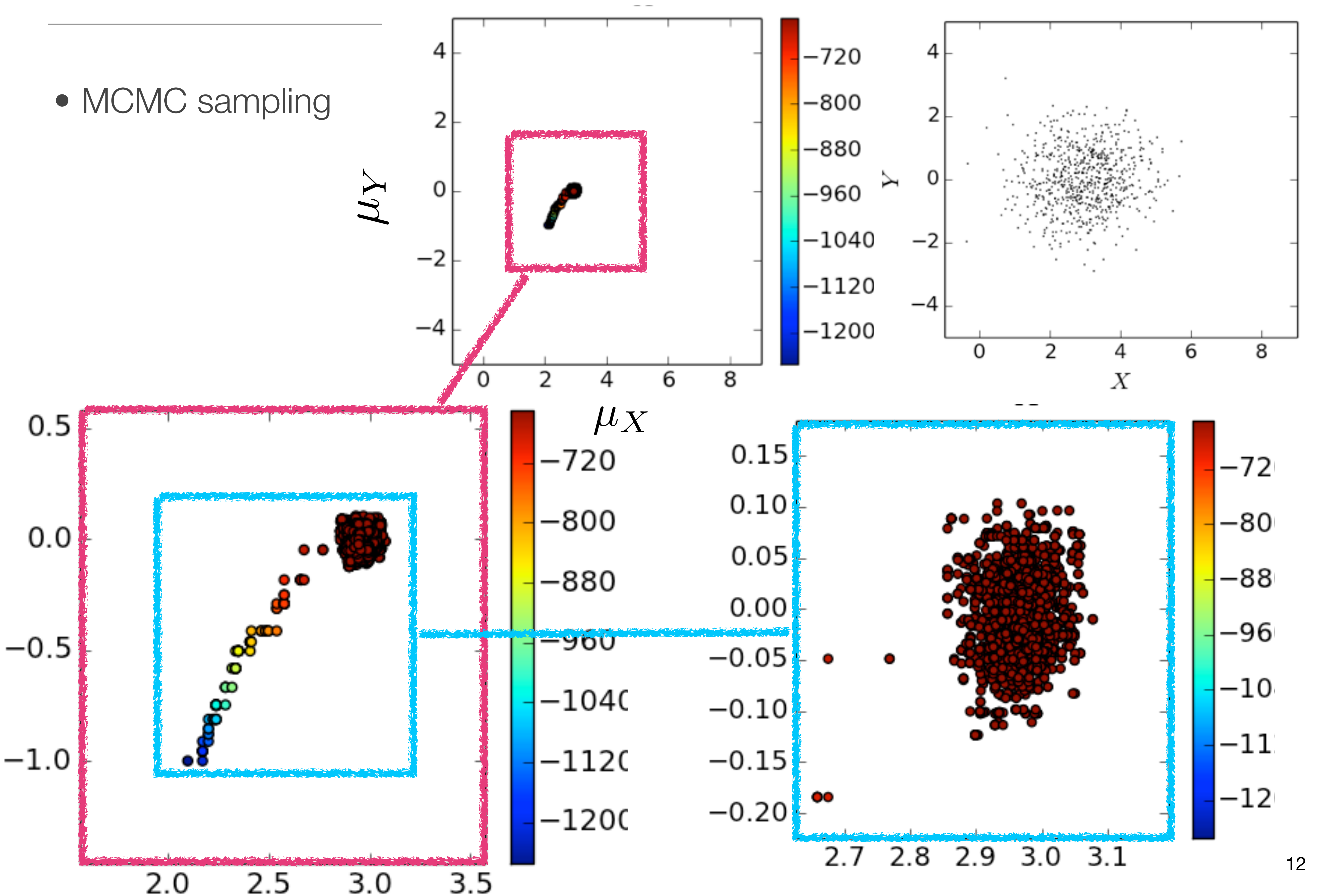


during this early stage the chain is not in equilibrium yet, it has memory of its path

this is the ‘burn-in’ stage, this points must be discarded

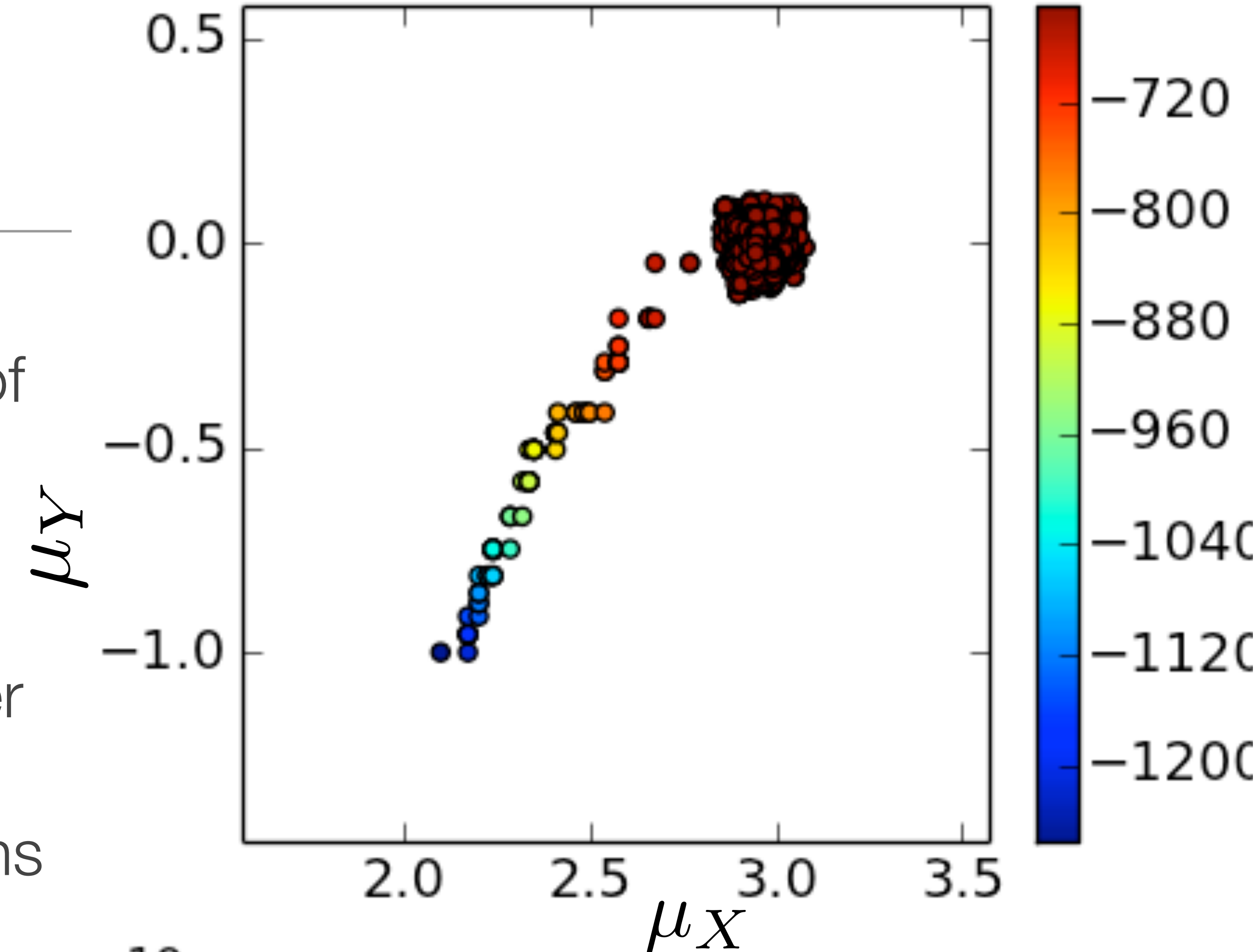
A two-parameter problem

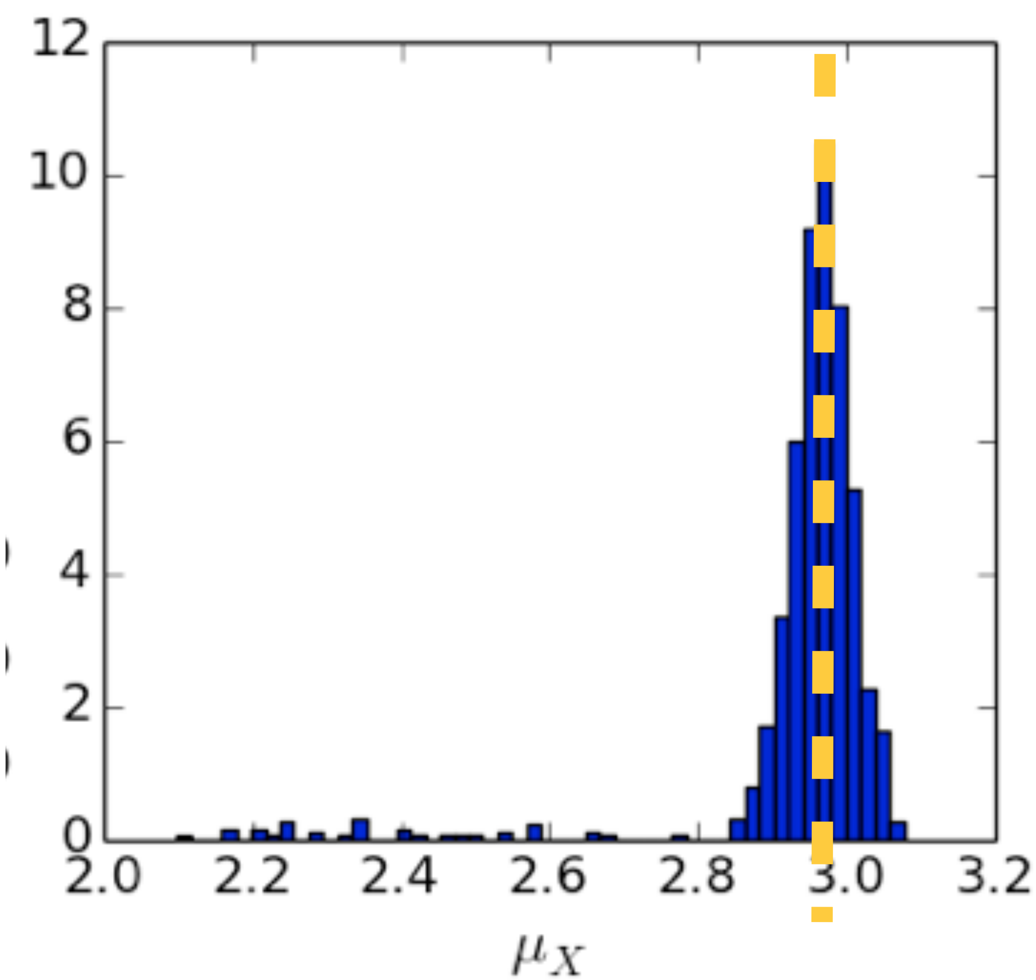
- MCMC sampling

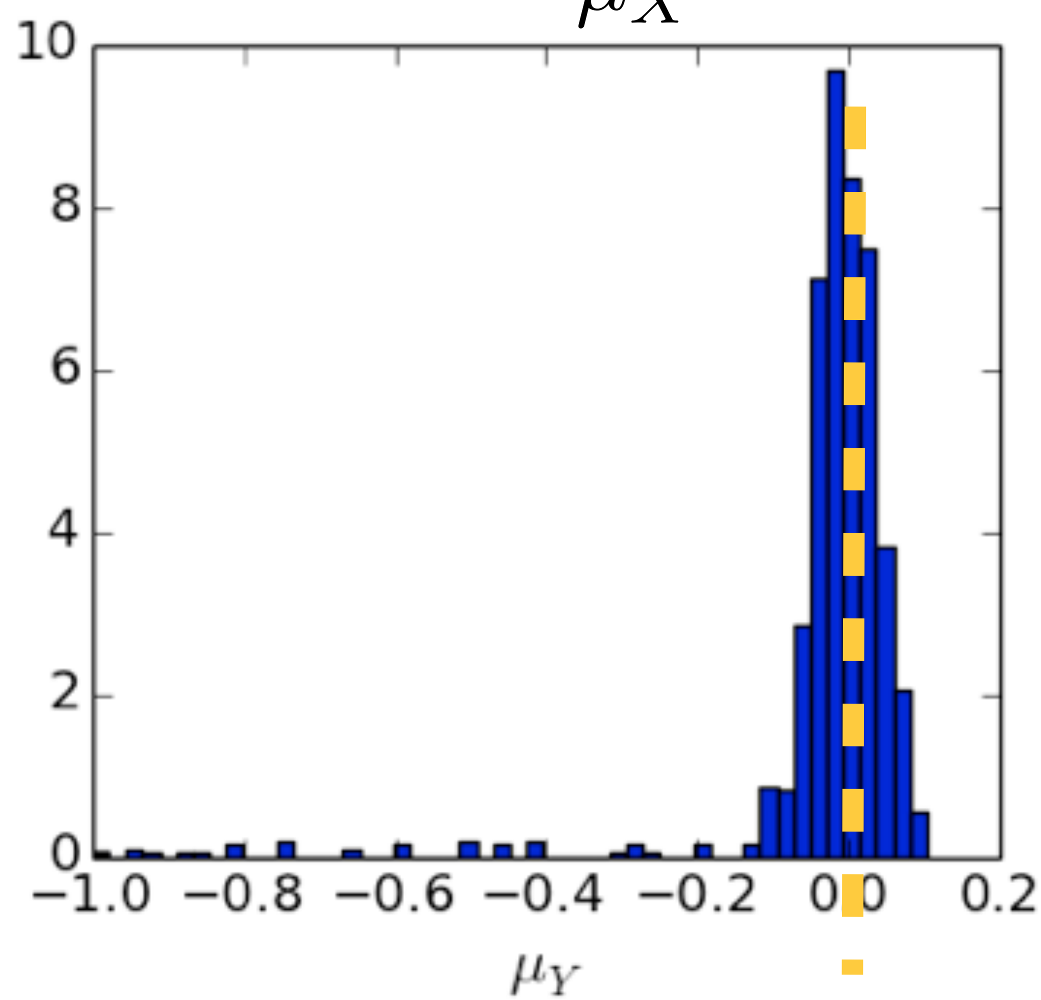


MCMC samples

- The set of points obtained in the final Markov Chain is a random realization of the Posterior PDF
- The mode of the histogram gives the most probable value of each parameter
- The percentiles give the credible regions

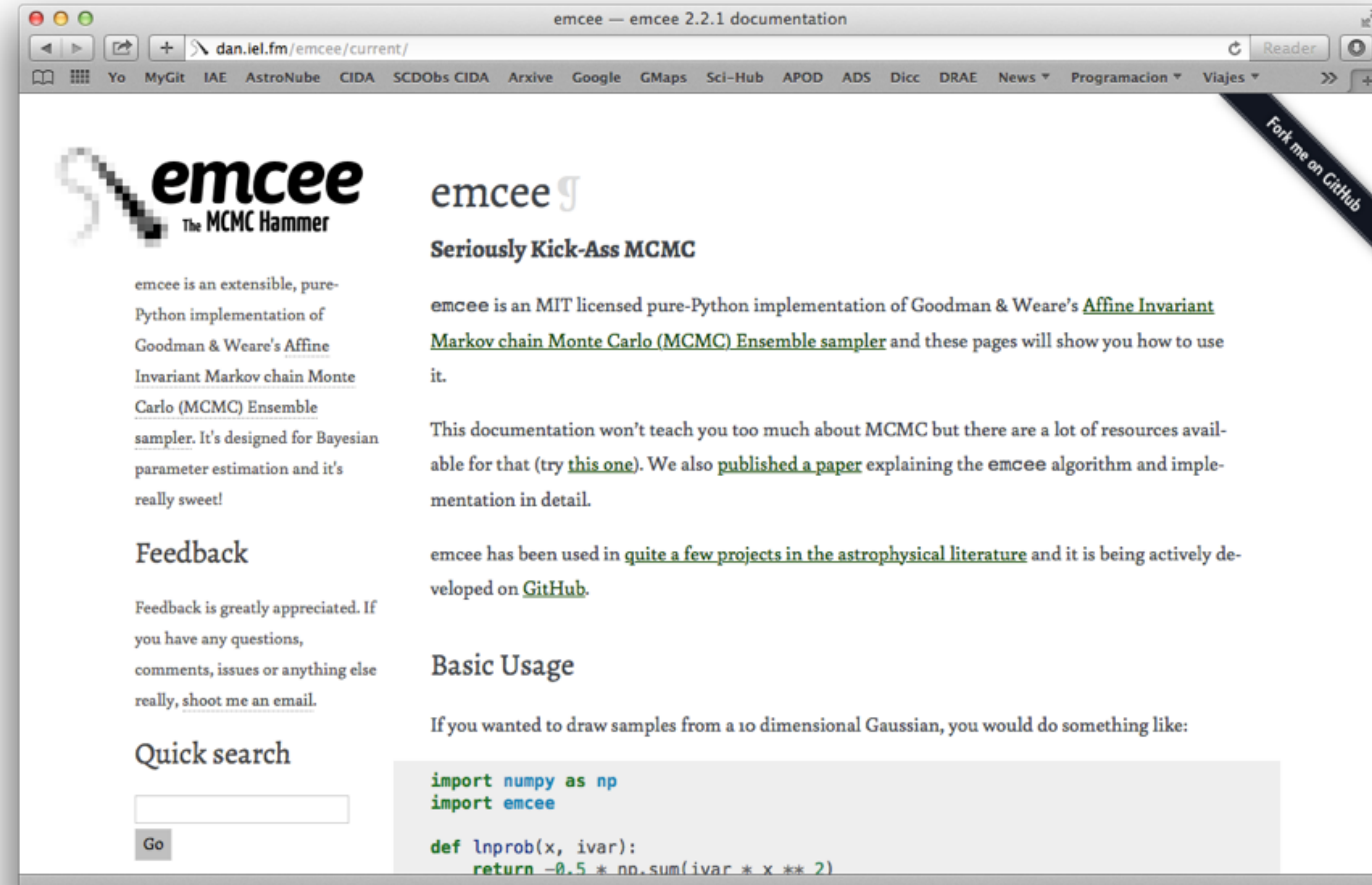






Suggested MCMC sampler: **emcee** (Python)

- Foreman-Mackey et al. 2013



More Suggested Bibliography

- Hogg, Bovy & Lang (2010)

Data analysis recipes:
Fitting a model to data*

David W. Hogg
Center for Cosmology and Particle Physics, Department of Physics, New York University
Max-Planck-Institut für Astronomie, Heidelberg

Jo Bovy
Center for Cosmology and Particle Physics, Department of Physics, New York University

Dustin Lang
Department of Computer Science, University of Toronto
Princeton University Observatory

Approximate Bayesian Computation (ABC)

Approximate Bayesian Computation (ABC)

- Option for cases where there's no analytic likelihood, but there is enough knowledge about the problem to do forward modelling

Basic ABC algorithm

For the observed data $y_{1:n}$, prior $\pi(\theta)$ and distance function ρ:

Algorithm*

1. Sample θ^* from prior $\pi(\theta)$
2. Generate $x_{1:n}$ from forward process $f(y \mid \theta^*)$
3. Accept θ^* if $\rho(y_{1:n}, x_{1:n}) < \epsilon$
4. Return to step 1

Generates a sample from an approximation of the posterior:

$$f(x_{1:n} \mid \rho(y_{1:n}, x_{1:n}, \theta) < \epsilon) \cdot \pi(\theta) \approx f(y_{1:n} \mid \theta)\pi(\theta) \propto \pi(\theta \mid y_{1:n})$$

*Introduced in Pritchard et al. (1999) (population genetics)

"Though there be no such thing as Chance in the world; our ignorance of the real cause of any event has the same influence on the understanding"

-David Hume (1748)