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

Cecilia Mateu J.
Instituto de Astronomía, UNAM, Ensenada

Universidad de Barcelona
6 de noviembre de 2014

More Examples

Coordinate Transformations

- Lets say we have a problem in which we know how to write our posterior P(x) in a variable x, but in reality we are interested in knowing the posterior P(y) for variable y=f(x)

- We know the integral of the probability must be conserved

$$\int P(x)dx = \int P(y)dy \qquad\qquad P(x)dx = P(y)dy$$

- This makes it easy, we know how to change variables inside an integral

$$P(y) = \left|\frac{dx}{dy}\right| P(x)$$

Jacobian of the transformation f(x)

The Parallax Example

- Lets illustrate this with an example. Say we have N measurements of the parallax for a star from Gaia

- Lets assume we have an error model and CU(?)/DPAC gives us an estimate on the parallax error and tells us that its safe to assume these errors are gaussian

- What we’d really like to get ultimately is the posterior on the distance to the star

The Parallax Example

- Lets write the posterior on the parallax first

$$P(\varpi|\{\varpi_i\}) = P(\{\varpi_i\}|\varpi)P(\varpi) \qquad |I$$

- Our likelihood is

$$P(\{\varpi_i\}|\varpi) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}}$$

- So the posterior is

$$P(\varpi|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}} P(\varpi) \qquad |I$$

no coordinate transformations yet...

The Parallax Example

- Now we want the posterior for the distance D=1/ϖ

$$P(\varpi|\{\varpi_i\})d\varpi = P(D|\{\varpi_i\})dD$$

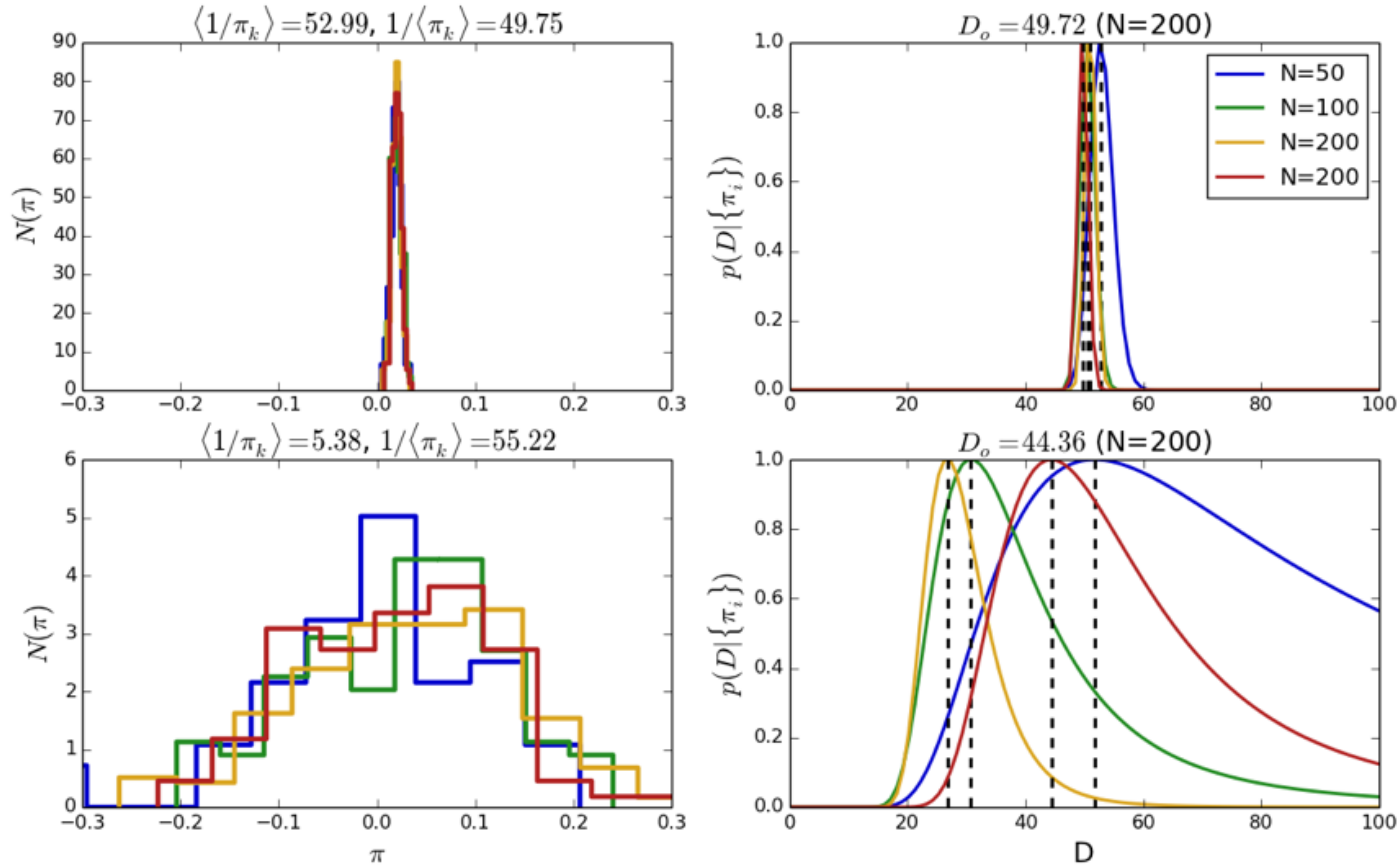
$$P(D|\{\varpi_i\}) = P(\varpi|\{\varpi_i\})\left|\frac{d\varpi}{dD}\right|$$

$$P(\varpi|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(\varpi-\varpi_i)^2}{2\sigma_\varpi^2}} P(\varpi)$$

$$P(D|\{\varpi_i\}) = P(\varpi|\{\varpi_i\})\frac{1}{D^2}$$

$$P(D|\{\varpi_i\}) = \prod_{i=1}^{N} e^{-\frac{(1/D-\varpi_i)^2}{2\sigma_\varpi^2}} \frac{1}{D^2} P(\varpi(D))$$

The Parallax Example



More Examples

Fitting a density profile to data

- Another example. Lets say you have a catalogue of a particular stellar tracer, e.g. RR Lyrae stars, Red Clump stars, Cefeids or whatever.
- We'd like to use these tracers to study how stars are distributed in space in the different components in the Galaxy, i.e. their density profile ρ
- How do we compare and fit this function to a sample of stars?

Thick Disc and Halo RRL Density Profiles

- Thick Disc density profile

$$\rho_{\rm DG} = C_{\rm DG}\, e^{-\frac{R-R}{h_R}}\, e^{-\frac{|z|}{h_z}}$$

- Halo density profile

$$\rho_{\rm H} = \frac{C_{\rm H}}{R_\odot^n}\left[R^2 + \left(\frac{z}{q}\right)^2\right]^{n/2}$$

... but the directly observable quantity is not the density ρ, but the number N_{RR} of RRLs in the survey volume V_S

$$N_{RR} = \iiint_{V_S} \rho(\vec{r})dV \;= \iiint_{V_S} [\rho_{\rm H}(R,z) + \rho_{\rm DG}(R,z)]\, R\, dR\, dz\, d\varphi$$

Density Profiles: A Bayesian approach

- Our free parameters are:

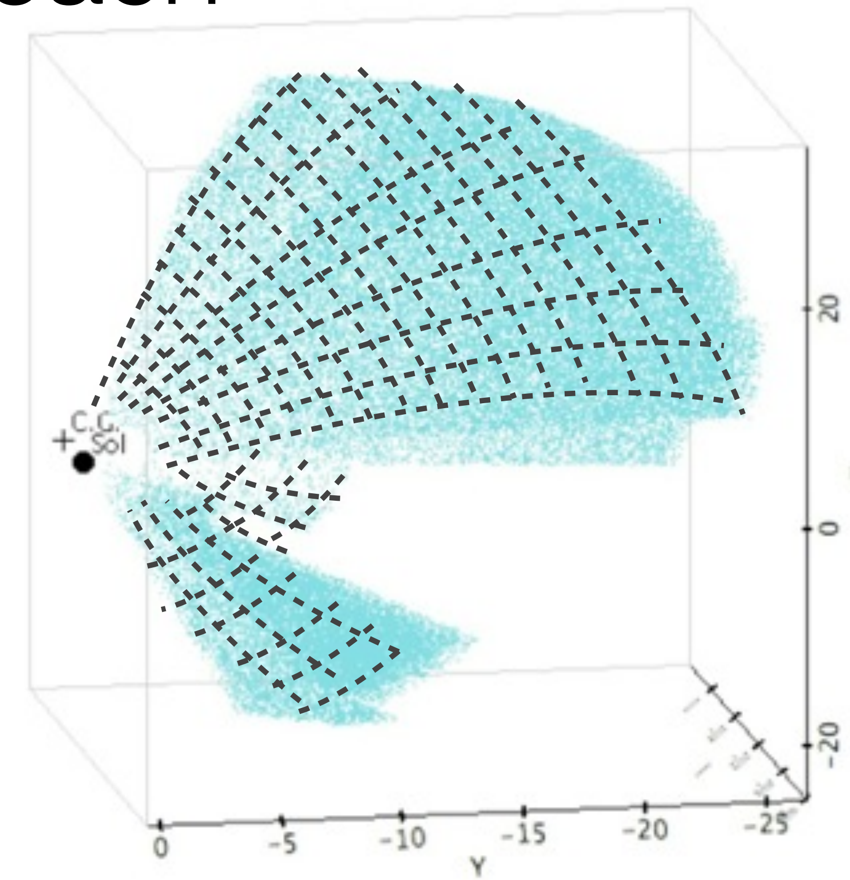
$$\vec{\theta} = (h_z, h_r, C_{tkd}, n, C_h)$$

- We build an imaginary grid restricted *only* to the survey volume Vs

- Our likelihood function is then

$$L \equiv p(\{\eta\}|\vec{\theta}) = \prod_{i \in V_S} p(\eta_i|\vec{\theta}) = \prod_{i \in V_S} \frac{\mu_i^{\eta_i} e^{-\mu_i}}{\eta_i!}$$

$$\ln L = \sum_{i \in V_S} \eta_i \ln \mu_i - \mu_i$$



η_i - ***observed*** number or RRLS on i-th bin

μ_i - ***predicted*** number or RRLS on i-th bin

- If now we make the grid cell size tend to 0

$$\mu_i(\vec{\theta}, \vec{r}_i^{RRLS}) \rightarrow \rho(\vec{\theta}, \vec{r}_i^{RRLS})$$

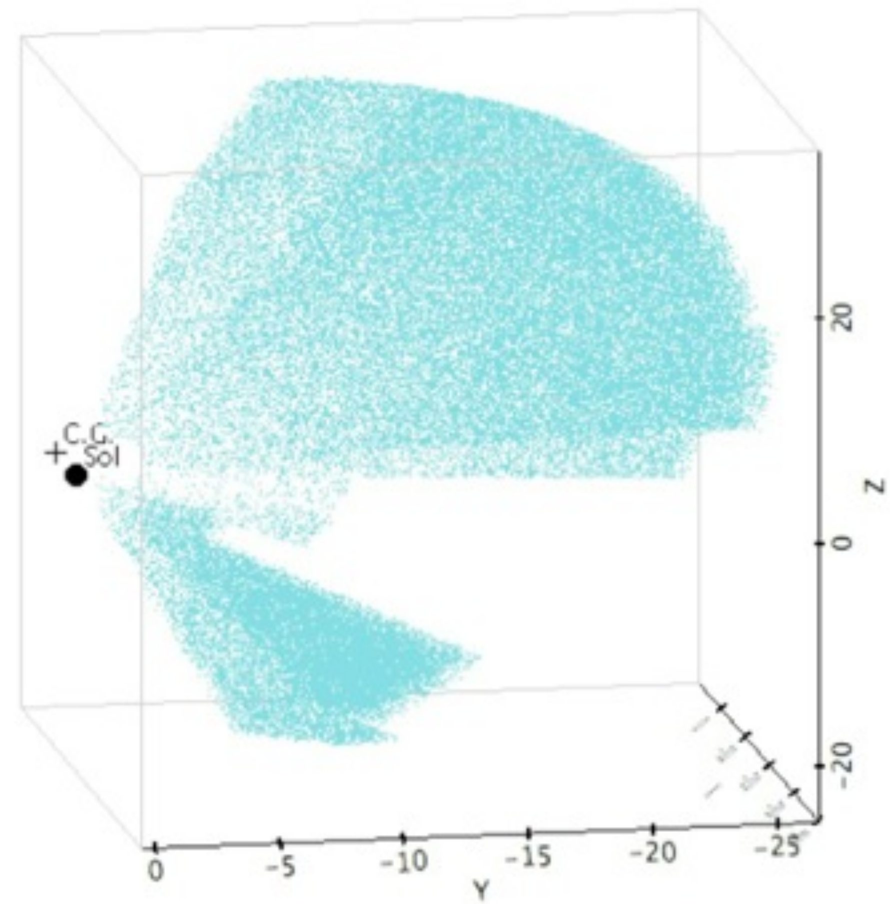
- We finally get

$$\ln L = \sum_{i=1}^{N_{obs}^{RRL}} \ln \rho(\vec{\theta}, \vec{r}_i^{RRLS}) - N_{model}^{RRL}(\vec{\theta})$$

Lombardi et al. (2013)

Density Profiles: A Bayesian approach

- This framework allows us to account for the inhomogeneities of the survey volume due to the variable extinction
- We could also include an incompleteness function for example

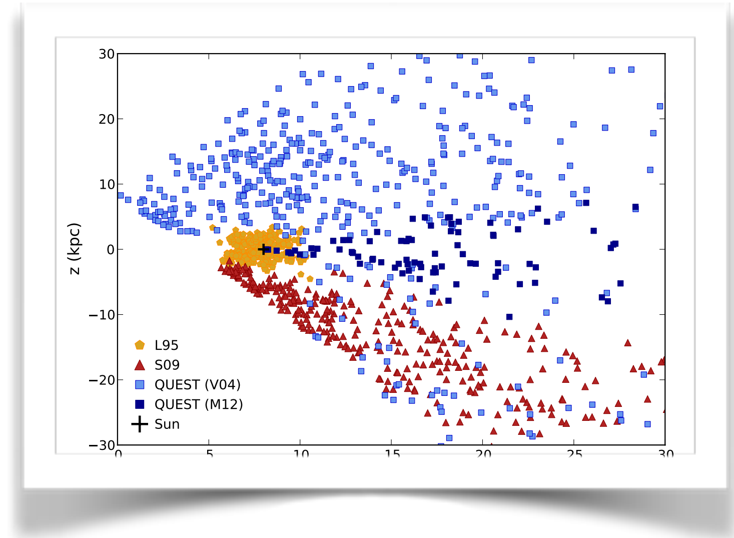


$$\ln L = \sum_{i=1}^{N_{obs}^{RRL}} \ln \rho(\vec{\theta}, \vec{r}_i^{RRLS}) - N_{model}^{RRL}(\vec{\theta})$$

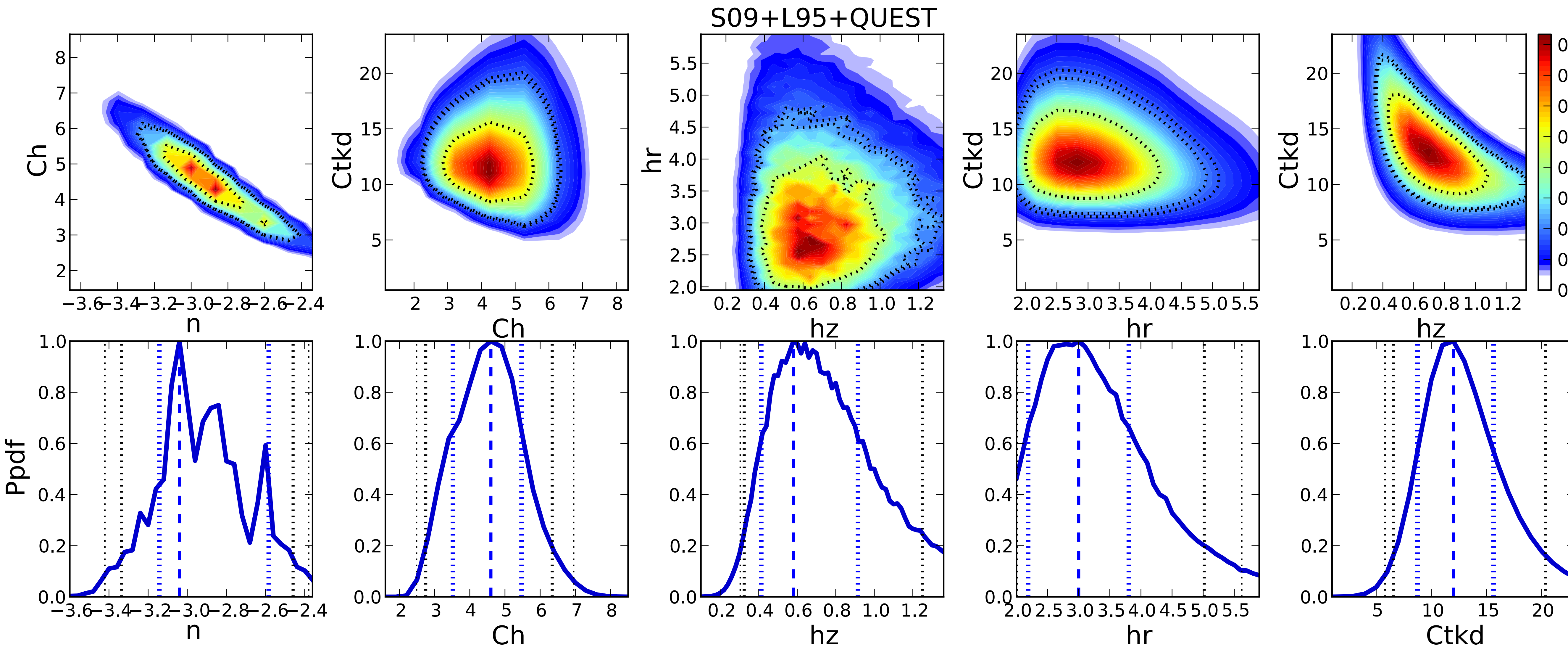
This is the computationally intensive part

$$N_{RR} = \iiint\limits_{V_S} \rho(\vec{r}) dV = \iiint_{V_S} [\rho_{\rm H}(R, z) + \rho_{\rm DG}(R, z)] R \, dR \, dz \, d\varphi$$

Density Profiles: Combined samples



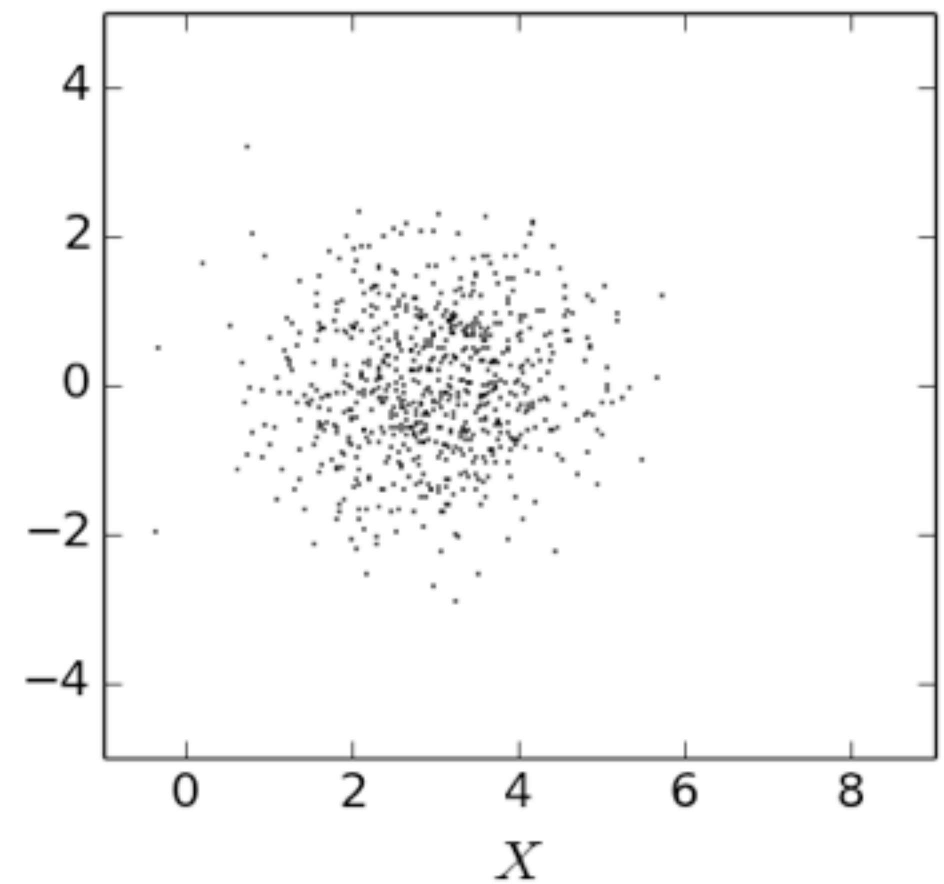
Combining three different samples we find the following parameters for the Halo and normal Thick Disk:



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

A two-parameter problem

- N Points in 2D
- 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 μ_X,μ_Y the unknown means in the X and Y directions

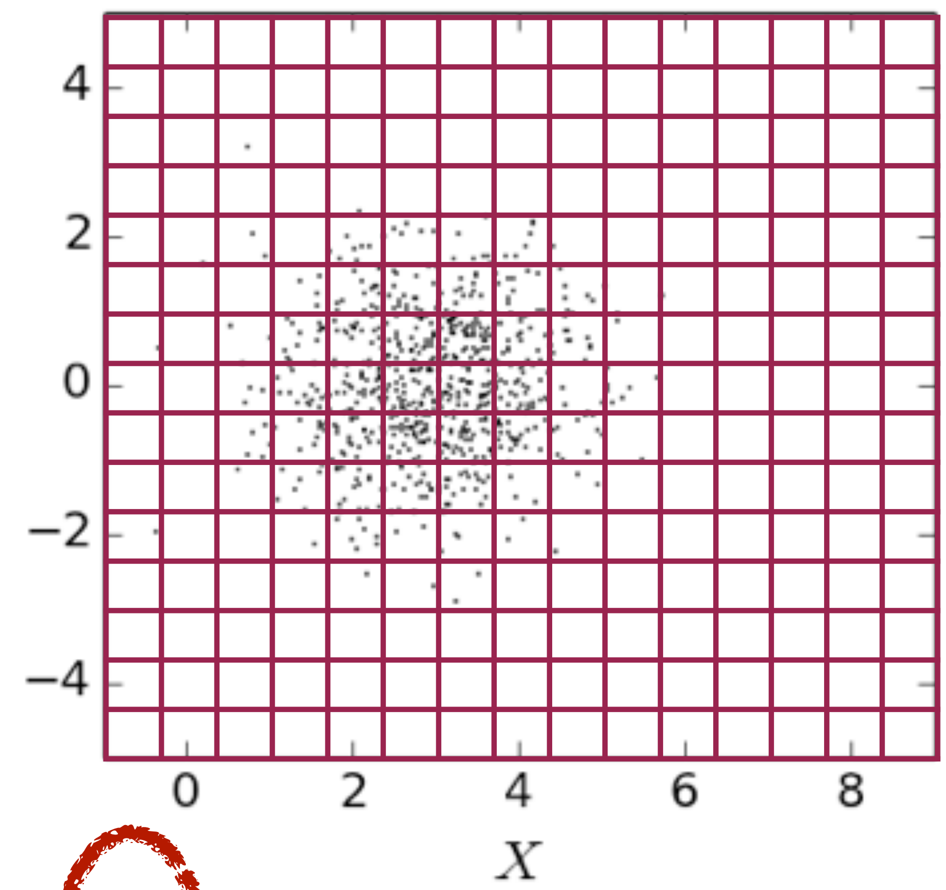


- The likelihood is expressed as

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

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

A two-parameter problem



- lets assume a uniform prior probability for μ_X,μ_Y
- The posterior is therefore given by

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

- The Posterior is a function of μ_X,μ_Y , we want to explore the parameter space efficiently, i.e. spend more time computing the Posterior in high probability areas than in low probability ones
- This is called Importance Sampling and a way of doing it is Markov Chain Monte Carlo sampling

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 $R<P(\mu_{i+1})$, 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 !**

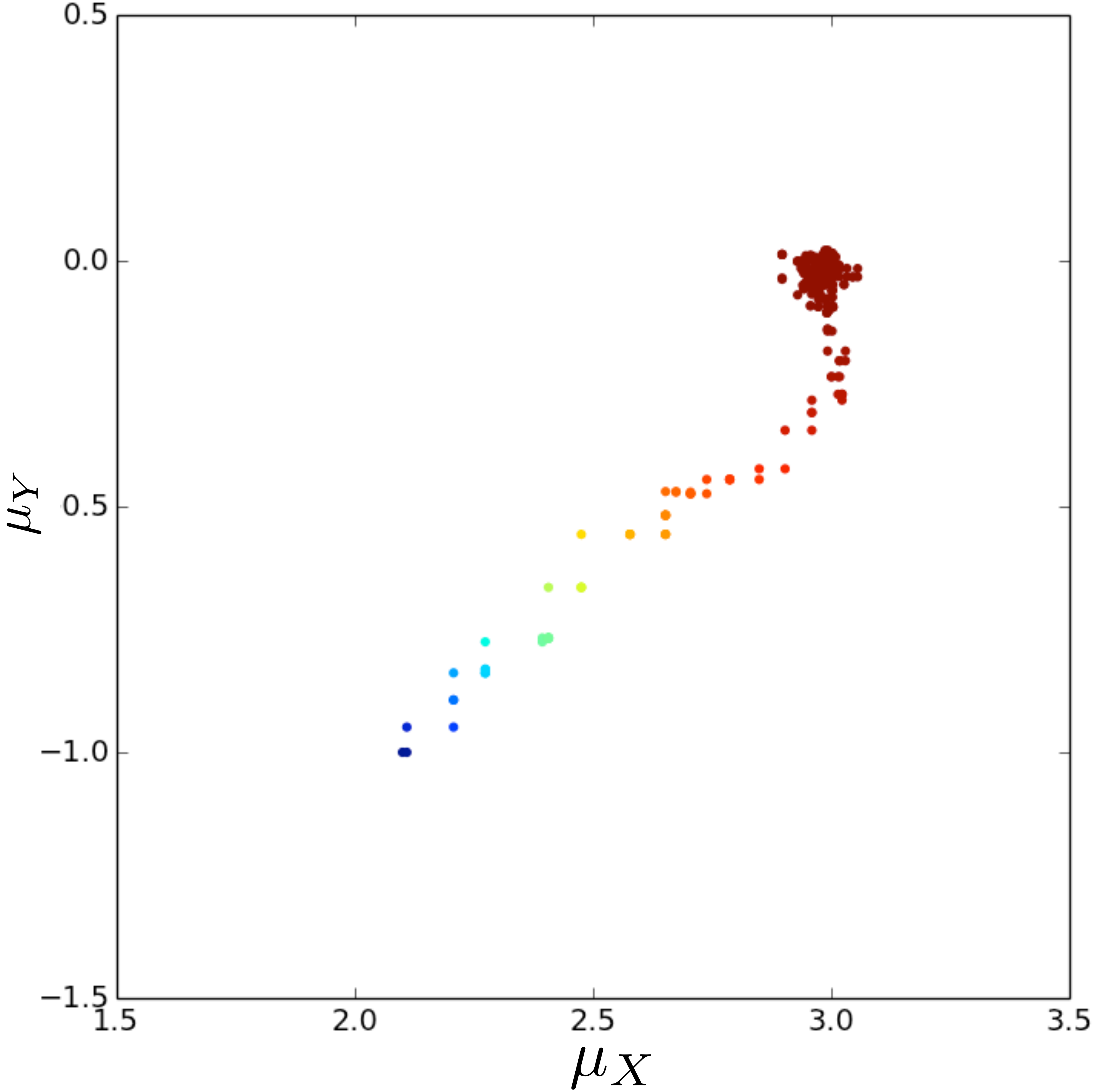
MCMC: The Metropolis-Hastings Recipe

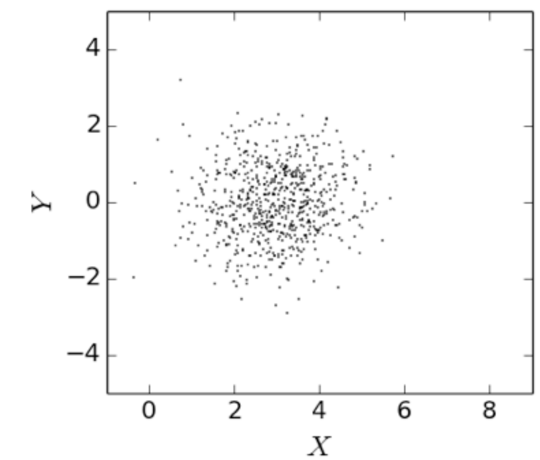
- The step size must be chosen so that the acceptance fraction (fraction of points accepted in the chain) lies between ~0.2 and ~0.5 (see Hogg et. al. 2010 and Foreman-Mackey et al. 2013)

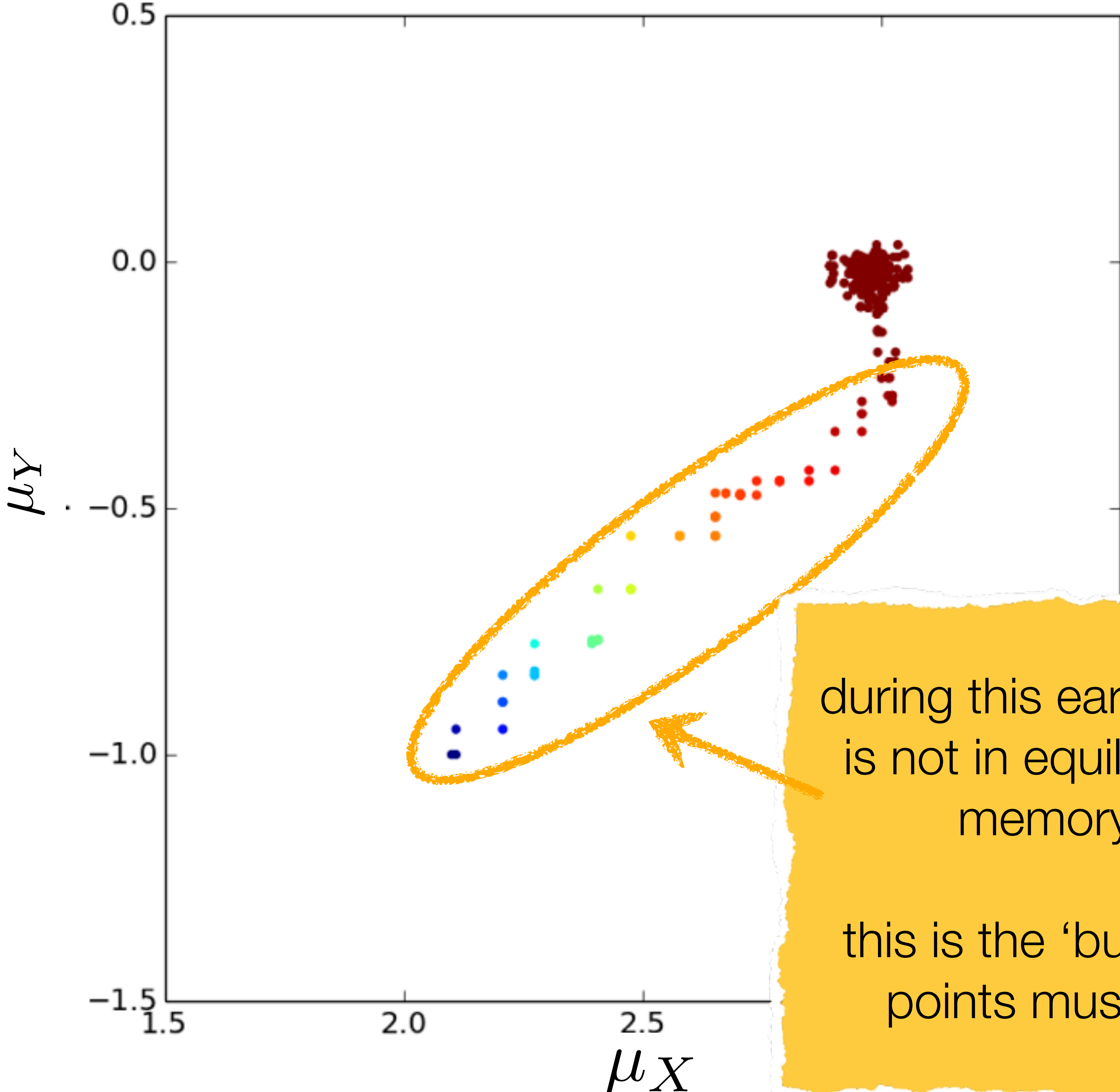
- This algorithm is a piece of cake to write, excellent for playing around to develop some intuition as to how the MCMC works

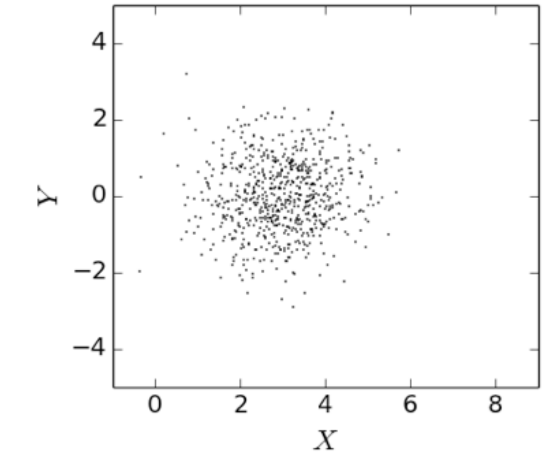
- The problem is that fine-tunning the chain when the number of parameters is large is highly non-trivial! (there's no way of guessing it a priori)

- This is solved by MCMC implementations like ***emcee*** (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



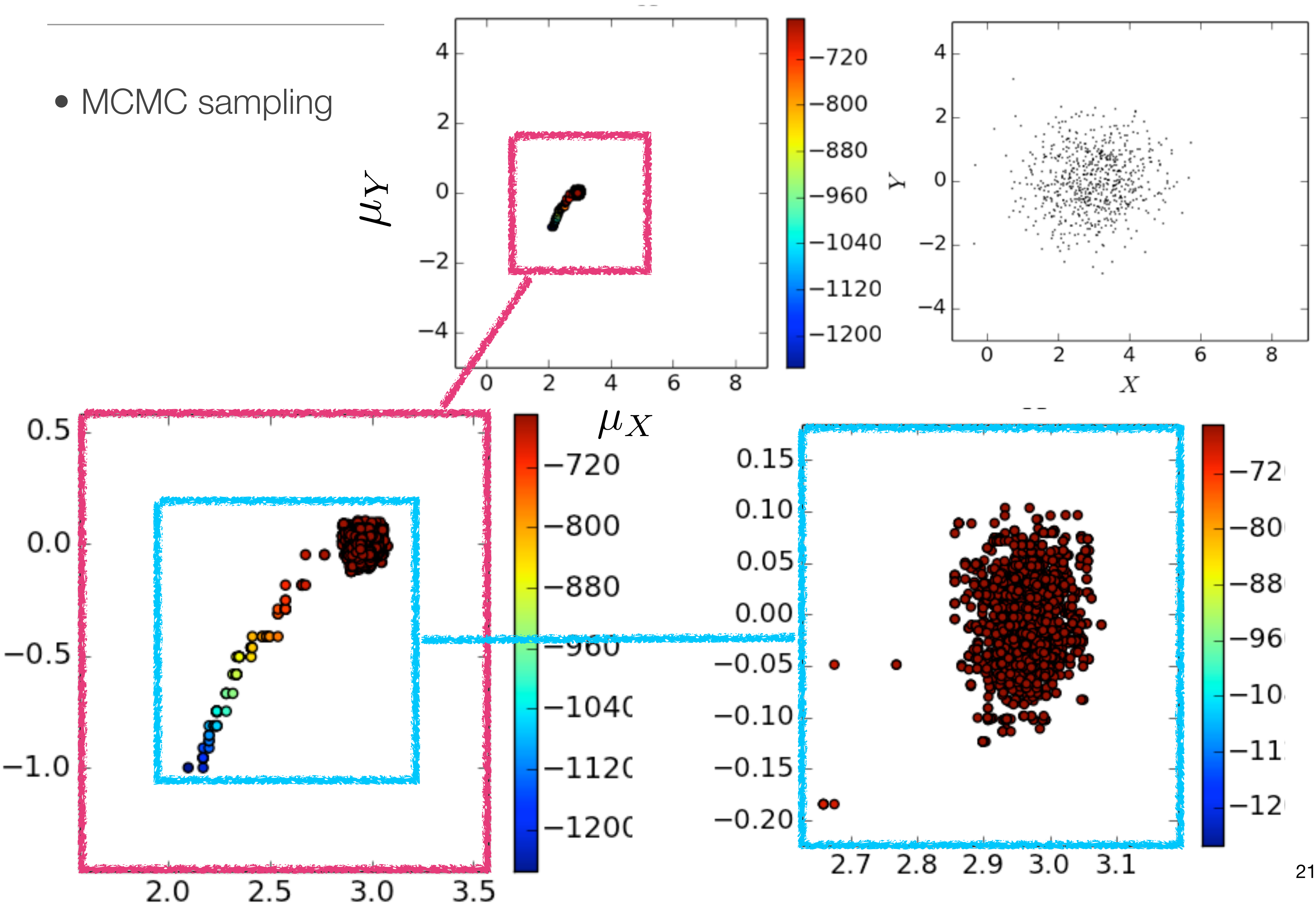


during this early stage the chain is not in equilibrium yet, it has memory of its path

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

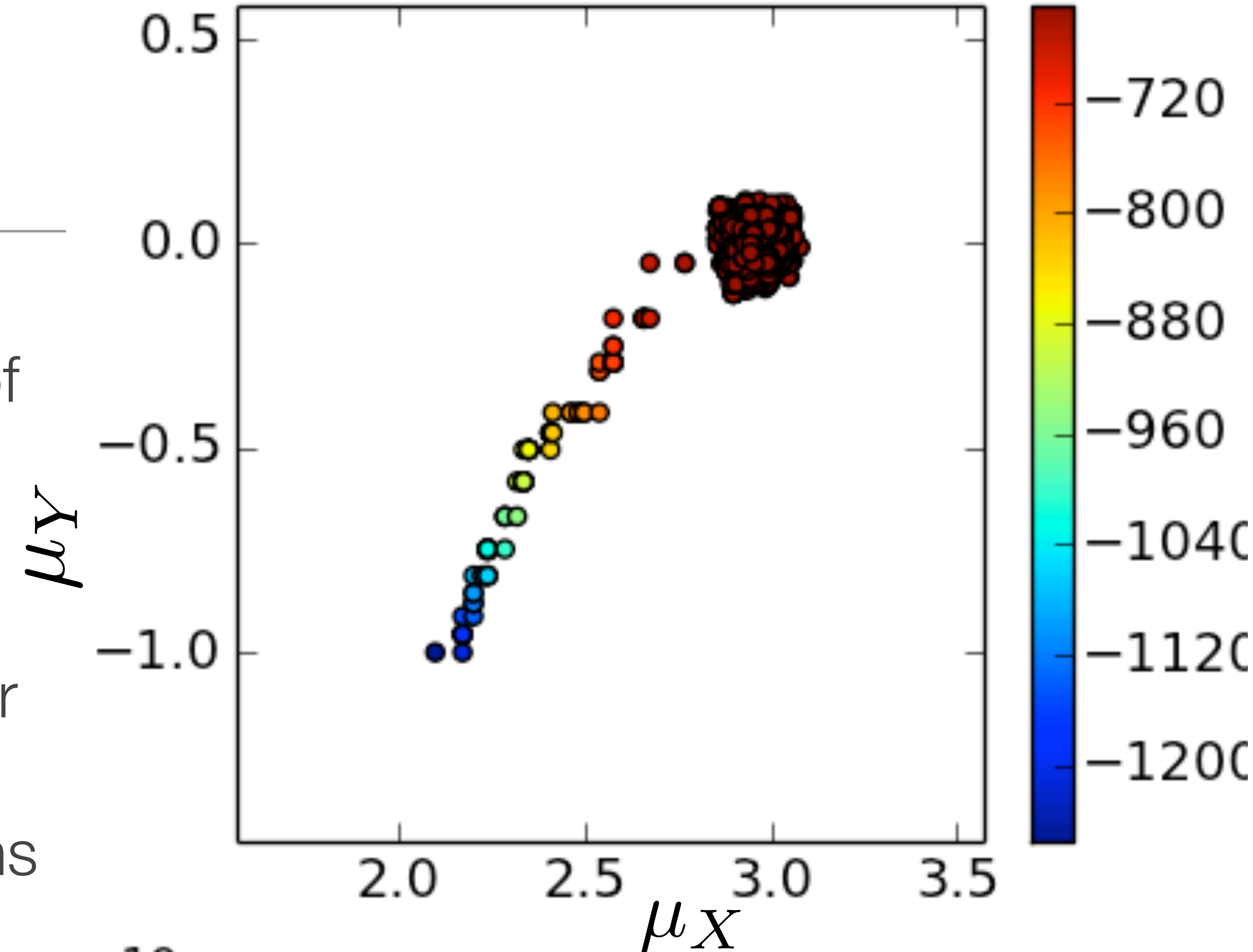
A two-parameter problem

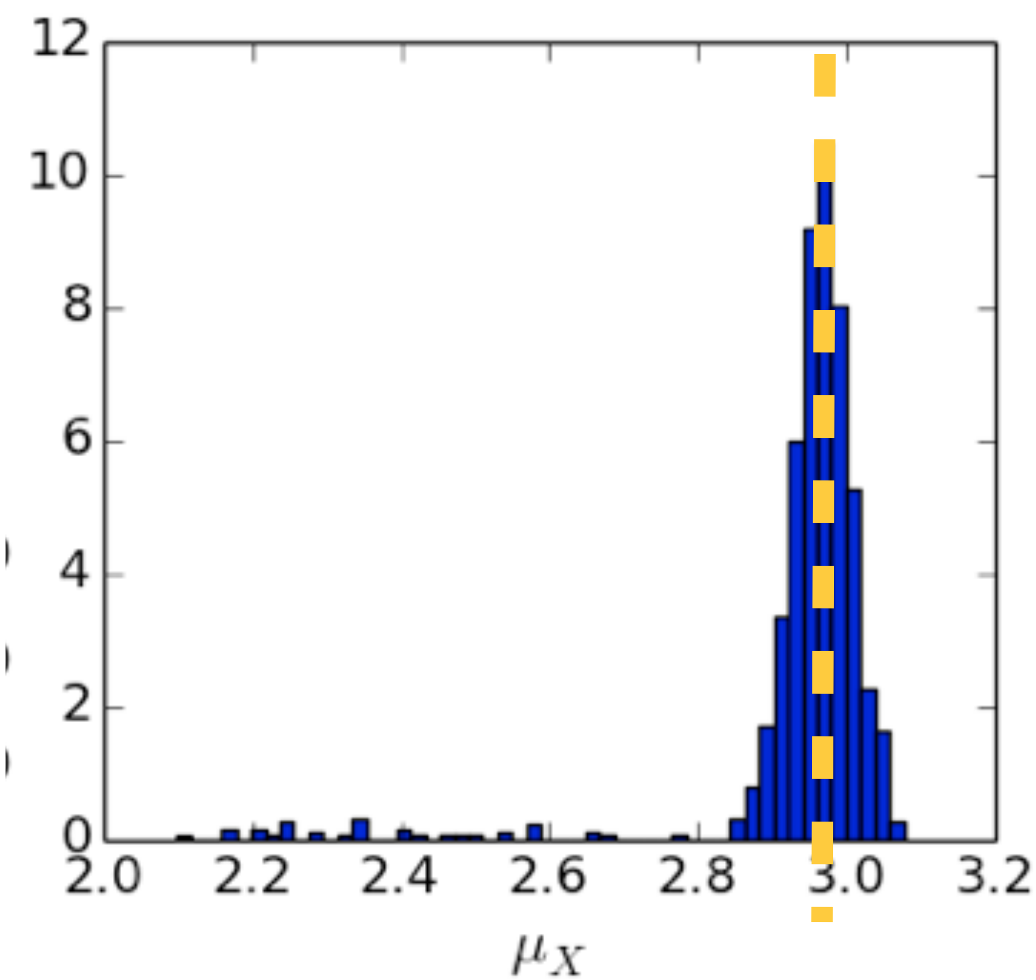
- MCMC sampling

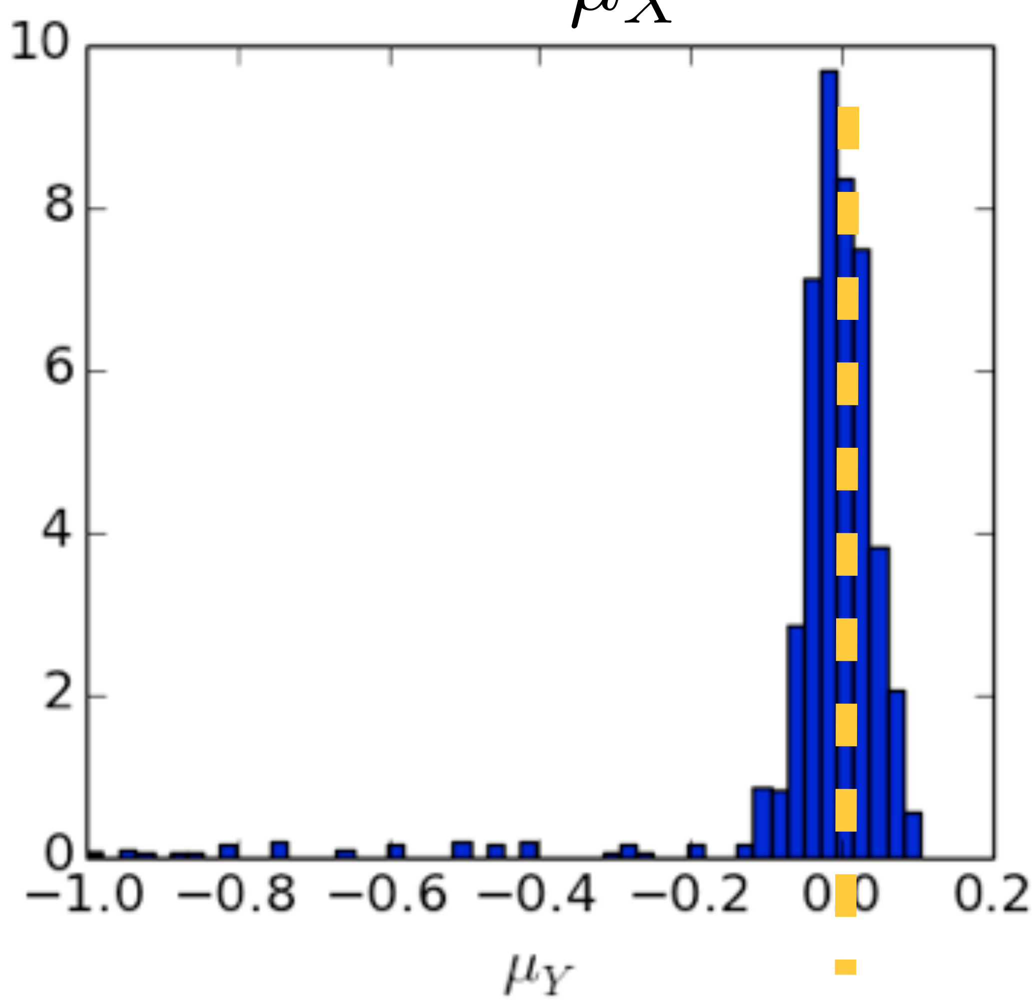


MCMC samples

- 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







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)