Astrostatistics: Friday 22 Feb 2019

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2019

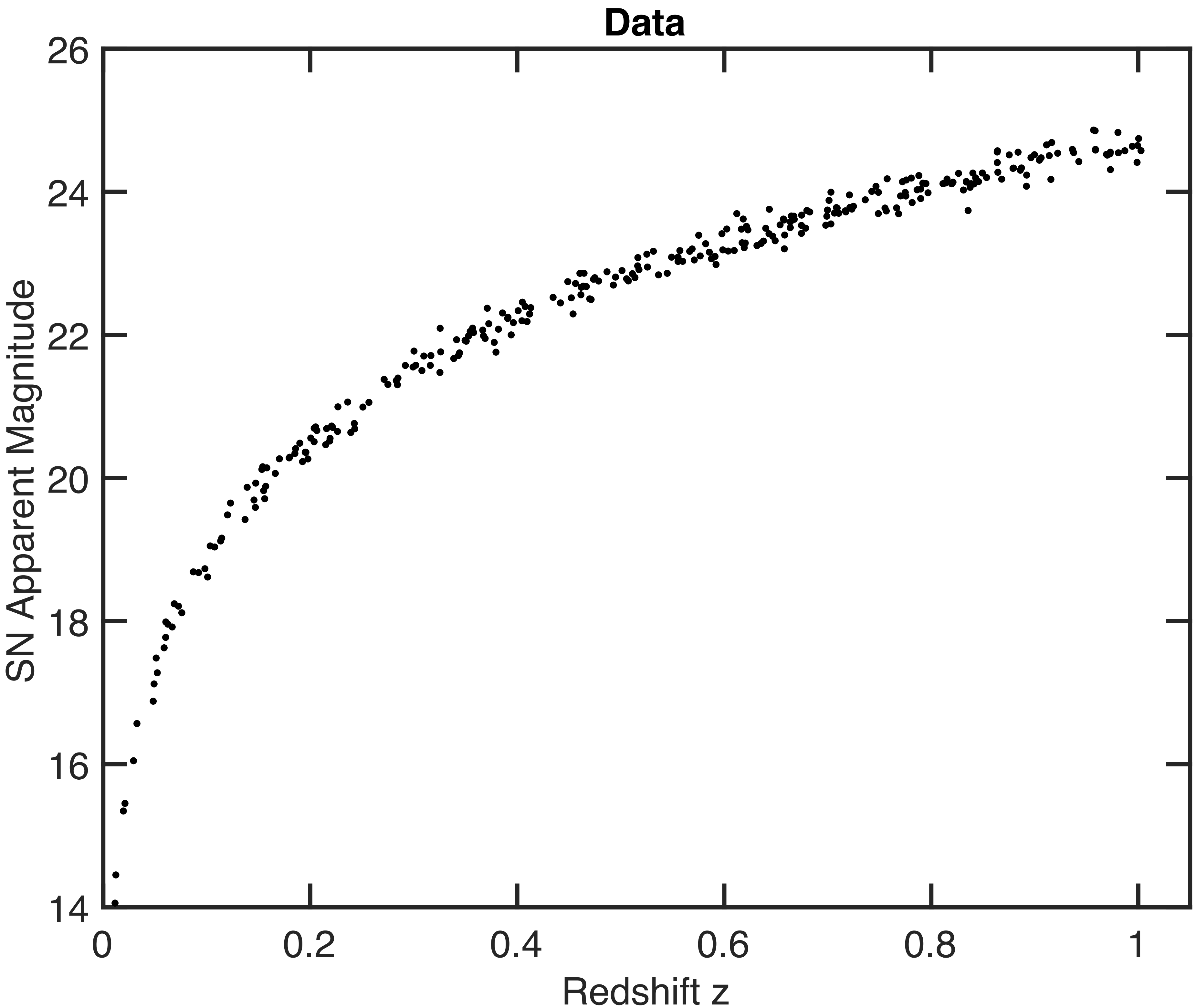
- Next example sheets soon
    - For 5 Mar, 12 Mar (TBC), +1 + revision class in Easter term.
- Today: continue MCMC
- MacKay: Ch 29-30; Bishop: Ch 11; Gelman
- Givens & Hoeting "Computational Statistics"
  (Free download through Cambridge Library iDiscover)
- Hogg & DFM, 2017 "Data analysis recipes: Using Markov Chain Monte Carlo." https://arxiv.org/abs/1710.06068

Markov Chain Monte Carlo (MCMC)

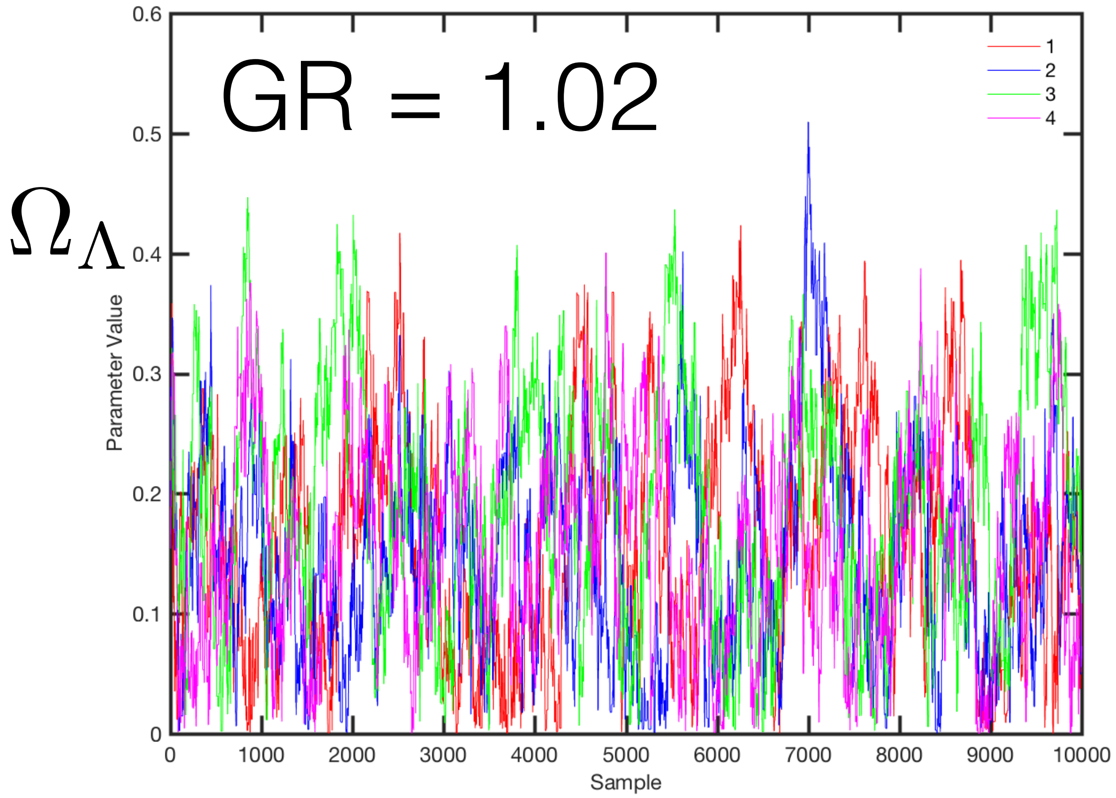
Markov Chain Monte Carlo (MCMC) to Map the Posterior P(θ |D)

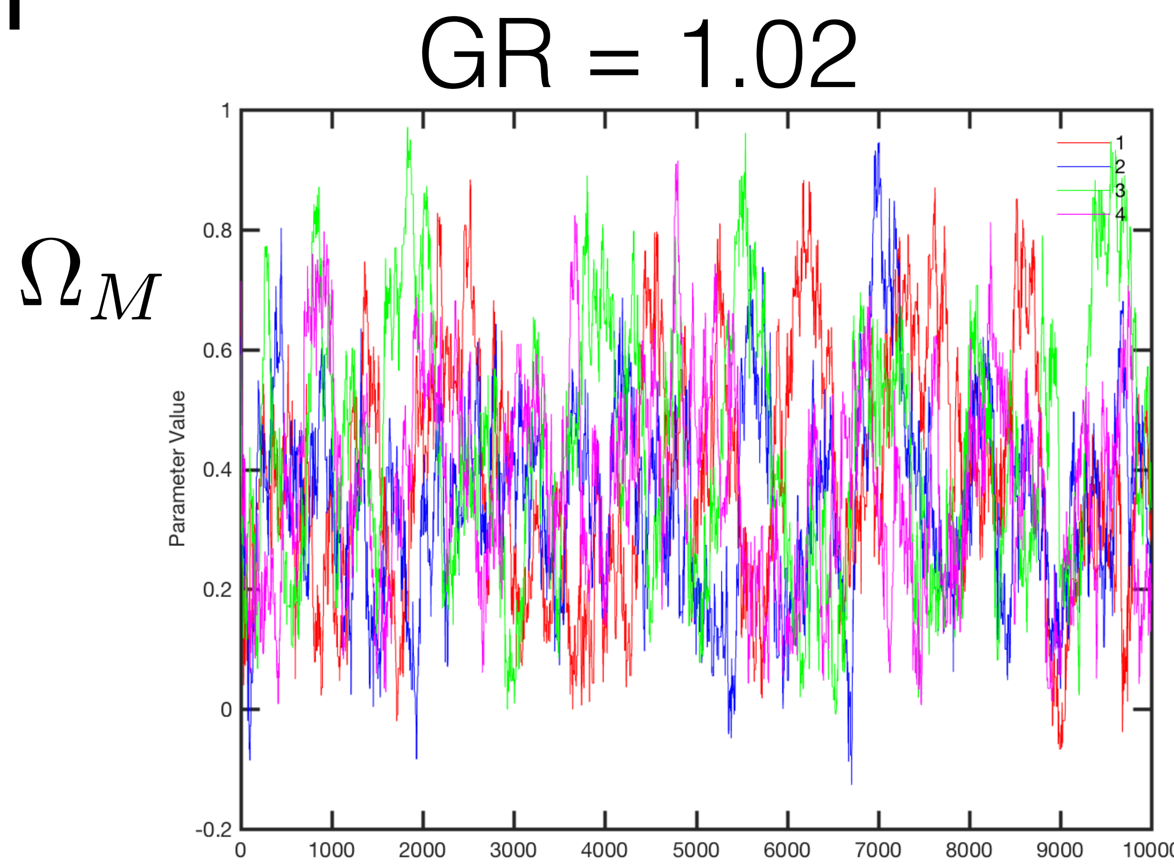
- Markov Chain Monte Carlo (MCMC)
    - Metropolis-Hastings algorithm
    - Gibbs sampling
    - Metropolis-within-Gibbs

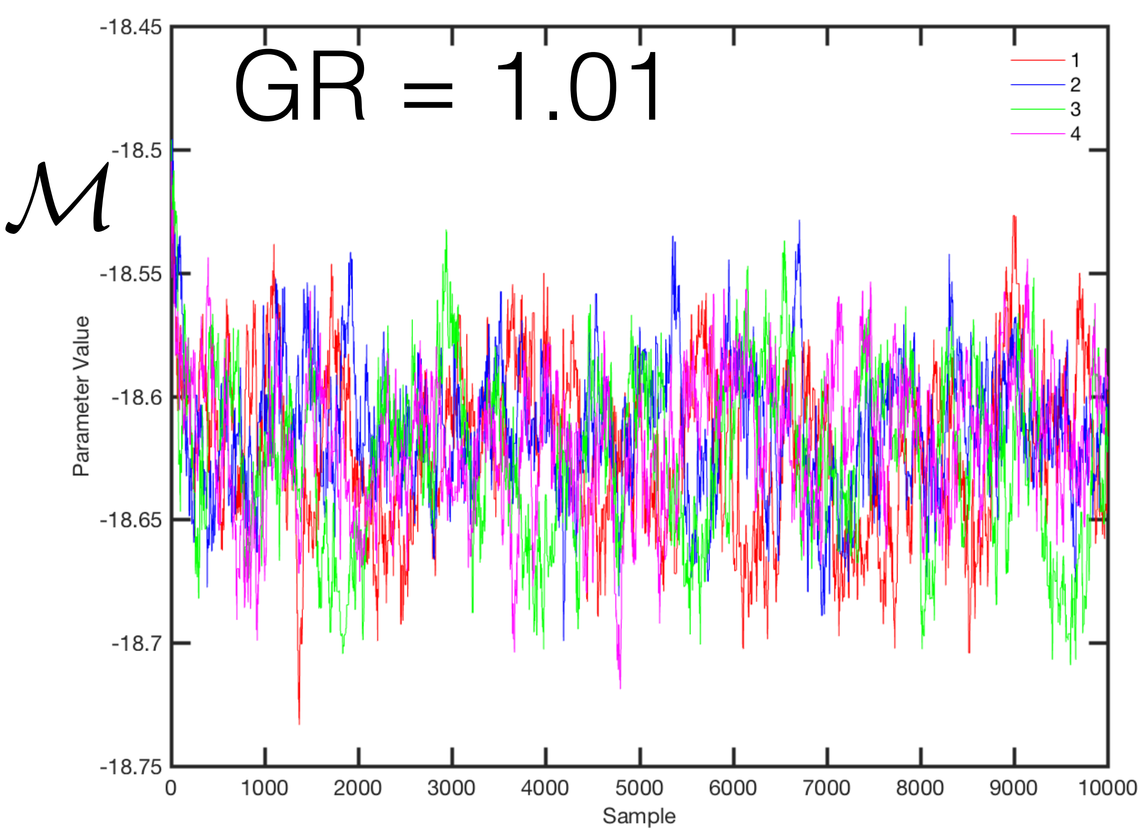
Idealised supernova dataset

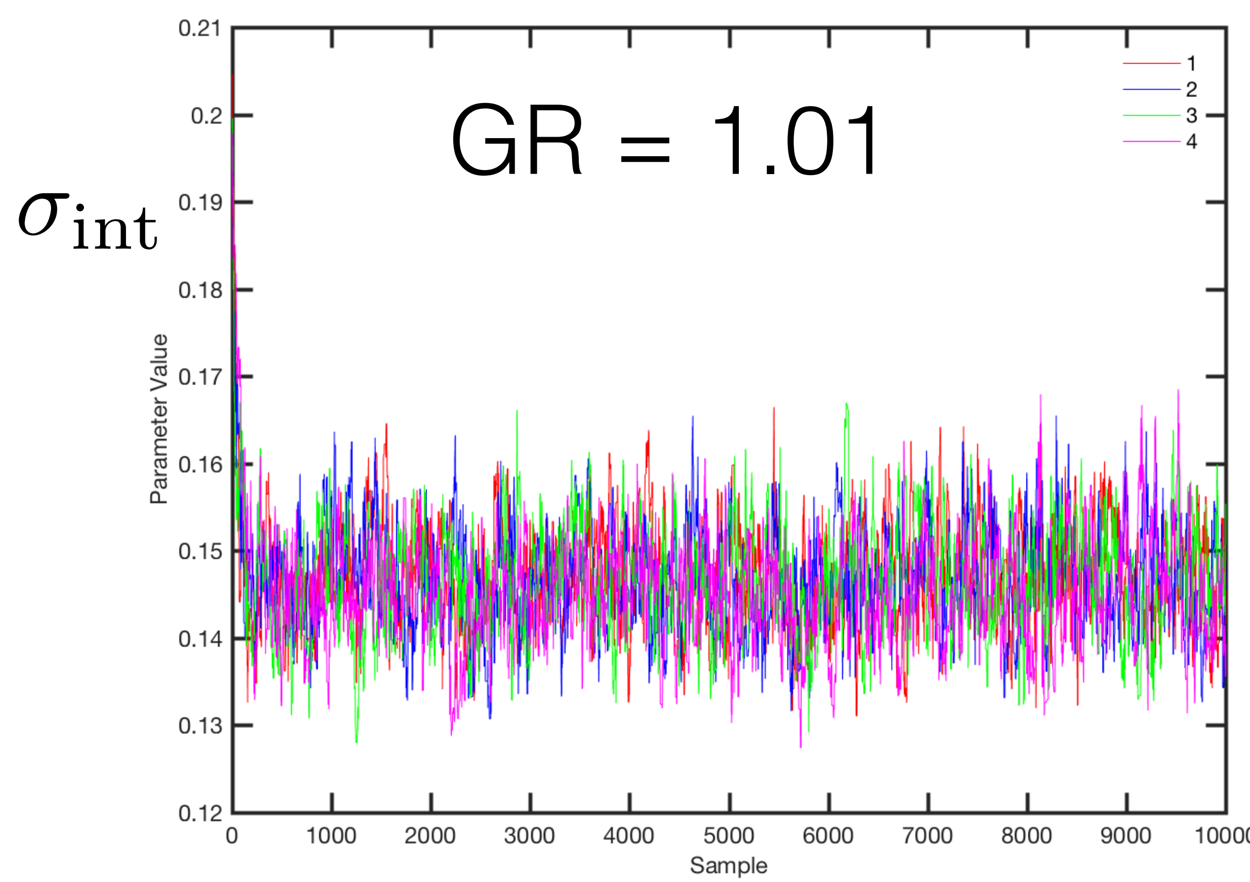


Multiple Independent Chains

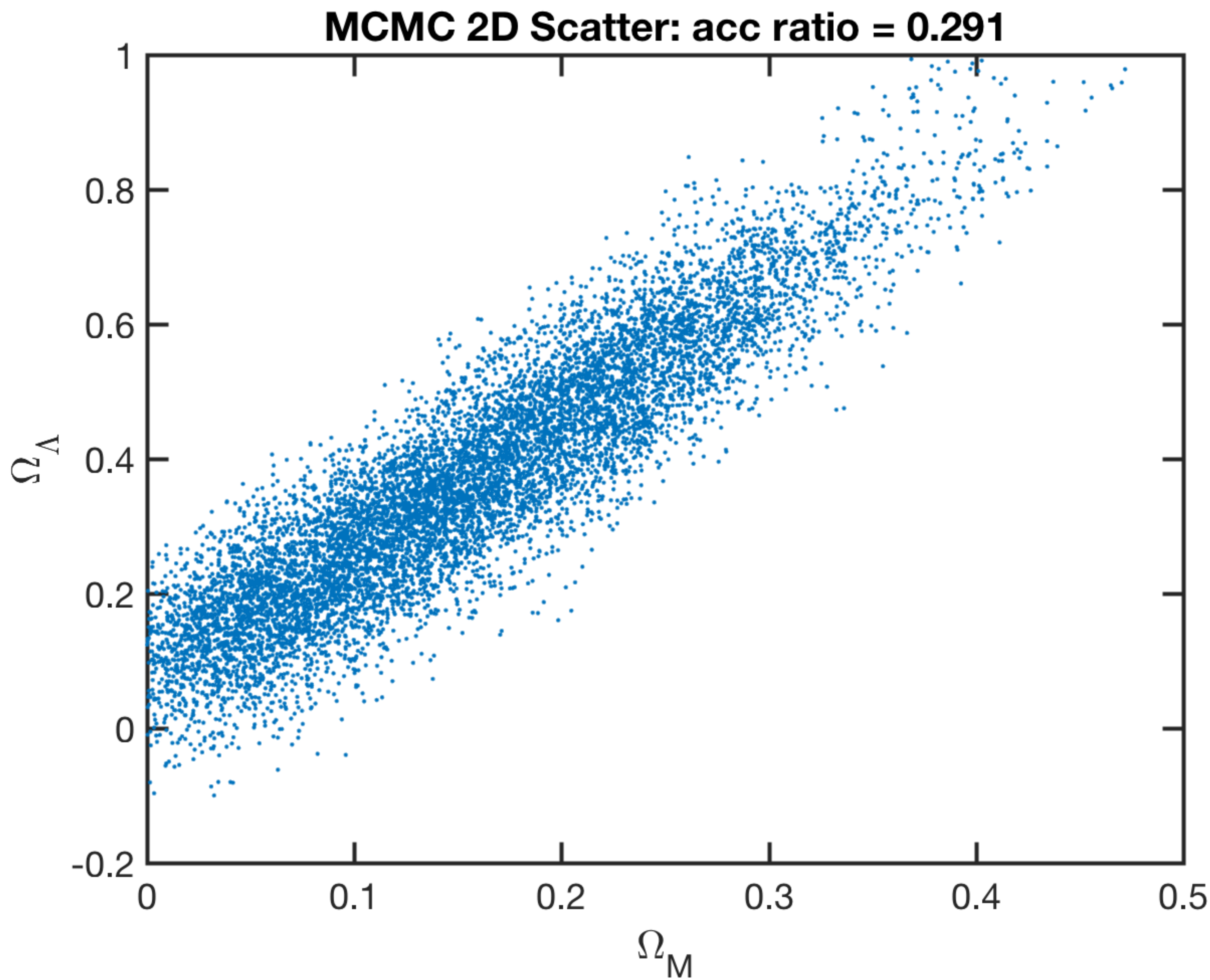




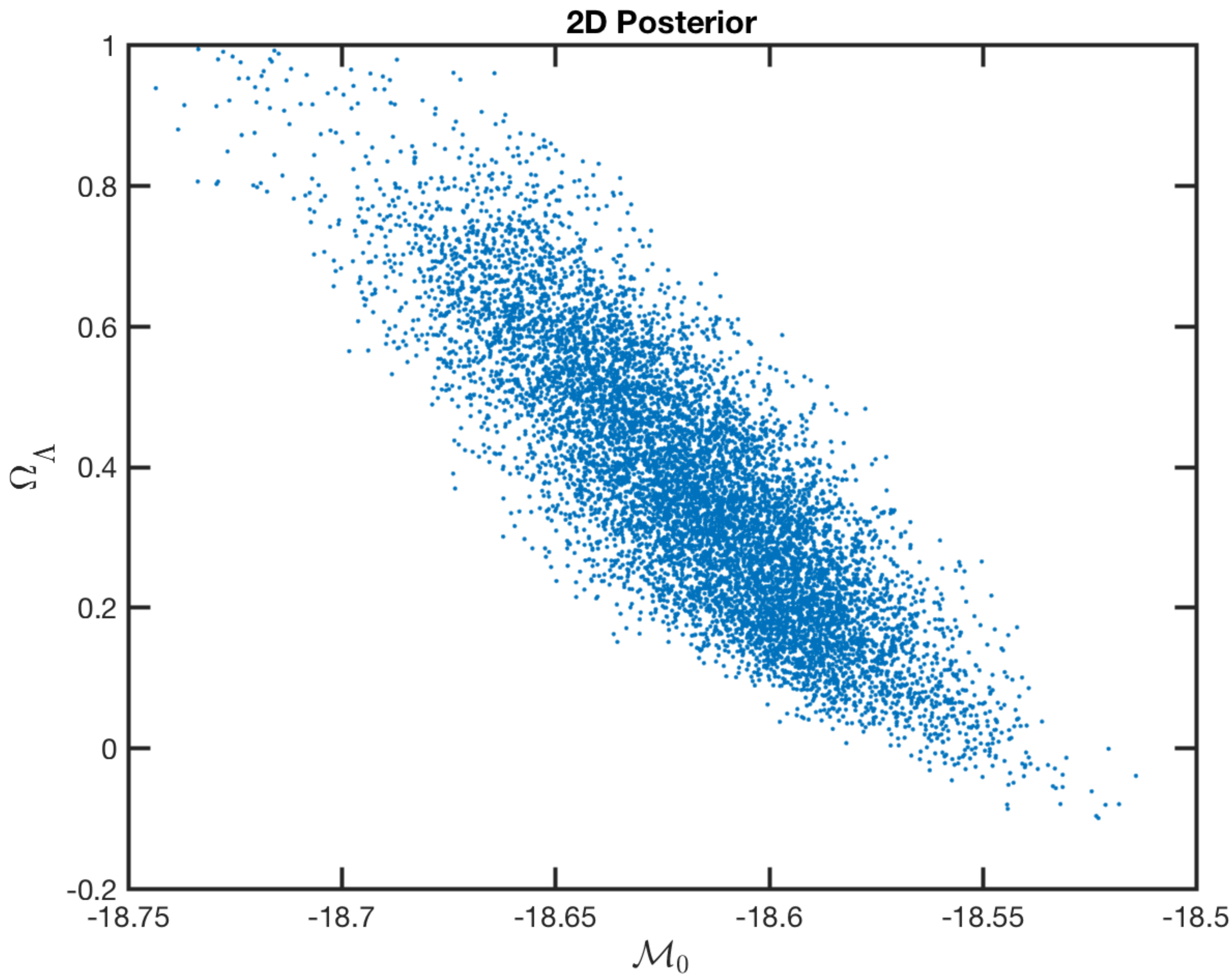




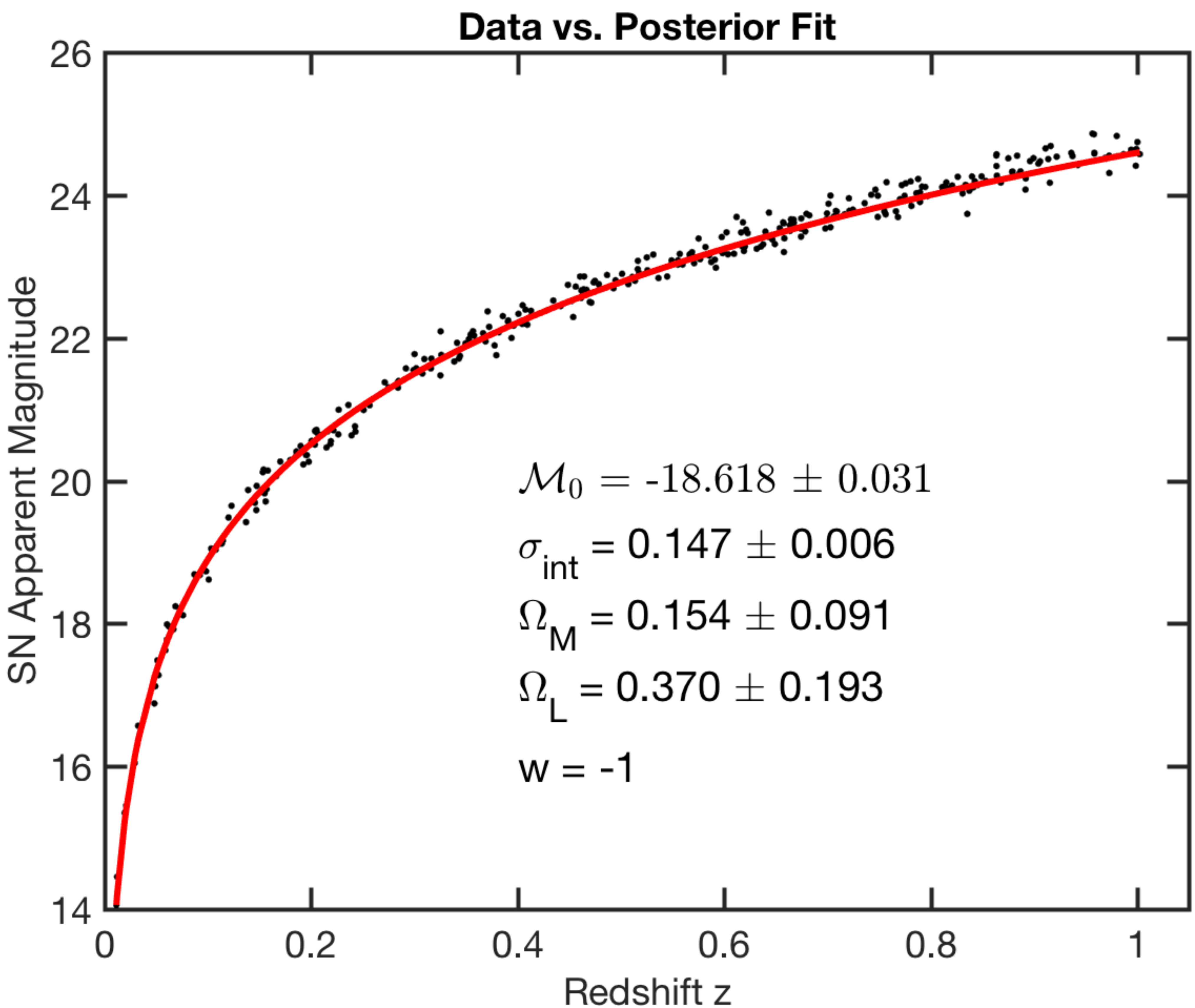
Posterior Scatter plot



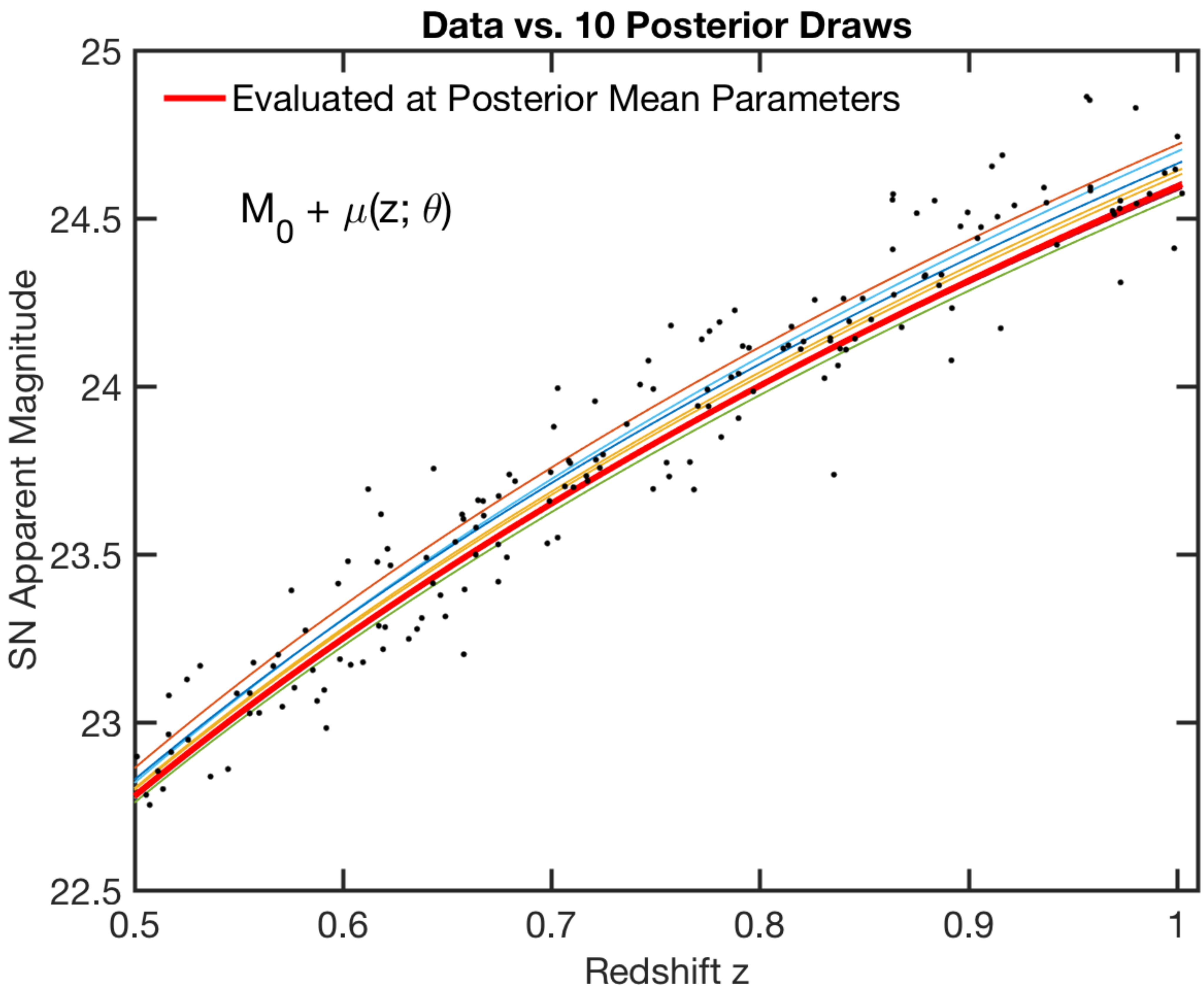
Posterior Scatter plot



Fit with posterior mean



Random Posterior Draws of parameters from chain



Metropolis-Hastings Algorithm: More General Jumping Rule: $J(\theta^*|\theta_i)$ [Need not be symmetric: $J(\theta_a \mid \theta_b) \neq J(\theta_b \mid \theta_a)$]

1. Choose a random starting point θ_0
2. At step i = 1…N, propose a new parameter value: $\theta^* \sim J(\theta^*|\theta_{i-1})$
3. Evaluate M-H ratio of posteriors at proposed vs current values.
   $r = [P(\theta^* \mid \mathbf{y}) / J(\theta^*|\theta_{i-1})] / [P(\theta_{i-1} \mid \mathbf{y}) / J(\theta_{i-1}|\theta^*)]$
4. Accept θ^* with probability min(r,1): $\theta_i = \theta^*$. If not accept, stay at same value $\theta_i = \theta_{i-1}$ & include in chain.
5. Repeat steps 2-4 until reach some measure of convergence and gather enough samples to compute your inference

Metropolis-Hastings Algorithm: More General Jumping Rule: $J(\theta^*|\theta_i)$ [Need not be symmetric: $J(\theta_a \mid \theta_b) \neq J(\theta_b \mid \theta_a)$]

- d-dim Metropolis is just a special case, where $J(\theta^*| \theta_i) = N(\theta^* \mid \theta_i, \mathbf{\Sigma_p}) = N(\theta_i \mid \theta^{\boldsymbol{*}}, \mathbf{\Sigma_p}) = J(\theta_i| \theta^*)$ is a symmetric proposal distribution
- More general asymmetric proposals, allow “biased” proposals —> more probable to propose towards a certain direction
- With some knowledge of structure of the posterior, can sometims engineer a clever proposal $J(\theta^*|\theta_i)$

Gibbs Sampling

- Multi-dimensional sampling, when you can utilise the set of conditional posterior distributions.
- If joint posterior is $P(\theta, \phi | \boldsymbol{D})$
- And you can solve for tractable conditionals:

$$P(\theta | \, \phi, \boldsymbol{D})$$

$$P(\phi | \, \theta, \boldsymbol{D})$$

- Jump along each parameter-dimension one at a time

2-dim Gibbs Sampler

1. Choose a random starting point (θ^0, $\boldsymbol{\phi}^0$)

2. At cycle t, update $\theta^t \sim P(\theta|\,\phi^{t-1}, \boldsymbol{D})$

3. Then update $\phi^t \sim P(\phi|\,\theta^t, \boldsymbol{D})$

   (Each complete set (pair) of updates is called a Gibbs cycle)

4. Record current values of chain (θ^t, $\boldsymbol{\phi}^t$)

5. Repeat steps 2-4, until reach some measure of convergence (G-R) and gather enough independent samples to compute your inference (reduce Monte Carlo error)

d-dim Gibbs Sampler

Parameter vector $\boldsymbol{\theta} = (\theta_1 \ldots \theta_d)$

Current state at j-th update within cycle t: $(\boldsymbol{\theta}_j^t, \boldsymbol{\theta}_{-j}^t)$

$$\boldsymbol{\theta}_{-j}^t \equiv (\theta_1^{t+1}, \ldots, \theta_{j-1}^{t+1}, \theta_{j+1}^t, \ldots, \theta_d^t)$$

1. Choose a random starting point θ_0.

2. In each cycle t, update through the d-parameters:
   For each j = 1… d, move jth parameter to

$$\theta_j \sim P(\theta_j | \boldsymbol{\theta}_{-j}^t, \boldsymbol{D})$$

   (update θ_j conditional on current values of all other parameters)

3. After updating all d parameters, record current state: $\boldsymbol{\theta}^{t+1}$

4. Repeat steps 2-3 until reach convergence and enough samples

Gibbs Sampling: Example (Gelman BDA Section 11.1)

Likelihood: $$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$ ρ known

Priors: $$P(\theta_1) = P(\theta_2) \propto 1$$

Posterior: $$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \Big| \boldsymbol{y} \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

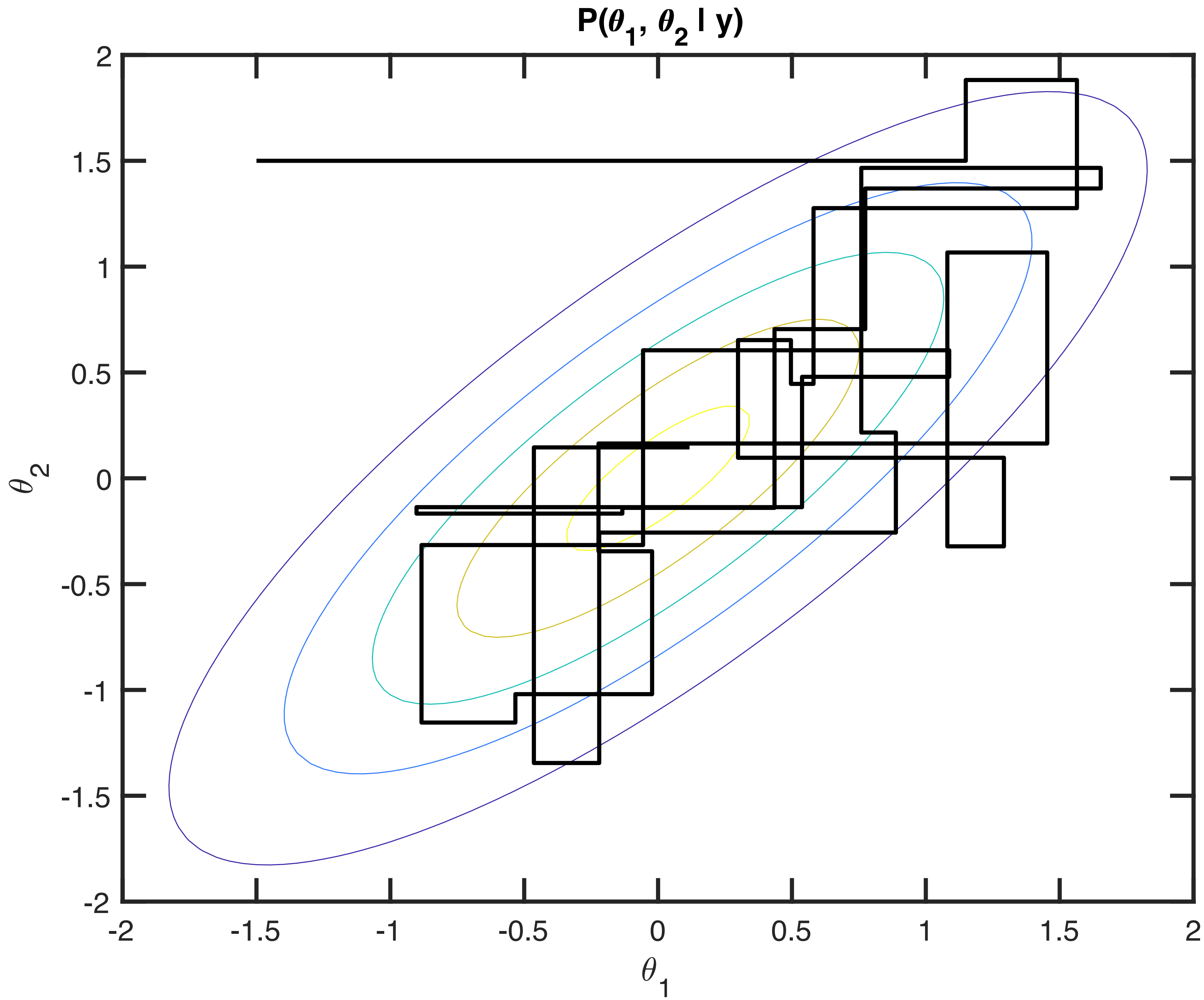
$$P(\boldsymbol{\theta}|y) = P(\theta_1|\theta_2, y)P(\theta_2|y) = P(\theta_2|\theta_1, y)P(\theta_1|y)$$

Conditional Posteriors:

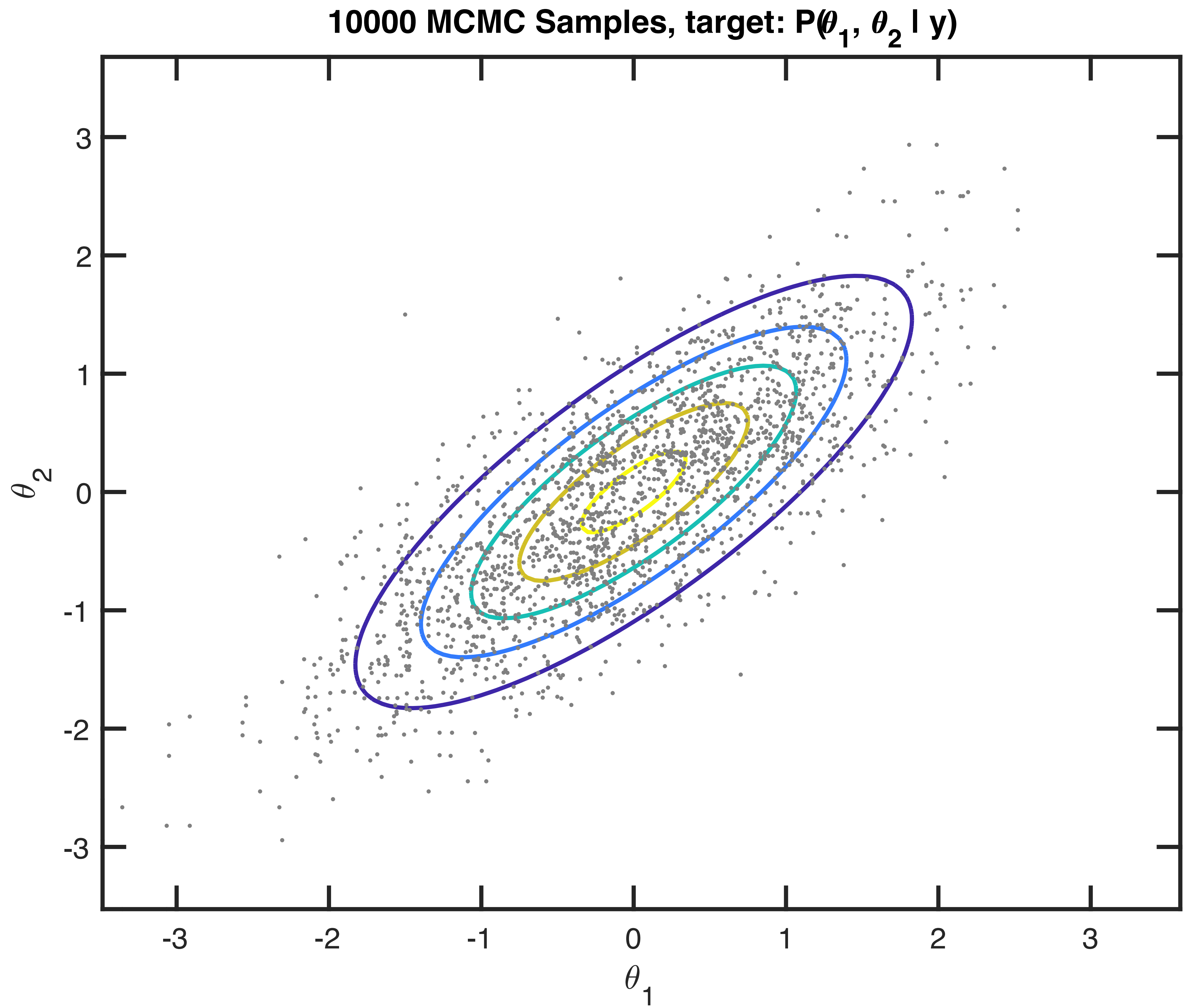
$$\theta_1|\theta_2, \boldsymbol{y} \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$

$$\theta_2|\theta_1, \boldsymbol{y} \sim N(y_2 + \rho(\theta_1 - y_1), 1 - \rho^2)$$

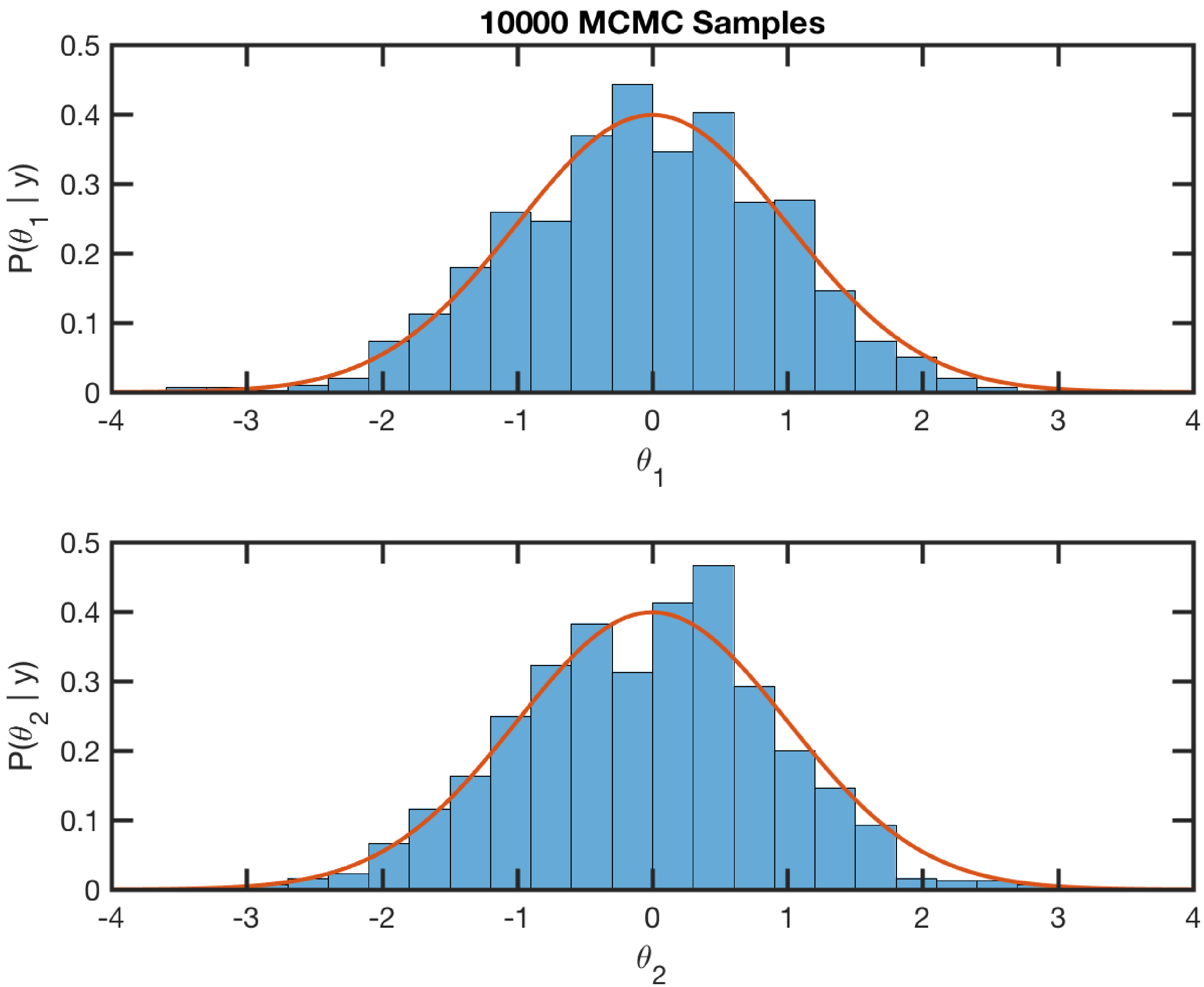
Gibbs Sampling: demo
gibbs_example.m
2D Trace Paths for 50 iterations



Gibbs Sampling: demo
gibbs_example.m (demo different ρ's)
Joint Posterior Densities



Gibbs Sampling: demo
gibbs_example.m
Marginal Posterior Densities



Metropolis-within-Gibbs

$$\boldsymbol{\theta} = (\theta_1 \ldots \theta_d) \qquad\qquad \boldsymbol{\theta}_{-j} \equiv (\theta_1, \ldots \theta_{j-1}, \theta_{j+1} \ldots \theta_d)$$

- When you can't solve for tractable conditional distributions for all θ_j: $P(\theta_j | \boldsymbol{\theta}_{-j}, \boldsymbol{D})$
- Replace each substep for updating each jth parameter θ_j with a separate Metropolis rule, compute Metropolis ratio, and accept/reject
- Cycle through all parameters, and repeat all for N MCMC steps

d-dim Metropolis-within-Gibbs Sampler

$$\boldsymbol{\theta} = (\theta_1 \ldots \theta_d) \qquad \boldsymbol{\theta}^t_{-j} \equiv (\theta_1^{t+1}, \ldots, \theta_{j-1}^{t+1}, \theta_{j+1}^t, \ldots, \theta_d^t)$$

1. Choose a random starting point θ_0

2. At cycle t = 1…N, cycle through the d-parameters:

   A. For each j = 1… d, propose a new j-th parameter value from a 1-Dimensional Gaussian: $\theta_j^* \sim N(\theta_j^t, \tau_j^2)$

   B. Evaluate ratio of posteriors at proposed vs current values:

   $$r = P(\theta_j^*, \boldsymbol{\theta}^t_{-j} \,|\, \boldsymbol{D}) / P(\theta_j^t, \boldsymbol{\theta}^t_{-j} |, \boldsymbol{D})$$
   $$= P(\theta_j^* \,|\, \boldsymbol{\theta}^t_{-j}, \boldsymbol{D}) / P(\theta_j^t \,|\, \boldsymbol{\theta}^t_{-j}, \boldsymbol{D})$$

   C. Accept $\theta_j^{t+1} = \theta_j^*$ with prob min(r,1), otherwise $\theta_j^{t+1} = \theta_j^t$

3. After full cycle, record current values $\boldsymbol{\theta}^{t+1}$

4. Repeat steps 2 for all parameters until convergence and enough samples

Metropolis-within-Gibbs Sampling: Example: Gelman BDA Section 11.1)

Likelihood: $$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right) \qquad \rho \text{ known}$$

Priors: $$P(\theta_1) = P(\theta_2) \propto 1$$

Posterior: $$\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \Big| \boldsymbol{y} \sim N\left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

Suppose we can't solve for the Conditional Posteriors.

Metropolis-within-Gibbs Sampling: Code Demo metropolisgibbs_example.m

demo different ρ’s

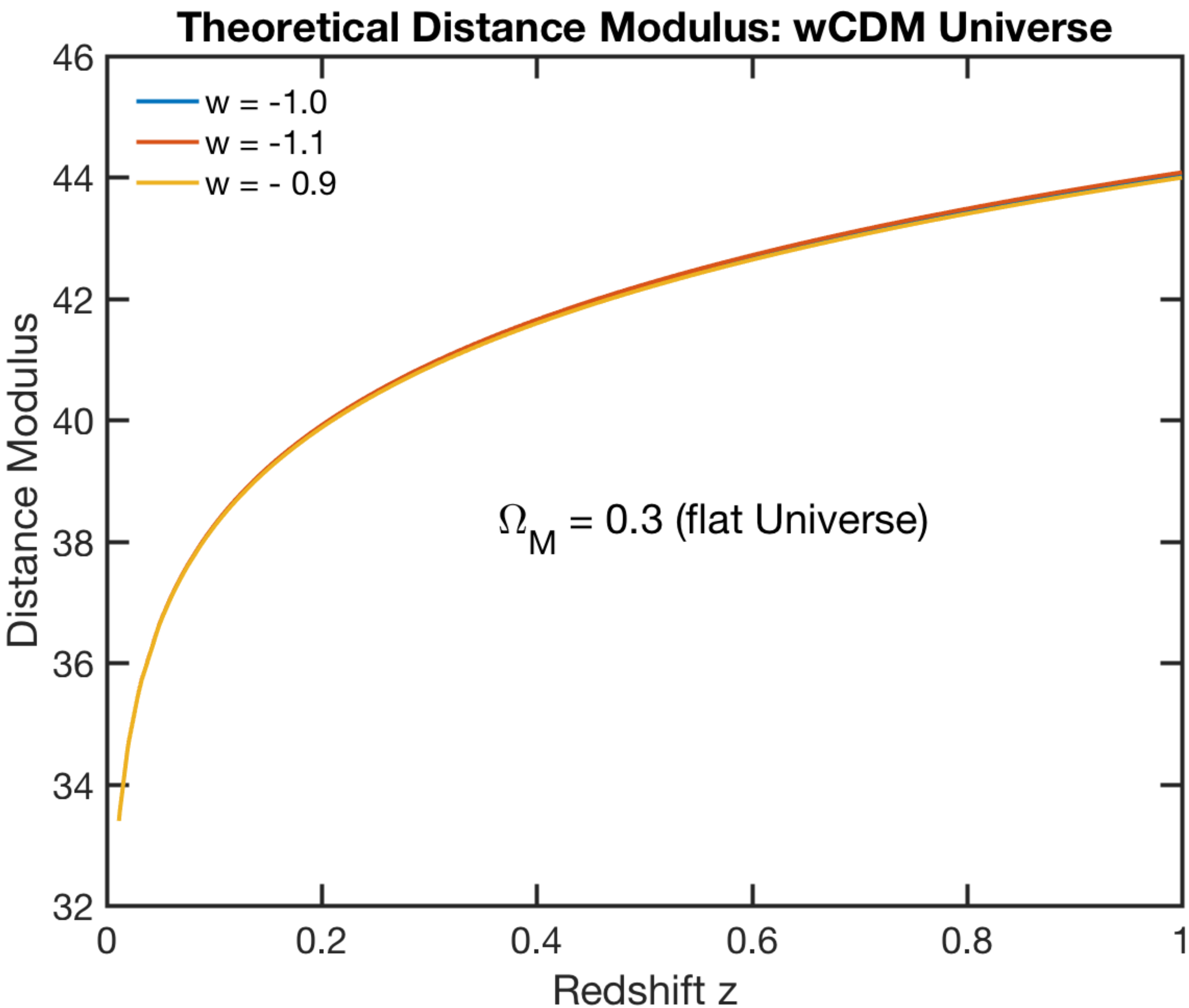
Supernova Cosmology Case Study:

Now assume flat Universe

$$\Omega_L = 1 - \Omega_M$$

but unknown w

Write down model & posterior

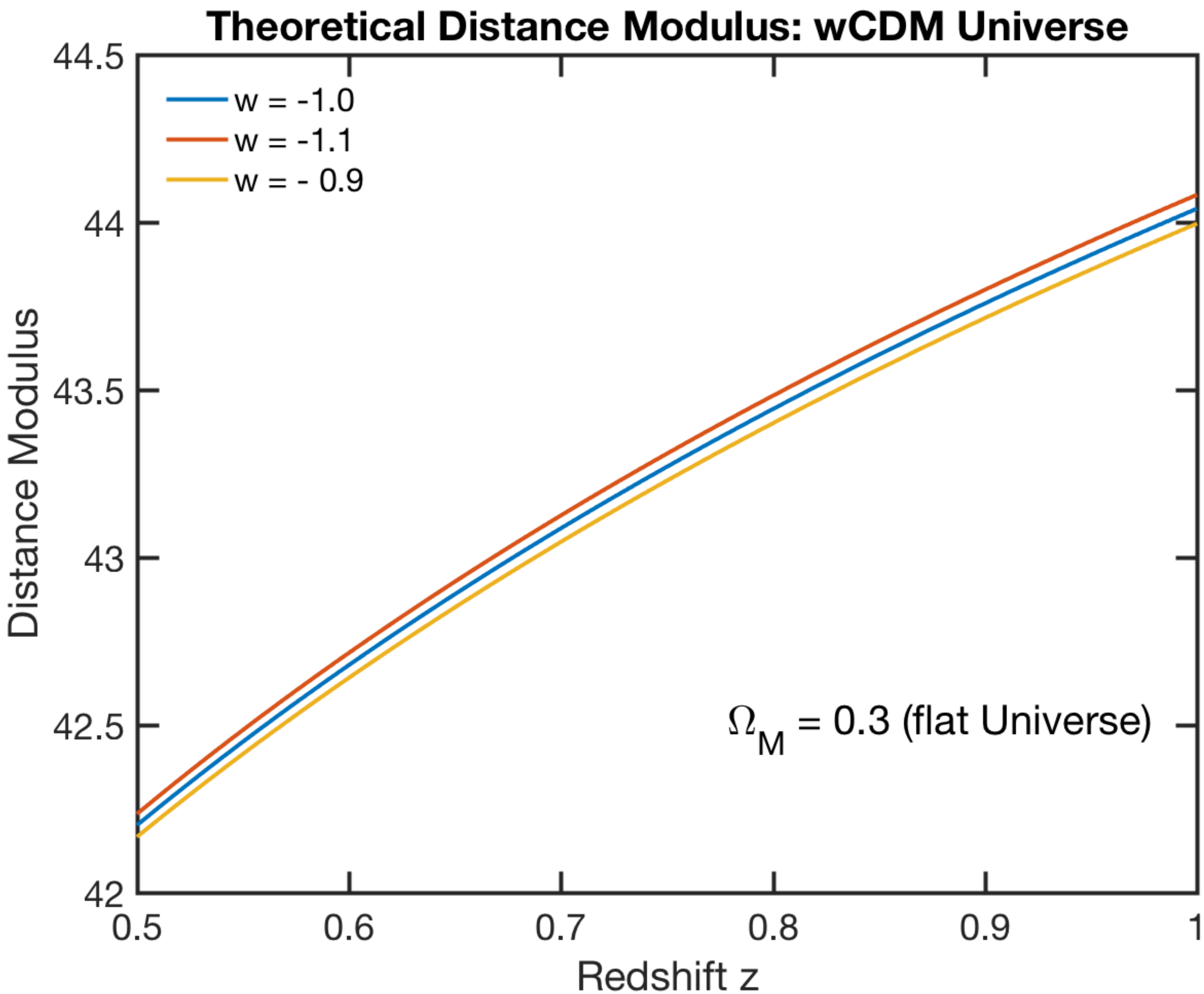


Supernova Cosmology Case Study:

Now assume flat Universe

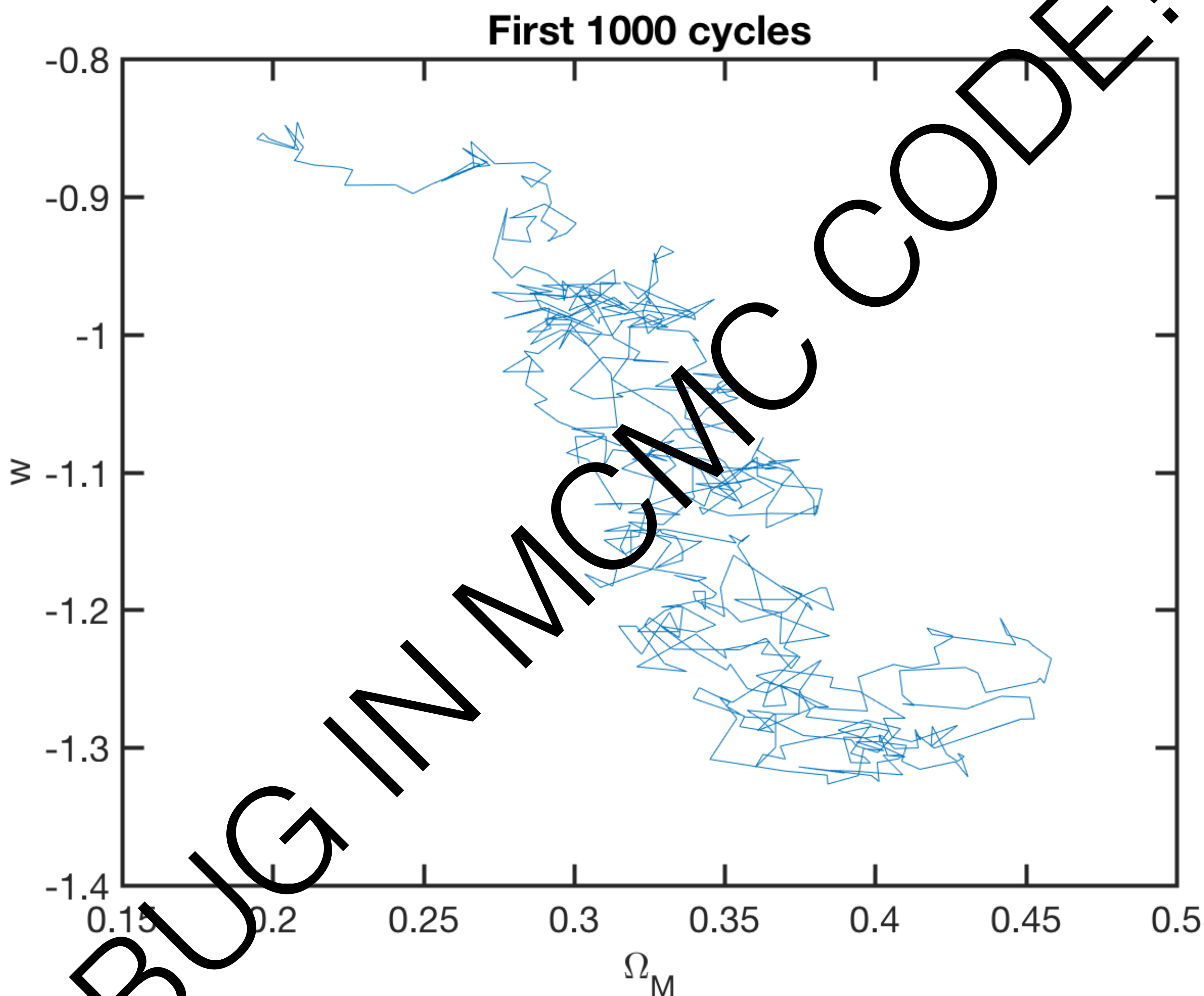
$$\Omega_L = 1 - \Omega_M$$

but unknown w



Metropolis-within-Gibbs: 2D trace plot

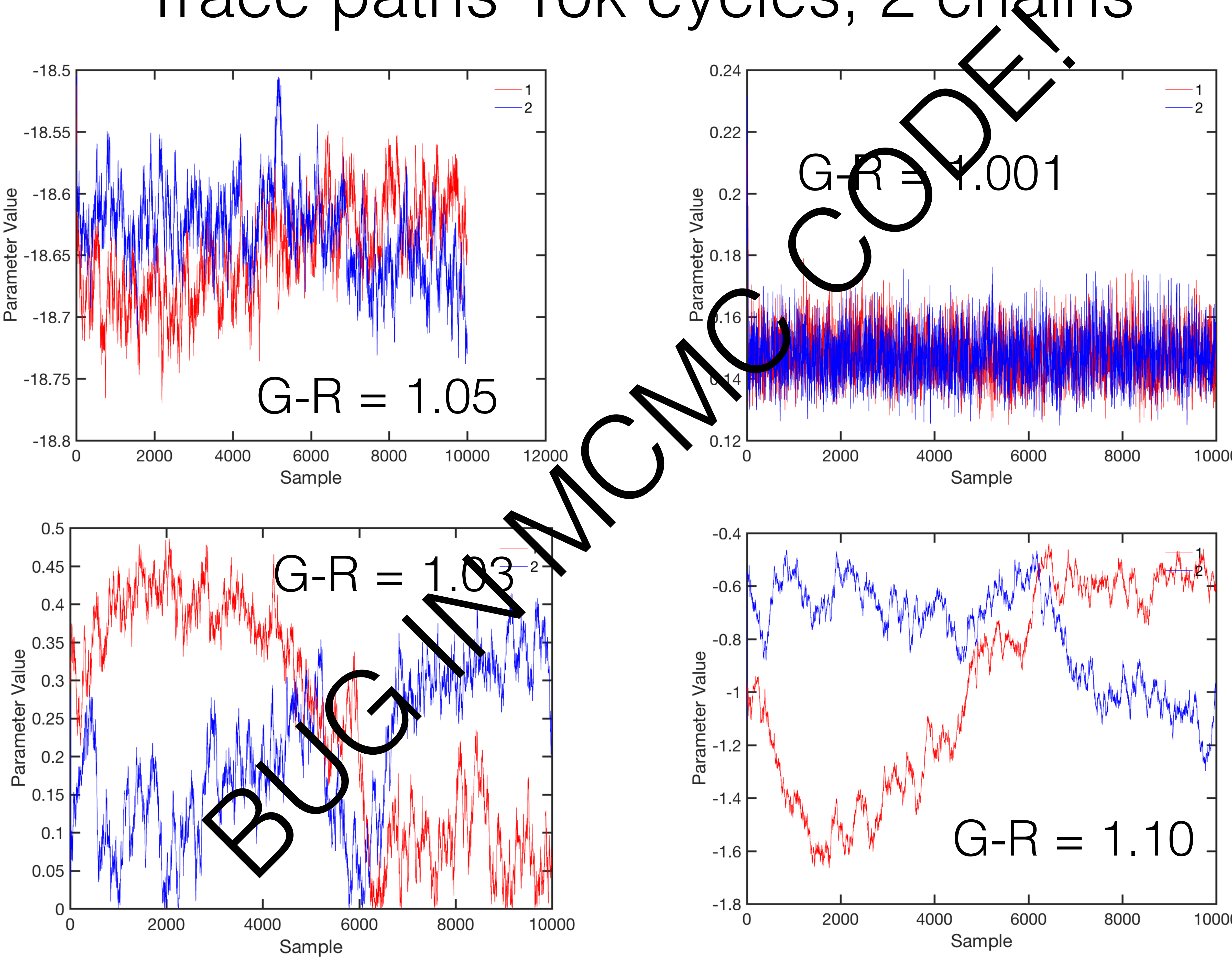
Dark Energy EoS parameter



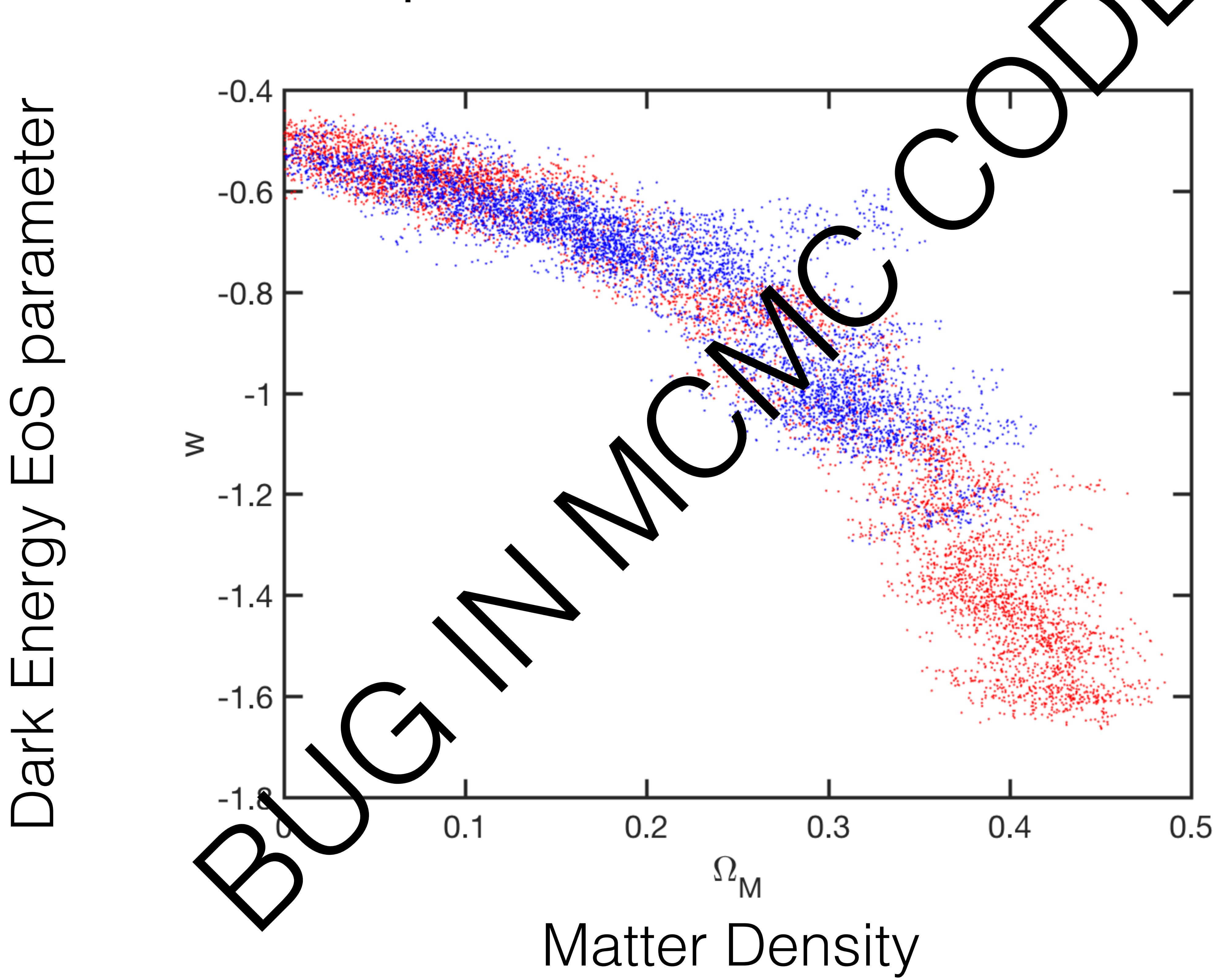
Matter Density

BUG IN MCMC CODE!

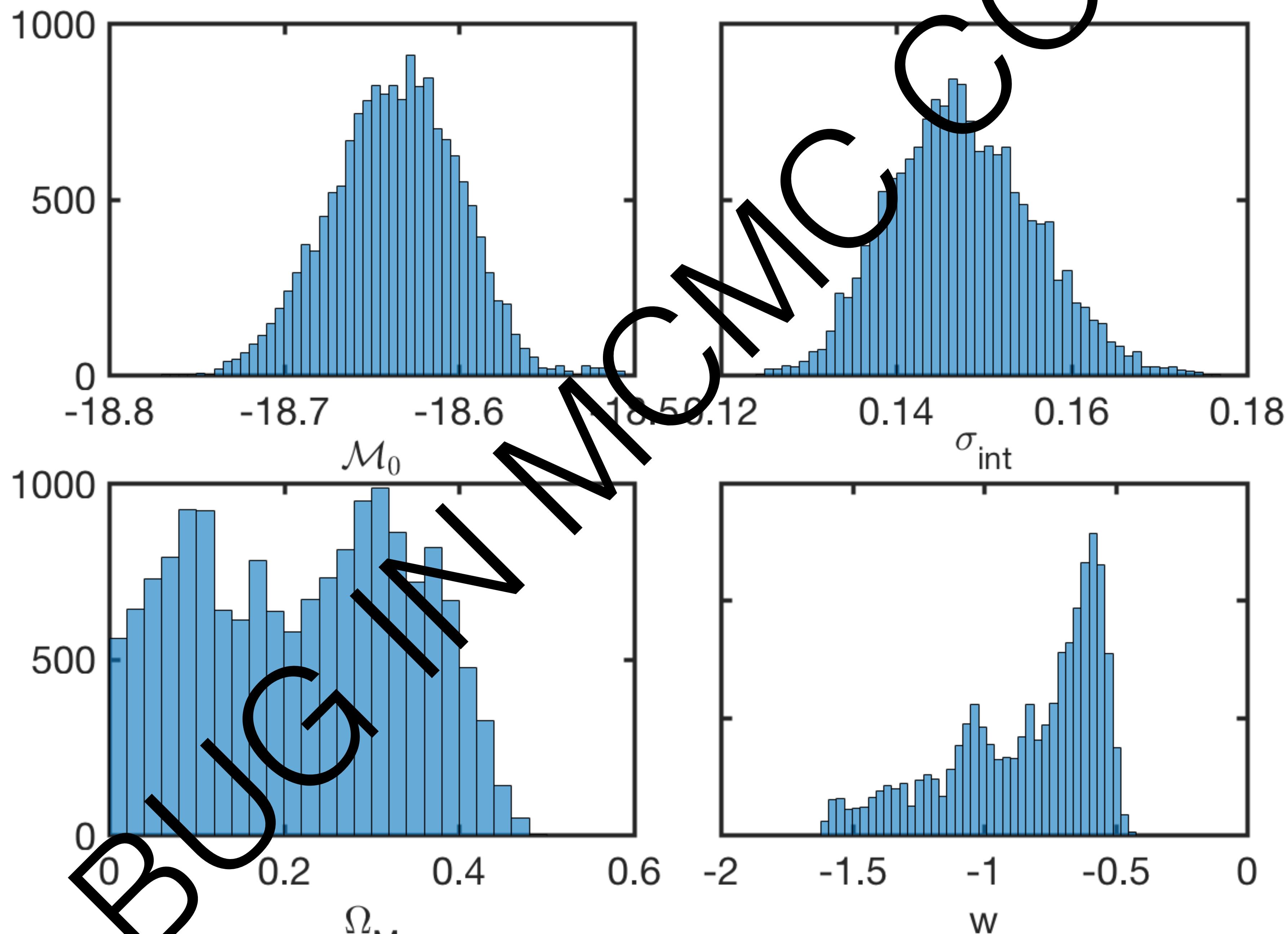
Trace paths 10k cycles, 2 chains



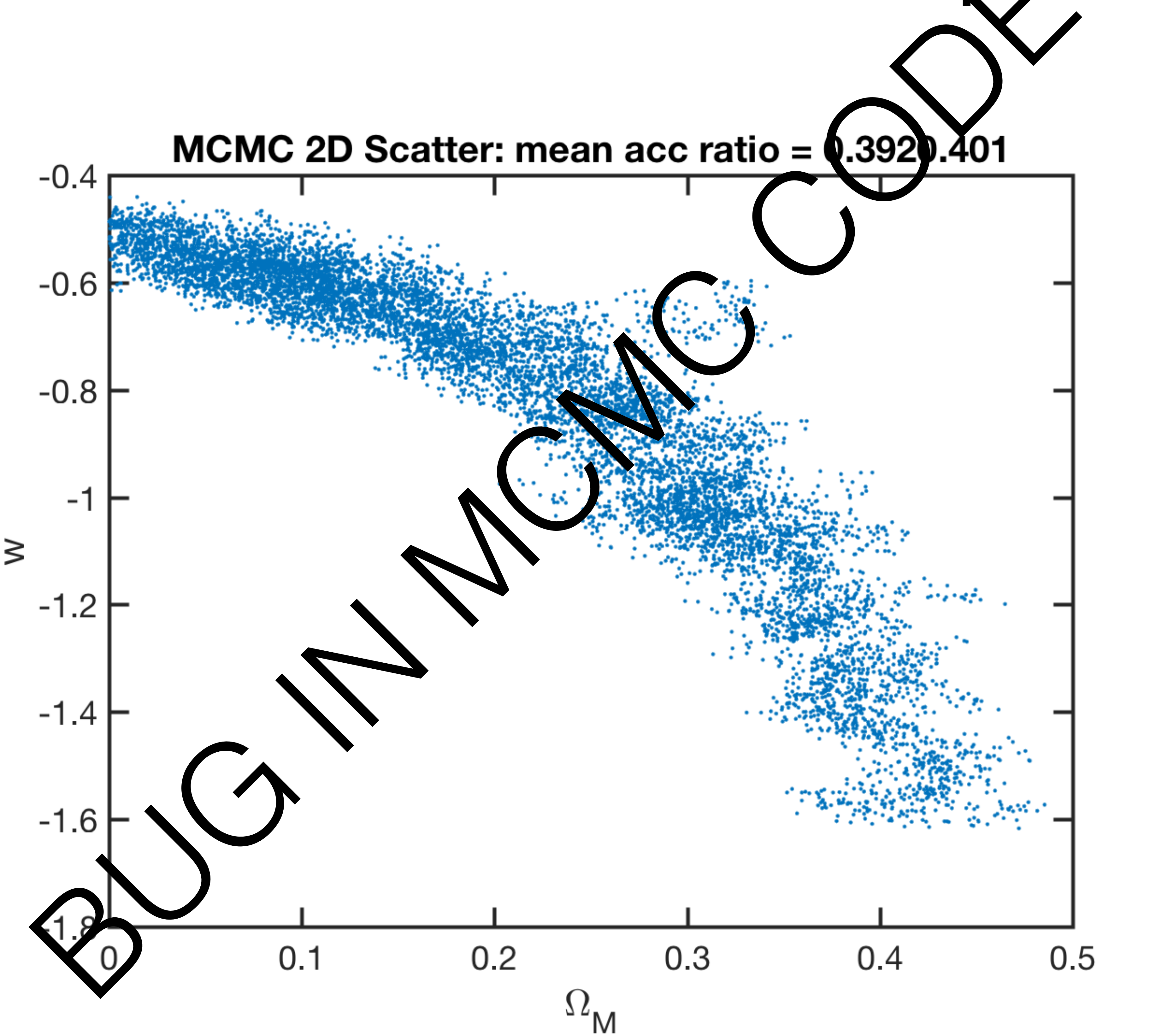
Metropolis-within-Gibbs: 2D trace plot
10,000 steps, 2 chain: Not well mixed!



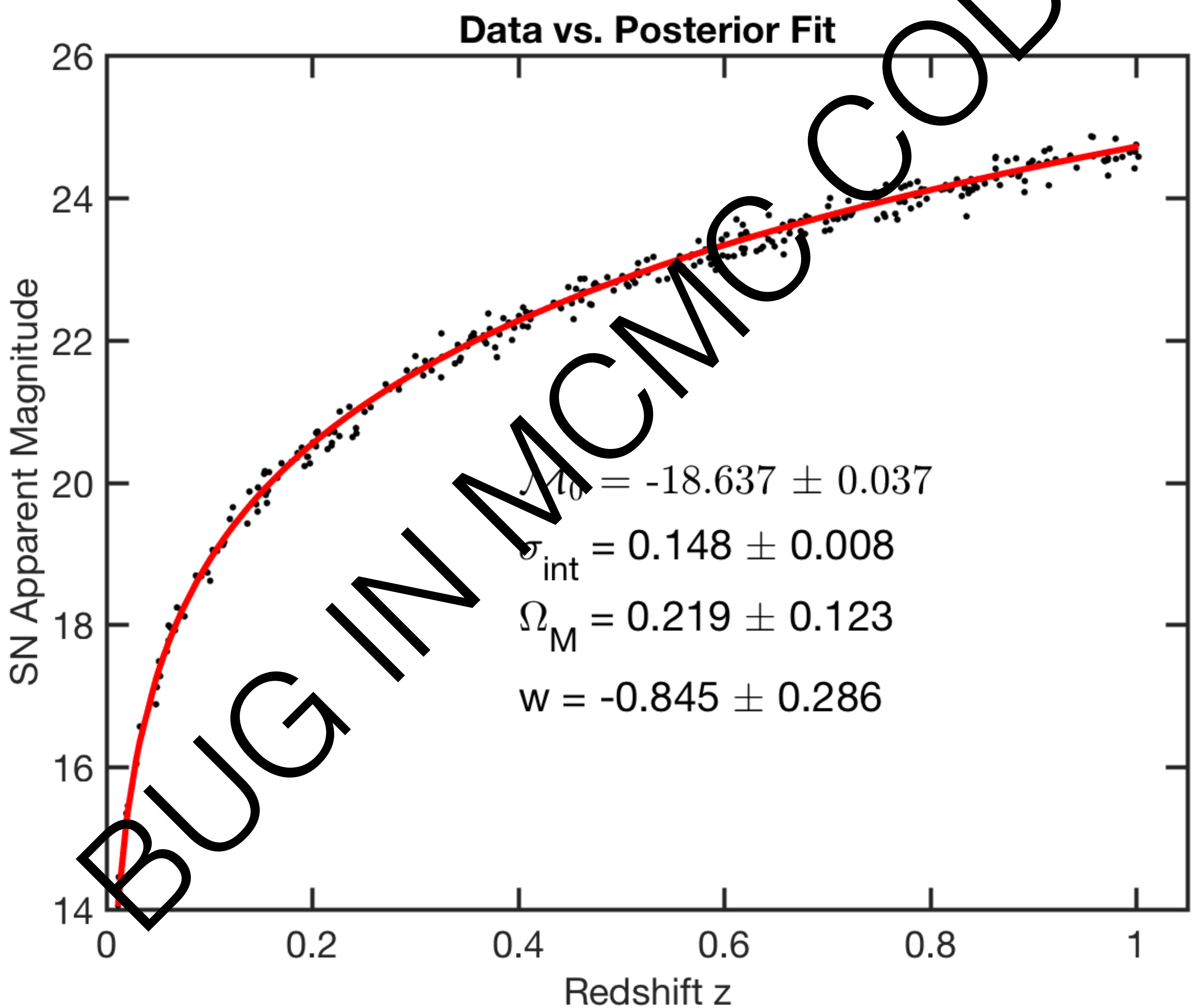
Posterior Histograms



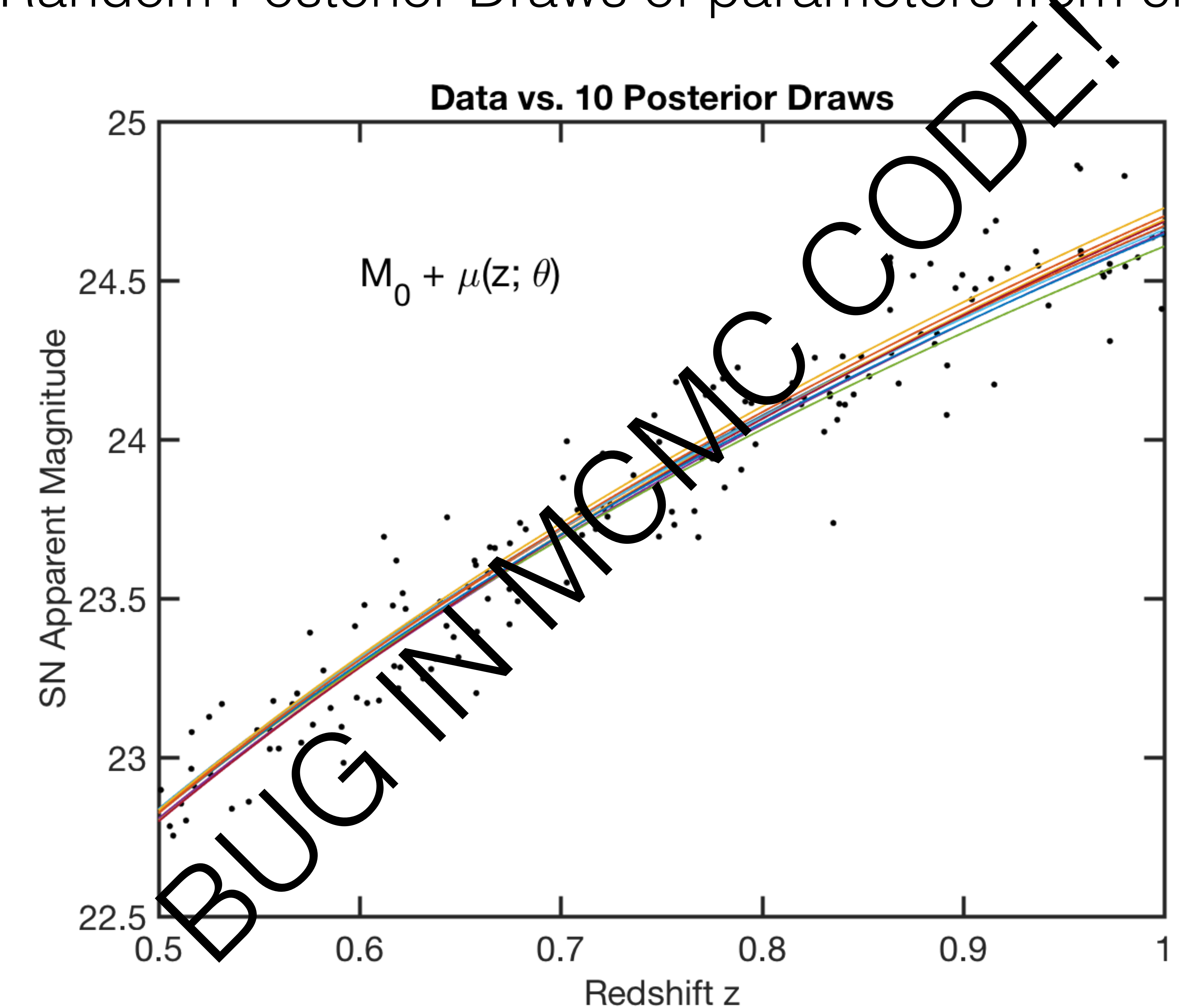
Posterior Scatter plot



Fit with posterior mean



Random Posterior Draws of parameters from chain



MCMC Animations

- The Markov-chain Monte Carlo Interactive Gallery
- http://chi-feng.github.io/mcmc-demo/