Astrostatistics: Monday 25 Feb 2019

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

- Next example classes:
  - Tue, 5 Mar, 12-2pm, MR 9 (confirm)
  - Tue, 12 Mar, 2 - 4pm, MR14 (confirm)
- Today: continue with 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) to Map the Posterior P(θ |D)

- Markov Chain Monte Carlo (MCMC)
  - Metropolis-Hastings algorithm
  - Gibbs sampling
  - Metropolis-within-Gibbs
- Comparing performance of MCMC algorithms
  - Autocorrelation time
  - Effective Sample Size
- Gibbs Sampling as Metropolis-Hastings
- Detailed Balance & theoretical considerations

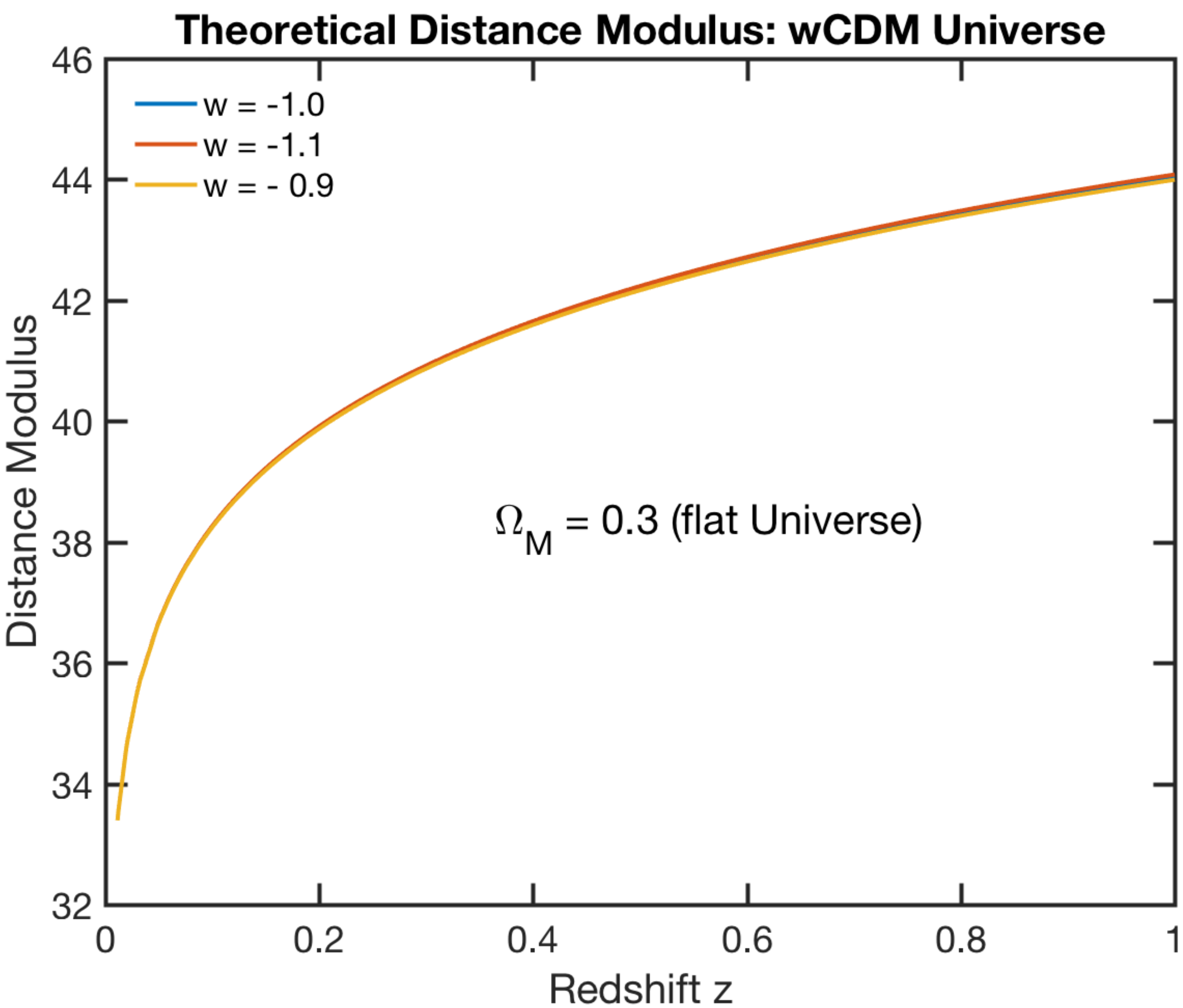
Supernova Cosmology Case Study:

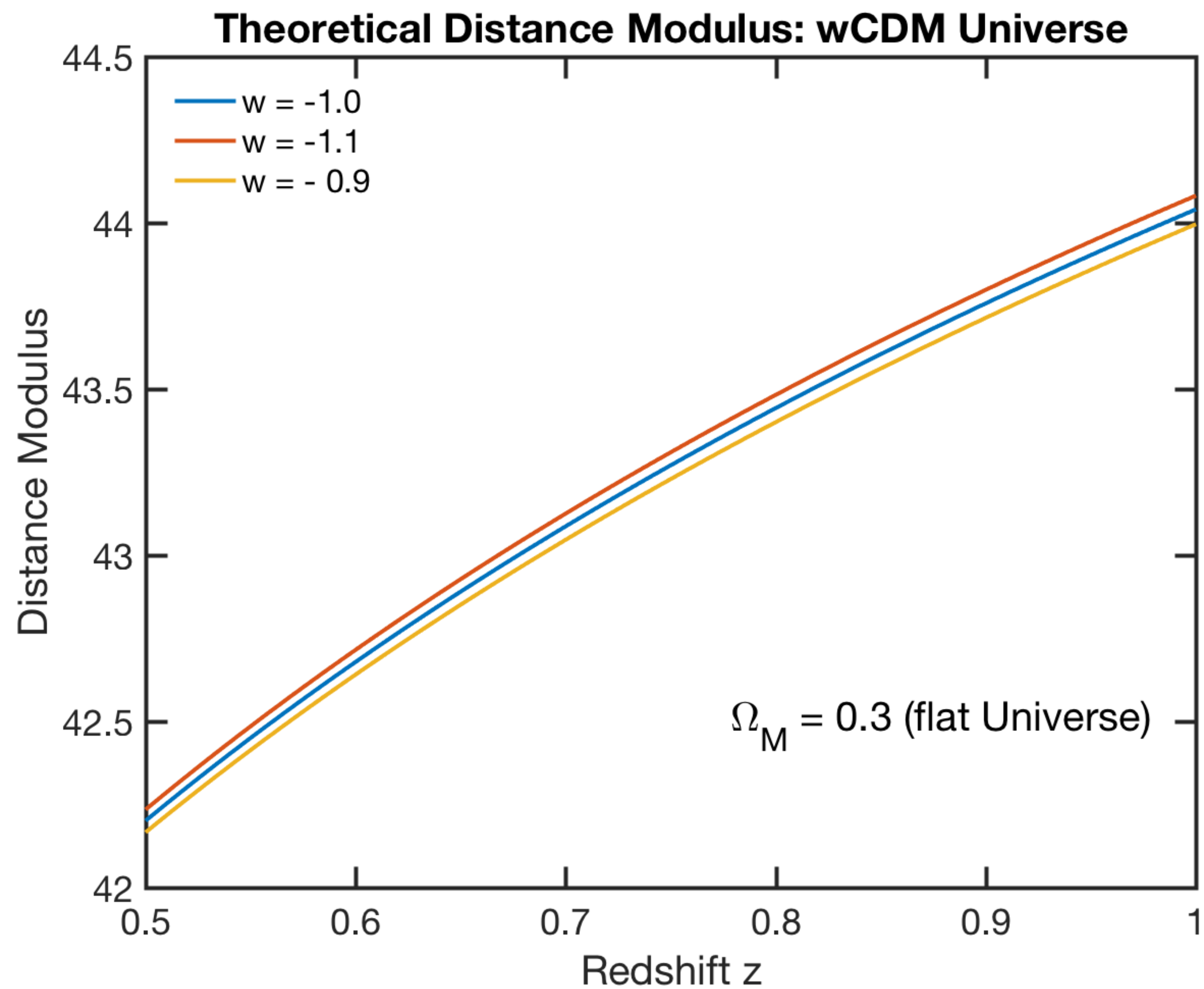
Assume flat Universe

$$\Omega_L = 1 - \Omega_M$$

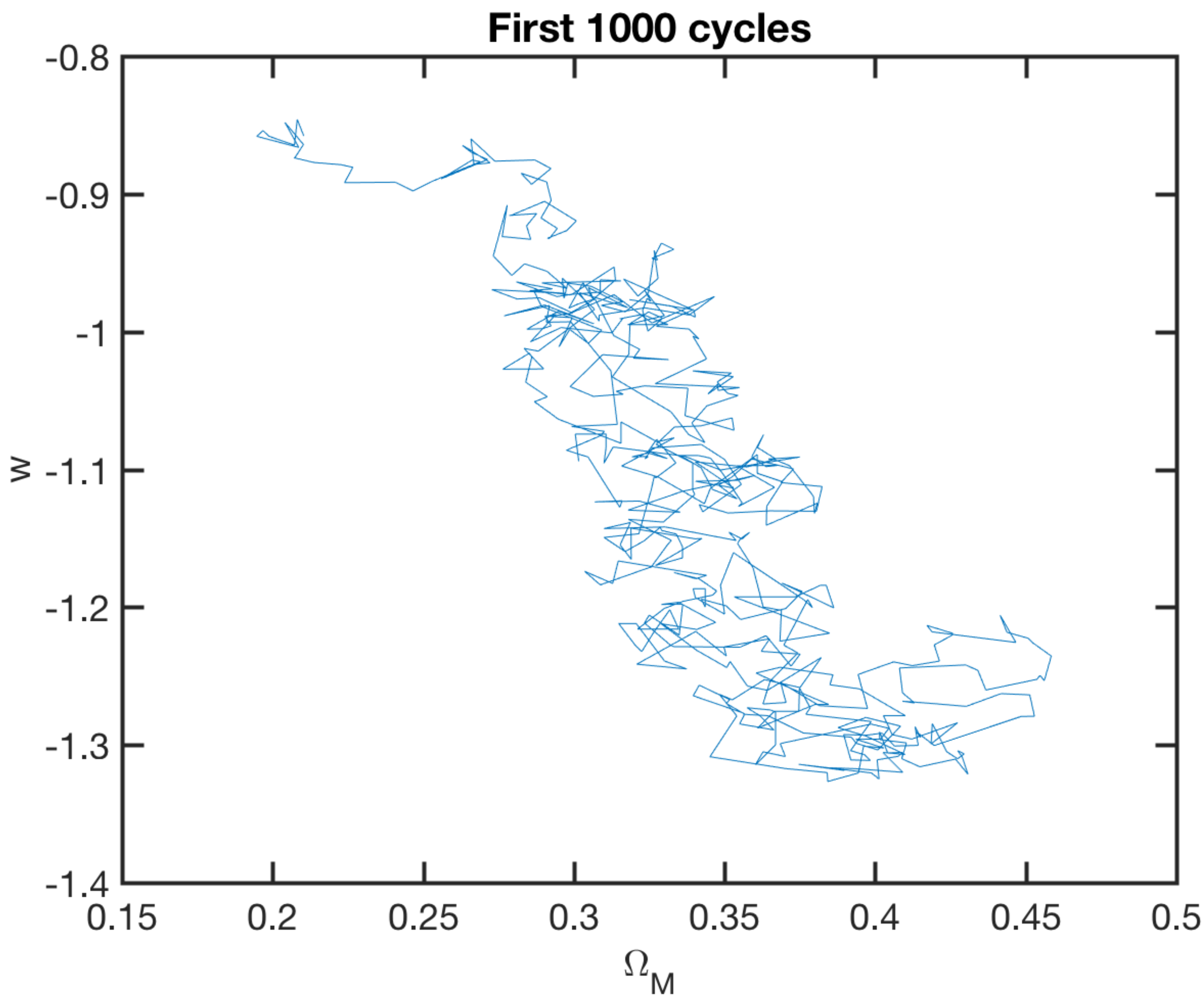
but unknown w

$$\boldsymbol{\theta} = (\mathcal{M}_0, \sigma_{\rm int}, \Omega_M, w)$$





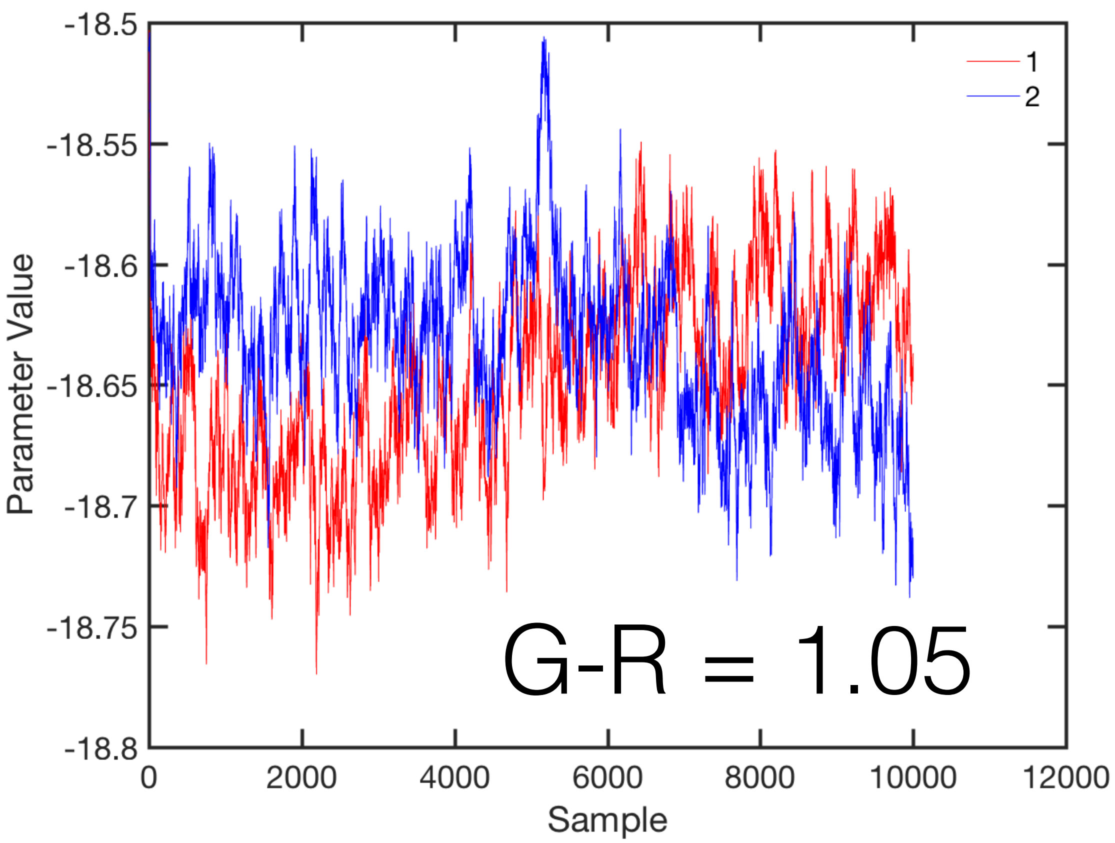
Metropolis-within-Gibbs: 2D trace plot

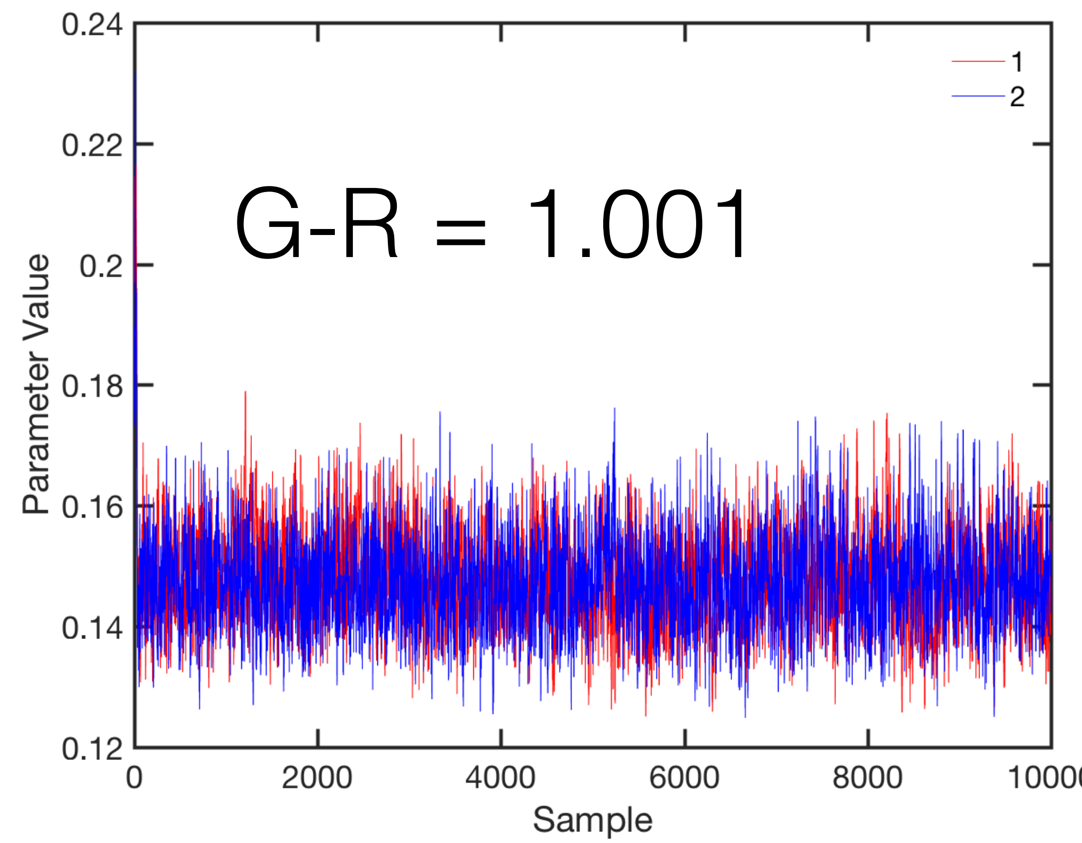


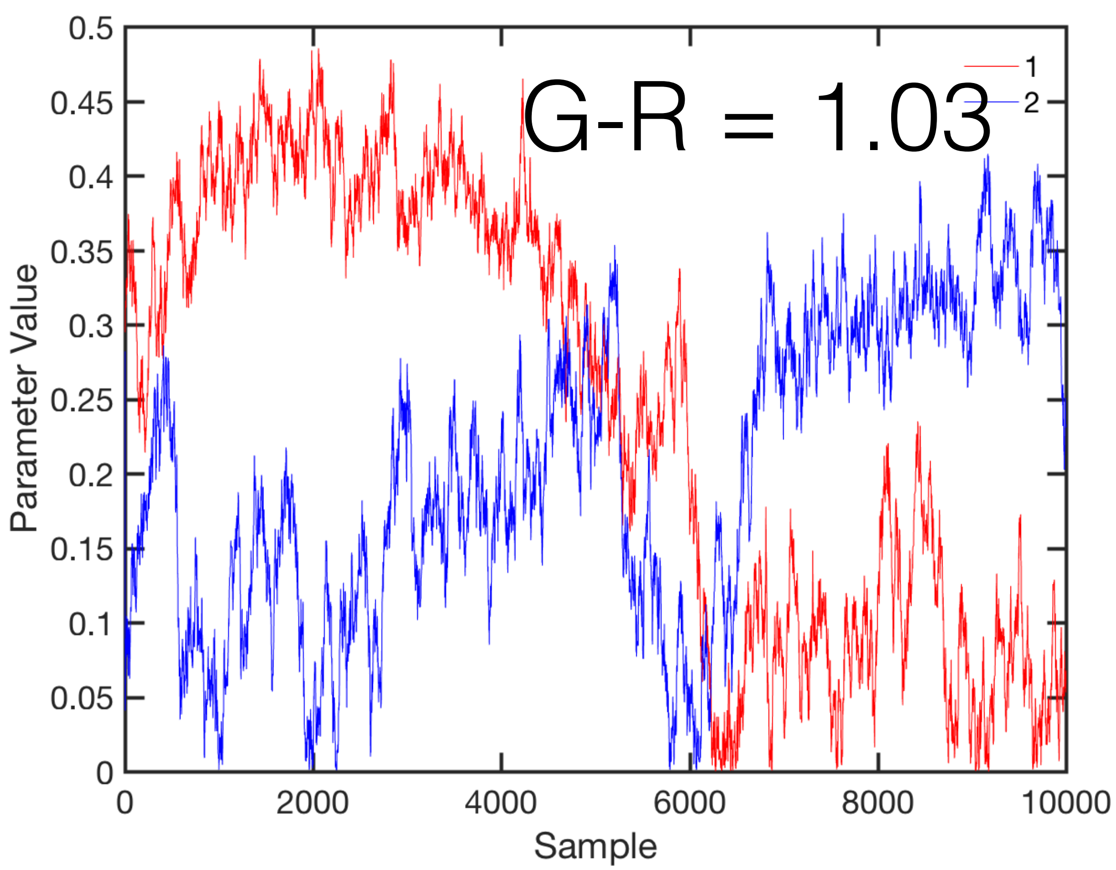
Dark Energy EoS parameter

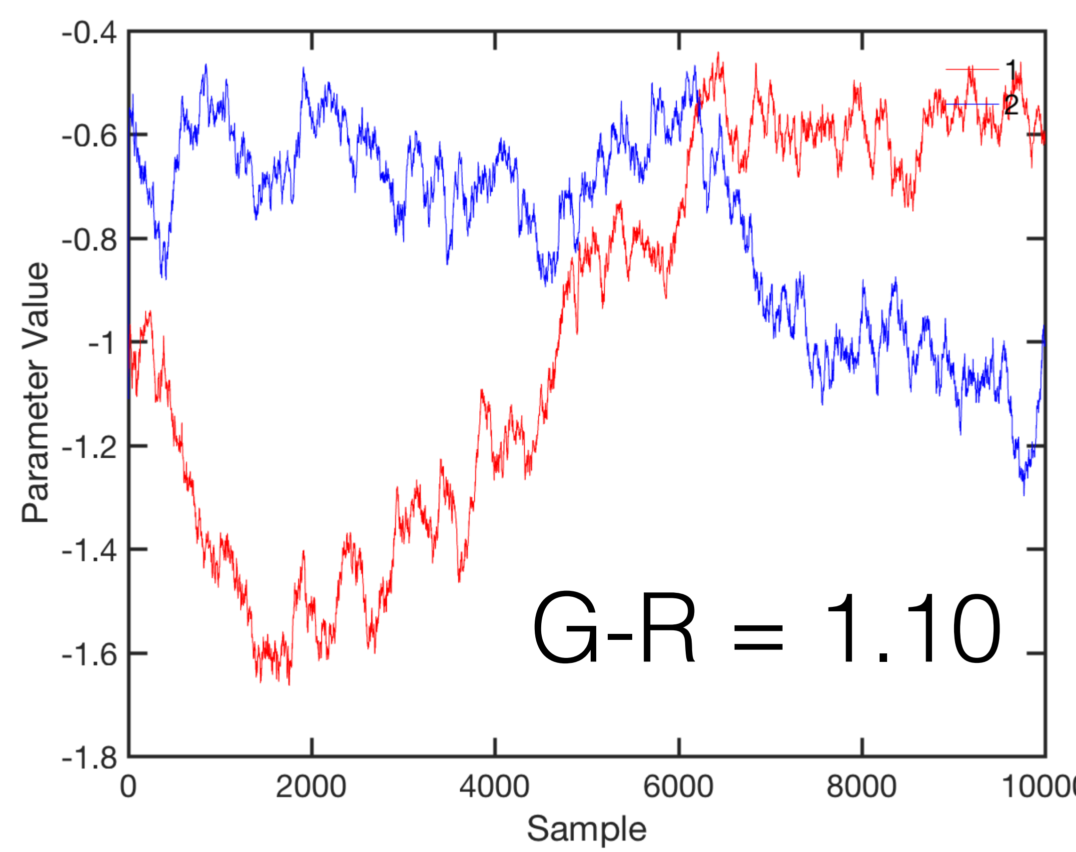
Matter Density

Trace paths 10k cycles, 2 chains



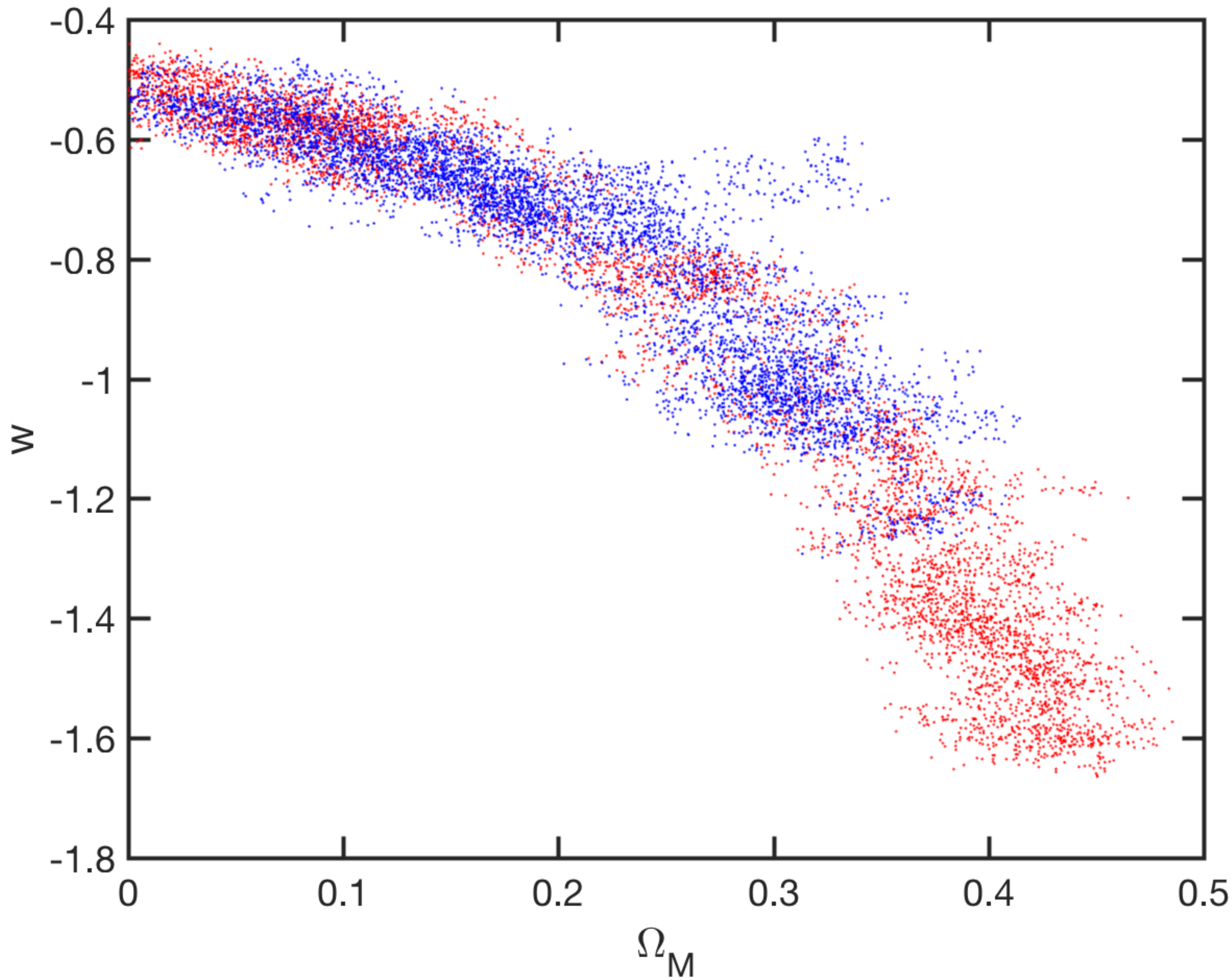






Metropolis-within-Gibbs: 2D trace plot
10,000 steps, 2 chain: Not well mixed!

Dark Energy EoS parameter

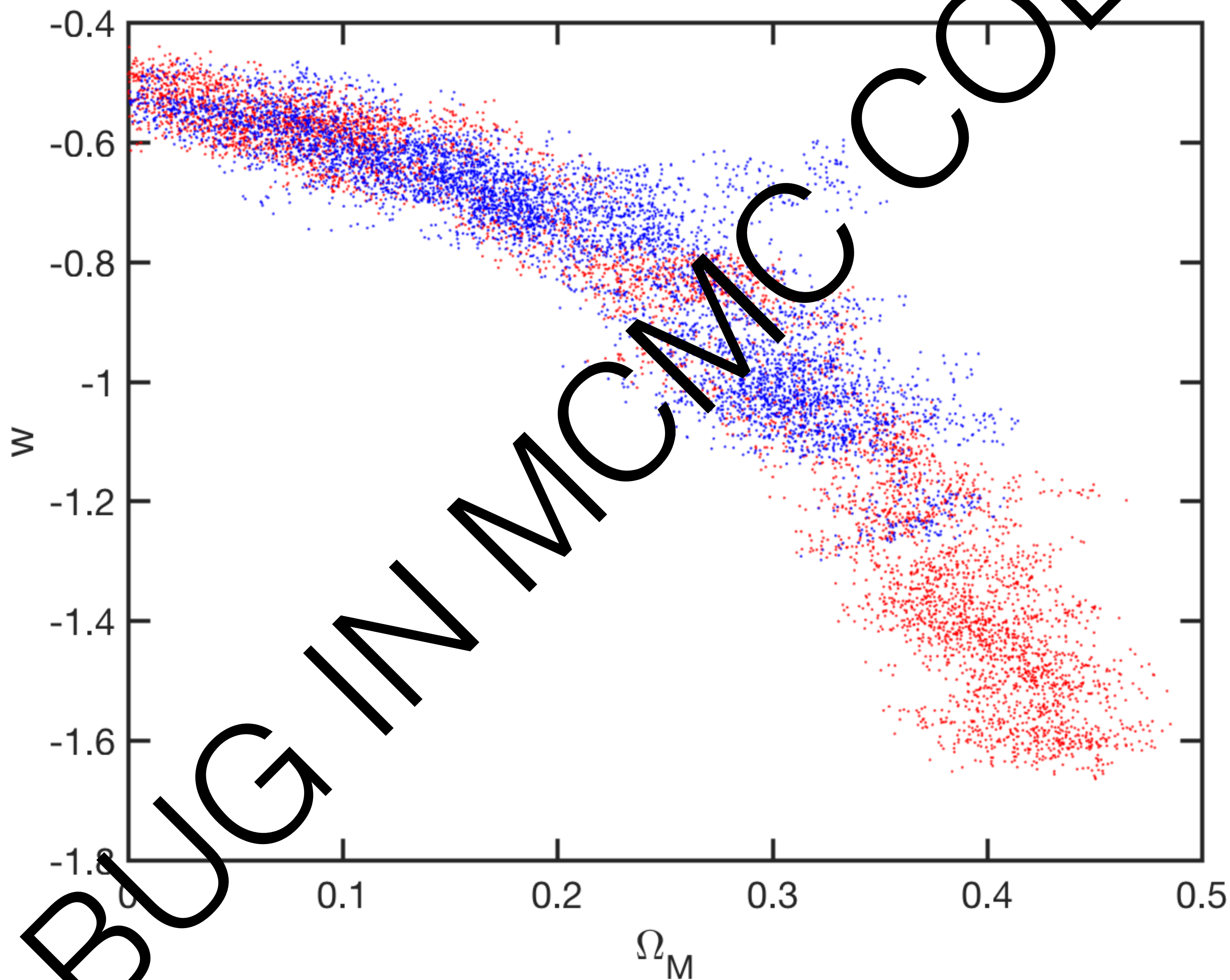


Matter Density

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

BUG IN MCMC CODE!

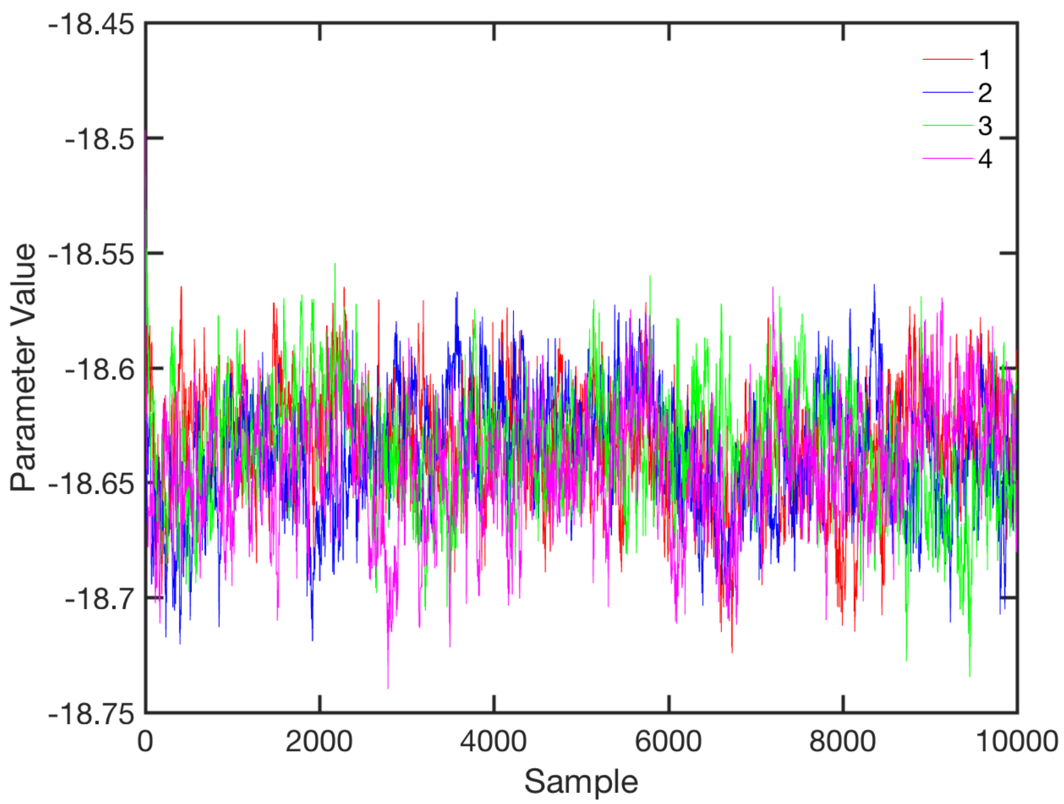
Dark Energy EoS parameter



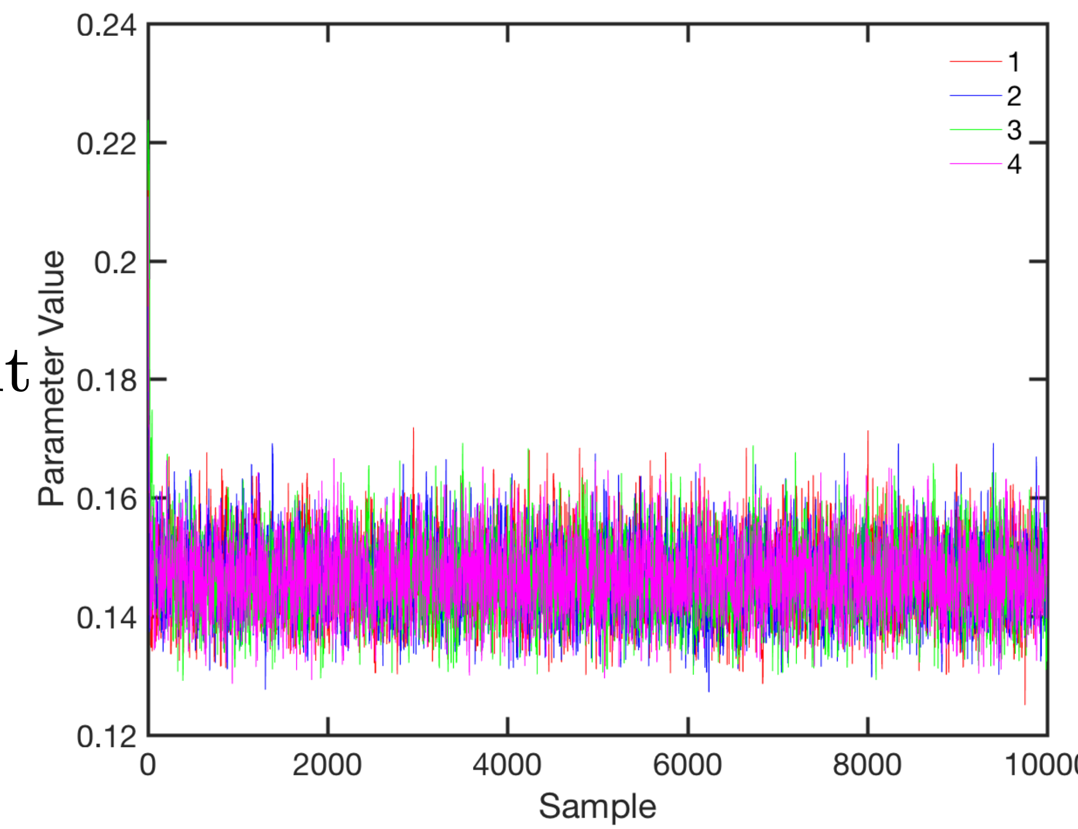
Matter Density

Corrected MwG: Trace paths 10k cycles, 4 chains
Still pretty bad (especially w)!

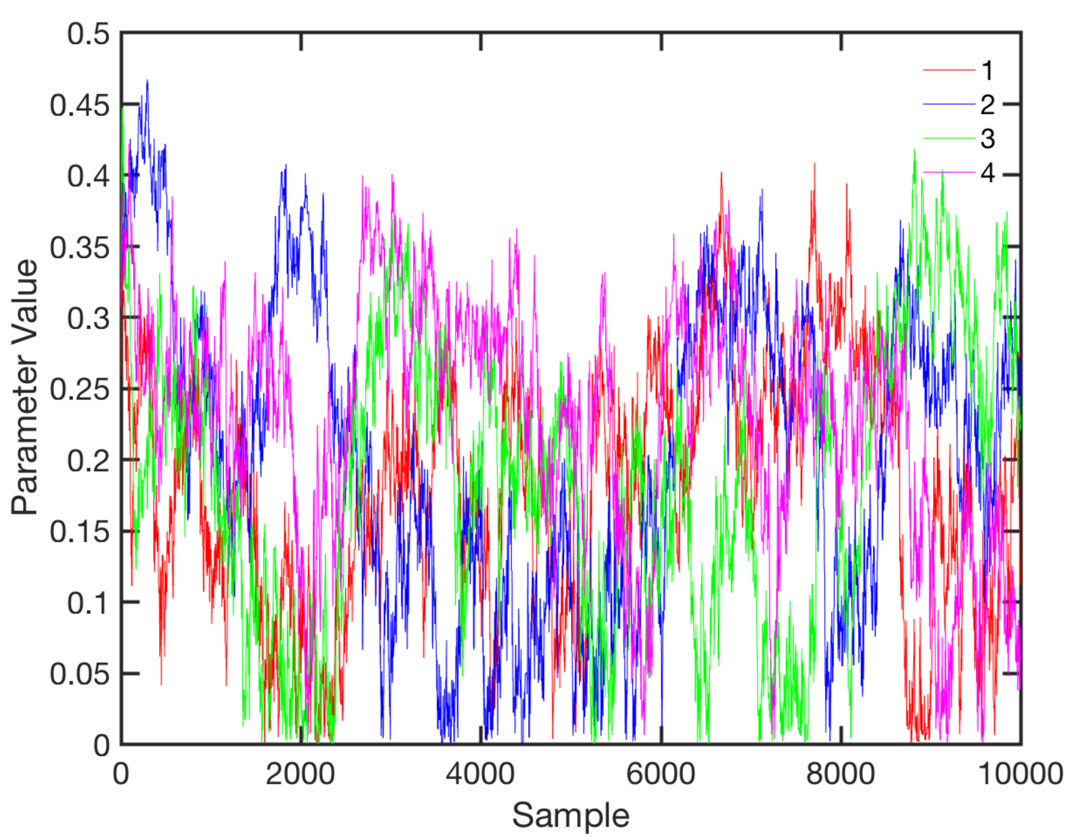
$\mathcal{M}_0$



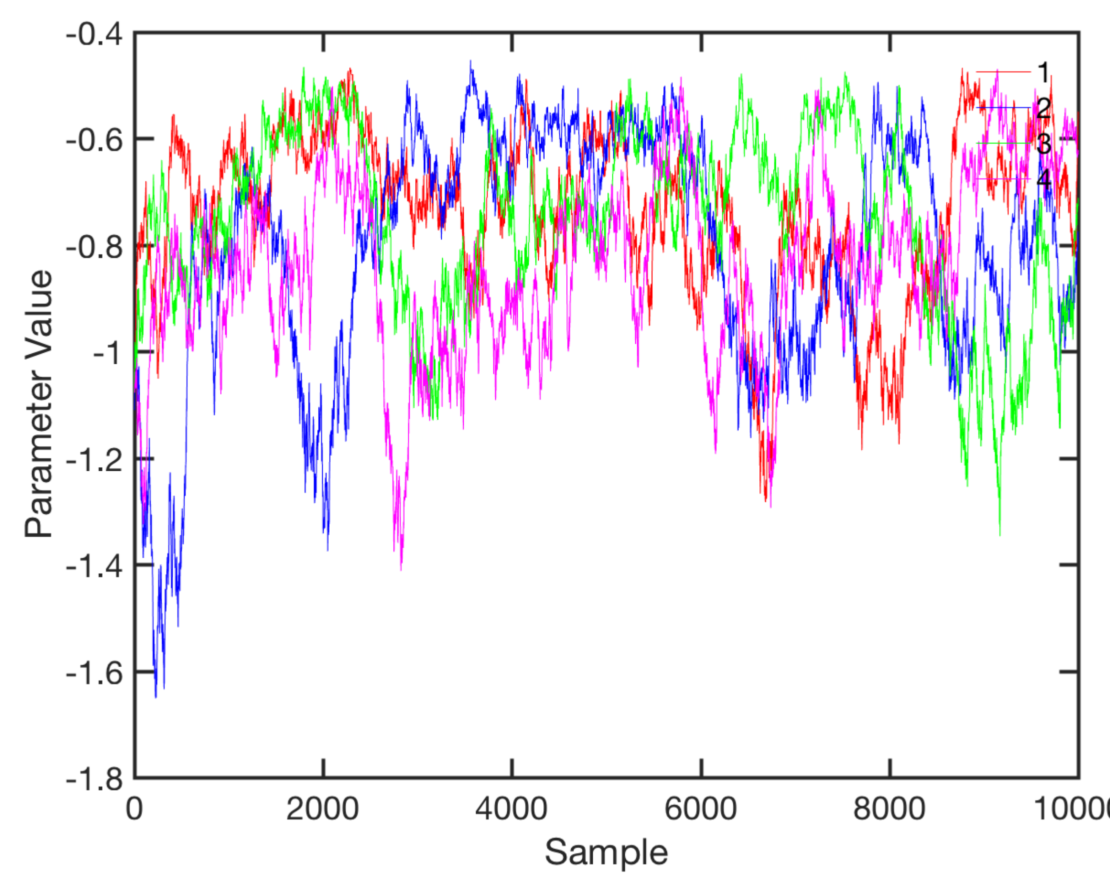
σ_{int}



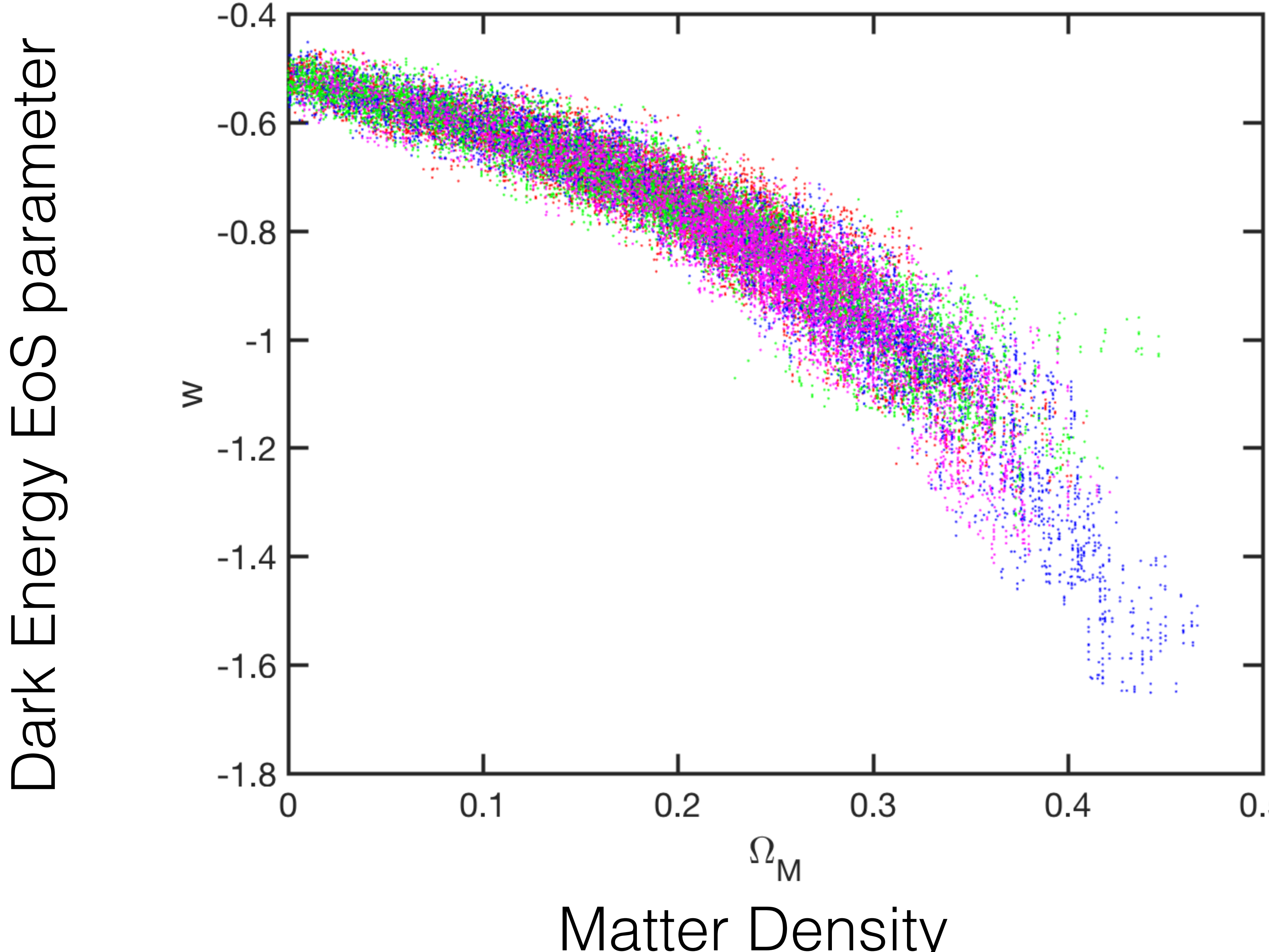
Ω_M



w



Metropolis-within-Gibbs: 2D trace plot
10,000 steps, 4 chain



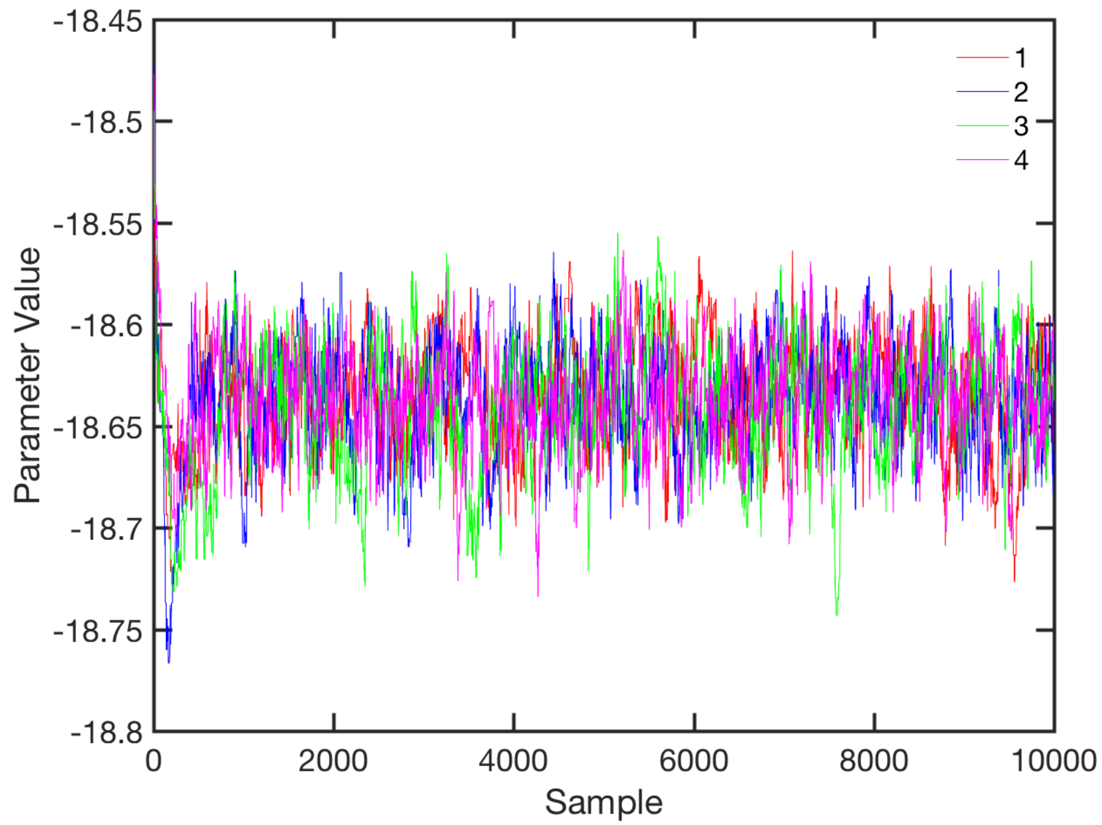
Highly correlated parameters

```
>> std(mc)

ans =

    0.0249    0.0061    0.0953    0.1742

>> corr(mc)

ans =

    1.0000   -0.0132   -0.4534    0.6653
   -0.0132    1.0000    0.0290   -0.0282
   -0.4534    0.0290    1.0000   -0.9337
    0.6653   -0.0282   -0.9337    1.0000
```

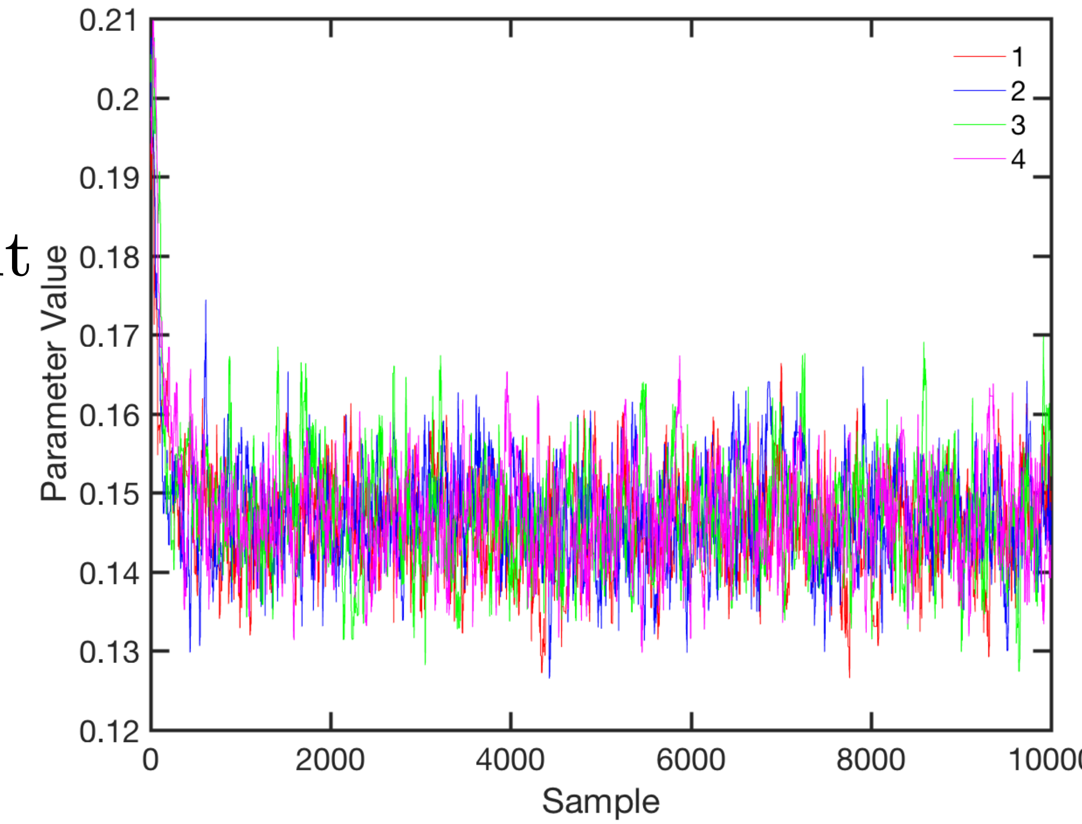
But can use as a correlated proposal distribution in 4D Metropolis! (show code, why better?)

4D Metropolis: Trace paths 10k cycles, 4 chains
Better!

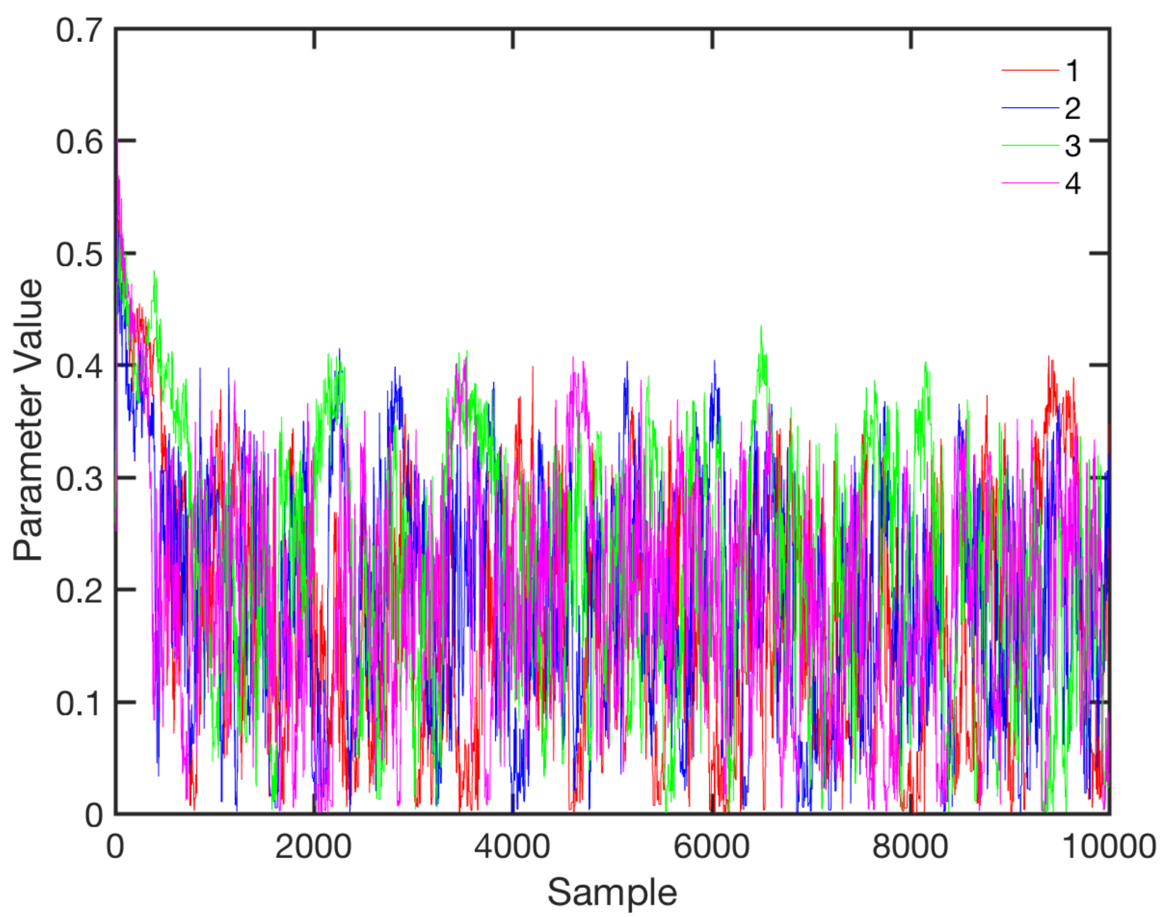
$\mathcal{M}_0$



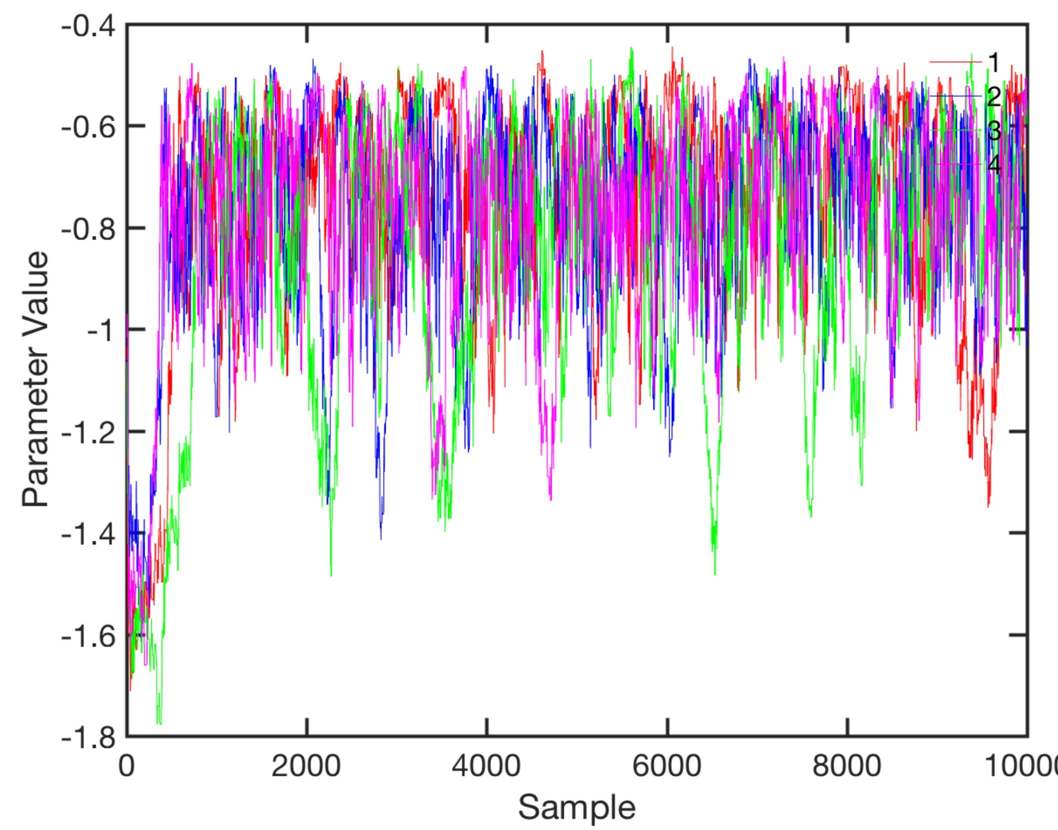
σ_{int}



Ω_M



w



Autocorrelation function

For each scalar parameter θ

$\hat{\rho}_t = C_t / C_0$ Sample Autocorrelation

$$C_t = \frac{1}{N-1} \sum_{i=1}^{N-t} (\theta_i - \bar{\theta})(\theta_{i+t} - \bar{\theta})$$

$$C_0 = \text{Sample Variance of } \theta$$

$\tau = 1 + 2 \sum_{t=1}^{\infty} \rho_t$ Autocorrelation time

Effective Sample Size

$$\hat{\tau} = 1 + 2\sum_{t=1}^{T} \hat{\rho}_t$$ Estimated Autocorr time

$$\text{Truncate at } T \text{ lags, so that } \hat{\rho}_T \approx 0.1$$

Effective number of independent samples

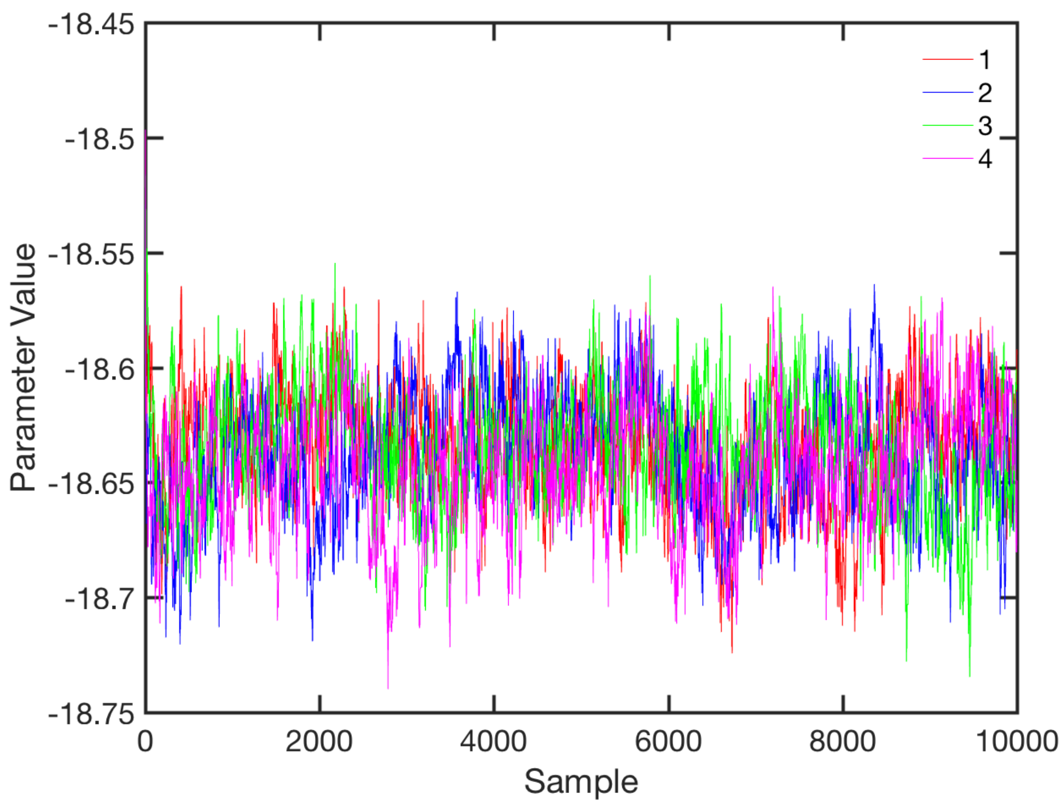
$$N_{\text{eff}} = N/\hat{\tau}$$

Can be computed for a single chain
(see Gelman BDA 3rd for multi-chain generalisation)

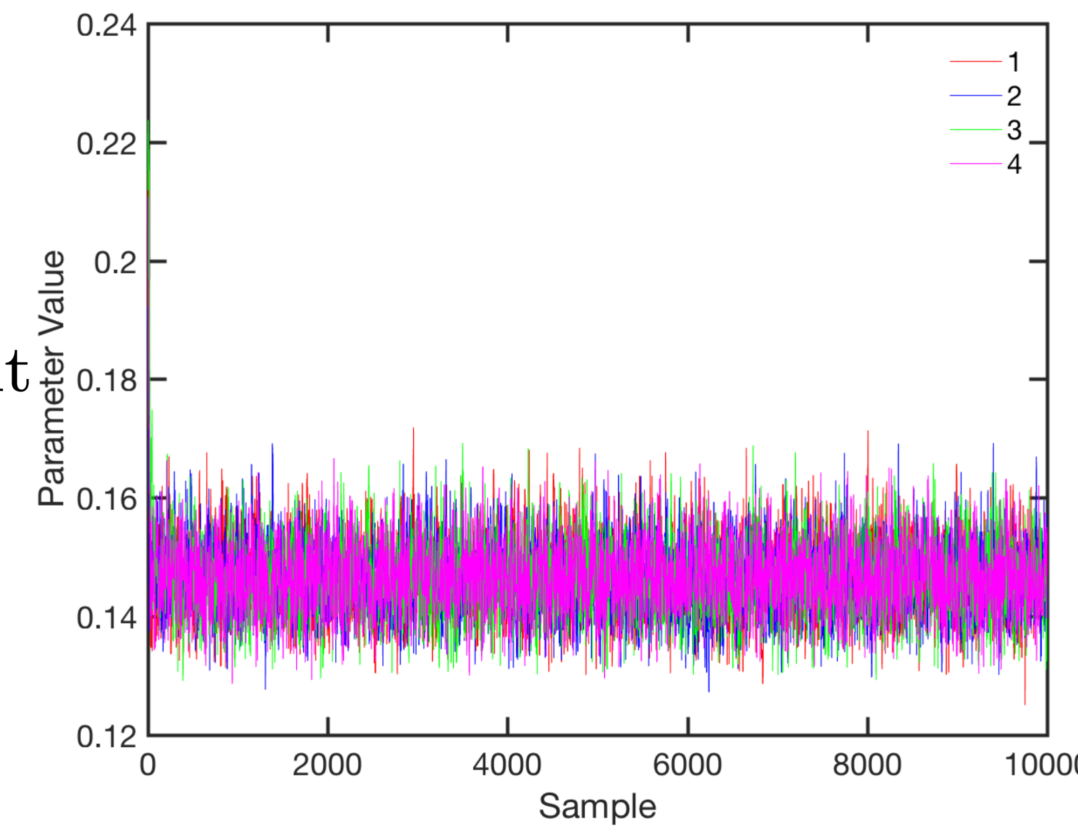
Slowest parameter is the limiting one!

Corrected MwG: Trace paths 10k cycles, 4 chains
Still pretty bad (especially w)!

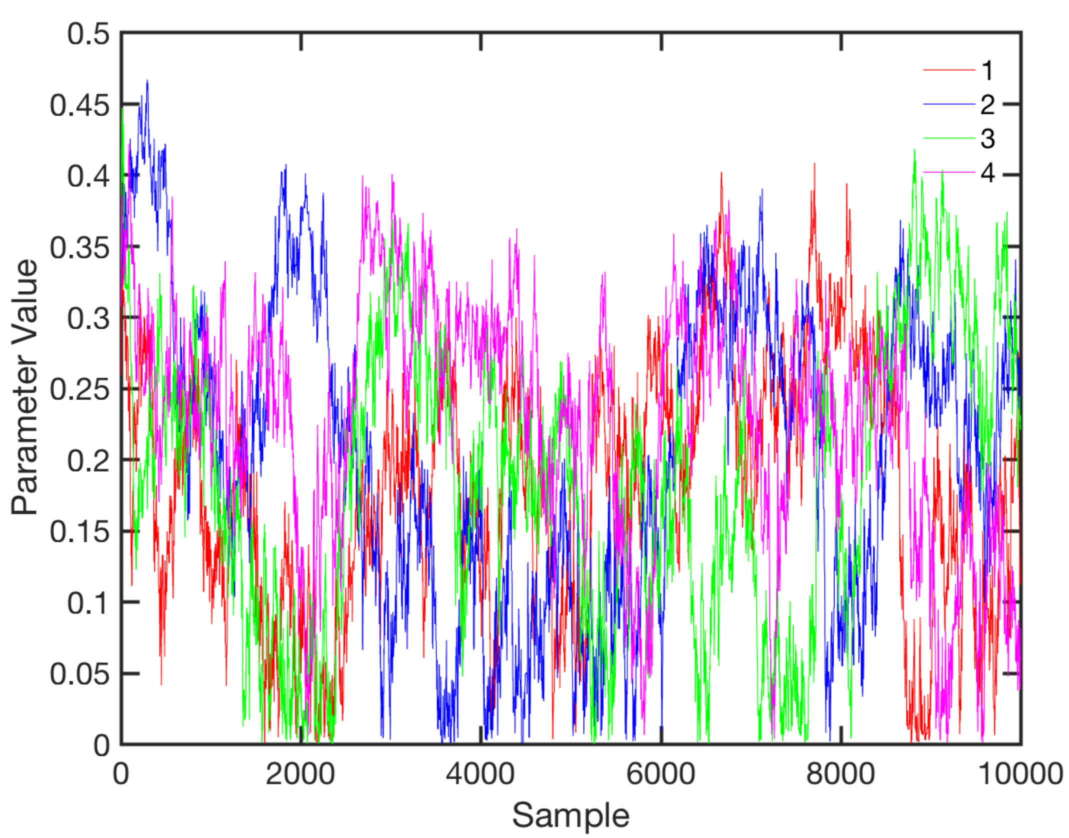
$\mathcal{M}_0$



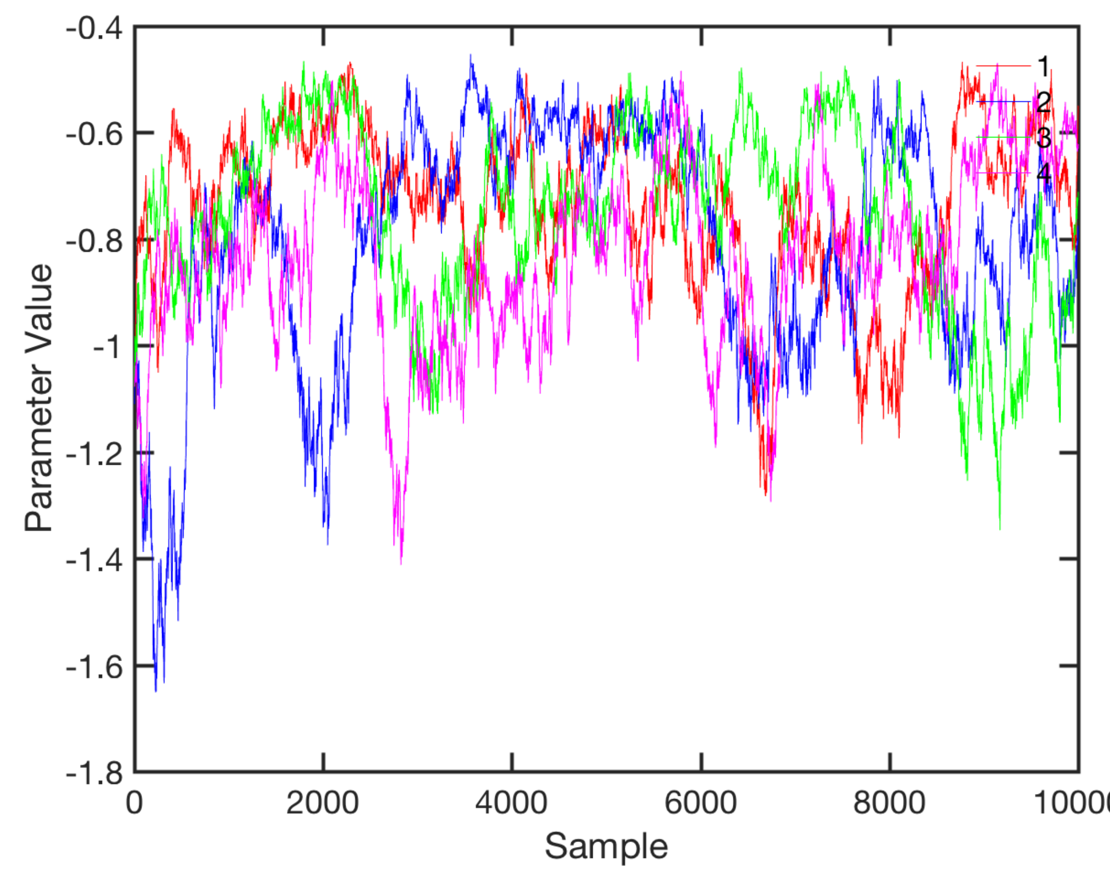
σ_{int}



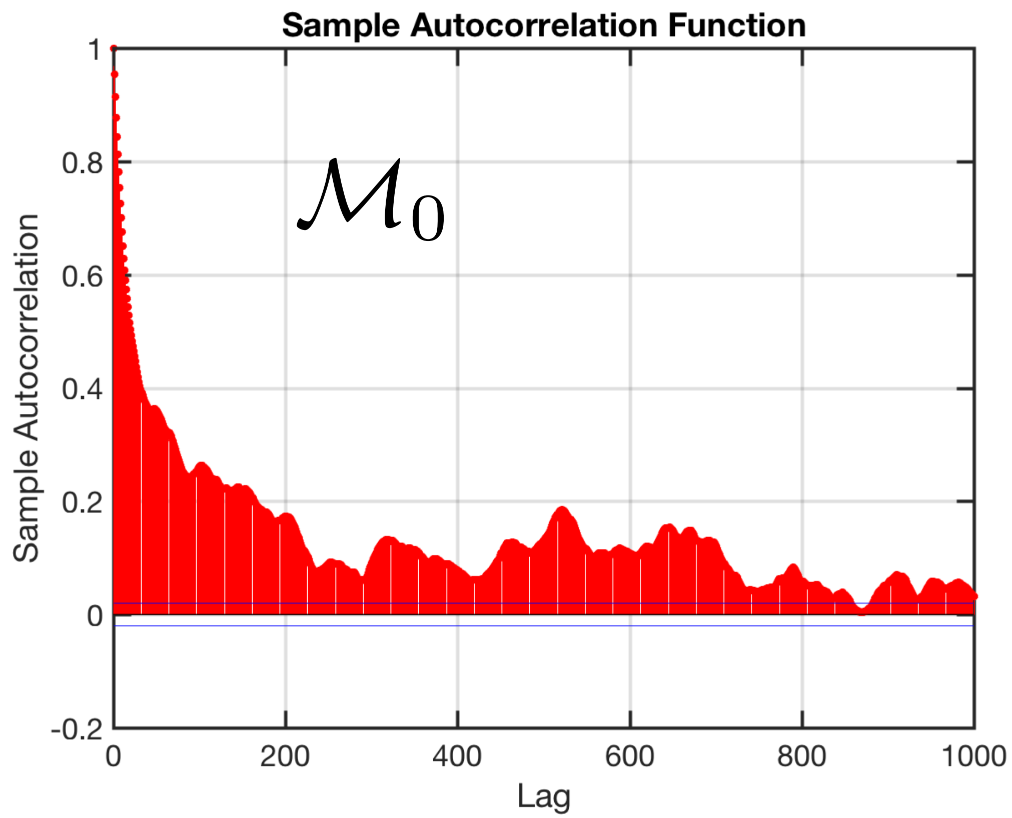
Ω_M

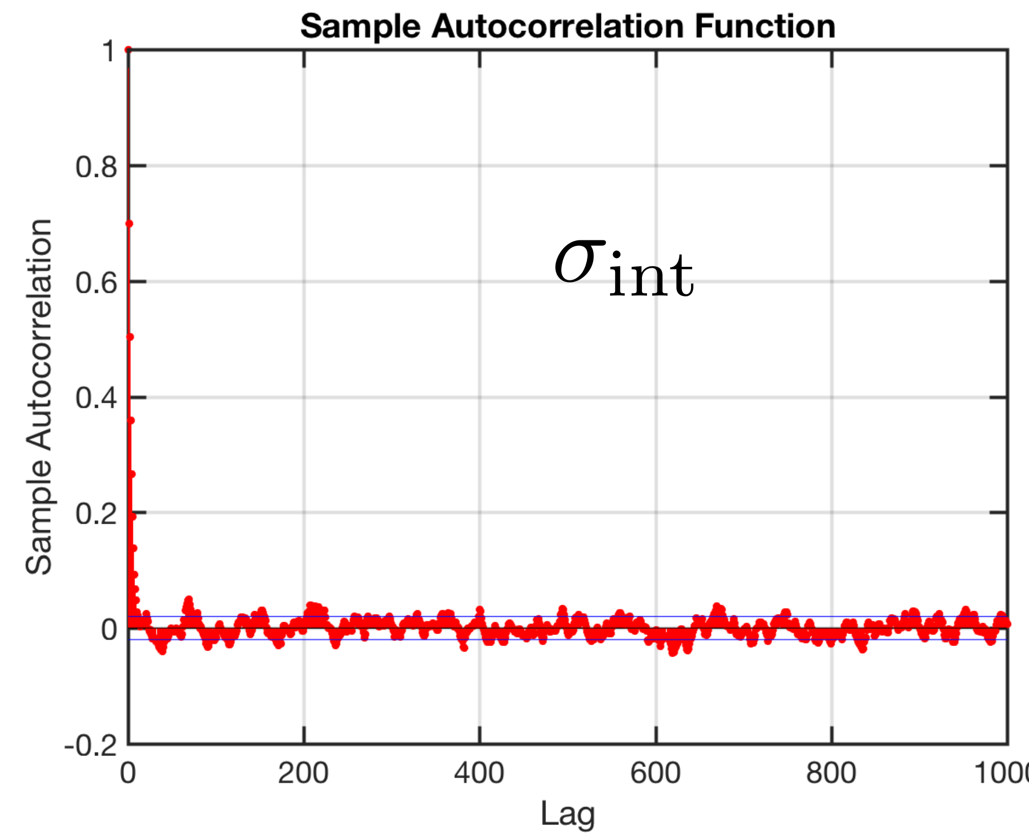


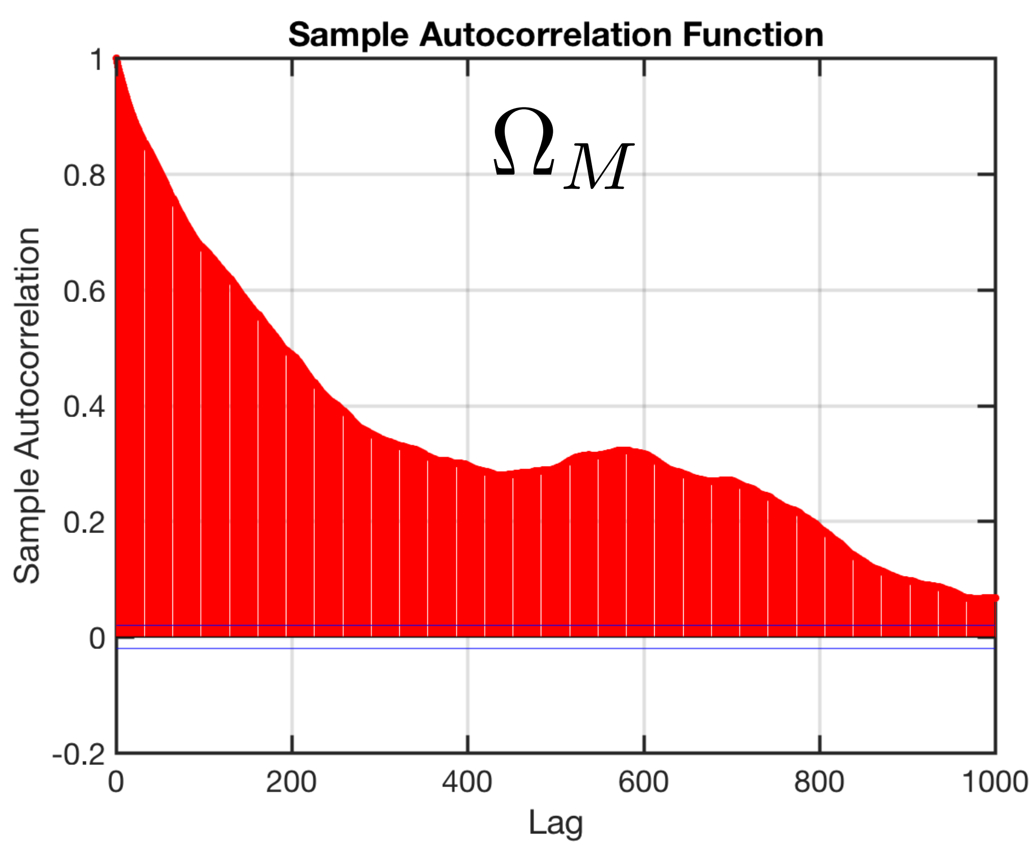
w

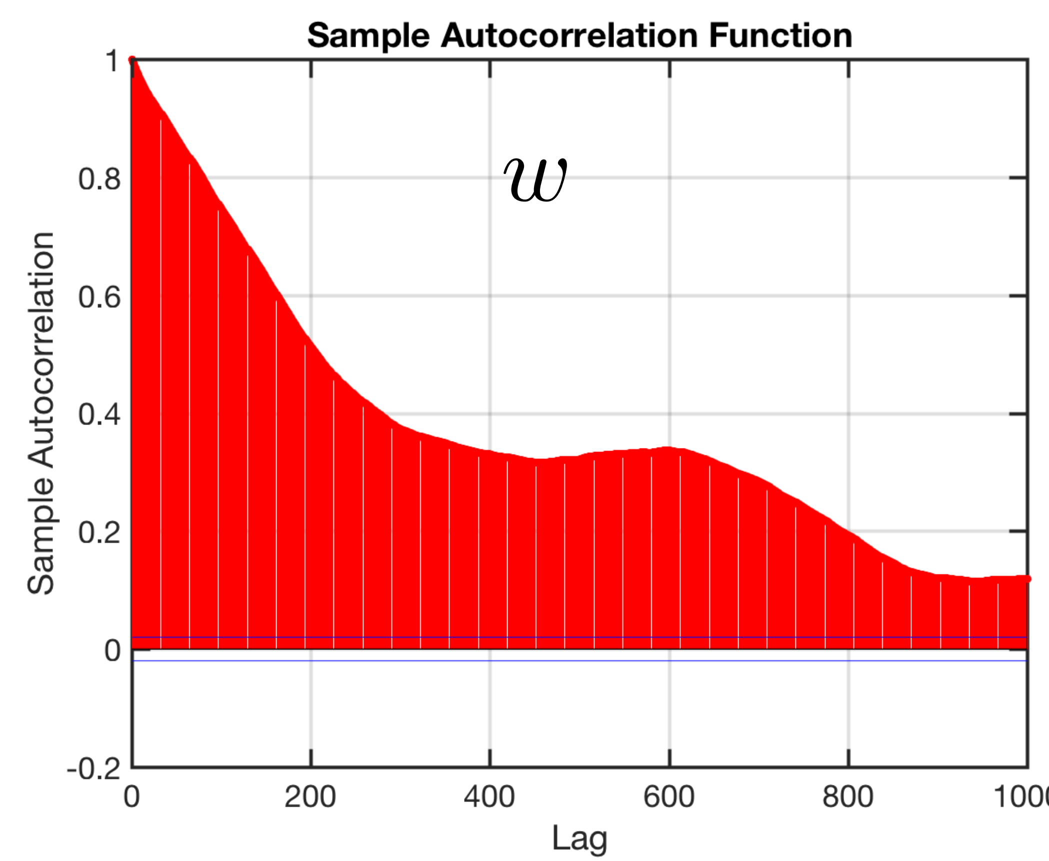


Corrected MwG: Autocorr over 1000 lags

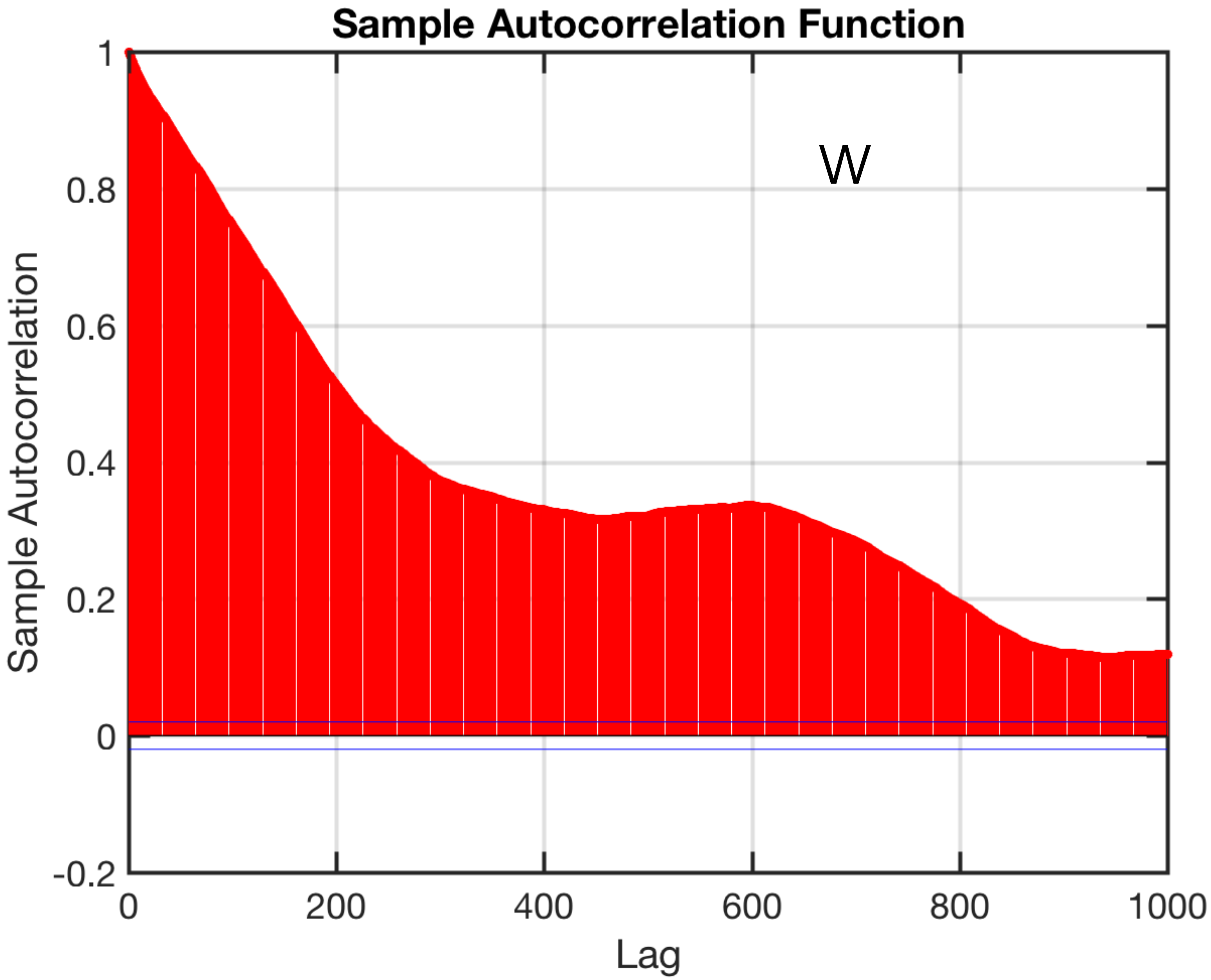








Metrop-w/in-Gibbs: autocorrelation fcn of w (slowest mixing parameter)



$$\hat{\tau} = 1 + 2\sum_{t=1}^{T} \hat{\rho}_t$$

$$\hat{\tau} = 751 \text{ cycles}$$

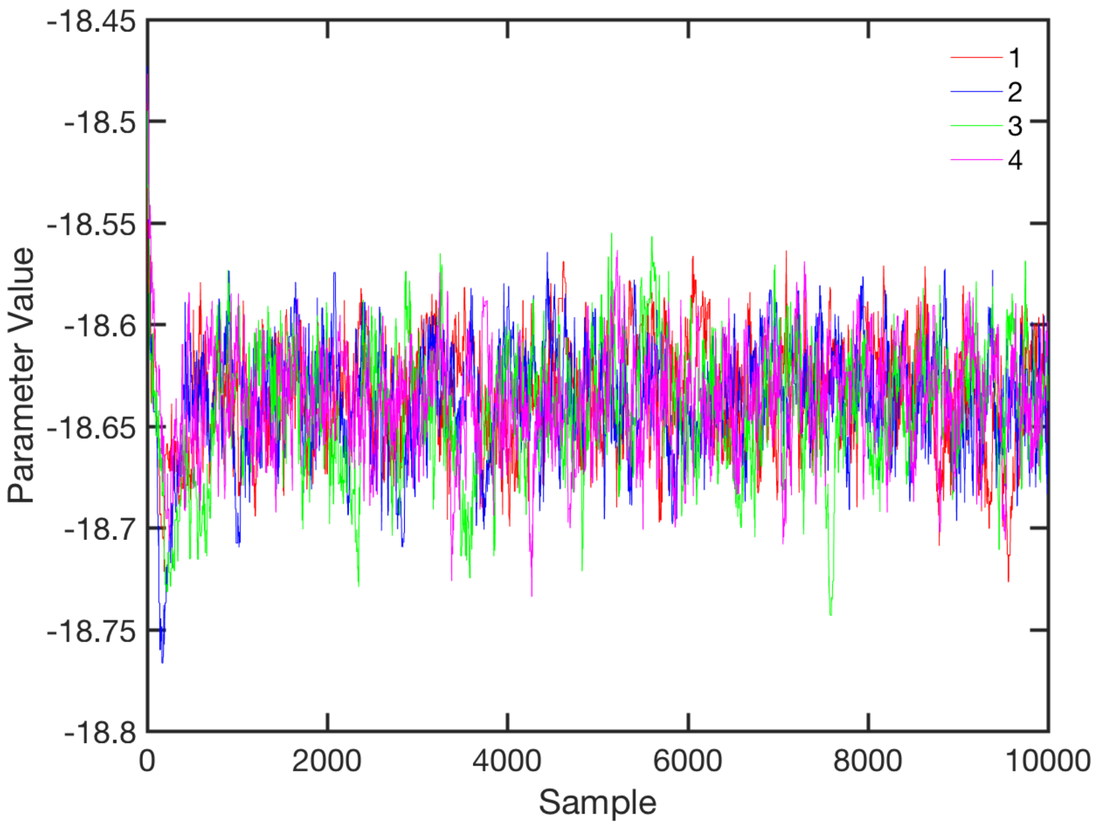
(single chain)

$$N_{\text{eff}} = 10,000/751 = 13 \text{ indep. samples}$$

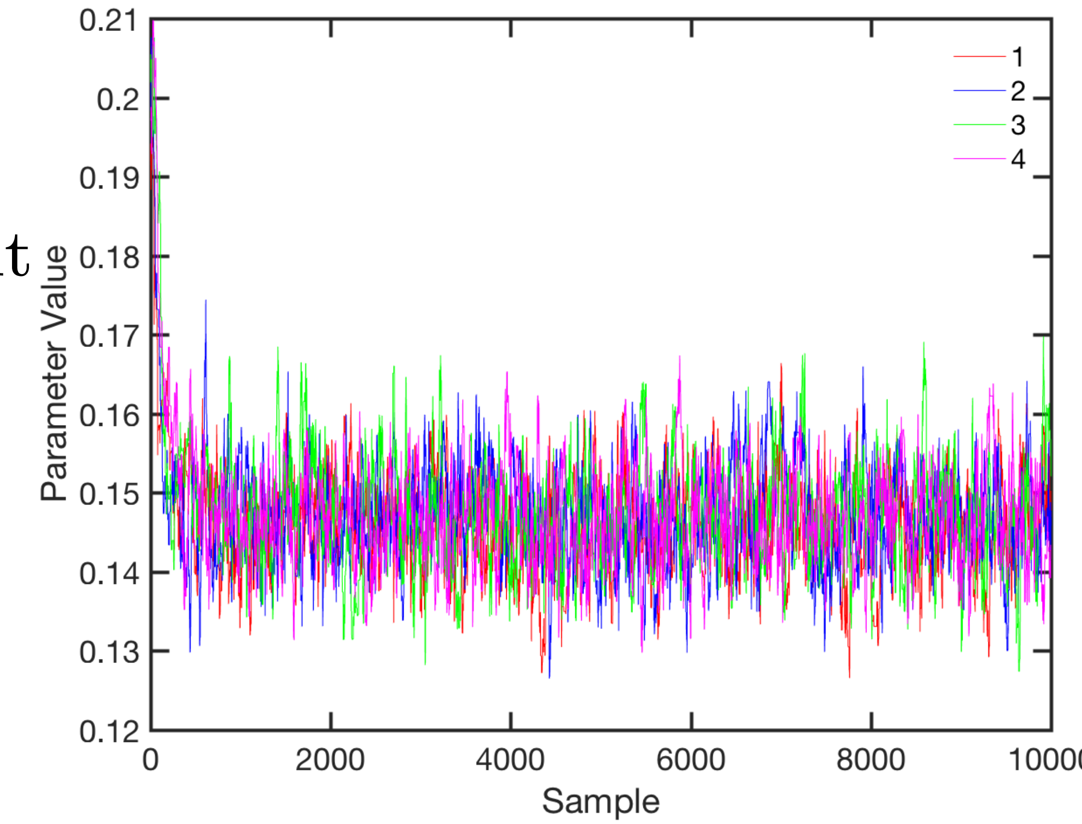
Rate = 552 sec / 13 = 42.5 sec per independent sample

4D Metropolis: Trace paths 10k cycles, 4 chains
Better!

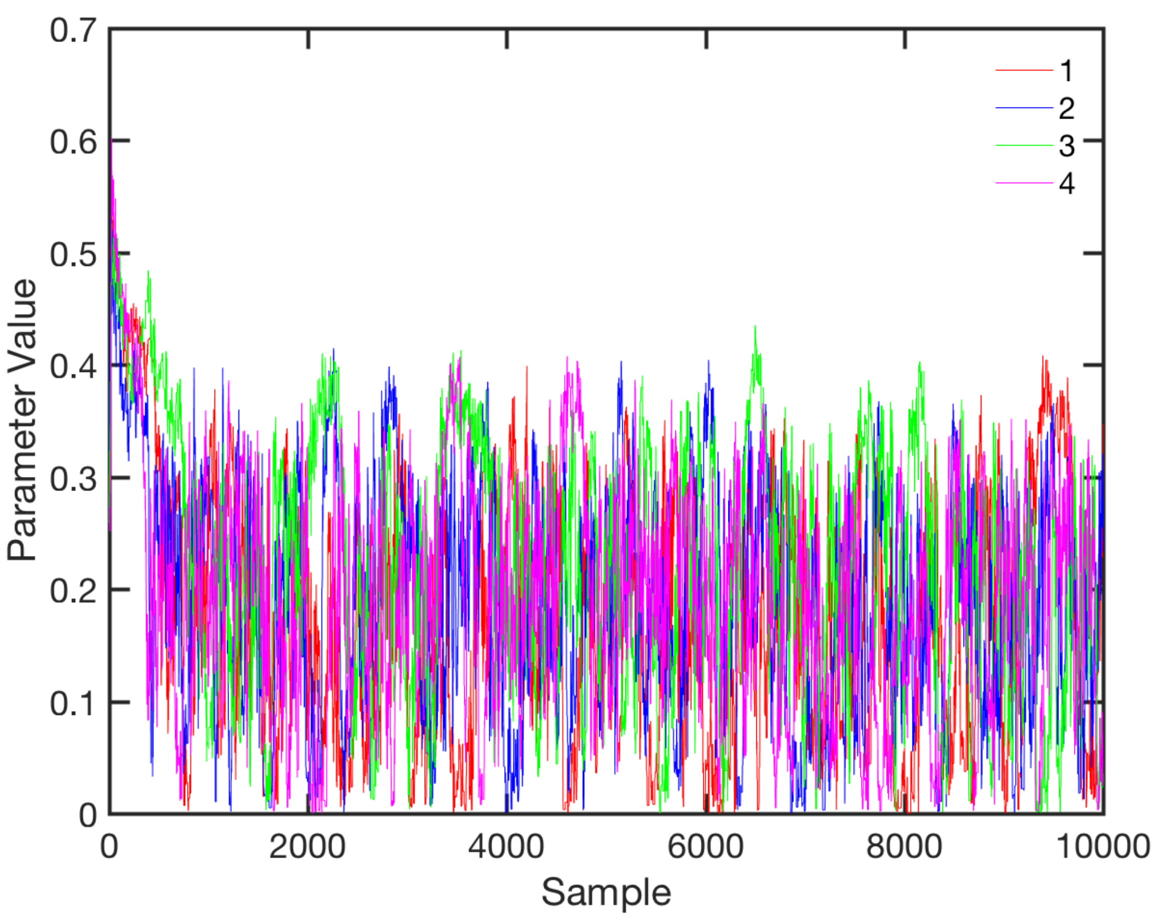
$\mathcal{M}_0$



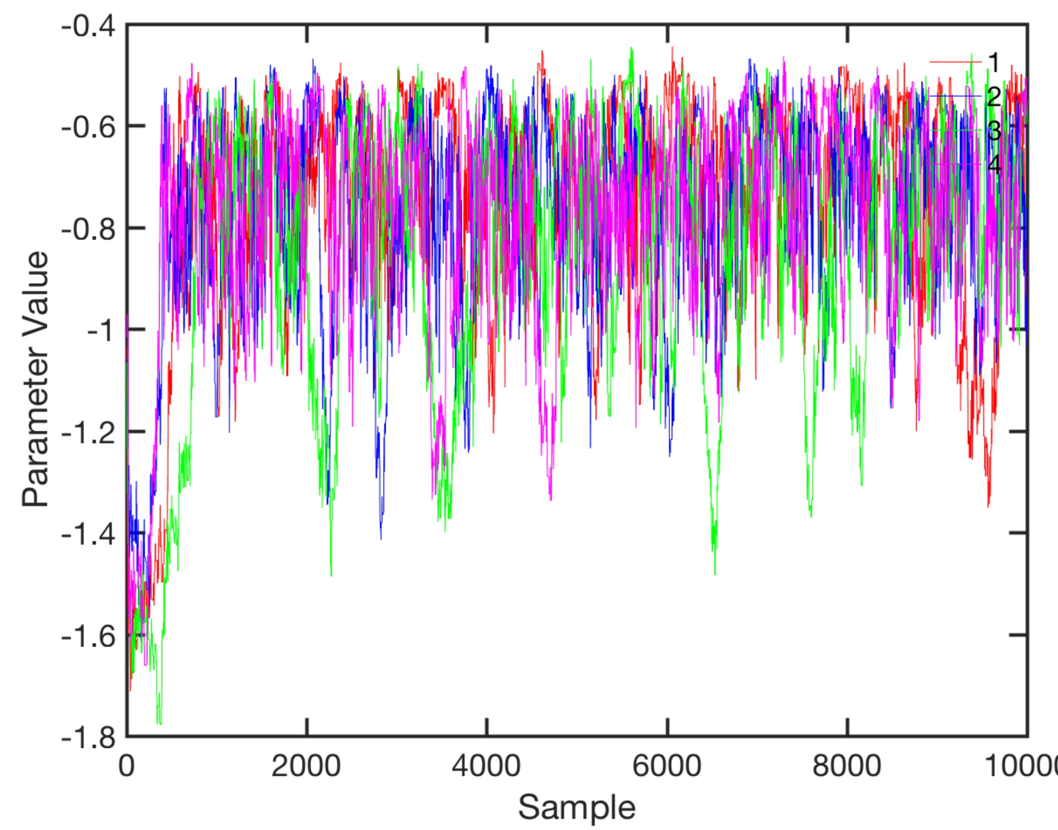
σ_{int}



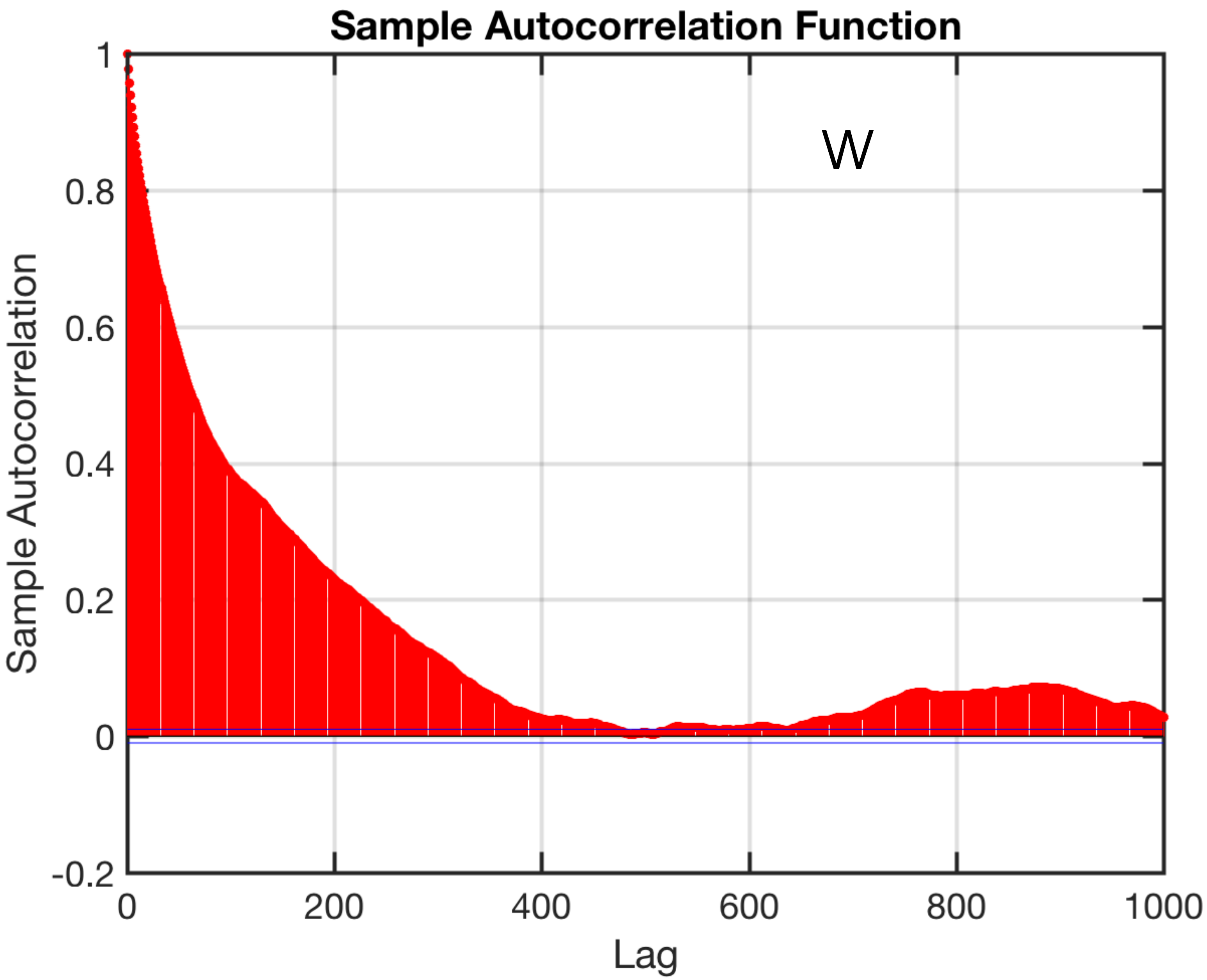
Ω_M



w



4D Metropolis: autocorrelation fcn of w (slowest mixing parameter)



$$\hat{\tau} = 1 + 2\sum_{t=1}^{T} \hat{\rho}_t$$

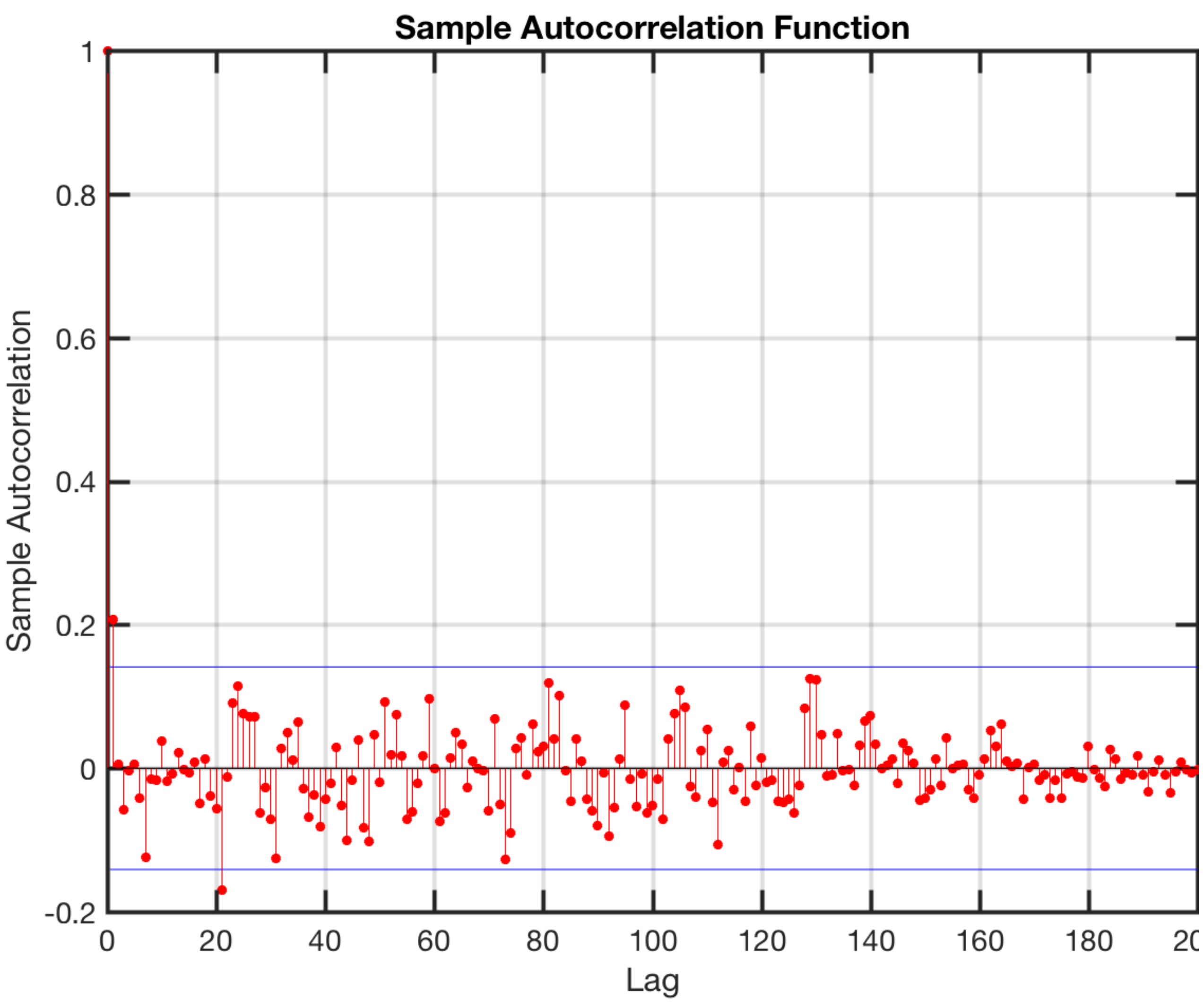
$$\hat{\tau} = 215 \text{ iterations}$$

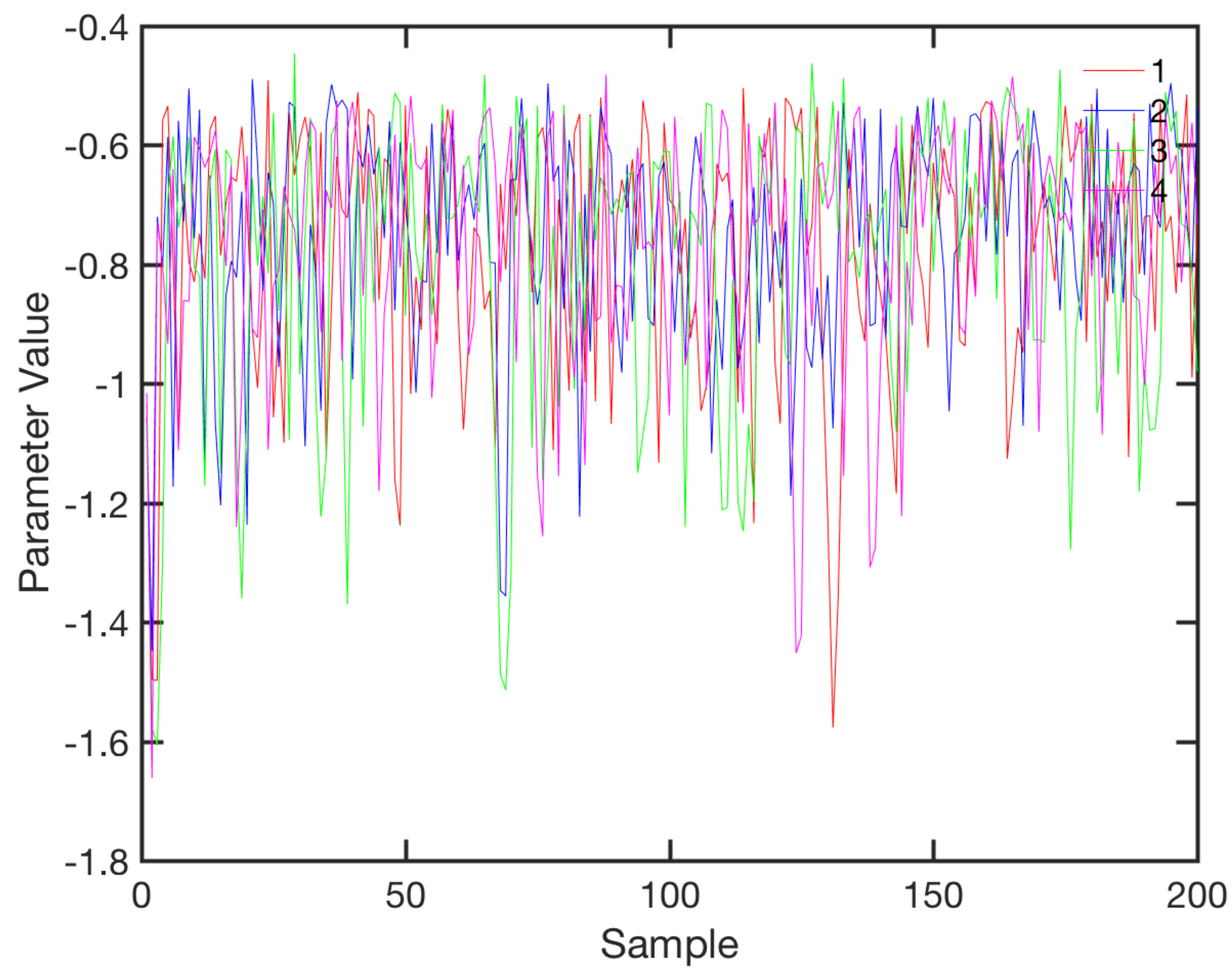
(single chain)

$$N_{\text{eff}} = 10{,}000/215 = 46 \text{ indep. samples}$$

Rate = 140 sec / 46 = 3.0 sec per independent sample
14x faster than MwG!

4D Metropolis: autocorrelation fcn of w after thinning by 200

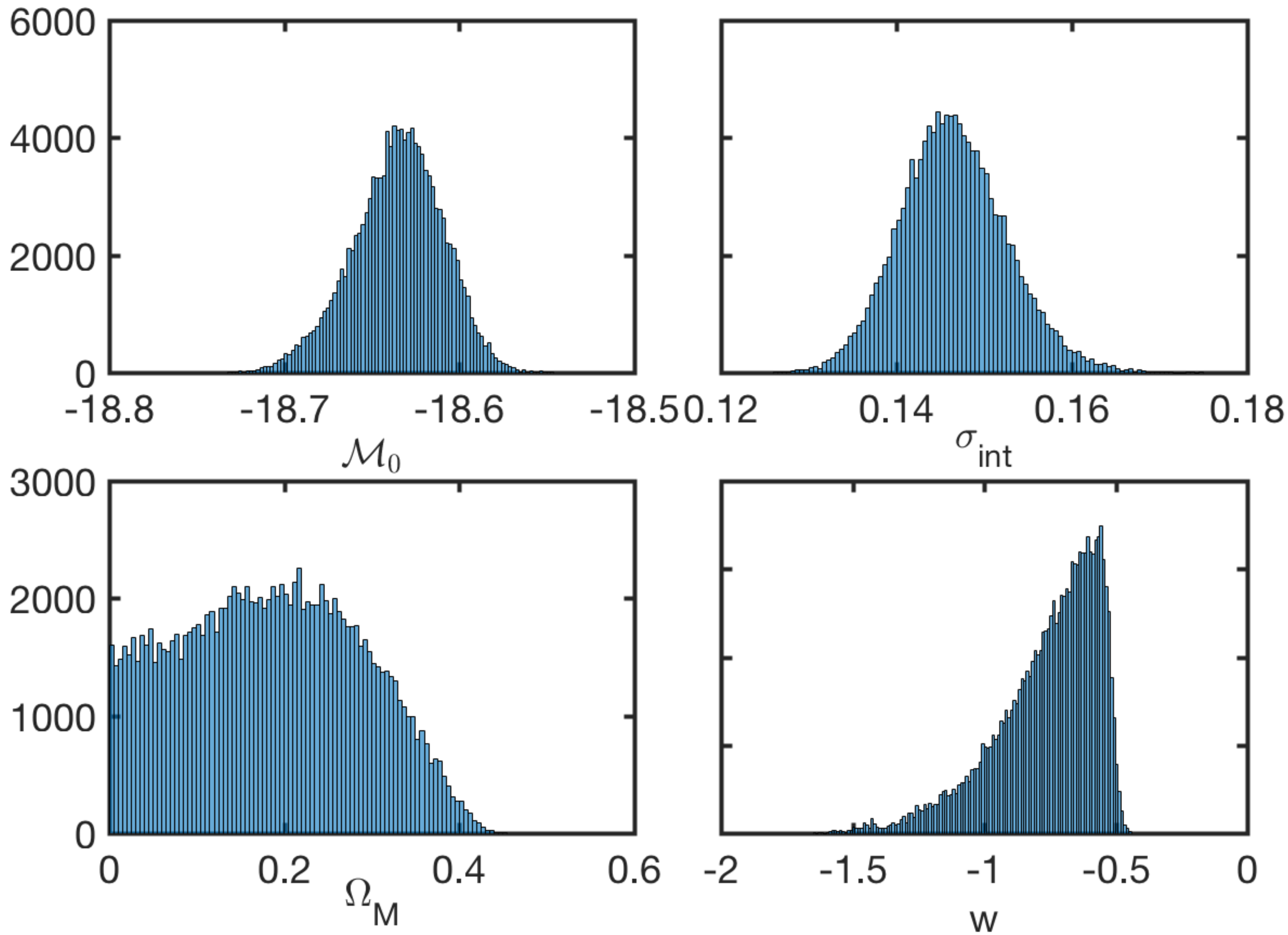




w

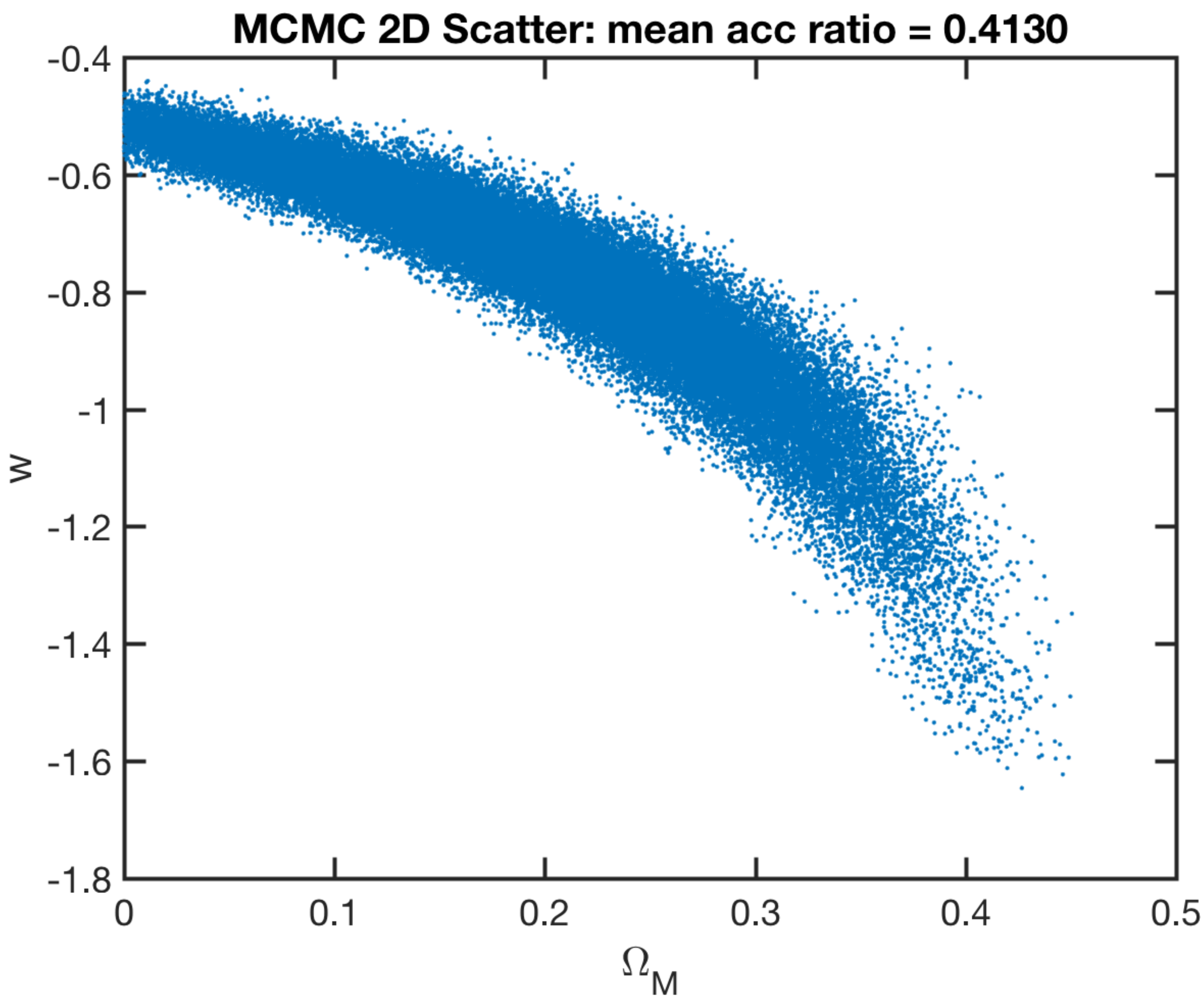
4D Metropolis run with 4x10k chains

Marginal Posterior Histograms



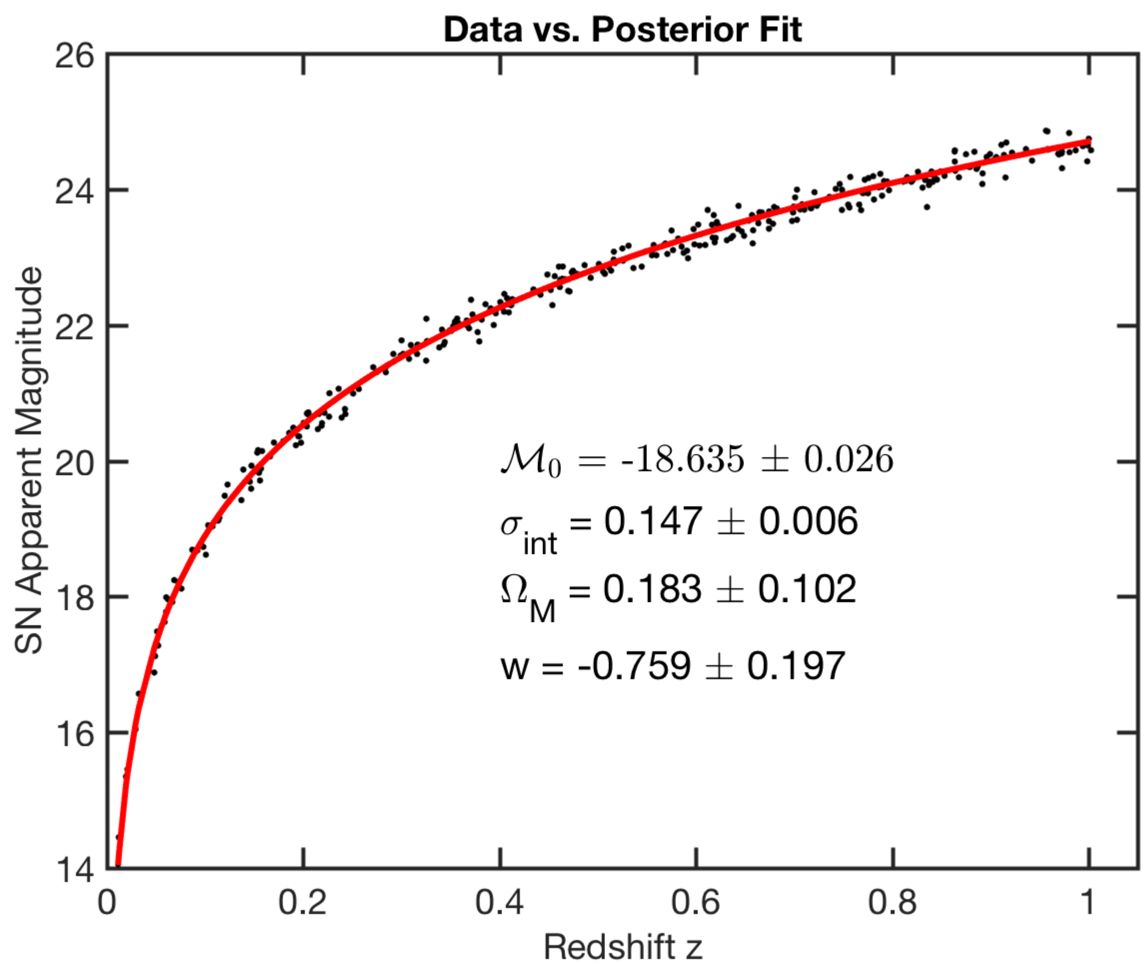
4D Metropolis run with 4x10k chains

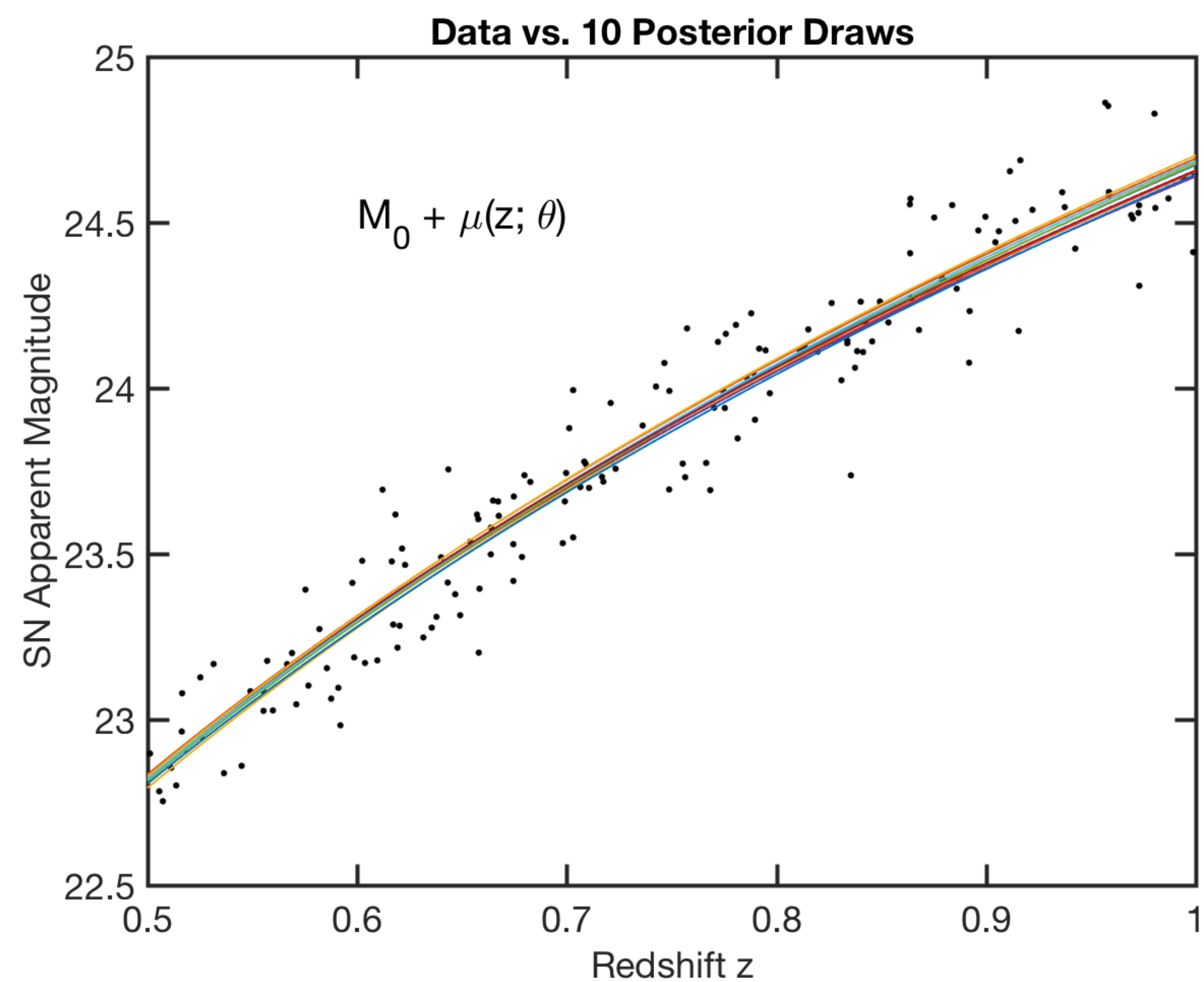
Dark Energy EoS parameter



Matter Density

4D Metropolis run with 4x10k chains





MCMC Animations

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

Review: 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

Review: 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
  $J(\theta^* | \theta_i) = N(\theta^* \mid \theta_i, \mathbf{\Sigma_p}) = N(\theta_i \mid \theta^*, \mathbf{\Sigma_p}) = J(\theta_i | \theta^*)$
  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 sometimes engineer a clever proposal $J(\theta^*|\theta_i)$

Gibbs sampling as special case of Metropolis-Hastings

Mix-n-Match Metropolis/Gibbs

A sketch of MCMC theory