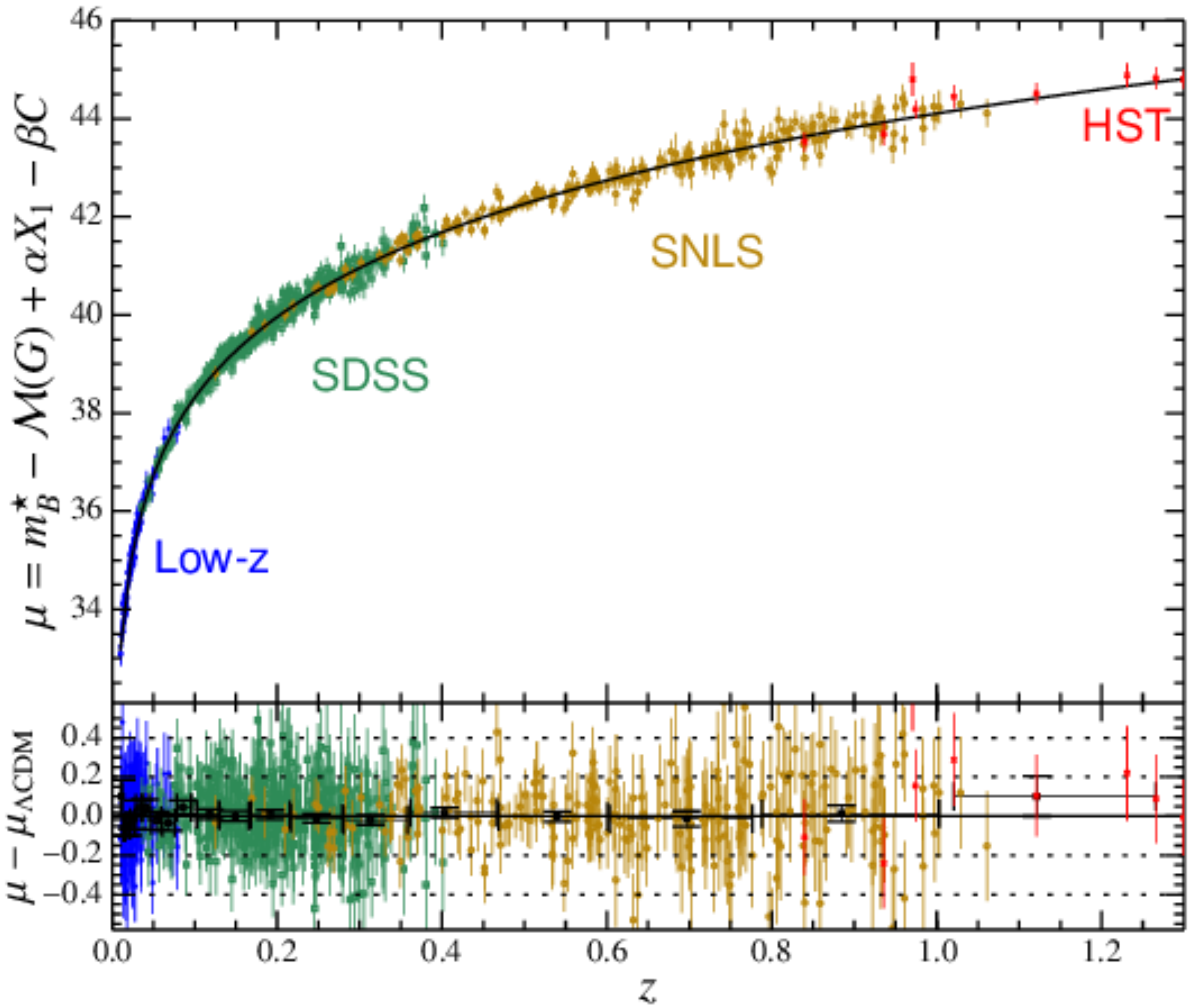
Astrostatistics: Wed 26 Feb 2020

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2020

- Today: continue Bayesian computation / MCMC
  - MacKay: Ch 29-30; Bishop: Ch 11; Gelman BDA
  - Givens & Hoeting. "Computational Statistics"
    (Free online through Cambridge Library iDiscover)
- Example Sheet 1 solutions & code posted
- Lecture MCMC code examples online: lecture_codes/
- Example Class 2, Thu Feb 27, 3:30pm MR13

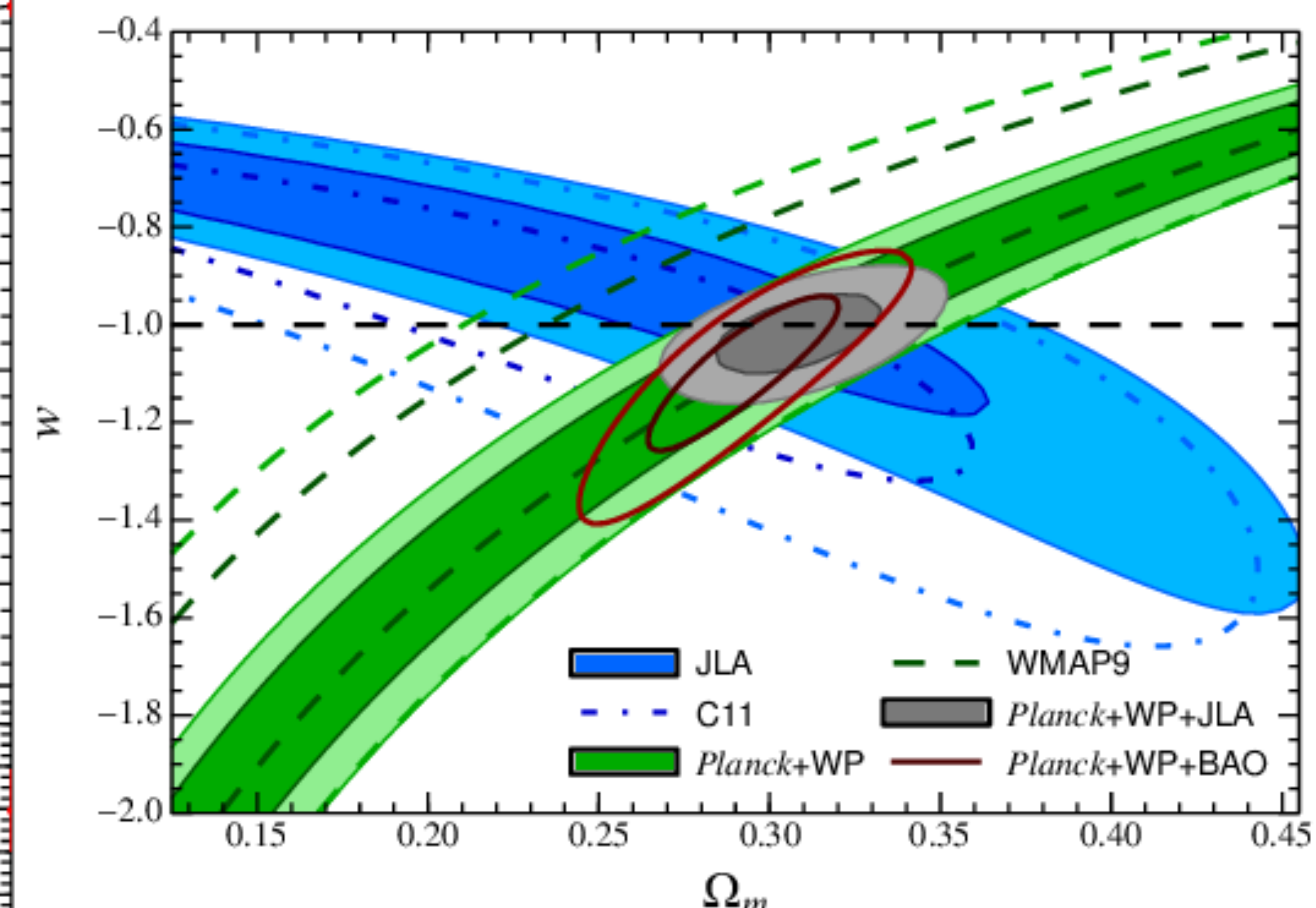
Type Ia SN Cosmology

Hubble Diagram
Modern SN Ia Surveys



Joint Lightcurve Analysis
(JLA, Betoule et al. 2014)

Cosmological Constraints



$w = -1.027 \pm 0.055$ (stat+sys)

Cosmology with standard candles

For us: Simplify:
Let's say Supernovae are *standard* candles

$$M_s \sim N(M_0, \sigma^2_{\mathrm{int}})$$ SN Population Distribution

$$m_s = M_s + \mu(z_s; H_0, \mathbf{\Omega})$$ (Log) Inverse Square Law

$$\mathbf{\Omega} = (\Omega_M, \Omega_\Lambda, w)$$

Cosmological Parameters:

- Hubble Constant H_0
- Matter Density Ω_M
- Dark Energy Density Ω_Λ
- DE EoS parameter w

Assume apparent magnitudes and redshifts
$\{m_s, z_s\}$ measured perfectly

Theoretical Distances

1 Comoving Distance

The dimensionless comoving distance $\tilde{d}$ to an object with redshift z is:

$$\tilde{d}(z;\Omega_M,\Omega_\Lambda,w) = \int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3+\Omega_k(1+z')^2+\Omega_\Lambda(1+z')^{3(w+1)}}} \tag{1}$$

where Ω_M is the matter density, Ω_Λ is the dark energy density, the curvature is $\Omega_k \equiv 1-\Omega_M-\Omega_\Lambda$, and w is the equation-of-state parameter of dark energy.

In terms of the scale factor $a \equiv (1+z)^{-1}$, the integral can be expressed as:

$$\tilde{d}(z;\Omega_M,\Omega_\Lambda,w) = \int_{(1+z)^{-1}}^1 \frac{da'/a'}{\sqrt{\Omega_M/a'+\Omega_k+\Omega_\Lambda/a'^{(1+3w)}}} \tag{2}$$

(derived from standard FLRW metric)

Theoretical Luminosity Distances (Arbitrary Curvature)

2 Luminosity Distance

The theoretical cosmological distance to an object with redshift z inferred from the standard candle method are given by the luminosity distance d_L. The dimensionless distance is:

$$\tilde{d}_L(z;\Omega_M,\Omega_\Lambda,w) = (1+z)\begin{cases} |\Omega_k|^{-\frac{1}{2}}\sinh[\sqrt{|\Omega_k|}\tilde{d}(z;\Omega_M,\Omega_\Lambda,w), & \Omega_k > 0 \\ \tilde{d}(z;\Omega_M,\Omega_\Lambda,w), & \Omega_k = 0 \\ |\Omega_k|^{-\frac{1}{2}}\sin[\sqrt{|\Omega_k|}\tilde{d}(z;\Omega_M,\Omega_\Lambda,w), & \Omega_k < 0 \end{cases} \tag{3}$$

For a given value of the Hubble constant, the dimensionful luminosity distance is

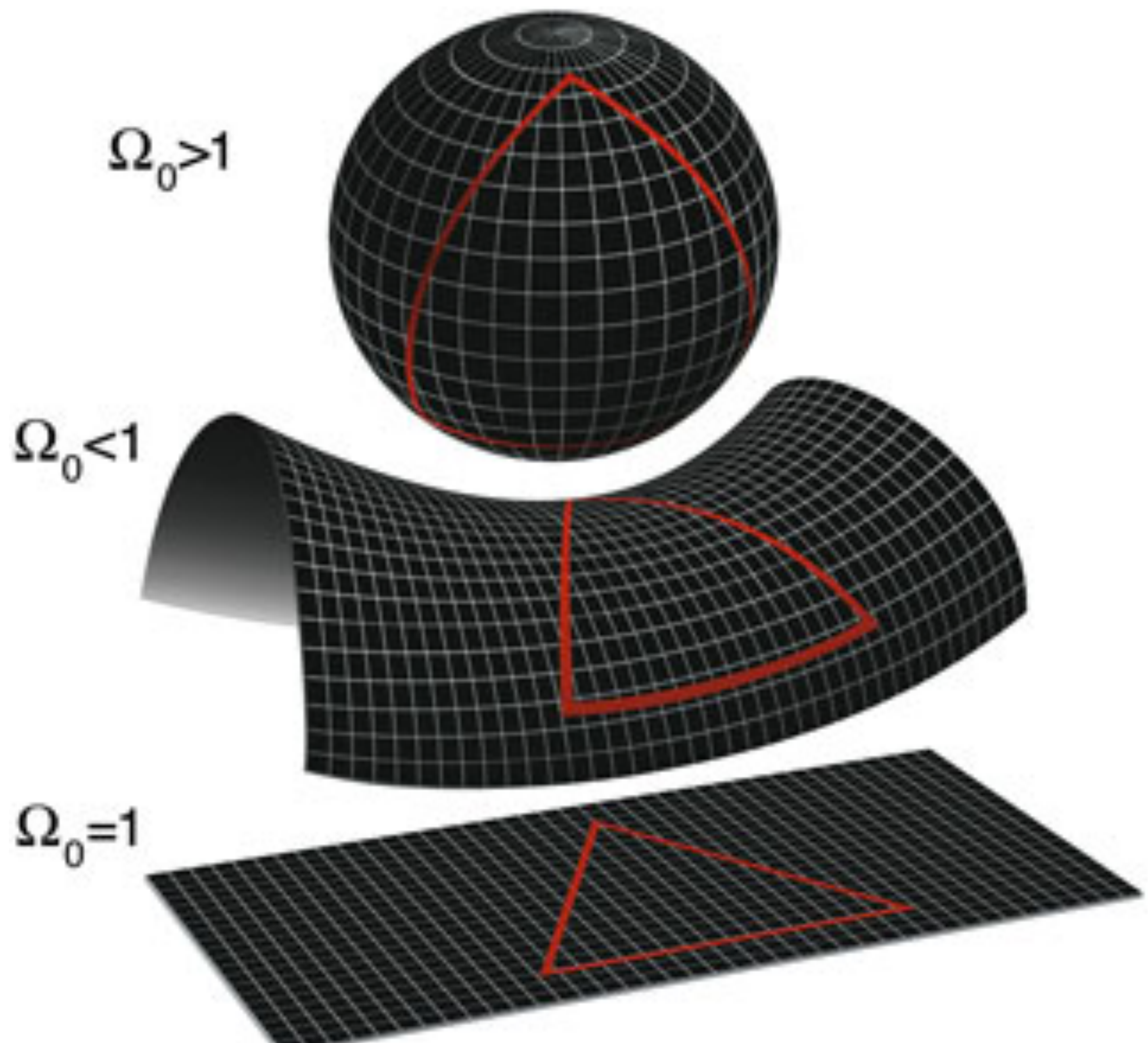
$$d_L(z;\Omega_M,\Omega_\Lambda,w,H_0) = \frac{c}{H_0}\,\tilde{d}_L(z;\Omega_M,\Omega_\Lambda,w) \tag{4}$$

Distance Modulus:

$$\mu(z;H_0,\mathbf{\Omega}) = 25 + 5\log_{10}\left(\frac{c}{H_0}\tilde{d}_L(z;\mathbf{\Omega})\right)$$

Curvature of the Universe

$$\Omega_K + \Omega_M + \Omega_\Lambda = 1$$

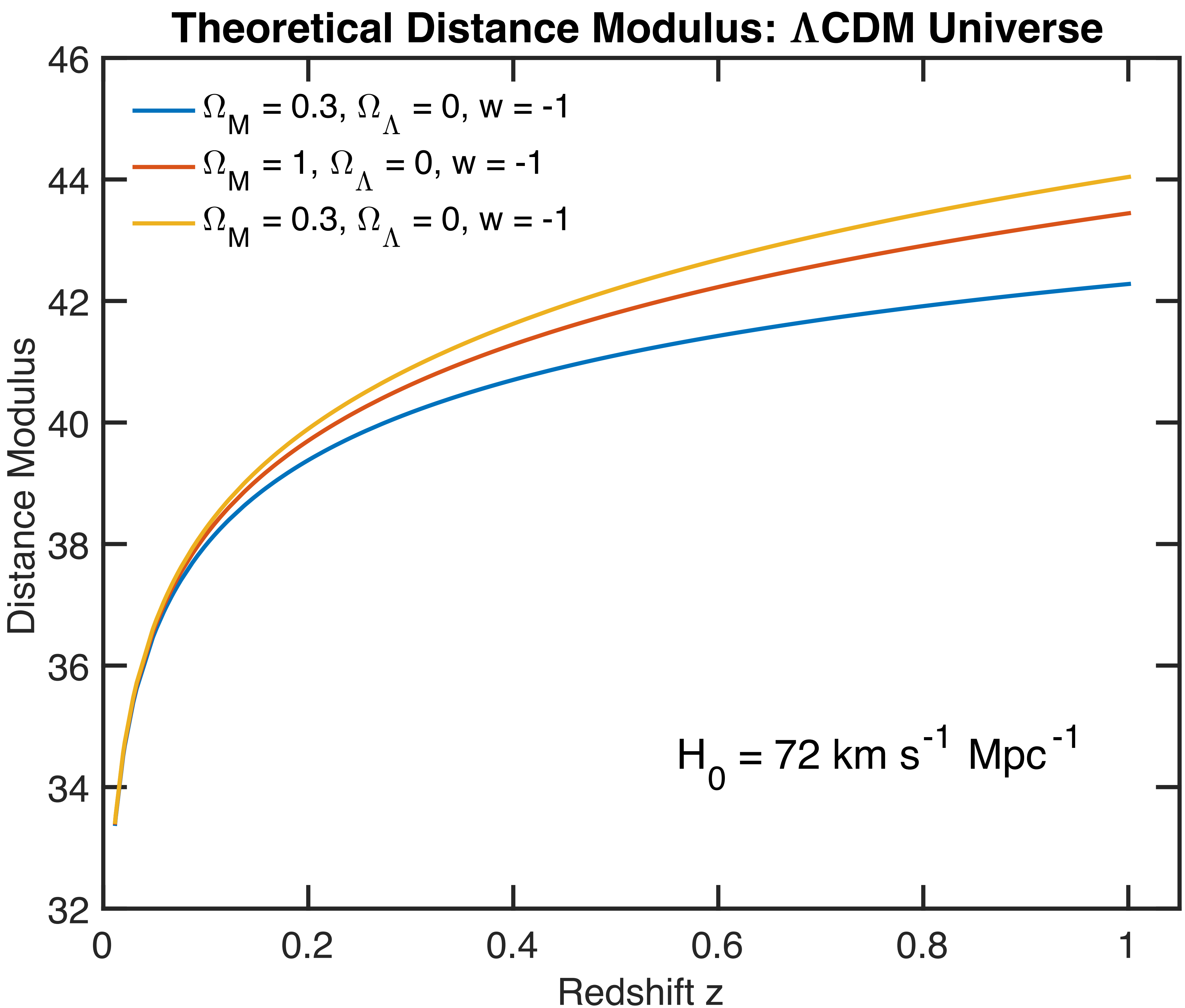


$$\Omega_k > 0$$

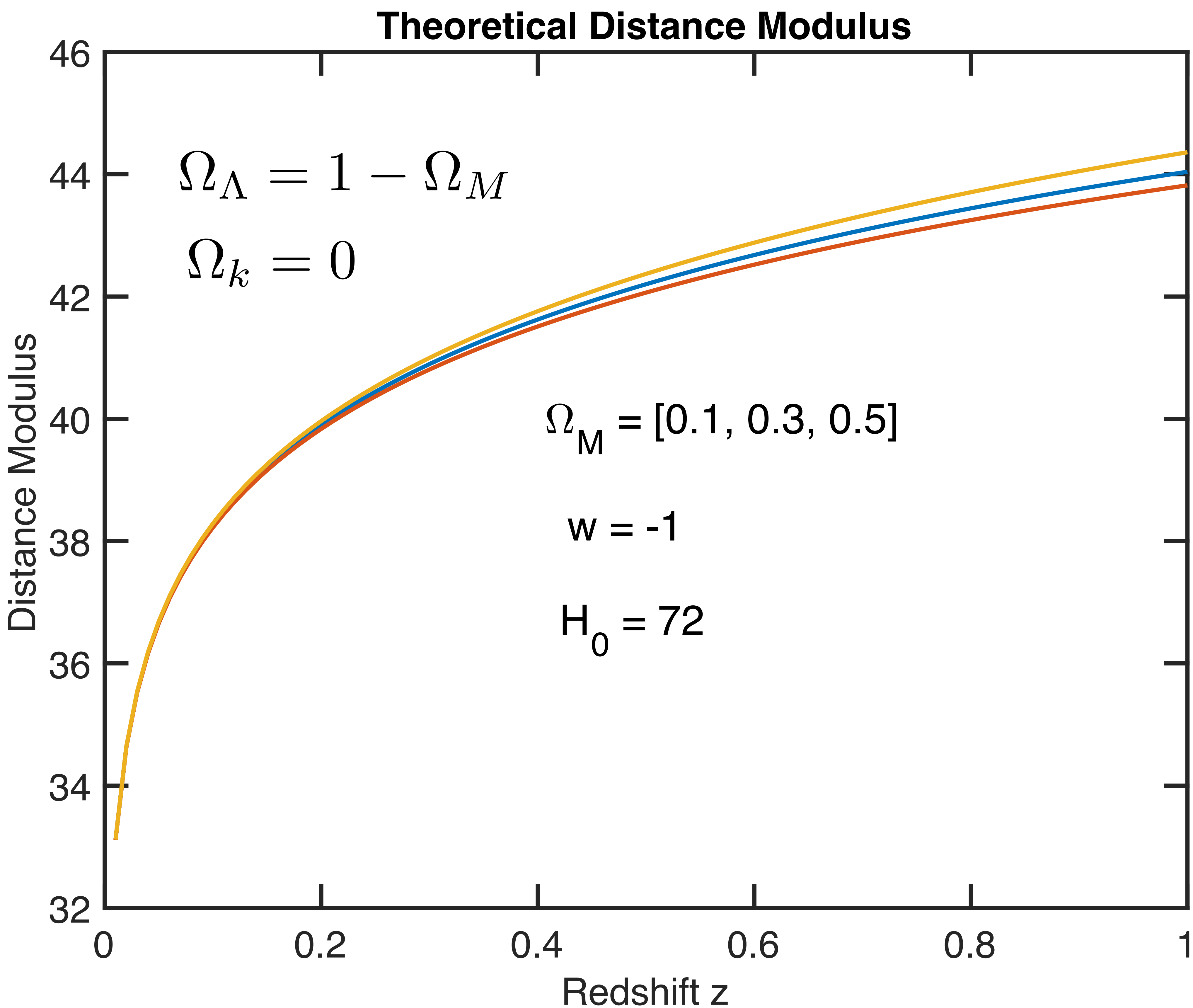
$$\Omega_k < 0$$

$$\Omega_k = 0$$

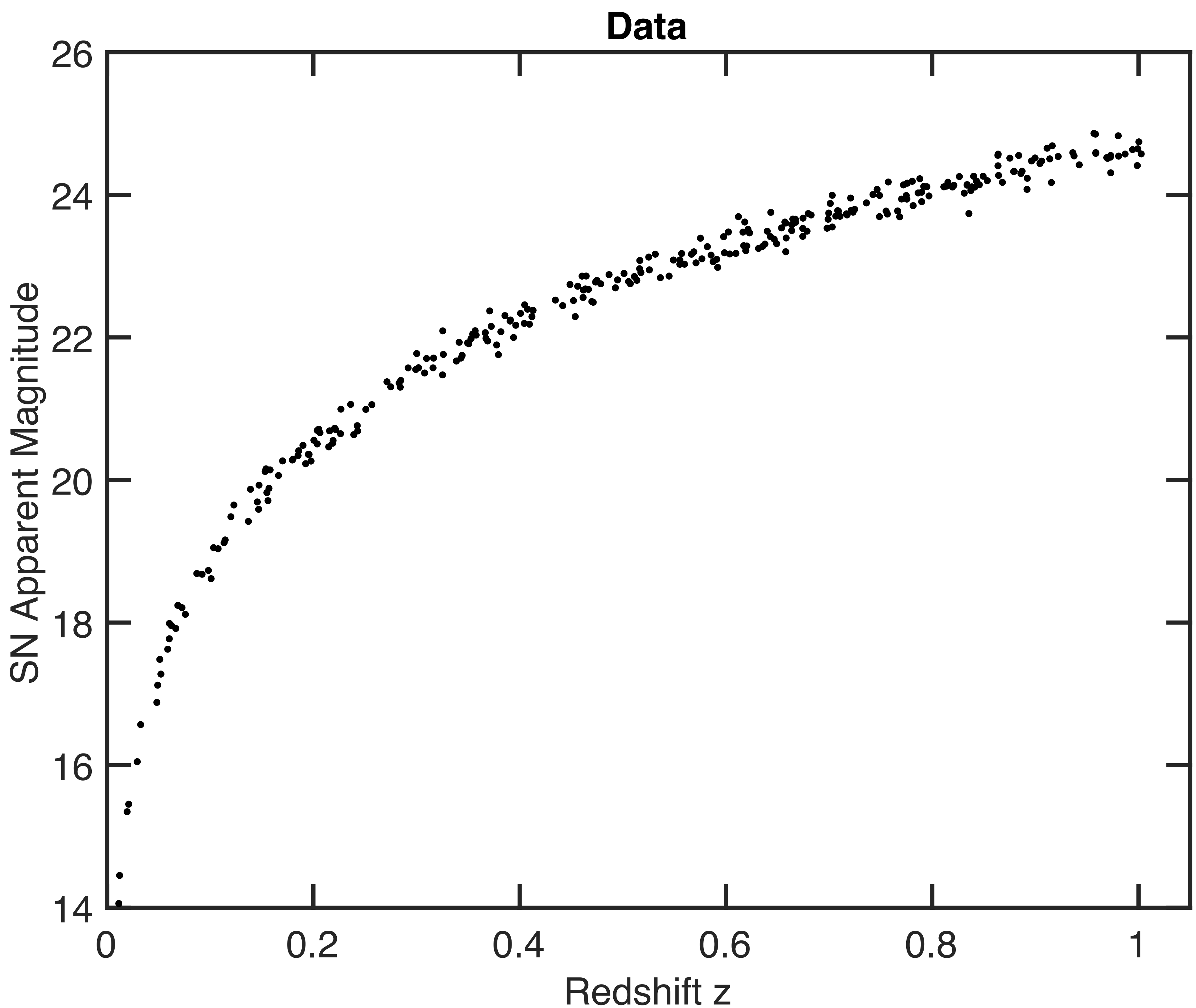
Theoretical Distance Modulus
(Arbitrary Curvature)



Flat Universe: $\Omega_k = 0$



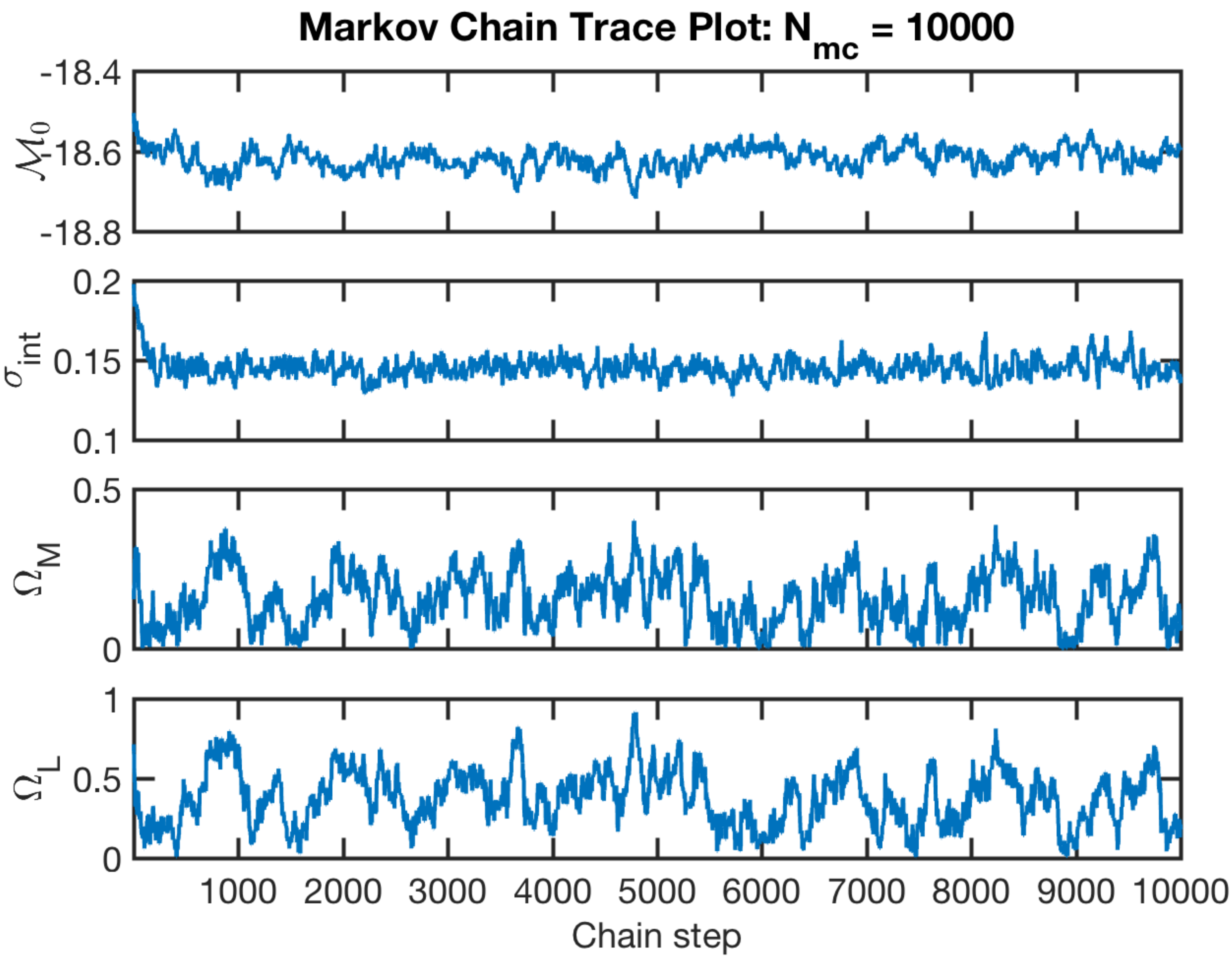
Idealised supernova dataset



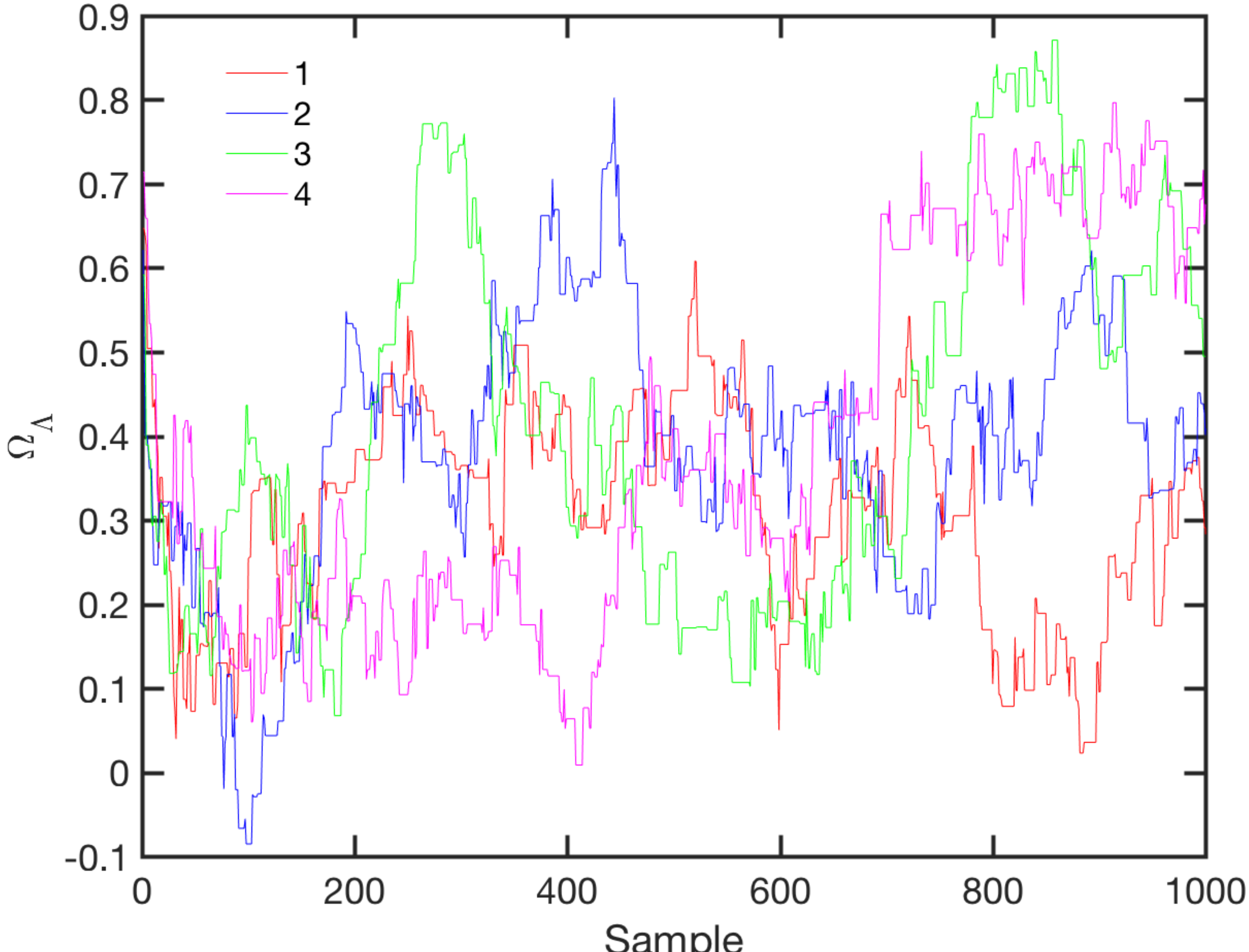
First assume $w = -1$

Write down model & posterior

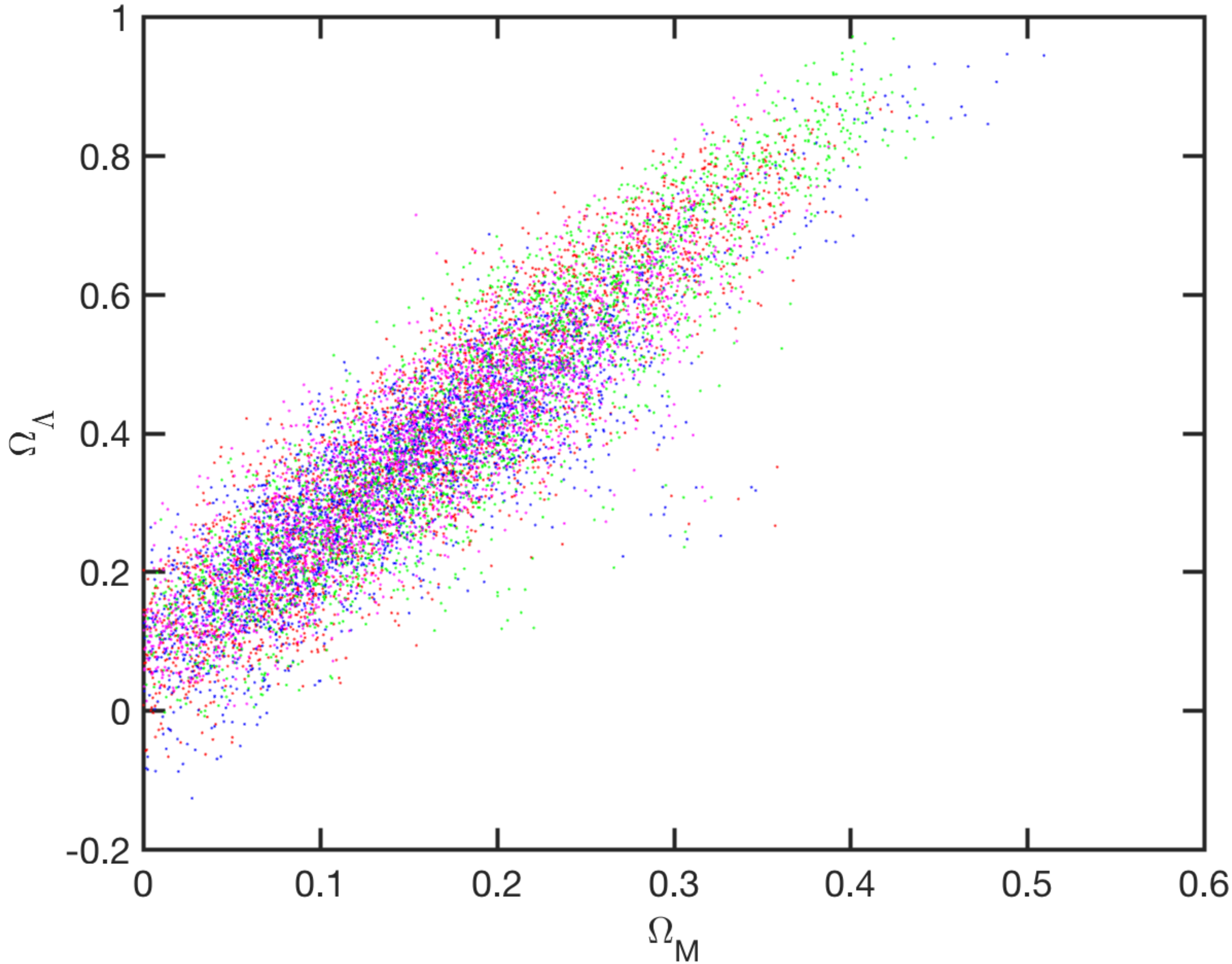
4D Metropolis: one chain



Multiple Independent Chains



Multiple Independent Chains



Assessing Convergence with multiple chains: Gelman-Rubin (G-R) ratio

Monitoring convergence of each scalar estimand

Suppose we have simulated m parallel sequences, each of length n (after discarding the first half of the simulations). For each scalar estimand ψ, we label the simulation draws as ψ_{ij} $(i = 1, \ldots, n; j = 1, \ldots, m)$, and we compute B and W, the between- and within-sequence variances:

$$B = \frac{n}{m-1}\sum_{j=1}^{m}(\overline{\psi}_{.j} - \overline{\psi}_{..})^2, \quad \text{where} \quad \overline{\psi}_{.j} = \frac{1}{n}\sum_{i=1}^{n}\psi_{ij}, \quad \overline{\psi}_{..} = \frac{1}{m}\sum_{j=1}^{m}\overline{\psi}_{.j}$$

$$W = \frac{1}{m}\sum_{j=1}^{m}s_j^2, \quad \text{where} \quad s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\psi_{ij} - \overline{\psi}_{.j})^2.$$

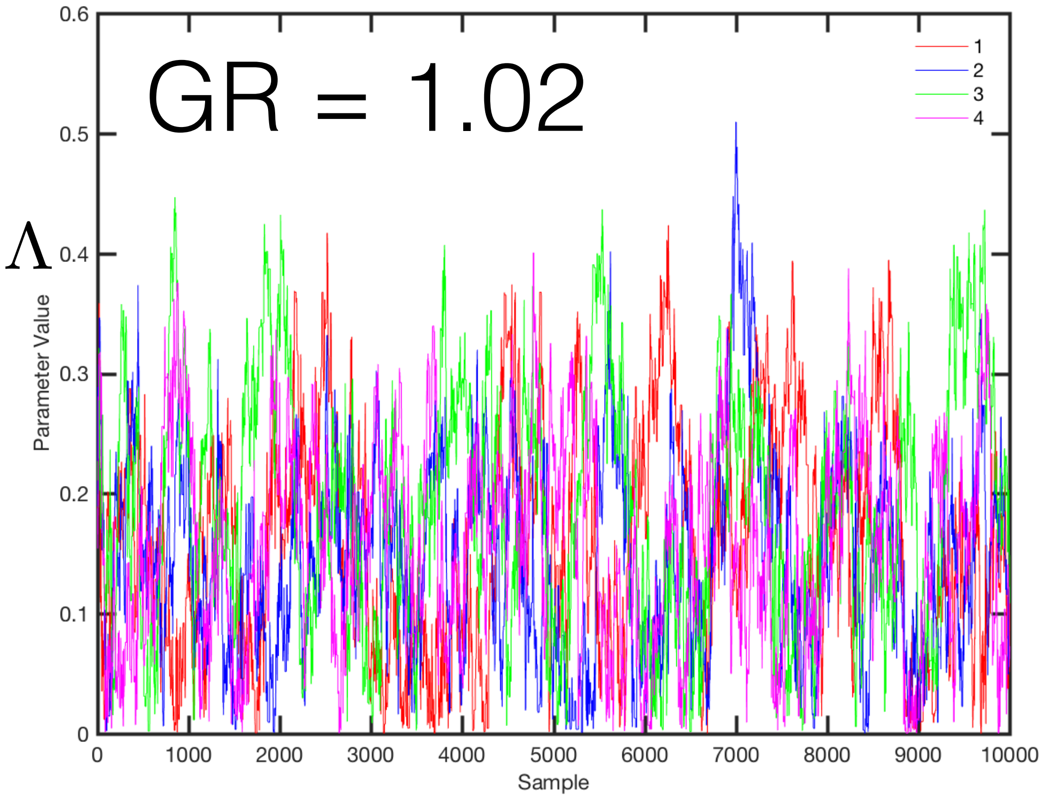
We can estimate $\text{var}(\psi|y)$, the marginal posterior variance of the estimand, by a weighted average of W and B, namely

$$\widehat{\text{var}}^+(\psi|y) = \frac{n-1}{n}W + \frac{1}{n}B. \tag{11.3}$$

G-R ratio: $\hat{R} = \sqrt{\frac{\widehat{\text{var}}^+(\psi|y)}{W}}, \approx 1$

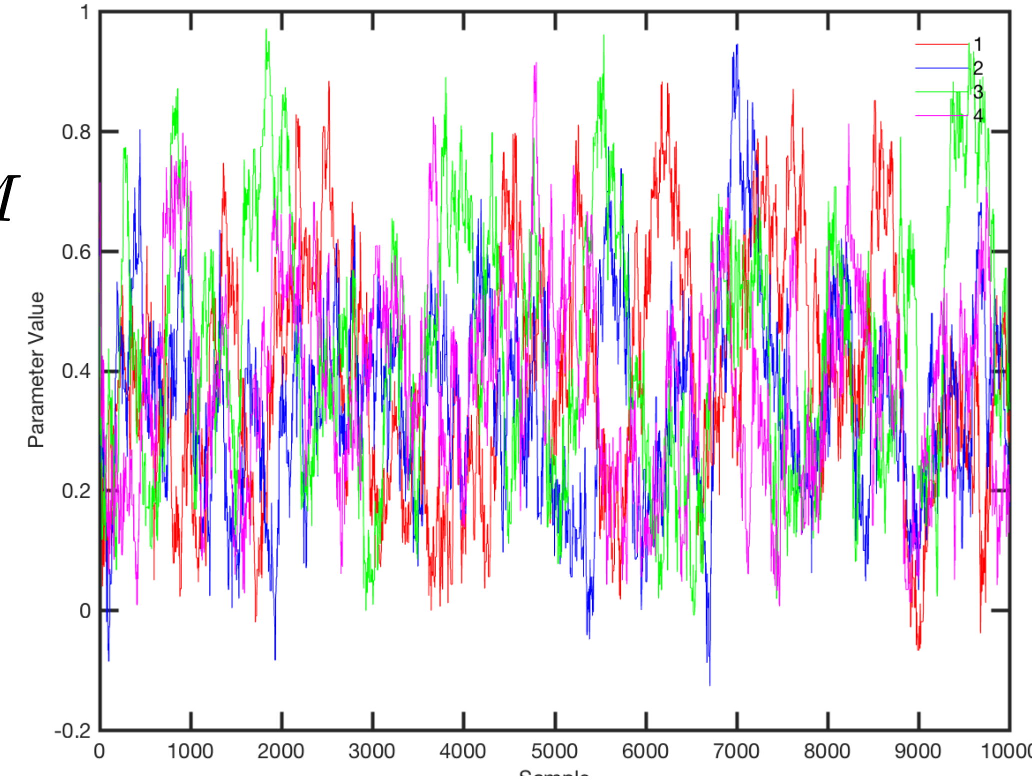
4 Independent Chains (10000 iters)

Ω_Λ

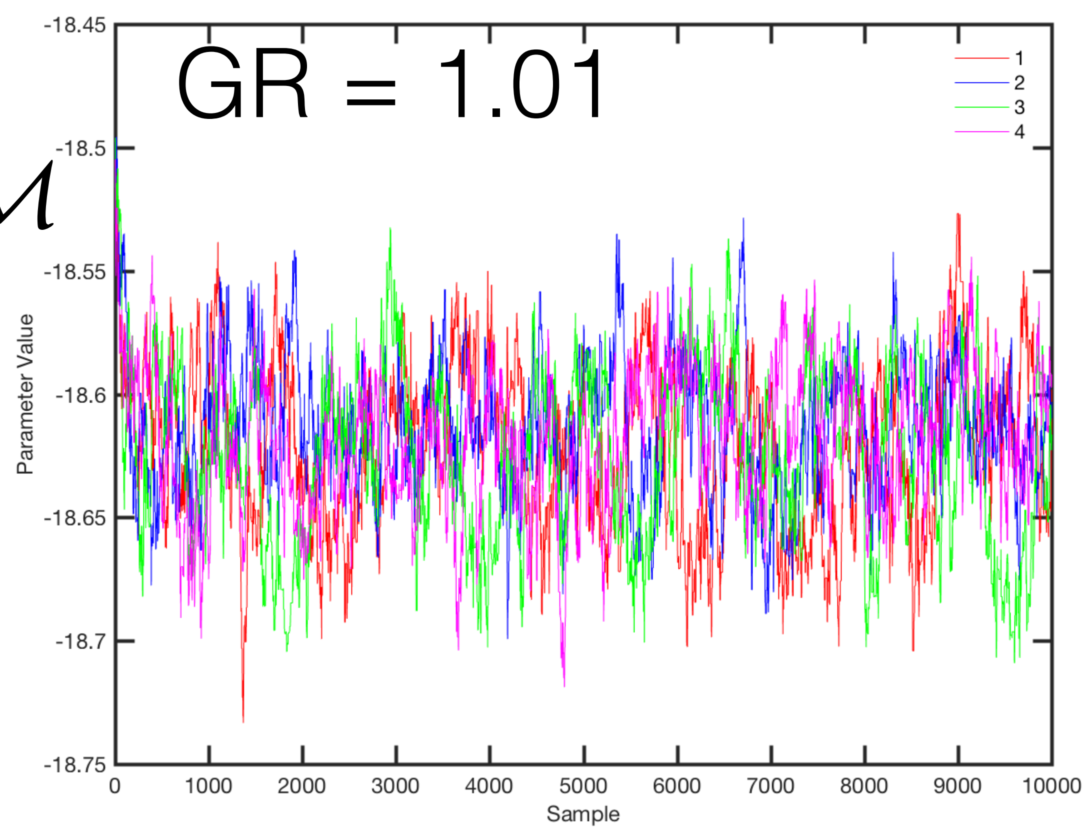


GR = 1.02

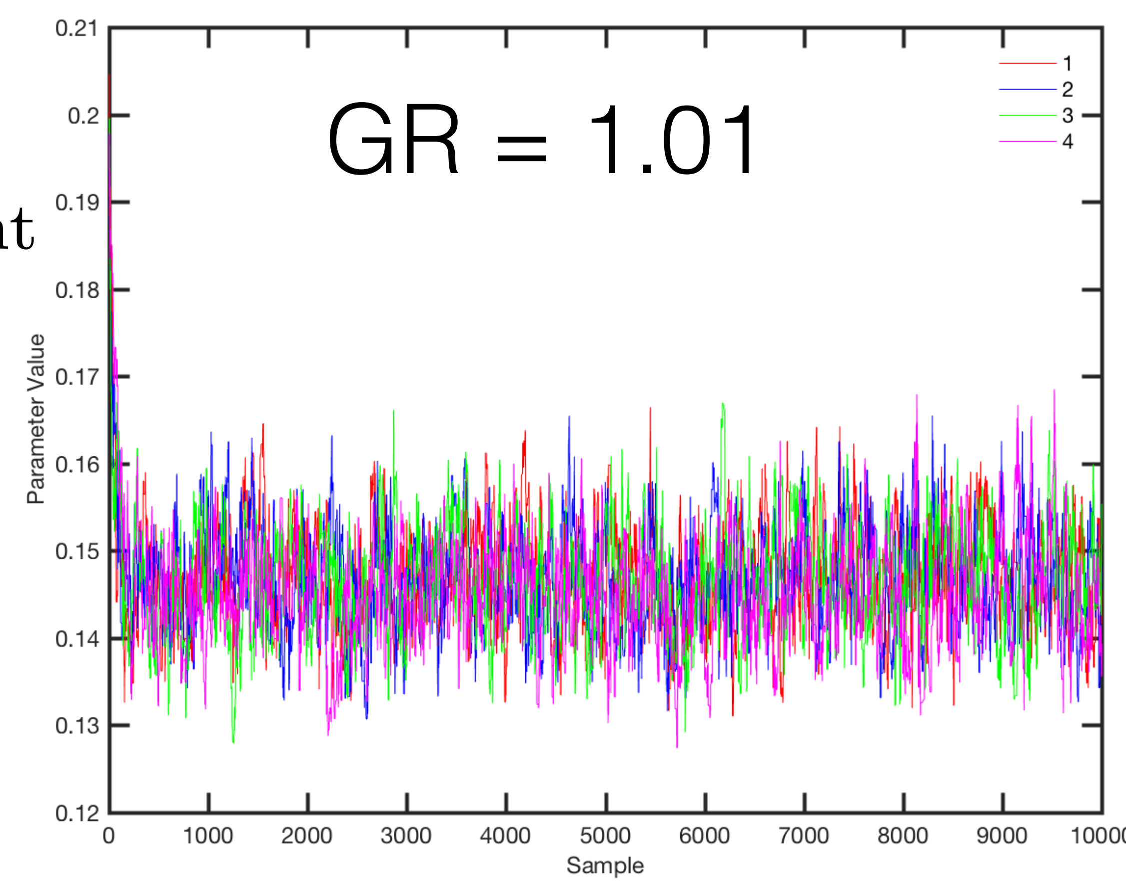
Ω_M



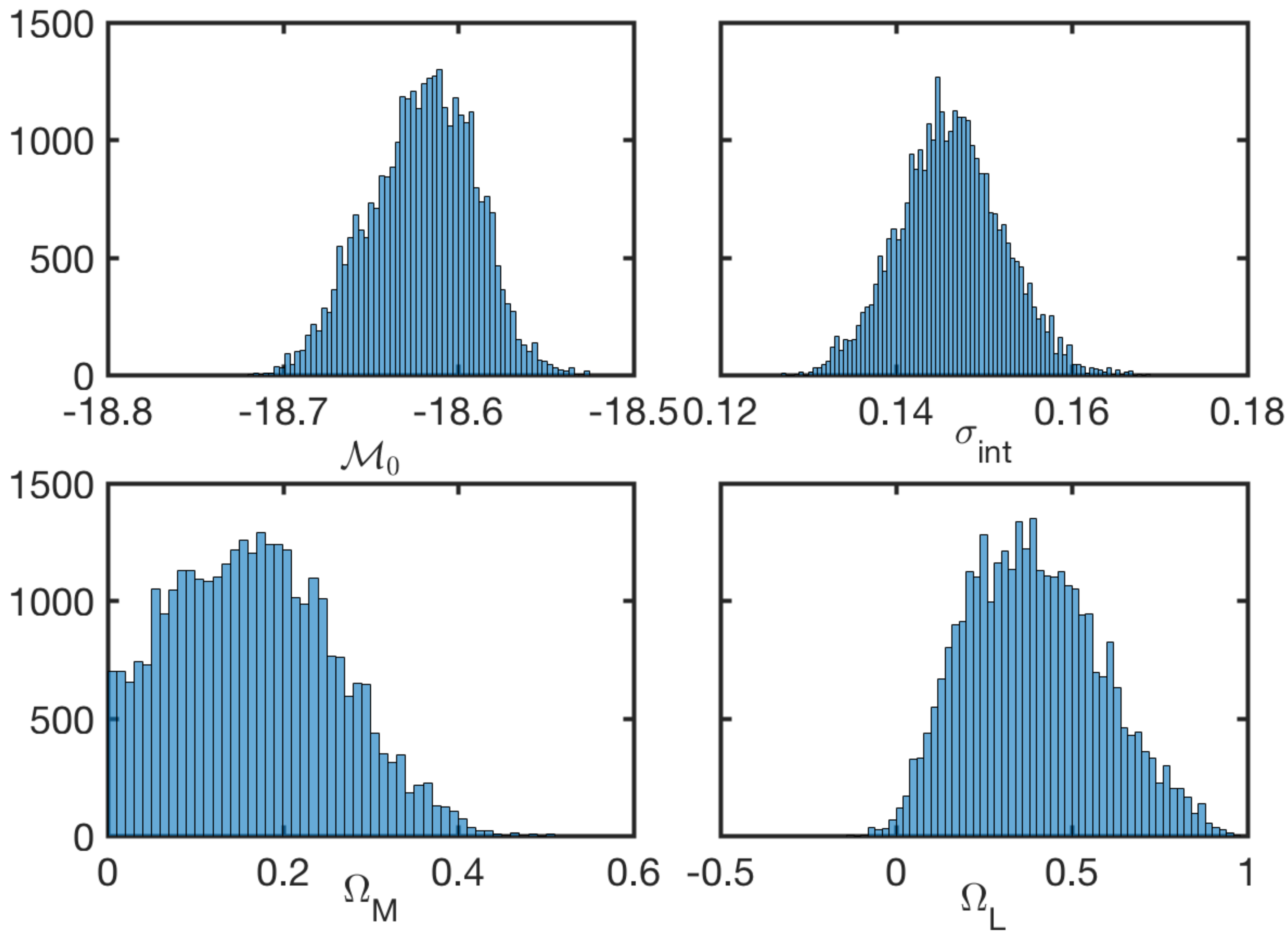
$\mathcal{M}$



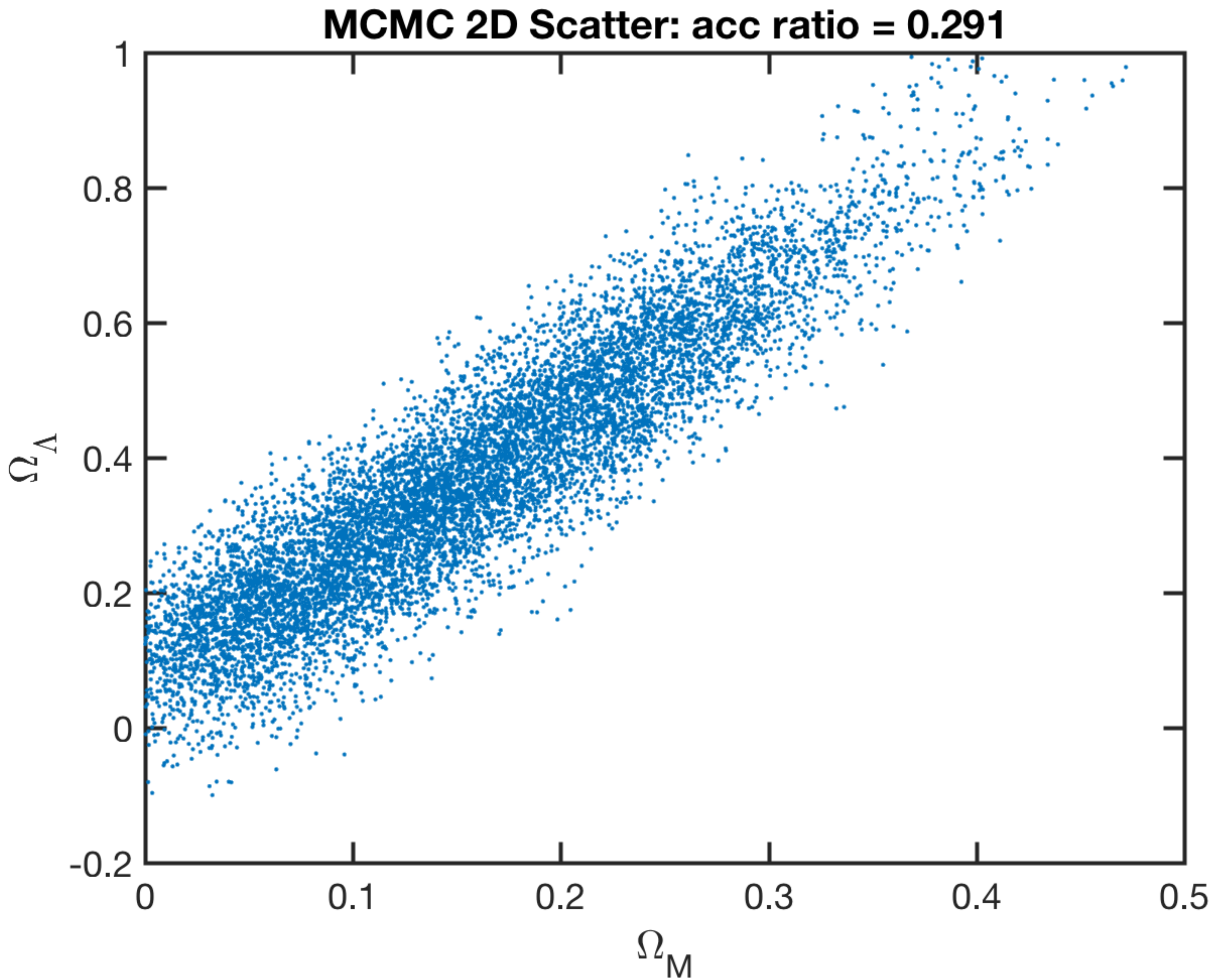
$\sigma_{\rm int}$



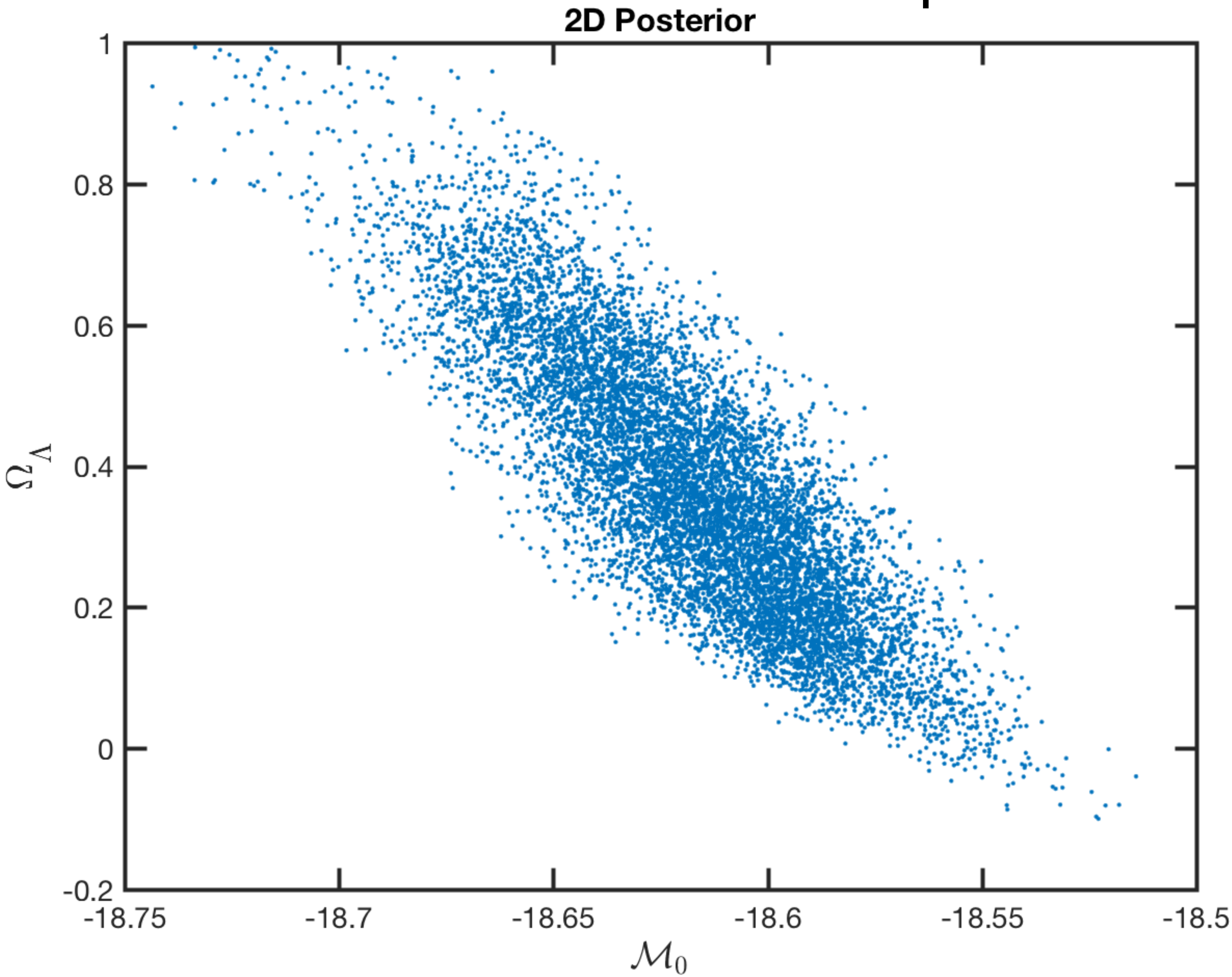
Cut burn-in (20%) and combine chains
Posterior Histograms



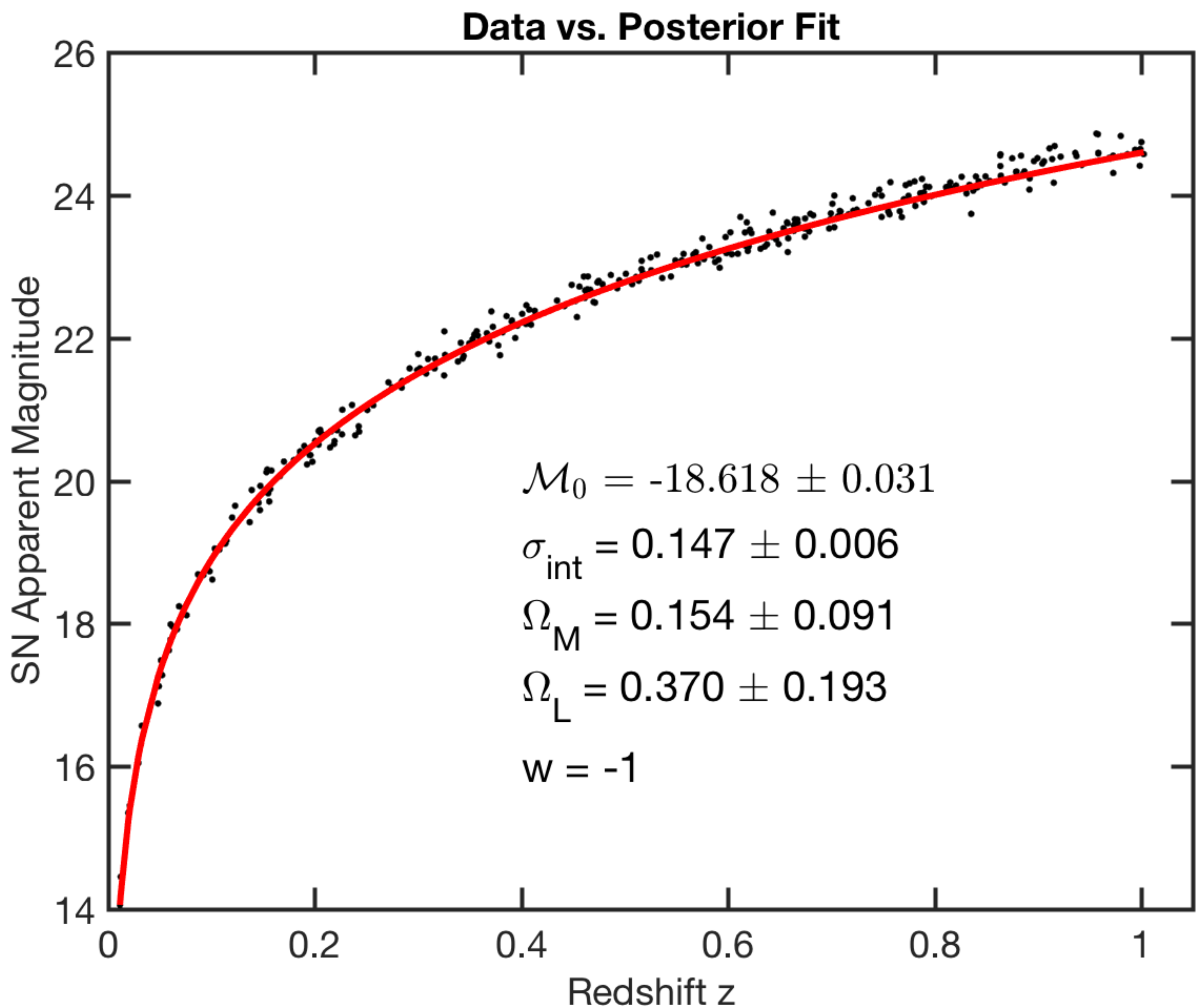
Posterior Scatter plot



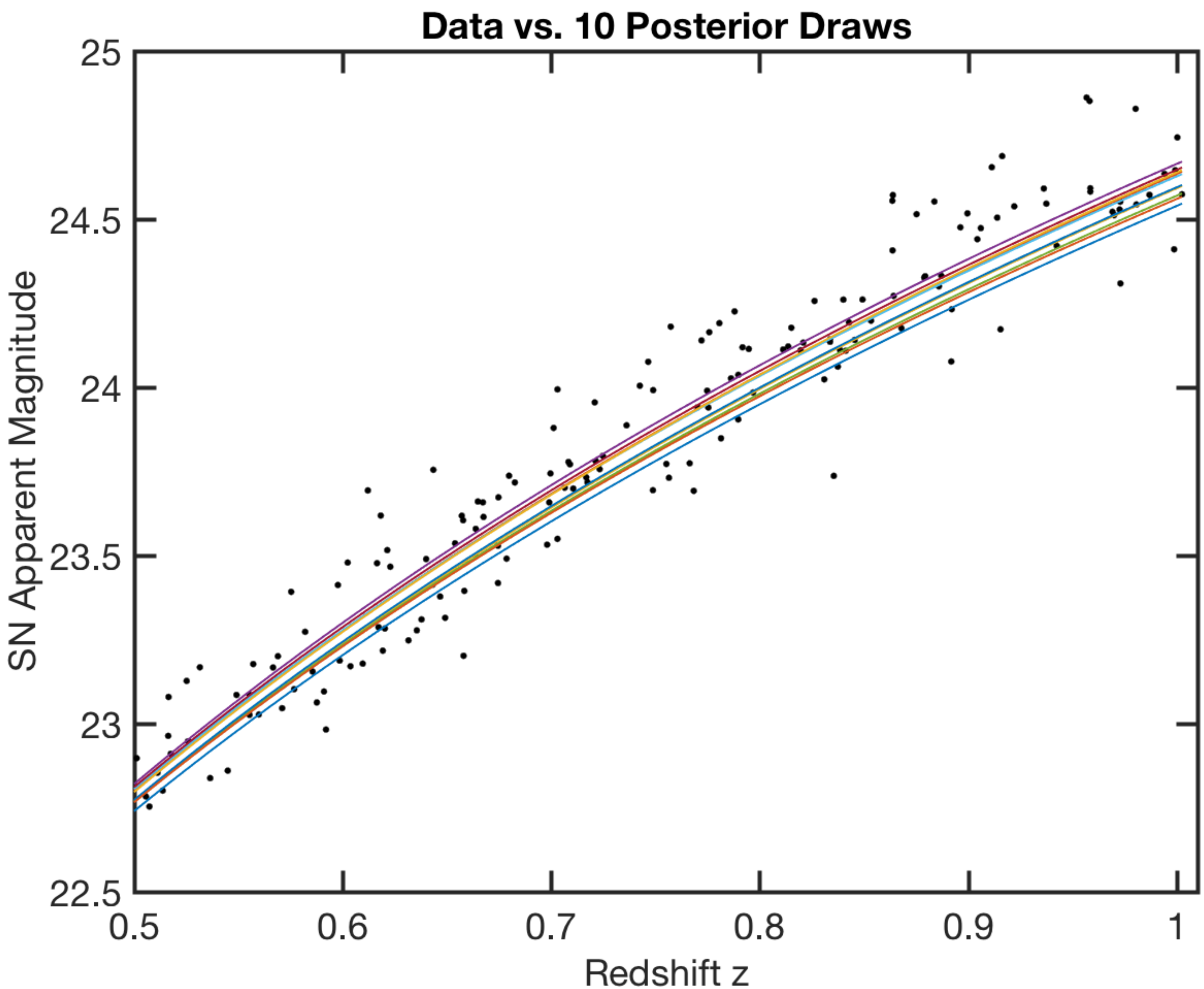
Posterior Scatter plot



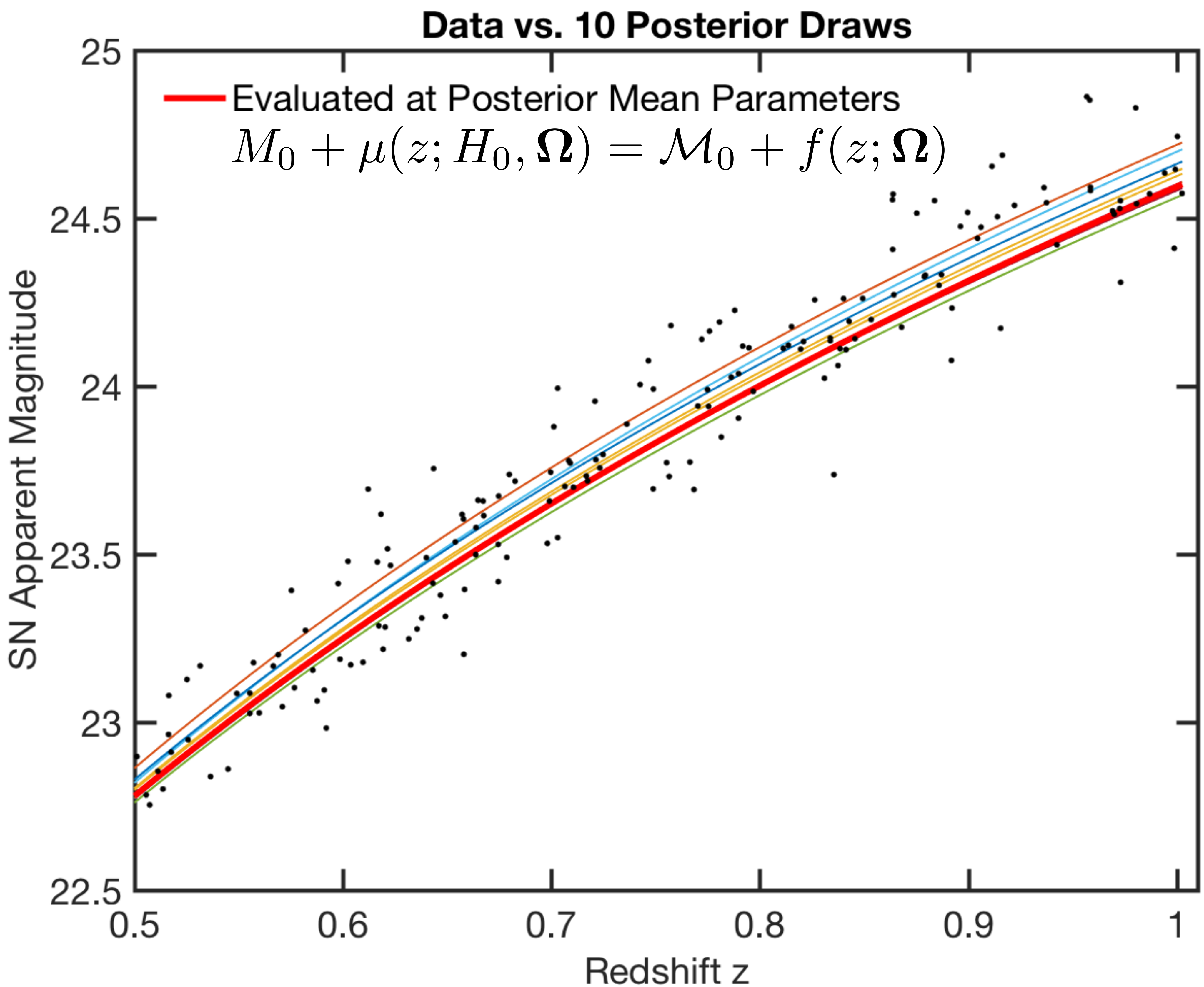
Fit with posterior mean



Random Posterior Draws of parameters from chain



Random Posterior Draws of parameters from chain



The deceleration parameter q_0

of the cosmological models. "It is a striking and slightly puzzling fact that almost all current cosmological observations can be summarized by the simple statement: The jerk of the Universe equals one" [25].

Let us make Taylor expansion of the scale factor in time using the above introduced parameters

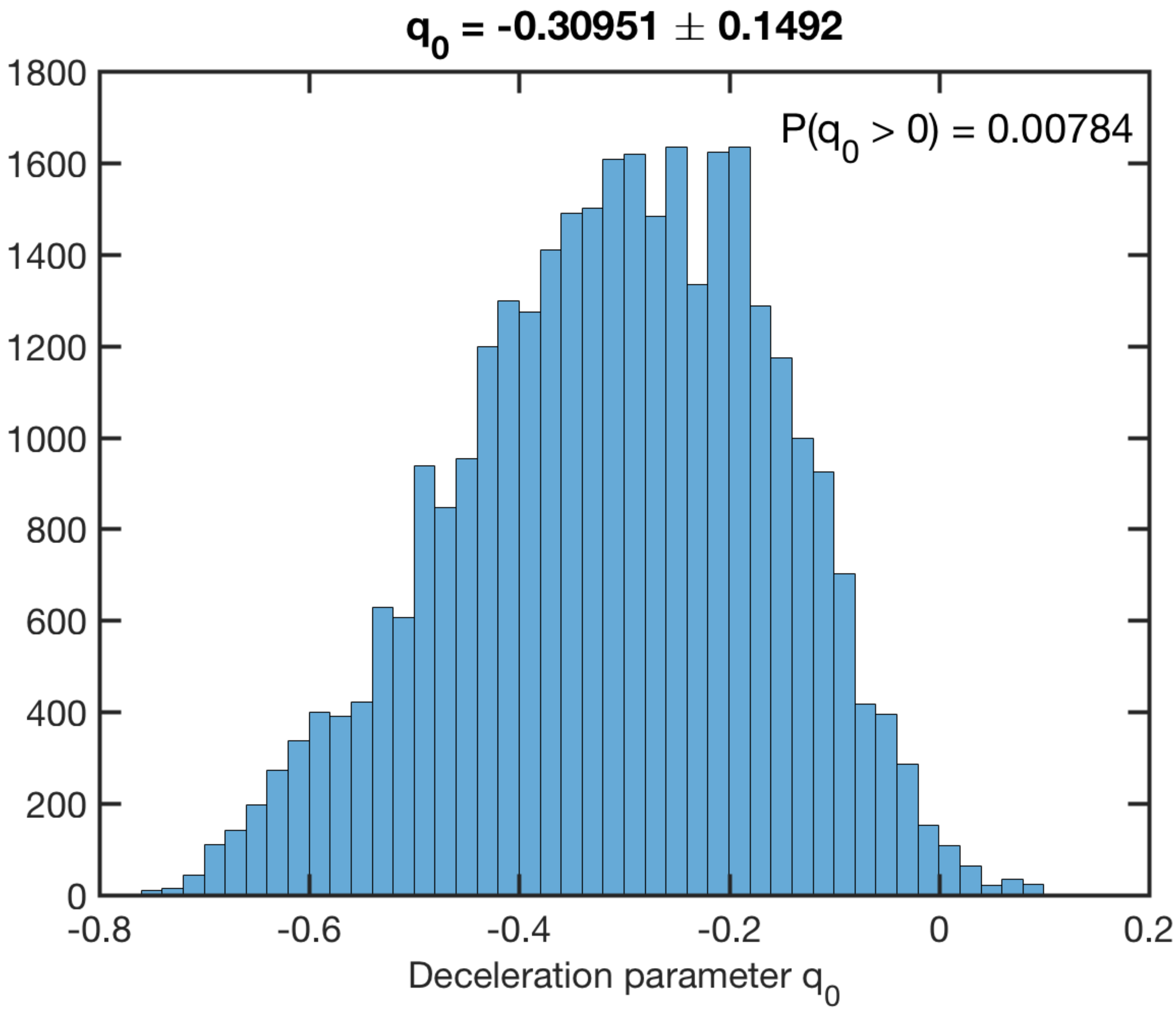
$$\begin{aligned} a(t) =& a_0 \Bigg[1 + H_0(t-t_0) - \frac{1}{2} q_0 H_0^2 (t-t_0)^2 + \frac{1}{3!} j_0 H_0^3 (t-t_0)^3 \\ &+ \frac{1}{4!} s_0 H_0^4 (t-t_0)^4 + \frac{1}{5!} l_0 H_0^5 (t-t_0)^5 + O\left((t-t_0)^6\right) \Bigg]. \end{aligned} \tag{2.49}$$

a(0) = 0 (Big Bang)
a(1) = 1 (Today)

$$q_0 = \Omega_M/2 - \Omega_\Lambda$$

The posterior of derived quantities

Deceleration Parameter: $q_0 = \Omega_M/2 - \Omega_\Lambda$



$q_0 < 0$
Expansion of Universe is accelerating!

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 $J(\theta^* \mid \theta_i) = N(\theta^* \mid \theta_i, \mathbf{\Sigma_p}) = N(\theta_i \mid \theta^*, \mathbf{\Sigma_p}) = J(\theta_i \mid \theta^*)$ 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, ϕ^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, ϕ^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 $\boldsymbol{\theta} = (\theta_1 \dots \theta_d)$

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

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

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

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

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

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

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

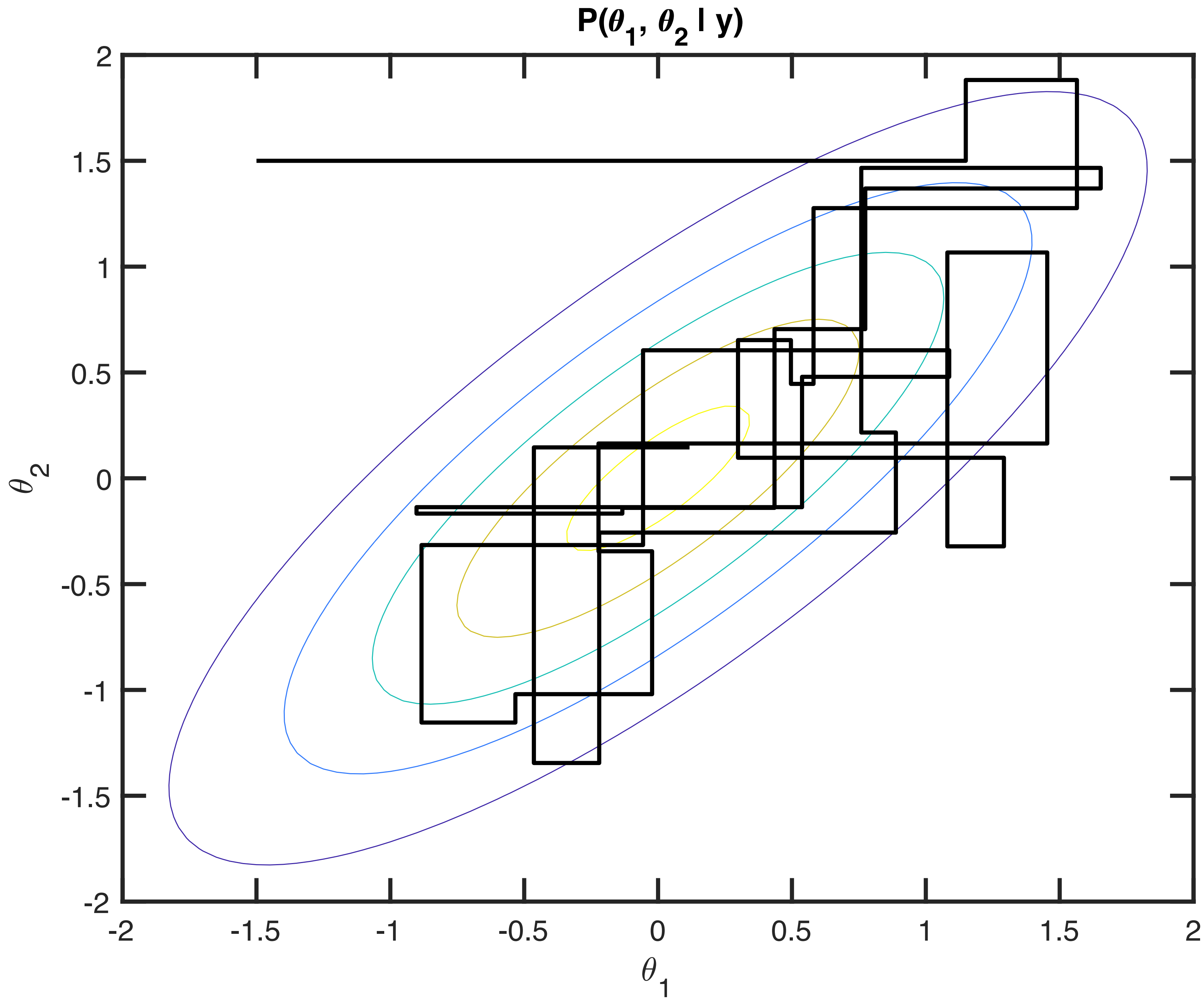
$$P(\boldsymbol{\theta}|y) = P(\theta_1|\theta_2, y)P(\theta_2|y) = P(\theta_2|\theta_1, y)P(\theta_1|y)$$

Conditional Posteriors:

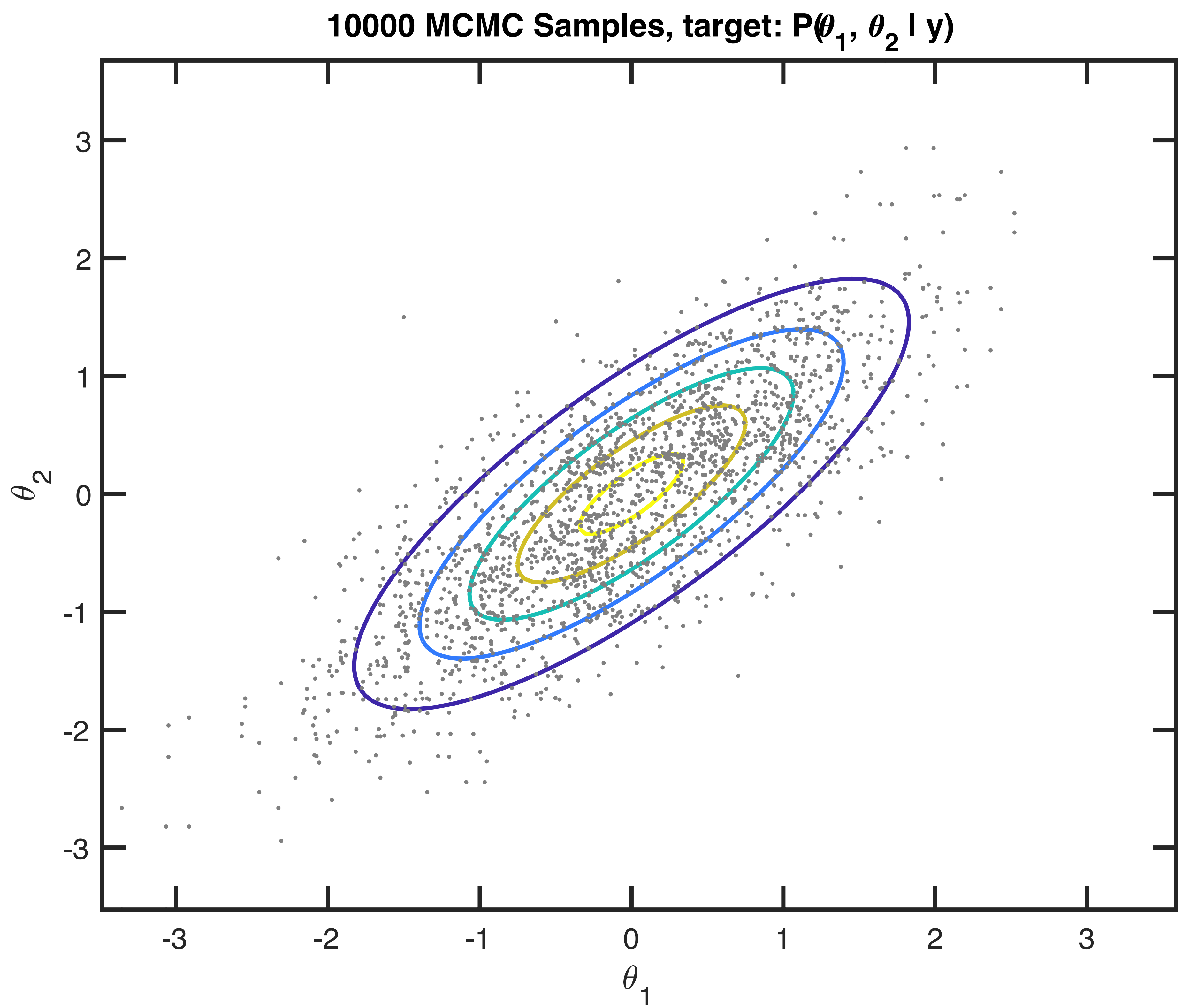
$$\theta_1|\theta_2, \boldsymbol{y} \sim N(y_1 + \rho(\theta_2 - y_2), 1 - \rho^2)$$

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

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