# Astrostatistics: Thu 22 Feb 2017

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

- Example Class 2: Friday, Mar 2 2:30pm, MR 5
- Fitting Statistical Models to Astronomical Data
    - Markov Chain Monte Carlo
    - Refs: Not much in Ivezic, Ch 5, F&B Ch 3
    - Gelman et al. Bayesian Data Analysis (Ch 11 & 12)
    - Givens & Hoeting. “Computational Statistics” (Ch 7 & 8)
    - Roberts & Casella. “Monte Carlo Statistical Methods” (theory) (Ch 6 & 7)
    - Hogg & DFM, 2017 “Data analysis recipes: Using Markov Chain Monte Carlo.” https://arxiv.org/abs/1710.06068

# What if you can't directly sample the posterior: $\theta_i \sim P(\theta\,|D)$?

$$\mathbb{E}[f(\boldsymbol{\theta})|\, D] = \int f(\boldsymbol{\theta}) P(\boldsymbol{\theta}|\, D)\, d\boldsymbol{\theta} \approx \frac{1}{m}\sum_{i=1}^{m} f(\boldsymbol{\theta}_i)$$

- Posterior simulation - Markov Chain Monte Carlo:
- Generate a correlated sequence (chain) of random variates (Monte Carlo) that (in a limit) are draws from the posterior distribution. The next value in the sequence only depends on the current values (Markov).

# Mapping the Posterior P($\theta$ |D)

- Markov Chain Monte Carlo (MCMC)
- Last time: 1D Metropolis algorithm
- Today:
    - Drawing Multivariate Gaussian random variables
    - N-D Metropolis Algorithm
    - Rules of thumb for proposal scale
    - assessing convergence (G-R Ratio)
    - Metropolis-Hastings algorithm
    - Gibbs sampling

# N-dim Metropolis Algorithm: Posterior P($\theta$ |D)

1. Choose a random starting point $\theta_0$

2. At step i = 1…N, propose a new parameter value $\theta_{prop} \sim N(\theta_i, \mathbf{\Sigma})$. The proposal scale $\mathbf{\Sigma}$ is chosen cleverly.

3. Evaluate ratio of posteriors at proposed vs current values. r = $P(\theta_{prop} \mid \mathbf{y}) / P(\theta_i \mid \mathbf{y})$.

4. Accept $\theta_{prop}$ with probability min(r,1): $\theta_{i+1} = \theta_{prop}$. If not accept, stay at same value $\theta_{i+1} = \theta_i$ & include in chain.

5. Repeat steps 2-4 until reach some measure of convergence and gather enough samples to compute your inference

# Multi-parameter Bayesian inference: Gaussian example: Gelman BDA Sec 3.2 - 3.3

Sampling distribution: $y_i \sim N(\mu, \sigma^2)$ $\quad i = 1 \dots n$

Likelihood Function: $P(\boldsymbol{y}|\mu, \sigma^2) = \prod_{i=1}^{n} N(y_i|\mu, \sigma^2)$

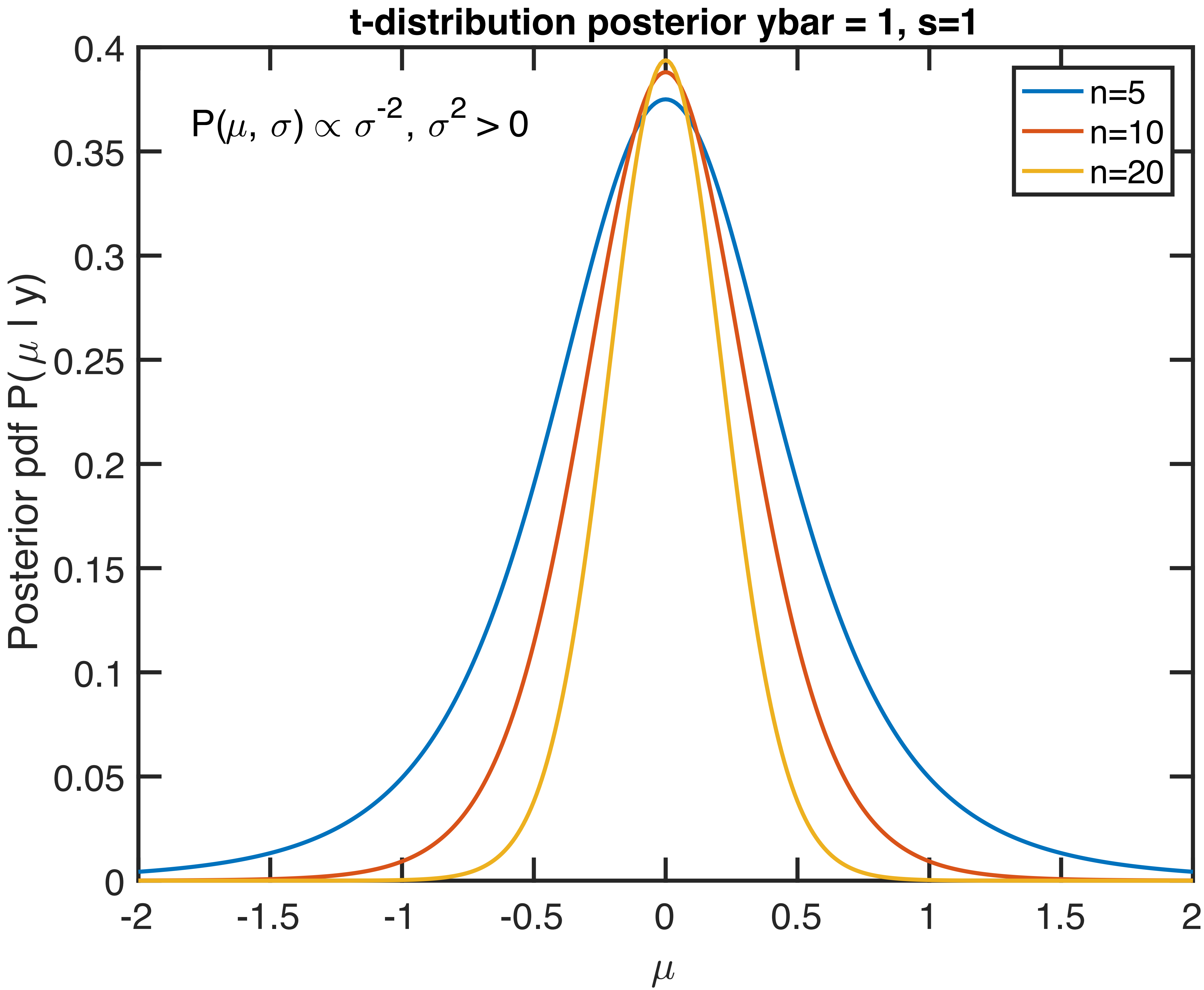
Prior: $P(\mu) \propto 1 \qquad P(\sigma^2) \propto \sigma^{-2}, \sigma^2 > 0$

<u>Posterior</u>

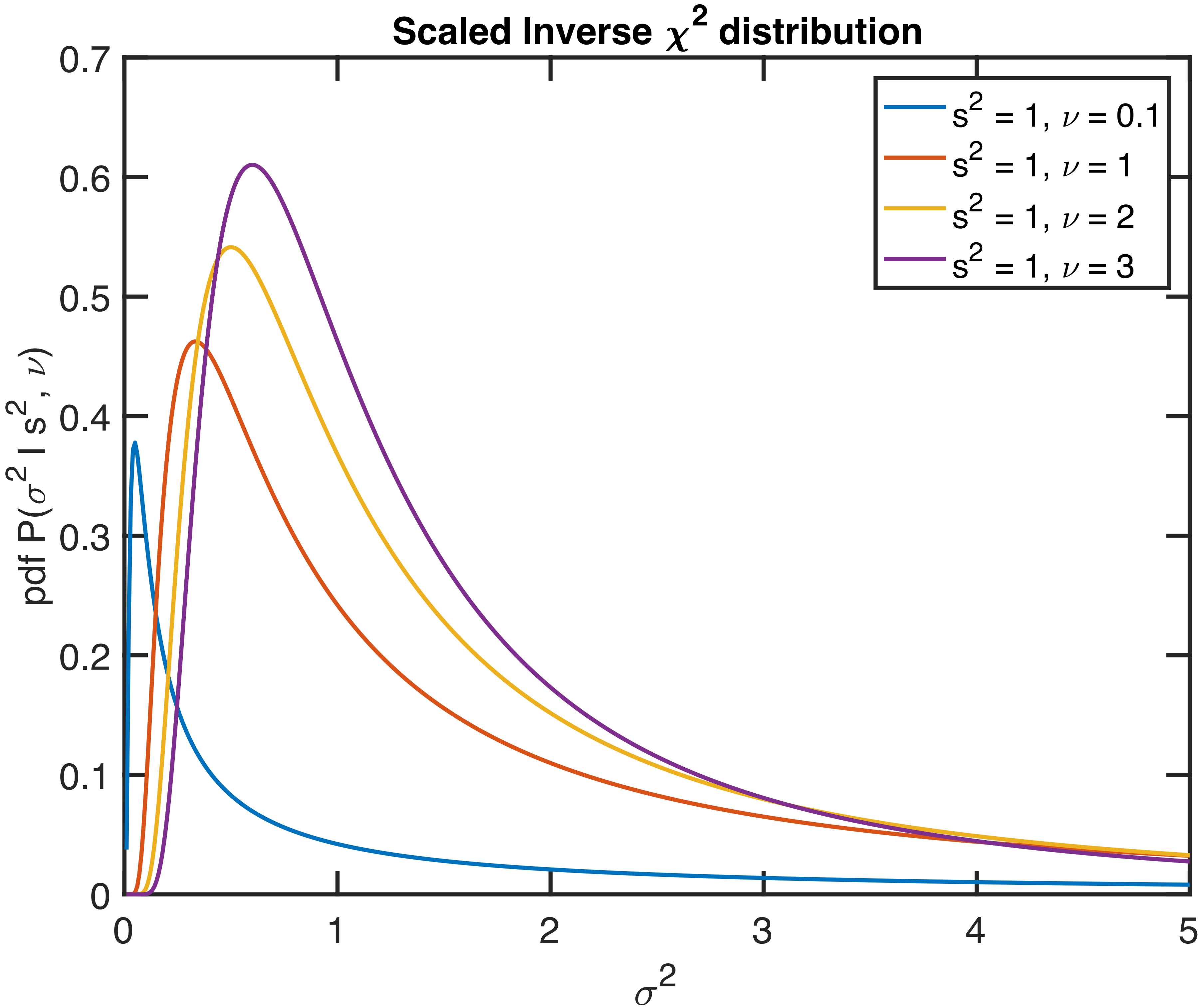
$$P(\mu, \sigma^2|\boldsymbol{y}) \propto (\sigma^2)^{-(n+2)/2} \exp\left(-\frac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y}-\mu)^2]\right)$$

Sufficient Statistics: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

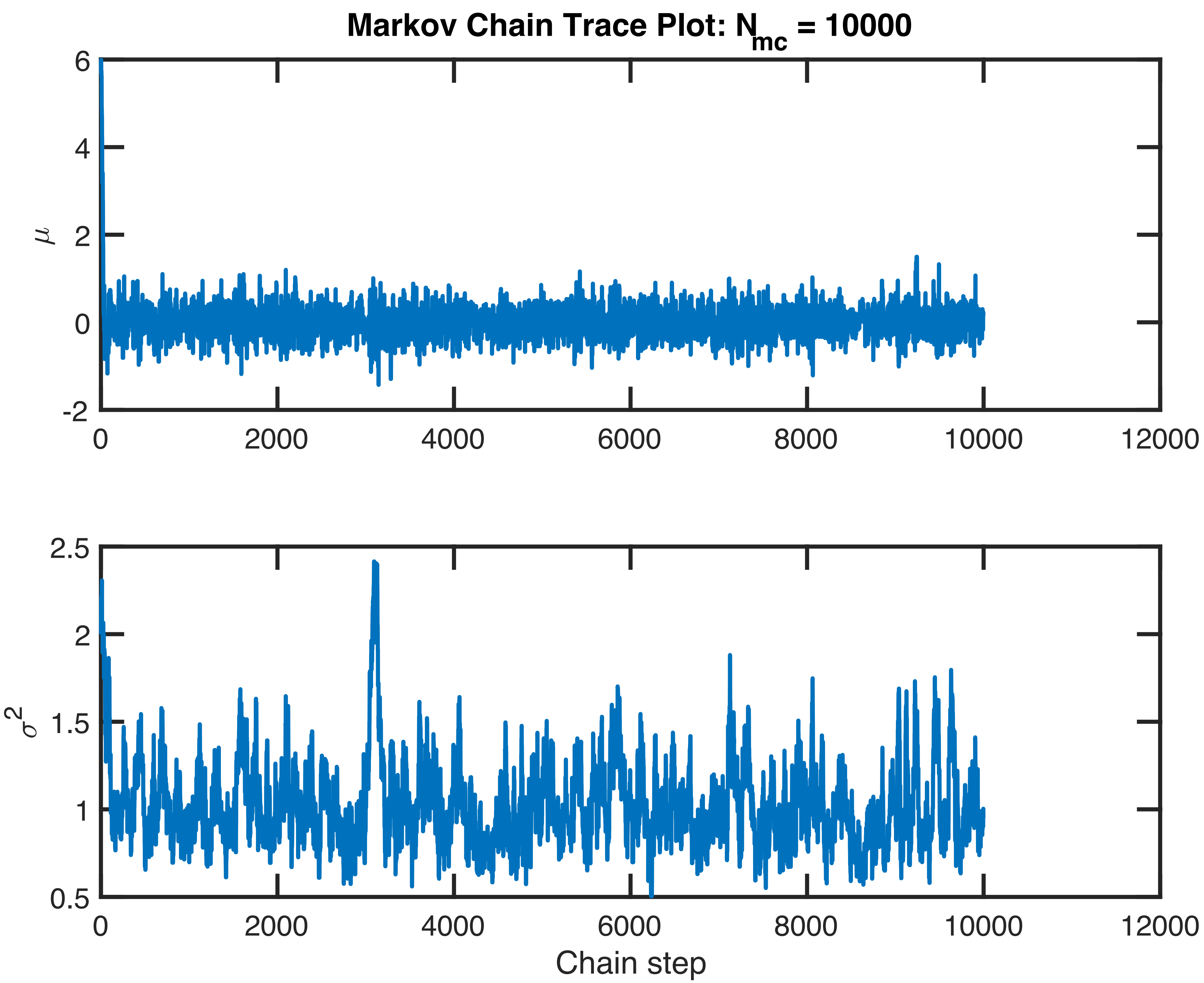
# Last Time: Posterior Distribution of a Gaussian Mean
# Analytic Result



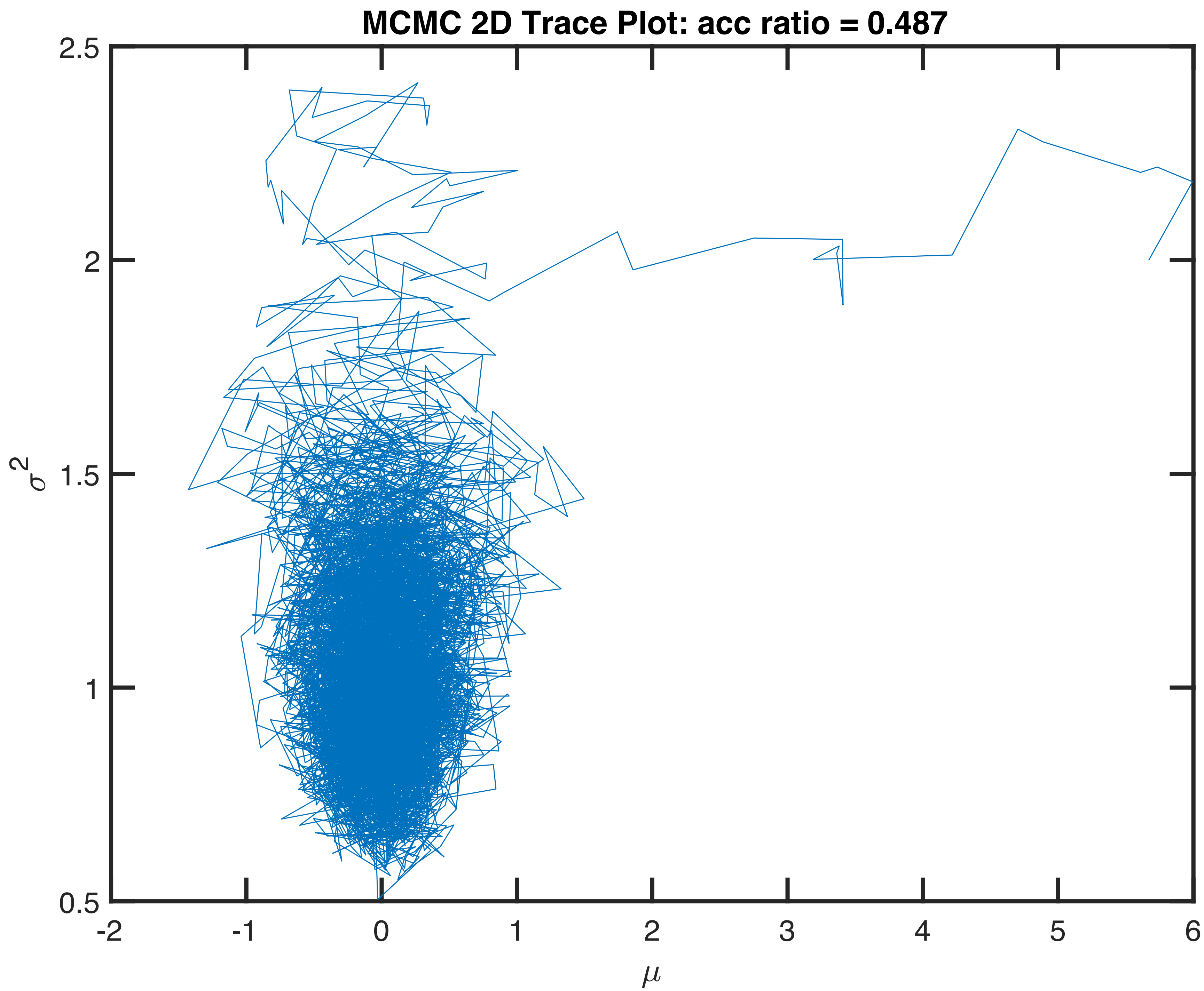
# Posterior Distribution $P(\sigma^2 \mid y)$ follows a Scaled Inverse $\boldsymbol{\chi}^2$ distribution (~ inverse gamma)



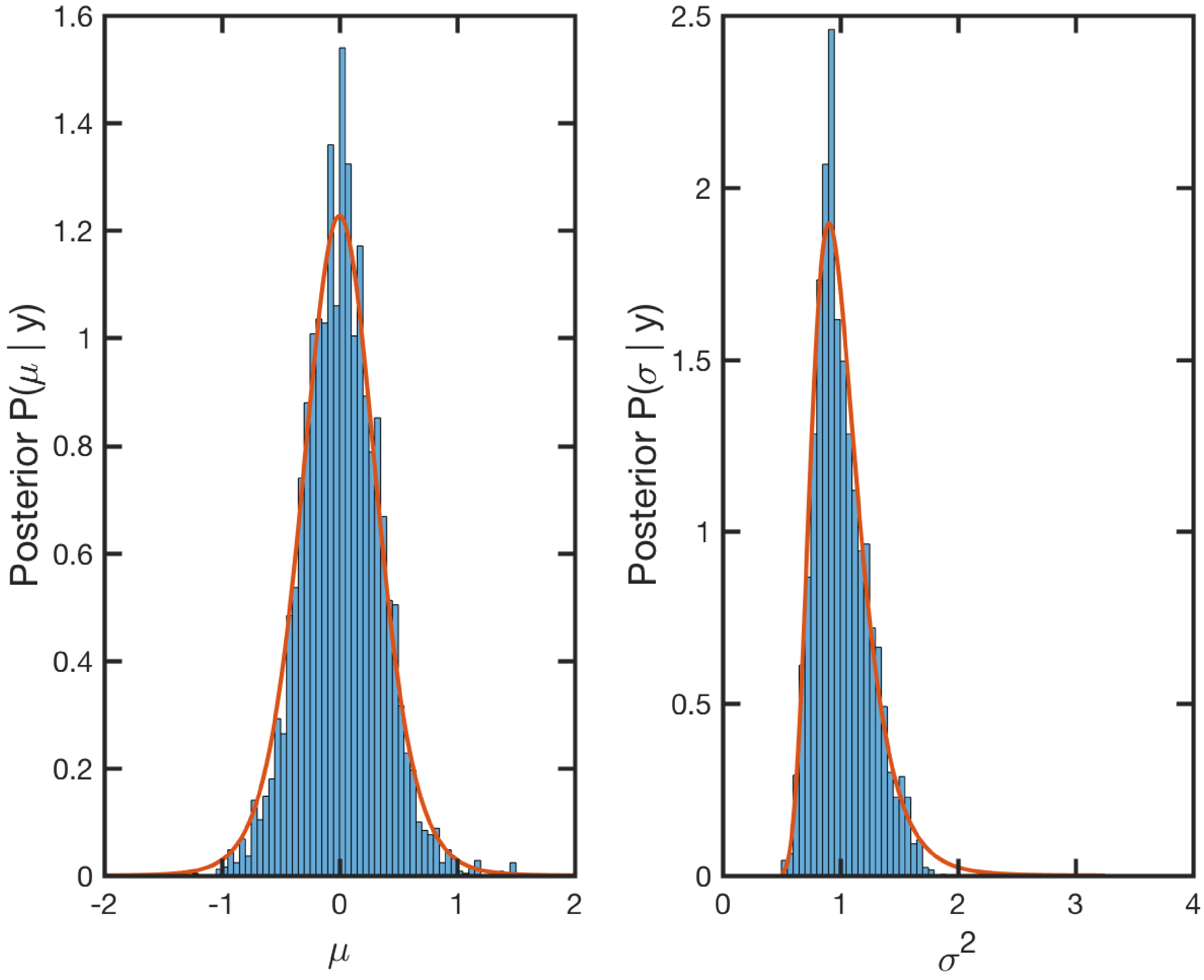
# Metropolis 2D Code Example (metropolis2.m)



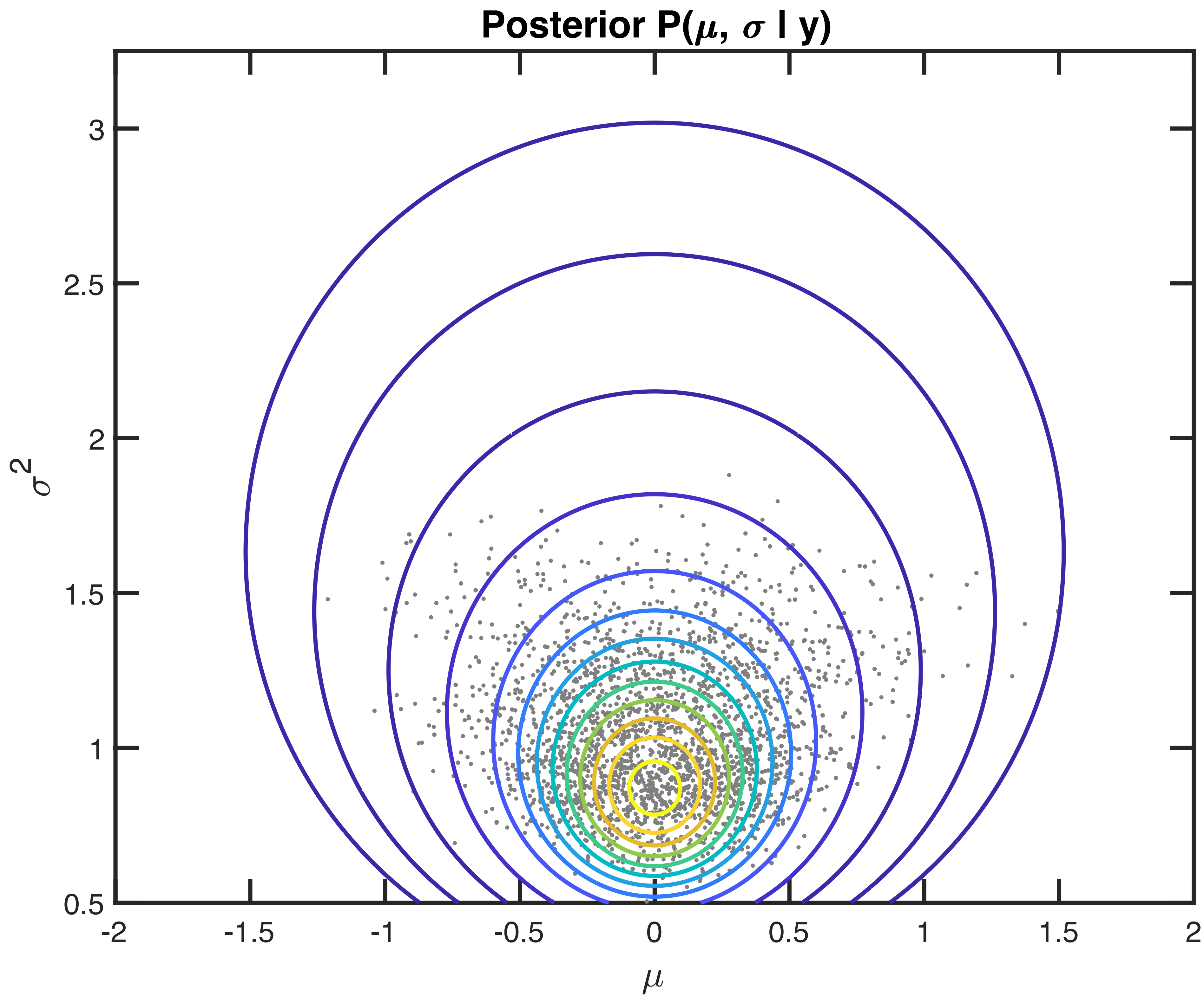
# Metropolis 2D Code Example (metropolis2.m)



# Metropolis 2D Code Example (metropolis2.m)



# Metropolis 2D Code Example (metropolis2.m)



# MCMC Animations

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

# 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 $\psi$, we label the simulation draws as $\psi_{ij}$ ($i = 1, \ldots, n; j = 1, \ldots, m$), 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: $\widehat{R} = \sqrt{\frac{\widehat{\text{var}}^{+}(\psi|y)}{W}}, \quad \approx 1$

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

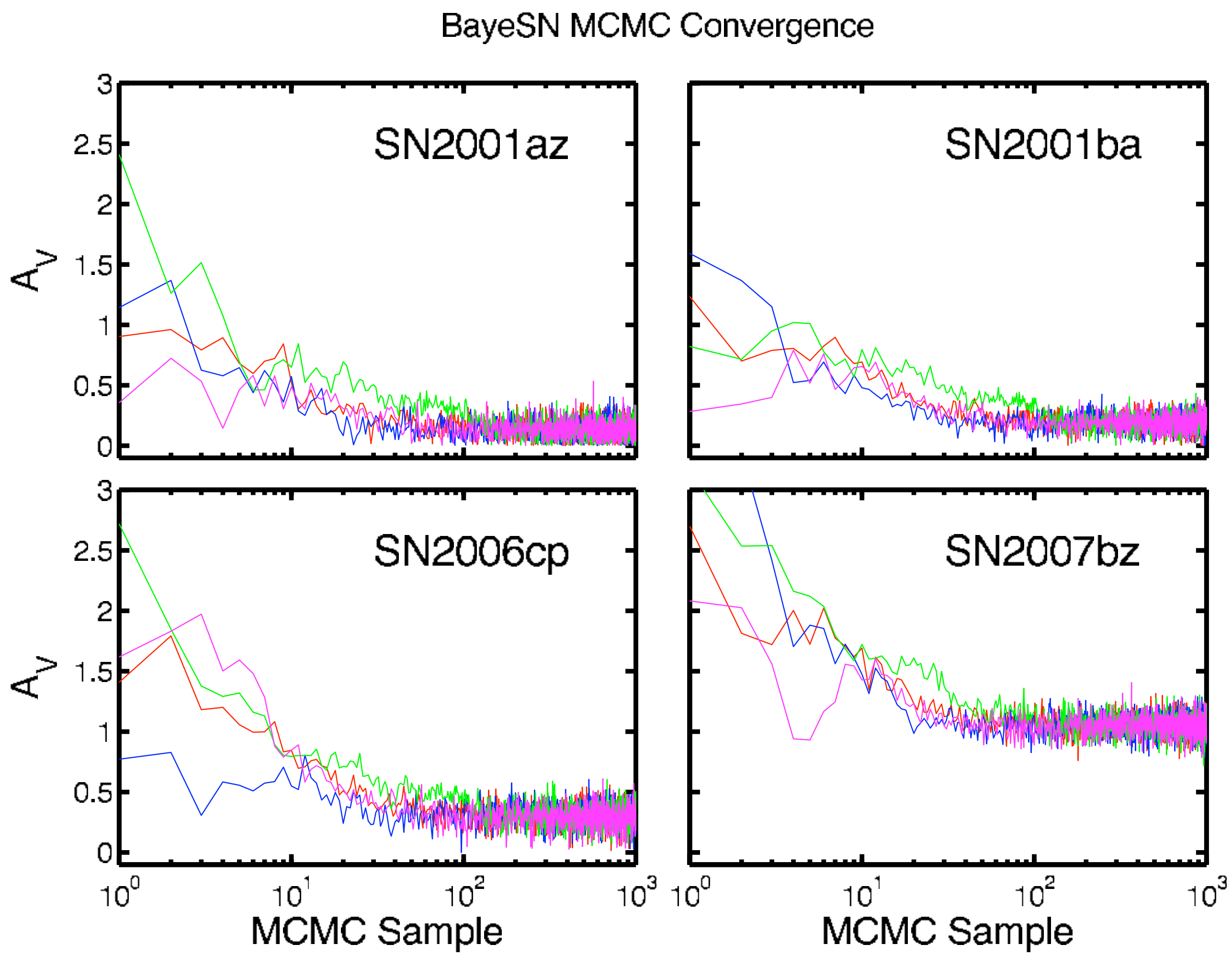



**Figure 2.** Example sample paths of Markov Chain Monte Carlo (MCMC) chains generated by the BayeSN MCMC sampling code. The full chain stochastically

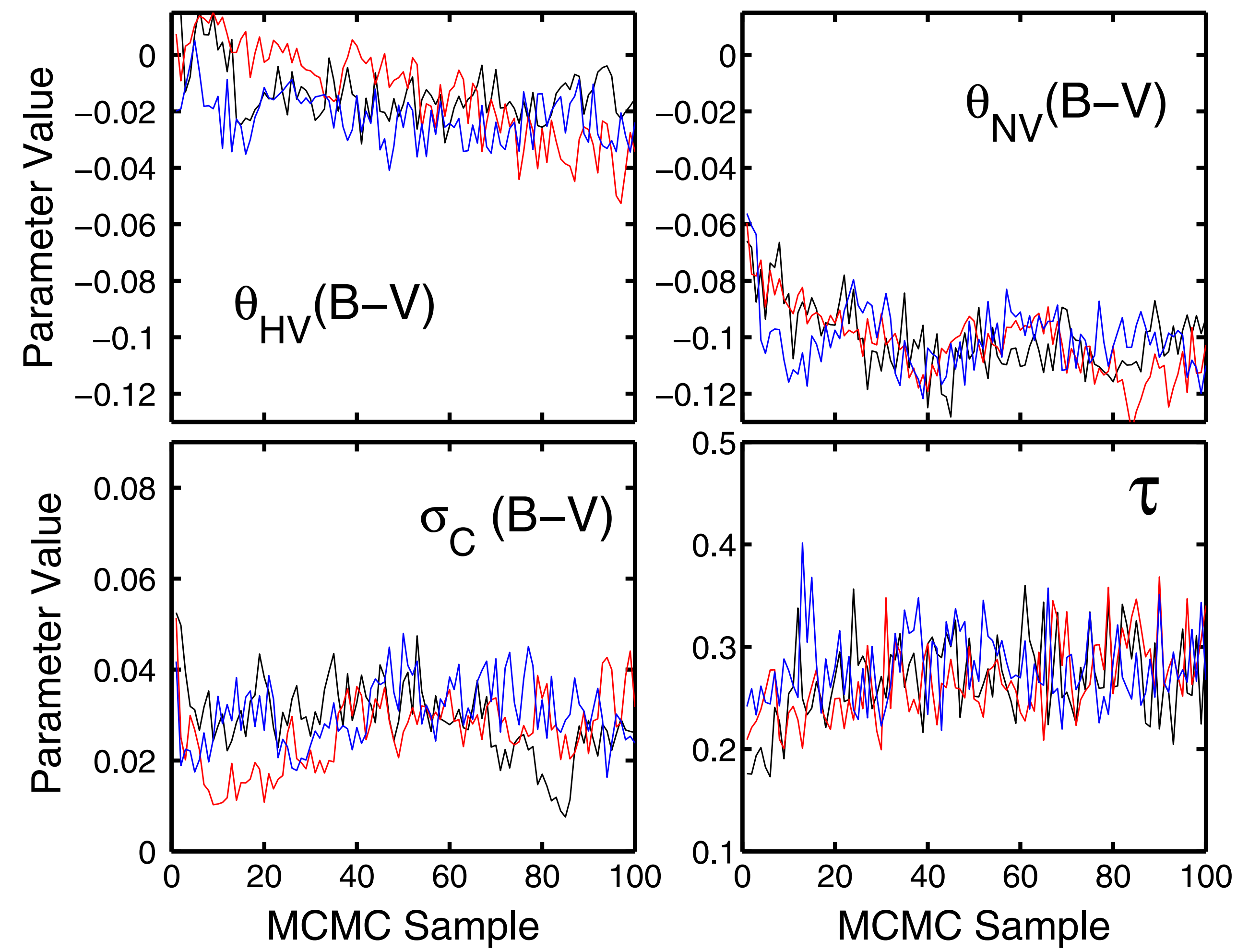


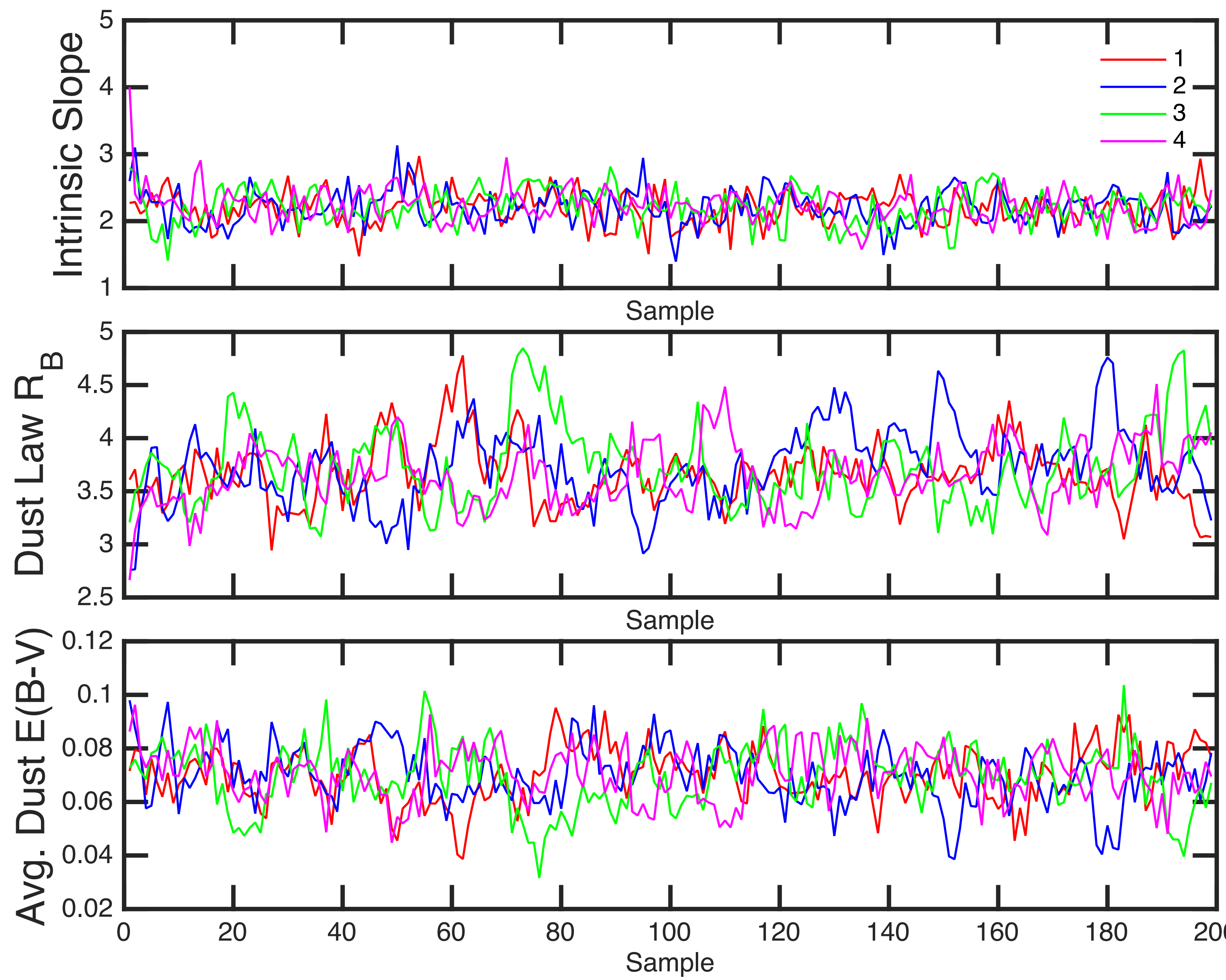

**Figure 10.** Trace plot for a run of MCRC for 1000 cycles with 3 independent chains. The current values of all parameters were recorded every 10 cycles.

Mandel et al. 2011

Mandel et al. 2014

# Assessing Convergence with multiple chains example, Mandel et al. 2017

- Estimate Intrinsic Relation, Dust Law, Dust Population, etc.
- Gibbs Sampling utilizes conditionals of full posterior to update MCMC steps
- Explore joint posterior probability of all parameters




Four Parallel MCMC Chains