# Astrostatistics: Friday 21 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 through Cambridge Library iDiscover)
- Example Class 2, Thu Feb 27, 3:30pm MR13

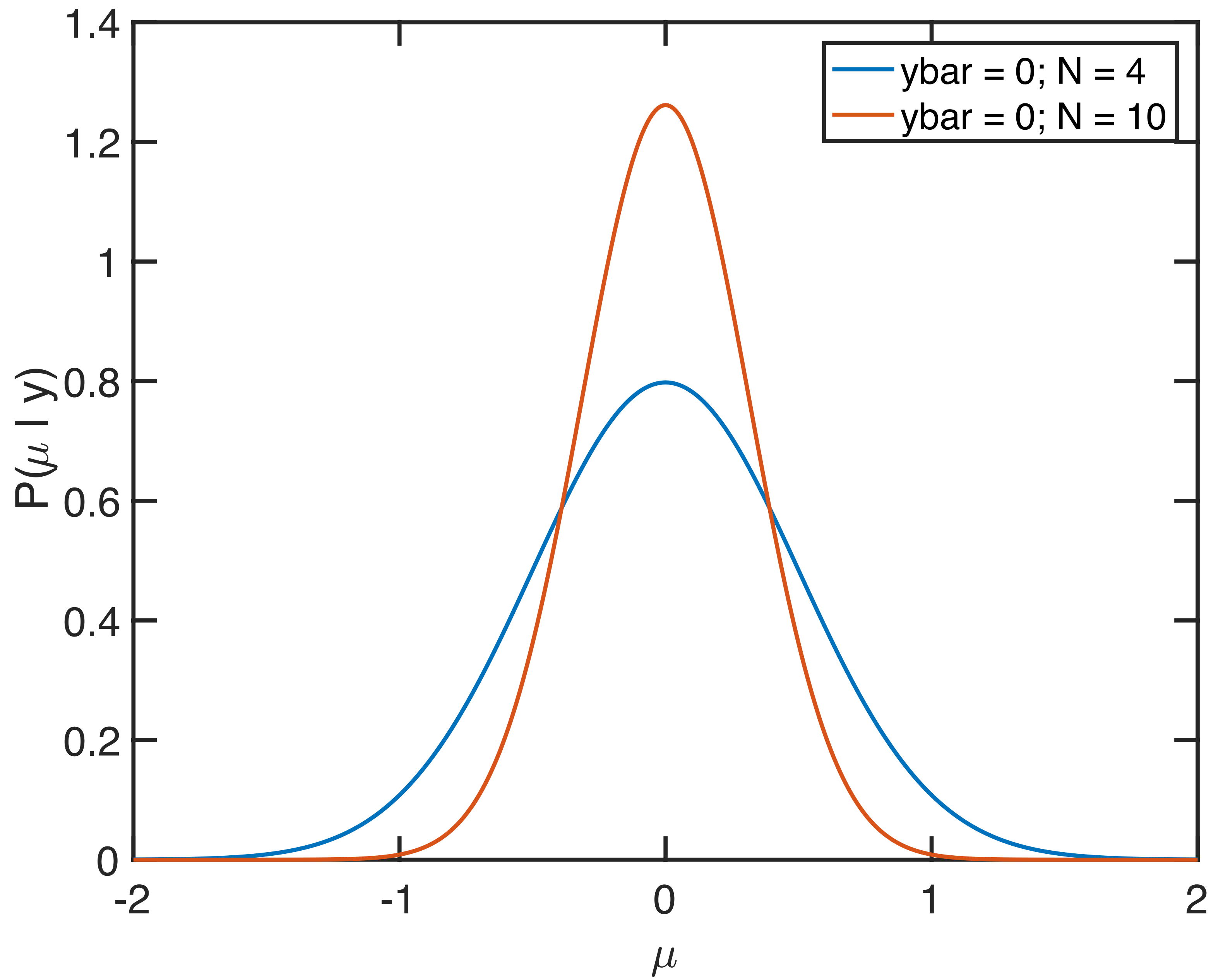
# Simplest MCMC: Metropolis Algorithm

1. Choose a random starting point $\mu_0$
2. At step i = 1…N, propose a new parameter value $\mu_{prop} \sim N(\mu_i, \tau^2)$. The proposal scale $\tau$ is chosen cleverly.
3. Evaluate ratio of posteriors at proposed vs current values. Metropolis Ratio $r = P(\mu_{prop} \mid \mathbf{y}) / P(\mu_i \mid \mathbf{y})$.
4. If $\mu_{prop}$ is a better solution (higher posterior), $r > 1$, accept the new value $\mu_{i+1} = \mu_{prop}$. Else accept with probability r (i.e. accept with probability min(r,1). **[If not accept, stay at same value $\mu_{i+1} = \mu_i$ & include in chain]**.
5. Repeat steps 2-4 until reach some measure of convergence and gather enough samples to compute your inference

# Simple Gaussian mean μ (where we know the answer)

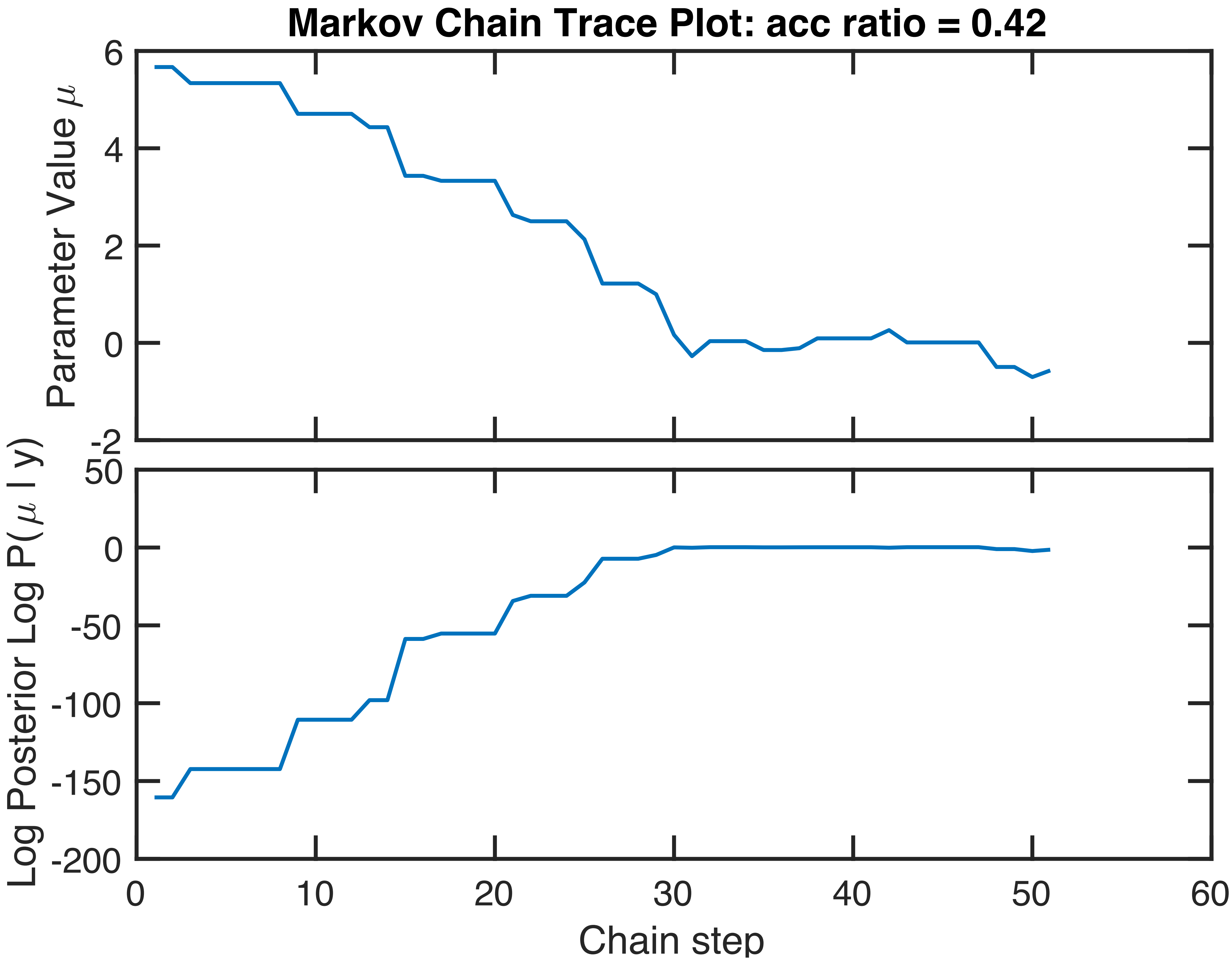
$$y_i \sim N(\mu, \sigma^2 = 1), i = 1 \dots N$$

$$P(\mu) \propto 1 \qquad P(\mu | \boldsymbol{y}) = N(\mu | \bar{y}, \sigma^2 / N)$$



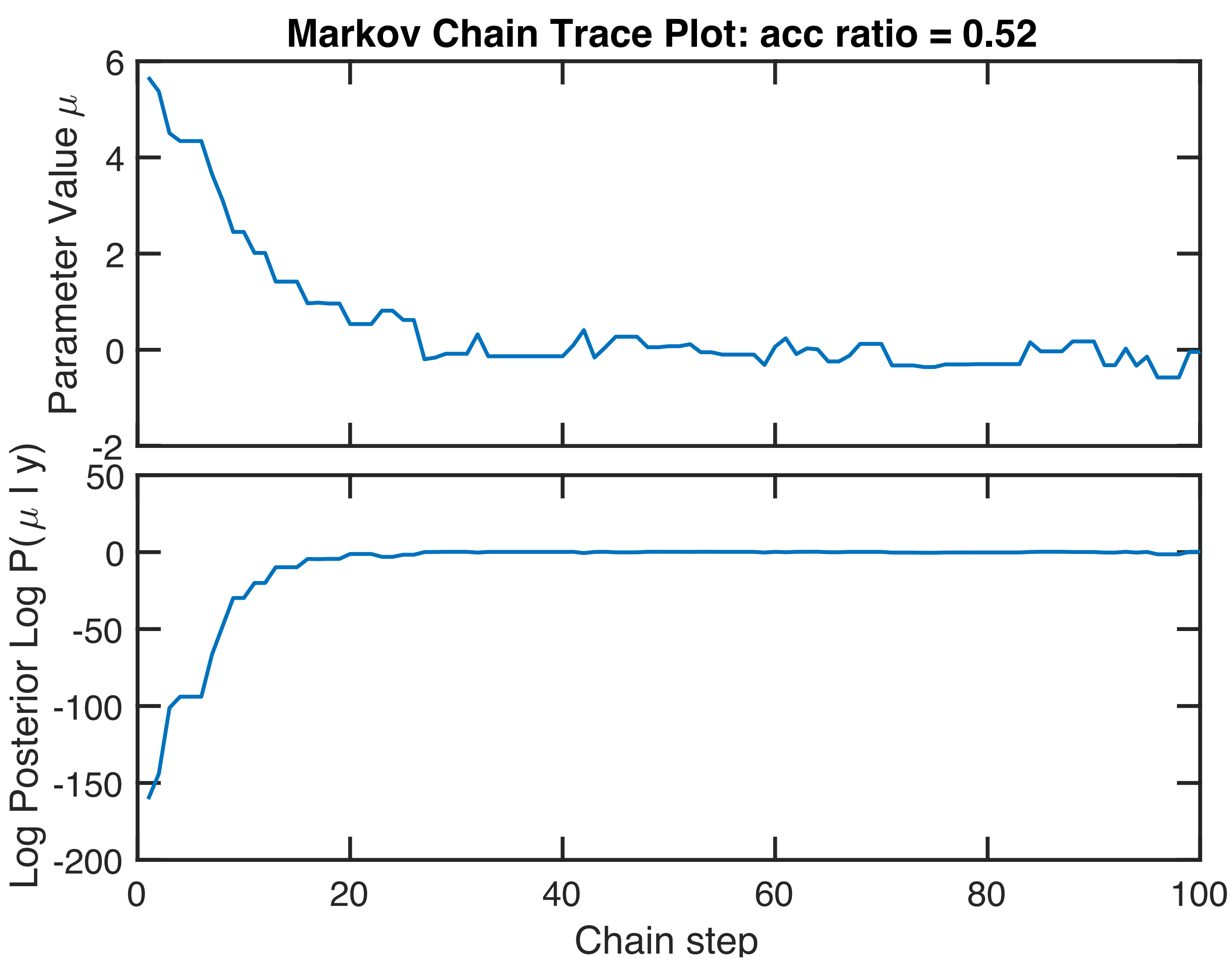
# Code demo: metropolis1.m

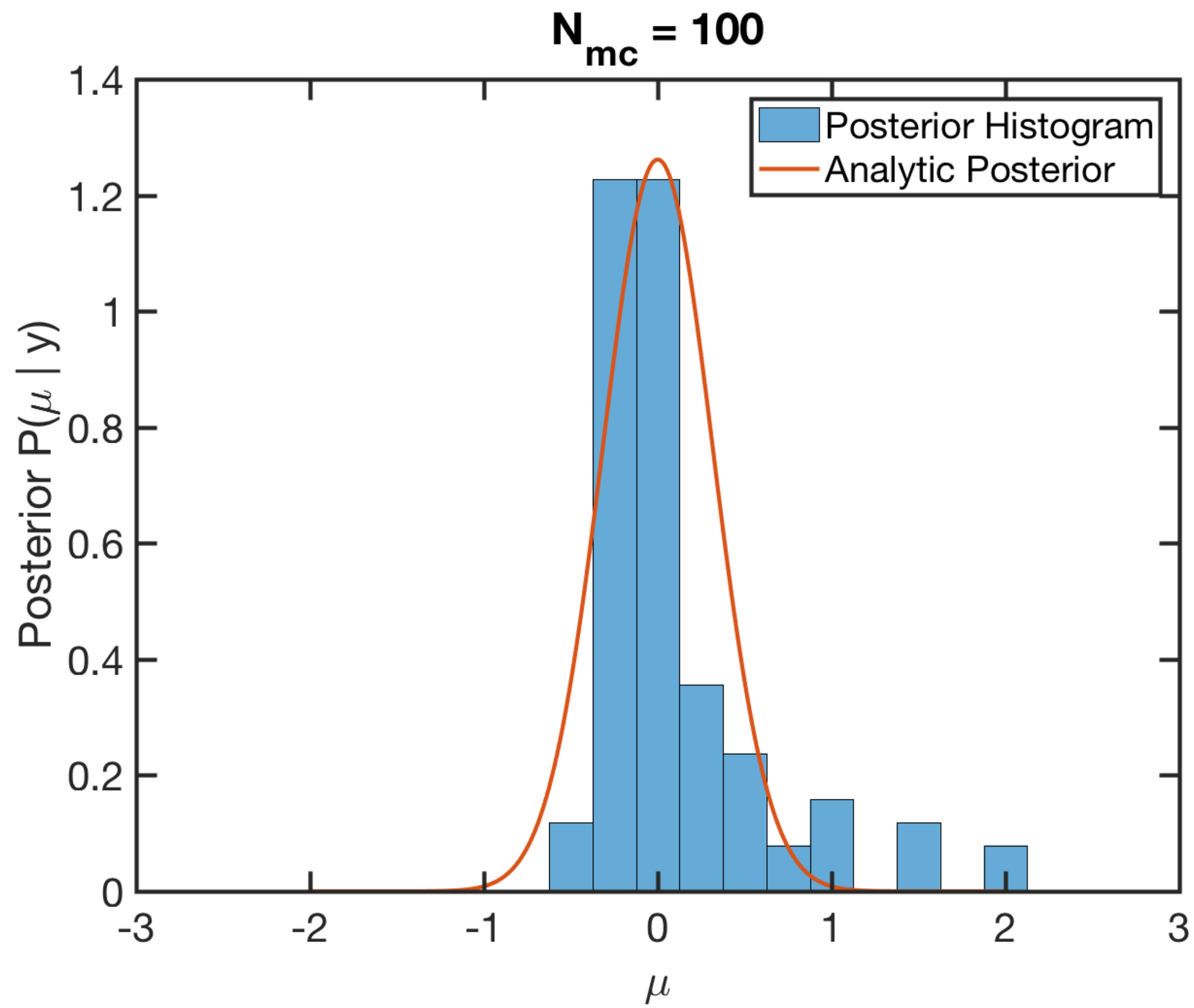
## First 50 iterations



# Code demo: metropolis1.m

## First 100 iterations



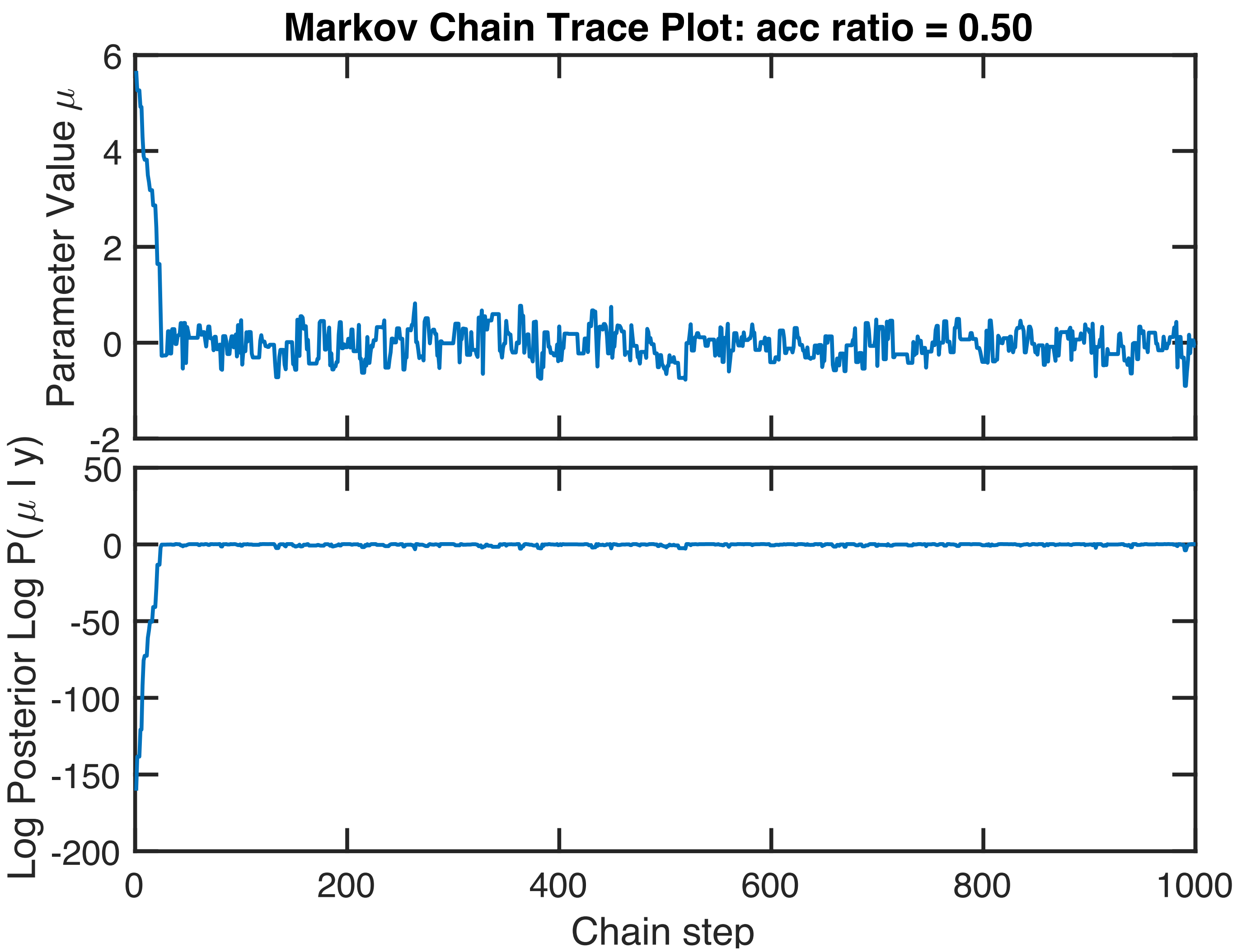




Trace Plot

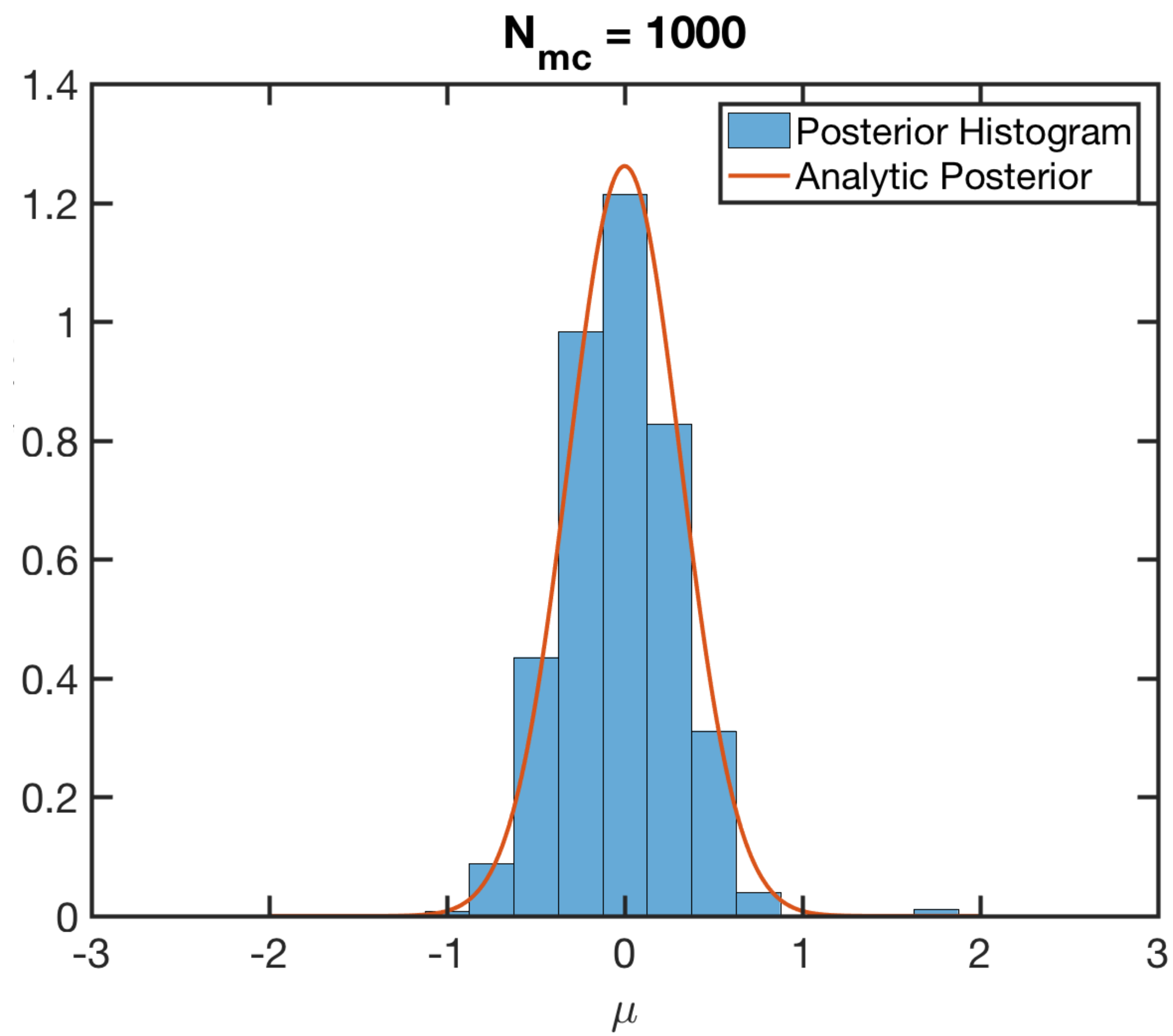
Posterior

# Code demo: metropolis1.m

## First 1000 iterations



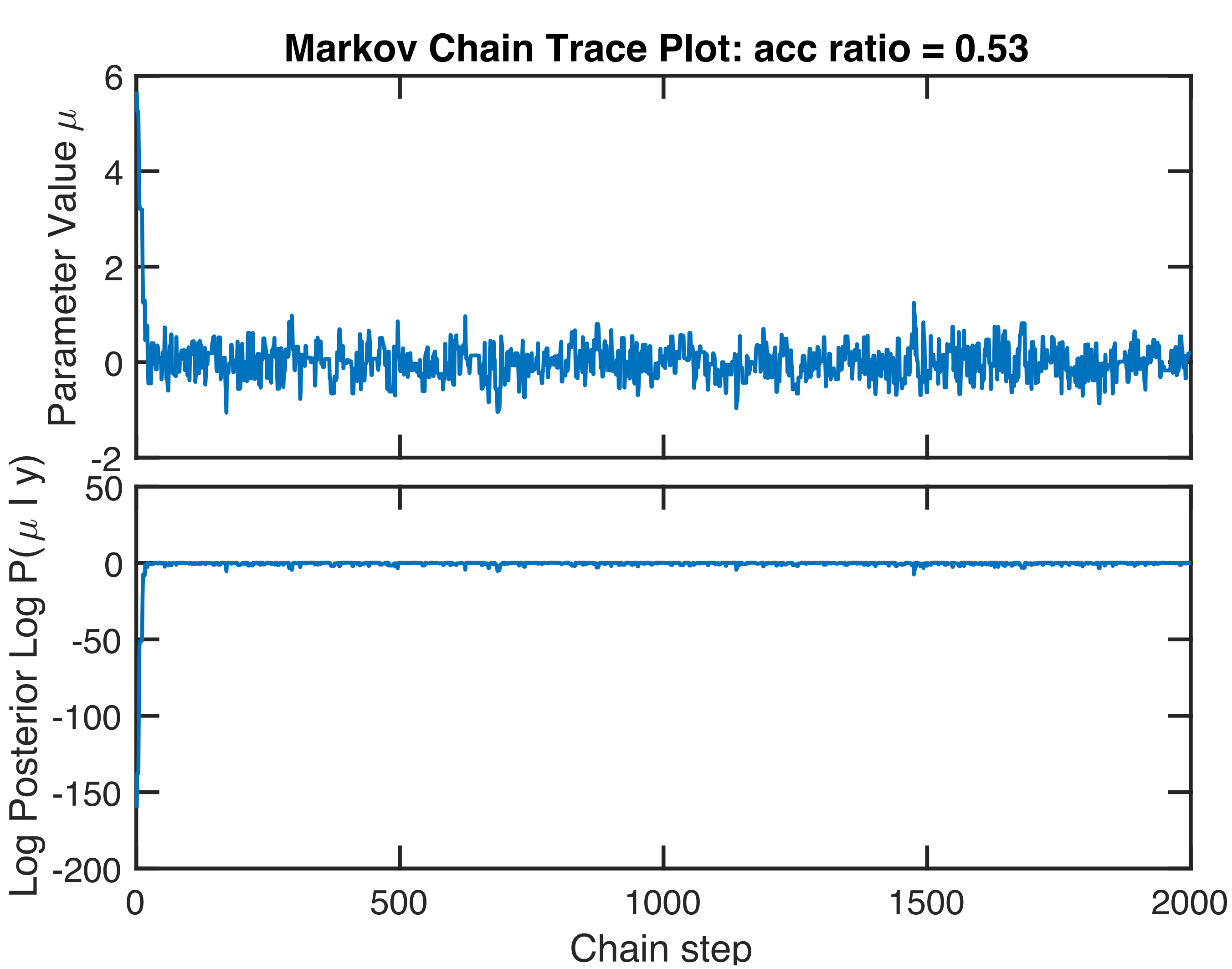

Trace Plot

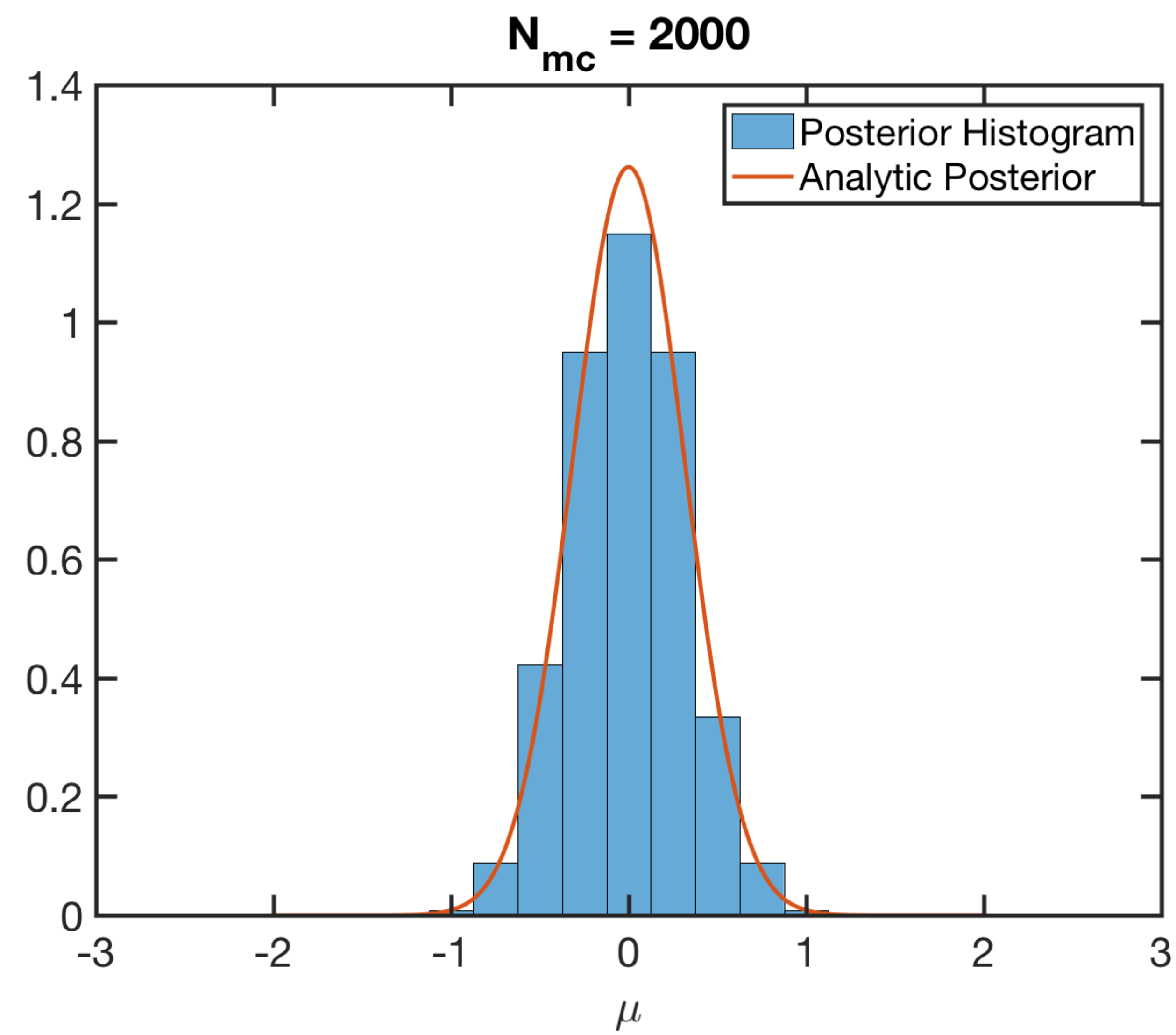



Posterior

# Code demo: metropolis1.m

## 2000 iterations



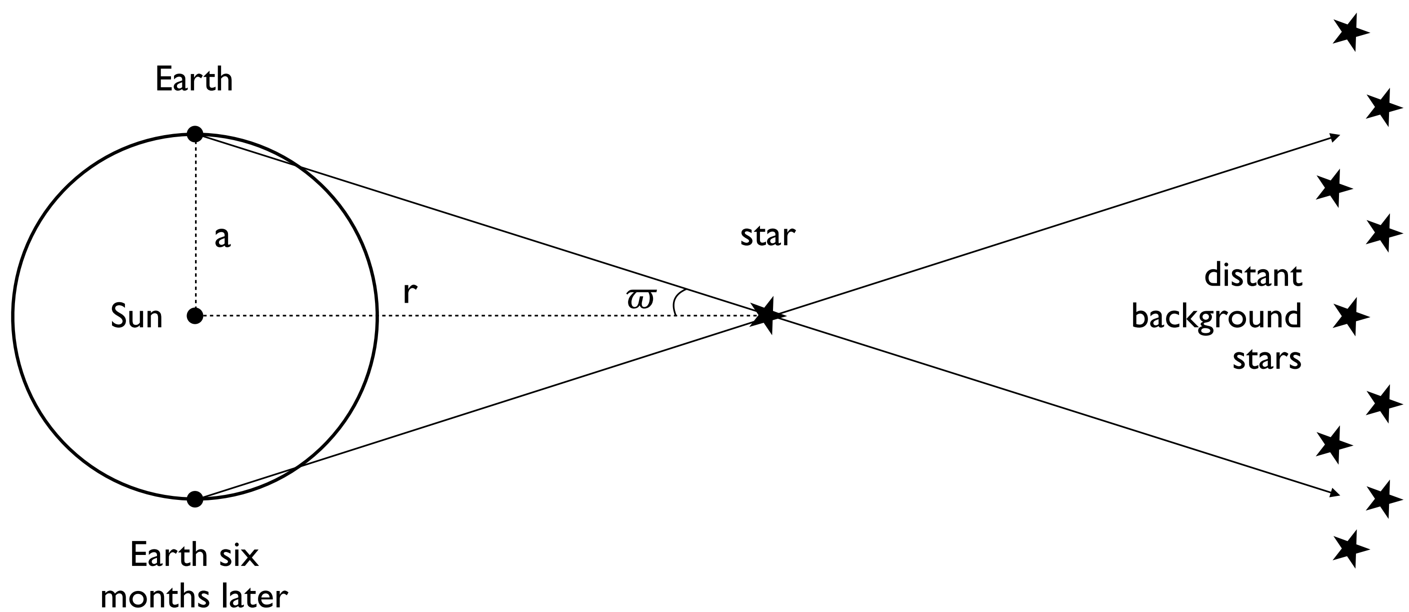




Trace Plot

Posterior histogram of 500 samples after cutting 50% burn-in & thinning by 2

# Parallax Example
# Likelihood:

$$P(\varpi\,|\,r) \;=\; \frac{1}{\sigma_\varpi\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi-\frac{1}{r}\right)^2\right] \quad \text{where} \quad \sigma_\varpi > 0,$$




The parallax $\varpi$ of a star is the apparent angular displacement of that star (relative to distant background stars) due to the orbit of the Earth about the Sun. More precisely, the parallax is the angle subtended by the Earth's orbital radius $a$ as seen from the star. As parallaxes are extremely small angles ($\varpi \ll 1$), $\varpi = a/r$ to a very good approximation. When $\varpi$ is 1 arcsecond, $r$ is defined as the *parsec*, which is about $3.1\times10^{13}$ km. In this sketch the size of the Earth's orbit has been greatly exaggerated compared to the distance to the star, and the distance to the background stars in reality is orders of magnitude larger again.

Parallax Angle

$$\frac{\omega}{\text{arcsec}} = \frac{\text{parsec}}{r}$$

Distance

# Introducing physical constraints into the prior

$$P(r) = \begin{cases} \dfrac{1}{2L^3}\, r^2 e^{-r/L} & \text{if } \ r > 0 \\ 0 & \text{otherwise} \end{cases}$$

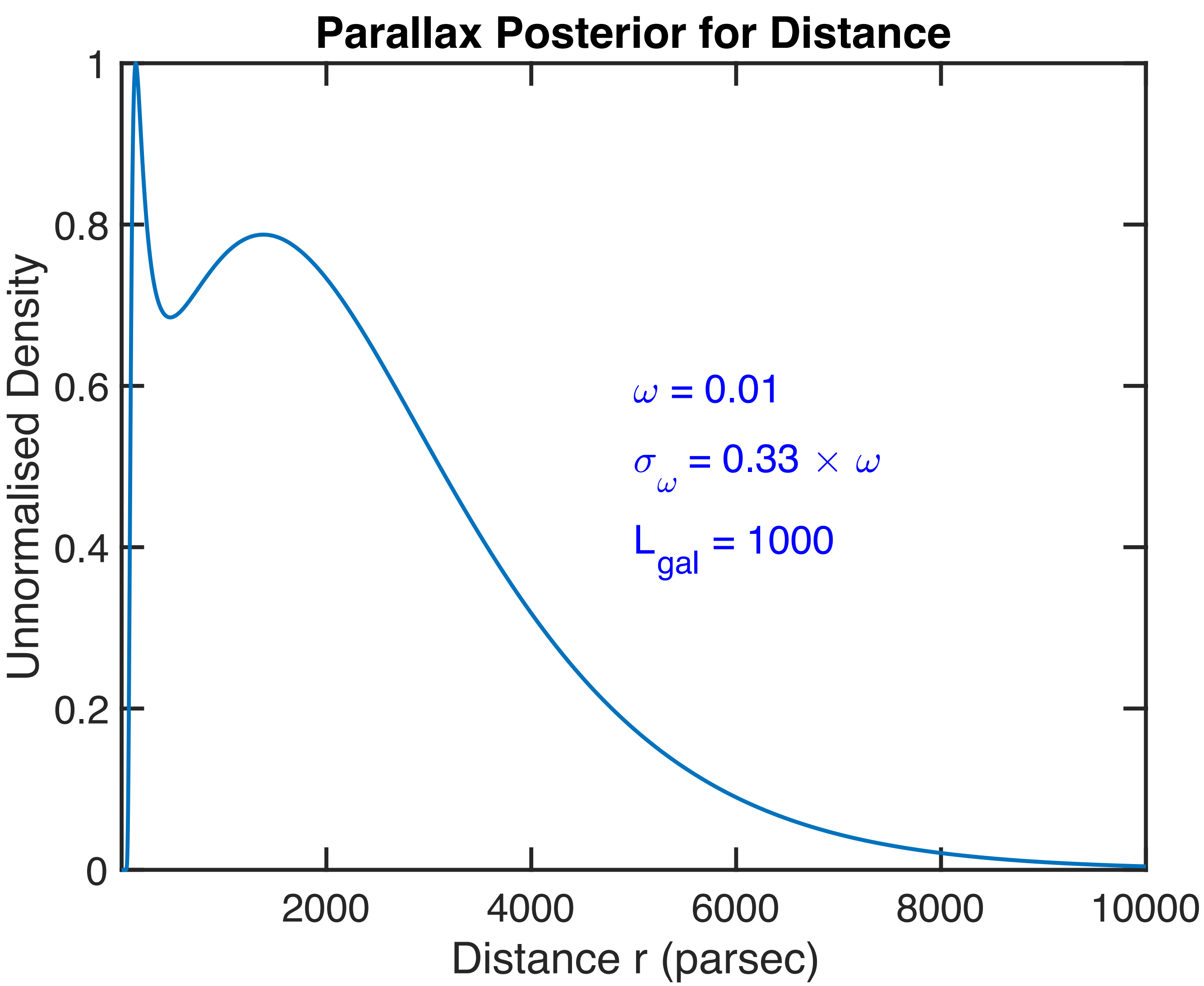
Exponential decrease in density of stars with Galactic length scale L

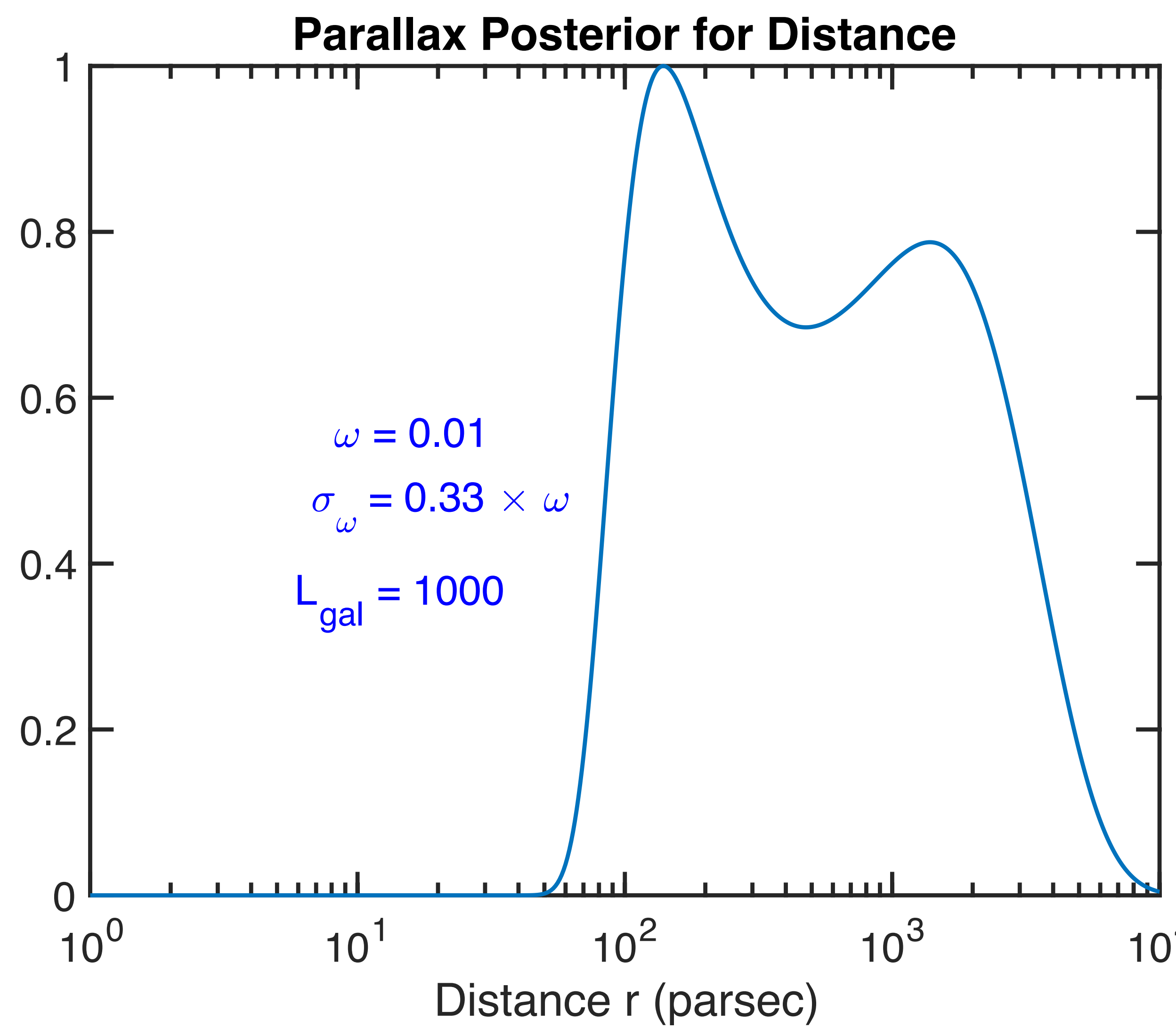
$$P(r|\,\omega) \propto P(\omega|r) \times P(r)$$

**Unnormalised** Posterior:

$$P^*_{r^2 e^{-r}}(r|\varpi, \sigma_\varpi) = \begin{cases} \frac{r^2 e^{-r/L}}{\sigma_\varpi} \exp\left[-\frac{1}{2\sigma_\varpi^2}\left(\varpi - \frac{1}{r}\right)^2\right] & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases}.$$

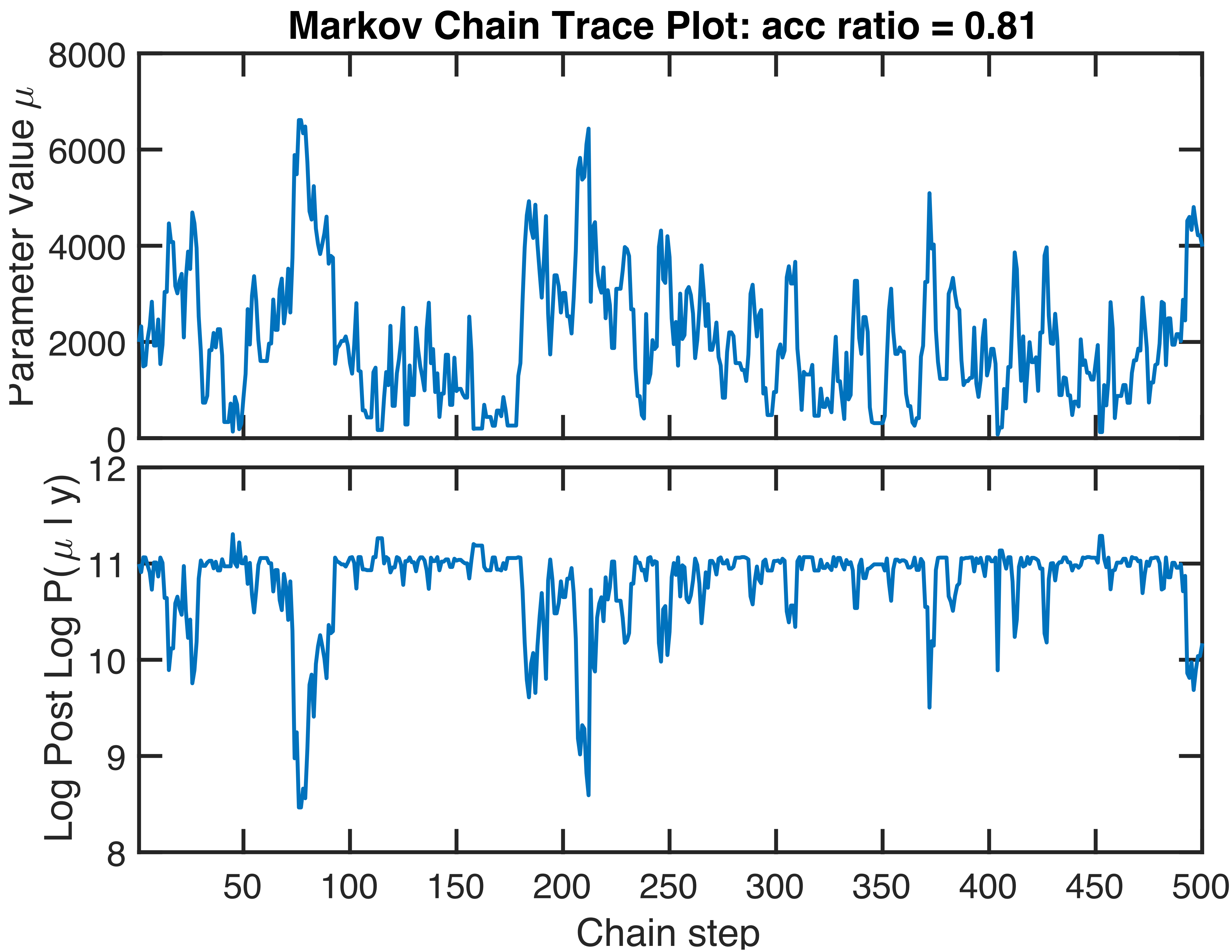
# Parallax Example



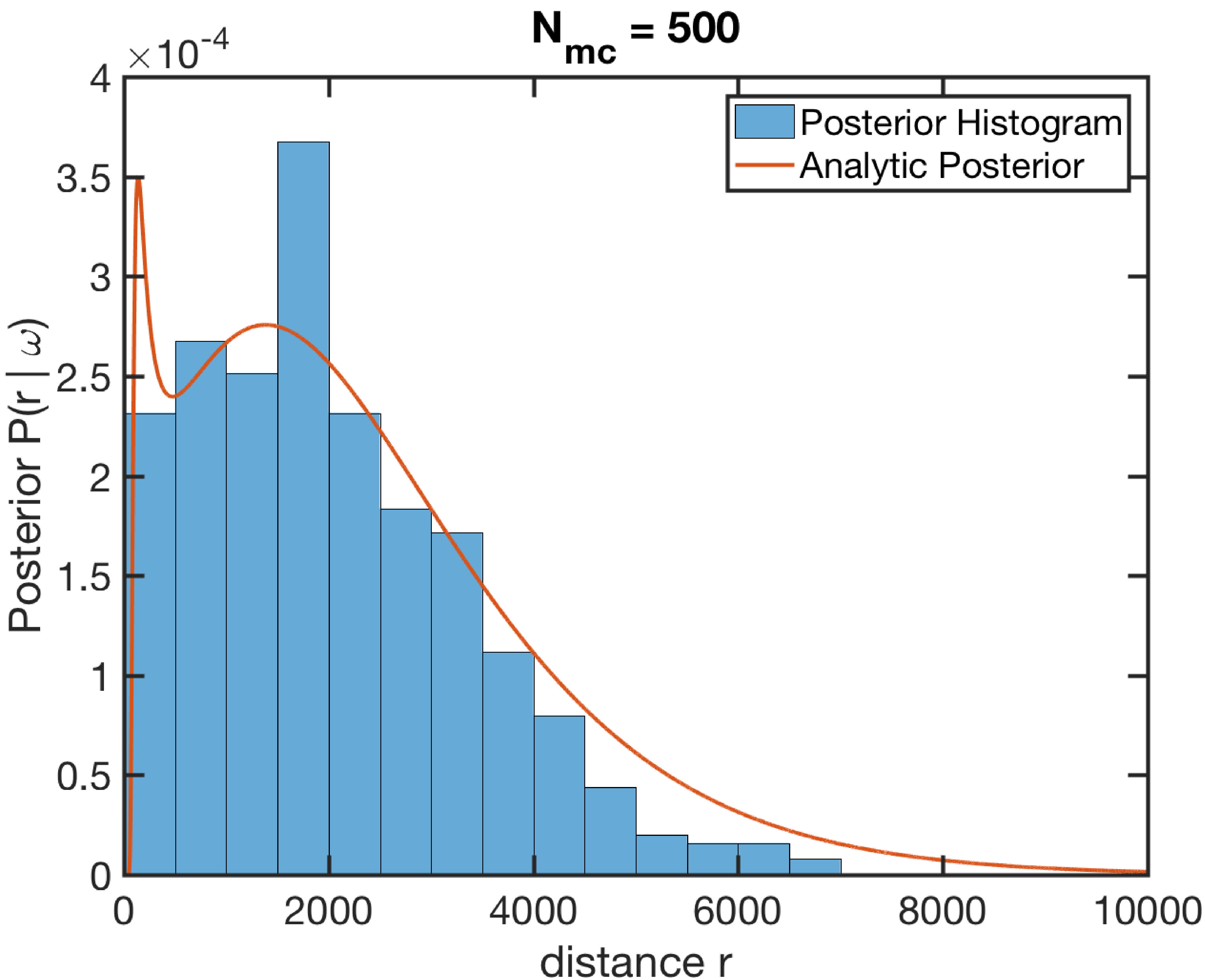




Log Scale

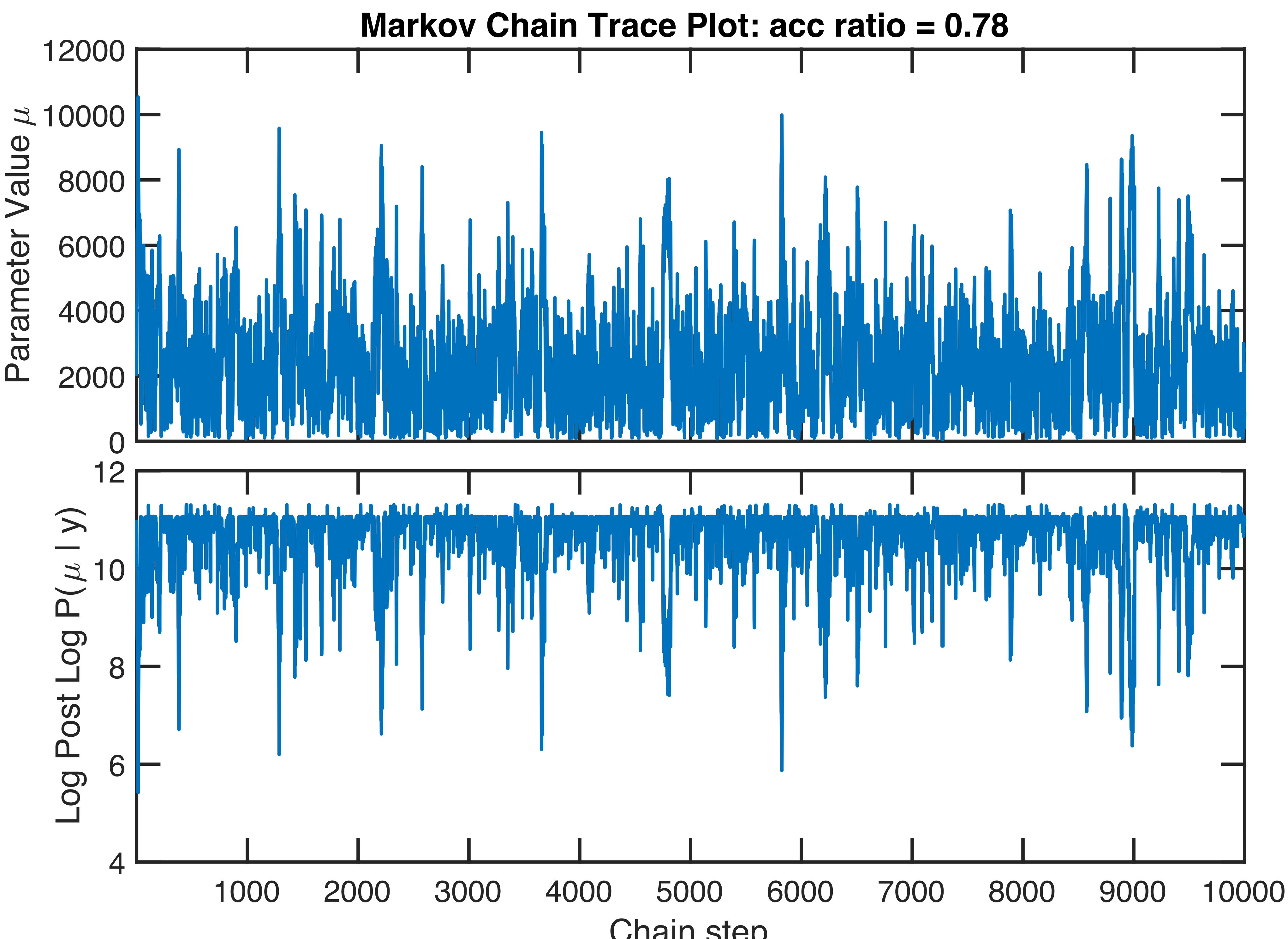
# Parallax Example



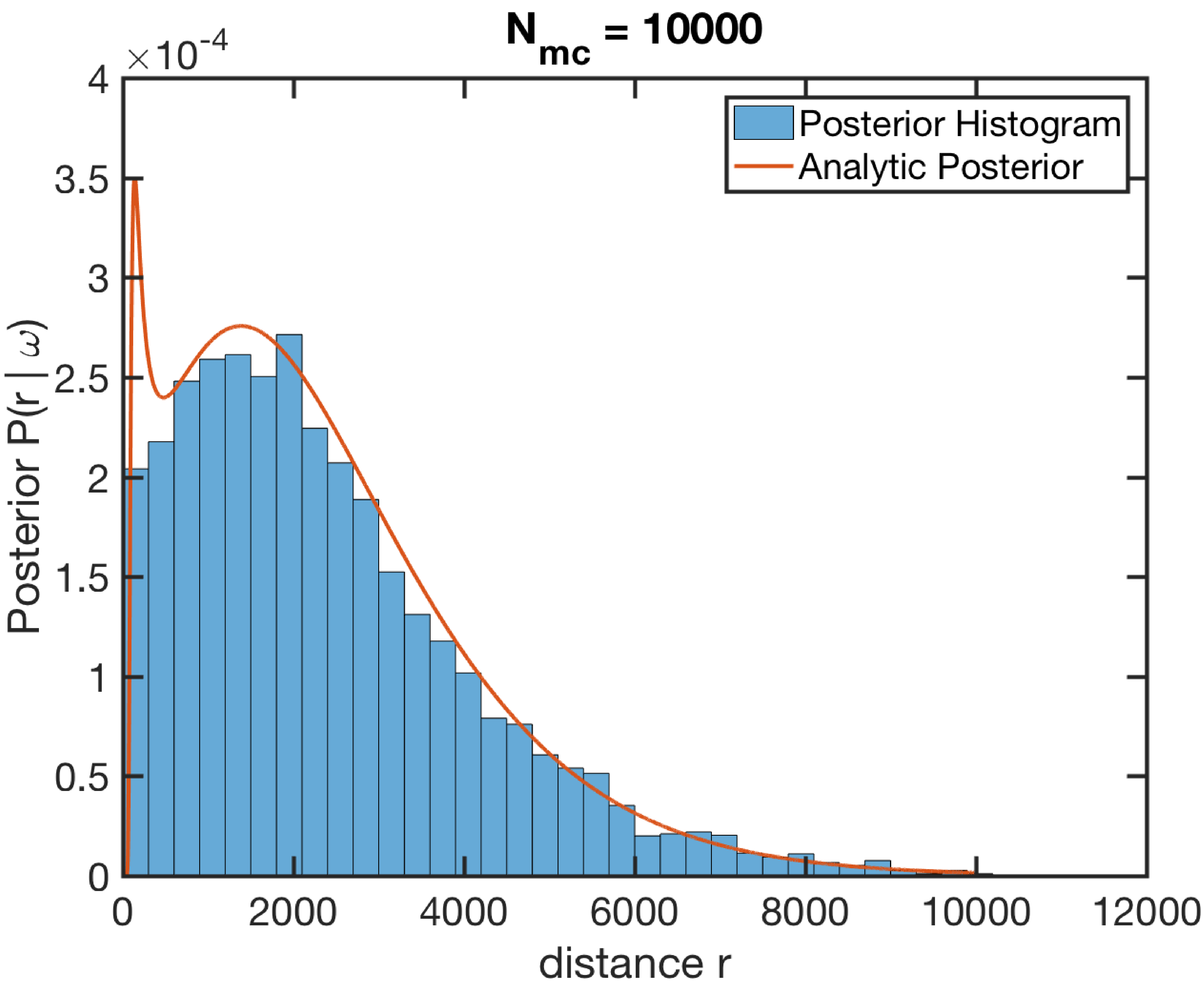
# Parallax Example



# Parallax Example



# Parallax Example



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

- Markov Chain Monte Carlo (MCMC)
- Just did: 1D Metropolis algorithm
- Now:
    - 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

# d-dim Metropolis Algorithm: Posterior $P(\theta | D)$, Symmetric Proposal/Jump dist'n $J(\theta^*| \theta) = J(\theta| \theta^*)$

1. Choose a random starting point $\theta_0$

2. At step i = 1…N, propose a new parameter value $\theta^* \sim N(\theta_{i-1}, \boldsymbol{\Sigma_p})$. The proposal distr'n is $J(\theta^*| \theta_{i-1}) = N(\theta^*| \theta_{i-1}, \boldsymbol{\Sigma_p})$

3. Evaluate ratio of posteriors at proposed vs current values. r = $P(\theta^* | \mathbf{y}) / P(\theta_{i-1} | \mathbf{y})$.

4. Accept $\theta^*$ with probability min(r,1): $\theta_i = \theta^*$. If not accept, stay at same value $\theta_i = \theta_{i-1}$ for the next step & include in chain.

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)

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

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

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

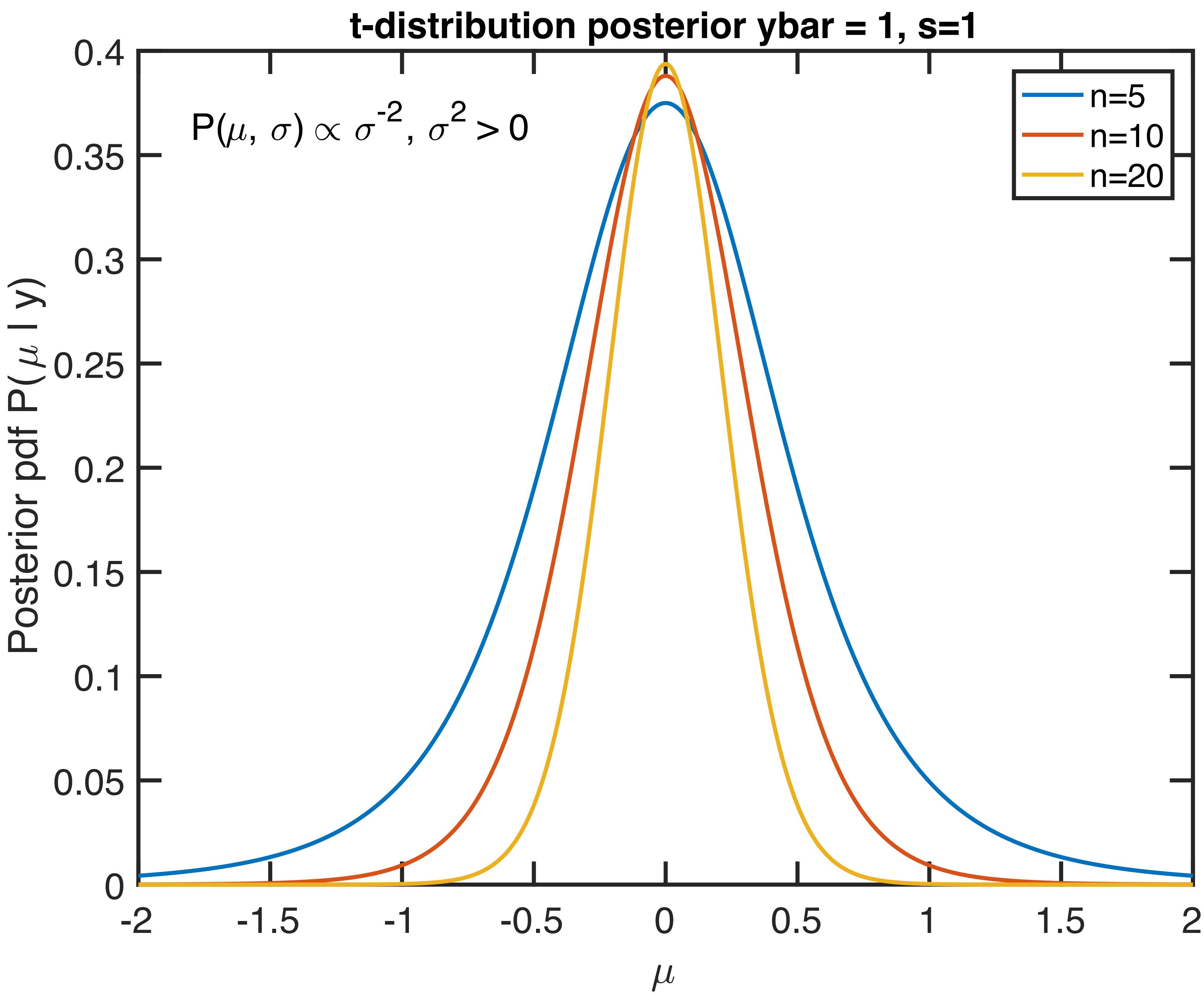
Prior: $P(\mu) \propto 1 \qquad P(\sigma^2) \propto \sigma^{-2}, \sigma^2 > 0$

Posterior

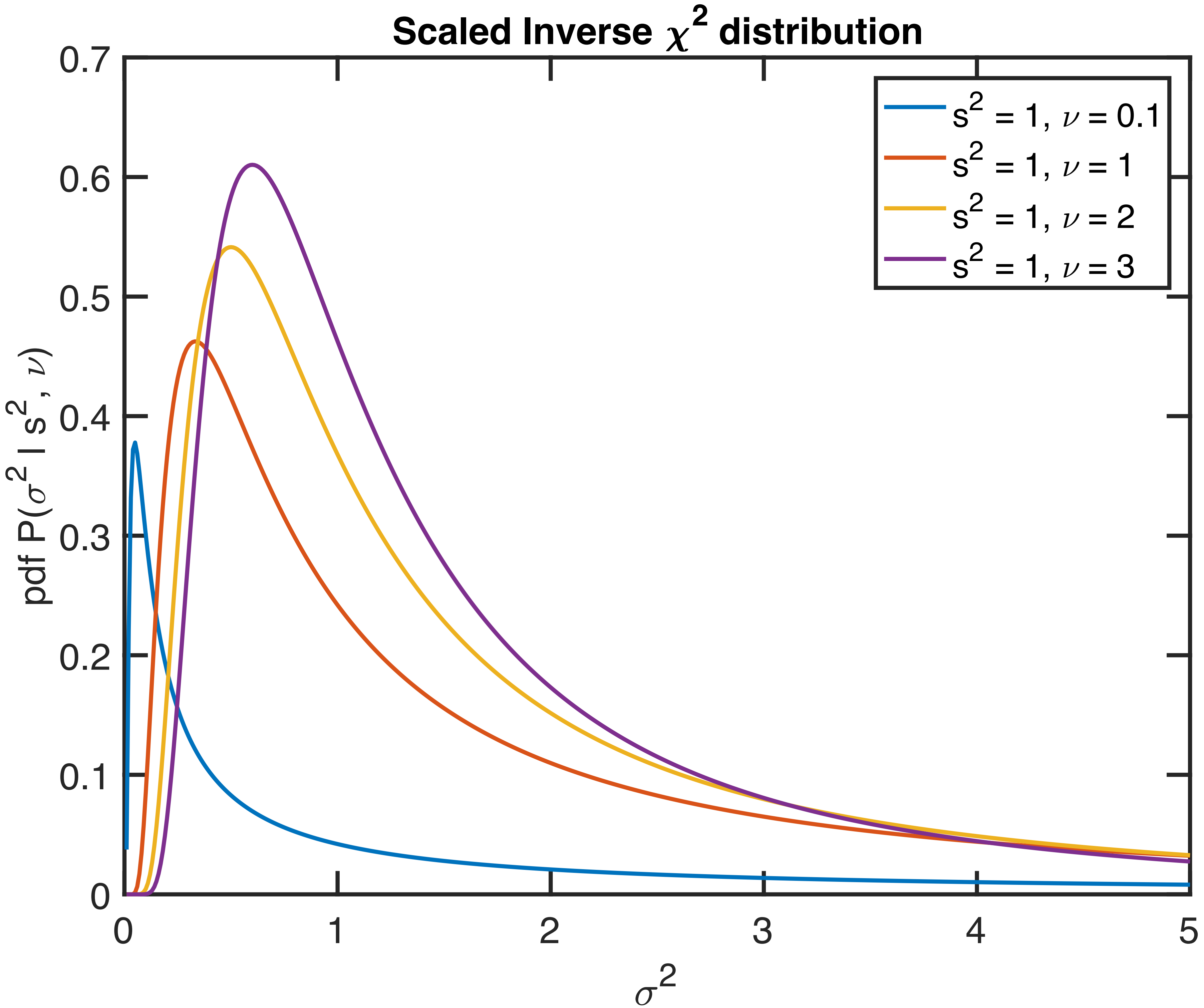
$$P(\mu, \sigma^2|\boldsymbol{y}) \propto (\sigma^2)^{-(n+2)/2} \exp\left(-\tfrac{1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\right)$$

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

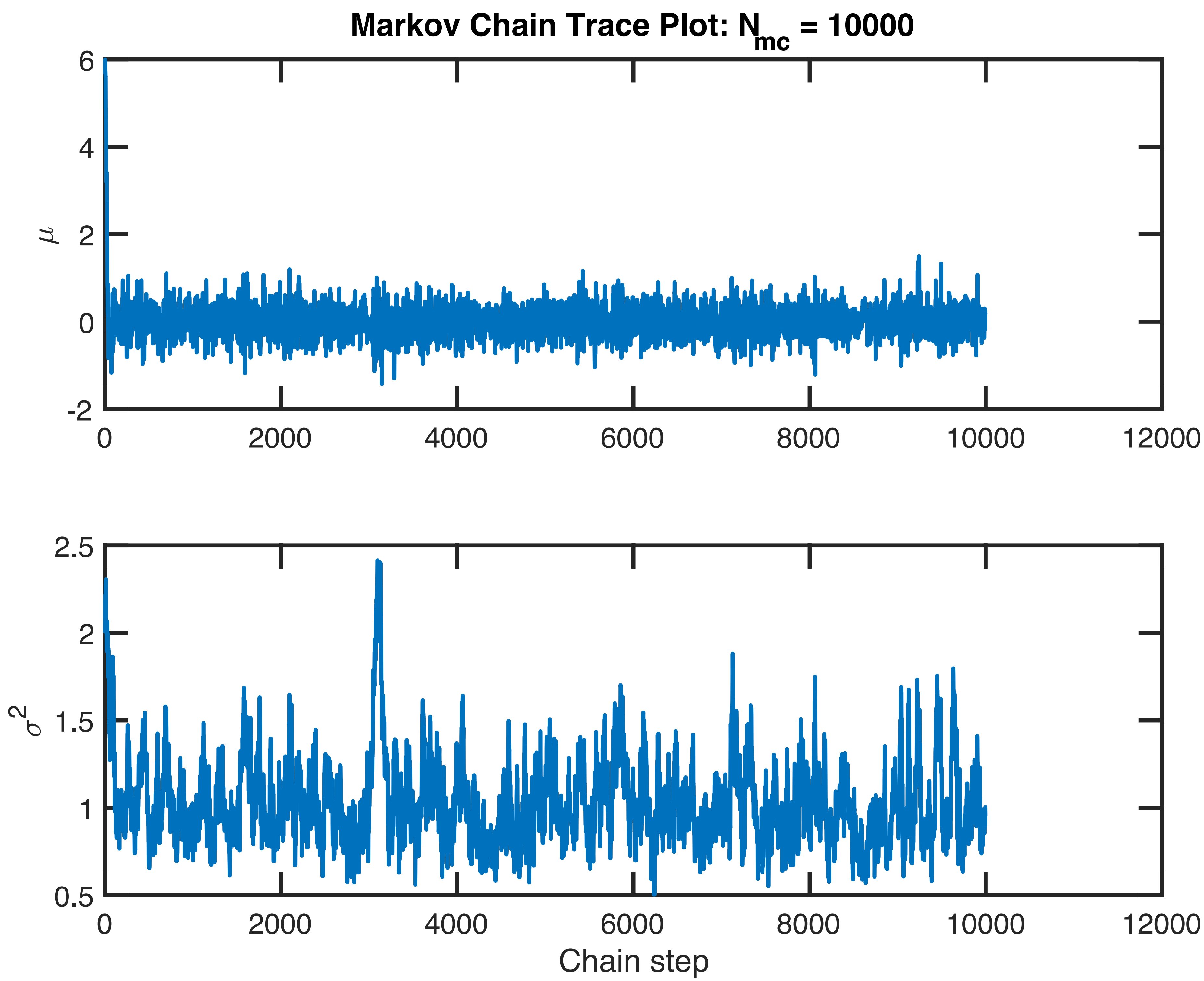
# Before: Posterior Distribution of a Gaussian Mean
## Analytic Result



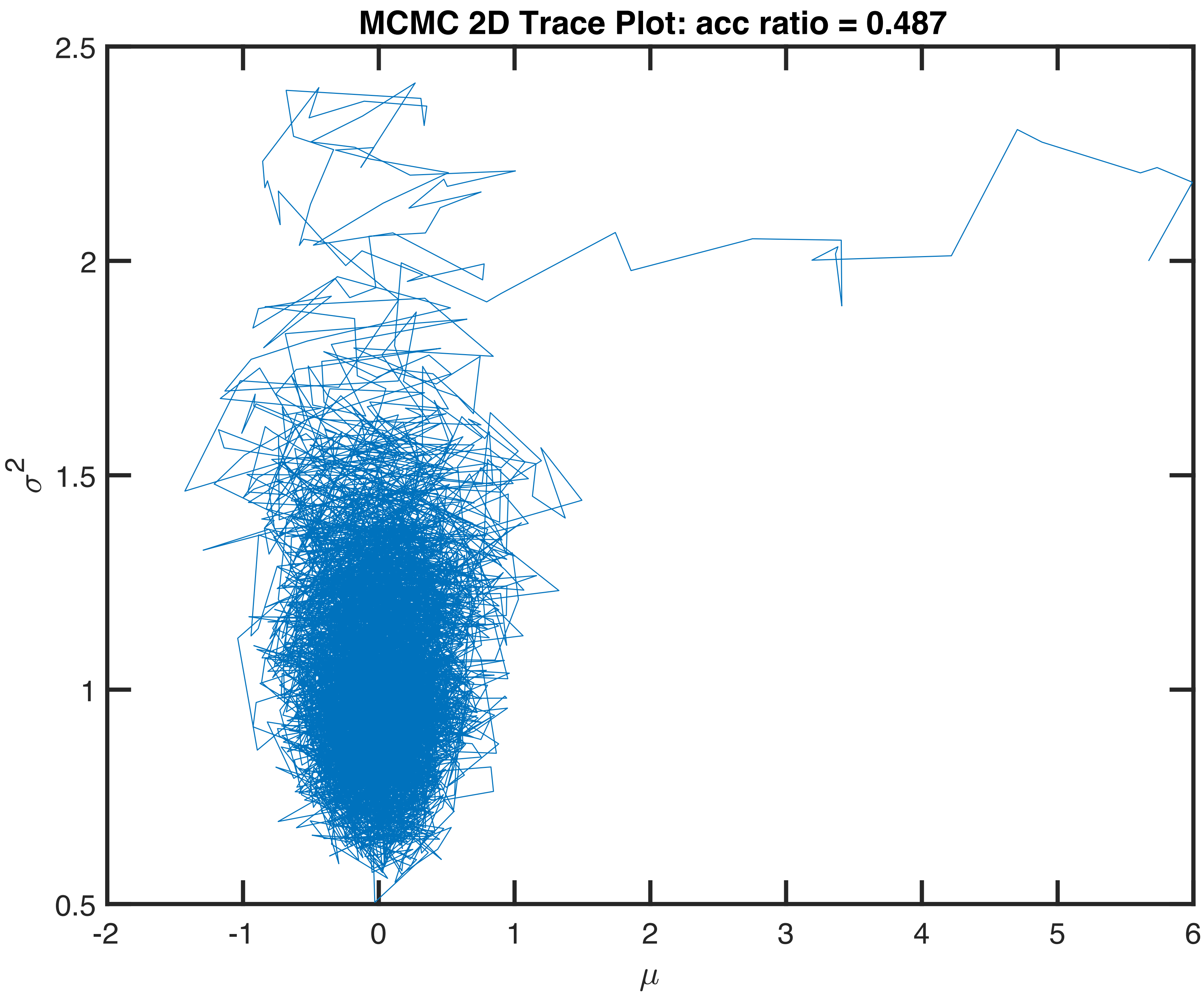
# Posterior Distribution P($\sigma^2$ | y) follows a Scaled Inverse $\boldsymbol{\chi}^2$ distribution (~ inverse gamma)



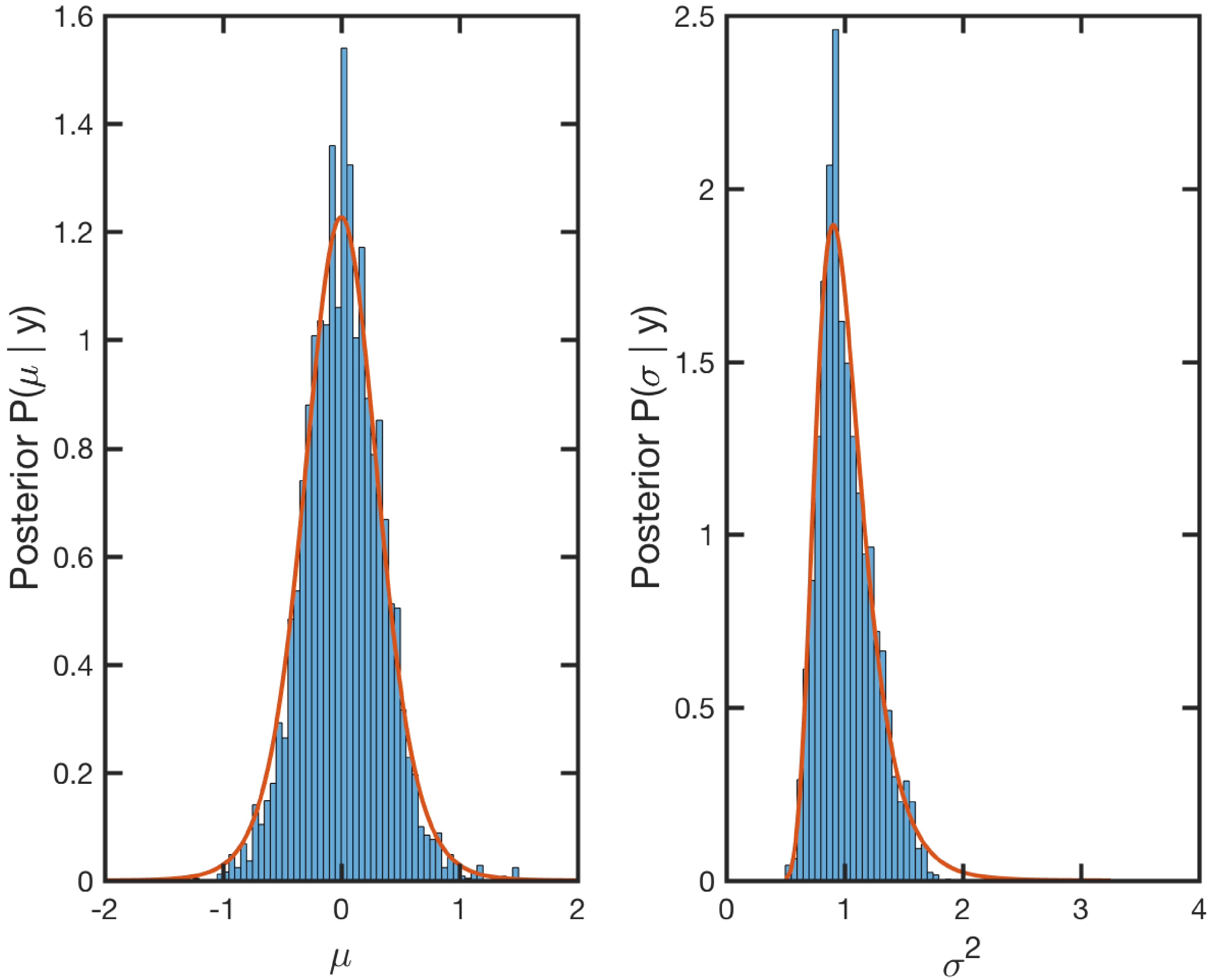
# Metropolis 2D Code Example (metropolis2.m)



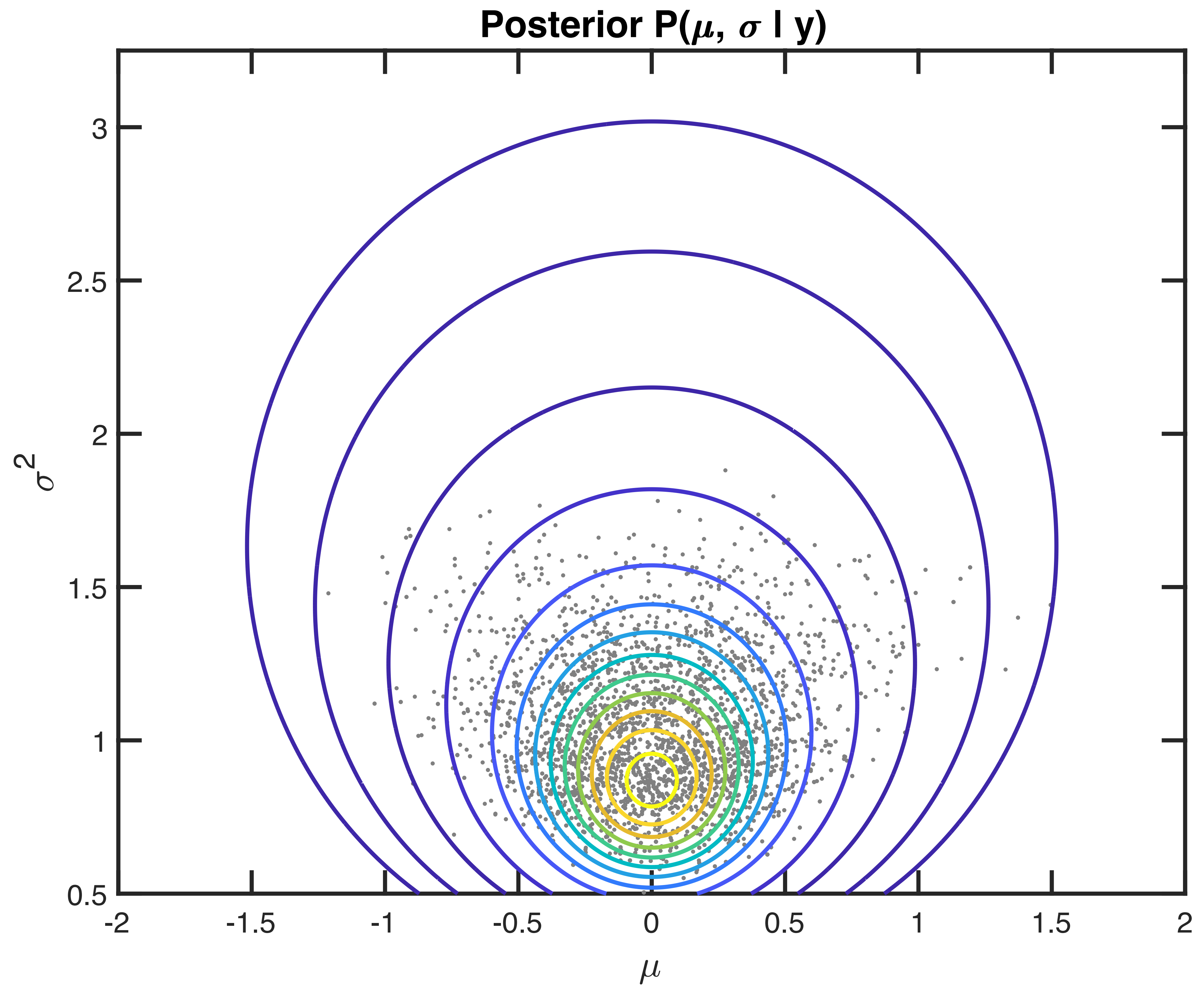
# Metropolis 2D Code Example (metropolis2.m)



# Metropolis 2D Code Example (metropolis2.m)



# Metropolis 2D Code Example (metropolis2.m)



# Tuning d-dim Metropolis

- $\theta^* \sim N(\theta_i, \mathbf{\Sigma_p})$ : if proposal scale $\mathbf{\Sigma_p}$ is too large, will get too many rejections and not go anywhere. If proposal scale too small, you will accept very many small moves: inefficient random walk

- Laplace Approximation: $$P(\boldsymbol{\theta}|\boldsymbol{D}) \approx N(\boldsymbol{\theta}|\,\hat{\boldsymbol{\theta}}, \boldsymbol{\Sigma})$$

$$\hat{\boldsymbol{\theta}} = \text{ posterior mode} \qquad \left(\boldsymbol{\Sigma}^{-1}\right)_{ij} = -\frac{\partial^2 \log P(\boldsymbol{\theta}|\,\boldsymbol{D})}{\partial\theta_i \partial\theta_j}\Big|_{\hat{\boldsymbol{\theta}}}$$

- Choose $\mathbf{\Sigma_p} = c^2\,\mathbf{\Sigma}$ : $$c \approx 2.4/\sqrt{d}$$

- Scale Proposal to aim for an acceptance ratio of 44% in 1D, 23% in d > 5