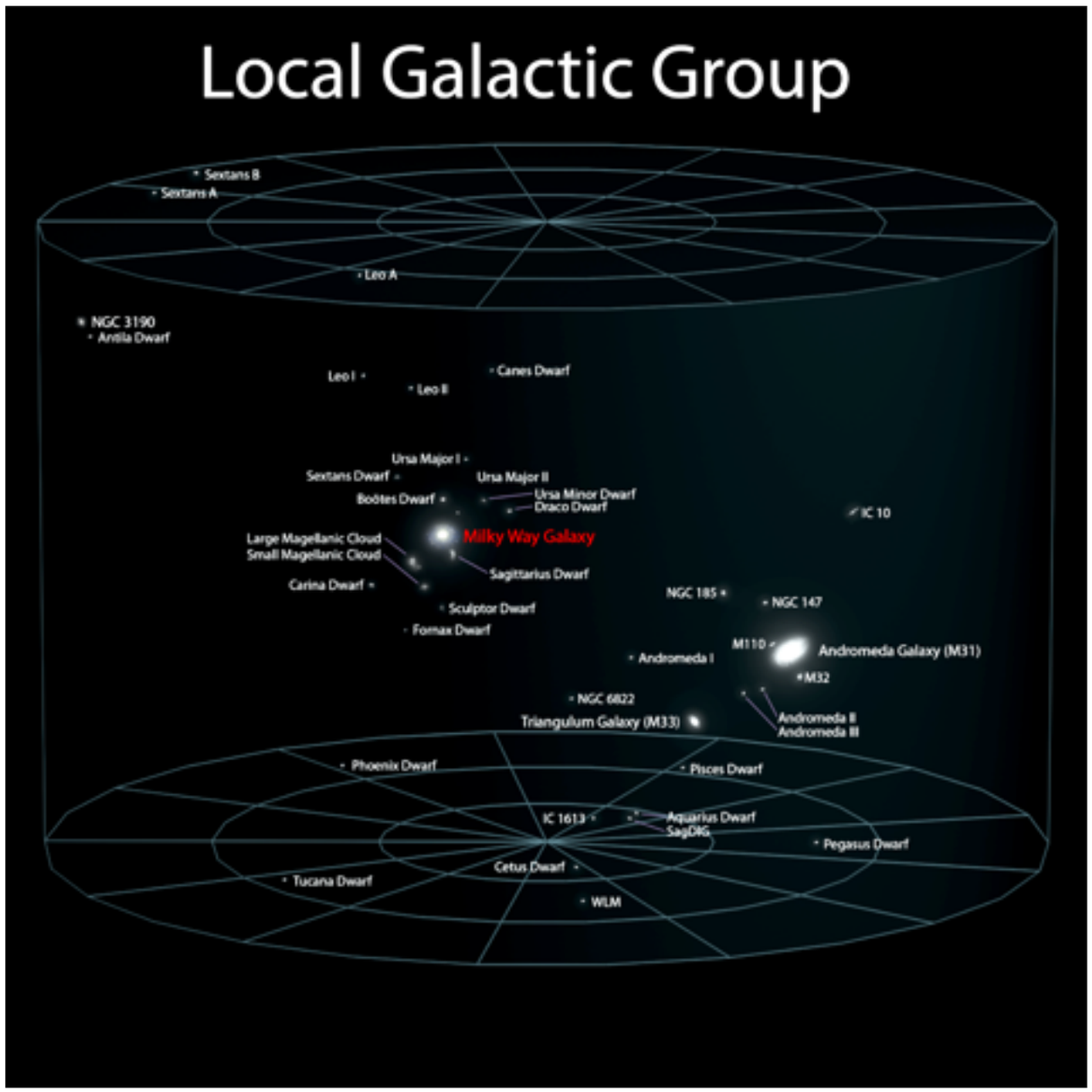
Astrostatistics: Tue 13 Feb 2017

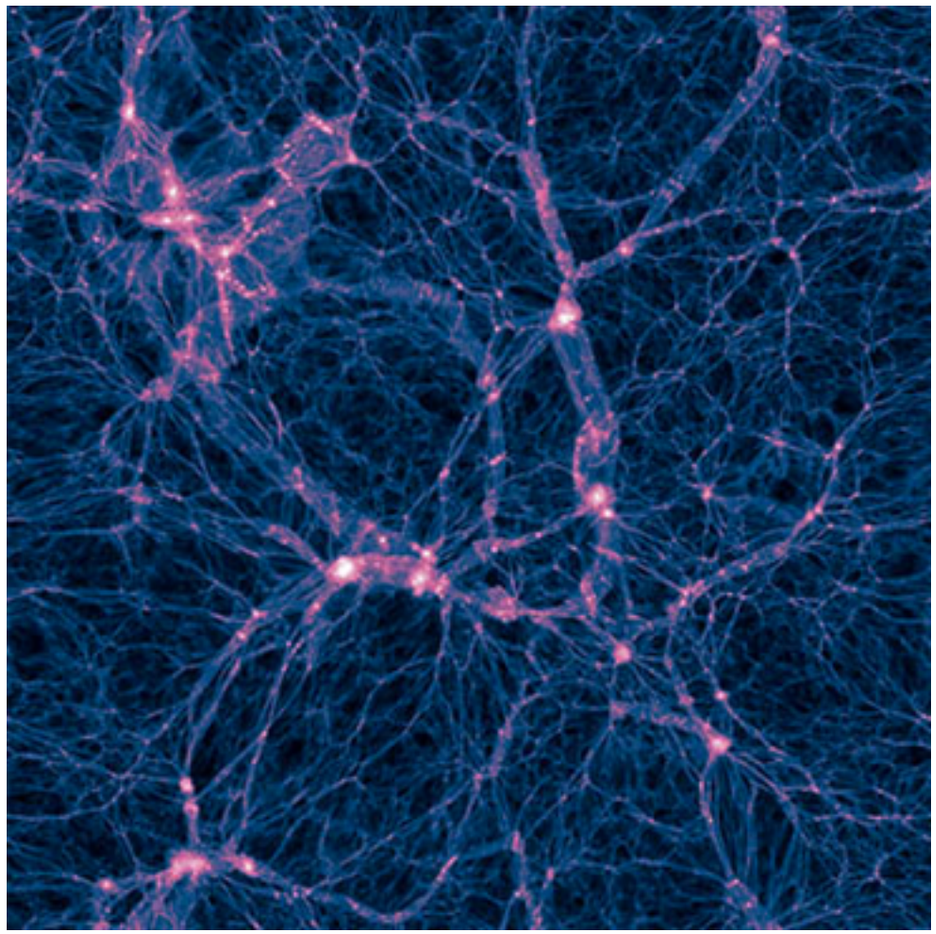
https://github.com/CambridgeAstroStat/PartIII-Astrostatistics

- Example Sheet 1 & datasets uploaded
  - Fri Feb 16 (2:30 pm, Room MR5)
- Fitting Statistical Models to Astronomical Data
- Bayesian Inference —> Sampling, Computation, examples
- References: Ivezic, Ch 5, F&B Ch 3, Gelman BDA
- Hogg, D., 2012. “Data analysis recipes: Probability calculus for inference.” https://arxiv.org/abs/1205.4446
- Hogg & DFM, 2017 “Data analysis recipes: Using Markov Chain Monte Carlo.” https://arxiv.org/abs/1710.06068

Next time: Astrostatistics Case Study 3:

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations (Patel et al. 2017, arXiv:1703.05767)

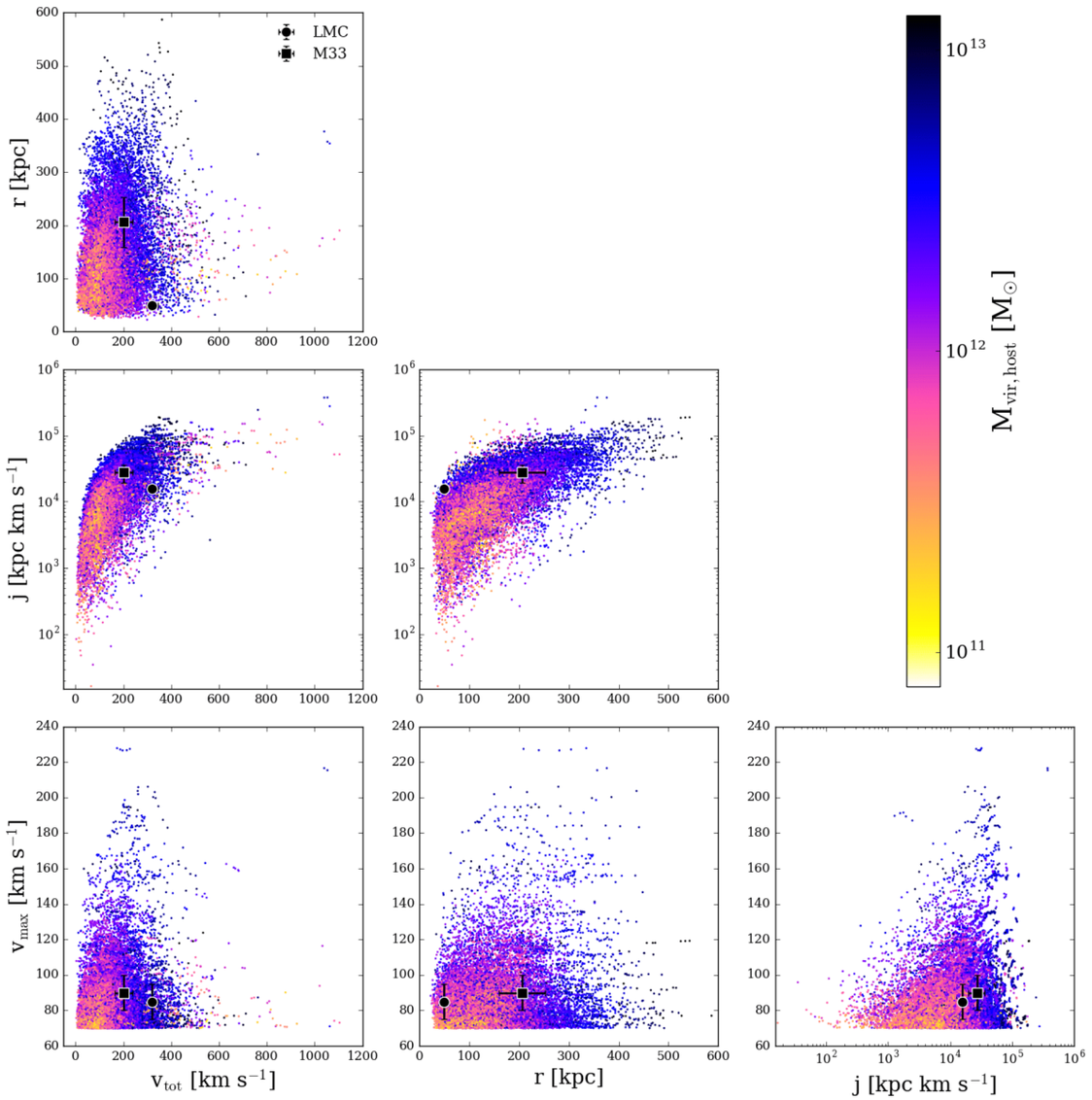




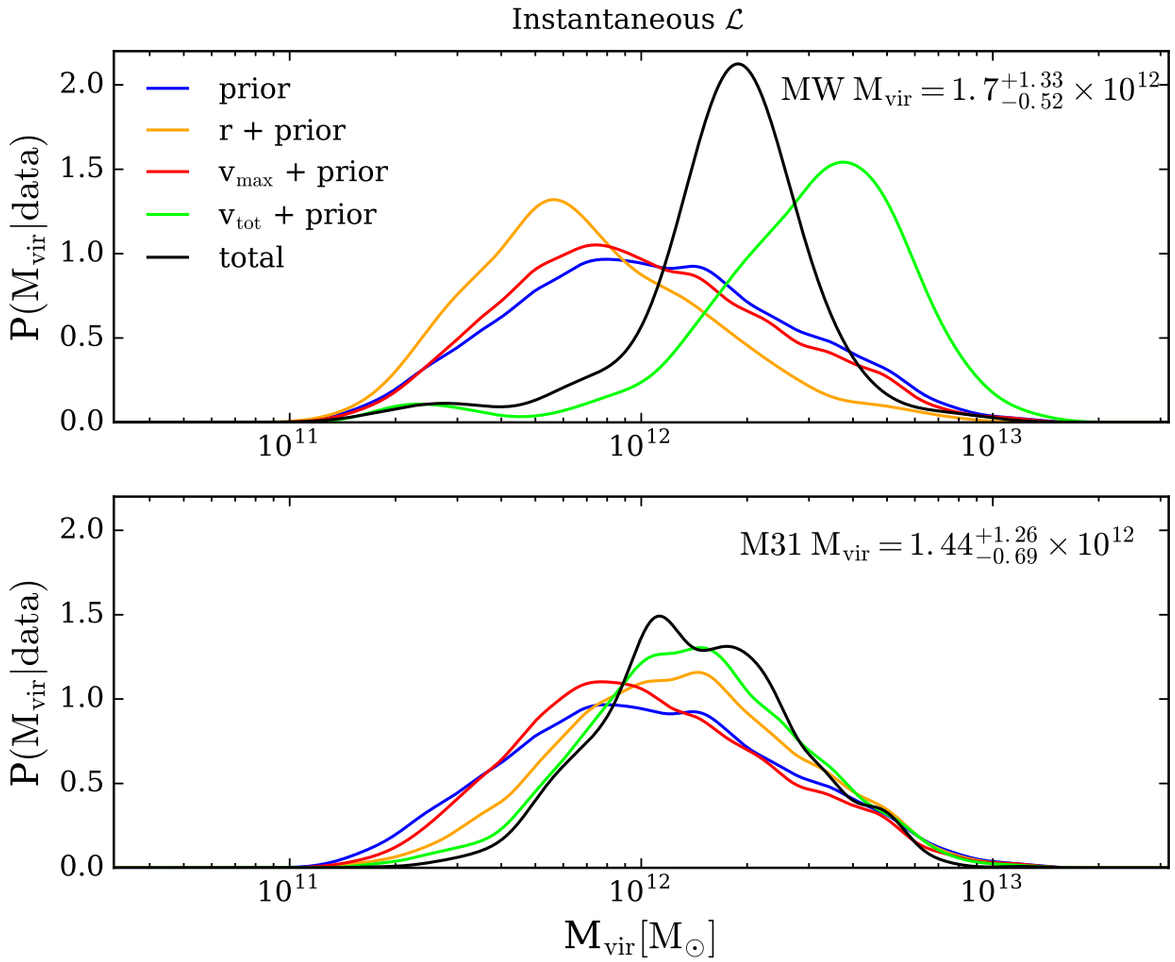
Illustris
Cosmological Simulation of Galaxy Formation

Next time: Astrostatistics Case Study 3:

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations (Patel et al. 2017, arXiv:1703.05767)

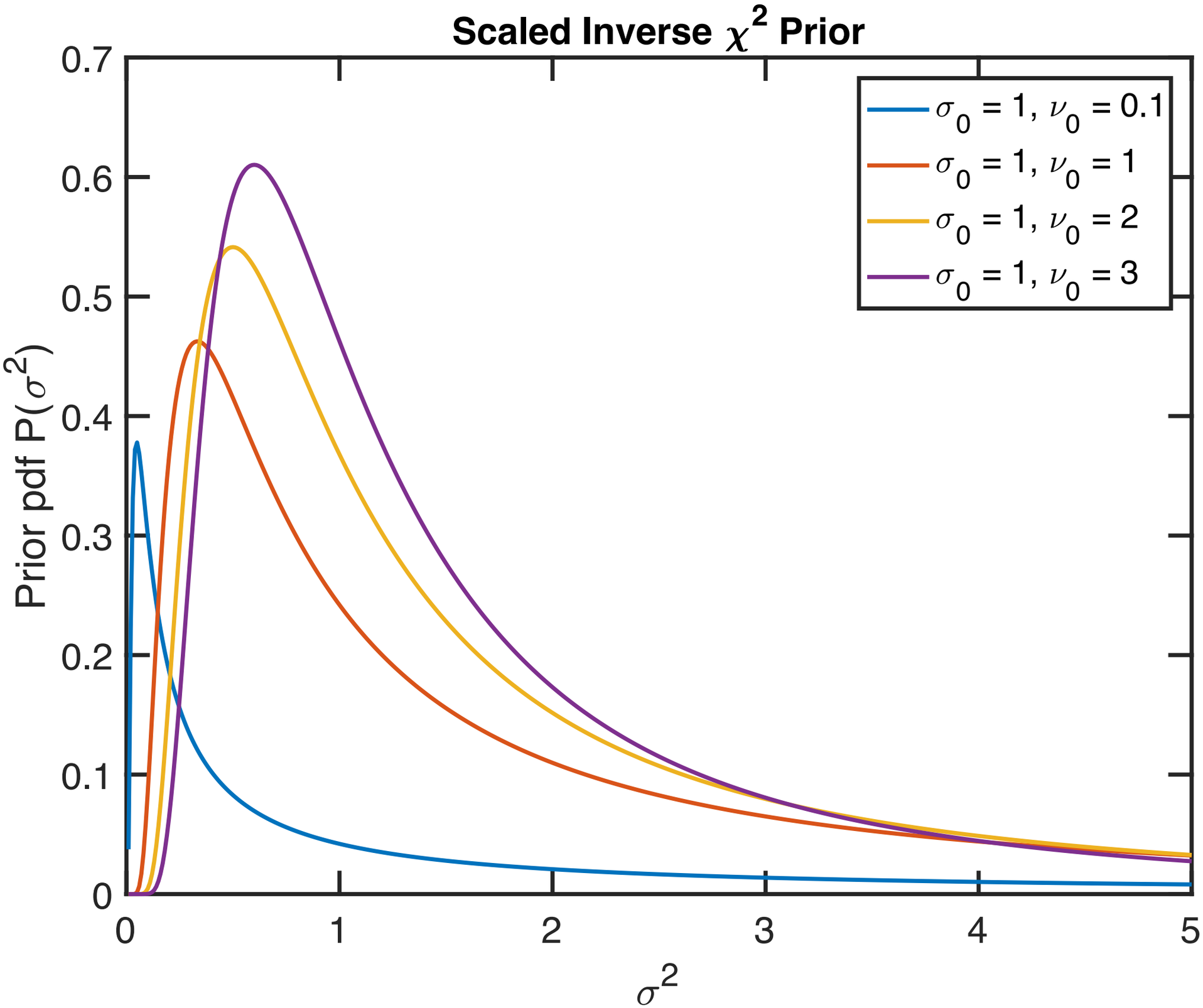


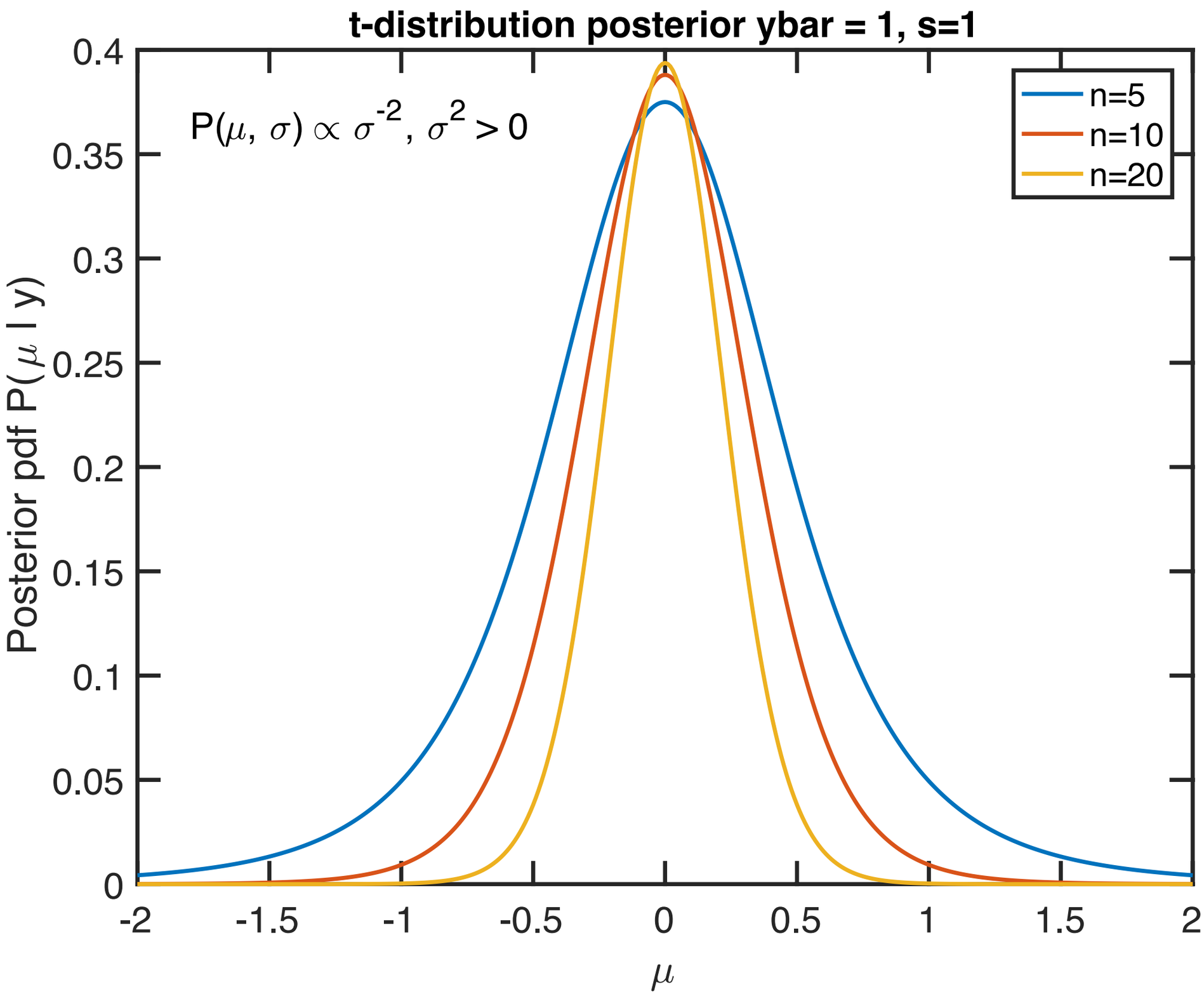
Simulation —> Prior

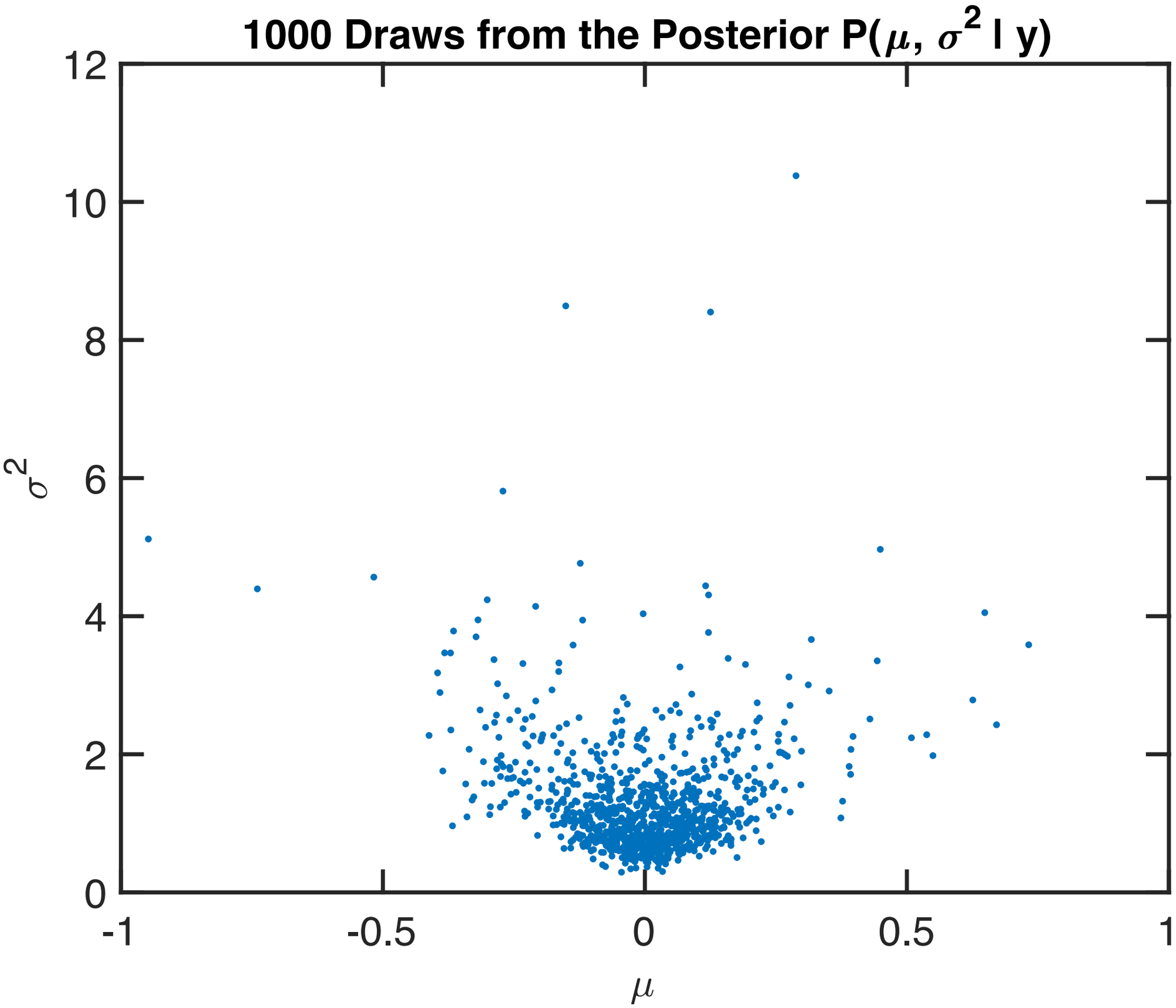


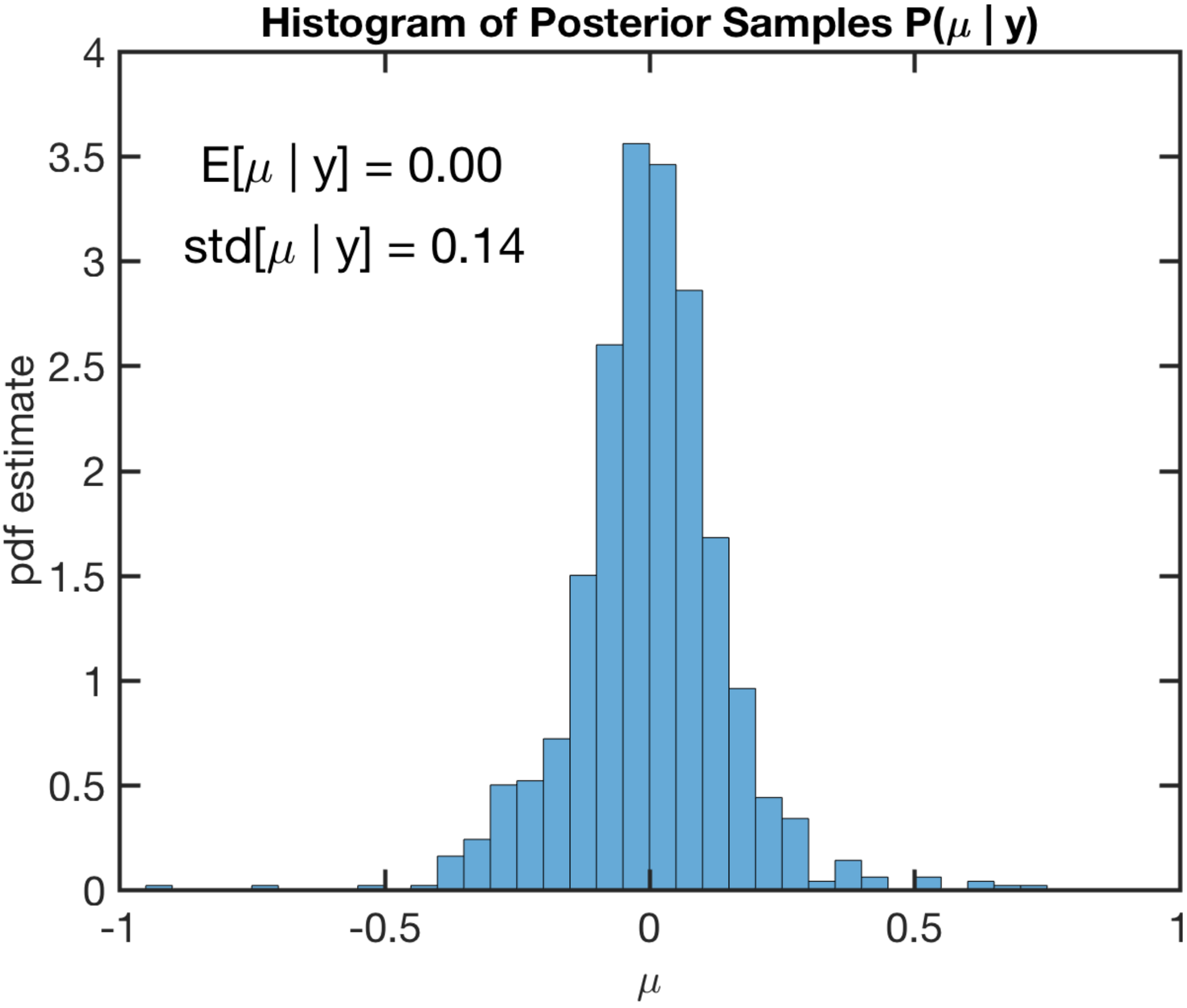
- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

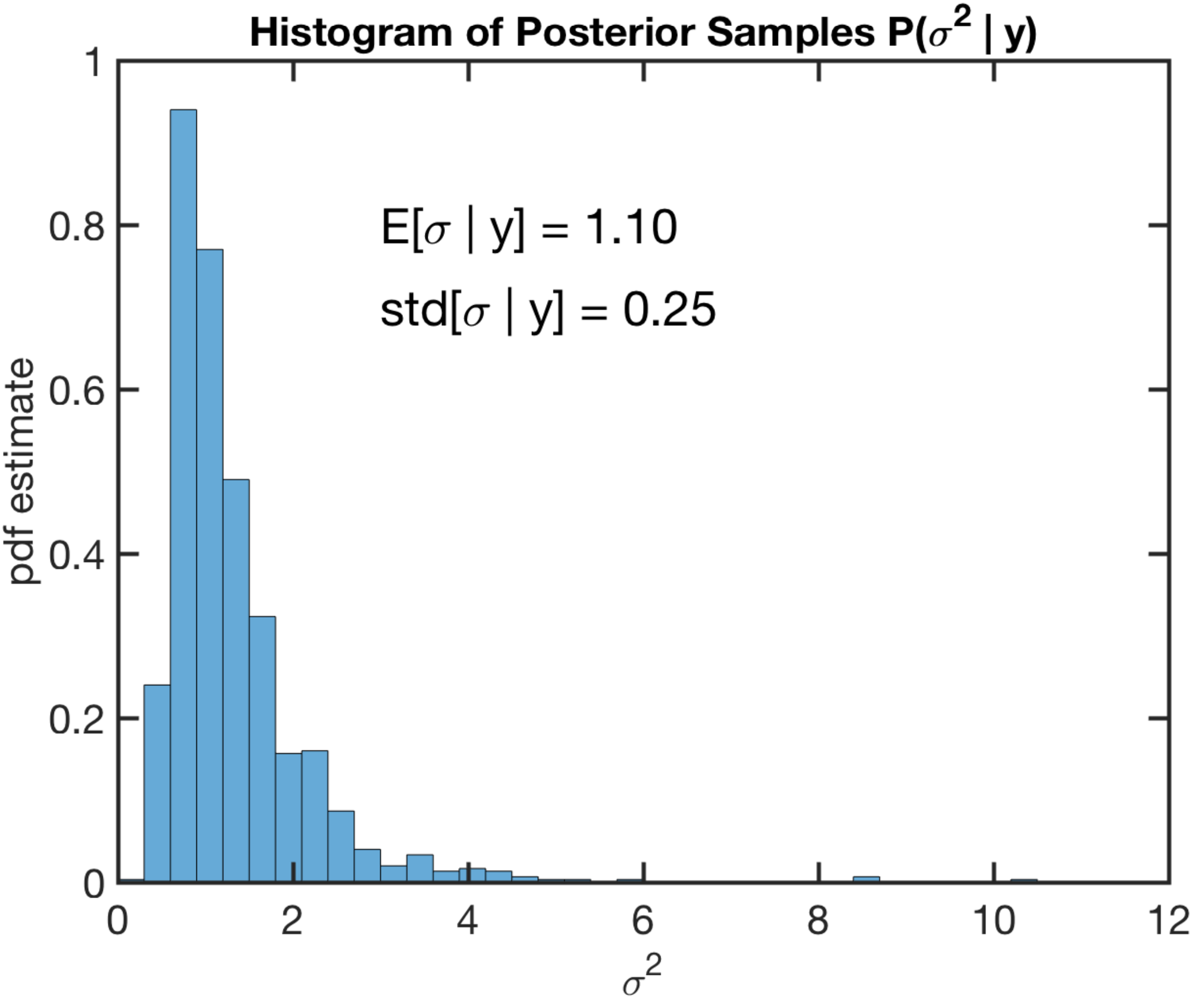
Multi-parameter Bayesian inference: Gaussian example: Gelman BDA Sec 3.2 - 3.3


Scaled Inverse χ^2 Prior
$\sigma_0 = 1, \nu_0 = 0.1$
$\sigma_0 = 1, \nu_0 = 1$
$\sigma_0 = 1, \nu_0 = 2$
$\sigma_0 = 1, \nu_0 = 3$
Prior pdf $P(\sigma^2)$
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
σ^2

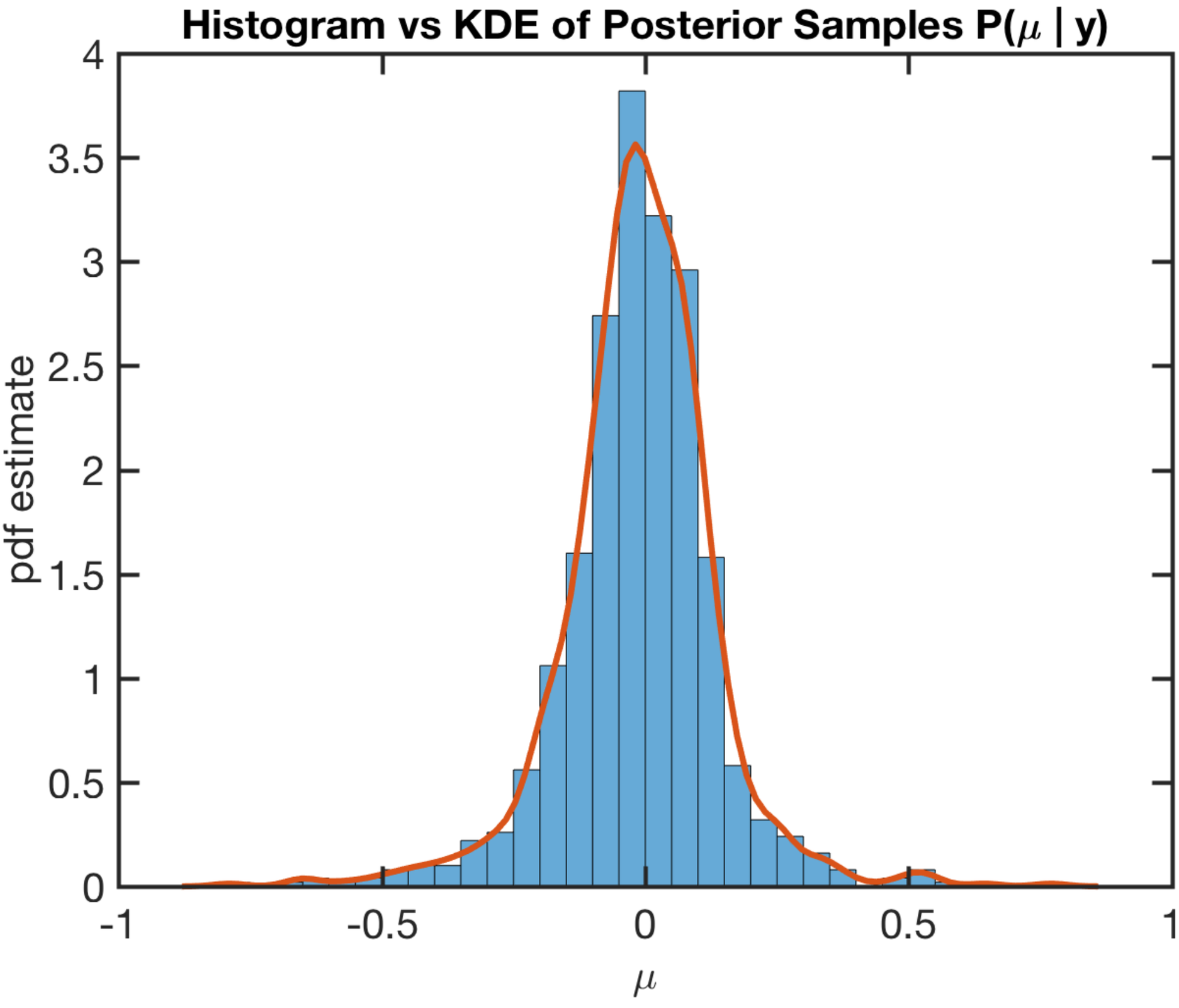

t-distribution posterior ybar = 1, s=1
$P(\mu, \sigma) \propto \sigma^{-2}, \sigma^2 > 0$
n=5
n=10
n=20
Posterior pdf P(μ | y)
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
μ


1000 Draws from the Posterior P(μ, σ^2 | y)
12
10
8
6
4
2
0
σ^2
-1
-0.5
0
0.5
1
μ


Histogram of Posterior Samples P(μ | y)
E[μ | y] = 0.00
std[μ | y] = 0.14
pdf estimate
4
3.5
3
2.5
2
1.5
1
0.5
0
-1
-0.5
0
0.5
1
μ


Histogram of Posterior Samples P(σ^2 | y)
E[σ | y] = 1.10
std[σ | y] = 0.25
pdf estimate
0
0.2
0.4
0.6
0.8
1
0
2
4
6
8
10
12
σ^2

Kernel Density Estimate =
estimate a smooth density from samples



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

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

- Posterior simulation - Markov Chain Monte Carlo, Nested Sampling, etc. generates draws
- Importance Sampling - draw from an easier ("tractable") distribution $\theta_i \sim Q(\theta)$ and weight the samples by $w_i = P(\theta_i | D) / Q(\theta_i)$

$$\int f(\boldsymbol{\theta})P(\boldsymbol{\theta}|D)d\boldsymbol{\theta} = \int f(\boldsymbol{\theta})\frac{P(\boldsymbol{\theta}|D)}{Q(\boldsymbol{\theta})}Q(\boldsymbol{\theta})d\boldsymbol{\theta} \approx \sum_{i=1}^{K} f(\boldsymbol{\theta}_i)w_i$$