

Astrostatistics: Thu 15 Feb 2017

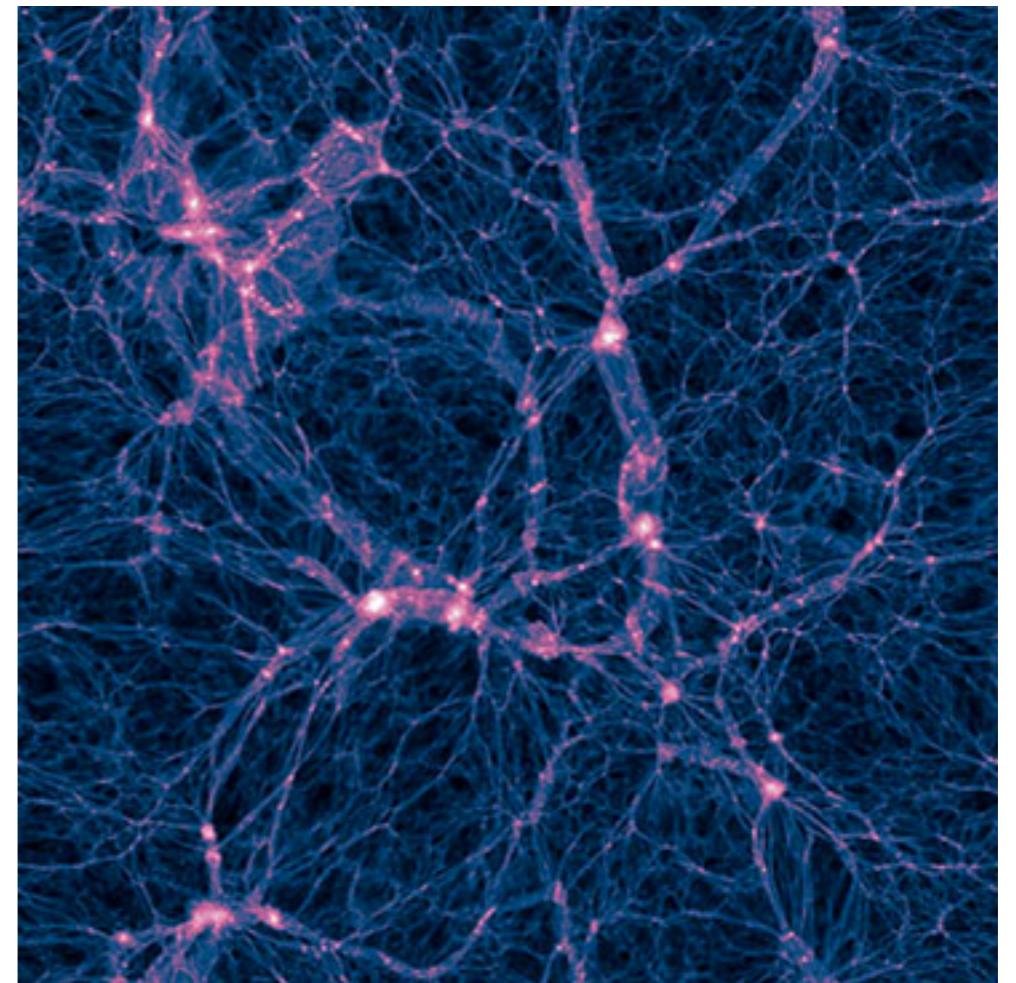
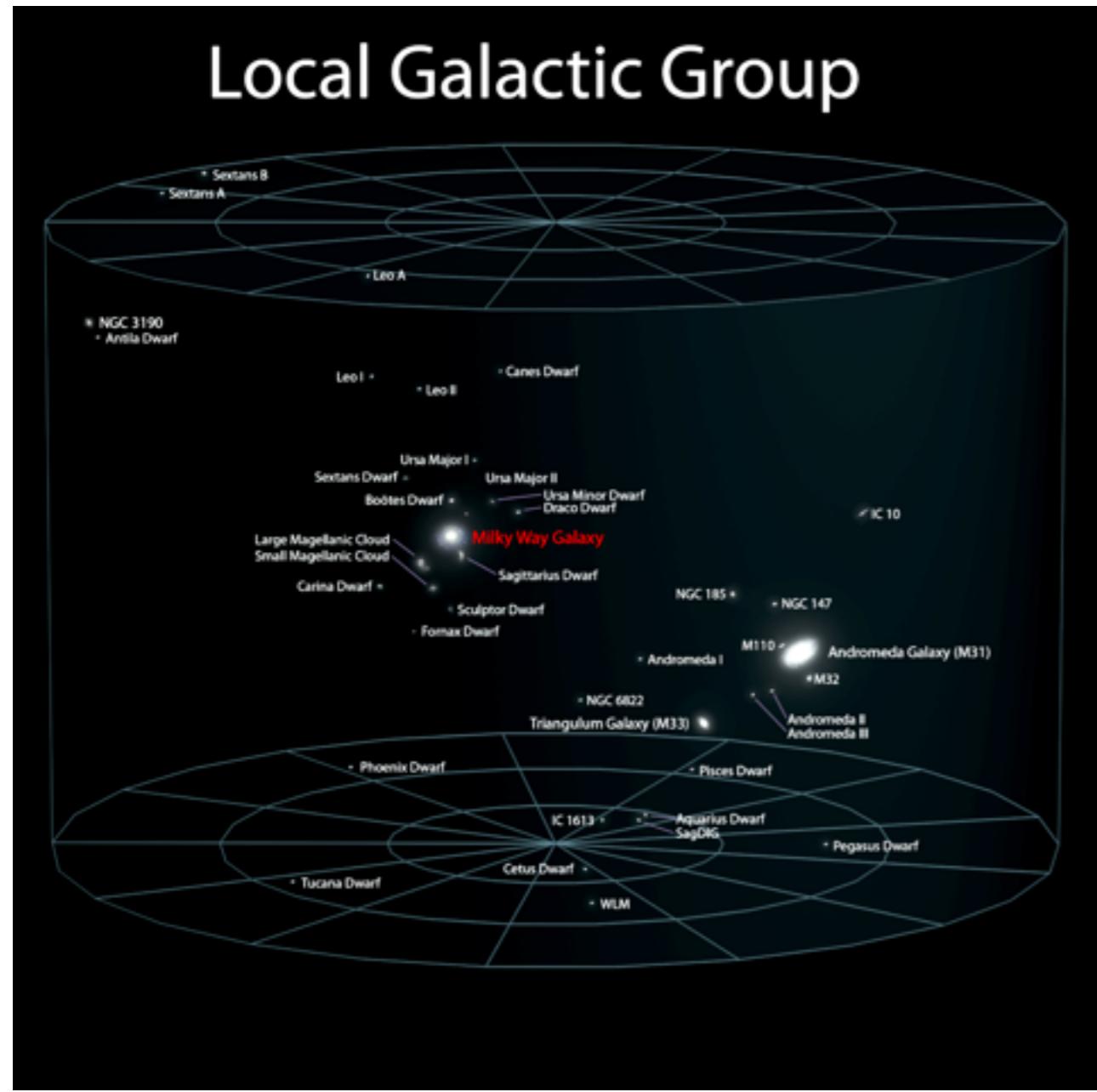
<https://github.com/CambridgeAstroStat/PartIII-Astrostatistics>

- Example Sheet 1 & datasets uploaded
 - This Fri Feb 16 (2:30 pm, Room MR5)
- Fitting Statistical Models to Astronomical Data
 - Bayesian Computation, Examples, Case Studies
 - Refs: Ivezic, Ch 5, F&B Ch 3 (not much), Gelman BDA
 - Givens & Hoeting “Computational Statistics”
 - Roberts & Casella “Monte Carlo Statistical Methods” (theory)
 - Hogg & DFM, 2017 “Data analysis recipes: Using Markov Chain Monte Carlo.” <https://arxiv.org/abs/1710.06068>

Later today: 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)

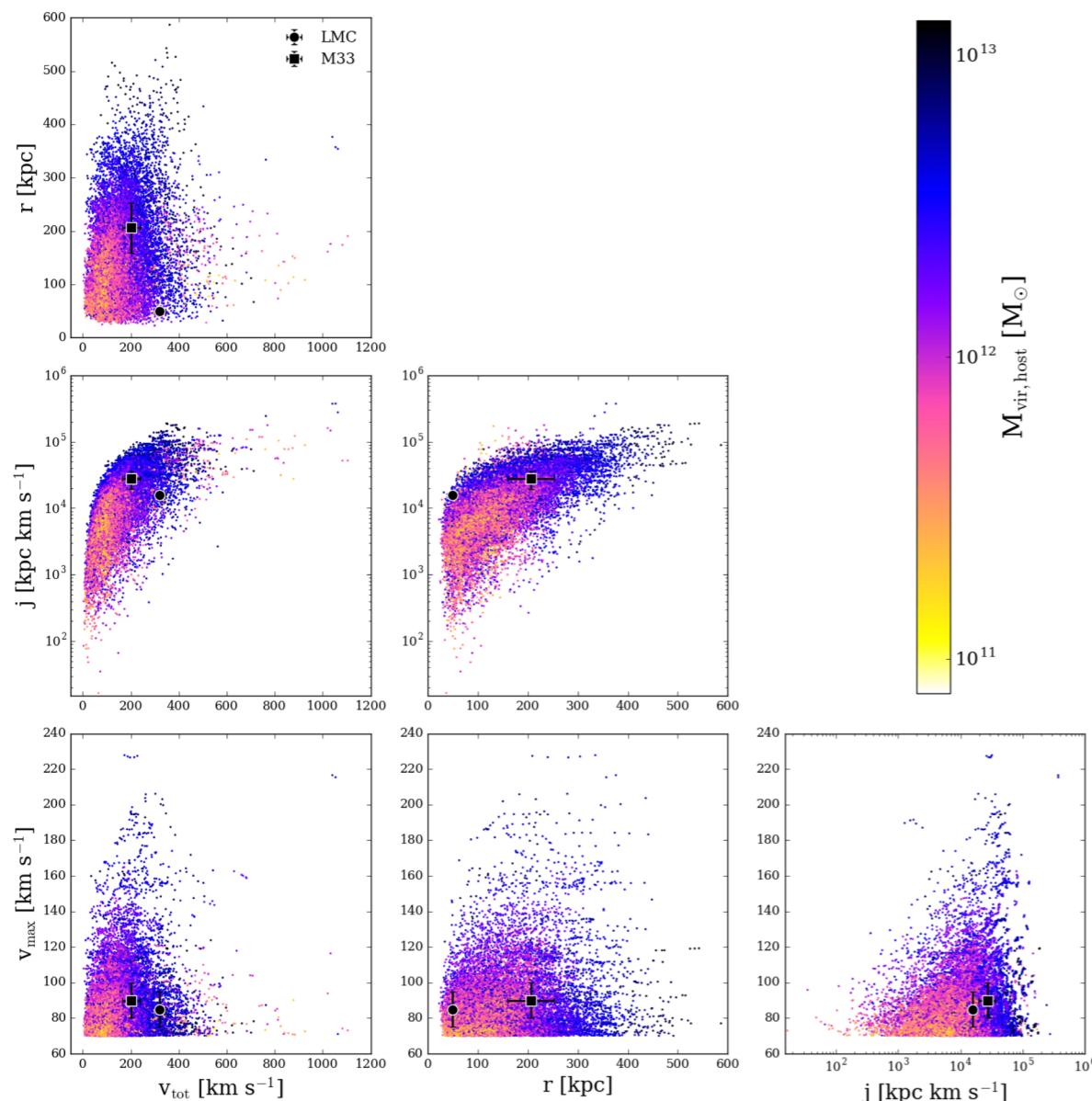


Illustris
Cosmological Simulation of
Galaxy Formation

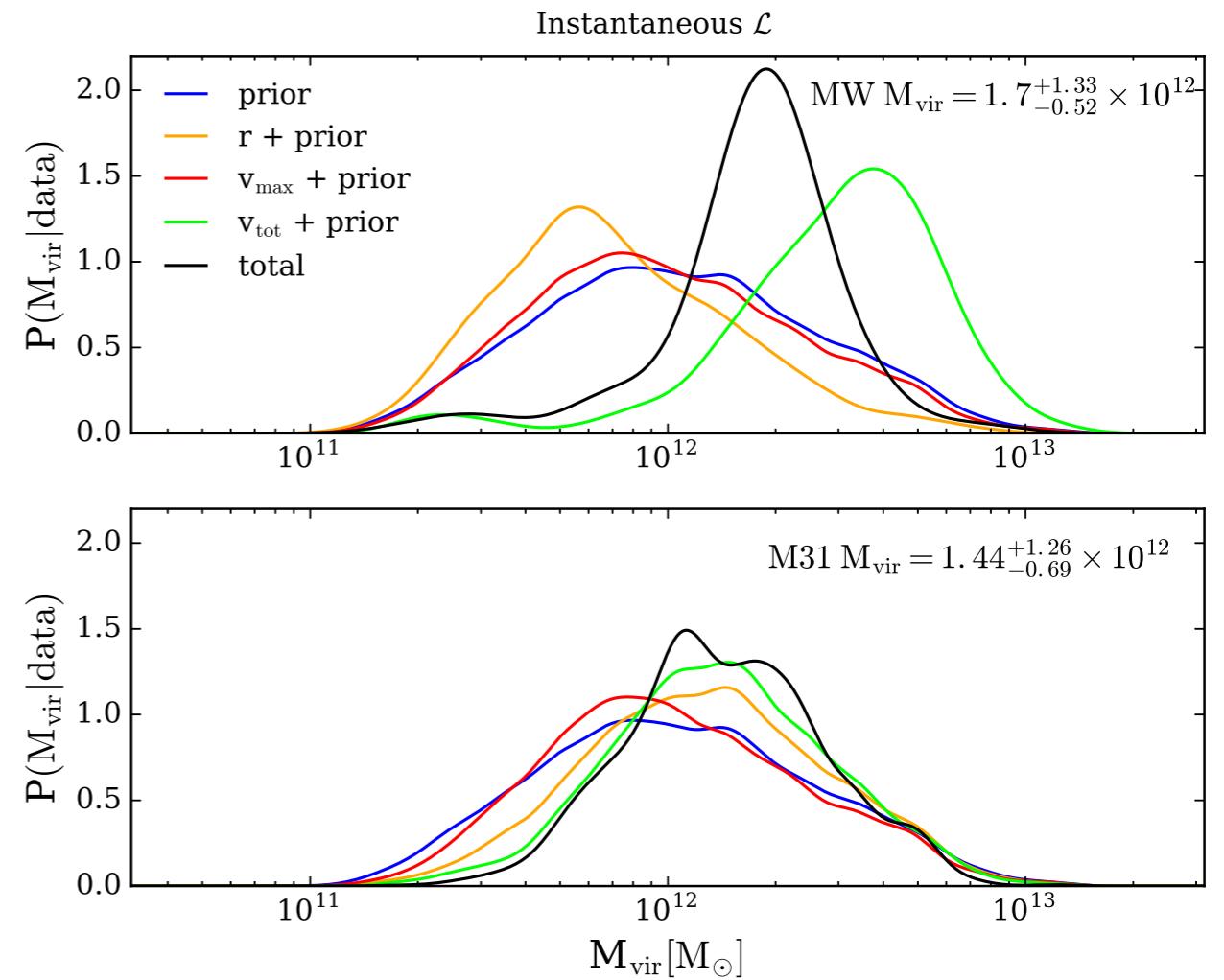
Later today: Astrostatistics Case Study 3:

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations

(Patel et al. 2017, arXiv:1703.05767)

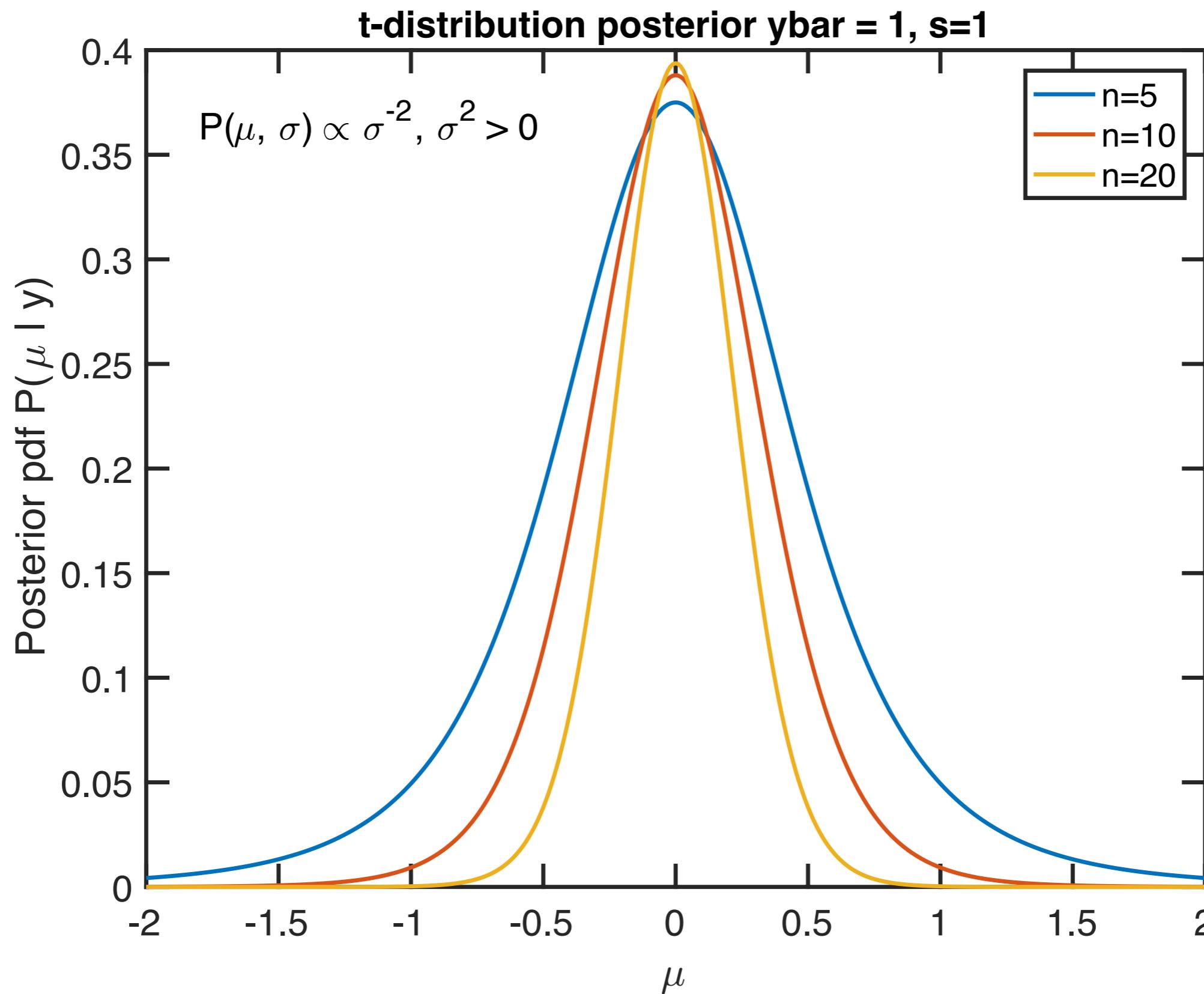


Simulation \rightarrow Prior

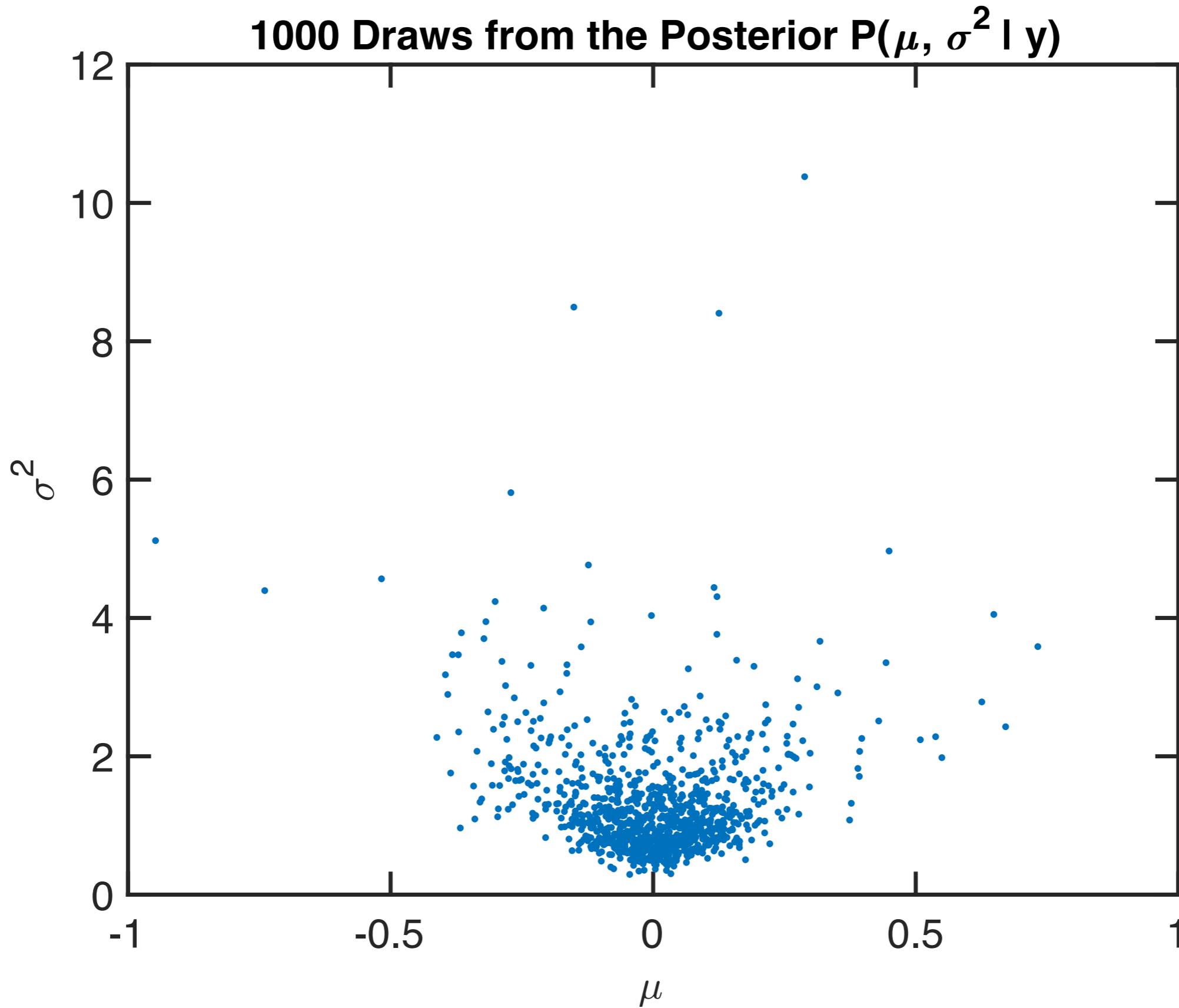


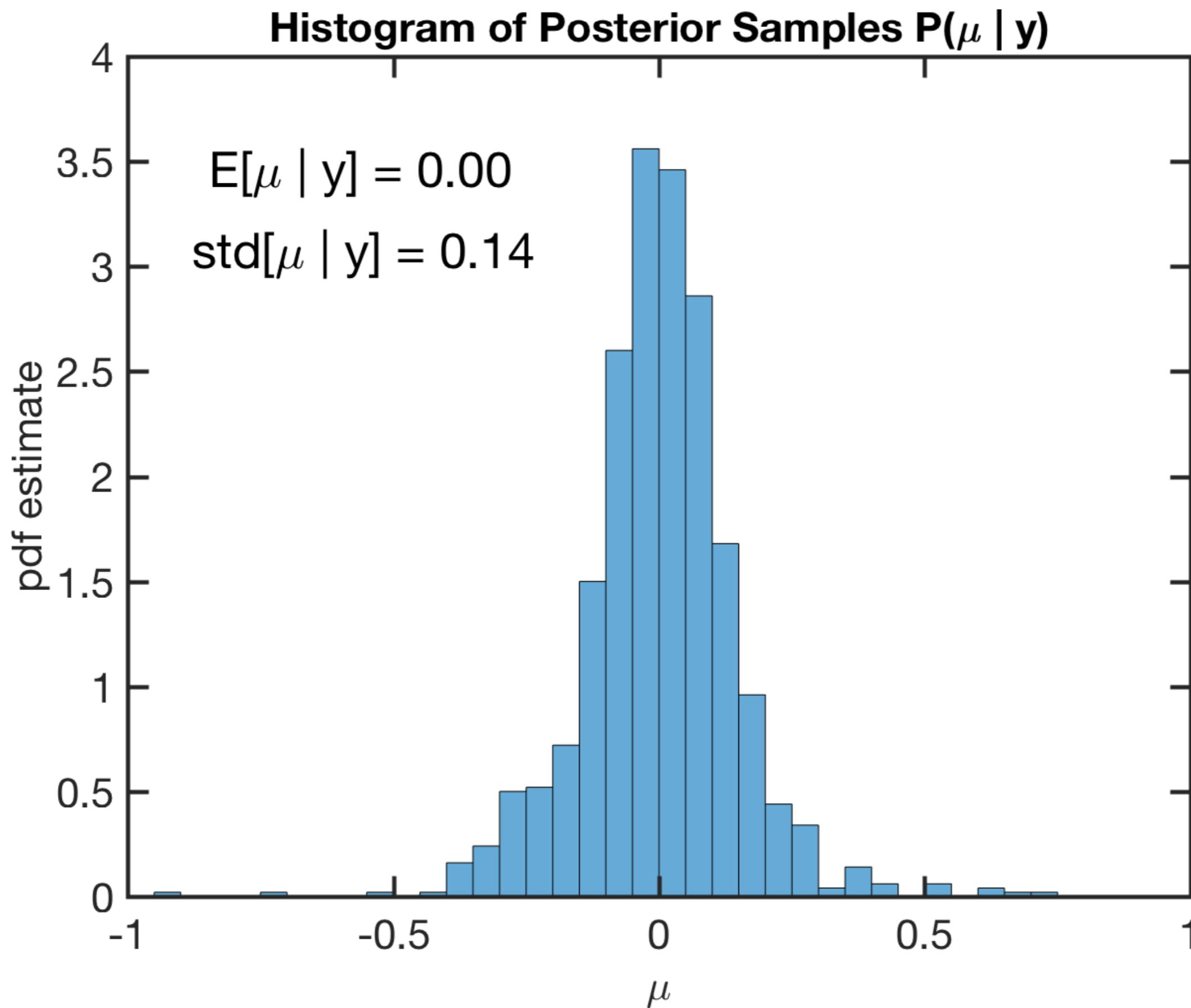
- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

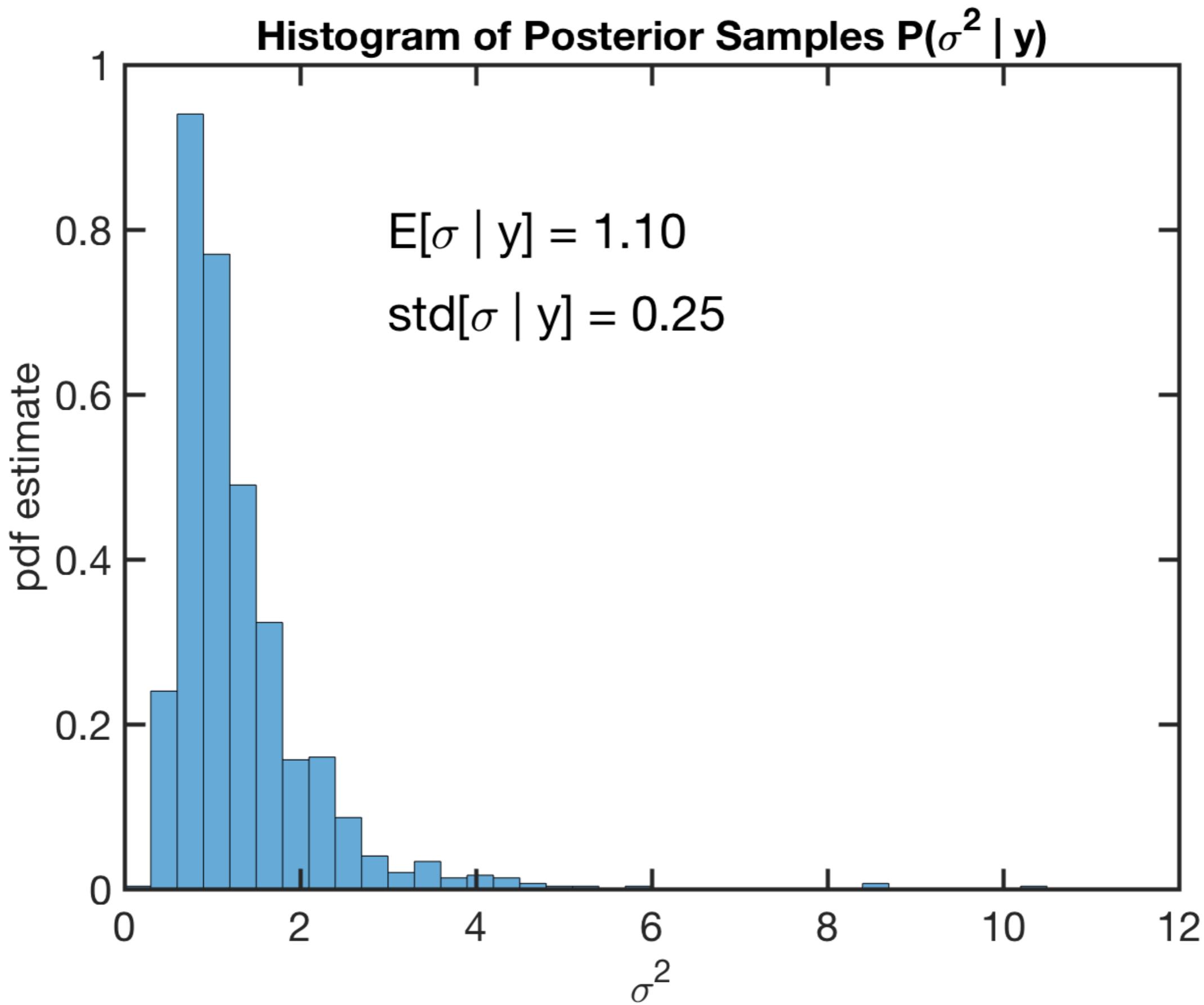
Last Time: Posterior Distribution of a Gaussian Mean Analytic Result



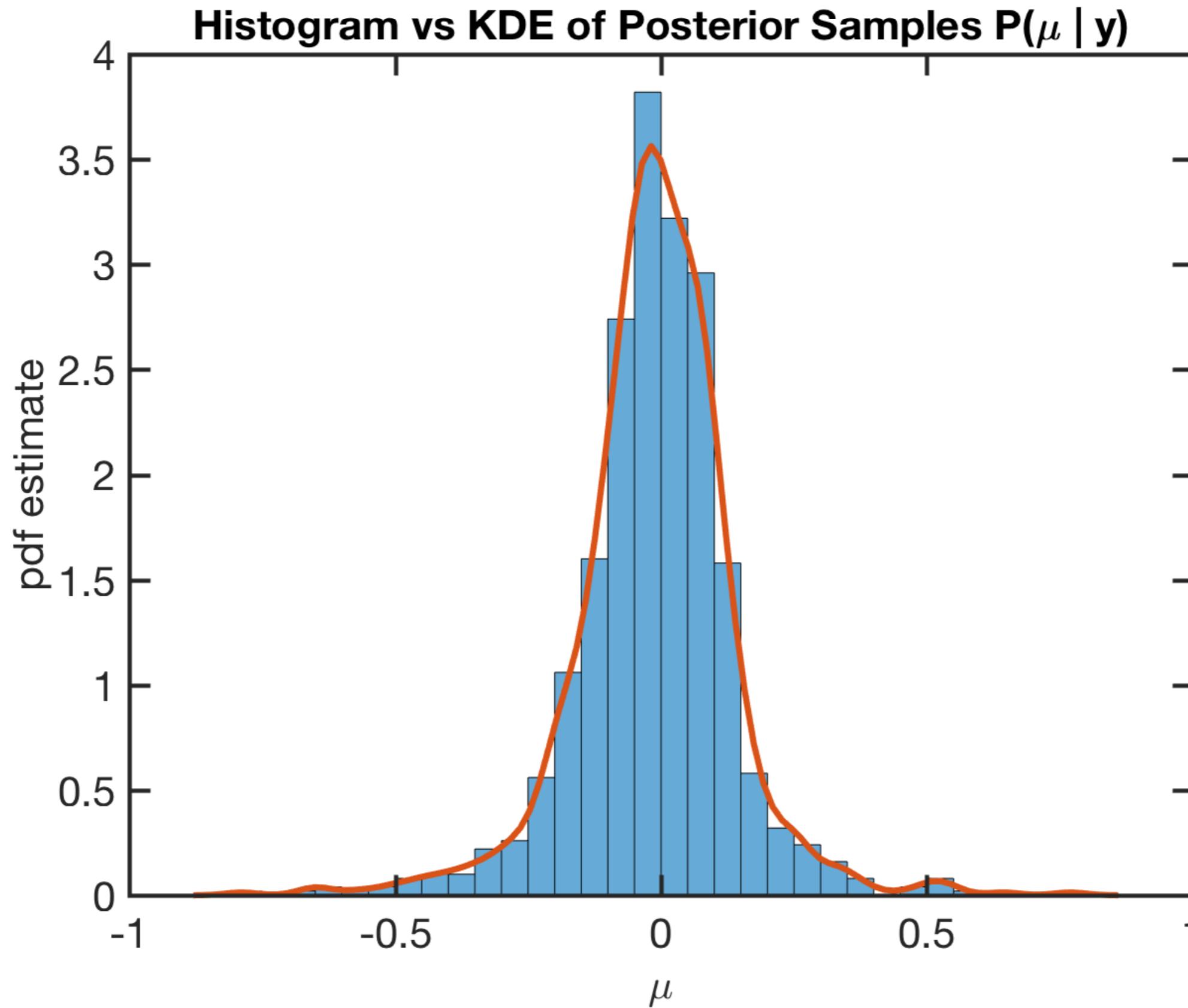
Monte Carlo Sampling





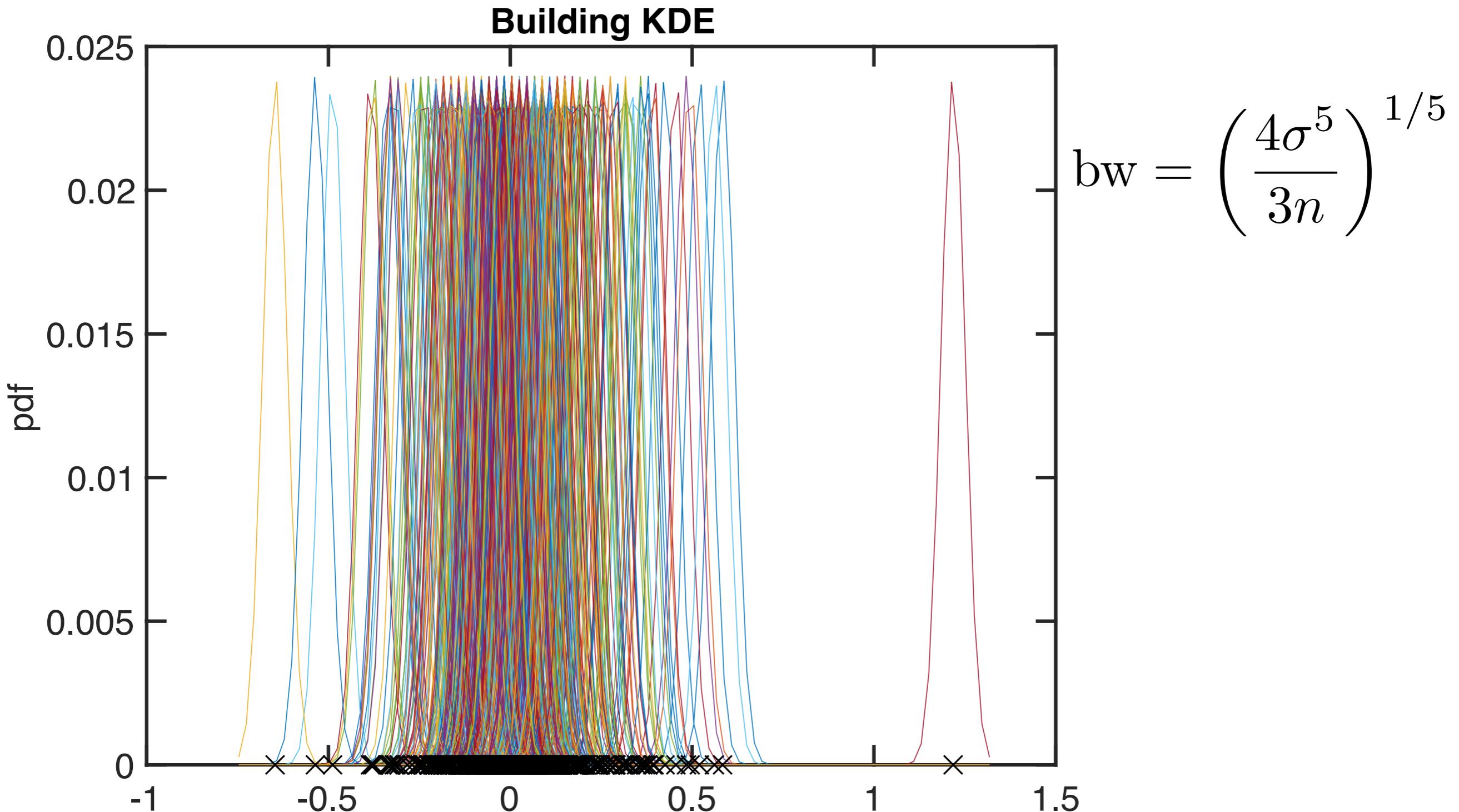


Kernel Density Estimate =
estimate a smooth density from samples



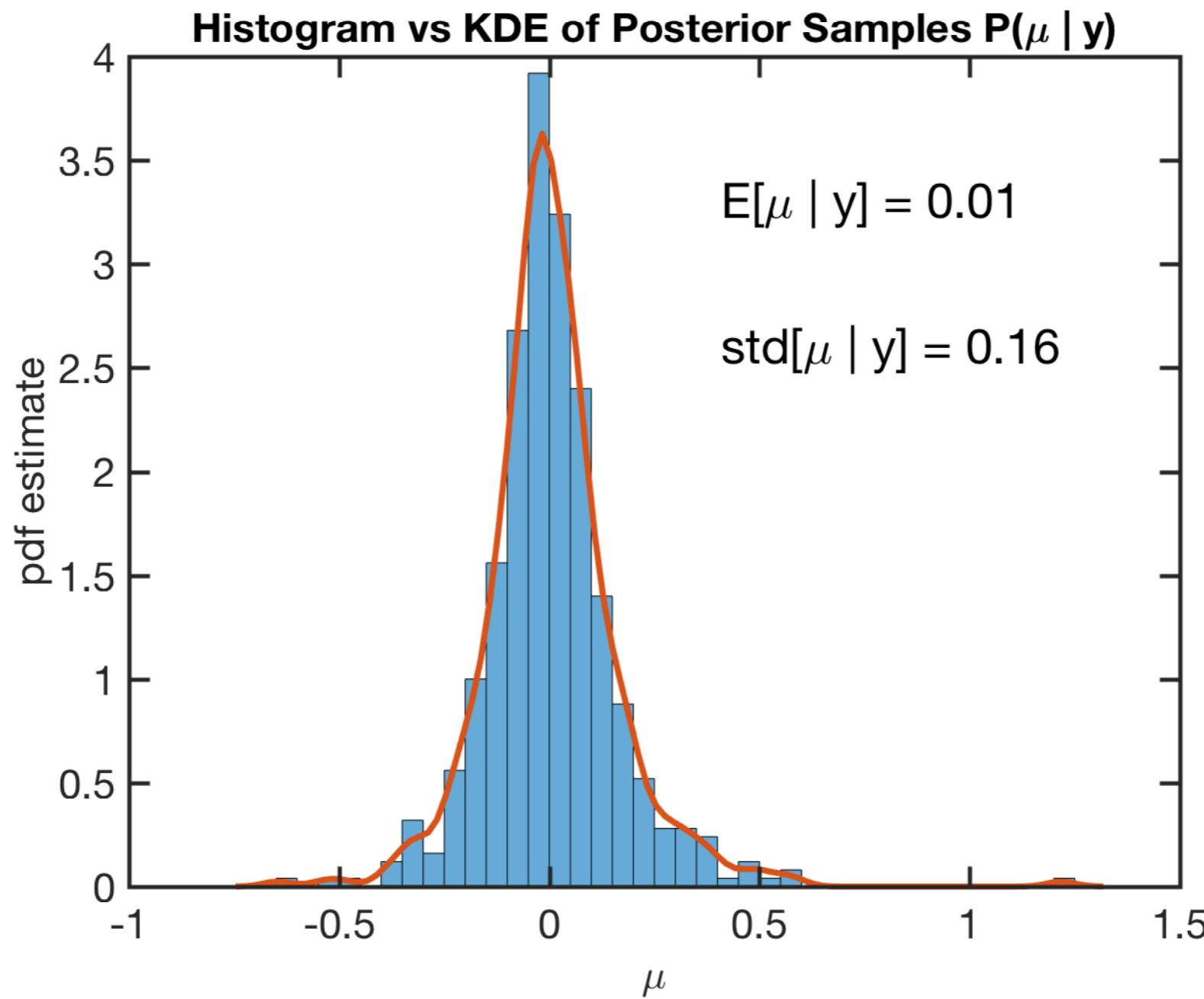
Kernel Density Estimation (KDE) (Smooth Histogram)

Each sample gets a Gaussian at the sample point
with an “optimal” bandwidth bw (rule of thumb)



Kernel Density Estimation (KDE) (Smooth Histogram)

Then add them up and normalise pdf to 1



Monte Carlo Integration

Typically, we want to compute expectations of the form:

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

Using m samples from the posterior:

$$\boldsymbol{\theta}_i \sim P(\boldsymbol{\theta} | D)$$

How many posterior samples $i = 1 \dots m$
do you need to approximate $\mathbb{E}[f(\boldsymbol{\theta} | D)]$
to some error tolerance?

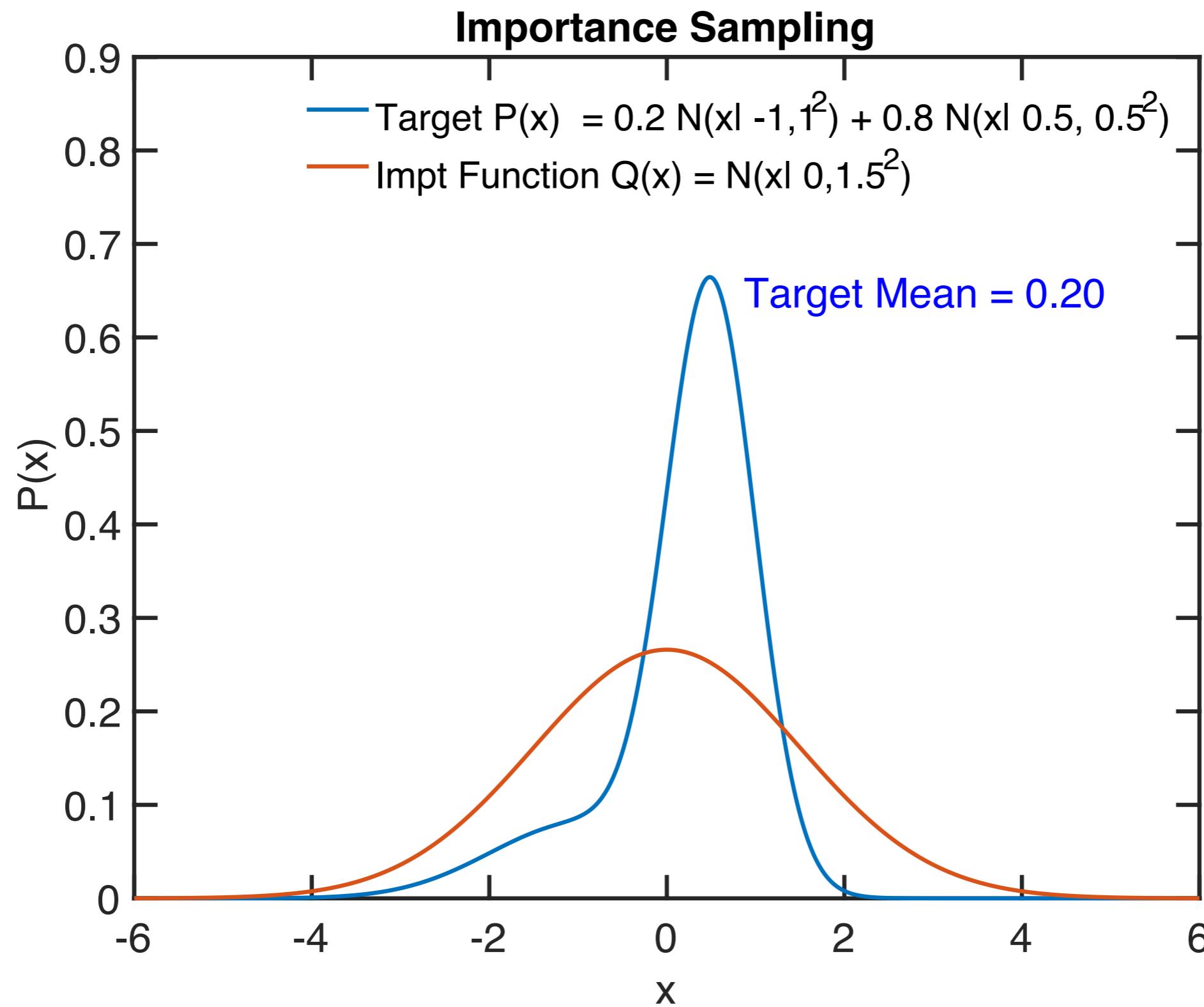
What if you can't directly sample the posterior: $\theta_i \sim P(\theta | D)$?

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

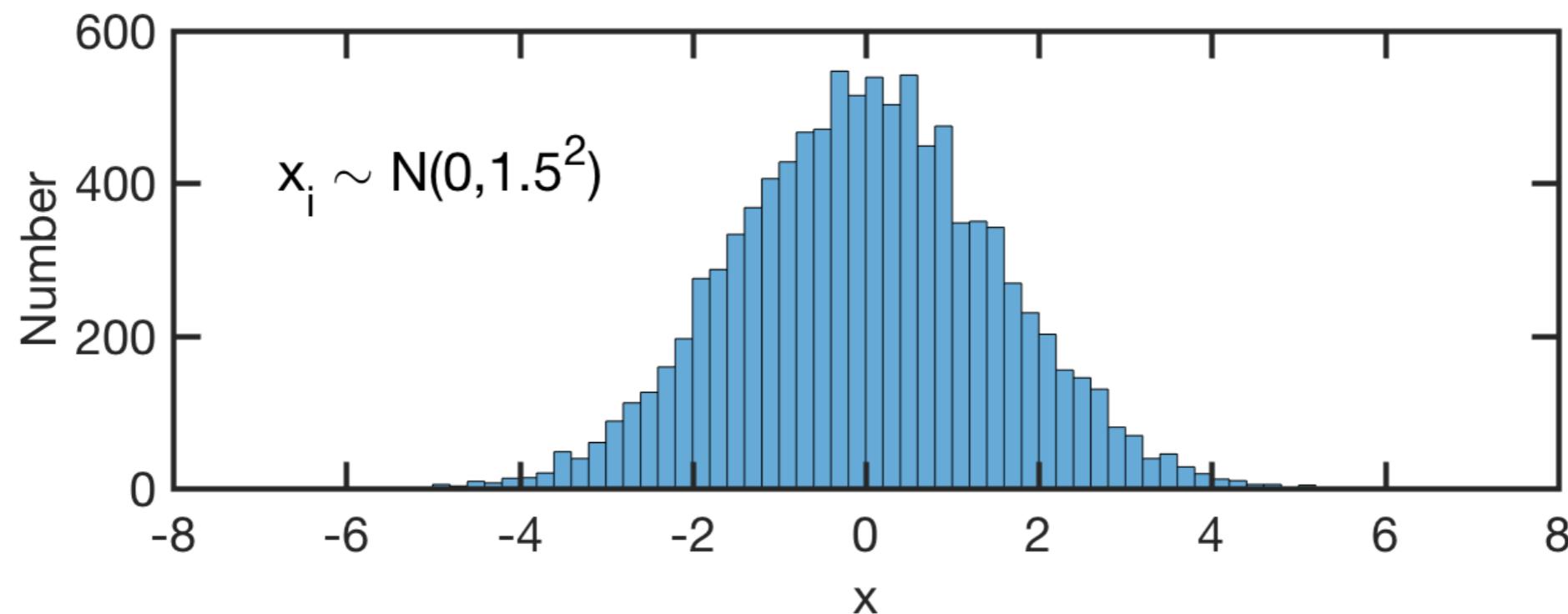
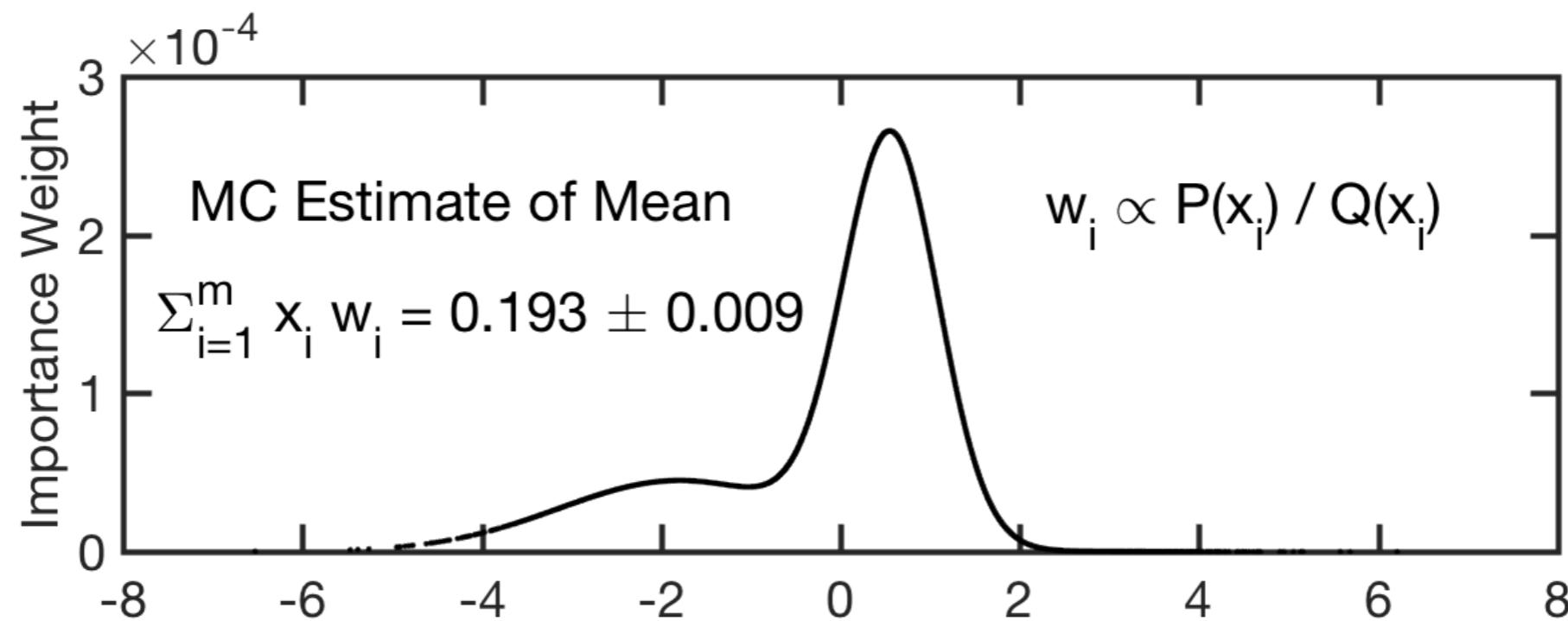
- Posterior simulation - Markov Chain Monte Carlo, Nested Sampling, etc. generates draws
- Importance Sampling - draw from an easier (“tractable”) distribution $\theta_i \sim Q(\theta)$ and weight the samples by $w_i = P(\theta_i | D) / Q(\theta_i)$

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

Importance Sampling Example



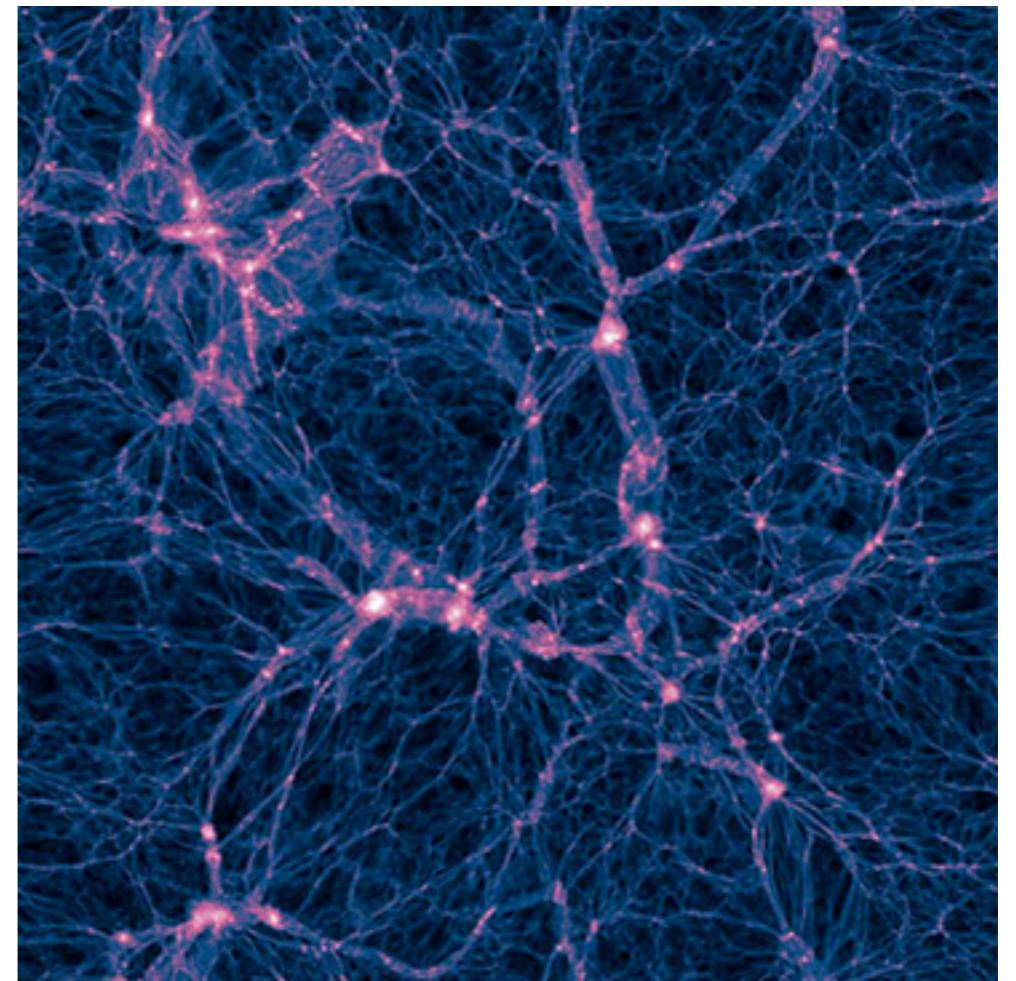
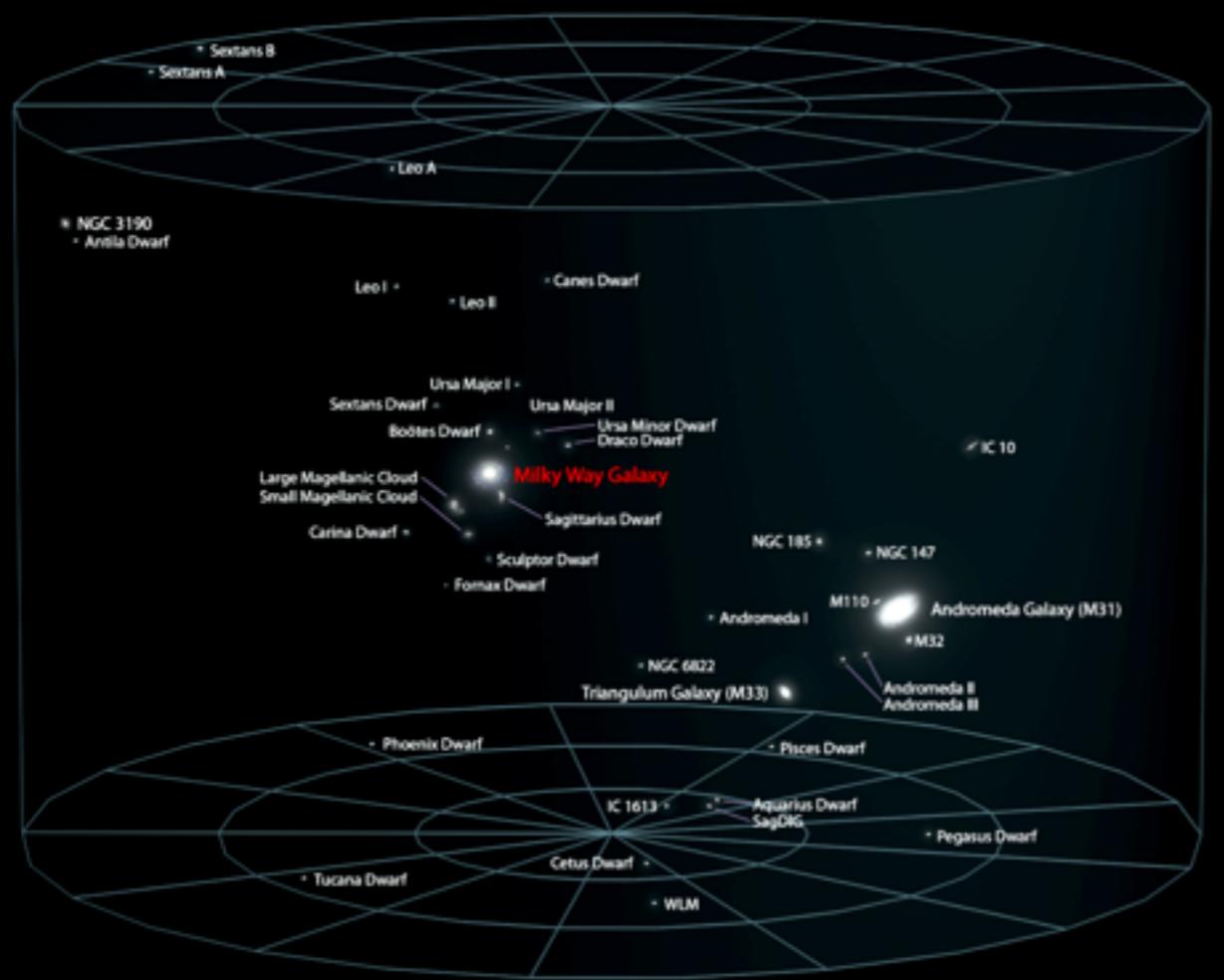
Importance Sampling Example



Now: Astrostatistics Case Study 3:

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations
(Patel, Besla & Mandel, MNRAS 468, 3428, arXiv:1703.05767)

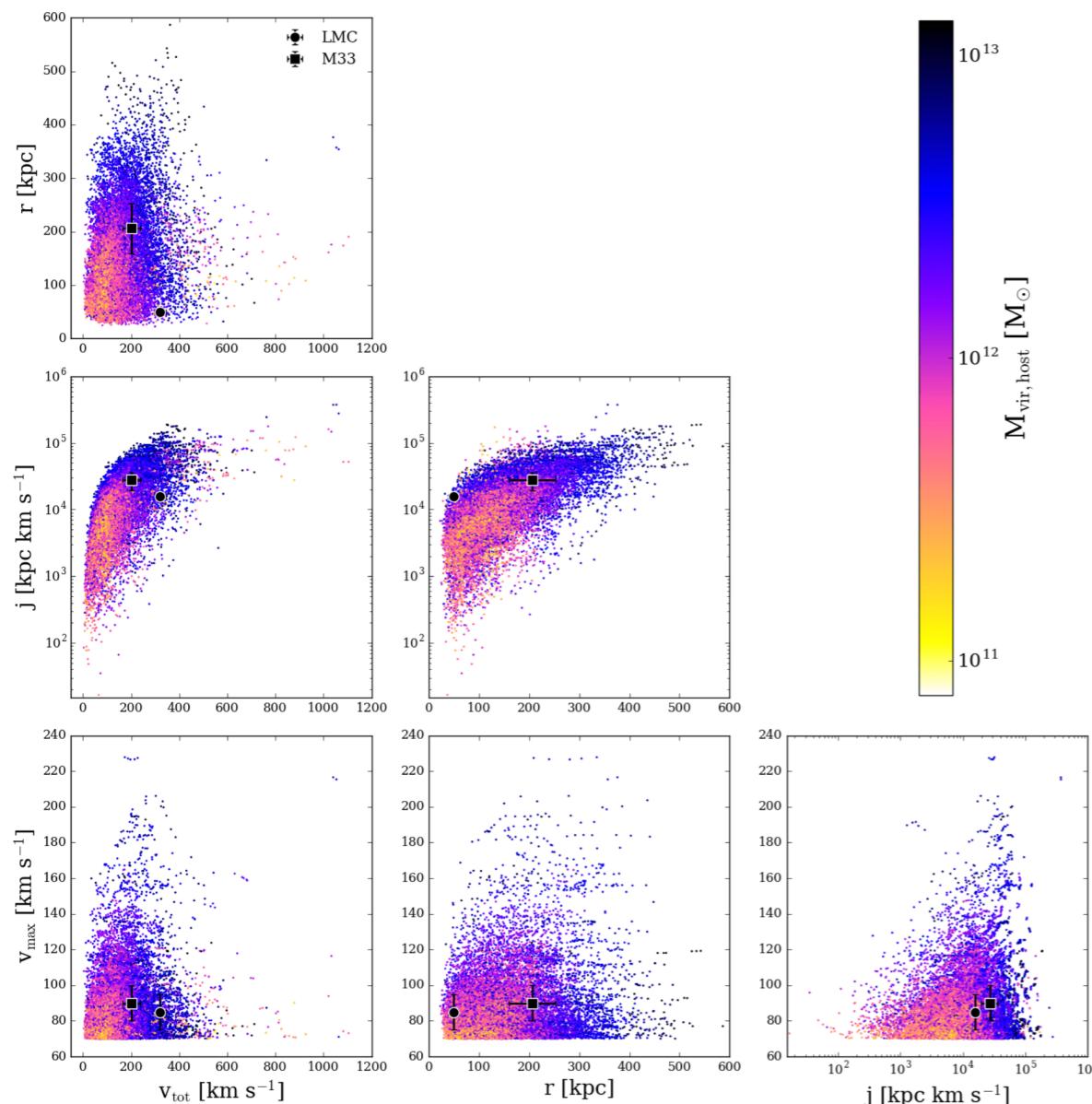
Local Galactic Group



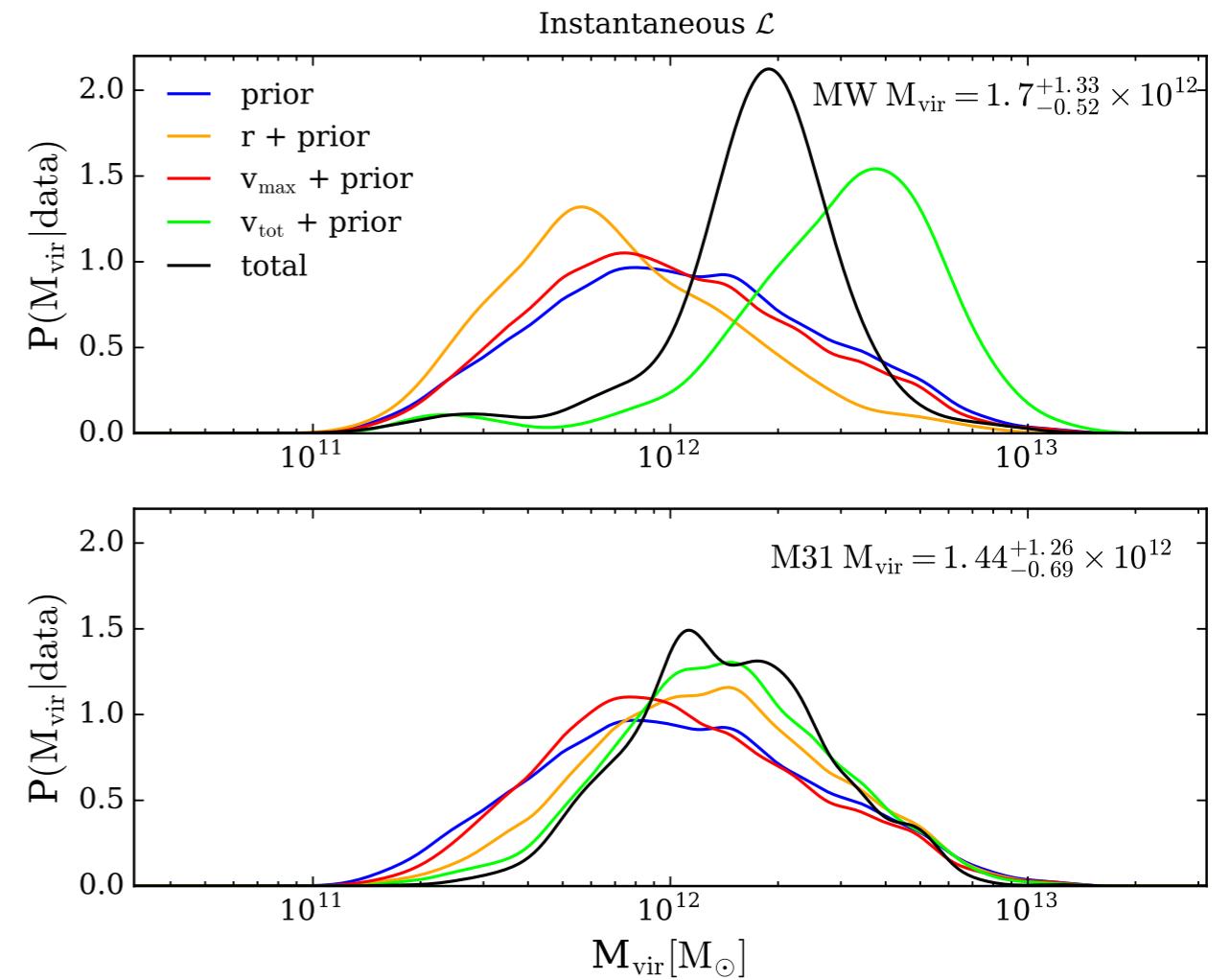
Illustris
Cosmological Simulation of
Galaxy Formation

Now: 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 \rightarrow Prior

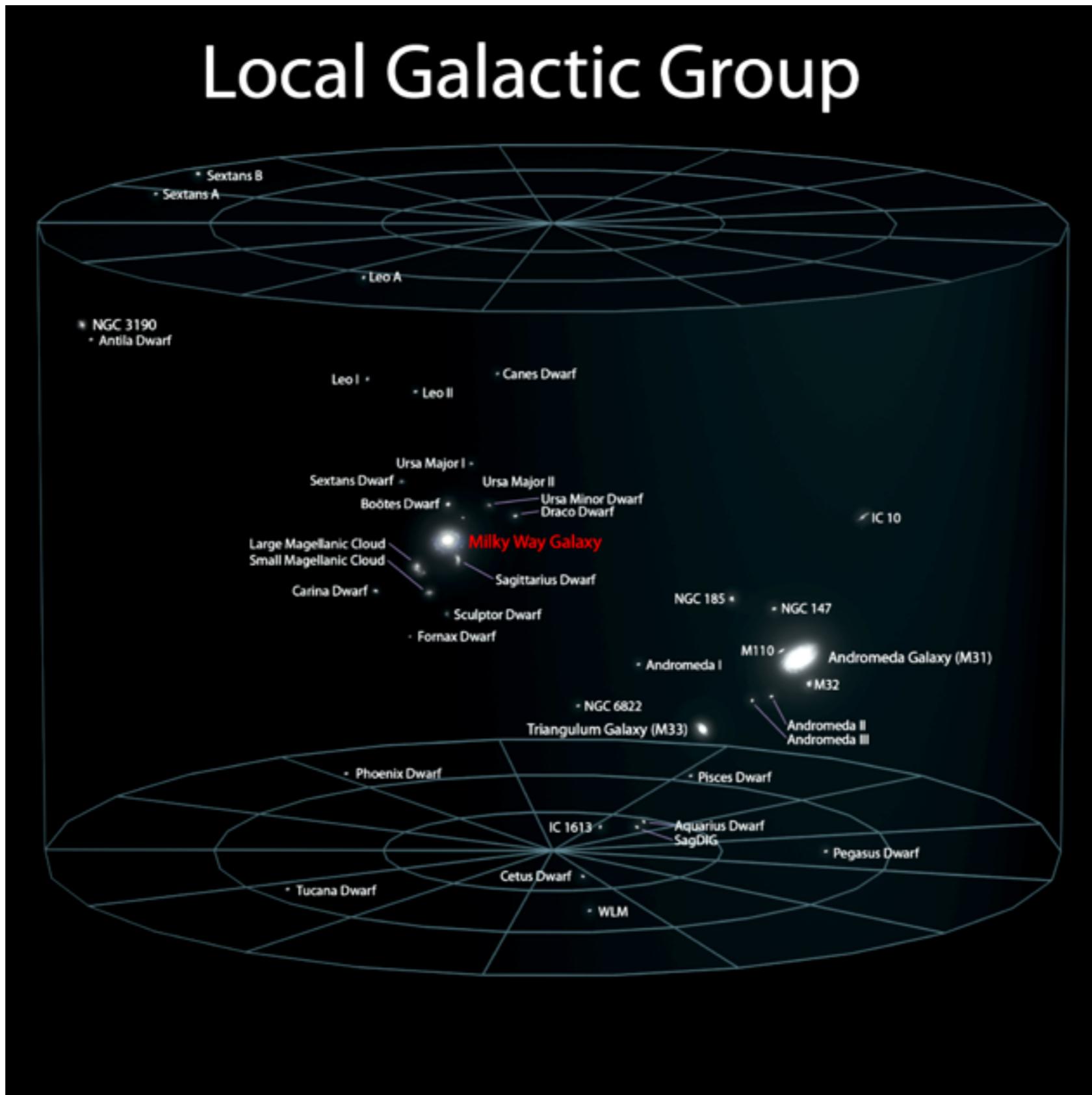


- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

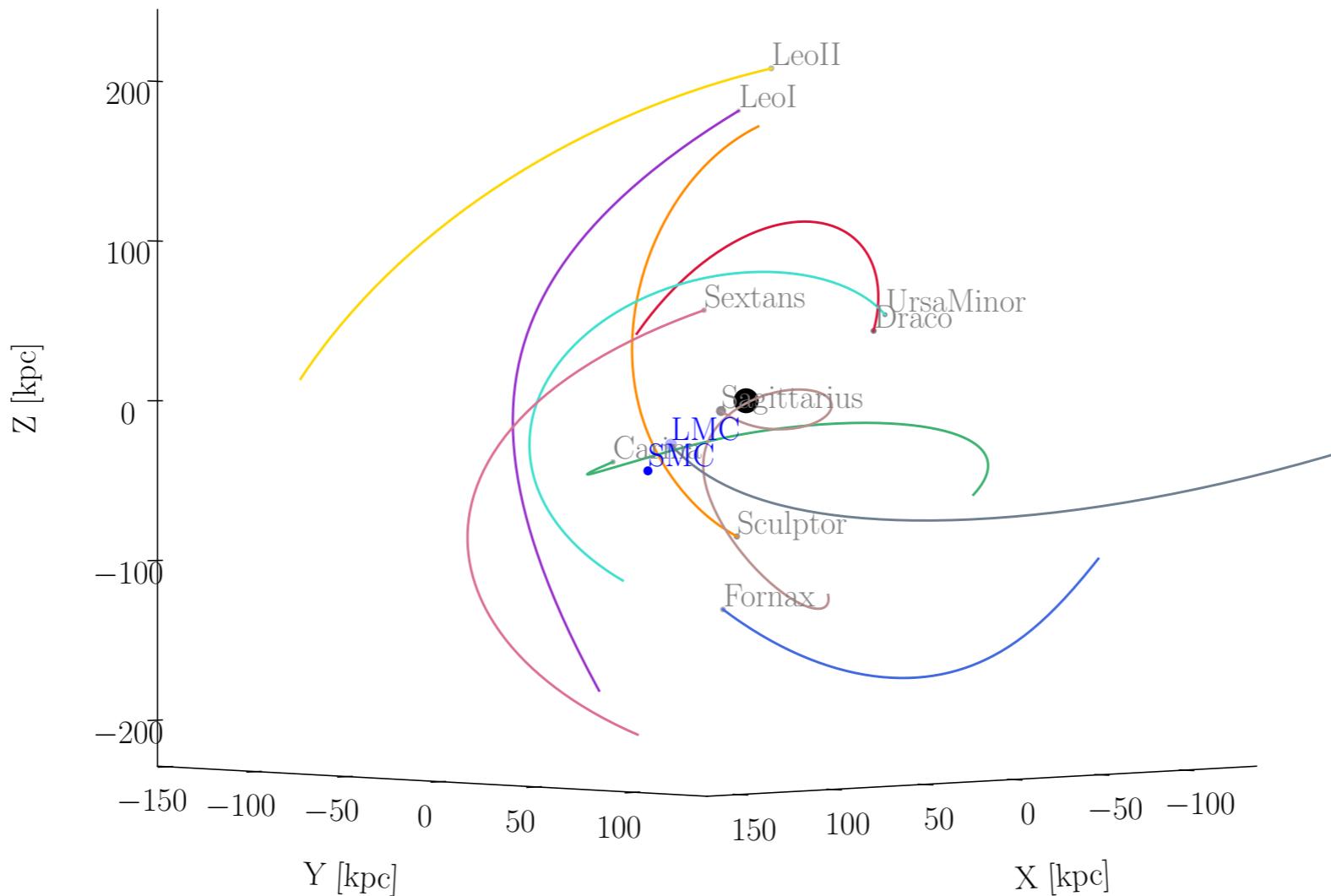
Illustris Cosmological Simulation Movie

[http://www.illustris-project.org/movies/
illustris_movie_cube_sub_frame.mp4](http://www.illustris-project.org/movies/illustris_movie_cube_sub_frame.mp4)

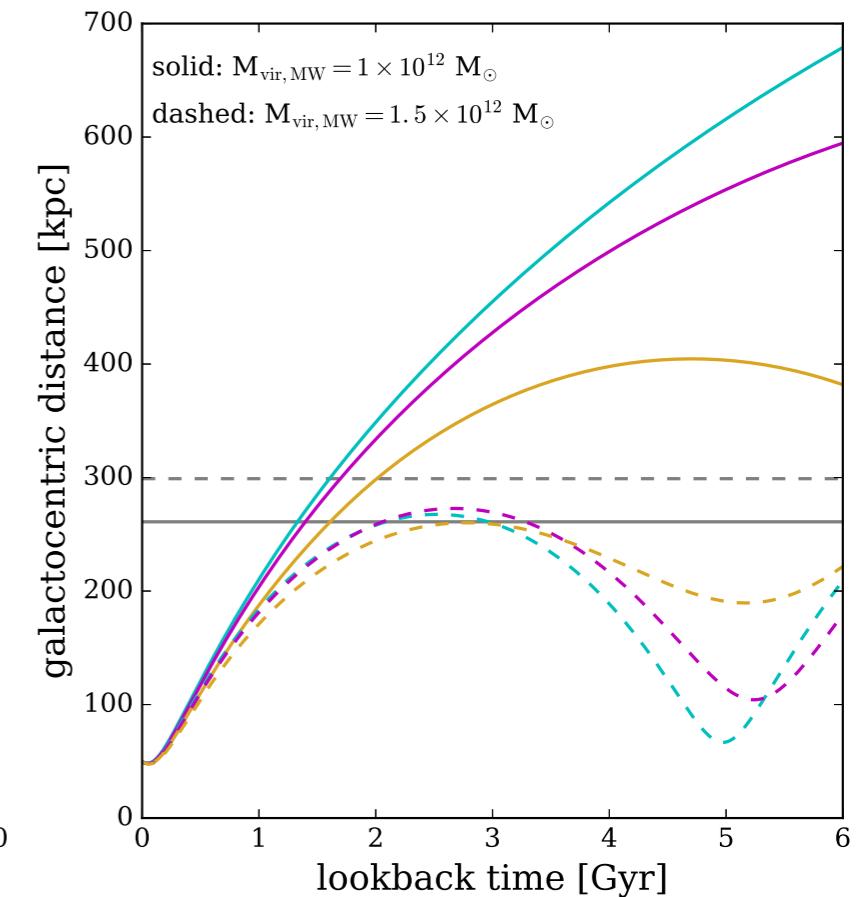
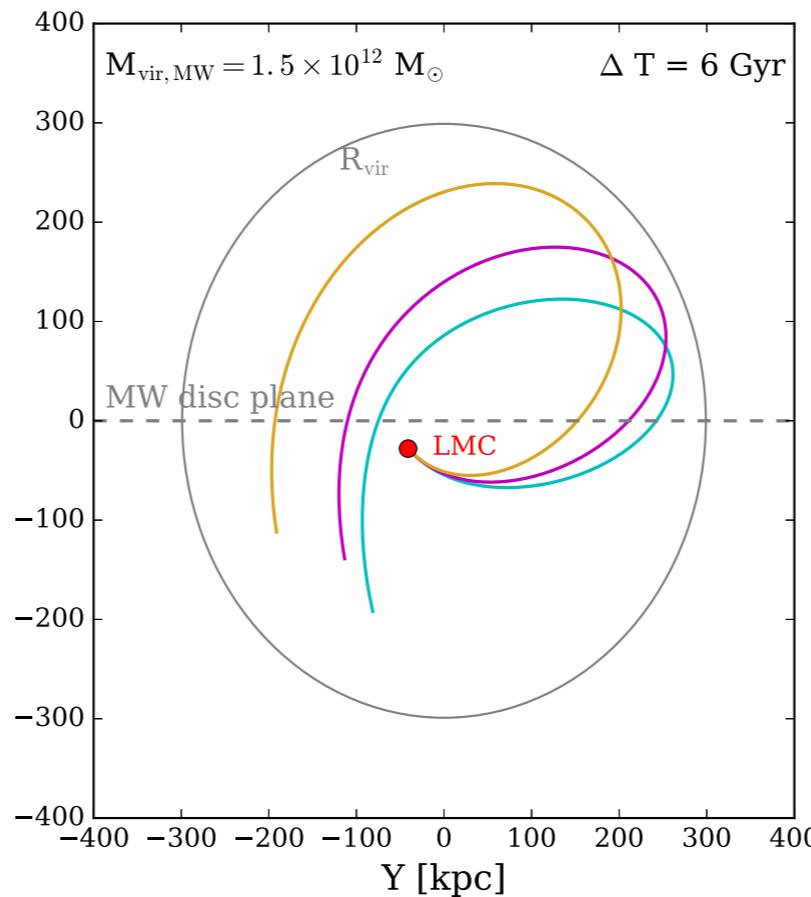
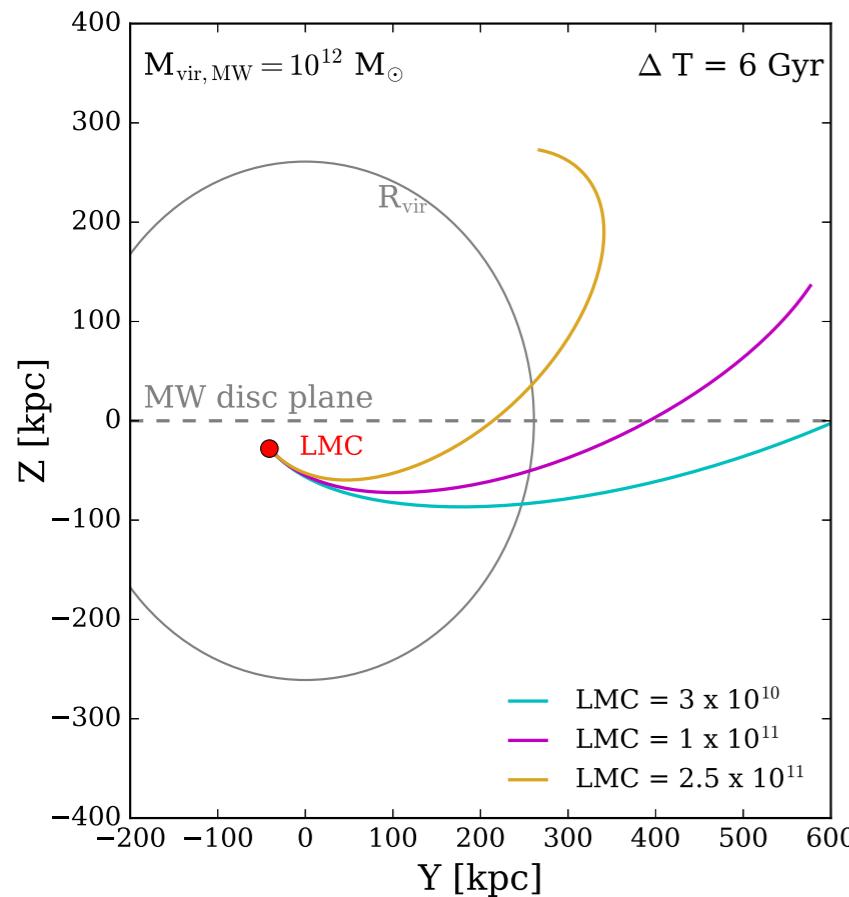
Milky Way has satellite galaxies



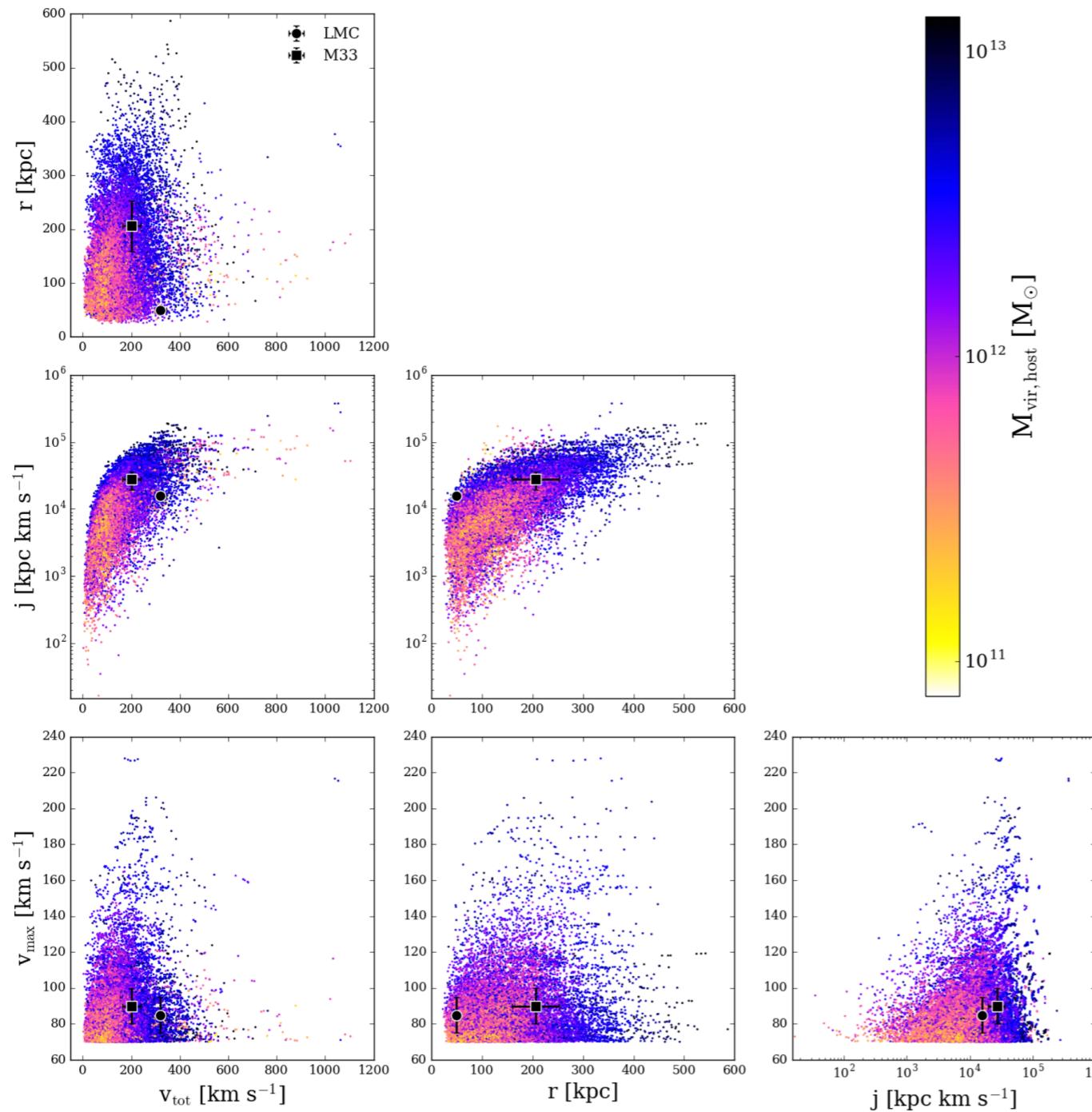
They are moving around



Their trajectories dependent on the Milky Way Mass



Velocities (v), positions (r), momenta (j),
of satellites are correlated with mass via
galaxy formation physics in simulations (Prior)



$x = \text{latent (true) values}$
 $\text{of } v, r, j$

$M_{\text{vir}} = \text{Mass of Galaxy}$

Parameters are:
 $\theta = (x, M_{\text{vir}})$

We can measure the (v , r , j) of MW's biggest satellite, Large Magellanic Cloud (LMC)

Table 1. Observational data (\mathbf{d}) for the LMC and M33 used to build likelihoods in the Bayesian inference scheme include the maximum circular velocity, current separation from the host galaxy and total velocity relative to the host galaxy.

	LMC μ	LMC σ	M33 μ	M33 σ
v_{\max}^{obs} (km s $^{-1}$)	85 ^a	10	90 ^b	10
r^{obs} (kpc)	50	5	203	47
$v_{\text{tot}}^{\text{obs}}$ (km s $^{-1}$)	321	24	202	38
j^{obs} (kpc km s $^{-1}$)	15 688	1788	27 656	8219

Notes. ^aThe maximal circular velocity of the LMC's halo rotation curve is adopted from Besla et al. (2012).

^bM33's halo rotation curve maximum is duplicated from van der Marel et al. (2012b).

M33's position, velocity and their errors are adopted from Paper I (table 1), and references within.

$$\mathcal{L}(\mathbf{x}|\mathbf{d}) = N(v_{\max}^{\text{obs}}|v_{\max}, \sigma_v^2) \times N(r^{\text{obs}}|r, \sigma_r^2) \times N(v^{\text{obs}}|v_{\text{tot}}, \sigma_v^2), \quad (8)$$

where

$$N(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y-\mu)^2}{2\sigma^2} \right] \quad (9)$$

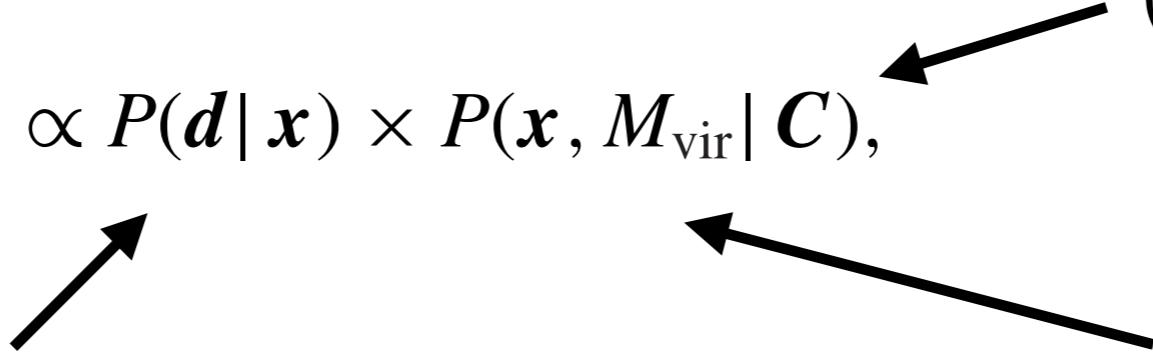
How do we combine these measurements (likelihood) with the joint prior on $P(v, r, j, M)$ from the Simulations?

d = measurements
 x = latent (true) values
 M_{vir} = Mass of Galaxy

3.2.3 *Importance sampling*

Now that the prior and likelihood have been defined, we return to Bayes' theorem:

$$P(x, M_{\text{vir}} | d, C) \propto P(d | x) \times P(x, M_{\text{vir}} | C), \quad (11)$$

Likelihood (observations) Prior (samples from Simulation) 

Importance Sampling

Parameters are: $\theta = (\mathbf{x}, M_{\text{vir}})$

measured data are: \mathbf{d}

Expectations of functions of the physical parameters under the posterior PDF are approximated as sums over the n samples as follows:

$$\begin{aligned} \int f(\boldsymbol{\theta}) P(\mathbf{x}, M_{\text{vir}} | \mathbf{d}, \mathbf{C}) d\boldsymbol{\theta} &= \frac{\int f(\boldsymbol{\theta}) P(\mathbf{d} | \mathbf{x}) P(\mathbf{x}, \underline{M}_{\text{vir}} | \mathbf{C}) d\boldsymbol{\theta}}{\int P(\mathbf{d} | \mathbf{x}) P(\mathbf{x}, M_{\text{vir}} | \mathbf{C}) d\boldsymbol{\theta}} \\ &\approx \frac{\sum_j^n f(\boldsymbol{\theta}_j) P(\mathbf{d} | \mathbf{x}_j)}{\sum_j^n P(\mathbf{d} | \mathbf{x}_j)}. \end{aligned} \quad (12)$$

The denominator of this equation is the normalization constant. If

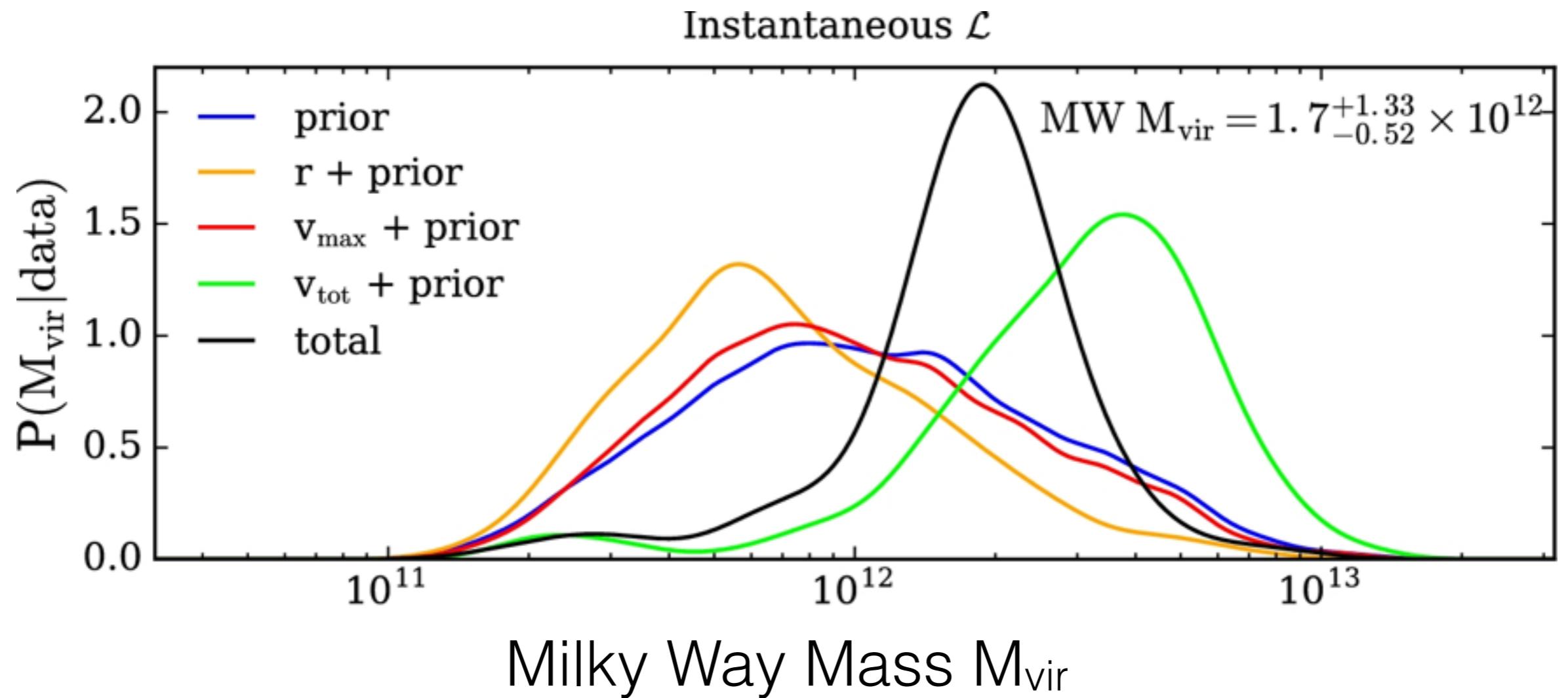
Importance Sampling

$$\begin{aligned} & \int f(M_{\text{vir}}) P(M_{\text{vir}} | \mathbf{d}, \mathbf{C}) dM_{\text{vir}} \\ &= \int f(M_{\text{vir}}) P(\mathbf{x}, M_{\text{vir}} | \mathbf{d}, \mathbf{C}) d\mathbf{x} dM_{\text{vir}} \\ &\approx \frac{\sum_j^n f(M_{\text{vir}}^j) P(\mathbf{d} | \mathbf{x}_j)}{\sum_j^n P(\mathbf{d} | \mathbf{x}_j)} \\ &= \sum_j^n f(M_{\text{vir}}^j) w_j, \end{aligned}$$

where $w_i = P(\mathbf{d} | \mathbf{x}_i) / \sum_j^n P(\mathbf{d} | \mathbf{x}_j)$ are importance weights.

Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations

Posterior Density of Milky Way Galaxy Mass with KDE



Is the estimate robust?
 Kinematic properties are changing as LMC moves:
 momentum inference more invariant with time

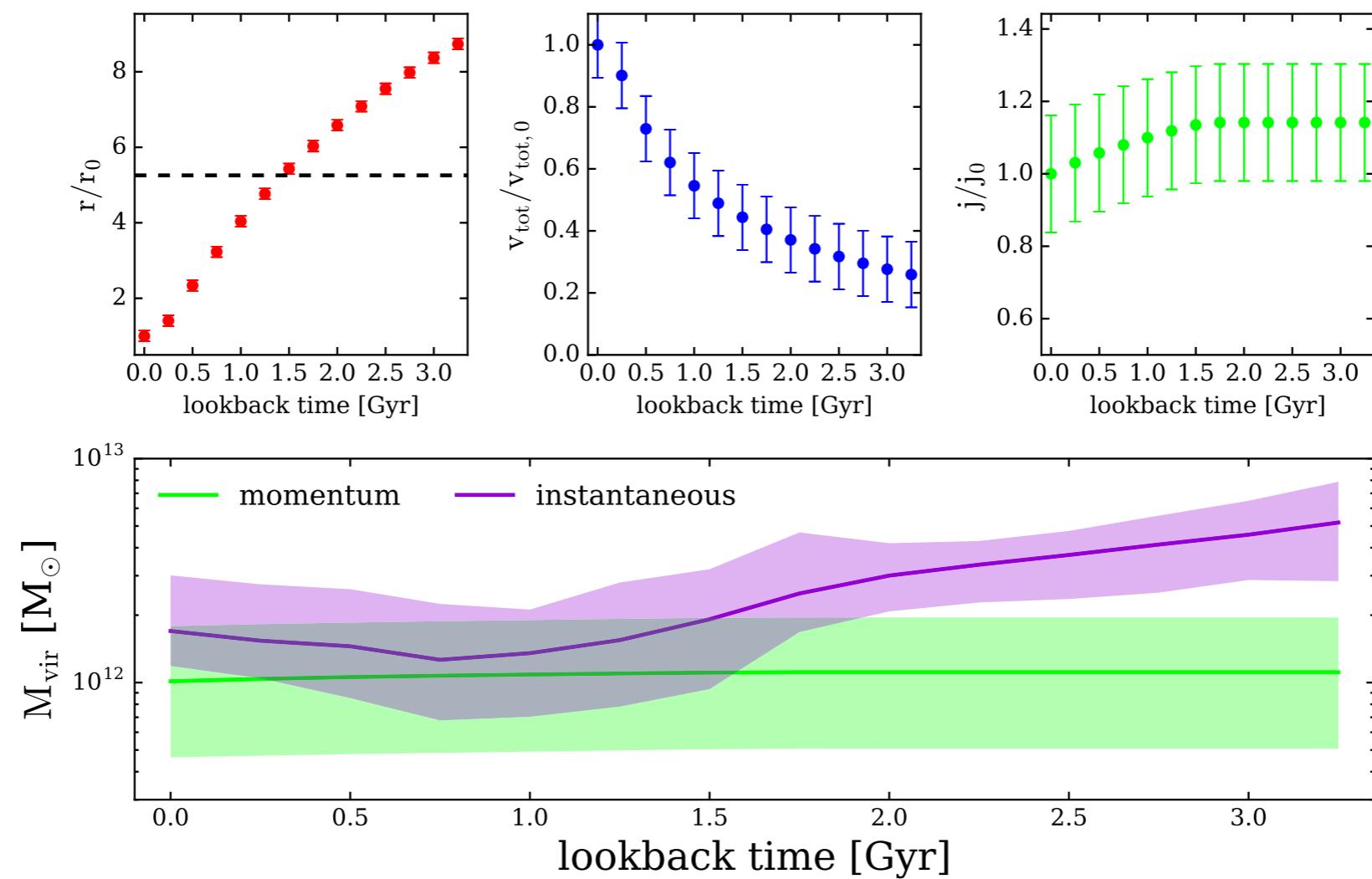


Figure 5. Posterior mean mass estimates for the MW based on the orbital history of the LMC using the two likelihood functions. The top left-hand panel

(Show movie)