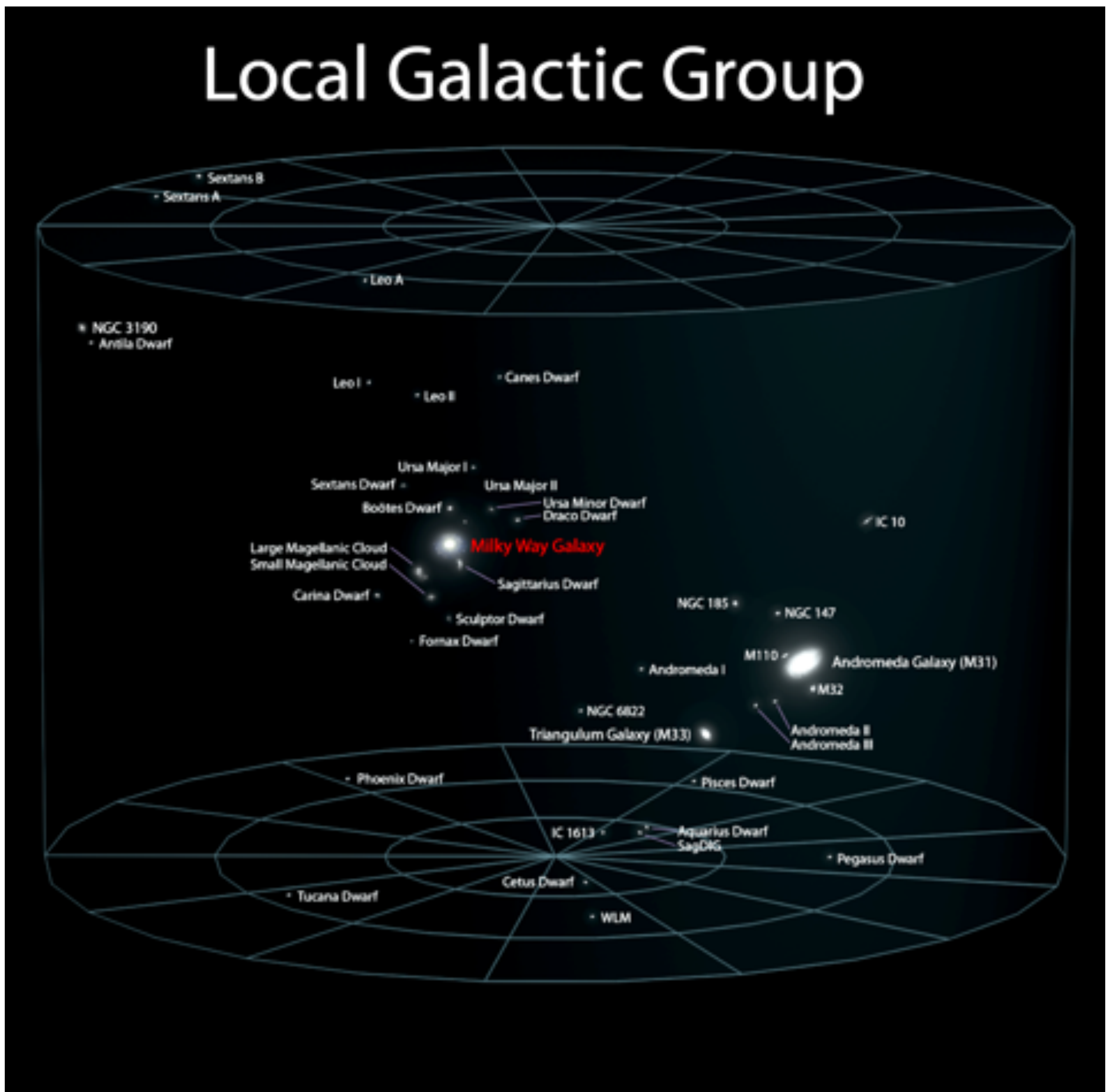
# Astrostatistics: Sat 17 Feb 2017

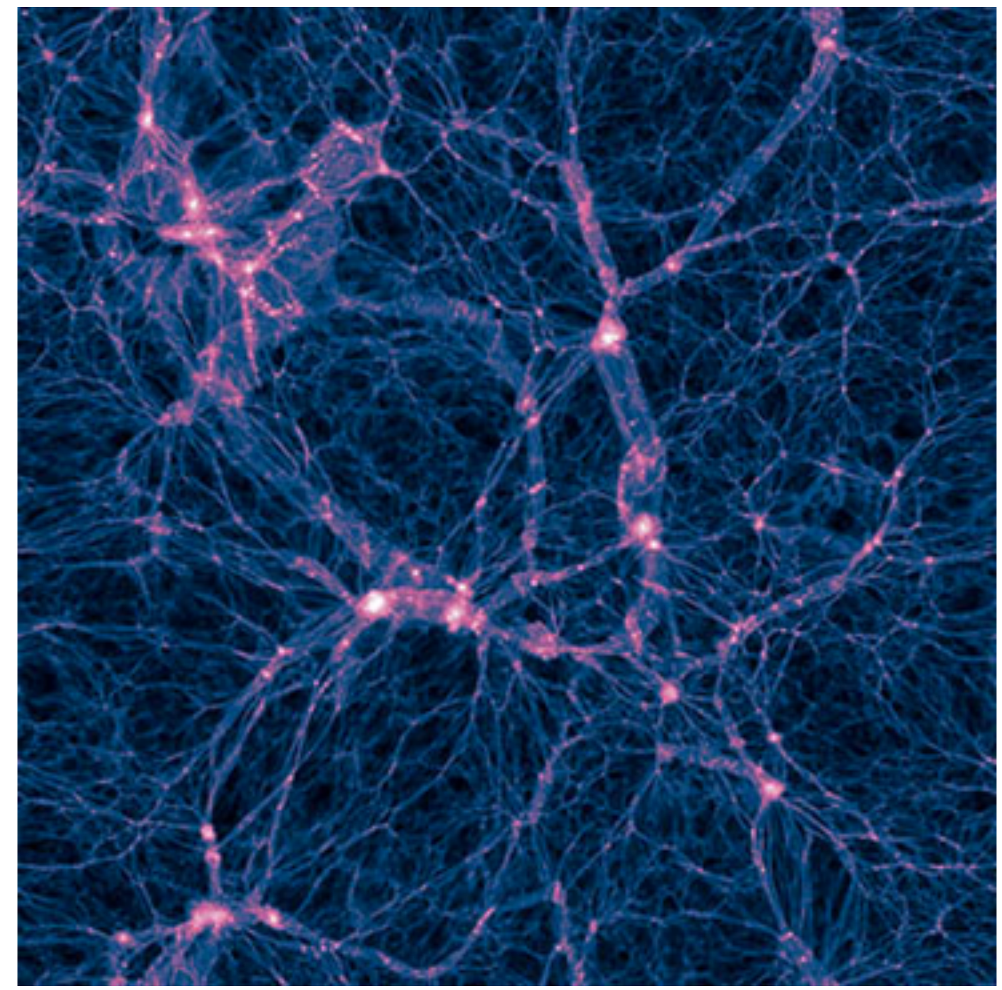
https://github.com/CambridgeAstroStat/PartIII-Astrostatistics

- 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

*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)

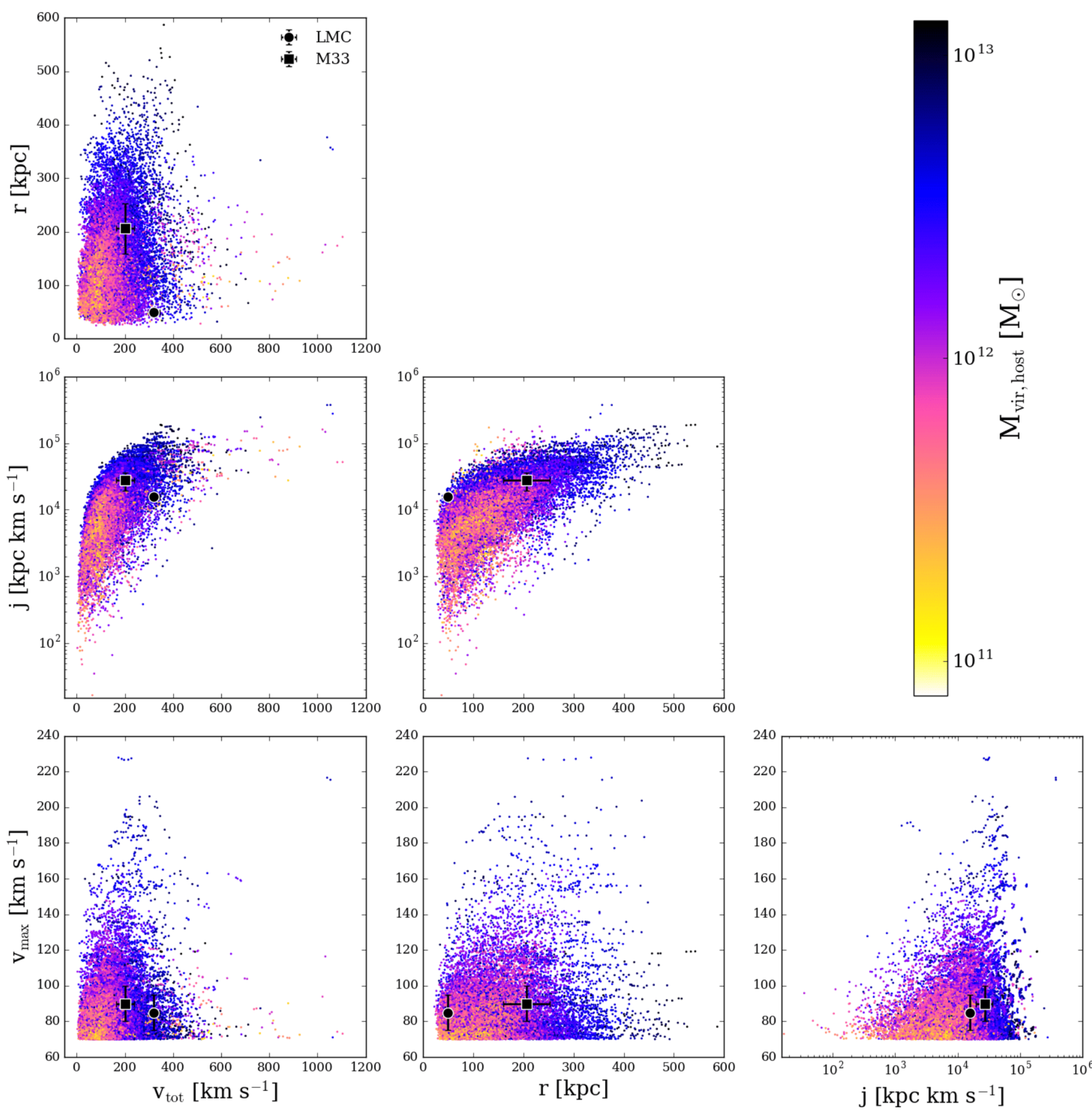


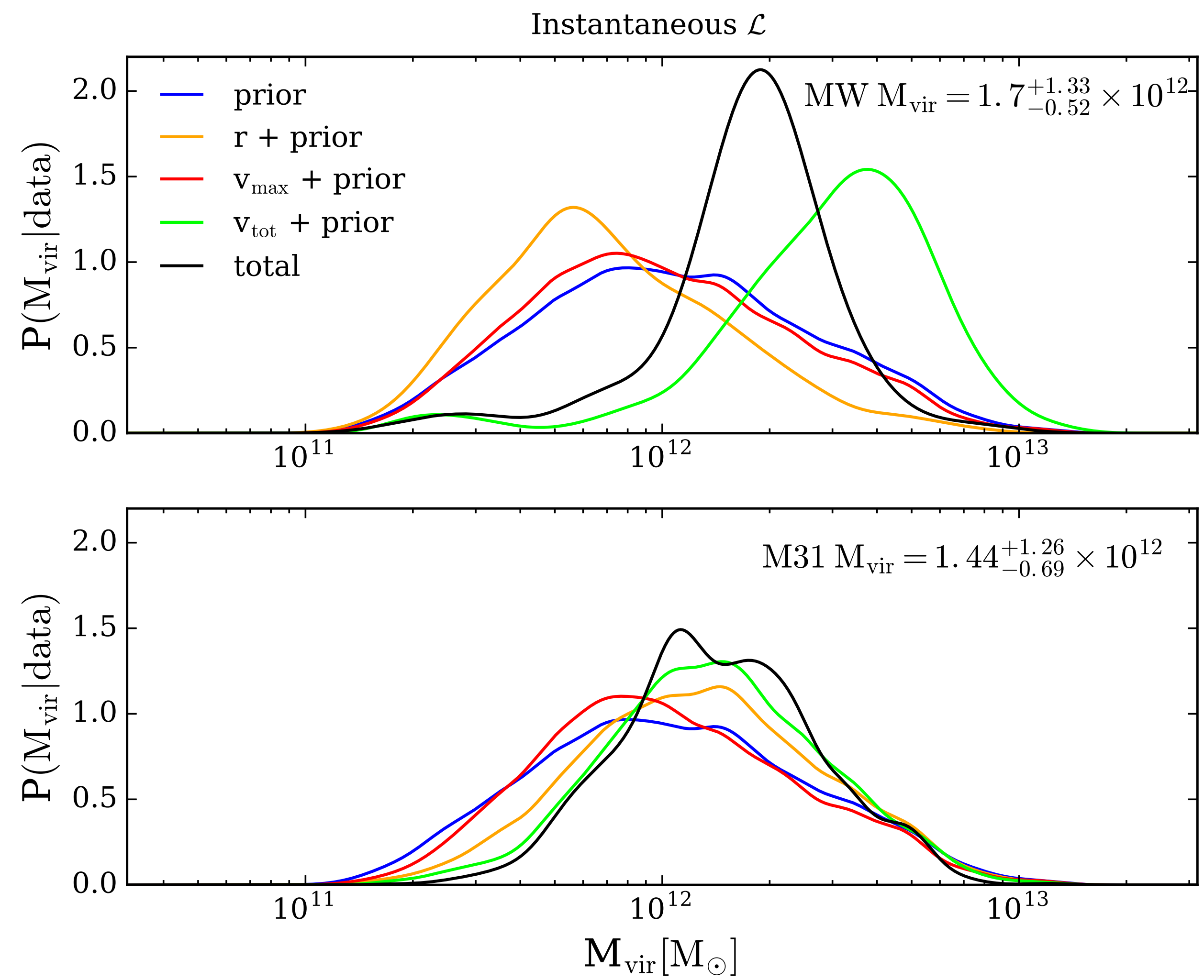
Illustris
Cosmological Simulation of Galaxy Formation

# *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)



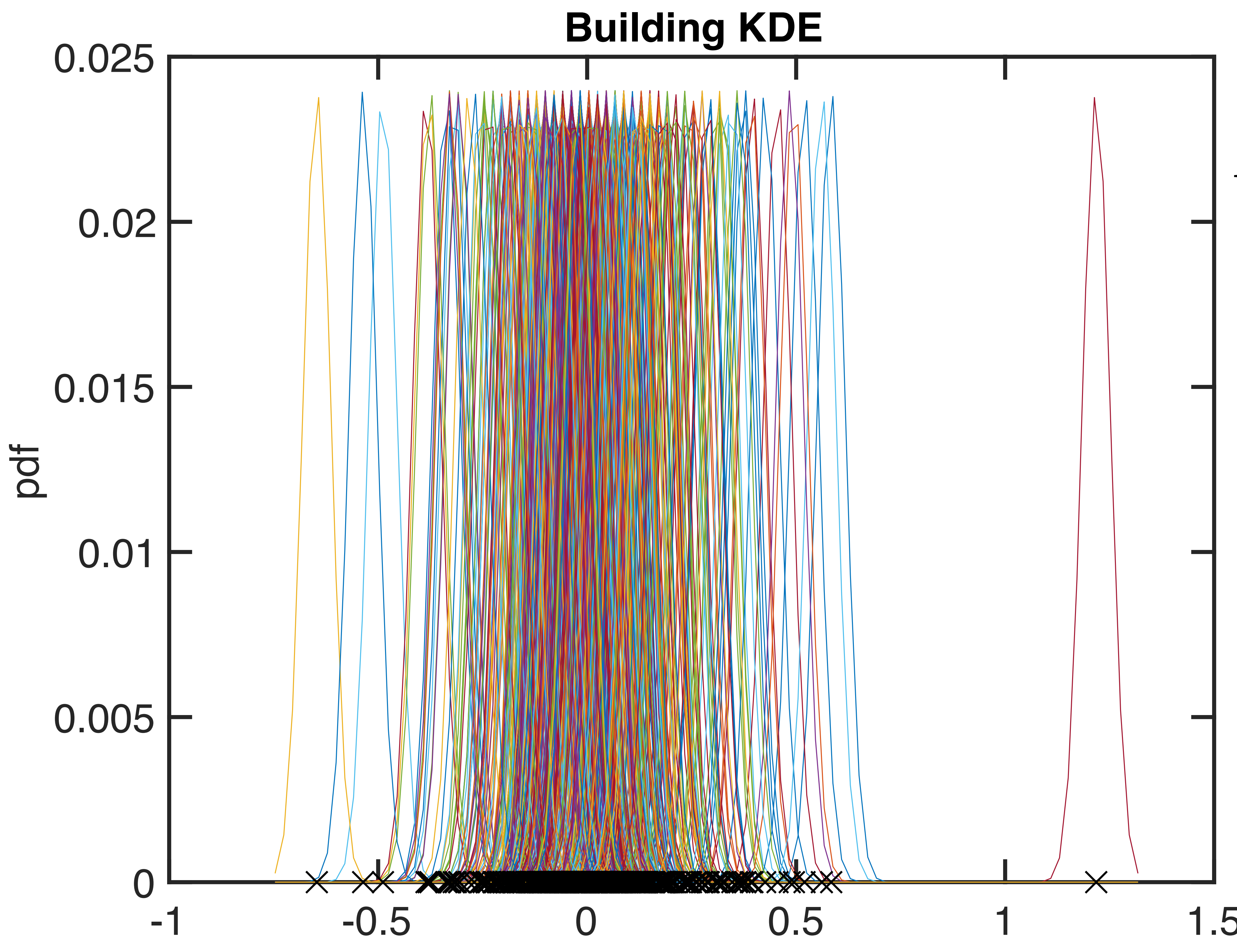

Simulation —> Prior




- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

# Kernel Density Estimation (KDE) (Smooth Histogram)

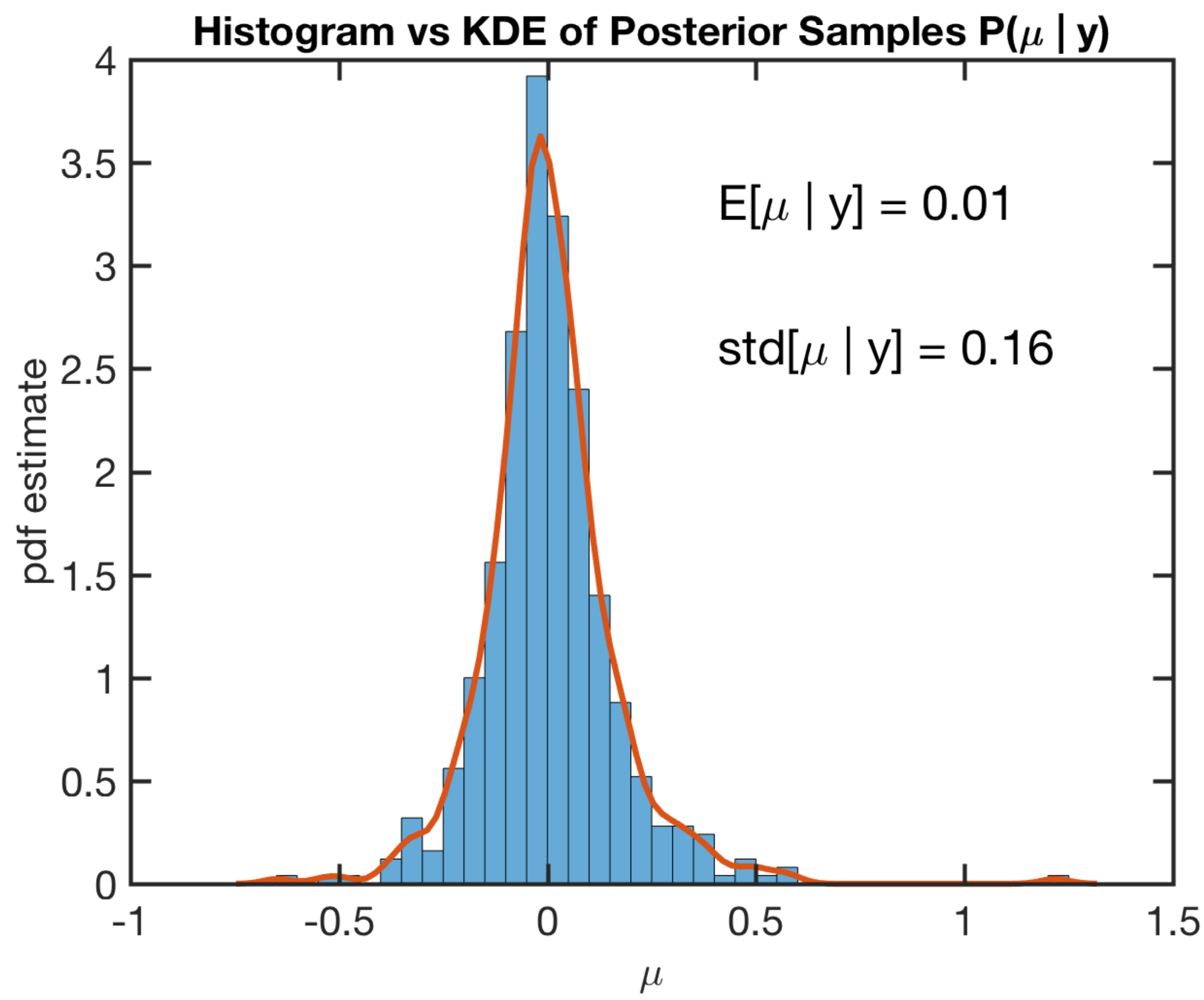
Each sample gets a Gaussian at the sample point
with an "optimal" bandwidth bw (rule of thumb)




$$\mathrm{bw} = \left(\frac{4\sigma^5}{3n}\right)^{1/5}$$

# Kernel Density Estimation (KDE) (Smooth Histogram)

Then add them up and normalise pdf to 1



# Weighted KDE

$$\mathrm{wkde}(\theta) = \sum_{s=1}^{m} w_s \times N(\theta|\, \theta_s, \mathrm{bw})$$

Silverman's Rule of Thumb: $\mathrm{bw} = \left(\frac{4\sigma^5}{3n}\right)^{1/5}$

if equal weights: $w_i = 1/m$

$$\mathrm{kde}(\theta) = \frac{1}{m}\sum_{s=1}^{m} N(\theta|\, \theta_s, \mathrm{bw})$$

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

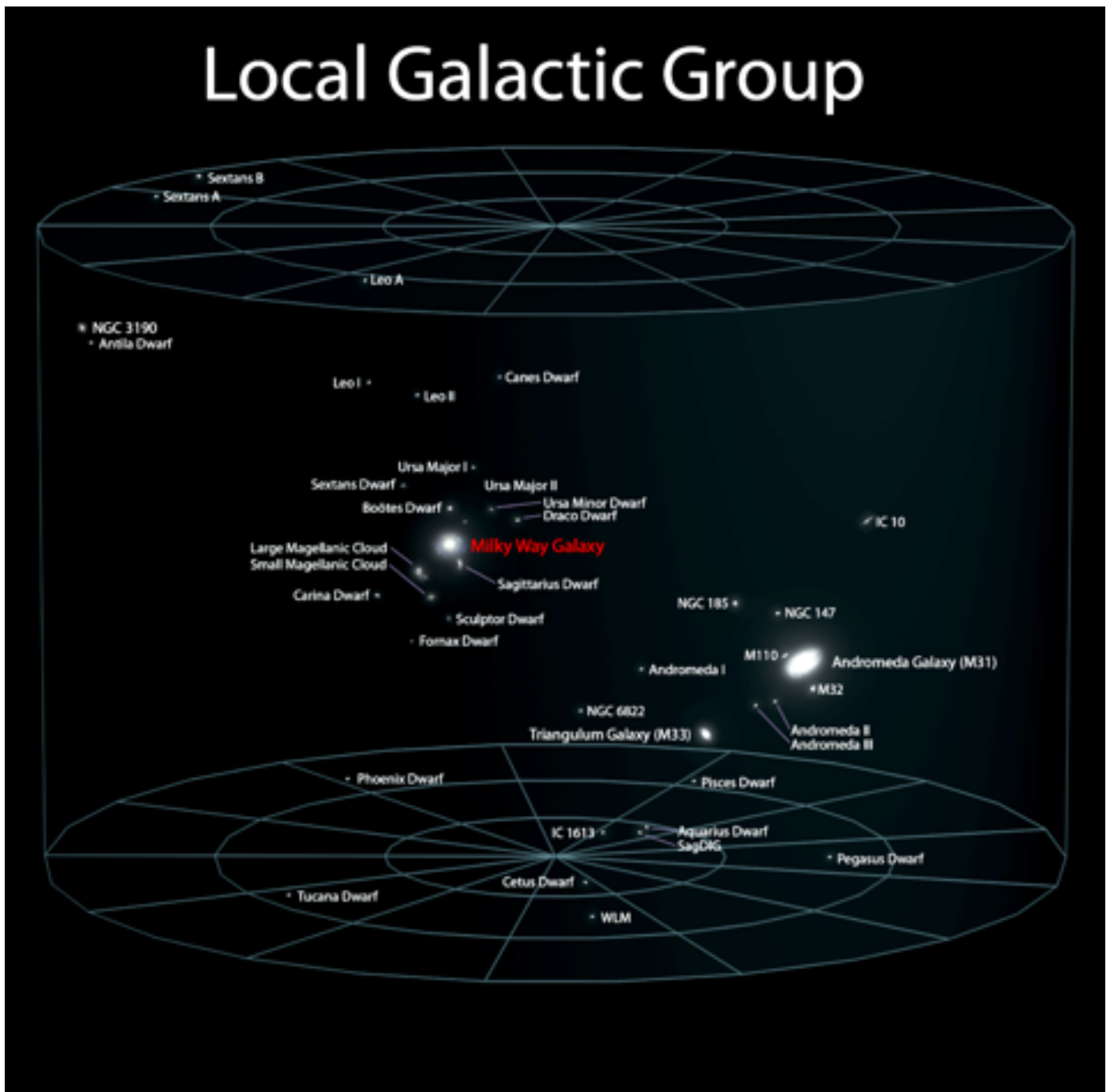
$$\mathbb{E}[f(\boldsymbol{\theta})|\, D] = \int f(\boldsymbol{\theta}) P(\boldsymbol{\theta}|\, D)\, d\boldsymbol{\theta} \approx \frac{1}{m} \sum_{i=1}^{m} f(\boldsymbol{\theta}_i)$$

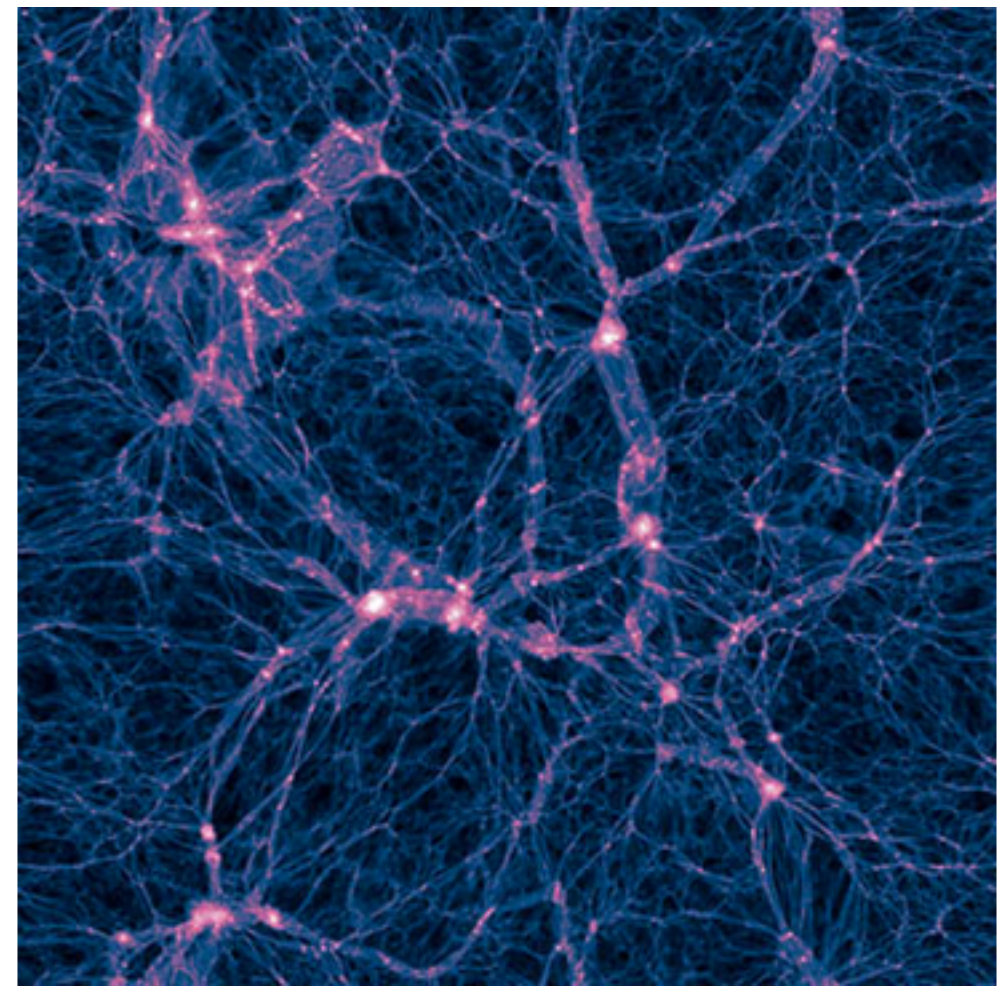
- Posterior simulation - Markov Chain Monte Carlo, Nested Sampling, etc. generates draws
- Importance Sampling - draw from an easier ("tractable") distribution $\theta_i$ ~ Q($\theta$) and weight the samples by $w_i$ = P($\theta_i$ | D) / Q($\theta_i$)

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

*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)

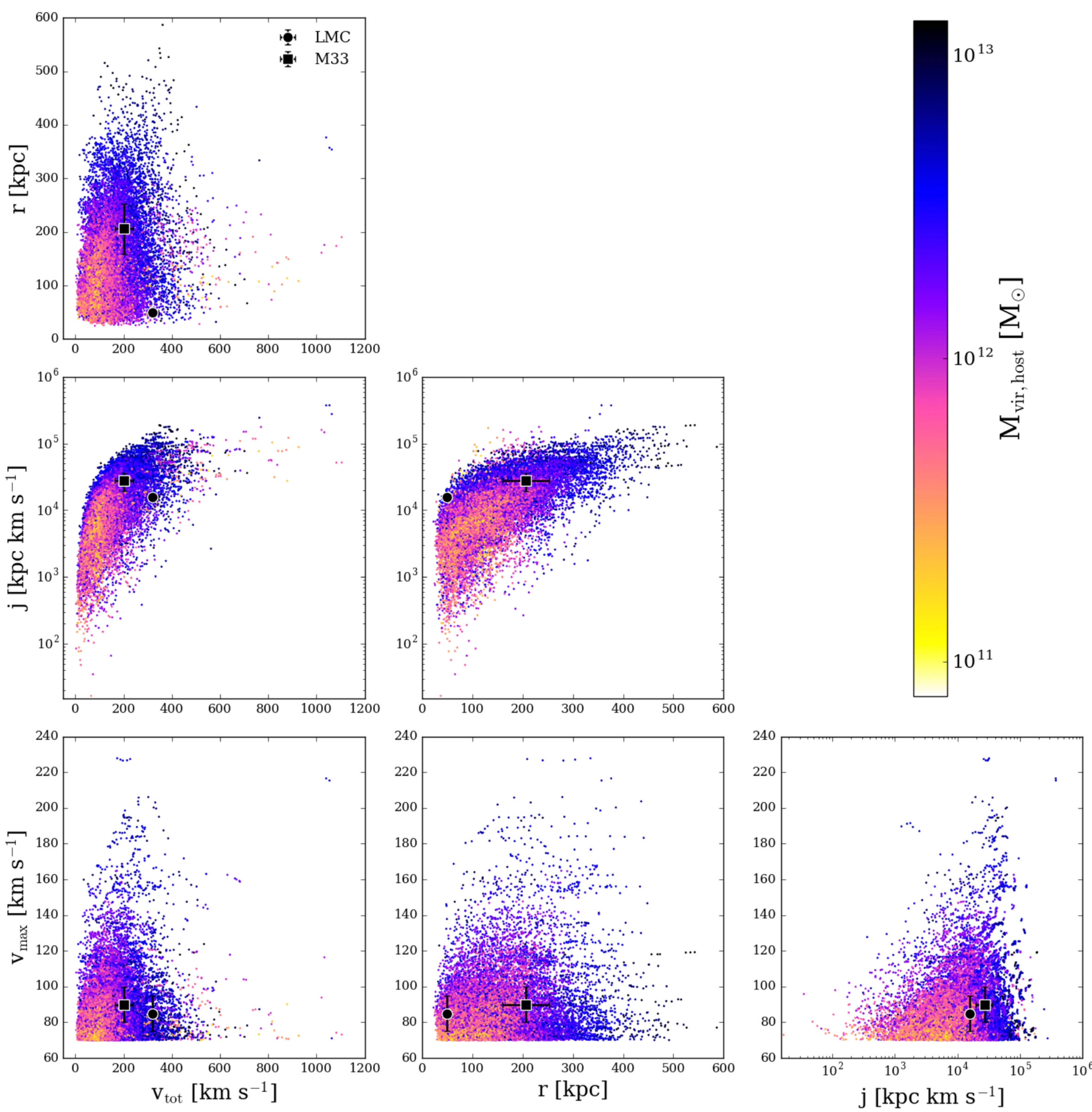


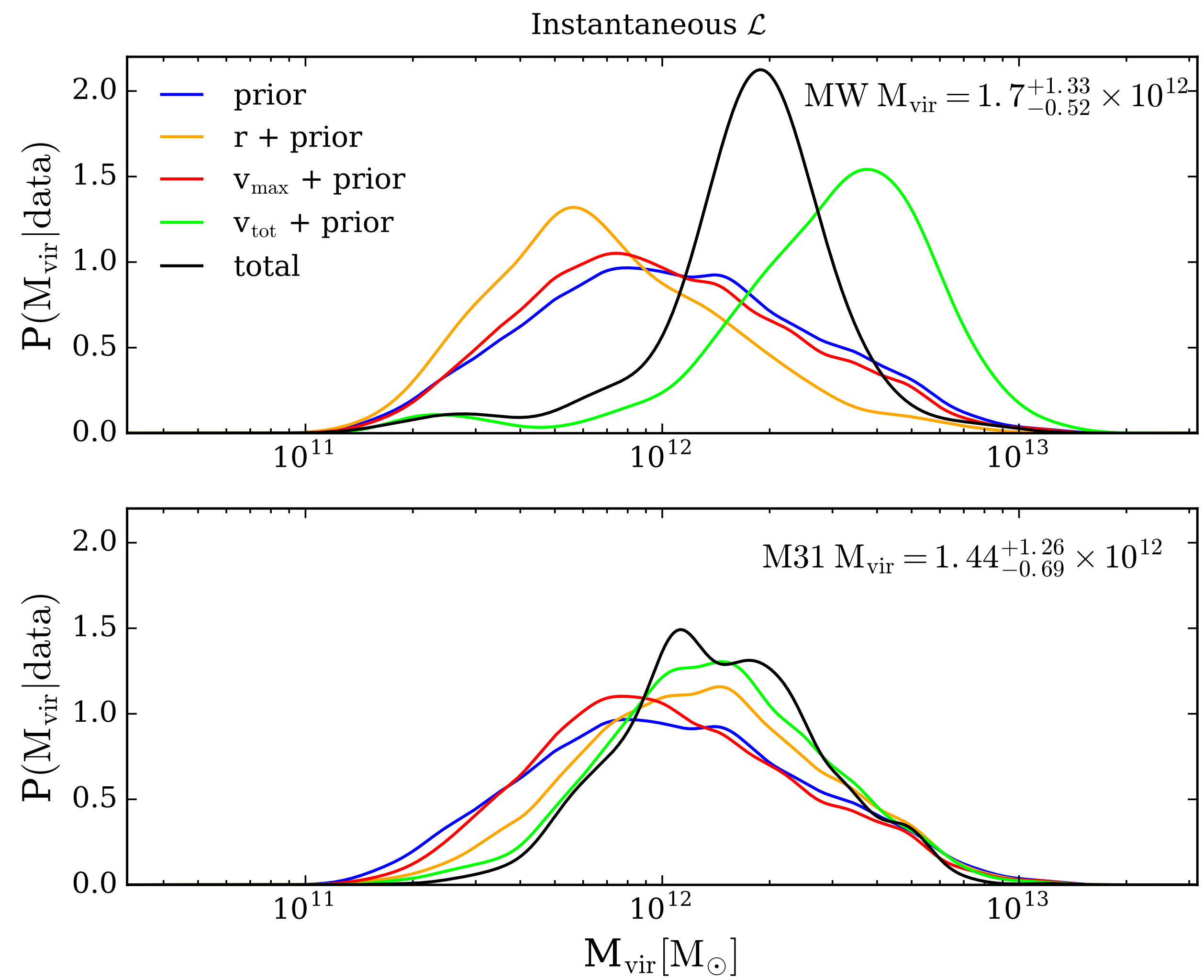
Illustris
Cosmological Simulation of Galaxy Formation

# *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)

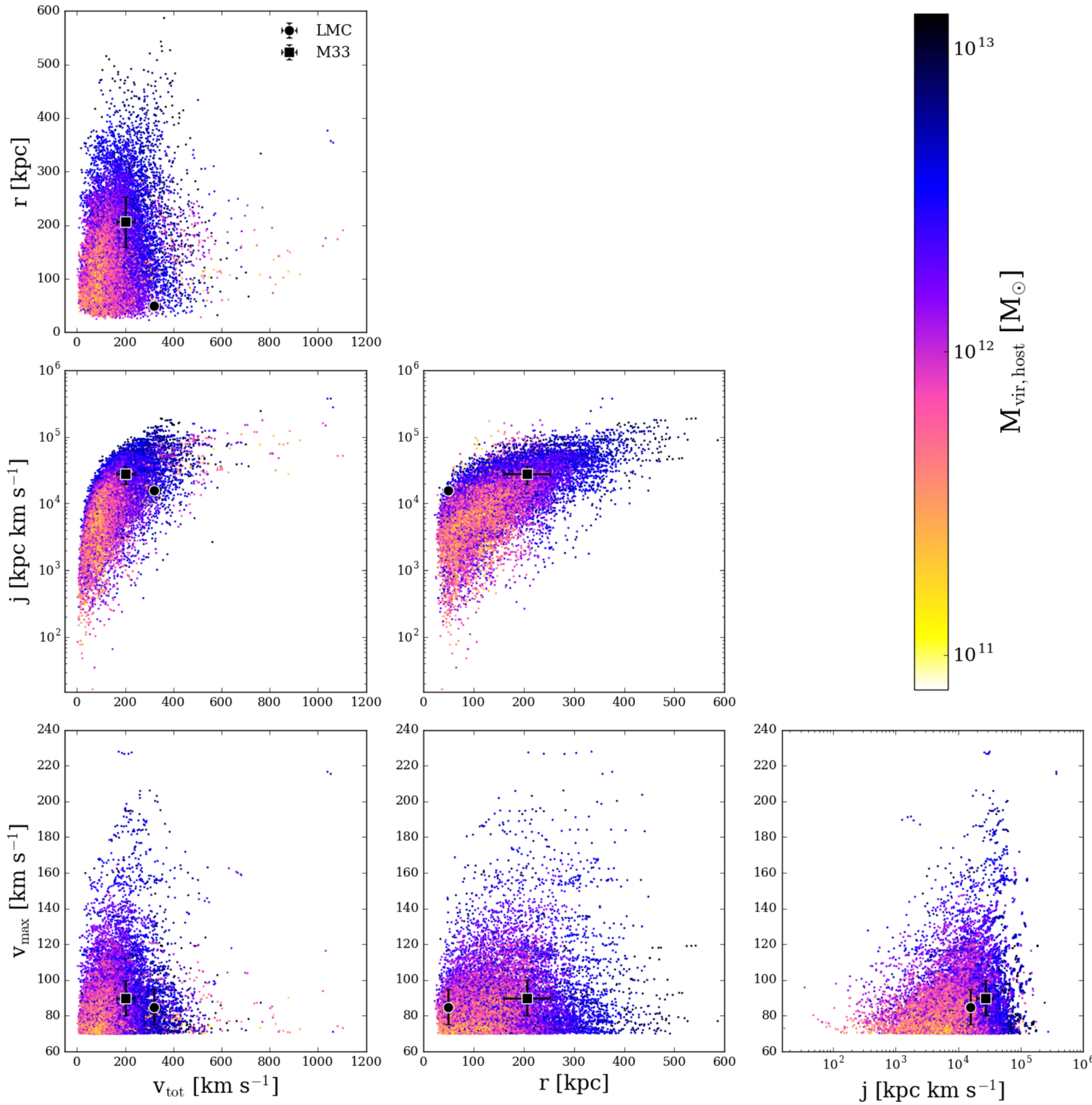



Simulation —> Prior




- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

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




x = latent (true) values of v, r, j

$M_{vir}$ = Mass of Galaxy

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

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

**Table 1.** Observational data ($\boldsymbol{d}$) 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 $\mu$ | LMC $\sigma$ | M33 $\mu$ | M33 $\sigma$ |
|---|---|---|---|---|
| $v_{\rm max}^{\rm obs}$ (km s$^{-1}$) | 85[a] | 10 | 90[b] | 10 |
| $r^{\rm obs}$ (kpc) | 50 | 5 | 203 | 47 |
| $v_{\rm tot}^{\rm obs}$ (km s$^{-1}$) | 321 | 24 | 202 | 38 |
| $j^{\rm obs}$ (kpc km s$^{-1}$) | 15 688 | 1788 | 27 656 | 8219 |

*Notes.* [a]The maximal circular velocity of the LMC's halo rotation curve is adopted from Besla et al. (2012).
[b]M33'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.

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

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:

(Ignore C)

$$P(\boldsymbol{x}, M_{\mathrm{vir}} | \boldsymbol{d}, C) \propto P(\boldsymbol{d} | \boldsymbol{x}) \times P(\boldsymbol{x}, M_{\mathrm{vir}} | \boldsymbol{C}), \tag{11}$$

Likelihood (observations)     Prior (samples from Simulation)

# Importance Sampling

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

measured data are: d

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

$$\int f(\boldsymbol{\theta}) P(\boldsymbol{x}, M_{\mathrm{vir}} | \boldsymbol{d}, \boldsymbol{C})\, \mathrm{d}\boldsymbol{\theta} = \frac{\int f(\boldsymbol{\theta}) P(\boldsymbol{d} | \boldsymbol{x}) P(\boldsymbol{x}, \underline{M}_{\mathrm{vir}} | \boldsymbol{C})\, \mathrm{d}\boldsymbol{\theta}}{\int P(\boldsymbol{d} | \boldsymbol{x}) P(\boldsymbol{x}, M_{\mathrm{vir}} | \boldsymbol{C})\, \mathrm{d}\boldsymbol{\theta}}$$

$$\approx \frac{\sum_j^n f(\boldsymbol{\theta}_j) P(\boldsymbol{d} | \boldsymbol{x}_j)}{\sum_j^n P(\boldsymbol{d} | \boldsymbol{x}_j)}. \quad (12)$$

The denominator of this equation is the normalization constant. If

# Importance Sampling

$$\int f(M_{\rm vir})P(M_{\rm vir}|\,\boldsymbol{d},\boldsymbol{C})\,{\rm d}M_{\rm vir}$$

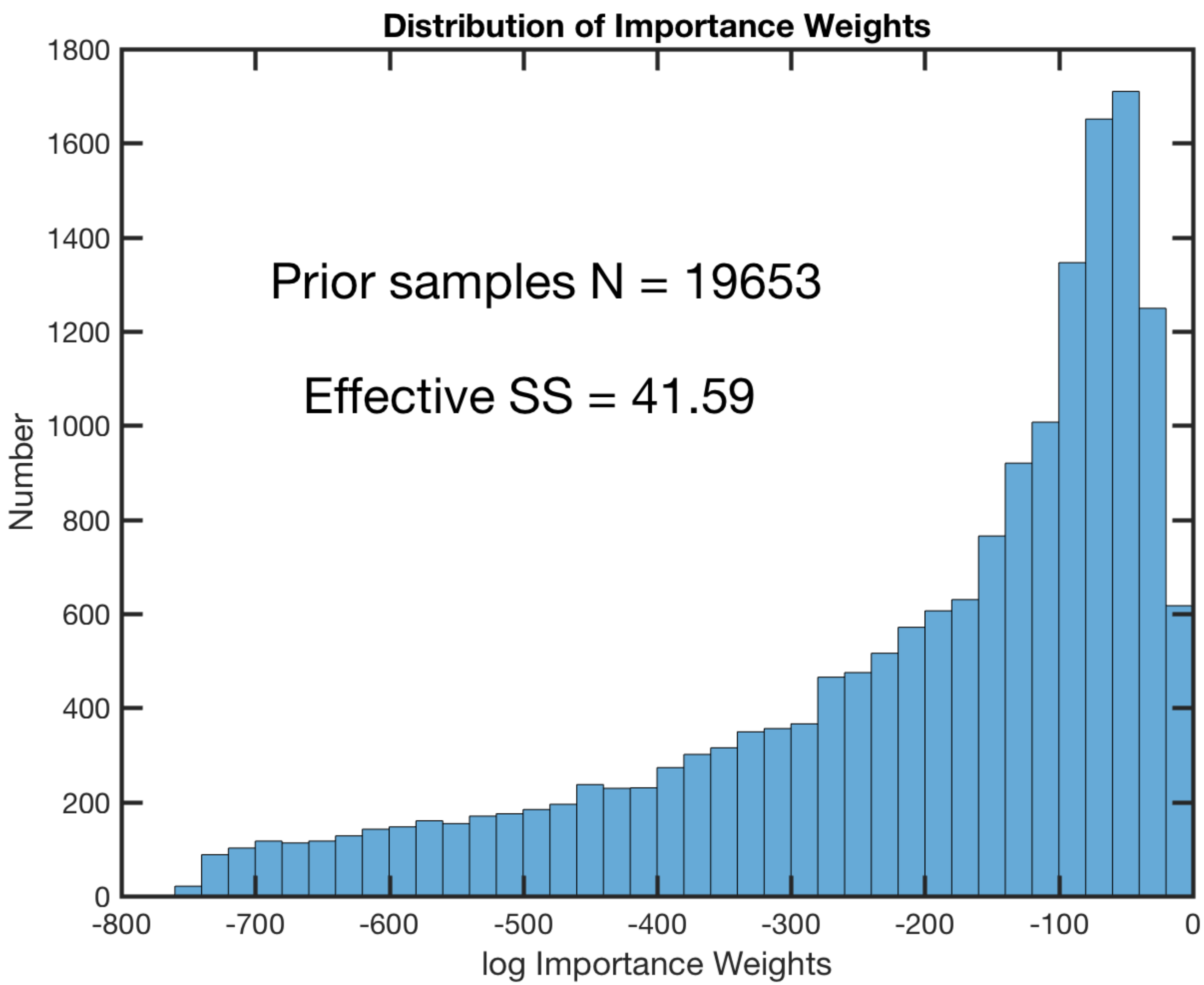
$$= \int f(M_{\rm vir})\,P(\boldsymbol{x},M_{vir}|\,\boldsymbol{d},\boldsymbol{C})\,{\rm d}\boldsymbol{x}\,{\rm d}M_{vir}$$

$$\approx \frac{\sum_j^n f(M^j_{\rm vir})\,P(\boldsymbol{d}|\,\boldsymbol{x}_j)}{\sum_j^n P(\boldsymbol{d}|\,\boldsymbol{x}_j)}$$

$$= \sum_j^n f(M^j_{\rm vir})\,w_j,$$

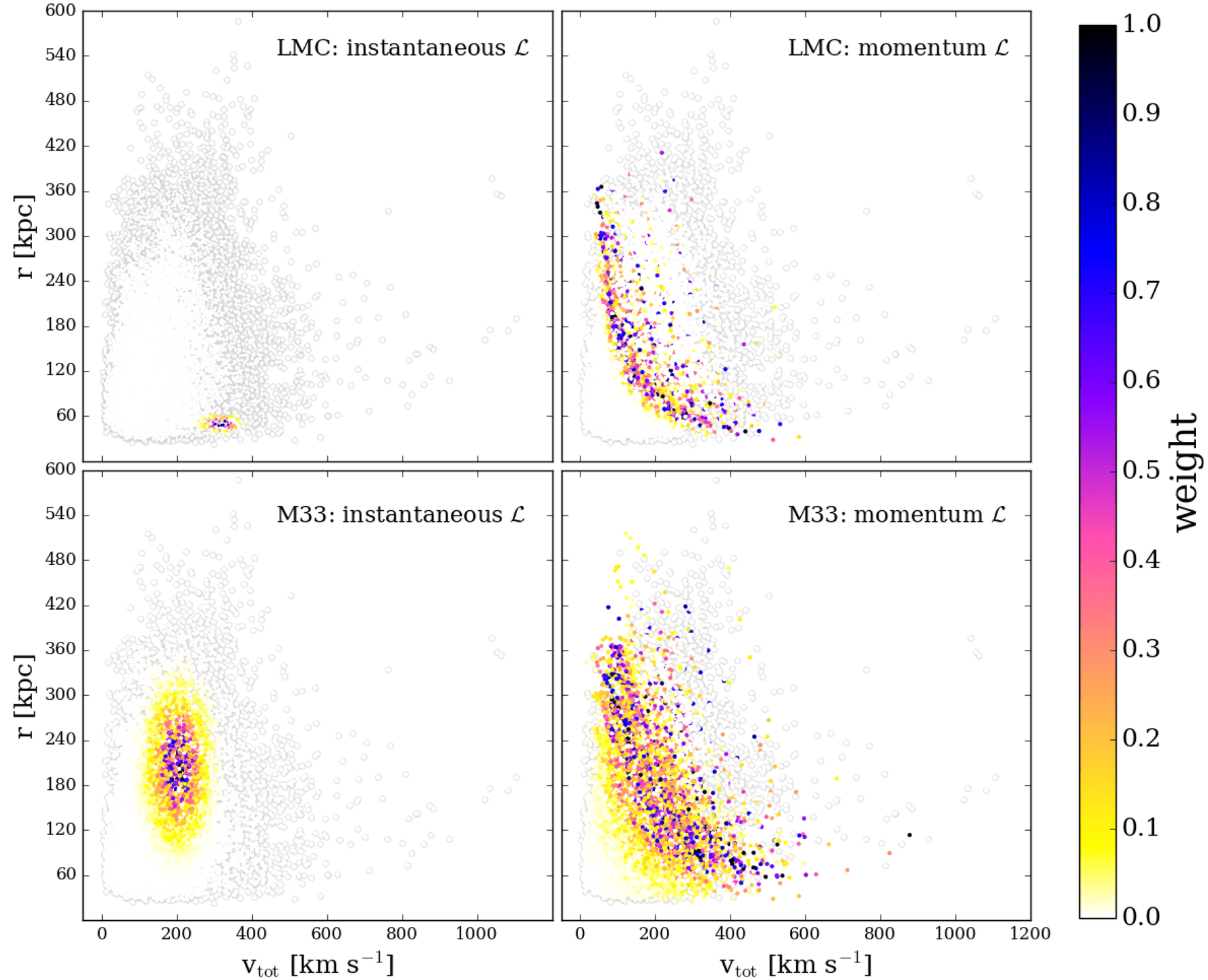
where $w_i = P(\boldsymbol{d}|\,\boldsymbol{x}_i)/\sum_{\rm j}^{\rm n} P(\boldsymbol{d}|\,\boldsymbol{x}_j)$ are importance weights.

# Run code demonstration

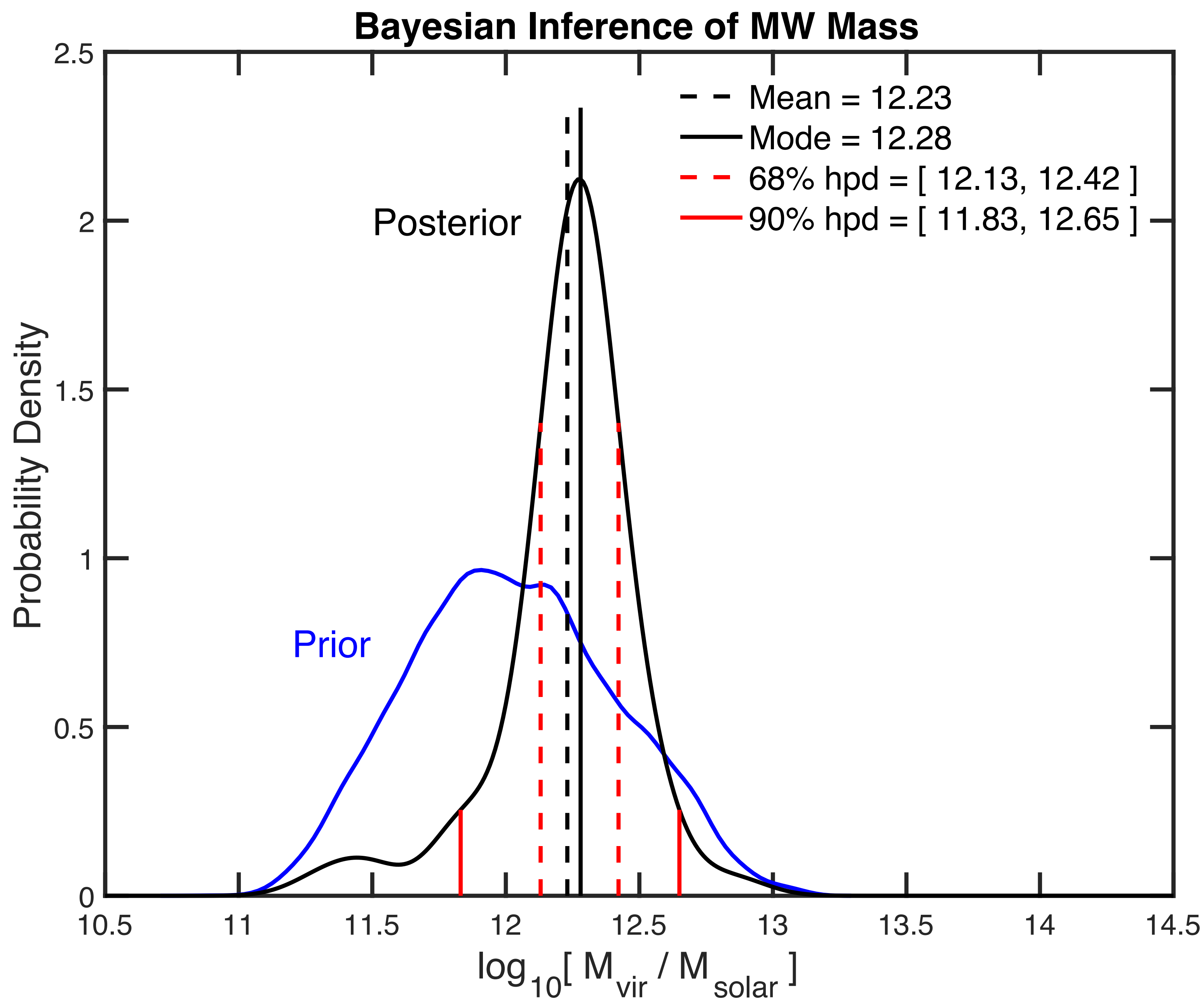



Distribution of Importance Weights

# Distribution of Importance Weights



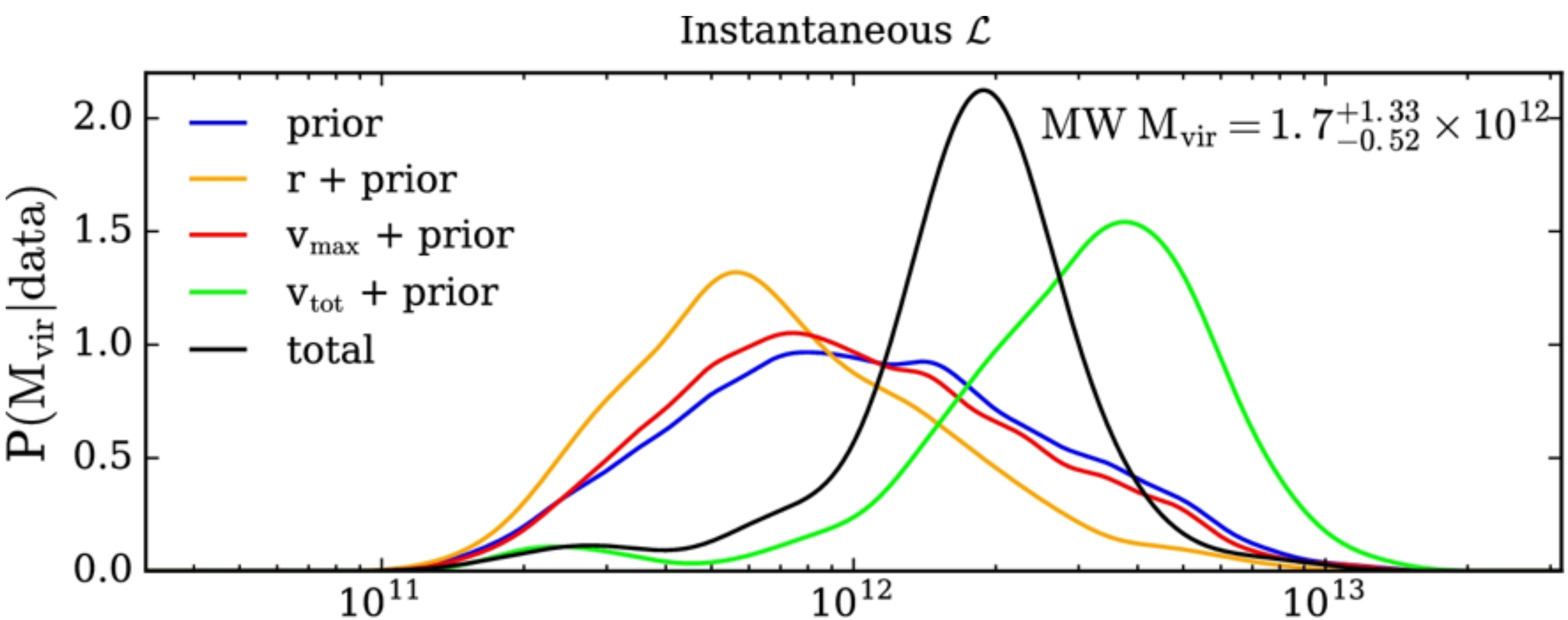
# Code Demo: Posterior using weighted KDE




x% HPD = Highest Posterior Density x% credible region
= interval(s) with highest density containing x% of posterior

# 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




Milky Way Mass $M_{vir}$