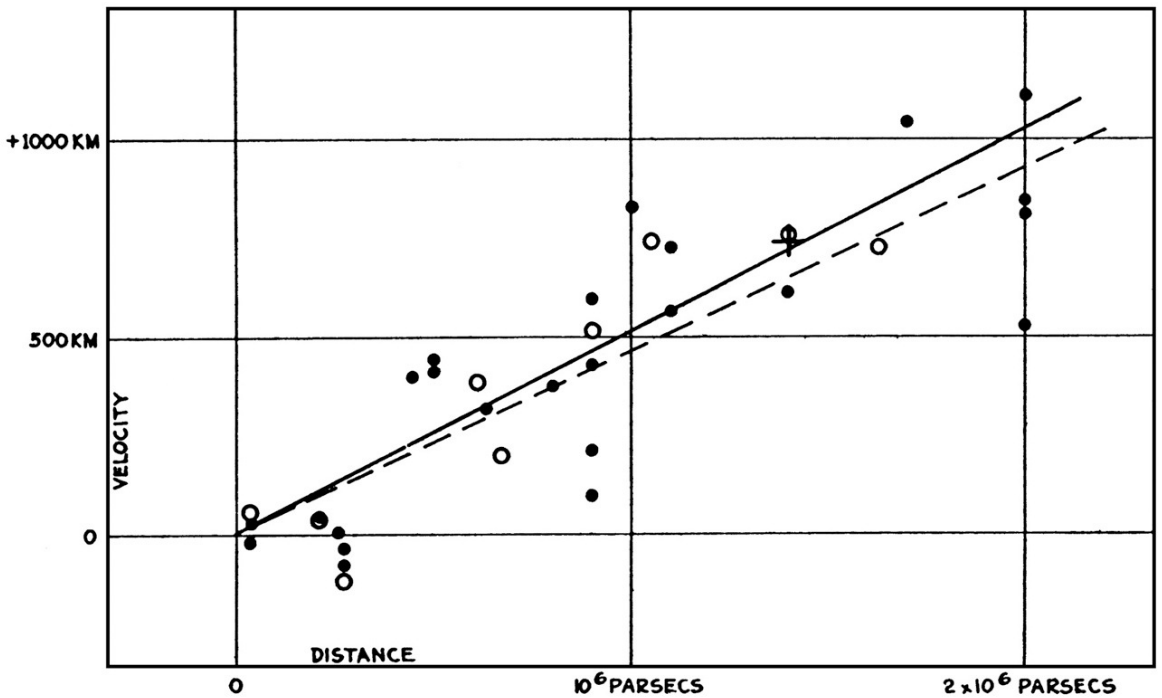
Astrostatistics: Wed 13 Feb 2017

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2019

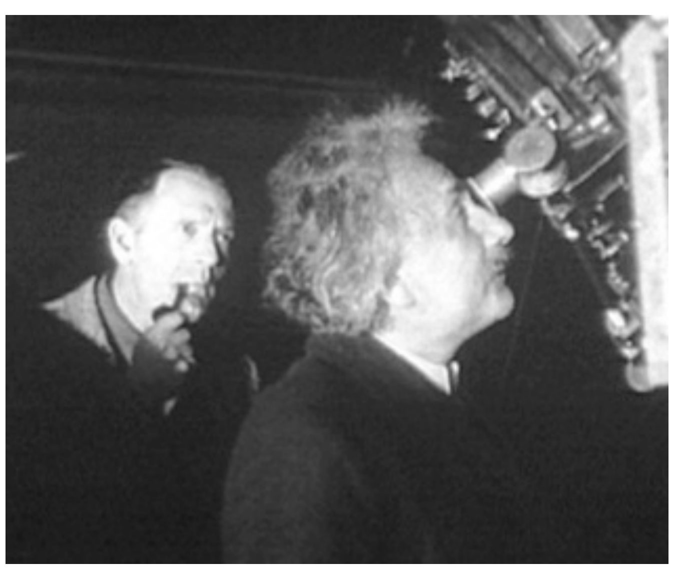
- Today: continue Bayesian computation / Monte Carlo Methods (e.g. MacKay: Ch 29-30; Bishop: Ch 11; Gelman)
    - Importance Sampling
    - Case Study: Bayesian Estimates of the Mass of the Milky Way Galaxy
- Example Sheet online, Ex Class Tue Feb 19, 1pm MR5
    - Problem on Estimating the Hubble Constant using supernova data from Dhawan et al. 2018, A&A, 609, A72

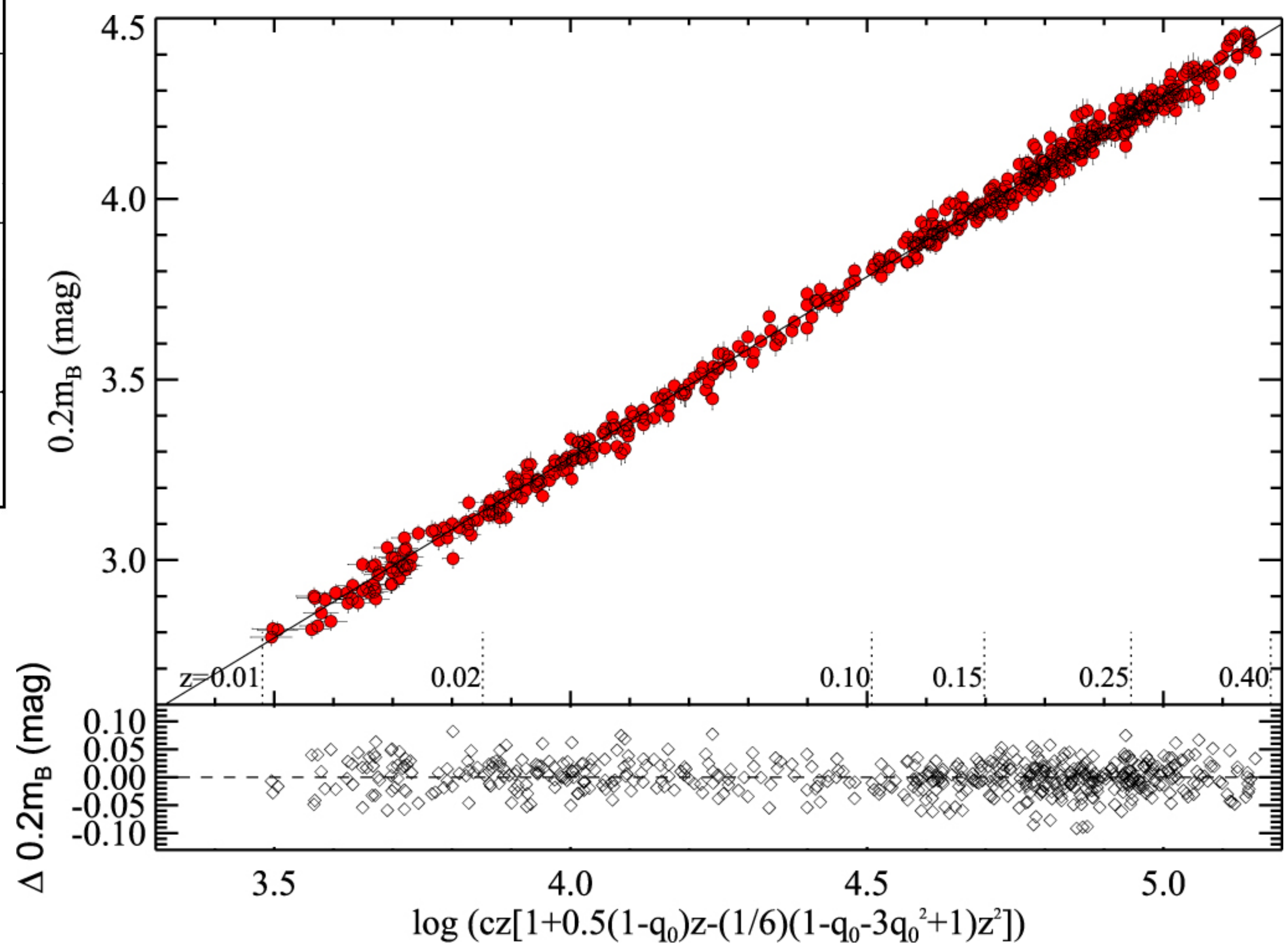
Hubble Constant

Velocity = $H_0 \times$ Distance



Hubble (1929)

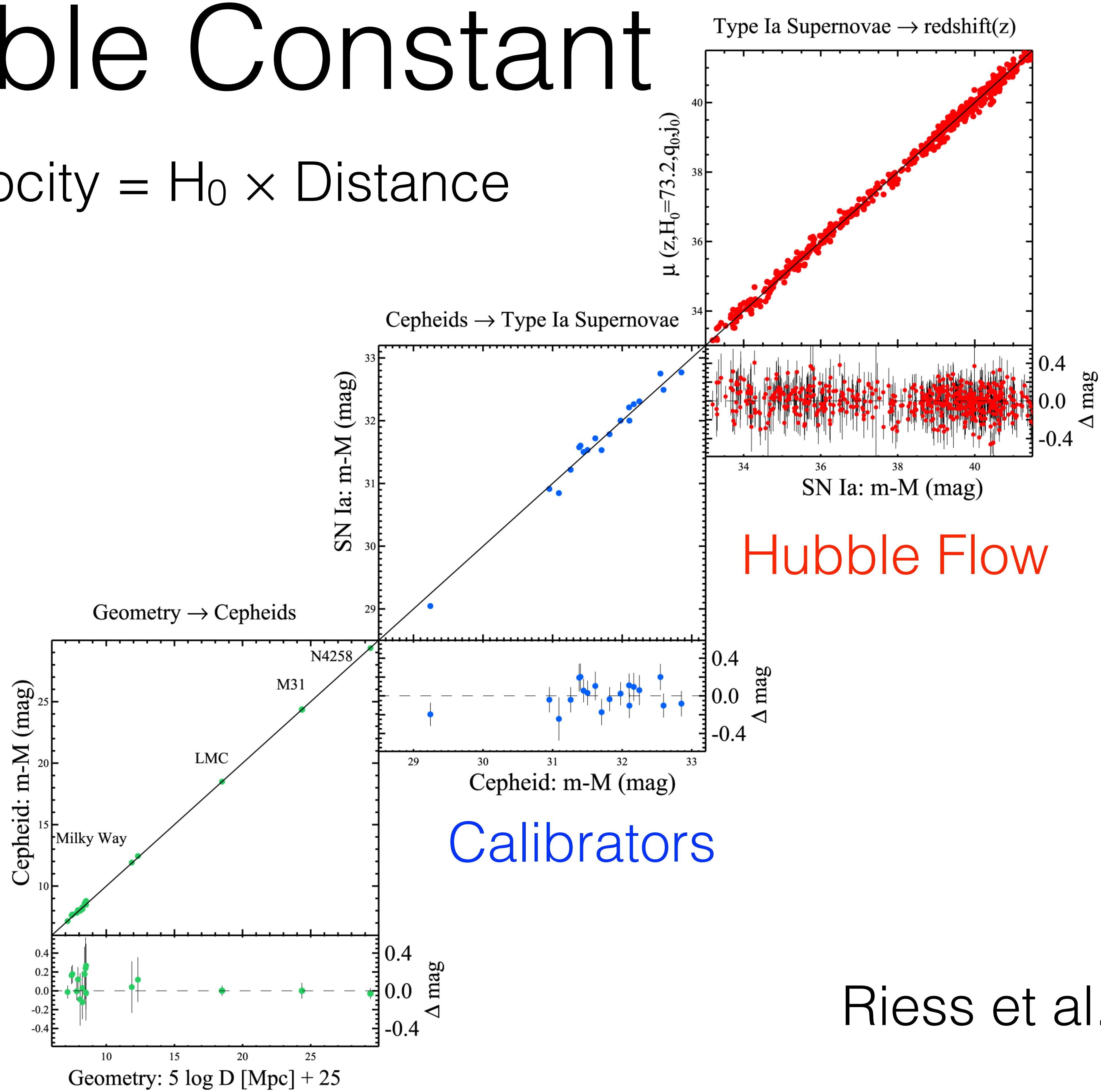




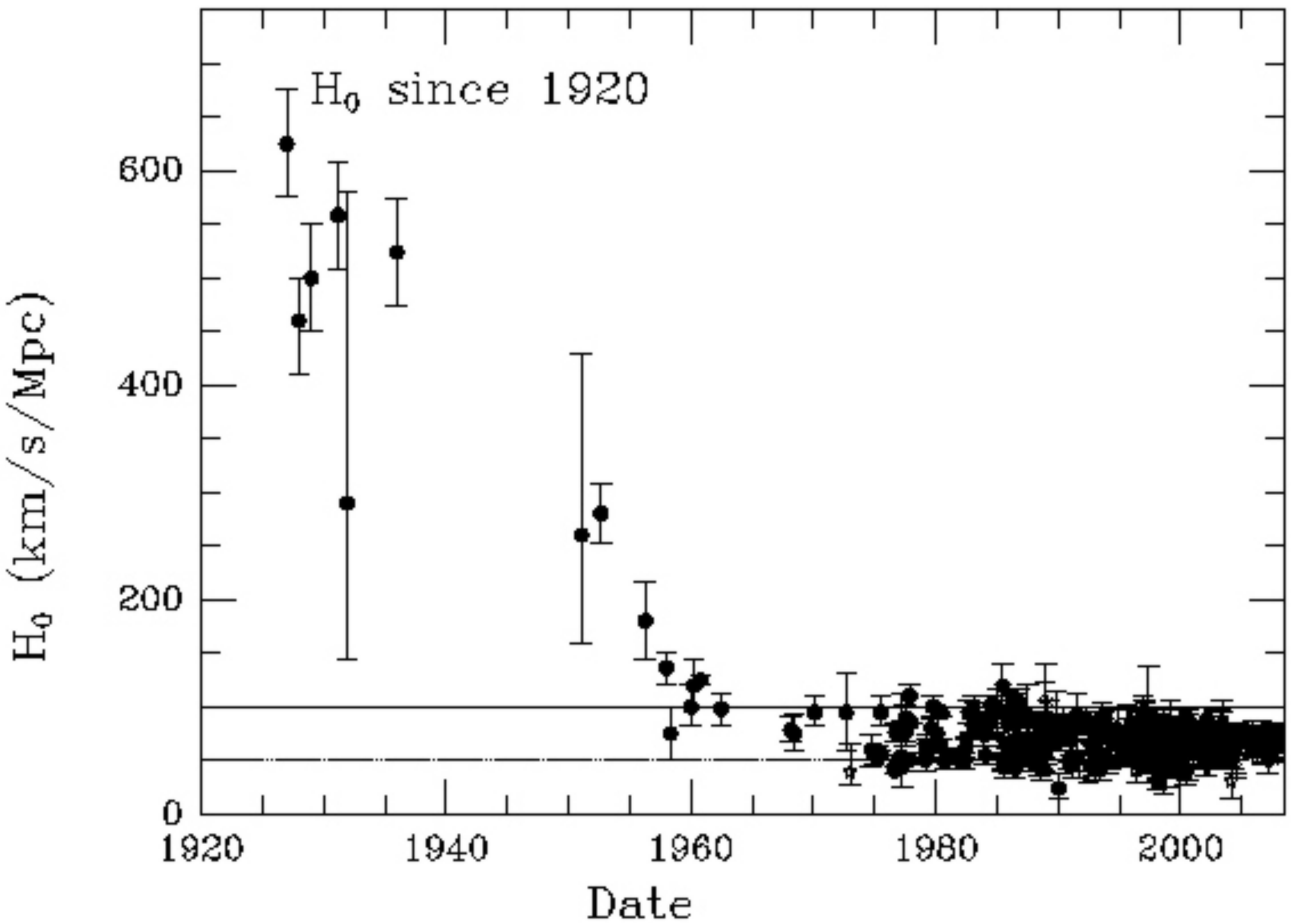
Riess et al. 2016

Hubble Constant

Velocity = $H_0 \times$ Distance

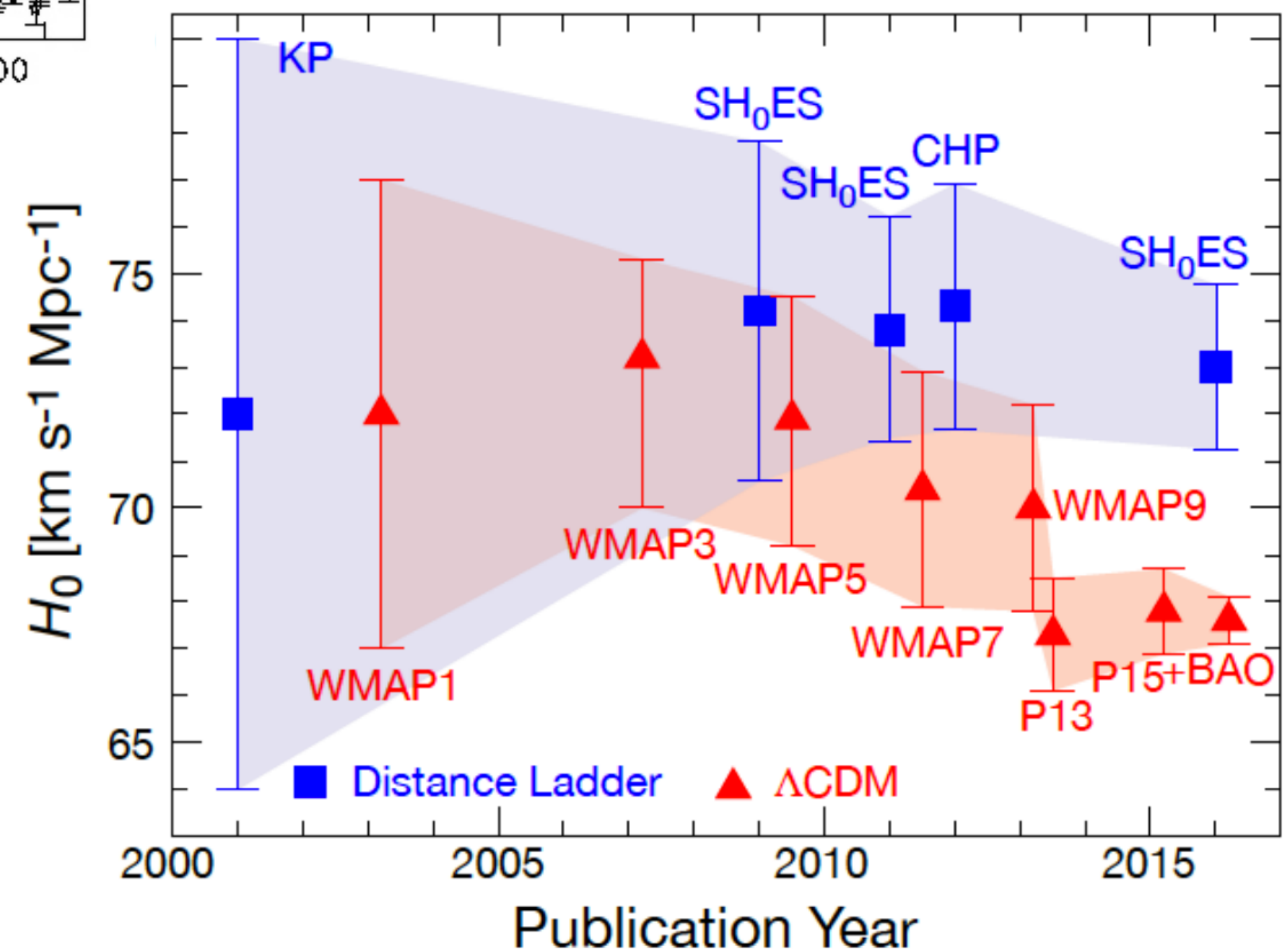


Riess et al. 2016

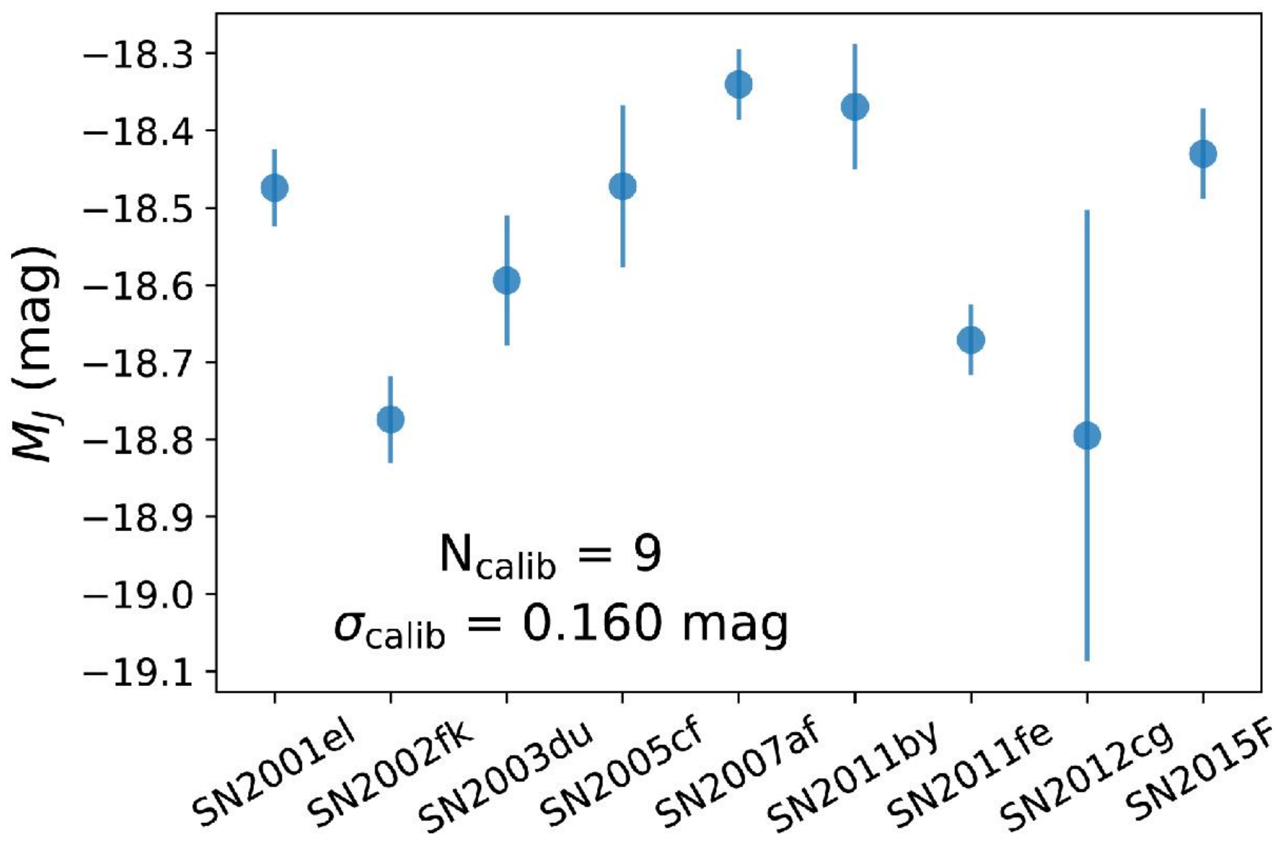


The Hubble Constant (estimates) Over Time

Not So Constant!

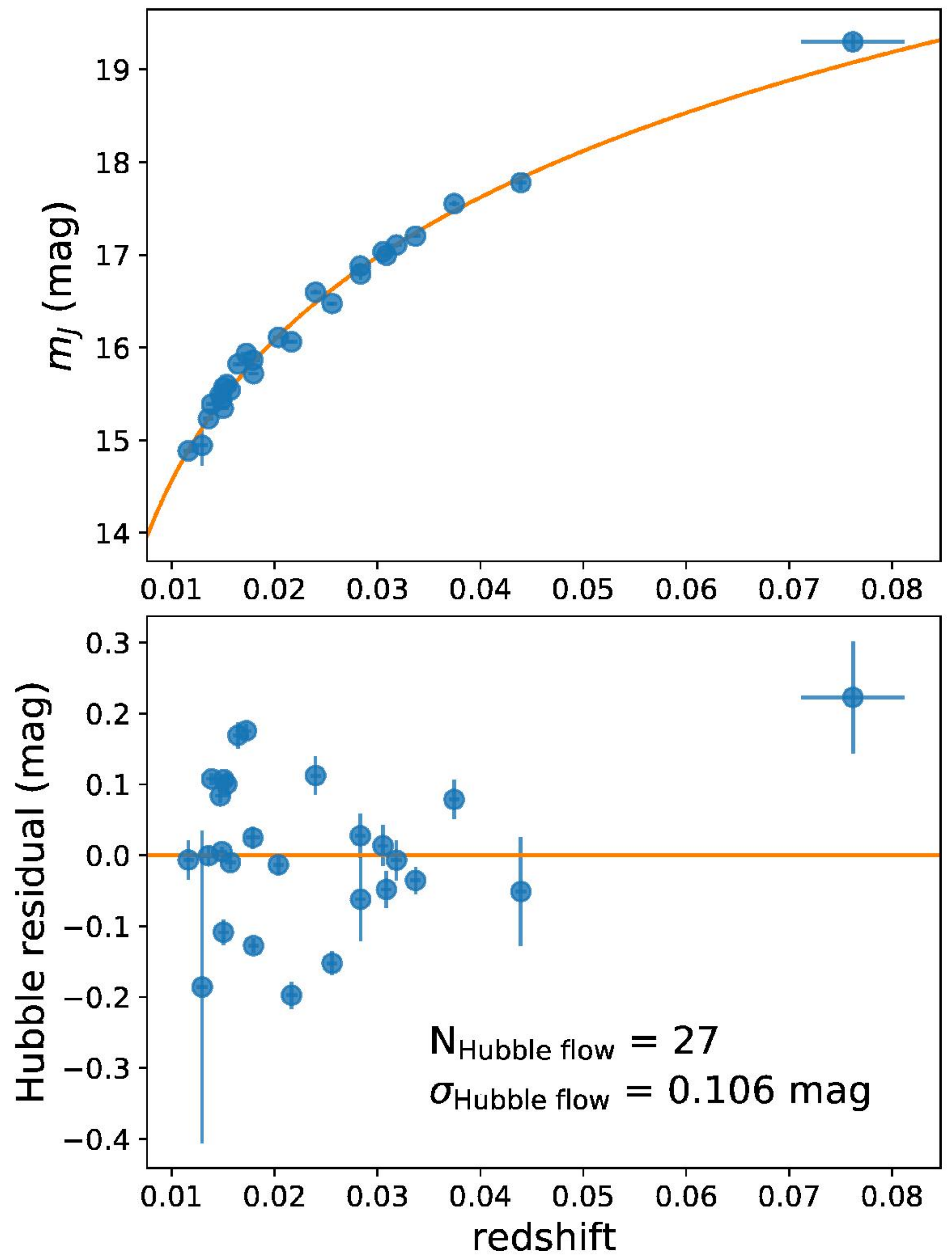


Example sheet: Estimating H_0 using supernovae



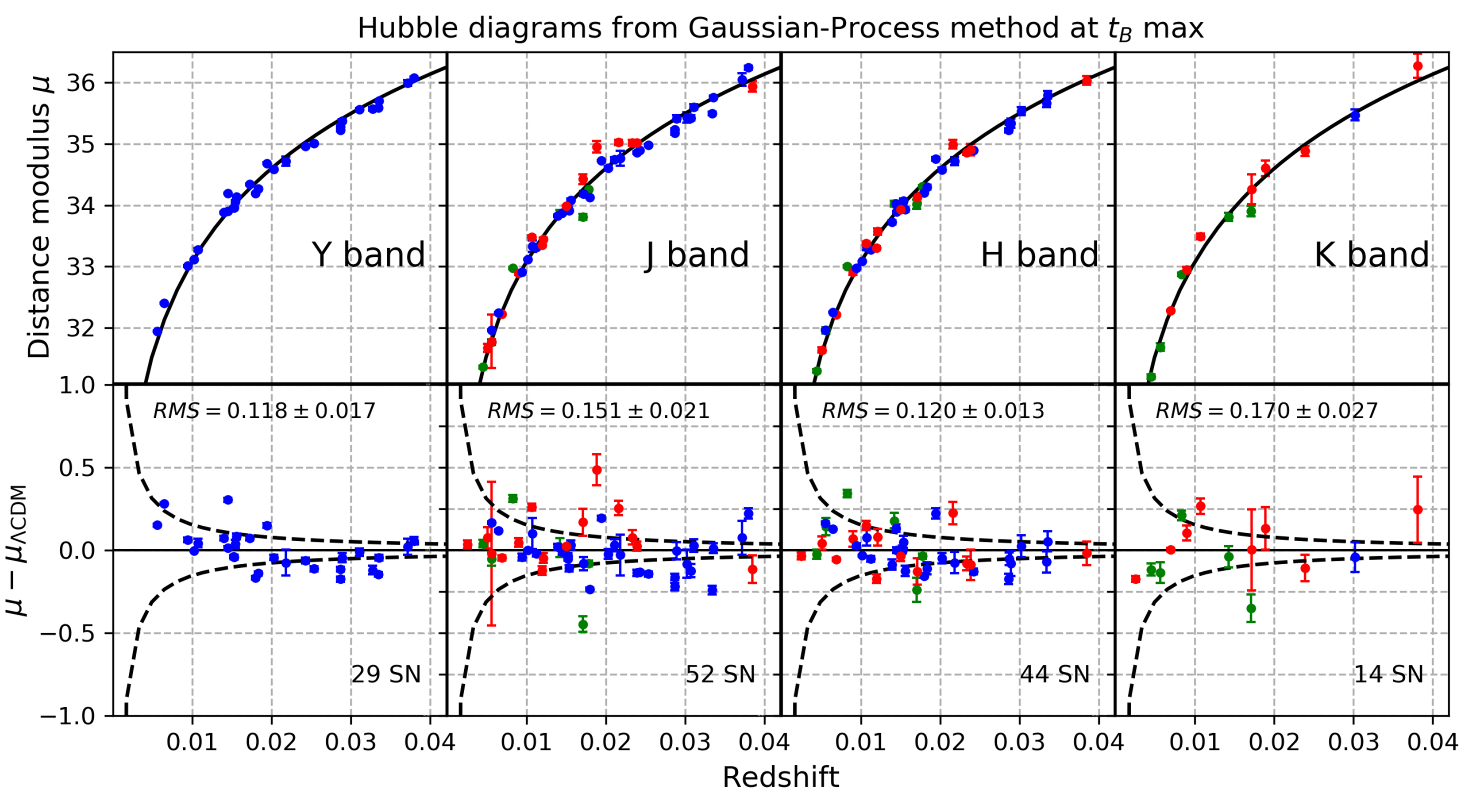
Calibrator Sample

"Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles."
Dhawan, Jha, Leibundgut 2018

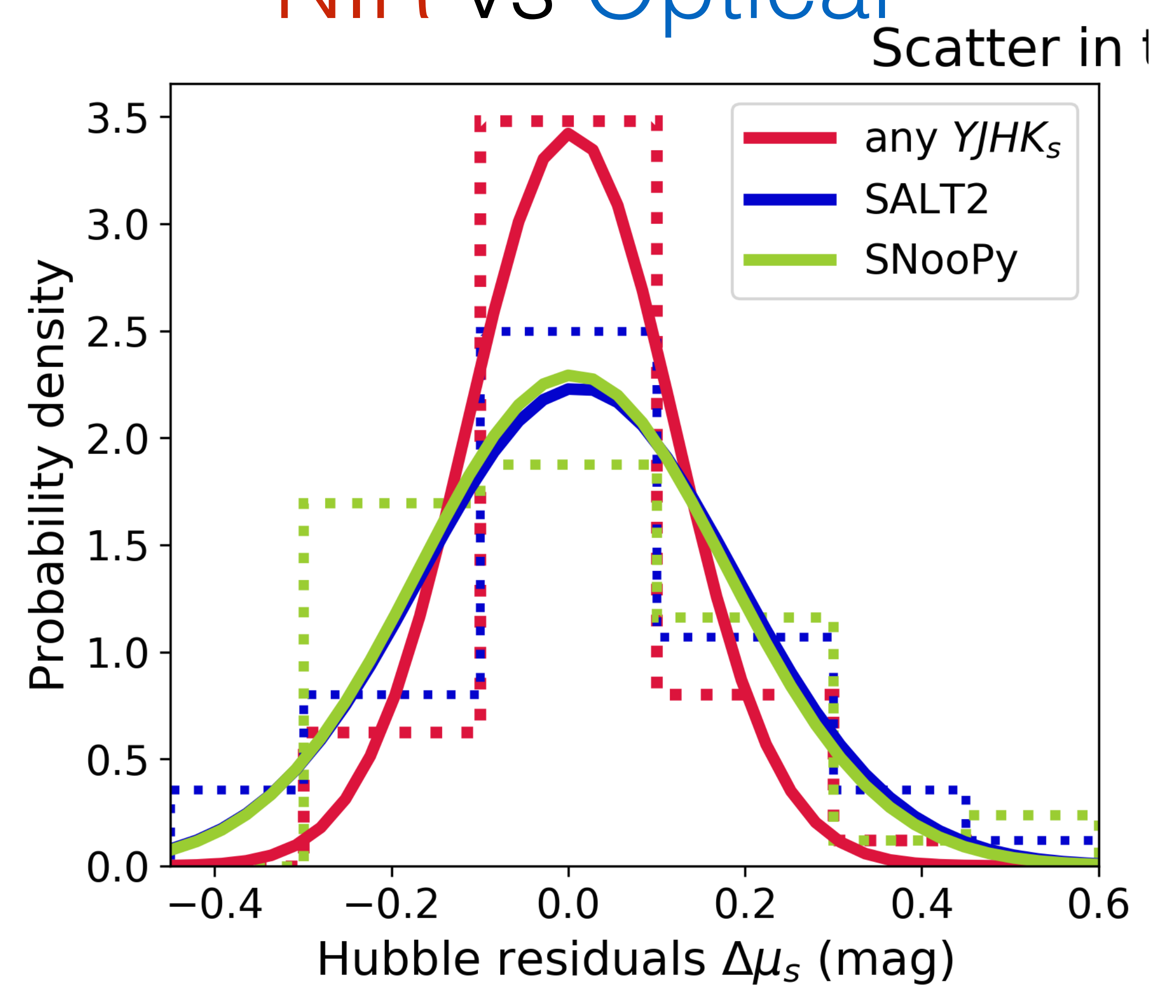


Hubble Flow Sample

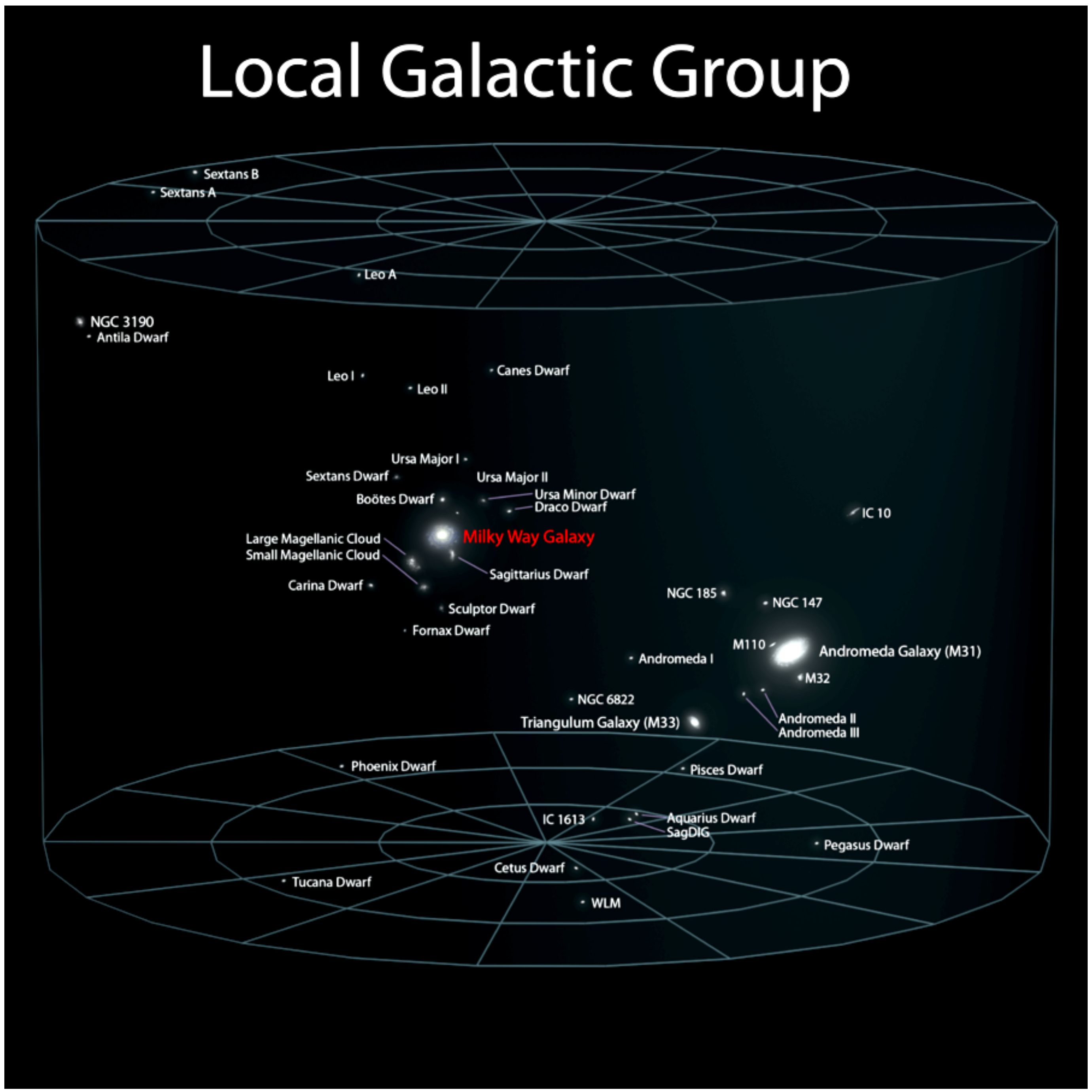
New Paper!
"Type Ia supernovae are excellent standard candles in the near-infrared."
Avelino, Friedman, Mandel+2019: arXiv:1902.03261

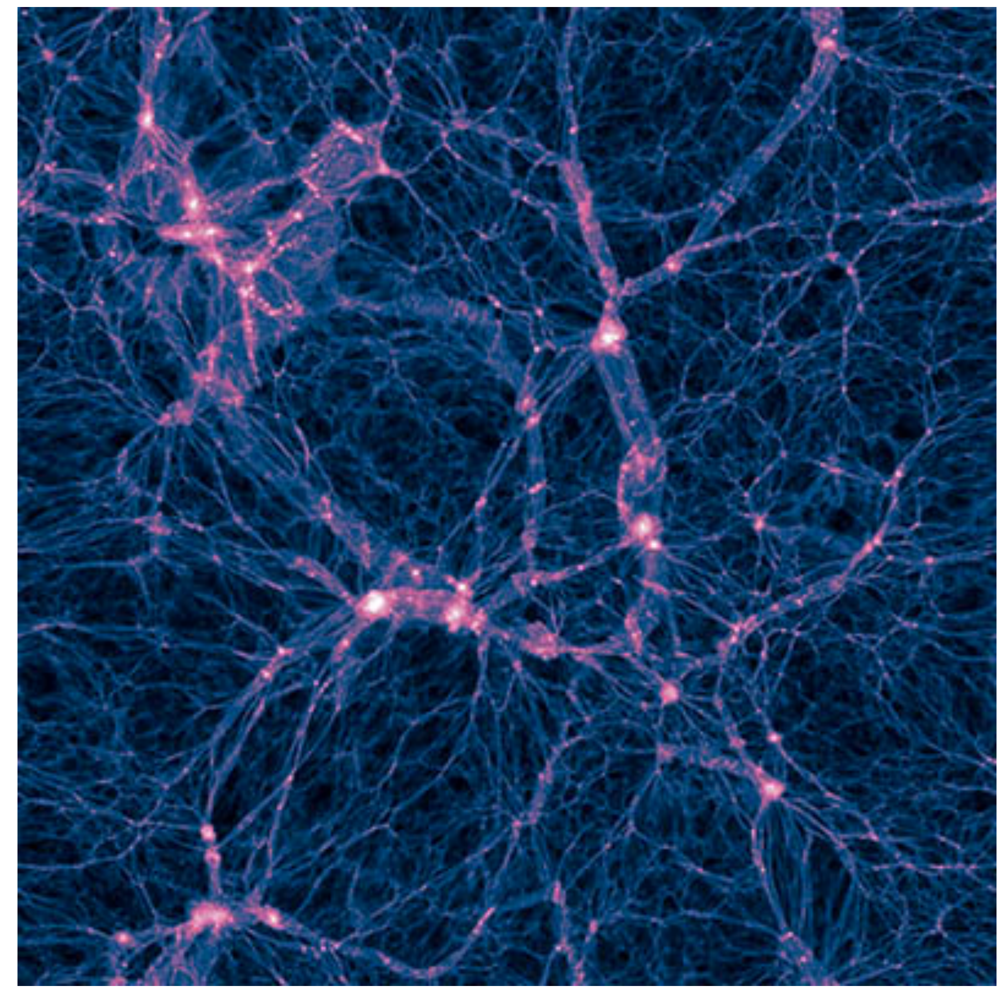


NIR vs Optical



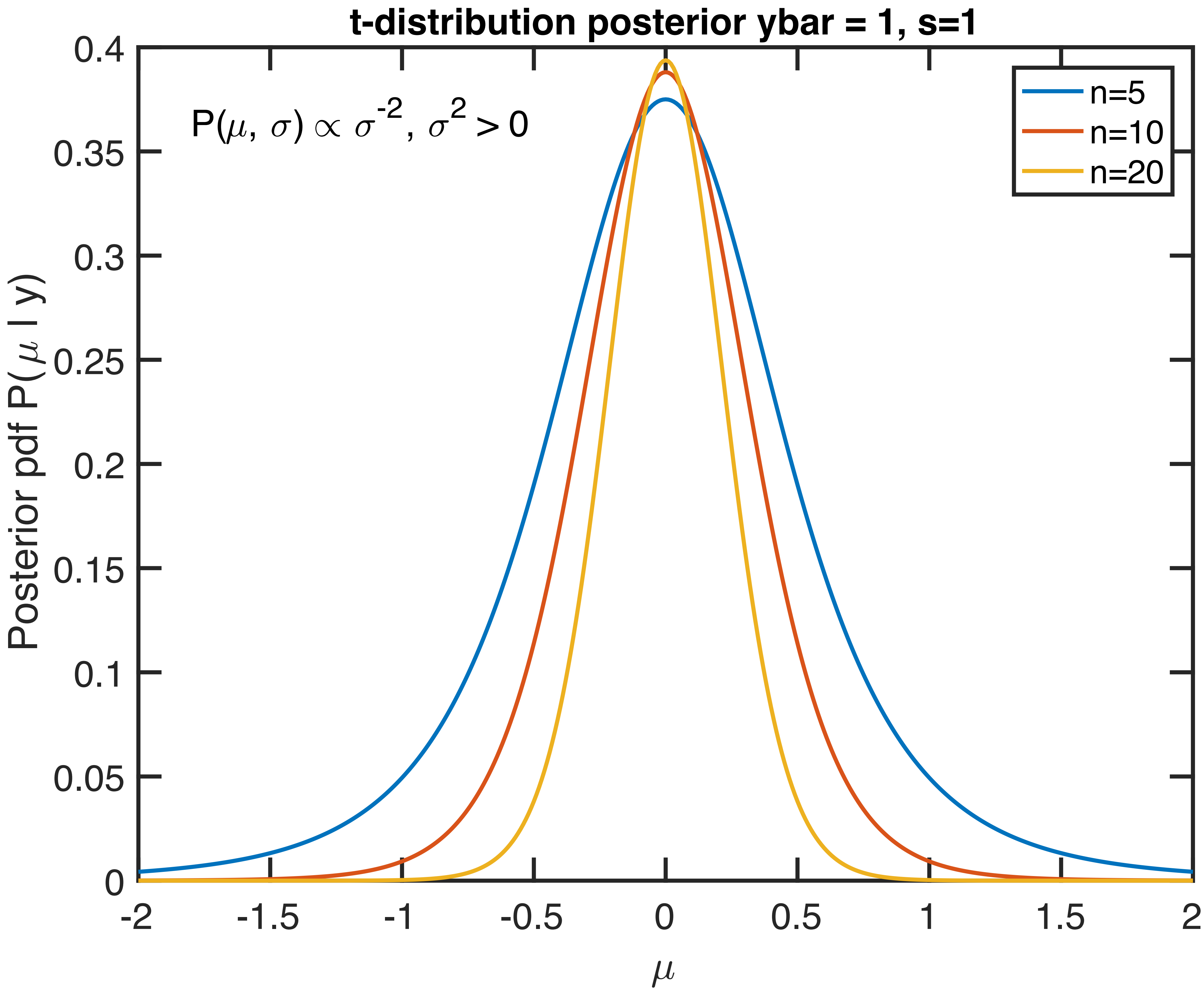
Later Today: Astrostatistics Case Study:
Bayesian estimates of the Milky Way Galaxy mass using high-precision astrometry and cosmological simulations
(Patel, Besla, & Mandel, 2017, 2018, arXiv:1703.05767, 1803.01878





Illustris
Cosmological Simulation of Galaxy Formation

Last Time:
Analytic Posterior Density for a Gaussian(μ, σ²) model



Monte Carlo Integration

Typically, we want to summarise the posterior and compute expectations of the form:

$$I = \mathbb{E}[f(\boldsymbol{\theta})|\boldsymbol{D}] = \int f(\boldsymbol{\theta})\, P(\boldsymbol{\theta}|\, \boldsymbol{D})\, d\boldsymbol{\theta}$$

Using m samples from the posterior:

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

$$\hat{I} = \frac{1}{m}\sum_{i=1}^{m} f(\boldsymbol{\theta}_i) \longrightarrow \boldsymbol{I} \quad \text{(LLN for large m)}$$

Monte Carlo Error:

$$\mathrm{Var}[\hat{I}] = \frac{1}{m^2}\sum_{i=1}^{m}\mathrm{Var}[f(\boldsymbol{\theta})] = \frac{1}{m}\mathrm{Var}[f(\boldsymbol{\theta})] \approx \frac{1}{m}\widehat{\mathrm{Var}}[\{f(\boldsymbol{\theta}_i)\}]$$

Bayesian computation using sampling: Fundamental theorem of Monte Carlo

Posterior Expectation | Sample Average

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

Examples:

Posterior Mean μ $\quad f(\boldsymbol{\theta}) = \boldsymbol{\theta}$

Posterior Variance $\quad f(\theta) = (\theta - \mu)^2$

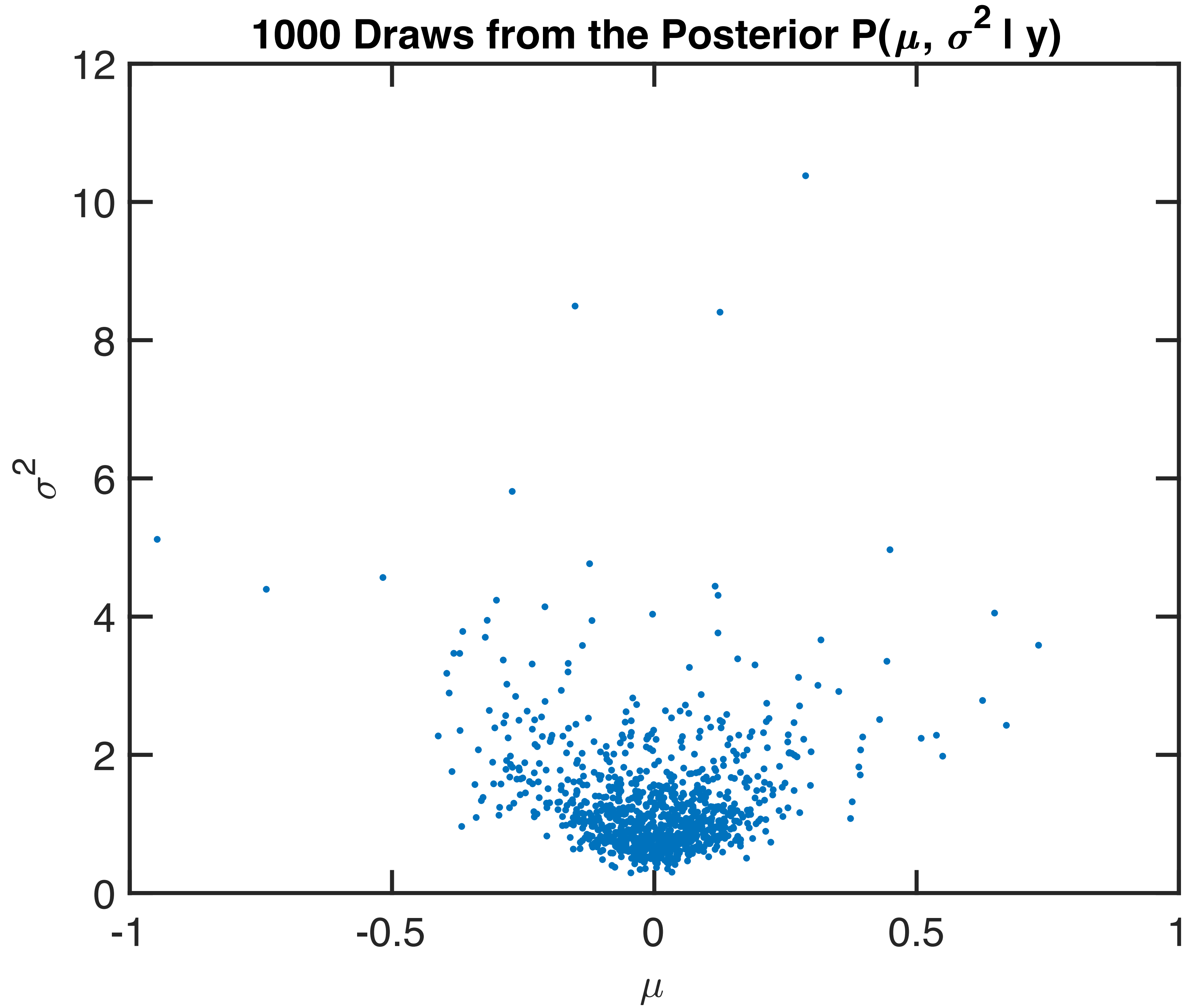
Posterior Probability in an interval $[a, b]$ $\quad f(\theta) = I_{[a,b]}(\theta)$

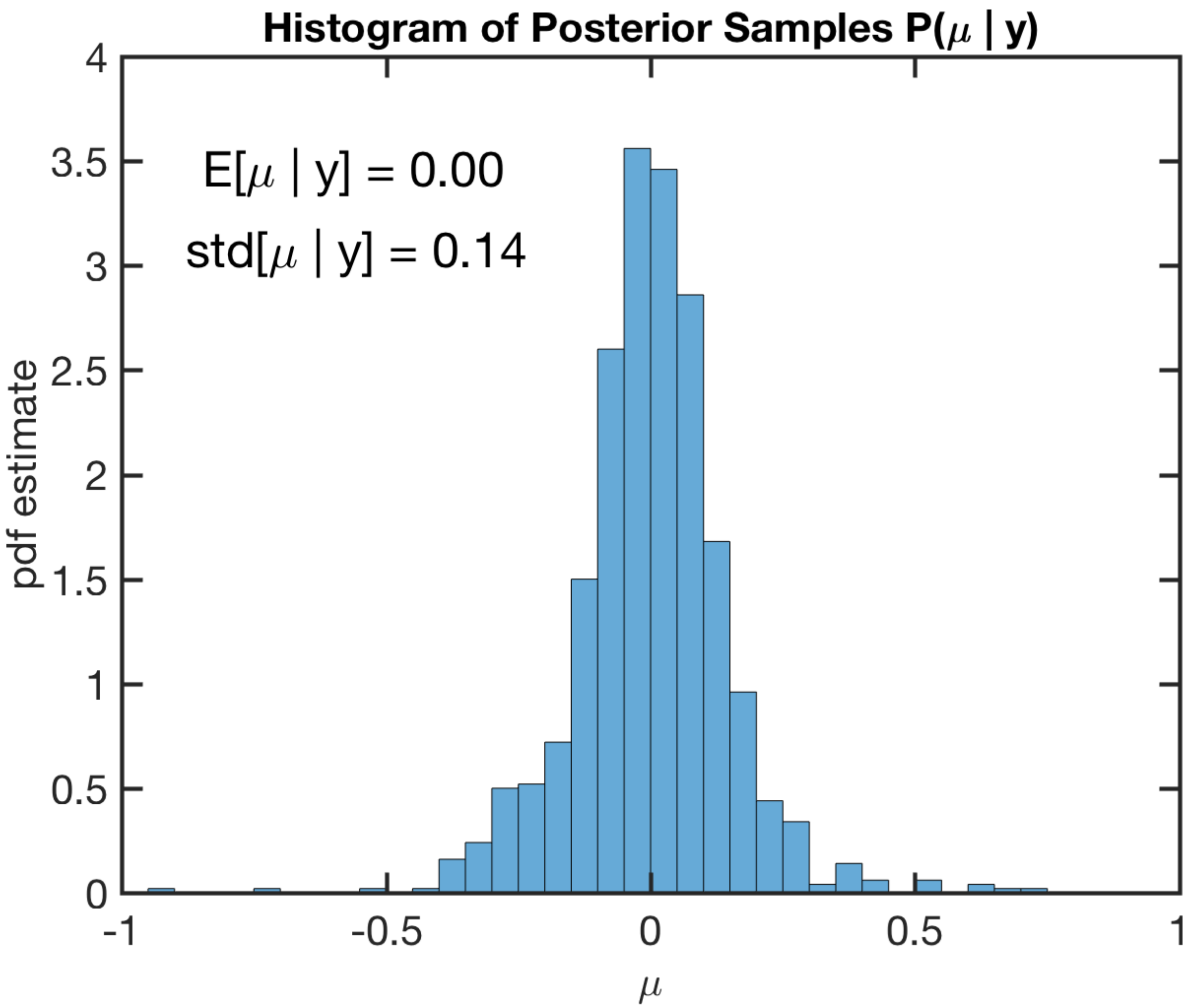
(indicator function)

Monte Carlo Direct Sampling

Factorise Posterior: $P(\mu, \sigma | \boldsymbol{y}) = P(\mu | \sigma^2, \boldsymbol{y}) \times P(\sigma^2 | \boldsymbol{y})$

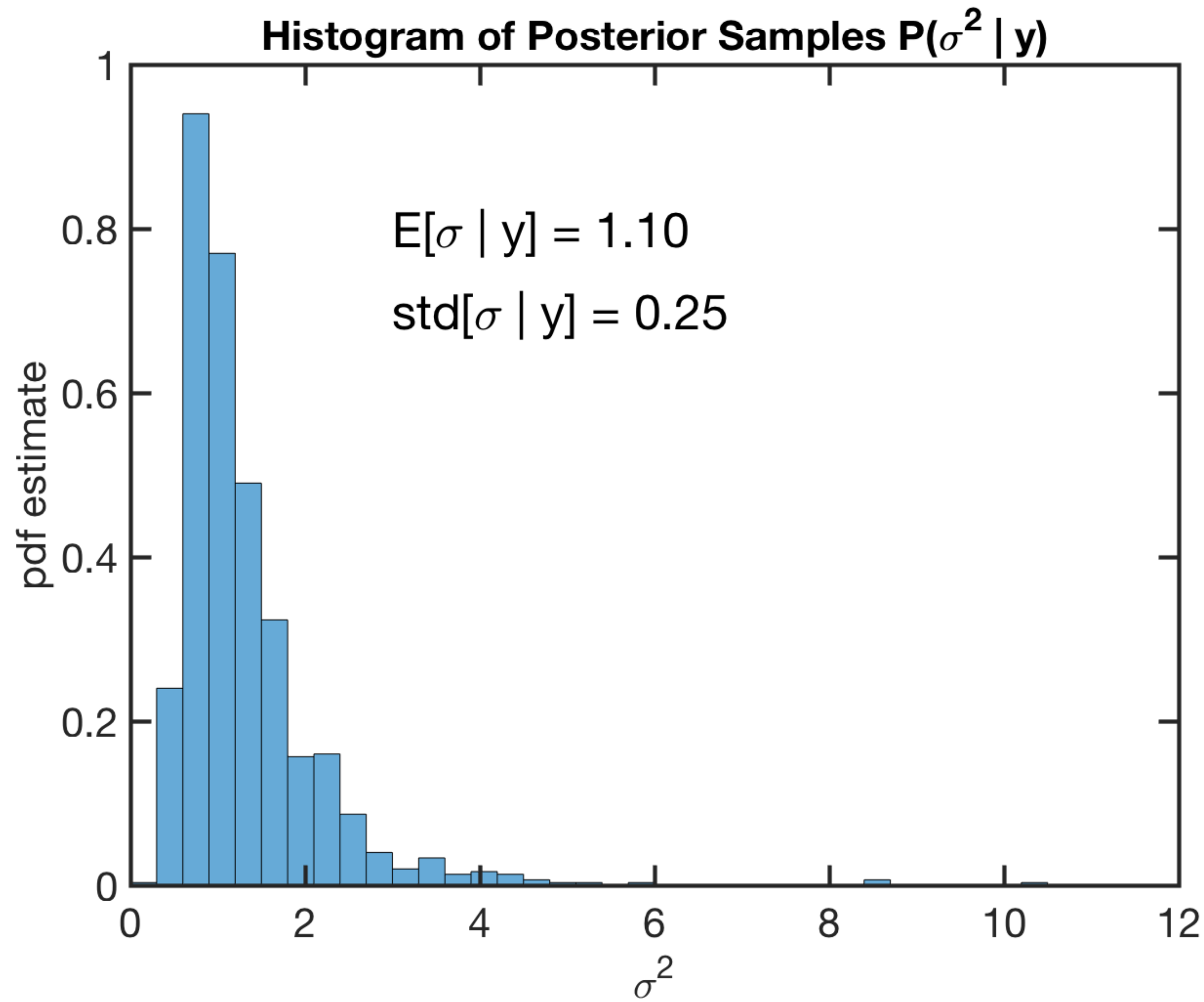
1. $\sigma_i^2 \sim P(\sigma^2 | \boldsymbol{y})$ $[\text{Inv-}\chi^2]$
2. $\mu_i | \sigma_i^2 \sim P(\mu | \sigma^2, \boldsymbol{y})$ $[\text{Normal}]$



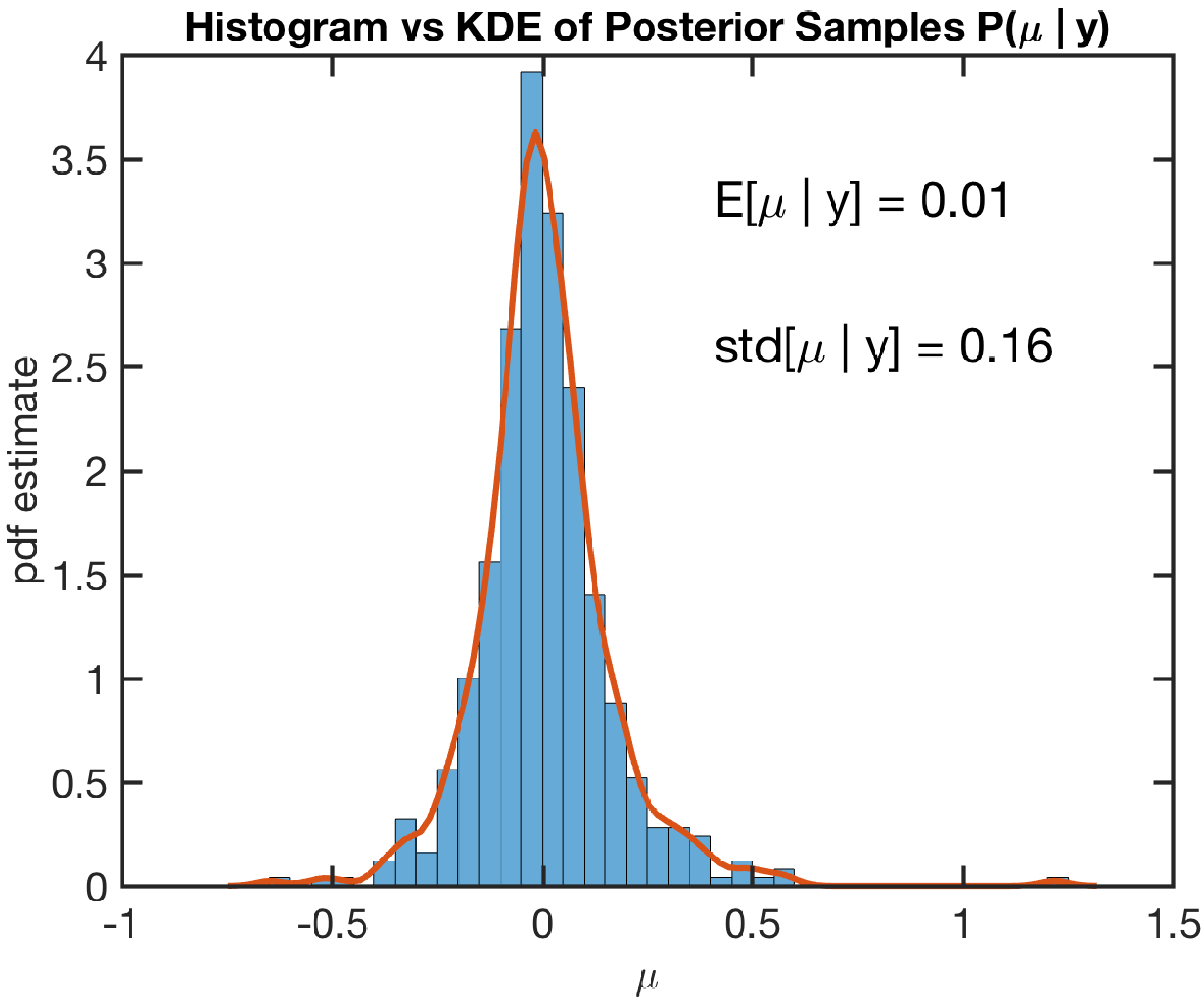


Marginal Posterior (μ)

Marginal Posterior (σ^2)



Kernel Density Estimate = estimate a smooth density from samples



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

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

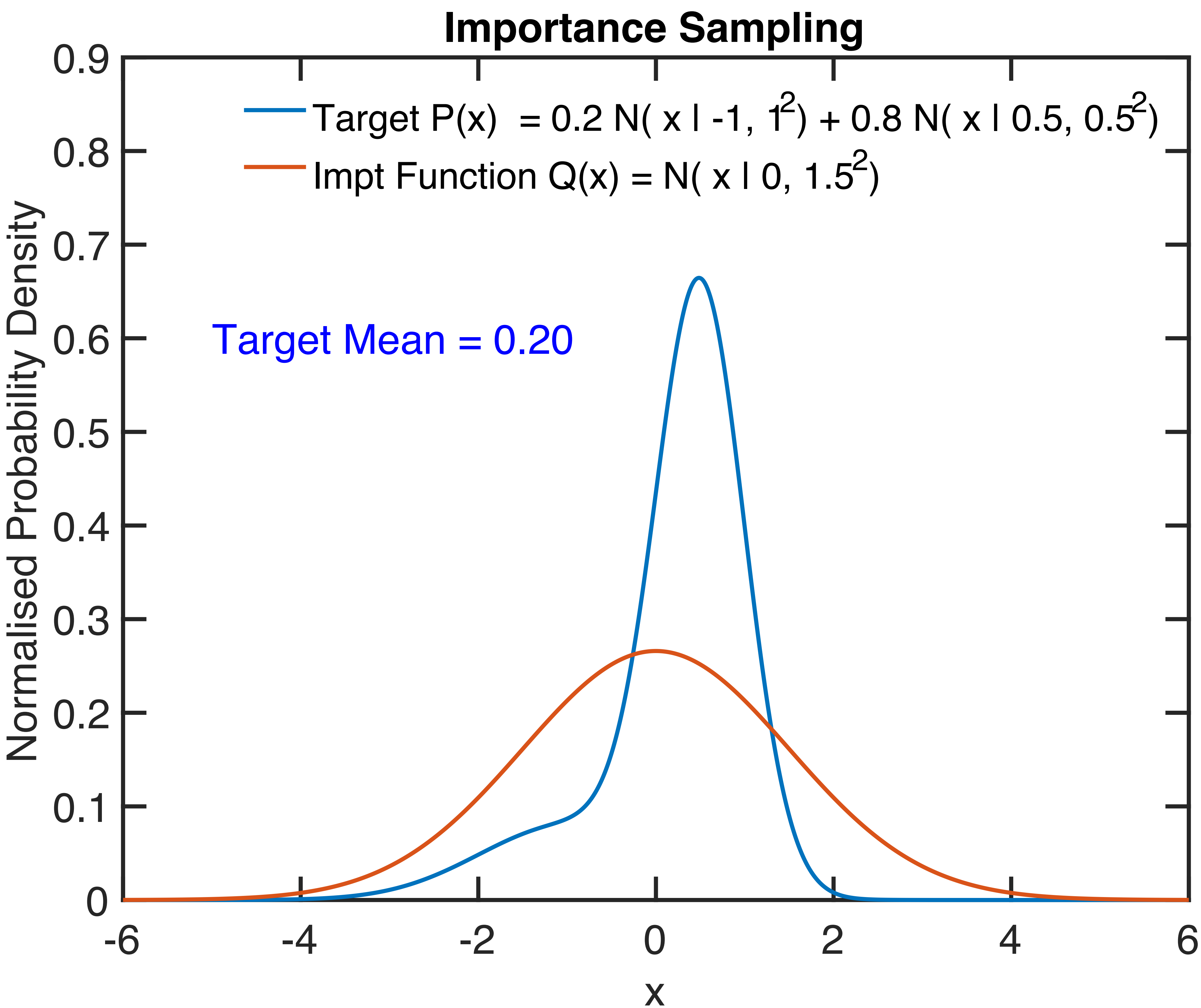
- Posterior simulation - Markov Chain Monte Carlo, Nested Sampling, etc. generates draws

- Importance Sampling - draw from an easier ("tractable") distribution (importance function) $\theta_i \sim Q(\theta)$ and weight the samples by

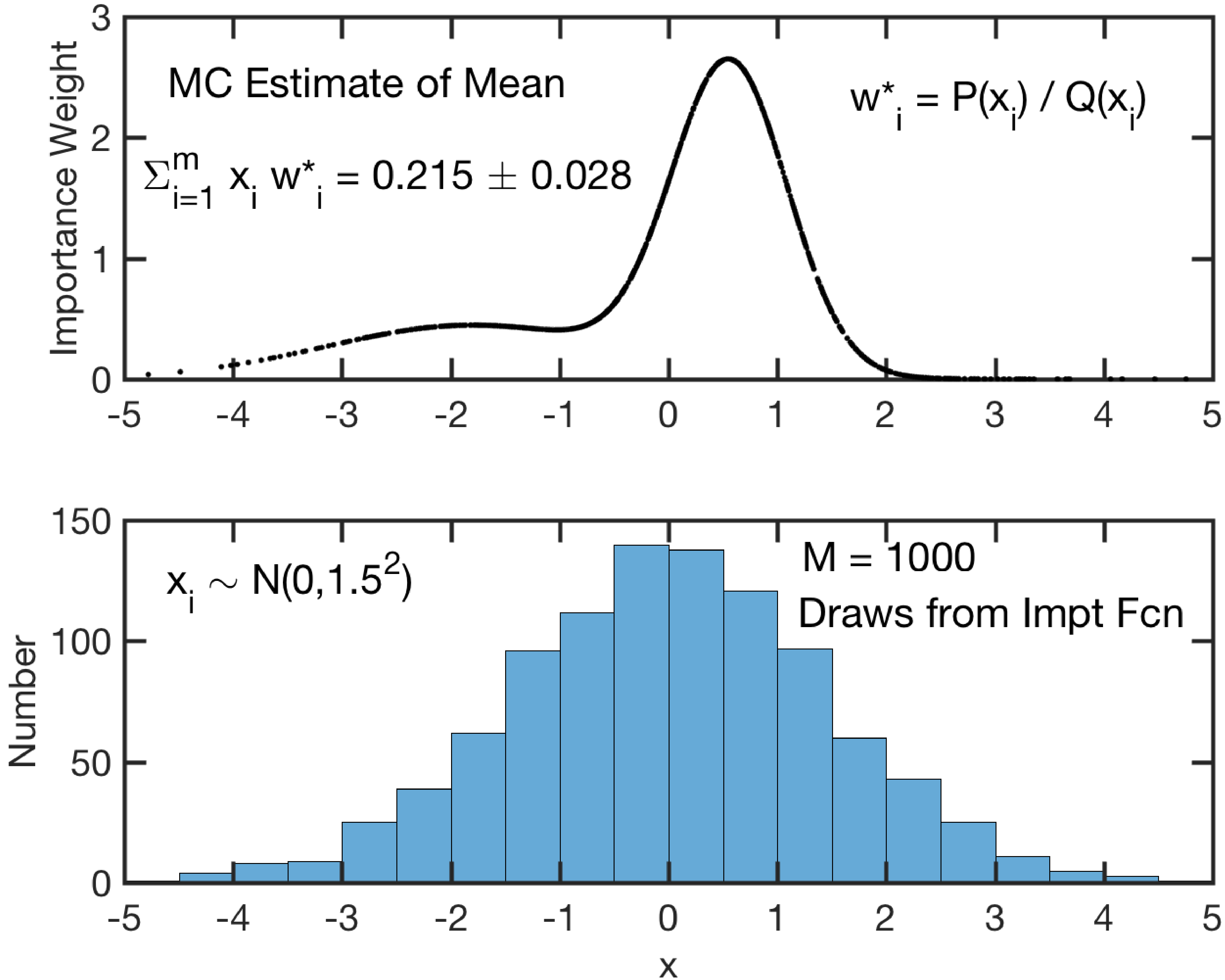
  $w_i = P(\theta_i | D) / Q(\theta_i)$

Two Cases of Importance Sampling

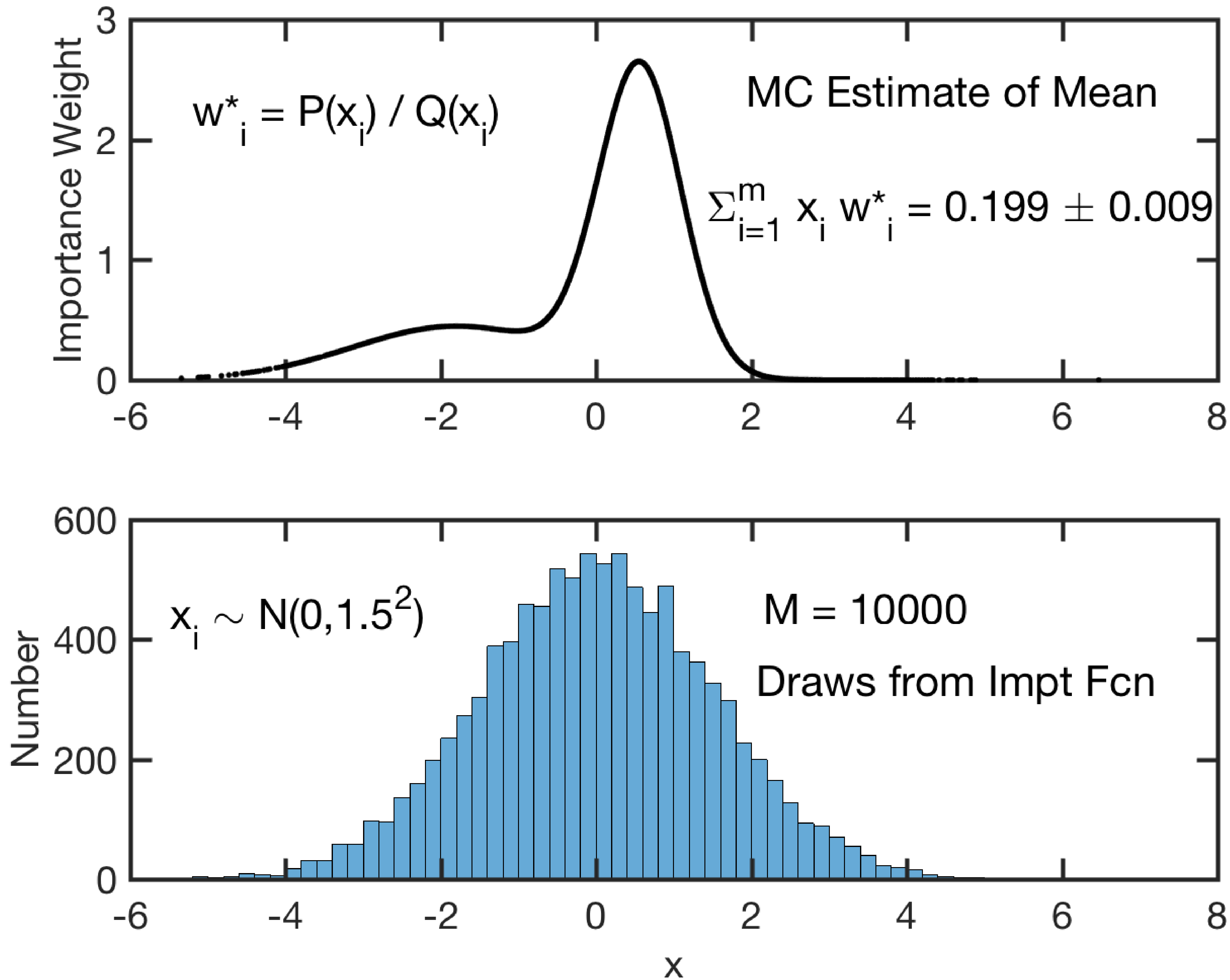
Contrived Example [Gaussian Mixture, Normalised PDF]



Importance Sampling Example
m = 1,000 Draws

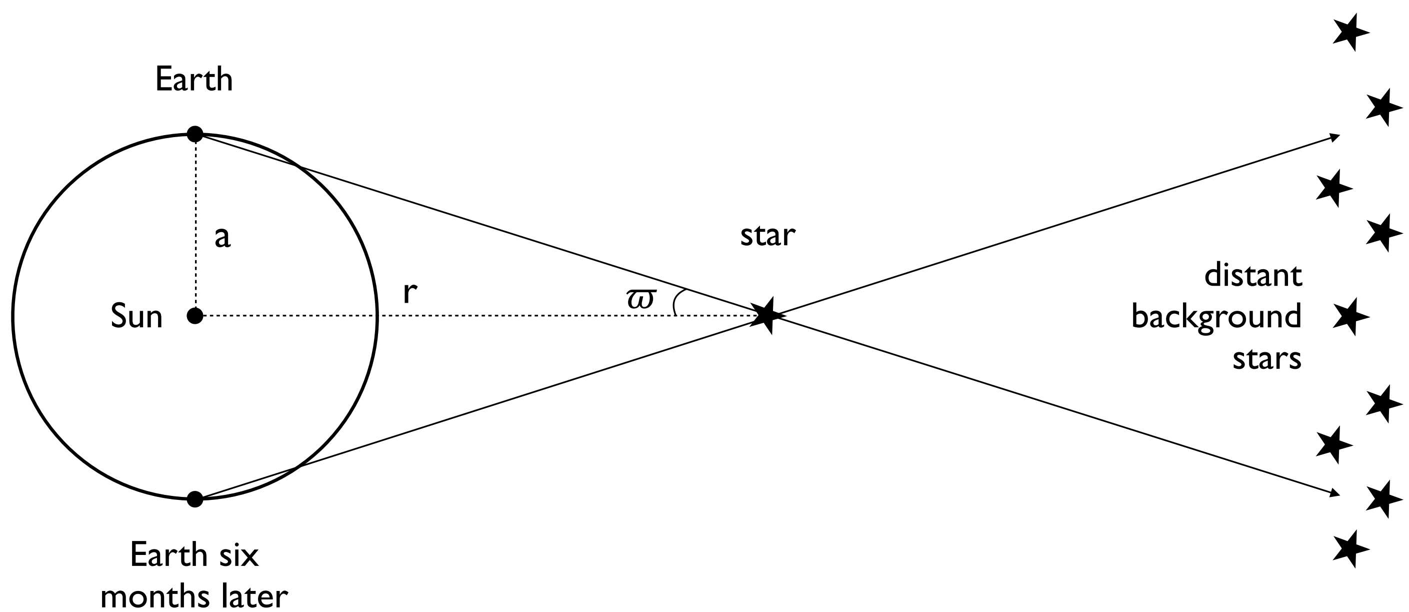


Importance Sampling Example
m = 10,000 Draws



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 ϖ 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 ϖ is 1 arcsecond, r is defined as the *parsec*, which is about 3.1×10^{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:

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

Introducing physical constraints into the prior

Exponential decrease in stellar density with Galactic length scale L

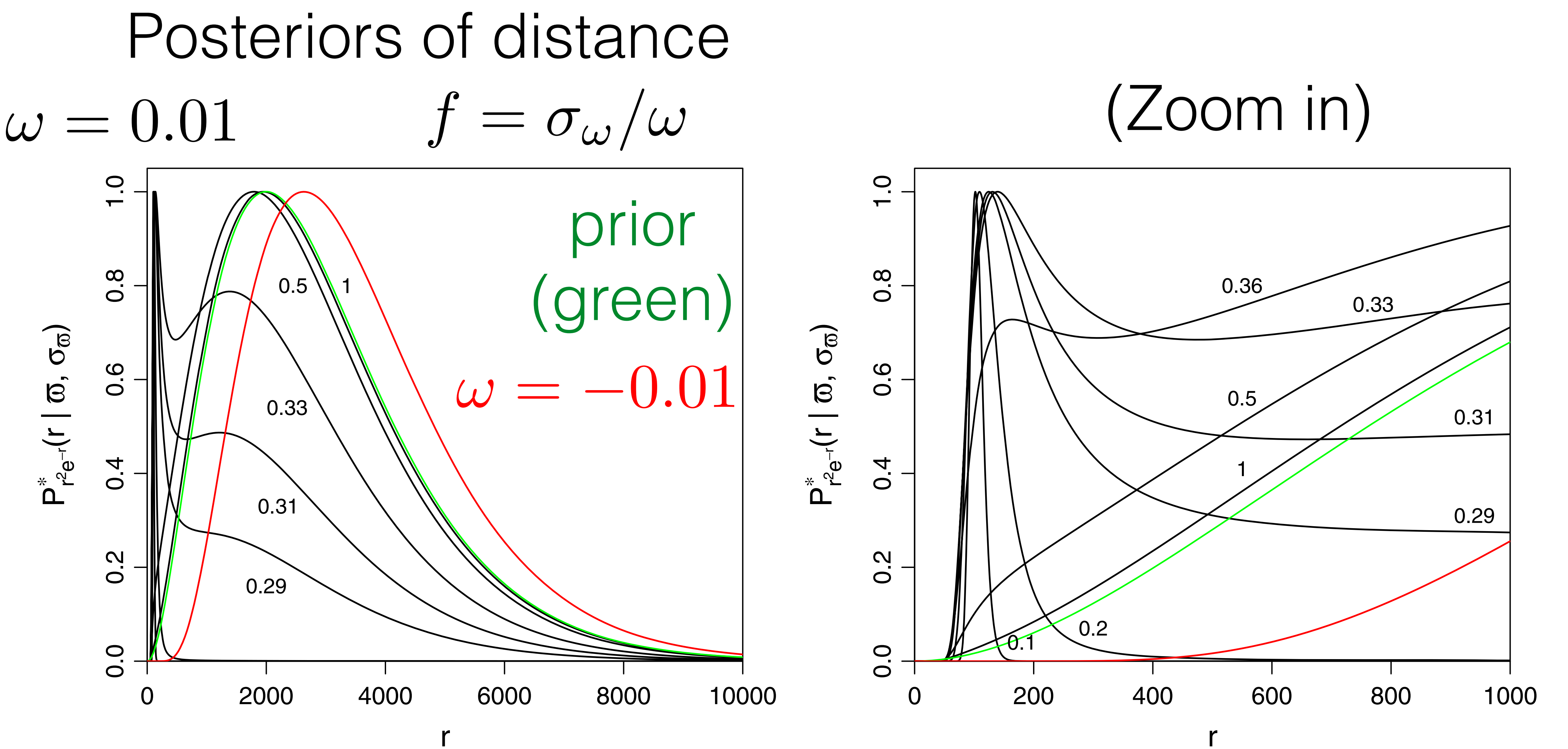
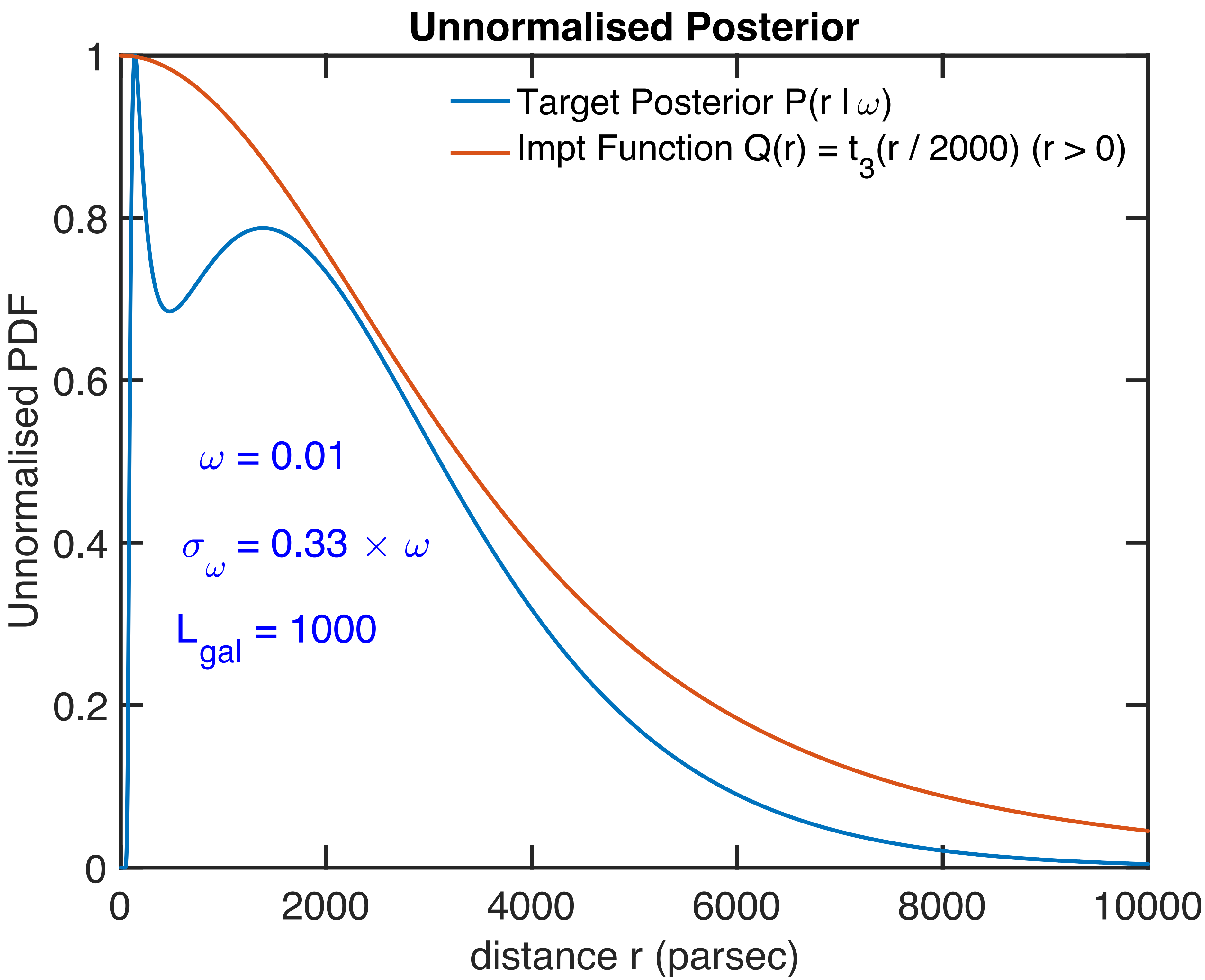


FIG. 12.—The *black lines* in the left panel show the unnormalized posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) for $L = 10^3$, $\varpi = 1/100$ and seven values of $f = (0.1, 0.2, 0.29, 0.31, 0.33, 0.5, 1.0)$. The *red line* is the posterior for $\varpi = -1/100$ and $|f| = 0.25$. The *green curve* is the prior. The right panel is a zoom of the left one and also shows an additional posterior for $f = 0.36$. All curves have been scaled to have their highest mode at $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi) = 1$ (outside the range for some curves in the right panel). See the electronic edition of the *PASP* for a color version of this figure.

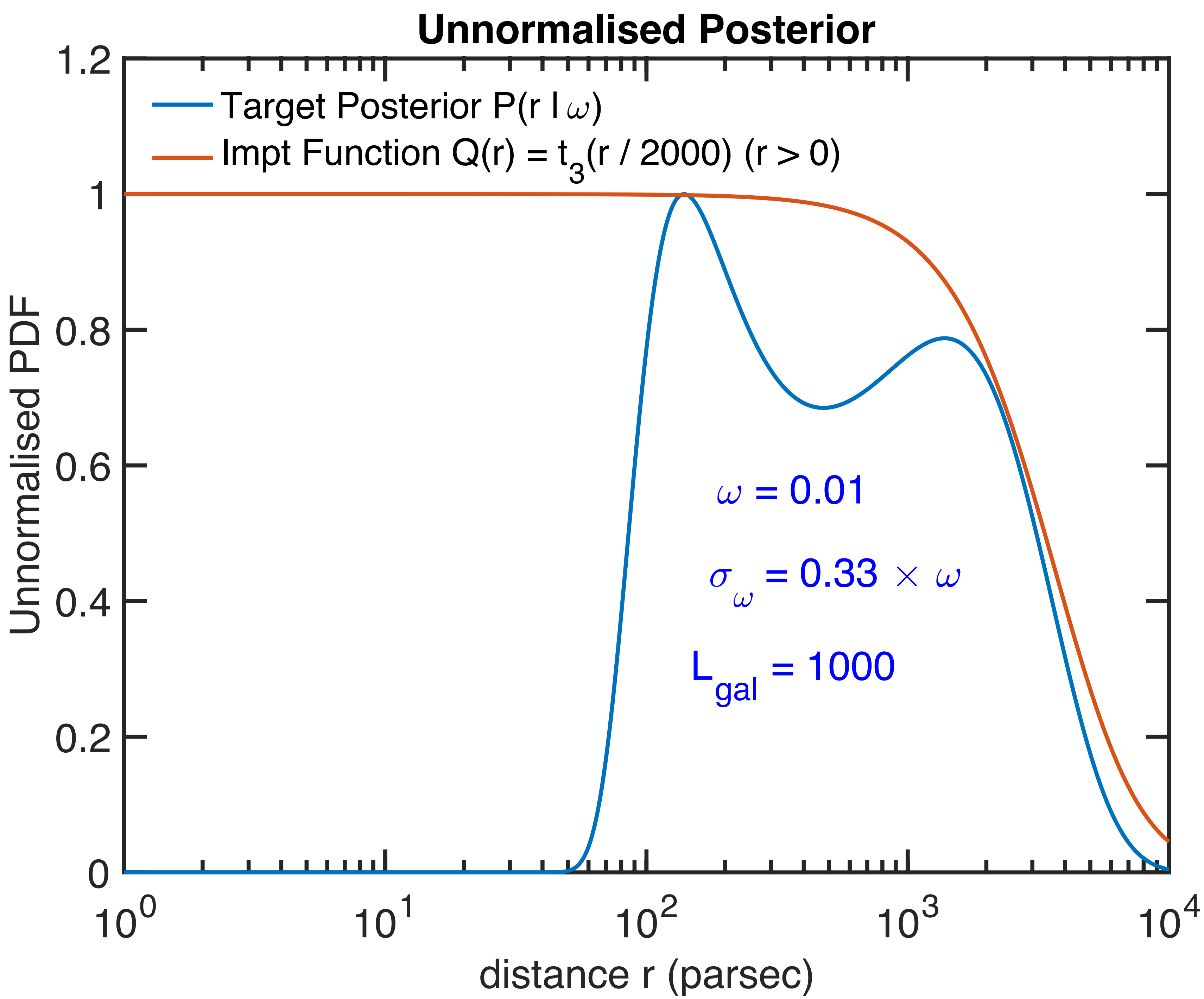
Parallax Example



Importance Function is a Student t with v = 3

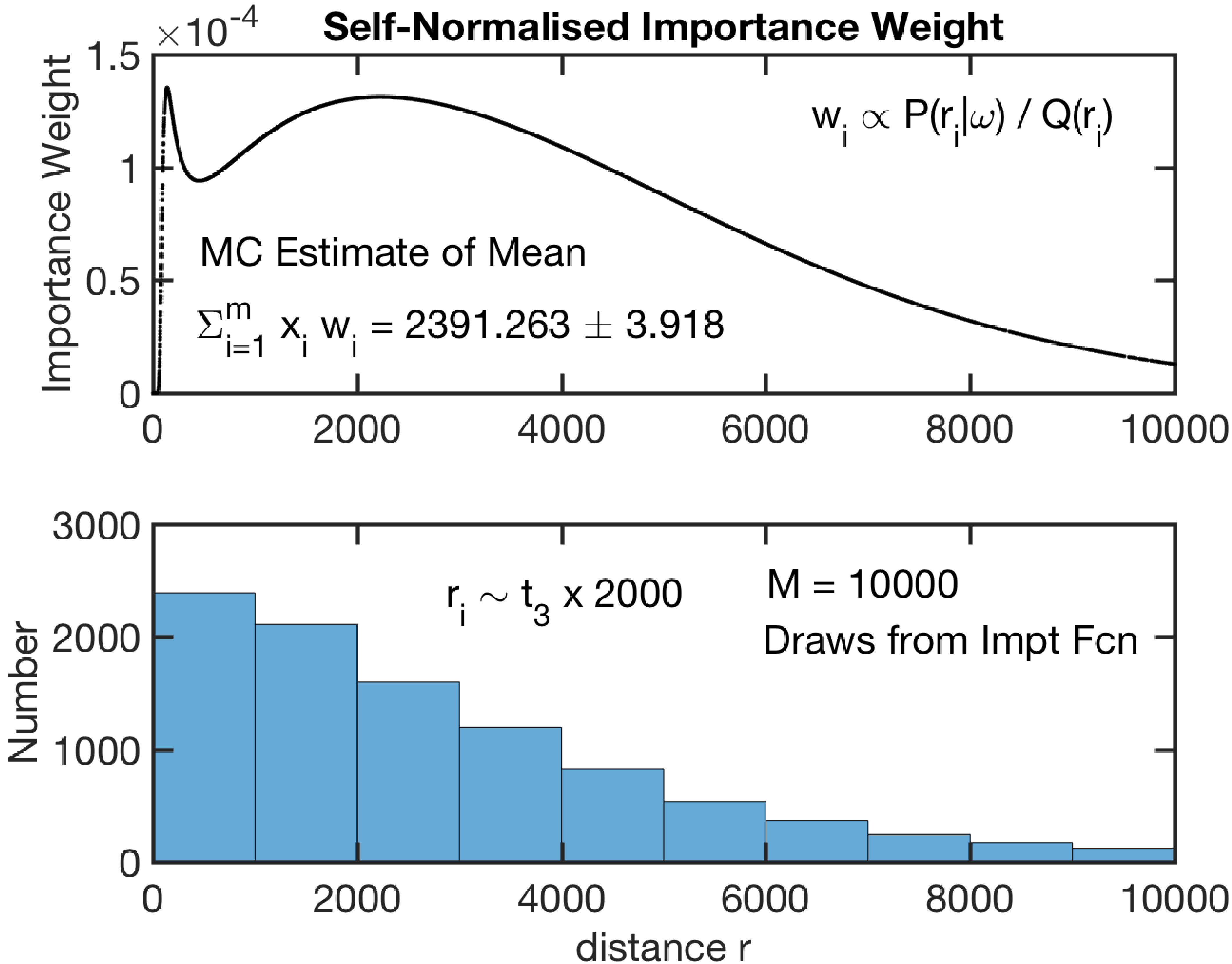
Parallax Example

Importance Function is a Student t with v = 3



Log Scale

Parallax Example m = 10^4



Choosing a good Importance Function

- Can be shown that theoretically optimal (minimum variance) importance function is:

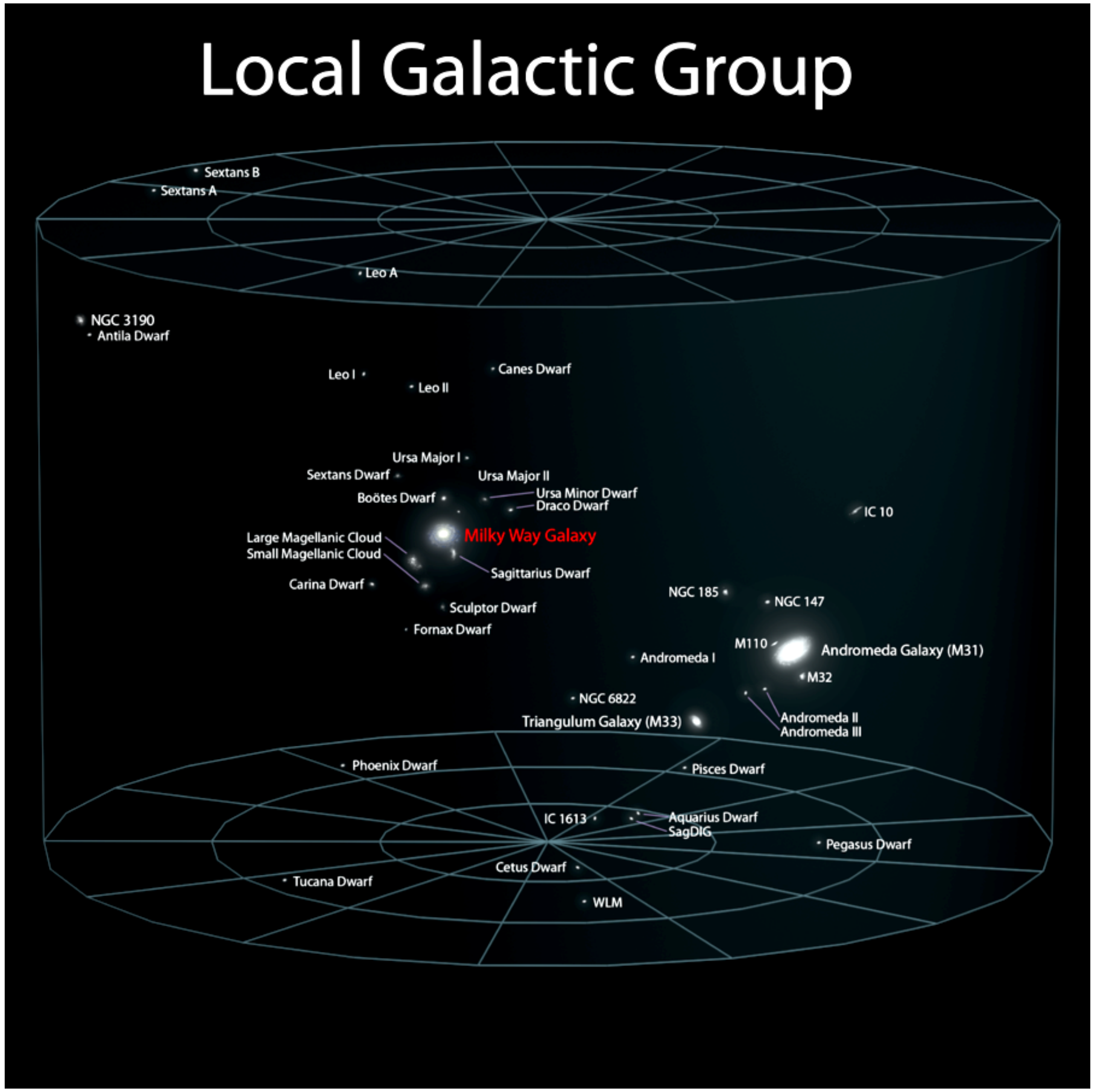
$$Q^*(\boldsymbol{\theta}) = \frac{|f(\boldsymbol{\theta})|P(\boldsymbol{\theta})}{\int |f(\boldsymbol{\theta})|P(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

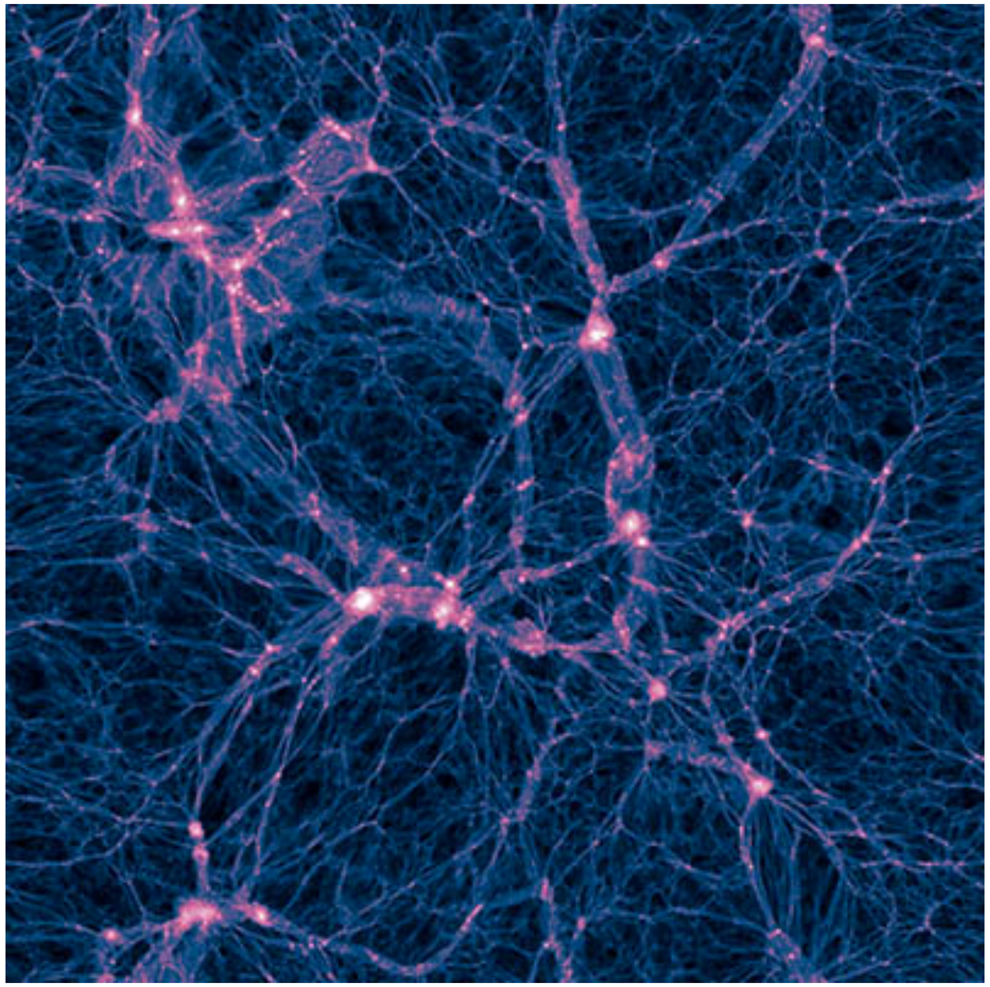
- However, if we can't directly sample from P(**θ**), then we probably can't sample from $|f(\boldsymbol{\theta})|P(\boldsymbol{\theta})$
- Want to keep the importance weights roughly constant, otherwise large variations in P(**θ**) / Q(**θ**) will lead to high variance of estimate, smaller ESS
- Effective Sample Size: $\dfrac{m}{1 + \widehat{\mathrm{Var}}[\{w^*(\boldsymbol{\theta_i})\}]}$
- In practice, find thick-tailed distribution Q(**θ**) that is positive everywhere and similar in shape to | f(**θ**) | P(**θ**)
- Don't want Q(**θ**) small when | f(**θ**) | P(**θ**) large!

Astrostatistics Case Study:

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

(Patel, Besla, & Mandel, 2017, 2018, arXiv:1703.05767, 1803.01878

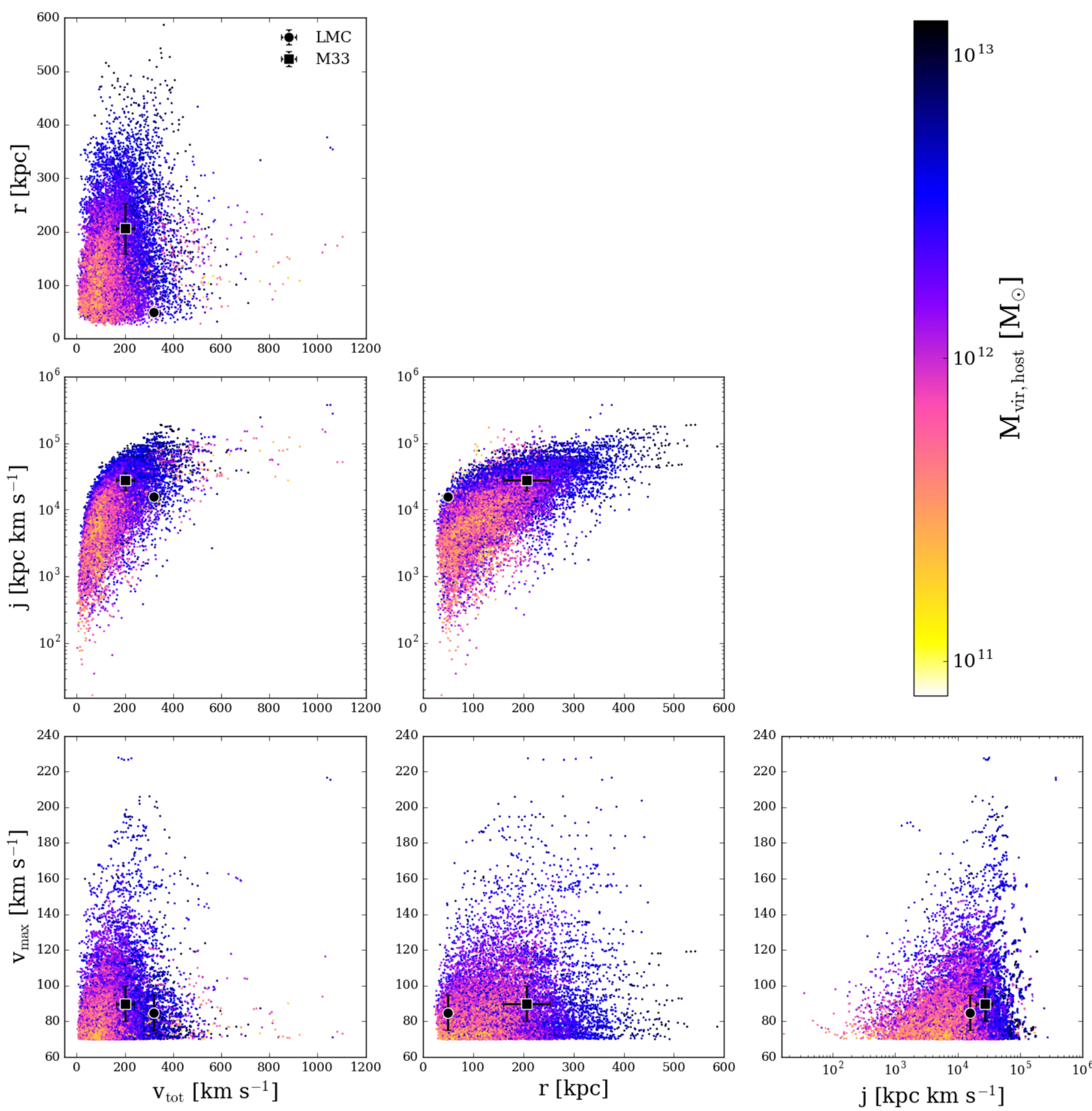




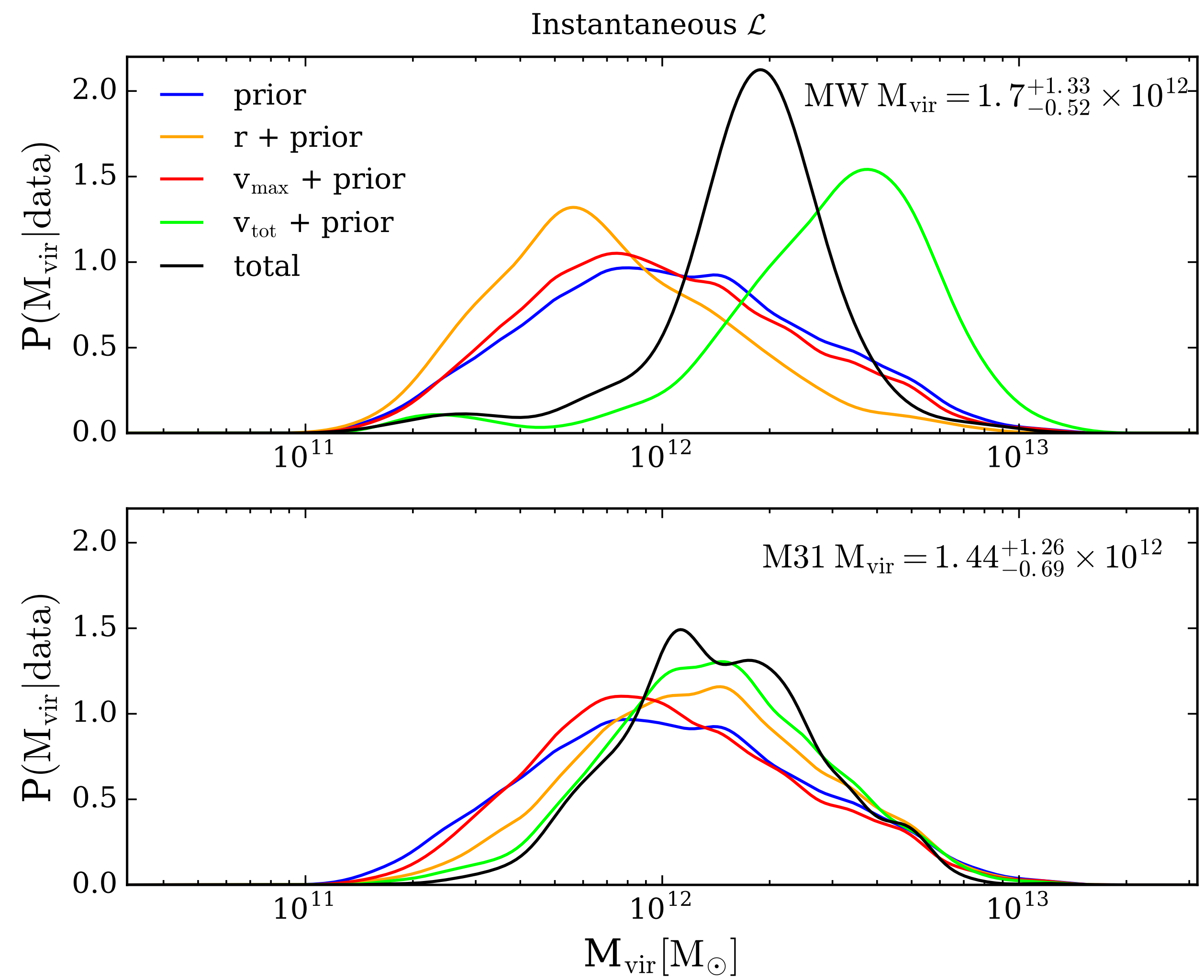
Illustris
Cosmological Simulation of Galaxy Formation

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

(Patel, Besla, & Mandel, 2017, 2018, arXiv:1703.05767, 1803.01878)



Simulation —> Prior

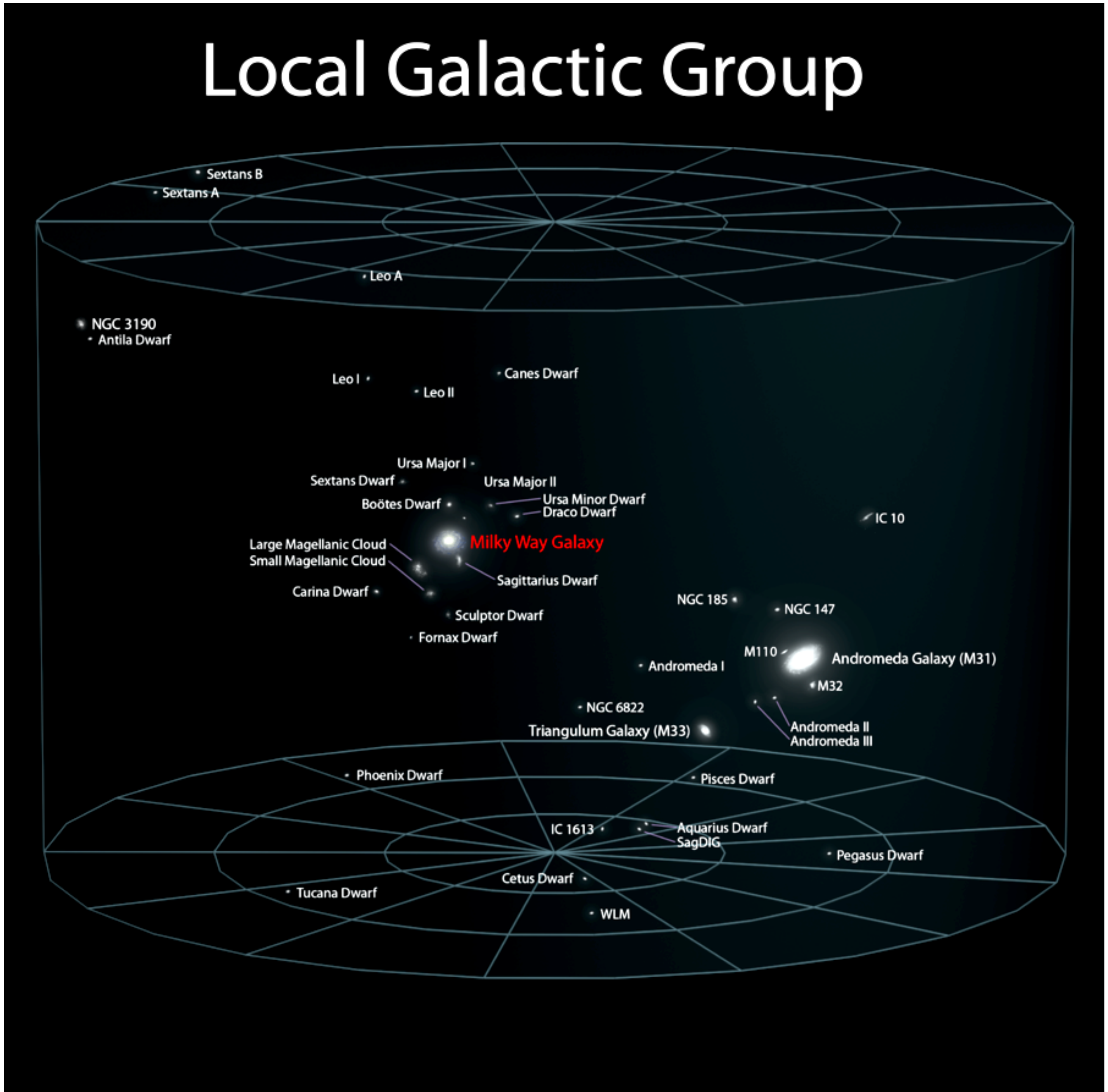


- Bayesian Inference
- Importance Sampling
- Kernel Density Estimation

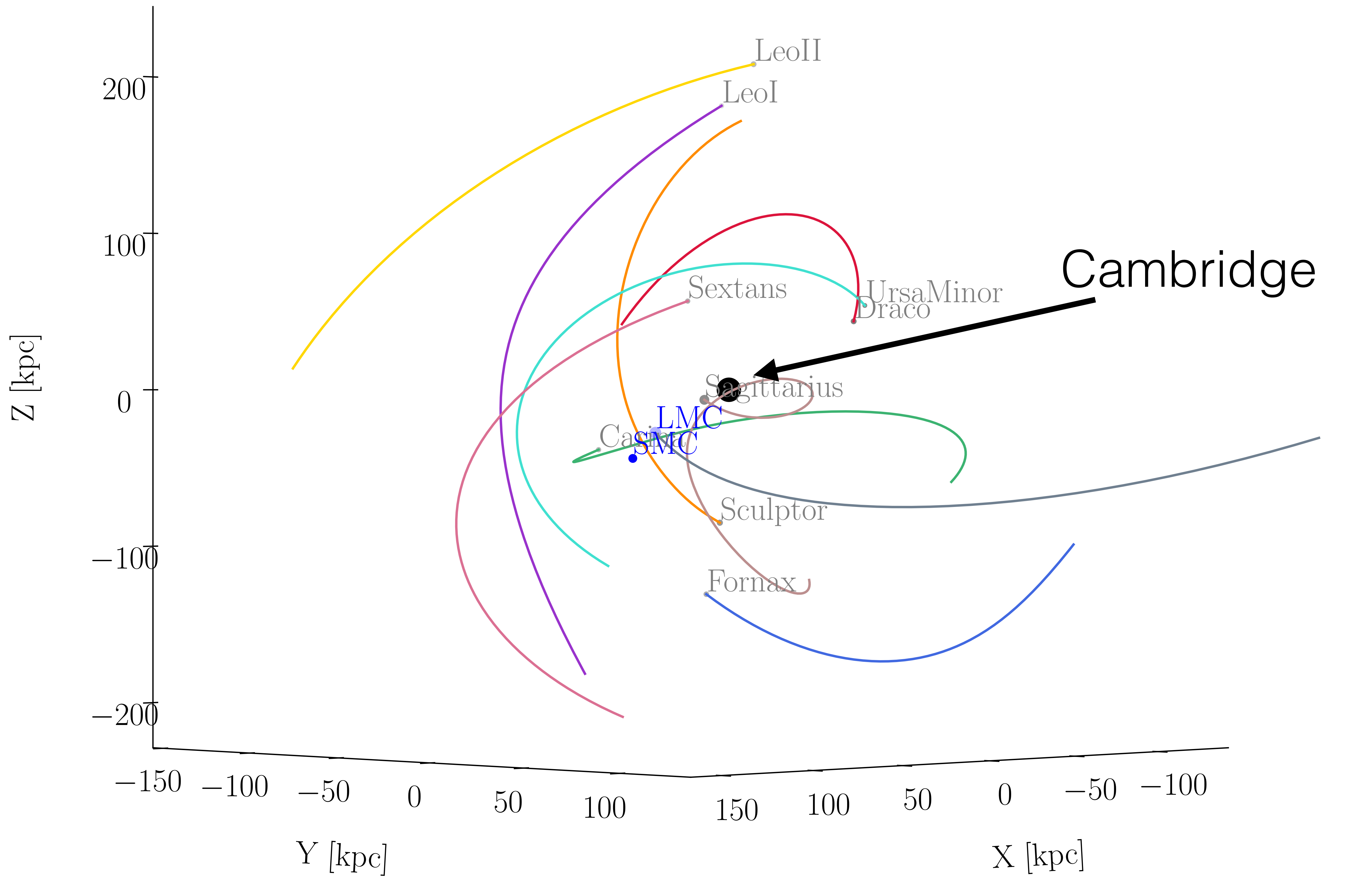
Illustris Cosmological Simulation
Movie

http://www.illustris-project.org/movies/illustris_movie_cube_sub_frame.mp4

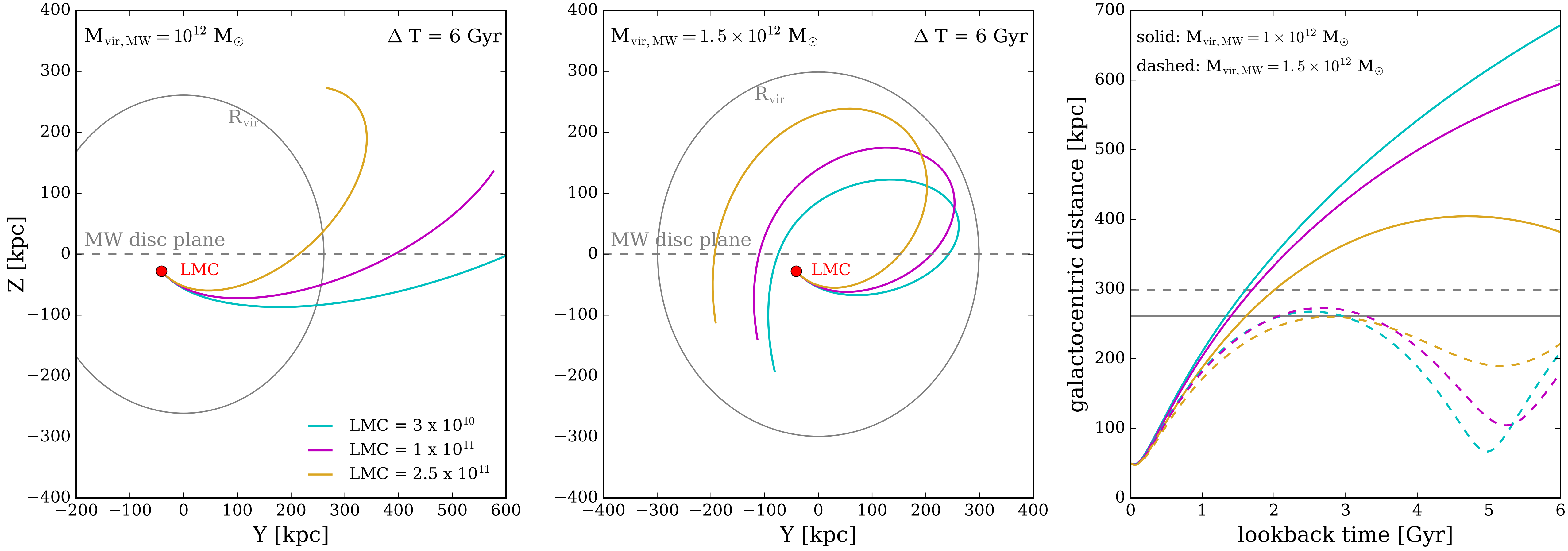
Milky Way has satellite galaxies



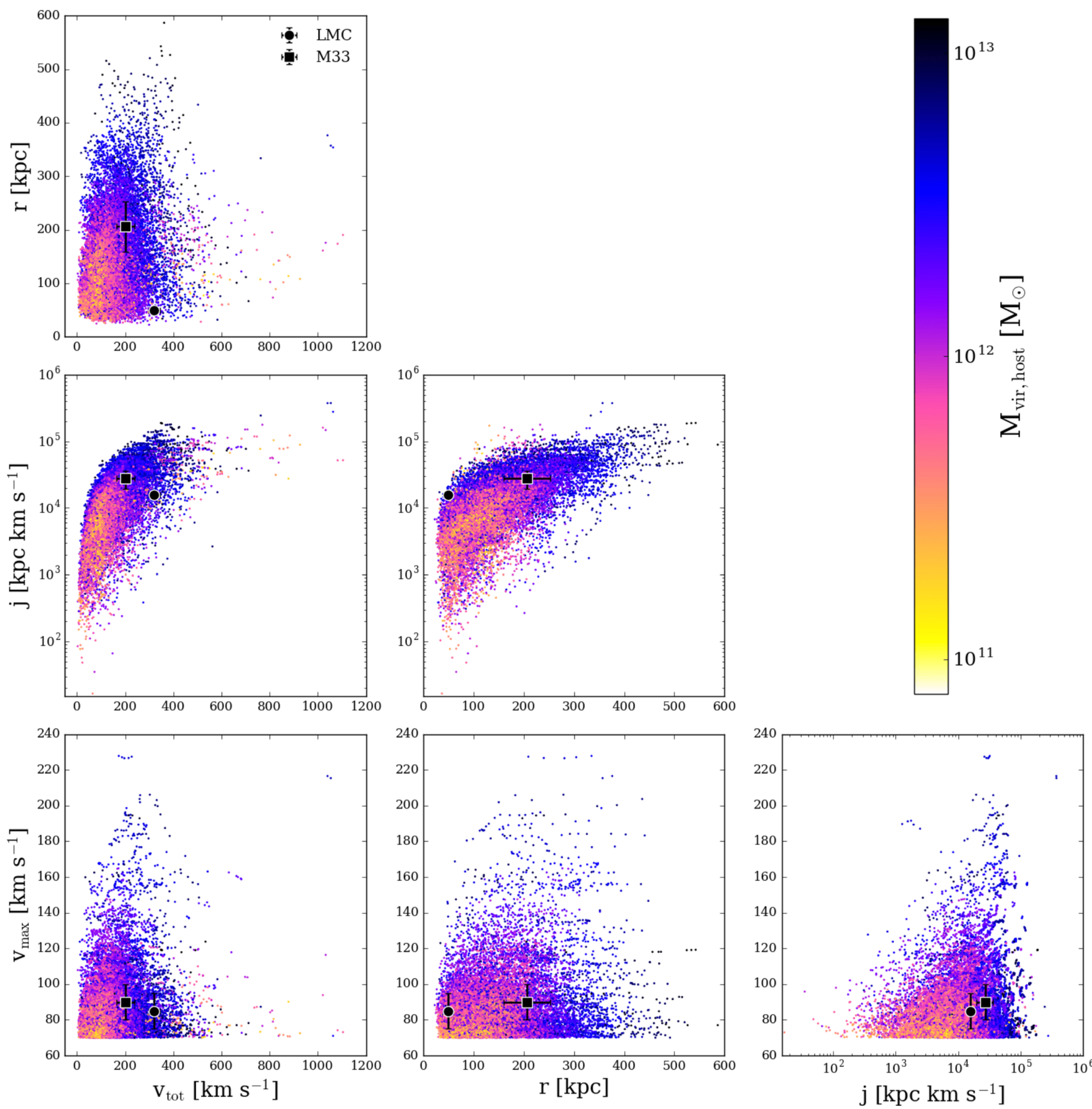
Satellite Galaxies are moving around (us)



Their trajectories depend on the Milky Way Mass



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



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

M_{vir} = Mass of Galaxy

Parameters are:
θ = (x, M_{vir})

Prior cannot be evaluated
Only Sampled!

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

Table 1. Observational data (***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
$v_{\rm max}^{\rm obs}$ (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

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

Likelihood Function

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

d = measurements of satellite properties
x = latent (true) values of satellite properties
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)

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

Likelihood (observations)   Prior (samples from Simulation)

Importance Sampling

Parameters are: θ = (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:

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

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

The denominator of this equation is the normalization constant. If

Sum over Samples from Prior P(x, M_{vir})

Importance Sampling

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

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

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

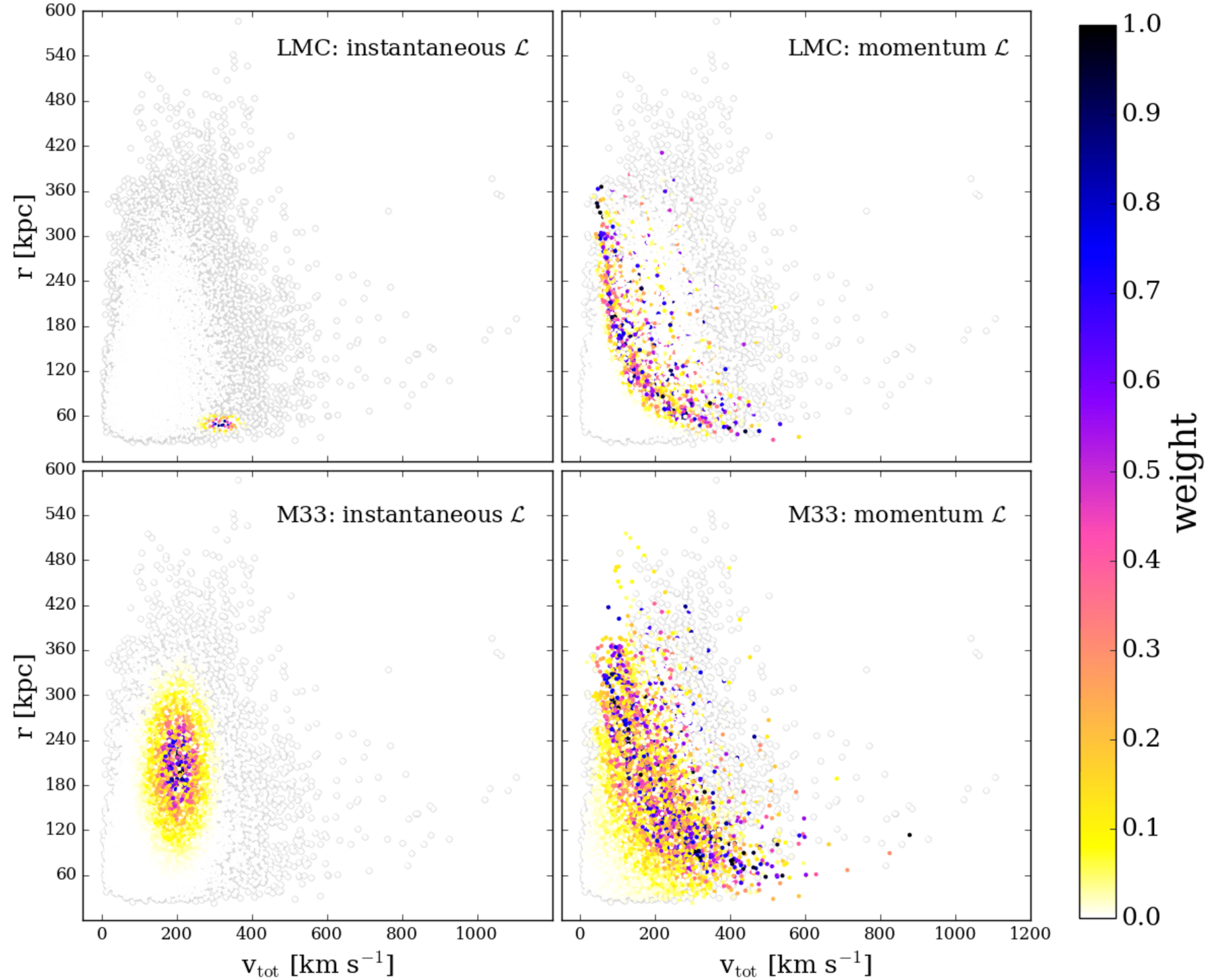
Impt Weight is proportional to likelihood of each sample

$$= \sum_j^n f(M_{\rm vir}^j)\, w_j,$$

(Self-normalised)

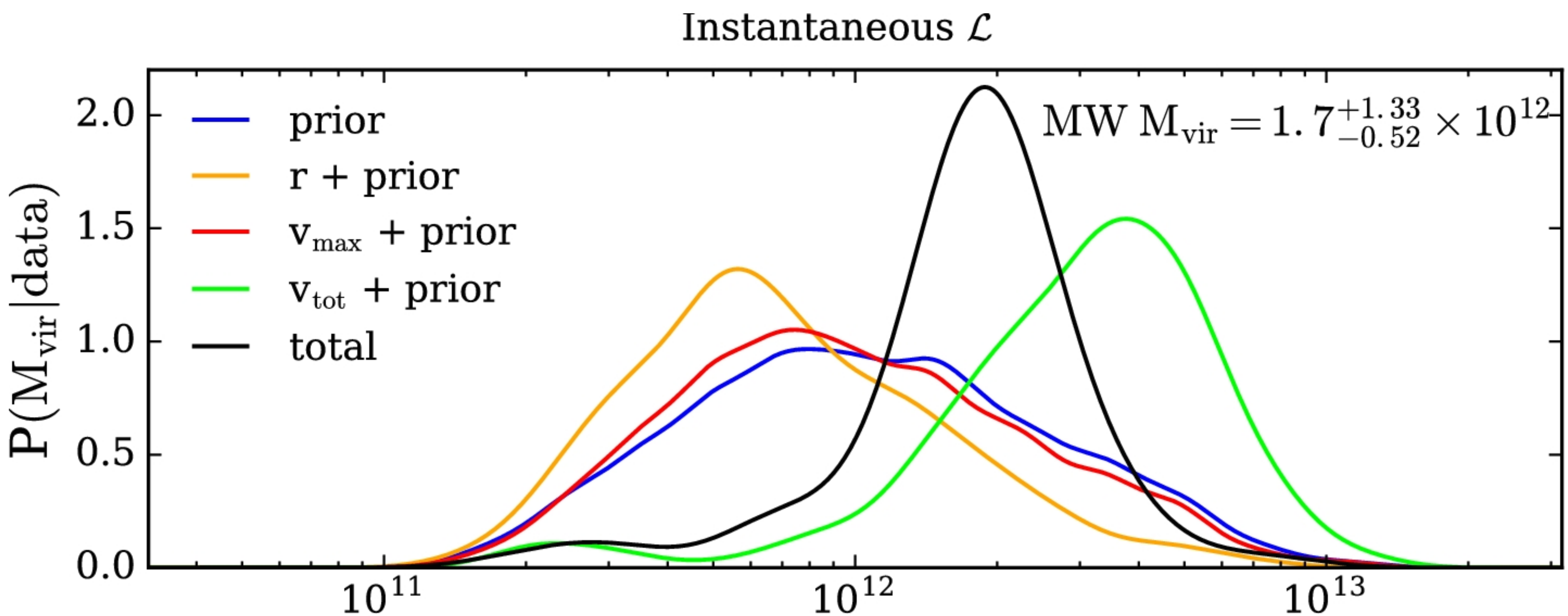
where $w_i = P(\boldsymbol{d} | \boldsymbol{x}_i) / \sum_{\rm j}^{\rm n} P(\boldsymbol{d} | \boldsymbol{x}_j)$ are importance weights.

Distribution of Importance Weights



Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations (Patel, Belsa & Mandel 2017)

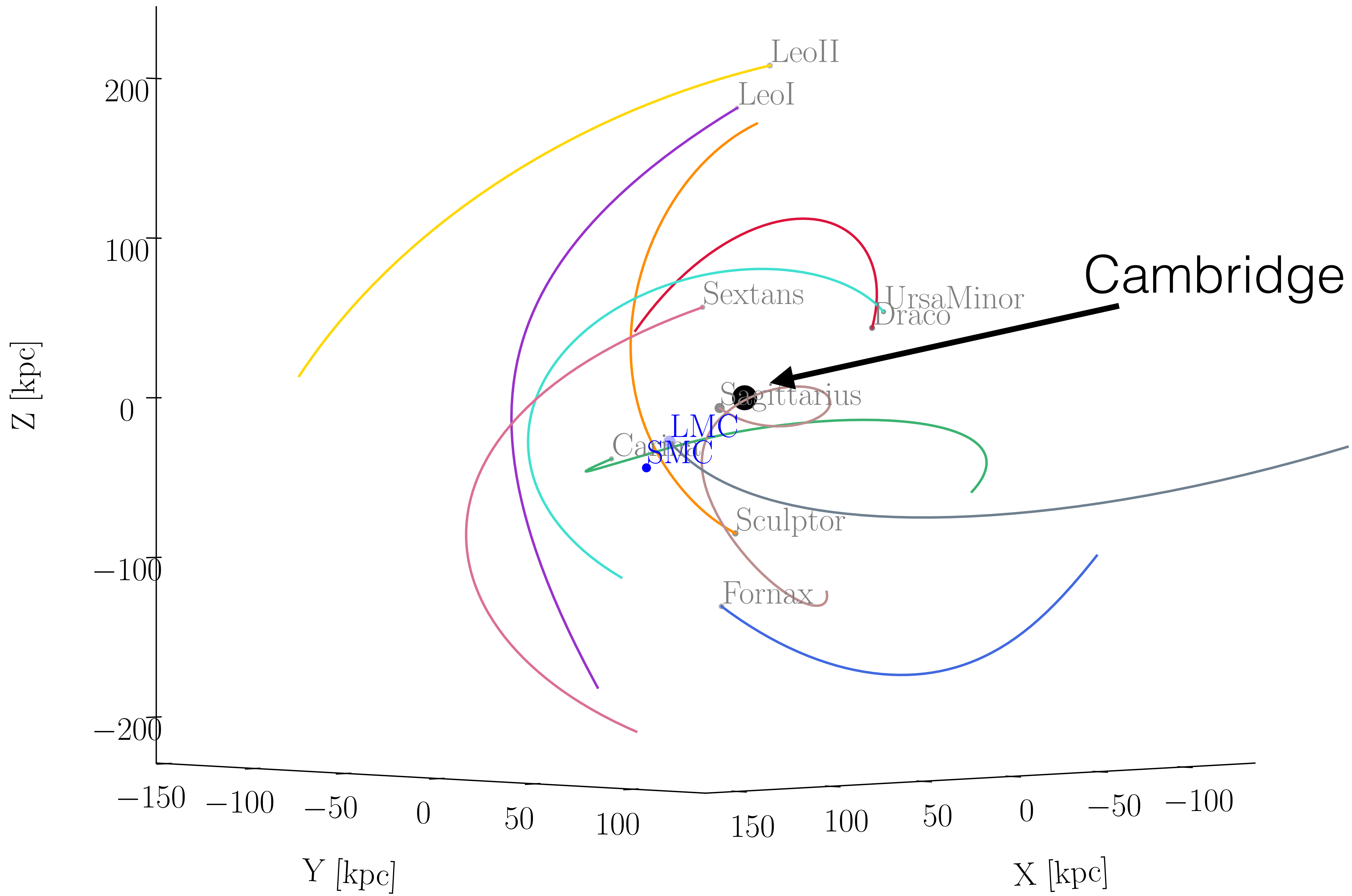
Posterior of Milky Way Galaxy Mass with weighted KDE



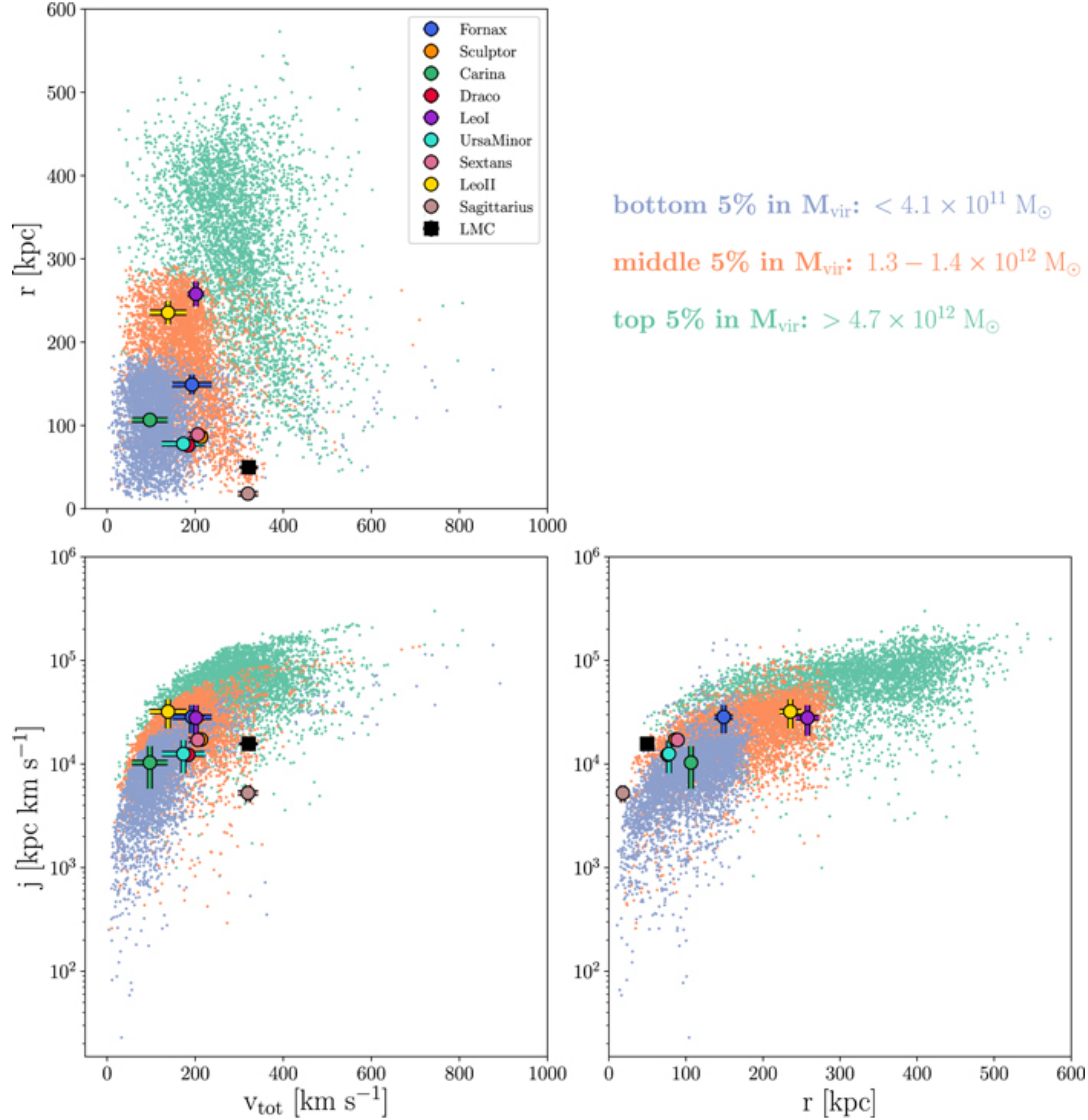
Milky Way Mass M_{vir}

Using only the biggest satellite (Large Magellanic Cloud)

There are many dwarf satellite galaxies



Prior Distribution of dynamical properties with central galaxy mass



bottom 5% in M_{vir}: $< 4.1 \times 10^{11}\ M_\odot$

middle 5% in M_{vir}: $1.3 - 1.4 \times 10^{12}\ M_\odot$

top 5% in M_{vir}: $> 4.7 \times 10^{12}\ M_\odot$

Combined Posterior of Milky Way Mass

Patel, Belsa, Mandel & Sohn 2018