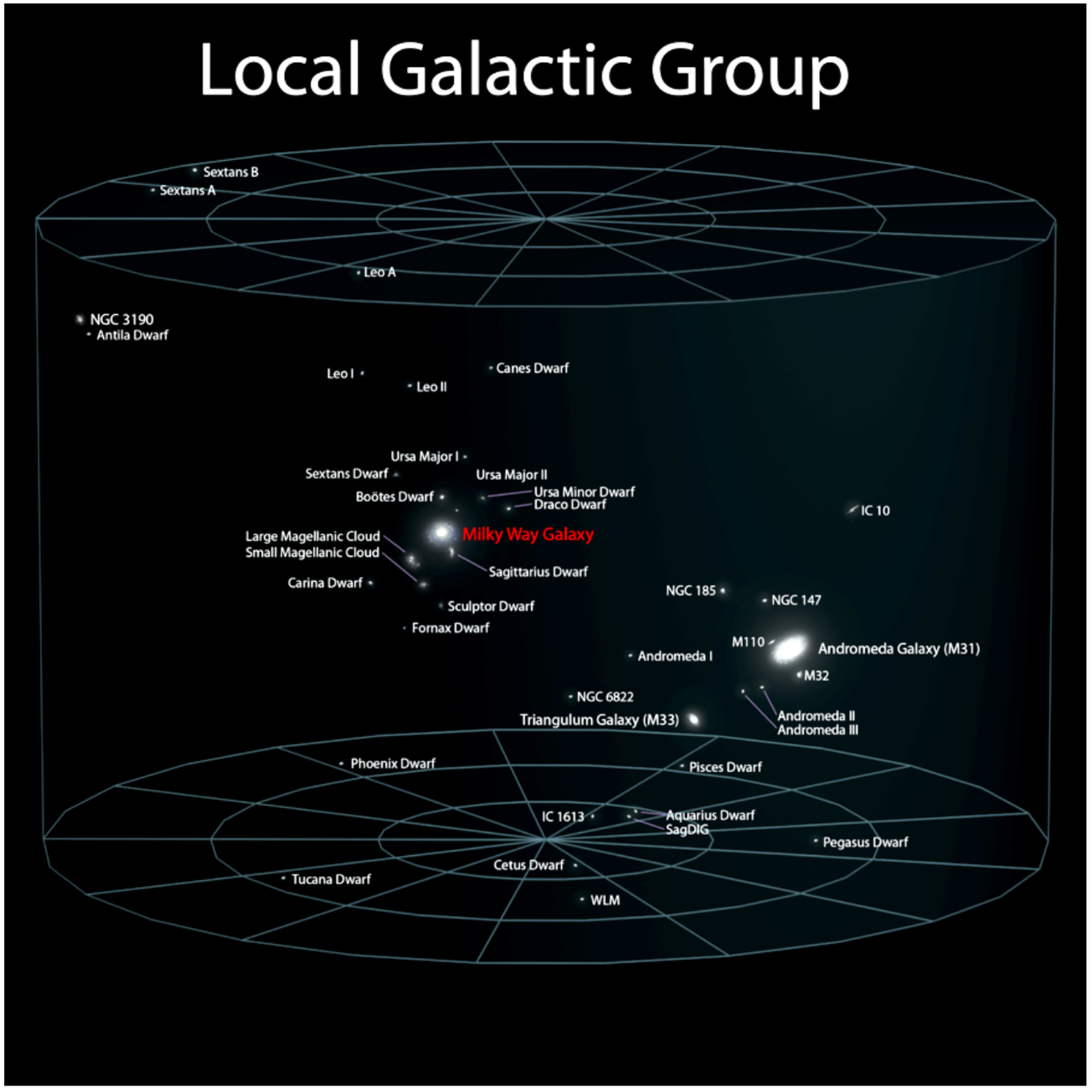
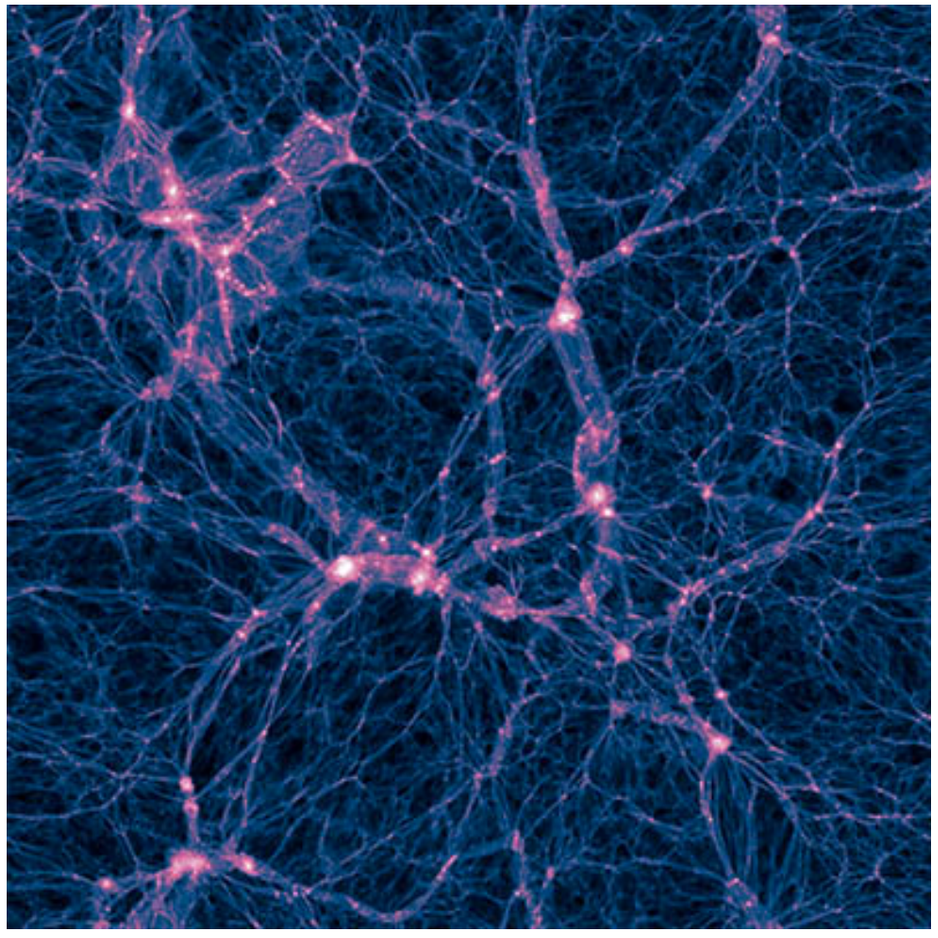
Astrostatistics: Monday 17 Feb 2020

https://github.com/CambridgeAstroStat/PartIII-Astrostatistics-2020

- Today: continue Bayesian computation / Monte Carlo Methods:
  - MacKay: Ch 29-30; Bishop: Ch 11; Gelman BDA
  - Givens & Hoeting. "Computational Statistics" (Free through Cambridge Library iDiscover)
  - Importance Sampling
  - Case Study: Bayesian Estimates of the Mass of the Milky Way Galaxy
- Example Class 2, Thu Feb 27, 3:30pm MR13

Later Today & Example Sheet 2: 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)

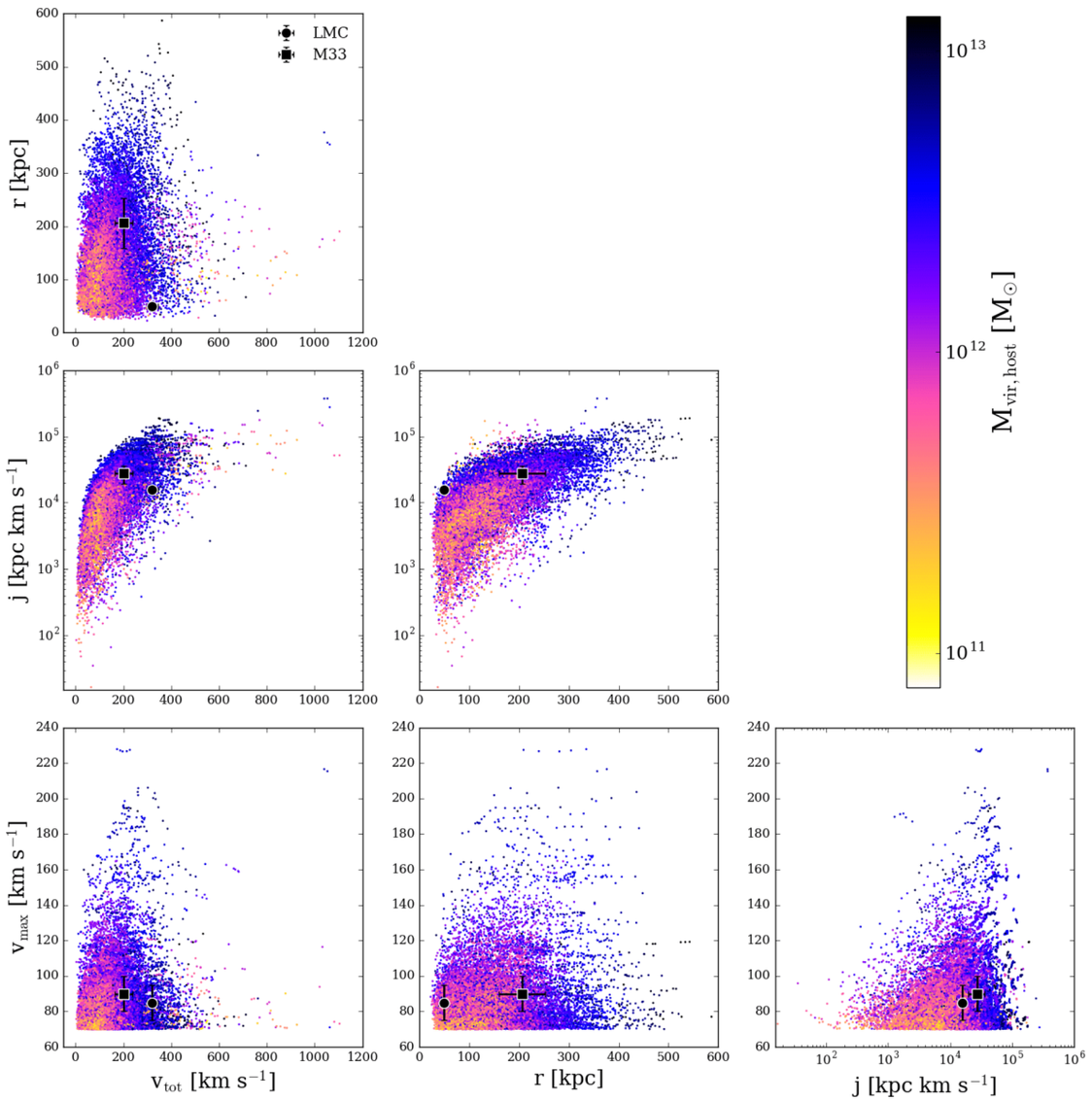




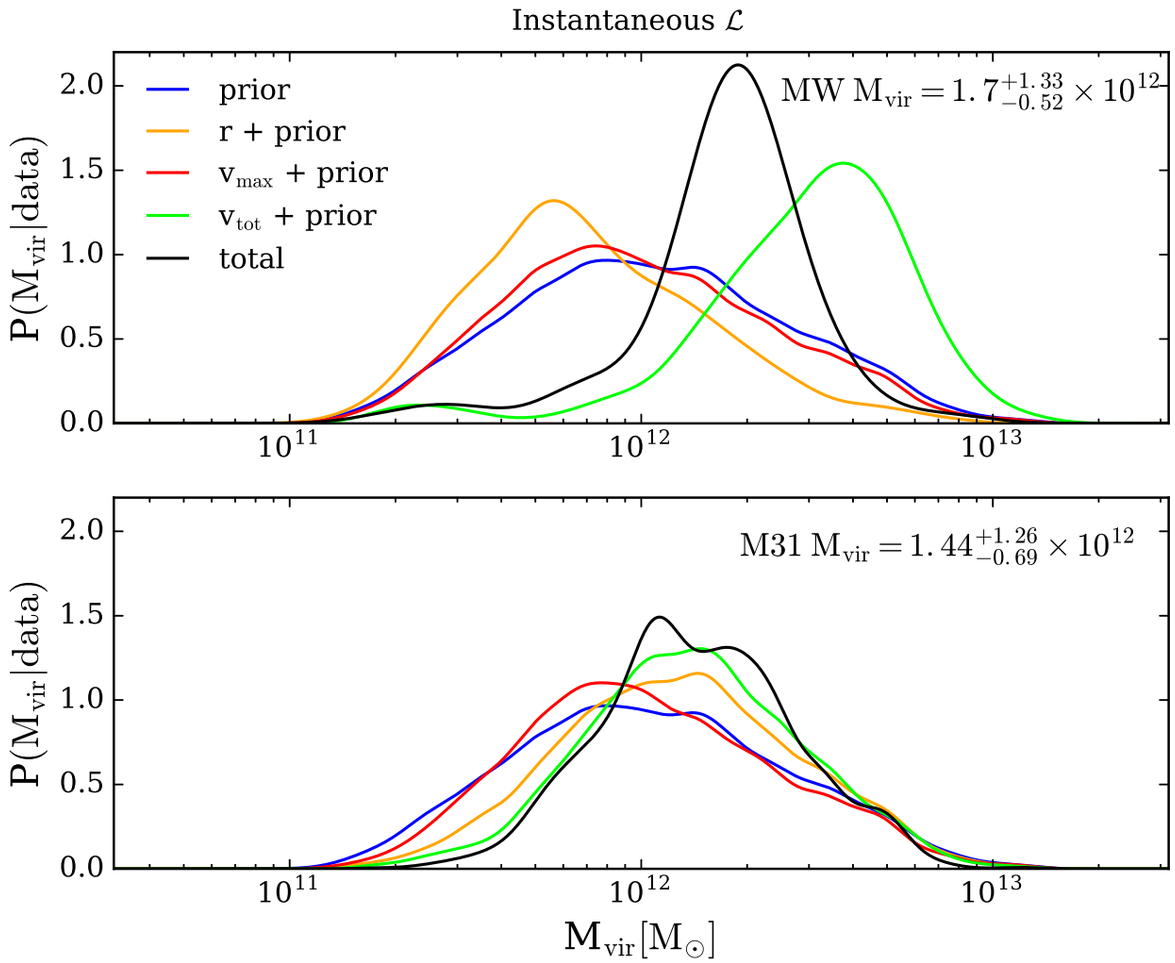
Illustris
Cosmological Simulation of Galaxy Formation

Later Today: 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

Last Time & Example Sheet 2:
Analytic Posterior Density for a Gaussian(μ, σ²) model

Likelihood: $$y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2) \qquad i = 1, \ldots, n$$

(improper) Prior: $$P(\mu) \propto 1$$

$$P(\log \sigma^2) \propto 1 \text{ or } P(\sigma^2) \propto \sigma^{-2} \qquad (\sigma^2 > 0)$$

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

sufficient statistics

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

Example Sheet 2:
Analytic Posterior Density for a Gaussian(μ, σ²) model

Joint Posterior: $(\sigma^2 > 0)$

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

Conditional Posterior

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

Marginal Posteriors

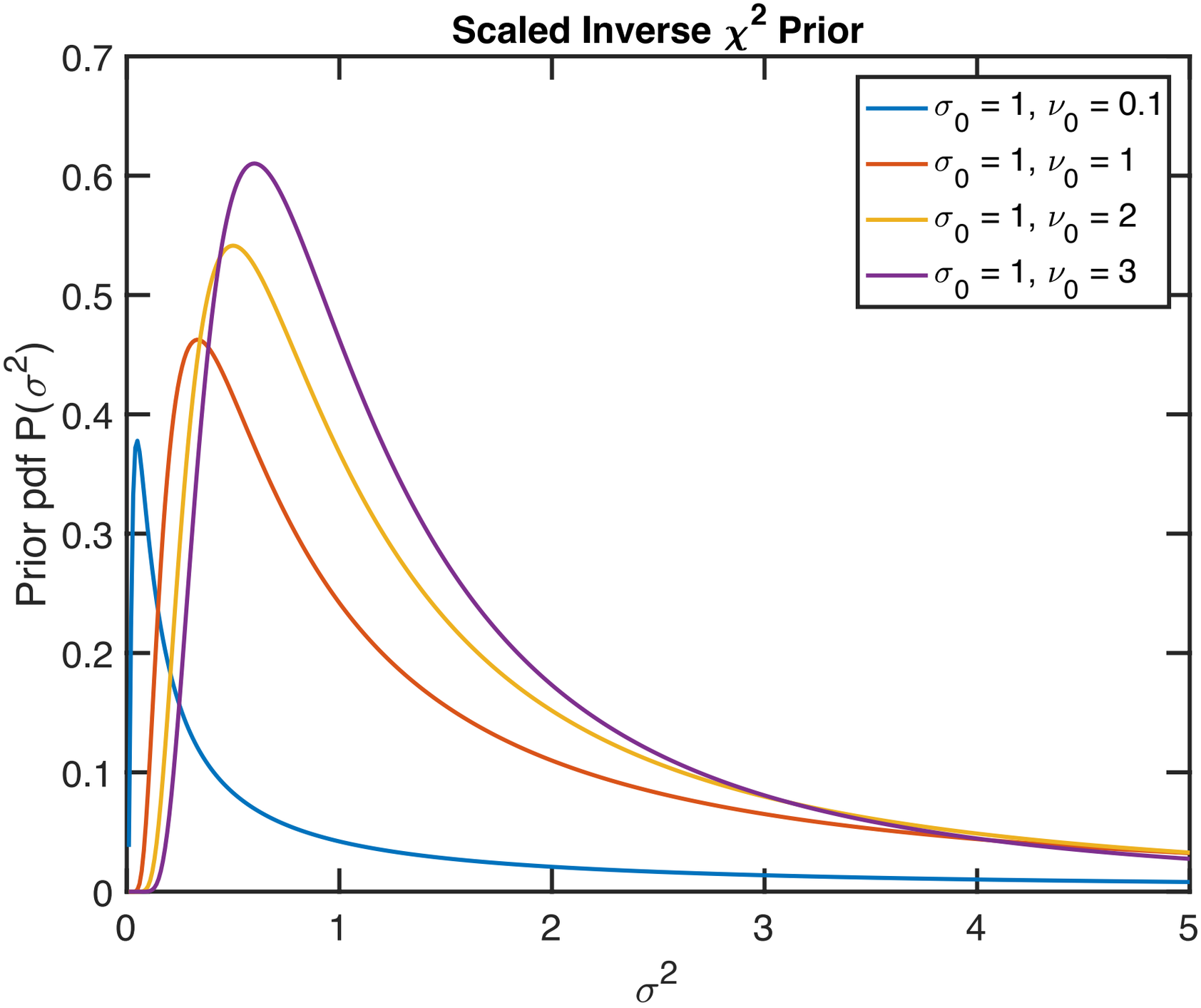
$$P(\sigma^2 | \, \boldsymbol{y}) = \int P(\mu, \sigma^2 | \boldsymbol{y}) \, d\mu = \text{Inv-}\chi^2(\sigma^2 \, | \, n-1, s^2)$$

$$P(\mu | \, \boldsymbol{y}) = \int P(\mu, \sigma^2 | \boldsymbol{y}) \, d\sigma^2$$

$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2} = t_{n-1}(\mu | \, \bar{y}, s^2/n)$$

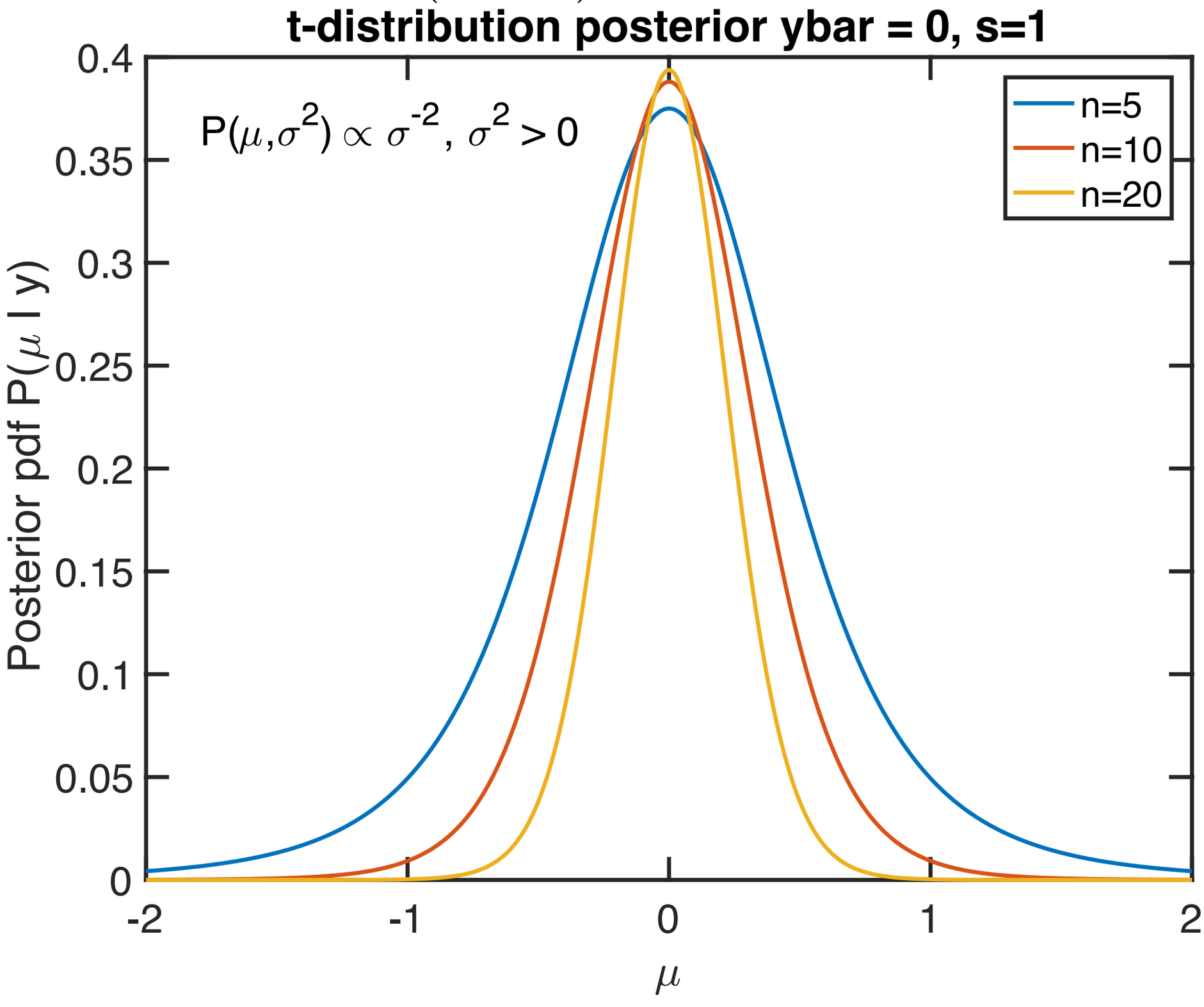
inverse chi^2 (or inv-gamma)

$$\text{Inv}-\chi^2(\sigma^2|\nu_0,\sigma_0^2) \propto (\sigma^2)^{(-\nu_0/2+1)} \exp(-\frac{\nu_0\sigma_0^2}{2\sigma^2})$$



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

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



What if you can’t compute marginals / expectations analytically?

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:

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 $f(\boldsymbol{\theta}) = \boldsymbol{\theta}$

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

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

(indicator function)

Posterior Expectation ≠ Expectation over Data!

Monte Carlo Direct Sampling

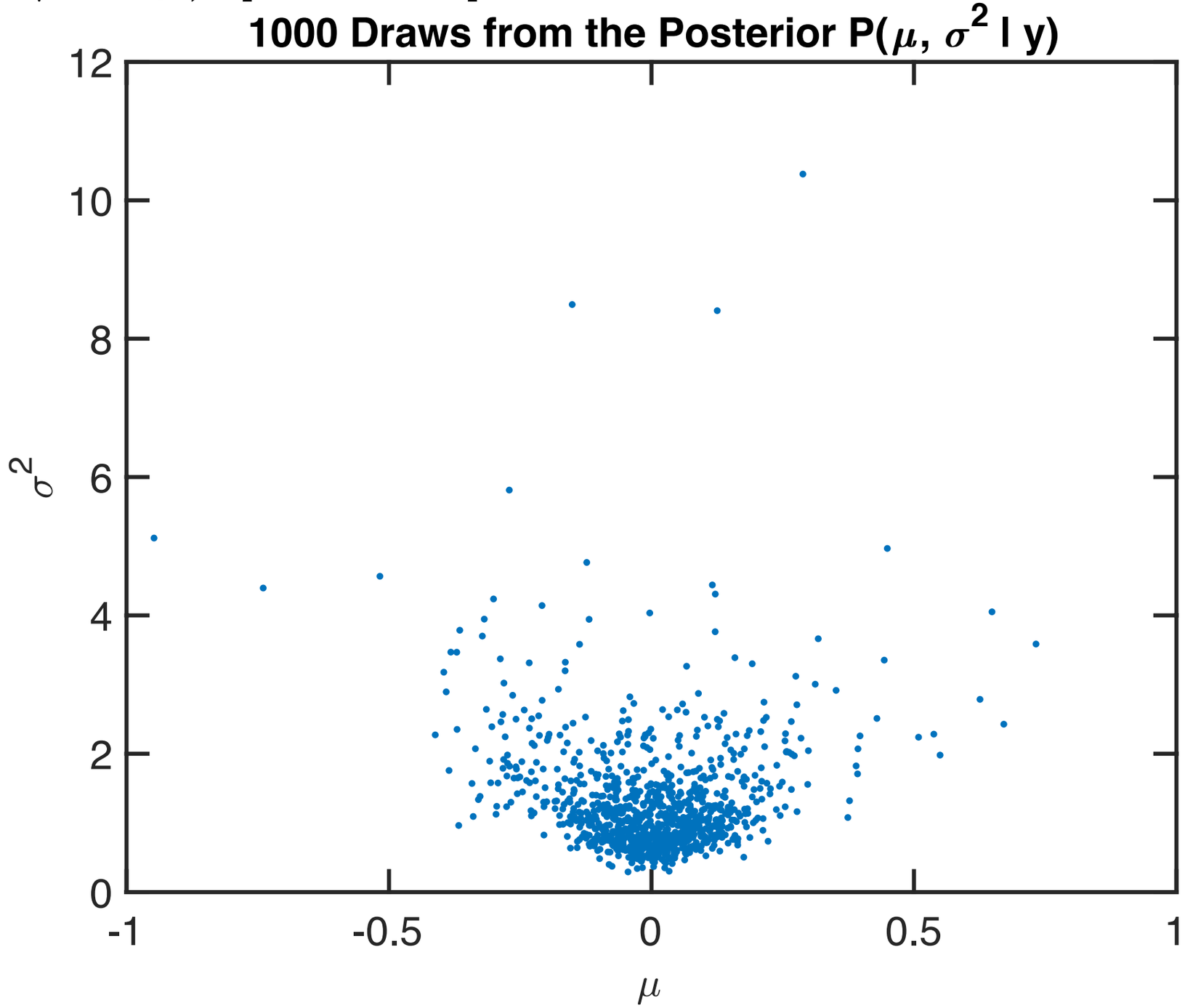
Factorise Posterior: $P(\mu, \sigma | \boldsymbol{y}) = P(\mu | \sigma^2, \boldsymbol{y}) \times P(\sigma^2 | \boldsymbol{y})$

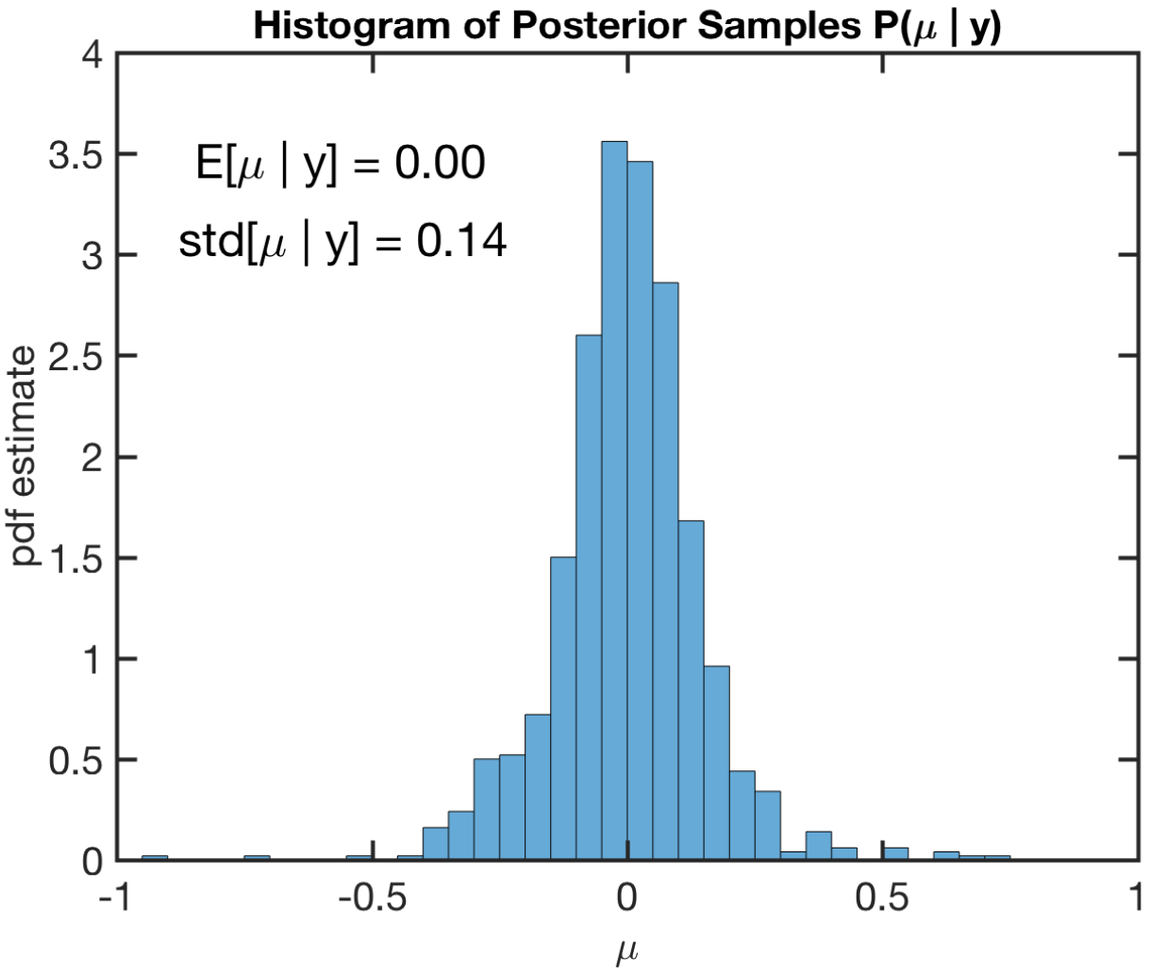
1. $\sigma_i^2 \sim P(\sigma^2 | \boldsymbol{y})$ [Inv-χ^2]
2. $\mu_i | \sigma_i^2 \sim P(\mu | \sigma^2, \boldsymbol{y})$ [Normal]

$\bar{y} = 0$

$s^2 = 1$

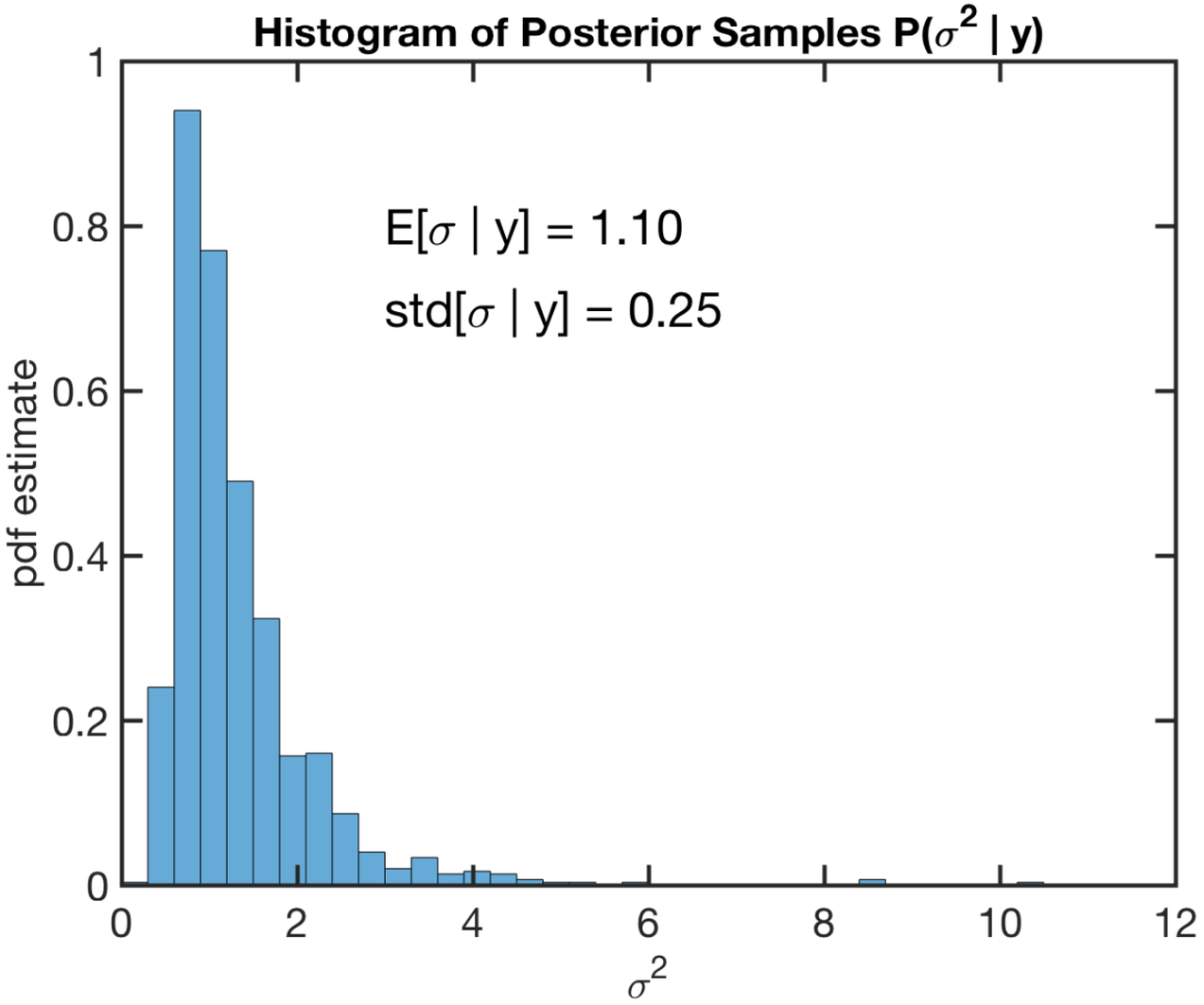
$n = 10$



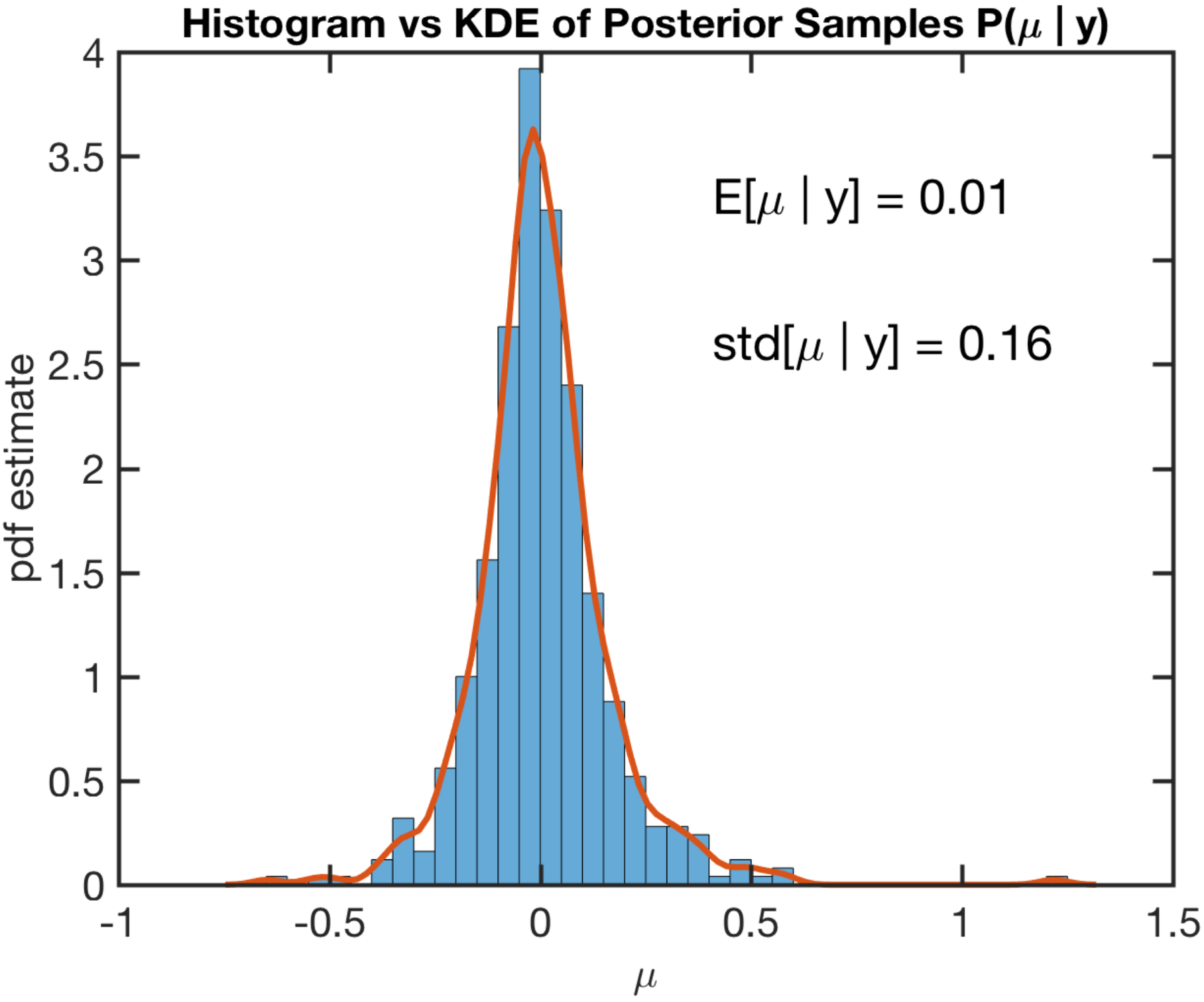


Marginal Posterior (μ)

Marginal Posterior (σ^2)

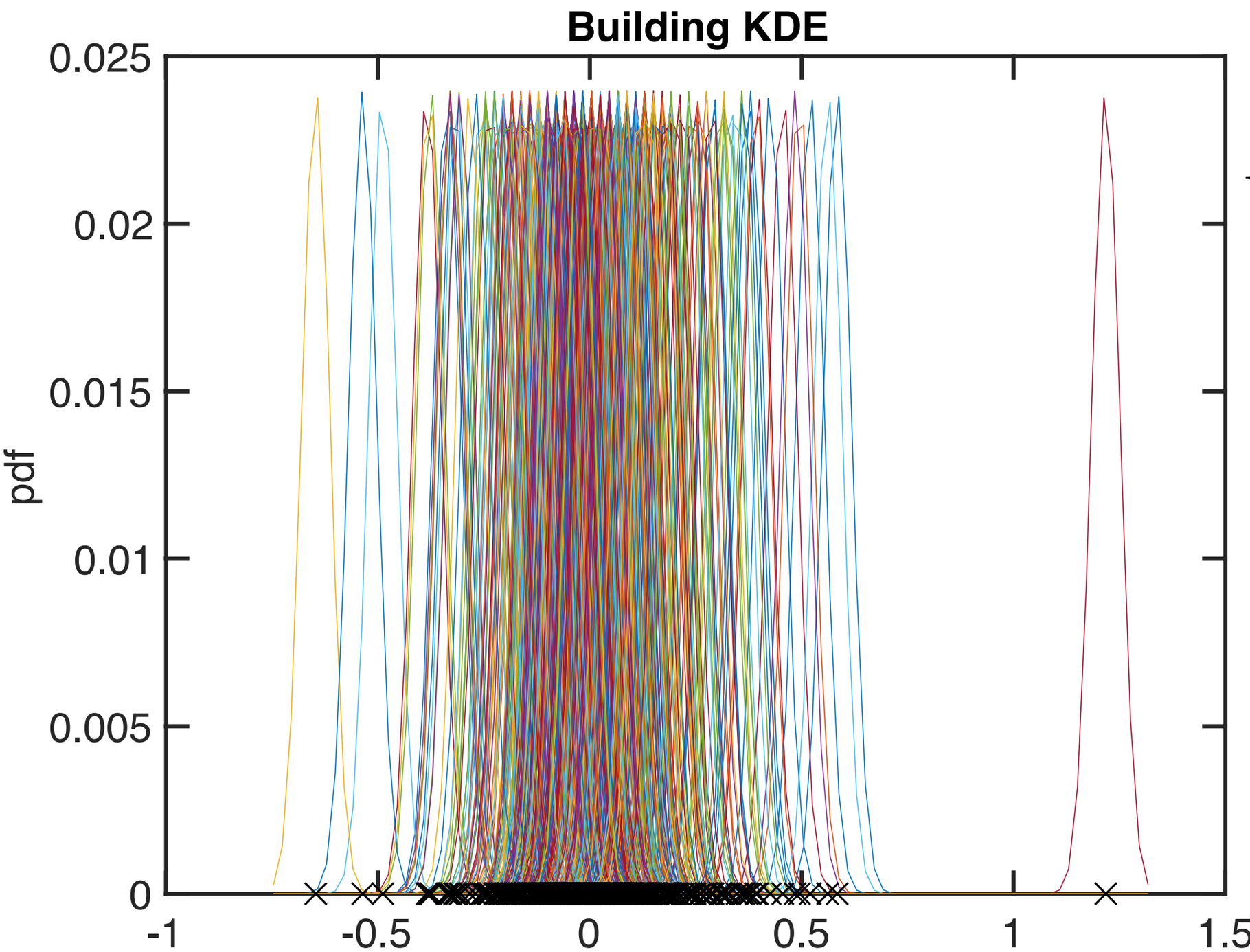


Kernel Density Estimate = estimate a smooth density from samples



Kernel Density Estimation (KDE) (Smooth Histogram)

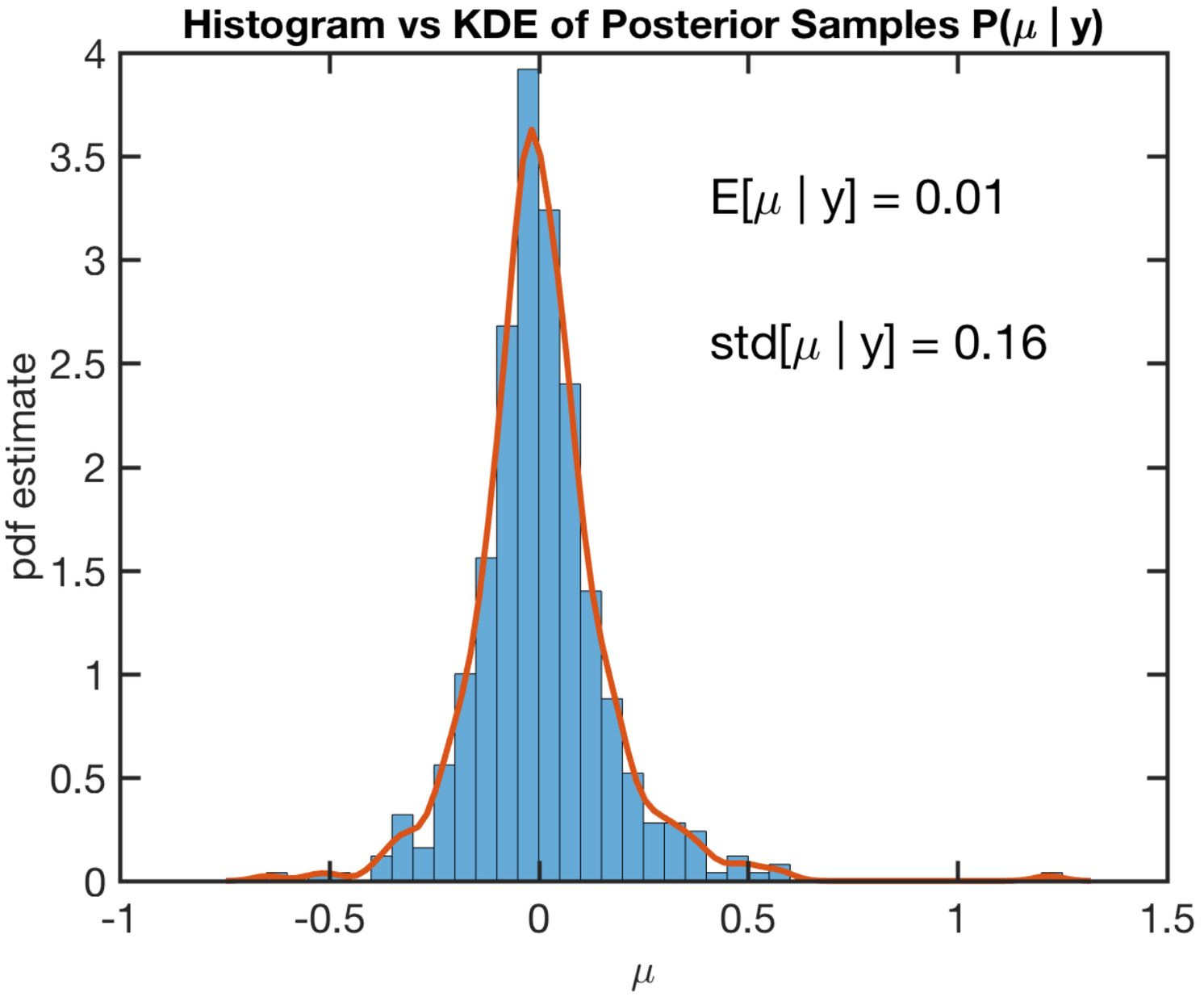
Each sample gets a Gaussian at the sample point
with an “optimal” bandwidth bw (rule of thumb)



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

Kernel Density Estimation (KDE) (Smooth Histogram)

Then add them up and normalise pdf to 1



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

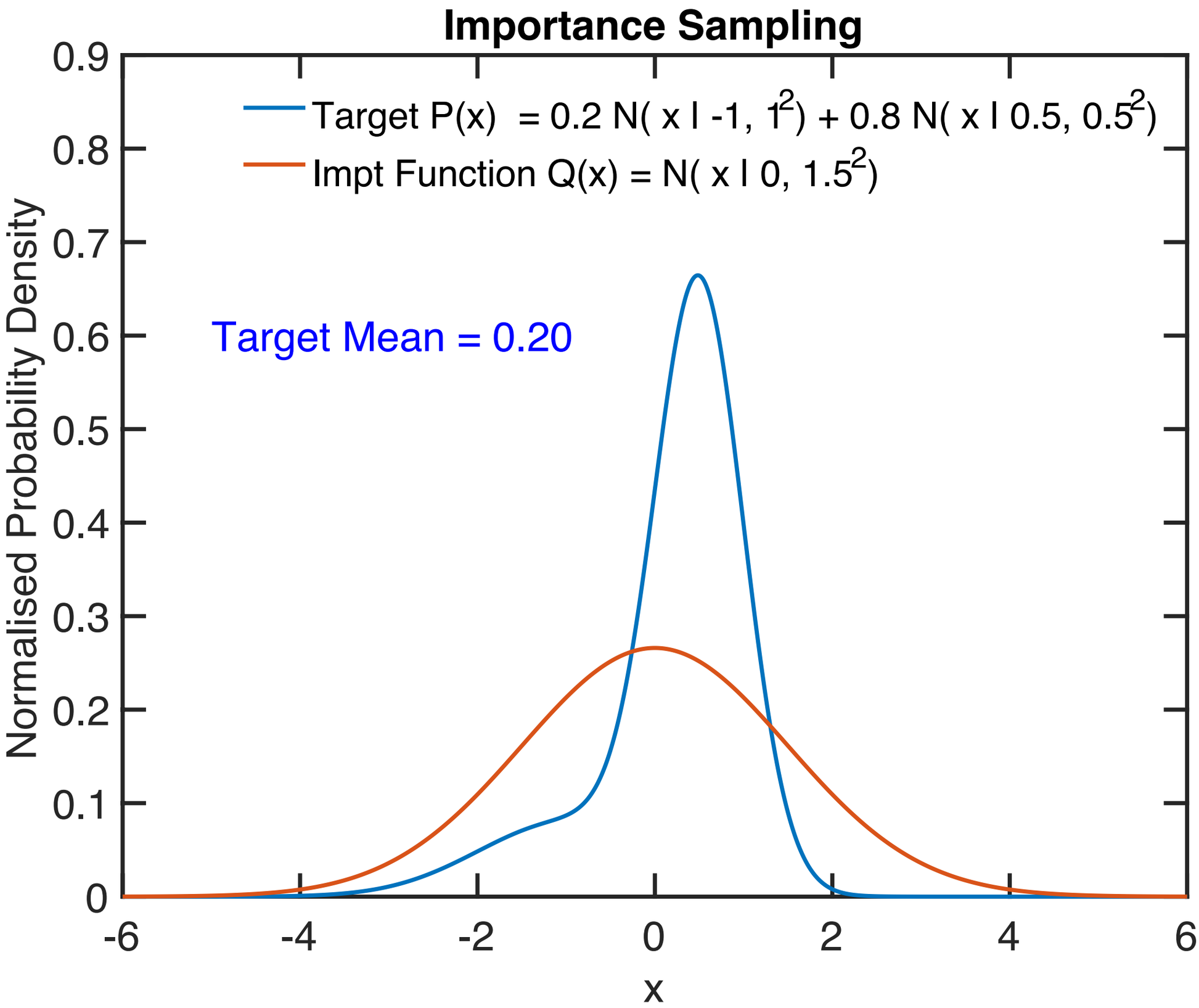
$$\mathbb{E}[f(\boldsymbol{\theta})|\, D] = \int f(\boldsymbol{\theta}) P(\boldsymbol{\theta}|\, D)\, d\boldsymbol{\theta} \approx \frac{1}{m} \sum_{i=1}^{m} f(\boldsymbol{\theta}_i)$$

- Posterior simulation - Markov Chain Monte Carlo, Nested Sampling, etc. generates draws from the posterior density iteratively in long-run

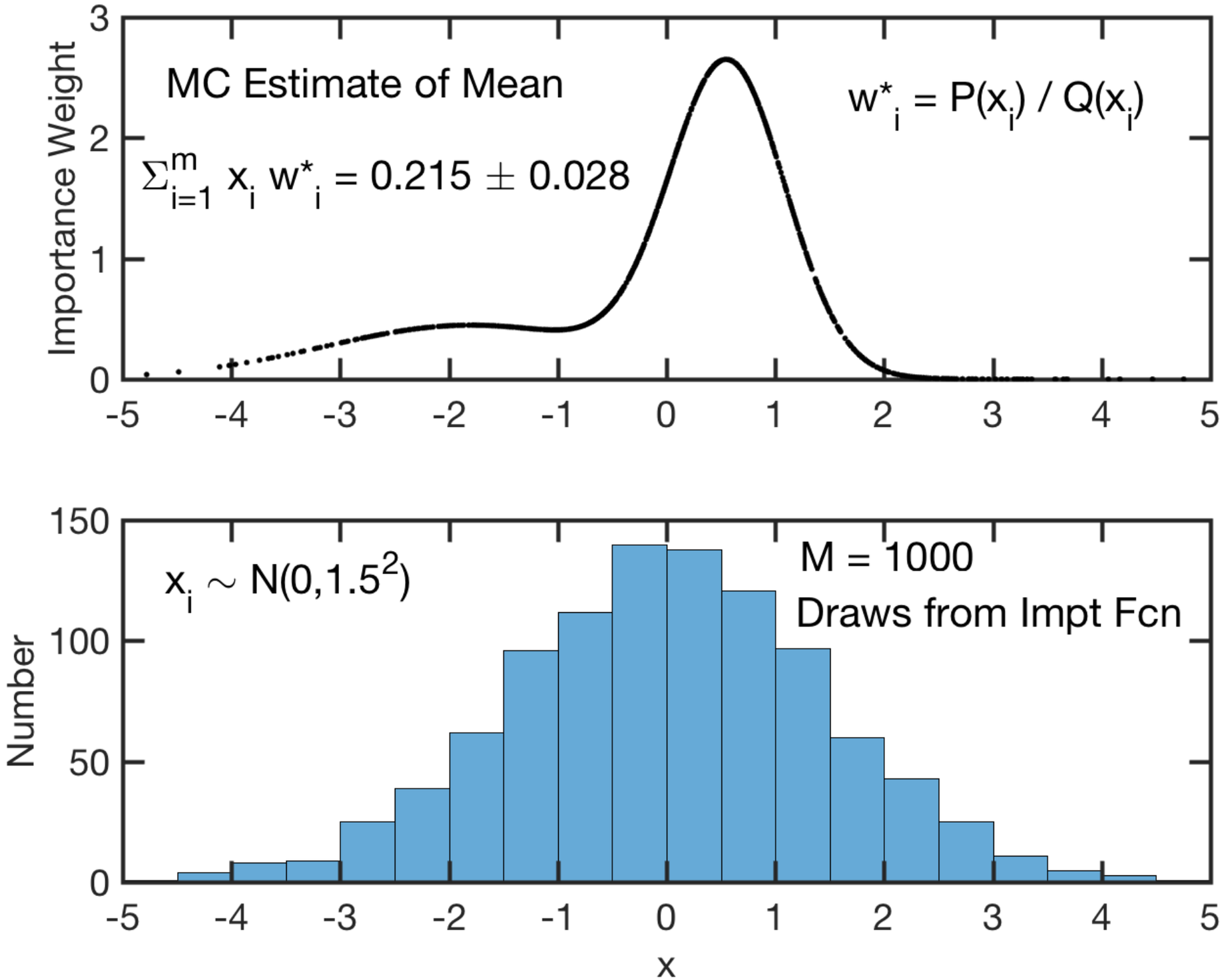
- Importance Sampling - draw from an easier ("tractable") distribution (importance function)
  $\theta_i \sim Q(\theta)$ and weight the samples by
  $w_i = P(\theta_i | D) / Q(\theta_i)$ to compute expectations

Two Cases of Importance Sampling

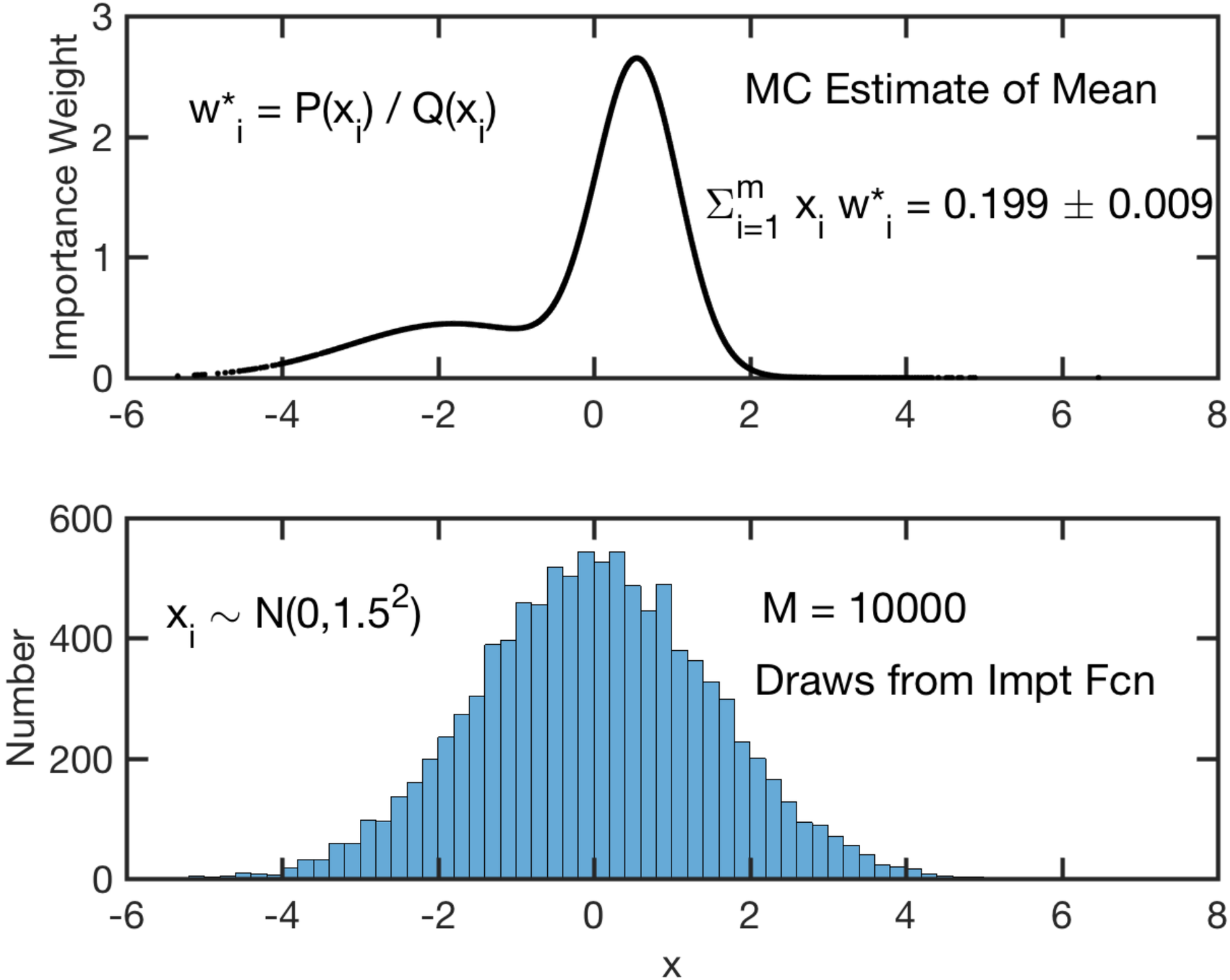
Contrived Example [Gaussian Mixture, Normalised PDF]



Importance Sampling Example
m = 1,000 Draws



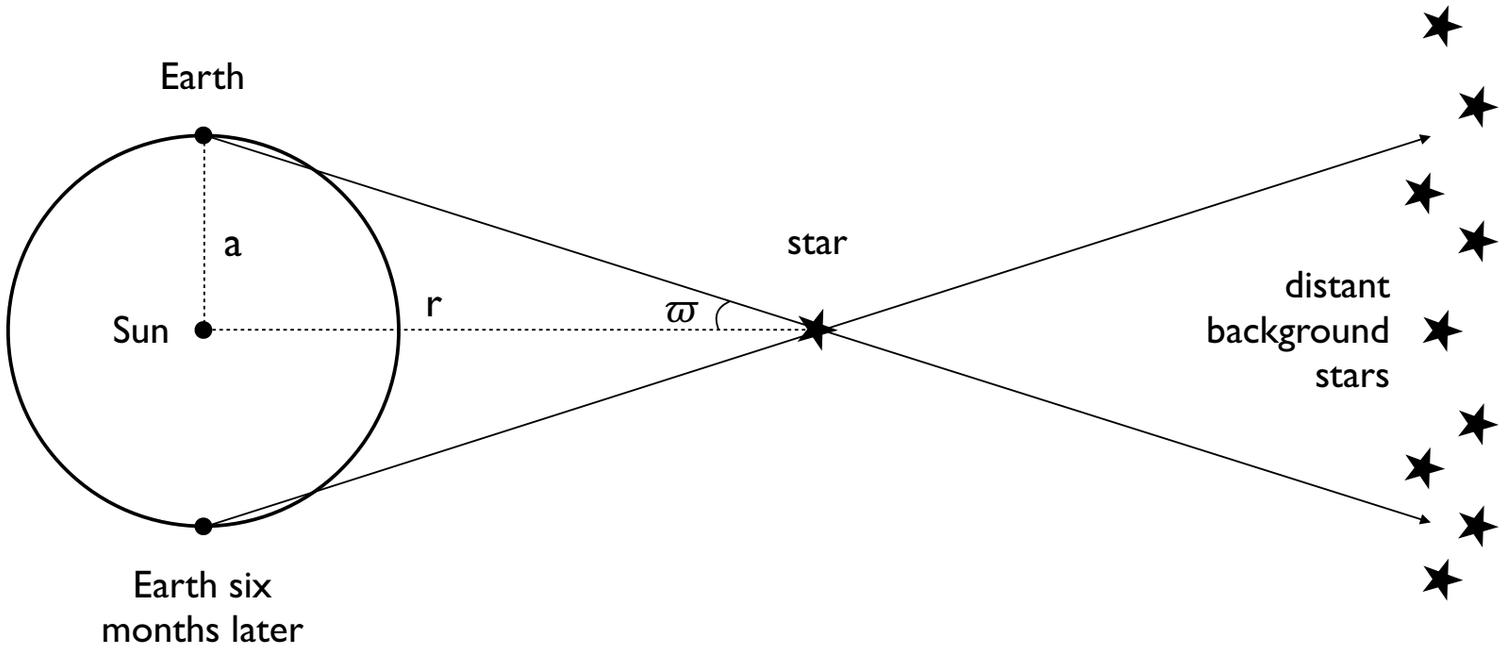
Importance Sampling Example
m = 10,000 Draws



Parallax Example

Likelihood: $$\omega \sim N\left(\frac{1}{r}, \sigma_\omega^2\right)$$

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



True relation (no errors)

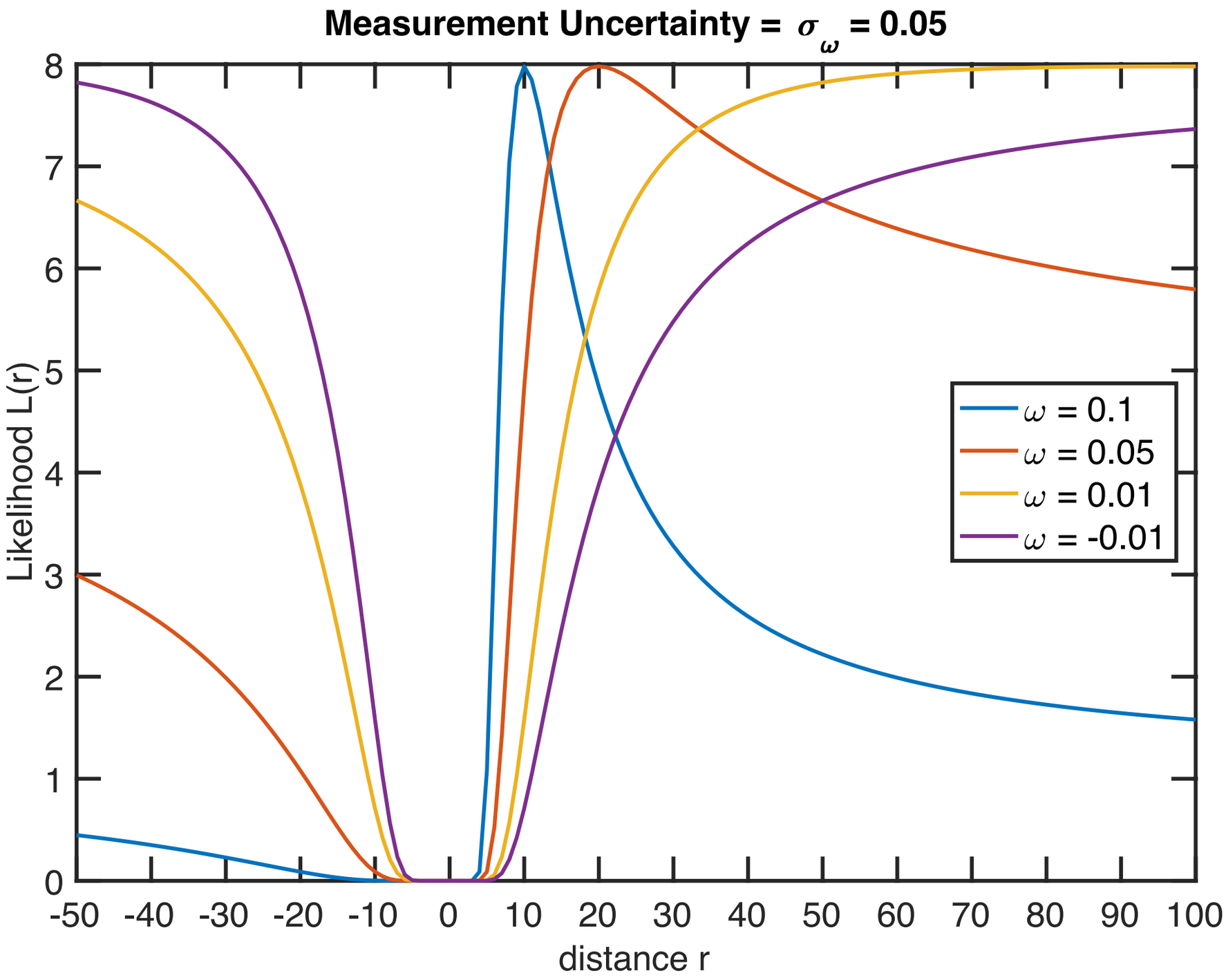
Parallax Angle

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

Distance

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.

The Likelihood Function



- Likelihood is positive on negative values of distance (unphysical)
- Negative Measurements have no mode / MLE

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:

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

Can’t compute normalisation analytically

Introducing physical constraints into the prior

Exponential decrease in stellar density with Galactic length scale L

Posteriors of distance: want to compute posterior mean

$\omega = 0.01$ $f = \sigma_\omega/\omega$

(Zoom in)

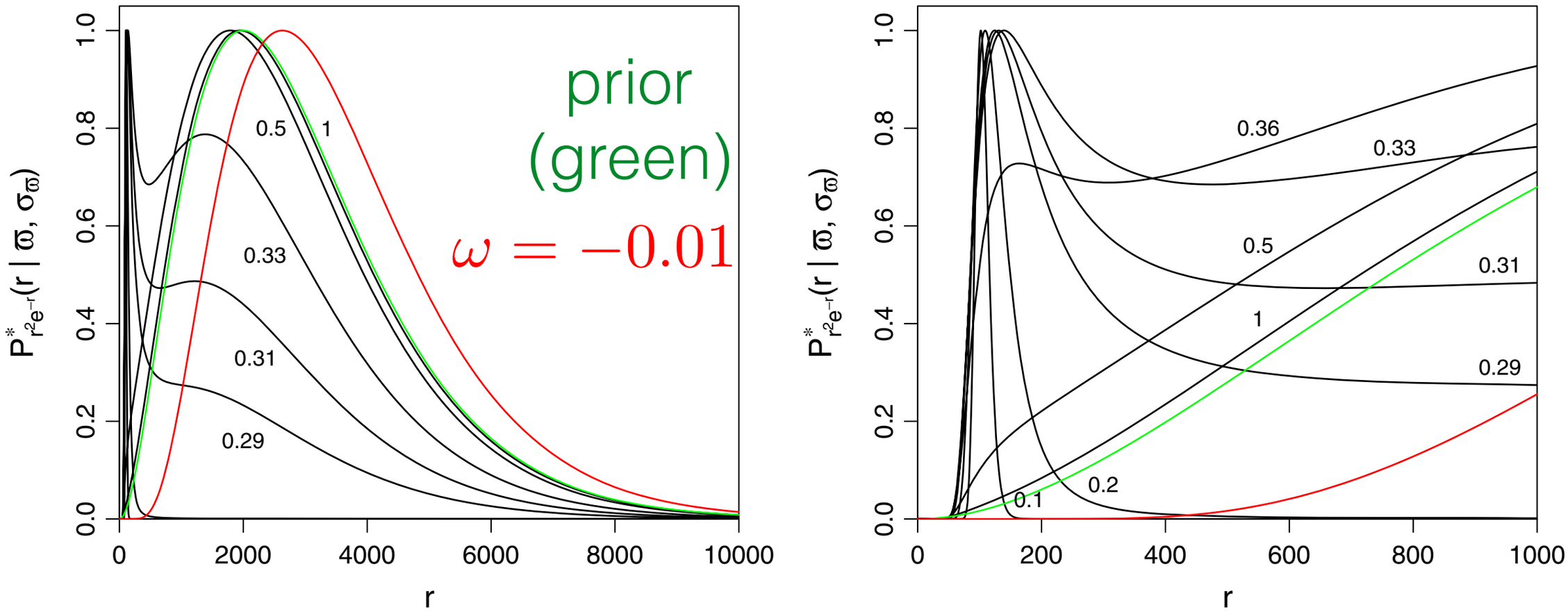
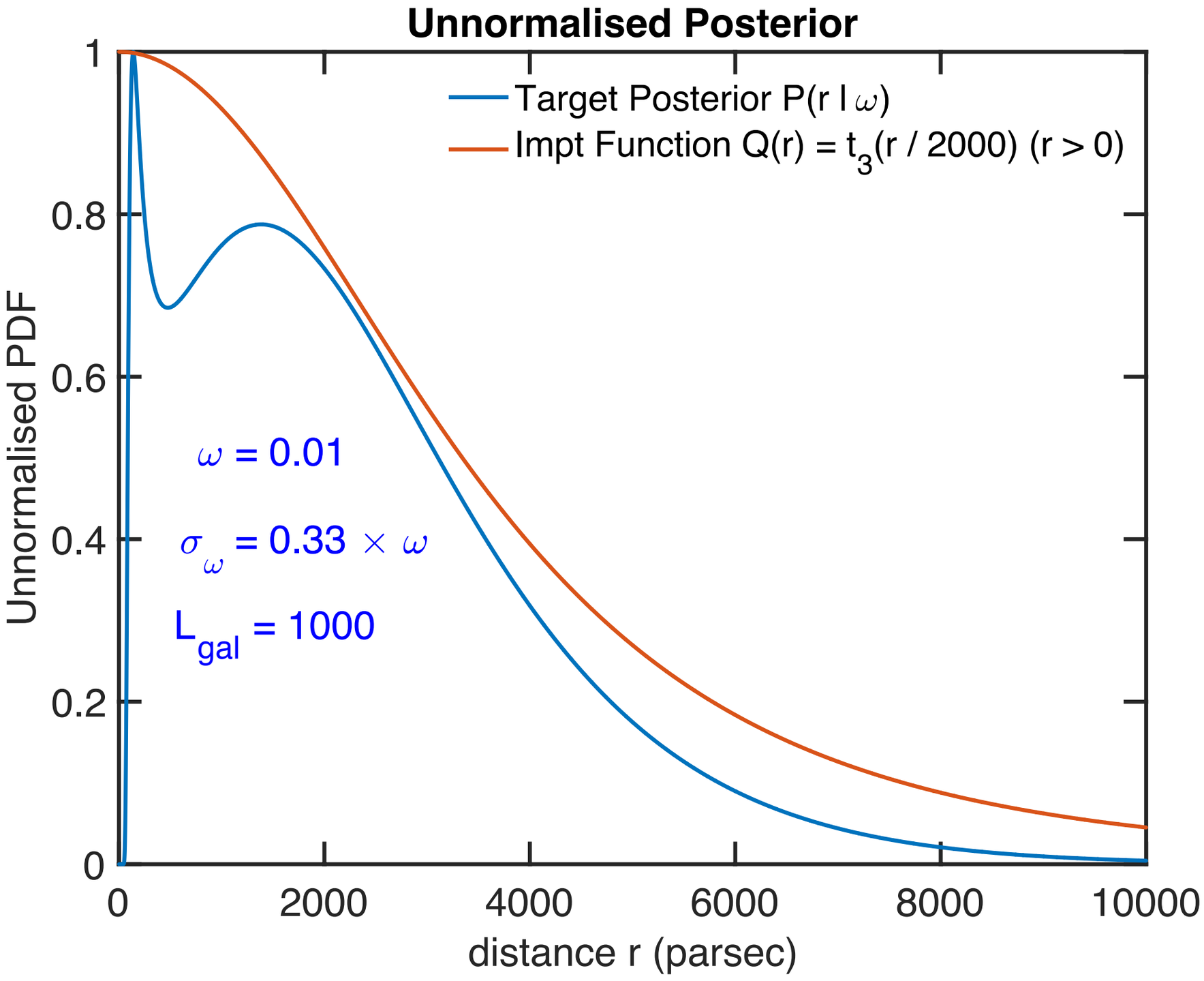


FIG. 12.—The *black lines* in the left panel show the unnormalized posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) for $L = 10^3$, $\varpi = 1/100$ and seven values of $f = (0.1, 0.2, 0.29, 0.31, 0.33, 0.5, 1.0)$. The *red line* is the posterior 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.

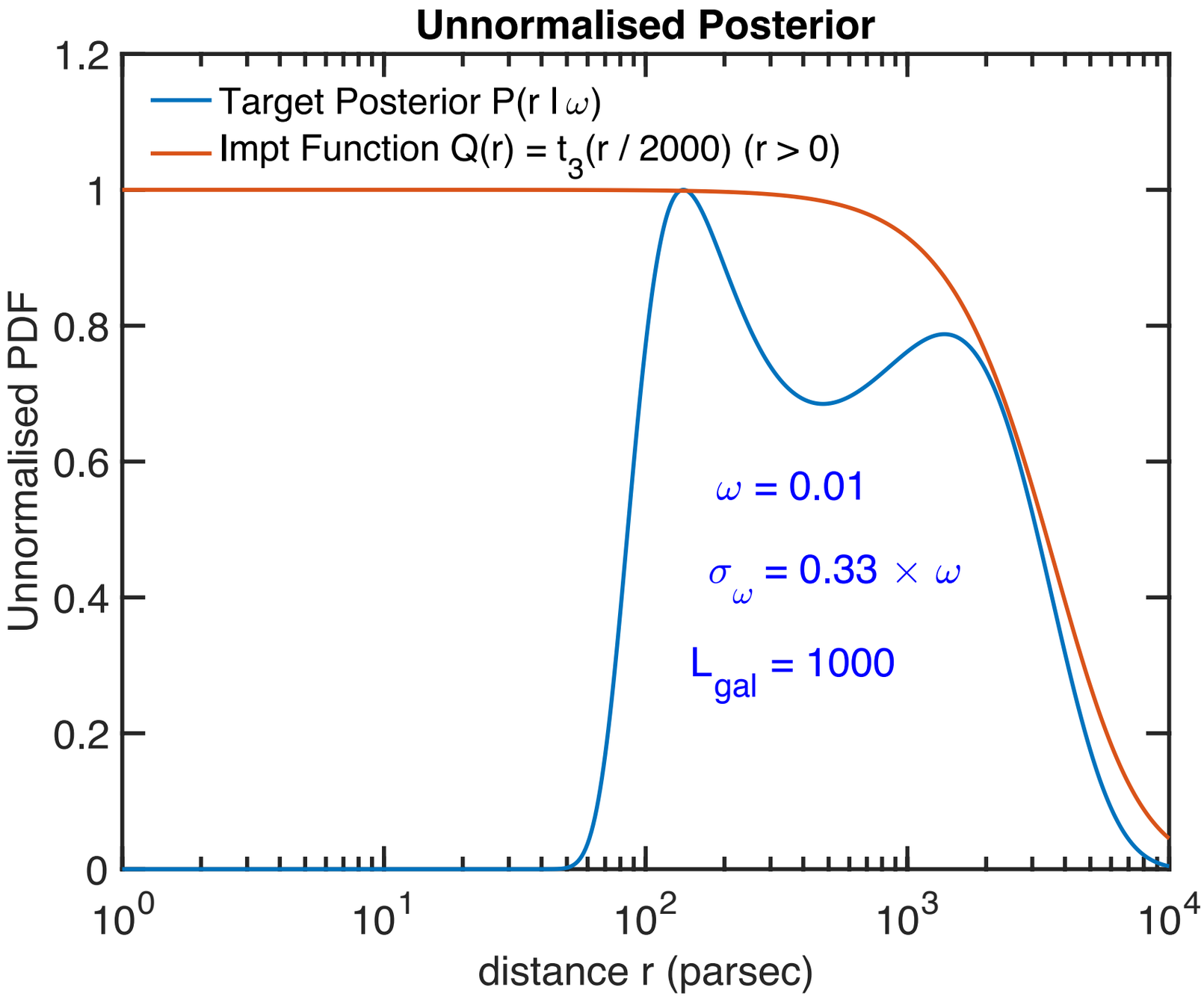
Parallax Example



Importance Function is a half Student-t with ν = 3

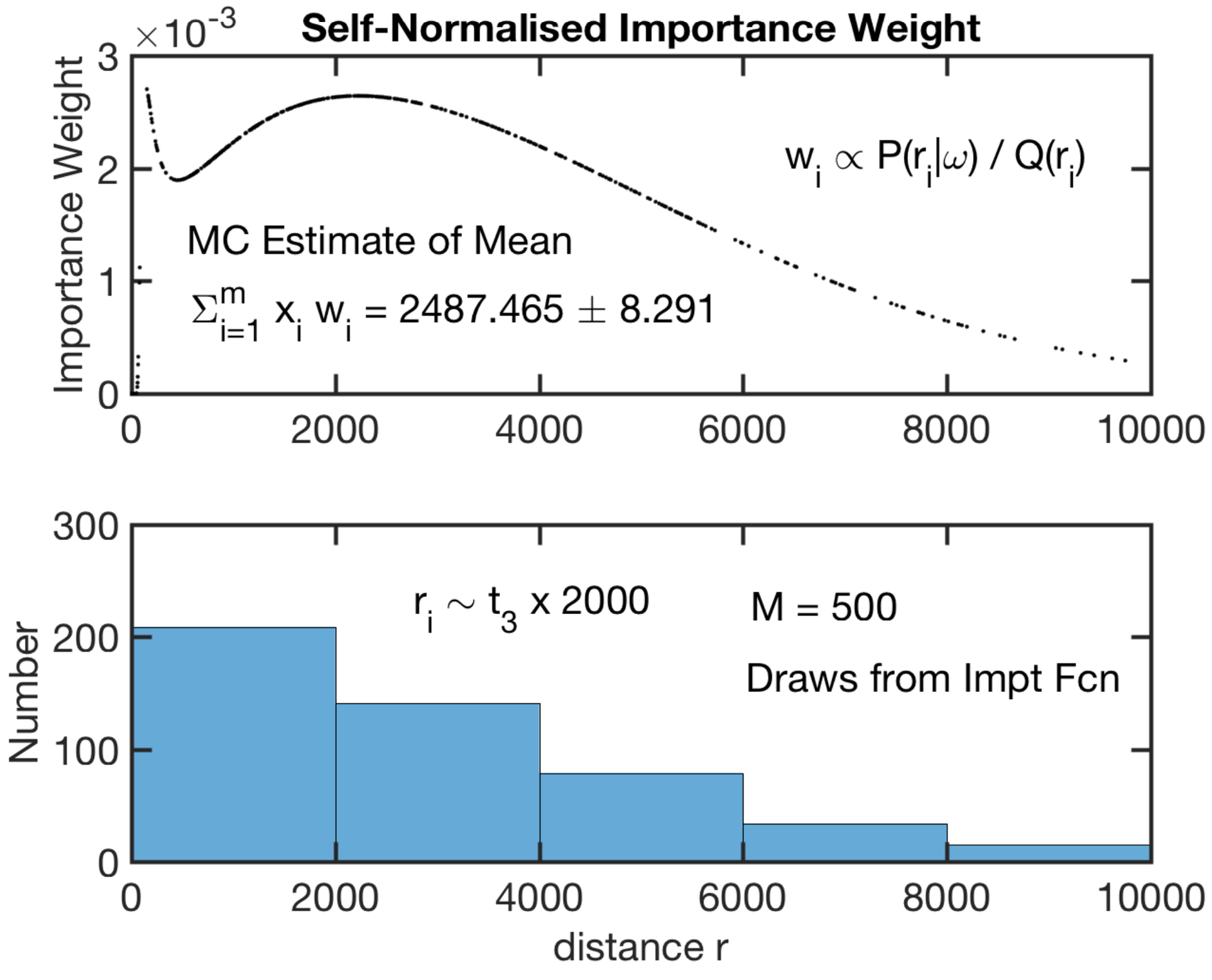
Parallax Example

Importance Function is a Half Student-t with v = 3

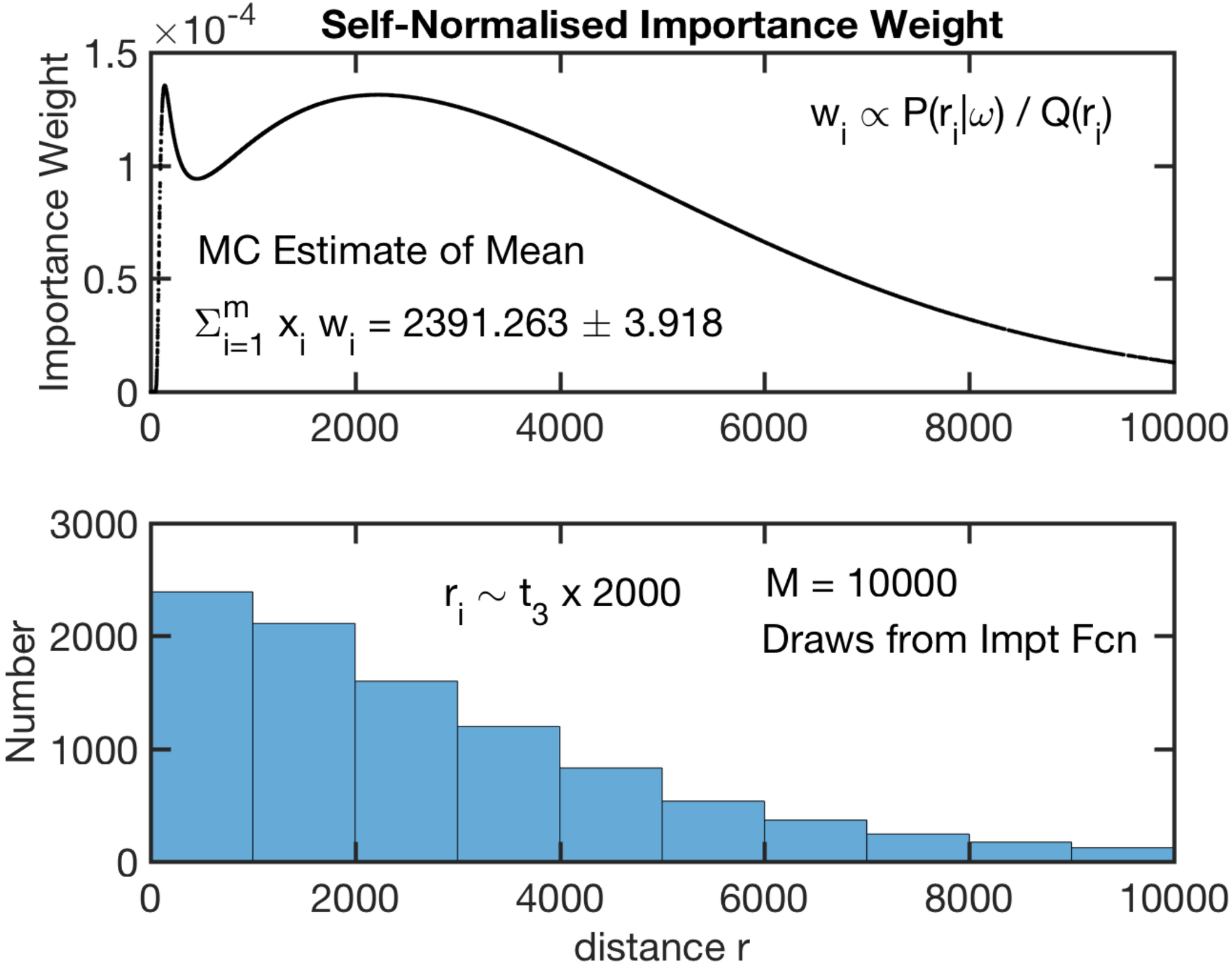


Log Scale

Parallax Example m = 500



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 $|f(\boldsymbol{\theta})|P(\boldsymbol{\theta})$

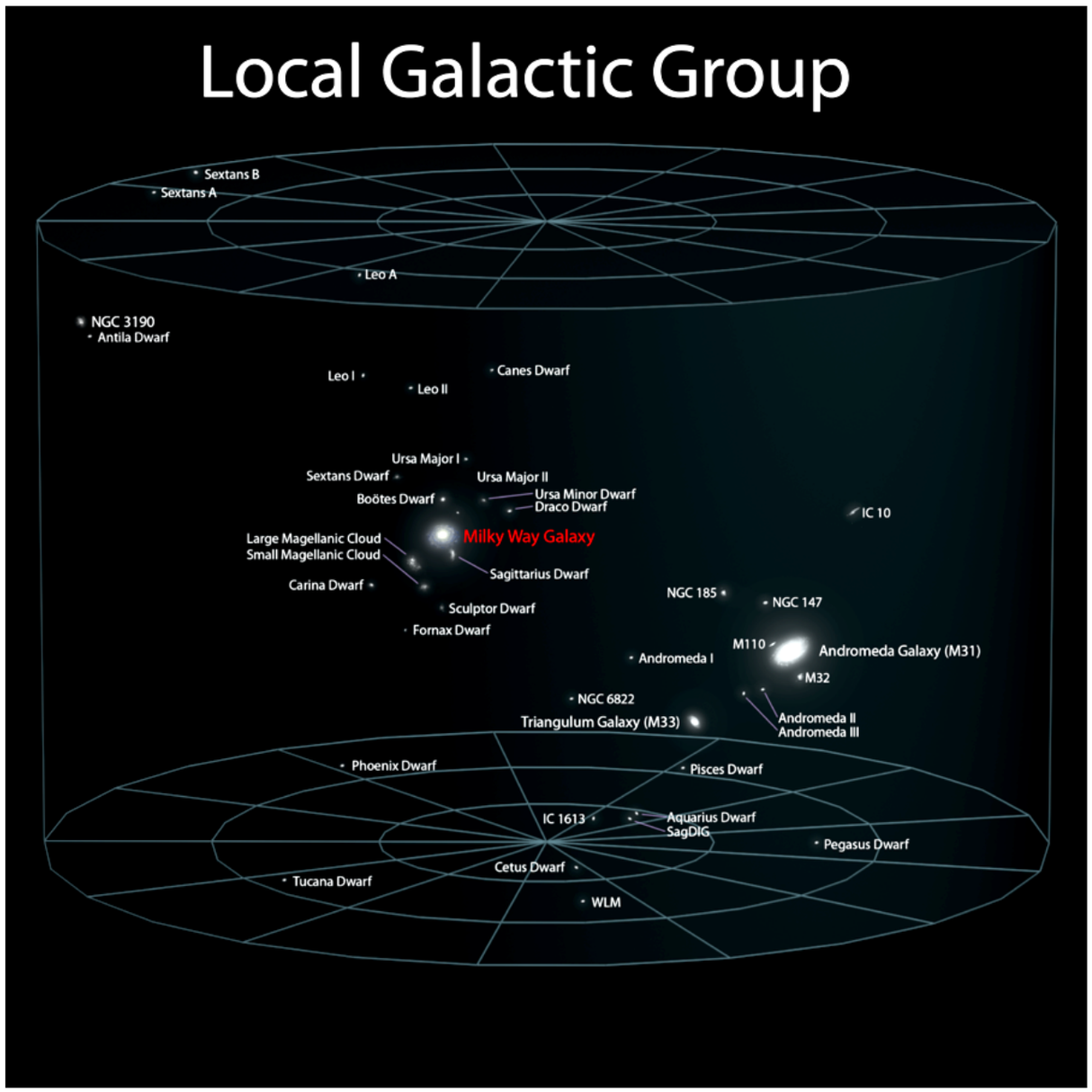
- Want to keep the importance weights roughly constant, otherwise large variations in P(**θ**) / Q(**θ**) will lead to high variance of estimate, smaller ESS
- Effective Sample Size: $\frac{m}{1 + \widehat{\mathrm{Var}}[\{w^*(\boldsymbol{\theta_i})\}]}$
- In practice, find thick-tailed distribution Q(**θ**) that is positive everywhere and similar in shape to | f(**θ**) | P(**θ**)

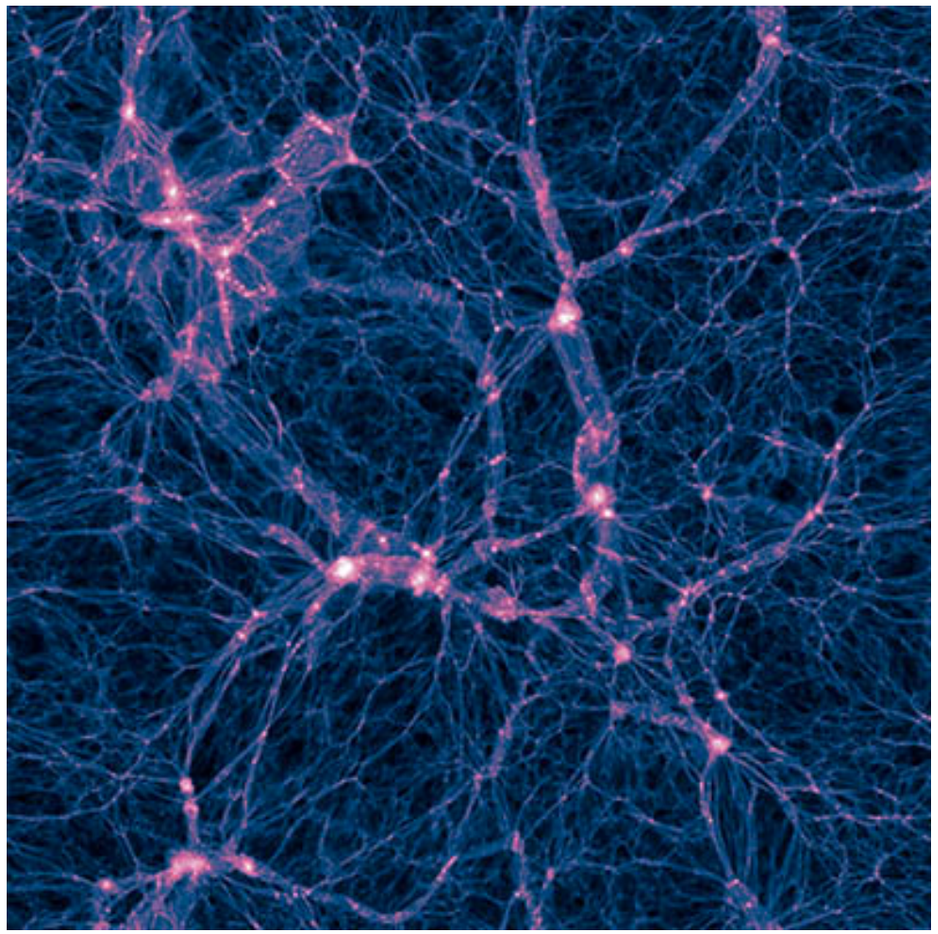
- Don't want Q(**θ**) small when | f(**θ**) | P(**θ**) large!

Astrostatistics Case Study:

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

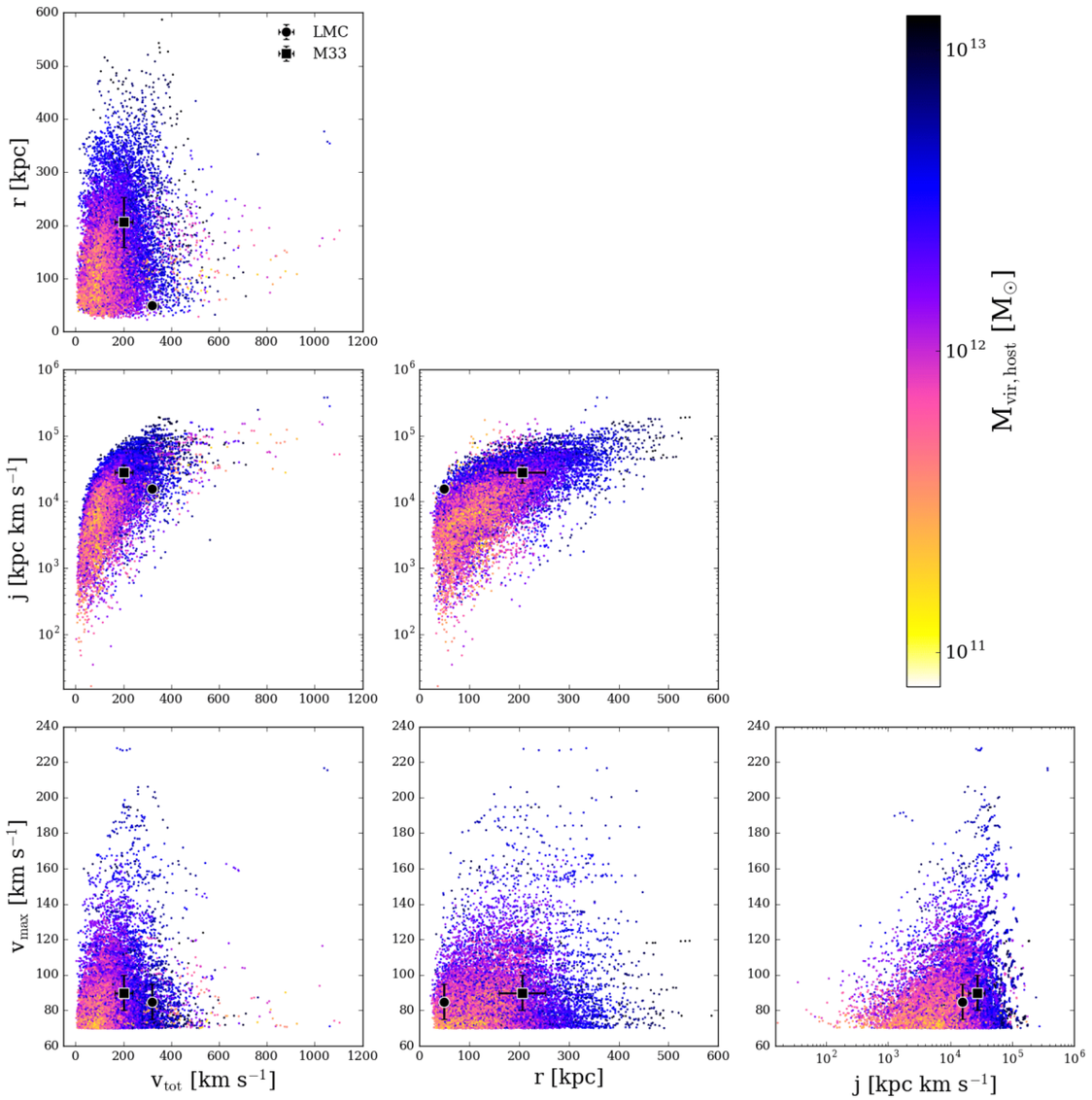
(Patel, Besla, & Mandel, 2017, 2018, arXiv:1703.05767, 1803.01878



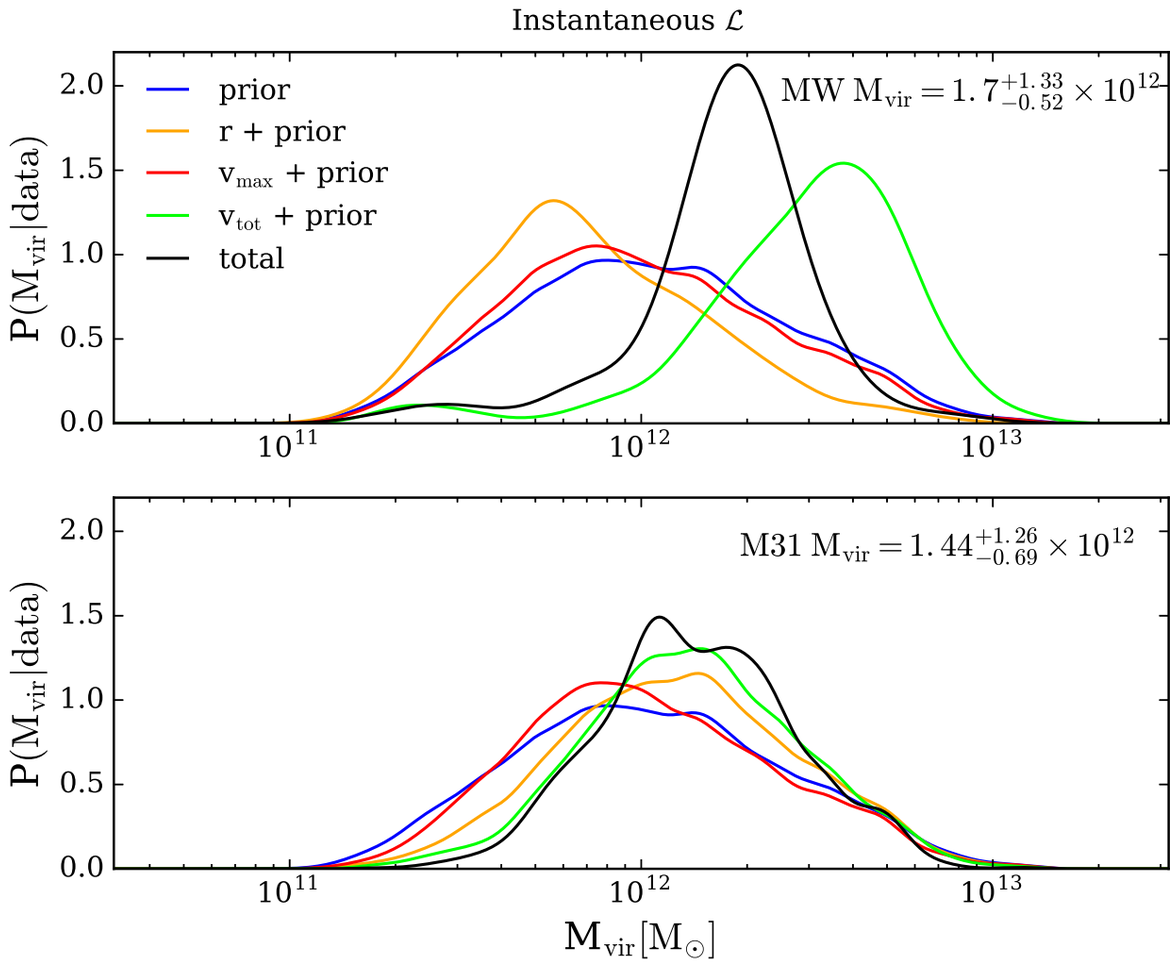


Illustris
Cosmological Simulation of Galaxy Formation

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