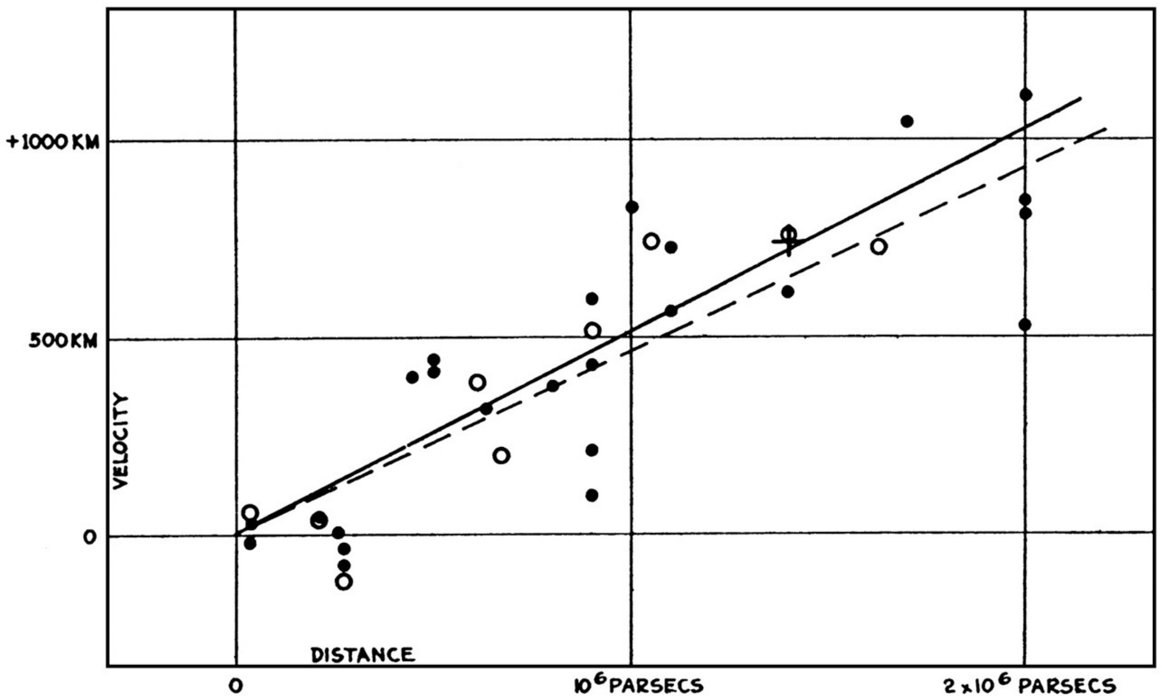
Astrostatistics: Fri 14 Feb 2020

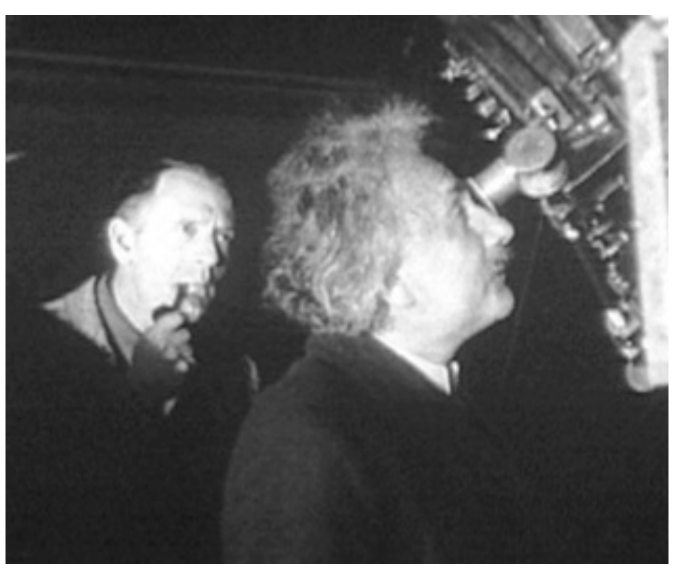
- Remark on Hubble Constant estimation (Ex Sheet 1)
- Bayesian Inference in Astronomy (F&B 3.8, Ivezic 5)
    - finish discussion of:
      C. Bailer-Jones.“Estimating Distances from Parallaxes.”
      2015, PASP,127, 994
      https://arxiv.org/abs/1507.02105
    - Multi-Parameter Bayesian inference
    - Posterior Simulation

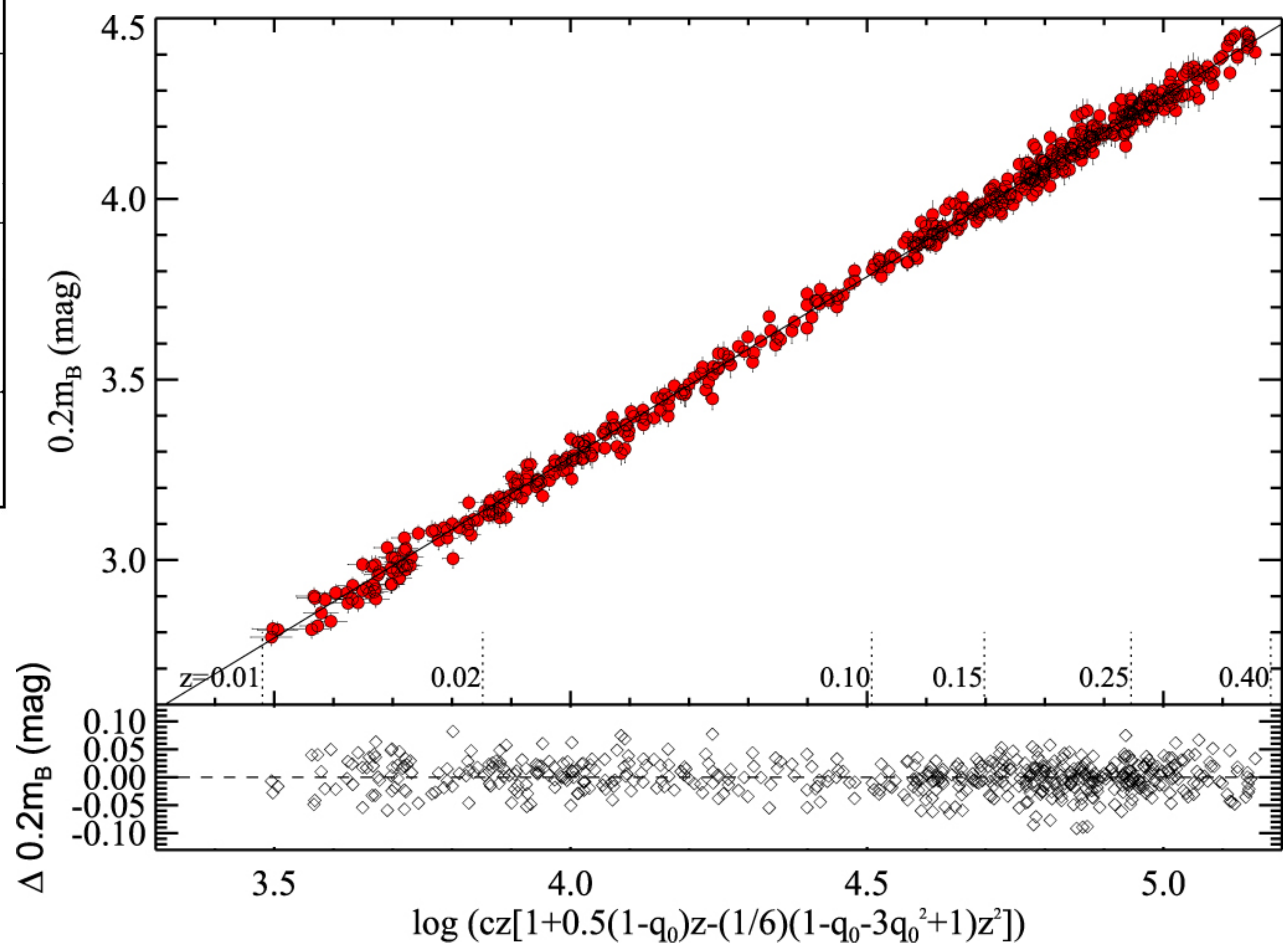
Hubble Constant

Velocity = $H_0 \times$ Distance



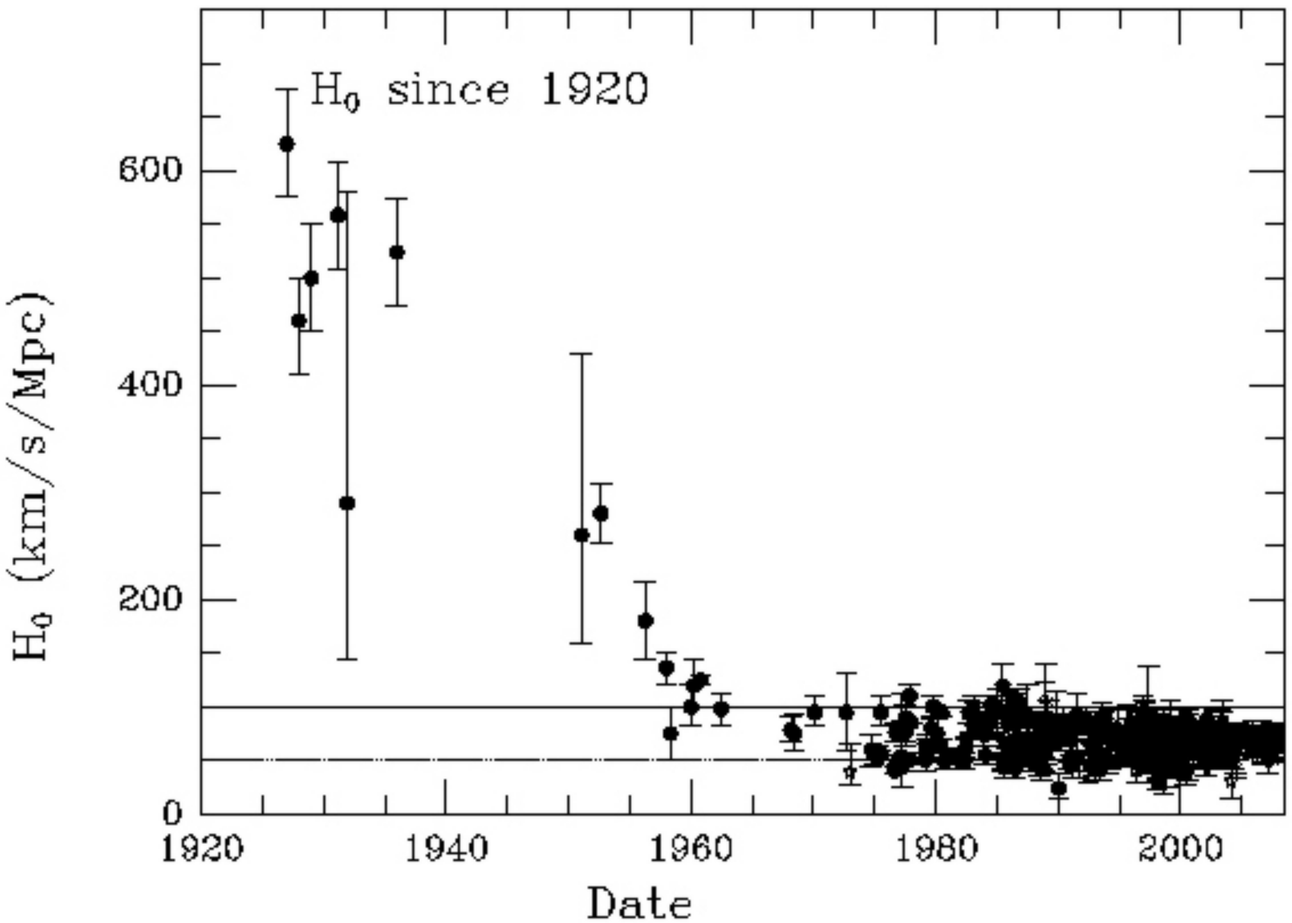
Hubble (1929)



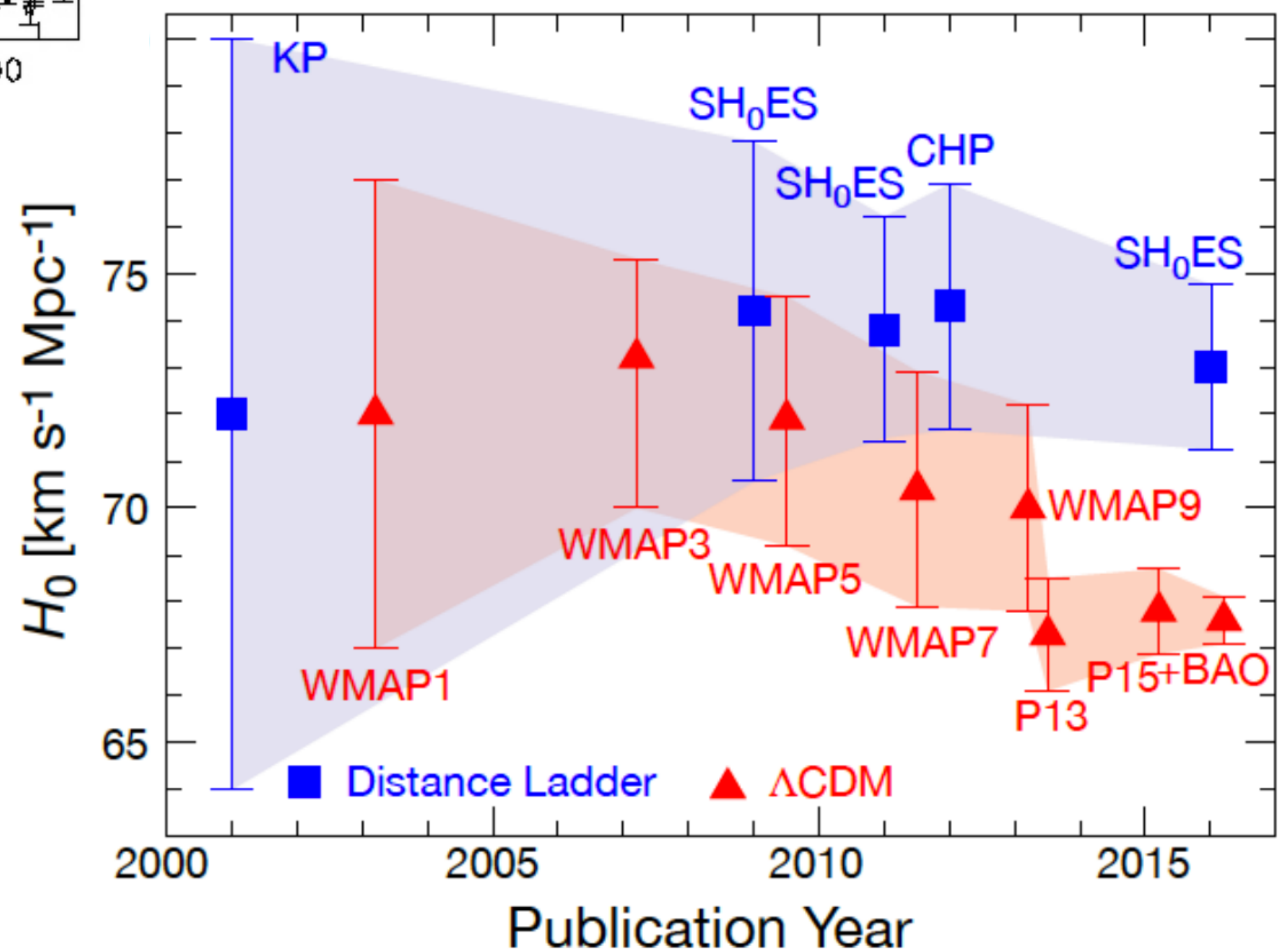


Riess et al. 2016

The Hubble Constant (estimates) Over Time

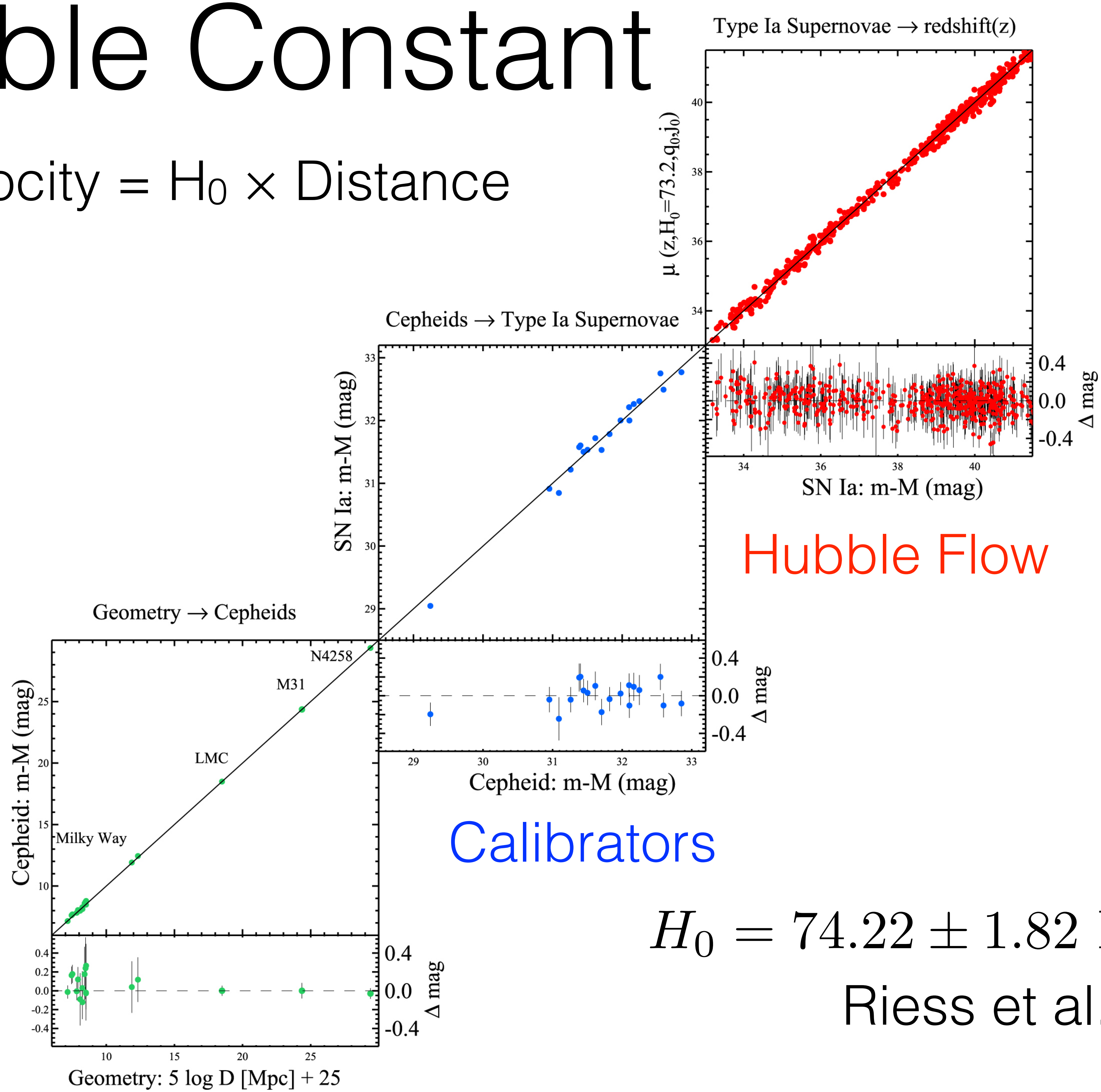


Not So Constant!



Hubble Constant

Velocity = $H_0 \times$ Distance



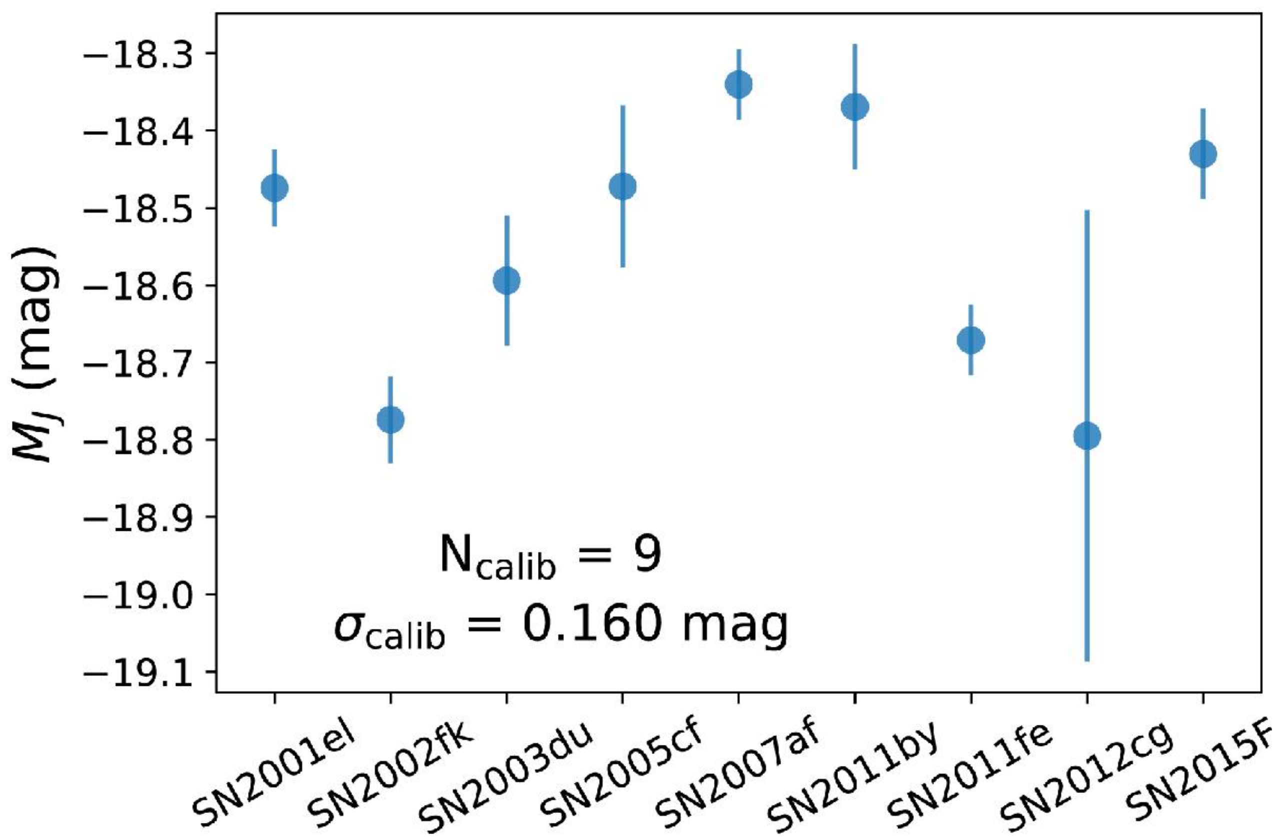
Hubble Flow

Calibrators

$$H_0 = 74.22 \pm 1.82 \text{ km s}^{-1}$$

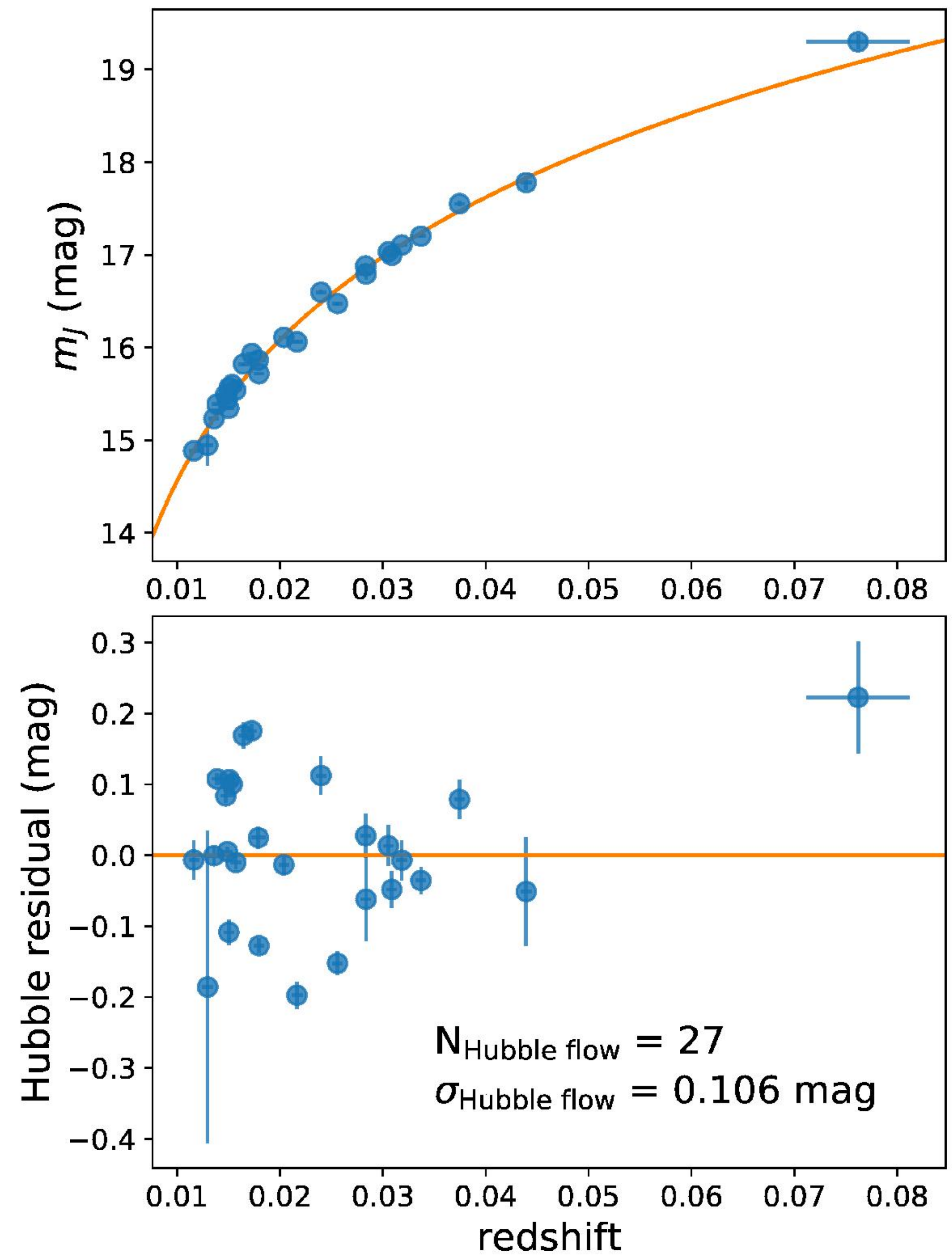
Riess et al. 2019

Example sheet 1: Estimating H_0 using supernovae



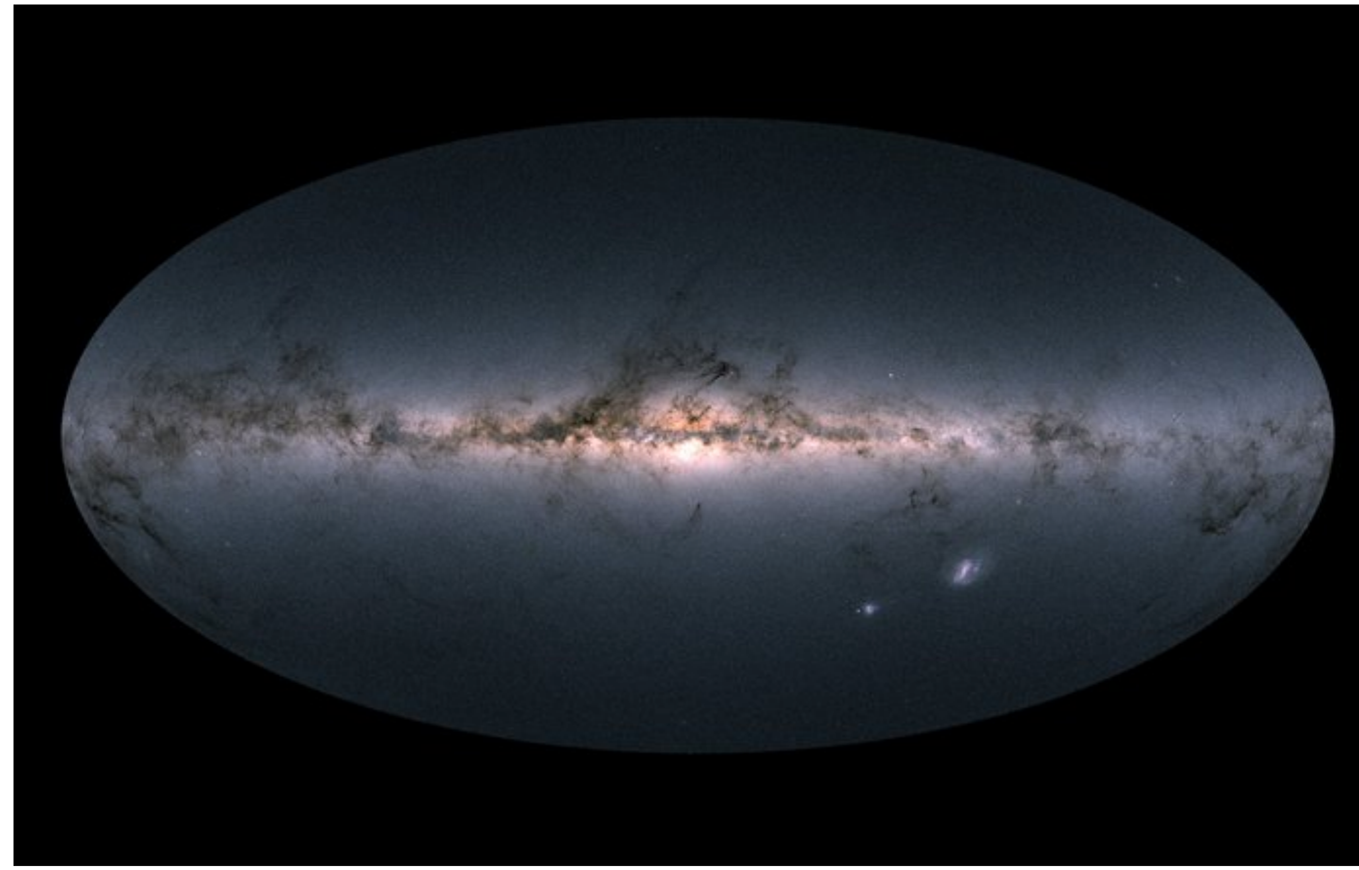
Calibrator Sample

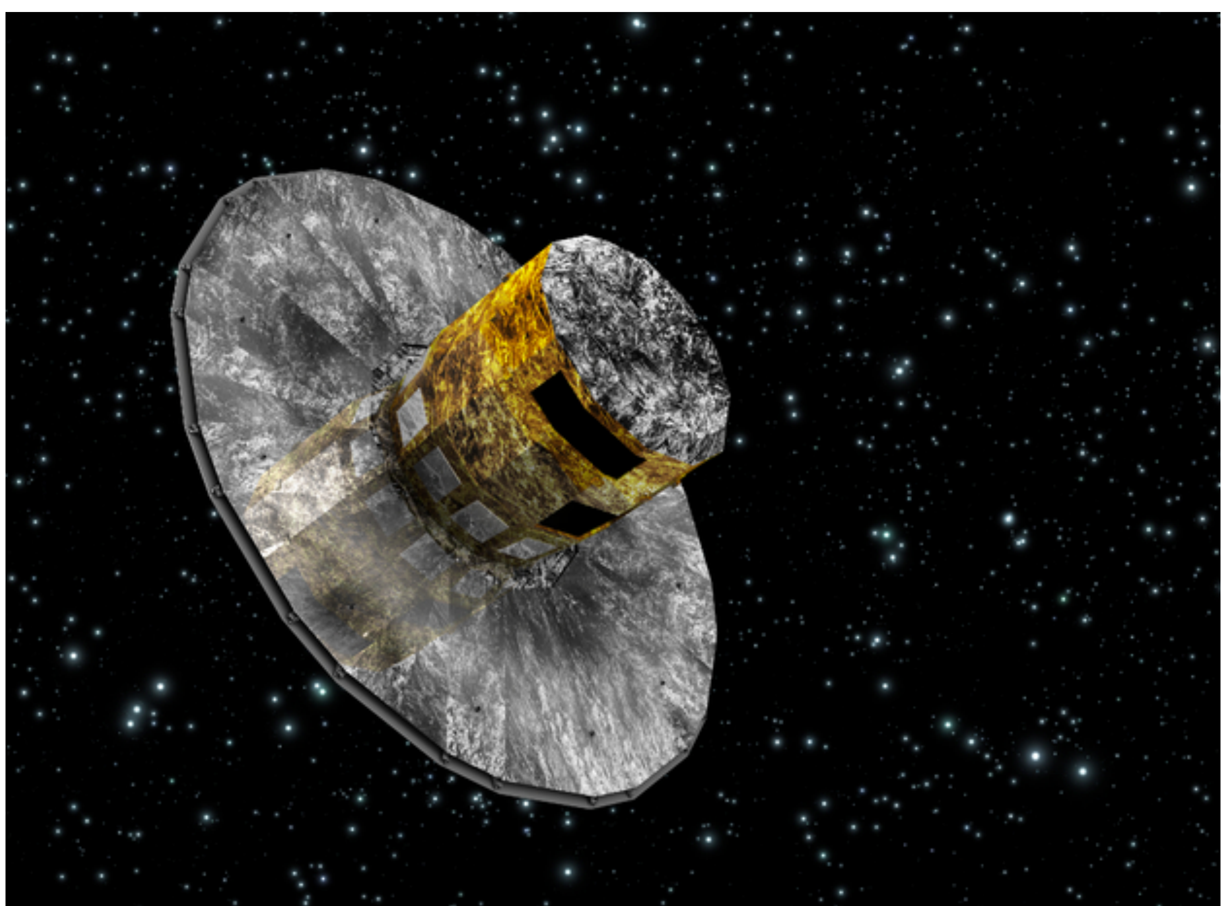
“Measuring the Hubble constant with Type Ia supernovae as near-infrared standard candles.”
Dhawan, Jha, Leibundgut 2018



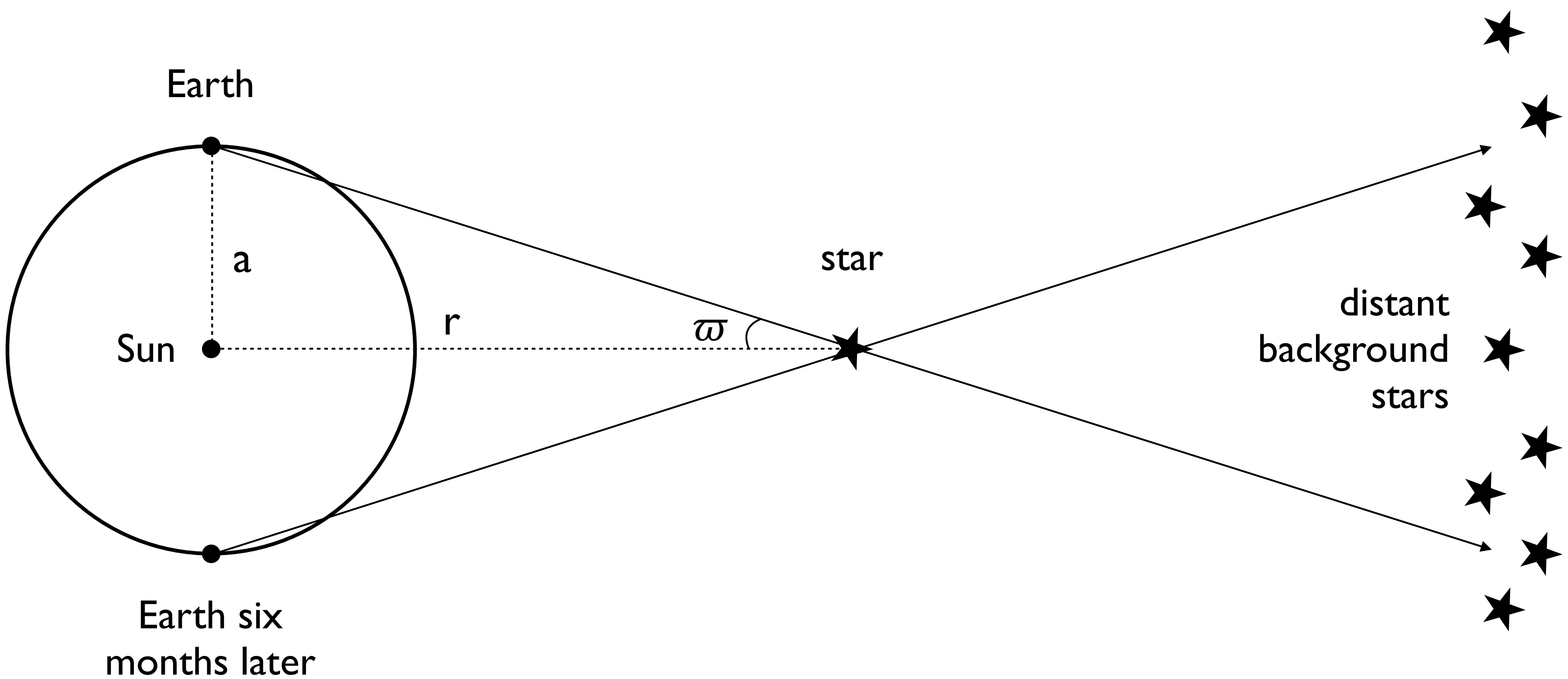
Hubble Flow Sample

Gaia Parallaxes





Parallax Example



True relation (no errors)

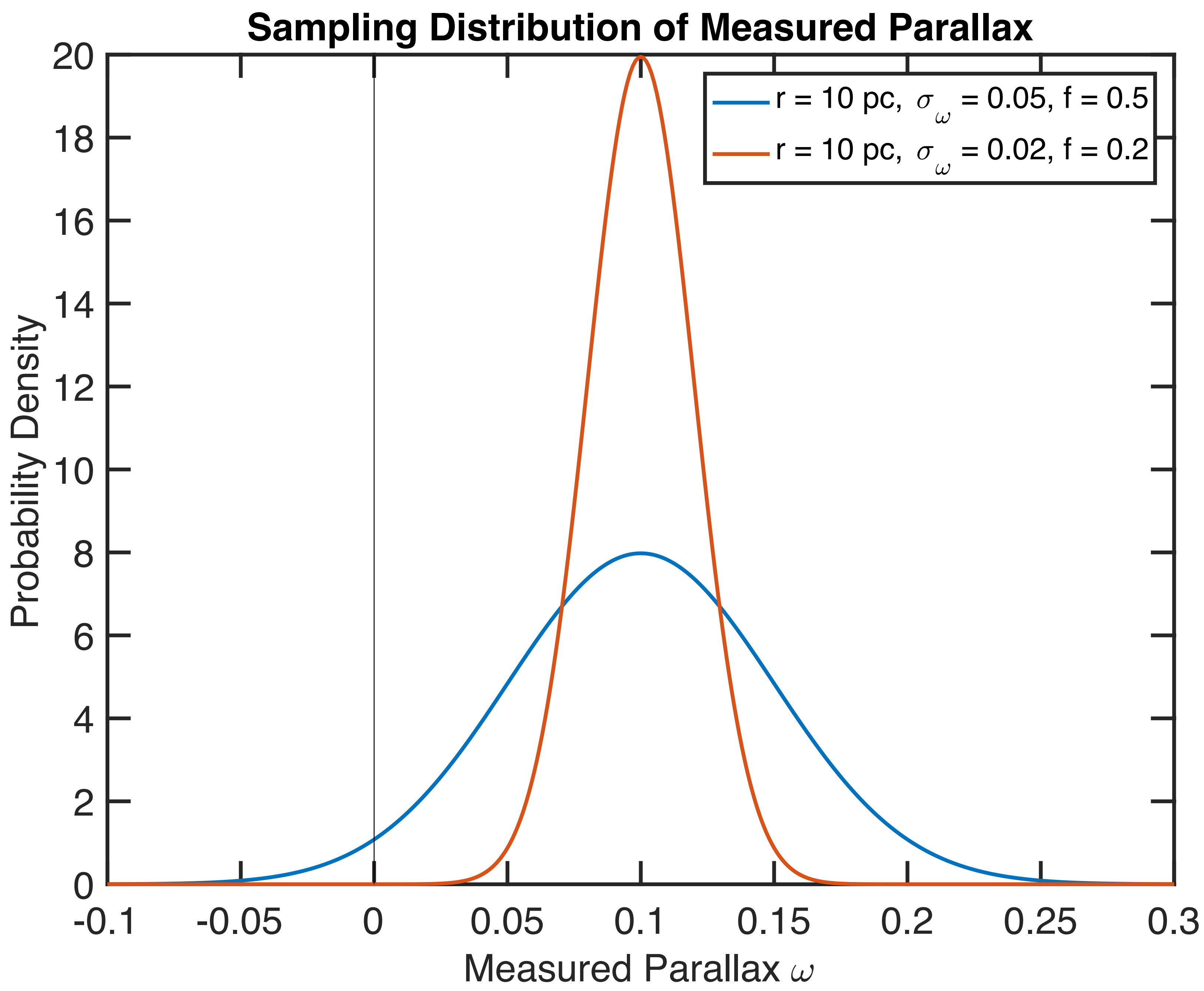
$$\frac{\omega}{\text{arcsec}} = \frac{\text{parsec}}{r}$$

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.

C. Bailer-Jones. "Estimating Distances from Parallaxes." 2015, PASP, 127, 994, https://arxiv.org/abs/1507.02105

Parallax Measurement Error (Gaussian)

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

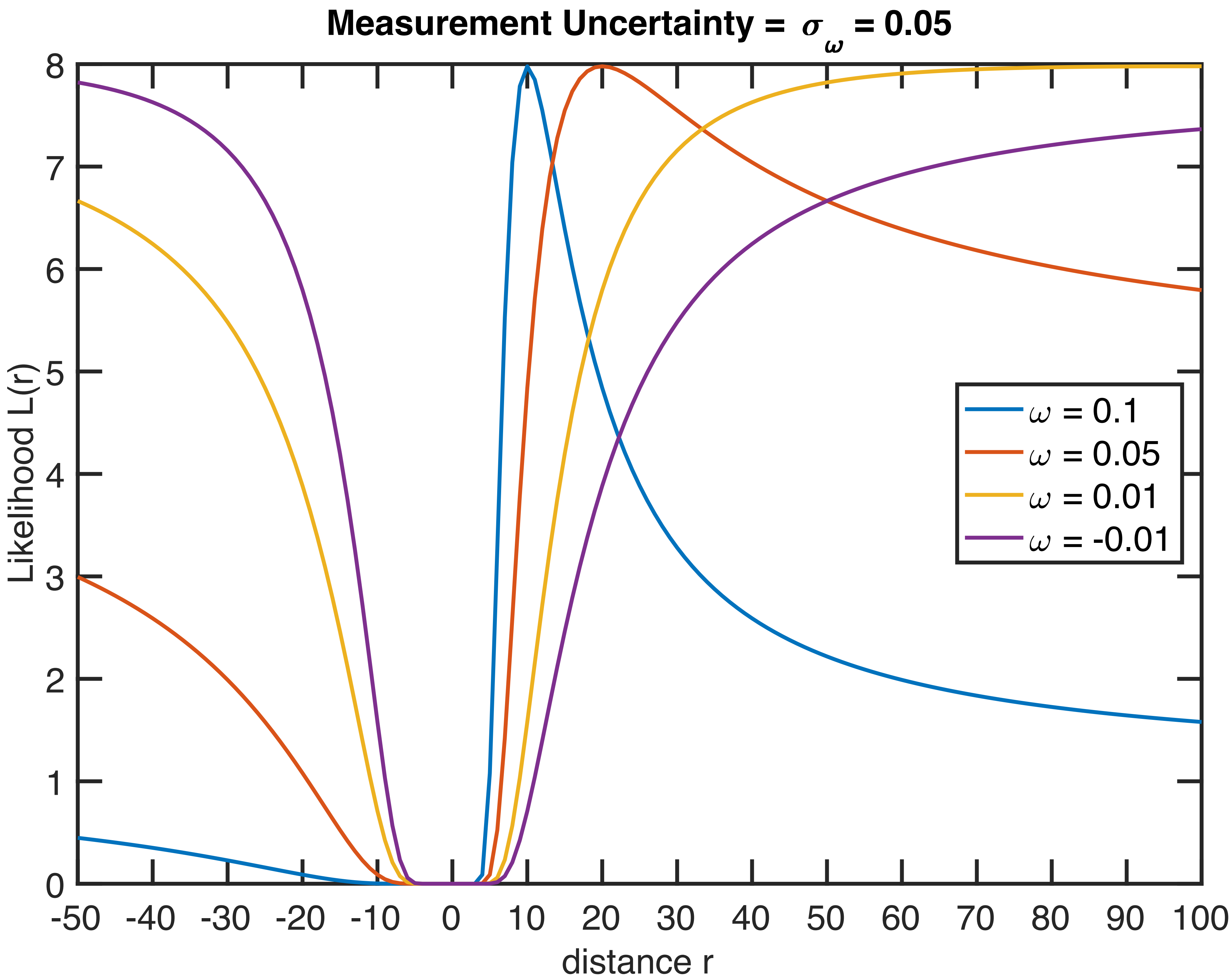


fractional measurement error

$$f = \frac{\sigma_\omega}{\omega}$$

- Negative parallax measurement possible due to noise
- Indicates that distance is likely to be large (true parallax close to zero)
- Contains information

The Futility Function



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

Introducing physical constraints into the prior

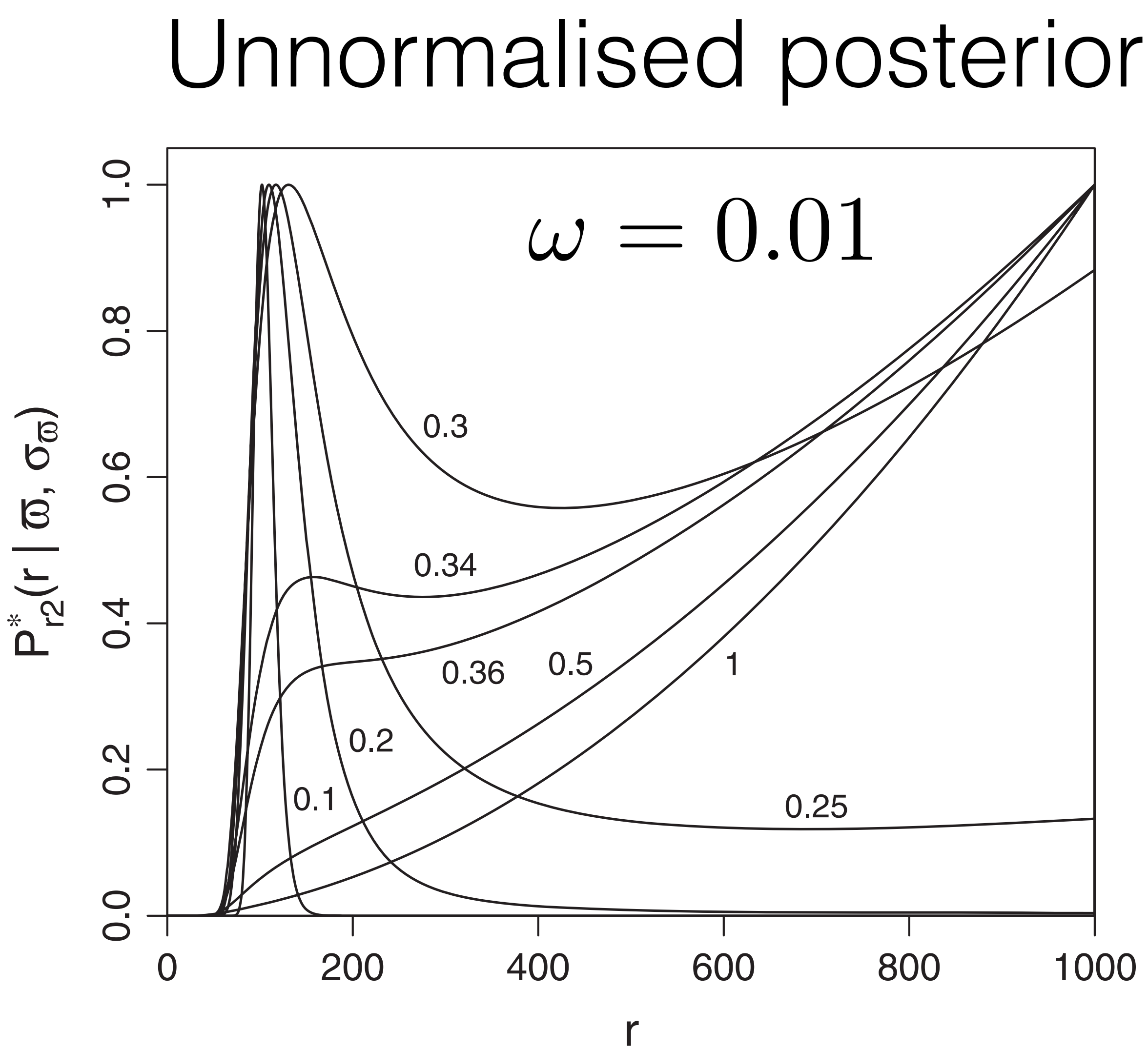
$$P(r) = \begin{cases} \frac{3}{r_{\text{lim}}^3} r^2 & \text{if } 0 < r \leq r_{\text{lim}} \\ 0 & \text{otherwise} \end{cases}$$

Implies a uniform volume density of stars
up to maximum r_{lim}

$$P(r|\omega) \propto P(\omega|r) \times P(r)$$

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

Introducing physical constraints into the prior: Uniform volume density of stars



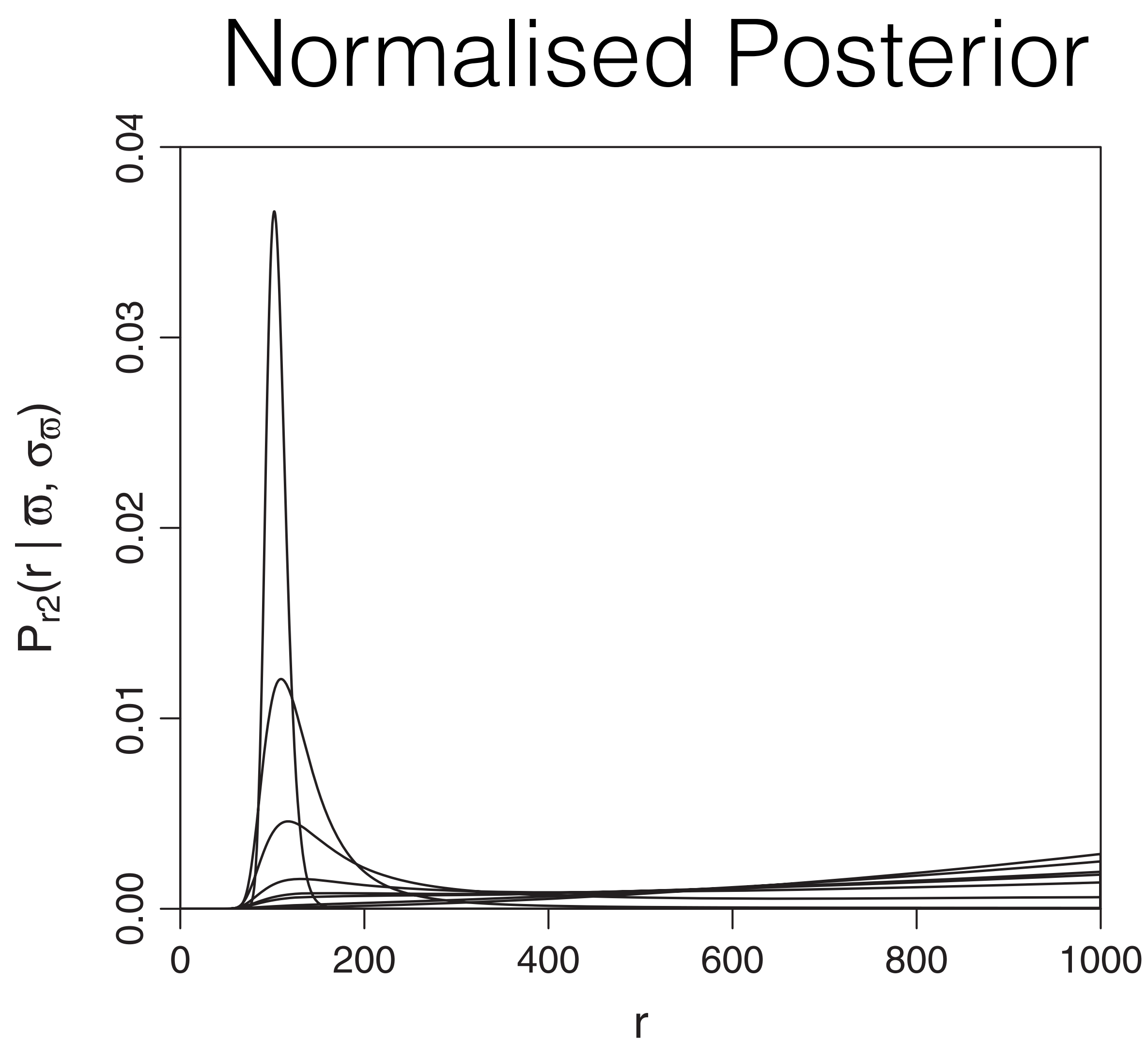
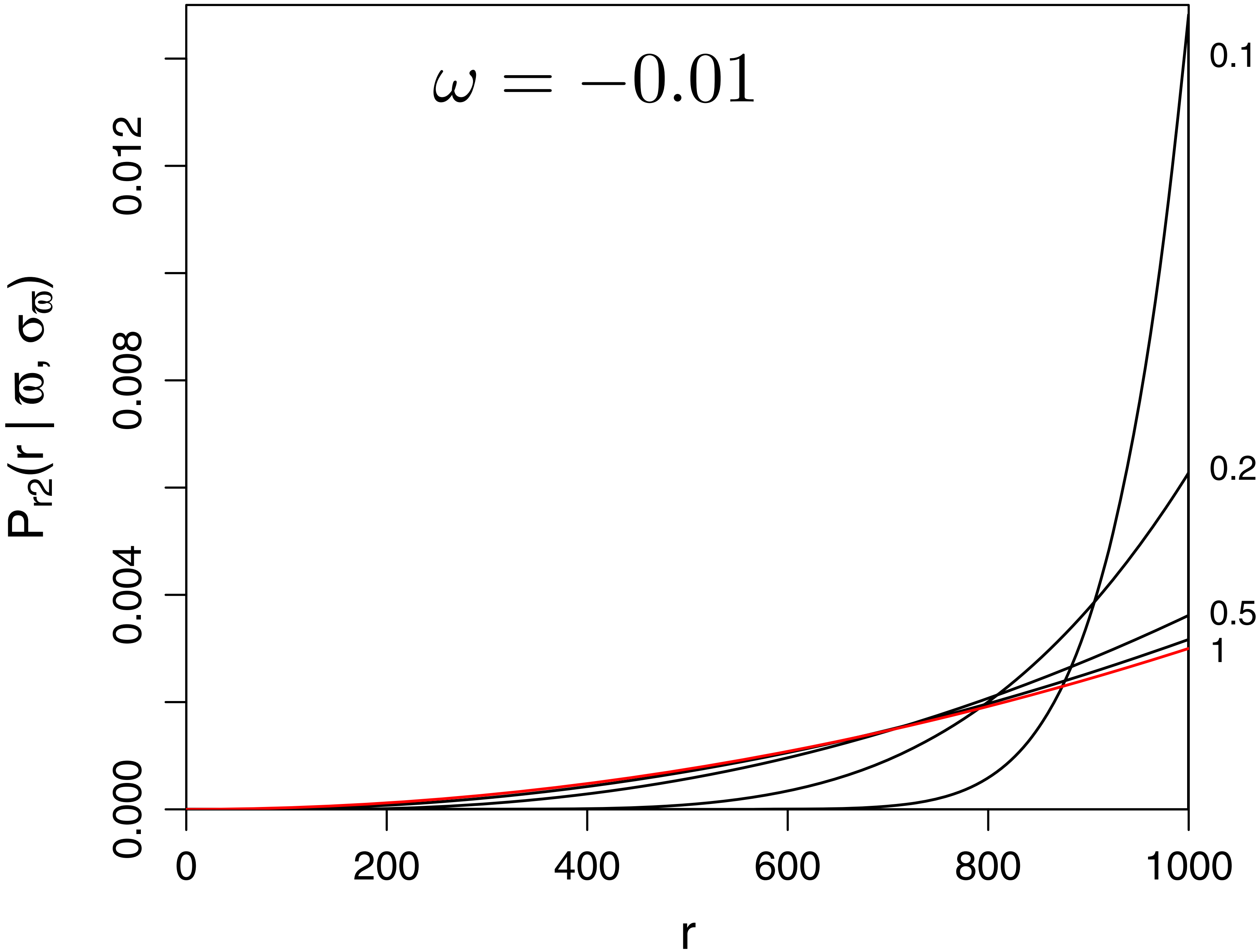


FIG. 7.—Left: the unnormalized posterior $P^*_{r^2}(r|\varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\rm lim} = 10^3$) for $\varpi = 1/100$ and eight values of $f = (0.1, 0.2, 0.25, 0.3, 0.34, 0.36, 0.5, 1.0)$. The posteriors have been scaled to all have their mode at $P^*_{r^2}(r|\varpi, \sigma_\varpi) = 1$. Right: the same posterior PDFs but now normalized. The curves with the clear maxima around $r = 100$ are $f = (0.1, 0.2, 0.25, 0.3)$ in decreasing order of the height of the maximum.

Introducing physical constraints into the prior: Uniform volume density of stars

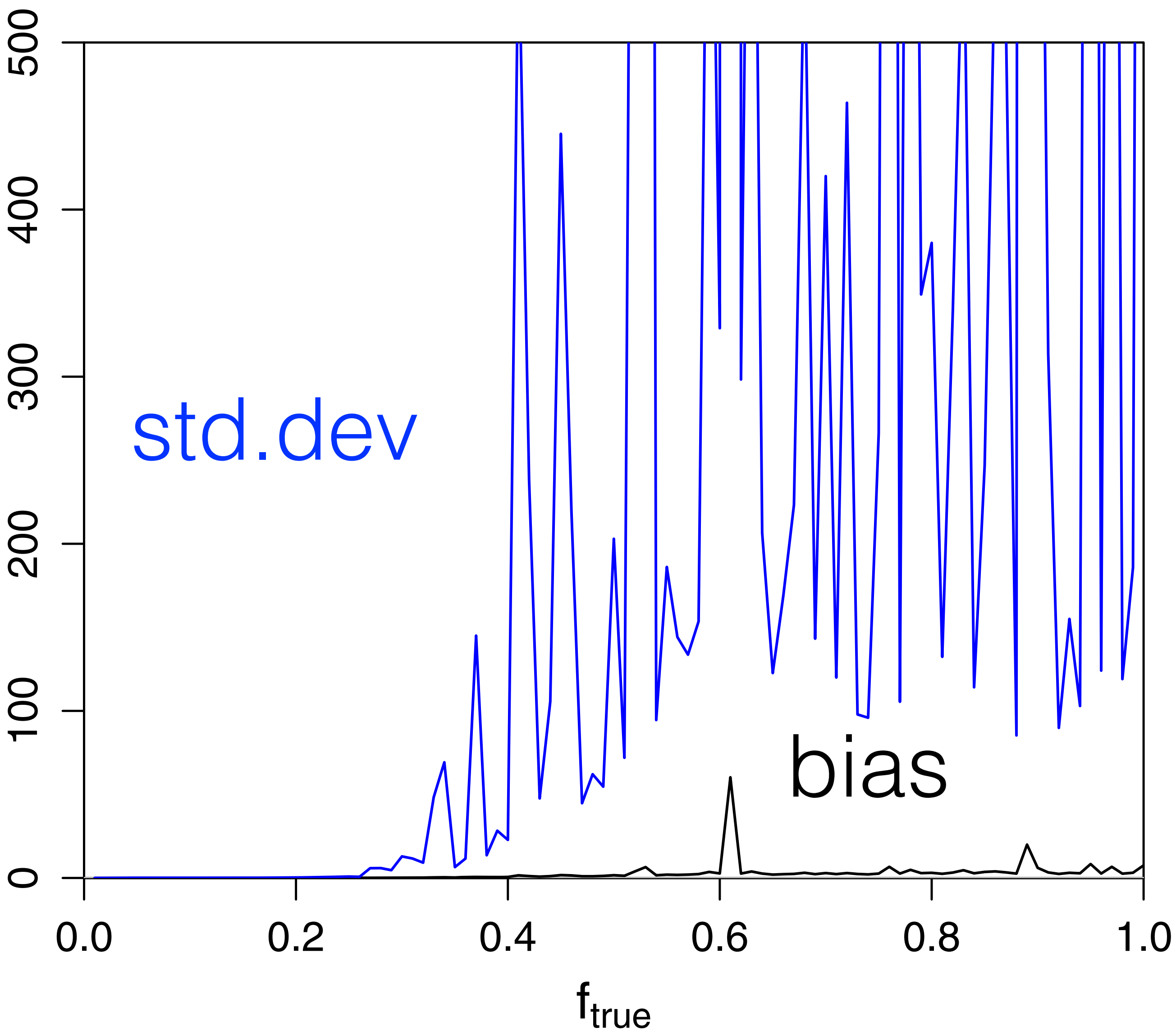


Works for negative and zero parallaxes **ω** but again cutoff at hard limit

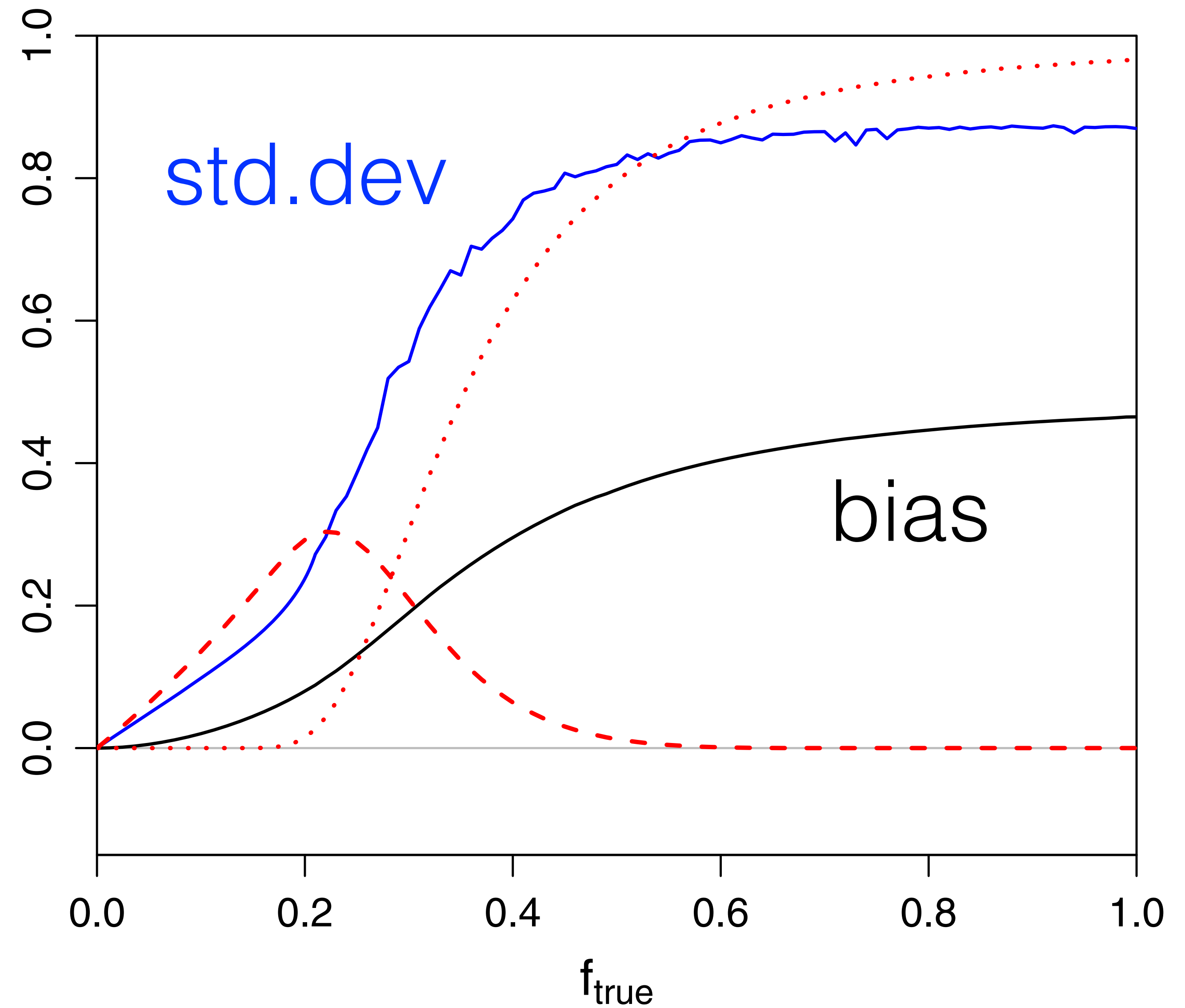
FIG. 9.—The normalized posterior $P^*_{r^2}(r|\varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\rm lim} = 10^3$) for $\varpi = -1/100$ and four values of $|f| = (0.1, 0.2, 0.5, 1.0)$ (*black lines*). The red line shows the posterior for $\varpi = 0$ for $\sigma_\varpi \gg 1/r_{\rm lim}$ (f is then undefined). See the electronic edition of the *PASP* for a color version of this figure.

Frequentist Properties of Posterior Mode

using improper uniform prior



using const vol density prior



sim r_{true} drawn from constant volume density dist'n

$$P_{r^2}(r) = \begin{cases} \frac{3}{r_{\mathrm{lim}}^3} r^2 & \text{if } 0 < r \leq r_{\mathrm{lim}} \\ 0 & \text{otherwise} \end{cases}$$

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 stars with Galactic length scale L

$$P(r|\,\omega) \propto P(\omega|r) \times P(r)$$

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

Introducing physical constraints into the prior

Exponential decrease in stars with Galactic length scale L

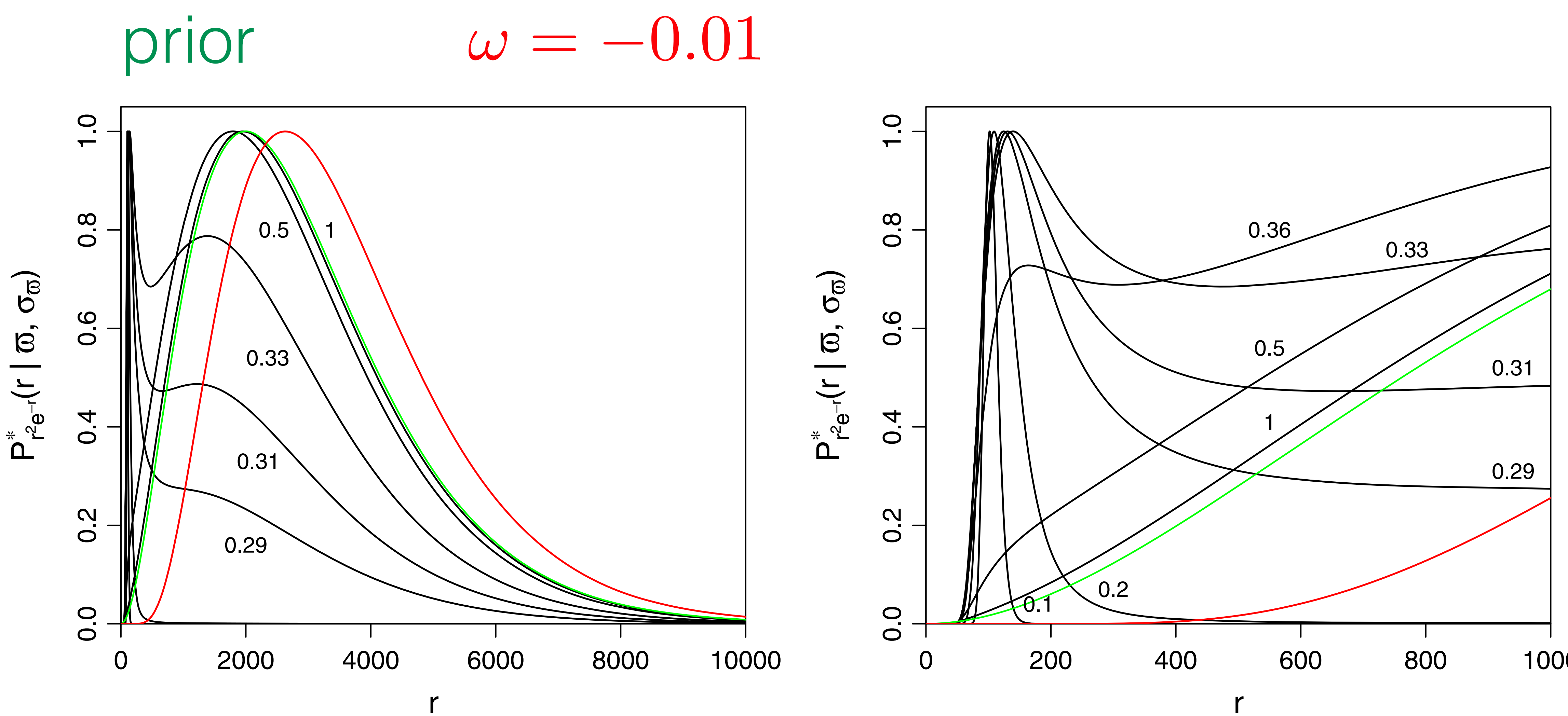


FIG. 12.—The *black lines* in the left panel show the unnormalized posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) for $L = 10^3$, $\varpi = 1/100$ and seven values of $f = (0.1, 0.2, 0.29, 0.31, 0.33, 0.5, 1.0)$. The *red line* is the posterior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

Smooth solutions : no hard edge in prior

Frequentist Properties of Posterior Mode

using exponential density prior

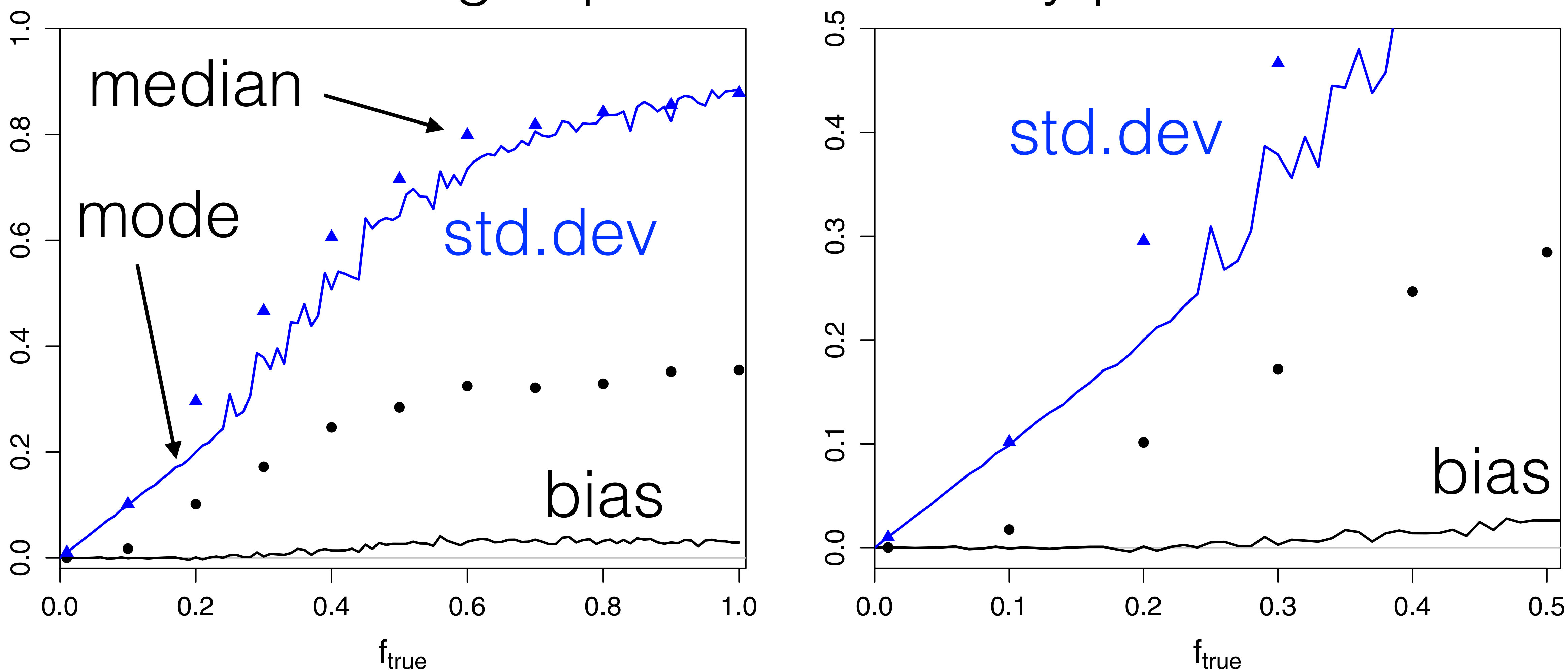


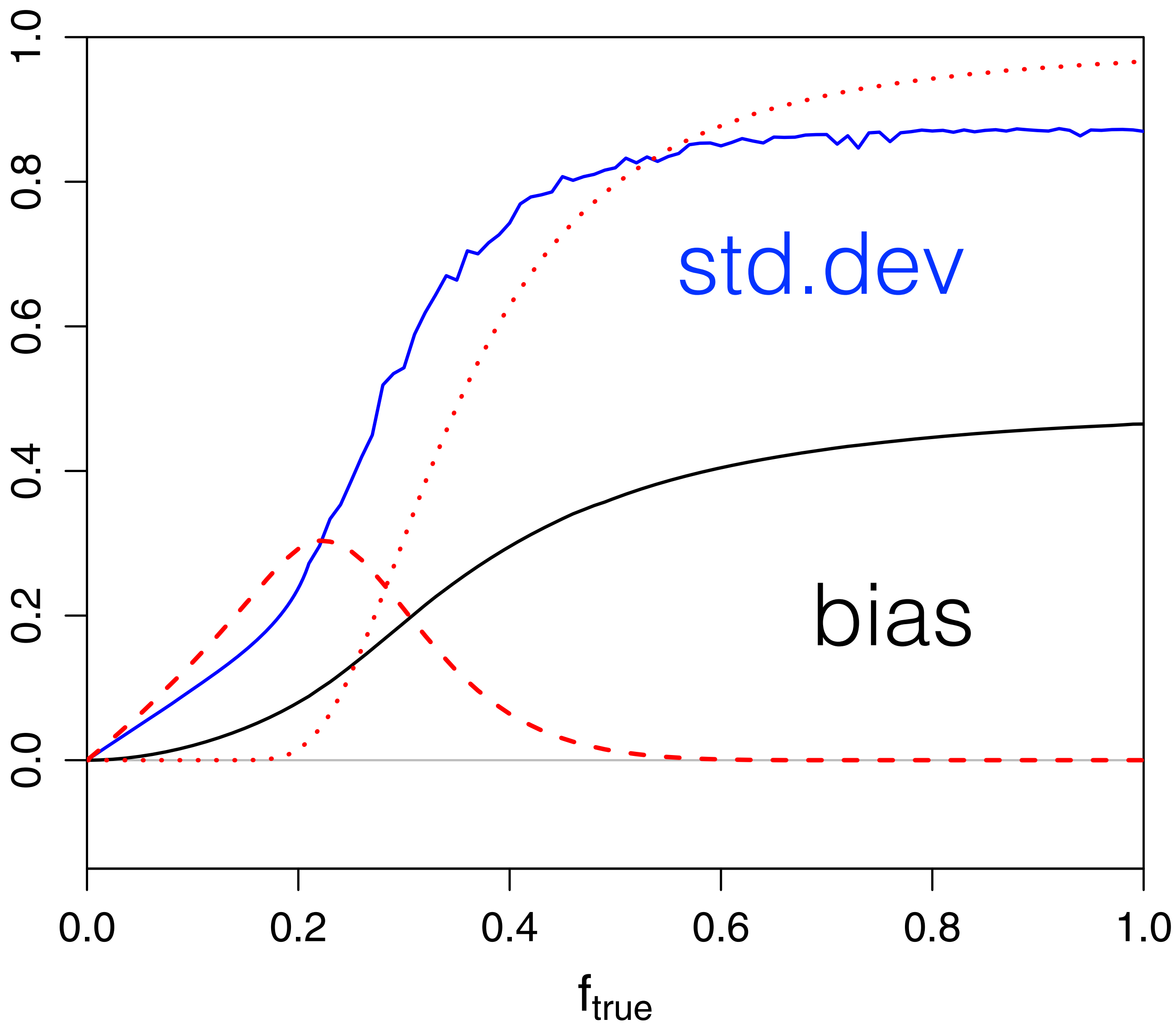
FIG. 14.—The bias (*black line*) and standard deviation (*blue line*) as a function of $f_{\rm true}$ for the mode distance estimator of the posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) with $L = 10^3$ for data drawn from the same prior. The *black circles* and *blue triangles* are the bias and standard deviation, respectively, of the median of the posterior. The right panel is a zoom of the left panel. See the electronic edition of the *PASP* for a color version of this figure.

sim $r_{\rm true}$ drawn from

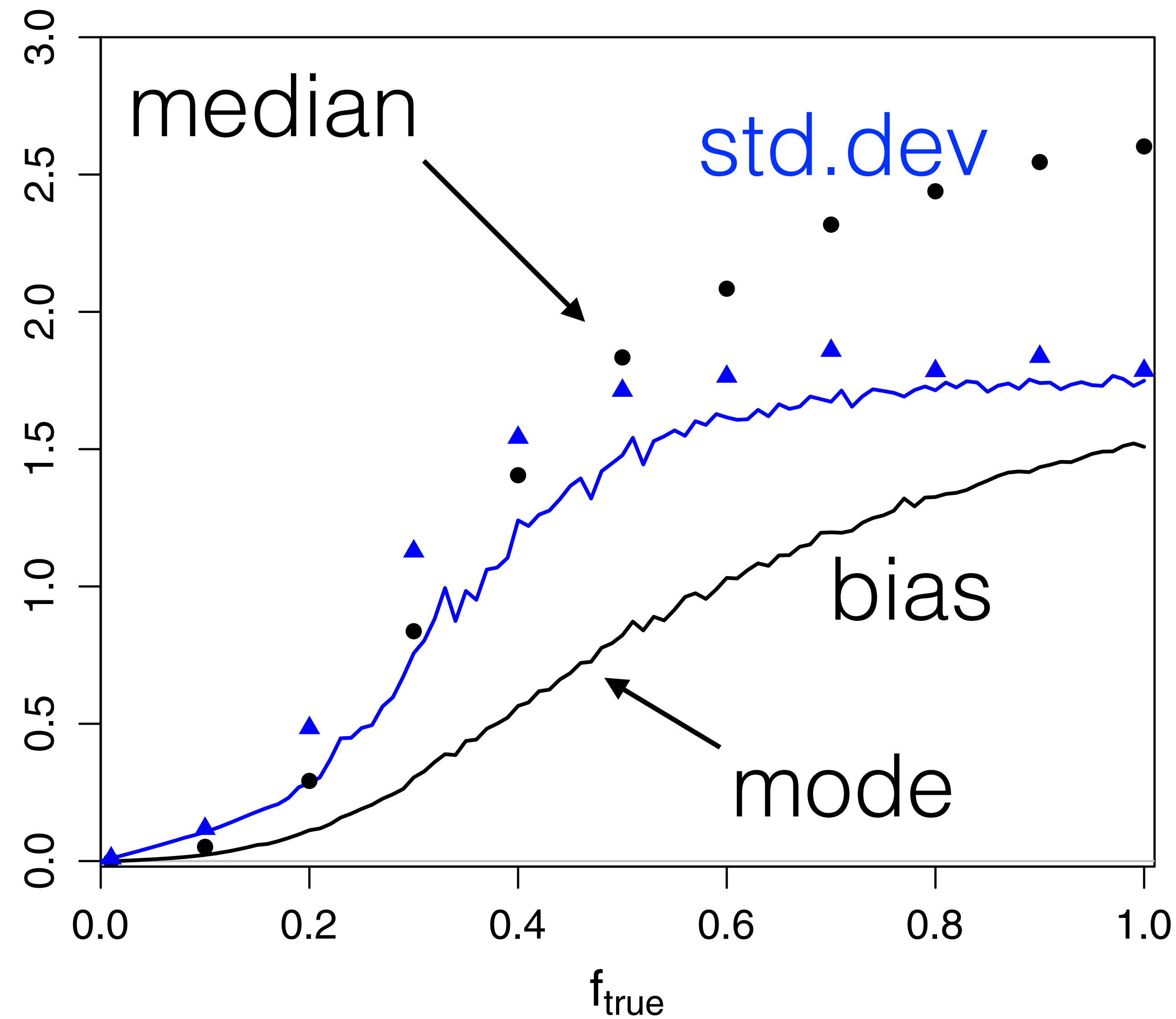
$$P(r) = \begin{cases} \frac{1}{2L^3} r^2 e^{-r/L} & \text{if } r > 0 \\ 0 & \text{otherwise} \end{cases}$$

Frequentist Properties of Posterior Mode

using const vol density prior



using exp density prior



sim r_{true} drawn from constant volume density dist'n

$$P_{r^2}(r) = \begin{cases} \frac{3}{r_{\rm lim}^3} r^2 & \text{if } 0 < r \leq r_{\rm lim} \\ 0 & \text{otherwise} \end{cases}$$

Priors

- Priors can be used to be encoded background information / external knowledge about parameters
- Estimators derived from Bayesian posteriors can be evaluated for their frequentist properties
- Should test sensitivity of your inferences to the priors(s) under various assumptions of the model (including the likelihood)

Mo' Bayes, Mo' problems

- Bayesian answer is the full posterior density P(θ | D), quantifying the "state of knowledge" after seeing the data. Any numerical estimates are attempts to (imperfectly) summarise the posterior. e.g. posterior mean, modes, 95% Highest Posterior Density (HPD) region(s).

- Often these are posterior expectations:
$$\mathbb{E}[f(\boldsymbol{\theta})|D)] = \int f(\boldsymbol{\theta})P(\boldsymbol{\theta}|D)d\boldsymbol{\theta}$$
which are often computationally difficult

- Bayesian computation: Algorithms to ``map out'' and/or sample the posterior density P(θ | D) and compute expections **E**[f(θ) | D]

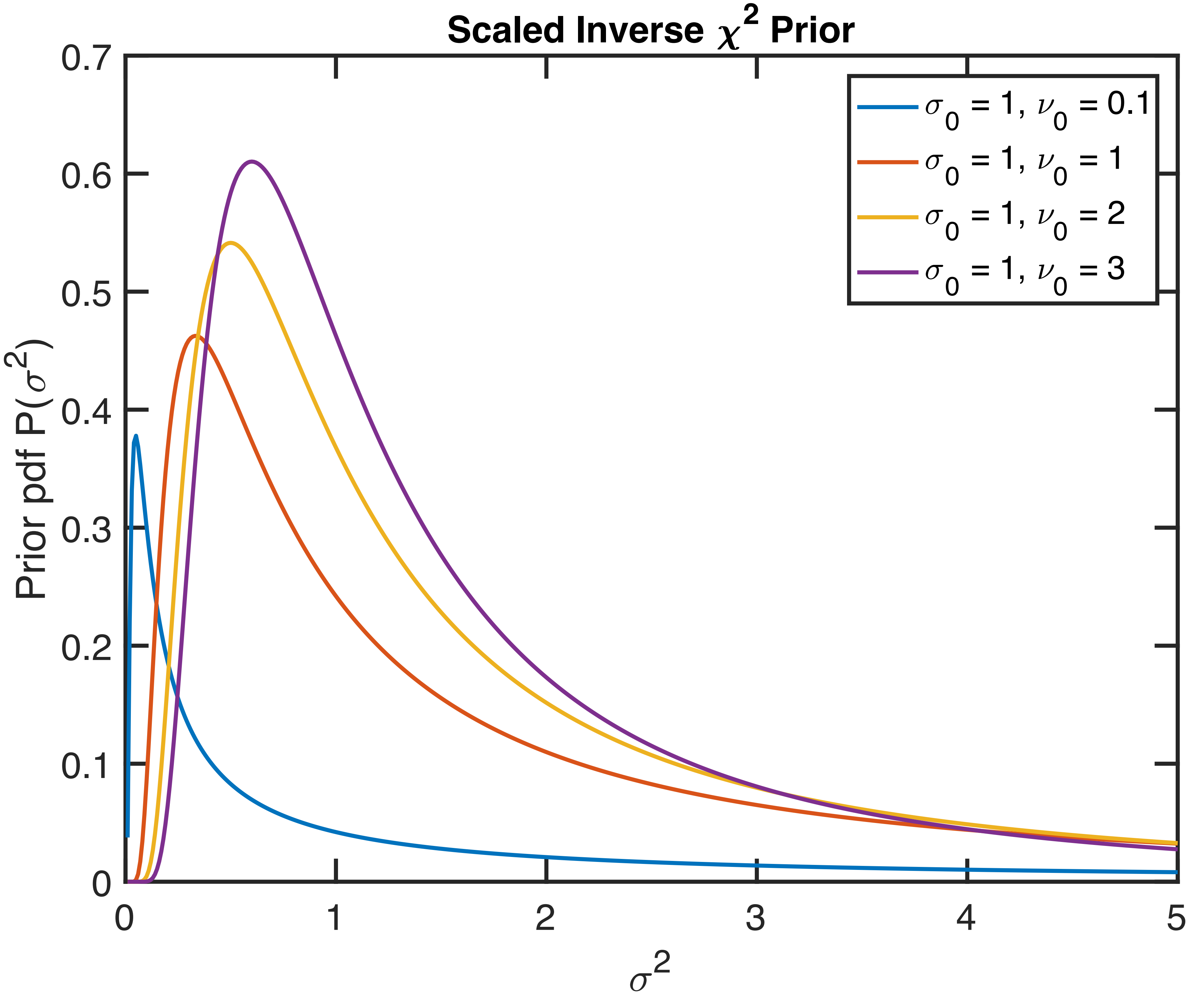
- Markov Chain Monte Carlo [Metropolis, Gibbs, Ensemble/emcee, HMC/Stan], Nested Sampling, Particle Filtering/Population MC, Importance Sampling

- All models are wrong, some are useful!

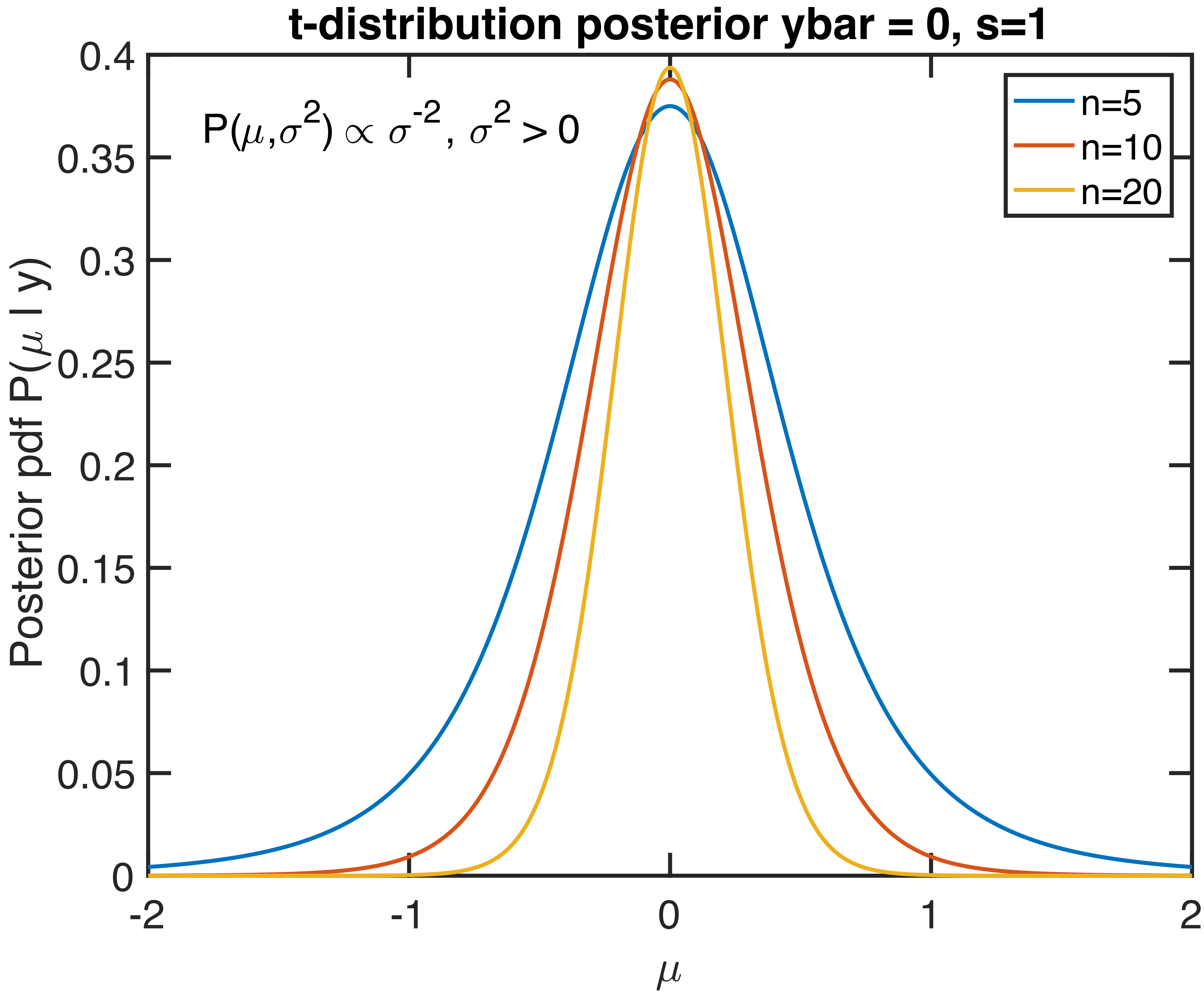
- Testing model fit, predictive checks, model comparison

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

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



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



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

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

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

(indicator function)

Monte Carlo Direct Sampling

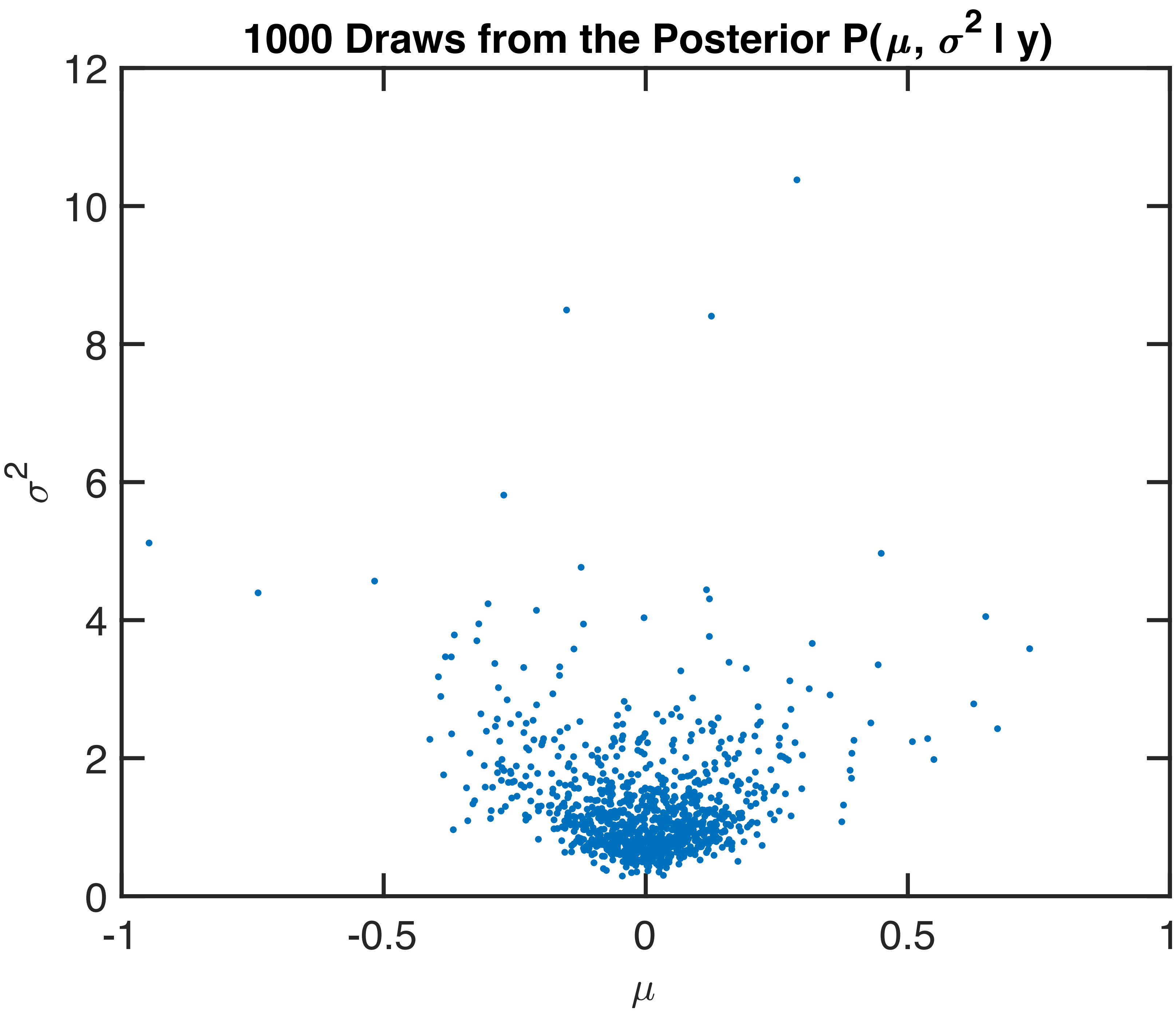
Factorise Posterior: $P(\mu, \sigma | \boldsymbol{y}) = P(\mu | \sigma^2, \boldsymbol{y}) \times P(\sigma^2 | \boldsymbol{y})$

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

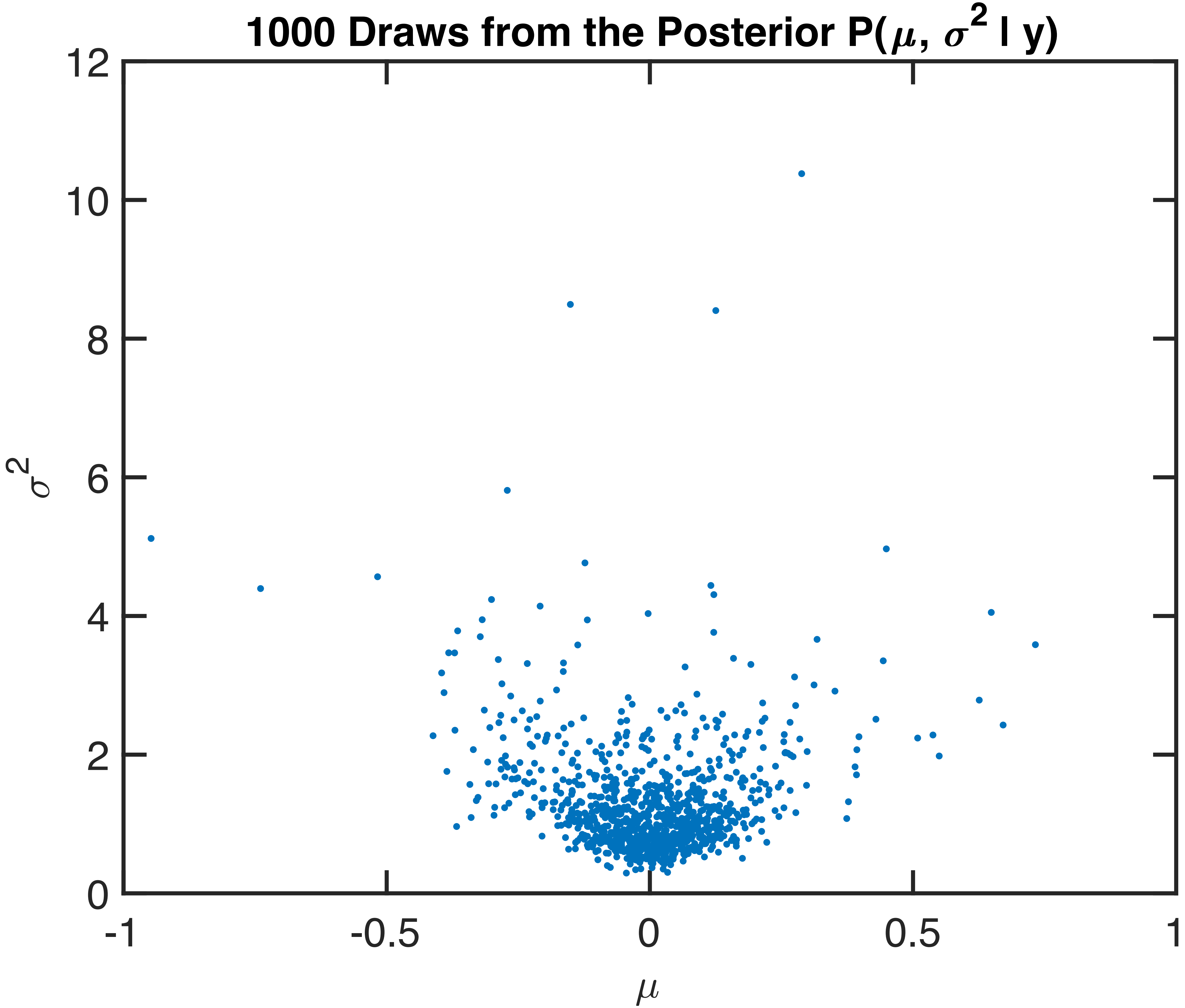
$\bar{y} = 0$

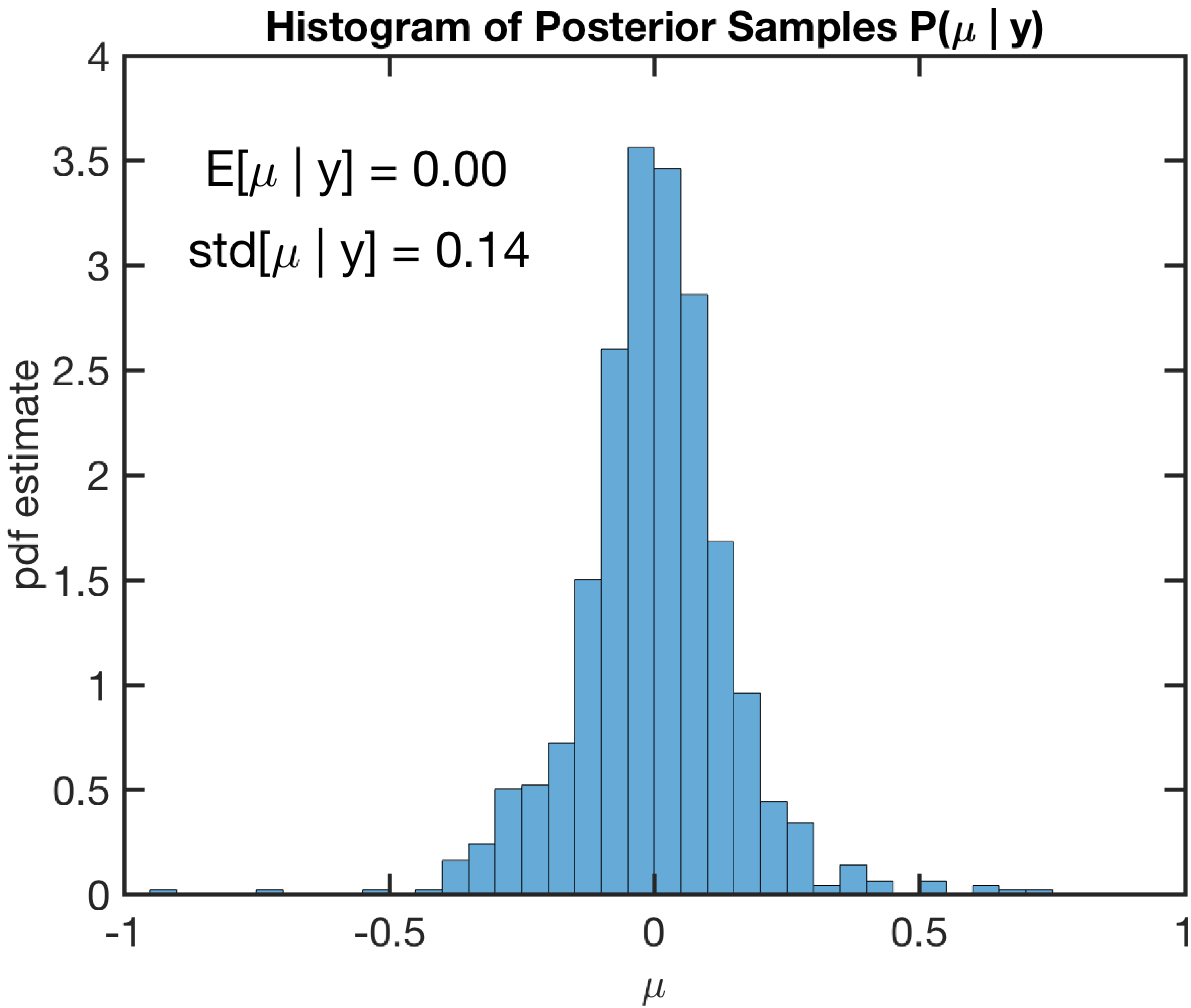
$s^2 = 1$

$n = 10$

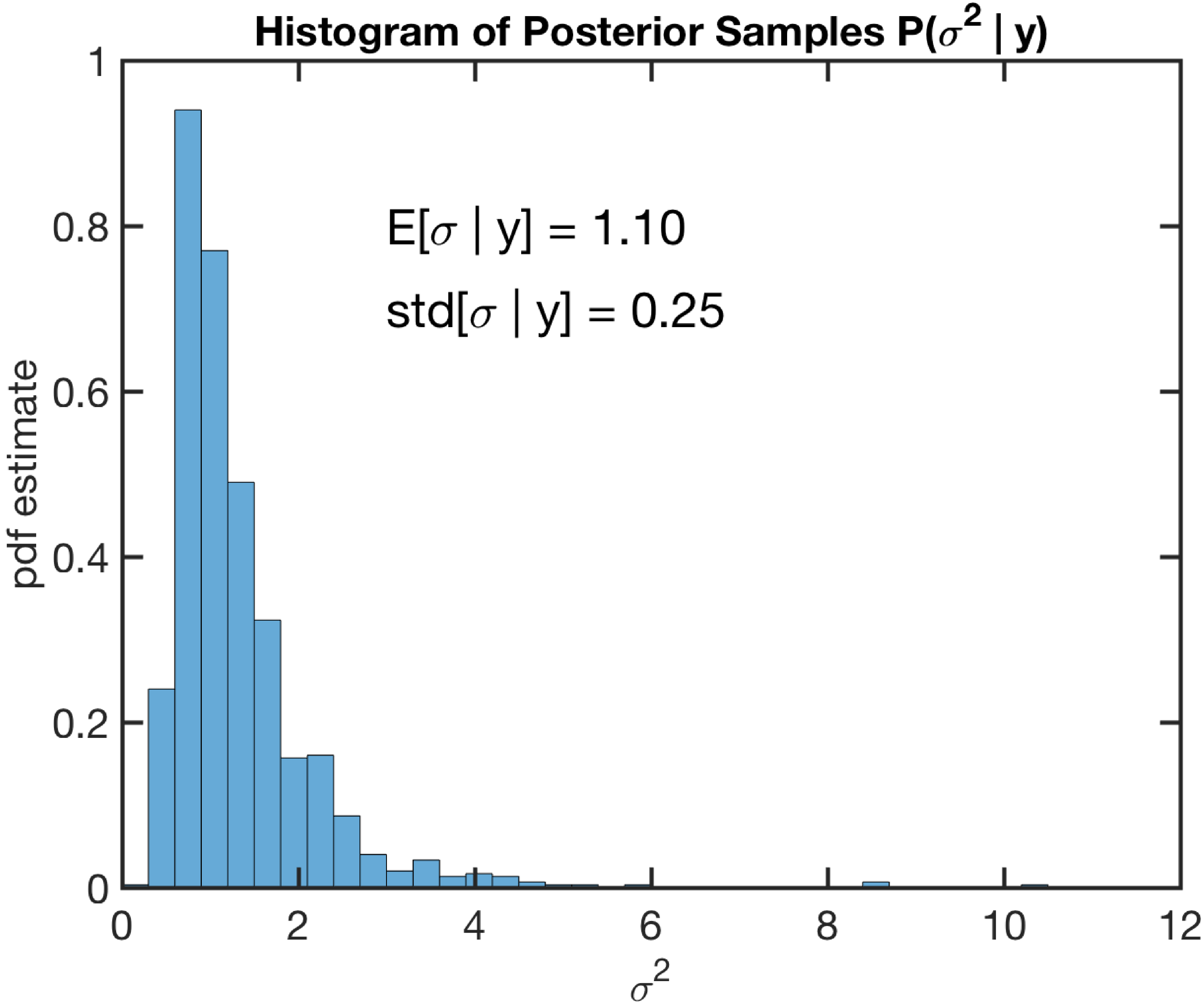


$\bar{y} = 0$ $s^2 = 1$ $n = 10$





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


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

Kernel Density Estimate =
estimate a smooth density from samples