Astrostatistics: Wed 12 Feb 2020

- Bayesian Inference in Astronomy (F&B 3.8, Ivezic 5)
  - C. Bailer-Jones.“Estimating Distances from Parallaxes.” 2015, PASP,127, 994 https://arxiv.org/abs/1507.02105

Simple Gaussian Example
Frequentist Confidence vs. Bayesian credible intervals

Bayesian credible interval

$$Y_1, \ldots, Y_4 \text{ iid} \sim N(\mu, 1)$$

$$\bar{Y} \sim N(\mu, \sigma^2/4)$$

$$y_{\text{obs}} = (-0.64, -0.93, 0.16, -0.88)$$

$$\bar{y} = -0.57, \sigma_{\bar{y}} = 0.5$$

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

$$p(\mu | \boldsymbol{Y} = \boldsymbol{y}_{\text{obs}}) \propto P(\mu) \times \prod_{i=1}^{4} N(y_{\text{obs},i} | \mu, 1^2)$$

(*Derive on board*) $$= N(\mu | \bar{y}, 1^2/4) = N(\mu | -0.57, 0.5^2)$$

This DOES mean that μ is within [-1.07, -0.07] with 68% probability! (degree of belief)

In this simple experiment, confidence and credible intervals are numerically identical, but not always the case

Frequentist vs. Bayes

- Frequentists make statements about the data (or statistics or estimators= functions of the data), conditional on the parameter: $P(D \mid \theta)$ or $P(f(D) \mid \theta)$

- Often goal is to get a "point estimate" or confidence intervals with good properties/coverage under "long-run" repeated experiments in Asymptopia. Arguments are based on datasets that could've happened, but didn't. **Example: Null Hypothesis testing.**

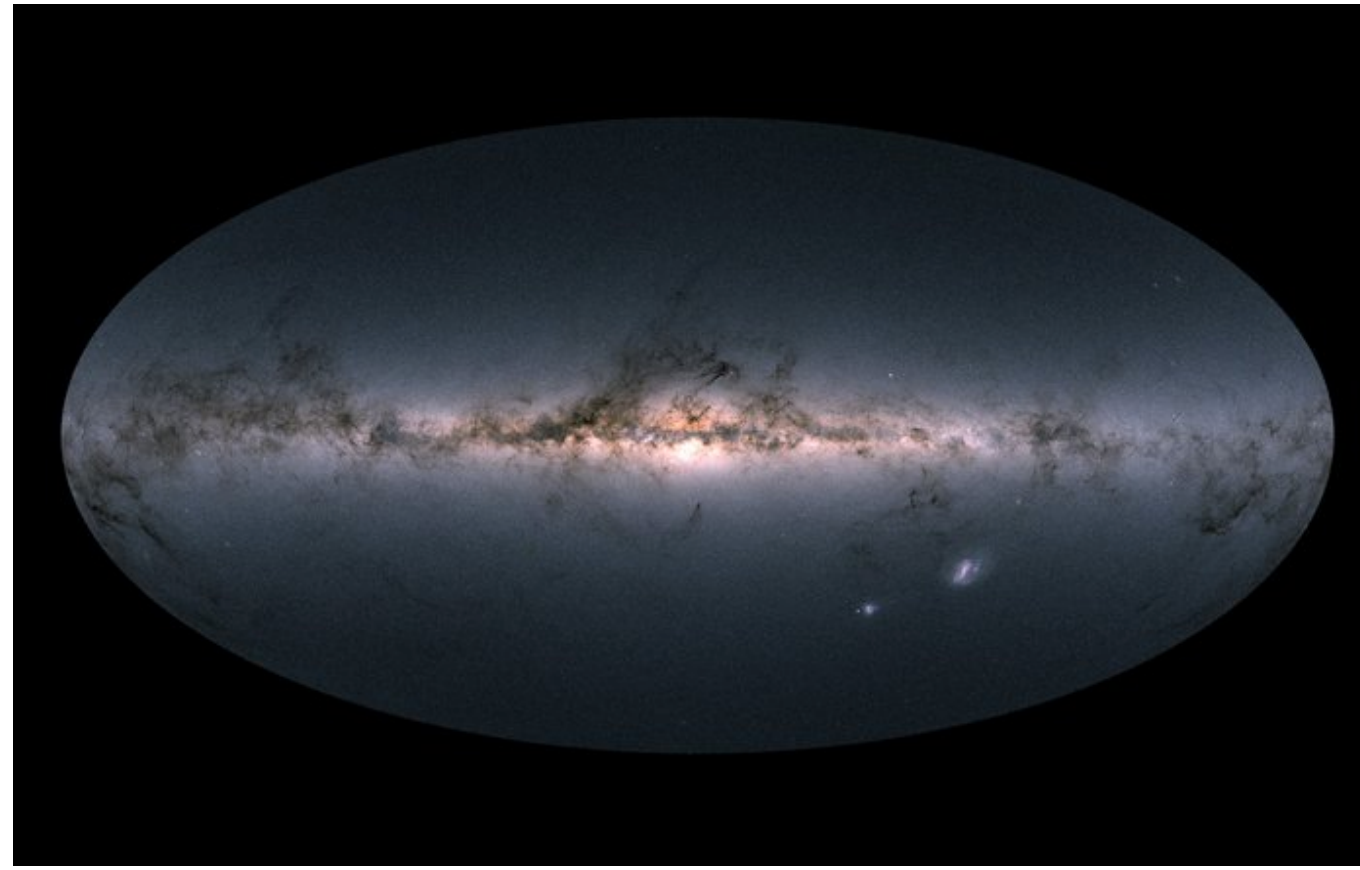
- Bayesians make statements about the probability of parameters, conditional on the dataset D that you actually observed: $P(\theta \mid D)$. This requires an interpretation of probability as a quantifying a "degree of belief" in a hypothesis.

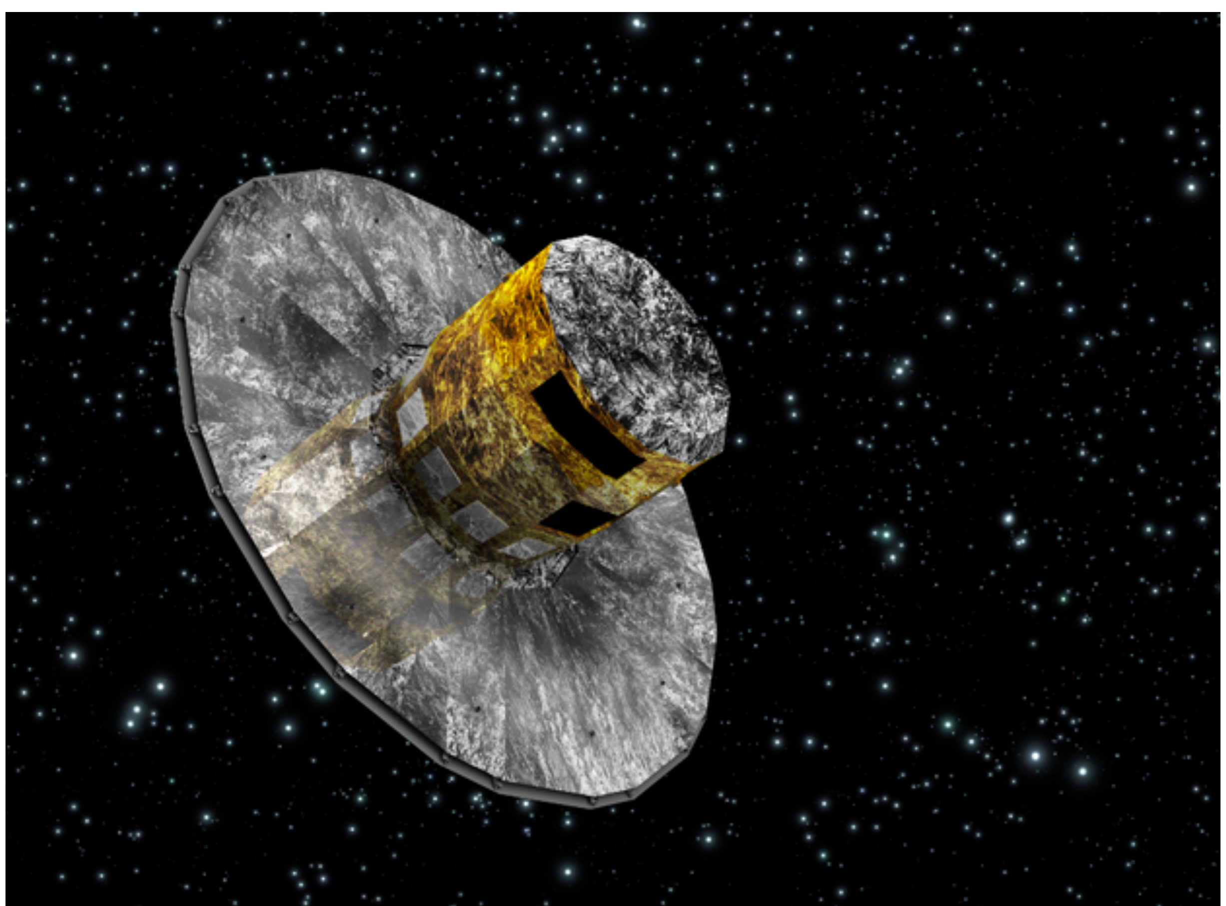
- Bayesian answer is the full posterior density $P(\theta \mid D)$, quantifying the "state of knowledge" after seeing the data. Any numerical estimates are attempts to (imperfectly) summarise the posterior.

Bayes advantages

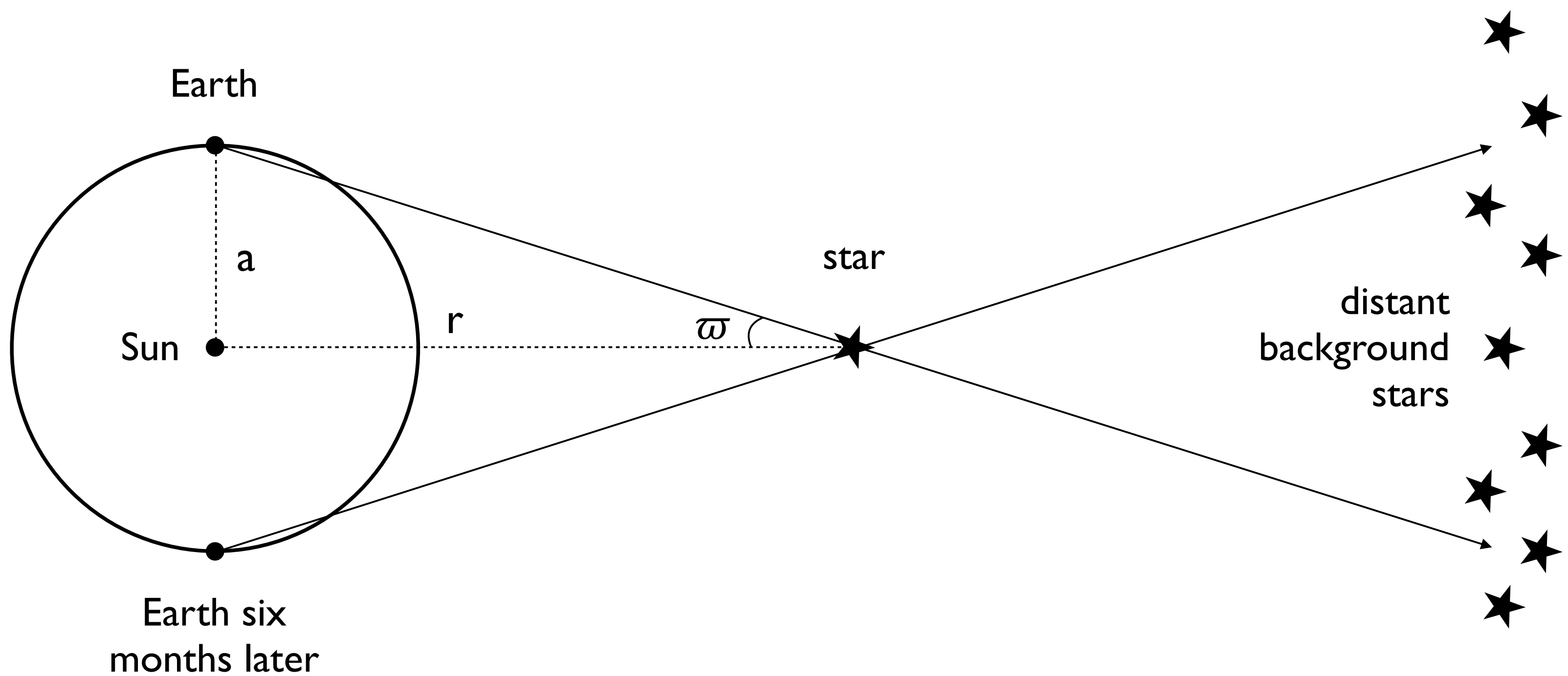
- Ability to include prior information $P(\theta)$
  - External datasets: $P(\theta)$ is really the posterior from some other data $P(\theta \mid D_{ext})$
  - Regularisation: Penalises overfitting data with complex model, e.g. Gaussian process prior
  - "Noninformative" / weakly informative priors / default priors when you don't have / want to use much prior information
- Likelihood is not a probability density in the parameters. But multiply by a prior (even flat), and the posterior is a probability density that obeys clear rules: conditional, marginal probabilities
- Ability to deal with high-dimensional parameter space, e.g. many latent variables or nuisance parameters, and marginalise them "out"
- Estimators derived from Bayesian arguments can still be evaluated in a Frequentist Basis (e.g. James-Stein estimators)

Gaia Parallaxes





Parallax Example



True relation
(no errors)

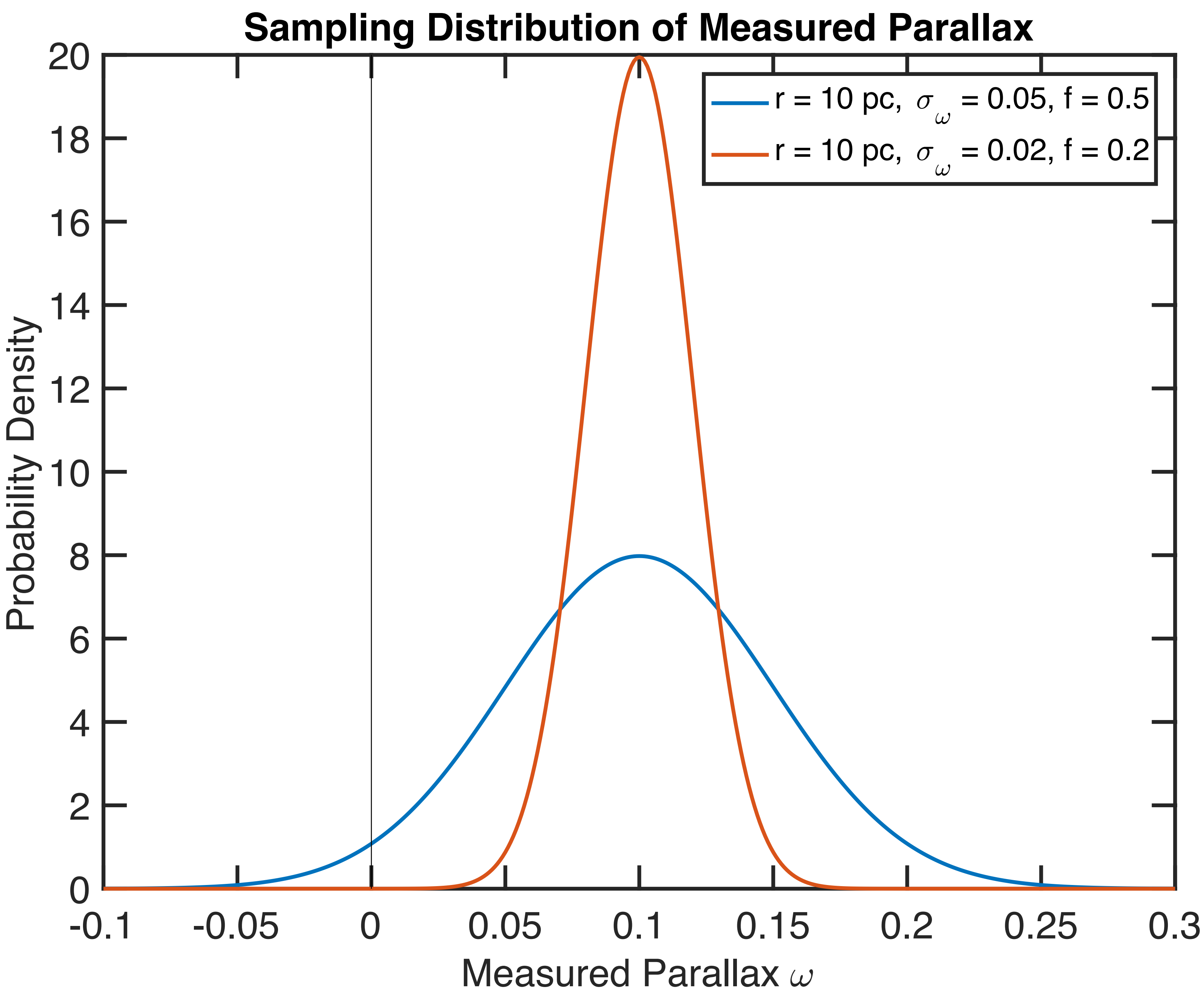
$$\frac{\omega}{\text{arcsec}} = \frac{\text{parsec}}{r}$$

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

Inaccuracy of Gaussian Approx / Propagation of Error

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

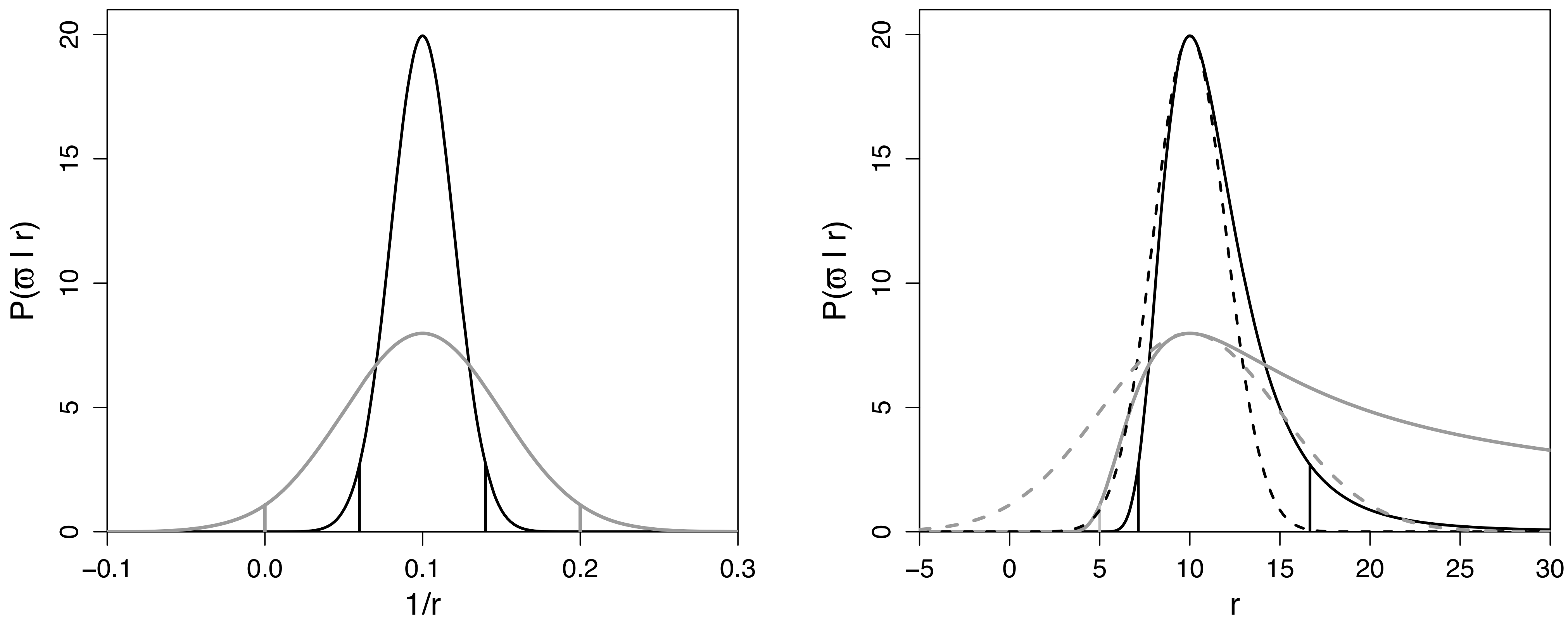
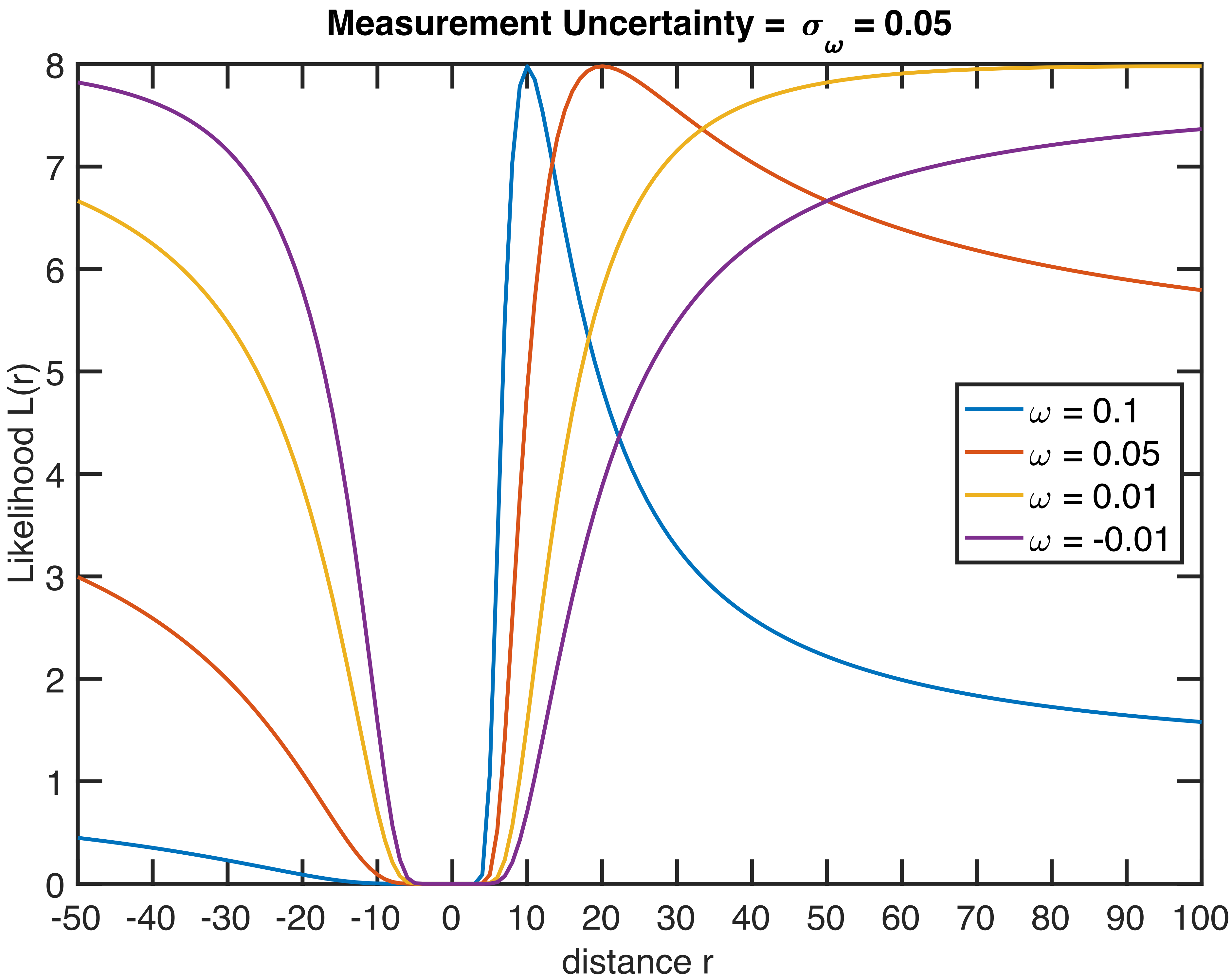


Fig. 3.4 The solid lines show the likelihood (equation 3.20) in the parallax problem for a measured parallax of $\varpi = 0.1$ as a function of $1/r$ in the left panel (a Gaussian) and as a function of r in the right panel. Note that the functions in the right panel are not PDFs over r, so they are not normalized. The black line is for $\sigma_\varpi = 0.02$ (so $f \equiv \sigma_\varpi/\varpi = 1/5$) and the grey line is for $\sigma_\varpi = 0.05$ ($f = 1/2$). The vertical lines denote the upper and lower 2σ limits around $1/\varpi$; the upper limit for the grey curve in the right panel is at $r = \infty$. The dashed lines in the right panel correspond to a Gaussian with mean $1/\varpi$ and standard deviation σ_ϖ/ϖ^2. Each of these Gaussians has been multiplied by the ratio of its standard deviation to that of the likelihood, $(\sigma_\varpi/\varpi^2)/\sigma_\varpi = 100$, in order to put them on the same vertical scale as the likelihood.

The Futility Function



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

An improper distance prior on r

Bayes' Thorem:

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

$$P(r) = \begin{cases} 1 & r > 0 \\ 0 & \text{otherwise} \end{cases}$$

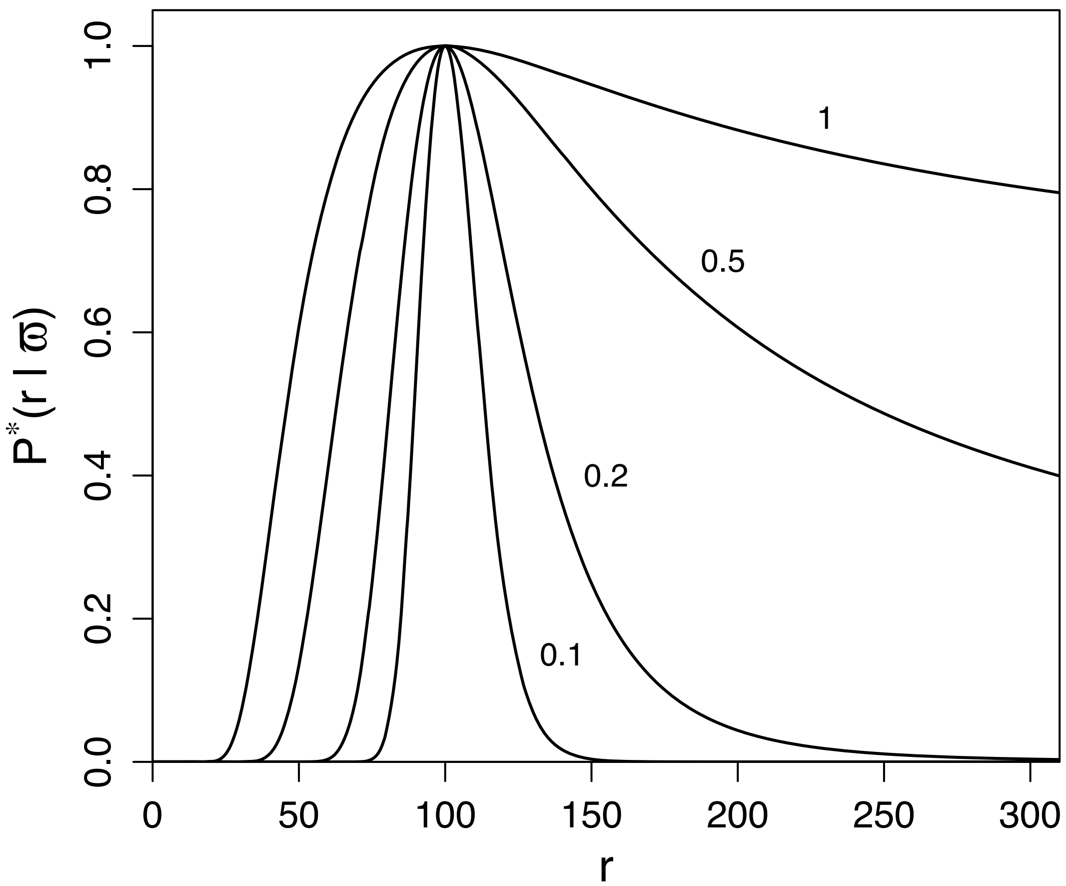
Require Positive Distances r

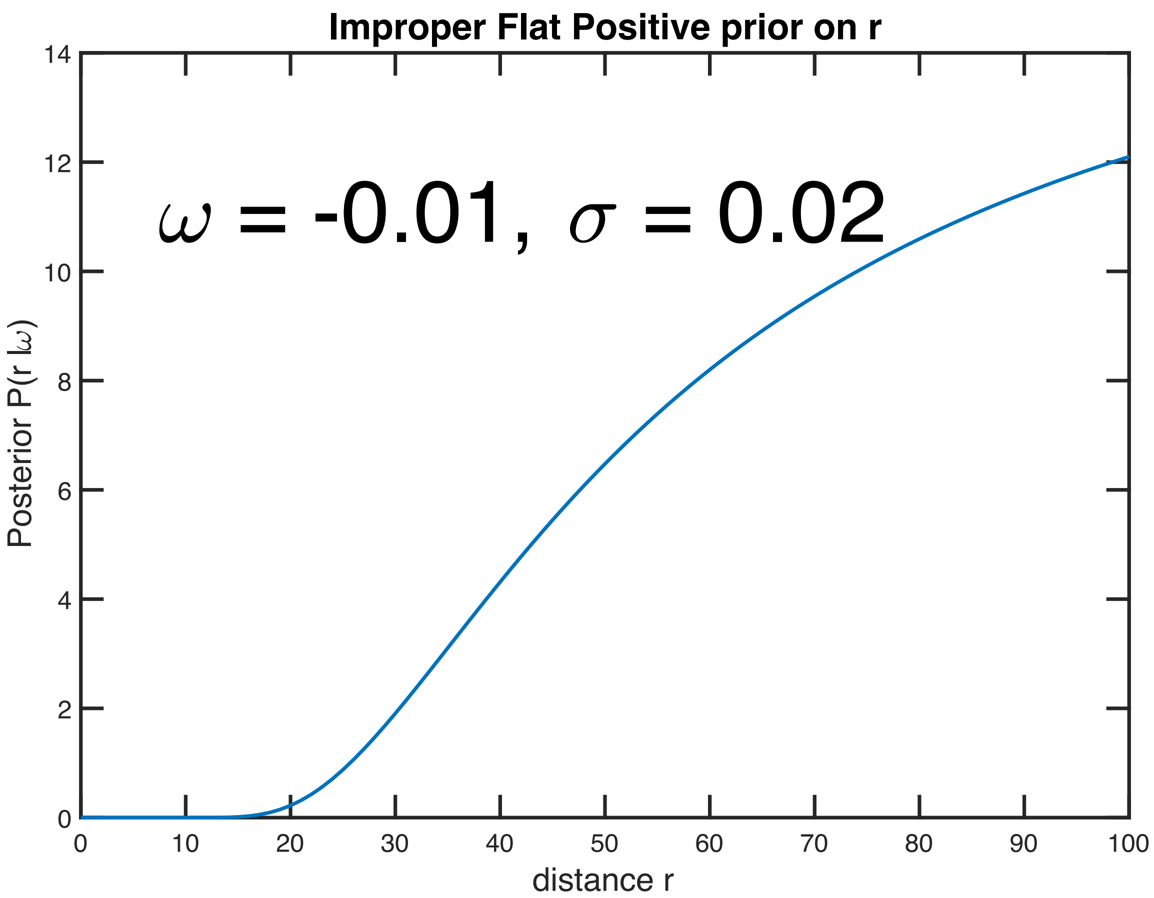
Improper = not normalisable

Posterior of r

$$f = \frac{\sigma_\omega}{\omega} = 0.1,\ 0.2,\ 0.5,\ 1.0$$

$$\omega = 0.01$$





The unnormalized posterior using the improper uniform prior (equation 3.25) for $\varpi = 1/100$ and four values of $f = 0.1, 0.2, 0.5, 1.0$. The unnormalized posteriors have been scaled to all have their mode at $P^*(r\,|\,\varpi) = 1$. Figure reproduced from Bailer-Jones (2015).

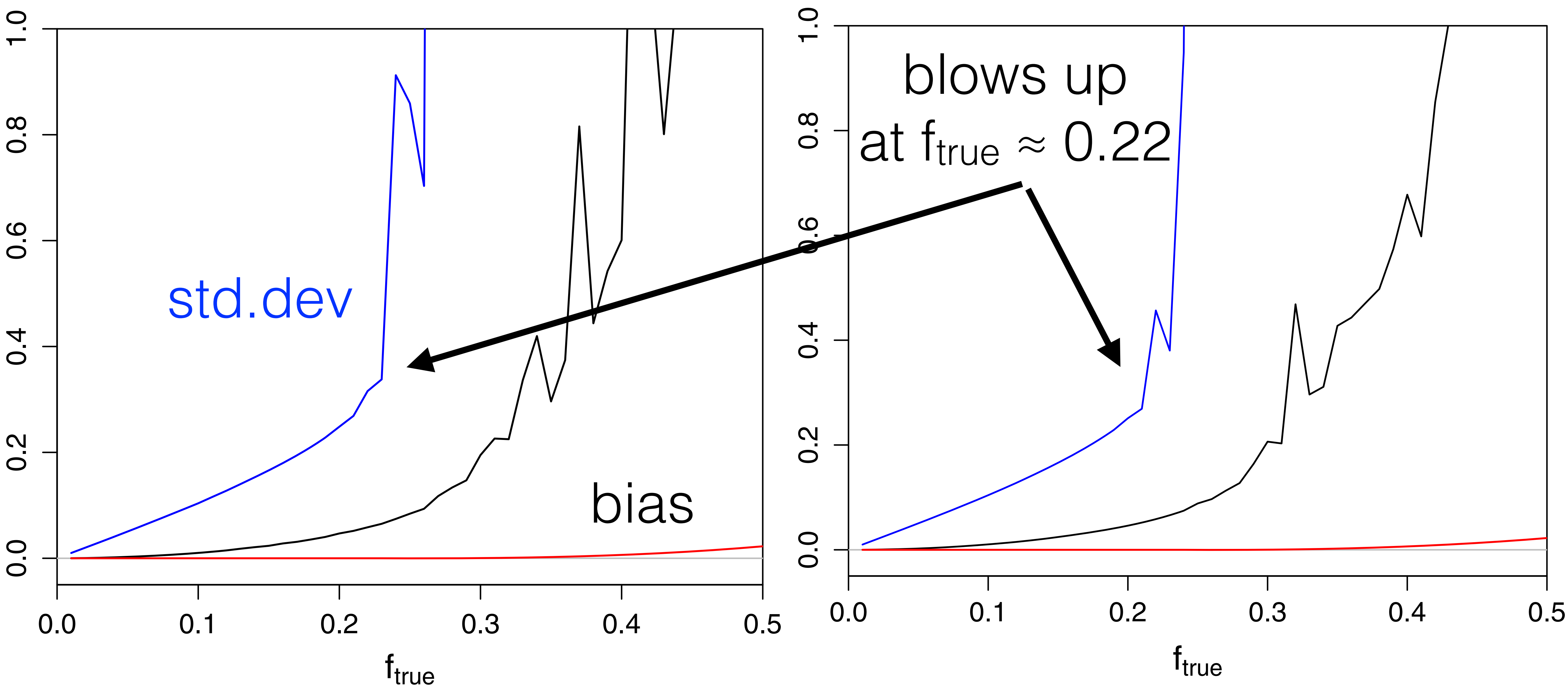
Improper Posterior: not-normalisable, no mean, variance, etc.

Mode (r = 1/$\boldsymbol{\omega}$) exists for positive $\boldsymbol{\omega}$, but undefined for negative $\boldsymbol{\omega}$ (r = ∞)

Evaluating Frequentist Properties with Simulations

1. Assign a fractional measurement error f_{true}
2. Randomly generate a true distance r_{true} (determining $\sigma_{\boldsymbol{\omega}}$)
3. Simulate parallax measurement $\boldsymbol{\omega} \sim P(\boldsymbol{\omega} \mid r = r_{true})$
4. Use $\boldsymbol{\omega}$ in the posterior to get distance estimator r_{est}
5. Calculate scaled residual $x = (r_{est} - r_{true})/r_{true}$
6. Repeat 2-5 to get many sims $\{ x_i \}$ for each f_{true}
7. Compute the sample mean (bias) and std.dev of the $\{ x_i \}$
8. Repeat 1-7 for different values of f_{true}

Frequentist Properties of Posterior Mode



sim r_{true} drawn from

$$P_{r^2}(r) = \begin{cases} \frac{3}{r_{\rm lim}^3} r^2 & \text{if } 0 < r \leq r_{\rm lim} \\ 0 & \text{otherwise} \end{cases} \qquad (9)$$

which corresponds to stars being uniformly distributed in three-dimensional space, $P(V) = \text{const.}$[3] I use $r_{\rm lim} = 10^3$. For each

sim r_{true} drawn from

$$P_u(r) = \begin{cases} \frac{1}{r_{\rm lim}} & \text{if } 0 < r \leq r_{\rm lim} \\ 0 & \text{otherwise} \end{cases}.$$

uniform dist'n in r

Proper Distance prior (truncated)

$$P(r) = \begin{cases} \dfrac{1}{r_{\text{lim}}} & \text{if } 0 < r \le r_{\text{lim}} \\ 0 & \text{otherwise} \end{cases}$$

r_{lim} = maximum possible distance of a star in our survey

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

$$P_u^*(r|\varpi, \sigma_\varpi) = \begin{cases} \frac{1}{r_{\text{lim}}} P(\varpi|r, \sigma_\varpi) & \text{if } 0 < r \le r_{\text{lim}} \\ 0 & \text{otherwise} \end{cases} .$$

Proper Distance prior

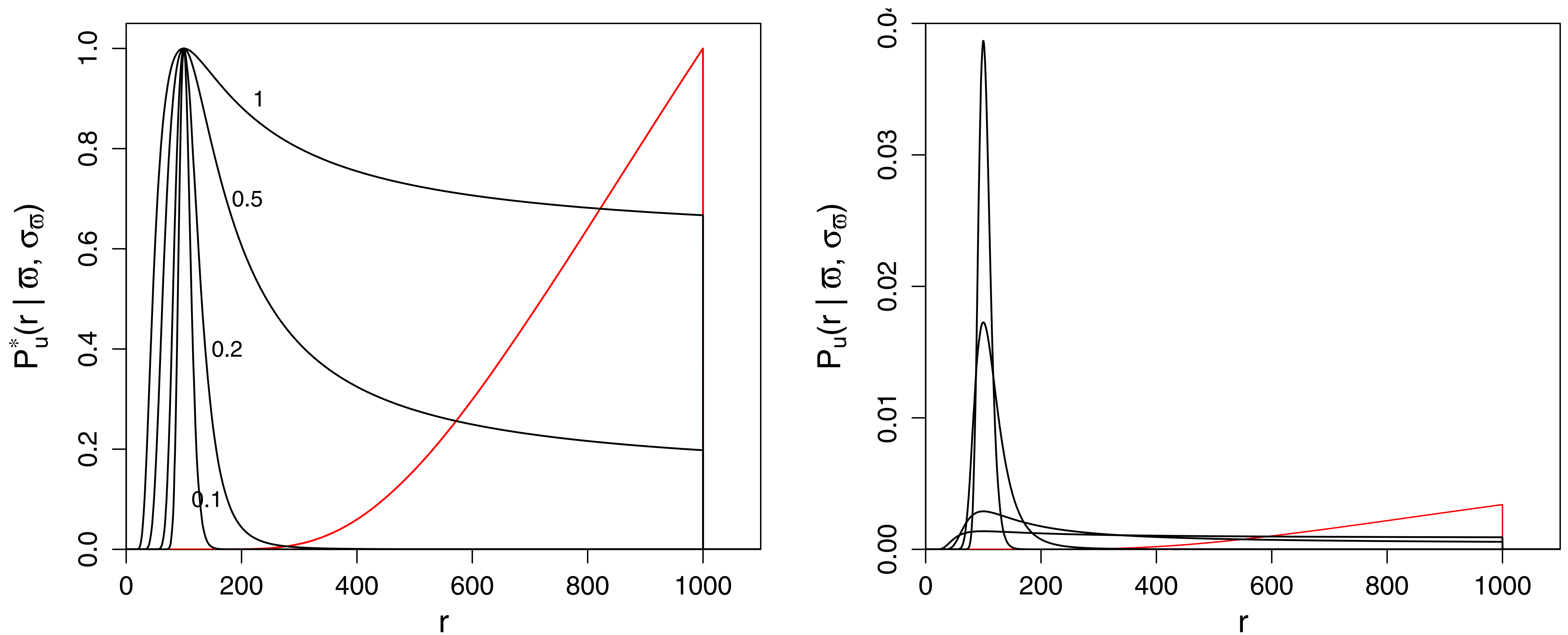


FIG. 4.—Left: the unnormalized posterior $P^*_u(r|\varpi, \sigma_\varpi)$ (truncated uniform prior with $r_{\rm lim} = 10^3$) for $\varpi = 1/100$ and four values of $f = (0.1, 0.2, 0.5, 1.0)$ (*black lines*). The red line shows the posterior for $\varpi = -1/100$ and $|f| = 0.25$. The posteriors have been scaled to all have their mode at $P^*_u(r|\varpi, \sigma_\varpi) = 1$. Right: the same five posterior PFDs but now normalized. See the electronic edition of the *PASP* for a color version of this figure.

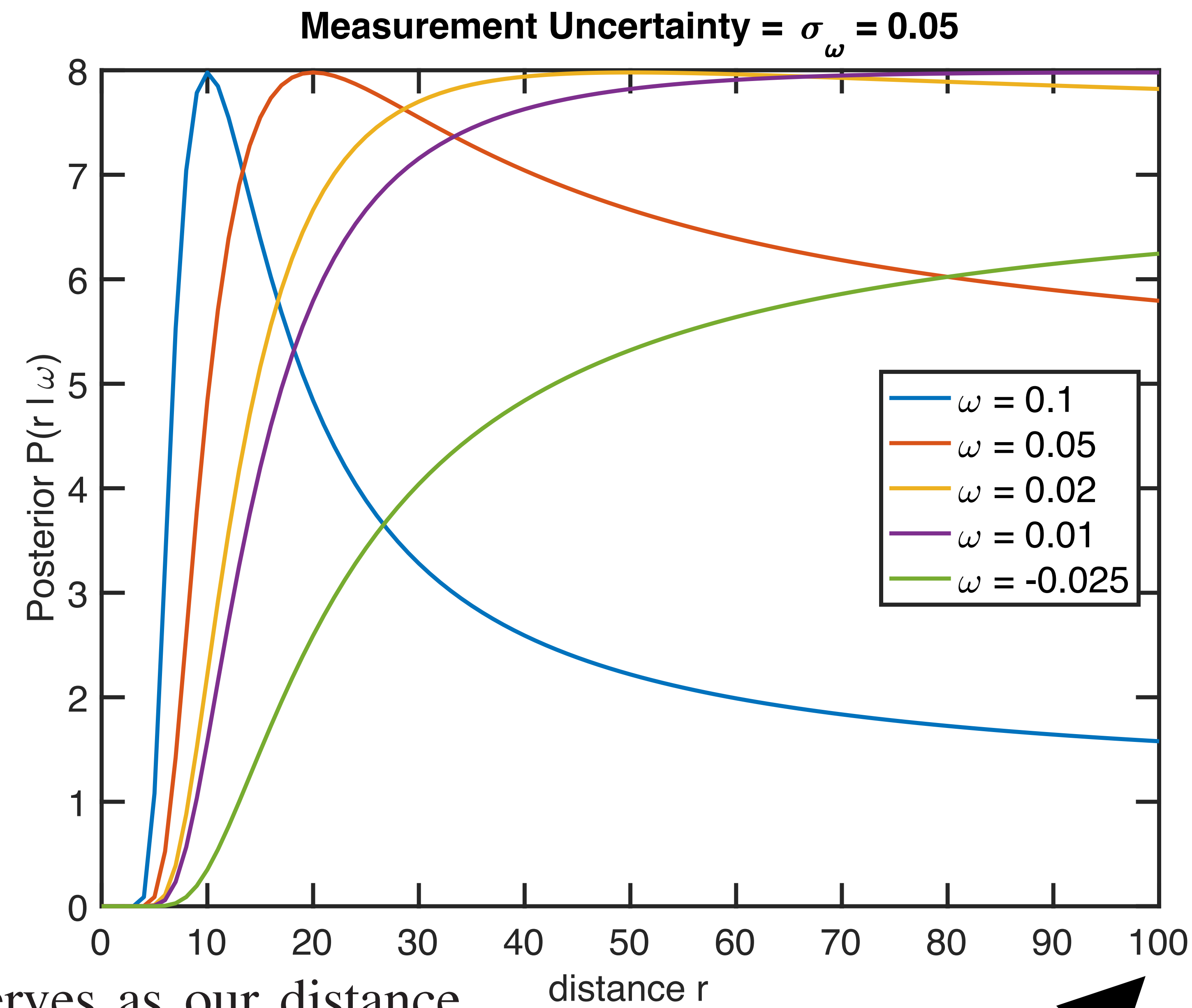
Normalisable posterior works for negative **ω**

"Cutoff" at upper limit for small measured parallax (**ω** < σ)

Proper Uniform Distance prior

$r_{\rm lim} = 100$

(Implies unrealistic Stellar density ~ $1/r^2$)



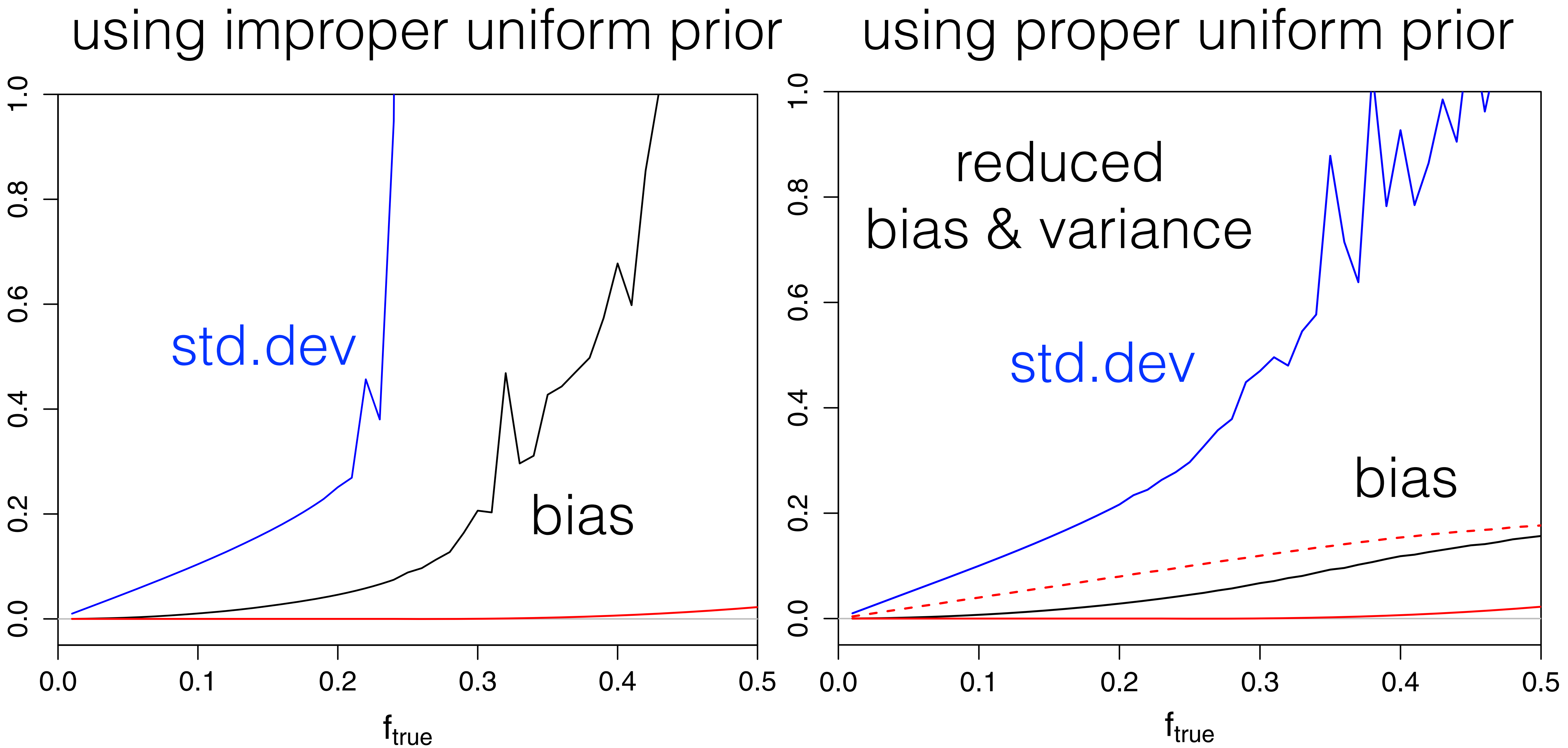
cutoff @ $r_{\rm lim}$

The mode of the resulting posterior serves as our distance estimator, and is

$$r_{\rm est} = \begin{cases} \frac{1}{\varpi} & \text{if } 0 < \frac{1}{\varpi} \le r_{\rm lim} \\ r_{\rm lim} & \text{if } \frac{1}{\varpi} > r_{\rm lim} \\ r_{\rm lim} & \text{if } \varpi \le 0 \end{cases} \quad \text{(extreme mode)} . \qquad (12)$$

Cutoff of modes at edge $r_{\rm lim}$ for small measured parallax ($\boldsymbol{\omega} < \sigma$) [but better than going to infinity!]

Frequentist Properties of Posterior Mode



sim r_{true} drawn from uniform dist'n in r

$$P_u(r) = \begin{cases} \frac{1}{r_{\text{lim}}} & \text{if } 0 < r \leq r_{\text{lim}} \\ 0 & \text{otherwise} \end{cases}.$$

Introducing physical constraints into the prior

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

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

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

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

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

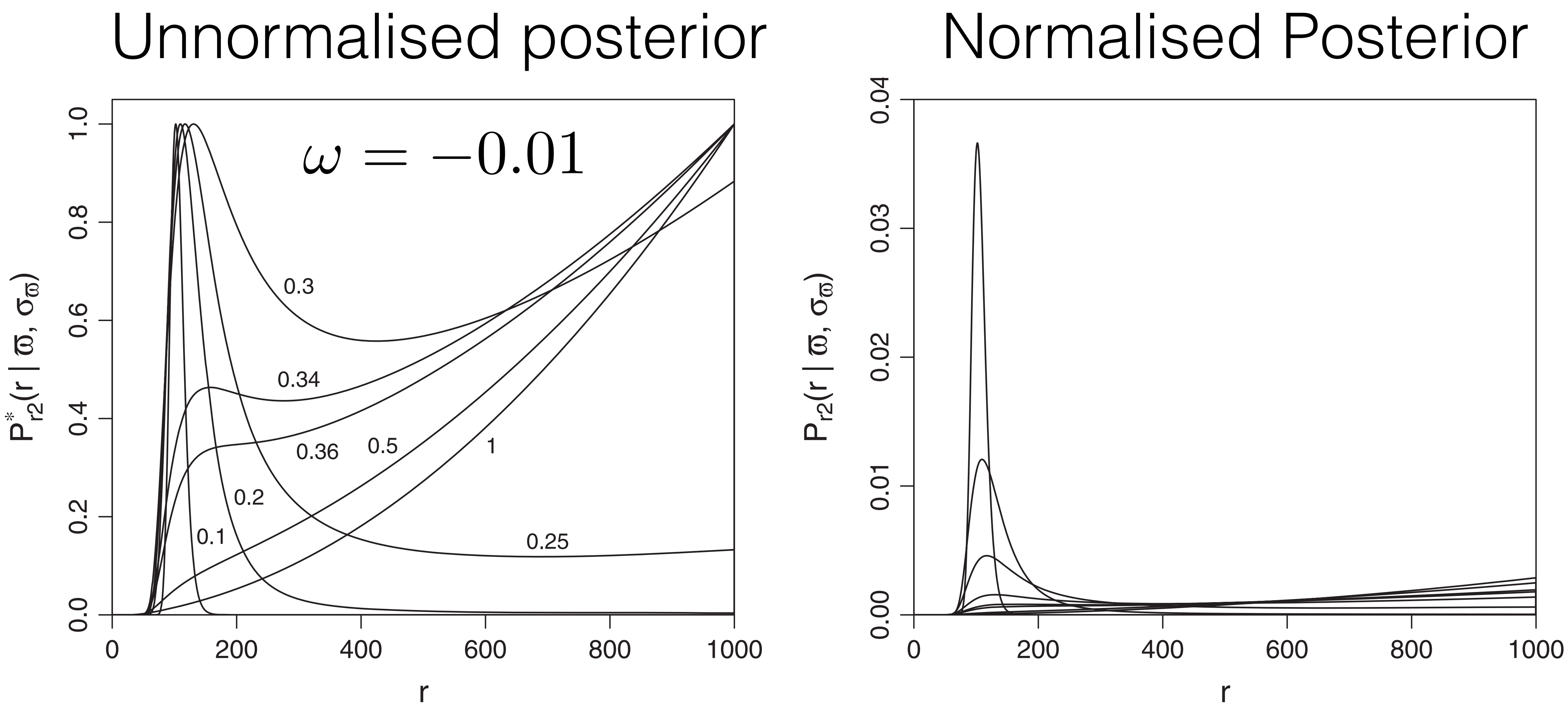
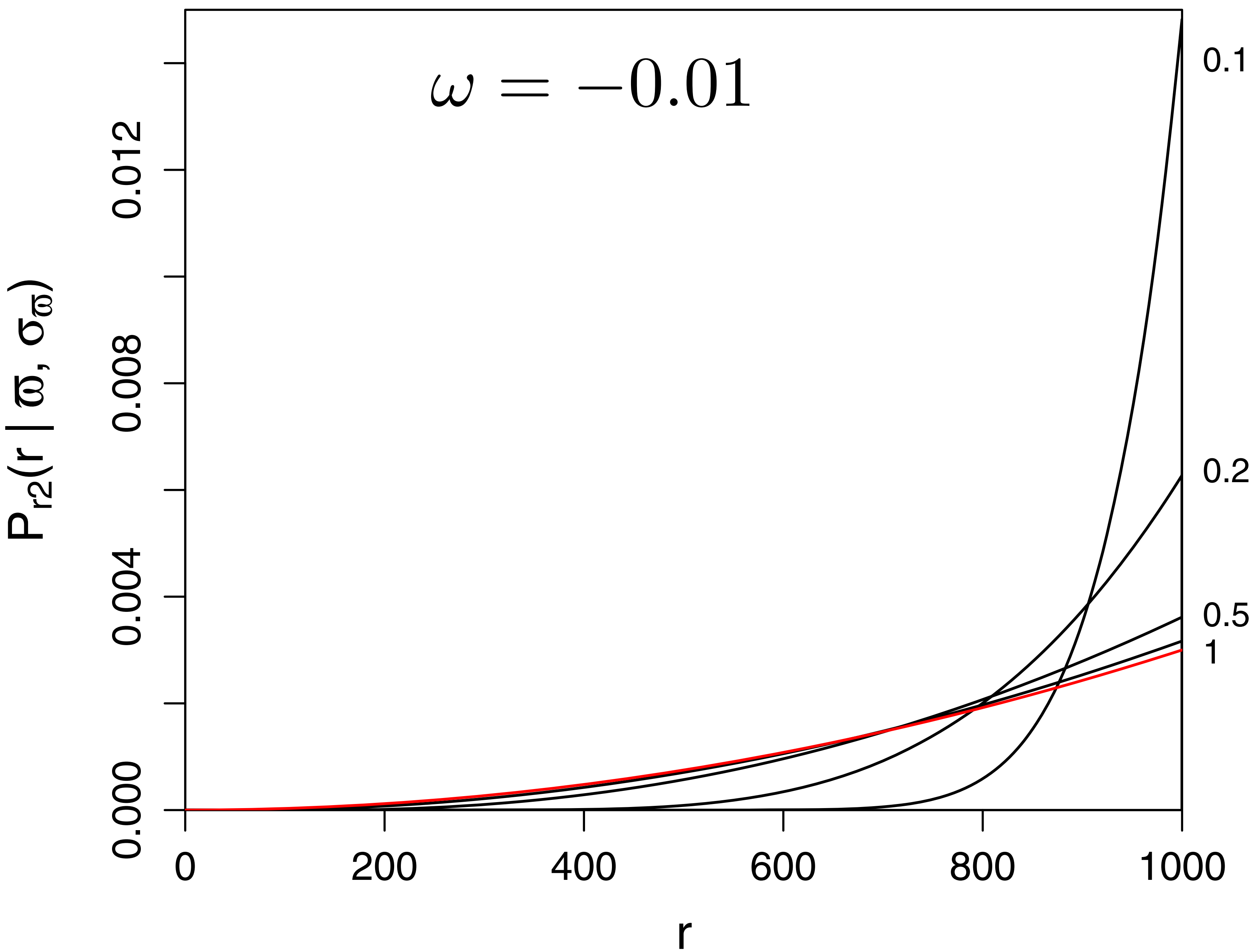


FIG. 7.—Left: the unnormalized posterior $P^*_{r^2}(r|\varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\mathrm{lim}} = 10^3$) for $\varpi = 1/100$ and eight values of $f = (0.1, 0.2, 0.25, 0.3, 0.34, 0.36, 0.5, 1.0)$. The posteriors have been scaled to all have their mode at $P^*_{r^2}(r|\varpi, \sigma_\varpi) = 1$. 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

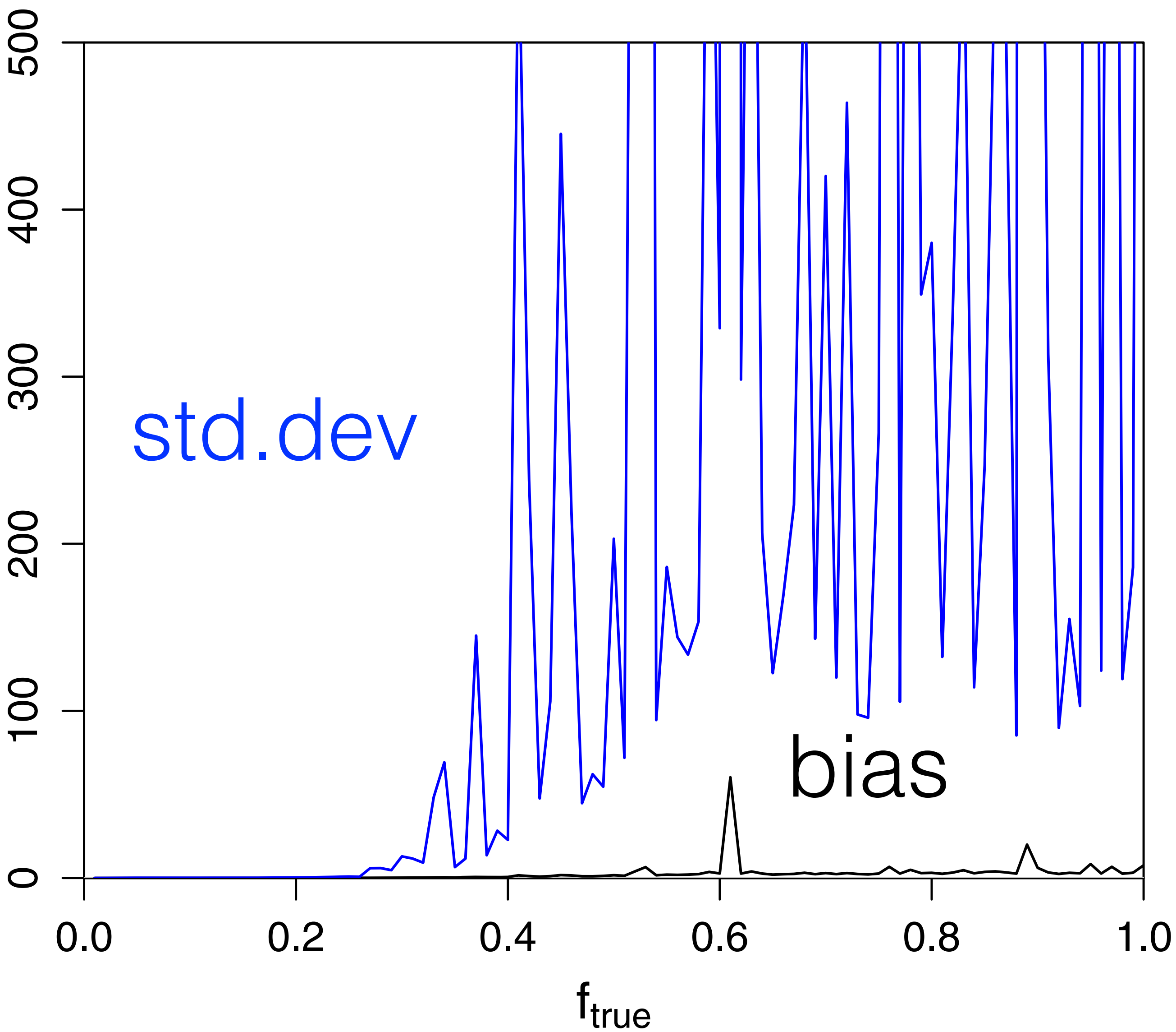


Works for negative and zero parallaxes **ω** but again cutoff at hard limit

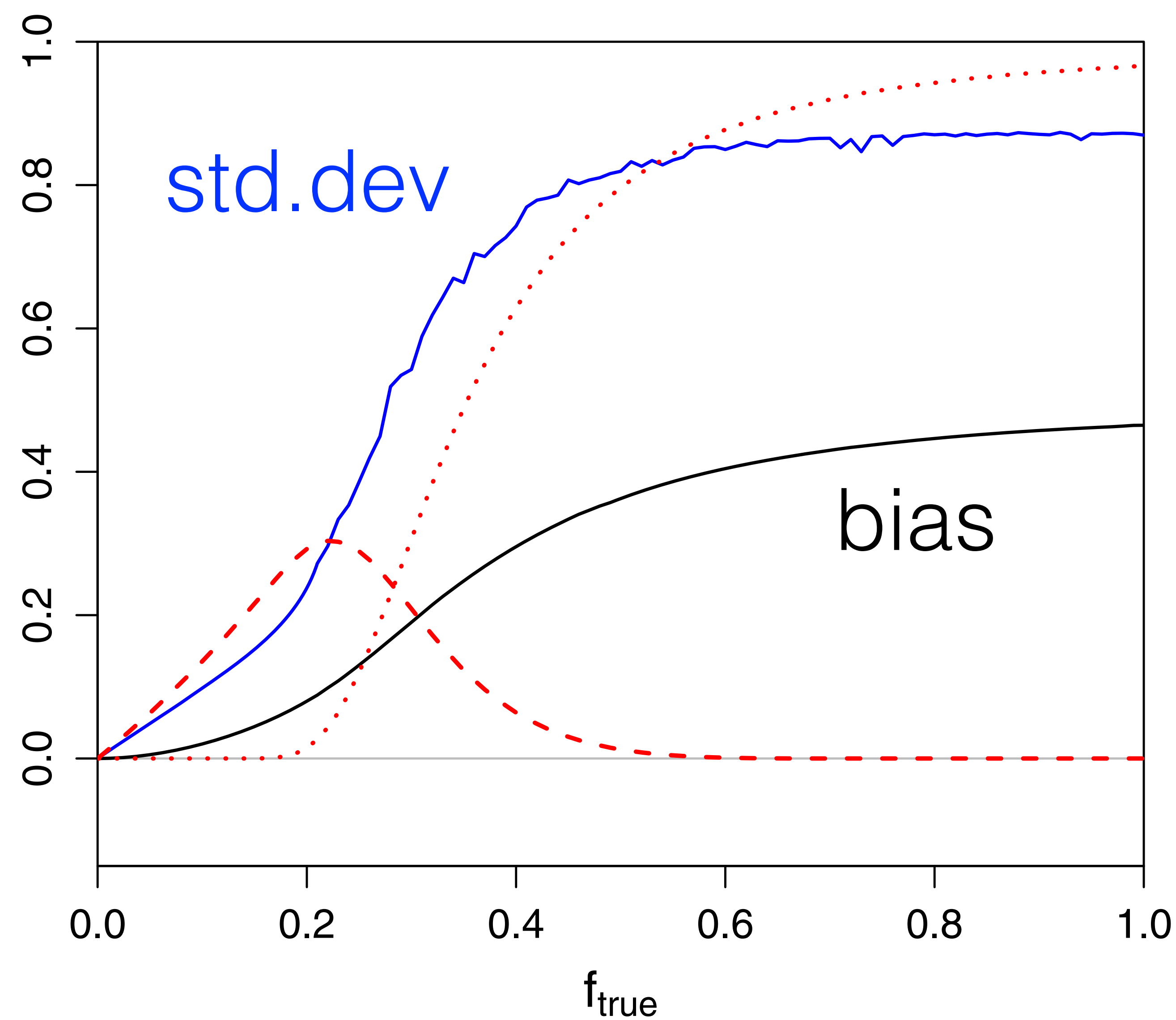
FIG. 9.—The normalized posterior $P^*_{r^2}(r|\varpi, \sigma_\varpi)$ (truncated constant volume density prior with $r_{\rm lim} = 10^3$) for $\varpi = -1/100$ and four values of $|f| = (0.1, 0.2, 0.5, 1.0)$ (*black lines*). The red line shows the posterior for $\varpi = 0$ for $\sigma_\varpi \gg 1/r_{\rm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

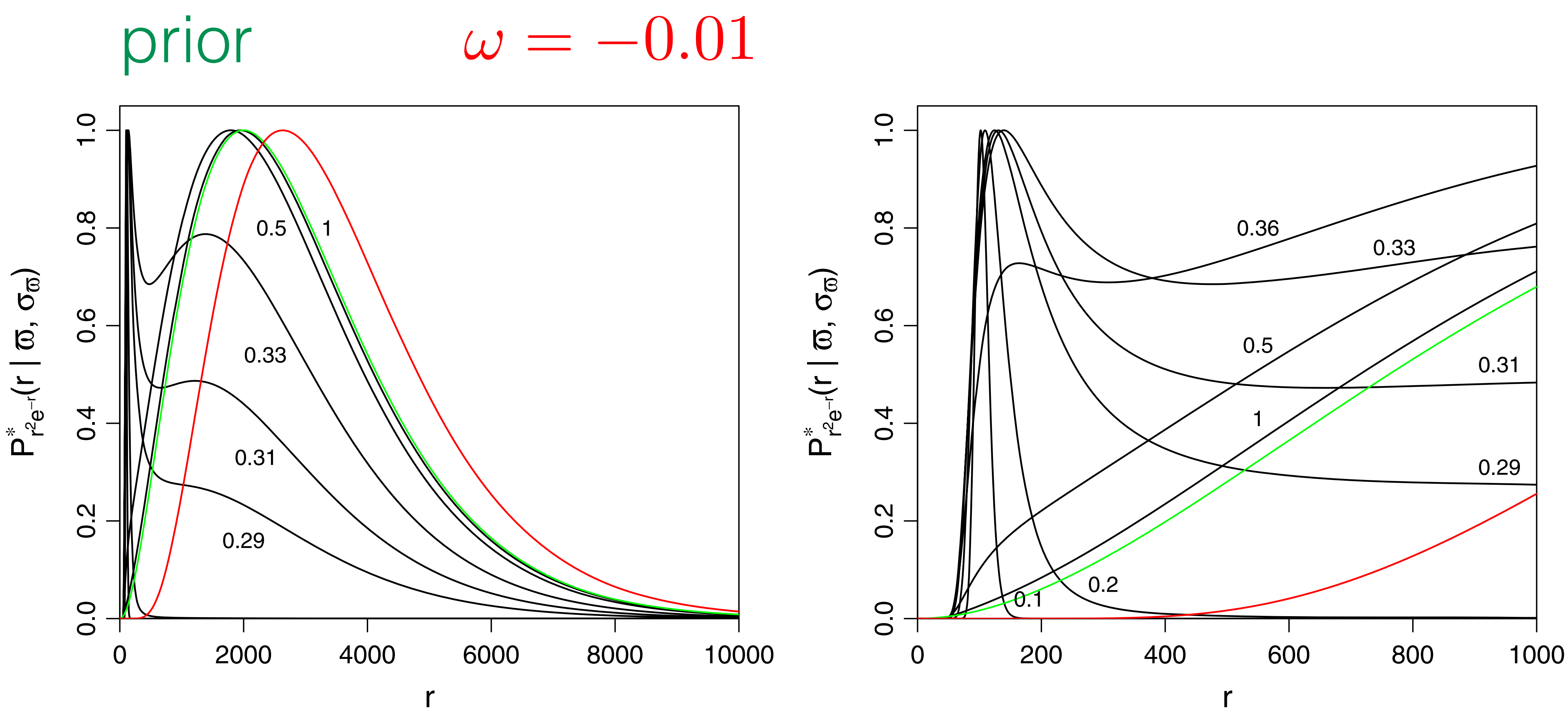


FIG. 12.—The *black lines* in the left panel show the unnormalized posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) for $L = 10^3$, $\varpi = 1/100$ and seven values of $f = (0.1, 0.2, 0.29, 0.31, 0.33, 0.5, 1.0)$. The *red line* is the posterior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

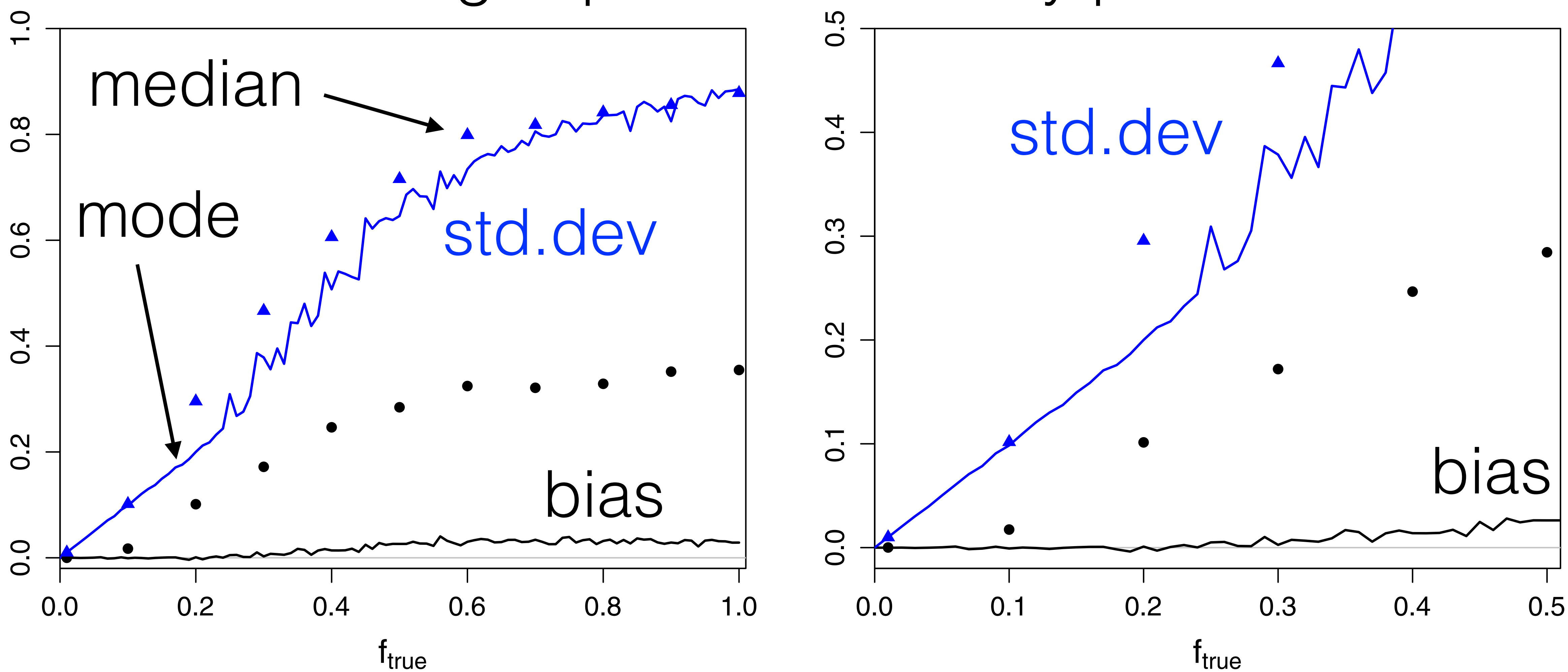
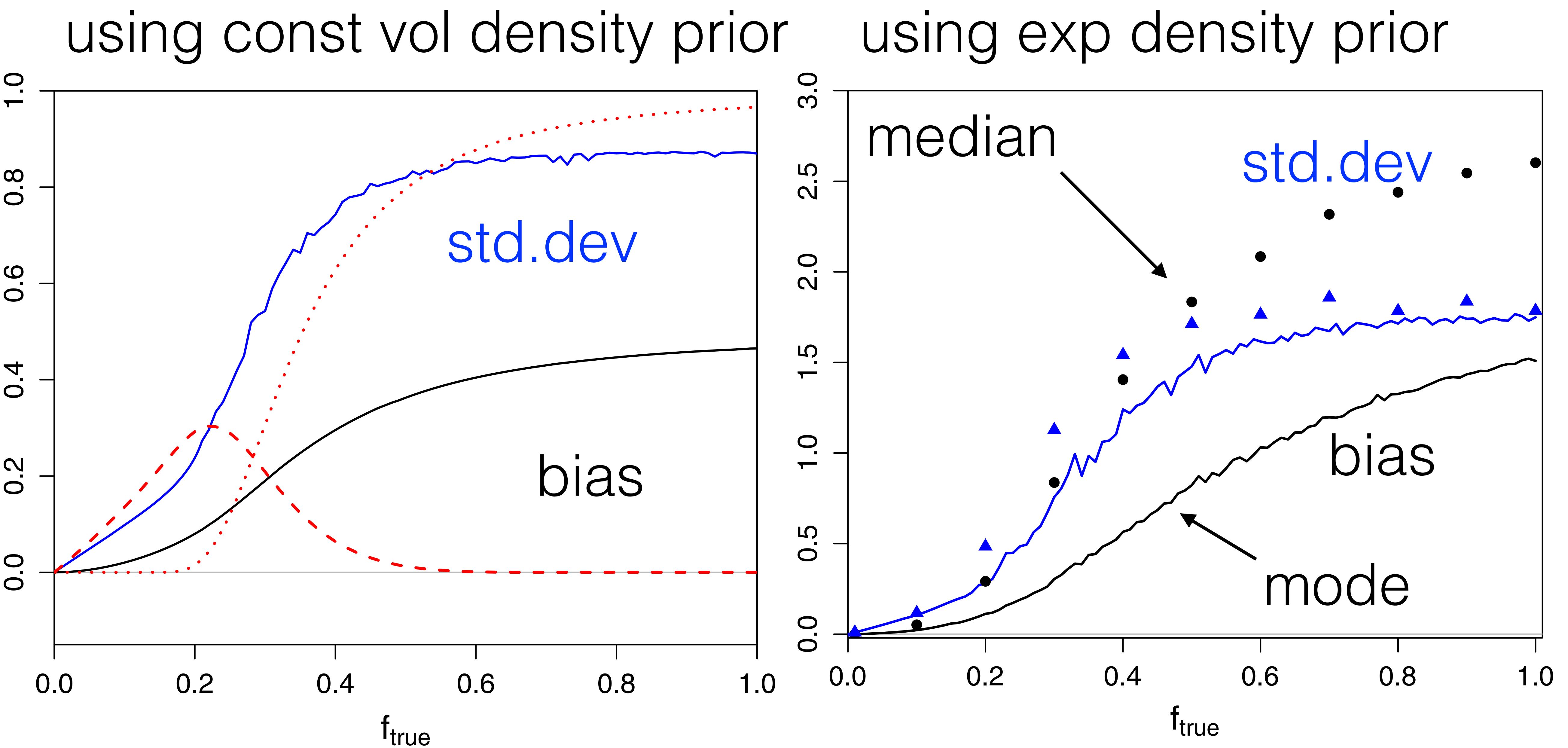


FIG. 14.—The bias (*black line*) and standard deviation (*blue line*) as a function of $f_{\rm true}$ for the mode distance estimator of the posterior $P^*_{r^2e^{-r}}(r|\varpi, \sigma_\varpi)$ (exponentially decreasing volume density prior; eq. [18]) with $L = 10^3$ for data drawn from the same prior. 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_{true} drawn from

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

Frequentist Properties of Posterior Mode



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

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

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)