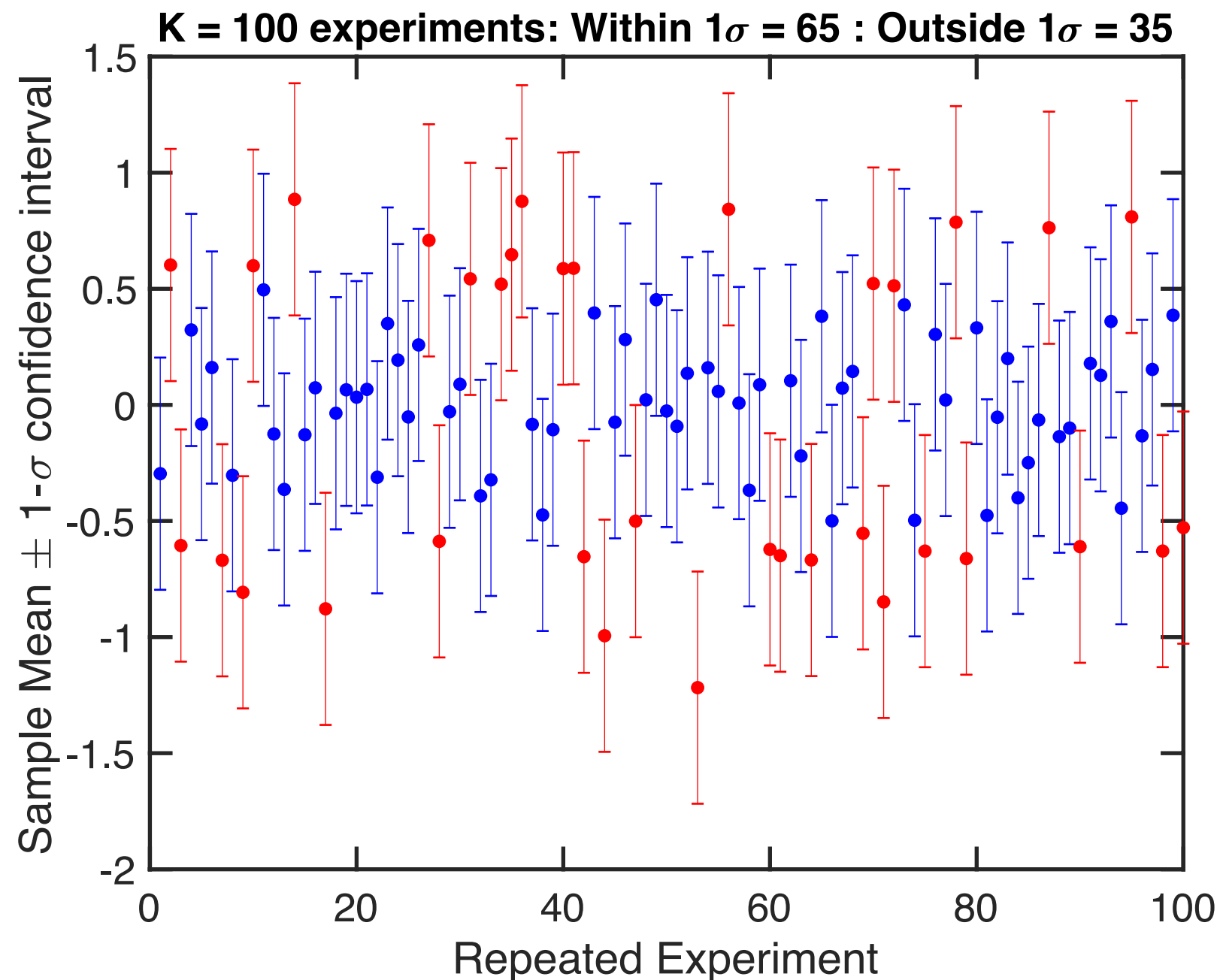


Astrostatistics: Wed 06 Feb 2019

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

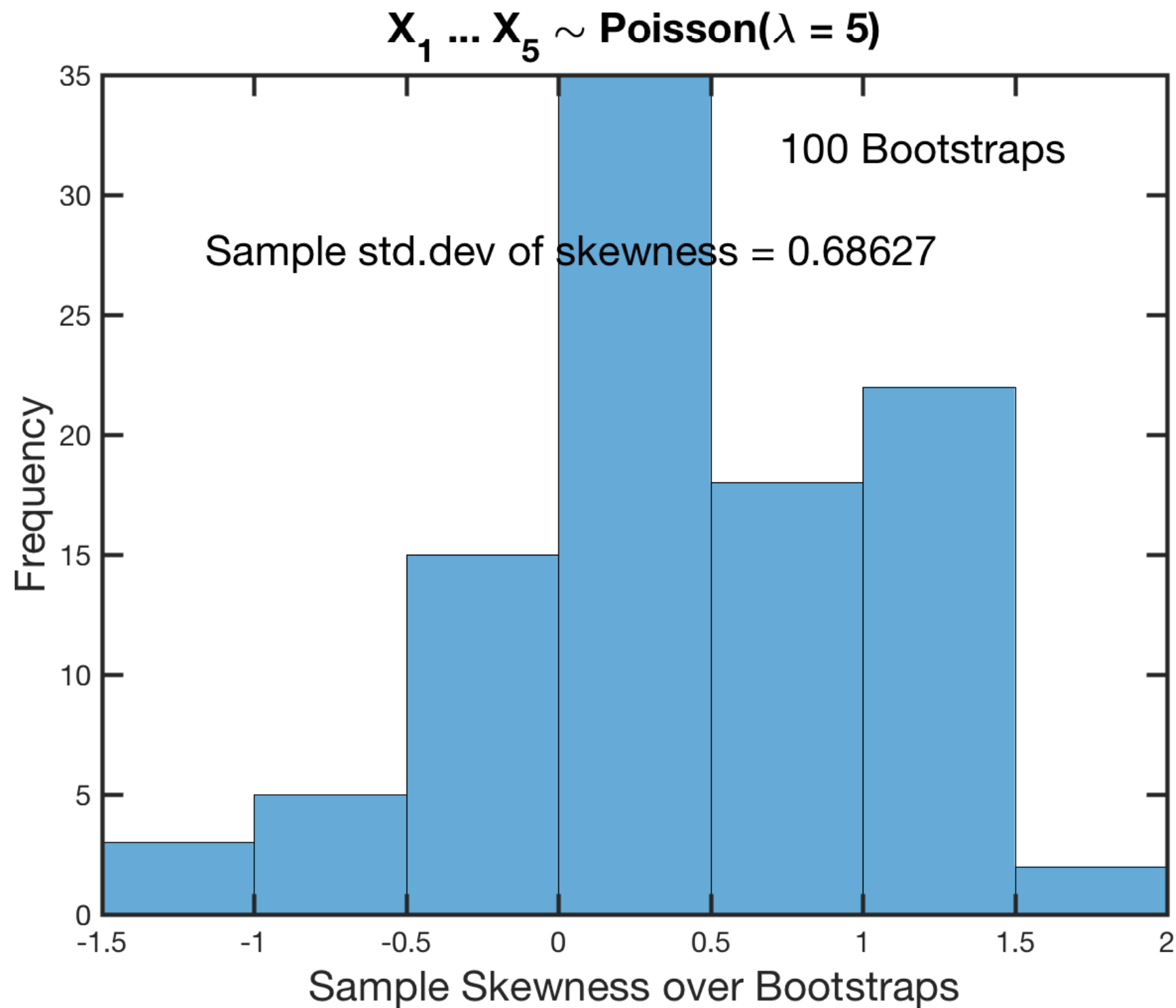
- Quantifying Uncertainty: Bootstrap
- Fitting Statistical Models to Astronomical Data
 - Linear Regression Approaches (F&B Ch 7, Ivezić, Ch 8)
 - Generative / Forward Modeling with Latent Variables
 - Linear Regression with intrinsic dispersion and heteroskedastic (x,y) measurement error
 - Kelly et al. “Some Aspects of Measurement Error in Linear Regression of Astronomical Data.” 2017, The Astrophysical Journal, 665, 1489 **(500 citations!)**
- Preview: Bayesian Inference in Astronomy
 - C. Bailer-Jones. “Estimating Distances from Parallaxes.” 2015, PASP, 127, 994
<https://arxiv.org/abs/1507.02105>
 - Patel, Besla & Mandel. “Bayesian estimates of the Milky Way and Andromeda masses using high-precision astrometry and cosmological simulations.” MNRAS, 468, 3428. <https://arxiv.org/abs/1803.01878>
 - Patel, Besla, Mandel & Sohn. “Estimating the Mass of the Milky Way Using the Ensemble of Classical Satellite Galaxies.” The Astrophysical Journal, 857, 78. <https://arxiv.org/abs/1703.05767>

Quantifying Uncertainty: Frequentist Meaning of Confidence Intervals

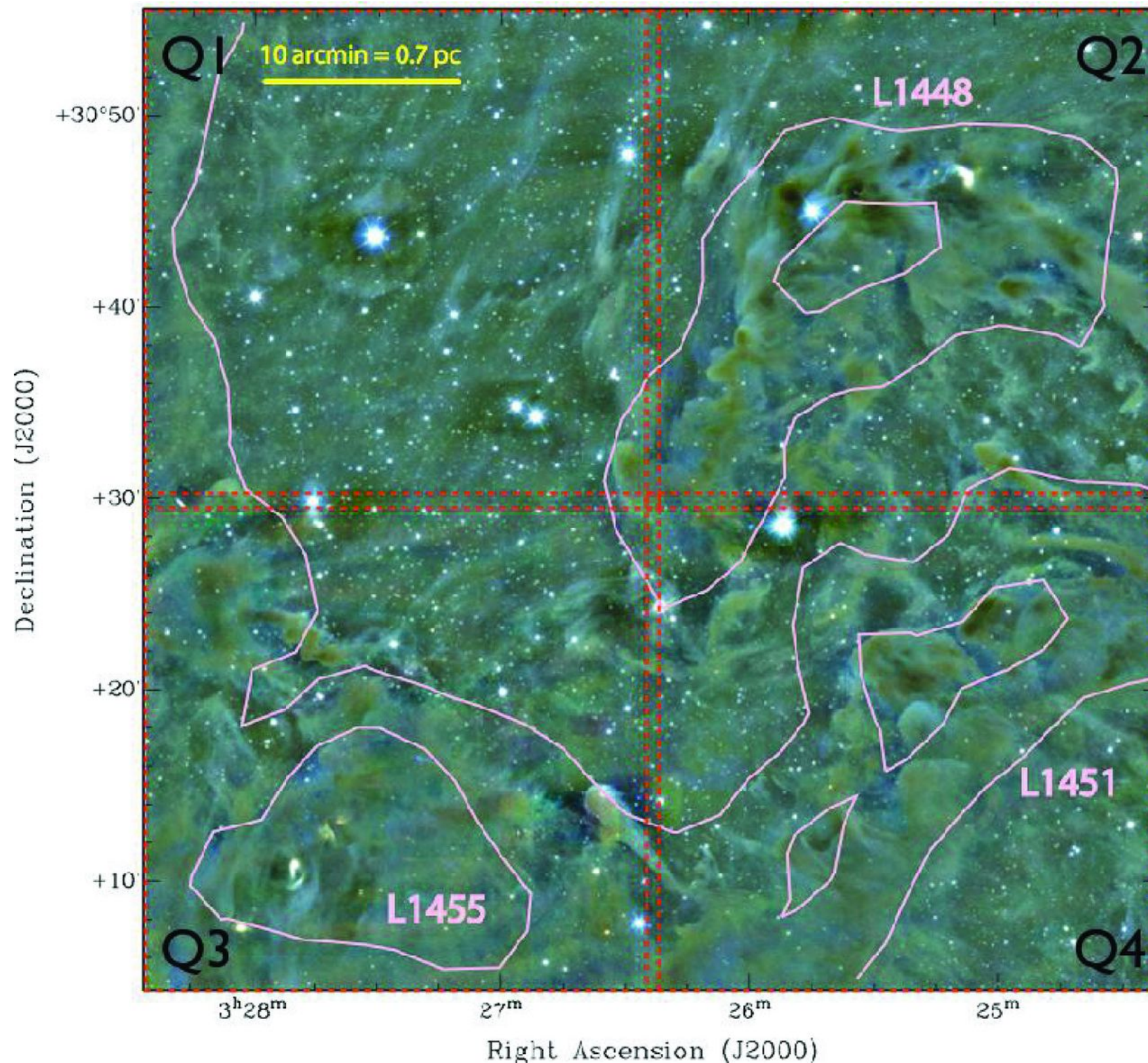


>68% of confidence intervals for repeated experiments
will contain true mean (= 0 here)

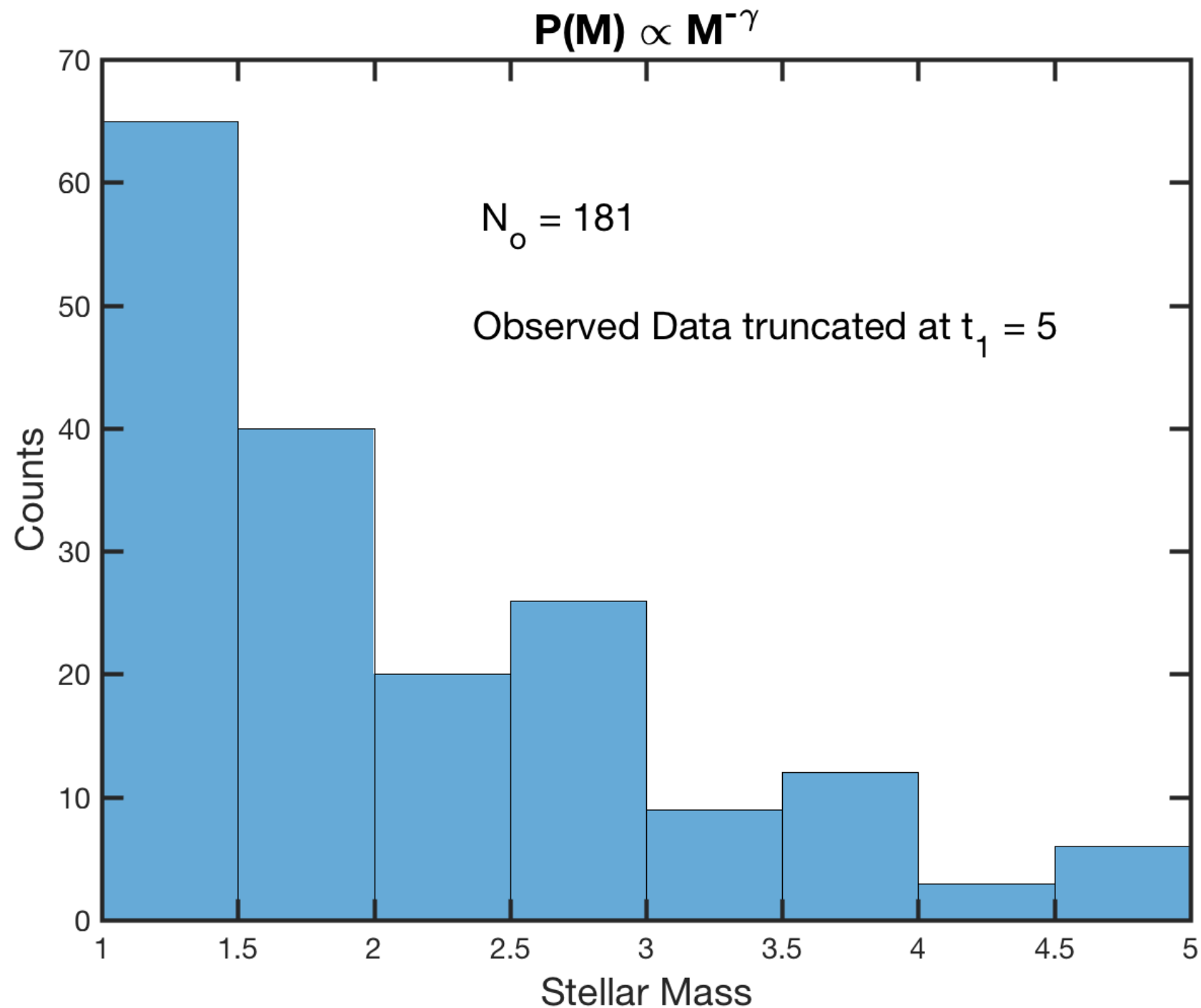
Bootstrap Standard Error



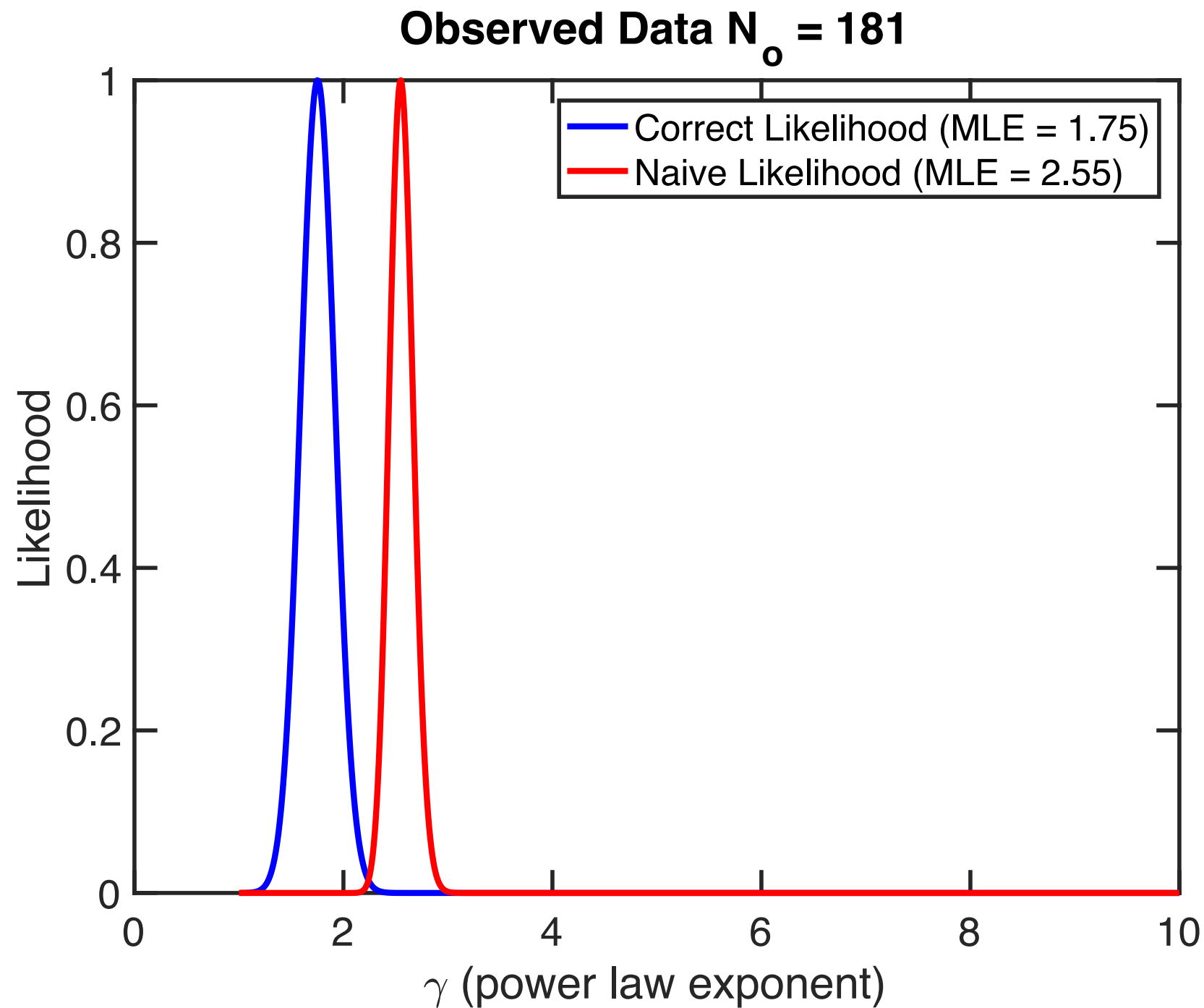
Star Formation in Perseus



Observed Stellar Mass Distribution (with truncation selection effect)

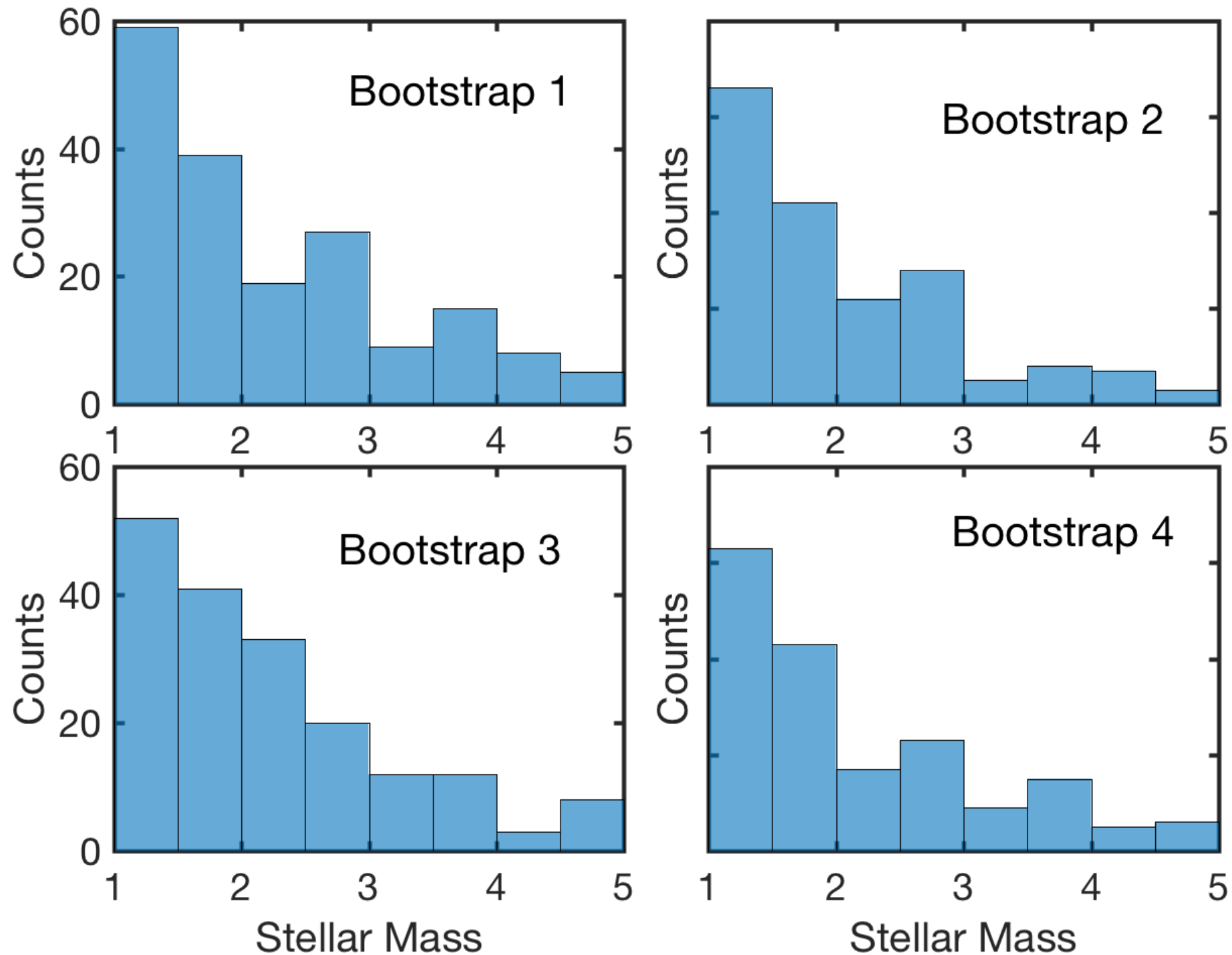


Maximum Likelihood for Power Law given Observed Stellar Mass Data

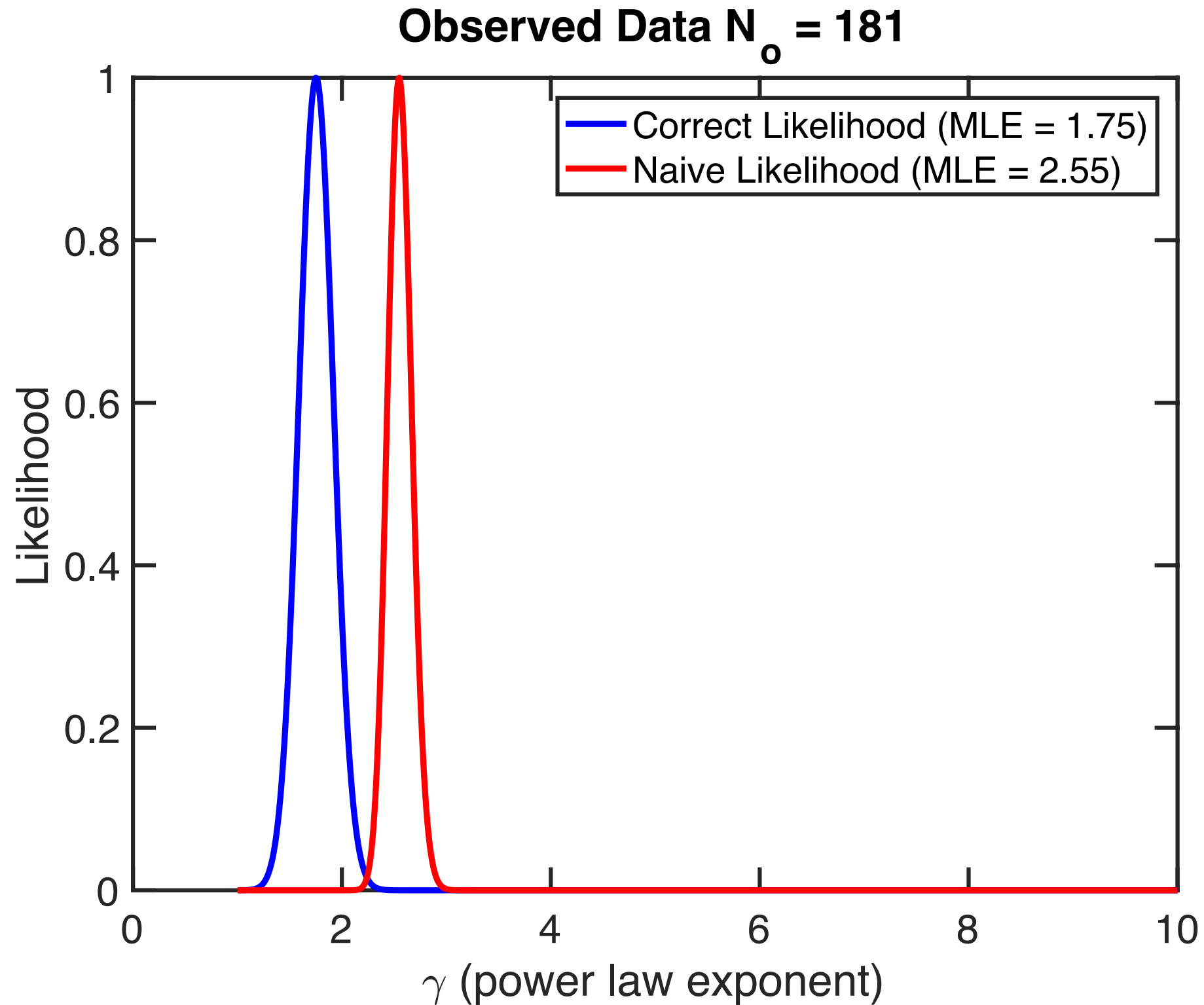


Standard Error on exponent?

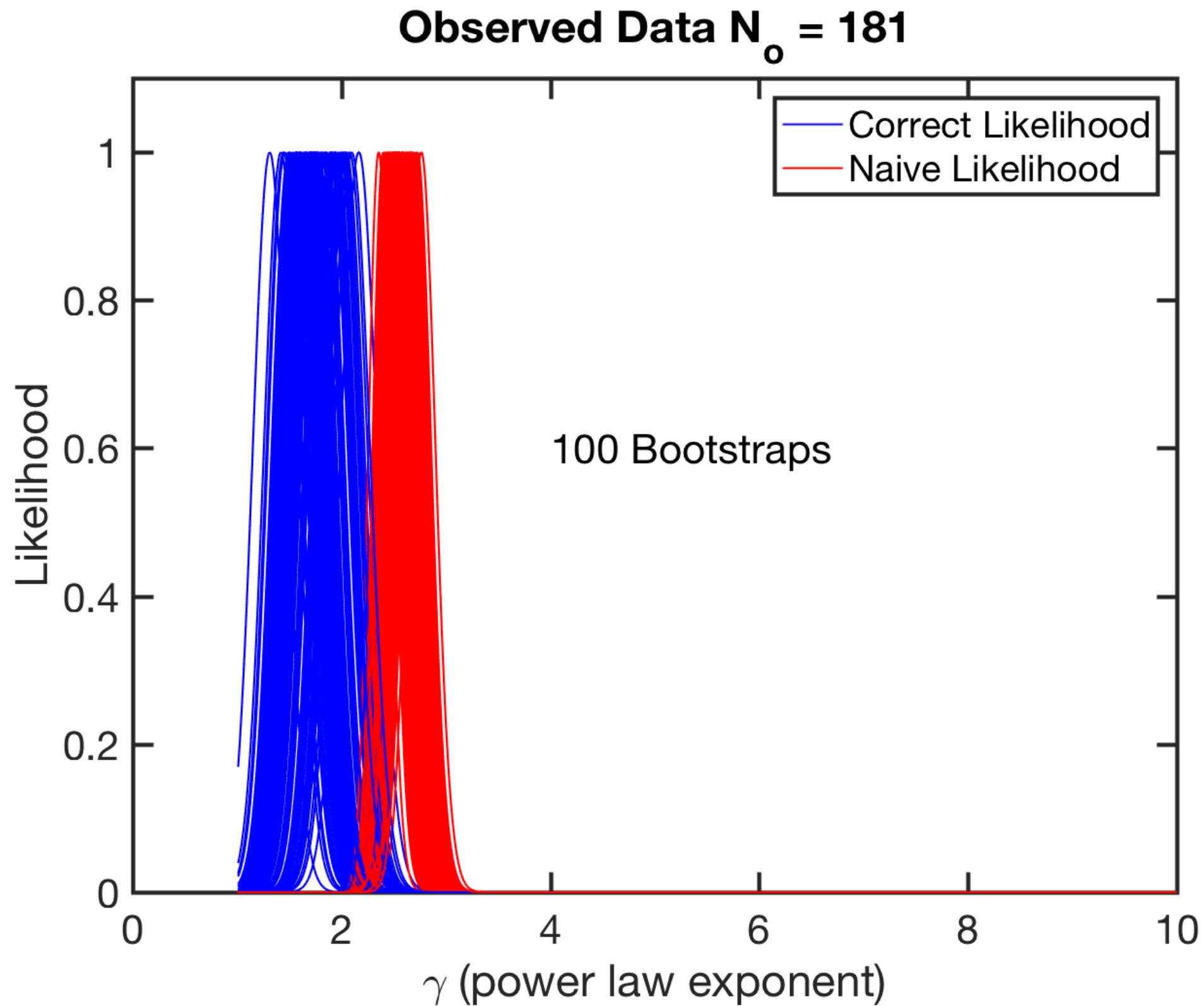
Bootstrapping the Observed Stellar Mass Data



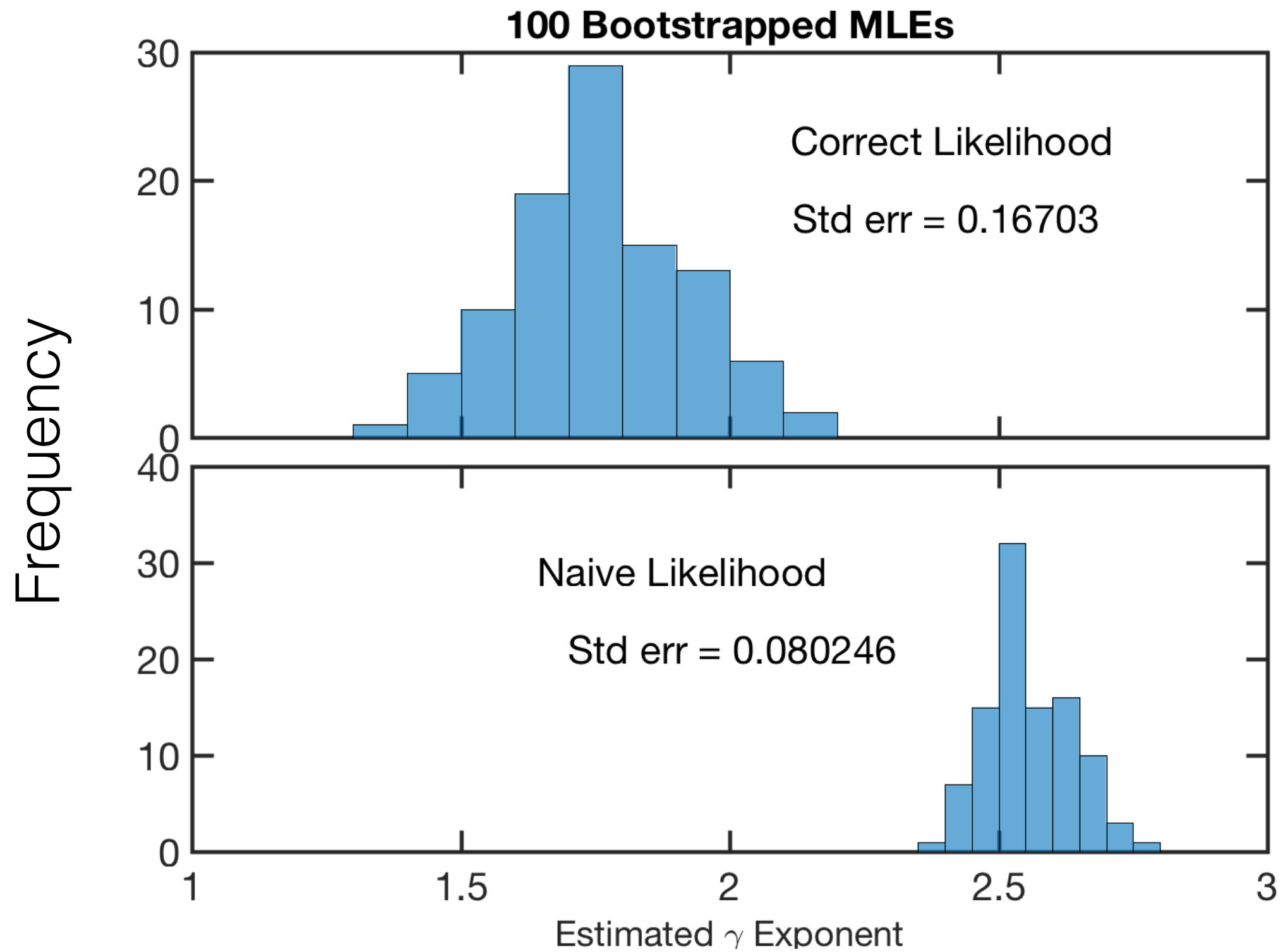
Maximum Likelihood for Power Law given Observed Stellar Mass Data



Bootstrapping the Maximum Likelihood Estimates



Bootstrapping the Maximum Likelihood Estimates



Regression

- Fitting a function $E[y \mid x] = f(x; \theta)$ for the mean relation between y and x
- Zoo of Methods (see F&B Ch 7, Ivezic Ch 8)
- Basic Approaches
 - Ordinary Least Squares (homoskedastic scatter)
 - Generalised Least Squares (heteroskedastic, correlated scatter)
 - Weighted Least Squares (minimum χ^2) (known variance)
 - Maximum Likelihood
- Real data problems require more complex statistical modelling