

# Astrostatistics: Sat 10 Feb 2017

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

- Example Sheet 1 uploaded
  - Fri Feb 16 (2:30 pm, Room MR5)
- Fitting Statistical Models to Astronomical Data
  - Bayesian Inference —> Computation, examples
  - References: Ivezić, Ch 5, F&B Ch 3, Gelman BDA
  - Hogg, D., 2012. “Data analysis recipes: Probability calculus for inference.” <https://arxiv.org/abs/1205.4446>

# Posterior Probability of the Dimensionality of Spacetime $P(D | D)$

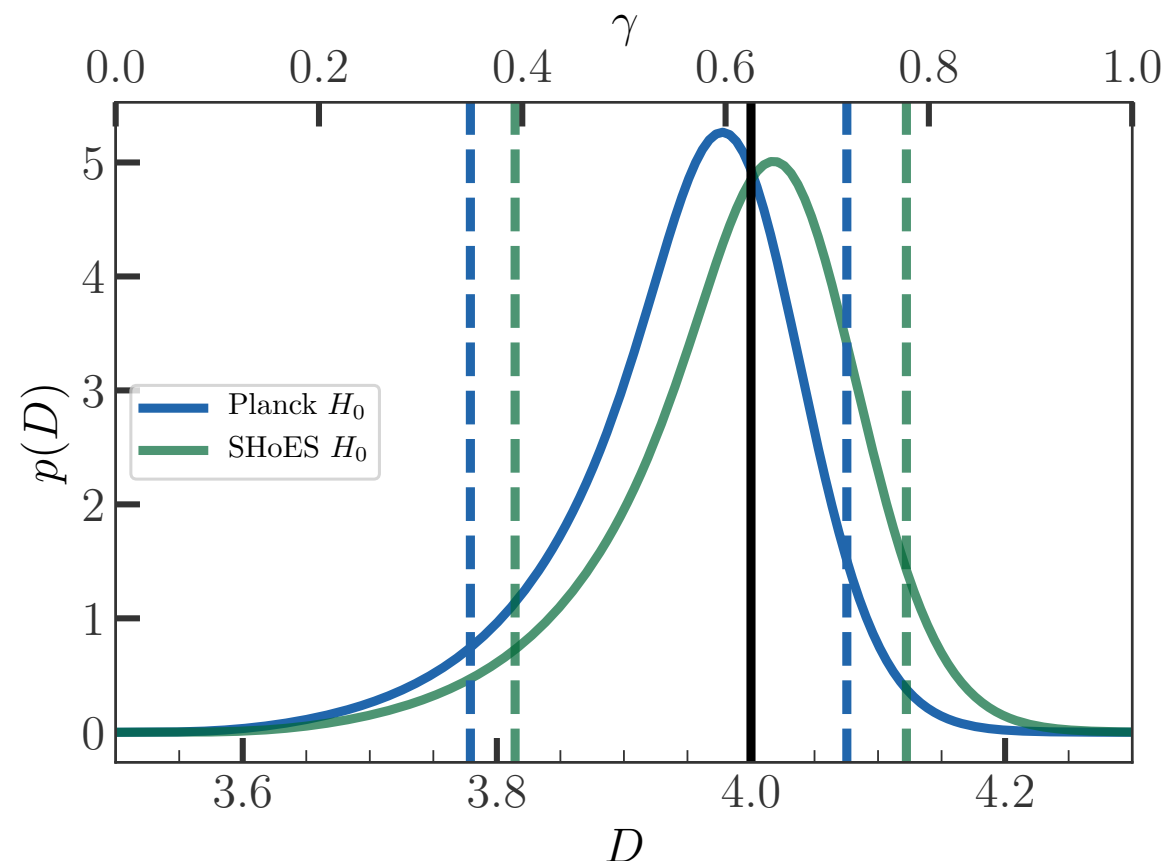


FIG. 1. Posterior probability distribution for the number of spacetime dimensions,  $D$ , using the GW distance posterior to GW170817 and the measured Hubble velocity to its host galaxy, NGC 4993, assuming the  $H_0$  measurements from [21] (blue curve) and [22] (green curve). The dotted lines show the symmetric 90% credible intervals. The equivalent constraints on the damping factor,  $\gamma$ , are shown on the top axis. GW170817 constrains  $D$  to be very close to the GR value of  $D = 4$  spacetime dimensions, denoted by the solid black line.

Pardo, Fishbach, Holz & Spergel. “Limits on the number of spacetime dimensions from GW170817”. arXiv:1801.08160

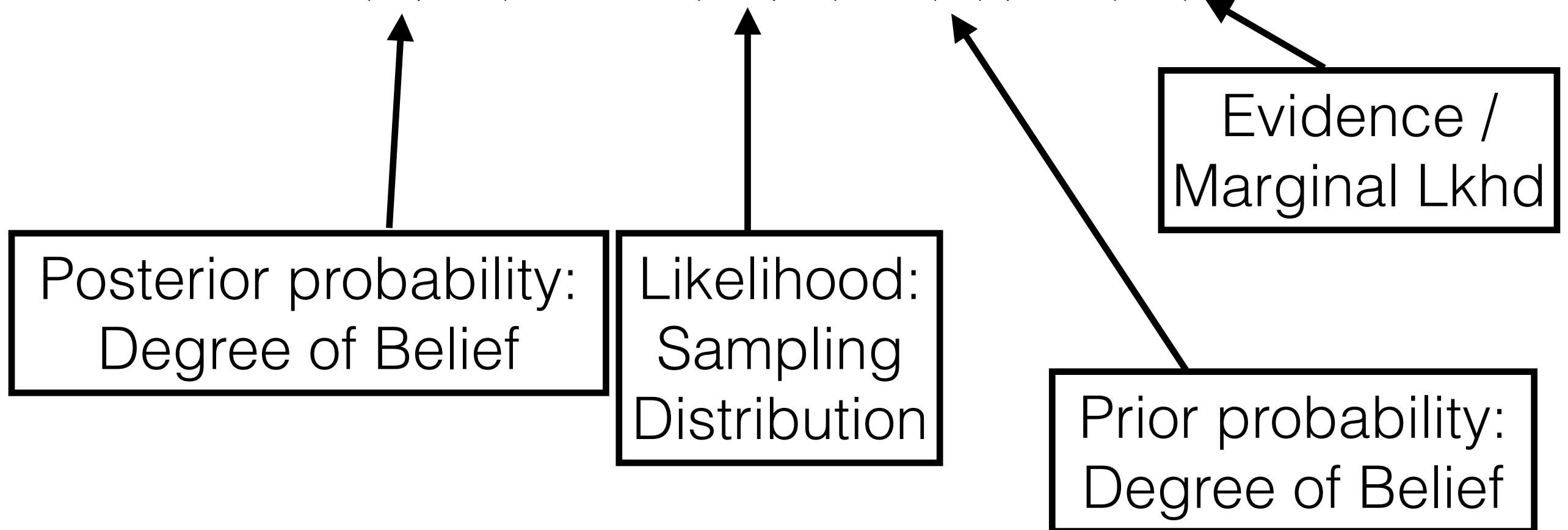
# Bayes' Theorem

Joint Probability of Data and Parameters:

$$P(D, \theta) = P(D | \theta) P(\theta) = P(\theta | D) P(D)$$

Probability of Parameters Given Data:

$$P(\theta | D) = P(D | \theta) P(\theta) / P(D)$$



# Simple Gaussian Example: Gelman BDA, Section 2.5

# 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  $P(\theta \mid D_{\text{ext}})$
  - 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)

# Mo' Bayes, Mo' problems

- Bayesian answer is the full posterior density  $P(\theta | D)$ , 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  $P(\theta | D)$  and compute expectations  $\mathbf{E}[f(\theta) | D]$
- 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