Astrostatistics: Mon 10 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

Bayesian Inference & Data Analysis

Recall: Probabilistic Generative Model for Linear Regression with (x,y) measurement errors

1. Population Distribution $\xi \sim N(\mu | \tau^2)$

Population Parameters: $\boldsymbol{\psi} = (\mu, \tau)$

2. Regression: $\eta_i | \, \xi_i \sim N(\alpha + \beta \xi_i, \sigma^2)$

Regression Parameters: $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$

Latent (true) Variables: (ξ_i, η_i)

3. Measurement Error: $[x_i, y_i] | \xi_i, \eta_i \sim N([\xi_i, \eta_i], \Sigma])$

Observed Data: (x_i, y_i)

Formulating Likelihood Function: Marginalising (integrating out) latent variables

"Complete Data Likelihood" (one datum)

$$P(x_i, y_i, \xi_i, \eta_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = P(x_i, y_i|\xi_i, \eta_i)P(\eta_i|\xi_i, \boldsymbol{\theta})P(\xi_i|\boldsymbol{\psi})$$

Measurement Error

Regression

Population Distribution

"Observed Data Likelihood" (one datum): integrate out latent variables

$$P(x_i, y_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = \int\int P(x_i, y_i|\xi_i, \eta_i)P(\eta_i|\xi_i, \boldsymbol{\theta})P(\xi_i|\boldsymbol{\psi})\, d\xi_i\, d\eta$$

Observed Data Likelihood (all data):

$$P(\boldsymbol{x}, \boldsymbol{y}|\,\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{N} P(x_i, y_i|\,\boldsymbol{\theta}, \boldsymbol{\psi})$$

Knowns and Unknowns

Population Parameters: $\boldsymbol{\psi} = (\mu, \tau)$

Regression Parameters: $\boldsymbol{\theta} = (\alpha, \beta, \sigma^2)$

Latent (true) Variables (ξ_i, η_i)

Observed Data: (x_i, y_i)

In Frequentist Statistics, distinction btw data and parameters: parameters are fixed and unknown, but not “random”. Only “data” are random realisations of random variables

Knowns and Unknowns

What is the nature of the latent variables (ξ_i, η_i) ?

They have a probability distribution:

$$(\xi_i, \eta_i) \sim P(\xi_i, \eta_i | \boldsymbol{\theta}, \boldsymbol{\psi}) = P(\eta_i | \, \xi_i, \boldsymbol{\theta}) P(\xi_i | \, \boldsymbol{\psi})$$

Often called "nuisance parameters"
parameters you need to introduce to complete the model
but are not the parameters of interest: $(\boldsymbol{\theta}, \boldsymbol{\psi})$

Also referred to as "missing data"
quantities that you didn't observe, but wish you had!
but relate to actual measurements (x,y)

Are the latent variables "data" or "parameters"?

Bayesian viewpoint

- There is a symmetry between data D and parameters θ - both are random variables described by probability distributions
- Actually they are described by a joint probability $P(D, \theta)$
- Data are random variables whose realisations are observed, parameters are RVs not observed
- Goal is to infer the unobserved parameters from the observed data using the rules of probability:
- Conditional Probability: $P(\theta | D) = P(D, \theta)/P(D)$
- Bayes' Theorem: $P(\theta | D) = P(D|\theta)P(\theta)/P(D)$
- Probability interpreted as degree of belief / uncertainty in hypotheses

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:

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

Normalisation Constant

Posterior probability: Degree of Belief

Likelihood: Sampling Distribution

Prior probability: Degree of Belief

Bayesian Inference of the Dimensionality of Spacetime

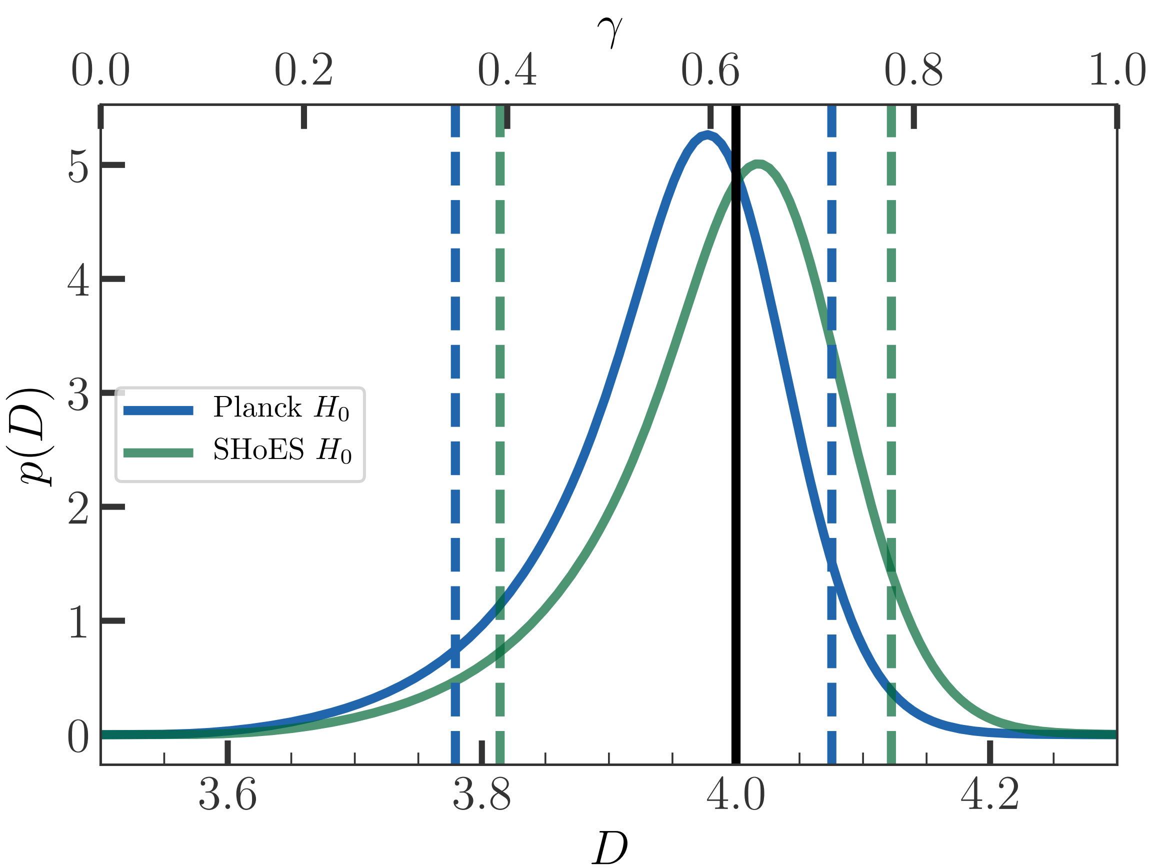

FIG. 1. Posterior probability distribution for the number of spacetime dimensions, D, using the GW distance posterior to GW170817 and the measured Hubble velocity to its host galaxy, NGC 4993, assuming the H_0 measurements from [21] (blue curve) and [22] (green curve). The dotted lines show the symmetric 90% credible intervals. The equivalent constraints on the damping factor, γ, 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.

Simple Gaussian Example
Frequentist Confidence vs. Bayesian credible intervals

Review: Frequentist Confidence Interval

$$Y_1, \ldots, Y_4 \text{ iid} \sim N(\mu, 1) \qquad \bar{Y} \sim N(\mu, \sigma^2/4)$$

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

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

$$[\bar{Y} - 0.5, \bar{Y} + 0.5] \text{ is a } 68\% \text{ confidence interval}$$

Under repeated experiments, 68% of the confidence intervals constructed this way will contain (cover) μ

This does NOT mean that μ is within [-1.07, -0.07] with 68% probability!

Simple Gaussian Example
Frequentist Confidence vs. Bayesian credible intervals

Bayesian credible interval

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

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

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

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

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