How to be a Bayesian?

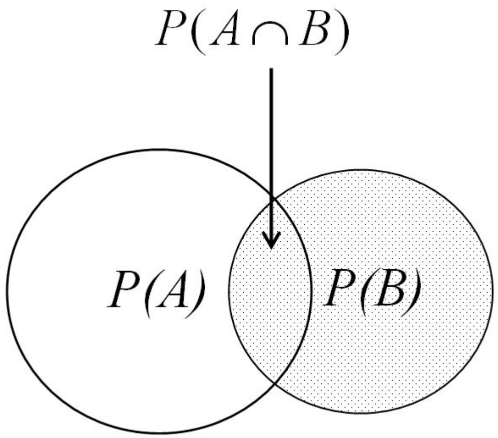
Christopher J. Moore

25/08/2020

STFC Introductory Solar System Plasmas Summer School

UNIVERSITY OF BIRMINGHAM | GRAVITATIONAL WAVE ASTRONOMY

Bayes' Theorem



The **product rule** as a law of probability

$$P(A \cap B) = P(A|B)P(B)$$

The **and** conjunction is symmetric between the two events/propositions A,B - therefore we have

$$P(A|B)P(B) = P(B|A)P(A)$$

Rearranging gives **Bayes' theorem**

$$\boxed{P(A|B) = P(B|A)\frac{P(A)}{P(B)}}$$

Bayes' Theorem

Despite its apparent simplicity, **Bayes' theorem** is extremely important - it is used in both frequentist and Bayesian statistics

We will use **Bayes' theorem** in the form

$$\boxed{P(M|D,I) = \frac{P(D|M,I)P(M|I)}{P(D|I)}}$$

D = our data, or observations, that we are going to analyse

M = the model, usually with some free parameters, θ

I = any other prior information or assumptions used/made (conceptually extremely important and always present; however, We usually omit to write this out)

Bayes' Theorem

$$P(M|D,I) = \frac{P(D|M,I)P(M|I)}{P(D|I)}$$

This equation arises so often we give each term a name (this is all just jargon but can get confusing - remember, each term is a probability)

$P(D|M,I)$ = **likelihood**; probability of getting the data from the model

$P(M|I)$ = **prior**; probability we assign to model being correct **before** performing the experiment

---

$P(M|D,I)$ = **posterior**; probability of model after performing experiment

$P(D|I)$ = **evidence;** acts as a normalisation constant.
usually not useful when estimating parameters for one model, but extremely important for comparing models

Bayes' Theorem

$$P(M|D,I) = \frac{P(D|M,I)P(M|I)}{P(D|I)}$$

This equation arises so often we give each term a name (this is all just jargon but can get confusing - remember, each term is a probability)

INPUTS

$P(D|M,I)$ = **likelihood**; probability of getting the data from the model

$P(M|I)$ = **prior**; probability we assign to model being correct **before** performing the experiment

OUTPUTS

$P(M|D,I)$ = **posterior**; probability of model after performing experiment

$P(D|I)$ = **evidence;** acts as a normalisation constant, usually not useful when estimating parameters for one model, but extremely important for comparing models

The Simplest Possible Bayesian Analysis

Our **data** is a single measurement, $D = 10$

Our prior assumptions I are that this measurement was made by an instrument subject to Gaussian noise with known variance $\sigma^2 = 3$

We have two competing models:

- **Null Hypothesis** the data is just noise, $D \sim \mathcal{N}(0, \sigma^2)$
- **Signal Hypothesis** there is a source of size A, so $D \sim \mathcal{N}(A, \sigma^2)$

We will treat these two hypotheses as being *a priori* equally likely - we will let the data decide which is prefered

For the signal hypothesis there is a free parameter A, for which we need to choose a prior; $P(A|I) = \mathcal{U}(-30, 30)$

The Simplest Possible Bayesian Analysis

The Signal Hypothesis:

From the statement of the problem we know that the likelihood is given by Gaussian our prior P(A)=1/60 is uniform over the allowed range

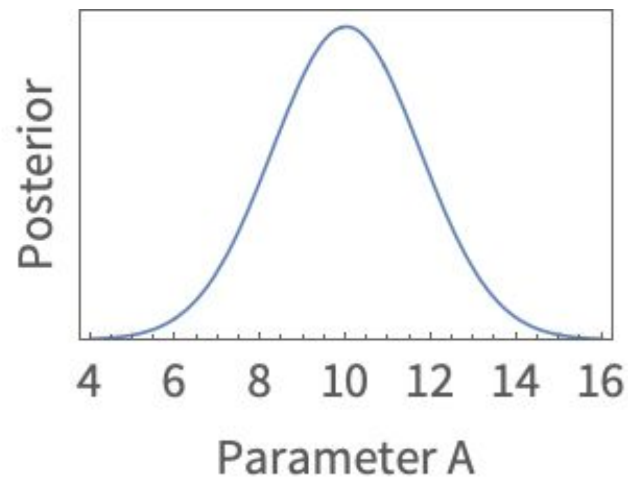
Using Bayes' theorem we find the **posterior**

$$P(A|D,I) = \frac{1}{Z}\frac{1}{60}\frac{\exp\left(\frac{-(D-A)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \quad \text{if } A \in (-30, 30)$$

We can also find the **evidence** $Z \equiv P(D|I)$ by properly normalising this distribution

$$Z = \int_{-30}^{30} \mathrm{d}A \, \frac{1}{60}\frac{\exp\left(\frac{-(D-A)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$

$$\approx 0.0167$$



The Simplest Possible Bayesian Analysis

The Null Hypothesis

For the null hypothesis, there are no free parameters.

Therefore, we do not need to integrate to find the evidence; we simply evaluate the likelihood

$$Z = \frac{\exp\left(\frac{-D^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}} \approx 5.78 \times 10^{-8}$$

The Bayes' Factor

The ratio of the evidences tells us the **signal hypothesis** is preferred

$$\frac{Z_{\text{signal}}}{Z_{\text{null}}} \approx 2.88 \times 10^5 \gg 1$$

Summary

$$P(M|D,I) = \frac{P(D|M,I)P(M|I)}{P(D|I)}$$

Bayes' theorem provides a possible starting point for a wide range of possible analyses comparing models to experimental data.

A Bayesian analysis typically takes as an "input" some **data** D, **model**(s) M, a **likelihood** and a **prior** function

For a single model, the aim of a Bayesian analysis is usually to find the **posterior** distribution on the model parameters

For multiple model, we may also want to compute the **evidence** for each model so we can compare models by computing **Bayes' factors**

The remainder of this session will use the resources in the public repository
https://github.com/cjm96/SASP_SkillsSession_Bayesian/blob/master/Bayesian.ipynb