Probability and Bayes' Theorem
Notes from: Bayesian Cosmology

Thomas Kitching

March 17, 2005

"Probability theory is nothing but common sense reduced to calculation"$_{Laplace}$

The general problem

The notion of probability is *central* to cosmology. It allows us to apply logical reasoning to data in order to deduce, with a particular – and calculable – degree of certainty, properties of that data. The general question that one could pose is: How do we reason in situations where it is not possible to argue with certainty?

In other words, we want to apply the *deductive* reasoning of everyday experience, and pure maths (Boolean logic), to problems which require *inductive* reasoning: that is going from general principles to particulars (or effects/observations to possible causes). This is an example of what is known as the 'epistemic' philosophical stance on probability:

- *Aleatory* is the concept of probability used for example in games of chance. It deals with predicting the future outcome of random physical processes. It can further be subdivided into phenomenon which are in principle predictable and those which are inherently unpredictable.
- *Epistemic* is the concept of probability used to describe the property or effect of a system when the causative circumstances for that property are unknown or uncertain.

There are a number of interpretations that one can use when dealing with the notion of probability. These can be categorised under the following, broad, headings:

- *Relative Frequency.* The relative frequency (probability) of an event arises from the number of times this event would occur defined relative to an infinite ensemble of 'identical' experiments.
- *Propensity.* Here probabilities are objective properties of a system.
- *Bayesian Probability.* Takes probability theory as a 'logic of inference', probabilities are the 'likelihood' of an event being such given all available information (data).

The formulation

Axioms

The axioms of probability theory stem directly, and are simply an expression of, Boolean logic. They were first formulated by Cox (1946) (and later by Kolmogorov):

$$0 \leq \; P(X \mid I) \; \leq 1$$

A probability is a number between 1 and 0.

$$P(X \mid I) + P(\bar{X} \mid I) = 1$$

The probability of X being true *or* false is 1.

$$P(X, Y \mid I) = P(X \mid Y, I) \times P(Y \mid I)$$

The probability of X being true is the product of the probability of X being true given that Y is true multiplied by the probability that Y is true.

Bayes' Theorem

Bayes theorem can be easily derived from the product rule and is written:

$$P(X \mid Y, I) = \frac{P(Y \mid X, I) \times P(X \mid I)}{P(Y \mid I)}$$

The power of the Bayesian approach is manifest when, for example, X is replaced by *hypothesis* and Y with *data*. It can then be seen that Bayes' theorem relates the probability of the hypothesis and the data to the probability of the data given the hypothesis. For example if you have some data $\{D\}$ and a hypothesis H parameterised by a set of parameters $\{\Theta\}$ then one usually wants the probability of the parameters having certain values given the data: $P(\Theta \mid D, H)$. This in general is not readily calculable, however the transposition $P(D \mid \Theta, H)$ usually is using a theoretical construct for the noise. For example assuming a Gaussian the probability of a single variable, x given $\Theta = (\mu, \sigma)$ (mean and variance) is:

$$P(x \mid \Theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x-\mu)^2}{2\sigma^2}$$

Bayes' theorem relates the two:

$$P(\Theta \mid D, H) = \frac{P(D \mid \Theta, H) \times P(\Theta \mid H)}{P(D \mid H)}$$

There are various formal names for the terms:

- $P(D \mid \Theta, H)$ is the 'likelihood', $L(D; \Theta)$ (of the data given a hypothesis).
- $P(\Theta \mid D, H)$ is the 'posterior' probability, which represents the state of knowledge of about the system inlight of the data.
- $P(\Theta \mid H)$ is the 'prior', this contains all the information we know about the probability of a given parameter before the experiment.
- $P(D \mid H)$ is the 'evidence'. This is usually taken to be a normalitive constant, although can be used to distinguish between competing hypotheses.

Useful Formulae

Here is listed a few useful formulae that will be discussed in the lecture.

The Fisher (information) matrix is defined as:

$$F_{ij} = \langle \frac{\partial^2(-\ln(\mathrm{L}))}{\partial\Theta_i \partial\Theta_j} \rangle$$

and for the Gaussian case:

$$F_{ij} = \frac{1}{2}\mathrm{Tr}(\mathrm{C}^{-1}\mathrm{C}_{,\mathrm{i}}\,\mathrm{C}^{-1}\mathrm{C}_{,\mathrm{j}} + \mathrm{C}^{-1}\mathrm{M}_{\mathrm{ij}})$$

For evidence analysis in the Gaussian case:

$$\epsilon \approx P(D \mid \Theta_0, H)\frac{V_{posterior}}{V_{prior}}$$

Further Reading

- Data Analysis: A Bayesian Tutorial
  D.S. Sivia
- Karhunen-Loeve eigenvalue problems in cosmology
  Max Tegmark, Andy Taylor, Alan Heavens astro-ph 9603021
- Revealing the Nature of Dark Energy Using Bayesian Evidence
  T.D.Saini, J.Weller, S.L. Bridle astro-ph 0305526