Introduction to Bayes

Alan Heavens

September 3, 2018

ICIC Data Analysis Workshop

Overview

1. Inverse Problems
2. The meaning of probability
   - Probability rules and Bayes theorem
   - $p(x|y)$ is not the same as $p(y|x)$
3. Parameter Inference
   - The posterior $p(parameters|data)$
   - How to set up a problem
   - Priors
4. Sampling
   - Case study: Eddington 1919 Eclipse expedition
   - The perils of not marginalising
5. The Monty Hall problem

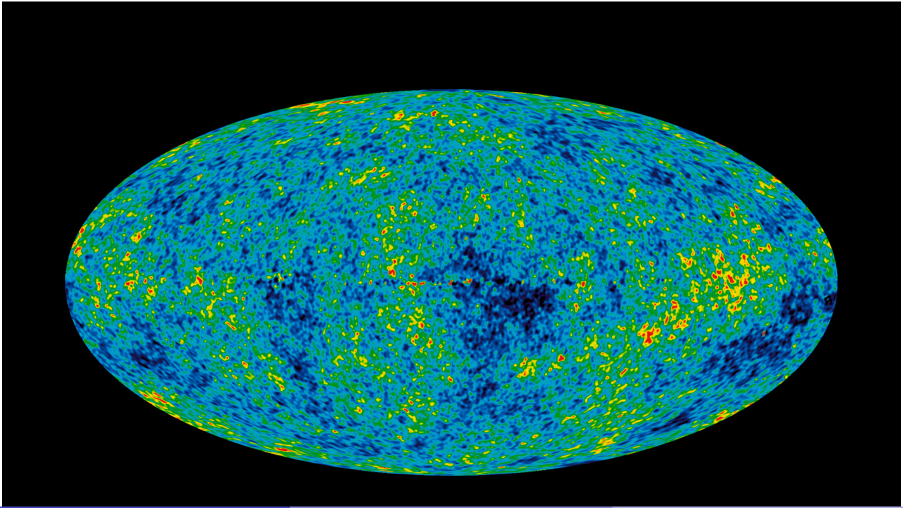
Some books for further reading

- D. Silvia & J. Skilling: Data Analysis: a Bayesian Tutorial (CUP) P. Saha: Principles of Data Analysis. (Capella Archive) `http://www.physik.uzh.ch/~psaha/pda/pda-a4.pdf`
- T. Loredo: Bayesian Inference in the Physical Sciences `http://www.astro.cornell.edu/staff/loredo/bayes/`
- M. Hobson et al: Bayesian Methods in Cosmology (CUP)
- D. Mackay: Information Theory, Inference and Learning Algorithms. (CUP) `http://www.inference.phy.cam.ac.uk/itprnn/book.pdf`
- A. Gelman et al: Bayesian Data Analysis (CRC Press)

Inverse Problems

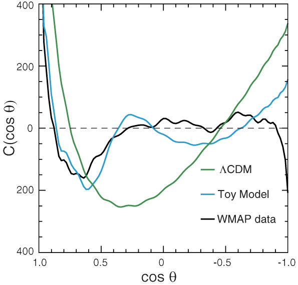
- Analysis problems are *inverse problems*: given some data, we want to infer something about the process that generated the data
- Generally harder than predicting the outcome, given a physical process
- The latter is called *forward modelling*, or a *generative model*
- Typical classes of problem:
- *Parameter inference*
- *Model comparison*

Typical problem: analyse WMAP Cosmic Microwave Background Data



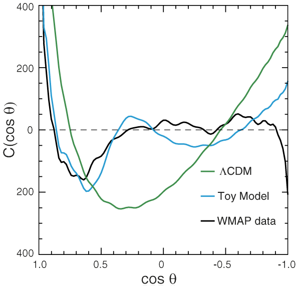
WMAP Cosmic Microwave Background Data

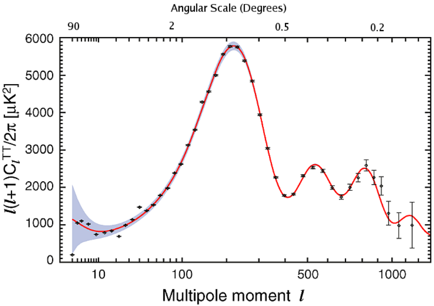
ΛCDM fits WMAP data well:



WMAP Cosmic Microwave Background Data

ΛCDM fits WMAP data well:





Bayesian Inference

What questions do we want to answer?

Parameter Inference:

- I have a set of (x, y) pairs, with errors. If I assume $y = mx + c$, what are m and c?
- I have detected 5 X-ray photons. What is the luminosity of the source and its uncertainty?
- Given LIGO gravitational wave data, what are the masses of the inspiralling objects?
- Given the Planck CMB map, how much Dark Matter is there?

Bayesian Inference

What questions do we want to answer?

Model Comparison:

- Do data support General Relativity or Newtonian gravity?
- Is ΛCDM more probably than alternatives?

The meaning of probability

- Probability describes the *relative frequency of outcomes in infinitely long trials* (Frequentist view)
- Probability expresses a *degree of belief* (Bayesian view)
- *Logical proposition:* a statement that could be true or false
- $p(A|B)$ is the degree to which truth of a logical proposition B implies that A is also true
- The Bayesian view expresses what we often want to know, e.g.
- *given the Planck CMB data, what is the probability that the density parameter of cold dark matter is between 0.3 and 0.4?*

Probability rules

- $p(x) + (\text{not } x) = 1$ (sum)
- $p(x,y) = p(x|y)\,p(y)$ (product)
- $p(x) = \sum_k p(x, y_k)$ (marginalisation over all possible discrete y_k values)
- $p(x) = \int p(x,y)\,dy$ (marginalisation, continuous variables. p = pdf)
- $p(x,y) = p(y,x) \rightarrow$ Bayes theorem:

Bayes Theorem

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

-

Conditional probabilities

Avoid the probability 101 mistake

$p(x|y)$ is not the same as $p(y|x)$

e.g.

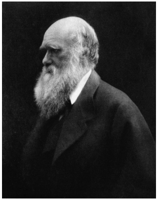


Figure: Julia Margaret Cameron

- $x =$ is male; $y =$ has beard
- $p(y|x) \sim 0.1$
- $p(x|y) \sim 1$

Example of $p(x|y) \neq p(y|x)$ errors

Medical test

An allergy test gives a positive result (T) in allergic patients (A) with $p = 0.8$, and has a false positive rate of 0.1. You get a positive result. If 0.01 of the population are allergic, what is the probability that you are?

- Rule 1: write down what you want to know. It is ...
- $p(A|T)$
- What we know are: $p(T|A) = 0.8$, $p(T|\sim A) = 0.1$, $p(A) = 0.01$
- Use Bayes theorem:

$$p(A|T) = \frac{p(T|A)p(A)}{p(T)}$$

- Marginalisation in the denominator: $p(T) = p(T, A) + p(T, \sim A)$
- and write $p(T, A) = p(T|A)p(A)$ and similarly for $p(T, \sim A)$:
-

$$p(A|T) = \frac{p(T|A)p(A)}{p(T|A)p(A) + p(T|\sim A)p(\sim A)}$$

Allergy problem

Medical test

An allergy test gives a positive result (T) in allergic patients (A) with $p = 0.8$, and has a false positive rate of 0.1. You get a positive result. If 0.01 of the population are allergic, what is the probability that you are?

- $$p(A|T) = \frac{p(T|A)p(A)}{p(T|A)p(A) + p(T|\sim A)p(\sim A)}$$
- Put in numbers:
$$p(A|T) = \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} = 0.075$$
- So there is still a 92.5% chance that you do not have the allergy.

Notation

- **Data** d; **Model** M; Model **parameters** θ
- Rule 1: write down what you want to know
- Usually, it is the probability distribution for the parameters, given the data, and assuming a model.
- $p(\theta|\mathrm{d}, M)$
- This is the **Posterior**
- To compute it, we use Bayes theorem:

$$p(\theta|d, M) = \frac{p(d|\theta, M)p(\theta|M)}{p(d|M)}$$

- where the **Likelihood** is $\mathcal{L}(\mathrm{d}|\theta) = p(\mathrm{d}|\theta, M)$
- and the **Prior** is $\pi(\theta) = p(\theta|M)$
- $p(\mathrm{d}|M)$ is the **Bayesian Evidence**, which is important for Model Comparison, but not for Parameter Inference.
- Dropping the M dependence

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{p(d)}$$

The Posterior

$$p(\theta|d, M)$$

If you just try long enough and hard enough, you can always manage to boot yourself in the posterior. A.J. Liebling.

It is all probability

The Posterior

Everything is focussed on getting at $p(\theta|\mathrm{d})$.

Computing the posterior

$p(\theta|\mathrm{d}) \propto \mathcal{L}(\theta)\,\pi(\theta)$.

We need to analyse the problem:

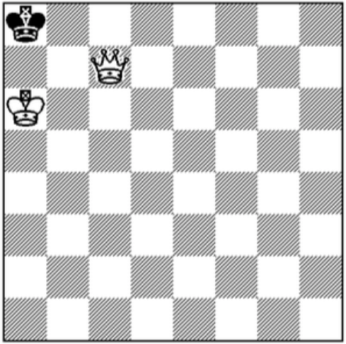
What are the data, d?
What is the model for the data?
What are the model parameters?
What is the likelihood function $\mathcal{L}(\theta)$?
What is the prior $\pi(\theta)$?

Priors

Bayesian: prior = (usually) the state of knowledge before the new data are collected.
For parameter inference, the prior becomes unimportant as more data are added and the likelihood dominates.
For model comparison, the prior remains important.
Issues:

- One usually wants an 'uninformative' prior, but what does this mean?
- Typical choices: $\pi(\theta) =$ constant; $\pi(\theta) \propto 1/\theta$ - so-called Jeffreys prior (by Astronomers)

Priors



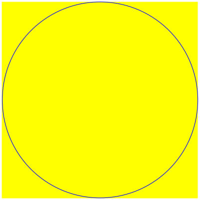
Credit: Daniel Mortlock

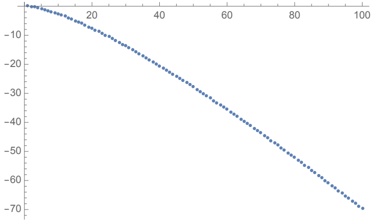
Uninformative prior

A flat prior seems natural, but consider this problem. Imagine cartesian coordinates in N dimensions, with the prior range being $(-\frac{1}{2}, \frac{1}{2})$ for all coordinates. The prior probability of being inside the N-sphere which just fits inside the prior volume is

$$\frac{\pi^{N/2}}{2^N \Gamma(1 + N/2)}$$

$\log_{10} p$ vs N



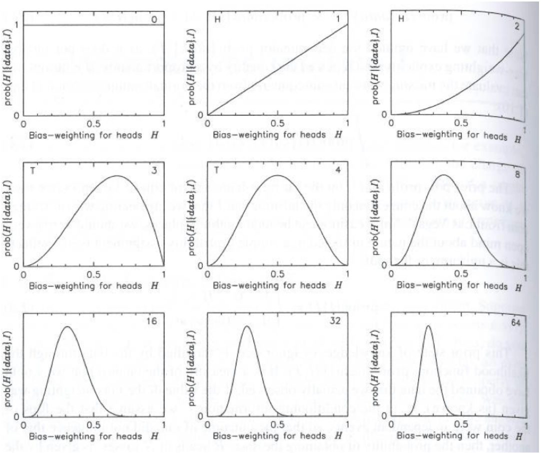


An apparently uninformative prior may be *highly informative* when viewed a different way.

Case Study: fair coin?

From Sivia & Skilling.

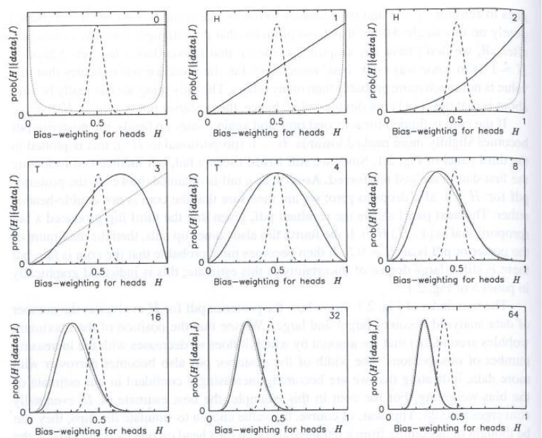
Model: probability of a head is θ. Uniform prior in θ assumed. Sequence is HHTT...



Second panel: $p(\theta|H) \propto p(H|\theta)\,\pi(\theta) = \theta$

Case Study: fair coin?

3 different priors, shown in first panel.

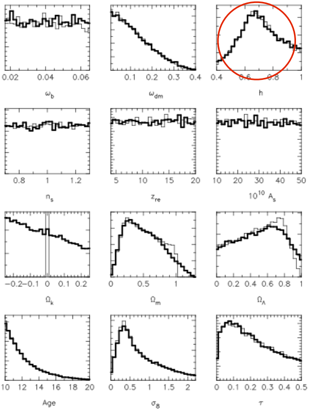


After many data are collected, the posterior becomes insensitive to the prior. In this case, the highly informative prior that supposes the coin is almost fair needs more data to overrule the prior.

High-dimensional priors

An interesting non-experiment with the VSA (Slozar et al. 2003)

Priors: $\Lambda \geq 0$, $10 \leq \mathrm{Age/Gyr} \leq 20$.

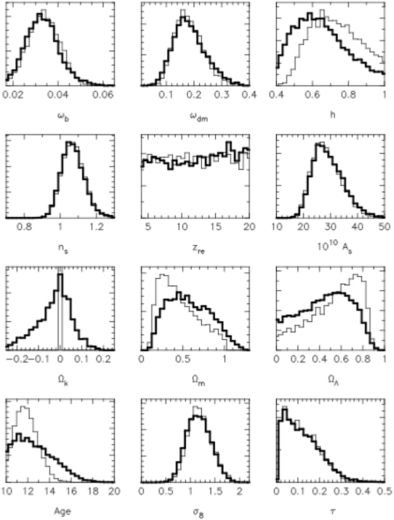


Hubble parameter $h \sim 0.7$, in line with expectations.

BUT...there are NO data here. This is the prior, marginalised over an odd shape.

High-dimensional priors

Now with the data



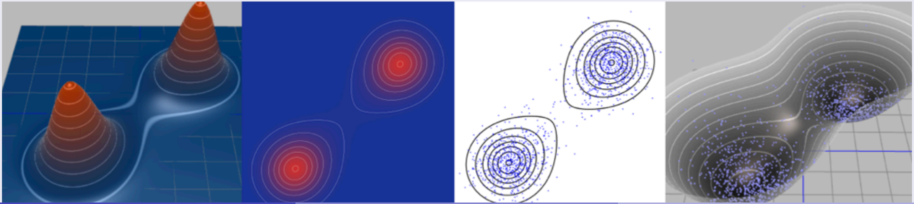
Sampling

The posterior is rarely an analytic function, and evaluating it on a grid in parameter space is usually prohibitively expensive if there are more than 2 or 3 parameters.

MCMC

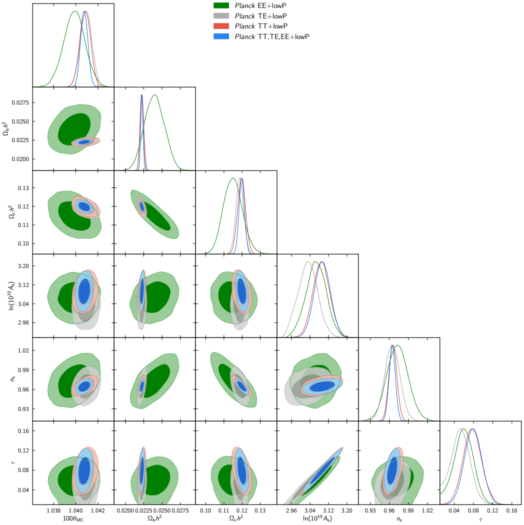
Standard technique is MCMC (Markov Chain Monte Carlo), where random steps are taken in parameter space, according to a proposal distribution, and accepted or rejected according to the Metropolis-Hastings algorithm. This gives a chain of samples of the posterior (or the likelihood), with an expected number density proportional to the posterior.

MCMC example



Planck parameter inference

Assuming ΛCDM



Case study: Eddington 1919 Eclipse expedition

In General Relativity, light is bent by the Sun through an angle $\frac{4GM}{rc^2}$.
In Newtonian theory, the bend angle is $\frac{2GM}{rc^2}$.

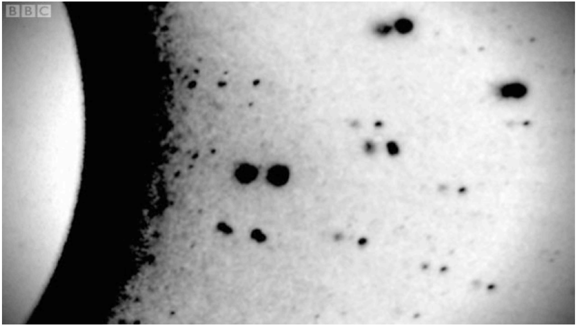


Figure: Illustration of the lensing effect (highly magnified).

If we treat this as a parameter inference problem: bend angle at the limb of the Sun $= \alpha$, and we will infer α.

Case study: Eddington 1919 Eclipse expedition

Analyse the experiment:

- **What are the data**?
- Measurements of displacements (Dx, Dy) of (7) stars, between eclipse plate(s) and a reference plate.
- **What is the model?**
- Displacements are radial, with magnitude α (arcsec) for light grazing the Sun.
- **What are the model parameters?**
- α
- **What is the likelihood function?**
- Measurement errors are Gaussian (assumption!)
- **What prior should we choose?**
- Uniform $\pi(\alpha)$ =constant.

Case study: Eddington 1919 Eclipse expedition. Nuisance Parameters

Hang on.

- Reference plate may not be centred correctly
- Plates may be rotated with respect to each other
- Eclipse plate may have be scaled in size (thermal effects/different instruments)
- Model for the data is

$$\begin{aligned} Dx &= ax + by + c + \alpha E_x \\ Dy &= dx + ey + f + \alpha E_y. \end{aligned} \qquad (1)$$

- 7 parameters, including 6 *nuisance parameters*
- Likelihood

$$\mathcal{L} \propto \prod_{\text{stars } i} \exp\left\{ -\frac{\left[Dx_i - (ax_i + by_i + c + \alpha E_{xi})\right]^2}{2\sigma_i^2} \right\}$$

Results from Plate II

Using only displacements in R.A. (4 parameters)

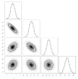


Figure: Hamiltonian MCMC samples of distribution of nuisance parameters a, b, c and bending at Solar limb α (in arcsec; GR predicts 1.75).

Marginalising over a, b, c gives $p(\alpha) = \int p(a, b, c, \alpha)\, da\, db\, dc$.

Marginalise over nuisance parameters!

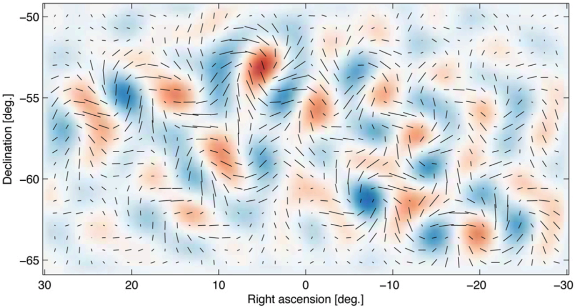


Figure: Exquisite BICEP B-mode CMB map (Credit: BICEP team).

BICEP and losing the Nobel Prize

Primordial Gravitational Waves: BICEP2 Announces First Direct Evidence Of Cosmic Inflation

By News Staff | March 17th 2014 09:47 AM | Print | E-mail

News Staff

Search This Blog

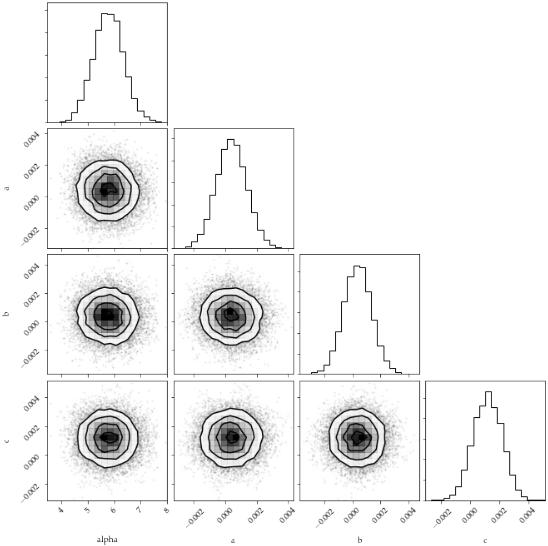
Researchers from the BICEP2 collaboration today announced the first direct evidence for this cosmic inflation. Their data also represent the first images of gravitational waves, or ripples in space-time. These waves have been described as the "first tremors of the Big Bang." Finally, the data confirm a deep connection between quantum mechanics and general relativity.

Figure: BICEP article (Science 2.0).

Dust contribution not marginalised over.

Back to Eddington 1919

What if we do not marginalise over a, b, c?



The Monty Hall Problem

The Monty Hall problem:

An exercise in using Bayes' theorem

You choose this one

Do you change your choice?

This is the Monty Hall problem

The Monty Hall Problem

The Monty Hall problem:

An exercise in using Bayes' theorem

Do you change your choice?

This is the Monty Hall problem