ICIC Data Analysis Workshop

Imperial College

5-8 September 2016

ICIC
Imperial Centre
for Inference & Cosmology

Sponsored by STFC
and Winton Capital

Science & Technology
Facilities Council

WINTON

Course Team

Alan Heavens

Andrew Jaffe

Daniel Mortlock

Jonathan Pritchard

Elena Sellentin

Roberto Trotta

Louise Hayward

ICIC

Logistics and events

- Fire exits
- I/O: Tea/coffee/lunch, toilets.
- Code of conduct (see website)
- ***Events:***
- **Breakfast** in Beit Hall 7-10 a.m. (Workshop starts 9.15)
- **Monday:** 5 p.m. Short talk by Geraint Harker (Winton Capital)
  5.30 p.m. Drinks reception, Level 8 Common Room and Roof, Blackett Lab.
- **Tuesday:** In Skempton LT201. Evening free
- **Wednesday:** 6 p.m. Barbecue, 58 Princes Gate
- **Thursday:** 1 p.m. Public engagement lunch (Roberto Trotta). End of Workshop 4 p.m.

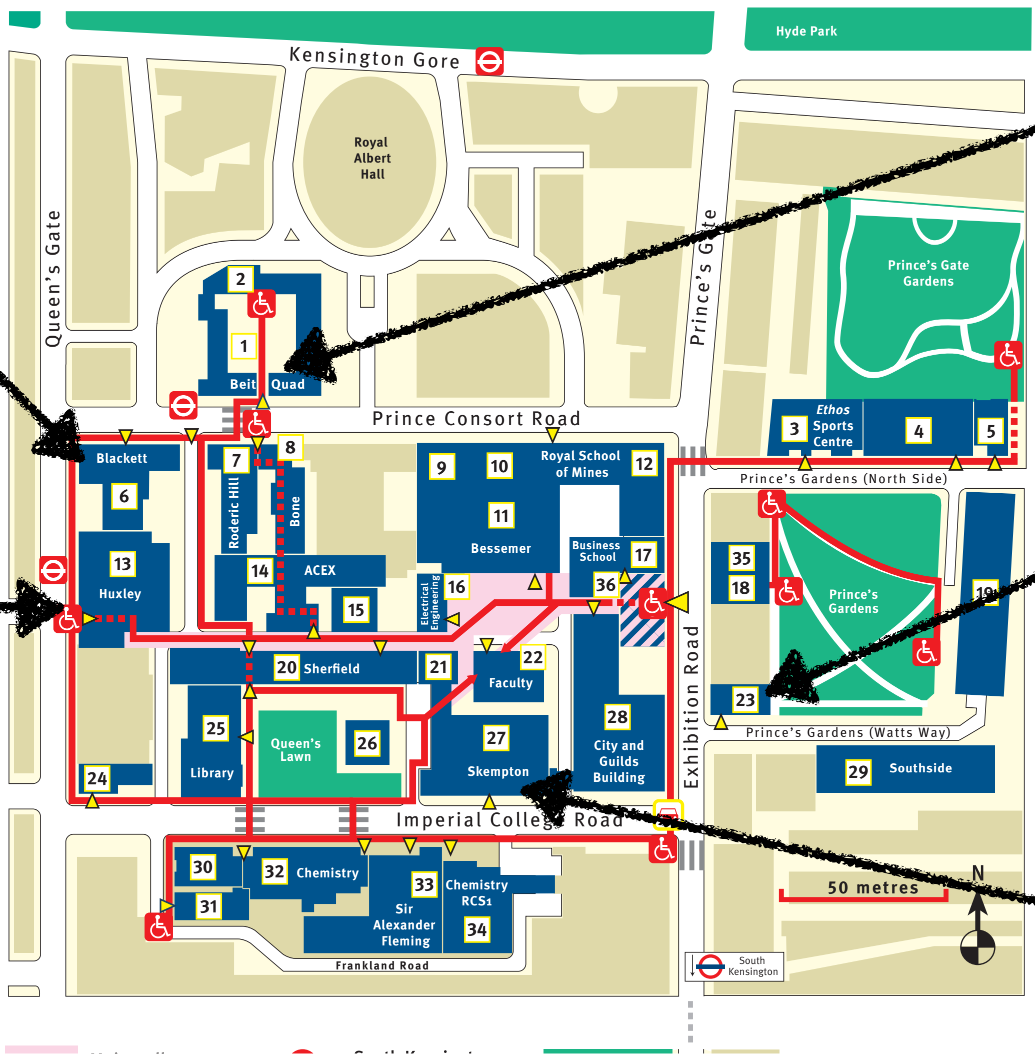
ICIC

Imperial College
London

South Kensington Campus

Drinks reception. Blackett Level 8 common room

We are here. Clore LT, Huxley Building



Accommodation: Beit Hall

Barbecue Wednesday 6 p.m 58 Princes Gate

Skempton Building LT201. Tuesday

Outline of course

- **Basic principles**
- **Sampling**
  - Numerical methods (Parameter inference)
  - Markov Chain Monte Carlo (MCMC)
  - Other samplers (Gibbs, HMC)
- **Generalised Linear Models**
- **Model comparison:** Bayesian Evidence
- **Bayesian vs Frequentist:** p-values
- **Bayesian Hierarchical Models**

Introduction to Bayesian Inference

Alan Heavens
Imperial College London

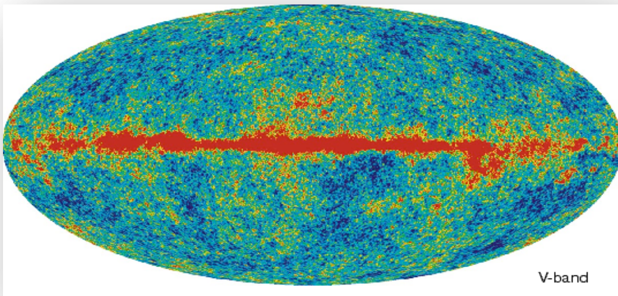
ICIC

Books

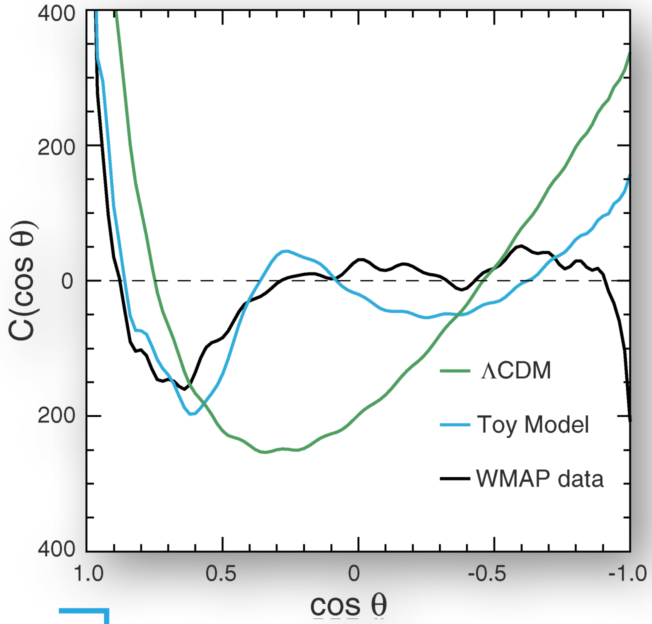
- **D. Silvia & J. Skilling**: *Data Analysis: a Bayesian Tutorial* (CUP)
- **P. Saha**: *Principles of Data Analysis.* (Capella Archive) *http://www.physik.uzh.ch/~psaha/pda/pda-a4.pdf*
- **T. Loredo**: *Bayesian Inference in the Physical Sciences* *http://www.astro.cornell.edu/staff/loredo/bayes/*
- **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)

ICIC

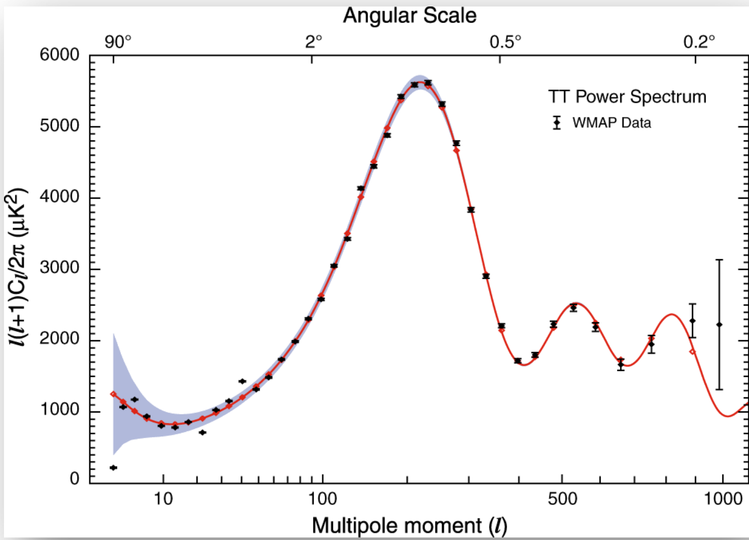
LCDM fits the WMAP data well.



Correlation Function $\langle \frac{\Delta T}{T}(\phi)\frac{\Delta T}{T}(\phi+\theta)\rangle$



Power Spectrum C_l



Inverse problems

- Most cosmological problems are *inverse problems*, where you have a set of data, and you want to infer something.
- - generally harder than predicting the outcomes when you know the model and its parameters
- Examples
  - Hypothesis testing
  - Parameter inference
  - Model selection

ICIC

Examples

- Hypothesis testing
  - Is the CMB radiation consistent with (initially) gaussian fluctuations?
- Parameter inference
  - In the Big Bang model, what is the value of the matter density parameter?
- Model selection
  - Do cosmological data favour the Big Bang theory or the Steady State theory?
  - Is the gravity law Einstein's General Relativity or a different theory?

ICIC

What is probability?

- Frequentist view: p describes the *relative frequency* of outcomes in infinitely long trials
- Bayesian view: p expresses our *degree of belief*
- p(A|B) = degree to which truth of some logical proposition B implies that the logical proposition A is also true.
- A *logical proposition* is a statement of fact that could be true or false
- The Bayesian view gives what we want: e.g. given Planck data, what is the probability that the density parameter of the Universe is between 0.9 and 1.1?

ICIC

Bayes' Theorem

- Rules of probability:
- $p(x) + p(\text{not } x) = 1$ — sum rule — $\text{not } x, \bar{x}, \sim x$
- $p(x,y) = p(x|y)\, p(y)$ — product rule
- $p(x) = \Sigma_k\, p(x,y_k)$ — marginalisation
- Sum$\longrightarrow$ integral — continuum limit (p=pdf)

$$p(x) = \int dy\, p(x,y)$$

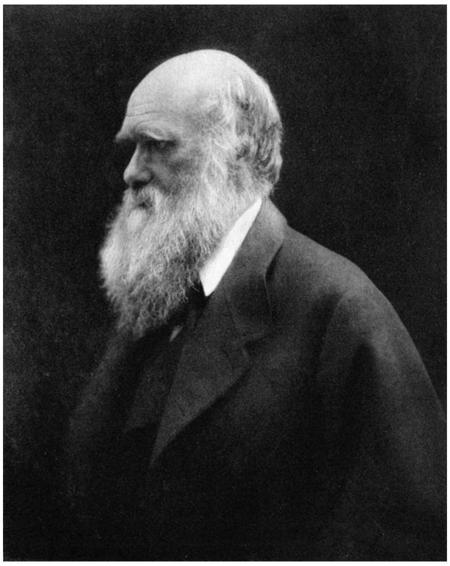
- p(x,y)=p(y,x) gives *Bayes' theorem*

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

ICIC

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

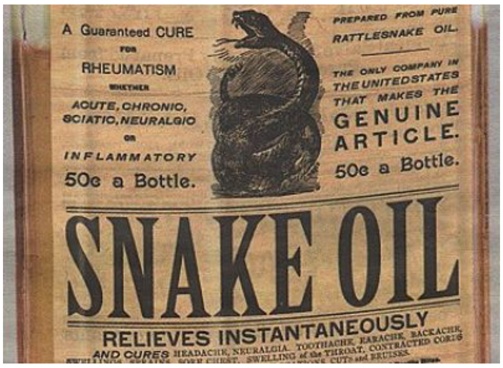
- x = is male; y = has beard
- $p(y|x) = 0.1$
- $p(x|y) = 1$



Julia Margaret Cameron

ICIC

Example



- Suppose a medical test for an allergy
  - gives a positive result with probability 0.8 in patients with the allergy
  - has a false positive probability (i.e. a positive result in patients without the allergy) of 0.1
  - And: The probability of having the allergy in the population is 0.01.
- You take the test, and it is positive (=T). What is the probability that you have the allergy (=A)?
- **RULE 1: WRITE DOWN WHAT IT IS YOU WANT:**
- $p(A \mid T)$
- RULE 2: THERE IS NO RULE n FOR n>1
- We know $p(T|A) = 0.8$; $p(T|{\sim}A) = 0.1$; $p(A) = 0.01$

ICIC

Solution

- We want p(A|T)
- We know p(T|A) = 0.8;   p(T|~A) = 0.1   ; p(A) = 0.0
- Bayes' theorem:

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

Marginalisation:

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

Product rule:

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

$$p(A|T) = \frac{0.8 \times 0.01}{0.8 \times 0.01 + 0.1 \times 0.99} = 0.075$$

ICIC

Not an effective test. A positive result = 7.5% chance of allergy.

The O J Simpson trial

- In 1994 the American football player O J Simpson was charged with murder
- Simpson was known to be violent
- The defence argued that only 1/2500 of violent people commit murder, *so the information that he is known to be violent is irrelevant*
- $p(M|A) = 0.0004$
- Statement was not challenged
- Simpson was acquitted

ICIC

Let us analyse this critically

- What key piece of information has been conveniently ignored by the defence?
- In this case, there is a body
- Given that the probability of being murdered in the USA is 0.00005, what is the probability that Simpson was the murderer given this evidence alone?
- Notation:
  - M = Simpson was the murderer
  - V = Simpson was violent
  - B = There is a body
- Exercise: apply Rule 1
- We want $p(M \mid B, V)$

ICIC

RULE 1: We want p(M|B,V)

$$p(M|B,V) = \frac{p(B|M,V)p(M|V)}{p(B|M,V)p(M|V) + p(B|M,\sim V)p(M|\sim V)}$$

- p(B|M,V) = 1
- p(M|V) = 0.0004
- p(~M|V) = 0.9996
- p(B|~M,V) < 0.00005

$$p(M|B,V) = \frac{1 \times 0.0004}{1 \times 0.0004 + 0.000005 \times 0.9996} = 0.89$$

A very different conclusion: 89% vs 0.04%

Bayes' Theorem and Inference

- If we accept *p* as a degree of belief, then what we often want to determine is (RULE 1)

$$p(\theta|x)$$

θ: model parameter(s), *x*: the data

To compute it, use Bayes' theorem $p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$

Note that these probabilities are all conditional on a) prior information *I*, b) a model *M*

$$p(\theta|x) = p(\theta|x, I, M) \text{ or } p(\theta|x\,I\,M)$$

ICIC

Posteriors, likelihoods, priors and evidence

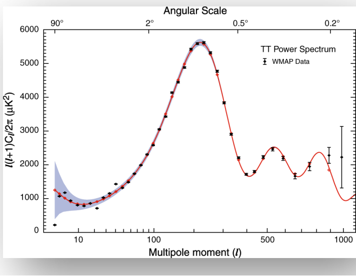
$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

Posterior — Likelihood L — Evidence or Marginal Likelihood — Prior

Reminder: we interpret these in the context of a model M, so all probabilities are conditional on M (and on any prior info I). E.g. $p(\theta) = p(\theta|M)$

The *evidence* looks rather odd – what is the *probability of the data*? For parameter inference, we can ignore it – it simply normalises the posterior. If you need it,

$$p(x) = \sum_k p(x|\theta_k)p(\theta_k) \text{ or } p(x) = \int d\theta\, p(x|\theta)p(\theta)$$



Noting that $p(x) = p(x|M)$ makes its role clearer.
In *model selection* (from *M* and *M'*), $p(x|M) \neq p(x|M')$

ICIC

Meaning

If we work within the framework of a single model,

$$p(\theta|x) \propto p(x|\theta)\, p(\theta)$$

Our state of knowledge after the experiment (the posterior)

The probability of getting the data we got (likelihood)

Our state of knowledge before the experiment (the prior)

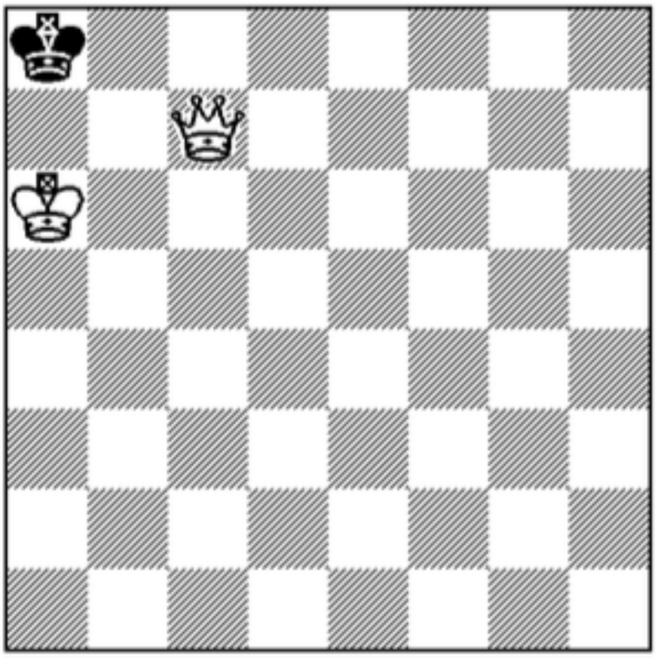
After the experiment, the posterior may act as the prior for the next experiment: we ‘update the prior’ with the information from the experiment

Self-consistent?

- Yes. Consider data from 2 experiments. We can do one of 3 things:
  - Define prior; obtain posterior from dataset 1; update the prior, then analyse dataset 2
  - As above, but swap 1 and 2
  - Define prior; obtain posterior from datasets 1 and 2 combined
- These have to (and do) give the same answers

A diversion on priors

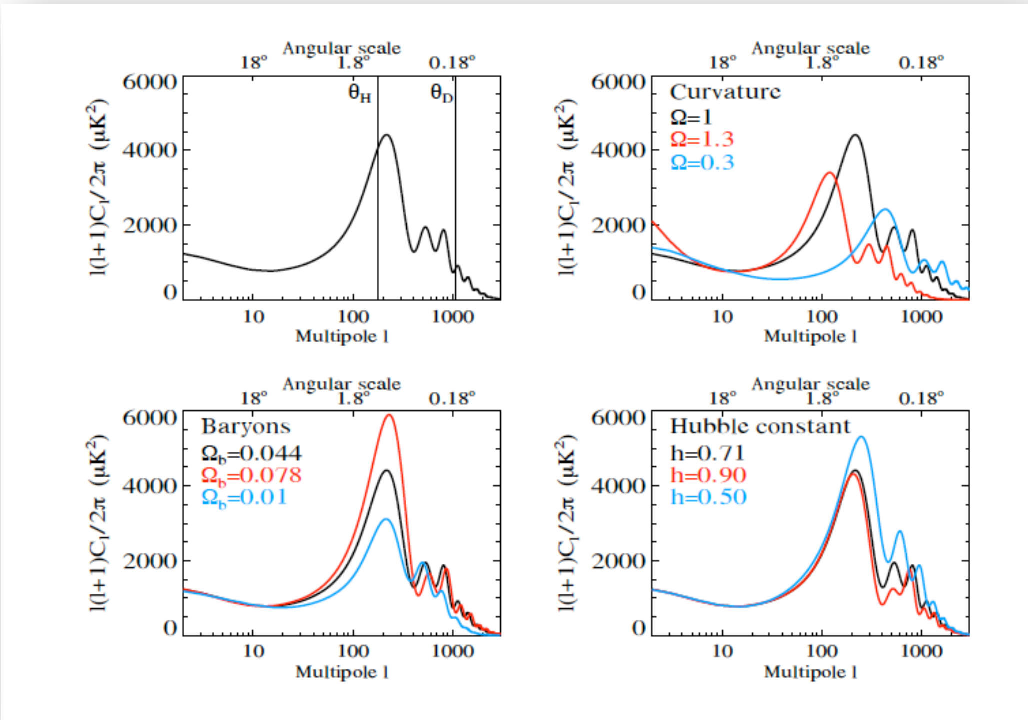
- You bring more prior information than you may think…

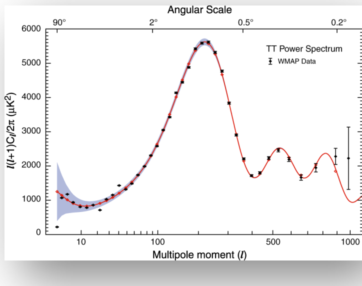


Credit: Daniel Mortlock

What is the probability that White wins?

Forward modelling $p(x|\theta)$





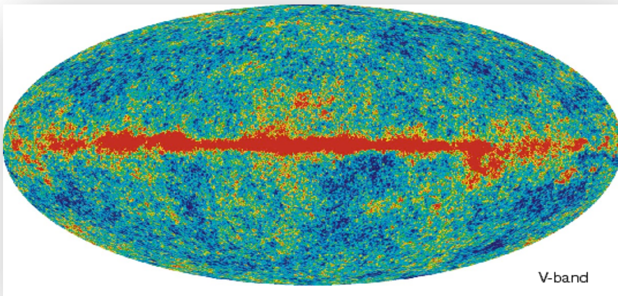
With noise properties we can predict the ***Sampling Distribution*** (the probability of obtaining a general set of data).

The ***Likelihood*** refers to the *specific* data we have) - as a function of θ It isn't a probability, strictly (not normalised)

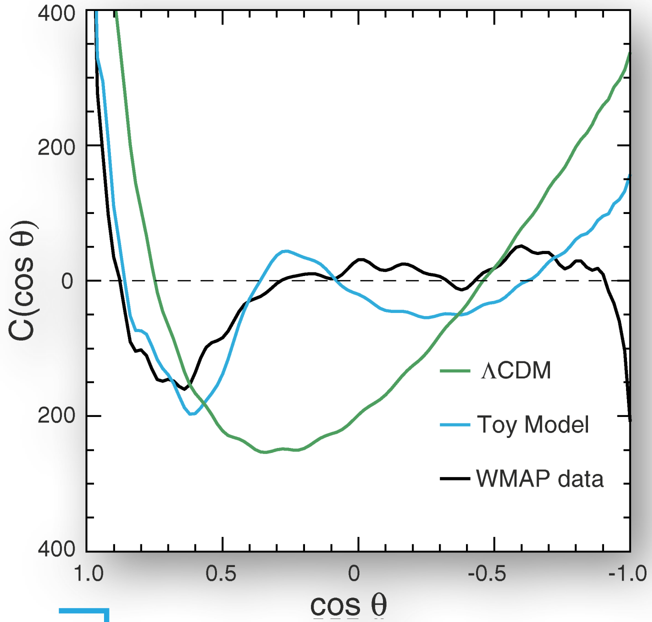
ICIC

Note: this is just the expectation value of x; the distribution is needed

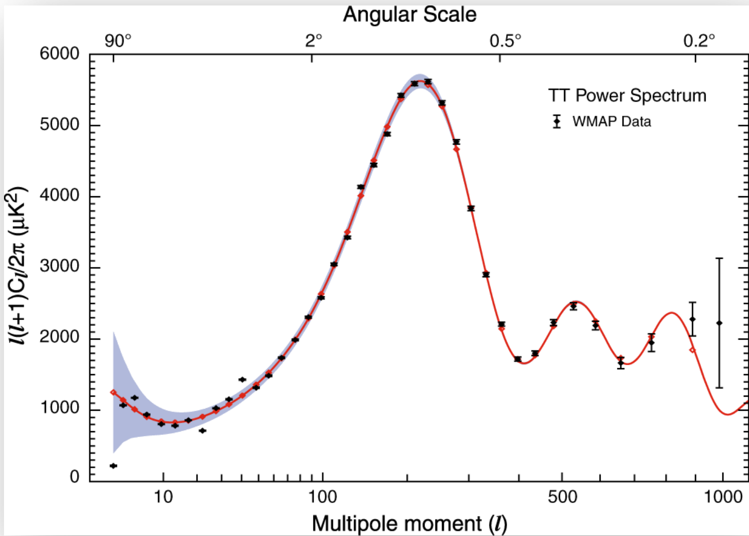
LCDM fits the WMAP data well.



Correlation Function $\langle \frac{\Delta T}{T}(\phi) \frac{\Delta T}{T}(\phi+\theta) \rangle$



Power Spectrum C_l



ICIC

State your priors

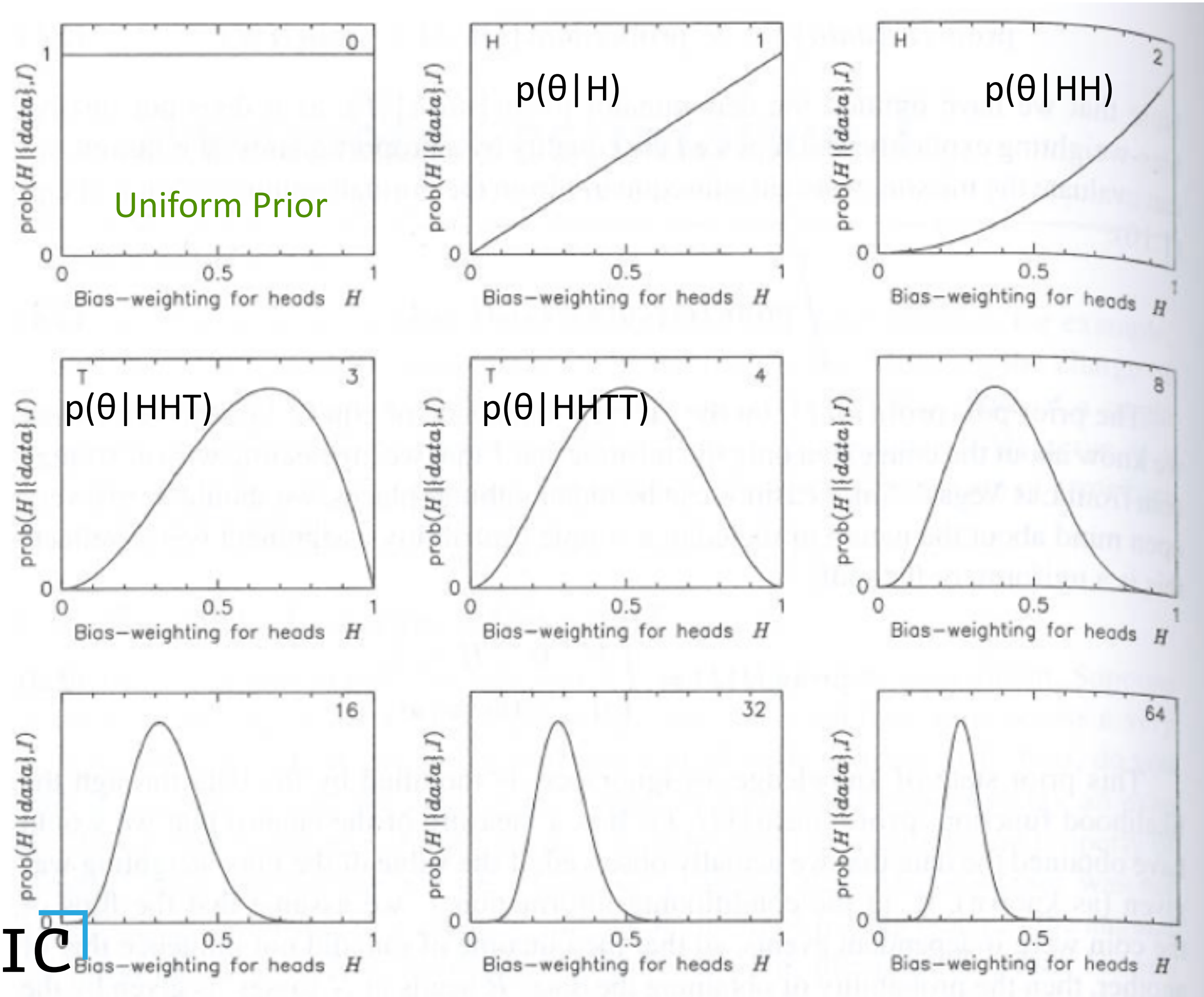
- In easy cases, the effect of the prior is simple
- As experiment gathers more data, the likelihood tends to get narrower, and the influence of the prior diminishes
- Rule of thumb: if changing your prior† to another reasonable one changes the answers a lot, you could do with more data
- Reasonable priors? Noninformative* – constant prior (can usually do this even if infinite interval - the normalisation is not important so may not need to be set).
- scale parameters in $[0, \infty)$; uniform in log of parameter (Jeffreys' prior)
- Bayesian reasoning is NOT subjective - posterior is determined unambiguously from the prior and likelihood

† I mean the raw theoretical one, not modified by an experiment

* Actually, it's better not to use these terms – other people use them to mean different things – just say what your prior is! Uniform priors can in fact be very informative.

ICIC

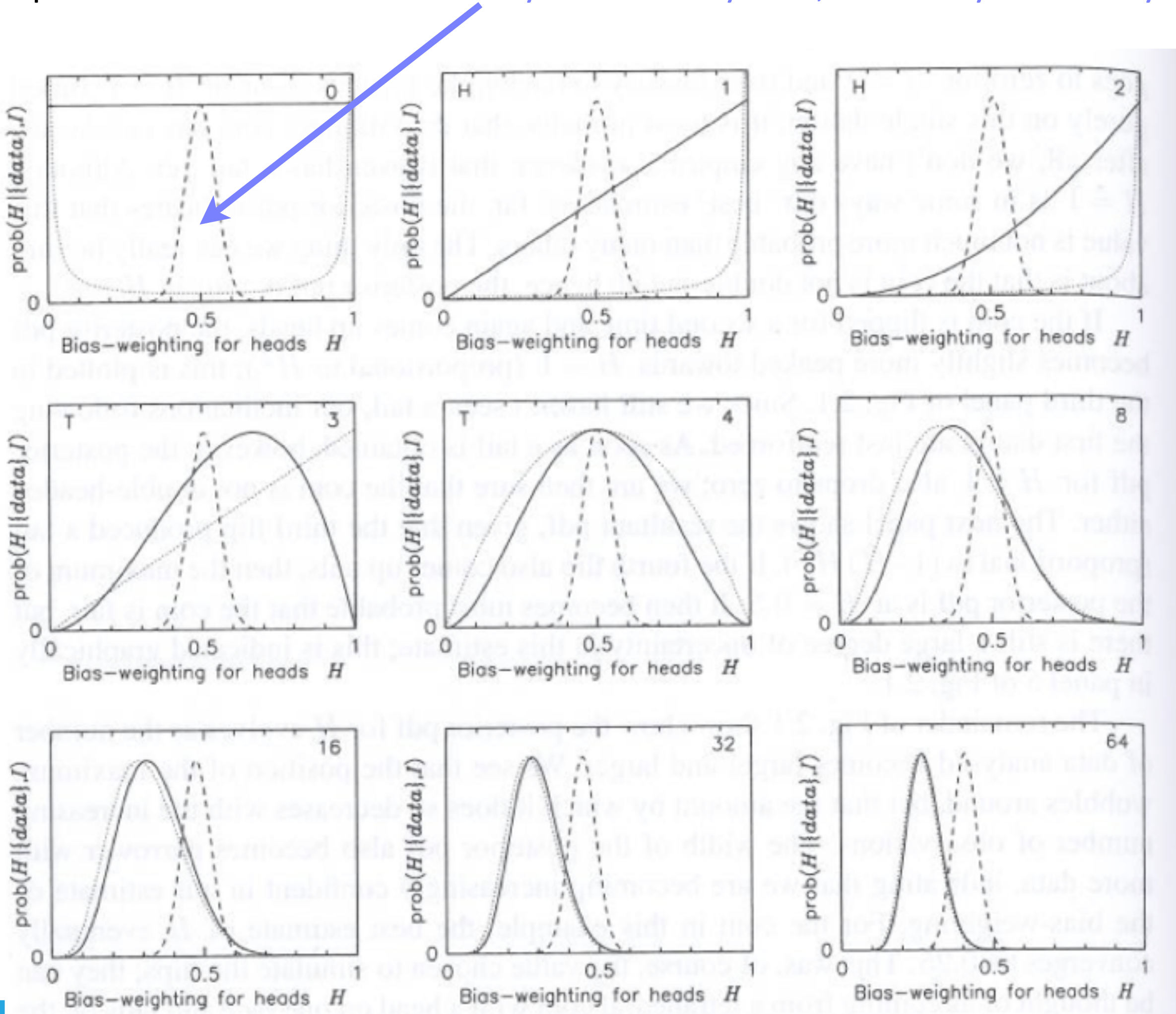
From Sivia & Skilling's *Data Analysis* book. IS THE COIN FAIR?
Model: independent throws of coin. Parameter θ = probability of H



ICIC

The effect of priors

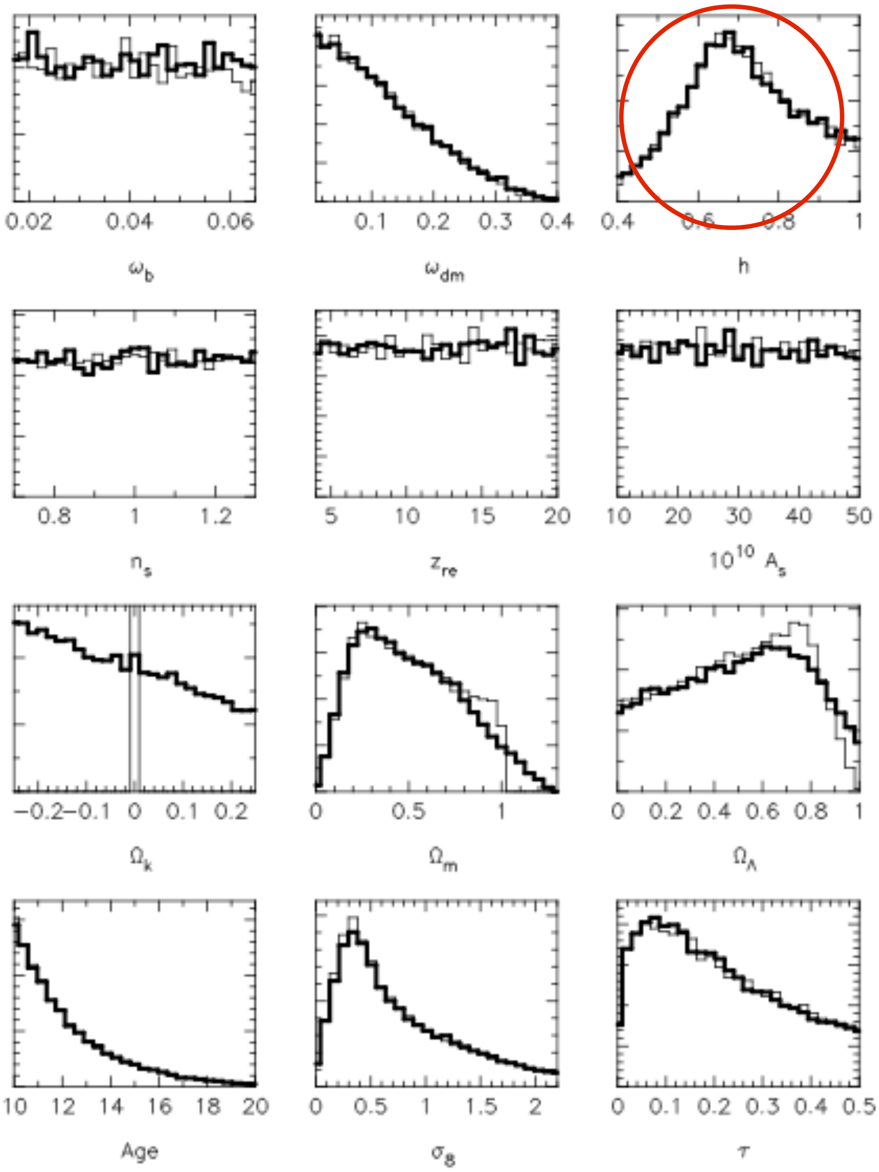
Priors = “It’s likely to be nearly fair”, “It’s likely to be very unfair”



Sivia & Skilling

- VSA CMB experiment

(Slosar et al 2003)



Priors: $\Omega_\Lambda \geq 0$

$10 \leq$ age ≤ 20 Gyr

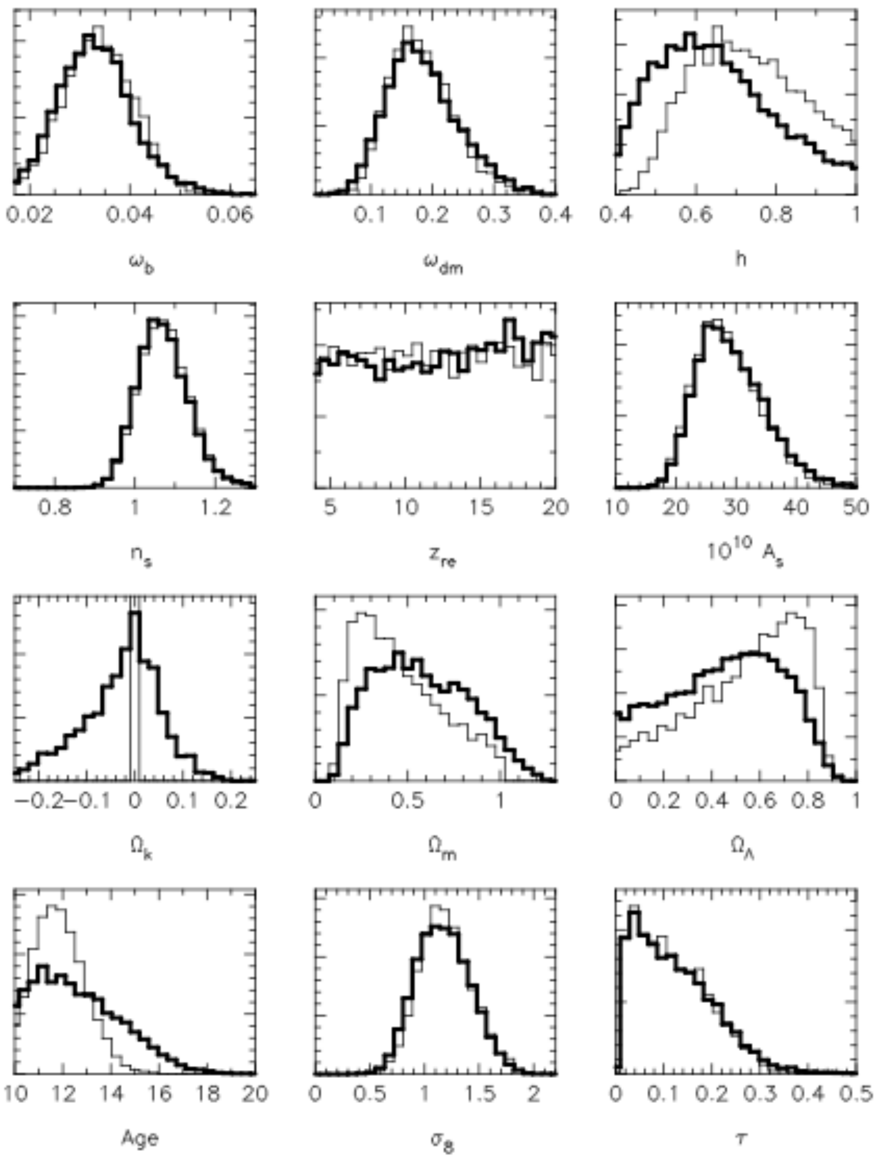
h ≈ 0.7 ± 0.1

There are no data in these plots – it is all coming from the prior!

$$p(\theta_1) = \int d\theta_{j\neq 1}\, p(x|\theta)\, p(\theta)$$

ICIC

VSA posterior



ICIC

Summary

- Write down what you want to know. For *parameter inference* it is typically:

$$p(\theta|x, I, M)$$

- What is *M* ?
- What is/are x?
- What is/are θ ?
- What is *I* ?
- You might want p(*M*|*x I*)…this is *Model Selection* - see later

ICIC

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

ICIC