Inference

Purpose of data analysis

A model is a physical/mathematical construct intended to represent some aspects of the real world. The predictions of the model normally depend on some unknown parameters. Most commonly, the aim of data analysis is to work out what values of those parameters are compatible with the data. More ambitiously, one may ask what choice of parameters is indicated by the data.

Useful models should

- Fit the present data acceptably.
- Have the ability to make predictions for future data.

Useful data should

- Be predictable by the models we aim to test.
- Have modelable intrinsic randomness and experimental error.

What is inference?

We are interested in a particular branch of inference (inductive inference) which is inference in the presence of uncertainty. This forces us to adopt notions of probability.

Inference: a conclusion reached on the basis of evidence and reasoning.
Oxford American Dictionary

Characteristically, any inductive conclusion will involve

1) Prior assumptions

2) Experimental/observational data

3) Inferential calculations

If any of these three items proves to be incorrect, the inference may fail to match reality.

Logic: The art of thinking and reasoning in strict accordance with the limitations and incapacities of the human misunderstanding.
Ambrose Bierce, The Devil's Dictionary

Probability

Typically, both theoretical predictions and observational outcomes may be subject to uncertainty, but in such a way that the uncertainty can be modelled in terms of a probability of different outcomes.

Simple example: toss of a coin

Here a theoretical prediction of the outcome can only be in terms of probability.
However the actual experimental outcome can presumably be precisely stated.

Another example: gene mutation rates in populations

Here there are several sources of possible error, eg an inevitable error from only being able to measure mutation in a finite number of organisms, possible errors in individual determinations, possible bias in sample selection, etc.

What is Bayesian inference?

Bayesian inference is a system of logical deduction which assigns probabilities to all quantities of interest. The probabilities are updated in light of new information according to a set of mathematical rules centred around Bayes' theorem (published in 1763).

Likelihood

Prior

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Posterior

θ = parameter value

D = data

In 1946, Cox showed that Bayesian inference is the unique consistent generalization of Boolean algebra.

What is Bayesian inference?

It can also be derived from a set of `requirements':

- Degrees of plausibility are real-valued.
- An increase in the plausibility of **A** results in a decrease in the plausibility of not-**A**.
- Plausibility depends only on the data and not on the order that the data are taken.
- The joint plausibility of **A** and **B** is related in a sensible way to the plausibility of **A** and **B**.

Terminology:

Parameters are quantities that can be varied in an attempt to best fit the data to hand.

A **model** is a choice of the ***set*** of parameters to be varied.

Posterior versus likelihood

Failure to distinguish posterior and likelihood, instead identifying them as equal, is one of the most common errors in inference.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

For example, P(pregnant|female) is about 0.03,
but P(female|pregnant) is 1.

Another example:

You've been to the doctor's with a broken wrist. She decides to give you a blood test. It comes back indicating you are HIV-positive. The doctor tells you the test is 95% reliable. What is your chance of having HIV?

Subsidiary question:

How would your answer change had you been to the doctor's because of swollen lymph glands?

Parameter estimation

If we have decided what parameters we want to constrain from the data, the task in front of us is parameter estimation.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

We need to decide our prior information P(θ), and we need to compute the likelihood $P(D|\theta)$. We can ignore $P(D)$ which just gives an irrelevant overall normalization.

The next lecture explains some ways to go about this.

Priors!!

Bayesian inference requires that the prior probabilities be specified, giving the state of knowledge *before* the data was acquired to test the hypothesis.

- Priors are to be chosen. Different people may not agree on their choice.
- Priors are where physical intuition comes in.
- In my view, one shouldn't seek a single `right' prior. Rather, one should test how robust the conclusions are under reasonable variation of the priors.
- Eventually, sufficiently good data will overturn incorrect choice of prior.
- If you don't know enough to set a prior, why did you bother getting the data?

Basic model structure

Model inputs: Prior

Model structure | Known parameters | Unknown parameters

θ

Predictions: Predictions for observables

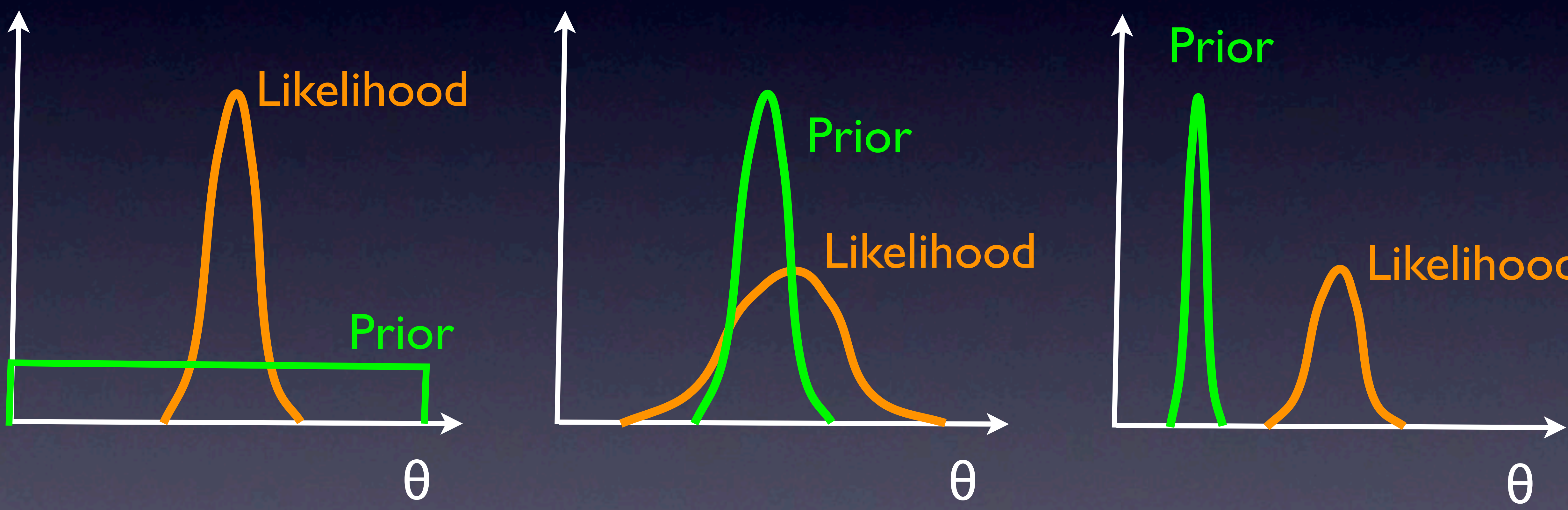
Likelihood: $P(D|\theta)$

Observations: Measurements of observables, with uncertainties D

Prior information versus information from data

Bayes theorem breaks up your final answer, the posterior, into the part from your prior assumptions and the part from the data. If you are wise, you will check how much comes from each one.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$



Good

Not good

Either terrible, or earthshaking new scientific discovery.

Levels of Bayesian inference

Parameter Estimation

I've decided what the correct model is.

Now I want to know what values of the parameters are consistent with the data.

I can do this using e.g. Markov Chain Monte Carlo.

Model Selection

Now I think about it, I don't actually know what the correct model is. It could be one of several.

Now I want to know what the best model is.

How am I going to do that?

The Bayesian evidence

Bayes theorem again, but conditioned on a model.

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \quad \Rightarrow \quad P(\theta|D,M) = \frac{P(D|\theta,M)P(\theta|M)}{P(D|M)}$$

Posterior model probability!

Bayesian evidence

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

How do we calculate it? $P(D|M) = \int P(D|\theta,M)\, P(\theta|M)\, d\theta$

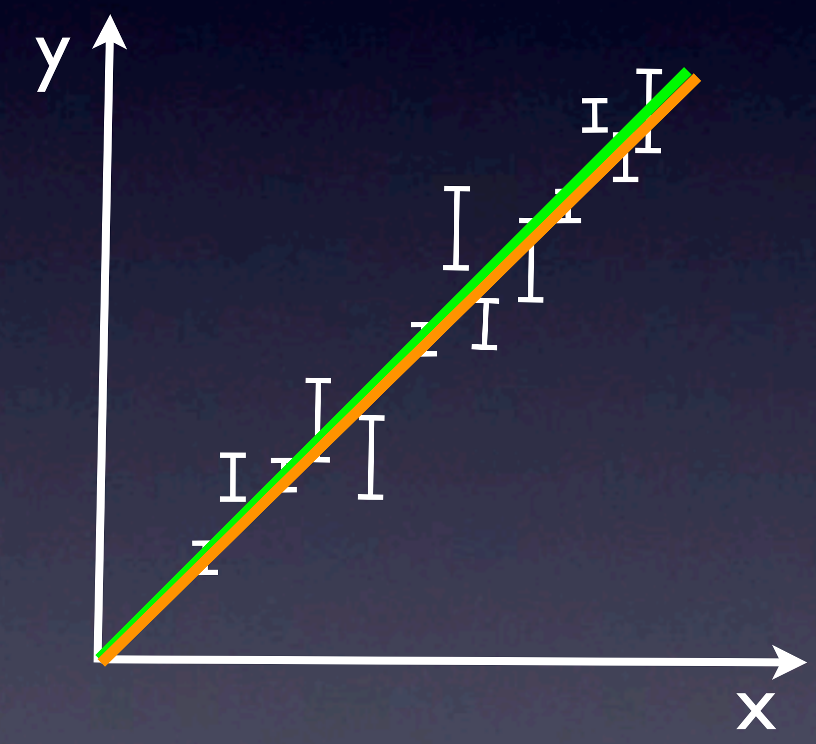
This can be evaluated in a number of ways: in my group we use a Monte Carlo integration method called nested sampling.

What does it reward? **Model predictiveness**

On predictiveness

Consider two different models, let's say a linear fit to a set of data. In one model, we consider the gradient of the line to be a free parameter, adjustable to give the best fit. In the other, we decide based on some prior intuition that the gradient should be one.

The fits turn out to be nearly identical. Which model do you think is better?

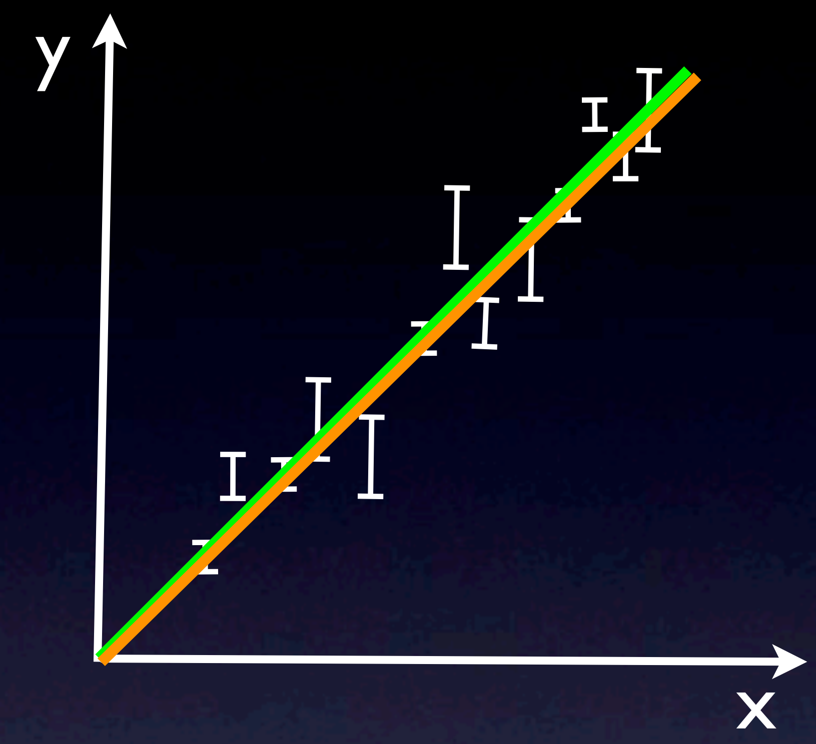


Most people would feel the model with fixed gradient is better, as it is simpler and fits the data well. The person who devised it seemed to have good intuition as to how the data would turn out. But of course it is the more general model which fits the data (marginally) better. We like the fixed gradient model because it is **predictive**.

Predictiveness and evidence

The Bayesian evidence is the average likelihood over the prior space. It answers the question "Over the parameter values which I thought were reasonable before the data came along, what was the average likelihood?"

$$E(M) = \int L(\theta)\, Pr(\theta)\, d\theta$$



Three possibilities:

- Model doesn't fit the data for any parameter value, hence low evidence.
- Model fits the data well for some parameter choice.
  - Model is predictive; then the likelihood will remain high across much of its parameter space.
  - Model is not predictive; then the evidence will be pulled down by regions where the predictions are different and hence the fit is poor.

Interpretational scale

Computing the evidence is often challenging, but feasible due to recent algorithm developments. For guidance in interpretting the evidence, people usually appeal to the Jeffreys' scale.

Jeffreys' Scale:

$\Delta \ln E < 1$	Not worth more than a bare mention
$1 < \Delta \ln E < 2.5$	Substantial evidence
$2.5 < \Delta \ln E < 5$	Strong to very strong evidence
$5 < \Delta \ln E$	Decisive evidence

The most useful divisions are 2.5 (odds ratio of 12:1) and 5 (odds ratio of 150:1).

Applications of Bayesian model selection

- **Analysis of data**

  What does existing data tell us about our various models?

- **Model selection forecasting**

  How well will a future experiment answer questions of model selection type?

- **Bayesian survey design**

  How can I maximize the chances of my experiment generating a successful model selection outcome?

Levels of Bayesian inference

Parameter Estimation

I've decided what the correct model is.

Now I want to know what values of the parameters are consistent with the data.

I can do this using e.g. Markov Chain Monte Carlo.

Model Selection

Now I think about it, I don't actually know what the correct model is. It could be one of several.

Now I want to know what the best model is.

I can do this by computing the Bayesian Evidence. I can then do parameter estimation using the best model.

Multi-model Inference

Mmm, I did the model selection thing, but there wasn't a single best model.

But I still want to know how probable the parameter values are.

I can do this by combining the parameter likelihoods using Bayesian Model Averaging, adding them together weighted by the model probabilities.

Multi-model inference: philosophy

- The top level of inference is deciding the models to be compared with data, i.e. what different choices of parameter sets are we interested in.
- There is a conceptual difference between a model where a quantity is fixed at some value, versus a model where that quantity can vary and its value just happens to have that particular value. Although for that parameter value the predictions are identical, overall the model predictiveness is different. Model selection compares models, not specific parameter values.
- Computing the evidence tells us how the probability of each model has been modified by the data.
- If only one model survives, proceed to standard parameter estimation.
- If several models survive, use multi-model inference, e.g. Bayesian model averaging.

Something like 95% of all 95% confidence "detections" turn out to be wrong. Why?

- **Statistical fluke:** By definition important only if people do their error analysis wrongly (actual errors, incorrectly modelled systematics, error distributions not Gaussian, etc).

- **Publication bias:** Only positive results get published, enhancing their apparent statistical significance (recognised as a major problem in clinical trials).

  Ioannidis (2005) actually claims a proof that most published results are false.

- **Inappropriate "a posteriori" reasoning:** choosing "interesting" features from the data and assessing their significance via Monte Carlo analyses.

- **Neglect of model dimensionality:** using parameter estimation rather than model selection.

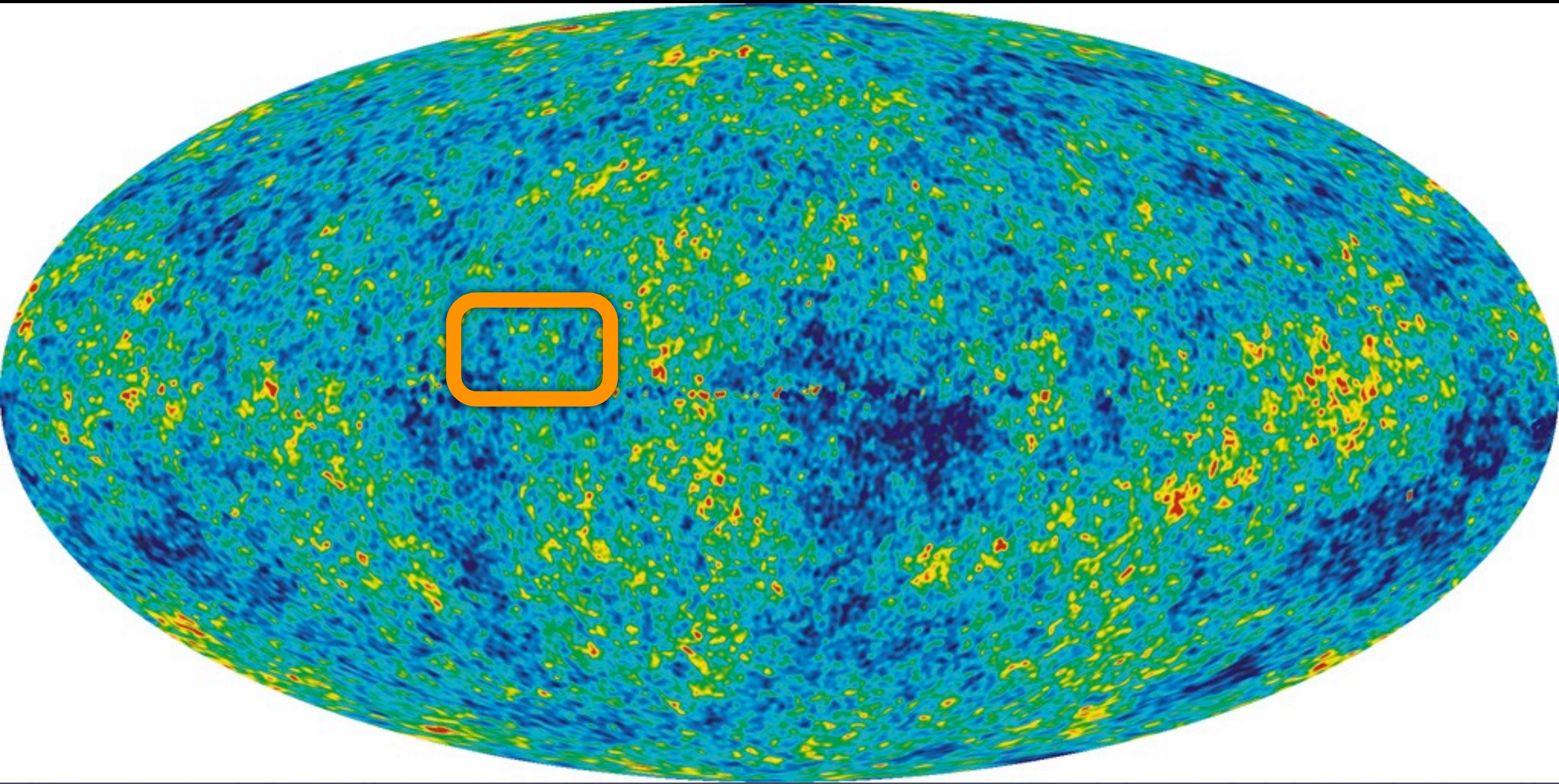
Some data

1	1	6	4	2	6	1	2	3	6
6	2	1	5	1	6	3	4	4	6
1	1	4	4	3	5	3	5	1	6
2	3	2	3	6	1	2	2	4	2
4	5	5	3	2	2	4	6	6	6

Eleven 6s and only five 5s. How unlikely is that?

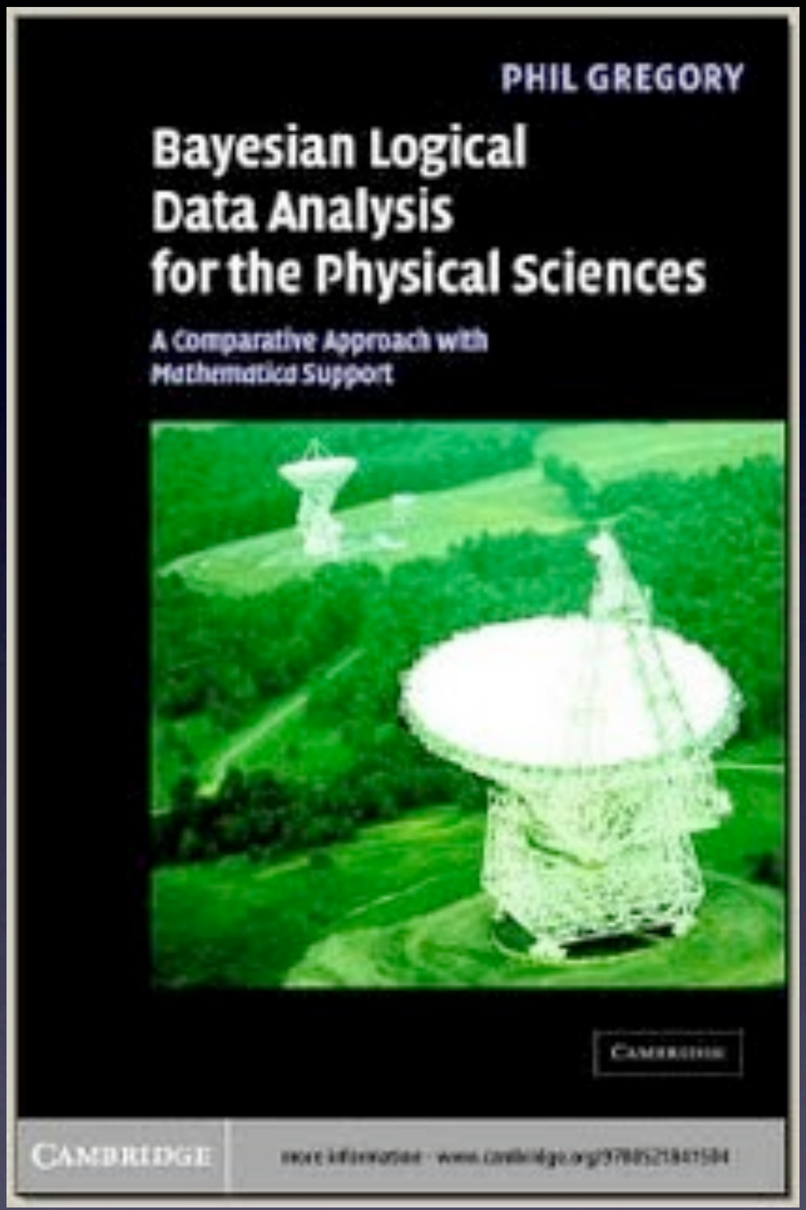
A Monte Carlo of 10
indicates that we get
only 4% of the time.
aren't random?

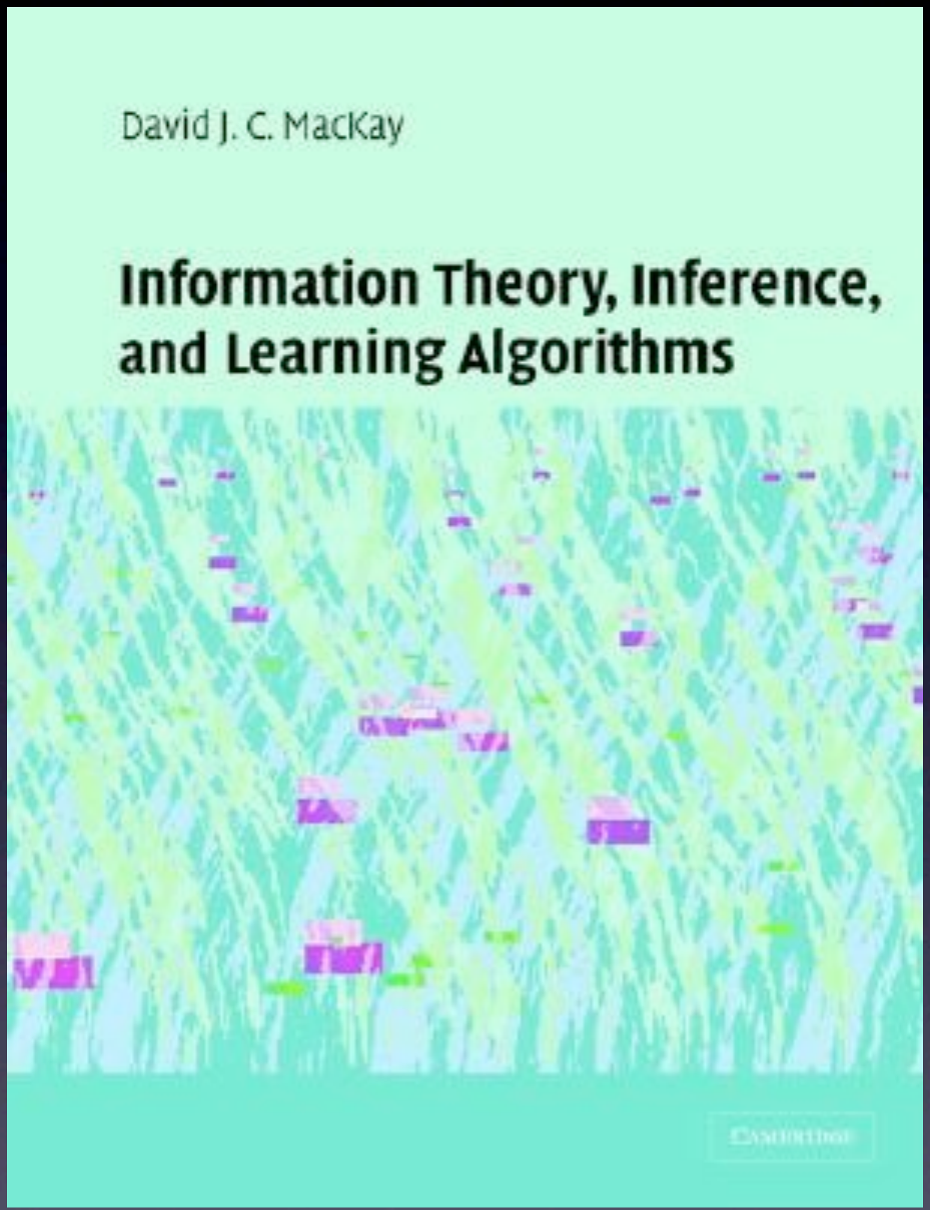
```
      count = 0
      do 50 k=1,100000
         fives=0
         sixes=0
         do 60 j=1,50
            throw = 1+int(ran1(iseed)*6.)
            if (throw.eq.5) fives=fives+1
            if (throw.eq.6) sixes=sixes+1
60          continue
         if ((fives.le.5).and.(sixes.ge.11)) count=count+1
50       continue
      write(6,*) count
```

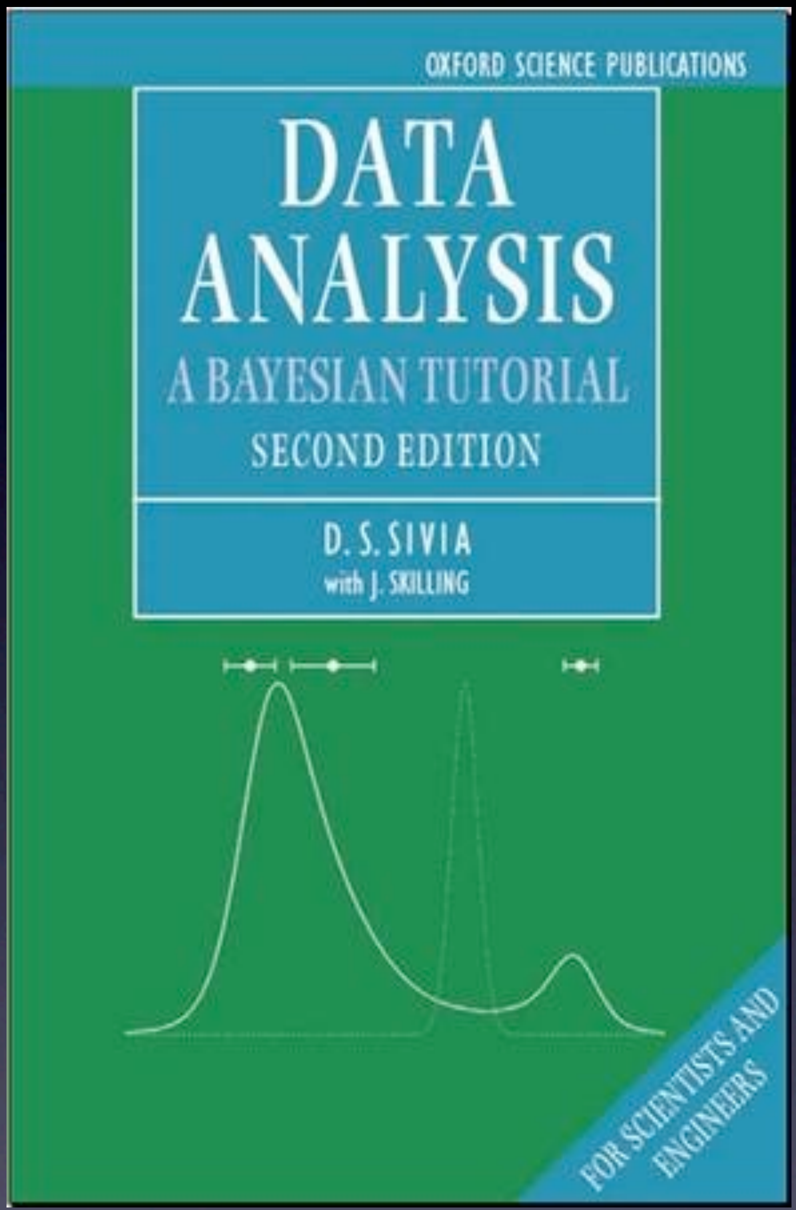


A more serious example is the various alignment anomalies in the large-angle CMB, whose significance is very difficult to assess.

Textbooks


PHIL GREGORY
Bayesian Logical Data Analysis for the Physical Sciences
A Comparative Approach with Mathematica Support
CAMBRIDGE


David J. C. MacKay
Information Theory, Inference, and Learning Algorithms


OXFORD SCIENCE PUBLICATIONS
DATA ANALYSIS
A BAYESIAN TUTORIAL
SECOND EDITION
D. S. SIVIA
with J. SKILLING
FOR SCIENTISTS AND ENGINEERS