STATS 331

Introduction to Bayesian Statistics
Semester 2, 2016

Lecture 3

A Brief History
and
The Basic Ideas

A Brief History

- Notions of chance and uncertainty have been around for ages
- (Mathemetical) probability theory was developed beginning in the 17th century
- Applications were mostly to games of chance – aka gambling
- Some important results were established. e.g. law of large numbers (Jacob Bernoulli)

From Gambling to Science

- Many people started to realise that probability theory would be useful in scientific applications
- In 1763, Thomas Bayes' famous paper “*An Essay towards solving a Problem in the Doctrine of Chances*” was published (after his death)
- In it, he solved a specific problem – one that we will look at in the next lecture!

Laplace

- In the 19th century, Pierre-Simon Laplace developed Bayes' ideas into a general theory
- He used them in many scientific applications
- e.g. the mass of Saturn: "...it is a bet of 11,000 to 1 that the error in this result is not 1/100th of its value"
- “probability theory is common sense reduced to calculation”
- Most of what Laplace did would today be called Bayesian Statistics

Laplace's Rule of Succession

- Laplace worked out a result called the *rule of succession*
- After observing r successes out of n trials, what is the probability that the next trial will be a success?
- Answer: prob = $(r + 1)/(n + 2)$
- Famously, Laplace worked out **the probability that the sun will rise tomorrow (it was pretty close to 1)**

Enter the Critics

- In the late 19th century many people (e.g. von Mises) criticised Laplace's work, particularly the rule of succession
- Many thought it didn't make sense to say a **non-repeatable event that is either true or false** could have a probability

Ronald Aylmer Fisher

- R. A. Fisher was a **major** player in the history of statistics
- He argued vigorously against Laplace's “inverse probability” and developed his own methods
- A lot of the stats you've learned was developed by Fisher!
- e.g. p-values, randomisation, maximum likelihood, …
- **He also worked out how evolution by natural selection works together with Mendelian Genetics!**

Show of Hands

- Have you heard of or used *maximum likelihood* before?

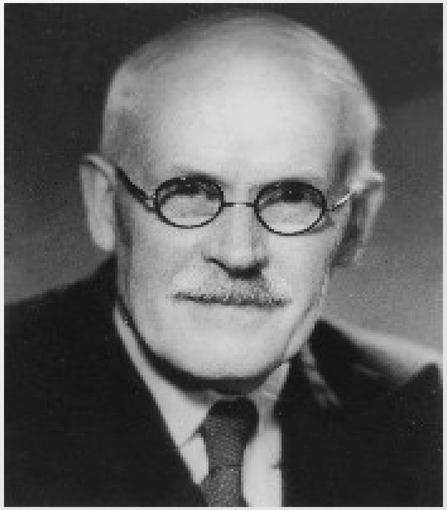
More Frequentists

- Type I and II errors, confidence intervals

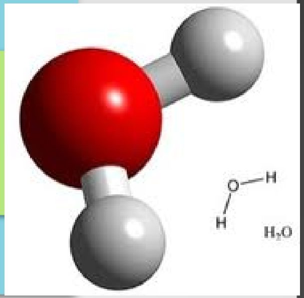
- Where did the Bayesians go?

Harold Jeffreys

- Jeffreys was a British geophysicist who worked on many topics
- In 1939 he published a classic book: “Theory of Probability”.

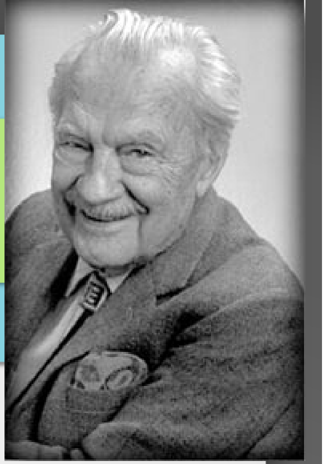


Meanwhile, over in physics...

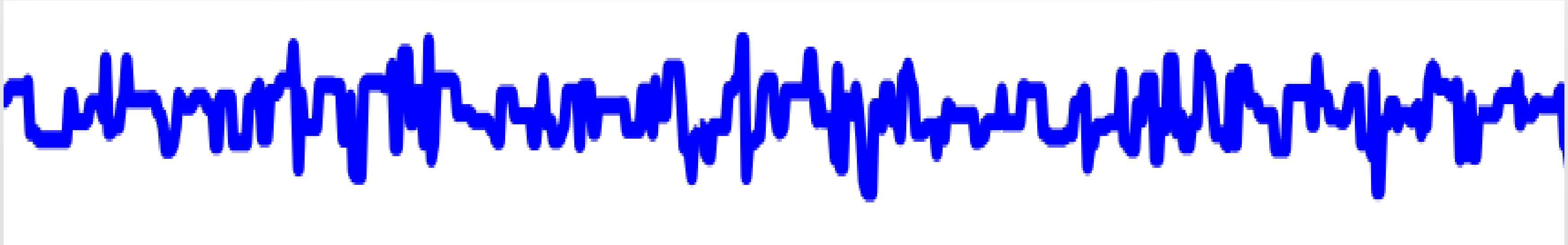


- With computers on the rise, physicists were thinking about algorithms they could use in the area of **statistical mechanics**
- Stat mech is basically how they can work out the *macroscopic* properties of a substance (e.g. water) based on its *microscopic* details
- e.g. by knowing that water is a bunch of H_2O molecules, statmech can tell you that it freezes at 0˚C, boils at 100˚C, etc.

Metropolis



- Nicholas Metropolis, and some other physicists, worked out a great algorithm for statistical mechanics
- It became known as the **Metropolis algorithm**
- Physicists used it a lot, but the statisticians didn't really notice



The Underground Bayesians

- Bayesian stats was on the fringes during most of the 20th century
- There were a few passionate advocates
- Bruno de Finetti, Jack Good, Ed Jaynes, Jimmie Savage, Dennis Lindley
- But it was not part of the mainstream

The Bayesians' Points

- Bayesian stats is more like “common sense”
- It satisfies a bunch of principles that seem like they should be true: e.g. likelihood principle
- It directly answers the question we (usually) really want answered:

Given the information that we have, how plausible is this hypothesis, as opposed to that one?

The Frequentists' Comebacks

- The Bayesian calculations often require more assumptions in order to get going
- These assumptions can be subjective – **different people might get different results**
- Even if the Bayesian theory were more attractive philosophically, in practice some problems couldn't be solved, usually because certain integrals couldn't be done

MCMC and the Bayesian Revival

- In 1990, there was a famous paper by Gelfand and Smith, showing how the old computer method by **Metropolis** (and other similar methods) could be used to solve Bayesian problems
- This was Markov Chain Monte Carlo (MCMC)
- Since this time, there has been a **massive explosion** in the use of Bayesian statistics

The Basic Bayesian Ideas

1. Plausibility is Confidence is Probability

- Say there is some statement/proposition and we are not sure whether it is true or not.

e.g. A = **New Zealand win more bronze medals than Australia at the Rio 2016 Olympics**.

- We can describe the *degree of plausibility* for this by a probability value between 0 and 1.

P(A) = 1	--> Certainty true
P(A) = 0	--> Certainly false
P(A) = 0.95	--> Probably true
P(A) = 0.5	--> As certain as a coin flip

Bayesian Probability

- Plausibility
- Degree of confidence
- Degree of belief
- The degree to which one statement implies another

- etc...

Probability is subjective

- Consider the hypothesis that **the population of New Zealand is greater than 5 million**
- For me, the probability is 0 (or very close to 0)
- For someone who doesn't know much about NZ, the probability might be 0.5
- **Different people may have different probabilities for the same proposition**
- This is okay. It means they have *different information* and/or *different assumptions*.

2. Probabilities Get Updated

- Probabilities are different if you have different information
- The key of Bayesian stats is that **when you get new information (e.g. some data), your probabilities should get updated**
- If your initial (“prior”) probability is P(A), then your probability after taking into account information A is P(H|A)
- Got more information B? Calculate P(H|B, A)
- Got more information C? Calculate P(H|C, B, A)

Bayes' Rule

- How exactly do the probabilities get updated?
- Bayes' Rule (follows from product rule) tells us how:

Original probability
"**prior**"

Prob. that we would have got this
data if *H* were true – "**likelihood**"

$$P(H|D) = \frac{P(H) \times P(D|H)}{P(D)}$$

Updated probability
"**posterior**"

Prob. that we would have got this
data whether *H* were true or not
"marginal likelihood"

Interpretation of Bayes' Rule

- The posterior probability depends on:
- **The prior probability**. If the hypothesis was plausible before the data, then that helps it to be plausible after the data!
- **The likelihood**. If the hypothesis assigned a high probability to the data (i.e. it *predicted the data*) then it is rewarded!
- **The P(D) in the denominator (marginal likelihood)***.* If the data would have happened anyway *even if H were false*, then a high likelihood doesn't mean much!

More Than One Hypothesis

- We will now solve our first Bayesian problem
- There will be **two** possible mutually exclusive hypotheses

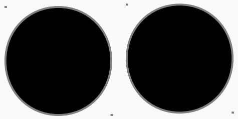
- This is always the case – **when there's just one H, there's always the negation ¬H**.

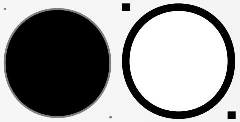
A First Bayesian Problem

Two Balls in a Bag

The Question

- There are two balls in a bag
- We don't know which of the following is true

Both balls are black 

One ball is black and the other is white 

- Call these hypotheses/statements BB and BW
- Note BW =

Use Your Intuition

- One of the two balls is picked from the bag “at random”. It is **black**. What happens to the hypothesis that *both balls are black (aka BB)*?

a) It gets more plausible

b) It gets less plausible

c) It stays the same

Now Let's Calculate It

- There are two possible hypotheses, and we don't know which is true
- We describe our uncertainty using **probabilities**
- Let's start by assigning *50/50* probabilities

- This is a good description of “prior ignorance”

A Black Ball is Drawn

- One of the two balls is picked from the bag at random. It is **black**.
- To apply Bayes' theorem, we need the likelihoods

P(black ball drawn | BB) = ?

P(black ball drawn | BW) = ?

And the marginal likelihood P(black ball drawn)

The Bayes Box

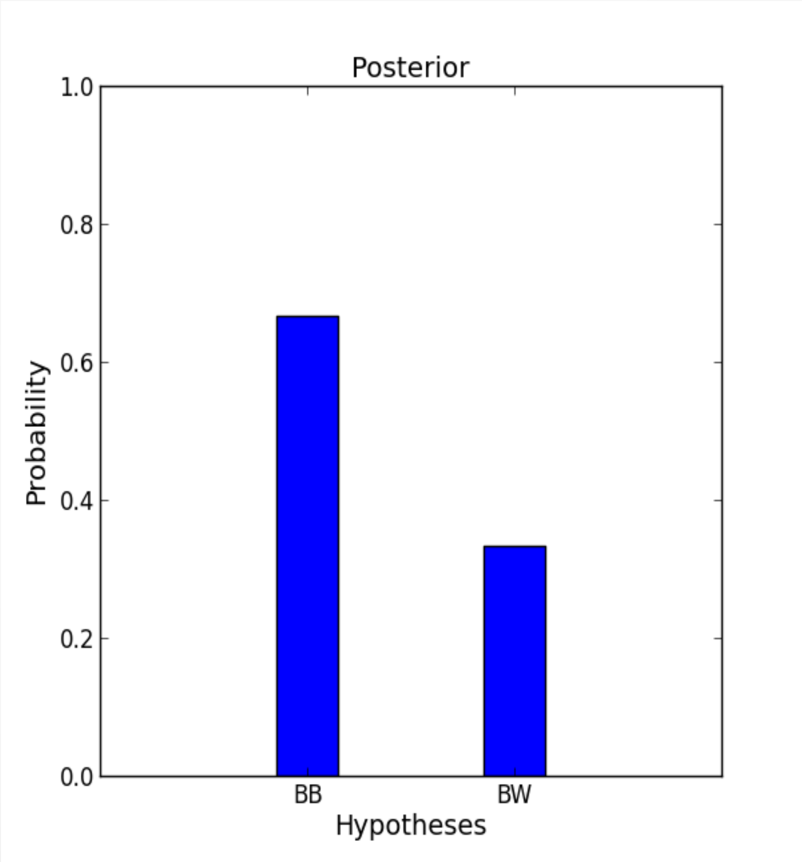
- The Bayes Box is a neat way of doing the calculation using a table. We will be using them a lot!

Possible Answers	**Prior**	**Likelihood**	**Prior x Likelihood**	**Posterior**
BB	0.5	1.0	0.5	0.666667
BW	0.5	0.5	0.25	0.333333
Totals	*1*		*0.75*	*1*

The calculation in R – try it out!

```
prior = c(0.5, 0.5)
lik = c(1, 0.5)
h = prior*lik
Z = sum(h)
posterior = h/Z
print(posterior)
```

Posterior



More Data

- So we got a black ball on the first draw. What if we then replaced the black ball, shook the bag, and repeated the experiment a further 4 times (making a total of 5 times)

- **The ball drawn was black every time**
- Seems like compelling evidence in favour of BB and against BW!
- What are the probabilities of BB and BW now?

The Priors

- Let's start with 0.5 and 0.5 prior probabilities again

Possible Answers	Prior	Likelihood	Prior x Likelihood	Posterior
BB	0.5			
BW	0.5			
Totals	*1*			

The Likelihoods

- If BB is true, what was the probability of getting black every single time (5/5)?

$$P(\text{Data} \mid BB) = 1$$

- If BW is true, what was the probability of getting black every single time (5/5)?

$$P(\text{Data} \mid BW) = \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = 1/32$$

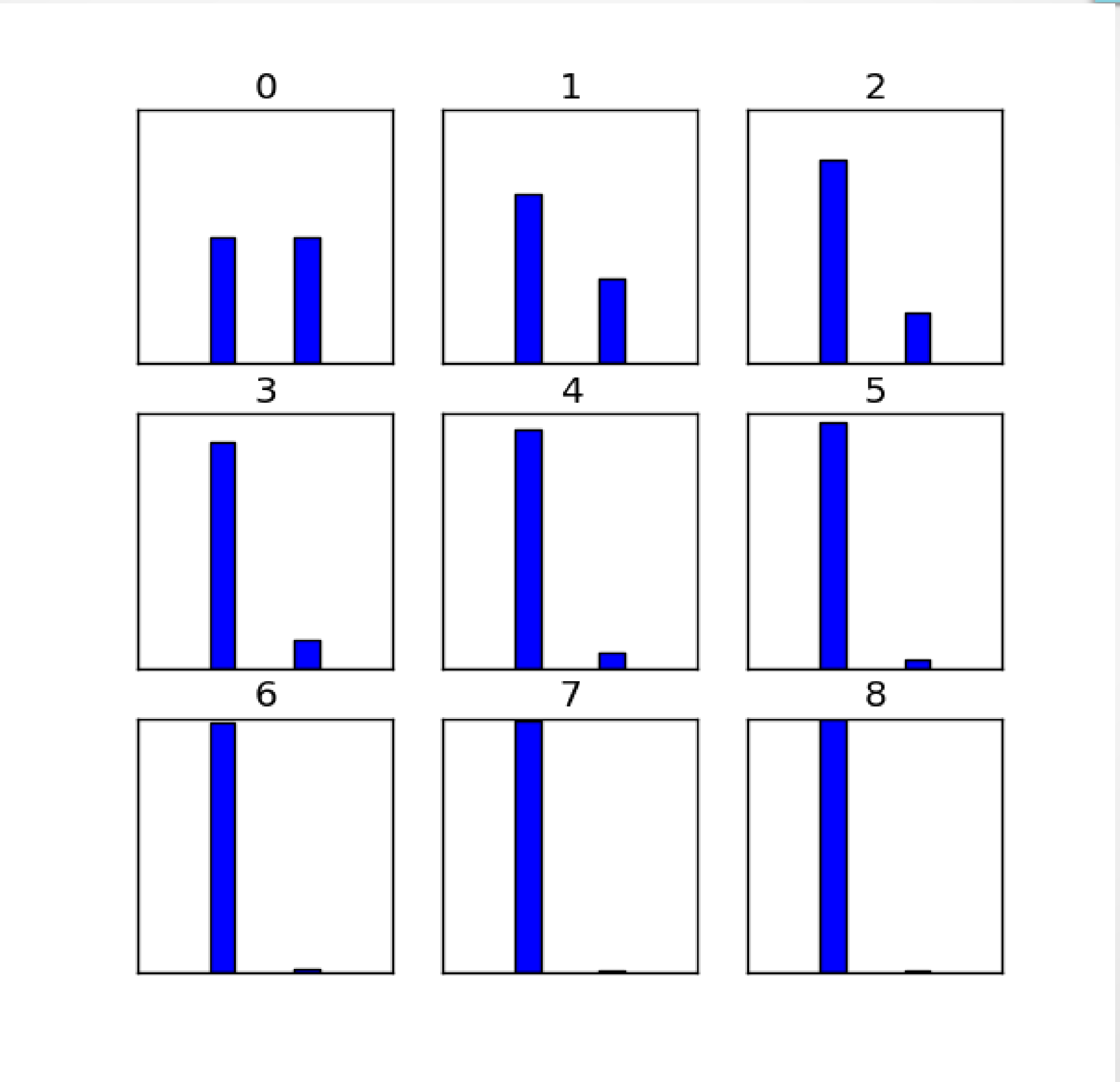
Filling in the Bayes Box

Possible Answers	Prior	Likelihood	Prior x Likelihood	Posterior
BB	0.5	1.0	0.5	0.9696...
BW	0.5	0.03125	0.015625	0.0303...
Totals	*1*		*0.515625*	*1*

Sequential Updating

- To solve the inference with the “5 black draws” example, we went back to the beginning, back to a 50/50 prior and using the likelihood for 5 black draws
- We could have started with the 67/33 prior and then used the likelihood for 4 extra black draws
- The answers would have been the same
- **The posterior after analysing one data set becomes the prior for analysing the next**

Changing Probabilities



Different Priors?

- What if we didn't start with a 50/50 prior?
- Suppose we were initially *sceptical* of BB and had a prior probability of 0.01
- Therefore a prior probability of 0.99 for BW

Bayes Box, Sceptical Prior

Possible Answers	Prior	Likelihood	Prior x Likelihood	Posterior
BB	0.01	1.0	0.01	0.2443
BW	0.99	0.03125	0.0309375	0.7557
Totals	*1*		*0.0409375*	*1*

Different Priors?

- If we start with a low prior probability for one of the hypotheses, then it takes **more data** to convince us that it is true

"Extraordinary claims require extraordinary evidence"

Carl Sagan. Image credit: NASA/JPL

Drawing White

- If we ever drew a white ball from the bag, we would immediately become 100% certain of BW.
- What in Bayes' rule would make this happen?

An Extreme Prior

- What if we start with an extreme prior that says

P(BB) = 0

P(BW) = 1

- What will be the posterior probability of BB after drawing the black ball **1 million times**?

Cromwell's Rule

“*I beseech you, in the bowels of Christ, think it possible that you may be mistaken.*” - Oliver Cromwell

"*leave a little probability for the moon being made of green cheese; it can be as small as 1 in a million, but have it there since otherwise an army of astronauts returning with samples of the said cheese will leave you unmoved.*"

- Dennis Lindley

This Lecture Was Important

- We looked at Bayes' rule and the notion that probabilities can measure degrees of certainty
- We saw how Bayes' rule can be used to update probabilities when new information is available
- Bayesian updating was illustrated with a simple example with two possible hypotheses.

Next Lecture

- We will look at *parameter estimation* from a Bayesian point of view
- The “Bayes Box” will be extended
- We will solve the famous problem of estimating a proportion
- We will see if the sun really does rise tomorrow.