Probability Theory: The Logic Of Science

Based on the book of *E.T. Jaynes* by the same name, and lecture notes from *R. Prix*

Logical statements

First let us define a logical statements by some examples

- A: 5 O'Clock is pub time
- B: This boring talk started at 4 O'Clock
- C: This is the first slide

We seamlessly combine this information in our heads to deduce that

$$A \text{ and } B \text{ and } C \Rightarrow D$$

where

- D: It is one hour until pub-time

Note that a logical statement can be thought of as *either* data or a statement

How do we do science?

Deductive reasoning

Many people believe that we do science in the follow way: first we make a statement such as

$$A \textit{ is true then } B \textit{ is true}$$

Then we say

$$A \textit{ is true} \Rightarrow B \textit{ is true}$$

This constitutes a *rigorous proof*, but it is not how we do science.

How do we do science?

Plausible Reasoning

Instead, we often take the statement

$$A \text{ is true then } B \text{ is true}$$

Then consider that

$$B \text{ is true, therefore } A \text{ is `more plausible'}$$

The evidence does not prove that A is true, but verifying one of A's consequences does give us more confidence in A.

How do we do science?

Example: GR and Gravitational Waves

Given the statement

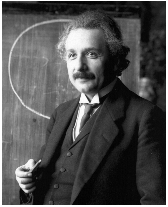
If GR is true, then GWs exist

We can't make deductive statements such as

GR is true so GWs exist

Instead we must do plausible reasoning

GWs exist, so GR is more plausible



How do we do science?

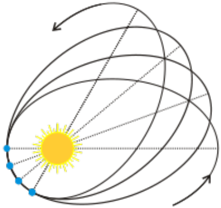
Example: Newton's theory and 'Falsification'

In science, the only time we really use deductive statements is in falsification. For example:

If Newton's theory is correct, then orbits are ellipses

This allows us to 'falsify' a statement: a deductive statement

Orbits are not ellipses (perihelion advance of Mercury) $\Rightarrow$ *Newton wrong*



Quantitative rules for plausible reasoning

Desiderata

- R1: We should represent out degree of plausibility by a real number:
  $P(A|B) =$ plausibility of A given B
- R2: Reasoning with 'common sense'
- R3: Consistency

From these desiderata, Cox 1946 derived *Three laws of Probability*

- P1:
$$P(A|I) \in [0, 1]$$
- P2: Sum rule
$$P(A|A) + P(\neg A|I) = 1$$
- P3: Product rule
$$P(AB|I) = P(A|I)P(B|AI) = P(B|I)P(A|BI)$$

Understanding probability

$P(A|B)$ quantifies out state of knowledge about the truth of the logical statement A given the assumptions/background B

- Not: related to random occurrences or repeated experiments
- Not: a property of the observed system, but a property of the observers state of knowledge - *Mind Projection Fallacy*

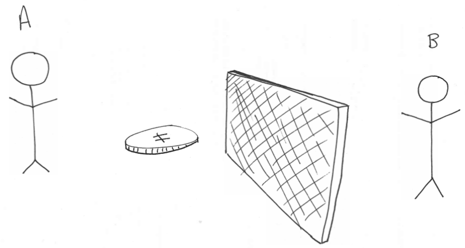
Example: Probability of a coin

- For person A:

$$P(H|I_A) = 1$$

- For person B:

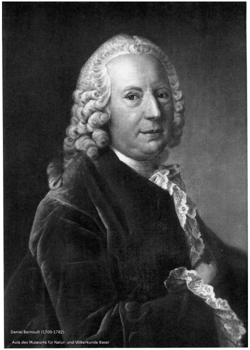
$$P(H|I_B) = ?$$



A simple example

Bernoulli's Urn:

Take an urn with T balls, we are told that N of them are red with the rest being white.



Bayes theorem

From $P3$ which was

$$P(AB|I) = P(A|I)P(B|AI) = P(B|I)P(A|BI)$$

we can derive *Bayes theorem*:

$$P(A|BI) = P(B|AI)\frac{P(A|I)}{P(B|I)}$$

This is a powerful tool in data analysis but very familiar since we use it all the time and often think of it as common sense.

- I: This is our previous information about the problem
- A: This is a logical statement we want to know more about
- B: This is a logical statement (data) that will hopefully inform us about A
- $P(A|BI)$: The *posterior* probability of A given B and I
- $P(B|AI)$: The likelihood of seeing B if A is true
- $\frac{P(A|I)}{P(B|I)}$: Our *prior* state of knowledge of the system

A simple application of Bayes theorem

Q: Given two coins, one is biased to always give heads and the other fair. We pick one up, toss it twice and get HH. What is the probability that we picked the biased coin?

Deriving an underlying probability distribution: More Coins

Imagine that we are asked to determine if a particular coin is fair. We could imagine that the chance to get H is some value p where a fair coin has $p = 0.5$.

Having flipped the coin n times and observed N_H heads, the likelihood can be modelled with the *Binomial sampling model*

$$P(N_H|p) = \mathrm{Bin}(N_H|, n, p) = \left(\begin{array}{c} n \\ N_H \end{array} \right) p^{N_H}(1-p)^{n-N_H}$$

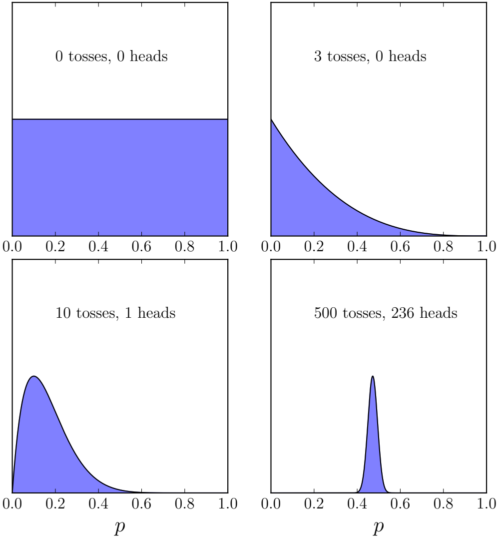
We now need a *prior* on p, here we have a choice: either use some past knowledge OR we have a non-informative prior such that

$$P(p) = \mathrm{Uniform}(0, 1)$$

Applying Bayes rule we have

$$\begin{aligned} P(p|N_H) &\propto P(N_H|p)P(p) \\ &\propto p^{N_H}(1-p)^{n-N_h} \end{aligned}$$

More Coins cont.



Hypothesis testing I

Here is an example of why we need a quantitative method to test hypotheses:

Cancer

Ann goes to the doctor and has a cancer test which comes back POS. We're given the following information:

- In the US 0.8% of people have this type of cancer
- When the patient has cancer the test returns a correct POS 98% of the time.
- When the patient doesn't have cancer the test returns a correct NEG 97% of the time.

What is the probability that Ann has cancer?

Hypothesis testing I

First lets turn it into probabilities

$$
\begin{aligned}
P(cancer) &= 0.008 & P(\neg cancer) &= 0.992 \\
P(POS|cancer) &= 0.98 & P(POS|\neg cancer) &= 0.03 \\
P(NEG|cancer) &= 0.97 & P(NEG|\neg cancer) &= 0.02
\end{aligned}
$$

Then we apply Bayes theorem

$$
\begin{aligned}
P(cancer|POS) &\propto P(POS|cancer)P(cancer) & &= 0.0078 \\
P(\neg cancer|NEG) &\propto P(NEG|\neg cancer)P(\neg cancer) & &= 0.0298
\end{aligned}
$$

So actually she *probably* doesn't have cancer

Hypothesis testing II

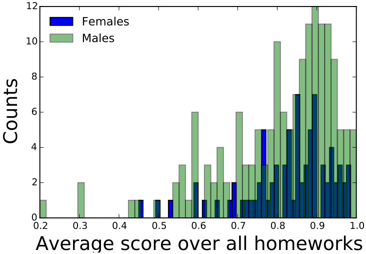
Lets test the hypothesis that

$$\mathcal{H} : \text{Females are better at maths than males}$$

Now we need some data which we denote y, lets use this years scores in the MATH1052 course. We could take a mean and standard deviation of each:

	Mean y	Standard dev y
Boys	0.809	0.146
Girls	0.829	0.114

This is a "point estimate": useful but it doesn't really give us a full picture.



Hypothesis testing II

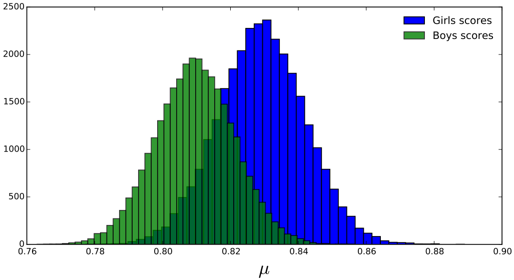
Instead, imagine that their scores both follow normal distributions

$$P(y|\mu_F, \tau_F) = N(\mu_F, \tau_F) \qquad P(y|\mu_M, \tau_M) = N(\mu_M, \tau_M)$$

Now we estimate the prob

$$P(\mu, \tau|y) \propto P(y|\mu\tau)P(\mu, \tau)$$

Use non-informative priors on μ and τ then consider the distribution of μ



This is the posterior distribution for μ for both sets of data. Calculated by Markov-Chain Monte-Carlo method which approximates the PDF $P(\mu|y)$ using samples.

Hypothesis testing II

Now we can test the hypothesis that "Females do better than males" from the means

$$P(\mu_F > \mu_M) = \frac{\text{Number of samples where girls do better}}{\text{Number of samples}}$$

From which we find

$$P(\mu_F > \mu_M) \approx 0.87$$

So we can say that we are 87% confident that girls do do better than boys!

Naive Bayes

Given some conditions $\{X_1, X_2, \dots X_N\}$ if they are independent then we can say

$$P(C|X_1, X_2, \dots, X_N) \ \propto P(X_1|C)P(X_2|C)\dots P(X_n|C)$$

M/F	Hair Colour	Drink	Person	Group
M	Dark	Coffee	Stu	GR
M	Light	Tea	Will	Strings
F	Dark	Tea	Ariana	Strings
M	Light	Coffee	Derek	Applied
M	Dark	Tea	Yafet	GR
F	Dark	Tea	Marta	GR
F	Light	Tea	Vanessa	GR
M	Light	Tea	John	GR
F	Light	Coffee	Alice	GR
M	Dark	Coffee	Kostas	GR
M	Dark	Coffee	Paco	GR
M	Dark	Tea	Johnathan	Applied

So if someone walks in we can try to classify them

$$P(GR|M, Dark, Tea) \propto P(M, D, T|GR) \propto P(M|GR)P(D|GR)P(T|GR) = \frac{5}{8}\frac{5}{8}\frac{4}{8}$$

Naive Bayes

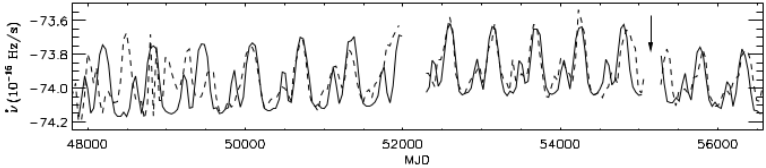
In the end we calculate the following

$$
\begin{aligned}
P(GR|M, D, T) &\propto 0.195 \\
P(Applied|M, D, T) &\propto 0.25 \\
P(Strings|M, D, T) &\propto 0.25
\end{aligned}
$$

Maybe this isn't such a great way to classify people...

Motivation to study this

- At a deep level the public don't understand how we do science. For example we can't 'prove' climate change.
- Future job prospects? Most of the so called 'big-data' is really just applied *Naive Bayes*.
- We need a framework to quantify our ideas
  For example: *Perrera et al 2014* claim that they have modelled some double-spin down variability in a pulsar



what does this even mean?

Conclusions

- Science is done by deductive reasoning and falsification
- Deductive reasoning can be formalised by Bayes theorem
- Bayesian data analysis is very hard
- Nevertheless, we should do better to avoid hand-waving and eyeballing
- $P(\text{pub time}|\text{given you are not Yafet or Will}) = 0.9999999)$

More information

E.T. Jaynes Probability theory: the logic of science