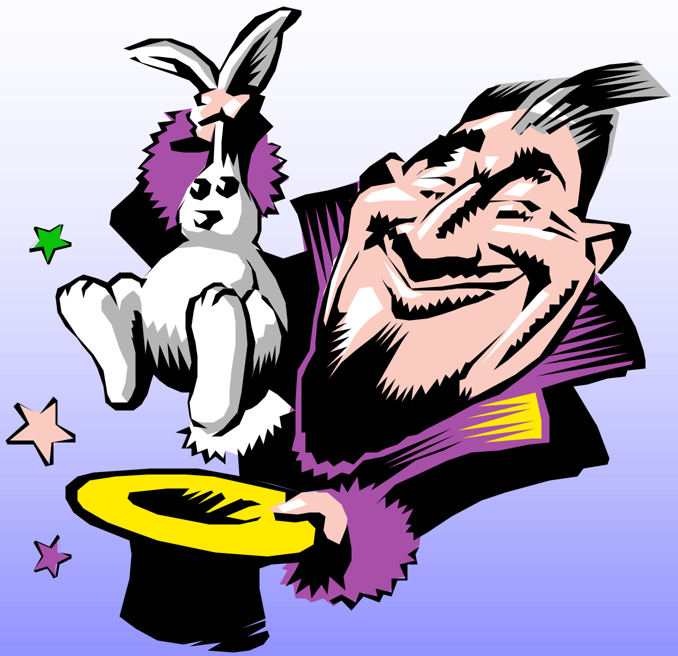
Statistics for HEP

Roger Barlow
Manchester University

Lecture 1: Probability

Definition 1: Mathematical



P(A) is a number obeying the Kolmogorov axioms

$$P(A) \geq 0$$

$$P(A_1 \vee A_2) = P(A_1) + P(A_2)$$

$$\sum P(A_i) = 1$$

Problem with Mathematical definition

No information is conveyed by P(A)

Definition 2: Classical

The probability P(A) is a property of an object that determines how often event A happens.

It is given by symmetry for equally-likely outcomes

Outcomes not equally-likely are reduced to equally-likely ones

Examples:

Tossing a coin:

P(H)=1/2

Throwing two dice

P(8)=5/36

Problems with the classical definition...

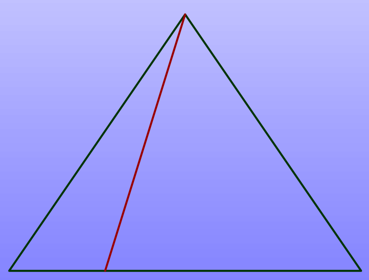
1. When are cases 'equally likely'?

- If you toss two coins, are there 3 possible outcomes or 4?

Can be handled

2. How do you handle continuous variables?

- Split the triangle at random:



Cannot be handled

Bertrand's Paradox

A jug contains 1 glassful of water and between 1 and 2 glasses of wine

Q: What is the most probable wine:water ratio?

A: Between 1 and 2 $\rightarrow$3/2

Q: What is the most probable water:wine ratio?

A: Between 1/1 and 1/2 $\rightarrow$3/4

$$(3/2)^{1}(3/4)^{-1}$$

Definition 3: Frequentist

- The probability P(A) is the limit (taken over some ensemble)

$$P(A) \underset{N\to\infty}{=} N(A) \big/ N$$

Problem (limitation) for the Frequentist definition

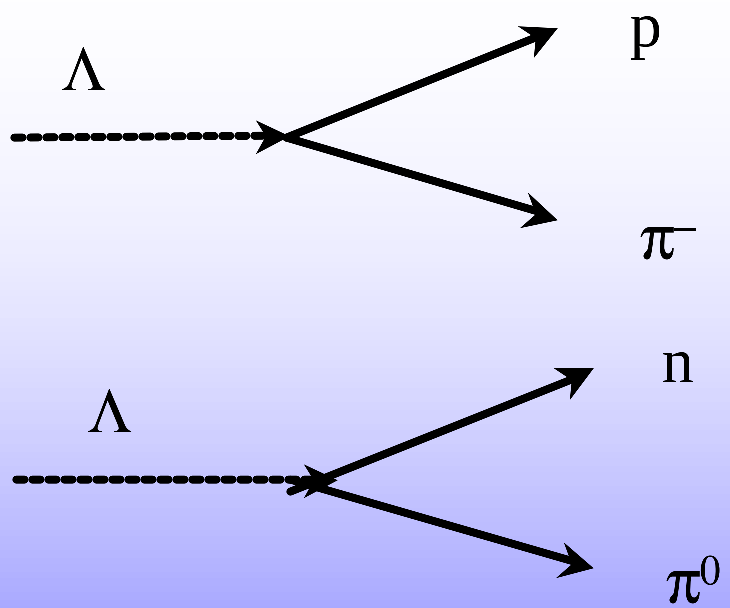
P(A) depends on A and the ensemble

Eg: count 10 of a group of 30 with beards.

P(beard)=1/3

Aside: Consequences for Quantum Mechanics

- QM calculates probabilities
- Probabilities are not ‘real’ – they depend on the process and the ensemble



PDG:

$P(p\pi^-)=0.639$,

$P(n\pi^0)=0.358$

Big problem for the Frequentist definition

Cannot be applied to unique events

~~'It will probably rain tomorrow'~~

Is unscientific

`The statement

"It will rain tomorrow"

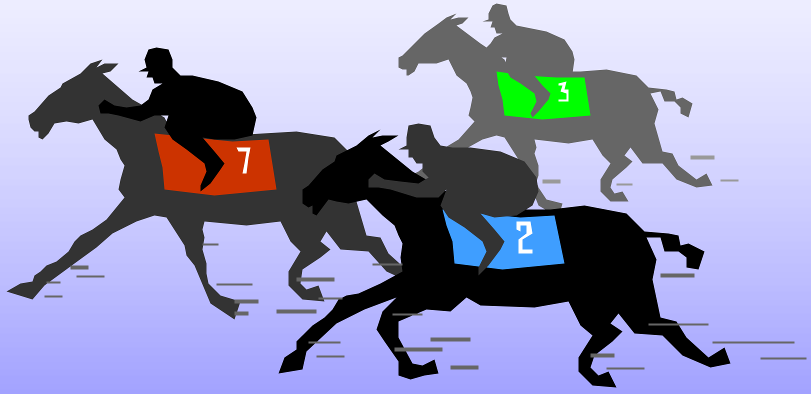
is probably true.'

Is quite OK

But that doesn’t always work

- Rain prediction in unfamiliar territory
- Euler’s theorem
- Higgs discovery
- Dark matter
- LHC completion

Definition 4: Subjective (Bayesian)



P(A) is your degree of belief in A;

You will accept a bet on A if the odds are better than

1-P to P

A can be Anything :

Beards, Rain, particle decays, conjectures, theories

Bayes Theorem
Often used for subjective probability

Conditional Probability P(A|B)

P(A & B)= P(B) P(A|B)

P(A & B)= P(A) P(B|A)

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

Example:

W=white jacket B=bald

P(W&B)=(2/4)x(1/2)

or (1/4)x(1/1)

P(W|B)

$= \frac{1}{(2/4)} \times (1/4) = 1/2$

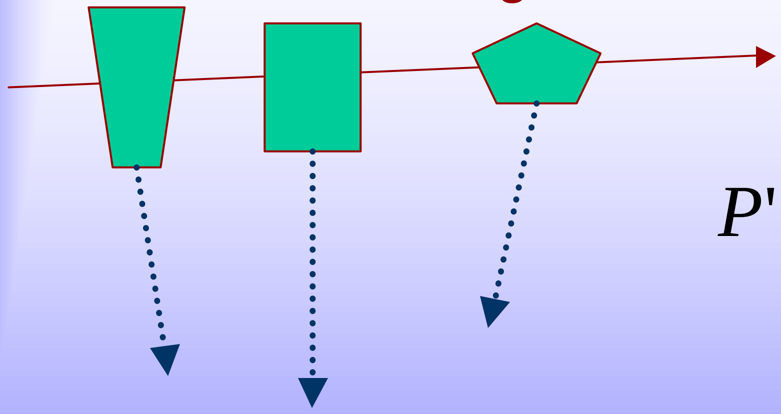
Start digression

Frequentist Use of Bayes Theorem

Example: Particle Identification

Particle types e,π,μ,K,p

Detector Signals: DCH,RICH,TOF,TRD



$$P'(e) = P(e \mid DCH) = \frac{P(DCH \mid e)}{P(DCH)} P(e)$$

$$P(DCH) = P(DCH \mid e)P(e) + P(DCH \mid m)P(m) + P(DCH \mid p)P(p)\ldots$$

Then repeat for P(e|RICH) using P'(e) etc

Warning Notice

To determine P'(e) need P(e), P(μ) etc

('*a priori* probabilities')

If/when you cut on Probability, the Purity depends on these *a priori* probabilities

Example: muon detectors.

P(track| μ)≈0.9 P(track|π) ≈0.015

But P(μ)≈0.01 P(π) ≈1

Quantities like

$$\frac{P(data\,|\,m)}{P(data\,|\,e)+P(data\,|\,m)+P(data\,|\,p)+P(data\,|\,K)}$$

Have no direct meaning –

use with care!

End digression

Bayes' Theorem and subjective probability

$$P(Theory|Result) = \frac{P(Result|Theory)}{P(Result)} P(Theory)$$

Your (posterior) belief in a Theory is modified by experimental result

If P(Result|Theory)=0 belief is killed

Large P(Result|Theory) increases belief, modified by general P(Result)

Applies to successive results

Dependence on prior P(Theory) eventually goes away

Problem with subjective probability

It is subjective

My P(A) and your P(A) may be different

Scientists are supposed to be objective

Reasons to use subjective probability:

- Desperation ✓
- Ignorance ✗
- Idleness ?

Can Honest Ignorance justify P(A)=flat?

Argument:

- you know nothing
- every value is as believable as any other
- all possibilities equal

How do you count discrete possibilities?

SM true or false?

SM or SUSY or light Higgs or Technicolor?

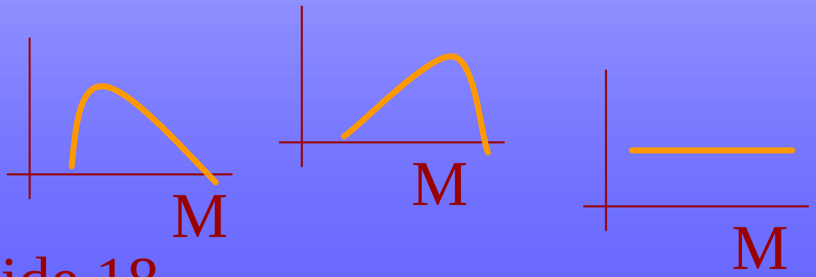
For continuous parameter (e g. M_{higgs})

Take. $P(M_{higgs})$ as flat

Actually has to be zero as $\int P(M)dM=1$ but never mind… ‘improper prior’

Working with ÖM or lnM will give different results

Real Statisticians accept this and test for robustness under different priors.



‘Objective Prior’ (Jeffreys)

Transform to a variable q(M) for which the Fisher information is constant

$$I(q) = -\left\langle \frac{\partial^2 \ln P(x;q)}{\partial q^2} \right\rangle = const$$

For a location parameter with P(x;M)=f(x+M) use M

For scale parameter with P(x;M)=Mf(x) use ln M

For a Poisson λ use prior $1/\sqrt{\lambda}$

For a Binomial with probability p use prior $1/\sqrt{p(1-p)}$

This has never really caught on

Conclusion
What is Probability?

4 ways to define it

- Mathematical
- Classical
- Frequentist
- Subjective

Each has strong points and weak points

None is universally applicable

Be prepared to understand and use them all -