An introduction to statistical inference

Ben Lambert[1]
ben.c.lambert@gmail.com

[1]University of Oxford

Tuesday 6th July, 2021

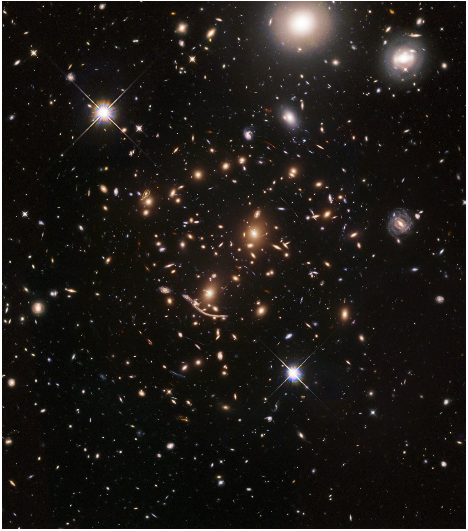
Course plan

- 9.30am-10.15am: lecture, “An introduction to statistical inference”
- 10.45am-1pm: practical

Outline

What is the aim of scientific inquiry?

Understand how the universe works.



What does it mean to understand the universe?

Understand causality.

How do we test our understanding?

Make predictions. Compare reality with our predictions.

What can statistical inference help with?

- Generating understanding from data
- Testing our scientific hypotheses versus data

But why do we need statistics?

- The universe is complex
- Its mechanisms are not directly observable
- Our data contain information both about the mechanisms and other nuisance factors

Statistics provides a way to separate the signal of the mechanisms from the noise.

How statistics separates signal from noise?

$$\text{observations} = \text{signal} + \text{noise} \tag{1}$$

- Signal contains our interesting scientific mechanism
- Noise contains a bunch of things not of interest

Since we do not know or observe the exact noise processes, in statistics, it is assumed that the noise is represented as being *random*.

But random does not mean unstructured. In statistics, making assumptions about the nature of the random process allows us to bound its influence on the observed data.

Example 1: flipping a coin

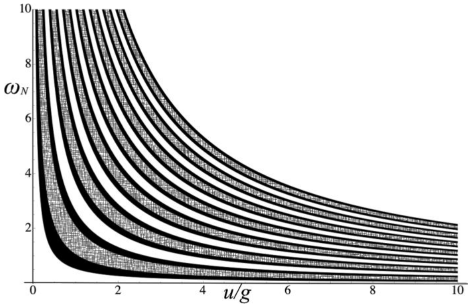
Suppose we flip a coin twice. We could obtain:

- Two tails
- One head; one tail
- Two heads

Why can we get different outcomes each time the coin is flipped?

Different initial conditions

Precessional frequency: ω_N; Magnitude of upward velocity, u.[1]



White indicates heads; hatched indicates lands on sides; hatched indicates tails.

[1] From Probability, geometry, and dynamics in the toss of a thick coin, Yong and Mahadevan (2011)

Coin flip dynamics: physics approach

Solve complex equations of motion (making assumptions about flipping process). One part of the system:

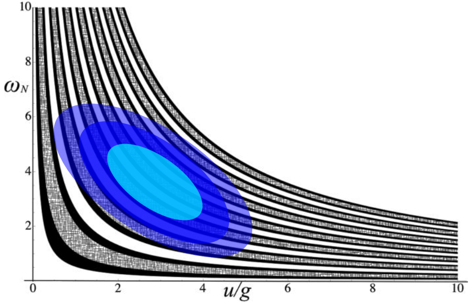
$$\frac{d\boldsymbol{N}}{dt} = \Omega \times \boldsymbol{N} \tag{2}$$

This system determines outcome given a set of initial conditions: precessional frequency and magnitude of upward velocity.

But we still don't know how different people throw a coin, so we'd need to measure this and likely represent this using randomness!

Coin flip dynamics: statistical approach

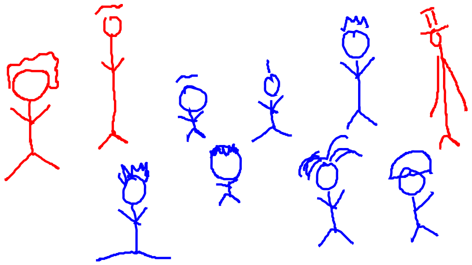
Assume outcome of a coin flip is a random variable with a probability of landing heads up (binomial distribution[2]). Implicitly:



[2] If we forget the landing on side situation.

Example 2: determining COVID-19 seropositivity

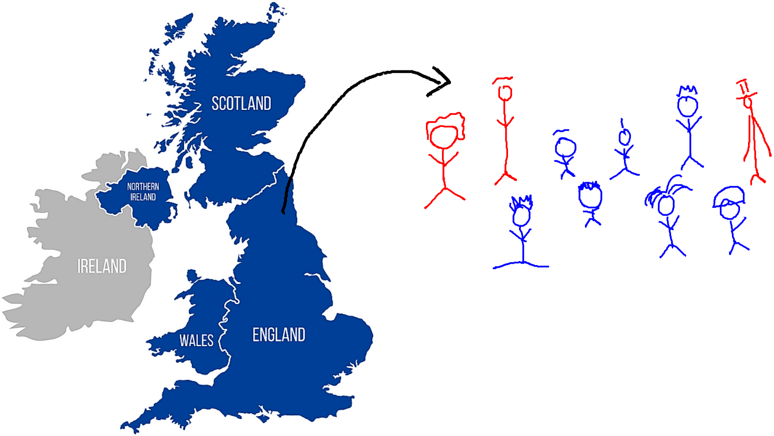
Imagine we want to determine the proportion of the UK population who have COVID-19 antibodies. To do this, we find 10 individuals and test their blood.



Does this mean $3/10 = 30\%$ of the UK population have these antibodies?

The sampling process yields variation in outputs

No.



What probability model to use?

Again, we could use a *binomial* model.

Note, very different problems often can share the same probability model.

Questions?

That's it!

Questions?