Course introduction

Outline:

1) Syllabus
2) Deductive vs inductive reasoning
3) The problem of induction
4) Coin flipping

Deductive vs inductive reasoning

Deduction:

Drawing inferences that follow logically from premises

Ex: (1) All real, symmetric matrices have real eigenvalues
(2) A is a symmetric matrix
Therefore, A has real eigenvalues

Ex: (1) No one in my daughter's preschool class has a peanut allergy
(2) Zoe is in my daughter's preschool class
Therefore, Zoe is not allergic to peanuts

Risk-free:

Valid arguments + true premises $\rightarrow$ true conclusions
(Conclusion could be wrong if a premise is wrong)

Deductive reasoning can involve probability:

(1) This die has six faces: 1, 2, .., 6
(2) Each side is equally likely
Therefore, the chance of rolling 4 is $1/6$

<u>Induction</u> : Observations $\longrightarrow$ general claims

<u>Risky!</u> We can (and will sometimes) be wrong

<u>Ex.</u> (1) I ate a free sample strawberry at the supermarket
(2) It was ripe and delicious.
Therefore, I should (probably) buy a whole carton.
(Would be more convincing with a random sample)

<u>Ex:</u> (1) Water at 1 atm pressure has always been observed to boil at 100°C
Therefore, all water at 1 atm (probably) boils at 100°C

<u>Ex:</u> (1) I flipped this coin 1000 times and got 502 heads
Therefore, it (probably) has about a 50% chance of landing heads

<u>Statistics</u> : mathematical science of inductive reasoning

The Problem of Induction

Inductive reasoning is not logically valid

<u>Hume</u>: All inductive reasoning requires a presumption
that unobserved cases will be like observed cases

How can we justify this?

Logical proof? There is none!*

Past observation? Circular!

Hume admitted we must reason inductively all the time
Called it "custom" or "habit"

Statistics evades this problem in one of two ways

<u>Idea 1</u>: Bayesian reasoning

Whatever our prior beliefs, we know how to update them in light of experience.

<u>Idea 2</u>: "Inductive behavior" (frequentist statistics)

If observations are from a reasonable experiment, we can design methods that give correct conclusions with high probability (provably!)

Coin Flipping

Diaconis, Holmes, & Montgomery (2007):

A coin is a bit more likely to land on the side it started on

Based on physical model (precession)

Predicted $\sim 51\%$ for typical human flipper

Bartoš et al. (2023):

350,757 coin flips

48 flippers using coins from 46 countries

Found 178,079 same-side outcomes ($\approx 50.8\%$)

Frequentist analysis

95% confidence interval [50.6%, 50.9%]

Based on binomial model:

- Every flip has same probability θ
- Flips are independent

$$\Rightarrow \mathbb{P}_\theta(X \text{ same side flips}) = \frac{n!}{x!(n-x)!} x^\theta (n-x)^{1-\theta}$$

Confidence interval construction mathematically ensures

$$\mathbb{P}_\theta(CI(X) \text{ covers } \theta) \geq 95\% \quad \text{(here, almost =)}$$

Bayesian analysis

95% credible interval [50.6%, 50.9%]

Same binomial model, plus Unif[0,1] prior on θ

(Question: whose prior opinion was that?)

Binomial model was wrong!

- Different flippers had different probabilities
- Most flippers improved (got closer to 50%) over time

Kinds of questions we'll ask in this course

Bayesian and frequentist frameworks

What are pros & cons of each?

Where does the prior come from? (Does it matter?)

What does probability mean in each framework?

Sufficiency:

Both analyses summarized data as "$X = 178{,}079$"

Did we lose anything? (Not under binomial model)

What about the model structure lets us do this?

<u>Estimation</u>: What's a good way to estimate:

1) Overall θ in binomial model

   no single best estimator for all θ

   X/n "obvious choice" if $n = 350k$ (unless $\theta = 10^{-6}$)

   less obvious for $n = 35$

2) θ_i for individual flipper i

   new model: $X_i \overset{ind.}{\sim} \text{Binom}(n_i, \theta_i)$ for $i = 1, \dots, m$

   how to use data from other flippers?

3) How fast $\theta_{i,t}$ changes from flip 1 to flip n_i

   variety of possible models

   parametric & nonparametric

<u>Testing</u>: How do we <u>efficiently</u> test hypotheses like

1) $H_0: \theta \leq 50\%$ vs $H_1: \theta > 50\%$ (one-sided)

   Unique best test exists

2) $H_0: \theta = 50\%$ vs. $H_1: \theta \neq 50\%$ (two-sided)

   Natural choice exists

3) $H_0: \theta_i = \theta$ for all flippers vs. $H_1: \theta_i$ varies

   <u>Nuisance parameter</u> θ affects null distribution of any test statistic

   Many ways for $(\theta_1, \dots, \theta_m)$ to be non-null should affect choice of test!

4) H_0: θ_i constant through time for all flippers vs H_1: $\theta_{i,t}$ tends to be decreasing in t

   (Can test this nonparametrically)

Asymptotics: No one calculated 350,757!

Actual model: $X \sim N(n\theta, \; n\theta(1-\theta))$

(or $X_i \overset{ind}{\sim} N(n_i\theta_i, \; n_i\theta_i(1-\theta_i))$)

Want good asymptotic approximations to other models