Course Introduction

ISTA 410 / INFO 510 - Bayesian Modeling and Inference

University of Arizona School of Information

August 23, 2021

Course Overview

Bayesian Modeling

- Represent our knowledge in the form of a *probability distribution*; probabilities measure confidence or belief
- Establish a model indicating the dependence of observations on parameters (and the dependence of those parameters on one another)
- Use Bayes' theorem to update the distribution(s) of the model parameter(s) based on data

Inference

- Estimate values of the model parameters (point estimates, interval estimates, etc.)
- Predict future observations

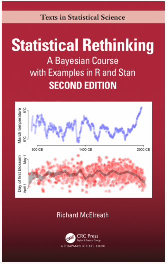
Topics

We'll cover the following topics, in something like this order:

- Foundations: probability theory and the Bayesian interpretation
- Simple one- and multi-parameter models
- Hierarchical and graphical models
- Model checking and evaluation
- Computational methods
    - Approximate inference with Markov chain Monte Carlo (MCMC)
- Mixture models and expectation-maximization algorithms
- Inference for dynamical systems using Kalman filters
- Other topics as time allows

Textbook

Statistical Rethinking, 2nd ed., by Richard McElreath



Available online at UA Library

Other sources as needed:

- *Bayesian Data Analysis*, 3rd ed., by Andrew Gelman, et al.
- *Causality*, 2nd ed., by Judea Pearl

Software requirements

In order to complete the course, especially the programming components, you'll need the following software on your computer:

- Zoom (videoconference software).
- Python, with the following packages: NumPy, SciPy, matplotlib, **PyMC3**, and Jupyter (notebook app or JupyterLab)

(PyMC3 is highlighted because it's the one you're least likely to already have!)

Expectations

- Attend videoconference class meetings
- Prepare for meetings by completing assigned readings, attempting exercises (sometimes)
- Complete several homework assignments over the course of the semester
- Complete a take-home midterm and final

Goals and methods

What I hope you will come out of this course with:

- Practical skills in building models from the ground up and critically analyzing them
- Enough philosophy of statistics to ground the above
- A high degree of comfort with confusion and uncertainty

Instructor and contact information

Dylan Murphy, Ph.D.
Lecturer in the iSchool
Please call me: Dylan or Dr. Murphy
Office: Harvill 444
Office hours: Tu 2-4 PM (drop-in office hours on Zoom or in the office) or by appointment

---

A Slack channel for this course will be available through a link on the D2L site.

Python resources

Resources:

- Python/PyMC3 translation of Rethinking: `https://github.com/pymc-devs/resources`
- GitHub repository for this course: `https://github.com/mpdylan/INFO510-public`

A first example

Flipping a coin

A researcher is trying to determine whether a coin is fair; she suspects it may be less likely to come up heads.

The data consist of a history of 12 flips:

$$T, T, H, T, T, T, H, T, T, T, T, H$$

What do you do to decide whether the coin is fair?

Flipping a coin

A researcher is trying to determine whether a coin is fair; she suspects it may be less likely to come up heads.

The data consist of a history of 12 flips:

$$T, T, H, T, T, T, H, T, T, T, T, H$$

What do you do to decide whether the coin is fair?

Let's start by thinking from the classical (frequentist) statistical perspective.

Flipping a coin

Idea:

- The result is drawn from a population of theoretically possible experiments, where we flip the coin $n = 12$ times and get r heads
- Under the assumption that the probability of heads, p_H, is 0.5 (called the null hypothesis), we can calculate the probability that $r \leq 3$
- If this probability (called the p-value) is small (conventionally, $p < 0.05$), then we conclude the null is inconsistent with the data

Flipping a coin

- flips are independent trials $\rightarrow$ r has a binomial distribution

Flipping a coin

- flips are independent trials $\rightarrow$ r has a binomial distribution
- If $r \sim \mathrm{Binomial}(n, p_H)$, then

$$P(r = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p_H^k (1 - p_H)^{1-k}$$

Flipping a coin

- flips are independent trials $\rightarrow r$ has a binomial distribution
- If $r \sim \mathrm{Binomial}(n, p_H)$, then

$$P(r = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p_H^k (1 - p_H)^{1-k}$$

- So,

$$\begin{aligned} P(r \leq k | p_H = 0.5) &= 0.5^{12} \left(\left(\begin{array}{c} n \\ 0 \end{array} \right) + \left(\begin{array}{c} n \\ 1 \end{array} \right) \right. \\ &\quad \left. + \left(\begin{array}{c} n \\ 2 \end{array} \right) + \left(\begin{array}{c} n \\ 3 \end{array} \right) \right) \\ &\approx 0.07 \end{aligned}$$

Flipping a coin

- flips are independent trials $\rightarrow$ r has a binomial distribution
- If $r \sim \mathrm{Binomial}(n, p_H)$, then

$$P(r = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p_H^k (1 - p_H)^{1-k}$$

- So,

$$\begin{aligned} P(r \leq k | p_H = 0.5) &= 0.5^{12} \left(\left(\begin{array}{c} n \\ 0 \end{array} \right) + \left(\begin{array}{c} n \\ 1 \end{array} \right) \right. \\ &\quad \left. + \left(\begin{array}{c} n \\ 2 \end{array} \right) + \left(\begin{array}{c} n \\ 3 \end{array} \right) \right) \\ &\approx 0.07 \end{aligned}$$

So we announce to our colleague that there is not statistically significant evidence of bias... right?

Flipping a coin

Our colleague comes back and examines the calculation. She says: "I'm sorry, but you've misunderstood my experimental design. r is not a random variable; I decided ahead of time that I would flip the coin until I got three heads, so n is the random variable."

What's different?

Flipping a coin

Our colleague comes back and examines the calculation. She says: "I'm sorry, but you've misunderstood my experimental design. r is not a random variable; I decided ahead of time that I would flip the coin until I got three heads, so n is the random variable."

What's different?

- The universe of possible experiments is different

Flipping a coin

Our colleague comes back and examines the calculation. She says: "I'm sorry, but you've misunderstood my experimental design. r is not a random variable; I decided ahead of time that I would flip the coin until I got three heads, so n is the random variable."

What's different?

- The universe of possible experiments is different
- Now, the probability of n flips and r heads is given by

$$P(n|p_H = 0.5) = \left(\begin{array}{c} n-1 \\ r-1 \end{array} \right) 0.5^n$$

Flipping a coin

Our colleague comes back and examines the calculation. She says: "I'm sorry, but you've misunderstood my experimental design. r is not a random variable; I decided ahead of time that I would flip the coin until I got three heads, so n is the random variable."

What's different?

- The universe of possible experiments is different
- Now, the probability of n flips and r heads is given by

$$P(n|p_H = 0.5) = \binom{n-1}{r-1} 0.5^n$$

- So,

$$P(n \geq 12|r = 3, p_H = 0.5) = \sum_{k=12}^{\infty} \binom{k-1}{r-1} 0.5^k \approx 0.03$$

Flipping a coin

Our colleague comes back and examines the calculation. She says: "I'm sorry, but you've misunderstood my experimental design. r is not a random variable; I decided ahead of time that I would flip the coin until I got three heads, so n is the random variable."

What's different?

- The universe of possible experiments is different
- Now, the probability of n flips and r heads is given by

$$P(n|p_H = 0.5) = \left(\begin{array}{c} n-1 \\ r-1 \end{array} \right) 0.5^n$$

- So,

$$P(n \geq 12 | r = 3, p_H = 0.5) = \sum_{k=12}^{\infty} \left(\begin{array}{c} k-1 \\ r-1 \end{array} \right) 0.5^k \approx 0.03$$

So, there is statistically significant evidence of bias... right?

Against tests

This should make you at least a little bit uneasy with statistical tests.

- Statistical tests: pre-packaged procedures
- Most are old; can be fragile, sensitive to opaque or unstated assumptions
- Based on falsification of an often arbitrary null model

Bayesian modeling

- Use probability to describe uncertainty
  - Rank possible explanations by how likely they are to produce the data
  - Extends conventional logic of true/false to a continuous scale of *plausibility*
- Computationally difficult
  - Few closed form solutions
  - Modern MCMC methods save us
- Requires a model for how the data might arise

The cookie problem

Suppose we have two barrels of cookies.[1] Bowl 1 contains a 3 each of chocolate and vanilla cookies. Bowl 2 has 4 vanilla and 2 chocolate.

We select a bowl at random and, without looking at which one we picked, pull 2 cookies at random from it. Both of them are chocolate.

What does this tell us about which bowl we're drawing from?

[1] This example is from *Think Bayes* by Allen Downey.

A Bayesian model

Let's make a model of the cookie-drawing process.

- a *model* just means a simplified representation of reality
- in Bayesian modeling, the reality we are trying to represent is the *process that created the data*

A Bayesian model

Let's make a model of the cookie-drawing process.

- a *model* just means a simplified representation of reality
- in Bayesian modeling, the reality we are trying to represent is the *process that created the data*

Our model is:

- We are working with one of the two bowls
- At the time of cookie selection, the cookie is randomly assigned to be chocolate or vanilla, with probabilities based on the identity of the barrel

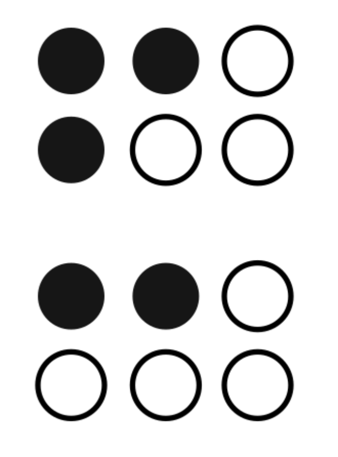
A Bayesian model

We have two competing ideas:

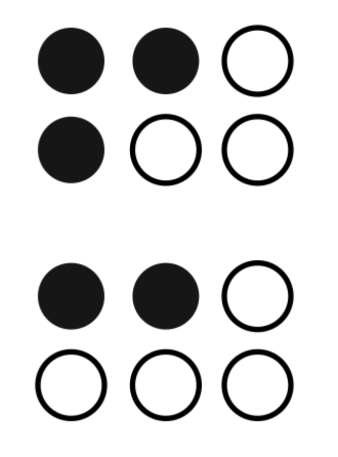
- We are drawing from bowl 1, and we chose two chocolate cookies
- We are drawing from bowl 2, and we chose two chocolate cookies

Approach: count the ways the data could arise

A Bayesian model

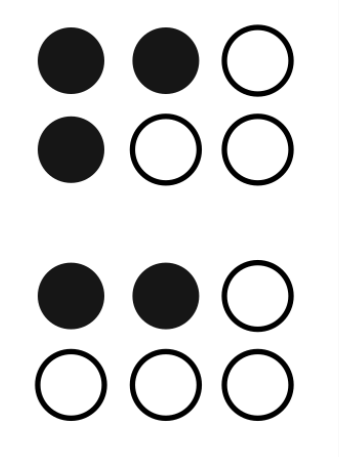


A Bayesian model



- 3 possible ways to get 2 chocolate cookies drawing from bowl 1

A Bayesian model



- 3 possible ways to get 2 chocolate cookies drawing from bowl 1
- Only 1 way to get 2 chocolate cookies drawing from bowl 2

How can we extend this to the coin flip?

Returning to the coin flip

Returning to the coin flip:

- Problem with the null/alternative hypothesis approach
  - Null: $p_H = 0.5$
  - Alternative: $p_H \neq 0.5$

Returning to the coin flip

Returning to the coin flip:

- Problem with the null/alternative hypothesis approach
  - Null: $p_H = 0.5$
  - Alternative: $p_H \neq 0.5$
- How can these hypotheses possibly be on equal footing?
- Why are $p_H = 0.9$ and $p_H = 0.2$ both consistent with the "alternative hypothesis"?

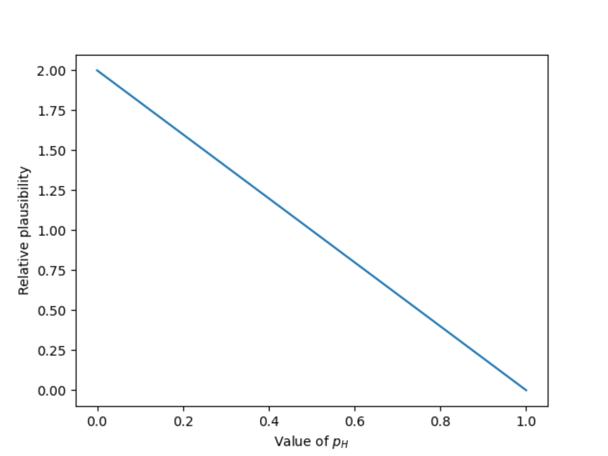
Bayesian model for the coin flip

The Bayesian perspective:

- Treat p_H as a continuous variable
  - For any value of p_H, we can calculate the probability of our observed sequence
- Then we can rank our values of p_H by their relative plausibility, as measured by how likely they are to create the observed data
- This "ranking" then becomes a continuous function of p_H

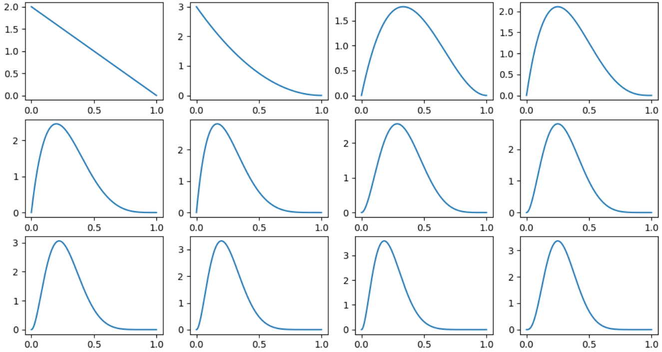
Ranking the plausibility

The first flip we got was a tails; the plausibility of any given value of p_H is just proportional to its likelihood of producing tails:



Ranking the plausibility

Sequentially after each flip:



For next time

Main task for the first week: get your software environment set up.

- I strongly recommend using the Anaconda distribution of Python
- Simplest thing: create a conda environment using the environment file on D2L
- If running Windows, you may need:
    - `conda install m2w64-toolchain`
    - `conda install -c anaconda libpython`

Confirm that you can `import pymc3` without errors.

One more task: read chapter 1 in Statistical Rethinking