Introduction & Probability Review

Stat 340: Bayesian Statistics

Welcome back!

1. Welcome

2. Probability review

3. Logistics

4. Comparing the philosophies

Reading:

Sections 7.1 - 7.2

Homework for Friday:

Review probability as needed

You should know how to solve the review problems

About me

- 5th year at Carleton
- Taught in Wisconsin for 4 years
- Ph.D., M.S. from Iowa State
- B.A. from Luther College
- Originally from Maryland
- Pragmatist (not a rabid Bayesian)

Image from Unsplash

COVID-19 related policies

To ensure the well-being of each other and the broader community, we will

- stay home when sick. (Even if you don't have COVID-19, you should stay home if you aren't feeling well.)
- follow CDC guidance on testing, quarantine, and isolation.
- follow the College mask-wearing policy.
- not eat in class while the college is requiring mask wearing.

What if Adam needs to miss class?

- I'll send a note through both email and Slack
- In most cases, class will still be held, but in a synchronous online format using Zoom

What if I need to miss class?

- Let me know
- Check for materials (slide, handouts, etc.), assignments, and announcements on the course webpage
- Check out the collaborative notes for that day
- Stop by virtual student hours or make an appointment to ask questions

Collaborative notes

- Each day, 2-3 of you will collaborate on notes to share with the class
- Creates a crowd-sourced version of what we do in class
- Helps anyone who needs to miss class
- You'll do this 2x throughout the course
- Sign up here (link is also on Week 01 webpage)
- Notes are due 24 hours after class
- Together, your 2 contributions count as a HW assignment

Probability review

Main results to review

Be sure that you are comfortable with the following concepts/results from probability:

- Probability mass/density functions
- Joint probability/distributions
- Marginal probability/distributions
- **Conditional probability/distributions**
- **Bayes' rule**
- **Law of total probability**

Example

A spam filter is designed by looking at commonly occurring phrases in spam. Suppose that 80% of email is spam. In 10% of the spam emails, the phrase “free money” is used, whereas this phrase is only used in 1% of non-spam emails.

1. What is the probability an email contains the phrase “free money” and is spam?

2. Write the event that an email contains the phrase “free money” as the union of disjoint events.

3. What is the probability that an email contains the phrase “free money?”

4. A new email has just arrived, which does mention “free money.” What is the probability that it is spam?

Need more review?

- Chapters 1-6 of *Probability and Bayesian Modeling*
- Section 2.1 of *Bayesian Ideas and Data Analysis*
- My probability lectures from fall 2020
- Stat 110 lectures by Joe Blitzstein at Harvard

Syllabus highlights

Please read the full syllabus

Scheduled ~~office~~ student hours

In person, CMC 307

Day	Time
Monday	2:30-3:30
Tuesday	12:30-1:30
Wednesday	1-2
Thursday	12:30-1:30

On Slack

Day	Time
Monday	8-9pm
Thursday	8-9pm

I'll also have appointment slots posted via Calendly

Types of "assignments"

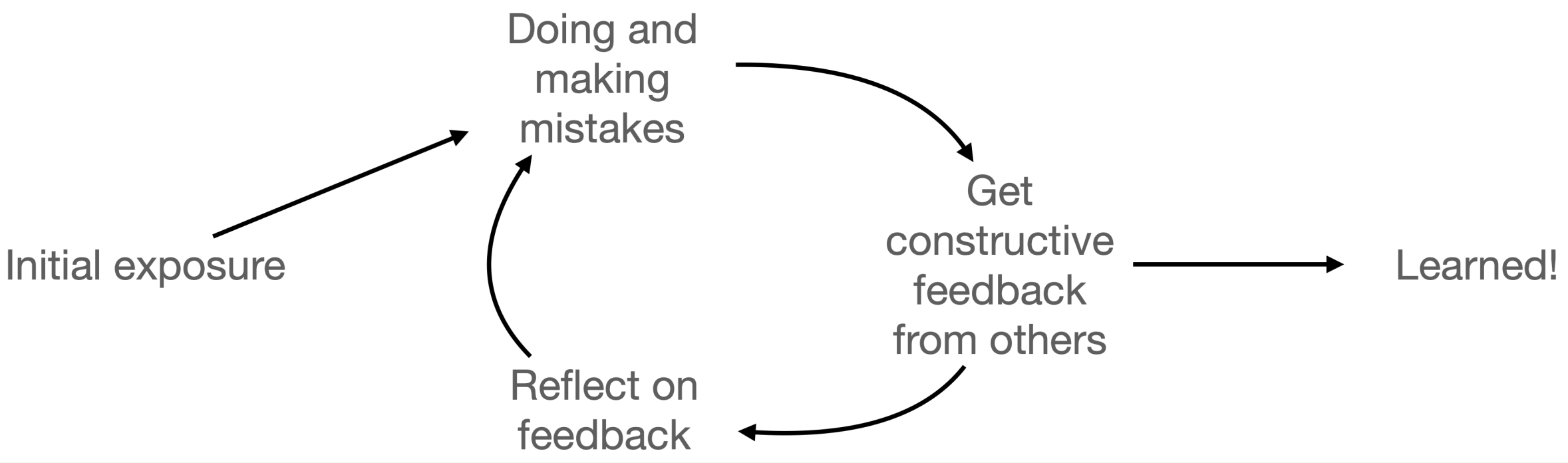
GRADED WORK

- Homework (10)
- Writing prompts (5)
- Problems (18 across 3 exams)
- Final project

UNGRADED WORK

- Attendance and engagement
- Reading and review

Learning takes place over time through feedback loops



Specifications grading

This courses uses a “specifications” (or “specs”) grading system

- Your work is evaluated on a pass/not yet basis, depending on whether or not it meets predefined specifications
- There are no points
- There is no partial credit
- Specs are clearly laid out and available from day 1
- There are opportunities to revise work that doesn't meet the specs
- Your grade is based on what you learn

Perceived benefits

- It will be very clear which topics you need to work on
- Avoids one-and-done tests
- You decide which grade you're aiming for, and you know what you have to do to earn it
- Aligns your grade more directly with how much course material you have mastered, and in what depth
- Mirrors expectations in industry

Possible drawbacks

- Unfamiliar system $\rightarrow$ possible stress!
- Revisions take time
- If you have relied on partial credit $\rightarrow$ shift mindset

Baseline grades

Assignment/Grade	A	B	C	D
Homework (of 10)	9	8	7	6
Writing (of 5)	5	4	3	2
Problems (of 18)	18	15	12	9
Final project	Success	Success	Not required	Not required

+/- based on your progress toward the next letter grade

Specifications are outlined on the resources page

Tokens

You can use a token to rewrite work that doesn't meet specifications

- Revision of a homework or writing assignment that did not earn a "Success"
- Revision of a writing assignment where you missed the rough draft or peer review (2 tokens, 1 for the revision, 1 for an exception to the intermediate deadlines)
- 48-hour extension on a homework assignment (the request must be submitted before the deadline)
- Revision of a Problem that earned a "Not yet"

Tokens

- I will track token balances in the Moodle gradebook (updated weekly)
- You start the term with 4 tokens
- You can earn up to 3 more

Why limit the number of tokens?

- You can't continually turn in substandard work without penalty
- Reflects reality in industry
- I am grading everything and you need prompt feedback

Questions?

Please read the syllabus for many more details

A Bayesian "personality quiz"

Adapted from *Bayes Rules! An Introduction to Bayesian Modeling with R* by Johnson, Ott & Dogucu

Question 1

When flipping a fair coin, we say that "the probability of flipping Heads is 0.5." How do you interpret this probability?

1. If I flip this coin over and over, roughly 50% will be Heads.

2. Heads and Tails are equally plausible.

3. Both a and b make sense.

Question 2

An election is coming up and a pollster claims that "candidate A has a 0.9 probability of winning." How do you interpret this probability?

1. If we observe the election over and over, candidate A will win roughly 90% of the time.

2. Candidate A is much more likely to win than to lose.

3. The pollster's calculation is wrong. Candidate A will either win or lose, thus their probability of winning can only be 0 or 1.

Question 3

Consider two claims.

- Zuofu claims that he can predict the outcome of a coin flip. To test his claim, you flip a fair coin 10 times and he correctly predicts all 10.
- Kavya claims that she can distinguish natural and artificial sweeteners. To test her claim, you give her 10 sweetener samples and she correctly identifies each.

In light of these experiments, what do you conclude?

1. You're more confident in Kavya's claim than Zuofu's claim.

2. The evidence supporting Zuofu's claim is just as strong as the evidence supporting Kavya's claim.

Question 4

Suppose that during a recent doctor's visit, you tested positive for a very rare disease. If you only get to ask the doctor one question, which would it be?

1. What's the chance that I actually have the disease?

2. If in fact I don't have the disease, what's the chance that I would've gotten this positive test result?

Tally your points

Question 1:

- 1 = 1 points
- 2 = 3 points
- 3 = 2 points

Question 2:

- 1 = 1 points
- 2 = 3 points
- 3 = 1 points

Question 3:

- 1 = 3 points
- 2 = 1 points

Question 4:

- 1 = 3 points
- 2 = 1 points

What does your score mean?

- 4-5 $\rightarrow$ you're more of a frequentist thinker
- 6-8 $\rightarrow$ you see the merit in both (a pragmatist?)
- 9-12 $\rightarrow$ you're more of a Bayesian thinker

Question 1: Interpretting probability

When flipping a fair coin, we say that "the probability of flipping Heads is 0.5." How do you interpret this probability?

1. (Frequentist) If I flip this coin over and over, roughly 50% will be Heads.

2. (Bayesian) Heads and Tails are equally plausible.

3. Both a and b make sense.

Question 2: Interpretting probability

An election is coming up and a pollster claims that "candidate A has a 0.9 probability of winning." How do you interpret this probability?

1. (Frequentist) If we observe the election over and over, candidate A will win roughly 90% of the time.

2. (Bayesian) Candidate A is much more likely to win than to lose.

3. (Rabid frequentist) The pollster's calculation is wrong. Candidate A will either win or lose, thus their probability of winning can only be 0 or 1.

Question 3: Balancing prior info and observed data

Consider two claims.

- Zuofu claims that he can predict the outcome of a coin flip. To test his claim, you flip a fair coin 10 times and he correctly predicts all 10.
- Kavya claims that she can distinguish natural and artificial sweeteners. To test her claim, you give her 10 sweetener samples and she correctly identifies each.

In light of these experiments, what do you conclude?

1. (Bayesian) You're more confident in Kavya's claim than Zuofu's claim.

2. (Frequentist) The evidence supporting Zuofu's claim is just as strong as the evidence supporting Kavya's claim.

Question 4: Asking questions

Suppose that during a recent doctor's visit, you tested positive for a very rare disease. If you only get to ask the doctor one question, which would it be?

1. (Bayesian) What's the chance that I actually have the disease?

2. (Frequentist) If in fact I don't have the disease, what's the chance that I would've gotten this positive test result?