Inference & Causality
Week 1

20.10.2025

Lecturer: Narges Chinichian
IU University of Applied Science, Berlin

About Me

- Dr. Narges Chinichian
- PhD in Physics (Computational Neuroscience), with 8+ years experience in data science & machine learning.
- Hobby teacher, I also teach rope climbing at the neighboring gym Der Kegel!
- Fun fact about me: I’m a grand-grand-…-grand student of Gauss.

About You?

Please share briefly:

- 1. Your name & if you'd like: where you're from.
- 2. What languages you speak/read.
- 3. What is your educational background.
- 4. How would you rate your Python and Git skills?
- 5. What are your expectations for this course? What do you hope to learn?

- Bonus: If causality were a person, how would they look and what would they be like?

Course Overview

- 6 weeks | 2 day per week (12 sessions)| 4 × 45min units
- Hands-on, project-based learning

- Check the course hub on Notion for up-to-date information:
  https://tinyurl.com/mrcjp79s

The course will be covering:

- Statistical inference
- Introduction to causality
- Interventions
- Do-calculus
- Fallacies

Outline of Week 1

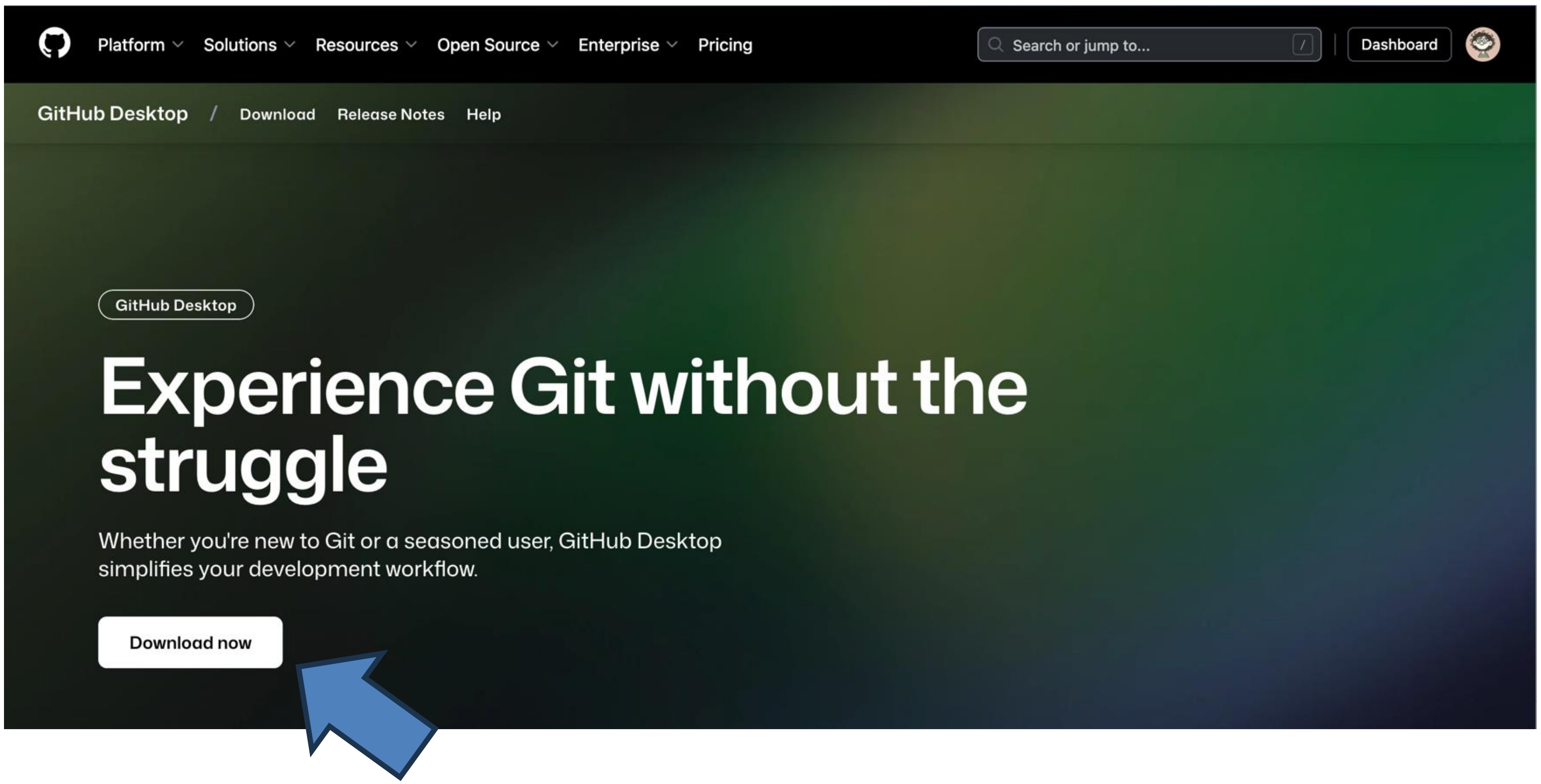
- Set up the infrastructure (GitHub).
- Explain the difference between frequentist and Bayesian inference
- Apply Bayes' theorem to simple problems
- Update beliefs with new evidence

Setting Up Your GitHub

- Please all add your GitHub handle and name here: https://tinyurl.com/4xxzhphf
- You handle is what you get in your url when you are on your profile page:
- So if you see:
- https://github.com/NoCh-Git
- Your handle is NoCh-Git.

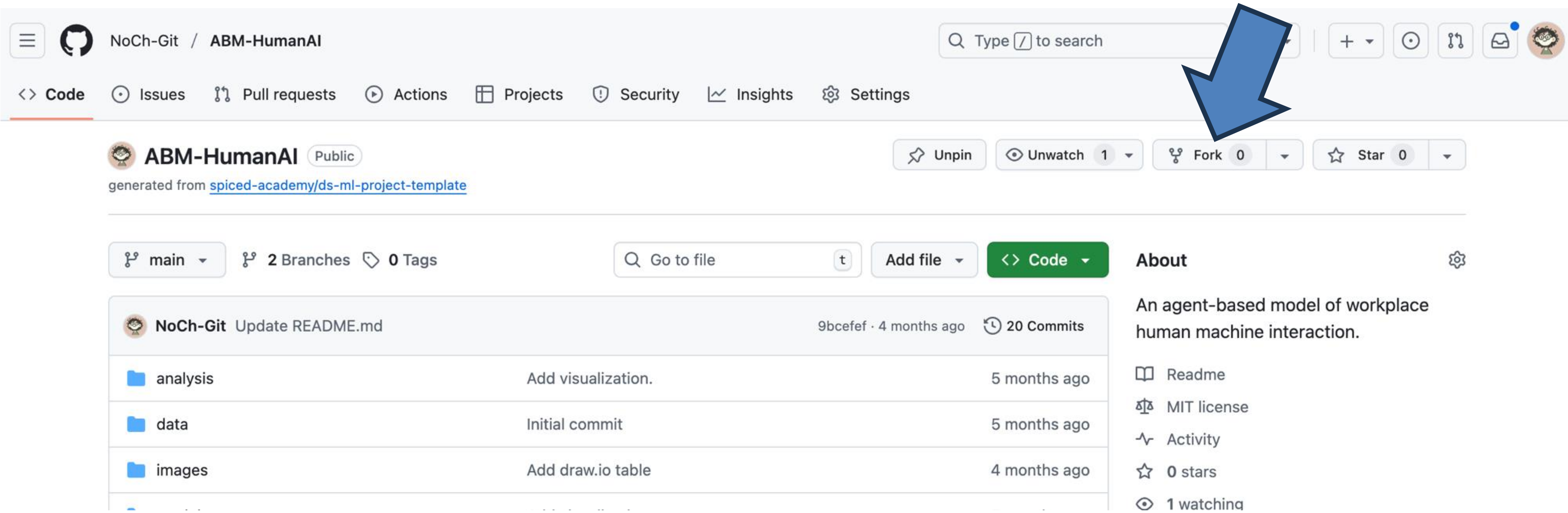
Install GitHub Desktop (Optional)

https://github.com/apps/desktop



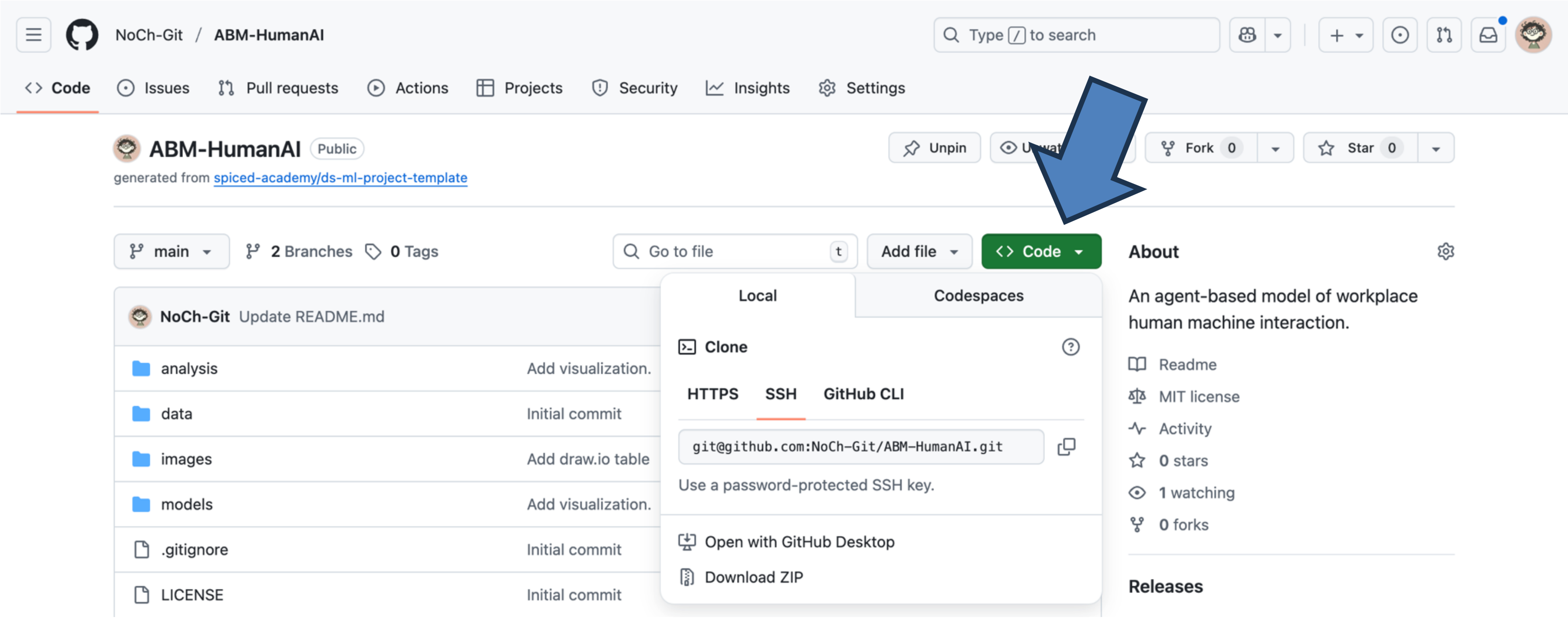
Fork Repo of Today

- Forking a repo would create a copy of that repo for you that you can play with.
- Choose yourself as the owner and untick the “Copy the main branch only” box.



Clone the Forked Repo to Your Machine Using GitHub Desktop or CLT

- You need to have a local copy of the Python notebooks.



Ready?

Notebook 1

Notebook 1 of week 1 in your “notebooks” folder.

What is Statistical Inference?

- Drawing conclusions about unknowns from data
- Two paradigms:
  - Frequentist: parameters fixed, data random
  - Bayesian: parameters uncertain, updated via evidence
- Both appear throughout this module; today we compare and practice

Want to play a quick game?

Two Paradigms:

Frequentist

- Probability = long-run frequency
- Parameters are fixed (unknown), data is random
- Confidence intervals, hypothesis tests

Bayesian

- Probability = degree of belief
- Parameters have distributions (prior → posterior)
- Bayes' theorem combines prior and likelihood

Bayes' Theorem

$$P(\mathrm{H}|\mathrm{D}) = \frac{P(\mathrm{D}|\mathrm{H})}{P(\mathrm{D})} P(\mathrm{H})$$

- Where:
- H = Hypothesis for example some "disease"
- D = Data for example a "positive test"
- $P(\mathrm{H})$ is the probability of the hypothesis before we see the data, called the **prior** probability, or just prior.
- $P(\mathrm{H}|\mathrm{D})$ is the probability of the hypothesis after we see the data, called the **posterior**.
- $P(\mathrm{D}|\mathrm{H})$ is the probability of the data under the hypothesis, called the **likelihood**.
- $P(\mathrm{D})$ is the **total probability** of the data, under any hypothesis.

Cookie Problem

From Allen Downey's version of Urn problem

- Suppose there are two bowls of cookies.
- Bowl 1 contains 30 vanilla cookies and 10 chocolate cookies.
- Bowl 2 contains 20 vanilla cookies and 20 chocolate cookies.
- Now suppose you choose **one of the bowls at random** and, without looking, choose **a cookie at random**. If the cookie is **vanilla**, what is the probability that it came from **Bowl 1**?

$$P(B_1|V) = \frac{P(B_1)P(V|B_1)}{P(V)}$$

$$\left.\begin{aligned} P(V|\mathrm{B}_1) &= \frac{3}{4} \text{ and } P(\mathrm{B}_1) = \frac{1}{2} \\ P(V) &= P(\mathrm{B}_1)P(V|\mathrm{B}_1) + P(\mathrm{B}_2)P(V|\mathrm{B}_2) \\ P(V) &= \frac{1}{2}\times\frac{3}{4} + \frac{1}{2}\times\frac{1}{2} = \frac{5}{8} \end{aligned}\right\} \quad P(\mathrm{B}_1|V) = \frac{\frac{1}{2}\times\frac{3}{4}}{\frac{5}{8}} = \frac{3}{5}$$

Bayes Tables

Useful tools to calculate Bayesian probabilities.

Normalized multiplication of prior and likelihood = Posterior

	prior	likelihood	unnorm	posterior
Bowl 1	0.5	0.75	0.375	0.6
Bowl 2	0.5	0.50	0.250	0.4

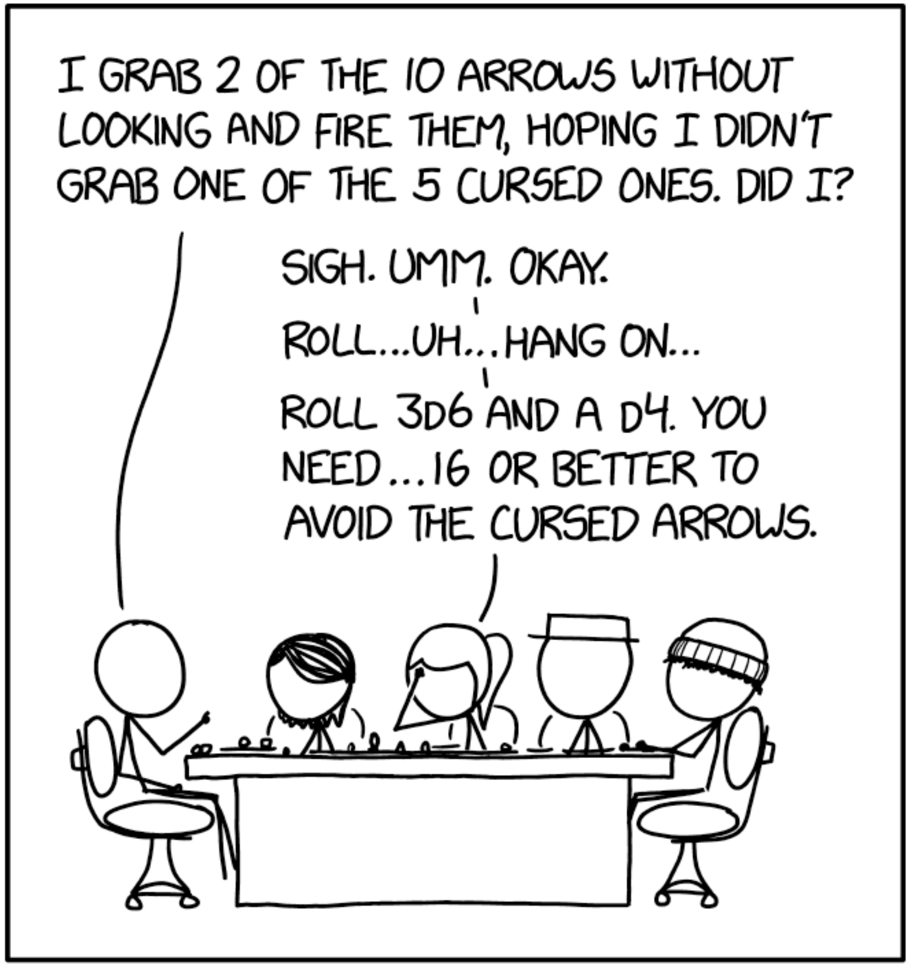
Unnormalized multiplication of prior and likelihood

Dice Problem (more than two distributions)

From Allen Downey's book, Think Bayes

- Suppose there are three dice:
- A 6 sided, an 8 sided and a 12 sided.
- We choose one of the dice at random, roll it, and report that the outcome is a 1.
- What is the probability that I chose the 6-sided die?

	prior	likelihood	unnorm	posterior
6	1/3	1/6	1/18	4/9
8	1/3	1/8	1/24	1/3
12	1/3	1/12	1/36	2/9



I GOT *WAY* MORE ANNOYING TO PLAY D&D WITH ONCE I LEARNED THAT OUR DM HAS A COMBINATORICS DEGREE AND CAN'T RESIST PUZZLES.

Source: https://www.explainxkcd.com/wiki/index.php/3015:_D%26D_Combinatorics

Monty Hall Problem

Selvin, S. (1975)

- Based on a show called “Let’s Make a Deal”.
- You are on a game show and have to choose one of three doors:
- Behind one door is a car, and behind the other two are goats.
- You pick a door (say, Door 1).
- The host, who knows what’s behind each door, opens another door (say, Door 3) revealing a goat.
- You are then asked:

“Do you want to stick with your choice, or switch to the other unopened door?”

What should you do to maximize your chance of winning the car?

- If the car is behind Door 1, Monty chooses Door 2 or 3 at random, so the probability he opens Door 3 is .
- If the car is behind Door 2, Monty has to open Door 3, so the probability of the data under this hypothesis is 1.
- If the car is behind Door 3, Monty does not open it, so the probability of the data under this hypothesis is 0.

	prior	likelihood	unnorm	posterior
Door 1	1/3	1/2	1/6	1/3
Door 2	1/3	1	1/3	2/3
Door 3	1/3	0	0	0

Notebook 2

Let’s get some hands-on experience using Python.

Chain Rule of Probability

- What if we want to expand our Bayes calculation and calculate the probability of the intersection of many, not necessarily independent, events?

Here is where we use Chain Rule of Probability:

$$P(A, B, C, D)$$

$$= P(D \mid C, B, A)\, P(C, B, A)$$
$$= P(D \mid C, B, A)P(C \mid B, A)P(B, A)$$
$$= P(D \mid C, B, A)P(C \mid B, A)P(B|A)P(A)$$

Just to give you some physical activity:

Please calculate:

P(A,B,C,D,F)

Bayesian Networks

- Problem: Directly applying Bayes' theorem to all variables quickly becomes intractable.
- Solution: Bayesian Networks (BNs) use a graph to represent dependencies efficiently.
- Structure:
- Nodes → random variables
- Directed edges → conditional dependencies (often causal)
- Graph is acyclic (no loops)
- Semantics: Each node is associated with a Conditional Probability Distribution (CPD) given its parents.

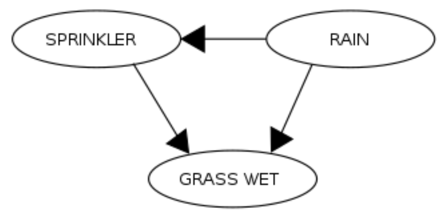
$$P(\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n}) = \prod_\mathrm{i} P(\mathrm{X}_1 | \mathrm{Pa}(\mathrm{X}_1))$$

- Inference: Given some evidence, we can update beliefs about other variables using the network.
- Motivation: Real-world systems involve many uncertain variables that influence each other.
- Interpretation: Bayesian Networks = "Bayes' theorem for many variables.

Bayesian Networks

Consider the following Bayesian network:

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

Image source: chegg.com

Notebook 3

Let’s get some hands-on experience using Python.

Recap & Quiz

joinmyquiz.com • 64793738

- Two views: frequentist vs Bayesian (both useful)
- Bayes' theorem links prior, evidence, posterior
- Bayesian networks are a convenient way to deal with more variables

Homework

- Exercise: Fill out the excercises on Jupyter notebooks 1 and 2 for this week, commit your answers and submit .
- Optional: Come up with a new example of where Frequentist/Bayesian approaches are the best.