Cognitive Modeling: Homework Assignment 1 Technical Stack

Team Members: Caleb Carr, Maryellen Marino

February 3, 2024

All answers and solutions to non-programming questions should be submitted to LMS as a **legible** write-up (either fully digital or a scan). All code should be committed to and merged into the main branch of your team's GitHub repository.

Problem 1: True-False Questions (4 points)

Mark all statements which are **FALSE**.

1. **False.** A random variable is discrete if its support is countable and there exists an associated probability density function (pdf).

Explanation: A discrete random variable is associated with a probability mass function (PMF), not a probability density function (PDF). PDFs are used for continuous random variables, where the outcomes are uncountably infinite and the probability of any single outcome is zero.

2. **True.** Probability mass functions have a lower bound of 0 and an upper bound of 1.

Explanation: Probability mass functions (PMFs) assign a probability to each outcome in the support of a discrete random variable. These probabilities must be between 0 and 1, inclusive. The total sum of these probabilities over all possible outcomes equals 1.

3. **False.** The set of all possible realizations of a random variable is called its probability density.

Explanation: The set of all possible realizations of a random variable is called its support, not its probability density. Probability density is a function that describes the likelihood of a random variable taking on a given value in the case of continuous variables.

4. **False.** The expected value of a discrete random variable is always part of its support, that is, $\mathbb{E}[X] \in R_X$.

Explanation: The expected value (or mean) does not necessarily have to be a value in the variable's support. It's possible for the expected value to be a number that the random variable never actually assumes.

5. **True.** Continuous random variables are functions which map points from the sample space to the real numbers.

Explanation: Continuous random variables are functions that map outcomes from the sample space to real numbers. They differ from discrete random variables in that they take on values in a continuous range.

6. **False.** We can formulate most parametric Bayesian models as a generative process, by which we first sample from the likelihood and then use the synthetic data point to sample from the prior.

Explanation: The main issue with the statement is the sequence of sampling: it suggests that we first sample from the likelihood and then from the prior, which is not how Bayesian models typically operate. First a prior distribution is defined, then the likelihood function, then we combine them to obtain the posterior distribution.

7. **False.** The Bayesian posterior $p(\theta|y)$ for continuous parameter vectors $\theta \in \mathbb{R}^D$ is just another density function. That means, its integral $\int p(\theta|Y = y)d\theta \neq 1$ for some y.

Explanation: The Bayesian posterior p(θ | y) is a probability density function for the parameters θ given the data y. For a valid probability density function, the integral over its entire domain must equal 1. So, the integral of the posterior distribution over all θ should be 1, not something else.

8. **True.** Each realization of a continuous random variable has a probability of zero.

Explanation: In the case of continuous random variables, any specific realization (a precise value) has a probability of zero. This is because the range of values it can take is continuous and uncountably infinite, making the probability of any single exact value effectively zero. Instead, probabilities are determined over intervals and are represented by the area under the curve of the probability density function.

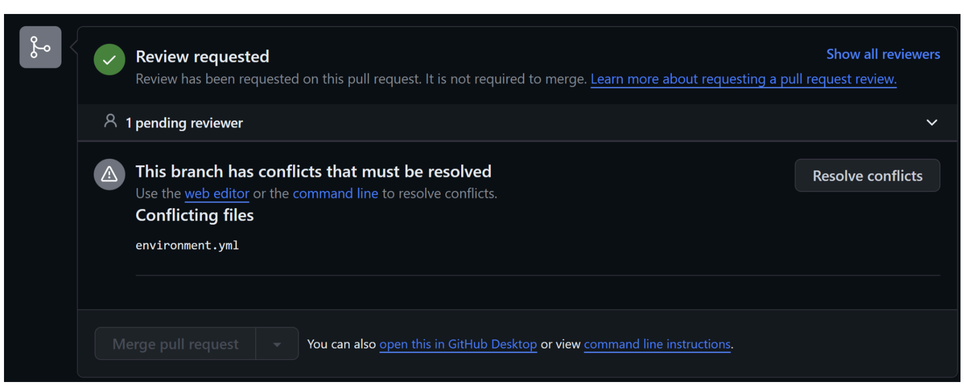
Problem 2: Git and GitHub (8 points)

1. Create a public GitHub repository, create and add a team logo to the **README** file, along with some basic introductory notes on why cognitive modeling is important for psychology and cognitive science. Create an **environment.yml** file and add all dependencies we have discussed so far. Then, in addition to the **main** branch, create separate branches for each of the two team members, from which you will be merging working code into the main branch.

GitHub Repository: https://github.com/maryellenmarino/Cognitive_ModelingS24

2. Create a *merge conflict* (either for some of the coding exercises or a mock conflict) and resolve it.

Merge conflict resolved in the following PR:
https://github.com/maryellenmarino/Cognitive_ModelingS24/pull/4



3. Explain the differences between the following git commands

   (a) **git restore**

   Purpose: `git restore` is used to restore working directory files to a specified state. It's a relatively new command introduced to split the functionality of `git checkout` into two separate commands (`git switch` and `git restore`). `git restore` focuses on changing the content of the working directory and the index (staging area).

   Example: Suppose you edited a `file.txt` and now want to discard those changes and revert it to its last committed state. You'd use:

   ```
   git restore file.txt
   ```

   (b) **git checkout**

   Purpose: `git checkout` has two main uses. Before more recent versions of git, it was used to switch branches and to restore files in the working directory to their state in a commit or branch. With more recent git, the part of switching branches has been taken over by `git switch`, and the part of restoring files by `git restore`. However, the functionality is still usable in more recent versions of git.

   Example: If you want to switch to a branch named `feature`, you'd use:

   ```
   git checkout feature
   ```

   And to restore `file.txt` to its state in the `feature` branch (in earlier versions of git):

   ```
   git checkout feature -- file .txt
   ```

   (c) **git reset**

   Purpose: `git reset` is used to reset the current HEAD to a specified state. It can affect the branch pointer, the staging area (index), and the working directory depending on the flags used (`--soft`, `--mixed`, `--hard`).

   Example: If you want to undo the last commit but keep the changes in your working directory, you'd use:

   ```
   git reset --soft HEAD~1
   ```

   If you want to undo the last commit and discard the changes:

   ```
   git reset --hard HEAD~1
   ```

   (d) **git revert**

   Purpose: `git revert` is used to create a new commit that undoes the changes made in a specific commit. This is a safe way to undo changes in a shared repository, as it doesn't rewrite commit history.

   Example: If you want to undo the changes made in commit `abc123`, but keep a record of the undoing:

```
git revert abc123
```

in terms of undoing changes to a repository by providing a minimal (actual or a synthetic) example.

Problem 3: Expectations I (4 points)

Suppose that you want to invest some money in the oil market. You believe that the probability of the market going up is 0.8 and the market going down is 0.2. Further, if the market goes up, oil prices will increase by 1%, if it goes down, oil prices will drop by 10%. What is your expectation? Would you invest in this market? Assuming the increase/drop in prices is fixed, what is the minimal probability of prices going up in order for you to invest rationally (according to expectation) in this market? Discuss a limitation of expectations when making single-shot, real-life decisions.

1. **Expected Change Calculation**:

   - **Given Data**:

     - Probability of market going up: 0.8

     - Probability of market going down: 0.2

     - Increase in prices if market goes up: 1% or 0.01

     - Decrease in prices if market goes down: -10% or -0.1

   - **Calculation**:

     - Expected change: (0.8 * 0.01) + (0.2 * -0.1) = -0.012 or -1.2%

   - **Conclusion**:

     - The expectation of a -1.2% change suggests a likely loss, making investing not advisable based on this expectation.

2. **Minimal Probability for Rational Investment**:

   - **Objective**: Find minimal probability of the market going up (p) for a neutral (zero) expected change.

   - **Equation**: 0 = p * 0.01 + (1 - p) * -0.1

   - **Solution**: Minimal probability p ≈ 0.909 or 90.9%

   - **Interpretation**:

     - For investment to be rational (according to expectation), the probability of the market going up must be at least 90.9%.

3. **Limitation of Expectations in Single-Shot Decisions**:

- Non-repeatability: Expected value calculations assume the ability to repeat an experiment (or decision) multiple times to achieve the average outcome. However, in real-life single-shot decisions, you only have one chance, and the actual outcome might be very different from the expected value.
- Impact of Outliers: In high-variance situations like the oil market, where changes can be drastic and unpredictable, relying solely on expectations can be misleading. The impact of a significant negative outcome (like a 10% drop) might outweigh the benefit of a more probable but smaller positive outcome (like a 1% increase).
- Risk Tolerance: Individual risk preferences play a significant role. Some investors might be risk-averse and unwilling to accept the possibility of a large loss for a small expected gain, even if the expected value is positive.
- Contextual Factors: The expectation calculation doesn't account for external factors such as political instability, environmental concerns, or technological changes, which can drastically alter market dynamics.
- Psychological Impact: Humans tend to react more strongly to losses than equivalent gains (loss aversion). A 10% loss might have a more significant psychological and financial impact than the benefit derived from a 1% gain, even if the latter is more probable.

Problem 4: Expectations II (4 points)

1. Show the following identity for the variance of a random variable X:

$$Var[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \qquad (1)$$

$$\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$ Use definition of variance

$$\mathbb{E}\left[X^2 - 2\mathbb{E}[X]X + \mathbb{E}[X]^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$ Expand Square

$$\mathbb{E}\left[X^2\right] - \mathbb{E}[2\mathbb{E}[X]X] + \mathbb{E}\left[\mathbb{E}[X]^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$ Use sum rule of expectations

$$\mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$ Take constants out

$$\mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$ Reqrite multiplication as squaring

$$\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$ Add like terms

2. Show the following property for the variance of a random variable X and a scalar α:

$$Var[\alpha X + \beta] = \alpha^2 Var[X] \qquad (2)$$

$$Var[\alpha X] + Var[\beta] = \alpha^2 Var[X]$$ Sum rule for variance

$$Var[\alpha X] = \alpha^2 Var[X]$$ Variance of constant is 0

$\mathbb{E}\left[\alpha^2 X^2\right] - (\mathbb{E}[\alpha X])^2 = \alpha^2\left(\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2\right)$ Rewrite variance using equivaland statement proven above

$\alpha^2\mathbb{E}\left[X^2\right] - (\alpha\mathbb{E}[X])^2 = \alpha^2\mathbb{E}\left[X^2\right] - \alpha^2\mathbb{E}[X]^2$ Take constants out of expectations

$\alpha^2\mathbb{E}\left[X^2\right] - \alpha^2\mathbb{E}[X]^2 = \alpha^2\mathbb{E}\left[X^2\right] - \alpha^2\mathbb{E}[X]^2$ square the term in paranthesis

3. Assume you are given a random variable X with a standard normal distribution (mean zero, variance one). We can write this as

$$X \sim Normal(\mu = 0, \sigma = 1) \tag{3}$$

One way to sample from this distribution is using NumPy's function **numpy.random.randn**. What transformation do you need to apply to the sampled values (i.e., the outputs of the function) such that they are now distributed according to

$$\tilde{X} \sim Normal(\mu = 3, \sigma = 5) \tag{4}$$

$\mathbb{E}[\alpha + \beta X] = \alpha + \beta\mathbb{E}[X]$ Use definition of expectation for normal distribution with linear shift

$3 = \alpha + \beta(0)$ plug in for E[X]

$3 = \alpha$ anything times 0 is zero

$Var[\alpha + \beta X] = \beta^2 Var[X]$ Use definition of variance for normal distribution with linear shift

$5 = \beta^2(1)$ plug in for Var[X]

$5 = \beta^2$ anything tumes one is itself

$\sqrt{5} = \beta$ take the square root of both sides

Problem 5: Simple Bayesian Inference (4 points)

The inhabitants of an island tell the truth one third of the time. They lie with probability 2/3. On an occasion, after one of them made a statement, you ask another "Was that statement true?" and he says "yes". What is the probability that the statement was indeed true?

$O = \{o^+, o^-\}$ where o^+ is original statement true and o^- is original statement lie

$S = \{s^+, s^-\}$ where s^+ is second statement true and s^- is second statement lie

$p(s^+|o^+) = \frac{1}{3}$ because given the original statement is true, the second person would say yes, tell the truth, one third of the time

$p(o^+) = \frac{1}{3}$ because the inhabitants tell the truth one third of the time

$p(s^+) = p(s^+|o^+) + p(s^+|s^-) = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{2}{3} = \left(\frac{1}{9} + \frac{4}{9}\right)$

If the person is telling the truth, then the probability of the second person saying true is ⅓ because they would be telling the truth. In the event the first person is lying (⅔) then the probability the second person lies and says true is ⅔

$$p\left(o^+|s^+\right) = \frac{p\left(s^+|o^+\right)p\left(o^+\right)}{p\left(s^+\right)} = \frac{\frac{1}{3}\cdot\frac{1}{3}}{\left(\frac{1}{9}+\frac{4}{9}\right)} = \frac{\frac{1}{9}}{\frac{5}{9}} = \frac{1}{5}$$

Problem 6: Murder Mystery Revised (6 points)

Construct a small story of your own choosing in the spirit of the "Murder Mystery" we considered in class (can be from a completely different domain). Define a "prior" distribution, a "likelihood", and justify your selection of probabilities. Create a first folder in your team repository and write a script containing fully vectorized code which simulates your story and outputs a table representing the approximate joint probabilities (which we approximated through the relative frequencies across *N* simulation runs). Compare this table to the analytic probabilities. How large should *N* be for the approximate probabilities to become almost indistinguishable from the analytic ones?

Solution on GitHub:

https://github.com/maryellenmarino/Cognitive_ModelingS24/blob/main/hw1/Murder_Mystery_Revised.ipynb

Problem 7: Priors, Sensitivity, Specificity (6 points)

Let's revisit the disease problem we tackled during class. Imagine you are a medical researcher analyzing the effectiveness of a new diagnostic test for a rare disease **X**. This disease affects 1% of the population. The probability of a true positive (the test correctly identifies an individual as having the disease) is 95%. This is also known as the *sensitivity* of the test. The probability of a true negative (the test correctly identifies an individual as not having the disease) is 90%. This is also known as the *specificity* of the test.

We will now consider a question of sensitivity analysis (not to be confused with the sensitivity of a test): How would the posterior probability change if the prior, the sensitivity, or the specificity of the test were to test. Write a Python program which produces three pretty and annotated 2D graphs depicting

1. The posterior probability (Y-axis) of actually having the disease given a positive test as a function of the prior probability (X-axis), assuming fixed sensitivity and specificity.
As the prior probability increases, the posterior increases logarithmicly. It only takes about 15% of the population to have the disease for there to be a 50% probability of actually having the disease given a positive test.
2. The posterior probability (Y-axis) of actually having the disease given a positive test as a function of the test's sensitivity (X-axis), assuming fixed prior and specificity.
As the sensitivity increases, the posterior increases linearly. It is a 1:1 ratio of sensitivity of test to posterior probability.
3. The posterior probability (Y-axis) of actually having the disease given a positive test as a function of the test's specificity (X-axis), assuming fixed prior and sensitivity.
As the specificity increases, the posterior increases expontntially. It only takes a whopping 99% of getting a true negative for there to be a 50% probability of actually having the disease given a positive test.

Briefly discuss how the posterior changes as a function of each of the quantities.

Bonus (4 points) Generate three 3D plots (either surface plots or scatter plots) depicting the same posterior probability as a function of the combination of two quantities (prior - sensitivity, prior - specificity, sensitivity - specificity).

Solution on GitHub:
https://github.com/maryellenmarino/Cognitive_ModelingS24/blob/main/hw1/Priors_Sensitivity_Specificity.ipynb

Problem 8: Monte Carlo Approximation (4 points)

Write a Python program that approximates the value of π using Monte Carlo approximation. Your program should generate a sequence of random points and use these points to estimate the value of π. The accuracy of the approximation should improve as the number of points increases. Here are some hints:

- Your program should generate random points with x and y coordinates ranging between -1 and 1. This will simulate points within a 2×2 square that circumscribes a unit circle centered at the origin $(0, 0)$.
- For each generated point, determine whether it falls inside the unit circle. A point $p = (x, y)$ is inside the circle if $x^2 + y^2 \leq 1$.
- Use the ratio of the number of points that fall inside the circle to the total number of generated points to approximate π. The formula is given by $\pi \approx 4 \times$ (number of points inside/total number of points).

Plot the approximation error of your Monte Carlo estimator as a function of the total number of points N. You can use np.pi as the ground truth.

Solution on GitHub:

https://github.com/maryellenmarino/Cognitive_ModelingS24/blob/main/hw1/Monte_Carlo_Approximation.ipynb

Problem 9: AI-Assisted Programming (4 points)

Use ChatGPT or any other large language model (LLM) to generate a function called **multivariate normal density(x, mu, Sigma)** which returns the density of a D-dimensional vector x given a D-dimensional mean (location) vector

mu and a $D \times D$-dimensional covariance matrix Cov. Compare the outputs of your

function with those obtained using SciPy's **scipy.stats.multivariate_normal**

for a few parameterizations including a spherical Gaussian (zero covariance, shared variance across dimensions), a diagonal Gaussian (zero covariance, different variance for each dimension), and a full-covariance Gaussian (non-zero covariance, different variance for each dimension). Describe briefly how the LLM performed.

Solution on GitHub:
https://github.com/maryellenmarino/Cognitive_ModelingS24/blob/main/hw1/AI_Assisted_Programming.ipynb