Computational Cognitive Science
9.66/9.660/6.804J
Problem Set 1
Fall 2023

1. Coin flipping

In lecture, we saw an example of a hypothesis test between a fair coin and a coin that always lands heads. In this problem you will experiment with the more general case of testing between the hypothesis that a coin is fair*, landing heads half the time and tails half the time, or that it is* unfair*, landing on one side more often than the other. This problem is intended to give you a taste of both the modeling and experimental sides of computational cognitive science in a simple but intuitive setting.*

(a) To start, you will explore human intuitions about random processes in more detail by conducting a mini-experiment on coin flipping described on the following page. We will ask you to run the experiment twice, in two conditions with slightly different instructions, so you will need to find two participants who are not in this class (*e.g.*, friends or roommates). You should also run both conditions on yourself. This will provide you with four datasets. Run the experiment on yourself first, before you have become too familiar with it or with the model. In total, the experiment should take fewer than ten minutes for each participant.

Conduct the experiment using the materials on the following page. The *cover story* below is intended to familiarize participants with the experimental procedure and to establish assumptions about the process that generated the stimuli. It should be read first. As discussed in class, the cover story plays a crucial role in setting up people's *priors*. If you tell people that a coin is from a magic store, they will probably assume it is more likely to be unfair than if the coin was obtained from a bank or a cash register. We will start off using a generic cover story and ask you to modify it later.

In the first condition of the experiment, present the attached instructions and stimuli to the participant. After the instructions have been read, the participant should rate how likely each flip sequence is to have been generated by a fair coin or an unfair coin, using the rating scale provided.

For the second condition, find a new participant. Re-run the experiment after modifying the cover story to induce a different prior over the fair and unfair hypotheses. This cover story should introduce a context within which unfair coins are either more or less probable *a priori* (*e.g.* the magic shop context above).

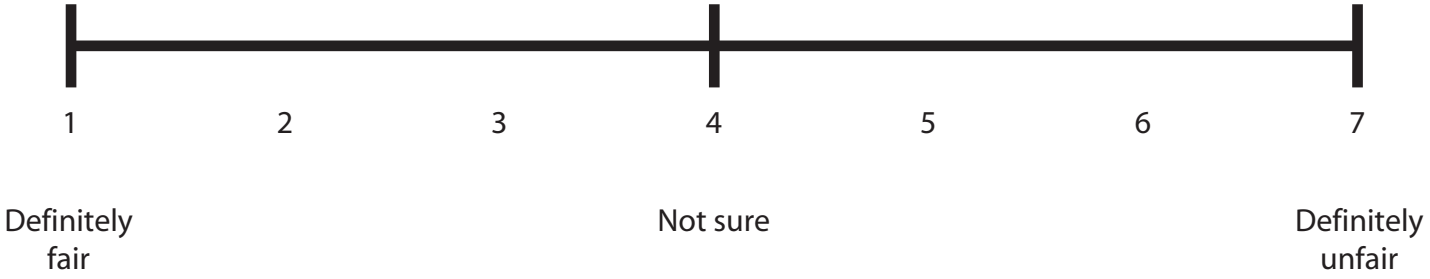
(i) Attach your modified cover story and all four data sets. What effect did you anticipate from the two cover stories you used?
(ii) Do you see any systematic difference in the ratings between the two conditions of the experiment?
(iii) Describe how the differences in the data across the conditions or lack thereof compared to your expectations.

One day you find a bag of strange-looking coins lying on the sidewalk. They have recognizable "heads" and "non-heads" (or "tails") sides, although they do not look like any coins you have ever seen before; perhaps they are from a foreign country? Always curious about coins, you decide to flip each coin a few times, and you observe the sequences of outcomes shown below. Each sequence was generated by flipping a *different* coin.

For each of the following sequences, please judge how likely you think the coin is to be a fair coin (tends to land heads half the time and tails half the time) or an unfair coin (tends to land on one side more often than the other). Use the 1-7 rating scale given below. Keep in mind that each sequence was generated by a different coin, so try not to let the information about one coin affect your judgments about another coin.

Sequences:

- Coin 1: H H T H T
- Coin 2: T H T T T
- Coin 3: H H H H H
- Coin 4: T H T T H T H T H T
- Coin 5: H H T H H H H H T H
- Coin 6: T T T T T T T T T T
- Coin 7: T H T T H T T H H T H T H T T H T H T T T H T T H T
- Coin 8: H H T H H H H T H H H T H H H T H H H H H T H H H H
- Coin 9: H H H H H H H H H H H H H H H H H H H H H H H H H H



(b) To model people's responses to this experiment, compare the following hypotheses:

- H_1: "fair coin", $P(\text{Heads}|H_1) = 0.5$. In this case, the probability of a sequence given H_1 only depends on the length of the sequence, H + T, because heads and tails are equally likely:

$$P(\mathcal{D}|H_1) = \frac{1}{2^{\text{H+T}}}.$$

- H_2: "weighted coin", $P(\text{Heads}|\theta) = \theta$; $p(\theta|H_2) = \text{Uniform}(0, 1)$. Computing $P(\mathcal{D}|H_2)$ requires marginalizing over the unknown coin weight θ:

$$P(\mathcal{D}|H_2) = \int_0^1 P(\mathcal{D}|\theta)P(\theta|H_2)d\theta.$$

  Later in the course, we will solve this integral analytically. For now, compute a discrete approximation (you can compute this in a language of your choice):

$$P(\mathcal{D}|H_2) \approx \sum_{n=1}^{100} P(\mathcal{D}|\theta_n)P(\theta_n|H_2)$$

To test between these hypotheses, use the log posterior odds ratio:

$$\log \frac{P(H_1|\mathcal{D})}{P(H_2|\mathcal{D})} = \log \frac{P(\mathcal{D}|H_1)}{P(\mathcal{D}|H_2)} + \log \frac{P(H_1)}{P(H_2)}.$$

Compute the log posterior odds ratio for each of the above coin flip sequences, assuming $P(H_1)/P(H_2) = 1$. To compare people's judgments to the models, we need to transform the log posterior odds ratio to the 7-point scale of the human data. Pass the log posterior odds ratio through a logistic function:

$$f(x) = \frac{1}{1 + \exp(-ax + b)},$$

where a and b are free parameters that you can tweak to fit the human data to the model predictions. Note that if $a = 1$ and $b = 0$ then the transformed value is just:

$$\frac{1}{1 + \exp(-\log \frac{P(H_1|\mathcal{D})}{P(H_2|\mathcal{D})})} = \frac{P(H_1|\mathcal{D})}{P(H_1|\mathcal{D}) + P(H_2|\mathcal{D})}.$$

The logistic transformation will transform the model predictions to a 0 to 1 scale; you can then scale the transformed model predictions appropriately.

(i) Plot the transformed model predictions against the human data and report a correlation (use `np.corrcoef` in Python or `corrcoef` in MATLAB) for each cover story. (ii) What settings of a and b seem to work best (you don't need to explicitly search for the best a and b; just try a few)? (iii) How did you assess goodness of fit for a and b? (iv) How well does the model qualitatively capture people's judgments in each condition? Are there any systematic differences between people and the model?

(c) One reason why the model might deviate from a participant's judgments is that we assumed equal priors: $P(H_1)/P(H_2) = 1$.
(i) What would the effect be of varying $P(H_1)$ on the model predictions (remember that $P(H_2) = 1 - P(H_1)$) and why? (ii) Can you draw any conclusions about which values of $P(H_1)$ fit your participants' judgments best in each condition?

(d) (i) What does the hypothesis space $\{H_1, H_2\}$ *not* capture about people's intuitions? (ii) Give two examples of coin flip sequences where the hypothesis test above will fail to predict human judgments.

2. The Number Game

In this problem, you will recapitulate some of the experiments and computations from the Number Game, an induction experiment described in Rules and Similarity in Concept Learning *(Tenenbaum,* NIPS *2000). The goals of this problem are to give you further exposure to the challenges involved in modeling a cognitive experiment and some exposure to Bayesian inference in discretely structured hypothesis spaces.*

Recall the setup from lecture or the paper. In each round of the Number Game, the computer selects some subset C from the 2^{100} subsets of the positive integers from 1 to 100. The computer then presents the subject with a sequence of randomly chosen members of that set, and (after each new number is presented) asks the subject to rate how likely they think various other numbers are to be in the set.

Tenenbaum modeled subjects' responses – their stated confidence that some number y is in C – as the posterior probability of that proposition under a simple model with a prior on a restricted space of hypotheses (including, for example, intervals, "odd numbers", "multiples of ten", etc) and a likelihood based on strong sampling with replacement from the set. In this problem set, we'll use a simple hypothesis space that includes two equally likely types of hypotheses: intervals and multiples-of-k (for integer k). Within a hypothesis type, we'll place a uniform prior over all hypotheses of that type. Lastly, the likelihood of an observed number y is $P(y|h) = \frac{1}{|h|}$ if $y \in h$ and 0 otherwise. Because we assume elements are sampled from a set independently, the probability of a set of numbers is just the product of the probabilities of individual members.

As a reminder, the probability that a number y is in the concept can given examples D can be written as the following:

$$P(y \in C|D) = \sum_{h \in H} P(y \in C|h)P(h|D)$$

(a) Manually compute the posterior probabilities of the hypotheses "all multiples of 10" and "all even numbers" given the data `10 70 30` (assuming those two are the only hypotheses). Show your work.

(b) Manually compute the probability the concept contains the number 40 given the data `10 70 30`. (Hint: This should be a simple calculation, the results from part (a) only trivially affect calculation here.)

(c) Write code to compute the log likelihood of a given dataset under a given hypothesis. You may use any programming language, but we have included Python and MATLAB function templates `number_game_likelihood` which can be completed. Using one of these templates will allow you to automate plotting in the following problem. (Note: If you are using the provided Python or MATLAB code, data is represented by binary vectors. This means that you cannot represent number sequences with multiple instances of the same number e.g. $[10, 10, 20]$. If you would like to do this, you will need to write your own code, otherwise stick to sequences of unique numbers). Attach your code to this report.

(d) If you implemented the Python or MATLAB function `number_game_likelihood` in the previous question, plots can be automatically generated using the function `number_game_plot_predictions(hypotheses, priors, data)`. You may construct `hypotheses` and `priors` using the provided Python or MATLAB

functions `number_game_simple_init(N, interval_prior_mass, math_prior_mass)` which initializes a hypothesis space and prior over interval and mathematical concepts on integers between 1 and N.

(i) Generate plots showing the predictive distribution for the dataset `[60, 52, 57, 55]` and one of your own choosing. (ii) Generate plots in sequence for `[80]`, `[80, 10]`, `[80, 10, 60]`, and `[80, 10, 60, 30]` demonstrating how new data changes the predictive distribution.

(iii) Experiment with three alternative settings of the prior, and show how the patterns of generalization change. Discuss and explain the effect of varying the prior.

(iv) Which settings for the prior best capture the human data? Explain what this implies. (Human data – average subject ratings – will appear as the second, labeled plot whenever the dataset corresponds to one which people were asked about.)

(e) (i) How do Marr's levels apply to the number game? For instance, what level of explanation does it aim for? What aspects of human concept learning does or doesn?t it capture?

(ii) Do you think the number game is an *ecologically relevant* task to study in cognitive psychology (*i.e.* Does it give us intuitions about how human cognition works outside of the lab?)?

(iii) If people play the number game by considering a hypothesis space similar to this one, where might this hypothesis space come from? How might it differ from the one above?

If you find the Number Game interesting, it could be extended, revised, or varied to produce a good final project, which could even lead to a publishable contribution. For example, one might try to use a richer hypothesis space allowing exceptions (e.g., "all multiples of 10 except 70") or combinations of basic hypotheses (e.g., "all multiples of 10 between 30 and 80"). The machinery in this paper could be useful:

A rational analysis of rule-based concept learning. N. D. Goodman, T. L. Griffiths, J. Feldman, and J. B. Tenenbaum (2007). Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society. http://web.mit.edu/cocosci/Papers/RRfinal3.pdf (Link valid as of September 2018)

Another possibility would be to attempt to explain individual differences in subjects' responses, perhaps via different hypothesis spaces, different priors, or different ways of approximating the large sums over hypotheses necessary for Bayesian generalization. Feel free to talk to us about these options. We will also discuss some of them in a few weeks when it comes time to formulate a project proposal.