Computational Cognitive Science
9.66/9.660/6.804J
Problem Set 1
Fall 2021

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 `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 provided a MATLAB function template `number_game_likelihood` which can be filled out. Using this template will allow you to automate plotting in the following problem. (Note: If you are using the 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 the MATLAB function `number_game_likelihood` was implemented in the previous question, plots can be automatically generated using `number_game_plot_predictions(hypotheses, priors, data)`. You may construct `hypotheses` and `priors` using the provided MATLAB function `[hypotheses priors] =`

`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.

3. Optimal Predictions in Everyday Cognition

In this problem, you will recapitulate some of the computations from Optimal Predictions in Everyday Cognition *(Griffiths & Tenenbaum,* Psychological Science*, 2006). The goals of the problem are to get your hands dirty making some basic Bayesian inferences and decisions (both analytically and numerically) in a simple but ecologically relevant context. We hope this problem will take you no more than 6 hours, though we expect that if you are comfortable with MATLAB, 18.01 level calculus and basic probability it should take you roughly 2 hours.*

Recall (from lecture and the paper) the basic setup for *Optimal Predictions.* An observer gets a sample t_{obs} from an interval of (unobserved) length t_{total}, drawn uniformly at random (*i.e.* nothing is special about the time of observation), so $p(t_{obs}|t_{total})$ is $\frac{1}{t_{total}}$ if $0 < t_{obs} < t_{total}$ and 0 otherwise (This is called an *anthropic* likelihood). The observer has background domain knowledge about which values of t_{total} are most likely *a priori*, captured in $p(t_{total})$. The observer is then asked to make an estimate of t_{total} based on this information. This estimate is modeled as the median of the posterior distribution

$$p(t_{total}|t_{obs}) = \frac{p(t_{total})p(t_{obs}|t_{total})}{p(t_{obs})}$$

where

$$p(t_{obs}) = \int_0^\infty p(t_{obs}|t_{total})p(t_{total})dt_{total}.$$

The posterior median is the value t^* such that half the cumulative mass of the posterior $p(t_{total}|t_{obs})$ is less than t^*, and half is greater than t^*; it is defined more formally below. We will compare the posterior median to the median of a sample of human subjects' responses.

Note that throughout this problem we will use the term *joint density* or *joint* to describe $P(H, D) = P(H)P(D|H) \propto P(H|D)$; this is consistent with standard use in Bayesian statistics. Also note that there are two integrals involved here: one in normalizing the joint density to compute the posterior and one in solving for the posterior median. The first integral is about *inference* - determining what one should believe about an unknown quantity given prior beliefs and data - while the second is about *decision making* - determining what estimate one should emit (or act on) to maximize some measure of performance. Both inference and decision making are distinct and important subproblems in rational action under uncertainty.[1]

Also note that throughout this problem set, we will freely use $p(x)$ and $Pr[X = x]$ to denote the probability density of a random variable X evaluated at value x (in some sense, 'as if' all densities were actually over fine discretizations, and therefore commensurable with probabilities of statements).

[1]The posterior median turns out to be one of several reasonable decision procedures in this setting. A possible extension of this work could consider other decision rules and compare them with other measures of human responses besides the sample median, possibly accounting for individual differences. Feel free to ask the instructor or TAs about this possibility.

(a) Analytically determine the posterior distribution $p(t_{total}|t_{obs})$ under a power-law prior distribution with exponent γ (i.e. $p(t_{total}) \propto t_{total}^{-\gamma}$ where for convenience we will assume $\gamma \geq 1$).

(b) Analytically determine the median of the posterior distribution: t^* *s.t.* $Pr[t^* > t_{total}|t_{obs}] = 0.5$) as a function of t_{obs} and γ. Show all your work. Hint: Do this in two steps. First, write $Pr[t^* > t_{total}|t_{obs}]$ as a function of t^* and t_{obs} (this involves integrating the posterior you computed in part a). Second, set that expression equal to 0.5 and solve for t^* as a function of t_{obs}. Also submit a plot of the posterior median (as a function of t_{obs}) over a reasonable range of observed timespans t_{obs}. Use $\gamma = 2.43$ for your plots and for below.

Now we will implement the same calculations numerically. This will allow us to apply a similar analysis to cases where the posterior doesn't admit a simple closed-form expression. It will also give you some experience comparing analytical to numerical calculations, a theme that may return later on in class when we study algorithmic issues of inference.

(c) In whatever language you choose, write a procedure according to the following MATLAB specification:

```
[thetavals postvals] = opt_compute_posterior(joint, theta_min, theta_max, num_steps)
```

where `thetavals` is a vector of theta (parameter) values. `postvals` is a vector of normalized posterior density values, such that `postvals(i)`$= Pr[H =$ `thetavals(i)`$|D = d]$. The first argument to `opt_compute_posterior` is a procedure `joint_density = joint(theta)` which evaluates $Pr[H = \theta, D = d]$ for some particular, pre-wired-in data value d.

`opt_compute_posterior` should work by forming a reasonable discrete approximation to

$$Z = Pr[D = d] = \int_{\theta_{min}}^{\theta_{max}} Pr[H = \theta, D = d]d\theta$$

(e.g. by a Riemann sum with `num_steps` rectangular elements, or possibly Simpson's rule[2], then returning a table with ordered entries ranging from θ_{min} to θ_{max} with values equal to `joint(theta)`$/Z$. Attach the code for your `opt_compute_posterior` function.

(d) (i) Write a procedure `opt_build_powerlaw_joint(t)` that takes as an argument the time t, where t is the time of the observed data, and returns a function, `joint(theta)`, that is suitable as the first argument of `opt_compute_posterior`. That is, `opt_build_powerlaw_joint` should return an anonymous function (function pointer) which is a function of `theta`. When evaluated on `theta` it should return the joint density of `theta` and t under the power-law prior and the anthropic likelihood. Attach your code.

[2]People with numerical analysis inclinations are welcome to learn about and use predictor-corrector methods, etc, being careful of the step in the likelihood. Do **not** use a Monte Carlo scheme unless you can justify it in this 1D context - and if you can, please let the TAs know how, as there might be a paper in it. For people not so inclined, a simple left-aligned Riemann sum will get full credit. A well written version of this procedure could be useful in your future lives as Bayesians, or as a memory of better days should you turn frequentist.

(e) Write a procedure `joint = opt_build_lifespan_joint(t)` (analogous to `opt_build_powerlaw_joint` which returns a procedure, of a form suitable as an argument to `opt_compute_posterior`, that computes the joint density of a timespan `t` under a Gaussian prior (with mean 75 and standard deviation 16). Attach your code.

(f) `opt_predictions_plot(@integrating_func, @build_joint_func, theta_min, theta_max)`, where the first argument can take `opt_compute_posterior` and the second argument can take `opt_build_powerlaw_joint`, should construct plots showing the posteriors and predictive medians with `theta_min = 0` and `theta_max = 300`.

Use your procedures from (c) - (e) along with `opt_prediction_plots` to generate plots showing posteriors and predictive medians for a range of observed lifespans. Submit these plots and comment on how the way the prediction changes as the observation approaches the mean expected lifespan compares to your own intuitions.