Exercises: Probabilistic models of cognition

Before starting these exercises please clone the `chdss2018` GitHub repository by typing:

```
git clone https://github.com/cskemp/chdss2018.git
```

at the command line, or by using the "clone or download" button at `https://github.com/cskemp/chdss2018`

1 Coughing friend

The script for this exercise is `day3_probmodels/coughingfriend/coughingfriend.R` in the `chdss2018`
The prior in the script indicates that $P(\text{cold}) = 0.46$, $P(\text{emphysema}) = 0.04$, and $P(\text{stomach upset}) = 0.05$.
The likelihood in the script indicates that $P(\text{coughs}|\text{cold}) = 0.4$, $P(\text{coughs}|\text{emphysema}) = 0.4$, and $P(\text{coughs}|\text{stomach upset}) = 0.05$.

1. Given the prior and likelihood specified above, use Bayes rule to calculate $P(\text{cold}|\text{coughs})$ by hand.

2. Run the script to generate plots showing the prior, the likelihood and the posterior. Use the plot to check your calculation from the previous question.

3. The prior and posterior distributions sum to 1, but the likelihood function does not. Should it? Why or why not.

4. Change the code so that the prior is uniform. Generate the plots again — does the posterior match your intuitions about what should happen if the prior were truly uniform?

5. Change the code to use the original prior but adjust the likelihood so that coughing is equally probable for each disease (e.g. $P(\text{coughs}|\text{cold}) = 0.4$, $P(\text{coughs}|\text{emphysema}) = 0.4$, and $P(\text{coughs}|\text{stomach upset}) = 0.4$). Generate the plots again—does the posterior match your intuitions about what should happen if the likelihood were truly uniform?

6. (ADVANCED) What if John might be healthy (ie John might have none of the diseases)? Change the model accordingly.

7. (ADVANCED) What if John might be healthy, but might also have multiple diseases? Change the model accordingly.

2 The Number Game

The script for this exercise is `day3_probmodels/numbergame/numbergame.R`

1. Run the script to generate plots showing generalizations for 4 models that represent all possible combinations of two priors (maths prior and uniform prior) and two likelihoods (strong and weak sampling). The plots show generalizations after the models observe three positive examples (4, 8 and 12). Compare the generalizations of the two models in the left column (strong maths and weak maths). Does the likelihood (strong vs sampling) make a qualitative difference? Explain any difference that you observe.

2. Compare the generalizations of the two models in the top row (strong maths and strong uniform). Does the prior make a qualitative difference? Explain any difference that you observe.

3. Try out some different sets of observations and see what generalizations the models with a uniform prior make (i.e. models in the right column). How would you describe the way in which the model generalizes to unobserved examples?

4. (ADVANCED) Consider a combined prior that is a weighted average of the uniform prior and the maths prior. This combined prior assigns some probability to all possible hypotheses but higher probability to the mathematical hypotheses. What kinds of inferences would this combined prior lead to? Adjust the code to add strong combined and weak combined models to the four that are already there.

5. (ADVANCED) The code does not include interval-based hypotheses (e.g. numbers between 2 and 5) like those considered by Tenenbaum. Implement a model with an interval-based prior. As a next step, implement a model with a prior based on a weighted combination of the maths prior, the interval-based prior, and the uniform prior.

6. (EVEN MORE ADVANCED) We specified the mathematical prior by writing down a "laundry list" of hypotheses. This seems a bit unsatisfying – can you think of a better way to specify a prior over mathematical concepts?

3 The Foodweb Problem: Inference by Enumeration

The script for this exercise is `day3_probmodels/foodweb/foodweb.R`. The script implements two inference methods:

1. Inference by enumerating the entire hypothesis space

2. Inference by sampling hypotheses from the prior

For this exercise set the flag at the top of `foodweb.R` to

```
inferencemethod <- "enumerate"
```

1. The function `p_h()` computes the prior probability of a hypothesis by multiplying 7 terms together. Make sure you understand what these seven terms are and why they are there.

2. Run the script to compute generalizations over the foodweb given that kelp does not have the disease but makos do. You should see that sandsharks are less likely to have the disease than dolphins and tuna – why?

3. Suppose that the observation vector is empty: i.e.

   ```
   obs <- list()
   ```

   What generalizations do you now expect? If you run the script you should see that makos are more likely to have the disease than kelp – why? You should also see that humans are less likely to have the disease than makos – why?

4. Suppose now that the observation vector specifies only that kelp have the disease: i.e.

   ```
   obs <- list(kelp = 2)
   ```

   Derive by hand the probability that herring have the disease. Check your answer by running the script.

5. Try playing around with the base rate and transmission rate parameters. Do the model's generalizations change as you might expect?

6. (ADVANCED) Suppose now that the observation vector specifies only that herring have the disease: i.e.

   ```
   obs <- list(herring = 2)
   ```

   Derive by hand the probability that kelp have the disease. Check your answer by running the script.

7. (MORE ADVANCED) The likelihood function in the script (`p_obs_given_h()`) assumes weak sampling. Implement a likelihood function that assumes strong sampling and try it out on some observation vectors that include only positive examples: e.g.

   ```
   obs <- list(kelp = 2, mako = 2)
   ```

8. (MORE ADVANCED) How could the model be adjusted to remove the assumption that the base rate and ttransmission rate are known in advance?

4 The Foodweb Problem: Inference by sampling from the prior

For this exercise set the flag at the top of `foodweb.R` to

```
inferencemethod <- "sample"
```

1. The function `sample_h()` samples a hypothesis h from the prior $P(h)$. First the value for kelp is sampled, then herring, and so on. Why do the values need to be sampled in this order?

2. After sampling 1000 samples the code only uses those that are consistent with the observation vector. Why?

3. Set the observation vector to

   ```
   obs <- list(kelp = 1, mako = 2)
   ```

   and run the script. Compare the output to what you saw when you computed the same set of predictions using enumeration as the inference method.

5 The Foodweb Problem: Inference by sampling from the posterior using JAGS

In this exercise we'll use a program called JAGS to sample directly from the posterior $P(h|obs)$.

The script for this exercise is `day3_probmodels/foodweb/foodweb_jags.R`. This script relies on a file called `foodweb.bug` which specifies the model of interest.

- Look at `foodweb.bug` – see if you can understand how this file specifies a distribution over all 7 variables in the model (ie all 7 species in the foodweb).
- Set the observation vector to

  ```
  obs <- list(kelp = 1, mako = 2)
  ```

  and run the script. Compare the output to what you saw when you computed the same set of predictions using enumeration as the inference method.

6 Sampling Frames: Spheres of Sodor

The files for this section are in the `day3_probmodels/samplingframes/model/code` directory of the `chdss2018` repository.

Throughout we'll assume that there are seven categories that include spheres of seven different sizes. The mean of category c_i is the proportion of spheres in that category that have a plaxium coating.

1. Run the `samplefromprior.R` script a few times to generate samples from the prior over category means. Does this prior capture the intuitive expectations that you bring to the spheres of Sodor problem? For example, the prior assigns some probability to U-shaped curves and other curves with one or more turning points: is this reasonable?

2. Just as we did for the foodweb problem, we could compute model predictions by sampling from the prior. But instead we'll compute predictions using JAGS. Run `simulations.R` to generate six sets of samples then run `plotsims.R` to plot the results. Compare the model generalizations to the analogous figure showing human generalizations from yesterday.

   `category.bug` and `property.bug` specify models that assume category sampling and property sampling respectively. Look first at `category.bug`. The model specifies how a vector of category means is generated, and how the actual observations (i.e. whether each observed rock has a plaxium coating) are generated given these category means. Make sure you understand the line

```
plaxium[i] ~ dbern(category_means[category[i]])
```

where `plaxium[i]` is the variable that indicates whether the ith observation has a plaxium coating.

3. (ADVANCED) See if you can understand the model specification in `property.bug`. In particular look at the lines

```
denom <- sum(category_means * base)
for(i in 1:ncat) {
  pcat[i] <- (category_means[i] * base[i]) / denom
}
```

   where `base` includes base rates of the categories, and `pcat[i]` indicates the probability that a sample from all plaxium-coated objects will belong to category `i` (c_i). This section implements the idea that

$$P(\text{item in } c_i|\text{item has plaxium}) \propto P(\text{item has plaxium}|\text{item in } c_i)P(\text{item in } c_i)$$

4. (MORE ADVANCED) Yesterday Dani showed that there are lots of interesting individual differences in the data set. How could the model be adapted to allow for individual differences?