Homework 5

Due: 1 Dec 2023
(late homeworks penalized 10% per day)

See the course web site for submission details. For each problem, show your work - if you only provide the answer, and it is wrong, then there is no way to assign partial credit! And, please don't procrastinate until the day before the due date... *start now*!

1. **Comparing two estimators**. A common method of estimating the size of biological populations is the "capture-mark-recapture" method. One proceeds by repeatedly capturing animals, putting a marker on them, and releasing them. After marking M animals, you then capture a new group of C animals and find that K of them are tagged with your mark. We will treat each of the C 2nd-round captures as *independent* of the other 2nd-round captures, i.e., as if you caught each animal and then re-released it immediately ("sampling with replacement"). If the full population size is N, then the proportion of the population that is marked is M/N and thus the expected value of the proportion marked in your 2nd sample (which is K/C) should be the same as the population proportion. Thus, an estimate of the population size is $\hat{N} = MC/K$.

    (a) Check whether $\hat{N}$ is a maximum likelihood estimator. For a few triplets $\{K, C, M\}$, plot the likelihood $L(N) = p(K|C, M, N)$ for a range of values of N (e.g., from $\hat{N}-5$ to $\hat{N}+5$. For example, try $\{K, C, M\} = \{10, 200, 4000\}$. Note that N has to be an integer, so note whether $\hat{N}$ should be rounded or truncated to the nearest lower integer to maximize likelihood, or whether your results suggest a consistent pattern with regard to the non-integer $\hat{N}$.

    (b) Check whether the estimator is unbiased. For a few triplets $\{C, M, N\}$, simulate the 2nd capture value K 1,000 times (i.e., generate appropriately distributed samples of K), and commpute the population estimate from each. Is the mean of thee population estimates close to the correct answer?

    (c) Write the precise distribution for samples K (when $\{C, M, N\}$ are known and fixed), so you can compute the exact probability distribution of estimates $\hat{N}$. Do that for $N = 1000$, $M = 100$, $C = 100$. Plot the distribution and compute its mean and standard deviation. Does thi calculation indicate that the estimator i biased? (Note: do this for the non-integer value of $\hat{N}$).

    (d) Some authors have suggested an alternative estimator: $\hat{N}' = \lfloor (M+1)(C+1)/(K+1) \rfloor$, where $\lfloor \cdot \rfloor$ indicates truncation to the next lower integer. Repeat part (c) for this estimator and compare the bias and variance of this estimator to the original one.

2. **Bayesian estimation.** Xinyuan and Sarah are looking for Ajay in a very large one-dimensional shopping mall. Location is specified by a coordinate X. They know that, all else being equal, Ajay likes to hang out near the center of the shopping mall. Specifically, the probability distribution of his location is Gaussian with mean $X = 75$ and variance 35. The only clue they have is a coffee cup at location X=40, containing the residue of a coffee that only Math Tools TAs drink (a naturally dried, triple-caffeine Ethiopian heirloom varietal, hand-extracted).

Given the location of the coffee cup, and the aroma and dryness of the coffee residue, they estimate the likelihood of his position, i.e., the probability of finding a cup at that location in that condition given Ajay's current position, to be a Gaussian with mean $X = 40$ and variance 80.

(a) Frame this problem as a problem in Bayesian estimation, using appropriate terminology. What is Ajay's posterior distribution? Draw his prior distribution, likelihood function and posterior distribution on a single plot. (Rather than `normpdf`, compute the probabilities from the formula for the Gaussian distribution.) What are the mean and variance of the posterior?

(b) Xinyuan and Sarah realize they over-estimated the sitting time of the coffee residue, and decide that Ajay's likelihood function has mean $X = 40$ but with a much smaller variance of 30. Redo part (a), and describe what happened to the posterior distribution, in terms of mean and variance. Does the change make sense?

(c) What would the posterior distribution in (a) be if the prior was nearly flat (e.g., variance 10^6). Compare this variance to that of the posterior in (a). How does the inclusion of prior information affect the variance?

3. **Bayesian inference of binomial proportions.** Poldrack (2006) published an influential attack on the practice of "reverse inference" in fMRI studies, i.e., inferring that a cognitive process was engaged on the basis of activation in some area. For instance, if Broca's area was found to be activated using standard fMRI statistical-contrast techniques, researchers might infer that the subjects were using language. In a search of the literature, Poldrack found that Broca's area was reported activated in 90 out of 840 fMRI contrasts involving engagement of language, but this area was also active in 215 out of 2754 contrasts not involving language.

(a) Assume that the conditional probability of activation given language, as well as that of activation given no language, each follow a Bernoulli distribution (i.e., like coin-flipping), with parameters x_l and x_{nl}. Compute the likelihoods of these parameters, given Poldrack's observed frequencies of activation. Compute these functions at the values `x=[0:.001:1]` and plot them as a bar chart.

(b) Find the value of x that maximizes each discretized likelihood function. Compare these to the exact maximum likelihood estimates given by the formula for the ML estimator of a Bernoulli probability.

(c) Using the likelihood functions computed for discrete x, compute and plot the discrete posterior distributions $P(x \mid data)$ and the associated cumulative distributions $P(X \leq x \mid data)$ for both processes. For this, assume a uniform prior $P(x) \propto 1$ and note that it will be necessary to compute (rather than ignore) the normalizing constant for Bayes' rule. Use the cumulative distributions to compute (discrete approximations to) upper and lower 95% confidence bounds on each proportion.

(d) *Are these frequencies different from one another?* Consider the joint posterior distribution over x_l and x_{nl}, the Bernoulli probability parameters for the language and non-language contrasts. Given that these two frequencies are independent, the (discrete) joint distribution is given by the outer product of the two marginals. Plot it (with `imagesc`). Compute (by summing the appropriate entries in the joint distribution) the posterior probabilities that $x_l > x_{nl}$ and, conversely, that $x_l \leq x_{nl}$.

(e) *Is this difference sufficient to support reverse inference?* Compute the probability $P(language \mid activation)$. This is the probability that observing activation in Broca's area implies engagement of language processes. To do this use the estimates from part (b) as the rel-

evant conditional probabilities, and assuming the prior that a contrast engages language, $P(language) = 0.5$. Hint: To calculate this probability, you will need to "marginalize", i.e., integrate over the unknown values of x_l and x_{nl}. Poldrack's critique said that we cannot simply conclude that activation in a given area indicates that a cognitive process was engaged without computing the posterior probability. Is this critique correct? To answer this, compute the posterior odds $\left(\frac{p(\text{language}|\text{activation})}{p(\text{not language}|\text{activation})}\right)$ using the maximum-likelihood estimates of x_l and x_{nl} from Poldrack's data of activation probabilities and compare the posterior odds to the prior odds before running your experiment $\left(\frac{p(\text{language})}{p(\text{not language})}\right)$.

4. **Signal Detection Theory.** Consider an experiment where a moving-dot visual stimulus is presented to a subject. The difficulty of detecting the motion is varied by changing the *coherence* of the moving dots, which is the fraction of dots moving to the right (at zero coherence, the dots move randomly, and at 100% coherence, all of the dots move to the right). Suppose we want to decide whether the stimulus is random or is moving to the right, based on the response of a single neuron that fires at a random rate, whose mean is 3 spikes/s in response to a 0% coherence noisy stimulus and 5 spikes/s for 10% coherence. Suppose also that the distribution of firing rates is Gaussian with a standard deviation of 1 spikes/s for both stimuli.

   (a) For the "no coherence"' stimulus, generate 1000 trials of the firing rate of the neuron in response to these stimuli (i.e., draw 1000 random samples from a Gaussian with $\mu = 3$ and $\sigma = 1$). Since we cannot have negative firing rates, set all rates that are below zero to zero. Now do the same thing for the 10% coherence stimulus. On the same figure, plot the histograms of the firing rates for each stimulus type.

   (b) The success of the decoder (assuming this model of Gaussian noise) is determined by two things, the separation of the mean firing rates and the standard deviation of the neuron. From class, we know that this is captured in the measure known as d'. Calculate d' for this task and pair of stimuli (ignoring the fact that you are clipping firing rates at zero).

   (c) Explain why the maximum likelihood decoder for this problem involves comparing the measurement to a threshold. For various thresholds t, calculate the hit and false-alarm rates using your sample data from (a), and plot these against each other (this is an ROC curve, defined in class). What threshold would you pick based on this curve to maximize the percentage-correct of the decoder, assuming that 0% and 10% coherence stimuli occur equally often. Plot this threshold as a point on the ROC curve and as a vertical line on your histogram from part (a). Next, suppose that 10% coherence stimuli occur 75% of the time. Determine and plot the threshold that maximizes percentage correct for this new prior.

   (d) Consider now a neuron with a more "noisy" response so that the mean firing rates are the same but the standard deviation is 2 spikes/s instead of 1 spike/s. What is the new value of d'. Recompute and plot the optimal (maximum accuracy) thresholds for this noisy neuron for both the 50-50 and 75-25 priors. How do they differ from those in the previous part?