PSYCH-GA.2211/NEURL-GA.2201 – Fall 2019
Mathematical Tools for Neural and Cognitive Science

Homework 5

Due: 26 Nov 2019
(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. **Dueling estimators**. In this problem, we use simulation to compare three estimators of the mean of a Normal (Gaussian) distribution.

   (a). First consider the *average*, which minimizes the sum of squared deviations, and is also the Maximum Likelihood estimator. Generate 10,000 samples, each of size 10, from the Normal(0,1) distribution (a 10x10000 matrix). Compute the average of each of the 10,000 samples. Plot a histogram of the resulting estimates (use 50 bins, and set the plot range to [-2.3,2.3]). What shape should the histogram have (explain why)? What is the (theoretical) variance of the average of 10 values drawn from a univariate Gaussian (derive this)? Is the empirical variance of your 10,000 estimates close to this?

   (b). Now consider the *median*, which minimizes the sum of absolute deviations. Compute the median of each of the 10,000 samples, and again plot a histogram. What shape does this one have? Compare it to a normal distribution using the function `normplot`, which plots the quantiles of a sample of data versus the normal quantiles (known as a Q-Q plot: if data are normally distributed, the points shuld fall nearly on a straight line.) Does the distribution of estimated values deviate significantly from a Normal distribution? Specifically, compare the Q-Q plot for the median estimator to that for the mean from part (a).

   (c). Finally, consider an estimator that computes the average of the minimum and maximum over the sample (as shown in class, this one minimizes the $L_\infty-$norm). Again, compute this estimate for each of your 10,000 samples, plot the histogram, and examine and comment on the Q-Q plot, just as in part (b).

   (d). All three of these estimators are unbiased (because of the symmetry of the distribution), so we can use variance as the sole criterion for quality. Generate a new set of 10,000 samples, this time of dimension 256. Apply each estimator to sub-matrices of samples of size $\{8, 16, 32, 64, 128, 256\}$, and compute the variance of each estimator for each. Plot these (on a single log-log plot), along with a line showing the theoretically-computed variance of the average estimator. Does the variance of all three estimators converge at the same rate ($1/N$)? How much larger is the variance of the median estimator than the average estimator? How large a sample would you need for the average and median estimators to achieve the same variance as the average-extrema estimator (from part (c)) on samples of size 256?

2. **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 103 out of 869 fMRI contrasts involving engagement of language, but this area was also active in 199 out of 2353 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 relevant conditional probabilities, and assuming the prior that a contrast engages language, $P(language) = 0.5$. 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, compare the Bayes factor (probability of language vs. not language) after taking Poldrack's data into account, compared to before having done so.

3. **Psychopathy.** You are interested in causes and treatment options for Psychopathy. You obtained a dataset, contained in the file `psychopathy.mat` obtained from a prison for violent offenders in upstate New York (not everyone in the prison is a psychopath, but they are more prevalent than in the general population). Each row represents data from one prisoner. All study participants underwent a structural scan with a mobile, truck-mounted MRI. The first data column contains the estimated cortical volume of paralimbic areas, relative to the population median, in cm^3. The second column contains the Hare Psychopathy Checklist (PCL-R) scores, which range from 0 to 40 (the higher the score, the more psychopathic traits someone exhibits). These scores are not distributed normally in the general population (median = 4) and definitely not normal in this subpopulation (median = 20). The third column indicates whether they already participated in an experimental treatment program known as âdecompression therapyâ (0 = did not yet participate, 1 = did already participate). To avoid self-selection effects, everyone in this dataset agreed to the therapy, but prisoners were randomly assigned to an earlier and a later treatment group, so that the untreated

prisoners could serve as a control group.

(a) You want to model PCL-R scores as a function of relative volume of paralimbic areas. Use polynomial regression to find a model that best explains the data using cross-validation. What degree does it have? (Note, you can use your code from HW2.)

(b) Use bootstrapping methods to estimate the 95% confidence interval of the average paralimbic volume of the decompression treatment group vs. the control group. If the random assignment worked, the confidence intervals should overlap. Do they? Also, do these data suggest that there is a statistically reliable difference to the general population in terms of paralimbic volume?

(c) Do a suitable t-test to compare the mean PCL-R score of prisoners who did and did not undergo decompression therapy. What is the p-value? Assuming an alpha-level of 0.05, is this difference significant? Can you reject the null hypothesis that decompression therapy is ineffective in terms of decreasing PCL-R scores?

(d) Do a permutation test to assess whether decompression therapy has an effect. Designate an appropriate test statistic and calculate its exact p value.

4. **Simulating a 2AFC experiment.** Consider a two-alternative forced choice (2AFC) psychophysical experiment in which a subject sees two stimulus arrays of some intensity on a trial and must say which one contains the target. (One and only one contains the target.) Her probability of being correct on a trial is:

$$p_c(I) = 1/2 + 1/2\Phi(I; \mu, \sigma)$$

where $\Phi(I; \mu, \sigma)$ is the cumulative distribution function of the Gaussian (`normcdf` in matlab) with mean μ and standard deviation σ evaluated at I. The function $p_c(I)$ is known as the *psychometric function.* (Minor note, somewhat subtle: This setup only makes sense if I is on a logarithmic scale, e.g., $I = k \log C$, where C is stimulus contrast.)

(a) Plot two psychometric functions, for $\{\mu, \sigma\}$ equal to $\{5, 2\}$ and $\{4, 3\}$. (Use $I = [1 : 10]$). Describe the difference between these. If you increase μ, how does the curve change? If you increase σ, how does the curve change? (If you are not sure, make more plots with different parameter values.) What is the range of $p_c(I)$? Explain why this range is appropriate.

(b) Write a function `C=simpsych(mu,sigma,I,T)` which takes two vectors `(I,T)` of the same length, containing a list of intensities and the number of trials for each intensity, respectively, simulates draws from $p_c(I)$, and returns a vector, C, of the same length as I and T, which contains the number of trials correct out of T, at each intensity I.

(c) Illustrate the use of `simpsych` with `T=ones(1,7)*100` and `I=1:7` for $\mu = 4$ and $\sigma = 1$. Plot `C ./ T` vs `I` (as points) and plot the psychometric function $p_c(I)$ (as a curve) on the same graph.

(d) Do the same with `T=ones(1,7)*10` and plot the results (including the psychometric function). What is the difference between this and the plot of the previous question?