PSYCH-GA.2211/NEURL-GA.2201 – Fall 2017
Mathematical Tools for Neural and Cognitive Science

Homework 5

Due: 1 Dec 2017
(late homeworks penalized 10% per day)

See the course web site for submission details. Please: don't wait 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 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 a Normal distribution? For comparison, you might want to look at the Q-Q plot for estimates from part (a).

   (c). Finally, consider an estimator that computes the average of the minimum and maximum over the sample. Again, compute this estimate for each of your 10,000 samples, plot the histogram, and examine and comment on the Q-Q plot.

   (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 size 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 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 in some contrast, 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., active or not with some probability as in the coin-flipping example in class). Compute the likelihoods of Poldrack's observed frequencies of activation as functions of the possible values of their respective Bernoulli probability parameters x_l and x_{nl}. 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?* Using the estimates from part (b) as the relevant conditional probabilities, and assuming the prior that a contrast engages language, $P(language) = 0.5$, compute the probability $P(language \mid activation)$ that observing activation in this area implies engagement of language processes. Is Poldrack's critique correct? How confident should you be in implicating language if you observe activity in Broca's area?

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 decompression therapy (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.

Find a polynomial model that best explains the data [use cross-validation]. What degree does it have?

(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.