PSYCH-GA.2211/NEURL-GA.2201 – Fall 2016
Mathematical Tools for Cognitive and Neural Science

Homework 4

Due: 1 Dec2016
(late homeworks penalized 10% per day)

See the course web site for submission details. Don't wait until the day before the due date... start *now*!

1. **Bayesian inference of eye color genes.** A male and female chimpanzee have blue and brown eyes, respectively. Assume a simple genetic model in which the gene for brown eyes is always dominant (so that the trait of blue eyes can only arise from two blue-eyed genes, but the trait of brown eyes can arise from two brown-eyed genes, or one of each). You can also assume that the apriori probability that the female has either of these gene configurations is 50%.

   (a) Suppose you observe that they have a child with brown eyes. Now what is the probability that the female chimp has a blue-eyed gene?

   (b) Suppose you observe that they have a second child with brown eyes. Now what is the probability?

   (c) Suppose they have N children with brown eyes... express the probability (as a function of N).

2. **Sums of random variables.** Suppose you are given a vector `p` of length `n` that contains a PDF over the integers $0 \ldots n-1$. (The values in `p` should sum to 1). Write a function `samples = randp(p, num)` that generates samples from the given PDF, as above.

   Test your function by choosing some arbitrary `p` of length 10, drawing 1,000 samples, plotting a histogram of how many times each value is sampled, and comparing this to the frequencies predicted by `p`.

   Next, write a function `psum(p, q)` that, given two such discrete PDFs, returns a vector encoding the PDF for the sum of a sample drawn from `p` and a sample drawn from q. (Hint: the output vector should have length $m+n-1$ when m and n are the lengths of the two input PDFs.)

   Test your function by letting `p = [1:6] / sum([1:6])` (a weighted die), and using repeated calls to `psum` to compute the PDF predicted for a sum of five rolls of the die. Use your function `randp` to generate 5 sets of 1000 samples from `p`; plot a histogram of how many times each sum is sampled; and compare this the frequencies predicted by your program.

3. **Multi-dimensional Gaussians.**

   (a) Write a function `samples = ndRandn(mean, cov, num)` that generates a set of samples drawn from a multidimensional Gaussian distribution with the specified `mean` (an N-vector) and `cov`ariance (an NxN matrix). The parameter `num` should be optional (defaulting to 1) and should specify the number of samples to return. The returned value should be a matrix with `num` rows each containing a sample of N elements. (Hint: use the MATLAB function `randn` to generate samples from an N-dimensional Gaussian with zero mean and identity covariance matrix, and then modify these appropriately. Recall that the covariance of $Y = MX$ is $E(YY^T) = MC_XM^T$ where C_X is the covariance of X).

(b) Test your function by plotting 1000 samples of a 2-dimensional Gaussian (choose an arbitrary nonzero mean and nonzero covariance). Measure the sample mean and covariance of your data points, comparing to the values that you requested when calling the function. Plot an ellipse on top of the scatterplot that traces out points that are two standard deviations away from the mean, according to the covariance matrix. Does this ellipse capture the shape of the data?

(c) Now consider the generalized marginal distribution of your 2-D Gaussian in which samples are projected onto a unit vector $\vec{u}$ to obtain a 1-D distribution. Using the equations presented in class, write a mathematical expression for the mean and variance of this marginal distribution as a function of $\vec{u}$ and check it for a set of 48 unit vectors spaced evenly around the unit circle. For each of these, compare the mean and variance predicted from your mathematical expression to the sample mean and variance estimated by projecting your 1,000 samples onto $\vec{u}$. Plot the mathematical mean and the sample mean (on the same plot), and also plot the mathematical variance and the sample variance. How would you, mathematically, estimate the direction (unit vector) that maximizes the variance of the marginal distribution? Compute this direction and verify that it is consistent with your plot.

4. **Comparing two estimators**. In World War II, Allied data analysts needed to estimate the number of German tanks based on daily surveillance reports of the serial numbers of observed tanks. Consider a simple version of this problem, where there are N tanks numbered from 1 to N (with N unknown), and K daily observations are drawn independently and uniformly from this set (with replacement - a tank might be observed more than once).

   (a) Examine the maximum likelihood estimator (MLE). The likelihood function of N is the probability of observing a given set of K values, drawn independently from the uniform distribution, $p(\{n_1, n_2, n_3, ...n_K\} \mid N) = (1/N)^K$, for $1 \leq n_k \leq N$. This expression decreases with increasing N, so it will be maximized by choosing N as small as possible. Thus, the MLE is simply the largest of the observed serial numbers: $\hat{N}_{\text{MLE}} = \max(\{n_1, n_2, n_3, ...n_K\})$. To examine the behavior of this estimator, we need to calculate the distribution of the maximum of K observed serial numbers. The probability that an observation is less than or equal to any given value n is just (n/N), so the probability that K (independent) observations are all less than or equal to n is $(n/N)^K$. If we place these values for different n into a vector, $c(n)$, this will represent the *cumulative* distribution of the max, and the difference, $c(n) - c(n-1)$, gives the probability that the maximum value is equal to n. Compute this distribution for $N = 40$ and $K = 16$, and plot it as a bar plot (use `bar()`). What is the probability that the maximum value is equal to N? What is the mean of the distribution of maximum values? What is the standard deviation of the distribution?

   (b) Now consider an intuitive estimator derived from the sample average. Since the mean of the distribution of observations is $(N + 1)/2$, we can take as our estimate of N twice the sample average minus 1. Using the expressions for the mean and standard deviation of a uniform distribution, and the fact that the standard deviation of a sample average falls with the number of samples as $1/\sqrt{K}$, write an expression for the mean and standard deviation of this estimator as a function of K and N . Plot a gaussian with this mean and variance, as a bar plot (noting that the values of this estimator can be larger than N, plot this from 1 to $2N - 1$). How does this distribution (and the mean and stdev values) compare to that of the previous estimator? Which estimator do you think is "better", and why?

(c) To verify your math, simulate the results for the previous two parts. First, write a function `observe(K, N)`, to generate K random observations of N tanks (use the matlab function `randi`. Then write functions `est1(obs)` and `est2(obs)` that compute each of the two estimates. Simulate 10,000 days worth of observations (again, assume $N = 40$ and $K = 16$) and collect the results of applying each estimator in a vector (length 10,000 each). Plot histograms of the estimates (use the function `hist`), and compare to the plots from parts (a) and (b) (note: normalize histograms to sum to one, and make the axes the same!). Compare the mean and standard deviations of the simulated data to the values you computed in parts (a) and (b).

5. **Gaussian estimation.** Catherine is looking for Alex in a very large one-dimensional shopping mall. Location is specified by a coordinate X. Catherine knows that, all else being equal, Alex prefers to be near the center of the shopping mall at location 50. He has a prior Gaussian distribution centered on 50 with variance 40. The only clue Catherine has is a coffee cup of a brand that only Alex drinks that she finds at location X=30. The coffee cup is cold and Alex has wandered off. Based on the location of the coffee cup, the likelihood function of his location is a Gaussian distribution with mean X=30 and variance 100.

   (a) Explain how you would frame this problem as a problem in Bayesian estimation, using appropriate terminology. What is Alex's posterior distribution? Draw his prior, likelihood and posterior distributions on a single plot. (Rather than `normpdf`, compute Gaussian probabilities from the formula for the Gaussian distribution.) What is the variance of the posterior?

   (b) The coffee cup was not that cold after all. Alex's likelihood function has mean X=30 but with a smaller variance of 20. Redo part a. Describe what happened to the posterior distribution. Has it moved? Does the change make sense?

   (c) What would the posterior distribution in (a) be if the prior had been uniform (and, thus, the posterior proportional to the likelihood). What would the variance of this distribution be? Compare this variance to that of the posterior in (a). How does the inclusion of prior information affect the variance?