Homework 6
Statistics 200B
Due Mar 14, 2019

1. Let $X_1, \ldots, X_n$ be *iid* with density $f(x;\beta) = \beta e^{-\beta x}$ for $x > 0$ and $\beta > 0$. Find the asymptotic (large sample) likelihood ratio test of size α for $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$.

2. Let X and Y be two random variables with joint distribution F. Suppose we observe pairs $(x_1, y_1), \ldots, (x_n, y_n)$, a random sample from F. Without making any assumptions about F, form a statistic for testing $H_0 : P(X > Y) = 0.5$. How would you calculate the p-value?

3. The file `pvals.R` contains `R` code for a simulation examining the proportion of times H_0 is actually true for tests within a specified range of p-values. The set-up (due to Sellke, Bayarri, and Berger, 2001) is as follows:

   Suppose for each dataset we observe $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. We form the Wald test statistic and calculate the p-value for testing $H_0 : \mu = 0$.

   The variable `nullprop` specifies the proportion of times in the simulation that H_0 is actually true. The rest of the time, the true value of μ is drawn from a pre-specified distribution. This is currently set to be $Unif(-1, 1)$. The last two lines of the file calculate and print the proportion of times, among tests for which $0.01 <$ p-value < 0.05 (what is usually called "strong evidence against H_0"), that the null hypothesis was actually true.

   (a) Experiment with the code to see what the effect is of changing the distribution of μ when it is not zero. For example, try changing the parameters of the uniform distribution, using a different distribution, or using a constant. Summarize your findings. (*Hint: If you change the file* `pvals.R`*, you can rerun everything in it using* `source("pvals.R")`*. You may need to add the directory to the beginning of the file if it is not in your working directory.*)

   (b) Now experiment with changing `nullprop`, the proportion of times in the simulation that H_0 is actually true. Again, summarize your findings.

4. Consider the Berkeley freshman admissions data on page 97 of the class notes.

   (a) Using the top 6 cells of the table, calculate a 2x3 table counting numbers of students according to the population (CA residents, non-residents, international) and admission status (yes, no).

   (b) Calculate the MLEs for the corresponding probabilities under the null hypothesis of independence of the two variables, and the MLEs with no restrictions. Report them in the same table format. (You may use the formulas given in class; you do not need to rederive the MLEs.)

   (c) Calculate the likelihood ratio test statistic and Pearson's χ^2 statistic. Also report the p-values for each, using the limiting distributions of the test statistics under the null hypothesis of independence.

5. (Multinomial tests of homogeneity) When Jane Austen died, her novel Sandition was incomplete. Someone else finished the novel and it was published. Morton (1978) examined word frequencies to see if the new author was distinguishable from Austen. The data are as follows:

Word	Sense and Sensibility	Emma	Sandition I (Austen)	Sandition II (New author)
a	147	186	101	83
an	25	26	11	29
this	32	39	15	15
that	94	105	37	22
with	59	74	28	43
without	18	10	10	4
Totals	375	440	202	196

   Treat each column as an independent Multinomial sample.

   (a) Construct the likelihood ratio statistic and calculate the p-value for the null hypothesis that the first three columns (by Austen) have the same set of probabilities for each column.

   (b) Now sum the first three columns to give a single column of counts for Austen and another for the new author. Construct the likelihood ratio statistic and calculate the p-value for the null hypothesis that probabilities are the same across authors.