Statistics 520: Assignment 5

Sam Olson

Assignment 5

The objectives of this assignment are to (1) ensure that you have a grasp on using the tools of basic likelihood in a data analysis and (2) help you to continue to develop the precise use of notation in presenting descriptions of analyses. On the course web page is a file in the Data folder called `gammadat.txt`. This file contains two columns of values with a header having labels `group1` and `group2`. Each column should be considered to contain values corresponding to a set of independent and identical gamma random variables. That is, the two columns are values from two groups that we wish to compare using a two-sample model with gamma distributions. Consider the first column to contain values for Group 1 and the second column to contain values for Group 2.

A number of resources are available to you to help you complete this assignment. Chapter 5 of the course notes contains a summary of likelihood methods. In the Computing folder of the course web page is a file `newtraph.txt` that contains a generic Newton-Raphson algorithm that you may use for maximum likelihood estimation. There is also a file called `newtraphexplain.txt` that describes the inputs needed, the syntax, and the output. Alternatively you may choose to make use of the built-in R functions `optim` or `nlm`. Any of these options (or others you might know of if you prefer Matlab or something else) are fine as long as you know what you are doing and can produce the quantities needed to conduct the analysis.

Your answer should contain complete and consistent notation using no undefined symbols. You should always clearly explain what you computed and the formulas used. Your answer should not contain computer code or material from a "screen dump." You will not be awarded any points for such material. If you want to report estimated values do so in the text, as a list, or construct a table.

Again, do not include copied computer function output. You will not get credit for anything presented in that way.

1.

Assume random variables $Y_{1,1}, \ldots, Y_{1,n_1}$ and $Y_{2,1}, \ldots, Y_{2,n_2}$ have been defined for the responses in this problem. These responses are strictly positive numbers, and an assumption of independence is reasonable. Formulate a two-sample model using gamma distributions. For one group, write the form of the log likelihood that will need to be computed to find estimates and other inferential quantities.

2.

Find maximum likelihood estimates and 95% Wald theory intervals for the parameters of each group. Recall that, in the data file, the first column of values is Group 1 and the second column of values is Group 2.

3.

Using a likelihood ratio test, determine whether you would reject a model having a common gamma distribution for both groups in favor of a model having separate gamma distributions for each of the two groups. Produce a plot of the estimated densities for each group (both densities on the same plot).

4.

Find maximum likelihood estimates and 95% Wald theory intervals for the expected value of each group. Also produce a 95% interval for the difference in expected values (Group 1 minus Group 2).

5.

Test whether the two groups should be considered significantly different using a two-sample t-test. (Take square roots if you think it makes the data look more symmetric for each group, though this is optional.) Does your result agree with the likelihood ratio test? Does it agree with the interval for difference in expected values?

6.

Find maximum likelihood estimates and 95% Wald theory intervals for the mode of each group. Also produce a 95% interval for the difference in modes (Group 1 minus Group 2).

7.

Although model assessment has not yet been covered formally, it is intuitive that the estimated distribution function (CDF) under our model and the empirical distribution function of the data should be similar. Produce plots of the estimated distribution function for each group with the empirical distribution function overlaid.

8.

Write a short paragraph giving your conclusions about this group comparison.