Empirical Methods for Computer Science (CS 5970) Homework 3

December 1, 2008

This homework assignment is due on Tuesday, December 2nd at 5:00pm. Your work may be handed in electronically (use the **Homework 3** digital dropbox on D2L) or in hardcopy form.

This assignment must be done individually: do not share/discuss your answers with others or look at the answers of others.

All data sets are contained within the hw3.mat file available on the main homework page.

Question 1

In this question, we will look at dealing with censoring of samples and at the use of the single sample bootstrap test. Implement a matlab function that shows the power of three distinct statistical tests as a function of the assumed mean of the underlying distribution. The three tests are:

- Bootstrap test that explicitly acknowledges censored values in the sampling process
- Bootstrap test that does not acknowledge censored values (i.e., censored elements are removed from the sample before the bootstrap test is performed).

- t-test (remove censored elements from the test before performing the test).

The following functions are provided:

- randn_bogotify($N, Q, prob$) produces an NxQ matrix of samples from a standard normal distribution, except: with probability *prob*, the samples are set to NaN.
- rand_multi_bogotify($N, Q, prob$) produces an NxQ matrix of samples from some distribution. As above, with probability *prob*, the samples are set to NaN.
- mean_safe(x): if x is a matrix, computes the mean of the columns of x; if a vector, computes the mean of the elements in the vector. In both cases, samples that are NaN are removed from the mean.

Make the following assumptions:

- $N = 7$ or $N = 30$: number of items in the sample.
- $Q = 1000$: number of times that we will take N samples and perform the statistical tests in order to estimate the power of the tests.
- $M = 1000$: the number resamples that we will take for the bootstrap tests.
- All tests are right-tailed.
- $\mu = [-5...0]$: The assumed mean of the underlying distribution for all three tests. (Note that $\mu = 0$ corresponds to the case in which there is no difference between the null and alternative hypothesis distributions.
- $\alpha = 0.05$.

1. (10pts) For the normal distribution, $N = 30$, and $prob = 0$, show power for each of the tests as a function of the assumed mean. Discuss the similarities/differences.

2. (10pts) For the normal distribution, $N = 7$, and $prob = 0$, show power for each of the tests as a function of the assumed mean. Discuss the similarities/differences.

3. (10pts) For the normal distribution, $N = 30$ and $N = 7$, and $prob = 0.5$, show power for each of the tests as a function of the assumed mean. Discuss the similarities/differences.

4. (10pts) For the non-normal distribution, $N = 30$ and $N = 7$, and $prob = 0$, show power for each of the tests as a function of the assumed mean. Discuss the similarities/differences. Compare to the normal case.

5. (10pts) For the non-normal distribution, $N = 30$ and $N = 7$, and $prob = 0.5$, show power for each of the tests as a function of the assumed mean. Discuss the similarities/differences. In the latter figure, explain the inflection point at $\mu = -2.5$

6. (10pts) Would you rather use the version of the bootstrap test that does or does not acknowledge censoring?

Question 2

dat_pre and **dat_post** are scores that are received for a set of 20 different tests. Each row of these matrices represents a set of samples corresponding to a single individual. Note, however, that individuals are not paired across the pre and post conditions. You may consider the scores to be drawn from continuous random variables for the purposes of our analysis (they are actually interval variables).

We wish to determine whether the post tests perform better than the pre tests.

1. (10pts) Use the appropriate t-test to compare the performance for each of the conditions. For which conditions is there a significant difference in performance (assume $\alpha = 0.15$)?

2. (20pts) Using the bootstrap randomization method (your own implementation), for which conditions is there a significant difference? (again, assume $\alpha = 0.15$)

3. (10pts) Under this experiment and analysis design, do we have a multiple comparisons problem?