CSCI 350: Data Analytics
Spring 2017
Proj#3
Confidence Intervals
Out: 3/27/2017
Due: 4/3/2017@11:59:59pm

Total: 100 points

Name: _____

Goal:

The goal of this assignment will be to gain experience with the intuition behind confidence interval through reduction to practice. This will be accomplished by testing if the proportion of confidence intervals that cover the ground truth parameter coincides with the probability associated with the confidence interval.

Assignment:

In the course module "Lecture # 13" we discussed the intuition behind confidence intervals and in "Lecture #14" we demonstrated normalization and discussed how to compute confidence intervals. In our discussion, we described confidence intervals as a method for describing the concept of "how far away" an estimator $\hat{\theta}$ is from the ground truth population parameter θ by using standard error as a way of expressing how "typical" (whether it belongs) a measurement is relative to a distribution. A confidence interval is associated with a probability. This probability describes the proportion of intervals [a,b] computed from a sample distribution that "cover" the ground truth population parameter θ. In this assignment your preparation will include:

1. Review the slides for module "Lecture 13" and "Lecture #14"
2. Make sure you have read the sections of Chapter 9 (Baron text) on confidence intervals.
3. Find on the course black board companion MATLAB code you will need for this assignment

    - `generateNormalSamples.m:` produces samples from normal distribution
    - `visualizeSamplingDistribution.m:` visualizes sampling distribution
    - `samplingDistTestNew.m:` driver program containing population parameters

4. Review and understand the companion MATLAB code, how it works, and what it does
5. Run the MATLAB code multiple times and note how the sampling distribution changes

You will conduct research into something called z-scores. The z-score represents the multiplier of standard deviation that governs the probability or percentage of probability density (or mass) associated with coverage of the ground truth parameter. A z-score represents the p-quantile of a standard Normal distribution. An example z-score table is found at https://www.ltcconline.net/greenl/courses/201/estimation/smallConfLevelTable.htm

Recall for example given the center of a sampling distribution for the confidence interval [a,b], the values of "a" and "b" are determined using $\mu_s \pm z\sigma_s$ where μ_s is the center of the sampling distribution, σ_s is he standard error of the sampling distribution, and z is the z-score. That is, the z-score describes the multiple of the standard error associated with the probability corresponding to a given confidence interval. For example, a 70% confidence level is associated with the interval [a,b] corresponding to $\mu_s \pm 1.04\ \sigma_s$. From the example z-score table, we find that a confidence level of 0.7 corresponds to a z-score of 1.04.

In your assignment, you will

1. Design a method for testing if the probability associated with a confidence interval (CI) coincides with the proportion of CI's that cover the ground truth population parameters.
2. You will do this test for CI's for probabilities of 10%, 20%,…,90%
3. Code has been provided to compute the sampling distribution. Moreover, this code contains the ground truth population parameter for mean and variance of a Normal distribution.
4. Design code to validate that
    a. Test confidence intervals for each of the aforementioned probabilities 10%,…,90%
    b. Generates a sampling distribution
    c. Uses appropriate z-score to compute confidence interval [a,b]
    d. Tests if confidence interval covers ground truth parameters
    e. Repeats steps b,c,d a number of times and measures the proportion of confidence intervals that covers ground truth
    f. Report your results about how well the the proportion of "covering" confidence intervals coincides with the probability.
    g. You will run this test (steps b,c,d,e,f) for 10%,…,90% for the sampling distribution of both the **MLE** estimate for **mean** and **variance**. These are provided in the example code.

5. Write a document (MS-Word, PDF only) describing your approach and results.
6. You will design and write MATLAB code for this assignment

Submitting your work

- Do not include a JPEG or image file. All images must be inserted into a single document along with any discussion.
- Late assignments will not be accepted. Please manage your time appropriately.

- Write a single MS-Word or PDF file for your report. If you submit a scan or image of handwritten work, it must be legible. Illegible work will receive a zero.
- Include every MATLAB file you write. Your program must run. Nonfunctioning programs will receive zero.
- Upload via Blackboard ONLY!!! Email will not be accepted.
- If you consult references and or other material from published work (book, paper, etc.), please cite your source.
- Submit all of your files as a single ZIP file (no Rar, Gzip, 7-ZIP, or TAR)

Assessment:

Confidence interval multipliers: 20 points
Code to compute confidence intervals from sampling dists: 20 points
Design of experiment: 20 points
Test for proportion of CI covering: 20 points
Approach and results/discussion: 20 points
Total: 100 points