CS 40003 : Data Analytics

Project Assignment #2   **29.08.2016**

Project objective:

In our theory classes, we have learned a number of probability distribution functions. Note that such a probability distribution is characterized with few parameters (e.g., normal distribution is characterized with the mean and standard deviation). One objective of the project is to draw a sample, which obey a given probability distribution function. If we have a large set of such samples (i.e., their probability distributions), then we shall be able to obtain the corresponding "sampling distributions of mean" and "sampling distributions of variance". The second objective of the project is to obtain such sampling distributions from a set of samples and draw the histogram for each.

The tasks of the project are precisely stated in the following.

1. You have to draw 10000 samples, each of size n using ***Binomial distribution***. You should assume the suitable values of the parameters, namely size of the sample (n), number of trials (s), and probability of success (p). In other words, you are to call a R function with 3 input parameters: n, s and p to draw a sample..

2. Create a matrix of size $10000 \times n$ where each row corresponds to one replication/sample of size n.

3. For each sample, do row-wise operations to compute mean and variance.

4. Plot the sampling distributions of means and variances. Use ***histogram*** for generating plot.

5. Repeat the aforementioned tasks for a) ***Normal distribution*** and b) ***Poisson's distribution***.

Submission procedure:

1. Submit three program files (all are executable) separately for a) Binomial, b) Normal and c) Poisson's distributions.
2. Each group should submit only one submission.
3. **Last date of submission is: 11.09.2016, 12:55 hours.**