jignesh.bhatt@iiitvadodara.ac.in

IIIT Vadodara
Autumn 2020-21
CS605 Data Analytics
Lab-2 | August 07, 2020.
Central limit theorem and Sampling distributions of mean

Q. 1: **Central limit theorem:** (A) Generate mean subtracted $X_1, X_2, ..., X_N$ uniformly distributed, i.e., $U[0,1]$, IID random variables. Generate at least 10000 samples for each of the random variables and let $N = 25$. Now, starting from X_1 start adding the random variables, i.e.,

$$Y_k = \sum_{i=1}^{k} X_i, \tag{1}$$

for $k = 1, 2, ..., N$. Now plot the PDFs for each of the Y_k. Hint: Shown demo in the class.

(B) Repeat the above experiment for mean subtracted independent random variables, with different distributions, e.g., Uniform, Laplacian, etc. Compare your results with the above experiment.

(C) Finally, perform the same experiment for *dependent* random variables and compare your results. For example, $X_1 \sim U[-1,1]; X_2 = 2X_1; X_3 = X_1 + X_2 + 3$, and likewise.

Q. 2: **Sampling distribution of mean:** This is as discussed in the class. A manufacturing process produces cylindrical component parts for the automobile industry. It is required to produce the parts with a mean diameter of 5.0 mm. It is known that the population standard deviation is $\sigma = 0.1$ mm. The engineer involved conjectures that the population mean is 5.0 mm.

(A) An experiment is conducted in which 100 parts produced are selected randomly and the diameter measured in each. Generate uniformly distributed 100 random numbers to represent the diameter using available information and compute a sample average diameter. Does this sample information appear to support or refute the engineer's conjecture? Draw necessary plots.

(B) Generate the sampling distribution of mean by repeating the above random experiment several times. Based on the sample mean obtained thus, comment on the engineer's conjecture? Draw necessary plots.

(C) Repeat the part (C) by repeating the random experiment for 10000 parts. Based on the sample mean obtained thus, comment on the engineer's conjecture? Draw necessary plots.

Q. 3: **Sampling distribution of the difference between two means:** Consider two populations with certain probability distributions $P_1(\Omega)$ and $P_2(\Omega)$. Draw two independent random samples of size n_1 and n_2 from each population, respectively. Now, generate the sampling distributions of differences of the two means under following cases, and comment on the Normal approximation of the sampling distribution of difference between two means (discussed in the class):

(A) For $P_1(\Omega) \sim N(5, 1)$ and $P_2(\Omega) \sim N(3, 1)$, and for $n_1 < 30$ and $n_2 < 30$.

(B) Repeat part (A) for non-normal populations while keeping other settings the same.

(C) For $P_1(\Omega) \sim N(5, 1)$ and $P_2(\Omega) \sim N(3, 1)$, and for $n_1 \geq 30$ and $n_2 \geq 30$.

(D) Repeat part (C) for non-normal populations while keeping other settings the same.

Draw all necessary plots for detail illustration.