

Assignment 2 -Normalization methods

- a. Generate 3 variables, 10000 samples each
 - i. B: Gaussian mean 5 sd 2
 - ii. I : Power law , use `scipy.stats.powerlaw.rvs` with `a= 0.3`,
 - iii. H : Geometric `p= 0.005`
- b. Compare above variables in single box plot
- c. Try following normalization methods
 - i. Divide each variable by max
 - ii. Divide variable by sum of its values
 - iii. Convert each variable into z score using respective mean and sd
 - iv. For each variable , convert the values in percentiles
 - v. Make medians of all the variables same
 - 1. Calculate median of each variable
 - 2. Calculate mean of these medians say `m1`
 - 3. Generate a multiplier for each of the variables so that median value of each variable becomes `m1`
 - vi. Quantile normalize the data using off the shelf library function
- d. For each of the above method,
 - i. Compare original distribution with its normalized version in single histogram
 - ii. Compare all the normalized variables in single box plot