Experiment 4

Laplace-De Moivre theorem

Instruction: Create a new .Rmd file and write code for each section in separate R code block

The Laplace-De Moivre theorem theorem states that the binomial distribution can be approximated by the normal distribution for a large number of trials (n). This approximation is useful because calculating binomial probabilities directly can be difficult for large values of n. It is also a good way to understand generation of Gaussian or Normal Random variable.

Dataset: You are given a dataset file, “bikerental.csv” with 12 columns. The Bike Rental UCI dataset records daily bike rentals with features like temperature, humidity, windspeed, and season, used for predicting demand. The 12th column is rentals/day. Columns 1-11 are various features.

1. Generate a binomial random variable b~B(10,0.4), where n=10 and p=0.4. Repeat the experiment for 10000 trials and store it in a vector ‘x’.

2. Convert x to a zero mean and unit variance random variable y. Draw a Density histogram for y.

3. Using the values in y, overlay a normal density function on the histogram in step 2.

4. Repeat step 1-3 by taking n as 50, 100 and 500. What are your observations?

5. Download “biketental.csv”. Define a variable (Z) as number of rentals/day. Transform Z to a zero mean and unit variance variable (Y). Write a code to plot the Density histogram for Y. In Density histogram x-axis is the variable and y-axis is the density (count/Total count). Choose number of bins as 40.

6. Using the values in Y, overlay a normal density function on the histogram in step 5.

7. Using correlation as metric, identify two features (out of columns 1-11) which have maximum correlation with Y.

8. Implement a linear regression (Y vs (Two features in step 7)). Split the dataset into training and test set. Find mean absolute error for the training and test set.

9. Plot a histogram for the variable e=Y-Yp, where Yp are the predictions on the training set instep 9. Fit a Gaussian distribution to e and show it on the histogram.

10. Exploratory: Implement linear regression in step 8 for all combinations of 2 features (55 combinations). Print Train MAE and Test MAE in each case.