CS 6464 - CSLT
SOFTWARE ASSIGNMENT 1

Task 1: Regression - Polynomial Fitting:

Consider the problem of fitting one-dimensional data with a polynomial. Write an R code to plot function y given in "Q1_fun_xx".

Randomly extract 100 points from the function and add normally distributed noise to the data points from sigma value range to get "noisy data", $\hat{y}$.

1. Fit polynomial of degree **"d"** (values given in Table 1 below) to the noisy data.
2. Compute the bias and variance for the models fitted.
3. Plot the bias-variance against model complexity.
4. Plot the bias and variance as a function of sigma (for plotting bias and variance against sigma, pick a value of sigma in 10 equal intervals in the range as given in the table below and add it to all the 100 data points and get the corresponding bias and variance values).

Functions: (Code for both functions for -10 <= x <= 10)

Q1_fun_01: y= $0.3\, cos(3\pi x) - 0.4\, cos(4\pi x) - 10/(x^2 + 1)$

Q1_fun_02: y= $0.4\, log\, (x^4 + log(x - 0.7) + e^{3x})$

Table 1:

S. No.	Roll Nos.	d values	Sigma values
1	CH17D406	1, 10, 25	0.5 - 1.5
2	CS16B044	2, 11, 24	1-2
3	CS18M016	3, 12, 23	0.2-1.2
4	CS18S038	4, 14, 26	1.5-2.5
5	CS18S040	5, 17, 28	0.1-1.1

6	CS19D002	1, 11, 21	0.9-1.9
7	CS19M010	2, 12, 28	1.1-2.1
8	CS19M011	3, 10, 24	0.3-1.3
9	CS19M016	4, 13, 22	0.8-1.8
10	CS19M017	5, 16, 27	1.2-2.2
11	CS19M020	1, 12, 23	0-1
12	CS19M023	2, 13, 26	0.4-1.4
13	CS19M024	3, 11, 25	0.7-1.7
14	CS19M028	4, 10, 21	1.3-2.3
15	CS19M029	5, 15, 26	0.6-1.6
16	CS19M030	1, 13, 22	0.3-1.3
17	CS19M031	2, 14, 22	1.4-2.4
18	CS19M033	3, 17, 26	0.7-1.7
19	CS19M036	4, 11, 23	0-1
20	CS19M038	5, 14, 25	0.4-1.4
21	CS19M039	1, 14, 24	0.8-1.8
22	CS19M042	2, 15, 21	1.5-2.5
23	CS19M044	3, 14, 27	1-2
24	CS19M045	4, 12, 28	0.9-1.9
25	CS19M047	5, 13, 24	0.6-1.6
26	CS19M048	1, 15, 26	1.2-2.2
27	CS19M049	2, 10, 23	0.5-1.5
28	CS19M050	3, 16, 28	1-2

29	CS19M051	4, 15, 24	1.5-2.5
30	CS19M052	5, 12, 23	0.4-1.4
31	CS19M055	1, 16, 28	1.1-2.1
32	CS19M060	2, 17, 25	0-1
33	CS19M061	3, 15, 21	0.5-1.5
34	CS19M062	4, 16, 25	1.2-2.2
35	CS19M064	5, 11, 22	0.3-1.3
36	CS19M066	1, 17, 27	1.3-2.3
37	CS19M067	2, 16, 27	0.2-1.2
38	ED18B001	3, 13, 22	0.8-1.8
39	ED18B032	4, 17, 27	1.4-2.4
40	ME19S016	5, 10, 21	0.1-1.1
41	CS15B017	1, 12, 19	0.7 - 1.7

Task 2: Exponential distribution in R

In this assignment you will investigate the exponential distribution (1 dimensional) in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with *rexp(n, lambda)* where lambda is the rate parameter. The mean of the exponential distribution is *1/lambda* and the standard deviation is also *1/lambda*. Set **lambda** according to the values given in Table 2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal using plot.

   In point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.
4. Calculate the parameters specified in Table 2 for the exponential distribution.

Output

- Sample Mean versus Theoretical Mean: Include figures with titles. In the figures, highlight the means you are comparing. Include text (upto 50 words maximum) that explains the figures and what is shown on them, and provides appropriate numbers.
- Sample Variance versus Theoretical Variance: Include figures (output from R) with titles. Highlight the variances you are comparing. Explain your understanding of the differences of the variances analytically using equations/expressions.
- Distribution: Via figures and text, explain how one can tell the distribution is approximately normal.

Note: If code is obtained from the internet (not encouraged), please specify the source.

Table 2:

S. No.	Roll Nos.	lambda values	Parameters to be calculated
1	CS19M036 CS19M028 CS18M016 CS19M045 CS19M051	0.1	Quantile, Entropy
2	CS19D002 CS19M010	0.2	Moment-generating function (MGF),

	CS19M011 CS19M016 CS19M017		Characterizing Function
3	CS19M020 CS19M023 CS19M024 CS19M029 CS16B044	0.3	Fisher infromation, KL divergence
4	CS19M030 CS19M031 CS19M033 CS19M038 CH17D406	0.4	Quantile, MGF
5	CS19M039 CS19M042 CS19M044 CS19M047 CS18S038	0.5	Entropy, Fisher information
6	CS19M048 CS19M049 CS19M050 CS19M052 CS18S040	0.6	KL divergence, Entropy
7	CS19M055 CS19M060 CS19M061 CS19M064 ED18B032	0.7	Fisher information, Characterizing Function
8	CS19M066 CS19M067 ED18B001 CS19M062 ME19S016 CS15B017	0.8	MGF, Quantile