Report for CS215 Assignment 2

AMEEYA RANJAN SETHY - 200050006

October 12, 2021

Question-2

Algorithm for Sample Generation

A random vector X of dimension D × 1 is said to be a multivariate Gaussian if it can be expressed as AW + µ where µ is a vector of dimension D × 1, A is some vector of dimension D × N and W is a random vector composed of independent identically distributed univariate standard normal Random Variables. The covariance matrix C can be defined as

$$C = AA^T$$

So when covariance matrix C is given, if we can find the matrix A, then we know that the matrix X = AW + µ is a multivariate Gaussian with the given mean µ and covariance C. Since the matrix C is symmetric and positive semidefinite, it has an eigen decomposition which is of the form:

$$C = VDV^T = VD^{1/2}D^{1/2}V^T = AA^T$$

where V is an orthonormal eigenvector matrix and D is a diagonal eigenvalue matrix with non-negative diagonal entries. Now if we take $A = VD^{1/2}$, then we can see that :

$$AA^T = VD^{1/2}(VD^{1/2})^T = VD^{1/2}D^{1/2}V^T = C$$

Here $(D^{1/2})^T = D^{1/2}$ because D is a diagonal matrix. The step by step algorithm is written below:

Step 1: First we find the eigen matrix V and the diagonal matrix D consisting of the eigen values by using **eig()** function in MATLAB. The exact code is **[V,D] = eig(C)**.

Step 2: Now we calculate the values of matrix A using the formula $A = VD^{1/2}$

Step 3: Now we will calculate the MLE estimates for mean and covariance using the formula :

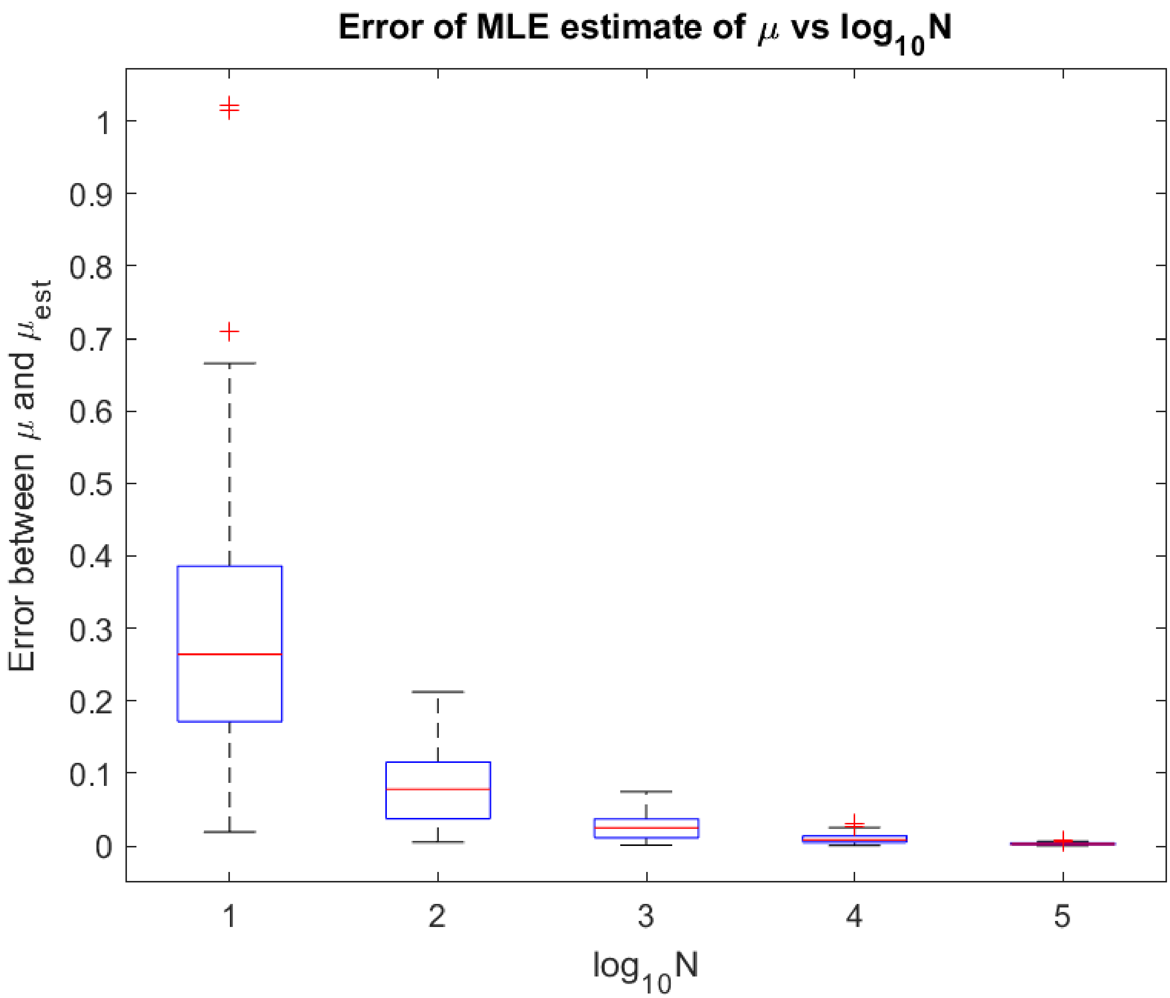
$$\mu_{est} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$C_{est} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_{est})(X_i - \mu_{est})^T$$

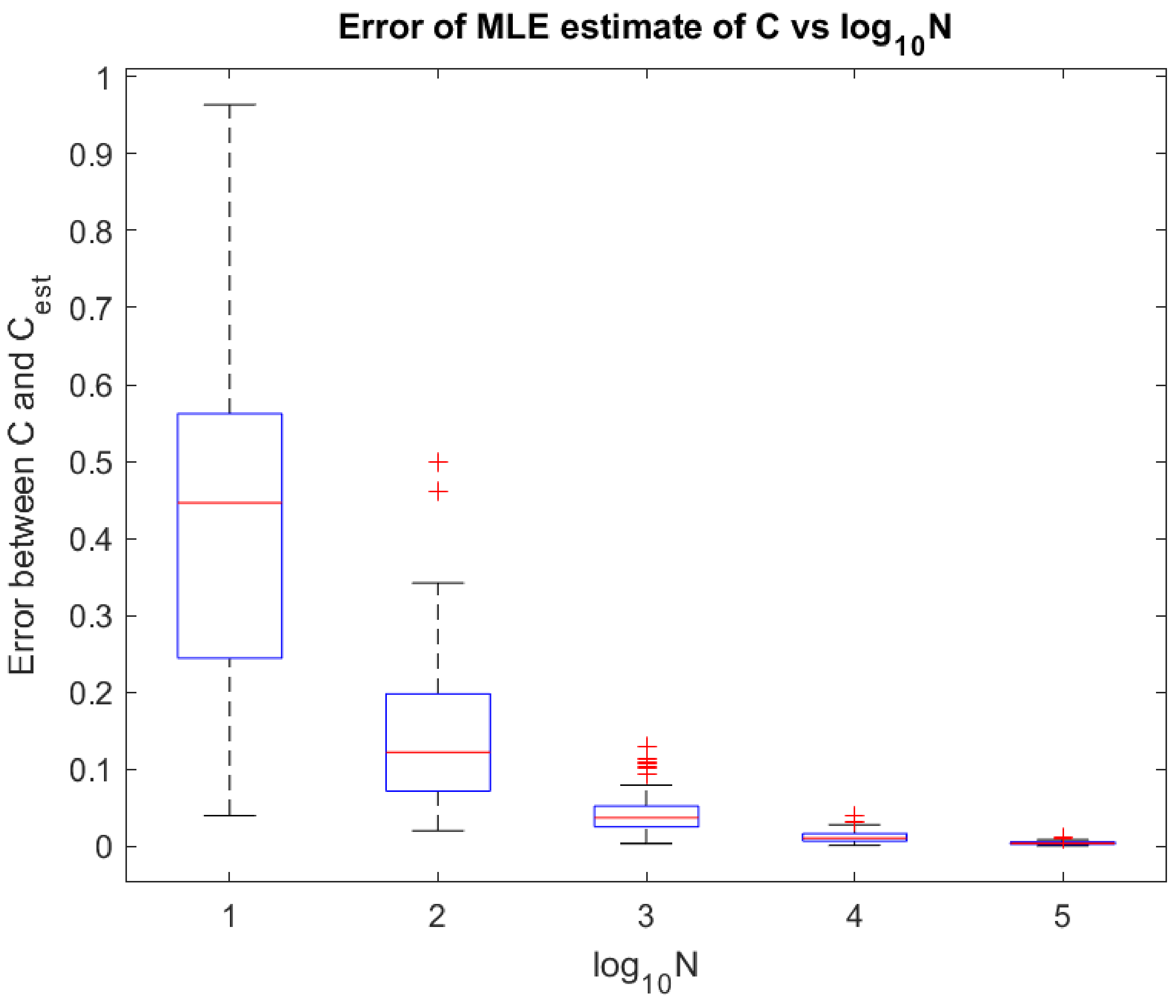
Step 4: Now we calculate the errors as per the formula given in problem statement using **norm()** function in MATLAB.

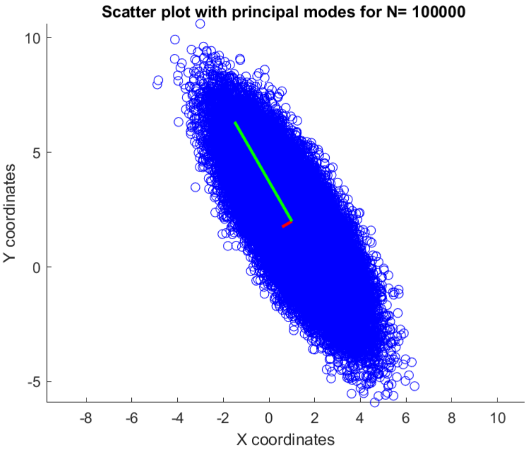
Step 5: Now for plotting the principal modes of variation, we can perform eigen decomposition of C_{est} and get V_{est} and D_{est} using the **eig()** function. The columns in V_{est} are the unit eigen vectors pointing in the direction of principal modes of variation and the corresponding diagonal values in D_{est} are the eigenvalues(also the variances). Thus using these, the required lines can be easily plotted.

Graph of Error of μ_{est} vs $log_{10}N$

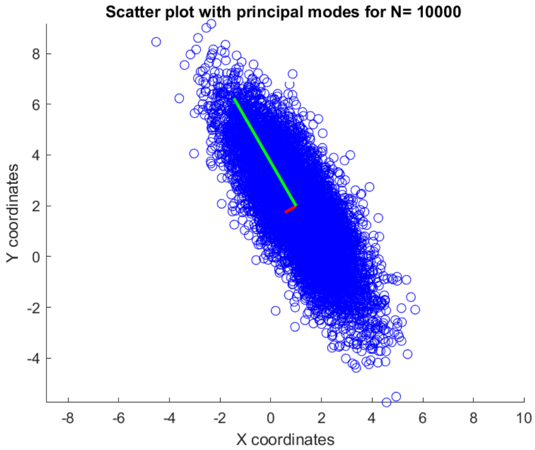


Graph of Error of C_{est} vs $log_{10}N$

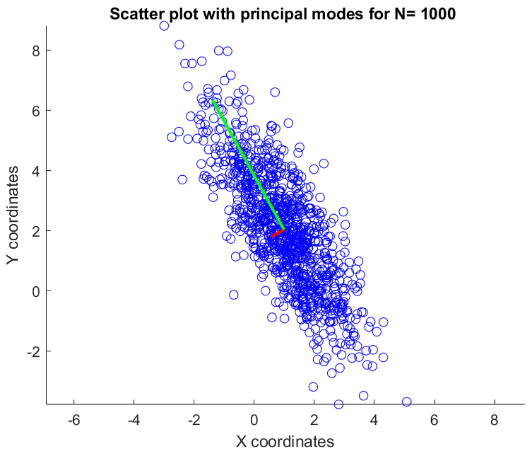




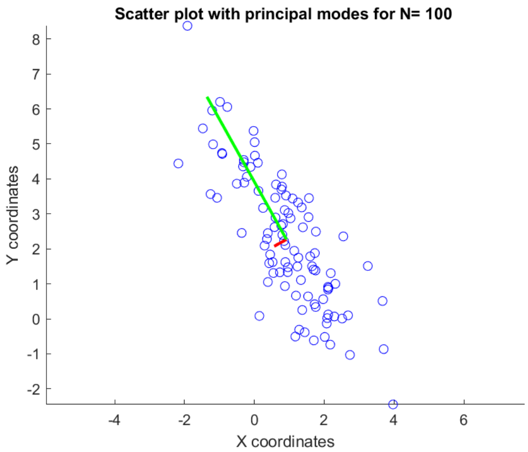
Scatter plot with 100000 points



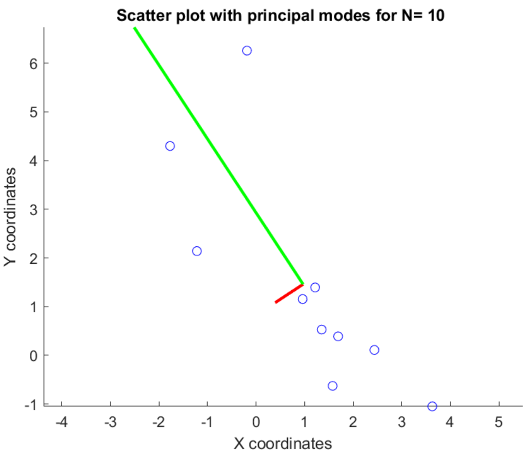
Scatterplot with 10000 points



Scatterplot with 1000 points



Scatterplot with 100 points



Scatterplot with 10 points