CS215 Assignment2 Problem 2

Parth Pujari and Anish Kulkarni

October 2022

1 Generating Samples for Multivariate Gaussians

Given a bivariate gaussian distribution with a given covariance matrix C and a mean vector μ, to sample N points with N varying from 10 to 10^5.
Method used:

- For a given value of N, generate N sets of standard bivariate Gaussian random variables (a pair of random variables) using the randn() function
- The bivariate gaussian distribution satisfies the probability density function

$$\frac{1}{2\pi|C|^{0.5}} exp(-(0.5)(X-\mu)^T C^{-1}(X-\mu))$$

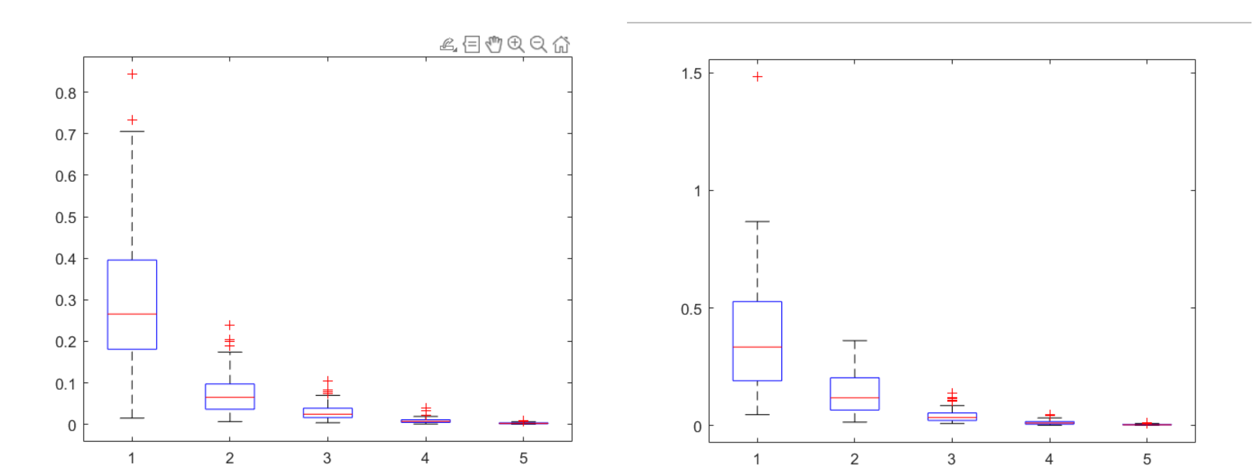
- Further, C being symmetric positive definite, we can write it as $P^{-1}DP$ where D is a diagonal matrix containing the eigen values of C and P is an orthogonal matrix (hence $P^{-1} = P^T$containing its eigenvectors and similarly C^{-1}
- Clearly the random variable $P^{-1}D^{-0.5}(X-\mu)$ is from a standard gaussian. Thus, our transform to find the random variable X is $D^{0.5}PY + \mu$ where Y is the standard bivariate gaussian that was generated randomly

2 Empirical Data

The maximum likelihood generated the likely mean and the covariance matrix of the transformed random variables.
For an experiment size of 100, the empirical mean and covariance were calculated and their errors were plotted for each value of N, the box plots of which are as follows:

3 Generating Scatterplots

Scatterplots were generated for each of the 5 values of N and plotted.
The empirical eigen values and eigen vectors were found and lines along the eigenvectors with lengths equal to the square root of the corresponding eigen values were drawn.
As is expected, the larger eigen value corresponds to the mode with the larger variance along the given eigenvector, hence the longer line and similarly for the smaller eigenvalue.
The scatter plots and the lines are as shown in the next page.



(a) Mean (b) Covariance

Figure 1: Boxplots

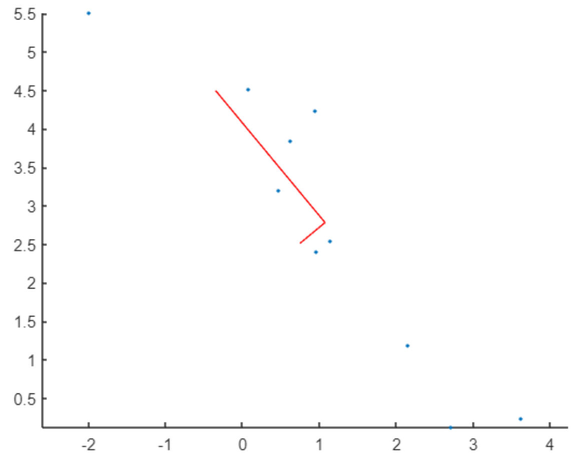


Figure 2: Scatter for N=10

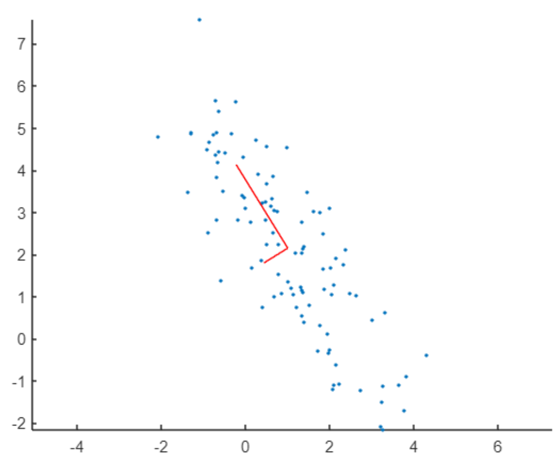


Figure 3: Scatter for N=10

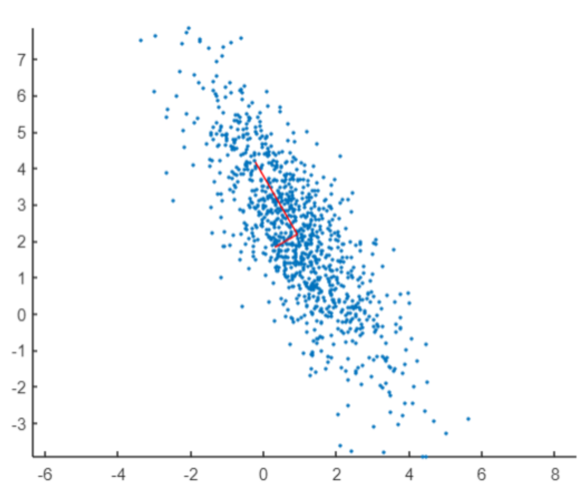


Figure 4: Scatter for N=100

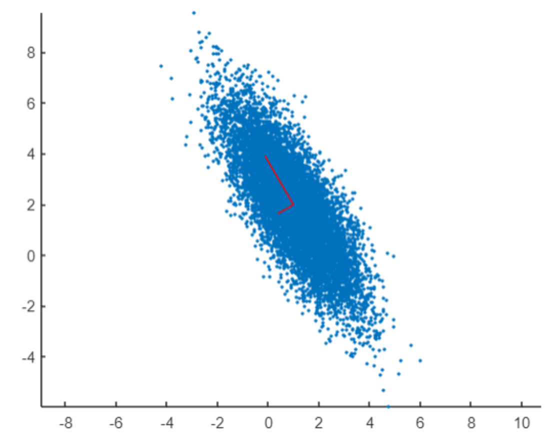


Figure 5: Scatter for N=1000

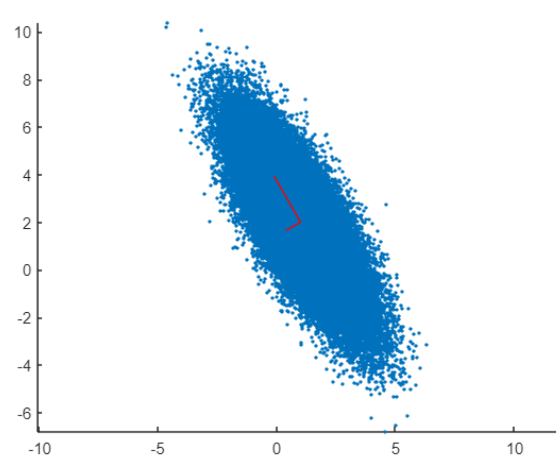


Figure 6: Scatter for N=10000