

COMP6245 : Lab 1 Report

Thanakorn Panyapiang(31446612)

October 12, 2019

1 Linear Algebra

The command `U @ U.T` produces a 3x3 identity matrix because of a symmetric property of B. Since B is a symmetric matrix, all eigenvectors of B are orthogonal to each other. Therefore, U, a matrix which each column represents a unit eigenvector of B (according to numpy specification), is an *orthogonal matrix*.

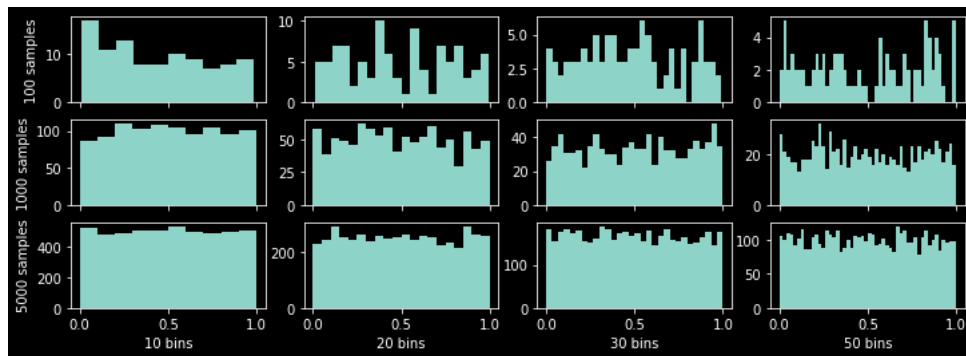
One property of an orthogonal matrix is its transpose is equal to its inverse ($A^T = A^{-1}$). Substituting this in the the command will get the result as follow:

$$U(U^T) = U(U^{-1})$$

$$U(U^{-1}) = I$$

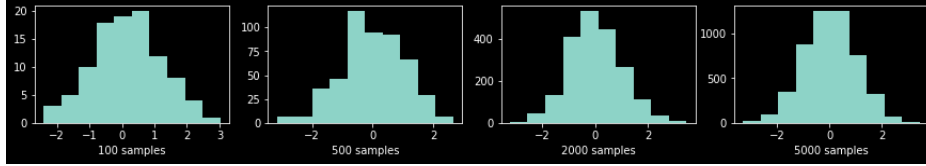
2 Random Numbers and Univariate Distributions

The distribution of uniform random samples with different numbers of samples and bins is as follow

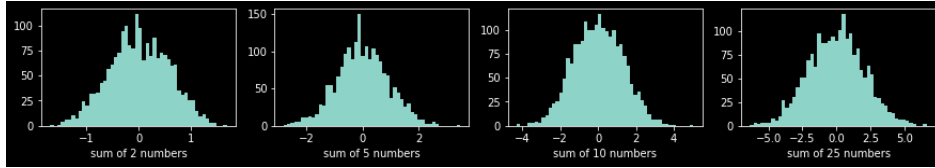


It can be observed that, when the number of samples is high enough, data is distributed equally regardless of the number of bins which causes the variation within bin counts to be low.

For samples drawn from a Gaussian distribution, the density of data, regardless of the number of samples, is higher when the value is close to mean and most of the data is located within 2 standard deviation(-2 to 2 in this case).

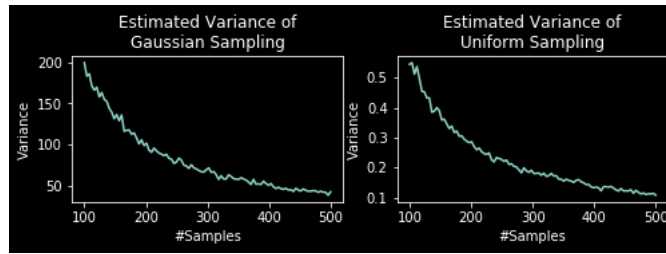


In the samples where data is generated by summation of uniform random numbers, the distribution of data is similar to a Gaussian distribution. What can be clearly observed is that when the number of numbers to add/subtract is increased, the distribution comes closer to a normal distribution.



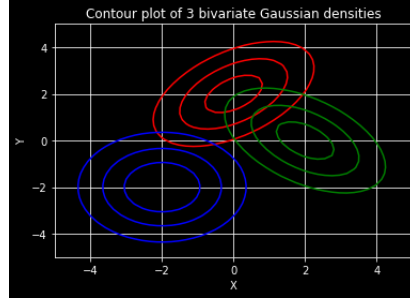
3 Uncertainty in Estimation

The number of samples and the variance has an inverse relation for both sampling methods. Although the trends of variance over the number of samples are similar in both samplings, it can be noticed that uniform sampling has a significantly lower variance compare to Gaussian sampling.



4 Bivariate Gaussian Distribution

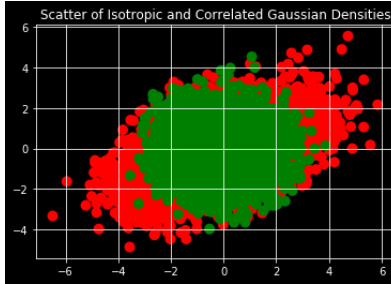
From the contour plot, \bar{X} and \bar{Y} are shifted by the value of the first and the second element of each matrix m as show in the figure below.



For the covariance matrices, the diagonal elements affect the distribution scale of each variables while the off-diagonal elements change the correlation between two variables. From the graph, the size of all three contours is scaled up by 2 since diagonal elements of all matrices are 2. However, the correlations between X and Y are different from each other. The correlation is direct variation and inverse variation for the covariance matrix C1(red) and C2(green) respectively while on C3(blue) both variables are independent from each other.

5 Sampling for Multivariate Gaussian Distribution

The scatter plot of X and Y is shown in the figure below.

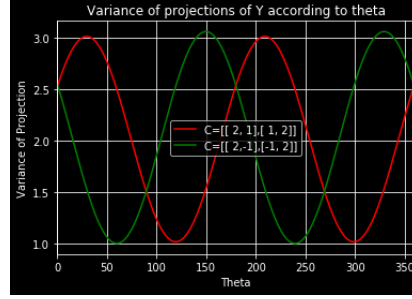


Although X(green) and Y(red) have a normal distribution, their correlations of two variables are different. Both variables are independent from each other in X as illustrated in the shape of the scatter plot which is a normal circle. On the other hand, both variables are positively correlated(the density is high when the first and the second variable is close to each other) in Y. This is the result of a linear transformation of X by the cholesky factor of the covariance matrix

$$C = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

6 Distribution of Projections

The figure below shows the projected variance of Y as a function of θ



From the graph, it can be observed that the maximum and minimum variance of the projected data is 3 and 1 for both datasets, but the directions which gives the maximum and minimum variance are different. Another thing that can be noticed is the eigenvalues of both covariance matrices are equal to the maximum and minimum projected variance. Also, when the data is projected into the direction of an eigenvector, the projected variance is equal to its corresponding eigenvalue.