Lab One: Exploring Properties of Multi-Variate Gaussian Densities

name: Huizhi Yu
number: ********
email: ********

Abstract

This experiment reviewed essential concepts from linear algebra, probability, and statistics, focusing on the properties of multi-variate Gaussian densities, particularly in the context of sampling and projection. The analysis was conducted using Python programming in a Jupyter Notebook environment.

1 Linear Algebra

1.1 Theoretical Background of vector and matrix

- **Scalar Product and Angle Between Vectors:** The scalar product (dot product) of two vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as:

$$a = \mathbf{x}^T \mathbf{y}$$

The angle θ between two vectors can be computed using the scalar product and the norms of the vectors:

$$\theta = \arccos\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}\right)$$

- **Transpose and Inverse of Matrices:** The transpose of a matrix $\mathbf{B}$, denoted as $\mathbf{B}^T$, is obtained by flipping the matrix over its diagonal. This operation transforms the rows of the matrix into columns. The inverse of a matrix $\mathbf{B}$ is a matrix $\mathbf{B}^{-1}$ such that:

$$\mathbf{B} \cdot \mathbf{B}^{-1} = \mathbf{I}$$

where $\mathbf{I}$ is the identity matrix. Not all matrices are invertible; a matrix must be square and have a non-zero determinant to possess an inverse.

- **Eigenvalues and Eigenvectors:** For a matrix $\mathbf{B}$, if there exists a scalar λ and a non-zero vector $\mathbf{u}$ such that:

$$\mathbf{B}\mathbf{u} = \lambda \mathbf{u}$$

then λ is the eigenvalue and $\mathbf{u}$ is the corresponding eigenvector.

1.2 Applications

1.2.1 Observations from the Command

To verify the properties of the matrix using Python, we defined the matrix B as follows:

$$B = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 6 & 5 \\ 1 & 5 & 9 \end{bmatrix}$$

Using the command `D, U = np.linalg.eig(B)`, we computed the eigenvalues D and eigenvectors U of the matrix. To verify the orthogonality of the eigenvectors, we calculated the scalar product of two eigenvectors using `print(np.dot(U[:,0], U[:,1]))`. The result was `-0.00`. This value is very close to zero, indicating that the eigenvectors are orthogonal to each other. Since B is symmetric ($B = B^T$), we expect its eigenvectors to be orthogonal to each other. The result is consistent with this expectation.

1.2.2 Proof of Orthogonality

For a symmetric matrix B, its eigenvalues λ_1, λ_2 and corresponding eigenvectors $\mathbf{u}_1, \mathbf{u}_2$ satisfy:

$$B\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$$

$$B\mathbf{u}_2 = \lambda_2 \mathbf{u}_2$$

We compute the following inner products:

$$\mathbf{u}_1^T B \mathbf{u}_2 = \mathbf{u}_1^T (\lambda_2 \mathbf{u}_2) = \lambda_2 \mathbf{u}_1^T \mathbf{u}_2$$

$$\mathbf{u}_2^T B \mathbf{u}_1 = \mathbf{u}_2^T (\lambda_1 \mathbf{u}_1) = \lambda_1 \mathbf{u}_2^T \mathbf{u}_1$$

The symmetry of matrix B implies:

$$\mathbf{u}_1^T B \mathbf{u}_2 = \mathbf{u}_2^T B \mathbf{u}_1$$

Thus, we have:

$$\lambda_2 \mathbf{u}_1^T \mathbf{u}_2 = \lambda_1 \mathbf{u}_2^T \mathbf{u}_1$$

The symmetry of inner products allows us to write:

$$\mathbf{u}_2^T \mathbf{u}_1 = \mathbf{u}_1^T \mathbf{u}_2$$

Therefore, we can rewrite the equation as:

$$\lambda_2 \mathbf{u}_1^T \mathbf{u}_2 = \lambda_1 \mathbf{u}_1^T \mathbf{u}_2$$

Since the eigenvalues are distinct (i.e., $\lambda_1 \neq \lambda_2$), for the above equality to hold, it must be that:

$$\mathbf{u}_1^T \mathbf{u}_2 = 0$$

This shows that the eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal.

2 Probability and Statistics

2.1 Applications

2.1.1 Histogram of 1000 uniform random numbers

In this section, we utilized the `NumPy` library in `Python` to generate 1000 uniform random numbers. The generated random numbers were used to create histograms with the `Matplotlib` library. We set two different bin configurations for the histograms: 4 bins and 40 bins, in order to compare the impact of different bin settings on the histogram representation. After executing the code, we obtained the following histograms **Figure 1:**

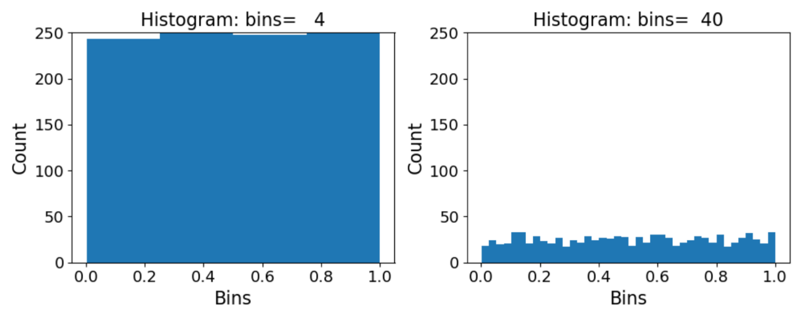


Figure 1: histograms of 1000 uniform random numbers

- **Reasons for Non-Flat Histogram:** This may be due to the natural fluctuations inherent in random number generation. In a finite sample, even if the data is derived from a uniform distribution, the differences among the samples can lead to slight fluctuations in the shape of the histogram.
- **Variations in Histogram with Each Run:** Each time the histogram is generated, it appears slightly different because the random numbers produced are not the same each time. Due to the variability of random numbers, the heights and shapes of the histogram can change with each run of the code.
- **Impact of Sample Size on Observations:** If we were to start with a larger dataset (for example, 10,000 random numbers), we would expect the histogram to appear more flat and closer to the theoretical uniform distribution, like **Figure 2**. This shows the Law of Large Numbers, which states that as the sample size increases, sample statistics (such as mean and variance) tend to converge to the true parameters.

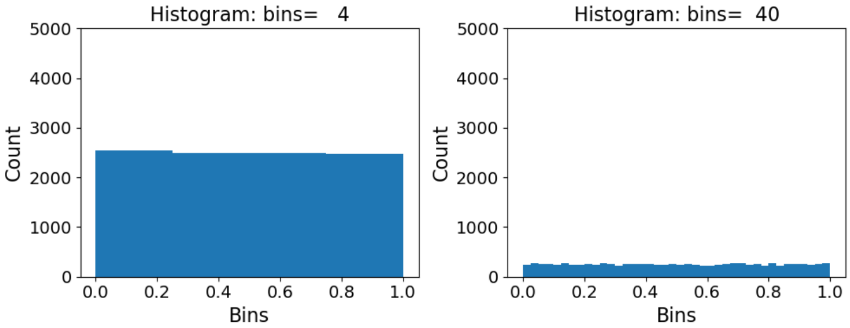


Figure 2: histograms of 10000 uniform random numbers

2.1.2 Effect of adding and subtracting uniform random numbers

In this section, we analyze the sum and difference of random numbers generated from a uniform distribution. We set $N = 1000$ and create an array x_1 to store the results. For each iteration, we sum 12 uniform random numbers between 0 and 1 and subtract another 12. The histogram of x_1 shows a bell-shaped curve, suggesting it follows a normal distribution. As we can see from the **Figure 3**, when changing the number of random numbers: With fewer numbers (e.g., 2), the histogram is more variable, with higher frequencies at the extremes. As the number increases (e.g., 12 or 30), the histogram smooths out, concentrating data toward the center and forming a bell shape. This indicates higher probabilities for values near the center and lower probabilities at the extremes.

This behavior is explained by the **Central Limit Theorem**. The theorem states that the sum of independent and identically distributed random variables approaches a normal distribution as the number of variables increases, regardless of their original distribution. In this experiment, using multiple uniformly distributed random numbers demonstrated that as their count increases, the histogram approximates a normal distribution, confirming the Central Limit Theorem's practical application.

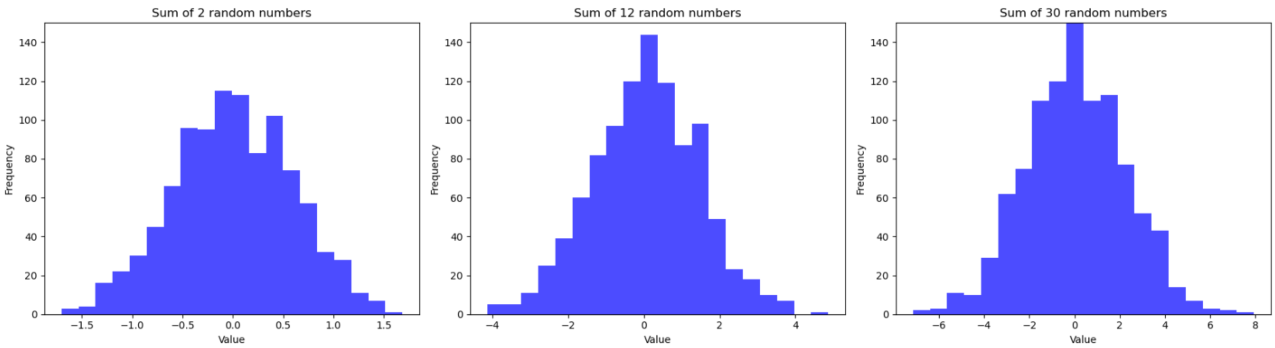


Figure 3: adding and subtracting uniform random numbers

2.1.3 Uncertainty in Estimation

The results demonstrate a clear downward trend in the variance of variance estimates as the sample size increases. This observation can be explained by understanding the properties of statistical estimators.

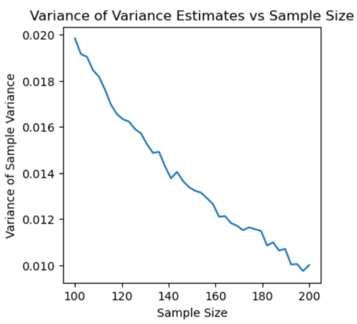


Figure 4: Variance of Variance Estimates vs Sample Size

When we compute the variance from a set of samples, each sample contributes to the estimate, introducing variability. However, as we increase the sample size, the estimates become more stable due to the Law of Large Numbers, which states that as the number of trials increases, the sample mean will converge to the expected value. Consequently, the variability in the estimated variance decreases like **Figure 4**, resulting in a lower variance of the variance estimates.

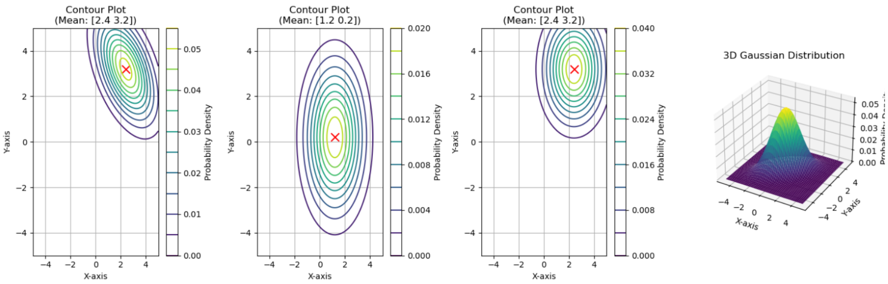
In practical terms, this means that larger sample sizes yield more reliable estimates of the population variance, reducing the uncertainty associated with our parameter estimation. Thus, our findings confirm that increasing the sample size leads to improved accuracy and consistency in statistical estimations.

3 Multi-variate Gaussian Distribution

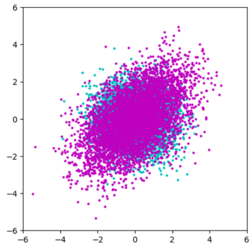
3.1 Appilications

3.1.1 Visualization of Gaussian Distributions

This **Figure 5(a)** presents three two-dimensional Gaussian distributions. The first distribution has a mean of $(2.4, 3.2)$ and shows a correlated spread. The second distribution, centered at $(1.2, 0.2)$, demonstrates a wider spread along the y-axis. The third distribution, with a mean of $(2.4, 3.2)$, exhibits an isotropic spread. It highlights the differences in variability and orientation among the distributions.
To generate multivariate Gaussian distribution data from a standard normal distribution, we begin by obtaining the lower triangular matrix of the covariance matrix through Cholesky decomposition. Next, we generate samples from the standard normal distribution. Finally, we transform these samples into multivariate Gaussian samples with the specified covariance matrix using matrix multiplication. **Figure 5(b)** illustrates the shapes and differences between the two distributions.



(a) Visualization of 2D and 3D Gaussian Distribution



(b) Comparison

Figure 5: Visualization of Gaussian Distribution

3.1.2 Distribution of Projections

To calculate the projected variance of the input data matrix Y along different directions. First, it generates a series of unit vectors, then computes the projected variance for each direction and stores the results in an array. Finally, it plots a curve showing how the variance changes with direction, illustrating the extent of data spread in various orientations. The maxima and minima of the plot from **Figure 6** represent the highest and lowest variances of the projected data along different directions. The maxima correspond to the directions of greatest spread in the data, while the minima correspond to directions of lesser spread.

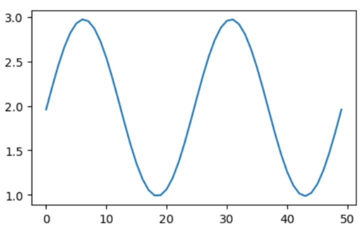


Figure 6: Visualization of Projected Variance Changes

The eigenvalues of the covariance matrix indicate the variance along the directions of their corresponding eigenvectors, which represent the principal directions of the data. There is a direct relationship: the maxima of the projected variance occur in the directions of the eigenvectors associated with the largest eigenvalues, while the minima occur in the directions of the eigenvectors associated with the smallest eigenvalues.
The relationship between eigenvalues and projected variance can be expressed through the formula $V(\theta) = \lambda_1 \cos^2(\theta - \theta_1) + \lambda_2 \cos^2(\theta - \theta_2)$, where λ_1 and λ_2 are the eigenvalues and θ_1 and θ_2 are the angles of the corresponding eigenvectors. This periodicity in variance, represented by the $\cos^2$ function, suggests that the projected variance changes in a sinusoidal manner as the angle θ varies, which aligns with the properties of two-dimensional Gaussian distributions.

4 Summary

In the linear algebra section, we discussed fundamental concepts such as scalar products, matrix transpose and inverse, as well as eigenvalues and eigenvectors. We validated the orthogonality of eigenvectors for a symmetric matrix, illustrating the theoretical foundations of linear algebra through computational methods.

In the probability and statistics section, we conducted experiments with uniform random numbers, analyzing histograms generated from samples of varying sizes. We observed that while smaller samples led to less stable histogram shapes, larger samples demonstrated convergence towards a uniform distribution, in accordance with the Law of Large Numbers. Furthermore, we applied the Central Limit Theorem to investigate the distribution of sums and differences of random numbers, showcasing the emergence of a normal distribution.

Finally, in the multivariate Gaussian distribution section, we visualized different Gaussian distributions and examined the projected variance along various directions. We demonstrated how eigenvalues correlate with variance and data spread, providing insight into the relationships between these statistical properties.