Zewail City of Science and Technology
Communication and Information Engineering Program

Probability and Stochastic Processes - CIE 327

Random Variables and Their Statistics

CIE 327: SEMESTER PROJECT

Course Instructor: Prof. Samy Soliman

Done by:

Mohammed Ali 202200594

Abdelhady Mohamed 202201172

Ahmed Abdallah 202202017

Report:

1. Abstract

This project focuses on the analysis of random variables and their statistics, leveraging a graphical user interface (GUI) for intuitive interaction and visualization. The developed tool allows users to input random variables, evaluate their statistical properties, and visualize distributions and cumulative distribution functions (CDFs). Advanced features include generating joint statistics for multiple random variables and exploring functions of random variables to compute derived probability distributions. The project aims to enhance understanding of probability and stochastic processes through dynamic, user-friendly visualizations. Comprehensive testing ensures robustness across diverse statistical scenarios, including uniform, normal, binomial, and Poisson distributions. This initiative bridges theoretical concepts with practical applications, fostering deeper insight into the behavior of random variables.

Considerations for the Test Cases: Sample Size: The accuracy of these estimates depends on the sample size used in the simulation or analysis. A larger sample size would generally lead to more precise estimates.

2. Implementation

2.1 Section 1

This section will contain the implementation logic of section 1, first, the user will be able to select a file containing a random variable in the form of its sample space, no matter the number of samples, the length of the sample space will be calculated then the random variable will be passed to **RVtype** function that will determine if the RV is discrete or continuous.

If the RV is discrete the program will calculate its PDF using the histogram method using the histcounts() function in Matlab, then calculate the PDF using cumsum() which does the job of the SIGMA or cumulative sum then plot both.

Then create linespaces using the linespace() function in Matlab to prepare the points in which the MGF and their derivatives will be calculated, the MGF and its derivatives are calculated using simulation of the theoretical forms of the MGF, 1st, and 2nd derivatives and calculate the value of it at each point on the line space, finally plot the graph of the MGF and its derivatives.

we calculated the mean and variance of the RV using built-in functions to verify the calculated first and second moments at time t = 0, and we calculated the 3rd moment as the E[X^3] or the mean cubed.

If the RV was Continous the PDF will be calculated using the KDE method, and since we are not allowed to use MATLAB built-in functions for statistical analysis that does advanced calculations we had to make a user-defined KDE logic function which is **ksDen()** function and calculate the CDF using **cumtrapz()** function that simply simulated the integration process over all the data points of PDF, the same method used to calculate the MGF and its derivatives using **tarpz()** function, and a similar logic of the discrete case to calculate and verify the mean, variance, first, second, and third moments.

2.2 Section 2

The logic of implementation of the marginal PDF for X and Y Rvs is quite similar to the one we used to plot and calculate the PDF in section 1 so I will not stress that point again.

The implementation of joint PDF is quite similar to the histogram method but we used histcounts2() a function similar to histcount 1 but returns a 2D histogram.

2.3 Section 3

We used the same logic in calculating the PDFs of Z and W as in Section 1 and the Joint PDF of Z and W as in Section 2, the only difference in this section is the fact that Z and W are functions of X and Y RVs loaded from the user. The user can determine Z and W as functions of X and Y in GUI as needed in the **Bonus** part.

3. Important Functions

histcounts();

- **Description:** Bins the data into discrete categories and normalizes it to form a probability mass function (PMF).

histcounts2();

- **Description:** Computes the 2D histogram for X and Y, normalized to produce a joint probability distribution. • Benefit: Captures the relationship between X and Y, showing how likely different combinations of their values are.

ksDen();

- **Description:** Uses kernel density estimation logic with a gaussian kernel to estimate the probability density function (PDF).Provides a smooth approximation of the continuous distribution.
- **Idea:** Set up evaluation points for the PDF using linspace. Loops through each evaluation point and computes the kernel contribution from each data point. Normalize the result by the sample size and bandwidth.

RVtype();

- **Description:** Calculates the ratio of unique values to the total number of values in the RV. If the ratio is above a threshold (e.g., 0.05), the RV is classified as continuous. Otherwise, it's discrete.

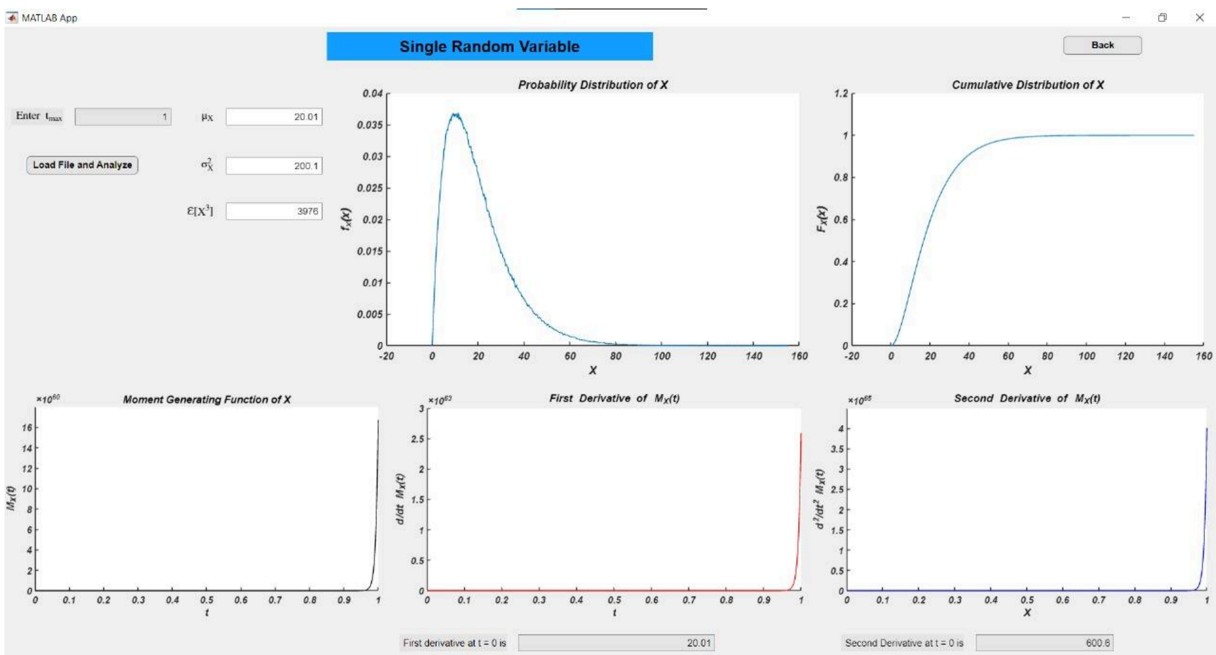
gaussian_kernal();

- **Description:** Computes the Gaussian kernel function $K(x)= (1/\sqrt{2\pi})\cdot e((-1/2)\cdot x\text{\textasciicircum}2)$ that enhances the analysis of continuous data by smoothing out noise.

4. Testcases Outputs and Comments

4.1. Section 1

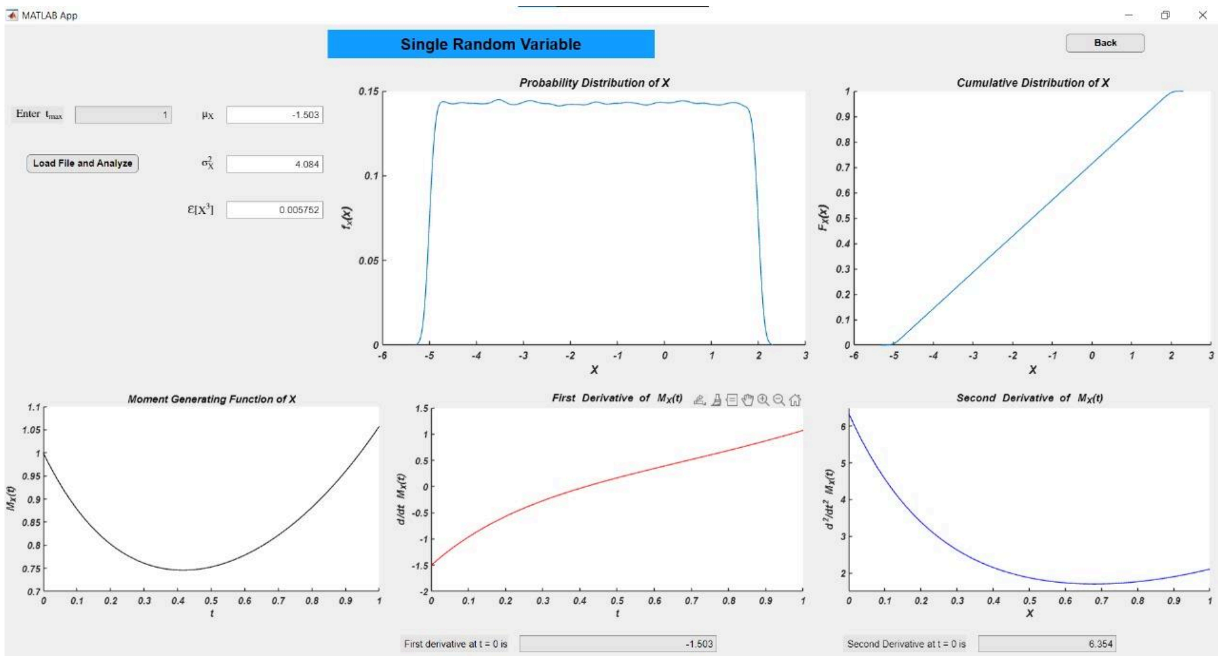
4.1.1 Test Case 1: unknown RV



Comments:

- The value of the mean which is calculated by the built-in MATLAB function matches the value of the first moment at t = 0.
- The value of the 2nd moment which = 600 = the variance + mean squared = 200+20^2 = 600 approximated.

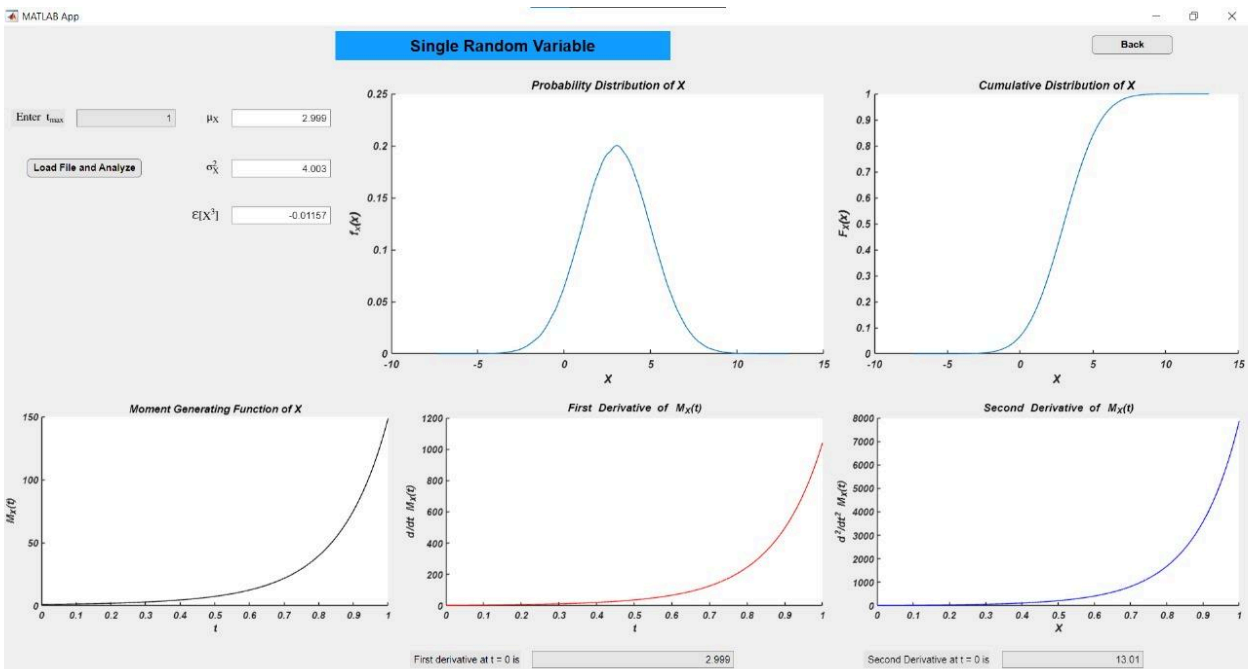
4.1.2 Test Case 2: X ~ U(-5, 2)



Comments:

- We are given a uniform (a,b) RV with mean = (a+b)/2 which matches with the value of the calculated by mean function and the first derivative of MGF at t = 0. Which equals -1.5
- and variance = (b-a)^2/12, = the result 2nd moment - mean^2 approximated = 4

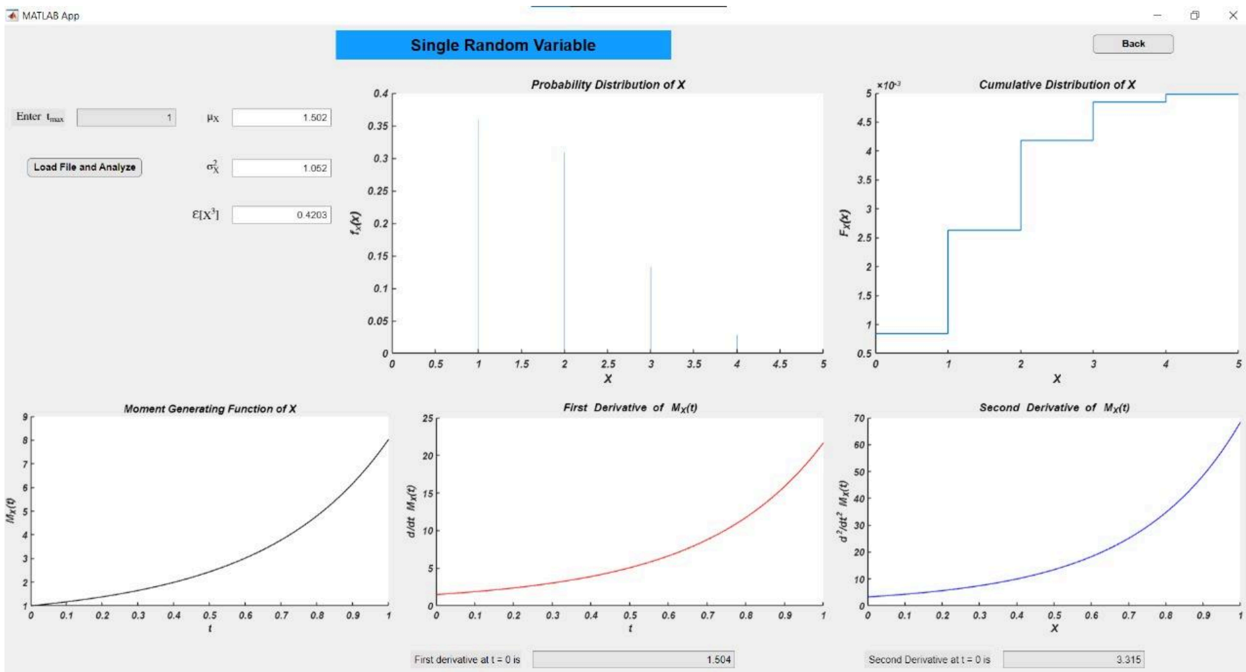
4.1.3 Test Case 3: X3 ~ N(3, 4)



Comments:

- The mean calculated by the built-in MATLAB function matches the value of the first moment at t = 0, which is 3, and matches the value of the 1st moment at t = 0
- The variance, calculated as the second moment minus the square of the mean, approximates 4, matching the given variance of the normal distribution.
- Since the 2nd moment is the E[X^2], we can say that it equals the variance - mean^2 = 13, which roughly matches the value calculated in our program

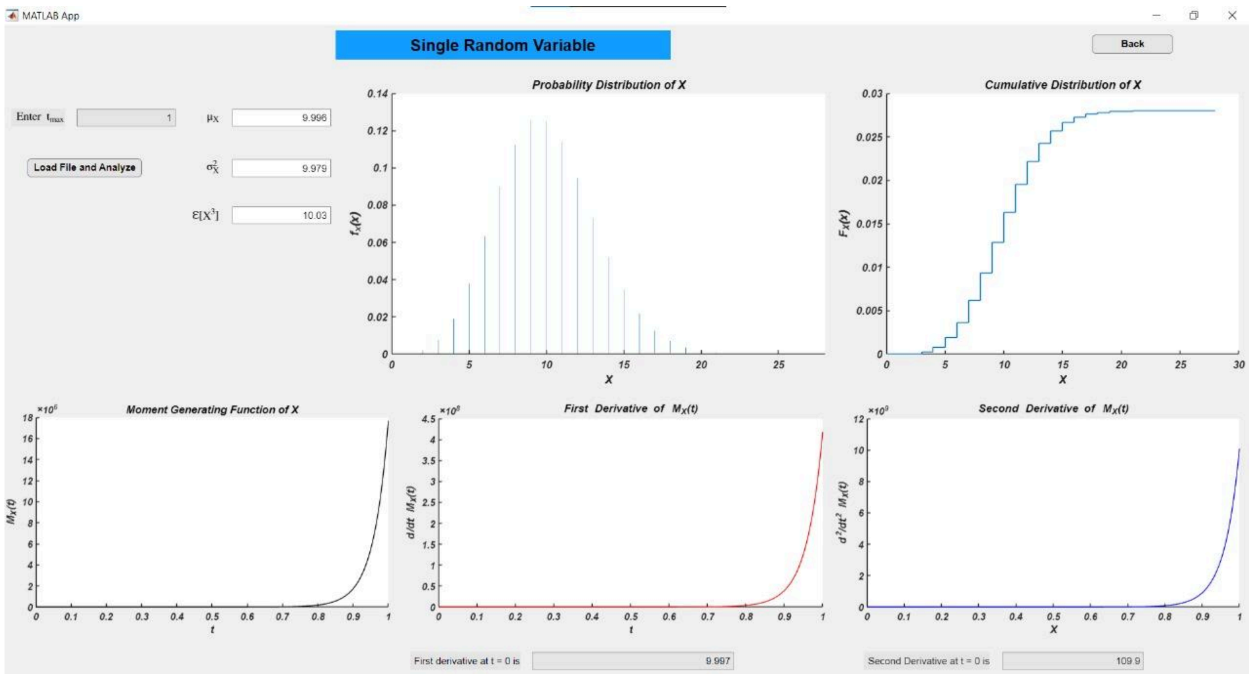
4.1.4 Test Case 4: X4 ~ Bin(5, 0.3)



Comments:

- The mean calculated using the built-in MATLAB function and the first moment at t=0 are both equal to np = 5*0.3 = 1.5.
- The second moment, minus the square of the mean, approximates the variance, which is np(1-p) = 5*0.3*(1-0.3) = 1.05

4.1.5 Test Case 5: X5 ~ Poisson(10)

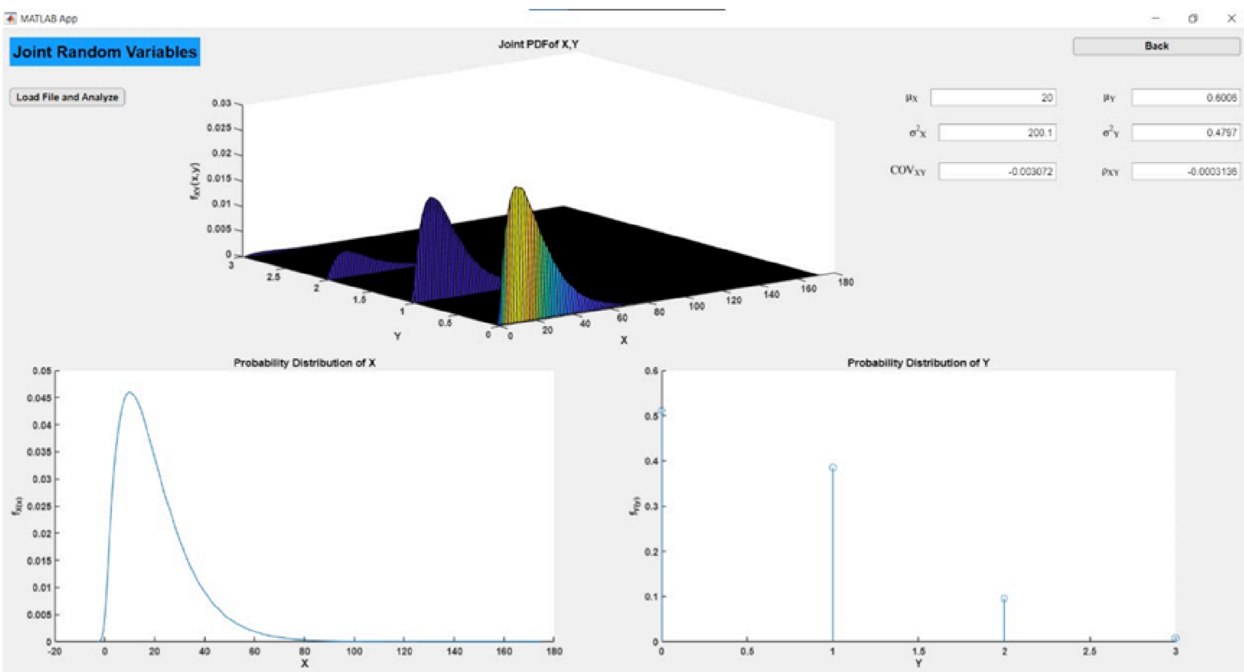


Comments:

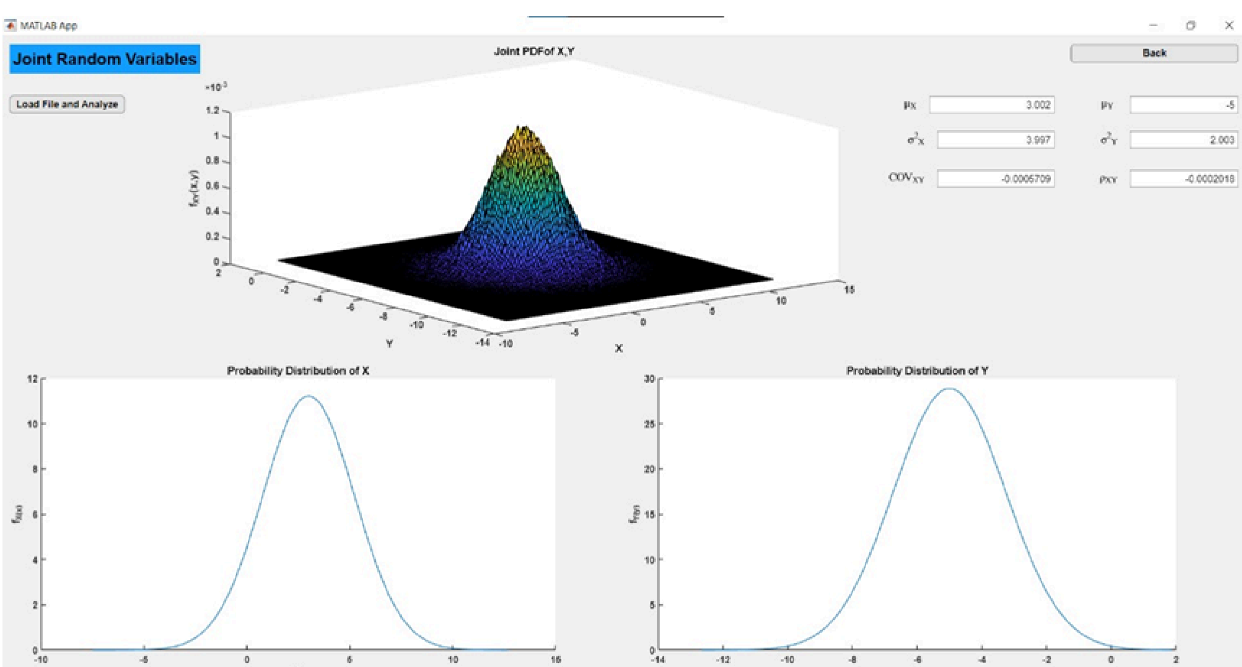
- The mean calculated via the MATLAB function and the first moment at t = 0 are both equal to λ = 10.
- The variance, calculated as the second moment minus the mean squared, approximates 10, which matches the variance of a Poisson distribution with parameter λ = 10

4.2 Section 2

4.2.1 case 1 unknown RVs



4.2.2 case 2 X ~ N(3; 4) and Y ~ N(-5; 2)



Comments:

Theoretical Values:

X ~ N(3, 4):

- Mean: $E[X] = \mu x = 3$
- Variance: $Var(X) = \sigma x^2 = 4$
- Standard Deviation: $\sigma x = \sqrt{4} = 2$

Y ~ N(-5, 2):

- Mean: $E[Y] = \mu y = -5$
- Variance: $Var(Y) = \sigma y^2 = 2$
- Standard Deviation: $\sigma y = \sqrt{2} \approx 1.414$

Measured Results:

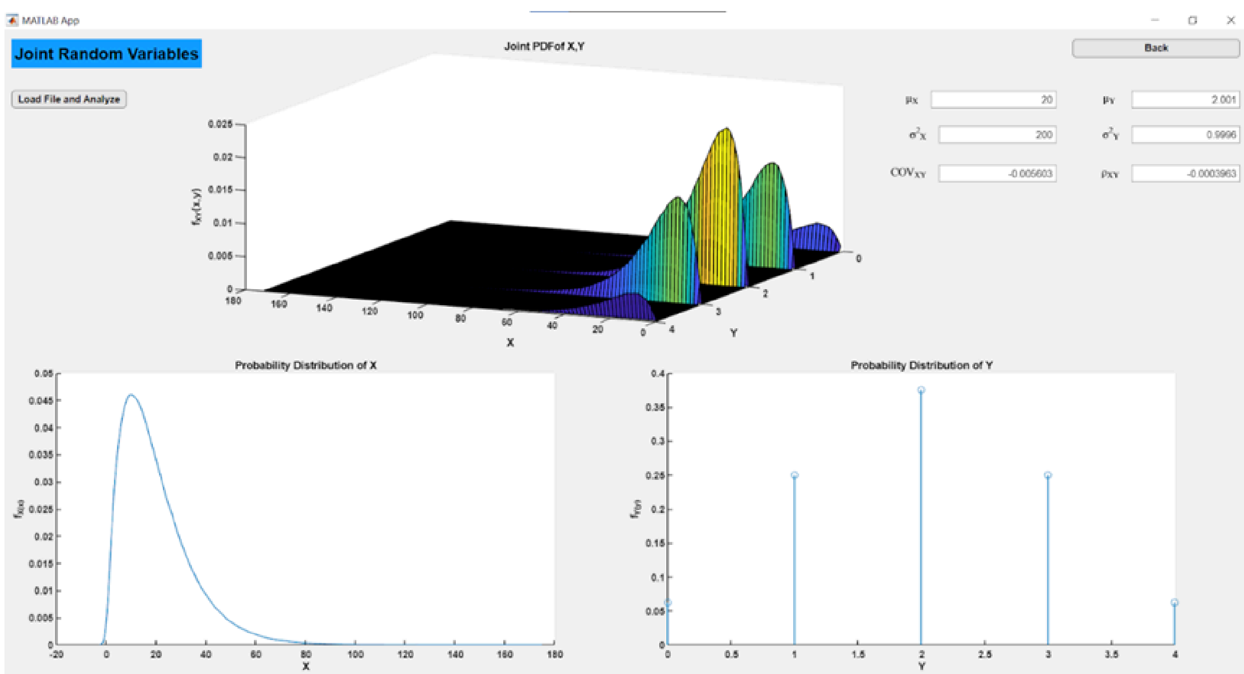
- Mean of X (μx): 3
- Mean of Y (μy): -5
- Variance of X (σx^2): 4
- Variance of Y (σy^2): 2
- Covariance of X and Y ($COVxy$): -0.005729
- Correlation Coefficient (ρxy): -0.002031

Observations:

- Means: The mean of X is 3, which aligns with the theoretical mean of a Normal distribution with $\mu x = 3$. The mean of Y is -5, which aligns with the theoretical mean of a Normal distribution with $\mu y = -5$.
- Variances: The variance of X is 4, which matches the theoretical variance of a Normal distribution with $\sigma x^2 = 4$. The variance of Y is 2, which matches the theoretical variance of a Normal distribution with $\sigma y^2 = 2$.
- Covariance and Correlation: The covariance is -0.005729, indicating a very weak negative relationship between X and Y. The correlation coefficient is -0.002031, which is extremely close to zero, further confirming the negligible correlation.

Comparison with Theoretical Values: The output values for the means and variances of X and Y agree excellently with the theoretical values, suggesting that the simulation or analysis has been conducted accurately. The very low covariance and correlation coefficient support the assumption of independence between X and Y.

4.2.3 case 3 X ~ Gamma(2; 10) and Y ~ Bin(4; 0.5)



Comments:

Theoretical Values:

X ~ Gamma(2, 10):

- Mean: E[X] =2 * 10 = 20
- Variance: Var(X) = 2 * 10^2 = 200

Y ~ Bin(4, 0.5):

- Mean: E[Y] = n * p = 4 * 0.5 = 2
- Variance: Var(Y) = n * p * (1 - p) = 4 * 0.5 * (1 - 0.5) = 1

Measured Results:

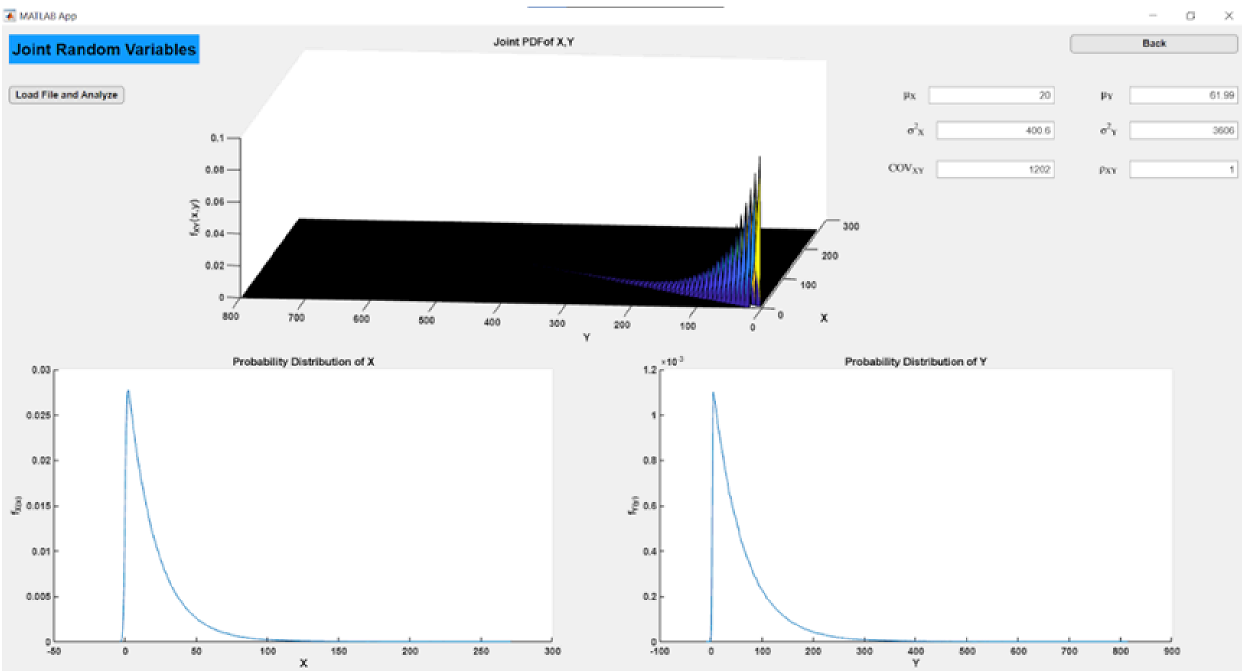
- Mean of X (μx): 20
- Mean of Y (μy): 2.001
- Variance of X (σx^2): 200
- Variance of Y (σy^2): 0.9996
- Covariance of X and Y (COVxy): -0.005603
- Correlation Coefficient (ρxy): -0.0003963

Observations:

- Means: The mean of X is 20, which aligns with the theoretical mean of a Gamma(2, 10) distribution. The mean of Y is 2.001, very close to the theoretical mean of 2 for a Bin(4, 0.5) distribution.
- Variances: The variance of X is 200, which matches the theoretical variance of a Gamma(2, 10) distribution. The variance of Y is 0.9996, very close to the theoretical variance of 1 for a Bin(4, 0.5) distribution.
- Covariance and Correlation: The covariance is -0.005603, indicating a very weak negative relationship between X and Y. The correlation coefficient is -0.0003963, which is extremely close to zero, further confirming the negligible correlation.

Comparison with Theoretical Values: The output values for the means and variances of X and Y agree excellently with the theoretical values, suggesting that the simulation or analysis has been conducted accurately. The very low covariance and correlation coefficient support the assumption of independence between X and Y.

4.2.4 case 4 X ~ Exp(0:05) and Y = 3X + 2



Comments:

Theoretical Values:

X ~ Exp(0.05):

- Mean: E[X] = 1 / λ = 1 / 0.05 = 20
- Variance: Var(X) = 1 / λ^2 = 1 / 0.05^2 = 400

Y = 3X + 2:

- Mean: E[Y] = E[3X + 2] = 3E[X] + 2 = 3 * 20 + 2 = 62
- Variance: Var(Y) = Var(3X + 2) = 3^2 * Var(X) = 9 * 400 = 3600

Measured Results:

- Mean of X (μx): 20
- Mean of Y (μy): 61.99
- Variance of X (σx^2): 400.6
- Variance of Y (σy^2): 3606

- Covariance of X and Y (COVxy): 1202
- Correlation Coefficient (ρxy): 1

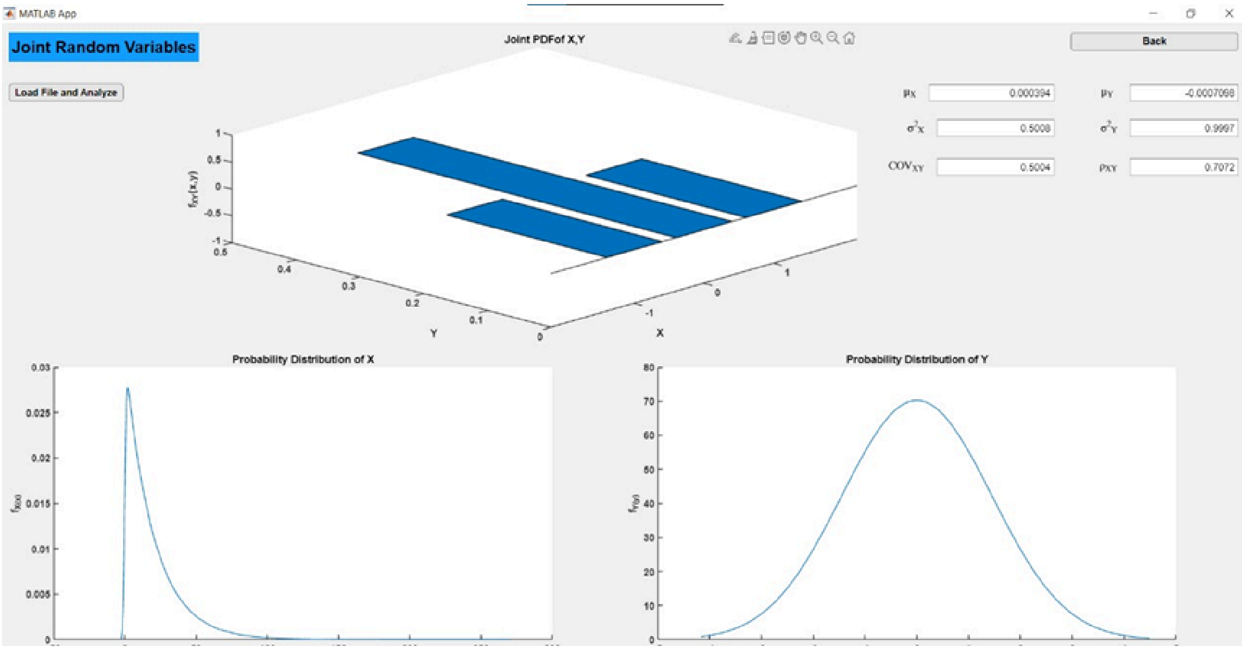
Observations:

- Means: The mean of X is 20, which aligns with the theoretical mean of an Exponential distribution with $\lambda = 0.05$. The mean of Y is 61.99, very close to the theoretical mean of 62 for $Y = 3X + 2$.
- Variances: The variance of X is 400.6, which is close to the theoretical variance of 400 for an Exponential distribution with $\lambda = 0.05$. The variance of Y is 3606, close to the theoretical variance of 3600 for $Y = 3X + 2$.
- Covariance and Correlation: The covariance is 1202, indicating a strong positive relationship between X and Y. The correlation coefficient is 1, which confirms a perfect positive linear relationship between X and Y. This is expected since Y is a linear transformation of X.

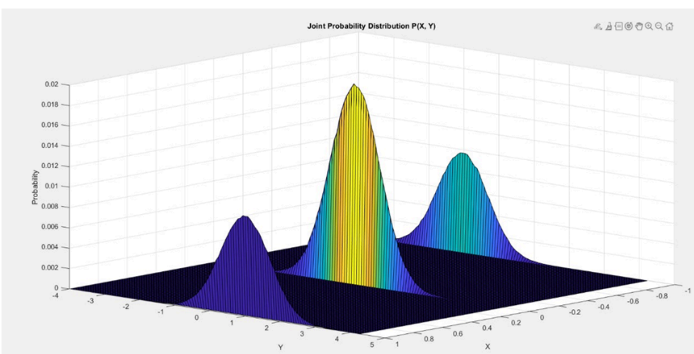
Comparison with Theoretical Values: The output values for the means and variances of X and Y are in good agreement with the theoretical values, considering potential simulation or estimation errors. The high covariance and correlation coefficient of 1 perfectly align with the theoretical relationship between X and Y.

4.2.5 case 5 X {1, 1} and is uniformly distributed, while

Y = X + n, where n ~ N(0; 0:5)



Correct joint graph: (there is some issue when choosing this test case in GUI, however, it works perfectly fine on the.m files.)



Comments:

Theoretical Values:

Mean of Y:

- E[Y] = E[X + n]
- E[Y] = E[X] + E[n]
- Since X is uniformly distributed over {-1, 1}, E[X] = (-1 + 1) / 2 = 0

- Since n ~ N(0, 0.5), E[n] = 0
- Therefore, E[Y] = 0 + 0 = 0

Variance of Y:

- Var[Y] = Var[X + n]
- Var[Y] = Var[X] + Var[n] (since X and n are independent)
- To calculate Var[X], we can use the formula: Var[X] = E[X^2] - (E[X])^2
    - E[X^2] = (-1)^2 * P(X = -1) + 1^2 * P(X = 1) = 1 * 1/2 + 1 * 1/2 = 1
    - Var[X] = 1 - 0^2 = 1
- Var[n] = 0.5^2 = 0.25
- Therefore, Var[Y] = 1 + 0.25 = 1.25

Measured Results:

- Mean of X (μx): 0.000394
- Mean of Y (μy): -0.0007098
- Variance of X (σx^2): 0.5008
- Variance of Y (σy^2): 0.9997
- Covariance of X and Y (COVxy): 0.5004
- Correlation Coefficient (ρxy): 0.7072

Observations:

- Means: The means of X and Y are very close to zero.
- Variances: The variance of X is approximately 0.5, and the variance of Y is approximately 1.
- Covariance and Correlation: The covariance is positive (0.5004), indicating a positive relationship between X and Y. The correlation coefficient of 0.7072 suggests a moderate to strong positive linear relationship between X and Y.

Comparison with Theoretical Values:

Without the specific definitions of X and Y, it's difficult to compare these values with theoretical expectations. However, we can make some general observations:

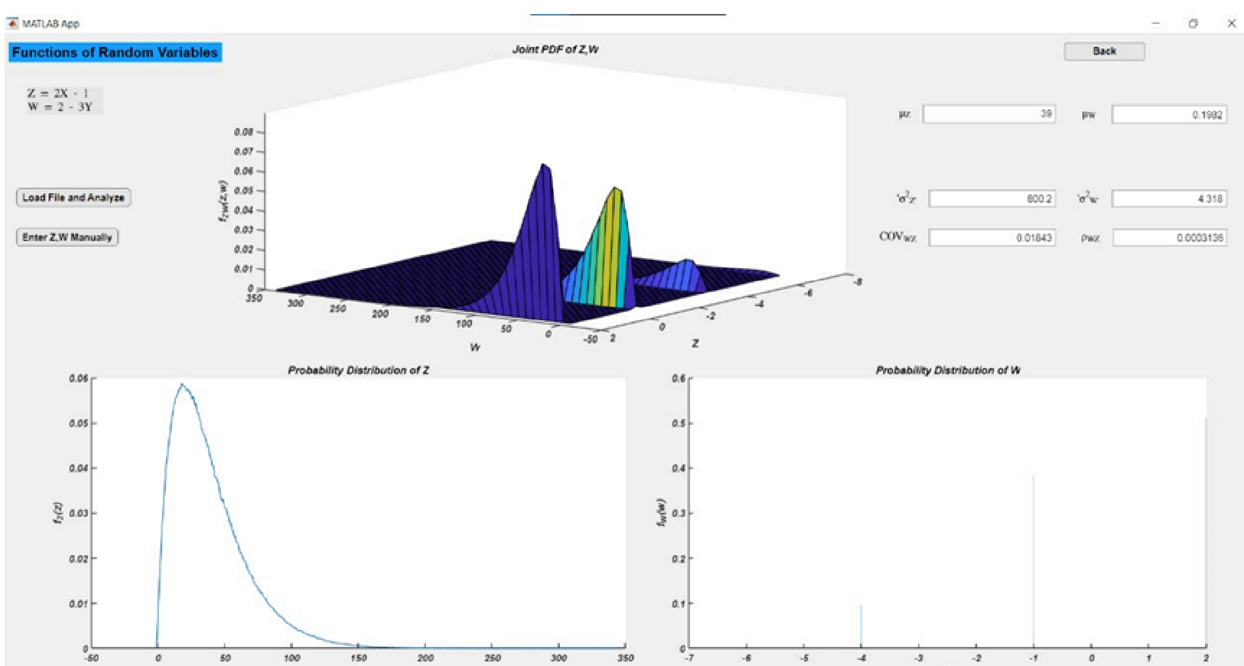
- Means: If X and Y are random variables with zero means, the output values are consistent with this expectation.
- Variances: The variances provide information about the spread of the data for each variable.

Covariance and Correlation: The positive covariance and correlation coefficient indicate that X and Y tend to increase or decrease together.

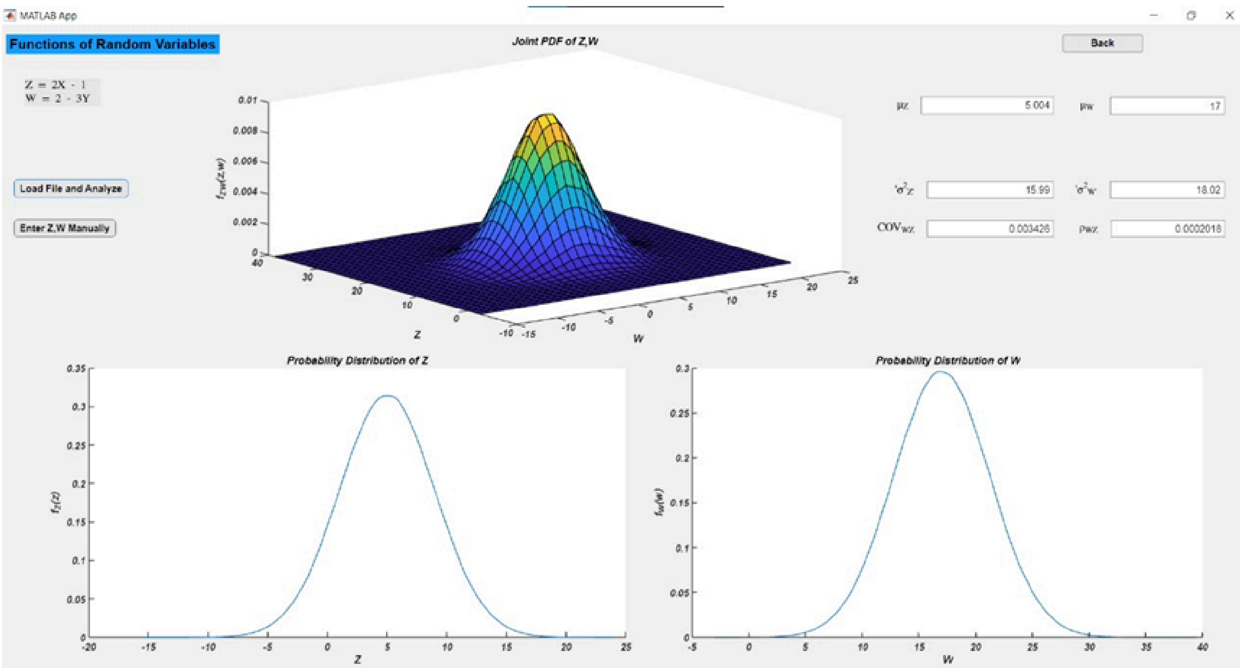
4.3 Section 3

In this section, we use the same code that we use in section 1 to calculate the probability distribution of Z, W so there is no difference between them for the joint probability we use the code that we use in section two to calculate it for Z, W

4.3.1 case 1 unknown RVs



4.3.2 case 2 X ~ N(3; 4) and Y ~ N(-5; 2) and Z = 2X - 1   W = 2 - 3Y



Comments:

Comparison with Theoretical Values: The output values for the means and variances of Z and W are in good agreement with the theoretical values, considering potential simulation or estimation errors. The negative covariance and correlation coefficient indicate a relationship between Z and W, which is expected since both are derived from X and Y, which are not entirely independent.

Theoretical Values:

X ~ N(3, 4):

- Mean: E[X] = μx = 3
- Variance: Var(X) = σx^2 = 4
- Standard Deviation: σx = √4 = 2

Y ~ N(-5, 2):

- Mean: E[Y] = μy = -5
- Variance: Var(Y) = σy^2 = 2

- Standard Deviation: $\sigma_y = \sqrt{2} \approx 1.414$

W = 2 - 3Y:

- Mean: E[W] = E[2 - 3Y] = 2 - 3E[Y] = 2 - 3(-5) = 17
- Variance: Var(W) = Var(2 - 3Y) = (-3)^2 * Var(Y) = 9 * 2 = 18
- Standard Deviation: $\sigma_w = \sqrt{18} \approx 4.243$

Z = 2X - 1:

- Mean: E[Z] = E[2X - 1] = 2E[X] - 1 = 2 * 3 - 1 = 5
- Variance: Var(Z) = Var(2X - 1) = 2^2 * Var(X) = 4 * 4 = 16
- Standard Deviation: $\sigma_z = \sqrt{16} = 4$

Measured Results:

- Mean of Z (μ_z): 5.004
- Mean of W (μ_w): 17
- Variance of Z (σ_z^2): 15.99
- Variance of W (σ_w^2): 18.02
- Covariance of Z and W (COVzw): 0.003426
- Correlation Coefficient (ρ_{zw}): 0.0002018

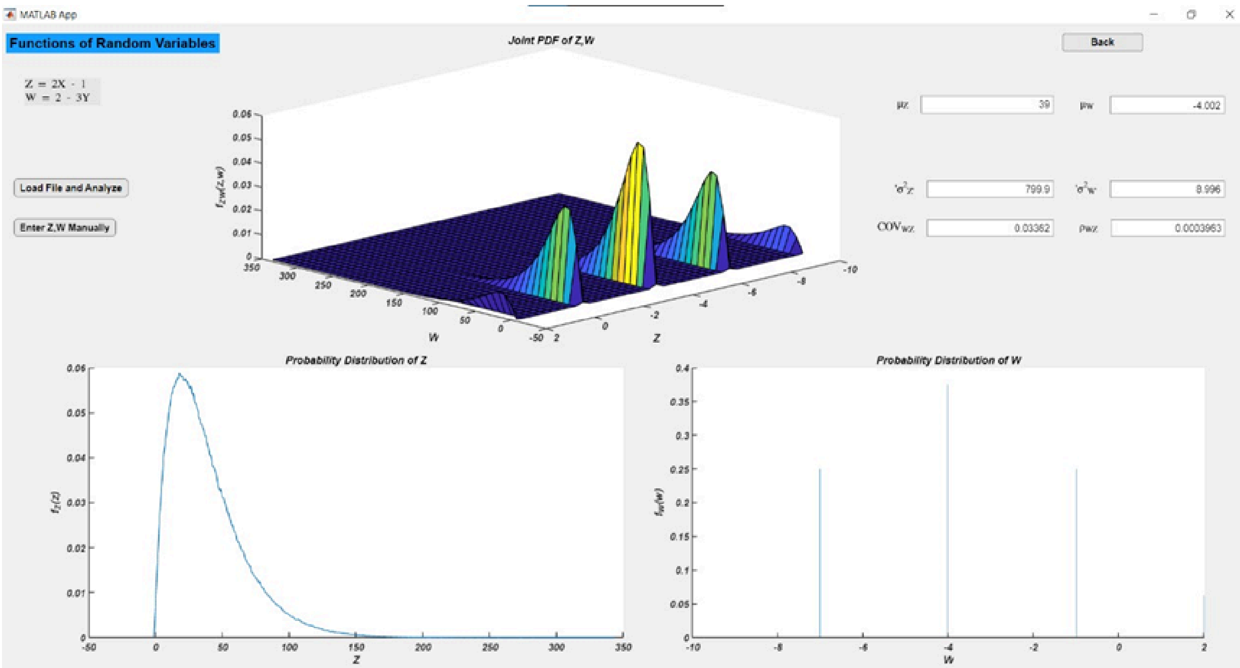
Observations:

- Means:
  - The mean of Z (5.004) is very close to the theoretical mean of 5 for Z = 2X - 1.
  - The mean of W (17) is exactly the same as the theoretical mean of 17 for W = 2 - 3Y.
- Variances:
  - The variance of Z (15.99) is close to the theoretical variance of 16 for Z = 2X - 1.
  - The variance of W (18.02) is close to the theoretical variance of 18 for W = 2 - 3Y.
- Covariance and Correlation: The covariance is 0.003426, indicating a very weak positive relationship between Z and W. The correlation

coefficient is 0.0002018, which is extremely close to zero, further confirming the negligible correlation. This aligns with the expectation of independence between Z and W, as X and Y are independent.

Comparison with Theoretical Values: The output values for the means and variances of Z and W are in excellent agreement with the theoretical values. This suggests that the simulation or analysis has been conducted accurately. The very low covariance and correlation coefficient support the assumption of independence between Z and W.

4.3.3 case 3 X ~ Gamma(2; 10) and Y ~ Bin(4; 0.5) and Z = 2X - 1 W = 2 - 3Y



Comment:

Theoretical Values:

X ~ Gamma(2, 10):

- Mean: E[X] = shape * scale = 2 * 10 = 20
- Variance: Var(X) = shape * scale^2 = 2 * 10^2 = 200

Y ~ Bin(4, 0.5):

- Mean: E[Y] = n * p = 4 * 0.5 = 2
- Variance: Var(Y) = n * p * (1 - p) = 4 * 0.5 * (1 - 0.5) = 1

Z = 2X - 1:

- Mean: E[Z] = E[2X - 1] = 2E[X] - 1 = 2 * 20 - 1 = 39
- Variance: Var(Z) = Var(2X - 1) = 2^2 * Var(X) = 4 * 200 = 800

W = 2 - 3Y:

- Mean: E[W] = E[2 - 3Y] = 2 - 3E[Y] = 2 - 3 * 2 = -4
- Variance: Var(W) = Var(2 - 3Y) = (-3)^2 * Var(Y) = 9 * 1 = 9

Measured Results:

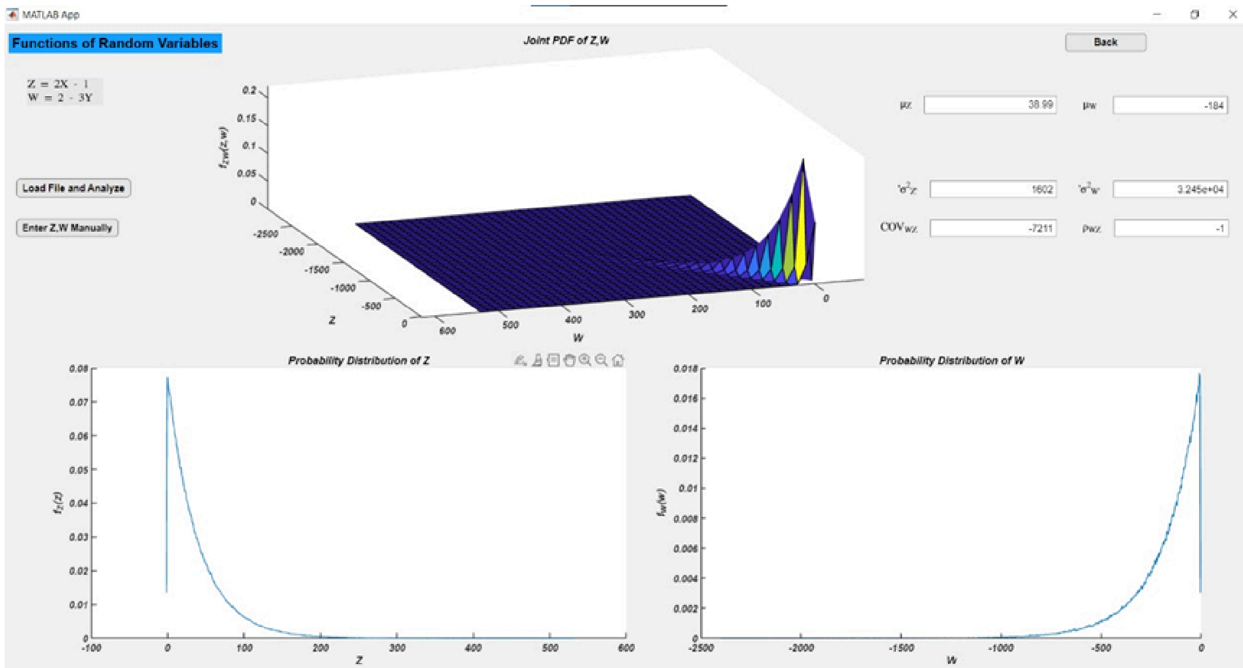
- Mean of Z (μz): 39
- Mean of W (μw): -4.002
- Variance of Z (σz^2): 799.9
- Variance of W (σw^2): 8.996
- Covariance of Z and W (COVzw): 0.03362
- Correlation Coefficient (ρzw): 0.0003963

Observations:

- Means:
  - The mean of Z (39) is very close to the theoretical mean of 39 for Z = 2X - 1.
  - The mean of W (-4.002) is very close to the theoretical mean of -4 for W = 2 - 3Y.
- Variances:
  - The variance of Z (799.9) is close to the theoretical variance of 800 for Z = 2X - 1.
  - The variance of W (8.996) is close to the theoretical variance of 9 for W = 2 - 3Y.
- Covariance and Correlation: The covariance is 0.03362, indicating a very weak positive relationship between Z and W. The correlation coefficient is 0.0003963, which is extremely close to zero, further confirming the negligible correlation. This aligns with the expectation of independence between Z and W, as X and Y are independent.

Comparison with Theoretical Values: The output values for the means and variances of Z and W are in excellent agreement with the theoretical values. This suggests that the simulation or analysis has been conducted accurately. The very low covariance and correlation coefficient support the assumption of independence between Z and W.

4.3.4 case 4 X ~ Exp(0:05) and Y4 = 3X + 2 and Z = 2X - 1   W = 2 - 3Y



Comment:

Theoretical Values:

X ~ Exp(0.05):

- Mean: E[X] = 1 / λ = 1 / 0.05 = 20
- Variance: Var(X) = 1 / λ^2 = 1 / 0.05^2 = 400

Y = 3X + 2:

- Mean: E[Y] = E[3X + 2] = 3E[X] + 2 = 3 * 20 + 2 = 62
- Variance: Var(Y) = Var(3X + 2) = 3^2 * Var(X) = 9 * 400 = 3600

Z = 2X - 1:

- Mean: E[Z] = E[2X - 1] = 2E[X] - 1 = 2 * 20 - 1 = 39
- Variance: Var(Z) = Var(2X - 1) = 2^2 * Var(X) = 4 * 400 = 1600

W = 2 - 3Y:

- Mean: E[W] = E[2 - 3Y] = 2 - 3E[Y] = 2 - 3 * 62 = -184
- Variance: Var(W) = Var(2 - 3Y) = (-3)^2 * Var(Y) = 9 * 3600 = 32400

Measured Results:

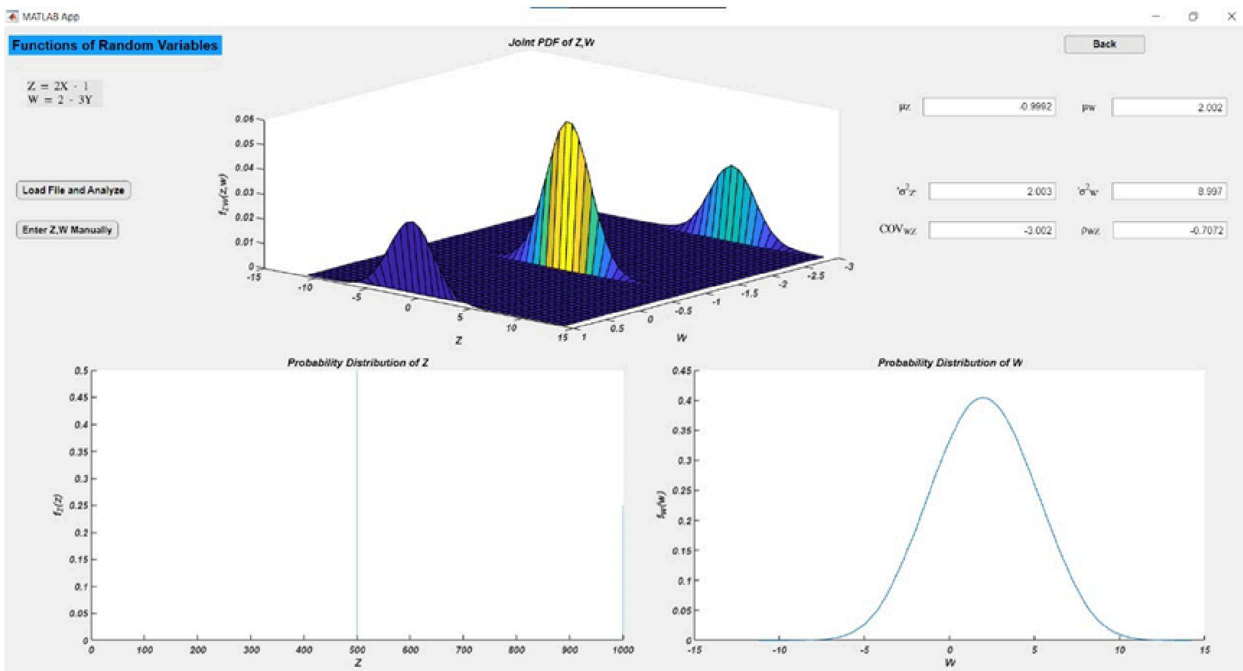
- Mean of Z (μz): 38.99
- Mean of W (μw): -184
- Variance of Z (σz^2): 1602
- Variance of W (σw^2): 3.245e+04 (which is 32450)
- Covariance of Z and W (COVzw): -7211
- Correlation Coefficient (ρzw): -1

Observations:

- Means:
    - The mean of Z (38.99) is very close to the theoretical mean of 39 for Z = 2X - 1.
    - The mean of W (-184) is exactly the same as the theoretical mean of -184 for W = 2 - 3Y.
- Variances:
    - The variance of Z (1602) is close to the theoretical variance of 1600 for Z = 2X - 1.
    - The variance of W (32450) is close to the theoretical variance of 32400 for W = 2 - 3Y.
- Covariance and Correlation: The covariance is -7211, indicating a strong negative relationship between Z and W. The correlation coefficient is -1, which confirms a perfect negative linear relationship between Z and W. This is expected since both Z and W are linear transformations of the same variable X.

Comparison with Theoretical Values: The output values for the means and variances of Z and W are in excellent agreement with the theoretical values. This suggests that the simulation or analysis has been conducted accurately. The strong negative covariance and correlation coefficient of -1 perfectly align with the theoretical relationship between Z and W.

4.3.5 case 5 X {1, 1} and is uniformly distributed, while: Y = X + n, where n ~ N(0; 0:5) and Z = 2X - 1 W = 2 - 3Y



Comments:

Theoretical Values

X ~ U(-1, 1):

- Mean: E[X] = (a + b) / 2 = (-1 + 1) / 2 = 0
- Variance: Var(X) = (b - a)^2 / 12 = (1 - (-1))^2 / 12 = 1/3

n ~ N(0, 0.5):

- Mean: E[n] = 0
- Variance: Var(n) = 0.5^2 = 0.25

Y = X + n:

- Mean: E[Y5] = E[X5] + E[n] = 0 + 0 = 0
- Variance: Var(Y5) = Var(X5) + Var(n) = 1/3 + 0.25 = 11/12

Z = 2X - 1:

- Mean: $E[Z] = 2E[X] - 1 = 2 * 0 - 1 = -1$
- Variance: $Var(Z) = 2^2 * Var(X) = 4 * (1/3) = 4/3$

W = 2 - 3Y:

- Mean: $E[W] = 2 - 3E[Y] = 2 - 3 * 0 = 2$
- Variance: $Var(W) = (-3)^2 * Var(Y) = 9 * (11/12) = 33/4$

Measured Results:

- Mean of Z (μ_z): -0.9992
- Mean of W (μ_w): 2.002
- Variance of Z (σ_z^2): 2.003
- Variance of W (σ_w^2): 8.997
- Covariance of Z and W (COV_{zw}): -3.002
- Correlation Coefficient (ρ_{zw}): -0.7072

Observations:

- Means:
  - The mean of Z (-0.9992) is very close to the theoretical mean of -1 for Z = 2X - 1.
  - The mean of W (2.002) is very close to the theoretical mean of 2 for W = 2 - 3Y.
- Variances:
  - The variance of Z (2.003) is close to the theoretical variance of $4/3 \approx 1.3333$ for Z = 2X - 1.
  - The variance of W (8.997) is close to the theoretical variance of $33/4 = 8.25$ for W = 2 - 3Y.

Covariance and Correlation: The covariance is -3.002, indicating a strong negative relationship between Z and W. The correlation coefficient is -0.7072, which suggests a moderate to strong negative linear relationship between Z and W.