Homework-01

Problem 1 (Simpson Paradox)

Machine 1	Wins	Losses
You	40	60
Friend	30	70

Machine 2	Wins	Losses
You	210	830
Friend	14	70

a. My winning probability in machine 1: $\frac{40}{40+60} = 40\%$

My winning probability in machine 2: $\frac{210}{210+830} = 20.1\%$

Friend's winning probability in machine 1: $\frac{30}{30+70} = 30\%$

Friend's winning probability in machine 2: $\frac{14}{14+70} = 16.67\%$

Observing both cases in machine 1 and machine 2, it seems that I have the highest probability to win than my friend.

b. My overall winning probability

$$= \frac{40+210}{40+60+210+830}$$

$= 21.92\%$

My Friend's overall winning probability

$$= \frac{30+14}{30+70+14+70}$$

$= 23.91\%$

In this case, my friend has the highest probability to win.

c. By comparing 1 and 2, we can find that the trend that appears in different groups of data (1), reverses when these groups are combined (2).

So, when the dataset of the slot machine are aggregated in problem 2, the probabily to win of my friend is higher than me. on the otherhand when the dataset is seperated into groups in problem 1. I am highly likely to win.

The problem here is that looking only at the percentages in the seperate data ignores many aspects. Each fraction shows the number of winnings out of the numbers played. on the other hand, the sizes of the groups, which are combined when the lurking variable is ignored, are very different.

d. The two conclusions are different because there are some hidden variables that split data into multiple seperate distributions. Such hidden variable is called lurking variable. If the lurking variables can be identified then the two conclusion will be same. After writing Addition probability equations we can identify which values should be updated.

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ [For machine 1]

$\Rightarrow 0.21 = 0.4 + 0.201 - P(A \cap B)$

$\Rightarrow P(A \cap B) = 0.391$

For machine 2, $0.24 = 0.3 + 0.17 - P(A \cap B) \Rightarrow P(A \cap B) = 0.23$

Problem 2 Linear algebra :

a. $a_1 = (1,0), b_1 = (2,0)$

$$a_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad W = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \quad b_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$\Rightarrow x_1.1 + y_1.0 = 2, \; x_2.1 + y_2.0 = 0$

so, $x_1 = 2, x_2 = 0.$

$a_2 = (0,1), b_2 = (3,2)$

$$a_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad W = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \quad b_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

$\Rightarrow x_1.0 + y_1.1 = 3, \; x_2.0 + y_2.1 = 2$

so, $y_1 = 3, \; y_2 = 2$

Therefore, 2x2 matrix $W = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$

2(b) $\tan(\alpha) = 2$

$\Rightarrow \alpha = \tan^{-1} 2$

$$V = \begin{bmatrix} \cos(\tan^{-1} 2) & \sin(\tan^{-1} 2) \\ -\sin(\tan^{-1} 2) & \cos(\tan^{-1} 2) \end{bmatrix}$$

$$= \begin{bmatrix} 0.45 & 0.89 \\ -0.89 & 0.45 \end{bmatrix}$$

$\Sigma = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ [This scales x axis by factor 4 and Y axis by factor 1]

$\tan(\beta) = \frac{1}{2}$

$\Rightarrow \beta = \tan^{-1}(\frac{1}{2})$

$$U = \begin{bmatrix} \cos(\tan^{-1} \frac{1}{2}) & -\sin(\tan^{-1} \frac{1}{2}) \\ \sin(\tan^{-1} \frac{1}{2}) & \cos(\tan^{-1} \frac{1}{2}) \end{bmatrix}$$

$$= \begin{bmatrix} 0.89 & -0.45 \\ 0.45 & 0.89 \end{bmatrix}$$

$U\Sigma V$

$$= \begin{bmatrix} 0.89 & -0.45 \\ 0.45 & 0.89 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.45 & 0.89 \\ -0.89 & 0.45 \end{bmatrix}$$

$$= \begin{bmatrix} 3.56 & -0.45 \\ 1.8 & 0.89 \end{bmatrix} \begin{bmatrix} 0.45 & 0.89 \\ -0.89 & 0.45 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3 \\ 0.02 & 2 \end{bmatrix}$$ which is similiar to W.

c.

$$W = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$$

$$W^T = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix}$$

$$W^T.W = \begin{bmatrix} 2 & 0 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}$$

λ is an eigenvalue of W^TW if and only if the determinant $(\lambda I - A) = 0$

$$\det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}\right) = 0$$

$$\Rightarrow \det\left(\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}\right) = 0$$

$$\Rightarrow \det\left(\begin{bmatrix} \lambda-4 & -6 \\ -6 & \lambda-13 \end{bmatrix}\right) = 0$$

$$\Rightarrow (\lambda-4)(\lambda-13) - 36 = 0$$

$$\Rightarrow \lambda^2 - 13\lambda - 4\lambda + 52 - 36 = 0$$

$$\Rightarrow \lambda^2 - 17\lambda + 16 = 0$$

$$\Rightarrow \lambda^2 - 16\lambda - \lambda + 16 = 0$$

$$\Rightarrow \lambda(\lambda-16) - (\lambda-16) = 0$$

$$\Rightarrow (\lambda-16)(\lambda-1) = 0$$

$\lambda = 16$ or $\lambda = 1$

So $A\vec{v} = \lambda\vec{v}$ can be satisfied when $\lambda = 16$ or $\lambda = 1$

Eigenvalues are: 16, 1 assuming non-zero eigenvectors

$A\vec{v} = \lambda\vec{v}$

$\Rightarrow 0 = \lambda\vec{v} - A\vec{v}$

$\Rightarrow 0 = \lambda I_n\vec{v} - A\vec{v}$

$= (\lambda I_n - A)\vec{v}$

for any eigenvalue λ eigenspace

$E_\lambda = N(\lambda I_n - A)$

when $\lambda = 16$

$E_{16} = N\left(\begin{bmatrix} 16 & 0 \\ 0 & 16 \end{bmatrix} - \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}\right) = N\left(\begin{bmatrix} 12 & -6 \\ -6 & 3 \end{bmatrix}\right)$

Now, $\begin{bmatrix} 12 & -6 \\ -6 & 3 \end{bmatrix}\vec{v} = \vec{0}$

where $\begin{bmatrix} 12 & -6 \\ -6 & 3 \end{bmatrix} = \begin{bmatrix} 12 & -6 \\ 0 & 0 \end{bmatrix}$ [reduced row echelon form]

$= \begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix}$

$$\begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$2v_1 - v_2 = 0$

$\Rightarrow 2v_1 = v_2$

$\Rightarrow v_1 = \frac{v_2}{2}$

$$E_{16} = \left\{ \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = t \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \right\} t \in \mathbb{R}$$

$$E_{16} = \text{Span}\left(\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix} \right)$$

$\lambda = 1$

$$E_1 = N\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) - \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}\Big) = N\left(\begin{bmatrix} -3 & -6 \\ -6 & 12 \end{bmatrix}\right)$$

$$= N\left(\begin{bmatrix} -3 & -6 \\ -0 & 0 \end{bmatrix}\right)$$ [reduced row echelon form]

$$= N\left(\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}\right)$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\Rightarrow v_1 + 2v_2 = 0$

$\Rightarrow v_1 = -2v_2$

Suppose $v_2 = t$, $v_1 = -2t$

$$\text{So, } E_1 = \left\{ \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = t \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\} = \text{span}\left(\begin{bmatrix} -2 \\ 1 \end{bmatrix} \right)$$

Every point on the unit circle gets transformed gets transformed by W.

The points on the unit circle are:

$a = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$c = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$

$d = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

$W = \begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix}$

$$\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = a'$$

$$\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = b'$$

$$\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix} = c'$$

$$\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -3 \\ -2 \end{bmatrix} = d'$$

The transformed shape is rotated ellipse, which is centerend in (0,0). The major and minor axis of the ellips is 4 and 1, which is square root of the eigen valus.

Distance between (0,0) and (-3,-2)

$$\sqrt{(0+3)^2+(0+2)^2}$$

$$=\sqrt{9+4}$$

≈ 4 (Semi-major axis)

Semi-minor axis $= 1$

Scanned with CamScanner

d. The determinant of w:

$$\begin{bmatrix} 2 & 3 \\ 0 & 2 \end{bmatrix} = 4-0 = 4$$

The area of the unit circle $= \pi R^2$ [R is the radius of circle]
$= \pi . 1 = \pi$

The area of the ellipse $= \pi ab$ [$a=4, b=1$]
$= 4\pi$ [a and b are semi major and minor axes respectively]

hypothesis:

If there is a region (in this case circle) in any domain and a transformation matrix is applied in that domain to get a new region, the new area is going to be the determinant of the transformation matrix times the area of original region.

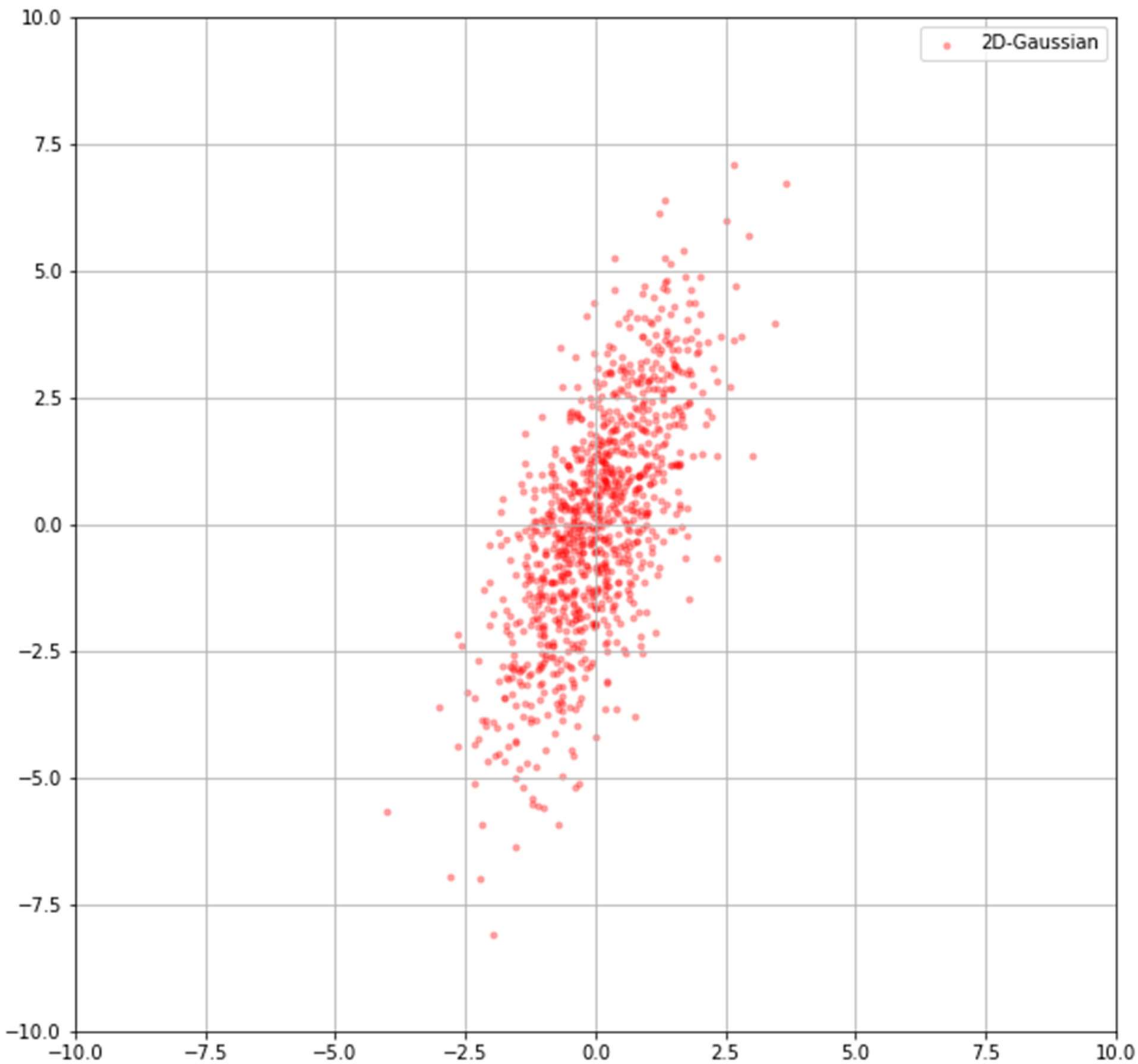
The determinant of the transformation matrix is essentially a scaling factor on the area of a certain region

Area $= |A \det(B)|$ [A = old region area, B = transformation matrix]

$\Rightarrow \det(AB) = \det(A)\det(B)$

Problem 3(a)

Given Multivariate Gaussian Distribution



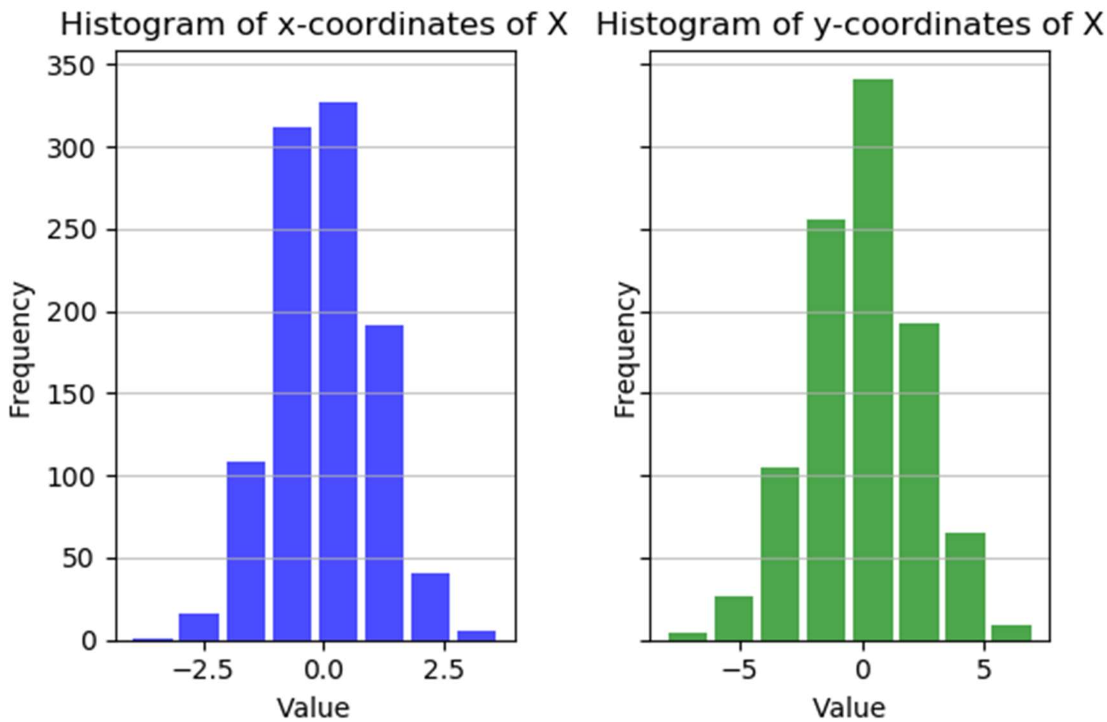
Mean and Variance of Sampled Points:

```
[[4.27808445 2.40991407 2.05275031 ... 2.97926342 2.1962627  3.06908942]
 [2.40991407 1.35754352 1.1563474  ... 1.6782672  1.23719025 1.72886764]
 [2.05275031 1.1563474  0.98496976 ... 1.42953791 1.05383122 1.47263906]
 ...
 [2.97926342 1.6782672  1.42953791 ... 2.07476281 1.52948012 2.13731776]
 [2.1962627  1.23719025 1.05383122 ... 1.52948012 1.12750693 1.57559457]
 [3.06908942 1.72886764 1.47263906 ... 2.13731776 1.57559457 2.20175875]]
```

```
0.03586188161776395
```

Problem 3(b)

Histogram for the x-coordinates of X and y-coordinates of X:



Problem 3(c)

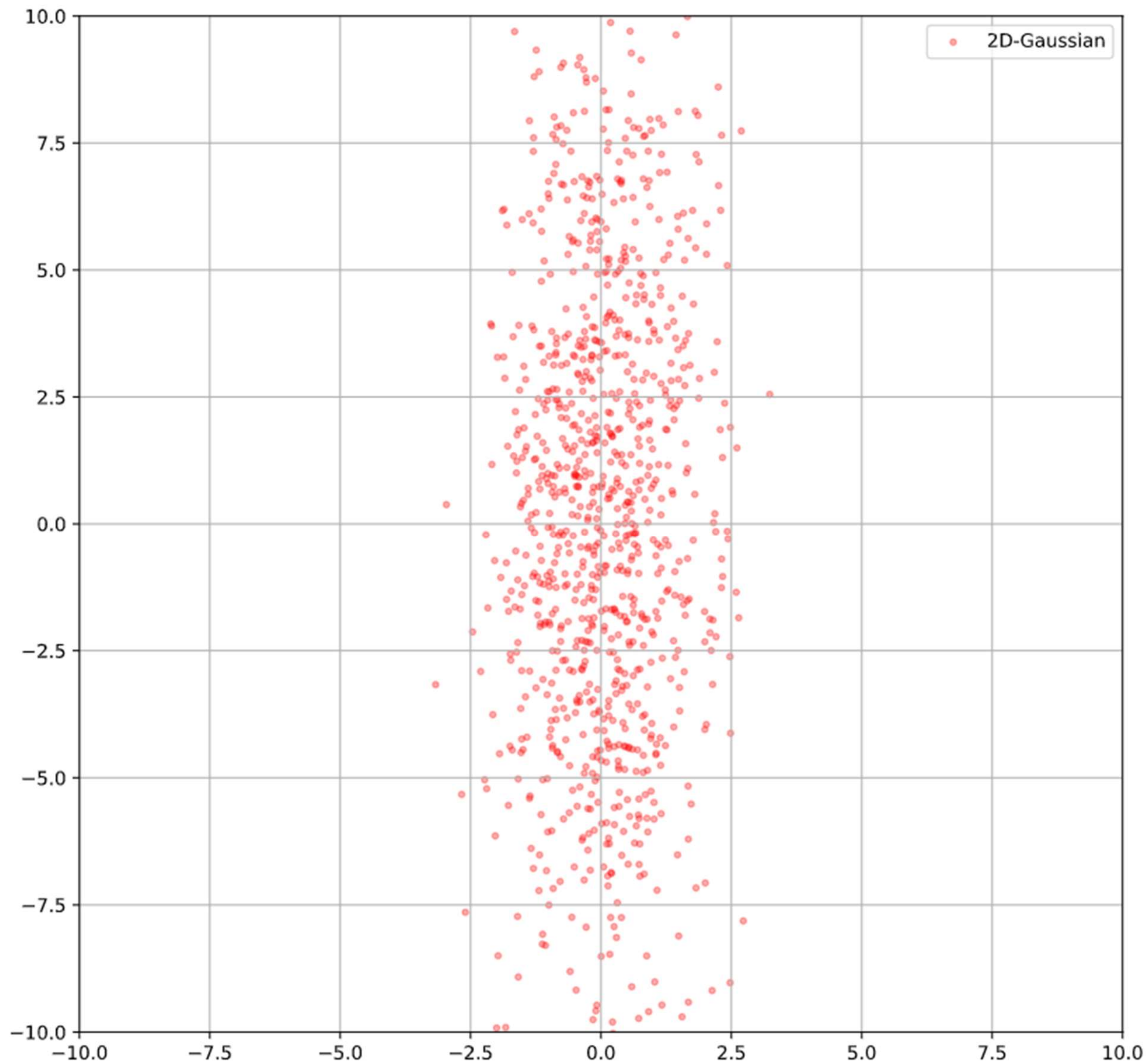
Both the x coordinates of X and y coordinates of X follow gaussian distribution.

Mean and Variance of x coordinate and y coordinate respectively:

```
x coordinate variance= 1.034708922316794
y coordinate variance = 5.075328237402045
x coordinate mean = 0.01047830014298007
y coordinate mean = 0.061245463092547824
```

Problem 3(d)

Sampling 1000 numbers from a 1D Gaussian distribution with the mean and variance of the x-coordinates and y coordinates and creating scatter plot with them:

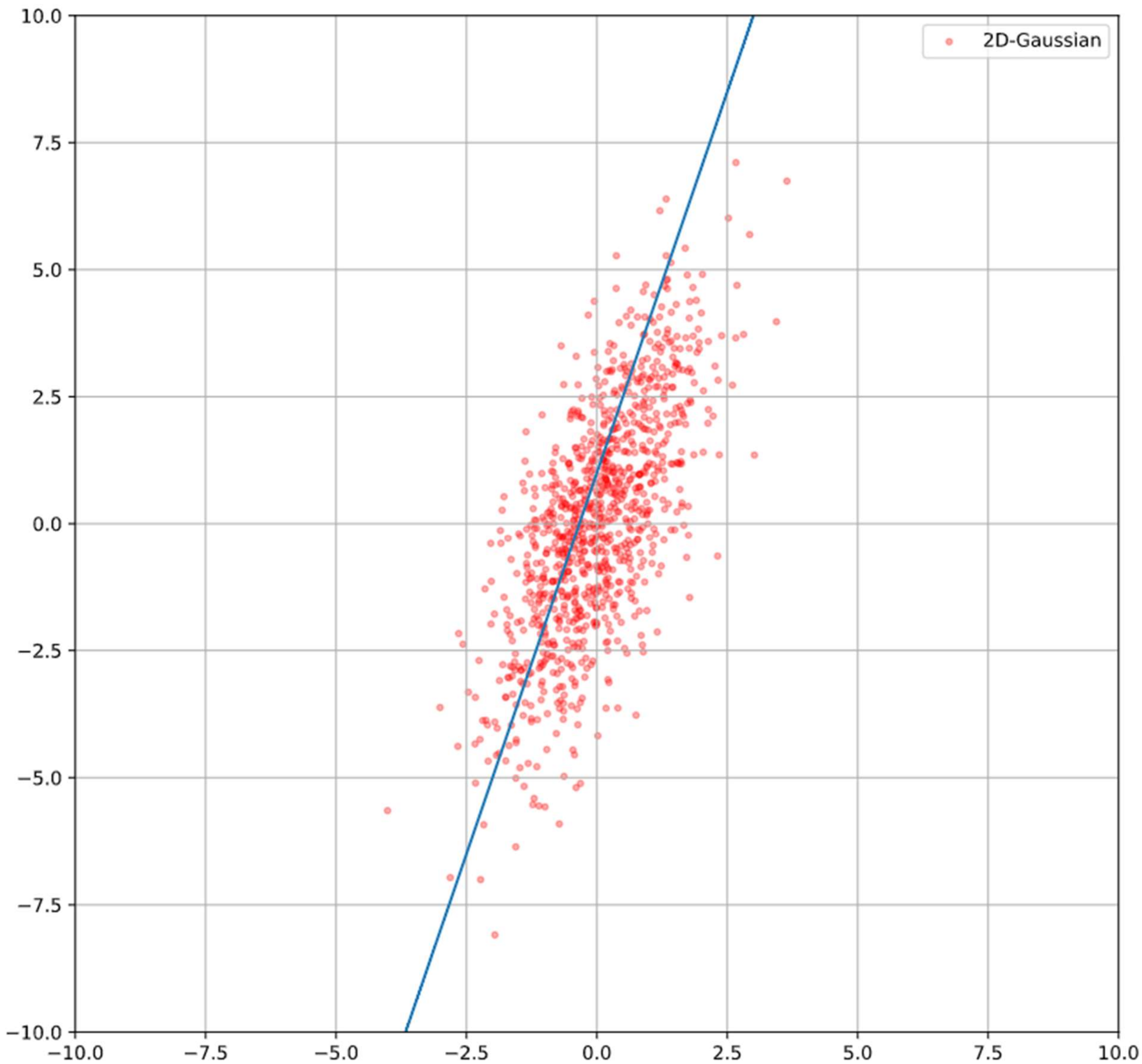


Difference:

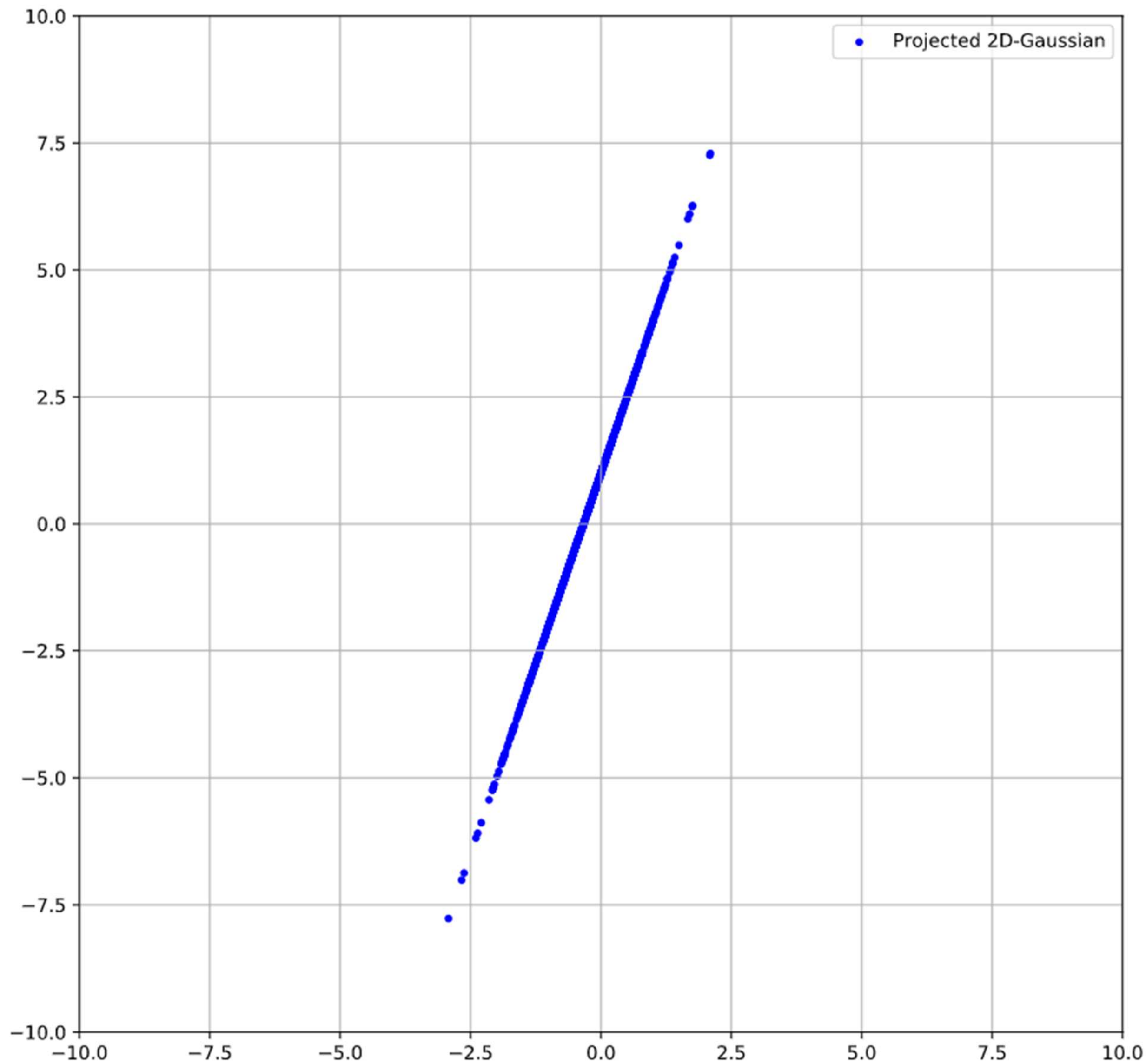
Comparing two scatter plots it is evident that there are some differences on the distributions of data. In the previous plot the two-dimension dataset represents arbitrary confidence ellipse which defines the region that contains 95%-97% of all samples that can be drawn from the underlying Gaussian distribution. Here the variances are parallel to what will become the major and minor axis of the confidence ellipse. But in the second distribution as the co-variance is zero the two-dimension dataset represents the axis-aligned confidence ellipse. Here variances are parallel to the x-axis and y-axis.

Problem 3(e)

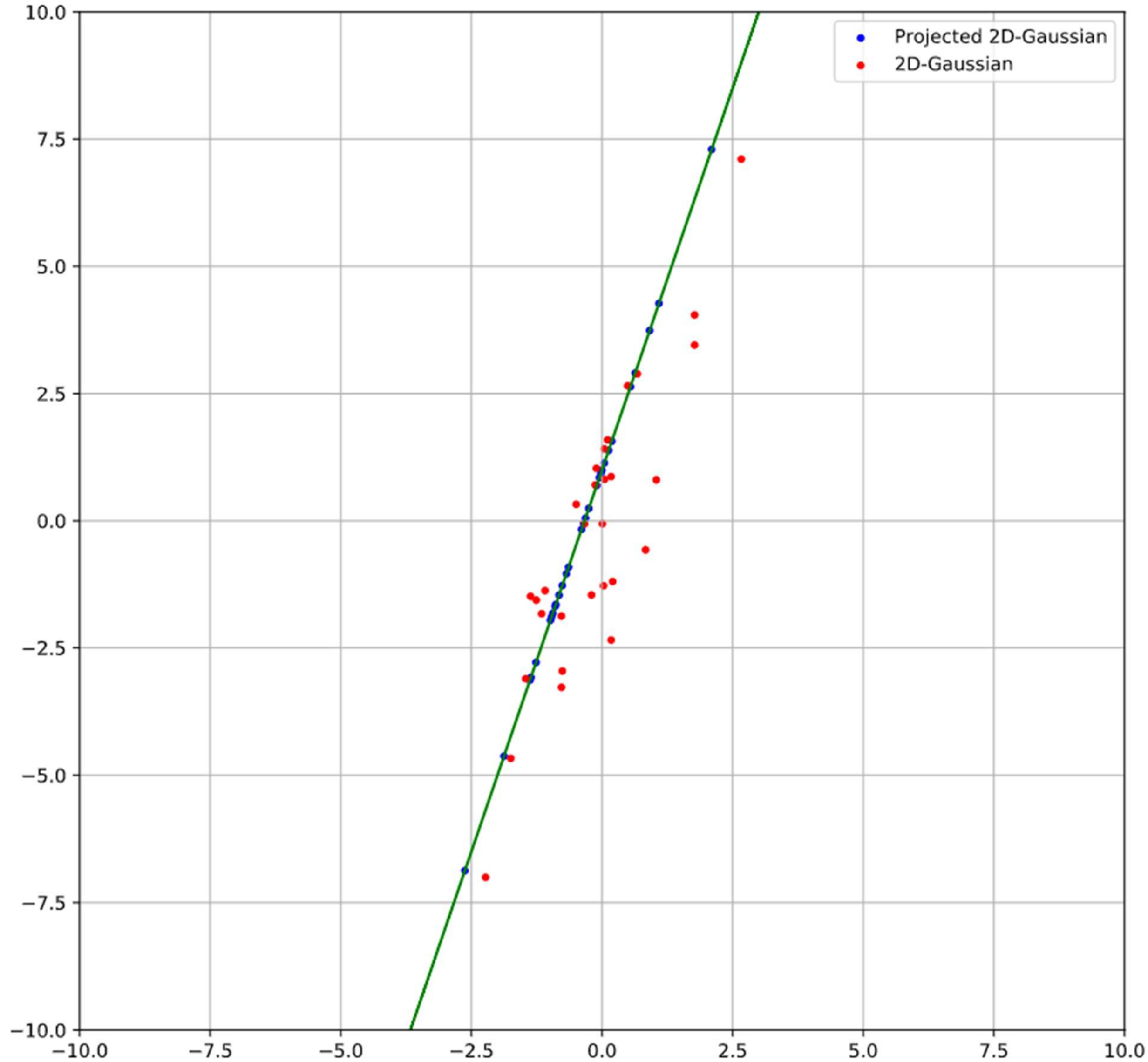
Plotting line over 2D Gaussian:



Projecting 2D Gaussian samples onto line:

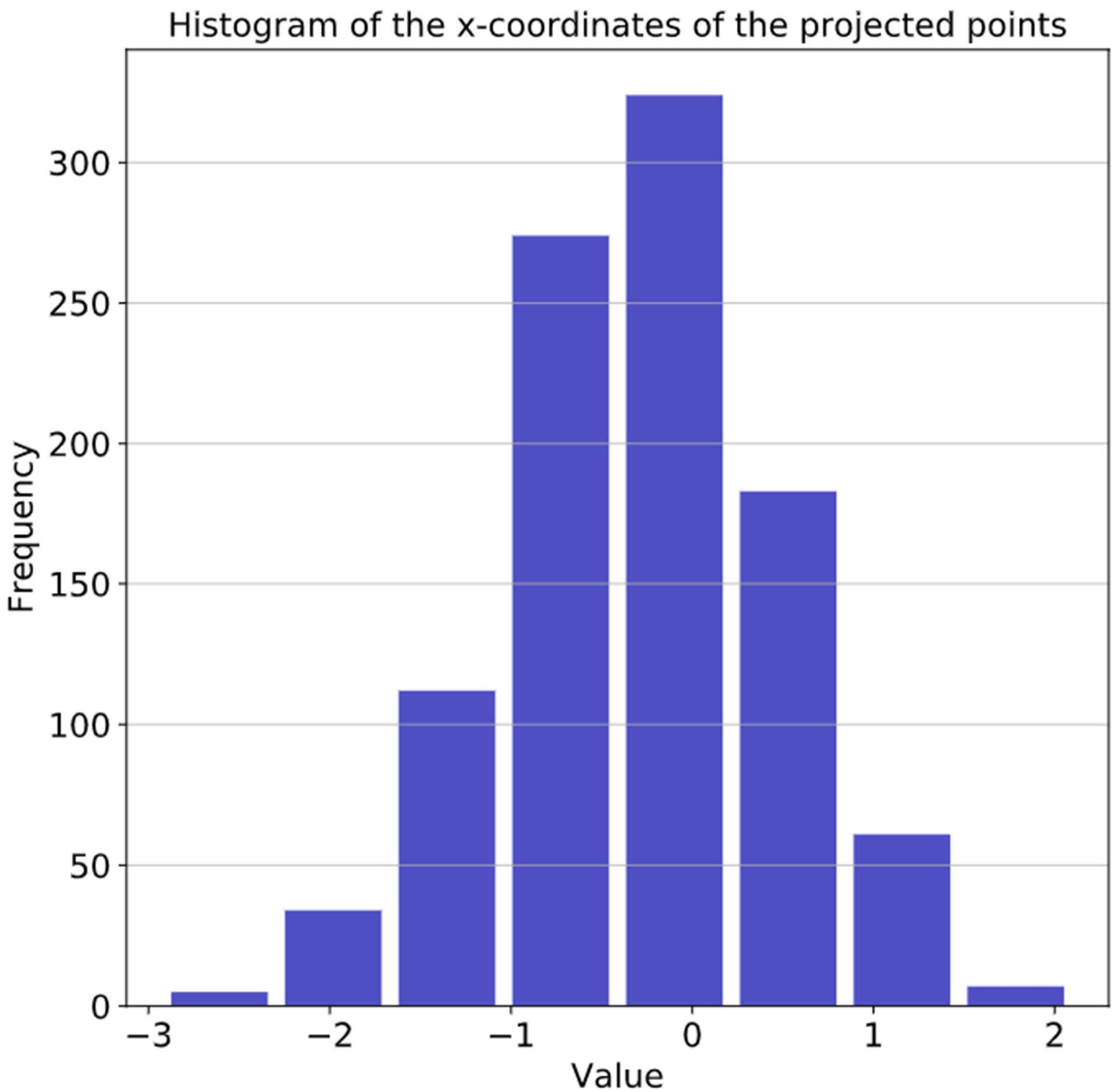


Projected 2D-Gaussian
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0



Problem 3(e):

Yes, the x-coordinates of the projected points sampled from some Gaussian distribution



Mean and Variance are respectively -0.2805785310579377 0.5612202950214332.