University Cote D'Azur

prabal5ghosh@gmail.com

Basic Probability Exam

Prabal Ghosh

October 18, 2023

Exercise 1

Let consider X a gaussian random variable with expectation -1 and variance 25.
Compute :

1. $P(X \geq -2)$
2. $P(-1.5 \leq X \leq -0.5)$
3. t such that $P(-1-t \leq X \leq -1+t) = 0.90$

Solutions

Here:

$$E[X] = -1$$

$$V[X] = 25$$

$$\sigma = \sqrt{25} = 5$$

. Here

$$X \sim N(-1, 25)$$

But for computation we need to convert this as standard Gaussian:

$$Z \sim N(0, 1)$$

. Using the properties of the standard normal distribution, we convert our Gaussian variable into the standard form to find the Z-score:

$$Z = \frac{X - \mu}{\sigma}$$

We convert our Gaussian variable into the standard form:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - (-1)}{5} = \frac{X + 1}{5}$$

1. To find $P(X \geq -2)$:

$$P(X \geq -2) = P(\frac{X+1}{5} \geq \frac{-2+1}{5}) = P(\frac{X+1}{5} \geq \frac{-1}{5}) = P(\frac{X+1}{5} \leq \frac{1}{5}) = P(\frac{X+1}{5} \leq 0.2) = 0.5793$$

Finally:

$$\boxed{P(X \geq -2) = 0.5793} \tag{1}$$

2. To find $P(-1.5 \leq X \leq -0.5)$: Thus we have:

$$P(-1.5 \leq X \leq -0.5) = P(\frac{-1.5+1}{5} \leq Z \leq \frac{-0.5+1}{5}) = P(\frac{-0.5+1}{5} \leq Z \leq \frac{0.5}{5}) = P(-0.1 \leq Z \leq .1)$$

Using the properties of the standard normal distribution,

$$P(-0.1 \leq Z \leq 0.1) = P(Z \leq 0.1) - P(Z \leq -0.1) = P(Z \leq 0.1) - (1 - P(Z \leq 0.1))$$

$$= 2P(Z \leq 0.1) - 1 = (2 * 0.5398) - 1 = 0.0796$$

Finally:

$$\boxed{P(-1.5 \leq X \leq -0.5) = 0.0796} \tag{2}$$

3. To find t such that $P(-1 - t \leq X \leq -1 + t)$:
We follow the same principle as before:

$$P(-1 - t \leq X \leq -1 + t) = P(\left(\frac{-1 - t - \mu}{\sigma}\right) \leq Z \leq \left(\frac{-1 + t - \mu}{\sigma}\right))$$

For the given values $\mu = -1$ and $\sigma = 5$:

$$P(-1-t \leq X \leq -1+t) = P(\left(\frac{-1 - t + 1}{5}\right) \leq Z \leq \left(\frac{-1 + t + 1}{5}\right)) = P(\frac{-t}{5} \leq Z \leq \frac{t}{5})$$

$$P(Z \leq \frac{t}{5}) - P(Z \leq \frac{-t}{5}) = P(Z \leq \frac{t}{5}) - (1 - P(Z \leq \frac{t}{5})) = 2P(Z \leq \frac{t}{5}) - 1$$

let's calculate:

$$P(-1 - t \leq X \leq -1 + t) = 0.9$$

$$\Rightarrow 2P(Z \leq \frac{t}{5}) - 1 = 0.9$$

$$\Rightarrow 2P(Z \leq \frac{t}{5}) = 1 + 0.9$$

$$\Rightarrow P(Z \leq \frac{t}{5}) = 1.9/2 = 0.95$$

To determine the value of t, we look up the z-value for which the cumulative probability is 0.95 in a standard normal table. The value is 1.65. We can determine t as

$$\frac{t}{5} = 1.65$$

$$\Rightarrow t = 1.65 * 5$$

$$\Rightarrow t = 1.65 * 5$$

$$\Rightarrow t = 8.25$$

Finally:

$$\boxed{t = 8.25} \tag{3}$$

Exercise 2

Let consider the following game : we take, at random and at the same time, 3 cards in a deck with 32 cards.
Let consider the random variable X given by :

$$X = \begin{cases} 7 & \text{if the 2 cards have the same value} \\ 0 & \text{if one card have a value in } \{7, 8, 9, 10\} \text{ and two in } \{V, D, R, 1\} \\ -5 & \text{otherwise} \end{cases}$$

Solution:

To compute the Variance of X, we first need to find the probabilities value for each of the random variables.
Generally, we can choose 3 cards from 32 decks $32\binom{}{3.}$

1.**Probability of getting 2 cards have the same value**:
There are 32 cards in the deck, and there are 8 different values.

To select 2 cards with the same value, we first choose one of the 8 values. We can do this $8\binom{}{1.}$

Then, we select 2 cards of the values we choose. We can do this in $4\binom{}{2.}$

Finally, we select the last card from the remaining 28 cards. We can do this in

$$\binom{7}{1} * \binom{4}{1}$$

.

Hence, P(X=7) =

$$P(X = 7) = \frac{\binom{8}{1} \times \binom{4}{2} \times \binom{7}{1} \times \binom{4}{1}}{\binom{32}{3}}$$

$$P(X = 7) = \frac{8 * (4 * 3/2) * 7 * 4}{(32 * 31 * 30)/(3 * 2)}$$

$$P(X = 7) = 0.27$$

2.**if one card have a value in {7, 8, 9, 10} and two in {V, D, R, 1}**:

To select 1 cards from {7,8,9,10} We can do this

$$\binom{4}{1} \times \binom{4}{1}.$$

Then, we select 2 cards from {V,D,R,1}. We can do this in $16\binom{}{2.}$

$$P(X=0) = \frac{\binom{4}{1} \times \binom{4}{1} \times \binom{16}{2}}{\binom{32}{3}}$$

$$P(X=0) = \frac{4*4*(16*15/2)}{(32*31*30)/(3*2)}$$

$$P(X=0) = 0.387$$

3. **Probability for all other outcomes**:

$$P(X=-5) = 1 - [P(X=7) + P(X=0)]$$

$$P(X=-5) = 1 - (0.27 + 0.387)$$

$$P(X=-5) = 0.343$$

Expected Values: Recall that the E(X) of a discrete random variable is defined as:

$$E(X) = \sum_i x_i P(X = x_i) for i = 1, 2, 3, ..., n.$$

Here, x$_i \in X(w) and X(w) = \{7, 0, -5\}$

Hence:

$$E(X) = 7 \times P(X=7) + 0 \times P(X=0) + (-5) \times P(X=-5)$$

$$E(X) = 7 \times 0.27 + 0 \times 0.387 - 5 \times 0.343$$

$$E(X) = 0.175$$

Finally, the variance:

$$Var(X) = E(X^2) - (E(X))^2$$

so we find E(X^2)$first.$

$$E(X^2) = 7^2(0.27) + 0 + (-5)^2(0.343)$$

$$E(X^2) = 21.805$$

Now V(X)=

$$Var(X) = E(X^2) - (E(X))^2$$
$$Var(X) = 21.805 - (-0.175)^2$$

$$\boxed{Var[x] = 21.77} \tag{4}$$

Thus, the variance of X is approximately 21.77.

Exercise 3

Given the random variable X whose density function is:

$$f_X(t) = \begin{cases} a * t^2 & \text{if } -2 \leq t < 1 \\ 0 & \text{otherwise} \end{cases}$$

1. What should be the value of a?
2. Compute the expectation and variance of X if they exist.
3. Determine the distribution function.

Solutions

1. To find the value of a:

   Using the properties of the probability density function:

   $$\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$$

   we can use this to find a by integrating over all the values of interest.

   $$\int_{-2}^{1} a * t^2\, dt = 1$$

   $$\begin{aligned} \int_{-2}^{1} a * t^2\, dt &= a \cdot \left.\frac{t^3}{3}\right|_{-2}^{1} \\ &= a * \frac{1+8}{3} \\ &= 3 * a \end{aligned}$$

Together we have

$$\int_{-2}^{1} a * t^2 \, dt = 1$$
$$\implies 3 * a = 1$$
$$\implies a = \frac{1}{3}$$

Finally:

$$\boxed{a = \frac{1}{3}} \tag{5}$$

2. Expectation and Variance:

Expectation computation

The expectation of a continuous random variable X with density function $f_X(t)$ is given by:

$$E(X) = \int_{-\infty}^{\infty} t \cdot f_X(t) \, dt$$

We have:

$$E(X) = 0 + \int_{-2}^{1} t \cdot \frac{t^2}{3} \, dt = \frac{1}{3} \cdot \left. \frac{t^4}{4} \right|_{-2}^{1}$$
$$= \frac{1 + 16}{3 * 4}$$
$$= \frac{17}{12}$$

$$\boxed{E(X) = \frac{17}{12}} \tag{6}$$

Variance Computation

The variance of a random variable X is given by:

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2$$

Where the second moment is:

$$E(X^2) = \int_{-\infty}^{\infty} t^2 \cdot f_X(t) \, dt$$

$$
\begin{aligned}
E(X^2) = \int_{-2}^{1} t^2 \cdot \frac{t^2}{3}\, dt = \int_{-2}^{1} \cdot \frac{t^4}{3}\, dt &= \frac{1}{3} \cdot \left.\frac{t^5}{5}\right|_{-2}^{1} \\
&= \frac{1+32}{3*5} \\
&= \frac{11}{5}
\end{aligned}
$$

$$
\boxed{E(X^2) = \frac{11}{5}} \tag{7}
$$

As we know that,

$$
Var(X) = E(X^2) - (E(X))^2
$$

Substituting in the formula for variance:

$$
\mathrm{Var}(X) = \frac{11}{5} - \left(\frac{17}{12}\right)^2
$$

So, Finally :

$$
\boxed{Var(X) = \frac{11}{5} - \left(\frac{17}{12}\right)^2} \tag{8}
$$

3. **Distribution Function: Given the density function**:

   The cumulative distribution function(CDF), $F_X(t)$, is defined as:

$$
F_X(t) = \int_{-\infty}^{t} f_X(u)\, du
$$

   Computing over different intervals:

   1. For $t < -2$:

$$
\boxed{F_X(t) = \int_{-\infty}^{-2} 0\, du = 0} \tag{9}
$$

2. For $-2 \leq t < 1$:

$$\begin{aligned} F_X(t) &= \int_{-\infty}^{-2} 0\,du + \int_{-2}^{t} \frac{u^2}{3}\,du \\ &= \frac{1}{3} \cdot \frac{t^4}{3}\Big|_{-2}^{t} \\ &= \frac{t^3+8}{3*3} \\ &= \frac{t^3+8}{9} \end{aligned}$$

$$\boxed{F_X(t) = \frac{t^3+8}{9}} \tag{10}$$

5. For $t \geq 1$:

$$F_X(t) = \int_{-\infty}^{-2} 0\,du + \int_{-2}^{1} \frac{u^2}{3}\,du + \int_{1}^{\infty} 0\,du$$

$$\boxed{F_X(t) = 1} \tag{11}$$

The cumulative distribution function $F_X(t)$ is:

$$\boxed{F_X(t) = \begin{cases} 0 & \text{if } t < -2 \\ \frac{1}{9}(t^3+8) & \text{if } -2 \leq t < 1 \\ 1 & \text{if } t \geq 1 \end{cases}} \tag{12}$$

Exercise 4

Let consider the dataset in data2.txt. Make a representation of it and propose a family of distribution that can model those observations.

Solutions

The data is distributed in subsections. One in around -4(MEAN) and another is around 4(MEAN). So here I can seperate the data with respect to point 0. and apply normal distribution. To do that I have used Python Programming. Also, I have implemented this using R programming.

PRABAL GHOSH MATHS EXAM CODING QUESTION

Exercise 4: Let consider the dataset in data2.txt. Make a representation of it and propose a family of distribution that can model those observa- tions.

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from math import exp
plt.rcParams["figure.figsize"] = (10, 6)
# Load the data
data = pd.read_csv("Data2.txt")

data.head()
```

Out[1]:

	3.895842
0	3.401982
1	2.945261
2	-3.628177
3	2.833800
4	5.135134

In [2]:
```
bin=int(1+3.22*(np.log10(len(data))))
bin
```

Out[2]:
```
12
```

In [3]:
```
X = data.iloc[:, 0]
X
```

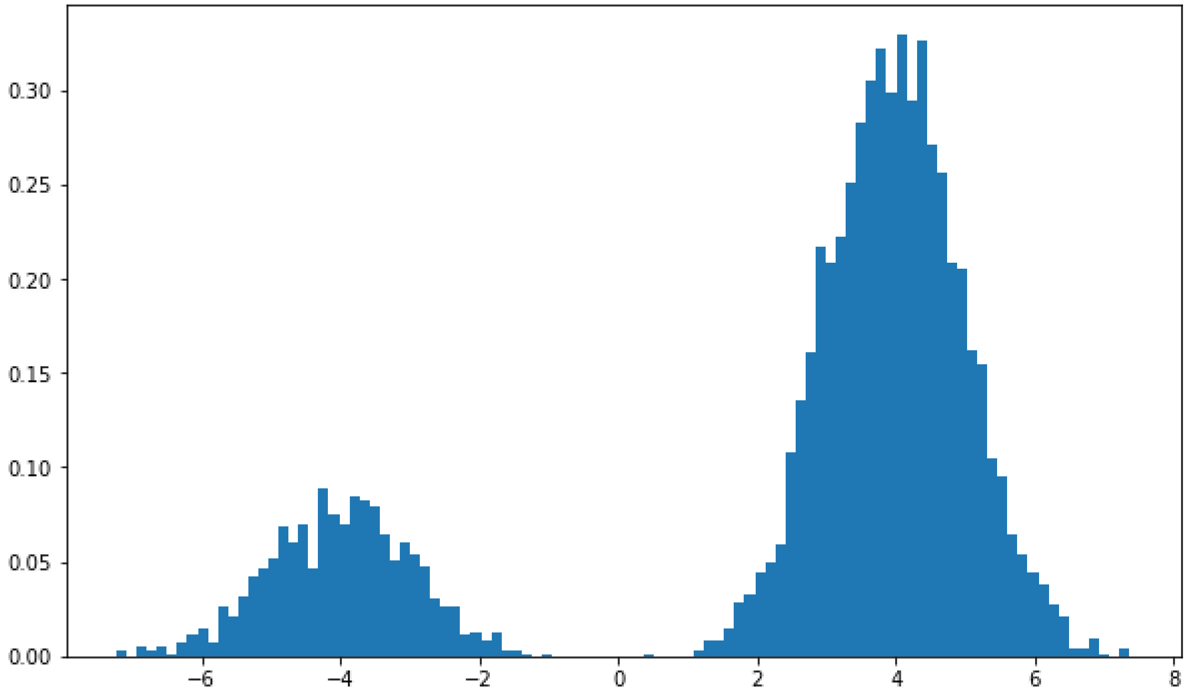
Out[3]:
```
0       3.401982
1       2.945261
2      -3.628177
3       2.833800
4       5.135134
          ...
4994    3.807107
4995    3.770204
4996   -3.642612
4997    1.995384
4998    2.417759
Name: 3.895842, Length: 4999, dtype: float64
```

In [4]:
```
x=np.linspace(0, 4998,num=4999)
len(x)
```

Out[4]:
```
4999
```

In [28]:
```
plt.hist(X,bins=100,density=True)
```

Out[28]:
```
(array([0.00274222, 0.        , 0.00548443, 0.00274222, 0.00548443,
        0.00137111, 0.00685554, 0.01096887, 0.0150822 , 0.00685554,
        0.02605107, 0.02056663, 0.0315355 , 0.04250437, 0.0466177 ,
        0.05210213, 0.06855544, 0.06032878, 0.06992655, 0.0466177 ,
        0.08912207, 0.07541098, 0.06992655, 0.08500874, 0.08226652,
        0.07952431, 0.06444211, 0.05073102, 0.06032878, 0.05347324,
        0.04798881, 0.03016439, 0.02605107, 0.02605107, 0.01096887,
        0.01233998, 0.00822665, 0.01233998, 0.00274222, 0.00274222,
        0.00137111, 0.        , 0.00137111, 0.        , 0.        ,
        0.        , 0.        , 0.        , 0.        , 0.        ,
        0.        , 0.        , 0.00137111, 0.        , 0.        ,
        0.        , 0.        , 0.00274222, 0.00822665, 0.00822665,
        0.0150822 , 0.02879328, 0.03290661, 0.04387548, 0.04935991,
        0.05895768, 0.10831759, 0.13573977, 0.16041972, 0.21663518,
        0.20840853, 0.22211962, 0.2509129 , 0.2824484 , 0.30438614,
        0.32221055, 0.29890171, 0.3290661 , 0.29478838, 0.32632388,
        0.27147953, 0.25639734, 0.20840853, 0.20566631, 0.16179083,
        0.15493529, 0.10420426, 0.0946065 , 0.06444211, 0.05347324,
        0.04387548, 0.03839104, 0.02742217, 0.02056663, 0.00411333,
        0.00411333, 0.00959776, 0.00137111, 0.        , 0.00411333]),
 array([-7.226604  , -7.08070747, -6.93481094, -6.78891441, -6.64301788,
        -6.49712135, -6.35122482, -6.20532829, -6.05943176, -5.91353523,
        -5.7676387 , -5.62174217, -5.47584564, -5.32994911, -5.18405258,
        -5.03815605, -4.89225952, -4.74636299, -4.60046646, -4.45456993,
        -4.3086734 , -4.16277687, -4.01688034, -3.87098381, -3.72508728,
        -3.57919075, -3.43329422, -3.28739769, -3.14150116, -2.99560463,
        -2.8497081 , -2.70381157, -2.55791504, -2.41201851, -2.26612198,
        -2.12022545, -1.97432892, -1.82843239, -1.68253586, -1.53663933,
        -1.3907428 , -1.24484627, -1.09894974, -0.95305321, -0.80715668,
        -0.66126015, -0.51536362, -0.36946709, -0.22357056, -0.07767403,
         0.0682225 ,  0.21411903,  0.36001556,  0.50591209,  0.65180862,
         0.79770515,  0.94360168,  1.08949821,  1.23539474,  1.38129127,
         1.5271878 ,  1.67308433,  1.81898086,  1.96487739,  2.11077392,
         2.25667045,  2.40256698,  2.54846351,  2.69436004,  2.84025657,
         2.9861531 ,  3.13204963,  3.27794616,  3.42384269,  3.56973922,
         3.71563575,  3.86153228,  4.00742881,  4.15332534,  4.29922187,
         4.4451184 ,  4.59101493,  4.73691146,  4.88280799,  5.02870452,
         5.17460105,  5.32049758,  5.46639411,  5.61229064,  5.75818717,
         5.9040837 ,  6.04998023,  6.19587676,  6.34177329,  6.48766982,
         6.63356635,  6.77946288,  6.92535941,  7.07125594,  7.21715247,
         7.363049  ]),
 <BarContainer object of 100 artists>)
```



The data is distributed in subsections. One in around -4 and another is around 4. So here I can seperate the data withrespect to pint 0. and apply normal distribution.

In [8]:

```
data_list=[]
for i in X:
    data_list.append(i)
```

In [10]:

```
data1=[] # is for right side distribution
data2=[] # is for left side distribution
for i in data_list:
    if i>0:
        data1.append(i)
    else:
        data2.append(i)
```

In [21]:

```
var1=np.var(data1)
var1
```

Out[21]: 0.9522098362932645

In [13]:

```
var2=np.var(data2)
```

In [14]:

```
mean1=np.mean(data1)
mean1
```

Out[14]: 3.9906373588606

In [16]:

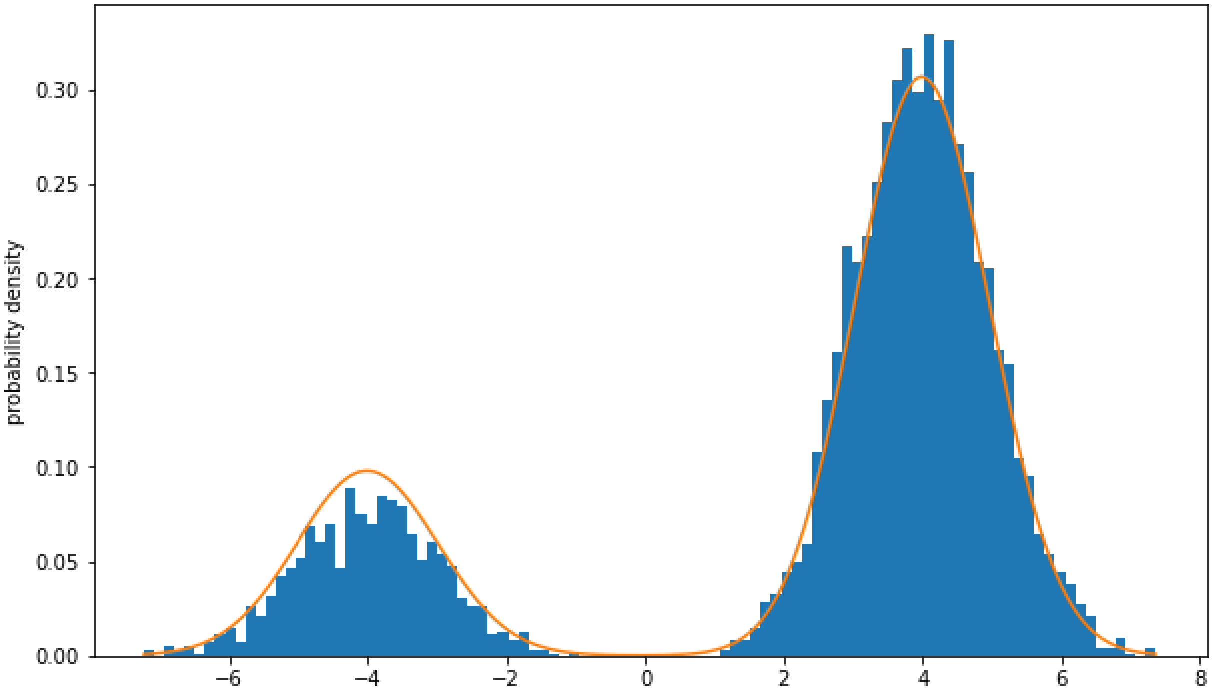
```
mean2=np.mean(data2)
mean2
```

Out[16]: -4.004748573643411

In [29]:
```
def function1(X,mean1,std1,mean2,std2):
    rest1= 1/np.sqrt(2*np.pi*std1**2)*np.exp((-1/(2*std1**2))*(X-mean1)**2)
    rest2= 1/np.sqrt(2*np.pi*std2**2)*np.exp((-1/(2*std2**2))*(X-mean2)**2)
    result= (rest1*3+rest2)/4    # WEIGHTED avg
    return result
```

In [30]:
```
plt.hist(X,bins=100,density=True)
xd2=np.linspace(np.min(X),np.max(X),1000)
plt.plot(xd2,function1(xd2,mean1,var1**(0.5),mean2,var2**(0.5)))
plt.ylabel("probability density")
```

Out[30]: Text(0, 0.5, 'probability density')



so its a normal distribution.

In [ ]:

R Notebook

This is an R Markdown (http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

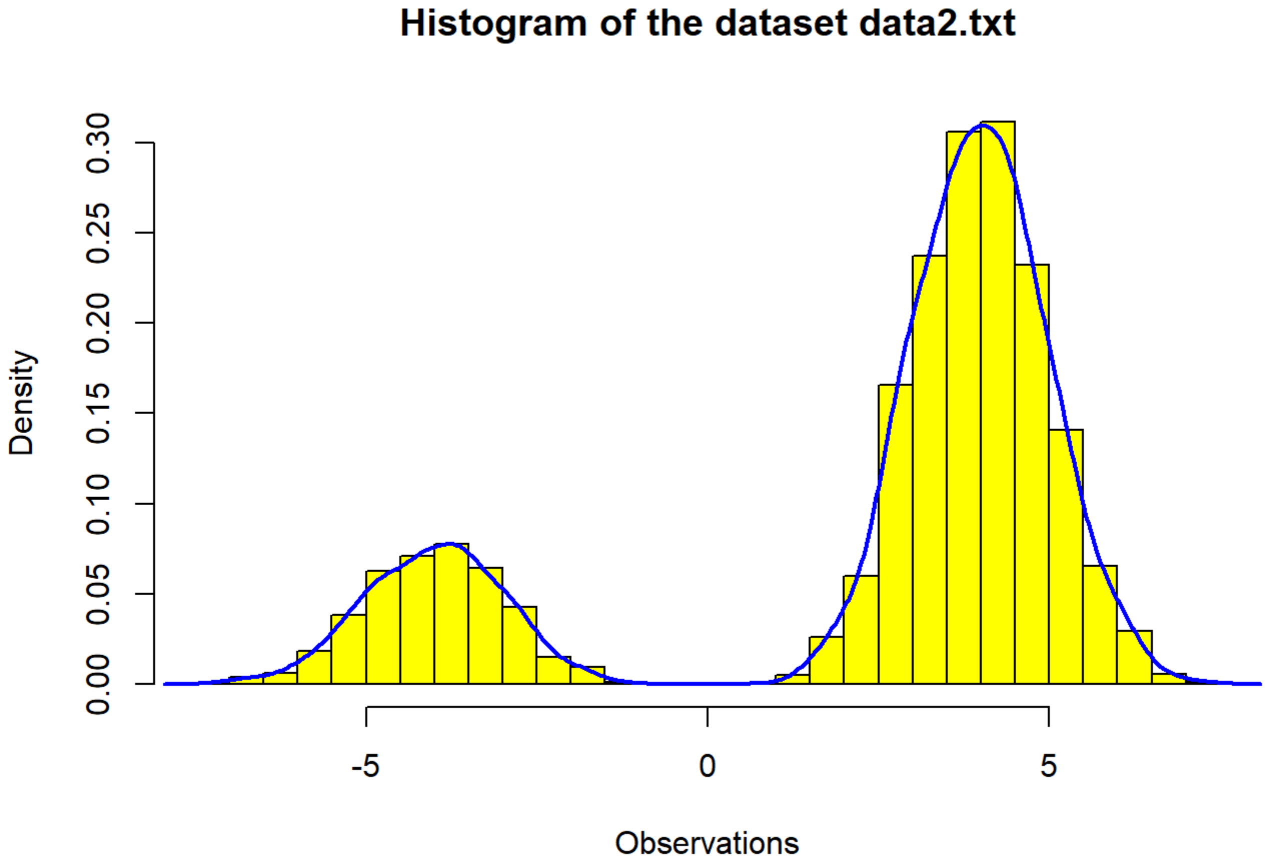
```
                      PRABAL GHOSH
```

QUESTION-4 Exercise 4: Let consider the dataset in data2.txt. Make a representation of it and propose a family of distribution that can model those observa- tions. DATASET- data2.txt ===

```
data<-read.table("C:\\Users\\praba\\Desktop\\uca1\\M1\\maths\\data2.txt", header = FALSE)

# histogram
data<-data$V1 #to use first  column
hist(data,
     main = "Histogram of the dataset data2.txt", #chart title
     xlab = "Observations",   #x-axis label
     ylab = "Density",      #y-axis label
     col = "yellow",         #desired colour
     border = "black",        #Border specification
     breaks = 50,             #Number of Bins
     probability = TRUE,      #Get the probabilities
     axes = FALSE,            #To overlay a density line
     )

# Add the axes
axis(1)
axis(2)
#Plot the density line
d <- density(data)
lines(d, col = "blue", lwd = 2)
```


Histogram of the dataset data2.txt
Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
-5
0
5
Observations