Assignment 6

Name: Rohit Chaudhary
Student ID Number: 112687906

Name: Prayag Goyal
Student ID Number: 112961923

Name: Nikhil Mangla
Student ID Number: 112820741

Name: Nikhil Mangla
Student ID Number: 112820741

Name: Abhay Goyal
Student ID Number: 112551276

Name: Sabrina Margetic
Student ID Number: 109898930

Q1a.

Date ___/___/______ Page No.: ______

(Q) $X_1, X_2, \ldots X_n$ are distributed Normally (θ, σ^2).
Prior for $\theta \longrightarrow$ Normal (a, b^2).

a.) Posterior $\Rightarrow f(\theta \mid \{X_1 \ldots X_n\}) = \dfrac{f(\{X_1, X_2 \ldots X_n\} \mid \theta) \cdot f(\theta)}{f(\{X_1, X_2 \ldots X_n\})}$

$$\vec{x} = \{X_1, X_2, \ldots X_n\}$$

$$f(\theta \mid \vec{x}) = \frac{f(\vec{x} \mid \theta) \cdot f(\theta)}{\underbrace{f(\vec{x})}_{\text{constant}}}$$

$$f(\theta \mid \vec{x}) \propto \underbrace{f(\vec{x} \mid \theta)}_{\text{likelihood}} \cdot \underbrace{f(\theta)}_{\text{prior}}$$

① Likelihood $f(\vec{x} \mid \theta) = \prod_{i=1}^{N} f(x_i \mid \theta) = \prod_{i=1}^{N} \underbrace{\dfrac{1}{\sigma\sqrt{2\pi}}}_{\text{constant}} e^{-\frac{(x_i-\theta)^2}{\underbrace{2\sigma^2}_{\text{constant}}}}$

$$= c \cdot e^{-\sum_{i=1}^{N}(x_i-\theta)^2}$$

② Prior $f(\theta) = \underbrace{\dfrac{1}{b\sqrt{2\pi}}}_{\text{constant}} e^{-\frac{(\theta-a)^2}{2b^2}}$

$$f(\theta \mid \vec{x}) \propto e^{-\sum_{i=1}^{N}(x_i-\theta)^2} \cdot e^{-\frac{(\theta-a)^2}{2b^2}}$$

We will just evaluate power of e and compare it with Normal distribution to get μ & σ^2.

$$\boxed{e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}}$$

"Failure comes only when we forget our ideals and objectives and principles." —Jawaharlal Nehru

Chitra

Date ___/___/______

$$\Rightarrow \quad \underbrace{-\sum_{i=1}^{N}\frac{(x_i-\theta)^2}{2\sigma^2}}_{\Downarrow \; ③} - \frac{(\theta-a)^2}{2b^2}$$

Power of e — Ⓐ

③ $\sum_{i=1}^{N}(x_i-\theta)^2 = \Sigma\left[(x_i-\bar{x})-(\theta-\bar{x})\right]^2 = \Sigma(x_i-\bar{x})^2+\Sigma(\theta-\bar{x})^2 - 2\Sigma(x_i-\bar{x})(\theta-\bar{x})$

$$= \Sigma(x_i-\bar{x})^2+\Sigma(\theta-\bar{x})^2-2(\theta-\bar{x})\left(\sum_{i=1}^{N}x_i-n\bar{x}\right)$$

$$= \Sigma(x_i-\bar{x})^2+\Sigma(\theta-\bar{x})^2-2(\theta-\bar{x})(\cancel{n\bar{x}}-\cancel{n\bar{x}})^{0}$$

$$= \Sigma(x_i-\bar{x})^2+n(\bar{x}-\theta)^2$$

Putting in A -

$$\underbrace{-\frac{\Sigma(x_i-\bar{x})^2}{2\sigma^2}} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-a)^2}{2b^2}$$

↳ We can ignore it as we will compare power of e to get distribution.

$$= \frac{-b^2n(\bar{x}-\theta)^2-\sigma^2(\theta-a)^2}{2\sigma^2b^2}$$

$$= \frac{-b^2n\bar{x}^2+2b^2n\bar{x}\theta-b^2n\theta^2-\sigma^2\theta^2-\sigma^2a^2+2\theta a\sigma^2}{2\sigma^2b^2}$$

$$= \frac{\theta^2\left(-b^2-\frac{\sigma^2}{n}\right)+2\theta\left(b^2\bar{x}+\frac{\sigma^2}{n}a\right)-\left(b^2n\bar{x}^2+\sigma^2a^2\right)\frac{1}{n}}{2\sigma^2b^2\times\frac{1}{n}}$$

$$= \frac{-\theta^2(b^2+s^2)+2\theta(b^2\bar{x}+s^2a)-(b^2\bar{x}^2+s^2a^2)}{2s^2b^2}$$

Putting back to e and $\propto$ being constant

$$= k e^{\frac{-(b^2+se^2)\left(\theta^2 - 2\theta\left(\frac{b^2\bar{x}+se^2a}{b^2+se^2}\right) + \frac{b^2\bar{x}^2+se^2a^2}{(b^2+se^2)}\right)}{2se^2b^2}}$$

Multiplying divide Numerator by $-(b^2+se^2)$

$$= k e^{\frac{-\left(\theta^2 - 2\theta\left(\frac{b^2\bar{x}+se^2a}{b^2+se^2}\right) + \frac{b^2\bar{x}^2+se^2a^2}{(b^2+se^2)}\right)}{\frac{2se^2b^2}{b^2+se^2}}}$$

= Multiplying and divide by $(b^2+se^2)(b^2\bar{x}+se^2a)^2$ and ignoring the constants we get.

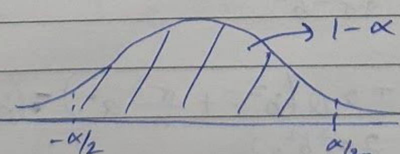
$$\propto e^{\frac{-\left(\theta - \frac{b^2\bar{x}+se^2a}{b^2+se^2}\right)^2}{2\,\frac{se^2b^2}{b^2+se^2}}}$$

as $x = \frac{b^2\bar{x}+se^2a}{b^2+se^2}$ & $y^2 = \frac{b^2se^2}{b^2+se^2}$

$f(\theta|\bar{x}) = \text{Nor}(x, y^2)$.

(b.) As posterior follows $\text{Nor}(x, y^2)$

Converting it to Standard Normal $Z(0,1)$: $\frac{\theta - x}{y}$



$1-\alpha$ posterior interval will be $(-Z_{\alpha/2}, Z_{\alpha/2})$

So, for θ it will be $(x - y.Z_{\alpha/2}, \; x + yZ_{\alpha/2})$

Q1b

Assignment 6 Question 2 Part A:

1-Initial distribution is Normal(0,1). Now variance is given in question for both parts.

2-By using the formula in question 1 part a, we can find out the new mean and variance. We have calculated expected mean and se square for each set of 100 observations to find out new mean and variance of posterior.

3-These new values are now considered as prior for the next step to calculate posterior.

```
import math
import argparse
import statistics
import numpy as np
import pandas as pd
import csv

with open("q2_sigma3.dat", "r") as myfile:
    priorMean=0
    priorVariance=1
    newMean=0
    newVariance=0
    givenSigmaSquare=9
    n=100
    seSquare=givenSigmaSquare/n
    for line in myfile:
        total=0
        expectedMean=0
        currentline = line.split(",")
        for value in currentline:
            total+=float(value)
        expectedMean=total/n
        newMean=(priorVariance*expectedMean+seSquare*priorMean)/(priorVariance+se
        newVariance=(priorVariance*seSquare)/(priorVariance+seSquare)
        print("New mean: ",  newMean, " New Variance: ", newVariance)
        priorMean=newMean
        priorVariance=newVariance
```

Table with estimates of the mean and variance of the posterior for all 5 steps:

Mean	Variance
4.590762414332327	0.08256880733944953
4.813523613446215	0.0430622009569378
4.921256878168492	0.02912621359223301
4.97283741207765	0.022004889975550123
4.983966097849453	0.01768172888015717

X-axis: value of observation x

Y-axis: probability value

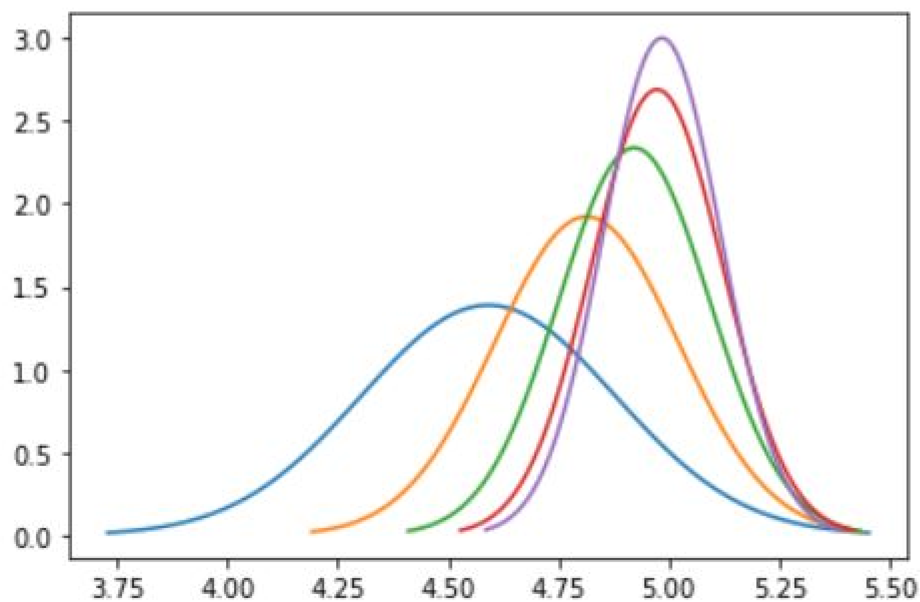
Distribution after 1st row: blue

Distribution after 2nd row: orange

Distribution after 3rd row: green

Distribution after 4th row: red

Distribution after 5th row: purple



Observation:

We can see that when standard deviation is low, the distribution is less confident of the mean value and the posterior interval (Theta) is updated quite substantially with each step, resulting in a converging mean and hence the distribution becomes more confident of it’s mean with ever more decreasing variance.

Assignment 6 Question 2 Part B:

```python
import math
import argparse
import statistics
import numpy as np
import pandas as pd
import csv

with open("q2_sigma100.dat", "r") as myfile:
    priorMean=0
    priorVariance=1
    newMean=0
    newVariance=0
    givenSigmaSquare=10000
    n=100
    seSquare=givenSigmaSquare/n
    for line in myfile:
        total=0
        expectedMean=0
        currentline = line.split(",")
        for value in currentline:
            total+=float(value)
        expectedMean=total/n
        newMean=(priorVariance*expectedMean+seSquare*priorMean)/(priorVariance+seSquare)
        newVariance=(priorVariance*seSquare)/(priorVariance+seSquare)
        print("New mean: ",  newMean, " New Variance: ", newVariance)
        priorMean=newMean
        priorVariance=newVariance
        print("\n")
```

Table with estimates of the mean and variance of the posterior for all 5 steps:

Mean	Variance
0.058716241479823114	0.9900990099009901
0.09500866961681816	0.9803921568627452
0.13822626152242073	0.970873786407767
0.17121883350740297	0.9615384615384617
0.2189182449674514	0.9523809523809524

X-axis: value of observation x

Y-axis: probability value

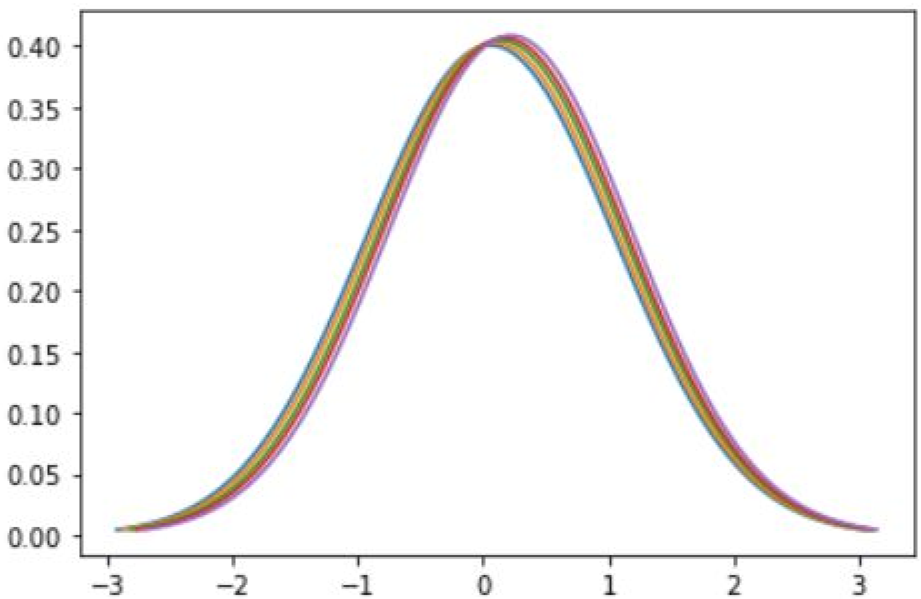
Distribution after 1st row: blue

Distribution after 2nd row: orange

Distribution after 3rd row: green

Distribution after 4th row: red

Distribution after 5th row: purple



Observation:

We can see that when variance is high, as compared to part a, with each progressing step, the distribution just moves slightly towards right (mean increases slightly every step with uniform amount) but there is no major sign of converging the mean and posterior confidence interval more or less remains the same.

Assignment 6 Question 2 Part C:

Conclusion:

In part a variance is low and the mean is converging to actual value and distribution is gaining confidence of true mean, but in part b variance is high and mean does not converge. Therefore, Bayesian inference works well for low variance and gets more confident (posterior interval decreases in such a case) of its mean with every step, whereas if the variance is high, the mean does not converge and posterior interval remains constant.

Q3.

Q3) a) Simple Linear Regression, $Y = \beta_0 + \beta_1 X + \epsilon_i$

where $E[\epsilon_i] = 0$

Show, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad \text{where } \bar{x} = \frac{\sum x_i}{n}$$

$$\bar{y} = \frac{\sum Y_i}{n}$$

Sum of Squared Errors $(S) = \sum_{i=1}^{n} (\hat{\epsilon}_i)^2$

$$= \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} \left(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2$$

We need to minimize S

$$\Rightarrow \frac{\partial S}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} \frac{\partial}{\partial \hat{\beta}_0} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2 = 0$$

$$\Rightarrow \sum_{i=1}^{n} -2 (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i)) = 0$$

$$\Rightarrow \sum Y_i = \sum (\hat{\beta}_0 + \hat{\beta}_1 X_i) = n \hat{\beta}_0 + \hat{\beta}_1 \sum X_i$$

Dividing by n on both sides,

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

$$\Rightarrow \boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$$

$$\frac{\partial S}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2\left(y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)(-x_i) = 0$$

$$\Rightarrow \sum x_i y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2$$

Substituting value of β_0 from ①

$$\Rightarrow \sum x_i y_i = (\bar{y} - \hat{\beta}_1 \bar{x}) \sum x_i + \hat{\beta}_1 \sum x_i^2$$

$$\Rightarrow \sum x_i y_i - \bar{y} \sum x_i = \hat{\beta}_1 \left(\sum x_i^2 - \bar{x} \sum x_i \right)$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum x_i y_i - \bar{y}(\sum x_i)}{\sum x_i^2 - \bar{x}(\sum x_i)} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i y_i - \bar{x}\bar{y})}{\sum_{i=1}^{n} (x_i^2 - \bar{x}^2)} = \frac{\sum (x_i y_i - \bar{x}\bar{y} + x_i \bar{y} - \bar{x} y_i)}{\sum (x_i^2 + \bar{x}^2 - 2\bar{x}^2)}$$

(Added and subtracted $x_i \bar{y}$, $\bar{x} y_i$, because $\sum x_i \bar{y} = \sum \bar{x} y_i$)

$$\Rightarrow \hat{\beta}_1 = \frac{\sum y_i (x_i - \bar{x}) - \bar{y}(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

$$\boxed{\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}}$$

Q3) b) Show that above estimators are unbiased

$\Rightarrow E[\hat{\beta}_1] = \beta_1$ for unbiased
and
$E[\hat{\beta}_0] = \beta_0$

$$E[\hat{\beta}_1] = E\left[\frac{\Sigma(X_i - \bar{X})(Y_i - \bar{Y})}{\Sigma(X_i - \bar{X})}\right] = E\left[\frac{\Sigma(X_i - \bar{X})Y_i}{\Sigma(X_i - \bar{X})^2}\right]$$

$$- E\left[\frac{\Sigma(X_i - \bar{X})\bar{Y}}{\Sigma(X_i - \bar{X})^2}\right] = 0$$

because $\Sigma(X_i - \bar{X}) = 0$

$$\Rightarrow E[\hat{\beta}_1] = E\left[\frac{\Sigma(X_i - \bar{X})(\hat{Y}_i)}{\Sigma(X_i - \bar{X})^2}\right]$$

Replacing $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \epsilon_i$

$$= E\left[\frac{\Sigma(X_i - \bar{X})(\hat{\beta}_0 + \hat{\beta}_1 X_i)}{\Sigma(X_i - \bar{X})^2}\right] \quad \left(\frac{\Sigma \epsilon_i}{\Sigma(X_i - \bar{X})} = 0\right)$$

$$= \underbrace{E\left[\frac{\hat{\beta}_0 \Sigma(X_i - \bar{X})}{\Sigma(X_i - \bar{X})^2}\right]}_{=0} + E\left[\frac{\hat{\beta}_1 \Sigma(X_i - \bar{X})X_i}{\Sigma(X_i - \bar{X})^2}\right]$$

Just subtracting 0

$$= E\left[\frac{\hat{\beta}_1 \Sigma((X_i - \bar{X})X_i) - \overbrace{\Sigma(\bar{X}(X_i - \bar{X}))}^{=0}}{\Sigma(X_i - \bar{X})^2}\right] = E\left[\hat{\beta}_1 \frac{\cancel{\Sigma(X_i - \bar{X})^2}}{\cancel{\Sigma(X_i - \bar{X})^2}}\right]$$

$$\boxed{E[\hat{\beta}_1] = \beta_1 \quad \therefore \hat{\beta}_1 \text{ is unbiased}}$$

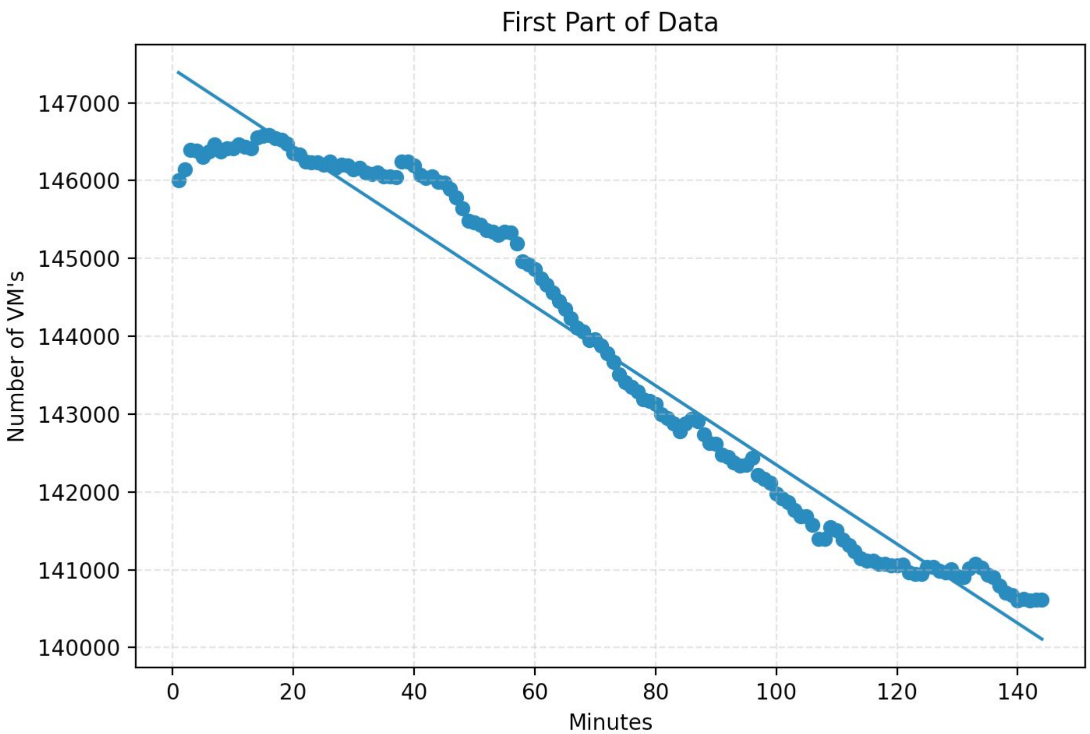
$$E[\hat{\beta}_0] = E[\bar{Y} - \hat{\beta}_1 \bar{X}_i] = E[\hat{\beta}_0 + \cancel{\hat{\beta}_1 \bar{X}} - \cancel{\hat{\beta}_1 \bar{X}_1}]$$

replacing $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \epsilon$ (From part a)

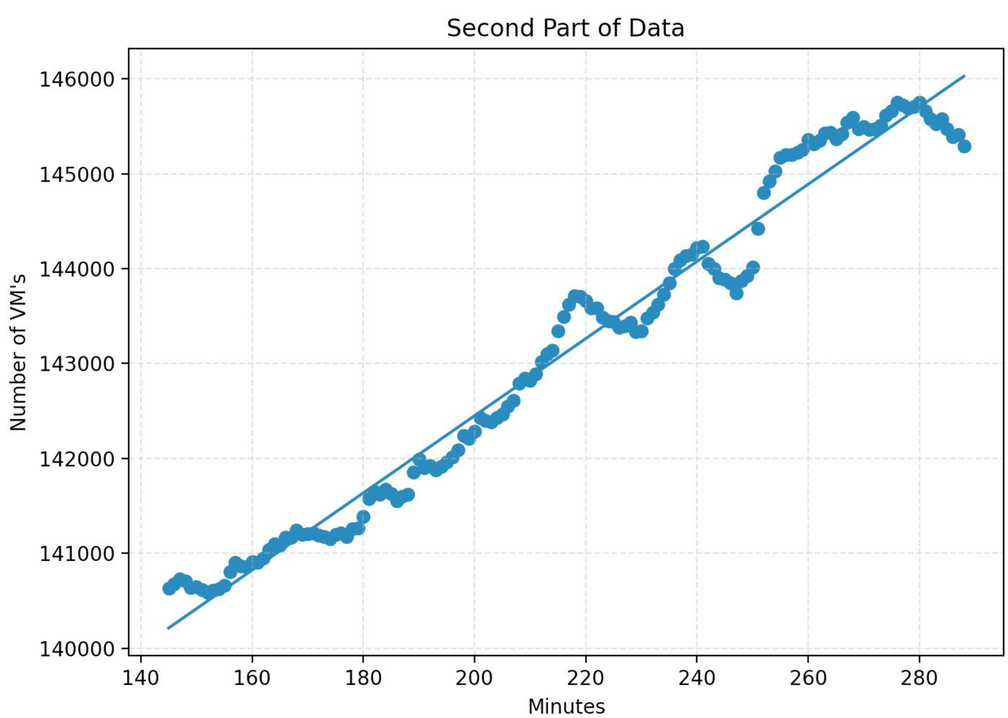
$$\Rightarrow \boxed{E[\hat{\beta}_0] = \beta_0 \quad \therefore \hat{\beta}_0 \text{ is unbiased}}$$

$E[\epsilon] = 0$

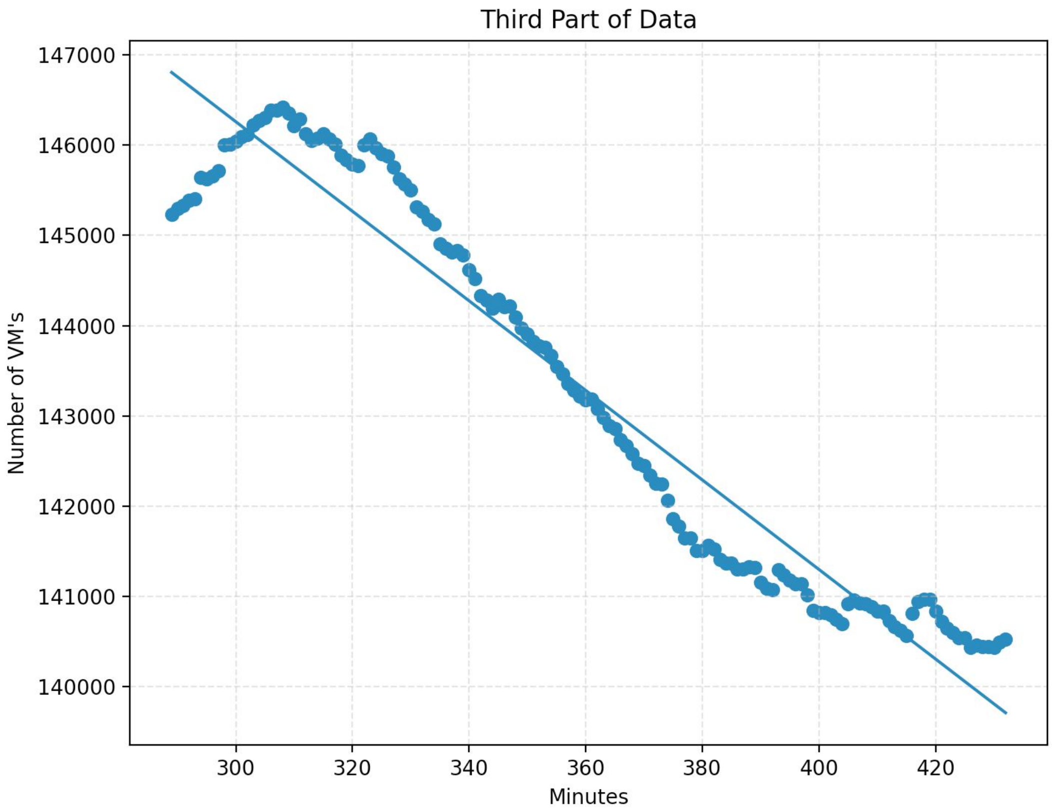
Question 4:

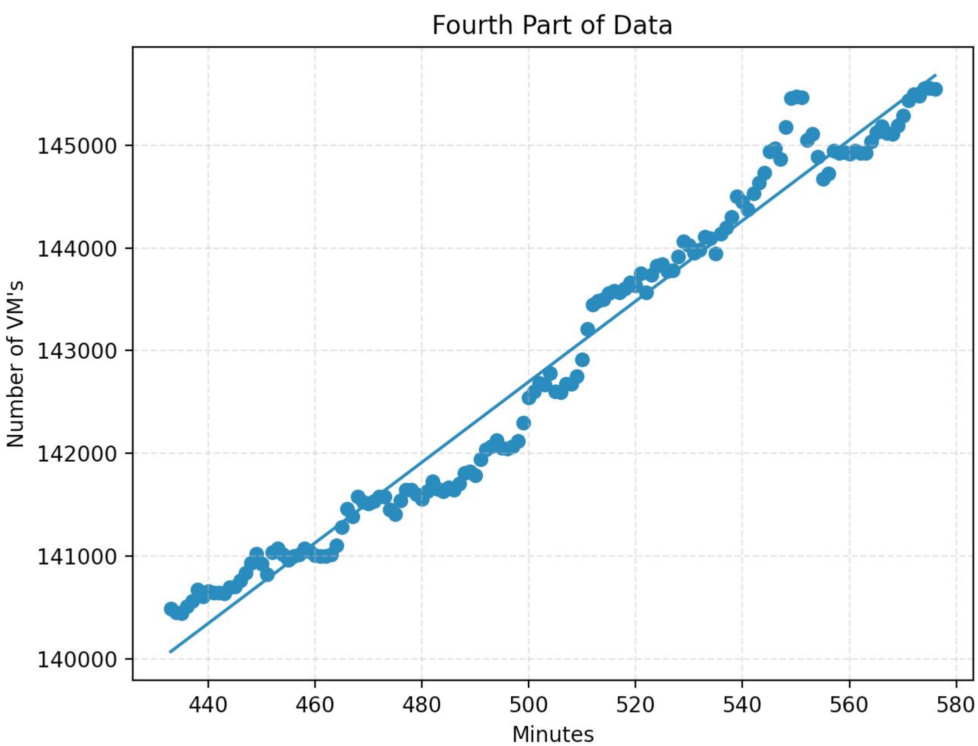


First Part of Data
Number of VM's
Minutes
147000
146000
145000
144000
143000
142000
141000
140000
0
20
40
60
80
100
120
140



Second Part of Data
Number of VM's
Minutes
146000
145000
144000
143000
142000
141000
140000
140
160
180
200
220
240
260
280


Third Part of Data
Number of VM's
Minutes
147000
146000
145000
144000
143000
142000
141000
140000
300
320
340
360
380
400
420


Fourth Part of Data
Number of VM's
Minutes
145000
144000
143000
142000
141000
140000
440
460
480
500
520
540
560
580

Sum of Squares for the first partition is: 28176831.90115946
Sum of Squares for the second partition is: 10656145.20472363
Sum of Squares for the third partition is: 43197449.811269216
Sum of Squares for the fourth partition is: 10948958.73407211

Q5

Q5.

We know that,

$f_W(w|H_0) = N(w; -\mu, \sigma^2)$ and

$f_W(w|H_1) = N(w, \mu, \sigma^2)$.

Also, as given

$$C = \begin{cases} 0 & \text{if } P(H=0|w) \geq P(H=1|w) \\ 1 & \text{o.w.} \end{cases}$$

Now, for choosing condition, we need to compare $Pr(H_0|w)$ and $Pr(H_1|w)$

$$\begin{aligned} Pr(H_0|w) &= Pr(w|H=0) \cdot Pr(H=0) \\ &= Pr(w|H=0) \cdot p \\ &= Pr(\{w_1 \dots w_n\}|H=0) \cdot p \\ &= \prod_{i=1}^{n} Pr(w_i|H=0)\, p \quad \text{--- (1)} \end{aligned}$$

Now, for $Pr(H_1|w)$

$$\begin{aligned} Pr(H_1|w) &= Pr(w|H=1)\, Pr(H=1) \\ &= Pr(w|H=1) \cdot (1-p) \\ &= \prod_{i=1}^{n} Pr(w_i|H=1)(1-p) \quad \text{--- (2)} \end{aligned}$$

Comparing ① & ②

$$\prod_{i=1}^{n} Pr(\omega_i \mid H=0) \cdot p \geq \prod_{i=1}^{n} Pr(\omega_i \mid H=1)(1-p)$$

$$p \cdot \prod_{i=1}^{n} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\omega_i-(-\mu)}{\sigma}\right)^2} \geq (1-p) \prod_{i=1}^{n} \frac{1}{\cancel{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\omega_i-\mu}{\sigma}\right)^2}$$

[∵ both are normally dis.]

$$p \cdot \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(\cancel{\omega_i^2}+\cancel{\mu^2}+2\mu\omega_i)} \geq (1-p)\, e^{-\frac{1}{2\sigma^2}(\cancel{\omega_i^2}+\cancel{\mu^2}-2\mu\omega_i)}$$

$$p \cdot \prod_{i=1}^{n} e^{-\frac{1}{2\sigma^2}(\cancel{2}\mu\omega_i)} \geq (1-p)\, e^{-\frac{1}{2\sigma^2}(-\cancel{2}\mu\omega_i)}$$

$$p \cdot e^{-\frac{1}{\cancel{2}\sigma^2}(\mu \sum_{1}^{n}\omega_i)} \geq (1-p)\, e^{-\frac{1}{\cancel{2}\sigma^2}(-\mu \sum_{1}^{n}\omega_i)}$$

$$e^{-\frac{2}{\sigma^2}(\mu \sum \omega_i)} \geq \frac{1-p}{p}$$

or $$\boxed{\sum \omega_i \leq -\frac{\sigma^2}{2\mu} \ln\left(\frac{1-p}{p}\right)}$$

Ans

(b)

```
import pandas as pd
import numpy as np

data = pd.read_csv('q5.csv')
prior = [0.1, 0.3, 0.5, 0.8]


for prior_val in prior:
    print("For P(Ho) = ",prior_val," the hypotheses selected are::", end = " ")
    for cols in data:
        sum_w = np.sum(data[cols].values)
        log_val = np.log(prior_val/(1 - prior_val))
        if sum_w < log_val:
            res = 0
        else:
            res = 1
        print(res, end = " ")
    print(" ")
```

(c) According to q5,

$AEP = Pr(C=0|H=1)\,Pr(H=1) + Pr(C=1|H=0)\,Pr(H=0)$

Now, we know from (a) that for $c=0$, the condition that needs to be satisfied is

$$\sum w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \quad \text{or}$$

$$Pr\left(\sum w_i \leq \frac{\sigma^2}{2\mu} \ln(p/1-p) \,\middle|\, H=1\right) Pr(H=1) +$$

$$Pr\left(\sum w_i > \frac{\sigma^2}{2\mu} \ln(p/1-p) \,\middle|\, H=0\right) Pr(H=0)$$

We know that $Pr(H=1) = 1-p$, $Pr(H=0) = p$.

thus,

$$Pr\left(\sum w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \,\middle|\, H=1\right) \cdot (1-p) +$$

$$Pr\left(\sum w_i > \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \,\middle|\, H=0\right) p.$$

Now, the condition given to us is

$f_{w_i}(w_i | H=0) \sim N(w_i; -\mu, \sigma^2)$. — ①

Each sample is conditionally independent.

$$\bar{w} = \frac{\Sigma w_i}{n}$$

So, ① becomes

$$f_{\bar{w}}(\bar{w} | H=0) \sim N(\bar{w}; -\mu, \sigma^2/n)$$

and

$$f_{\bar{w}}(\bar{w} | H=1) \sim N(\bar{w}; \mu, \sigma^2/n).$$

from the previous condition, we can say that $[H=1]$

$$Pr\left(\frac{\Sigma w_i}{n} \le \frac{\frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)}{n} \Big| H=1\right)$$

$$\Rightarrow Pr\left(\frac{\bar{w}-\mu}{\sigma^2/\sqrt{n}} \le \frac{\frac{\sigma^2}{2\mu n} \ln\left(\frac{p}{1-p}\right)}{\sigma^2/\sqrt{n}}\right)$$

or $\Pr\left(Z \le \frac{\sigma^{\cancel{2}}}{2\mu n} \ln\left(\frac{p}{1-p}\right) \frac{\sqrt{n}}{\cancel{\sigma}} - \frac{\mu\sqrt{n}}{\sigma}\right)$

$$= \Pr\left(Z \le \frac{\sigma}{2\mu n} \ln\left(\frac{p}{1-p}\right)\sqrt{n} - \frac{\mu\sqrt{n}}{\sigma}\right)$$

$$= \Phi\left(\frac{\sigma}{2\mu\sqrt{n}} \ln\left(\frac{p}{1-p}\right) - \frac{\mu\sqrt{n}}{\sigma}\right)$$

Now when $[H=0]$

$$\Pr\left(\sum w_i > \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \middle| H=0\right)$$

$$= \Pr\left(\frac{\sum w_i}{n} > \frac{\sigma^2}{2\mu n} \ln\left(\frac{p}{1-p}\right) \middle| H=0\right)$$

$$= \Pr\left(\frac{\bar{w}-(-\mu)}{\sigma/\sqrt{n}} > \frac{\frac{\sigma^2}{2\mu n} \ln\left(\frac{p}{1-p}\right)}{\sigma/\sqrt{n}} \middle| H=0\right)$$

$$= \Pr\left(Z > \frac{\sigma^{\cancel{2}}}{2\mu n} \ln\left(\frac{p}{1-p}\right) \frac{\sqrt{n}}{\cancel{\sigma}} + \frac{\mu\sqrt{n}}{\sigma}\right)$$

$$= \Pr\left(Z > \frac{\sigma}{2\mu\sqrt{n}} \ln\left(\frac{p}{1-p}\right) + \frac{\mu\sqrt{n}}{\sigma}\right)$$

$$= 1 - \Pr\left(Z < \frac{\sigma}{2\mu\sqrt{n}} \ln\left(\frac{p}{1-p}\right) + \frac{\mu\sqrt{n}}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{\sigma}{2\mu\sqrt{n}} \ln\left(\frac{p}{1-p}\right) + \frac{\mu\sqrt{n}}{\sigma}\right)$$

Combining,

$$ABP = \Pr\left(\sum w_i \le \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \Big| H=1\right)(1-p) +$$
$$\Pr\left(\sum w_i > \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \Big| H=0\right) \cdot p$$

$$ABP = \Phi\left(\frac{\sigma}{2\mu\sqrt{n}} \ln\left(\frac{p}{1-p}\right) - \frac{\mu\sqrt{n}}{\sigma}\right)(1-p) +$$
$$\left(1 - \Phi\left(\frac{\sigma}{2\mu\sqrt{n}} \ln\left(\frac{p}{1-p}\right) + \frac{\mu\sqrt{n}}{\sigma}\right)\right) p$$

Ans