CSE 544: Probability & Statistics for Data Science,
Spring 2021

Assignment - 6

Submitted by

Harmanpreet Singh Khurana, SBU ID: 113262379,
Mayank Jain, SBU ID: 113263864,
Rajadorai DS, SBU ID: 113259773,
Venkatesan Ravi, SBU ID: 113263484

Assignment - 6

1) Posterior for Normal :

(a) $X = \{X_1, X_2 \dots X_n\}$

σ^2 - Known

θ - mean $\Rightarrow$ unknown

$f(\theta|X) = f(\theta|\{X_1, X_2 \dots X_n\})$

$f(\theta) \sim Nor(a, b^2)$

$\Rightarrow f(\theta) = \frac{1}{b\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\theta - a}{b}\right)^2}$

Using Bayes

$$\Rightarrow \frac{f(\{X_1, X_2 \dots X_n\}|\theta) . f(\theta)}{f(\{X_1 \dots X_n\})}$$

$$\Rightarrow f(\theta|X) \propto f(\{X_1 \dots X_n\}|\theta) . f(\theta)$$

[Posterior $\propto$ likelihood $\times$ prior]

Likelihood

$$f(X|\theta) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x - \theta}{\sigma}\right)^2}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\sum_{i=1}^{n} \frac{1}{2}\left(\frac{X_i - \theta}{\sigma}\right)^2}$$

$$\Rightarrow f(\theta|X) \propto \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\sum_{i=1}^{n} \frac{1}{2}\left(\frac{X_i - \theta}{\sigma}\right)^2} \cdot \frac{1}{\sqrt{2\pi}b} \cdot e^{-\frac{1}{2}\left(\frac{\theta - a}{b}\right)^2}$$

Removing all the constants as we are considering proportionality → so only terms involving θ are relevant.

$$\Rightarrow f(\theta \mid x) \propto \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot \left(\frac{1}{b\sqrt{2\pi}}\right) \cdot e^{\left\{-\frac{1}{2}\left(\frac{\theta-a}{b}\right)^2 - \sum_{i=1}^{n} \frac{1}{2}\left(\frac{x_i-\theta}{\sigma}\right)^2\right\}}$$

$$f(\theta \mid x) \propto e^{-1/2 \underbrace{\left[\left(\frac{\theta-a}{b}\right)^2 + \sum_{i=1}^{n} \left(\frac{x_i-\theta}{b}\right)^2\right]}_{\text{let this be } t}}$$

$$\text{Let } f(\theta \mid x) \propto e^{-1/2 \cdot t}$$

$$t = \left(\frac{\theta-a}{b}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i-\theta}{b}\right)^2$$

$$t = \frac{\theta + \overset{\to \text{const}}{a^2} - 2a\theta}{b^2} + \frac{\overset{\to \text{const}}{\sum x_i^2} - 2\sum_{i=1}^{n} x_i\theta + \sum_{i=1}^{n}\theta^2}{\sigma^2}$$

Removing all constant terms that do not involve 'θ'.

$$t \propto \sigma^2(\theta^2 - 2a\theta) + b^2(-2\underbrace{\sum_{i=1}^{n} x_i}_{n\bar{x}}\theta + \theta^2 n)$$

$$\hookrightarrow t \propto \frac{\sigma^2\theta^2 - 2a\theta\sigma^2 + nb^2\theta^2 - 2n\bar{x}\theta b^2}{\sigma^2 b^2}$$

$$t \propto \frac{\theta^2(\sigma^2 + nb^2) - 2(\sigma^2 a + n\bar{x}b^2)\theta}{\sigma^2 b^2}$$

Dividing the numerator and denominator by

$(\sigma^2 + nb^2)$

$$t \propto \frac{\theta^2 - 2\theta\left(\frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}\right)}{\frac{\sigma^2 b^2}{\sigma^2 + nb^2}}$$

Now, in order to complete this as a square term, we add & subtract a constant term to t as follows:

$$t \propto \frac{\theta^2 - 2\theta\left(\frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}\right) + \left(\frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}\right)^2 - \underbrace{\left(\frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}\right)}_{\text{Constant independent of } \theta.}}{\frac{\sigma^2 b^2}{\sigma^2 + nb^2}}$$

$$t \propto \frac{\left(\left[\theta - \left(\frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}\right)\right]^2\right)}{\frac{\sigma^2 b^2}{\sigma^2 + nb^2}}$$

$$\Rightarrow f(\theta | x) \propto e^{-1/2}\left[\frac{\left(\theta - \frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}\right)^2}{\sigma^2 b^2 / \sigma^2 + nb^2}\right]$$

The constants can be adjusted with the proportionality sign to show that $f(\theta|x)$ follows a Normal distribution: Nor (X, y^2) with

the mean, $X = \dfrac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}$

and standard deviation, $y^2 = \dfrac{\sigma^2 b^2}{\sigma^2 + nb^2}$

Let $se = \sigma^2/n$,

$X = \dfrac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}$ (div by n) $= \dfrac{\frac{\sigma^2}{n} a + b^2 \bar{x}}{\frac{\sigma^2}{n} + b^2}$

$\therefore \boxed{X = \dfrac{se^2 a + b^2 \bar{x}}{se^2 + b^2}}$

For y^2, $y^2 = \dfrac{\sigma^2 + b^2}{\sigma^2 + nb^2}$ (div by n) $\Rightarrow \dfrac{\frac{\sigma^2}{n} \cdot b^2}{\frac{\sigma^2}{n} + b^2}$

$\boxed{y^2 = \dfrac{se^2 b^2}{se^2 + b^2}}$

∴ Posterior for Normal $= f(\theta|x)$

$= \text{Normal}(x, y^2)$

Hence, Proved.

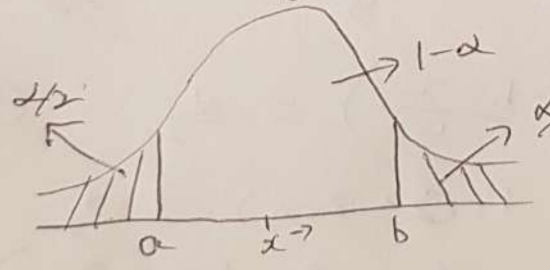
(b) Posterior interval of θ :-

Let $D = \{x_1, x_2 \ldots x_n\}$

To find the posterior $(1-\alpha)$ interval of θ is effectively to find $\theta \in [a,b]$ & find a and b.

$$\Rightarrow \Pr(\theta \in [a,b]) \geq 1-\alpha$$

$$\Pr(\theta < a|D) = \alpha/2, \quad \Pr(\theta > b|D) = \frac{\alpha}{2}$$



'Step A'

From result of (a) & converting it to standard normal by substracting the mean $-x$ and dividing by the standard deviation, y we get :-

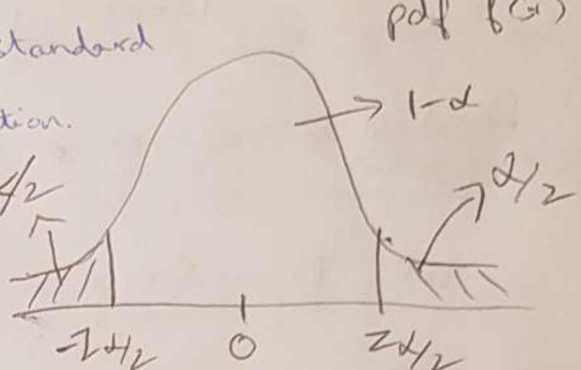
$$\Pr(\theta < a|D) = \Pr\left(\frac{\theta - x}{y} < \frac{a-x}{y} \middle| D\right) = \frac{\alpha}{2}$$

$$\rightarrow Pr\left(z < \frac{a-x}{y}\right) = \frac{\alpha}{2} \quad \cdots (3)$$

We know that

$$\rightarrow Pr\left(z < -z_{\alpha/2}\right) = \frac{\alpha}{2} \quad \cdots (4)$$

'Z' here is the standard normal distribution.



From (3) & (4), we get

$$\frac{a-x}{y} = -z_{\alpha/2}$$

$$\boxed{\therefore\ a = x - y\,z_{\alpha/2}}$$

$$Pr(\theta > b \mid D) = \alpha/2$$

Using 'Step A' from computing value of 'a',

we get $Pr\left(\frac{\theta - x}{y} > \frac{b-x}{y} \,\middle|\, D\right) = \alpha/2$

$$Pr\left(z > \frac{b-x}{y}\right) = \frac{\alpha}{2} \quad \cdots (5)$$

We know that : $Pr(z \leq z_{\alpha/2}) = 1 - \frac{\alpha}{2}$

$$\frac{\alpha}{2} = Pr\left(z > z_{\alpha/2}\right) \rightarrow (6)$$

From (5) & (6), $\frac{b-x}{y} = z_{\alpha/2}$; $\boxed{\therefore b = x + y\,z_{\alpha/2}}$

$\therefore (1-\alpha)$ Posterior interal for $\theta = [a, b] = [x - y\, Z_{\alpha/2},\ x + y\, Z_{\alpha/2}]$

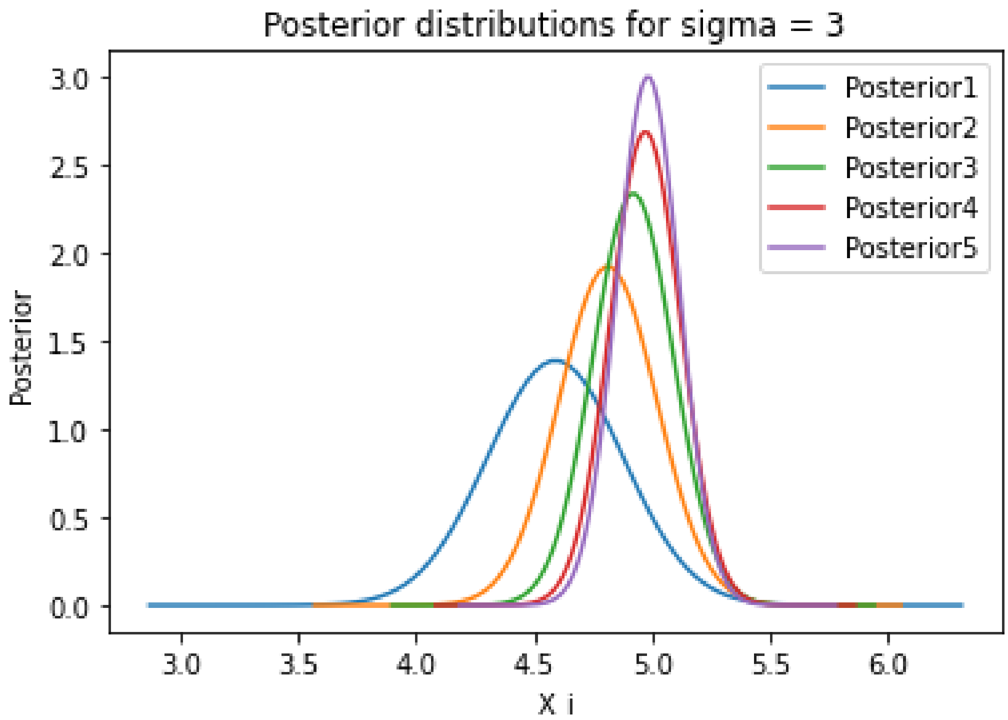
Substituting the values in x and y from part 1(a),

$$\left[\frac{b^2 \bar{x} + se^2 a}{b^2 + se^2} - Z_{\alpha/2}\left(\frac{b \cdot se}{\sqrt{b^2 + se^2}}\right),\ \frac{b^2 \bar{x} + se^2 a}{b^2 + se^2} + Z_{\alpha/2}\left(\frac{b \cdot se}{\sqrt{b^2 + se^2}}\right) \right]$$

↳ $\therefore$ This is the $(1-\alpha)$ Posterior interval for θ.

2)

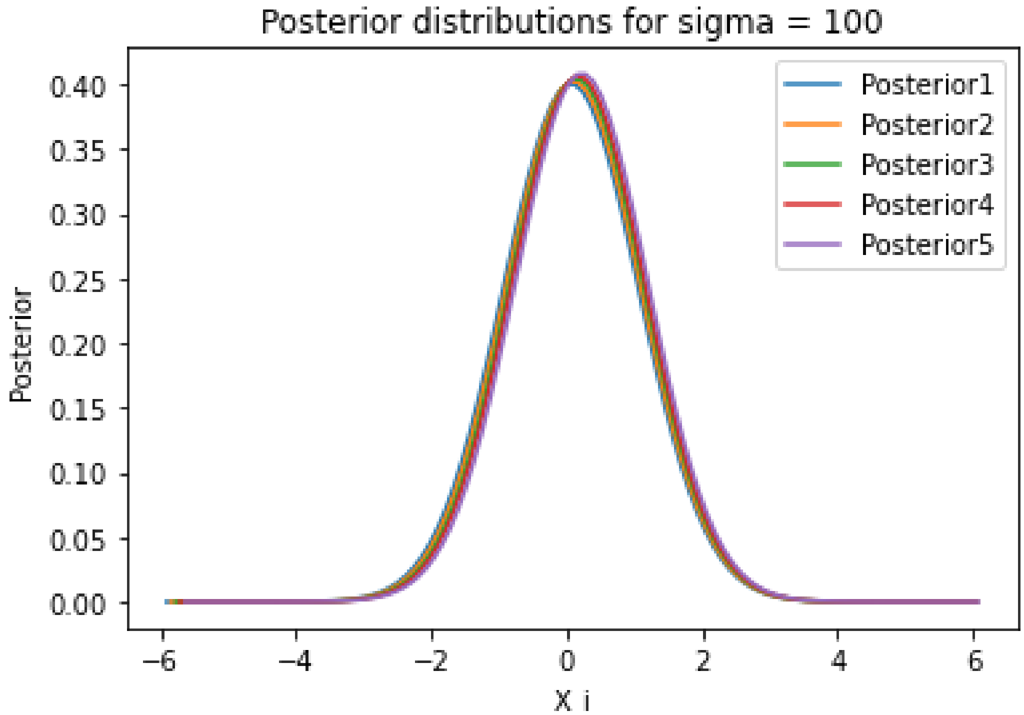
a)



Mean	Variance
4.590762	0.082569
4.813524	0.043062
4.921257	0.029126
4.972837	0.022005
4.983966	0.017682

Observation: When the variance decreases, the graph converges to the mean which is equal to 5.0

b)



Mean	Variance
0.058716	0.990099
0.095009	0.980392
0.138226	0.970874
0.171219	0.961538
0.218918	0.952381

Observation: When we start with a higher variance as our assumption for our trials/iterations, all posterior distribution converges to correct distribution or one distribution.

c) **Observation:** From the graphs we can observe that when the starting variance is higher the predicted distribution is close to MLE(part a), and when the starting variance is low, it is closer to prior.

3) Regression Analysis :

Given sample linear regression on 'n' sample points $(Y_1, X_1), (Y_2, X_2), (Y_3, X_3) \dots (Y_n, X_n)$ is

$$Y = \beta_0 + \beta_1 X + \varepsilon_i \text{, where } E[\varepsilon_i] = 0.$$

(a) We have

$$y_i | x_i = \beta_0 + \beta_1 x_i + \varepsilon_i \rightarrow (1)$$

$$E[y_i | x_i] = E[\beta_0 + \beta_1 x_i + \varepsilon_i]$$

$$= \beta_0 + \beta_1 x_i \text{ , (since } E[\varepsilon_i] = 0)$$

We need to get the estimates of β_0 and β_1 ,

we have $\hat{y}_i = \hat{E}[y_i | x_i] = \hat{\beta}_0 + \hat{\beta}_1 x_i \rightarrow (2)$

(plugin)

residual, $\hat{\varepsilon}_i = y_i - \hat{y}_i$

$$= y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

We can see that the RH.S is not an absolute error, so we square on both sides and take error for all the data samples, we have :

$$\text{Sum of square Errors (S)} = \sum_{i=1}^{n} (\hat{\varepsilon}_i)$$

$$(3) \leftarrow = \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$$

To find β_0 and β_1, minimize S,

Taking partial derivative on RHS wrt $\hat{\beta_0}$

$$\frac{\partial S}{\partial \hat{\beta_0}} = \sum_{i=1}^{n} \frac{\partial}{\partial \beta_0} (y_i - (\hat{\beta_0} + \hat{\beta_1} y_i))^2$$

$$= \sum_{i=1}^{n} 2(y_i - (\hat{\beta_0} + \hat{\beta_1} x_i))(0 - (1+0))$$

Here, β_0, x_i, y_i are constants.

$$\frac{\partial S}{\partial \hat{\beta_0}} = 0$$

$$\sum (y_i - (\hat{\beta_0} + \hat{\beta_1} x_i)) = 0$$

$$\sum y_i = \sum (\hat{\beta_0} + \hat{\beta_1} x_i)$$

$$\sum y_i = n\hat{\beta_0} + \hat{\beta_1} \sum x_i$$

$$\frac{\sum y_i}{n} = \hat{\beta_0} + \hat{\beta_1} \frac{\sum x_i}{n}$$

$$\boxed{\bar{y} = \hat{\beta_0} + \hat{\beta_1} \bar{x}} \Rightarrow 4$$

where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\sum_{i=1}^{n} \frac{y_i}{n}$.

$$\boxed{\therefore \hat{\beta_0} = \bar{y} - \hat{\beta_1} \bar{x}} \rightarrow I$$

Hence, proved.

Taking partial derivative of ③ w.r.t $\hat{\beta_1}$,

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

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

Substituting $\hat{\beta}_0$ from (4),

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

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

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

Simplifying further,

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \bar{y} \sum x_i \overbrace{- \bar{x} \sum y_i + \bar{x} \sum y_i}^{\text{adding \& subtracting same value}}}{\sum x_i^2 - \bar{x} \sum x_i \underbrace{- \bar{x} \sum x_i + \bar{x} \sum x_i}_{\text{adding \& subtracting same value.}}}$$

$$\hat{\beta}_1 = \frac{\sum x_i y_i - \sum \bar{y} x_i - \sum \bar{x} y_i + \bar{x} n . \bar{y}}{\sum x_i^2 - \sum \bar{x} x_i - \sum \bar{x} x_i + \bar{x} . n . \bar{x}}$$

Substituting $\bar{x} = \frac{\sum x_i}{n}$ and $\bar{y} = \frac{\sum y_i}{n}$,

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

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

Hence, Proved. $\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}}$ $\Rightarrow$ Shown in "I".

(b) Expectation of $\hat{\beta}_0$ and $\hat{\beta}_1$

$$E[\hat{\beta}_0] = E[\bar{y} - \hat{\beta}_1 \bar{x}]$$

Using $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$$E[\hat{\beta}_0] = E[\bar{y} - \hat{\beta}_1 \bar{x}] = E[\beta_0 + \beta_1 \bar{x} + \varepsilon_i - \hat{\beta}_1 \bar{x}]$$

using the following assumption : $E[\varepsilon_i] = 0$

$$E[\hat{\beta}_0] = \beta_0 + \beta_1 \bar{x} - \bar{x} E[\hat{\beta}_1] \longrightarrow (4)$$

$$\hat{\beta}_1 \Rightarrow E[\hat{\beta}_1] = E\left[\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right] = E\left[\frac{S_{xy}}{S_{xx}}\right]$$

Using $y_i = \beta_0 + \beta_1 x_i$ and $\bar{y} = \hat{\beta}_0 - \hat{\beta}_1 \bar{x}$ from (1)

we know that $\frac{\sum_{i=1}^{n} (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{S_{xx}} = 0$

Let $\frac{\sum_{i=1}^{n} (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = t$, then the expectation :

$$E[\hat{\beta}_1] = E\left[\sum t_i (y_i - \bar{y})\right]$$

$= \sum_{i=1}^{n} t_i \, E[y_i] - \underbrace{\sum t_i E[\bar{y}]}_{\to 0}$ $\quad t_i \to$ constant for given x_is

$= \sum_{i=1}^{n} t_i \, E[\beta_0 + \beta_1 x_i]$ $\quad \bar{x} \to$ constant for given x_is

$$\boxed{E[\hat{\beta}_1] = \beta_1} \Rightarrow (5)$$

$E[\hat{\beta}_0] = \beta_0 + \beta_1 \bar{x} - \bar{x} \, E[\hat{\beta}_1]$ [from (4)]

Substituting value of $E[\hat{\beta}_1]$ from (5),

$$E[\hat{\beta}_0] = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x}$$

$$\boxed{E[\hat{\beta}_0] = \beta_0}$$

Bias of $\hat{\beta}_0$: $\text{Bias}(\hat{\beta}_0) = E[\hat{\beta}_0] - \beta_0 = 0$

$$\boxed{\therefore \text{Bias}(\hat{\beta}_0) = 0}$$

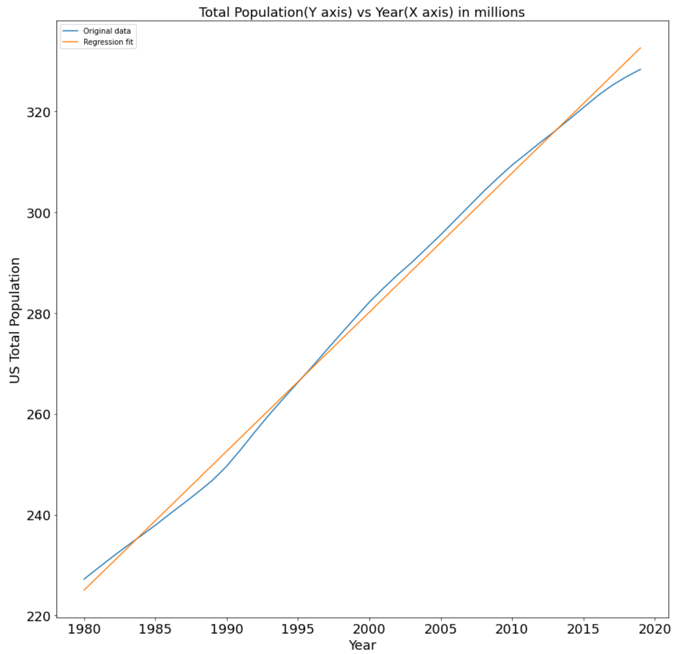
Bias of $\hat{\beta}_1$: $\text{Bias}(\hat{\beta}_1) = E[\hat{\beta}_1] - \beta_1 = 0$

$$\boxed{\therefore \text{Bias}(\hat{\beta}_1) = 0}$$

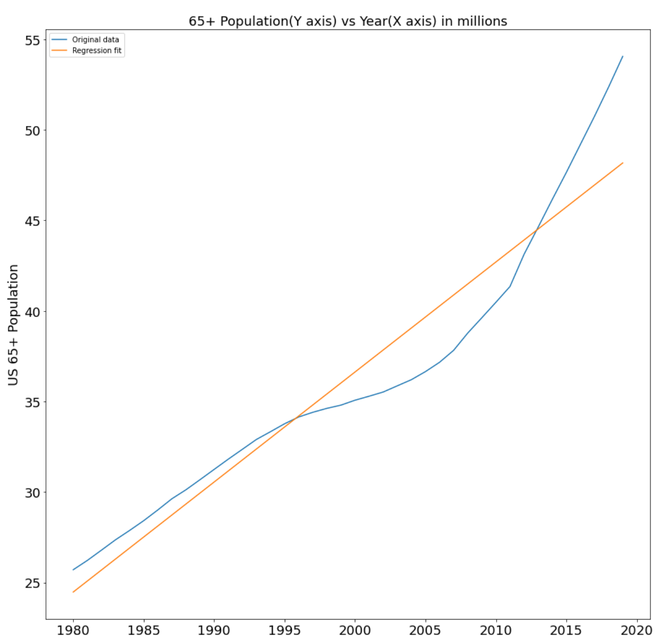
$\therefore$ Both estimators $\hat{\beta}_0, \hat{\beta}_1$ are unbiased.

Hence, Proved.

4. (a) For US Total Population vs Year:
Regression equation : **y = -5234.8564752 + 2.7575177 * x**
SSE : **123.8493735**



For US 65+ Population vs Year:
Regression equation : **y = -1178.5613646 + 0.6075945 * x**
SSE : **176.0353836**



US total population vs year is suitable for linear regression, while 65+ population vs Year is not.

4. (b) For 1980- 2018 data:

Regression equation : **y = -1131.0736031 + 0.5837632 * x**
SSE : **137.6926003**
Predicted 65+ population in 2060: **71.4786776 million**

For 2008- 2018 data:

Regression equation : **y = -2778.3894109 + 1.4025382 * x**
SSE : **2.008761**
Predicted 65+ population in 2060: **110.8394309 million**

We should trust the second prediction done using data from 2008 - 2018 because of it's low SSE. Media is right according to second prediction as the 65+ population is approximately doubling compared to 2018.

4. (c) Predicting ratio using method 1: **0.162362124**
SSE for method 1: **1.7980393**

Predicting ratio using method 2:
Ratio in 2019: **0.1617748**
Total Population in 2019**: 329.6888533 million**
65+ year population in 2019**: 53.3353617 million**
SSE: **2.008761**

Actual ratio as given in data: **0.1646461**

From the calculations done, prediction for ratio done using method 1 is more accurate compared to the second method as in that we used 65+ population for prediction which is non linear, while the total population which is used in method 1 is linear as can be seen from part a graph and also the SSE value for method 1 is lower compared to method 2 SSE value.

5. (a) Equation: [Output] = [Coefficients] * [Features]

    where Coefficients (excluding Beta o) :
    `[[-0.00291067],[0.00323154],[0.01990962],[0.00057609],[0.02319267],[0.1308982],[0.05682043]]`

    and Features: [GRE Score, TOEFL Score, University rating, SOP, LOR, GPA, Research]

    SSE: **0.31641141**

   (b) Equation: [Output] = [Coefficients] * [Features]

    where Coefficients (excluding Beta o): `[[0.00388655], [0.04187385], [0.04825699]]`

    and Features: [TOEFL Score, SOP, LOR]

    SSE: **0.64038876**

   (c) Equation: [Output] = [Coefficients] * [Features]

    where Coefficients (excluding Beta o) : `[[-0.00410631], [ 0.23571159]]`

    and Features: [GRE Score, GPA]

    SSE: **0.46380507**

   (d) We observe that when all the features are used in part (a), SSE is low so the prediction is more accurate compared to parts b and c. In Part b, SSE is more which means that these features should not be taken for prediction. For part c, SSE is not that high so these features are better as compared to part b for prediction.

Q6)

a) We have,

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

Derive a condition for choosing the hypothesis that soil is type 0.

i.e, $H=0$. (or) H_0 (or) $C=0$

When $C=0$, we must have,

$$P(H=0|w) \geq P(H=1|w)$$

We have to express this in terms of P, μ, σ.

Given,

$$H_0 : H=0$$
$$H_1 : H=1$$
$$P(H_0) = P(H=0) = P$$
$$P(H_1) = P(H=1) = 1-P$$

If RV for water content is W, then,

$$f_W(w|H=0) = N(w; -\mu, \sigma^2)$$

$$f_w(w|H=1) = N(w; \mu, \sigma^2)$$

Sample set of water $w = \{w_1, w_2, \dots, w_n\}$

We will calculate $P(H=0|w)$ and $P(H=1|w)$

By Baye's theorem

$$P(H=0|w) = \frac{P(w|H=0)\ P(H=0)}{P(w)}$$

Similarly,

$$P(H=1|w) = \frac{P(w|H=1)\ P(H=1)}{P(w)}$$

But since samples in w, i.e, $w_1, w_2, \dots w_n$ are conditionally independent of the original distribution (type of soil), we can say,

$$P(w|H=0) = \prod_{i=1}^{n} f_w(w|H=0)$$

Similarly,

$$P(w|H=1) = \prod_{i=1}^{n} f_w(w|H=1)$$

$$\therefore P(H=0|w) = \frac{\left[\prod_{i=1}^{n} f_w(w|H=0)\right] P(H=0)}{P(w)}$$

Since $f_w(w|H=0) = N(w; -\mu, \sigma^2)$,

$$\prod_{i=1}^{n} f_w(w|H=0) = \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{w_i+\mu}{\sigma}\right)^2}\right]$$

Similarly,

$$\prod_{i=1}^{n} f_w(w|H=1) = \prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{w-\mu}{\sigma}\right)^2}\right]$$

Also, $P(H=0) = p$, $P(H=1) = 1-p$. So,

$$P(H=0|w) \geq P(H=1|w)$$

$$\Rightarrow p\left[\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{w_i+\mu}{\sigma}\right)^2}\right]\right] \geq (1-p)\left[\prod_{i=1}^{n}\left[\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{w_i-\mu}{\sigma}\right)^2}\right]\right]$$

$$\therefore\ e^{\left(\sum_{i=1}^{n}\frac{(w_i-\mu)^2}{2\sigma^2} - \sum_{i=1}^{n}\frac{(w_i+\mu)^2}{2\sigma^2}\right)} \geq \frac{(1-p)}{p}$$

$$\Rightarrow e^{\frac{1}{2\sigma^2}\left[\cancel{\sum w_i^2} + \cancel{\sum \mu^2} - 2\mu\sum w_i - \cancel{\sum w_i^2} - \cancel{\sum \mu^2} - 2\mu\sum w_i\right]} \geq \frac{1-p}{p}$$

$$\Rightarrow e^{\frac{1}{2\sigma^2}\left[-4\mu\sum w_i\right]} \geq \frac{1-p}{p}$$

$$\Rightarrow e^{\frac{-2}{\sigma^2}\left[\mu\sum w_i\right]} \geq \frac{1-p}{p}$$

Taking ln on both sides

$$\frac{-2}{\sigma^2}\mu\sum_{i=1}^{n}w_i \geq \ln\left(\frac{1-p}{p}\right)$$

$$\Rightarrow \frac{2}{\sigma^2}\mu\sum_{i=1}^{n}w_i \leq -\ln\left(\frac{1-p}{p}\right)$$

$$\Rightarrow \frac{2}{\sigma^2}\mu\sum_{i=1}^{n}w_i \leq \ln\left(\frac{p}{1-p}\right)$$

$$\boxed{\therefore \sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2M} \ln\left(\frac{p}{1-p}\right)}$$

So thus the given data points $w_1, w_2, \ldots w_n$ this is the condition to check if H_0 holds or $C=0$ holds, in terms of p, σ, M.

6b)

For P(H_0) = 0.1, the hypotheses selected are :: [0, 1, 0, 0, 1, 0, 1, 1, 0, 1]
For P(H_0) = 0.3, the hypotheses selected are :: [0, 1, 0, 0, 1, 0, 1, 1, 0, 1]
For P(H_0) = 0.5, the hypotheses selected are :: [0, 1, 0, 0, 1, 0, 1, 1, 0, 1]
For P(H_0) = 0.8, the hypotheses selected are :: [0, 1, 0, 0, 1, 0, 1, 1, 0, 1]

c) $AEP = p\left(C=0 \mid H=1\right) p(H=1)$

$$+ \; p\left(C=1 \mid H=0\right) p\left(H=0\right)$$

Given $w = \{w_1, w_2, \cdots, w_n\}$ derive AEP in terms of $\mu, \sigma, \Phi()$ and p.

From the result of part a, $H_0 (H=0)$ (or) $C=0$ holds when,

$$\sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)$$

$\therefore \; p\left(C=0 \mid H=1\right)$

$$= p\left(\left(\sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)\right) \middle| H=1\right)$$

$\left[\text{Since } C=0 \equiv \sum_{i=1}^{n} w_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right)\right)$

$\sum_{i=1}^{n} w_i$ will also be a normal distribution,

(Since w_i's in $w = \{w_1, w_2, \cdots, w_n\}$ is normaly distributed and any operation on normal distribution gives us normal distribution.)

Now we can also find the mean and Variance of the normal distribution, $\sum_{i=1}^{n} w_i$.

By LOE, LOV

$$(\text{Mean}, \text{Variance}) = (n\mu, n\sigma^2)$$

$$\therefore \left(\sum_{i=1}^{n} w_i\right) \mid H=0 \sim N(-n\mu, n\sigma^2)$$

Similarly,

$$\left(\sum_{i=1}^{n} w_i\right) \mid H=1 \sim N(n\mu, n\sigma^2)$$

We know that if $X \sim N(\mu, \sigma^2)$,

$$\left(\frac{X-\mu}{\sigma}\right) \sim N(0,1).$$

This will help us represent $P(C=0 \mid H=1)$, $P(C=1 \mid H=0)$ in terms of $\Phi(\,)$

$\therefore$ We have,

$$P(C=0 \mid H=1) = \Phi\left(\frac{\frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) - n\mu}{\sqrt{n\sigma^2}}\right)$$

and,

$$P(C=1 \mid H=0) = 1 - \Phi\left(\frac{\frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) + n\mu}{\sqrt{n\sigma^2}}\right)$$

From these 2, we can write AEP as,

$$\Rightarrow AEP = P(C=0 \mid H=1) \cdot P(H=1) + P(C=1 \mid H=0) \cdot P(H=0)$$

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

$\left(\because P(H=0) = p,\ P(H=1) = 1-p\right)$

$\therefore$ AEP,

$$\boxed{\begin{gathered} AEP = (1-p)\,\Phi\left(\frac{S - n\mu}{\sigma\sqrt{n}}\right) + p\,\Phi\left(\frac{S + n\mu}{\sigma\sqrt{n}}\right) \\ \text{where } S = \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right) \end{gathered}}$$