CSE 544, Spring 2020: Probability and Statistics for Data Science

Assignment 6: Bayesian Inference and Regression Due: 04/29, 2:30pm, via google forms

(5 questions, 60 points total)

I/We understand and agree to the following:

(a) Academic dishonesty will result in an 'F' grade and referral to the Academic Judiciary.

(b) Late submission, beyond the 'due' date/time, will result in a score of 0 on this assignment.

(write down the name of all collaborating students on the line below)

---

1. Posterior for Normal

(Total 10 points)

Let $X_1, X_2, \ldots, X_n$ be distributed as Normal(θ, σ²), where σ is assumed to be known. You are also given that the prior for θ is Normal(a, b²).

(a) Show that the posterior of θ is Normal(x, y²), such that: (6 points)

$x = \frac{b^2\bar{X}+se^2 a}{b^2+se^2}$ and $y^2 = \frac{b^2 se^2}{b^2+se^2}$; where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $se^2 = \sigma^2/n$.

(Hint: less messier if you ignore the constants, but please justify why you can ignore them)

(b) Compute the (1-α) posterior interval for θ. (4 points)

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

$\sigma^2 \to$ known, $\theta \to$ unknown (mean)

$f(\theta|x) = f(\theta|\{x_1, x_2 \ldots x_n\})$

$\overset{\text{Bayes}}{=} \frac{f(\{x_1, x_2 \ldots x_n\}|\theta) \cdot f(\theta)}{f(\{x_1 \ldots x_n\})}$

$\Rightarrow f(\theta|x) \propto f(\{x_1 \ldots x_n\}|\theta) \cdot f(\theta)$

posterior $\propto$ likelihood $\times$ prior

$\hookrightarrow f(\theta) \sim \text{Nor}(a, b^2)$

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

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

Likelihood

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

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

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

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

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

$$f(\theta|x) \propto e^{-\frac{1}{2}\underbrace{\left[\left(\frac{\theta-a}{b}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i-\theta}{\sigma}\right)^2\right]}_{t}}$$

Let $f(\theta|x) \propto e^{-\frac{1}{2}\cdot t}$

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

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

Removing all constant terms (that do not involve θ).

$$t \propto \frac{\sigma^2(\theta^2 - 2a\theta) + b^2\left(\cancel{\sum X_i^2} - 2\overset{n\bar{X}}{\sum_{i=1}^{n} X_i}\theta + \theta^2(n)\right)}{\sigma^2 b^2}$$

Replacing $\sum_{i=1}^{n} X_i$ by $n\bar{X}$:

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

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

Dividing the numerator & denominator by $(\sigma^2 + nb^2)$:

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

Scanned with CamScanner

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

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

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

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

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, $x = \frac{\sigma^2 a + nb^2\bar{X}}{\sigma^2+nb^2}$

and std deviation, $y^2 = \frac{\sigma^2 b^2}{\sigma^2+nb^2}$

Let $se = \sigma^2/n$

$$X = \frac{\sigma^2 a + nb^2 \bar{x}}{\sigma^2 + nb^2}$$

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

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

$$y^2 = \frac{\sigma^2 b^2}{\sigma^2 + nb^2}$$

$$= \frac{\frac{\sigma^2}{n} \cdot b^2}{\frac{\sigma^2}{n} + b^2}$$

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

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

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

Scanned with CamScanner

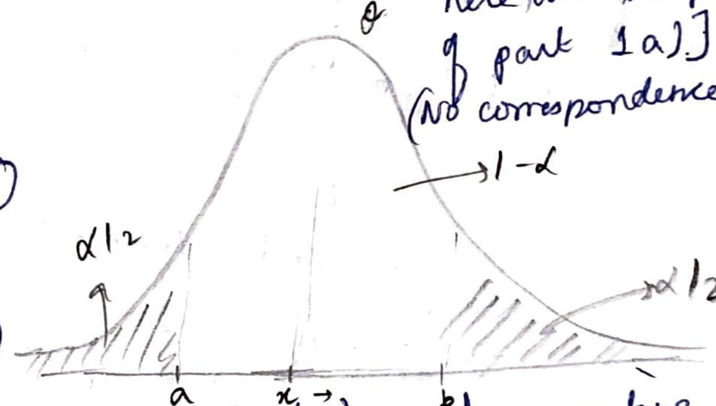
b) Compute the $(1-\alpha)$ posterior interval for θ

Let $D = \{X_1, X_2, \ldots X_n\}$

We want an ~~estimate~~ interval $[a, b]$ such that : [Note: a and b defined here are independent of part 1a)] (No correspondence)

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

$Pr(\theta < a | D) = \frac{\alpha}{2}$ ... (1)



$Pr(\theta > b | D) = \frac{\alpha}{2}$ ... (2)

Utilizing the result obtained in part 1a) and converting it to standard normal by subtracting the mean x and dividing by the standard deviation, y we get :-

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

$$\Rightarrow Pr\left(Z < \frac{a-x}{y}\right) = \frac{\alpha}{2} \quad \ldots (3)$$

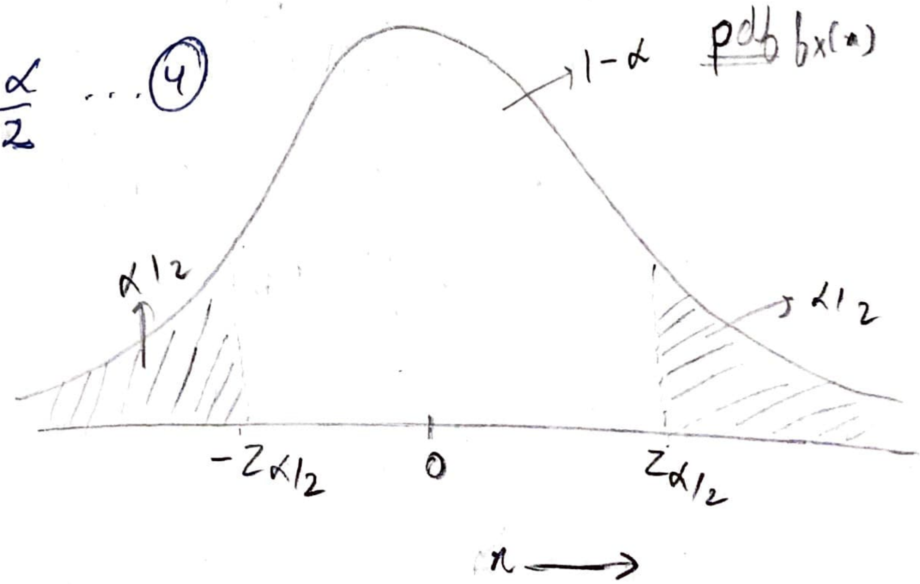
Z here refers to a std. normal distribution

We know that:

$\Rightarrow Pr(Z < -Z_{\alpha/2}) = \frac{\alpha}{2}$ ... (4)

From (3) and (4), we get :-

i.e $\frac{a-x}{y} = -Z_{\alpha/2}$



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

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

Converting to std normal by subtracting the mean x and dividing by std. deviation, y we get:-

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

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

We know that:

$$Pr(Z \leq Z_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

$$\frac{\alpha}{2} = 1 - Pr(Z \leq Z_{\alpha/2})$$

$$\frac{\alpha}{2} = Pr(Z > Z_{\alpha/2}) \quad \ldots (6)$$

from (5) and (6), we conclude.

$$\frac{b-x}{y} = Z_{\alpha/2}$$

$$\boxed{b = x + y Z_{\alpha/2}}$$

$(1-\alpha)$ Posterior interval for θ = $[a, b] = [x - yZ_{\alpha/2}, x + yZ_{\alpha/2}]$

Substituting values for x and y as obtained in part 1a), we get.

$(1-\alpha)$ Posterior interval for θ is:

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

↳ Answer

Page No.
Date

A2.

a) for $\sigma = 3$,

In Bayesian Inference, we are first assuming prior as Nor (0, 1) → given. We then see new data and update our estimates at each step.

So, in step 1,

we have prior = Nor (0, 1)

We see the first row & calculate posterior

Posterior (calculated) = Nor (4.59, 0.083)

Now, in step 2,

we take our prior as ~~step~~ posterior observed from step 1.

∴ Prior = Nor (4.59, 0.083)

Posterior (calculated from seeing row 2 data)
= Nor (4.813, 0.043)

Similarly, the steps are below:-

Step 3 Prior = Nor (4.813, 0.043)

Posterior = Nor (4.921, 0.029) - calculated

Step 4:- Prior :- Nor (4.921, 0.029)

Posterior ~~f4~~ :- Nor (4.972, 0.022)

Step 5:- Prior :- Nor (4.972, 0.022)

Posterior :- Nor (4.9839, 0.017)

Tabulating the results as below:-

Step	Mean	Variance
1	4.59	0.082
2	4.8135	0.043
3	4.9212	0.029
4	4.972	0.022
5	4.9839	0.0176

Looking at the graphs attached, we observe that as we see more data, the distributions shift closer to the true ~~mean~~ ~~& var.~~ distribution faster as our confidence increases with each step. The fluctuations at each step are large.

b) For $\sigma = 100$,
Again we start with Prior Nor(0, 1) and at each step, ~~esti~~ calculate the new posterior values seeing data row by row.

Step 1:-
Prior :- Nor (0, 1) → given
Posterior :- Nor (0.058, 0.990) → calculated

Step 2:-
Prior = Posterior of Step 1 = Nor (0.058, 0.99)
Posterior :- Nor (0.095, 0.9803)

Step 3:-
Prior = Nor(0.095, 0.9803)
Posterior = Nor(0.138, 0.97)

Step 4:-

Prior = Nor (0.138, 0.97)

Posterior = Nor (0.171, 0.961)

Step 5:-

Prior :- Nor (0.171, 0.961)

Posterior :- Nor (0.2189, 0.9523)

Tabulating the results of posterior after each step:-

Step	Mean	Variance
1	0.0587	0.99
2	0.0950	0.98
3	0.1382	0.971
4	0.1712	0.9615
5	0.2189	0.952

Looking at the graphs, we observe that the distributions at each step are almost the same (even the mean & variance values above do not change much). Hence the change from prior at step 1:- Nor (0,1) is very less at each step and finally posterior after step 5 is Nor (0.218, 0.95). Thus, it will take more time-steps to converge.

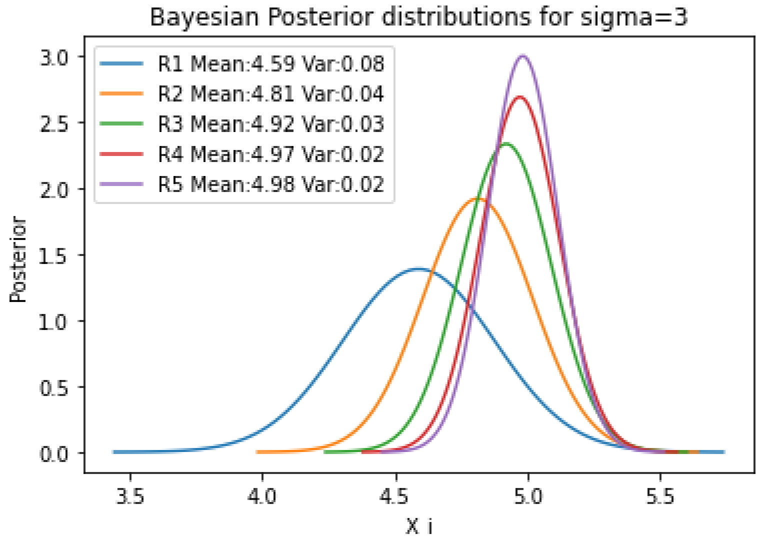
Scanned with CamScanner

Page No.
Date

A2.c After each step in Bayesian Inference, we keep refining our estimate as we see new data. From part (a), we observe that for small value of $\sigma = 3$, our estimates quickly converge to true distribution (large changes in mean & variance after each step) & our confidence increases (height of peaks of distributions increases). Whereas for part (b), for a high value of $\sigma = 100$, we observe low fluctuations in mean & variances & hence it will take more steps to converge (our confidence in estimates increased marginally).

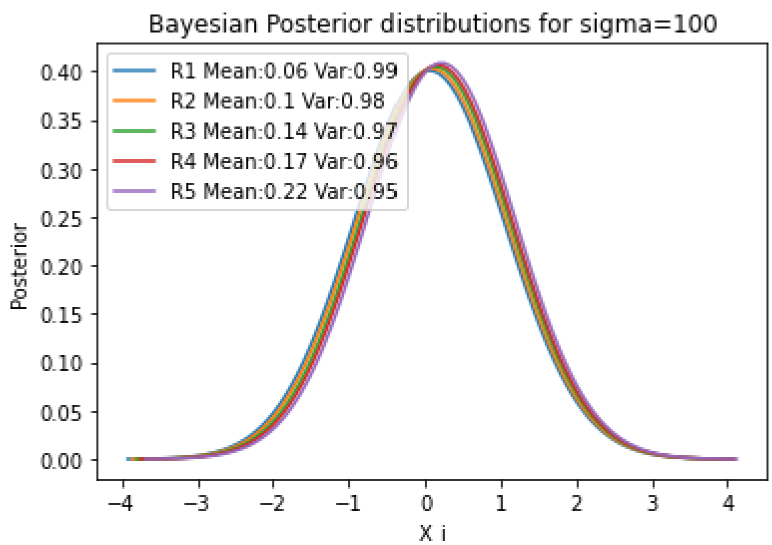
```
1 plotBayesianDistributions(sigma3, 3)
```

	Mean	Variance
0	4.590762	0.082569
1	4.813524	0.043062
2	4.921257	0.029126
3	4.972837	0.022005
4	4.983966	0.017682




```
1 plotBayesianDistributions(sigma100, 100)
```

	Mean	Variance
0	0.058716	0.990099
1	0.095009	0.980392
2	0.138226	0.970874
3	0.171219	0.961538
4	0.218918	0.952381



3. Regression Analysis **(Total 7 points)**

Assume Simple Linear Regression on n sample points $(Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)$; that is, $Y = \beta_0 + \beta_1 X + \varepsilon_i$, where $E[\varepsilon_i] = 0$.

(a) Derive the estimates of β when minimizing the sum of squared errors and show that:

$\widehat{\beta_1} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$ and $\widehat{\beta_0} = \bar{Y} - \widehat{\beta_1}\bar{X}$, where $\bar{X} = (\sum_{i=1}^{n} X_i)/n$ and $\bar{Y} = (\sum_{i=1}^{n} Y_i)/n$. (4 points)

(b) Show that the above estimators, given X_i's, are unbiased (Hint: Treat X's as constants) (3 points)

(a)

$$Y = \beta_0 + \beta_1 X + \varepsilon_i$$

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

calculating the expectation:

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

(assume $E[\varepsilon_i] = 0$)

$$E[y_i \mid x_i] = \beta_0 + \beta_1 x_i$$

(β_0 & $\beta_1 x_i$ are constants)

To estimate β_0 & β_1:

$$\hat{y}_i = E[y_i \mid x_i] = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad \text{(the residual)}$$

$$\Rightarrow \hat{\varepsilon}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i) \quad \text{—— ①}$$

using the OLS (Ordinary Least Squares method):

$$S = \sum_{i=1}^{n} (\hat{\varepsilon}_i)^2$$

goal: find $\hat{\beta}_0$ and $\hat{\beta}_1$ such that S is minimized.

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

(from ①)

Now, taking the partial derivative of S w.r.t. $\hat{\beta_0}$

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

$$\frac{\partial S}{\partial \hat{\beta_0}} = \sum_{i=1}^{n} 2\left(y_i - (\hat{\beta_0} + \hat{\beta_1} x_i)\right)(-1) = 0.$$

(Equating to zero to get the minima)

Since x_i & y_i are constants:

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

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

$$\sum_{i=1}^{n} y_i = n\hat{\beta_0} + \sum_{i=1}^{n} \hat{\beta_1} x_i$$

Dividing by n on both sides:

$$\Rightarrow \sum_{i=1}^{n} \frac{y_i}{n} = \hat{\beta_0} + \hat{\beta_1} \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = \hat{\beta_0} + \hat{\beta_1} \bar{x} \quad \text{--- (1)}$$

$$\Rightarrow \boxed{\hat{\beta_0} = \bar{y} - \hat{\beta_1} \bar{x}} \quad \text{--- (2)}$$

Now, taking the partial derivative of S wrt $\hat{\beta_1}$:

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

(equating to 0 to get the minima)

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

substituting $\hat{\beta_0} = \bar{y} - \hat{\beta_1} \bar{x}$ from (2):

$$\sum_{i=1}^{n} y_i x_i - (\bar{y} - \hat{\beta_1} \bar{x}) \sum_{i=1}^{n} x_i - \hat{\beta_1} \sum_{i=1}^{n} x_i^2 = 0.$$

$$\Rightarrow \sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i + \hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = 0.$$

$$\sum_{i=1}^{n} x_i y_i - \bar{y}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} y_i + \bar{x}\sum_{i=1}^{n} y_i = \hat{\beta}_1 \left[\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i - \bar{x}\sum_{i=1}^{n} x_i + \bar{x}\sum_{i=1}^{n} x_i \right]$$

Adding and subtracting $\bar{x}\sum_{i=1}^{n} y_i$ on LHS and $\hat{\beta}_1 \bar{x} \sum_{i=1}^{n} x_i$ on RHS :

$$\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + \bar{x}\sum_{i=1}^{n} y_i = \hat{\beta}_1 \left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \bar{x}\sum_{i=1}^{n} x_i \right]$$

On writing $\sum_{i=1}^{n} y_i$ as $n\bar{y}$ and $\sum_{i=1}^{n} x_i$ as $n\bar{x}$, we get :

$$\sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i - \bar{y}\sum_{i=1}^{n} x_i + \underbrace{n\bar{x}\bar{y}}_{\sum_{i=1}^{n} \bar{x}\bar{y}} = \hat{\beta}_1 \left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \underbrace{n\bar{x}\bar{x}}_{\sum_{i=1}^{n} \bar{x}^2} \right]$$

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

$$\Rightarrow \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \left(\sum_{i=1}^{n} (x_i - \bar{x})^2 \right)$$

Therefore,

$$\Rightarrow \boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}} = \frac{S_{xy}}{S_{xx}}$$

$$\text{and} \quad \boxed{\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}} \quad \text{—} \ (3)$$

(b) <u>Expectation of $\hat{\beta}_0$ and $\hat{\beta}_1$</u>

$\boxed{\hat{\beta}_0}$ $\quad E[\hat{\beta}_0] = E[\bar{Y} - \hat{\beta}_1 \bar{X}]$

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 : $E[\varepsilon_i] = 0$.

$$\boxed{E[\hat{\beta}_0] = \beta_0 + \beta_1 \bar{X} - \bar{X} E[\hat{\beta}_1]} \quad \text{—} \ (4)$$

$\boxed{\hat{\beta}_1}$ $\quad E[\hat{\beta}_1] = E\left[\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2}\right] = E\left[\frac{S_{xy}}{S_{xx}}\right]$

Using :

$Y_i = \beta_0 + \beta_1 X_i$ and $\bar{Y} = \hat{\beta}_0 - \hat{\beta}_1 \bar{X}$

from (3)

We know that:

Let $$\frac{\sum_{i=1}^{n}(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})}{S_{xx}} = 0$$

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

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

$t_i \to$ constant for given X_i's

$$= \sum_{i=1}^{n} t_i E[Y_i] - \cancel{\sum t_i E[\bar{Y}]}$$

$\bar{X} \to$ constant for given X_i's

$$= \sum_{i=1}^{n} t_i E[\beta_0 + \beta_1 X_i]$$

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

$$E[\hat{\beta}_0] = \beta_0 + \beta_1 \bar{X} - \bar{X} E[\hat{\beta}_1] \quad \text{from (4)}.$$

substituting the value of $E[\hat{\beta}_1]$ from (5):

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

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

Scanned with CamScanner

Bias of $\hat{\beta_0}$:

$$Bias(\hat{\beta_0}) = E[\hat{\beta_0}] - \beta_0$$

$$\boxed{Bias(\hat{\beta_0}) = 0}$$

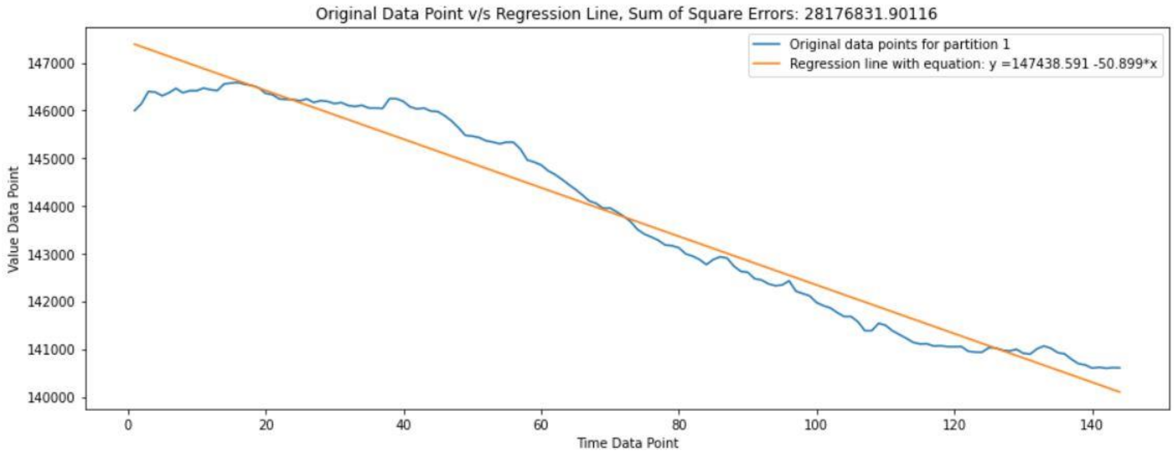
Bias for $\hat{\beta_1}$:

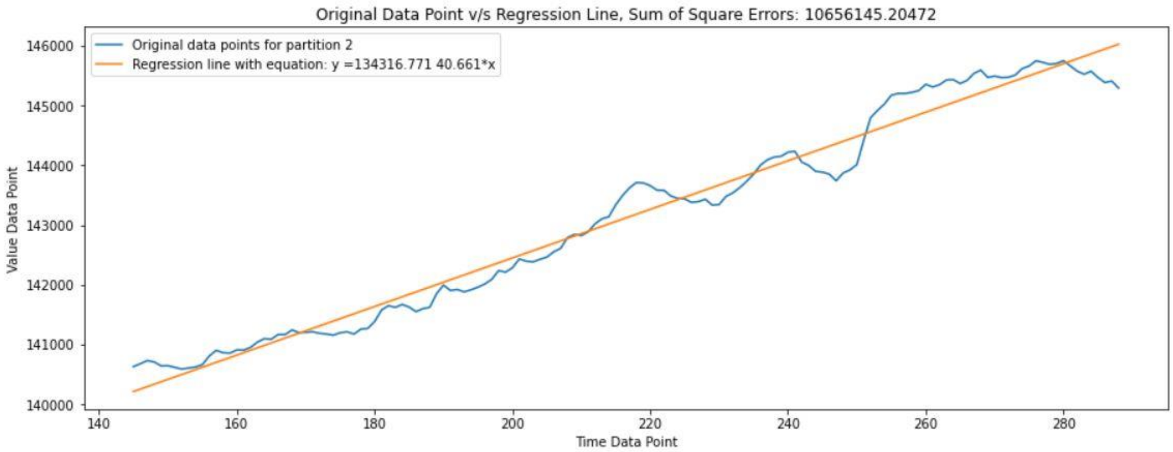
$$Bias(\hat{\beta_1}) = E[\hat{\beta_1}] - \beta_1$$

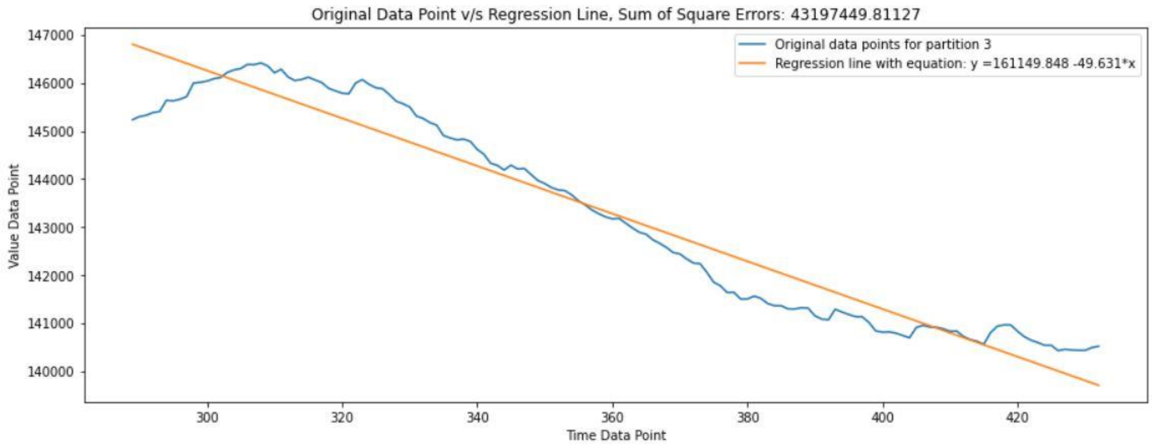
$$\boxed{Bias(\hat{\beta_1}) = 0}$$

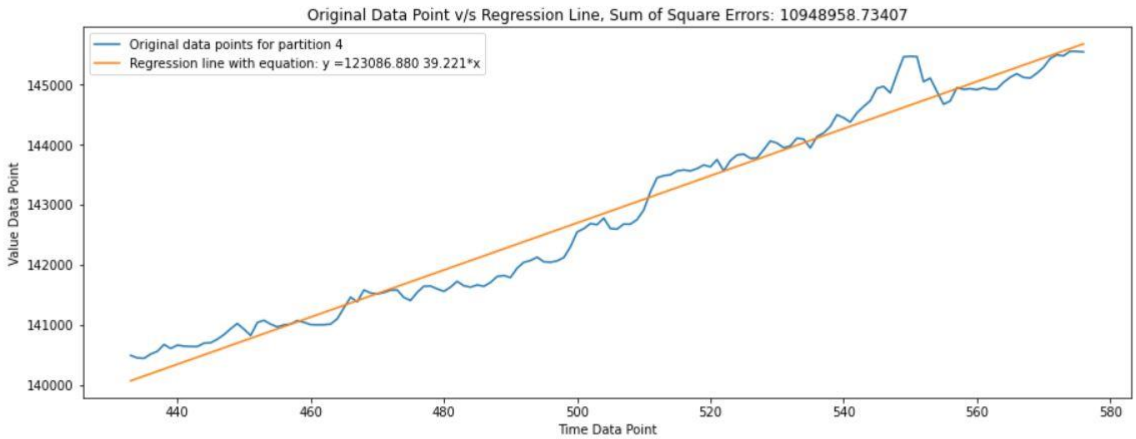
4a.

```
Sum of Square Errors in case: 1 is: 28176831.90116
Sum of Square Errors in case: 2 is: 10656145.20472
Sum of Square Errors in case: 3 is: 43197449.81127
Sum of Square Errors in case: 4 is: 10948958.73407
```






Original Data Point v/s Regression Line, Sum of Square Errors: 43197449.81127
Original data points for partition 3
Regression line with equation: y =161149.848 -49.631*x
Value Data Point
Time Data Point
147000
146000
145000
144000
143000
142000
141000
140000
300
320
340
360
380
400
420


Original Data Point v/s Regression Line, Sum of Square Errors: 10948958.73407
Original data points for partition 4
Regression line with equation: y =123086.880 39.221*x
Value Data Point
Time Data Point
145000
144000
143000
142000
141000
140000
440
460
480
500
520
540
560
580

5(a)

$C=0$ i.e. we choose $H=0$ (good soil) iff $P(H=0|\omega) \geq P(H=1|\omega)$

Now,

$$P(H=0|\omega) = \frac{P(\omega|H=0) \cdot P(H=0)}{P(\omega)} \quad \text{(by Bayes thm)}$$

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

$C=0$ iff, $\cancel{P(H=0|\omega) \cdot P(H=0)} \cancel{\geq P(\omega|H=1) \cdot P(H=1)}$

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

$$\frac{P(\omega|H=0) \cdot P(H=0)}{P(\omega)} \geq \frac{P(\omega|H=1) \cdot P(H=1)}{P(\omega)}$$

$$P(\omega|H=0) \cdot P(H=0) \geq P(\omega|H=1) \cdot P(H=1) \quad [\because p(\omega) \text{ is +ve}]$$

$$P(\omega|H=0) \cdot p \geq p(\omega|H=1) \cdot (1-p)$$

Now, $\omega = \{\omega_1, \omega_2, \omega_3 \dots \omega_n\}$

$$P(\omega_1 \dots \omega_n | H=0) \cdot p \geq P(\omega_1 \dots \omega_n | H=1)(1-p)$$

Given that samples are conditionally independent given hypothesis

$$\therefore \prod_{i=1}^{n} P(\omega_i | H=0) \cdot p \geq \prod_{i=1}^{n} P(\omega_i | H=1)(1-p)$$

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

this is because $f_W(\omega|H=0) = N(\omega_i; -\mu, \sigma^2)$

& $f_W(\omega|H=1) = N(\omega_i; \mu, \sigma^2)$

$\therefore$ We have,

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

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

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

$$e^{\frac{-\cancel{4}^{2}\mu\sum_{i=1}^{n}\omega_i}{\cancel{2}\sigma^2}} \geq \frac{(1-p)}{p}$$

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

taking ln on both sides

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

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

this is the condition for choosing the hypothesis H=0

5(c) Average Error Probability (AEP) $= P(C=0|H=1)P(H=1) + P(C=1|H=0).P(H=0)$

$P(C=0|H=1) = P(\text{Choosing } H=0 \mid H=1)$

from 5(a), we know that we choose $H=0$ iff,

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

$$\therefore P(C=0|H=1) = P\left(\sum_{i=1}^{n} \omega_i \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{P}{1-P}\right) \Big| H=1\right)$$

As ω_i's are normally distributed given $H=1$ with μ, σ^2

$W' = \sum_{i=1}^{n} \omega_i$ is also normally distributed with mean $n\mu$ &

std. deviation $n\sigma^2$ (linear transformation of normal distribution)

$$\therefore P(C=0|H=1) = P\left(W' \leq \frac{\sigma^2}{2\mu} \ln\left(\frac{P}{1-P}\right) \Big| H=1\right)$$

let $\frac{\sigma^2}{2\mu} \ln\left(\frac{P}{1-P}\right)$ be "c"

$$\therefore P(C=0|H=1) = P(W' \leq c \mid H=1)$$

$$= \Phi\left(\frac{c - n\mu}{\sqrt{n}\,\sigma}\right)$$

$$\therefore P(C=0|H=1).P(H=1) = \Phi\left(\frac{c-n\mu}{\sqrt{n}\,\sigma}\right)(1-P) \quad \text{--- (1)}$$

Similarly, $P(C=1 \mid H=0) = P\left(\sum_{i=1}^{n} \omega_i > \frac{\sigma^2}{2\mu} \ln\left(\frac{p}{1-p}\right) \mid H=0\right)$

$$= P(W' > c \mid H=0)$$

where, $W' = \sum_{i=1}^{n} \omega_i$ & W'_i is normally distributed given $H=1$ with mean $-n\mu$ & std. deviation σ.

$$P(C=1 \mid H=0) = P(W' > c \mid H=0)$$

$$= 1 - \Phi\left(\frac{c+n\mu}{\sqrt{n}\,\sigma}\right)$$

$$\therefore P(C=1 \mid H=0) \cdot P(H=0) = \left(1 - \Phi\left(\frac{c+n\mu}{\sqrt{n}\,\sigma}\right)\right) \cdot (p) \quad \text{--- (2)}$$

from (1) & (2)

$$AEP = \Phi\left(\frac{c-n\mu}{\sqrt{n}\,\sigma}\right)(1-p) + \left[1 - \Phi\left(\frac{c+n\mu}{\sqrt{n}\,\sigma}\right)\right] \cdot p$$