Probability and Statistics

Assignment 6

Sharfuddin Mohd.	112070993
Irfan Ahmed	113166464
Anurag Yepuri	113070893
Fatemeh Aslan Beigi	113870646

1.)

1.) a.) $x_1, x_2, \ldots, x_n \sim N(\theta, \sigma^2)$

Prior $f(\theta) \sim N(a, b^2)$

Posterior: $f(\theta \mid \{x_1, \ldots, x_n\}) \propto f(\{x_1, \ldots, x_n\} \mid \theta) \cdot f(\theta)$

$$f(\theta/D) \propto \prod_{i=1}^{n} f(x_i \mid \theta) \cdot f(\theta)$$

$$\propto \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_i - \theta)^2}{2\sigma^2}} \right) \cdot \frac{1}{\sqrt{2\pi b^2}} e^{-(\theta-a)^2/2b^2}$$

$$\propto \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{\frac{-1}{2\sigma^2}\sum (x_i - \theta)^2} \cdot f(\theta)$$

$$\propto e^{\frac{-1}{2\sigma^2}\sum (x_i - \theta)^2 - \frac{1}{2b^2}(\theta - a)^2}$$

($\because$ Ignoring constants as it can be obtained as a normalized const. later)

$$\propto e^{\frac{-1}{2\sigma^2}(\sum x_i^2 + n\theta^2 - 2\theta \sum x_i) - \frac{1}{2b^2}(\theta - a)^2}$$

$$\propto e^{-\frac{(\sum x_i^2 - 2\theta n \bar{x} + n\theta^2)}{2\sigma^2} - \frac{(\theta - a)^2}{2b^2}}$$

$\left(\bar{x} = \frac{1}{n}\sum x_i \quad se^2 = \sigma^2/n \right)$

Solving power term:

$$e^{-\left(\frac{\sum x_i^2 - (\bar{x})^2 + n\theta^2 - 2\theta n\bar{x} + \bar{x}^2}{2\sigma^2} \right) - \frac{(\theta - a)^2}{2b^2}}$$

$\frac{1}{n}\sum (x_i - \bar{x})^2 = se^2$

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

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

$se^2 = \sigma^2/n \Rightarrow$

$$-\frac{1}{2}\left(\frac{b^2nse^2 + b^2n\theta^2 - 2b^2\theta n\bar{x} + b^2\bar{x}^2) + (se^2)n(\theta^2 + a^2 - 2\theta a)}{(se^2)b^2n.}\right)$$

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

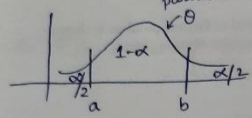
Removing $b^2se^2, b^2n\bar{x}^2, a^2se^2, \rightarrow$ constants (Merge into Normalizing const)
$b^2\bar{x}$.

$$= -\frac{1}{2}\left(\frac{\theta^2 - 2\theta\left(\frac{b^2\bar{x}+ase^2}{b^2+se^2}\right) + \left(\frac{b^2\bar{x}+ase^2}{b^2+se^2}\right)^2}{\frac{b^2se^2}{b^2+se^2}}\right)$$

$$= -\frac{1}{2}\left(\frac{\left(\theta - \frac{b^2\bar{x}+ase^2}{b^2+se^2}\right)^2}{\frac{b^2se^2}{b^2+se^2}}\right)$$

$\therefore$ It is equavalent to $\mathcal{N}\left(\underbrace{\frac{b^2\bar{x}+ase^2}{b^2+se^2}}_{x}, \underbrace{\frac{b^2se^2}{b^2+se^2}}_{y^2}\right)$

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



$$\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi y^2}} e^{-\frac{(\theta - x)^2}{2y^2}} = \alpha/2$$

$$\text{cdf}(P(\theta)) = Pr(\theta \le a)$$
$$= Pr(yZ + x \le a) \quad \text{where } Z \sim N(0,1)$$
$$= Pr(yZ \le a - x) = Pr\left(Z \le \frac{a-x}{y}\right)$$

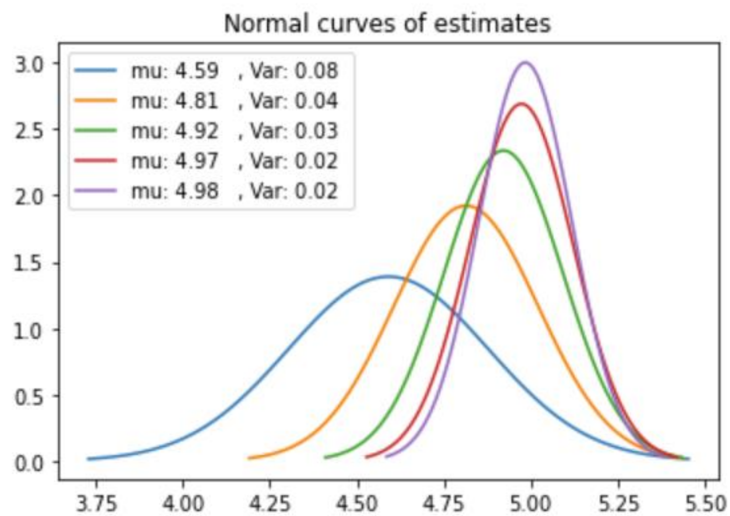
$$\therefore \quad \alpha/2 = \Phi\left(\frac{a-x}{y}\right)$$
$$a = x + y\,\Phi^{-1}(\alpha/2).$$

$$\int_{b}^{\infty}(\cdot) = \alpha/2 \Rightarrow Pr(\theta \ge b) = \alpha/2$$
$$1 - Pr(\theta \le b) = \alpha/2$$
$$1 - \alpha/2 = Pr(\theta \le b) = Pr(yZ + x \le b)$$

$$\Rightarrow b = x + y\,\Phi^{-1}(1 - \alpha/2).$$

2.a)

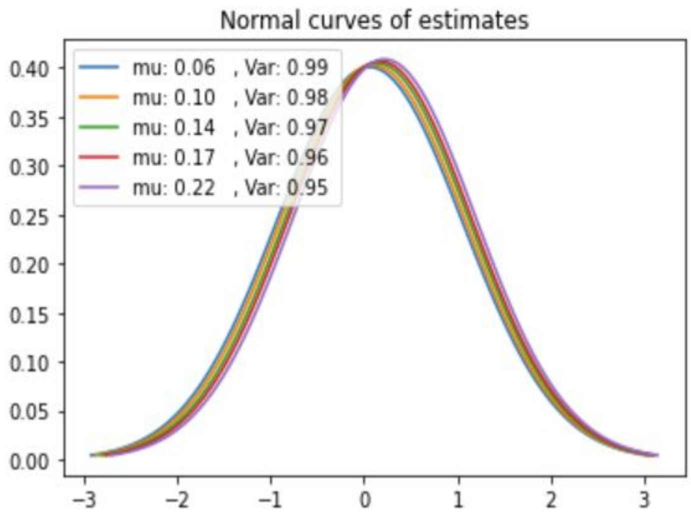
Mean	Variance
4.590762414332327	0.08256880733944953
4.813523613446215	0.0430622009569378
4.921256878168492	0.02912621359223301
4.97283741207765	0.022004889975550123
4.983966097849453	0.01768172888015717



With decreasing variance, the graph shifts towards 5.00 as mean.

2.b)

Mean	Variance
0.05871624147982311	0.9900990099009901
0.09500866961681816	0.9803921568627452
0.13822626152242073	0.970873786407767
0.17121883350740297	0.9615384615384617
0.2189182449674514	0.9523809523809524



With higher variance as assumption, the graphs converged to getting a correct data distribution.

2.c)

Since (b) has high variance, it is closer to prior. (a) has lower variance, it is closer to MLE.

3.a,b)

3.) a) Estimates of β:

$$\hat{\beta}_1 = \frac{\sum (x_i y_i) - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

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

$$= \sum x_i y_i - \bar{y}\sum x_i - \bar{x}\sum y_i + \bar{x}\bar{y}(n)$$

$$= \sum x_i y_i - 2\bar{x}\bar{y}(n) + \bar{x}\bar{y}(n) = \sum x_i y_i - n\bar{x}\bar{y}$$

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

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

$$= \sum x_i^2 + n\bar{x}^2 - 2n\bar{x}^2 = \sum x_i^2 - n\bar{x}^2$$

$$\hat{\beta}_1 = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

3.) b)
To show $E(\hat{\beta}_0) = \beta_0$ & $E(\hat{\beta}_1) = \beta_1$

$$E[\hat{\beta}_1 | x_i] = E\left[\frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2} \,\middle|\, x_i \right]$$

$$E[\hat{\beta}_1 | x_i] = E\left[\frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n(\bar{x})^2}\right]$$

[x's are constants]

$\hookrightarrow \sum x_i^2 \to$ const

$(\bar{x})^2 \to$ const

$\bar{y} \to$ const.

$$= \frac{E(\sum x_i y_i) - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$= \frac{\sum x_i E(y_i) - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \quad \text{—①}$$

$$E[\hat{\beta}_1] = \frac{\sum x_i E(y_i) - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \Rightarrow \cancel{\beta_0 +} \; y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$E(y_i) = \beta_0 + \beta_1 E(x_i) + E(\varepsilon_i)$$

$$E(y_i | x_i) = \beta_0 + \beta_1 x_i$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{\sum \beta_0 + \beta_1 x_i + \varepsilon_i}{n} = \frac{n\beta_0 + \beta_1 n\bar{x} + 0}{n} = \beta_0 + \beta_1 \bar{x}$$

From ①

$$= \frac{\sum x_i(\beta_0 + \beta_1 x_i) - n\bar{x}(\beta_0 + \beta_1 \bar{x})}{\sum x_i^2 - n\bar{x}^2} \qquad \sum x_i = n\bar{x}$$

$$= \frac{\beta_0 \sum x_i + \beta_1 \sum x_i^2 - n\bar{x}\beta_0 - n\beta_1 \bar{x}}{\sum x_i^2 - n\bar{x}^2} = \frac{\beta_1(\sum x_i^2 - n\bar{x})}{\sum x_i^2 - n\bar{x}} = \beta_1$$

$\therefore E(\hat{\beta}_1) = \beta_1$, $\hat{\beta}_1$ is unbiased.

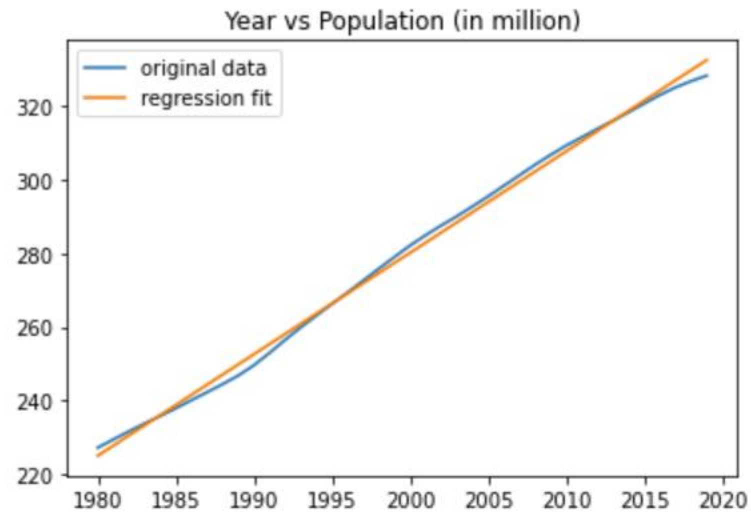
$$E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1 \bar{x}) = \bar{y} - \bar{x}E(\hat{\beta}_1) = \beta_0 + \beta_1 \bar{x} - \bar{x}\beta_1 = \beta_0.$$

$\therefore \hat{\beta}_0$ is unbiased.

4.a)

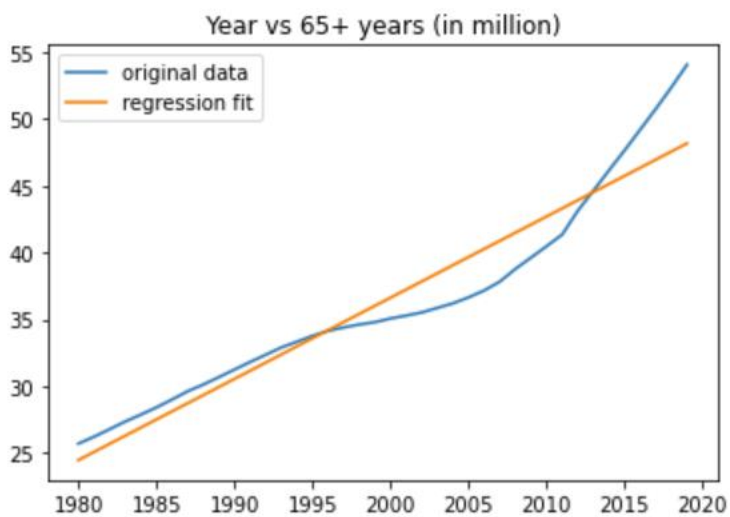
```
Regression line : y = -5234.856475245826 + 2.7575177477473503 * x
```

```
SSE : 123.8493735899118
```




```
Regression line : y = -1178.5613646415716 + 0.6075945051595757* x
```

```
SSE : 176.0353836252837
```



Since, 65+ years population data is not linear with Year data, it cannot be used for linear regression. But the Total population is linear with Year data, so it is suitable for linear regression.

4.b)

Data from 1980 – 2018:

Linear Regression Equation: y = -1131.0736031454455 + 0.5837632431173584 * x

```
Prediction result for 2060 (in million): 71.47867767631283

SSE : 137.69260032800727
```

Data from 2008 – 2018:

```
Linear Regression Equation:  y = -2778.38941090817 + 1.4025382727268152* x

Prediction result for 2060 (in million): 110.8394309090695

SSE : 2.008761070738738
```

We will be trusting the data from 2008- 2018 because it has low SSE. The media is right is saying so, because the predicted value(110.83 million) is double than 52 million.

4.c)

Predicting Ratio directly:

```
Ratio : 0.16236192897081914
```

Calculating Total population and 65+ year population and then calculating ratio:

```
Ratio: 0.16177484068121645

Predicted Total Population(in million): 329.6888533636047

Predicted 65+ year population( in million) : 53.33536172727008
```

The 1st method is more accurate because the ratio is linear with year data whereas 65+ year data is non-linear.

5.a)

```
SSE = 0.31641140910750554

Coefficients =  array([-0.00291067,  0.00323154,  0.01990962,  0.00057609,
0.02319267,0.1308982 ,  0.05682043]))
```

b)
```
SSE : 0.6403887599866744

Coefficients :  array([0.00388655, 0.04187385, 0.04825699]))
```

c)
```
SSE : 0.4638050678628288

Coefficients : array([-0.00410631,  0.23571159]))
```

d) We can see that using the whole data, we can predict accurately because of the low SSE value.

When we use TOEFL, SOP, LOR -> SSE increases indicating we have taken wrong features for prediction

Using GRE, GPA -> SSE is better than (b) , indicating these are better features for prediction.

6.a)

6)a.) $P(H_0) = p$, $P(H_1) = 1-p$

$f_W(w|H=0) = \mathcal{N}(w: -\mu, \sigma^2)$. $f_W(w|H=1) = \mathcal{N}(\mu, \sigma^2)$

Posterior: $P(H|W) \propto P(W|H) \cdot P(H)$

$$P(H=0|W) \propto P(W|H=0) \cdot P(H=0) \quad — ①$$

$$\propto P(\{w_1, \dots w_n\}|H=0) \cdot P(H=0) \propto \prod_{i=1}^{n} P(w_i|H=0) \cdot P(H=0)$$

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

$$\propto \exp\left(-\frac{\sum (w_i+\mu)^2}{2\sigma^2}\right) p$$

$$P(H=1|W) \propto \exp\left(-\frac{\sum (w_i-\mu)^2}{2\sigma^2}\right)(1-p) \quad — ②$$

$C=0$: ① $\geq$ ②

$$p \cdot e^{-\left(\frac{\sum (w_i+\mu)^2}{2\sigma^2}\right)} \geq (1-p)\, e^{-\left(\frac{\sum (w_i-\mu)^2}{2\sigma^2}\right)}$$

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

$$\left(e^{-\frac{4\mu \sum w_i}{2\sigma^2}}\right) \geq \frac{1-p}{p} \Rightarrow \frac{2\mu}{\sigma^2} \sum w_i \leq \ln \frac{p}{1-p}$$

$$\sum w_i \leq \frac{\sigma^2}{2\mu} \ln \frac{p}{1-p}$$

6.b)

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]

6.c)

6.) c) From 6.) a)

$$\sum w_i \le \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right)$$

$P(C=0 \mid H=1)$:

Since $(H=1)$, w follows $\mathcal{N}(\mu,\sigma^2)$

$\sum_i^n w_i$ follows $\mathcal{N}(n\mu, n\sigma^2) = (\sqrt{n}\sigma Z + n\mu)$

↑ Standard Normal

$$\Pr\left(\sum w_i \le \frac{\sigma^2}{2\mu}\ln\left(\frac{p}{1-p}\right)\right)$$

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

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

$P(C=1 \mid H=0)$:

Since $H=0$, w follows $\mathcal{N}(-\mu,\sigma^2)$

$\sum w_i$ follows $\mathcal{N}(-n\mu, n\sigma^2) = (\sqrt{n}\sigma Z - n\mu)$

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

$$= 1 - \Phi(\cdot)$$

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

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