Sec. 6.5 Ex 2

Let $Y_i = \beta x_i + \varepsilon_i$ where $\varepsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, \dots, n$ and $\varepsilon_1, \dots, \varepsilon_n$ are independent

This means that $Y_i \sim N(\beta x_i, \sigma^2)$.

a) Let us use the notation $\theta = \sigma^2$ in the following (the reason to do so is because we will be taking derivatives w.r.t σ^2 and so setting $\theta = \sigma^2$ will make it notationally convenient)

Define the likelihood function

$$L(\beta, \theta) = \prod_{i=1}^{n} f(y_i; \beta, \theta) = \prod_{i=1}^{n} \underbrace{\frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(y_i - \beta x_i)^2}{2\theta}}}_{\text{pdf of } N(\beta x_i, \theta)} =$$

$$= (2\pi\theta)^{-n/2} \cdot \exp\left(-\frac{1}{2\theta} \sum_{i=1}^{n} (y_i - \beta x_i)^2 \right)$$

Applying natural logarithm to $\ln(L(\beta,\sigma))$, we obtain:

$$\ln(L(\beta,\sigma)) = -\frac{m}{2}\ln(2\pi) - \frac{m}{2}\ln\sigma - \frac{1}{2\sigma}\sum_{i=1}^{m}(y_i-\beta x_i)^2$$

Let us now determine the first order conditions for $\ln(L(\beta,\sigma))$:

$$\begin{cases} \dfrac{\partial}{\partial\beta}\ln(L(\beta,\sigma)) = 0 \\[2ex] \dfrac{\partial}{\partial\sigma}\ln(L(\beta,\sigma)) = 0 \end{cases} \iff \begin{cases} +\dfrac{1}{\sigma}\displaystyle\sum_{i=1}^{m} x_i(y_i-\beta x_i) = 0 \\[2ex] -\dfrac{m}{2\sigma} + \dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{m}(y_i-\beta x_i)^2 = 0 \end{cases}$$

$$\iff \left.\begin{cases} \beta = \dfrac{\sum_{i=1}^{m} x_i y_i}{\sum_{i=1}^{m} x_i^2} \\[2ex] \sigma = \dfrac{1}{m}\displaystyle\sum_{i=1}^{m}(y_i-\hat{\beta}x_i)^2 \end{cases}\right\} \text{critical point of } \ln(L(\beta,\sigma)) \text{ and } L(\beta,\sigma)$$

To check that the critical pt just determined is indeed a maximizer of $\ln(L(\beta,\sigma))$ and $L(\beta,\sigma)$, we study the matrix of 2^{nd} derivatives of $\ln(L(\beta,\sigma))$:

$$\frac{\partial^2 \ln(L(\beta,\sigma))}{\partial \beta^2} = \frac{\partial}{\partial \beta}\left[\frac{1}{\sigma}\sum_{i=1}^{m} x_i(y_i - \beta x_i)\right] = -\frac{1}{\sigma}\sum_{i=1}^{m} x_i^2$$

$$\frac{\partial^2 \ln(L(\beta,\sigma))}{\partial \sigma\,\partial \beta} = \frac{\partial}{\partial \sigma}\left[\frac{1}{\sigma}\sum_{i=1}^{m} x_i(y_i - \beta x_i)\right] = -\frac{1}{\sigma^2}\sum_{i=1}^{m} x_i(y_i - \beta x_i)$$

$$\frac{\partial^2 \ln(L(\beta,\sigma))}{\partial \sigma^2} = \frac{\partial}{\partial \sigma}\left[-\frac{m}{2\sigma} + \frac{1}{2\sigma^2}\sum_{i=1}^{m}(y_i - \beta x_i)^2\right] = \frac{m}{2\sigma^2} - \frac{1}{\sigma^3}\sum_{i=1}^{m}(y_i - \beta x_i)^2$$

Let us now evaluate each of these 2^{nd} partial derivatives at the critical pt:

$$\hat{\beta} = \frac{\sum_{i=1}^{m} x_i y_i}{\sum_{i=1}^{m} x_i^2}, \quad \hat{\sigma} = \frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{\beta} x_i)^2$$

$$\left.\frac{\partial^2 \ln(L(\beta,\theta))}{\partial \beta^2}\right|_{\beta=\hat{\beta},\,\theta=\hat{\theta}} = -\frac{1}{\hat{\theta}} \sum_{i=1}^{m} x_i^2 < 0$$

} no simplification appears from replacing $\hat{\theta}$ by its expression

$$\left.\frac{\partial^2 \ln(L(\beta,\theta))}{\partial \theta\, \partial \beta}\right|_{\beta=\hat{\beta},\,\theta=\hat{\theta}} = -\frac{1}{\hat{\theta}^2} \sum_{i=1}^{m} x_i (y_i - \hat{\beta} x_i) = -\frac{1}{\hat{\theta}^2} \left(\sum_{i=1}^{m} x_i y_i - \hat{\beta} \cdot \sum_{i=1}^{m} x_i^2 \right)$$

$$= -\frac{1}{\hat{\theta}^2} \left(\sum_{i=1}^{m} x_i y_i - \underbrace{\frac{\sum_{i=1}^{m} x_i y_i}{\sum_{i=1}^{m} x_i^2}}_{\hat{\beta}} \cdot \left(\sum_{i=1}^{m} x_i^2\right) \right)$$

these two terms cancel

$$= -\frac{1}{\hat{\theta}^2} \left(\sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i y_i \right) = 0$$

$$\left.\frac{\partial^2 \ln(L(\beta,\theta))}{\partial \theta^2}\right|_{\beta=\hat\beta,\ \theta=\hat\theta} = \frac{m}{2\hat\theta^2} - \frac{1}{\hat\theta^3}\underbrace{\sum_{i=1}^{m}(y_i-\hat\beta x_i)^2}$$

since $\hat\theta = \frac{1}{m}\sum_{i=1}^{m}(y_i-\hat\beta x_i)^2$

then $\sum_{i=1}^{m}(y_i-\hat\beta x_i)^2 = m\hat\theta$

$$= \frac{m}{2\hat\theta^2} - \frac{1}{\hat\theta^3}\cdot m\hat\theta = \frac{m}{2\hat\theta^2} - \frac{m}{\hat\theta^2} = -\frac{m}{2\hat\theta^2} < 0$$

Hence, the matrix of 2nd derivatives of $\ln(L(\beta,\theta))$, when evaluated at the critical point $\hat\beta, \hat\theta$ is $\longrightarrow$

$$\begin{pmatrix} -\dfrac{\sum_{i=1}^{m} x_i^2}{\hat\theta} & 0 \\ 0 & -\dfrac{m}{2\hat\theta^2} \end{pmatrix}.$$

Since the matrix has two negative eigenvalues and the critical point is unique, then $L(\beta,\theta)$ is maximum at $\hat\beta, \hat\theta$.

Hence, we conclude that

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} y_i^2} \quad \text{and} \quad \widehat{\sigma^2} = \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta} x_i)^2$$

are the MLE of β and σ^2.

b) Let us start by finding the distribution of $\hat{\beta}$.

Since $Y_i \sim N(\beta x_i, \sigma^2)$, $i = 1, 2, \ldots, n$, are independent (this is because $Y_i = \beta x_i + \varepsilon_i$, $i = 1, 2, \ldots, n$ and $\varepsilon_1, \ldots, \varepsilon_n$ are independent), then

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$ is a linear combination of independent normal r.v.s and so, $\hat{\beta}$ is normally distributed.

We now observe that

$$E[\hat{\beta}] = E\left[\frac{\sum_{i=1}^m x_i Y_i}{\sum_{i=1}^m x_i^2}\right] \underset{\substack{\text{linearity}\\ \text{of expected value}}}{=} \frac{\sum_{i=1}^m x_i E[Y_i]}{\sum_{i=1}^m x_i^2} \underset{E[Y_i] = \beta x_i}{=} \frac{\sum_{i=1}^m x_i \beta x_i}{\sum_{i=1}^m x_i^2} = \beta \frac{\sum_{i=1}^m x_i^2}{\sum_{i=1}^m x_i^2} = \beta$$

and

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left(\frac{\sum_{i=1}^m x_i Y_i}{\sum_{i=1}^m x_i^2}\right) \underset{\text{independence of } Y_1, \dots, Y_m}{=} \frac{1}{\left(\sum_{i=1}^m x_i^2\right)^2} \sum_{i=1}^m \mathrm{Var}(x_i Y_i) = \frac{1}{\left(\sum_{i=1}^m x_i^2\right)^2} \sum_{i=1}^m x_i^2 \underbrace{\mathrm{Var}(Y_i)}_{\sigma^2} =$$

$$= \frac{\sum_{i=1}^m x_i^2 \sigma^2}{\left(\sum_{i=1}^m x_i^2\right)^2} = \sigma^2 \cdot \frac{\sum_{i=1}^m x_i^2}{\left(\sum_{i=1}^m x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^m x_i^2}.$$

We conclude that $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^m x_i^2}\right)$

We will now study the distribution of $\hat{\sigma}^2$:

Start by noting that since $Y_i \sim N(\beta x_i, \sigma^2)$, $i = 1, 2, \dots, m$, are independent then

$$\frac{Y_i - \beta x_i}{\sigma} \sim N(0,1) \Rightarrow \left(\frac{Y_i - \beta x_i}{\sigma}\right)^2 \sim \chi^2(1) \Rightarrow W = \sum_{i=1}^m \left(\frac{Y_i - \beta x_i}{\sigma}\right)^2 \sim \chi^2(m)$$

Observe also that:

$$\sum_{i=1}^{n} (Y_i - \beta x_i)^2 = \sum_{i=1}^{n} \Big(Y_i \overbrace{- \hat\beta x_i + \hat\beta x_i}^{=0} - \beta x_i \Big)^2 =$$

$$= \sum_{i=1}^{n} \Big[(Y_i - \hat\beta x_i) + (\hat\beta - \beta) x_i \Big]^2 =$$

$$= \sum_{i=1}^{n} (Y_i - \hat\beta x_i)^2 + 2(\hat\beta - \beta) \underbrace{\sum_{i=1}^{n} (Y_i - \hat\beta x_i) x_i}_{=0} + (\hat\beta - \beta)^2 \sum_{i=1}^{n} x_i^2$$

Moreover, the middle term in the expression above simplifies as follows:

$$\sum_{i=1}^{n} (Y_i - \hat\beta x_i) \cdot x_i = \sum_{i=1}^{n} Y_i x_i - \hat\beta \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} Y_i x_i - \frac{\sum_{i=1}^{n} Y_i x_i}{\sum_{i=1}^{n} x_i^2} \cdot \sum_{i=1}^{n} x_i^2$$

$$= \sum_{i=1}^{n} Y_i x_i - \sum_{i=1}^{n} Y_i x_i = 0$$

Hence, we obtain that

$$\sum_{i=1}^{m} (y_i - \beta x_i)^2 = \sum_{i=1}^{m} (y_i - \hat{\beta} x_i)^2 + (\hat{\beta} - \beta)^2 \sum_{i=1}^{m} x_i^2$$

Dividing by σ^2, yields:

$$\underbrace{\sum_{i=1}^{m} \left(\frac{y_i - \beta x_i}{\sigma} \right)^2}_{} = \underbrace{\sum_{i=1}^{m} \left(\frac{y_i - \hat{\beta} x_i}{\sigma} \right)^2}_{} + \underbrace{\left(\frac{\hat{\beta} - \beta}{\sigma / \sqrt{\sum_{i=1}^{m} x_i^2}} \right)^2}_{}$$

This is the random variable $W \sim \chi^2(m)$ introduced at the beginning of the exercise

This equals $\frac{m \hat{\sigma}^2}{\sigma^2}$, the random variable for which we are trying to find the distribution

We will denote this by $Z^2 \sim \chi^2(1)$ since $\frac{\hat{\beta} - \beta}{\sigma / \sqrt{\sum_{i=1}^{m} x_i^2}} \sim N(0,1)$ as we have proved that $\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^{m} x_i^2} \right)$

Hence, we obtain that

$$W = \frac{m\widehat{\sigma^2}}{\sigma^2} + Z^2 \quad , \quad \text{with} \quad W \sim \chi^2(m) \text{ and } Z^2 \sim \chi^2(1)$$

We will now determine the m.g.f of $\frac{m\widehat{\sigma^2}}{\sigma^2}$:

$$E\left[e^{tW}\right] = E\left[e^{t m\widehat{\sigma^2}/\sigma^2} \cdot e^{tZ^2}\right] = E\left[e^{t\frac{m\widehat{\sigma^2}}{\sigma^2}}\right] \cdot E\left[e^{tZ^2}\right]$$

independence of $\frac{m\widehat{\sigma^2}}{\sigma^2}$ and Z^2

Since $W \sim \chi^2(m)$, then $E\left[e^{tW}\right] = (1-2t)^{-m/2}$, for $t < 1/2$.

Similarly, since $Z^2 \sim \chi^2(1)$, then $E\left[e^{tZ^2}\right] = (1-2t)^{-1/2}$, for $t < 1/2$.

As a consequence, we obtain that:

$$(1-2t)^{-n/2} = E\left[e^{t\, n\widehat{\sigma^2}/\sigma^2}\right] \cdot (1-2t)^{-1/2}$$

and so

$$E\left[e^{t\, n\widehat{\sigma^2}/\sigma^2}\right] = (1-2t)^{-n/2} \cdot (1-2t)^{1/2} = (1-2t)^{-(n-1)/2}, \quad t < 1/2$$

Hence, the m.g.f of $\frac{n\sigma}{\sigma^2}$ is that of a $\chi^2(n-1)$ distribution and we conclude that

$$\boxed{\frac{n\widehat{\sigma^2}}{\sigma^2} \sim \chi^2(n-1)}$$

This is the relevant characterization for the distribution of $\hat{\sigma}^2$

If we really want the distribution of $\widehat{\sigma^2}$ (and not just that of $\frac{m\widehat{\sigma^2}}{\sigma^2}$), we may proceed as follows.

Start by finding the cdf of $\widehat{\sigma^2}$: for $y > 0$, we have

$$G(y) = P(\widehat{\sigma^2} \leq y) = P\Big(\underbrace{\frac{m\widehat{\sigma^2}}{\sigma^2}}_{\chi^2(m-1)} \leq \frac{my}{\sigma^2}\Big) = \int_0^{\frac{my}{\sigma^2}} \underbrace{\frac{x^{\frac{m-1}{2}-1} e^{-x/2}}{\Gamma\left(\frac{m-1}{2}\right) \cdot 2^{(m-1)/2}}}_{\text{pdf of } \chi^2(m-1) \text{ or Gamma with } \alpha = \frac{m-1}{2} \text{ and } \theta = 2} dx$$

Using the fundamental theorem of Calculus, we get that for $y > 0$:

$$g(y) = G'(y) = \frac{\left(\frac{my}{\sigma^2}\right)^{\frac{m-1}{2}-1} e^{-my/2\sigma^2}}{\Gamma\left(\frac{m-1}{2}\right) 2^{(m-1)/2}} \cdot \frac{m}{\sigma^2} = \underbrace{\frac{y^{\frac{m-1}{2}-1} e^{-y/(2\sigma^2/m)}}{\Gamma\left(\frac{m-1}{2}\right)\left(\frac{2\sigma^2}{m}\right)^{\frac{m-1}{2}}}}_{\text{pdf of Gamma with } \alpha = \frac{m-1}{2} \text{ and } \theta = \frac{2\sigma^2}{m}}, \text{ for } y > 0$$

Hence, we conclude that $\widehat{\sigma^2}$ follows a gamma distribution with parameters $\alpha = \frac{n-1}{2}$

and $\theta = \frac{2\sigma^2}{n}$.

$\frac{2}{25}$ $\frac{8}{25}$

c) The estimate for β is:

$$\hat{\beta} = \frac{\sum_{i=1}^{2} x_i y_i}{\sum_{i=1}^{2} x_i^2} = \frac{1.2 + 1.1}{1^2 + 2^2} = \frac{3}{5}$$

The regression line is then $y = \hat{\beta} x$, that is $y = \frac{3}{5} x$

Hence, for $x_1 = 1$, we have $\hat{Y}_1 = \frac{3}{5} x_1 = \frac{3}{5}$

and for $x_2 = 2$, we have $\hat{Y}_2 = \frac{3}{5} x_2 = \frac{6}{5}$.

Then $$\hat{\sigma}^2 = \frac{1}{2} \sum_{i=1}^{2} [y_i - \beta x_i]^2 = \frac{1}{2}(y_1 - \hat{Y}_1)^2 + \frac{1}{2}(y_2 - \hat{Y}_2)^2 = \frac{1}{2}\left(1 - \frac{3}{5}\right)^2 + \frac{1}{2}\left(2 - \frac{6}{5}\right)^2$$
$$= \frac{2}{5}$$

d)

$$\sum_{i=1}^{2} (y_i - \hat{y}_i) = \left(1 - \frac{3}{5}\right) + \left(1 - \frac{6}{5}\right) = 2 - \frac{9}{5} = \frac{1}{5} \neq 0$$