

Multivariate Parameter Estimation MLE

Matthew Seguin

1.

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ and τ be such that $\mathbb{P}[X < \tau] = 0.9$.

First we will calculate the MLE for $\theta = (\mu, \sigma^2)$:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}} \implies l(\theta) = \log L(\theta) = \sum_{i=1}^n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(x_i - \mu)^2}{2\sigma^2} \\ &= n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{n}{2} \log(\sigma^2) - \sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^2} \\ \nabla l(\theta) &= \left(\frac{\partial}{\partial \mu} l(\theta), \frac{\partial}{\partial \sigma^2} l(\theta) \right) = \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n \frac{\partial}{\partial \mu} (x_i - \mu)^2 \right), -\frac{\partial}{\partial \sigma^2} \frac{n}{2} \log(\sigma^2) - \sum_{i=1}^n \frac{\partial}{\partial \sigma^2} \frac{(x_i - \mu)^2}{2\sigma^2} \right) \\ &= \left(\frac{1}{\sigma^2} \left(\left(\sum_{i=1}^n x_i \right) - n\mu \right), \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^4} \right) - \frac{n}{2\sigma^2} \right) \end{aligned}$$

Setting this equal to $(0, 0)$ we get the simultaneous equations:

$$\begin{aligned} \frac{1}{\sigma^2} \left(\left(\sum_{i=1}^n x_i \right) - n\mu \right) &= 0 & \left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^4} \right) - \frac{n}{2\sigma^2} &= 0 \\ \sum_{i=1}^n x_i &= n\mu & \sum_{i=1}^n (x_i - \mu)^2 &= n\sigma^2 \\ \mu &= \frac{1}{n} \sum_{i=1}^n x_i & \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

So our critical point is:

$$\left(\hat{\mu}, \hat{\sigma}^2 \right) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{i=1}^n x_i \right)^2 \right)$$

Continued on next page.

To show this is indeed a maximum we take the determinant of the Hessian matrix:

$$\begin{aligned} \det H_l(\mu, \sigma^2) &= \begin{vmatrix} \frac{\partial^2}{\partial \mu^2} l(\theta) & \frac{\partial^2}{\partial \mu \partial \sigma^2} l(\theta) \\ \frac{\partial^2}{\partial \sigma^2 \partial \mu} l(\theta) & \frac{\partial^2}{\partial (\sigma^2)^2} l(\theta) \end{vmatrix} = \begin{vmatrix} \frac{\partial}{\partial \mu} \frac{1}{\sigma^2} \left(\left(\sum_{i=1}^n x_i \right) - n\mu \right) & \frac{\partial}{\partial \sigma^2} \frac{1}{\sigma^2} \left(\left(\sum_{i=1}^n x_i \right) - n\mu \right) \\ \frac{\partial}{\partial \mu} \left(\left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^4} \right) - \frac{n}{2\sigma^2} \right) & \frac{\partial}{\partial \sigma^2} \left(\left(\sum_{i=1}^n \frac{(x_i - \mu)^2}{2\sigma^4} \right) - \frac{n}{2\sigma^2} \right) \end{vmatrix} \\ &= \begin{vmatrix} -\frac{n}{\sigma^2} & \frac{1}{\sigma^4} \left(n\mu - \sum_{i=1}^n x_i \right) \\ \frac{1}{\sigma^4} \left(n\mu - \sum_{i=1}^n x_i \right) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \end{vmatrix} = \left(-\frac{n}{\sigma^2} \right) \left(\frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \right) - \left(\frac{1}{\sigma^4} \left(n\mu - \sum_{i=1}^n x_i \right) \right)^2 \end{aligned}$$

Then we evaluate this at our critical point:

$$(\hat{\mu}, \hat{\sigma}^2) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n \left(x_i - \frac{1}{n} \sum_{i=1}^n x_i \right)^2 \right) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \right)$$

$$\begin{aligned} \det H_l(\hat{\mu}, \hat{\sigma}^2) &= \left(-\frac{n}{\hat{\sigma}^2} \right) \left(\frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (x_i - \hat{\mu})^2 \right) - \left(\frac{1}{(\hat{\sigma}^2)^2} \left(n\hat{\mu} - \sum_{i=1}^n x_i \right) \right)^2 \\ &= \left(-\frac{n}{\hat{\sigma}^2} \right) \left(\frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (x_i - \hat{\mu})^2 \right) = \left(-\frac{n}{\hat{\sigma}^2} \right) \left(\frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} n\hat{\sigma}^2 \right) = \left(-\frac{n}{\hat{\sigma}^2} \right) \left(\frac{n}{2(\hat{\sigma}^2)^2} - \frac{n}{(\hat{\sigma}^2)^2} \right) = \frac{n^2}{2(\hat{\sigma}^2)^3} > 0 \end{aligned}$$

So the product of the eigenvalues of the Hessian matrix is positive, and since $\frac{\partial^2}{\partial \mu^2} l(\theta) = -\frac{n}{\sigma^2} < 0$ this means that both of the eigenvalues are negative and hence the Hessian matrix is negative definite. Showing that $l(\theta)$ at our critical point is indeed a maximum point.

So the MLE is indeed:

$$\hat{\theta}_n = \left(\hat{\mu}_n, \hat{\sigma}_n^2 \right) = \left(\frac{1}{n} \sum_{i=1}^n X_i, \frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} \sum_{i=1}^n X_i \right)^2 \right)$$

Which implies that the MLE of (μ, σ) is:

$$\left(\hat{\mu}_n, \hat{\sigma}_n \right) = \left(\frac{1}{n} \sum_{i=1}^n X_i, \sqrt{\frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} \sum_{i=1}^n X_i \right)^2} \right)$$

Now continuing with τ :

$$0.9 = \mathbb{P}[X < \tau] = \mathbb{P}\left[\frac{X - \mu}{\sigma} < \frac{\tau - \mu}{\sigma}\right] = \Phi\left(\frac{\tau - \mu}{\sigma}\right)$$

$$\frac{\tau - \mu}{\sigma} = \Phi^{-1}(0.9) \implies \tau = \mu + \sigma \Phi^{-1}(0.9)$$

Then the MLE of τ is:

$$\hat{\tau}_n = \hat{\mu}_n + \hat{\sigma}_n \Phi^{-1}(0.9) = \left(\frac{1}{n} \sum_{i=1}^n X_i \right) + \Phi^{-1}(0.9) \sqrt{\frac{1}{n} \sum_{i=1}^n \left(X_i - \frac{1}{n} \sum_{i=1}^n X_i \right)^2} \quad \square$$

Where Φ is the standard normal CDF and Φ^{-1} is its inverse.

2.

Let $Y_i \stackrel{\text{ind.}}{\sim} N(\beta_0 + \beta_1 x_i, \sigma^2)$ for each $i \in \{1, 2, \dots, n\}$.

a.

We are going to find the MLE of $\theta = (\beta_0, \beta_1, \sigma^2)$:

$$\begin{aligned}
 L(\theta) &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}} \implies l(\theta) = \log L(\theta) = \sum_{i=1}^n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \\
 &= n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{n}{2} \log(\sigma^2) - \sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \\
 \nabla l(\theta) &= \left(\frac{\partial}{\partial \beta_0} l(\theta), \frac{\partial}{\partial \beta_1} l(\theta), \frac{\partial}{\partial \sigma^2} l(\theta) \right) \\
 &= \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n \frac{\partial}{\partial \beta_0} (y_i - \beta_0 - \beta_1 x_i)^2 \right), -\frac{1}{2\sigma^2} \left(\sum_{i=1}^n \frac{\partial}{\partial \beta_1} (y_i - \beta_0 - \beta_1 x_i)^2 \right), -\frac{\partial}{\partial \sigma^2} \frac{n}{2} \log(\sigma^2) - \sum_{i=1}^n \frac{\partial}{\partial \sigma^2} \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right) \\
 &= \left(\frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i), \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i), \left(\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} \right) - \frac{n}{2\sigma^2} \right)
 \end{aligned}$$

Setting this equal to $(0, 0, 0)$ we get the simultaneous equations:

$$\begin{aligned}
 \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) &= 0 & \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) &= 0 & \left(\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} \right) - \frac{n}{2\sigma^2} &= 0 \\
 \left(\sum_{i=1}^n y_i \right) - n\beta_0 - \beta_1 \left(\sum_{i=1}^n x_i \right) &= 0 & \left(\sum_{i=1}^n x_i y_i \right) - \beta_0 \left(\sum_{i=1}^n x_i \right) - \beta_1 \left(\sum_{i=1}^n x_i^2 \right) &= 0 & \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 &= n\sigma^2 \\
 \beta_0 &= \left(\frac{1}{n} \sum_{i=1}^n y_i \right) - \beta_1 \left(\frac{1}{n} \sum_{i=1}^n x_i \right) & \beta_1 &= \frac{\left(\sum_{i=1}^n x_i y_i \right) - \beta_0 \left(\sum_{i=1}^n x_i \right)}{\left(\sum_{i=1}^n x_i^2 \right)} & \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2
 \end{aligned}$$

Plugging equation 1 into equation 2 we get:

$$\begin{aligned}
 \beta_1 &= \frac{\left(\sum_{i=1}^n x_i y_i \right) - \left(\left(\frac{1}{n} \sum_{i=1}^n y_i \right) - \beta_1 \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \right) \left(\sum_{i=1}^n x_i \right)}{\left(\sum_{i=1}^n x_i^2 \right)} \\
 &= \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \left(\left(\frac{1}{n} \sum_{i=1}^n y_i \right) - \beta_1 \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \right) \left(\frac{1}{n} \sum_{i=1}^n x_i \right)}{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)} \\
 \beta_1 \left(1 - \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2}{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)} \right) &= \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i y_i \right) - \left(\frac{1}{n} \sum_{i=1}^n y_i \right) \left(\frac{1}{n} \sum_{i=1}^n x_i \right)}{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right)}
 \end{aligned}$$

Continued on next page.

$$\beta_1 = \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i y_i\right) - \left(\frac{1}{n} \sum_{i=1}^n y_i\right) \left(\frac{1}{n} \sum_{i=1}^n x_i\right)}{\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - \left(\frac{1}{n} \sum_{i=1}^n x_i\right)^2}$$

Then plugging this into equation 1 we get:

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

Then plugging these in for σ^2 we get:

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

From here on out for simplicity I will use the following abbreviations of terms:

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i \quad s_{wz} = s_{zw} = \left(\frac{1}{n} \sum_{i=1}^n z_i w_i\right) - \left(\frac{1}{n} \sum_{i=1}^n w_i\right) \left(\frac{1}{n} \sum_{i=1}^n z_i\right) = \bar{z}\bar{w} - \bar{w}\bar{z}$$

Where z and w can be appropriately replaced with other variables as needed.

So our critical point is:

$$\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2\right) = \left(\bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}, \quad \frac{s_{xy}}{s_{xx}}, \quad \frac{1}{n} \sum_{i=1}^n \left((y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}} (x_i - \bar{x})\right)^2\right)$$

Continued on next page.

To show this is indeed a maximum we examine the Hessian matrix:

$$\begin{aligned}
H_l(\beta_0, \beta_1, \sigma^2) &= \begin{bmatrix} \frac{\partial^2}{\partial \beta_0^2} l(\theta) & \frac{\partial^2}{\partial \beta_0 \partial \beta_1} l(\theta) & \frac{\partial^2}{\partial \beta_0 \partial \sigma^2} l(\theta) \\ \frac{\partial^2}{\partial \beta_1 \partial \beta_0} l(\theta) & \frac{\partial^2}{\partial \beta_1^2} l(\theta) & \frac{\partial^2}{\partial \beta_1 \partial \sigma^2} l(\theta) \\ \frac{\partial^2}{\partial \sigma^2 \partial \beta_0} l(\theta) & \frac{\partial^2}{\partial \sigma^2 \partial \beta_1} l(\theta) & \frac{\partial^2}{\partial (\sigma^2)^2} l(\theta) \end{bmatrix} \\
&= \begin{bmatrix} \frac{\partial}{\partial \beta_0} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) & \frac{\partial}{\partial \beta_1} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) & \frac{\partial}{\partial \sigma^2} \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \\ \frac{\partial}{\partial \beta_0} \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) & \frac{\partial}{\partial \beta_1} \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) & \frac{\partial}{\partial \sigma^2} \frac{1}{\sigma^2} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) \\ \frac{\partial}{\partial \beta_0} \left(\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} \right) & \frac{\partial}{\partial \beta_1} \left(\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} \right) & \frac{\partial}{\partial \sigma^2} \left(\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^4} \right) - \frac{\partial}{\partial \sigma^2} \frac{n}{2\sigma^2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2} \bar{x} & -\frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) \\ -\frac{n}{\sigma^2} \bar{x} & -\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 & -\frac{1}{\sigma^4} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) \\ -\frac{1}{\sigma^4} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i) & -\frac{1}{\sigma^4} \sum_{i=1}^n x_i (y_i - \beta_0 - \beta_1 x_i) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2} \bar{x} & -\frac{n}{\sigma^4} (\bar{y} - \beta_0 - \beta_1 \bar{x}) \\ -\frac{n}{\sigma^2} \bar{x} & -\frac{n}{\sigma^2} (s_{xx} + \bar{x}^2) & -\frac{n}{\sigma^4} (\bar{x}\bar{y} - \beta_0 \bar{x} - \beta_1 (s_{xx} + \bar{x}^2)) \\ -\frac{n}{\sigma^4} (\bar{y} - \beta_0 - \beta_1 \bar{x}) & -\frac{n}{\sigma^4} (\bar{x}\bar{y} - \beta_0 \bar{x} - \beta_1 (s_{xx} + \bar{x}^2)) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{bmatrix}
\end{aligned}$$

Then we evaluate this at our critical point:

$$\begin{aligned}
(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) &= \left(\bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}, \frac{s_{xy}}{s_{xx}}, \frac{1}{n} \sum_{i=1}^n \left((y_i - \bar{y}) - \frac{s_{xy}}{s_{xx}} (x_i - \bar{x}) \right)^2 \right) = \left(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1, \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \right) \\
H_l(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) &= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{n}{\hat{\sigma}^2} \bar{x} & -\frac{n}{(\hat{\sigma}^2)^2} (\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}) \\ -\frac{n}{\hat{\sigma}^2} \bar{x} & -\frac{n}{\hat{\sigma}^2} (s_{xx} + \bar{x}^2) & -\frac{n}{(\hat{\sigma}^2)^2} (\bar{x}\bar{y} - \hat{\beta}_0 \bar{x} - \hat{\beta}_1 (s_{xx} + \bar{x}^2)) \\ -\frac{n}{(\hat{\sigma}^2)^2} (\bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}) & -\frac{n}{(\hat{\sigma}^2)^2} (\bar{x}\bar{y} - \hat{\beta}_0 \bar{x} - \hat{\beta}_1 (s_{xx} + \bar{x}^2)) & \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{n}{\hat{\sigma}^2} \bar{x} & 0 \\ -\frac{n}{\hat{\sigma}^2} \bar{x} & -\frac{n}{\hat{\sigma}^2} (s_{xx} + \bar{x}^2) & -\frac{n}{(\hat{\sigma}^2)^2} (\bar{x}\bar{y} - \bar{x}\bar{y} + \hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 (s_{xx} + \bar{x}^2)) \\ 0 & -\frac{n}{(\hat{\sigma}^2)^2} (\bar{x}\bar{y} - \bar{x}\bar{y} + \hat{\beta}_1 \bar{x}^2 - \hat{\beta}_1 (s_{xx} + \bar{x}^2)) & \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{n}{\hat{\sigma}^2} \bar{x} & 0 \\ -\frac{n}{\hat{\sigma}^2} \bar{x} & -\frac{n}{\hat{\sigma}^2} (s_{xx} + \bar{x}^2) & -\frac{n}{(\hat{\sigma}^2)^2} (s_{xy} - s_{xx} \frac{s_{xy}}{s_{xx}}) \\ 0 & -\frac{n}{(\hat{\sigma}^2)^2} (s_{xy} - s_{xx} \frac{s_{xy}}{s_{xx}}) & \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{bmatrix} \\
&= \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & -\frac{n}{\hat{\sigma}^2} \bar{x} & 0 \\ -\frac{n}{\hat{\sigma}^2} \bar{x} & -\frac{n}{\hat{\sigma}^2} (s_{xx} + \bar{x}^2) & 0 \\ 0 & 0 & \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{bmatrix}
\end{aligned}$$

Continued on next page.

Now we need to look at the principle sub-matrices of this matrix.

$$\text{First: } \mathbf{det}A_1 = -\frac{n}{\sigma^2} < 0$$

Then:

$$\begin{aligned} \mathbf{det}A_2 &= \begin{vmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2}\bar{x} \\ -\frac{n}{\sigma^2}\bar{x} & -\frac{n}{\sigma^2}(s_{xx} + \bar{x}^2) \end{vmatrix} = \frac{n^2}{(\hat{\sigma}^2)^2} (s_{xx} + \bar{x}^2) - \frac{n^2}{(\hat{\sigma}^2)^2} \bar{x}^2 = \left(\frac{n}{(\hat{\sigma}^2)^2}\right) \left(\left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) - \bar{x}^2\right) \\ &= \left(\frac{n^2}{(\hat{\sigma}^2)^2}\right) \left(\frac{1}{n} \sum_{i=1}^n (x_i^2 - 2\bar{x}^2 + \bar{x}^2)\right) = \left(\frac{n}{(\hat{\sigma}^2)^2}\right) \left(\sum_{i=1}^n (x_i - \bar{x})^2\right) > 0 \end{aligned}$$

Finally:

$$\begin{aligned} \mathbf{det}A_3 &= \mathbf{det}H_l(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) = \begin{vmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2}\bar{x} & 0 \\ -\frac{n}{\sigma^2}\bar{x} & -\frac{n}{\sigma^2}(s_{xx} + \bar{x}^2) & 0 \\ 0 & 0 & \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{vmatrix} \\ &= - \begin{vmatrix} 0 & 0 & \frac{n}{2(\hat{\sigma}^2)^2} - \frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \\ -\frac{n}{\sigma^2}\bar{x} & -\frac{n}{\sigma^2}(s_{xx} + \bar{x}^2) & 0 \\ -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2}\bar{x} & 0 \end{vmatrix} \\ &= \left(\left(\frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \right) - \frac{n}{2(\hat{\sigma}^2)^2} \right) \begin{vmatrix} -\frac{n}{\sigma^2}\bar{x} & -\frac{n}{\sigma^2}(s_{xx} + \bar{x}^2) \\ -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2}\bar{x} \end{vmatrix} \\ &= \left(\frac{n}{2(\hat{\sigma}^2)^2} - \left(\frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \right) \right) \begin{vmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2}\bar{x} \\ -\frac{n}{\sigma^2}\bar{x} & -\frac{n}{\sigma^2}(s_{xx} + \bar{x}^2) \end{vmatrix} \end{aligned}$$

We already saw the determinant of that 2x2 matrix was positive so $\mathbf{det}A_3 = \mathbf{det}H_l(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2) < 0$ if and only if:

$$\frac{n}{2(\hat{\sigma}^2)^2} - \left(\frac{1}{(\hat{\sigma}^2)^3} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \right) < 0 \text{ i.e. } \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 > \frac{n\hat{\sigma}^2}{2} \text{ i.e. } \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = n\hat{\sigma}^2 > \frac{n\hat{\sigma}^2}{2}$$

Which is true so we have shown that $\mathbf{det}A_1 < 0$, $\mathbf{det}A_2 > 0$, and $\mathbf{det}A_3 < 0$ and therefore we know $H_l(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}^2)$ is negative definite. Showing that $l(\theta)$ at our critical point is indeed a maximum point.

So the MLE is indeed:

$$\hat{\theta}_n = \left(\hat{\beta}_{0n}, \hat{\beta}_{1n}, \hat{\sigma}_n^2 \right) = \left(\bar{Y} - \frac{s_{xY}}{s_{xx}} \bar{x}, \frac{s_{xY}}{s_{xx}}, \frac{1}{n} \sum_{i=1}^n \left((Y_i - \bar{Y}) - \frac{s_{xY}}{s_{xx}} (x_i - \bar{x}) \right)^2 \right)$$

Where (with z and w appropriately replaced with other variables as needed):

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i \quad s_{wz} = s_{zw} = \left(\frac{1}{n} \sum_{i=1}^n z_i w_i \right) - \left(\frac{1}{n} \sum_{i=1}^n w_i \right) \left(\frac{1}{n} \sum_{i=1}^n z_i \right) = \bar{zw} - \bar{w}\bar{z}$$

b.

From before we know the Hessian matrix was:

$$H_l(\beta_0, \beta_1, \sigma^2) = \begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{n}{\sigma^2} \bar{x} & -\frac{n}{\sigma^4} (\bar{y} - \beta_0 - \beta_1 \bar{x}) \\ -\frac{n}{\sigma^2} \bar{x} & -\frac{n}{\sigma^2} (s_{xx} + \bar{x}^2) & -\frac{n}{\sigma^4} (\overline{xy} - \beta_0 \bar{x} - \beta_1 (s_{xx} + \bar{x}^2)) \\ -\frac{n}{\sigma^4} (\bar{y} - \beta_0 - \beta_1 \bar{x}) & -\frac{n}{\sigma^4} (\overline{xy} - \beta_0 \bar{x} - \beta_1 (s_{xx} + \bar{x}^2)) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 \end{bmatrix}$$

Where again we have the following:

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i \quad s_{wz} = s_{zw} = \left(\frac{1}{n} \sum_{i=1}^n z_i w_i \right) - \left(\frac{1}{n} \sum_{i=1}^n w_i \right) \left(\frac{1}{n} \sum_{i=1}^n z_i \right) = \overline{zw} - \bar{w} \bar{z}$$

And z and w can be appropriately replaced with other variables as needed.

First note that:

$$\mathbb{E}[\bar{x}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[x_i] = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

Quickly note $s_{xx} + \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2$ where $\mathbb{E}[s_{xx} + \bar{x}^2] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n x_i^2\right] = \frac{1}{n} \sum_{i=1}^n x_i^2 = \overline{x^2}$

Also note that $\mathbb{E}[\bar{y}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[y_i] = \frac{1}{n} \sum_{i=1}^n \beta_0 + \beta_1 x_i = \beta_0 + \beta_1 \bar{x}$

Finally note that $\mathbb{E}[\overline{xy}] = \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n x_i y_i\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[x_i y_i] = \frac{1}{n} \sum_{i=1}^n \beta_0 x_i + \beta_1 x_i^2 = \beta_0 \bar{x} + \beta_1 \overline{x^2}$

Therefore:

$$\begin{aligned} I_n(\theta) &= -\mathbb{E}[H_l(\beta_0, \beta_1, \sigma^2)] = \\ &= \begin{bmatrix} \mathbb{E}\left[\frac{n}{\sigma^2}\right] & \mathbb{E}\left[\frac{n}{\sigma^2} \bar{x}\right] & \mathbb{E}\left[\frac{n}{\sigma^4} (\bar{y} - \beta_0 - \beta_1 \bar{x})\right] \\ \mathbb{E}\left[\frac{n}{\sigma^2} \bar{x}\right] & \mathbb{E}\left[\frac{n}{\sigma^2} (s_{xx} + \bar{x}^2)\right] & \mathbb{E}\left[\frac{n}{\sigma^4} (\overline{xy} - \beta_0 \bar{x} - \beta_1 (s_{xx} + \bar{x}^2))\right] \\ \mathbb{E}\left[\frac{n}{\sigma^4} (\bar{y} - \beta_0 - \beta_1 \bar{x})\right] & \mathbb{E}\left[\frac{n}{\sigma^4} (\overline{xy} - \beta_0 \bar{x} - \beta_1 (s_{xx} + \bar{x}^2))\right] & \mathbb{E}\left[\frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2 - \frac{n}{2\sigma^4}\right] \end{bmatrix} \\ &= \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \bar{x} & 0 \\ \frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \overline{x^2} & 0 \\ 0 & 0 & \frac{n}{\sigma^6} \mathbb{V}[Y] - \frac{n}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \bar{x} & 0 \\ \frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \overline{x^2} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix} \end{aligned}$$

So we have our final answer:

$$I_n(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \bar{x} & 0 \\ \frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \overline{x^2} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix}$$

c.

We saw before that:

$$I_n(\theta) = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \bar{x} & 0 \\ \frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \bar{x}^2 & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

Where:

$$A = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n}{\sigma^2} \bar{x} \\ \frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \bar{x}^2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} \frac{n}{2\sigma^4} \end{bmatrix}$$

So to find the variance matrix we invert it:

$$V_n(\theta) = (I_n(\theta))^{-1} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}$$

First inverting B is easy since:

$$B^{-1} = \begin{bmatrix} \frac{2\sigma^4}{n} \end{bmatrix} \quad \text{results in} \quad B^{-1}B = \begin{bmatrix} \frac{2\sigma^4}{n} \end{bmatrix} \begin{bmatrix} \frac{n}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} \frac{n}{2\sigma^4} \end{bmatrix} \begin{bmatrix} \frac{2\sigma^4}{n} \end{bmatrix} = BB^{-1}$$

Then inverting A :

$$A^{-1} = \frac{1}{\left(\frac{n}{\sigma^2}\right)\left(\frac{n}{\sigma^2} \bar{x}^2\right) - \left(\frac{n}{\sigma^2} \bar{x}\right)^2} \begin{bmatrix} \frac{n}{\sigma^2} \bar{x}^2 & -\frac{n}{\sigma^2} \bar{x} \\ -\frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \end{bmatrix} = \frac{\sigma^4}{n^2 s_{xx}} \begin{bmatrix} \frac{n}{\sigma^2} \bar{x}^2 & -\frac{n}{\sigma^2} \bar{x} \\ -\frac{n}{\sigma^2} \bar{x} & \frac{n}{\sigma^2} \end{bmatrix} = \begin{bmatrix} \left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}^2}{s_{xx}}\right) & -\left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}}{s_{xx}}\right) \\ -\left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}}{s_{xx}}\right) & \left(\frac{\sigma^2}{n}\right)\left(\frac{1}{s_{xx}}\right) \end{bmatrix}$$

Therefore we have the variance matrix:

$$V_n(\theta) = (I_n(\theta))^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix} = \begin{bmatrix} \left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}^2}{s_{xx}}\right) & -\left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}}{s_{xx}}\right) & 0 \\ -\left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}}{s_{xx}}\right) & \left(\frac{\sigma^2}{n}\right)\left(\frac{1}{s_{xx}}\right) & 0 \\ 0 & 0 & \frac{2\sigma^4}{n} \end{bmatrix}$$

Then by the properties of the MLE we know $\hat{\theta}_n$ is asymptotically normal with mean θ and variance matrix $V_n(\theta)$.

Which implies that $\hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{\sigma^2}{n}\right)\left(\frac{\bar{x}^2}{s_{xx}}\right)\right)$ and $\hat{\beta}_1 \sim N\left(\beta_1, \left(\frac{\sigma^2}{n}\right)\left(\frac{1}{s_{xx}}\right)\right)$.

Then we can plug in the estimates of our unknown variables to get the $1 - \alpha$ confidence intervals:

$$\beta_0 : \left(\hat{\beta}_0 - z_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2 \bar{x}^2}{n s_{xx}}}, \hat{\beta}_0 + z_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2 \bar{x}^2}{n s_{xx}}} \right) \quad \text{and} \quad \beta_1 : \left(\hat{\beta}_1 - z_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n s_{xx}}}, \hat{\beta}_1 + z_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2}{n s_{xx}}} \right)$$

3.

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$, i.e. $f_{X_i}(x_i) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}}$ for $0 \leq x_i < \infty$.

a.

Here we assume α is known and find the MLE for β :

$$L(\beta) = \prod_{i=1}^n \frac{1}{\beta^\alpha \Gamma(\alpha)} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}} \implies l(\beta) = \log L(\beta) = \sum_{i=1}^n \alpha \log \frac{1}{\beta} - \log \Gamma(\alpha) + (\alpha - 1) \log x_i - \frac{x_i}{\beta}$$

Now I am actually going to find the MLE with respect to $\frac{1}{\beta}$

Taking the derivative with respect to $\frac{1}{\beta}$:

$$\begin{aligned} \frac{d}{d\frac{1}{\beta}} l(\beta) &= \frac{d}{d\frac{1}{\beta}} \sum_{i=1}^n \alpha \log \frac{1}{\beta} - \log \Gamma(\alpha) + (\alpha - 1) \log x_i - \frac{x_i}{\beta} = \sum_{i=1}^n \frac{d}{d\frac{1}{\beta}} \left(\alpha \log \frac{1}{\beta} - \frac{x_i}{\beta} \right) \\ &= \sum_{i=1}^n \frac{\alpha}{(\frac{1}{\beta})} - x_i = n \frac{\alpha}{(\frac{1}{\beta})} - \sum_{i=1}^n x_i \end{aligned}$$

Setting this equal to 0 we get:

$$n \frac{\alpha}{(\frac{1}{\beta})} - \sum_{i=1}^n x_i = 0 \implies \frac{1}{n} \sum_{i=1}^n x_i = \frac{\alpha}{(\frac{1}{\beta})} \implies \frac{1}{\beta} = \frac{\alpha}{\frac{1}{n} \sum_{i=1}^n x_i}$$

So our critical point is:

$$\frac{1}{\beta} = \frac{\alpha}{\frac{1}{n} \sum_{i=1}^n x_i}$$

To verify this is indeed a maximum point we take the second derivative:

$$\frac{d^2}{d(\frac{1}{\beta})^2} l(\beta) = \frac{d}{d\frac{1}{\beta}} \left(n \frac{\alpha}{(\frac{1}{\beta})} - \sum_{i=1}^n x_i \right) = -n \frac{\alpha}{(\frac{1}{\beta})^2} = -n \alpha \beta^2 < 0$$

Since $\alpha > 0$, Showing that $l(\beta)$ at our critical point is indeed a maximum point. So the MLE for $\frac{1}{\beta}$ is indeed:

$$\left(\frac{1}{\beta} \right)^\wedge = \frac{\alpha}{\frac{1}{n} \sum_{i=1}^n X_i}$$

Therefore the MLE of β is given by:

$$\hat{\beta} = \frac{1}{\left(\frac{1}{\beta} \right)^\wedge} = \frac{1}{\alpha} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \quad \square$$

b.

First I will show a result about the Gamma distribution, letting $X \sim \text{Gamma}(\alpha, \beta)$:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{x}{\beta}\right)^\alpha e^{-\frac{x}{\beta}} dx$$

Let $u = \frac{x}{\beta}$ then $\frac{du}{dx} = \frac{1}{\beta}$ and:

$$\mathbb{E}[X] = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{x}{\beta}\right)^\alpha e^{-\frac{x}{\beta}} dx = \frac{\beta}{\Gamma(\alpha)} \int_0^{\infty} u^\alpha e^{-u} du$$

Now let $v = u^\alpha$ and $\frac{dv}{du} = \alpha u^{\alpha-1}$ so that $\frac{dv}{du} = \alpha u^{\alpha-1}$ and $w = -e^{-u}$, then:

$$\mathbb{E}[X] = \frac{\beta}{\Gamma(\alpha)} \int_0^{\infty} u^\alpha e^{-u} du = \frac{\beta}{\Gamma(\alpha)} \left(-u^\alpha e^{-u} \Big|_0^{\infty} + \int_0^{\infty} \alpha u^{\alpha-1} e^{-u} du \right) = \frac{\Gamma(\alpha) \beta \alpha}{\Gamma(\alpha)} = \beta \alpha$$

Where we use the fact that the exponential grows faster than the polynomial term as $u \rightarrow \infty$ and the fact that the remaining integral is the definition of $\Gamma(\alpha)$.

Now computing $\mathbb{E}[X^2]$:

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\infty} x^2 \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \frac{\beta}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{x}{\beta}\right)^{\alpha+1} e^{-\frac{x}{\beta}} dx$$

Let $u = \frac{x}{\beta}$ then $\frac{du}{dx} = \frac{1}{\beta}$ and:

$$\mathbb{E}[X^2] = \frac{\beta}{\Gamma(\alpha)} \int_0^{\infty} \left(\frac{x}{\beta}\right)^{\alpha+1} e^{-\frac{x}{\beta}} dx = \frac{\beta^2}{\Gamma(\alpha)} \int_0^{\infty} u^{\alpha+1} e^{-u} du$$

Now let $v = u^{\alpha+1}$ and $\frac{dv}{du} = (\alpha+1)u^\alpha$ so that $\frac{dv}{du} = (\alpha+1)u^\alpha$ and $w = -e^{-u}$, then:

$$\begin{aligned} \mathbb{E}[X^2] &= \frac{\beta^2}{\Gamma(\alpha)} \int_0^{\infty} u^{\alpha+1} e^{-u} du = \frac{\beta^2}{\Gamma(\alpha)} \left(-u^{\alpha+1} e^{-u} \Big|_0^{\infty} + \int_0^{\infty} (\alpha+1) u^\alpha e^{-u} du \right) \\ &= \frac{\beta^2}{\Gamma(\alpha)} \left((\alpha+1) \int_0^{\infty} u^\alpha e^{-u} du \right) = \frac{(\alpha+1)\beta^2}{\Gamma(\alpha)} \alpha \Gamma(\alpha) = \alpha(\alpha+1)\beta^2 \end{aligned}$$

Where we use the fact that the exponential grows faster than the polynomial term as $u \rightarrow \infty$ and the fact that we just solved the preceding integral before.

Then we can find the variance:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2$$

Continued on next page.

The MLE is asymptotically normal so we just need to find the variance and we can construct a $1 - p$ confidence interval.

$$\mathbb{V}[\hat{\beta}] = \mathbb{V}\left[\frac{1}{\alpha}\left(\frac{1}{n}\sum_{i=1}^n X_i\right)\right] = \frac{1}{(n\alpha)^2}\sum_{i=1}^n \mathbb{V}[X_i] = \frac{1}{n^2\alpha^2}\sum_{i=1}^n \alpha\beta^2 = \frac{n\alpha\beta^2}{n^2\alpha^2} = \frac{\beta^2}{n\alpha}$$

Where I used the fact that all of the X_i are mutually independent and the variance of a sum of independent variables is just the sum of the variances.

Then plugging in our MLE we get the approximate $1 - p$ confidence interval:

$$\left(\hat{\beta} - z_{p/2}\frac{\hat{\beta}}{\sqrt{n\alpha}}, \hat{\beta} + z_{p/2}\frac{\hat{\beta}}{\sqrt{n\alpha}}\right)$$

So a 95% confidence interval is given by:

$$\left(\hat{\beta} - z_{0.05/2}\frac{\hat{\beta}}{\sqrt{n\alpha}}, \hat{\beta} + z_{0.05/2}\frac{\hat{\beta}}{\sqrt{n\alpha}}\right) = \left(\hat{\beta} - z_{0.025}\frac{\hat{\beta}}{\sqrt{n\alpha}}, \hat{\beta} + z_{0.025}\frac{\hat{\beta}}{\sqrt{n\alpha}}\right) \approx \left(\hat{\beta} - 1.96\frac{\hat{\beta}}{\sqrt{n\alpha}}, \hat{\beta} + 1.96\frac{\hat{\beta}}{\sqrt{n\alpha}}\right)$$

c.

Similar to before we have the likelihood function (just using $\theta = (\alpha, \beta)$ instead of β now):

$$L(\theta) = \prod_{i=1}^n \frac{1}{\beta^\alpha \Gamma(\alpha)} x_i^{\alpha-1} e^{-\frac{x_i}{\beta}} \implies l(\theta) = \log L(\theta) = \sum_{i=1}^n -\alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log x_i - \frac{x_i}{\beta}$$

We will use numerical approximation to maximize this, optim in R finds the minimum point so we will minimize the negative log likelihood which maximizes the log likelihood. I will also be using hessian = TRUE to approximate the observed fisher information.

From before we knew $\hat{\beta} = \frac{\bar{x}}{\alpha}$ was the MLE if α was known so I will be using $(\alpha, \beta) = (a, \frac{\bar{x}}{a})$ as my starting point where a is just some real number (I will try a number of different values for a).

```
lifetimes <- scan("guineapigdata.txt", sep=",", comment.char="#")

neg_log_lik <- function(theta, x = lifetimes){
  a <- theta[1]
  b <- theta[2]
  return(-1*sum(-a*log(b) - lgamma(a) + (a - 1)*log(x) - x/b))
}

find_mle <- function(){
  a <- 1
  mle <- optim(par = c(a, mean(lifetimes)/a), fn = neg_log_lik, hessian = TRUE)
  while (mle$convergence != 0){
    a <- a + 0.5
    mle <- optim(par = c(a, mean(lifetimes)/a), fn = neg_log_lik, hessian = TRUE)
  }
  a <- a + 1
  mle_alt1 <- optim(par = c(a, mean(lifetimes)/a), fn = neg_log_lik, hessian = TRUE)
  a <- a + 1
  mle_alt2 <- optim(par = c(a, mean(lifetimes)/a), fn = neg_log_lik, hessian = TRUE)
  return(list(mle1 = mle, mle2 = mle_alt1, mle3 = mle_alt2))
}

mles <- find_mle()
```

First we check the convergence codes (0 is a success)

```
mles$mle1$convergence
```

```
## [1] 0
```

```
mles$mle2$convergence
```

```
## [1] 0
```

```
mles$mle3$convergence
```

```
## [1] 0
```

Then we check the function values found

```
mles$mle1$value
```

```
## [1] 425.8014
```

```
mles$mle2$value
```

```
## [1] 425.8014
```

```
mles$mle3$value
```

```
## [1] 425.8014
```

Then we check the actual parameters

```
mles$mle1$par
```

```
## [1] 3.082674 57.358233
```

```
mles$mle2$par
```

```
## [1] 3.082983 57.345690
```

```
mles$mle3$par
```

```
## [1] 3.082983 57.345690
```

Just because I am also going to check the Hessian matrices

```
mles$mle1$hessian
```

```
##           [,1]      [,2]  
## [1,] 27.546261 1.25526878  
## [2,] 1.255269 0.06746554
```

```
mles$mle2$hessian
```

```
##           [,1]      [,2]  
## [1,] 27.543038 1.2555434  
## [2,] 1.255543 0.0675178
```

```
mles$mle3$hessian
```

```
##           [,1]      [,2]  
## [1,] 27.543038 1.2555434  
## [2,] 1.255543 0.0675178
```

Continued on next page.

We can see that mle2 and mle3 match each other basically exactly so I will take that as the true MLE. That is we are

taking $\hat{\theta}_n = \begin{pmatrix} \hat{\alpha}_n, \hat{\beta}_n \end{pmatrix} = \begin{pmatrix} 3.082983, 57.345690 \end{pmatrix}$ with the estimated observed Fisher information being:

$$I_n(\theta) = \mathbb{E}[H(\theta)] \approx H(\hat{\theta}_n) \approx \begin{bmatrix} 27.543038 & 1.2555434 \\ 1.255543 & 0.0675178 \end{bmatrix}$$

Then we can invert the estimated observed Fisher information to get the matrix:

$$\begin{aligned} V_n(\theta) &= I_n(\theta)^{-1} = \mathbb{E}[H(\theta)]^{-1} \approx H(\hat{\theta}_n)^{-1} \approx \begin{bmatrix} 27.543038 & 1.2555434 \\ 1.255543 & 0.0675178 \end{bmatrix}^{-1} \\ &= \frac{1}{(27.543038)(0.0675178) - (1.255543)(1.2555434)} \begin{bmatrix} 0.0675178 & -1.2555434 \\ -1.255543 & 27.543038 \end{bmatrix} \\ &= \begin{bmatrix} 0.238363 & -4.432535 \\ -4.432535 & 97.237174 \end{bmatrix} = V_n(\hat{\theta}) \end{aligned}$$

```
v <- solve(mles$mle2$hessian)
v
```

```
##           [,1]      [,2]
## [1,]  0.238363 -4.432535
## [2,] -4.432535 97.237174
```

Then by the properties of the MLE we know $\hat{\theta}_n$ is asymptotically normal with mean θ and variance matrix $V_n(\theta)$.

Which implies that $\hat{\alpha}_n \sim N(\alpha, V_n(\theta)_{1,1})$ and $\hat{\beta}_n \sim N(\beta, V_n(\theta)_{2,2})$.

Then we can use this to construct approximate $1 - p$ confidence intervals:

$$\alpha : \left(\hat{\alpha}_n - z_{p/2} V_n(\hat{\theta})_{1,1}, \hat{\alpha}_n + z_{p/2} V_n(\hat{\theta})_{1,1} \right) \quad \text{and} \quad \beta : \left(\hat{\beta}_n - z_{p/2} V_n(\hat{\theta})_{2,2}, \hat{\beta}_n + z_{p/2} V_n(\hat{\theta})_{2,2} \right)$$

Therefore our 95% confidence intervals are:

$$\alpha : \left(\hat{\alpha}_n - 1.96\sqrt{0.238363}, \hat{\alpha}_n + 1.96\sqrt{0.238363} \right) \quad \text{and} \quad \beta : \left(\hat{\beta}_n - 1.96\sqrt{97.2371}, \hat{\beta}_n + 1.96\sqrt{97.2371} \right)$$

Which by the R code below is:

$$\alpha : \left(2.126064, 4.039903 \right) \quad \text{and} \quad \beta : \left(38.01834, 76.67304 \right)$$

As a side note one might want to use 2 instead of 1.96 to make this more accurate due to the large amount of approximating used in this problem.

```
mles$mle2$par[1] + c(-1,1)*1.96*sqrt(v[1,1])  
## [1] 2.126064 4.039903  
mles$mle2$par[2] + c(-1,1)*1.96*sqrt(v[2,2])  
## [1] 38.01834 76.67304
```

Since the optimization function returned a success code of 0 we can say that the algorithm used found a maximum. This is also supported by the fact that we got very similar points for three optimizations with different start points (with two of them even being the same point) and very similar hessian matrices as well.