Question 1

Suppose you are tracking the value of two companies listed on the London Stock Exchange over the course of one week. Rather than record the actual values of the share prices, you record the increase or decrease in each share price value at the daily close to the nearest pound. You record the following table:

	Monday	Tuesday	Wednesday	Thursday	Friday
Company X	5	4	8	6	2
Company Y	3	2	7	4	-1

Table 1: Daily change in share price (£)

Without using a calculator:

(a) Compute the sample covariance between the two sequences to two decimal places.

(b) Compute the sample correlation between the two sequences. You may leave your answer as a fraction.

(c) Compute the sample correlation between the two sequences to two decimal places.

(d) Are the two sequences significantly correlated?

Solution to Question 1

Part (a):

Let $\mathbf{x} = (5, 4, 8, 6, 2)$ and let $\mathbf{y} = (3, 2, 7, 4, -1)$. Then

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}(25) = 5$$
$$\overline{y} = \frac{1}{5}\sum_{i=1}^{5} y_i = \frac{1}{5}(15) = 3$$

Then,

$$\mathbf{x} - \overline{x} = (0, -1, 3, 1, -3)$$
$$\mathbf{y} - \overline{y} = (0, -1, 4, 1, -4)$$

Which implies that the sample covariance is

$$\frac{1}{5-1}\sum_{i=1}^{5}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{4}\left[(0)(0) + (-1)(-1) + (3)(4) + (1)(1) + (-3)(-4)\right]$$

$$= \frac{1}{4}\left[0 + 1 + 12 + 1 + 12\right]$$

$$= \frac{26}{4} = \frac{13}{2} = 6.5$$

Part (b):

We need to compute the sample variances of $\mathbf{x}$ and $\mathbf{y}$, or at least the sum of squared deviations:

$$\sum_{i=1}^{5}(x_i-\overline{x})^2 = (0)^2+(1)^2+(3)^2+(1)^2+(-3)^2 = 0+1+9+1+9 = 20$$
$$\sum_{i=1}^{5}(y_i-\overline{y})^2 = (0)^2+(-1)^2+(4)^2+(1)^2+(-4)^2 = 0+1+16+1+16 = 34$$

Then the sample correlation is

$$r_{XY} = \frac{\sum_{i=1}^{5}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{5}(x_i-\overline{x})^2}\sqrt{\sum_{i=1}^{5}(y_i-\overline{y})^2}}.$$

$$= \frac{26}{\sqrt{20}\sqrt{34}} = \frac{26}{2\sqrt{5}\sqrt{34}} = \frac{13}{\sqrt{5}\sqrt{34}}$$

Part (c):

Using the computation in Part (b):

$$\begin{aligned} r_{XY} = \frac{13}{\sqrt{5}\sqrt{34}} = \frac{\sqrt{169}}{\sqrt{170}} = \sqrt{\frac{169}{170}} = \sqrt{\frac{170-1}{170}} &= \sqrt{1-\frac{1}{170}} \\ &> \sqrt{1-\frac{1}{100}} = \sqrt{0.99} \\ &> 0.99 \end{aligned}$$

The first inequality follows since

$$\begin{aligned} &170 > 100 \\ \Rightarrow &\frac{1}{170} < \frac{1}{100} \\ \Rightarrow &1-\frac{1}{170} > 1-\frac{1}{100} \\ \Rightarrow &\sqrt{1-\frac{1}{170}} > \sqrt{1-\frac{1}{100}} \end{aligned}$$

The second inequality follows because the functions $f(x)=\sqrt{x}$ and $g(x)=x$ on the interval (0,1) have the property $f(x)>g(x)$ (to see this, plot the functions).

Finally, although we always have $r_{XY}\le 1$, in this case $\frac{169}{170}<1 \Rightarrow \sqrt{\frac{169}{170}}<1$, and therefore

$$0.99 < r_{XY} < 1.$$

Part (d):

Although we do not have the distribution for r_{XY} in this case, and so cannot make an inference with any degree of confidence, a value of $r_{XY}>0.99$ is likely to be significant.

Question 2

Suppose that for every batch of lightbulbs produced in a factory an unknown proportion θ are defective. Suppose that a random sample of n lightbulbs is taken from a batch, and for $i = 1, 2, \ldots, n$ let the random variable $X_i = 1$ if the ith lightbulb is defective and let $X_i = 0$ otherwise. We assume that the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are independent, and we observe $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

(a) What distribution can we use to model each X_i?

(b) Given your answer in (a), write down the probability mass/density function $f(x_i|\theta)$.
Hint: the p.m.f./p.d.f. should be a polynomial in θ.

(c) Given your answer in (b), compute down the likelihood function $L(\theta|\mathbf{x})$.

(d) Compute the maximum likelihood estimate for θ, given the observations $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

(e) Write down the maximum likelihood estimator for θ, given the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$.

Solution to Question 2

Part (a):

Since the random variable X_i is equal to 1 with (unknown) probability θ, and 0 otherwise (i.e. with probability $1 - \theta$), we can model $X_i \sim \text{Bern}(\theta)$.

Part (b):

$$f(x_i|\theta) = \begin{cases} \theta^{x_i}(1-\theta)^{1-x_i}, & \text{if } x_i \in \{0,1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that this is equivalent to:

$$f(x_i|\theta) = \begin{cases} \theta, & \text{if } x_i = 1, \\ 1-\theta, & \text{if } x_i = 0, \\ 0, & \text{otherwise.} \end{cases}$$

This is the probability mass function of a random variable following a Bernoulli distribution with parameter θ.

Part (c):

Since the X_i are assumed to be independent, the joint probability mass function of $\mathbf{X}$ is

$$f(\mathbf{x}|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\left(\sum_{i=1}^{n} x_i\right)}(1-\theta)^{\left(\sum_{i=1}^{n}(1-x_i)\right)} = \theta^{n\overline{x}}(1-\theta)^{n-n\overline{x}}$$

Then,

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \theta^{n\overline{x}}(1-\theta)^{n-n\overline{x}}.$$

Part (d):

Although this was hinted at in Example 3.6.7 of the lecture notes, this is the complete solution.

Instead of trying to maximise the likelihood directly, it will be easier to maximise the log-likelihood.

$$\begin{aligned} \log L(\theta|\mathbf{x}) &= \log\left(\theta^{n\overline{x}}(1-\theta)^{n-n\overline{x}}\right) \\ &= \log\left(\theta^{n\overline{x}}\right) + \log\left((1-\theta)^{n-n\overline{x}}\right) \\ &= n\overline{x}\log\theta + (n-n\overline{x})\log(1-\theta) \end{aligned}$$

For the moment, assume that the x_i are not all 0 or not all 1. Taking the derivative,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\log L(\theta|\mathbf{x}) = \frac{n\overline{x}}{\theta} + \frac{n-n\overline{x}}{1-\theta}(-1) = \frac{n\overline{x}}{\theta} - \frac{n-n\overline{x}}{1-\theta}$$

Setting the derivative equal to 0,

$$0 = \frac{n\overline{x}}{\theta} - \frac{n-n\overline{x}}{1-\theta} = \frac{n\overline{x}(1-\theta)-(n-n\overline{x})\theta}{\theta(1-\theta)}$$

$$\begin{aligned}
&\Rightarrow 0 = n\overline{x}(1-\theta)-(n-n\overline{x})\theta = n\overline{x} - n\overline{x}\theta - n\theta + n\overline{x}\theta \\
&\Rightarrow 0 = n(\overline{x}-\theta) \\
&\Rightarrow \theta = \overline{x}
\end{aligned}$$

To check if this is a maximum or a minimum, we need to compute the second derivative evaluated at this value, $\theta = \overline{x}$:

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta|\mathbf{x}) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\left[\frac{n\overline{x}}{\theta} - \frac{n-n\overline{x}}{1-\theta}\right] \\
&= -\frac{n\overline{x}}{\theta^2} - \left(-\frac{n-n\overline{x}}{(1-\theta)^2}(-1)\right) \\
&= -\frac{n\overline{x}}{\theta^2} - \frac{n-n\overline{x}}{(1-\theta)^2} \\
&= \frac{n}{\theta^2(1-\theta)^2}\left(-\overline{x}(1-2\theta+\theta^2)-(1-\overline{x})\theta^2\right) \\
&= \frac{n}{\theta^2(1-\theta)^2}\left(-\overline{x}+2\overline{x}\theta-\theta^2\right)
\end{aligned}$$

Now,

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\log L(\theta|\mathbf{x})\right|_{\theta=\overline{x}} = \frac{n}{\theta^2(1-\theta)^2}\left(\overline{x}^2-\overline{x}\right) = \frac{n\overline{x}}{\theta^2(1-\theta)^2}\left(\overline{x}-1\right)$$

Therefore, the sign of the second derivative is the same as the sign of $x(\overline{x}-1)$. Since all $x_i \in \{0,1\}$, and we assumed that the x_i are not all 0 and not all 1, then

$$\begin{aligned}
&0 < \overline{x} < 1 \\
\Rightarrow &\overline{x}(\overline{x}-1) < 0 \\
\Rightarrow &\left.\log L(\theta|\mathbf{x})\right|_{\theta=\overline{x}} < 0
\end{aligned}$$

which shows that $\theta = \overline{x}$ is a maximum.

Now, we need to check the cases that (i) all the $x_i = 0$ and (ii) all the $x_i = 1$. If all the $x_i = 0$, then

$$\log L(\theta|\mathbf{x}) = (n-n\overline{x})\log(1-\theta) = n\log(1-\theta)$$

since $\overline{x} = 0$. Note that as θ increases, this function decreases, and there is no local maximum in the interval $\theta \in (0,1)$. Therefore, the maximum occurs at $\theta = 0$, and therefore $\theta = 0 = \overline{x}$.

For the other case, suppose that all $x_i = 1$. Then $\overline{x} = 1$ and

$$\log L(\theta|\mathbf{x}) = n\overline{x}\log\theta = n\log\theta$$

This is an increasing function of θ, with no local maximum in the interval $(0, 1)$. Therefore, the maximum occurs at $\theta = 1 = \overline{x}$.

Therefore, in all cases, $\widehat{\theta} = \overline{x}$ is the maximum likelihood estimate for θ. If we were being very careful, and wanted to emphasise that this depends on observing the data $\mathbf{x}$, we would write $\widehat{\theta}(\mathbf{x}) = \overline{x}$.

Part (e):

From Part (d), the maximum likelihood estimate is $\widehat{\theta}(\mathbf{x}) = \overline{x}$. Therefore, the maximum likelihood **estimator** is $\widehat{\theta}(\mathbf{X}) = \overline{X}$.

Question 3

Young's inequality states that if a and b are non-negative real numbers, and p and q are any positive numbers such that

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab.$$

Use Young's inequality to prove Hölder's Inequality: Let X and Y be random variables and let p and q be two positive numbers satisfying $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$|\mathrm{E}\,(XY)| \leq \left(\mathrm{E}\,(|X|^p)\right)^{1/p} \left(\mathrm{E}\,(|Y|^q)\right)^{1/q}.$$

(a) If Z is a non-negative random variable, i.e. $Z \geq 0$, prove that $\mathrm{E}\,(Z) \geq 0$.

(b) Prove that $\mathrm{E}\,(|XY|) \geq 0$.

(c) Prove that $|\mathrm{E}\,(XY)| \leq \mathrm{E}\,(|XY|)$.

(d) Use Young's inequality to prove $\mathrm{E}\,(|XY|) \leq \left(\mathrm{E}\,(|X|^p)\right)^{1/p} \left(\mathrm{E}\,(|Y|^q)\right)^{1/q}$.

(e) Conclude that Hölder's Inequality is true.

(f) Use Hölder's Inequality to prove the Cauchy-Schwarz Inequality: $|\mathrm{E}\,(XY)| \leq (\mathrm{E}(|X|^2))^{1/2}(\mathrm{E}(|Y|^2))^{1/2}$.

(g) Use the Cauchy-Schwarz inequality to prove Theorem 4.1.6: $|\mathrm{Cov}\,(X, Y)| \leq \sigma_X\sigma_Y$, where σ_X^2 and σ_Y^2 are the variances of X and Y, respectively.

Solution to Question 3

Part (a):

We consider the case that Z is a continuous random variable; the discrete case is similar.

Let f be the probablity density function of Z. Then $f(z) \geq 0$ for any $z \in \mathbb{R}$. Since $Z \geq 0$, the support of f is a subset of $\{z \in \mathbb{R} : z \geq 0\}$. Now,

$$\begin{aligned}
&\forall z \geq 0, f(z) \geq 0 \\
\Rightarrow &\forall z \geq 0, zf(z) \geq 0 \\
\Rightarrow &\int_0^\infty zf(z)\mathrm{d}z \geq 0 \\
\Rightarrow &\mathrm{E}\,(Z) \geq 0
\end{aligned}$$

Part (b):

For any random variables X and Y, define the random variable $Z = |XY|$. Then Z is a non-negative random variable, and from Part (a) $\mathrm{E}\,(|XY|) = \mathrm{E}\,(Z) \geq 0$.

Part (c):

First, we note that $-|XY| \leq XY \leq |XY|$. Now,

$$\begin{aligned}
&-|XY| \leq XY \leq |XY| \\
\Rightarrow &-\mathrm{E}\,(|XY|) \leq \mathrm{E}\,(XY) \leq \mathrm{E}\,(|XY|)
\end{aligned}$$

The second line follows from (for example) considering $f_{X,Y}$ to be the joint probability density function of X and Y and

$$\begin{aligned}
&\forall x, y \in \mathbb{R}, \qquad xy \leq |xy| \\
\Rightarrow &\forall x, y \in \mathbb{R}, \qquad xy f_{X,Y}(x,y) \leq |xy| \, f_{X,Y}(x,y) \\
\Rightarrow &\int_{-\infty}^{\infty} xy f_{X,Y}(x,y) \mathrm{d}x\mathrm{d}y \leq \int_{-\infty}^{\infty} |xy| \, f_{X,Y}(x,y) \mathrm{d}x\mathrm{d}y \\
\Rightarrow &\mathrm{E}\left(XY\right) \leq \mathrm{E}\left(|XY|\right)
\end{aligned}$$

This proves the second inequality; the first inequality $-\mathrm{E}\left(|XY|\right) \leq \mathrm{E}\left(XY\right)$ follows similarly, and using $\mathrm{E}\left(-|XY|\right) = -\mathrm{E}\left(|XY|\right)$.

Next, since for any $a \in \mathbb{R}$ and any $b \in \mathbb{R}, b \geq 0$,

$$\begin{aligned}
&-b \leq a \leq b \\
\Rightarrow &|a| \leq b
\end{aligned}$$

and since from Part (b) we have $\mathrm{E}\left(|XY|\right) \geq 0$, then

$$\begin{aligned}
&-\mathrm{E}\left(|XY|\right) \leq \mathrm{E}\left(XY\right) \leq \mathrm{E}\left(|XY|\right) \\
\Rightarrow &|\mathrm{E}\left(XY\right)| \leq \mathrm{E}\left(|XY|\right)
\end{aligned}$$

Part (d):

Starting with Young's Inequality,

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab.$$

for p, q such that $\frac{1}{p} + \frac{1}{q} = 1$. Now define

$$a = \frac{|X|}{(\mathrm{E}\left(|X|^p\right))^{1/p}}, \qquad b = \frac{|Y|}{(\mathrm{E}\left(|Y|^q\right))^{1/q}}.$$

Then, plugging these values into Young's Inequality,

$$\frac{1}{p}\frac{|X|^p}{\mathrm{E}\left(|X|^p\right)} + \frac{1}{q}\frac{|Y|^q}{\mathrm{E}\left(|Y|^q\right)} \geq \frac{|XY|}{(\mathrm{E}\left(|X|^p\right))^{1/p}(\mathrm{E}\left(|Y|^q\right))^{1/q}}$$

Now, taking expectations of both sides and using the linearity of expectation,

$$\begin{aligned}
&\frac{1}{p}\frac{\mathrm{E}\left(|X|^p\right)}{\mathrm{E}\left(|X|^p\right)} + \frac{1}{q}\frac{\mathrm{E}\left(|Y|^q\right)}{\mathrm{E}\left(|Y|^q\right)} \geq \frac{\mathrm{E}\left(|XY|\right)}{(\mathrm{E}\left(|X|^p\right))^{1/p}(\mathrm{E}\left(|Y|^q\right))^{1/q}} \\
\Rightarrow &\frac{1}{p}\left(1\right) + \frac{1}{q}\left(1\right) \geq \frac{\mathrm{E}\left(|XY|\right)}{(\mathrm{E}\left(|X|^p\right))^{1/p}(\mathrm{E}\left(|Y|^q\right))^{1/q}} \\
\Rightarrow &1 \geq \frac{\mathrm{E}\left(|XY|\right)}{(\mathrm{E}\left(|X|^p\right))^{1/p}(\mathrm{E}\left(|Y|^q\right))^{1/q}} \\
\Rightarrow &(\mathrm{E}\left(|X|^p\right))^{1/p}(\mathrm{E}\left(|Y|^q\right))^{1/q} \geq \mathrm{E}\left(|XY|\right)
\end{aligned}$$

as required.

Part (e):

Putting together Parts (c) and (d),

$$|\mathrm{E}\,(XY)| \le \mathrm{E}\,(|XY|) \le (\mathrm{E}\,(|X|^p))^{1/p}(\mathrm{E}\,(|Y|^q))^{1/q}$$

which proves Hölder's Inequality.

Part (f):

Taking $p = q = 2$, Hölder's Inequality becomes

$$|\mathrm{E}\,(XY)| \le (\mathrm{E}(|X|^2))^{1/2}(\mathrm{E}(|Y|^2))^{1/2}$$

Part (g):

If we define μ_X and μ_Y to be the means of X and Y, respectively, and σ_X^2 and σ_Y^2 to be the variances of X and Y, respectively, then starting with the Cauchy-Schwarz inequality for the random variables $X - \mu_X$ and $Y - \mu_Y$,

$$|\mathrm{Cov}\,(X,Y)| = |\mathrm{E}[(X-\mu_X)\,(Y-\mu_Y)]| \le (\mathrm{E}(|X-\mu_X|^2))^{1/2}(\mathrm{E}(|Y-\mu_Y|^2))^{1/2} = \sigma_X\sigma_Y.$$

Alternative proof:

An alternative approach is to prove the Cauchy-Schwarz Inequality in terms of inner products, i.e.

$$|\langle X, Y\rangle|^2 \le \langle X, X\rangle\langle Y, Y\rangle.$$

and then define the inner product $\langle X, Y\rangle = \mathrm{E}\,(XY)$ (you will need to show this is an inner product), and then defining μ_X and μ_Y as the means of X and Y, respectively, and σ_X^2 and σ_Y^2 as the variances of X and Y, respectively,

$$\begin{aligned}
(\mathrm{Cov}\,(X,Y))^2 &= |\mathrm{Cov}\,(X,Y)|^2 \\
&= |\mathrm{E}[(X-\mu_X)\,(Y-\mu_Y)]|^2 \\
&= |\langle X-\mu_X, Y-\mu_Y\rangle|^2 \\
&\le \langle X-\mu_X, X-\mu_X\rangle\langle Y-\mu_Y, Y-\mu_Y\rangle \qquad \text{(Cauchy-Schwarz Inequality)} \\
&= \mathrm{E}[(X-\mu_X)^2]\mathrm{E}[(Y-\mu_Y)^2] \\
&= \mathrm{Var}\,(X)\,\mathrm{Var}\,(Y) \\
&= \sigma_X^2\sigma_Y^2 \\
\Rightarrow \mathrm{Cov}\,(X,Y) &\le \sigma_X\sigma_Y
\end{aligned}$$

Question 4 (Knowledge of partial derivatives required)

Suppose the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are assumed to be independent and identically distributed as $\mathrm{N}\left(\mu, \sigma^2\right)$, where μ and σ^2 are unknown. Suppose further that $\mathbf{X}$ is observed as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

(a) Compute the likelihood $L(\mu, \sigma^2|\mathbf{x})$.

(b) Compute the log-likelihood $\log L(\mu, \sigma^2|\mathbf{x})$.

(c) Compute the partial derivative $\frac{\partial}{\partial \mu} \log L(\mu, \sigma^2|\mathbf{x})$.

(d) Set the partial derivative in (c) equal to 0 and solve for μ. Show that this value of μ maximises $L(\mu, \sigma^2|\mathbf{x})$ globally for fixed σ^2.

(e) Compute the partial derivative $\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2|\mathbf{x})$. **Hint**: it may help to set $z = \sigma^2$ and then compute the partial derivative with respect to z.

(f) Set the partial derivative in (e) equal to 0 and solve for σ^2. Show that this value of σ^2 is a (local) maximum for $L(\mu, \sigma^2|\mathbf{x})$, for fixed μ.

(g) Show that the value of σ^2 found in (f) maximises the likelihood $L(\mu, \sigma^2|\mathbf{x})$ globally for fixed μ.

(h) Write down the maximum likelihood estimators for μ and σ^2 given the random variables $\mathbf{X}$.

Solution to Question 4 (Knowledge of partial derivatives required)

Part (a):

The probability density function for each X_i is

$$f(x_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right).$$

Since the X_i are assumed to be independent, the joint distribution of $\mathbf{X}$ is

$$\begin{aligned} f(\mathbf{x}|\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i|\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \end{aligned}$$

Therefore the likelihood of μ and σ^2 given $\mathbf{x}$ is

$$L(\mu, \sigma^2|\mathbf{x}) = f(\mathbf{x}|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right).$$

Part (b):

$$\begin{aligned}
\log L(\mu, \sigma^2|\mathbf{x}) &= \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\right) + \log\left(\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)\right) \\
&= \log\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\right) + \left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) \\
&= \log\left(\left(2\pi\sigma^2\right)^{-n/2}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
&= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\
\Rightarrow \log L(\mu, \sigma^2|\mathbf{x}) &= -\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2
\end{aligned}$$

Part (c):

$$\begin{aligned}
\frac{\partial}{\partial\mu}\log L(\mu, \sigma^2|\mathbf{x}) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)(-1) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)
\end{aligned}$$

Part (d):

$$\begin{aligned}
&\frac{\partial}{\partial\mu}\log L(\mu, \sigma^2|\mathbf{x}) = 0 \\
\Rightarrow &\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \\
\Rightarrow &\sum_{i=1}^{n}(x_i-\mu) = 0 \\
\Rightarrow &\mu = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{x}
\end{aligned}$$

To show this $\mu = \overline{x}$ maximises the likelihood globally, recall Exercise 1.2.10 which showed that for any $\mu \in \mathbb{R}$,

$$\sum_{i=1}^{n}(x_i-\overline{x}) \leq \sum_{i=1}^{n}(x_i-\mu)$$

Therefore, for any $\sigma^2 > 0$,

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2 \geq -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\Rightarrow \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right) \geq \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$\Rightarrow \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right) \geq \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$\Rightarrow L(\overline{x},\sigma^2|\mathbf{x}) \geq L(\mu,\sigma^2|\mathbf{x})$$

Part (e):

One can compute the partial derivative with respect to σ^2 directly, but one needs to be careful not to compute the derivative with respect to σ (by accident). Solving for σ^2 for either $\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) = 0$ or $\frac{\partial}{\partial\sigma}\log L(\mu,\sigma^2|\mathbf{x}) = 0$ results in the same answer, but the second option is slightly more work, requiring the Chain Rule. Although we shall compute it directly below, if any of the steps are unclear, try setting $z = \sigma^2$ and then computing the partial derivative with respect to z.

The log-likelihood $\log L(\mu,\sigma^2|\mathbf{x})$ is

$$\log L(\mu,\sigma^2|\mathbf{x}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

Part (f):

Now,

$$\begin{aligned}\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) &= 0 - \frac{n}{2}\cdot\frac{1}{\sigma^2} - \frac{1}{2}\cdot\frac{-1}{(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\end{aligned}$$

Setting this partial derivative equal to 0,

$$\begin{aligned}&\frac{\partial}{\partial\sigma^2}\log L(\mu,\sigma^2|\mathbf{x}) = 0 \\ \Rightarrow & -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \\ \Rightarrow & -n\sigma^2 + \sum_{i=1}^{n}(x_i-\mu)^2 = 0 \\ \Rightarrow & \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\end{aligned}$$

Since the value $\mu = \overline{x}$ maximised the likelihood for any value σ^2, we set $\mu = \overline{x}$ here, so

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

To show this is a local maximum, we need to compute the second derivative $\frac{\partial^2}{\partial(\sigma^2)^2}\log L(\mu,\sigma^2|\mathbf{x})$ and evaluate it at $\sigma^2=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})^2$. In this case, it may be simpler to substitute values $z=\sigma^2$ first. Since

$$\log L(\mu,\sigma^2|\mathbf{x}) = -\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\left(\sigma^2\right)-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2$$
$$\Rightarrow \log L(\overline{x},z|\mathbf{x}) = -\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(z)-\frac{1}{2z}\sum_{i=1}^n(x_i-\overline{x})^2$$

And setting $A=\sum_{i=1}^n(x_i-\overline{x})^2$,

$$\log L(\overline{x},z|\mathbf{x}) = -\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(z)-\frac{A}{2z}$$

Now, we compute the second partial derivative with respect to z:

$$\frac{\partial^2}{\partial z^2}\log L(\overline{x},z|\mathbf{x}) = \frac{\partial}{\partial z}\left(\frac{\partial}{\partial z}\left[-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(z)-\frac{A}{2z}\right]\right)$$

$$=\frac{\partial}{\partial z}\left(-\frac{n}{2z}+\frac{A}{2z^2}\right)$$

$$=\frac{n}{2z^2}+\frac{A}{2z^3}(-2)$$

$$\Rightarrow \frac{\partial^2}{\partial z^2}\log L(\overline{x},z|\mathbf{x}) = \frac{n}{2z^2}-\frac{A}{z^3}$$

Evaluating this second derivative at $z=\sigma^2=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})^2=\frac{A}{n}$,

$$\left.\frac{\partial^2}{\partial z^2}\log L(\overline{x},z|\mathbf{x})\right|_{z=\frac{A}{n}} = \frac{n}{2(A/n)^2}-\frac{A}{(A/n)^3}$$

$$=\frac{n}{2}\cdot\frac{n^2}{A^2}-\frac{An^3}{A^3}$$

$$=\frac{n^2}{2A^2}-\frac{n^3}{A^2}=-\frac{n^2}{2A^2}<0$$

since $A>0$, if we assume the x_i are not all equal. This shows that $\log L(\overline{x},\sigma^2|\mathbf{x})$ has a local maximum at $\sigma^2=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})^2$.

Part (g):

We have already shown that $\sigma^2=\frac{1}{n}\sum_{i=1}^n(x_i-\overline{x})^2$ is a local maximum of $\log L(\overline{x},\sigma^2|\mathbf{x})$, and is therefore also a maximum of $L(\overline{x},\sigma^2|\mathbf{x})$. Furthermore, it is the only possible local maximum, since it is the only value of σ^2 which satisfies $\frac{\partial}{\partial\sigma^2}\log L(\overline{x},\sigma^2|\mathbf{x})=0$. However, we still need to check the boundary points for the range of σ^2, which is $(0,\infty)$. Setting $z=\sigma^2$ again and $A=\sum_{i=1}^n(x_i-\overline{x})^2$, this is equivalent to computing the limits

$$\lim_{z\to 0^+} z^{-n/2}\exp\left(-\frac{A}{2z}\right)$$

$$\lim_{z\to\infty} z^{-n/2}\exp\left(-\frac{A}{2z}\right)$$

(We ignore the coefficient of $(2\pi)^{-n/2}$ for now.) For the first limit, we can set $y = \frac{1}{z}$, and then consider the limit $y \to \infty$, and use L'Hôpital's rule:

$$\lim_{z\to 0^+} z^{-n/2} \exp\left(-\frac{A}{2z}\right) \lim_{y\to\infty} y^{n/2} \exp\left(-\frac{Ay}{2}\right) \lim_{y\to\infty} \frac{y^{n/2}}{\exp\left(\frac{Ay}{2}\right)}$$

There are two cases to consider for n, which is a positive integer.. Either n is an even integer, or n is an odd integer. If n is even, $n = 2k$ for some integer k, and then

$$\begin{aligned}
\lim_{y\to\infty} \frac{y^{n/2}}{\exp\left(\frac{Ay}{2}\right)} = \lim_{y\to\infty} \frac{y^{k}}{\exp\left(\frac{Ay}{2}\right)} &= \lim_{y\to\infty} \frac{k y^{k-1}}{\frac{A}{2}\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule)} \\
&= \lim_{y\to\infty} \frac{k(k-1) y^{k-2}}{\left(\frac{A}{2}\right)^2\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule a second time)} \\
&\vdots \\
&= \lim_{y\to\infty} \frac{k!}{\left(\frac{A}{2}\right)^k\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule k times)} \\
&= 0
\end{aligned}$$

If n is odd, the result is similar. Set $n = 2k - 1$ for some positive integer k. Then

$$\begin{aligned}
\lim_{y\to\infty} \frac{y^{n/2}}{\exp\left(\frac{Ay}{2}\right)} = \lim_{y\to\infty} \frac{y^{k-1/2}}{\exp\left(\frac{Ay}{2}\right)} &= \lim_{y\to\infty} \frac{(k-1/2) y^{k-1-1/2}}{\frac{A}{2}\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule)} \\
&= \lim_{y\to\infty} \frac{(k-1/2)(k-3/2) y^{k-2-1/2}}{\left(\frac{A}{2}\right)^2\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule a second time)} \\
&\vdots \\
&= \lim_{y\to\infty} \frac{(k-1/2)(k-3/2)\cdots(1/2) y^{-1/2}}{\left(\frac{A}{2}\right)^k\exp\left(\frac{Ay}{2}\right)} && \text{(L'Hôpital's Rule k times)} \\
&= \lim_{y\to\infty} \frac{(k-1/2)(k-3/2)\cdots(1/2)}{y^{1/2}\left(\frac{A}{2}\right)^k\exp\left(\frac{Ay}{2}\right)} \\
&= 0
\end{aligned}$$

Therefore, we have:

$$\lim_{z\to 0^+} z^{-n/2} \exp\left(-\frac{A}{2z}\right) = 0$$

For the second limit,

$$\lim_{z\to\infty} z^{-n/2} \exp\left(-\frac{A}{2z}\right) = \lim_{z\to\infty} \frac{\exp\left(-\frac{A}{2z}\right)}{z^{n/2}} = \frac{\lim\limits_{z\to\infty} \exp\left(-\frac{A}{2z}\right)}{\lim\limits_{z\to\infty} z^{n/2}} = \frac{\exp(0)}{\lim\limits_{z\to\infty} z^{n/2}} = \frac{1}{\lim\limits_{z\to\infty} z^{n/2}} = 0$$

Since for all $\sigma^2 > 0$,

$$L(\overline{x}, \sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2\right) > 0,$$

this implies that $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ maximises $L(\overline{x}, \sigma^2|\mathbf{x})$ globally.

Overall, we can now state that the likelihood $L(\mu, \sigma^2|\mathbf{x})$ is maximised when we use the following maximum likelihood estimates for μ and σ^2:

$$\widehat{\mu}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$
$$\widehat{\sigma^2}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}).$$

Part (h):

Given the answers summarised in Part (g), the maximum likelihood estimators of μ and σ^2 given the random variables $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are

$$\widehat{\mu}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},$$
$$\widehat{\sigma^2}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = S_b^2.$$