SM-4331 Exercise 1

1. (a) Let X be a random variable with mean μ and variance σ^2. Show that $\mathrm{P}(|X-\mu| > 2\sigma) \leq 0.25$. What does this inequality tell us about the distribution of X?

> **Solution:** Using Chebyshev's inequality directly,
>
> $$\mathrm{P}(|X-\mu| > 2\sigma) \leq \frac{1}{2^2} = 0.25.$$
>
> The probability that X takes values that are within 2 s.d. away from the mean is 0.75.

(b) Let $X_1, \dots, X_n$ be an iid sample from a population with mean μ and variance σ^2. Show that for any $\epsilon > 0$, $\mathrm{P}(|\bar{X}_n - \mu| > \epsilon\sigma) \leq \frac{1}{n\epsilon^2}$. Compare this bound with the approximation implied by the CLT when n is large.

> **Solution:** Using Markov's inequality,
>
> $$\mathrm{P}(|\bar{X}_n - \mu| > \epsilon\sigma) \leq \frac{\mathrm{E}[(\bar{X}_n - \mu)^2]}{\epsilon^2\sigma^2} = \frac{\sigma^2/n}{\epsilon^2\sigma^2} = \frac{1}{n\epsilon^2}.$$
>
> By the CLT, $\bar{X}_n \xrightarrow{\mathrm{D}} \mathrm{N}(\mu, \sigma^2/n)$, and thus the bound implied by this is
>
> $$\mathrm{P}(|\bar{X}_n - \mu| > \epsilon\sigma) = \mathrm{P}\left(\left| \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \right| > \sqrt{n}\epsilon \right) = 2\big(1 - \Phi(\sqrt{n}\epsilon)\big).$$

[Note: The conditions required for these inequalities are minimum, and you may assume that they are met.]

2. Let X and Y be two r.v. with positive and finite variances. The correlation coefficient of X and Y is defined as

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

(If $\rho = 0$, X and Y are said to be uncorrelated or linearly independent).

(a) Show that $|\rho| < 1$. *Hint: Use the Cauchy-Schwarz inequality.*

> **Solution:**
>
> $$\begin{aligned} |\mathrm{Cov}(X,Y)|^2 &= \big|\mathrm{E}\left[(X - \mathrm{E}\,X)(Y - \mathrm{E}\,Y)\right]\big|^2 \\ &\leq \mathrm{E}\left[(X-\mathrm{E}\,X)^2\right]\mathrm{E}\left[(Y-\mathrm{E}\,Y)^2\right] \\ &= \mathrm{Var}(X)\,\mathrm{Var}(Y) \end{aligned}$$
>
> $$\Rightarrow |\rho| := \left| \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \right| \leq 1$$

(b) If $Y = aX + b$ for some constant $a \neq 0$ and b, show that $|\rho| = 1$.

Solution:

$$\begin{aligned}
\operatorname{Cov}(X, aX+b) &= \mathrm{E}[X(aX+b)] - \mathrm{E}(X)\,\mathrm{E}(aX+b) \\
&= \mathrm{E}(aX^2 + bX) - a\,\mathrm{E}(X)^2 - b\,\mathrm{E}(X) \\
&= a\big(\mathrm{E}(X^2) - \mathrm{E}(X)^2\big) + \cancelto{0}{b\,\mathrm{E}(X) - b\,\mathrm{E}(X)} \\
&= a\operatorname{Var}(X)
\end{aligned}$$

Also, $\operatorname{Var}(aX+b) = a^2 \operatorname{Var}(X)$. So

$$|\rho| = \left| \frac{\operatorname{Cov}(X, aX+b)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(aX+b)}} \right| = \left| \frac{a\operatorname{Var}(X)}{\sqrt{\operatorname{Var}(X)a^2\operatorname{Var}(X)}} \right| = 1$$

[Note: In fact, $|\rho| = 1$ if and only if $Y = aX + b$ for some constants $a \neq 0$ and b.]

3. Let $X_1, \dots, X_n$ be a sample from a distribution with mean μ and variance $\sigma^2 \in (0,1)$. Let $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, where $\bar{X}_n$ is the sample mean.

(a) Show that $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}\bar{X}_n^2$.

Solution:

$$\begin{aligned}
S_n^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i^2 + \bar{X}_n^2 - 2\bar{X}_n X_i) \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^n X_i^2 + n\bar{X}_n^2 - 2\bar{X}_n \sum_{i=1}^n X_i\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^n X_i^2 + n\bar{X}_n^2 - 2n\bar{X}_n^2\right\} \\
&= \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}\bar{X}_n^2
\end{aligned}$$

(b) Using Slutzky's theorem, show that $S_n^2 \xrightarrow{\mathrm{P}} \sigma^2$.

Solution: By the WLLN, $\frac{1}{n-1}\sum_{i=1}^n X_i^2 \xrightarrow{\mathrm{P}} \mathrm{E}(X_1)^2$. Also, since $X_n \xrightarrow{\mathrm{P}} \mu$, Slutzky's theorem gives us $X_n^2 \xrightarrow{\mathrm{P}} \mu^2$ and $n/(n-1) \to 1$ as $n \to \infty$, $\frac{n}{n-1}\bar{X}_n^2 \xrightarrow{\mathrm{P}} \mu^2$. Another application of Slutzky's theorem gives us

$$S_n^2 = \left\{\frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}\bar{X}_n^2\right\} \xrightarrow{\mathrm{P}} \mathrm{E}(X_1)^2 - \mu^2 = \sigma^2$$

4. Let $X_1, \ldots, X_n$ be sample from Bern(p), and $Y_1, \ldots, Y_m$ be a sample from Bern(q), and the two samples are independent of one another.

   (a) Find a reasonable estimator for $p - q$ and its standard error.

   Solution: Let $\theta = p - q$, and $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{m}\sum_{i=1}^{m} Y_i$. We will show that $\hat{\theta}$ is unbiased and consistent.

   $$\mathrm{E}(\hat{\theta}) = \mathrm{E}(\bar{X}_n - \bar{Y}_m) = p - q$$

   hence it is unbiased. Now,

   $$\begin{aligned} \mathrm{MSE}(\hat{\theta} - \theta) &= \mathrm{Var}(\hat{\theta}) \quad \text{since it is unbiased} \\ &= \mathrm{Var}(\bar{X}_n - \bar{Y}_m) \\ &= \mathrm{Var}(X_1)/n + \mathrm{Var}(Y_1)/m \\ &= p(1-p)/n + q(1-q)/m \to 0 \end{aligned}$$

   as $n, m \to \infty$, hence $\hat{\theta} \xrightarrow{\mathrm{P}} \theta$. The standard error is given by

   $$\mathrm{SE}(\hat{\theta}) = \bar{X}_n(1 - \bar{X}_n)/n + \bar{Y}_m(1 - \bar{Y}_m)/m$$

   (b) Find an approximate 95% confidence interval for $p - q$ when both n and m are large.

   Solution: This relationship holds approximately true:

   $$\frac{\hat{\theta} - \theta}{\mathrm{SE}(\hat{\theta})} \sim \mathrm{N}(0, 1)$$

   So an approximate 95% confidence interval for $\theta = p - q$ is

   $$\hat{\theta} \pm 1.96 \times \mathrm{SE}(\hat{\theta})$$

5. 100 people are given a standard antibiotic to treat an infection and another 100 are given a new antibiotic. In the first group, 90 people recover, while in the second group, 85 people recover. Let p_1 be the probability of recovery with the standard antibiotic, and p be the probability of recovery with the new antibiotic. We are interested in estimating $\theta = p_1 - p_2$. Provide an estimate, standard error, an 80% confidence interval, and a 95% confidence interval for θ.

Solution: $\bar{X}_n = 90/100 = 0.9$ and $\bar{Y}_n = 85/100 = 0.85$. Therefore, $\hat{\theta} = 0.9 -$

0.85 = 0.05. The standard error for this estimate is

$$\mathrm{SE}(\hat{\theta}) = \sqrt{\frac{0.9(1-0.9)}{100} - \frac{0.85(1-0.85)}{100}} = 0.04663.$$

Using the answers to Q4, an 80% confidence interval is

$$0.05 \pm 1.282 \times 0.04663 = (-0.0094, 0.109)$$

while a 95% confidence interval is

$$0.05 \pm 1.96 \times 0.04663 = (-0.0414, 0.141)$$

6. Let $Y_1, \ldots, Y_n$ be a sample from a Poisson distribution with mean $\theta > 0$ unknown.

(a) Let $Y = Y_1 + \cdots + Y_n$. Find the mean, variance, and the distribution of Y.
Hint 1: Use the moment generating function (MGF) to solve this. The MGF of a random variable Y_i is given by $M_{Y_i}(t) = \mathrm{E}(e^{tY_i})$. Furthermore, use the fact that $M_Y(t) = \prod_{i=1}^n M_{Y_i}(t)$.
Hint 2: If X and Z are two random variables such that $M_X(t) = M_Z(t)$, then X and Z have the same probability distribution.

Solution: Using the hint, the MGF for a Poisson r.v. is

$$\begin{aligned} M_{Y_i}(t) &= \mathrm{E}\left[\exp(tY_i)\right] \\ &= \sum_{k=0}^{\infty} e^{tk} \cdot \frac{e^{-\theta}\theta^k}{k!} \\ &= e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta e^t)^k}{k!} \\ &= e^{-\theta} e^{\theta e^t} \\ &= e^{\theta(e^t-1)} \end{aligned}$$

So,

$$M_Y(t) = \prod_{i=1}^n M_{Y_i}(t) = e^{n\theta(e^t-1)}$$

which implies that $Y \sim \mathrm{Pois}(n\theta)$. Therefore, $\mathrm{E}(Y) = \mathrm{Var}(Y) = n\theta$.

(b) Obtain the MLE for θ and its standard error.

Solution: The log-likelihood for θ from n observations of Y_i is

$$
\begin{aligned}
l(\theta|Y_1,\ldots,Y_n) &= \sum_{i=1}^{n} \log\left(\frac{e^{-\theta}\theta^{Y_i}}{Y_i!}\right) \\
&= \sum_{i=1}^{n} \{-\theta + Y_i \log\theta - \log Y_i!\} \\
&= \text{const.} - n\theta + \log\theta \sum_{i=1}^{n} Y_i
\end{aligned}
$$

Taking first derivatives wrt θ and equating to zero we get

$$
\begin{aligned}
l'(\theta) = -n + \frac{n\bar{Y}_n}{\theta} &= 0 \\
\Rightarrow \hat{\theta} &= \bar{Y}_n
\end{aligned}
$$

(c) Suppose now that only the first m ($m < n$) observations of the sample are known explicitly, while for the other $n-m$ only their sum is known, determine the MLE for θ.

Solution: Let $Z = Y_{m+1}, \ldots, Y_n$. From part (a) we know that $Z \sim \text{Pois}(\lambda)$, where $\lambda = (n-m)\theta$. The log-likelihood due to the known data is

$$
\begin{aligned}
l(\theta|Y_1,\ldots,Y_m,Z) &= \log\left\{\prod_{i=1}^{m} f(y_i|\theta) f(z|(n-m)\theta)\right\} \\
&= \text{const.} - m\theta + \log\theta \sum_{i=1}^{m} Y_i - (n-m)\theta + \log((n-m)\theta)Z
\end{aligned}
$$

Taking first derivatives wrt θ and equation to zero we get

$$
\begin{aligned}
l'(\theta) &= -m + \frac{\sum_{i=1}^{m} Y_i}{\theta} - n + m + \frac{Z}{\theta} \\
&= -n + \frac{Z + \sum_{i=1}^{m} Y_i}{\theta} = 0 \\
\Rightarrow \hat{\theta} &= \bar{Y}_n
\end{aligned}
$$

7. Find the MLE for λ given a random sample from the gamma distribution with pdf

$$
f(x|\lambda, r) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1},
$$

where r is a known constant.

Solution: The log-likelihood for λ is

$$\begin{aligned} l(\lambda|x_1,\ldots,x_n) &= \log \prod_{i=1}^{n} \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x_i} x_i^{r-1} \\ &= \text{const.} + nr\log\lambda - \lambda \sum_{i=1}^{n} x_i \end{aligned}$$

Therefore the MLE is obtained by differentiating and equation to zero the above log-likelihood, as follows:

$$\begin{aligned} l'(\lambda) &= \frac{nr}{\lambda} - \sum_{i=1}^{n} x_i = 0 \\ &\Rightarrow \hat{\lambda} = r/\bar{x} \end{aligned}$$

8. Find the MLE for θ from a random sample from the population with density function

$$f(y|\theta) = \frac{2y}{\theta^2}$$

where $0 < y \leq \theta$, and $\theta > 0$. **Do not use calculus.** Draw a picture of the likelihood function.

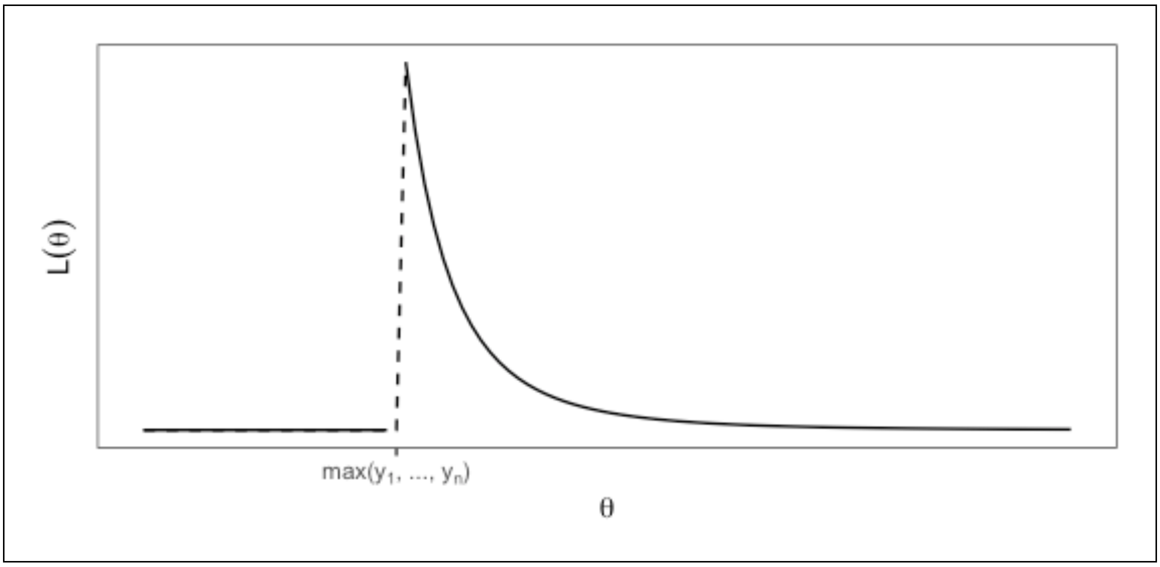
Solution: The key here is to realise that the value of θ determines the value of y_i that is observed, because the support of y is $(0, \theta]$. It is not possible to get values $y > \theta$ by definition. Therefore, the likelihood is

$$L(\theta|y_1,\ldots,y_n) = \begin{cases} \prod_{i=1}^{n} \frac{2y_i}{\theta^2} & \text{if } 0 < y_1,\ldots,y_n \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

In other words, the likelihood will take positive value in the case where $\theta > \max(y_1,\ldots,y_n)$, and zero otherwise. Therefore, $\hat{\theta} = \max(y_1,\ldots,y_n)$. We can write the likelihood as

$$L(\theta|y_1,\ldots,y_n) \propto \theta^{-2n}\, \mathbb{1}_{[\theta > \max(y_1,\ldots,y_n)]}(\theta)$$

which is plotted roughly as follows:



9. Let $X_1, \ldots, X_n$ be a sample from $\text{Unif}(0, \theta)$, where $\theta > 0$ is an unknown parameter. Find the MLE for θ. Derive the distribution for $\hat{\theta}$, and therefore, show that $\hat{\theta}$ is a consistent estimator in the sense that $\hat{\theta} \xrightarrow{\text{P}} \theta$ as $n \to \infty$. *Hint:* $\text{P}(\max_{1 \leq i \leq n} X_i \leq y) = \prod_{i=1}^n \text{P}(X_i \leq y)$.

Solution: This is similar to the previous question. Realise that the support of X_i is $(0, \theta)$, and therefore all values $X_1, \ldots, X_n$ must fall within $(0, \theta)$. This must mean that $\max(X_1, \ldots, X_n) \leq \theta$. In any case, the pdf of $X \sim \text{Unif}(0, \theta)$ is $f(x|\theta) = \frac{1}{\theta}$ if $x \in (0, \theta)$, and 0 otherwise. Thus,

$$
\begin{aligned}
L(\theta|X_1, \ldots, X_n) = \prod_{i=1}^n \frac{[X_i \in (0, \theta)]}{\theta} &= \theta^{-n}[X_1, \ldots, X_n \in (0, \theta)] \\
&= \theta^{-n}[\max(X_1, \ldots, X_n) \leq \theta]
\end{aligned}
$$

where $[\cdot]$ refers to the Iverson bracket (equals 1 if the event is achieved, and 0 otherwise). Again, by reasoning that the likelihood is positive when $\max(X_1, \ldots, X_n) \leq \theta$, it must be that the MLE is $\hat{\theta} = \max(X_1, \ldots, X_n)$.

To show that it is consistent, consider

$$
\begin{aligned}
\text{P}[\max(X_1, \ldots, X_n) \leq x] &= \prod_{i=1}^n \text{P}(X_i \leq x) \\
&= \begin{cases} 0 & \text{if } x < 0 \\ \left(\frac{x}{\theta}\right)^n & \text{if } x \in (0, \theta) \\ 1 & \text{if } x > \theta \end{cases}
\end{aligned}
$$

So for $\epsilon > 0$

$$\begin{aligned}
\mathrm{P}(|\hat{\theta} - \theta| > \epsilon) &= \mathrm{P}(\hat{\theta} - \theta < -\epsilon) + \mathrm{P}(\hat{\theta} - \theta > \epsilon) \\
&= \mathrm{P}(\hat{\theta} < \theta - \epsilon) + (1 - \cancelto{1}{\mathrm{P}(\hat{\theta} < \theta + \epsilon)}\ ) \\
&= \left(\frac{\theta - \epsilon}{\theta}\right)^n \to 0
\end{aligned}$$

as $n \to \infty$ (multiplying fraction which is less than one infinite number of times). Thus, $\hat{\theta} \xrightarrow{\mathrm{P}} \theta$, and is therefore consistent.

10. Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution, i.e.

$$X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1 - p \end{cases}$$

for $i = 1, \ldots, n$, where $p \in (0, 1)$ is unknown. Let $\theta = p^2$.

(a) Find the Cramér-Rao lower bound for the variance of unbiased estimators of θ.

Solution: First we find the Fisher information for p for a single observation. The derivative of the score function is

$$\begin{aligned}
S'(p) &= \frac{\partial^2}{\partial p^2}\left\{X_i \log p + (1 - X_i)\log(1 - p)\right\} \\
&= \frac{\partial}{\partial p}\left\{\frac{nX_i}{p} - \frac{1 - X_i}{1 - p}\right\} \\
&= -\frac{X_i}{p^2} + \frac{1 - X_i}{(1 - p)^2}
\end{aligned}$$

The Fisher information is then given by $\mathcal{I}(p) = -\mathrm{E}[S'(p)]$:

$$\begin{aligned}
\mathcal{I}_{X_i}(p) &= \mathrm{E}\left[\frac{X_i}{p^2} - \frac{1 - X_i}{(1 - p)^2}\right] \\
&= \frac{\mathrm{E}(X_i)}{p^2} - \frac{1 - \mathrm{E}(X_i)}{(1 - p)^2} \\
&= \frac{1}{p(1 - p)}
\end{aligned}$$

Of course, the full Fisher information is simply the sum of these unit Fisher information:

$$\mathcal{I}_{\mathbf{X}}(p) = \frac{n}{p(1 - p)}$$

Suppose $T = T(X_1, \ldots, X_n)$ is a statistic for which $\mathrm{E}(T) = g(p) = p^2 =: \theta$.

Now, the Cramér-Rao lower bound for $g(p) = p^2 =: \theta$ is

$$\begin{aligned}\mathrm{Var}(T) &\geq \frac{[g'(p)]^2}{\mathcal{I}_{\mathbf{X}}(p)} \\ &= \frac{4p^3(1-p)}{n}\end{aligned}$$

(b) Find the MLE $\hat{\theta}$ for θ.

Solution: We know that $\hat{p} = \bar{X}_n$ is the MLE for p so therefore by the invariance property of the MLE, $\hat{\theta} = \hat{p}^2 := (\bar{X}_n)^2$ is the MLE for $\theta = p^2$.

(c) Show that $\mathrm{E}(\hat{\theta}) \neq \theta$.

Solution:

$$\begin{aligned}\mathrm{E}(\hat{\theta}) = \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right] &= \frac{1}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i\neq j} X_i X_j\right] \\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n} \mathrm{E}(X_i^2) + \sum_{i\neq j} \mathrm{E}(X_i X_j)\right) \\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n} (p(1-p) + p^2) + \sum_{i\neq j} p^2\right) \\ &= \frac{1}{n^2}\left(np + n(n-1)p^2\right) \\ &= p^2 + \underbrace{\frac{p(1-p)}{n}}_{\text{bias}}\end{aligned}$$

In the above, we used the property that

$$\mathrm{E}(X_i^2) = \mathrm{Var}(X_i) + \mathrm{E}^2(X_i) = p(1-p) + p^2 = p$$

and also for two independent Bernoulli r.v. X_i and X_j, the product $X_i X_j$ has the distribution

$$X_i X_j = \begin{cases} 1 & \text{w.p. } p^2 \\ 0 & \text{w.p. } 1 - p^2 \end{cases}$$

Therefore, $\mathrm{E}(X_i X_j) = 1 \cdot p^2 + 0 \cdot (1 - p^2) = p^2$.

11. Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ be a sample from $\mathrm{N}(\mu, \sigma^2)$. Let $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$. Find the Fisher information matrix $\mathcal{I}_{\mathbf{X}}(\boldsymbol{\theta})$, i.e. the Fisher information using all n data points. *Hint:*

Use $\theta_2 = \sigma^2$ in your calculations.

Solution: Define $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$. The log-likelihood function for the normal distribution from a single observation is

$$l(\boldsymbol{\theta}) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\theta_2 - \frac{(X_i - \theta_1)^2}{2\theta_2}$$

The derivative w.r.t to θ_1 and θ_2 (which are the first and second components of the score function) are as follows:

$$\begin{aligned} S_1(\boldsymbol{\theta}) &:= \frac{\partial l(\boldsymbol{\theta})}{\partial\theta_1} = \frac{X_i - \theta_1}{\theta_2} \\ S_2(\boldsymbol{\theta}) &:= \frac{\partial l(\boldsymbol{\theta})}{\partial\theta_2} = -\frac{1}{2\theta_2} + \frac{(X_i - \theta_1)^2}{2\theta_2^2} \end{aligned}$$

Taking further derivatives gives us

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_1^2} &= -\frac{1}{\theta_2} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_2^2} &= \frac{1}{2\theta_2^2} - \frac{(X_i - \theta_1)^2}{\theta_2^3} \\ \frac{\partial^2 l(\boldsymbol{\theta})}{\partial\theta_1\partial\theta_2} &= -\frac{X_i - \theta_1}{\theta_2^2} \end{aligned}$$

Taking negative expectations of the above yield the entries of the unit Fisher information $\mathcal{I}_{X_i}(\boldsymbol{\theta})$:

$$\begin{aligned} \mathcal{I}(\boldsymbol{\theta})_{11} &= -\operatorname{E}(-1/\theta_2) = \theta_2^{-1} = \sigma^{-2} \\ \mathcal{I}(\boldsymbol{\theta})_{22} &= -\operatorname{E}\left[\frac{1}{2\theta_2^2} - \frac{(X_i - \theta_1)^2}{\theta_2^3}\right] = -\frac{1}{2\sigma^4} + \frac{1}{\sigma^4} = \frac{1}{2\sigma^4} \\ \mathcal{I}(\boldsymbol{\theta})_{12} = \mathcal{I}(\boldsymbol{\theta})_{21} &= -\operatorname{E}\left[-\frac{X_i - \theta_1}{\theta_2^2}\right] = 0 \end{aligned}$$

Therefore,

$$\mathcal{I}_{X_i}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & 1/2\sigma^4 \end{pmatrix}$$

and hence

$$\mathcal{I}_{\mathbf{X}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathcal{I}_{X_1}(\boldsymbol{\theta}) = \begin{pmatrix} n\sigma^{-2} & 0 \\ 0 & n/2\sigma^4 \end{pmatrix}$$