Mathematical Statistics Assignment 1

Liam Carroll - 830916

Due date: 24th April 2020

Q1. MGFs for $\sum \chi^2_{p_i}$

Let $X_1, \dots X_n$ be independent and $X_i \sim \chi^2_{p_i}$ for $i = 1, \dots, n$. Let $p = \sum_i p_i$. Consider the moment generating function (MGF) of X_i:

$$M_{X_i}(t) = \left(\frac{1}{1-2t}\right)^{p_i/2}$$

By properties of an MGF, we know that $M_{X+Y}(t) = M_X(t)M_Y(t)$, hence

$$\begin{aligned} M_{\sum_i X_i}(t) &= \prod_{i=1}^n M_{X_i}(t) \\ &= \prod_{i=1}^n \left(\frac{1}{1-2t}\right)^{p_i/2} \\ &= \left(\frac{1}{1-2t}\right)^{\sum_i p_i/2} \\ &= \left(\frac{1}{1-2t}\right)^{p/2} \end{aligned}$$

By the uniqueness of an MGF, we see that $M_{\sum_i X_i}(t) \sim M_{\chi^2_p}(t)$, hence we conclude that $\sum_{i=1}^n X_i \sim \chi^2_p(t)$ as required. $\square$

Q2. Minimising absolute error of a random sample

Let $x_1, \ldots, x_n$ be an observed sample. We wish to find the value of θ that minimises $S(\theta) = \sum_{i=1}^n |x_i - \theta|$. Firstly, without loss of generality, send $x_i \mapsto x_{(i)}$, its corresponding order statistic. Using the fact that $\frac{d}{dx}(|x|) = \text{sign}(x)$ for $x \neq 0$ (more on this assumption later), we can attempt to minimise $S(\theta)$ by finding its derivative

$$\begin{aligned}\frac{dS}{d\theta} &= -\sum_{i=1}^n \text{sign}(x_{(i)} - \theta) \\ &= -\sum_{i=1}^n \big(\mathbb{1}(x_{(i)} \geq \theta) - \mathbb{1}(x_{(i)} \leq \theta)\big)\end{aligned}$$

and then solving $\frac{dS}{d\theta} = 0$

$$\begin{aligned}\implies -\sum_{i=1}^n \big(\mathbb{1}(x_{(i)} \geq \theta) - \mathbb{1}(x_{(i)} \leq \theta)\big) = 0 \\ \implies \sum_{i=1}^n \mathbb{1}(x_{(i)} \leq \theta) = \sum_{i=1}^n \mathbb{1}(x_{(i)} \geq \theta)\end{aligned} \tag{2.1}$$

This suggests that θ must partition the ordered statistics $(x_{(1)}, \ldots, x_{(n)})$ such that the number of observed samples is the same "on both sides" of θ. We claim that $\hat{\theta} = \text{median}(\{x_{(i)}\}_{i=1}^n) = \frac{1}{2}(x_{(\lfloor \frac{n+1}{2} \rfloor)} + x_{(\lceil \frac{n+1}{2} \rceil)})$ is the appropriate minimser of θ. We will denote this as $\hat{\theta} = m_x$.

If n is even, then $\sum_{i=1}^n \mathbb{1}(x_{(i)} \leq m_x) = \frac{n}{2}$ and $\sum_{i=1}^n \mathbb{1}(x_{(i)} \geq m_x) = \frac{n}{2}$, hence m_x satisfies $\frac{dS}{d\theta}|_{\theta = m_x} = 0$. It is worth noting, however, that in the case of n being even, $\hat{\theta}$ need not be unique - indeed, any value $\theta \in (x_{(n/2)}, x_{(n/2+1)})$ would minimise $S(\theta)$.

If n is odd, then we must include the case where $\mathbb{1}(x_{(i)} = m_x)$ on both sides of our equality in (2.1). Hence, $\sum_{i=1}^n \mathbb{1}(x_{(i)} \leq m_x) = \frac{n}{2} + 1$ and $\sum_{i=1}^n \mathbb{1}(x_{(i)} \geq m_x) = \frac{n}{2} + 1$, hence m_x satisfies $\frac{dS}{d\theta}|_{\theta = m_x} = 0$ again as required.

We do notice that $\frac{d^2 S}{d\theta^2} = 0$, so this is not an appropriate measure of whether our claimed $\hat{\theta}$ is a minimum. Instead we notice that $\lim_{\theta \to -\infty} S(\theta) = \lim_{\theta \to \infty} S(\theta) = \infty$ which tells us that, since we have found a value of θ such that $\frac{dS}{d\theta}|_{\theta = m_x} = 0$, it must be a minimum. Thus, for all cases of n, $\hat{\theta} = m_x$ is the appropriate estimate of θ that minimses $S(\theta)$. $\square$

[N.B. It is worth pointing out that we stated that $|x|$ is not *differentiable at $x = 0$, however, since we are seeking to minimise $S(\theta)$, the contribution for $x_{(i)} = m_x$ to $S(\theta)$ is clearly 0 (i.e. $|m_x - m_x| = 0$). Hence we can assume without loss of generality that it is fine (!) to define the sign(x) function as the derivative of $|x|$ for algebraic purposes. (We have also taken the standard definition of sign(x) here where* $\text{sign}(0) = 0$*).]*

Q3. MME estimator for Gamma distribution

Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} \text{Gamma}(\lambda, r)$, $\lambda > 0$ and $r > 0$. Using the definition of the Gamma distribution as in the question, the MGF for this distribution is

$$M_{X_i}(t) = \left(\frac{1}{1 - t/\lambda}\right)^r$$

We can easily show by induction that the n^{th} derivative of $M_{X_i}(t)$ is

$$M_{X_i}^{(n)}(t) = \left(\frac{1}{\lambda}\right)^n r \dots (r + (n-1)) \left(\frac{1}{1 - t/\lambda}\right)^{r+n}$$

We then appeal to the fact that $\mu_n = \mathbb{E}[X^n] = M_X^{(n)}(0)$ to derive the first and second moments of the Gamma function.

$$\mu_1 = \frac{r}{\lambda} \qquad \mu_2 = \frac{r(r+1)}{\lambda^2}$$

Let $m_1 = \frac{1}{n}\sum_i X_i$ and $m_2 = \frac{1}{n}\sum_i X_i^2$ be the first and second sample moments respectively. Equating $\mu_1 = m_1$ and $\mu_2 = m_2$ and rearranging gives us $\frac{1}{\lambda} = \frac{m_1}{r}$. Substituting this into the equation for m_2 gives

$$\begin{aligned} m_2 &= \left(\frac{m_1}{r}\right)^2 r(r+1) \\ &= \frac{m_1^2(r+1)}{r} \\ \implies m_2 r &= m_1^2(r+1) \\ \implies r &= \frac{m_1^2}{m_2 - m_1^2} \\ &= \frac{\bar{X}_n^2}{\sigma_n^2} \end{aligned} \qquad \begin{aligned} \lambda &= \frac{r}{m_1} \\ &= \frac{\bar{X}_n}{\sigma_n^2} \end{aligned}$$

Hence, with $\bar{X}_n$ and σ_n^2 being the sample mean and sample (unbiased) variance, we see that the MME estimates for λ and r are

$$\tilde{\lambda} = \frac{\bar{X}_n}{\sigma_n^2} \qquad \tilde{r} = \frac{\bar{X}_n^2}{\sigma_n^2}$$

□

Q4. MLE of multi-mean normal distribution

Let $X_{i,j}$ for $i = 1, \dots, m$ and $j = 1, \dots, n$ be independently distributed as $N(\mu_i, \sigma^2)$. We wish to calculate the MLE of $\boldsymbol{\theta} = (\mu_1, \dots, \mu_m, \sigma^2)^T$. We can calculate the likelihood function $L(\boldsymbol{\theta})$ as follows

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= f(\mathbf{x}|\boldsymbol{\theta}) \\
&= \prod_{i=1}^{m} \prod_{j=1}^{n} f(x_{i,j}|\boldsymbol{\theta}) \\
&= \prod_{i=1}^{m} \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_{i,j}-\mu_i)^2}{2\sigma^2}} \\
&= \prod_{i=1}^{m} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_{i,j}-\mu_i)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{n+m} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-\mu_i)^2} \\
\implies \log L(\boldsymbol{\theta}) &= -\frac{(n+m)}{2}\log(2\pi) - \frac{(n+m)}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-\mu_i)^2 \\
&= -\frac{(n+m)}{2}\log(2\pi) - \frac{(n+m)}{2}\log(\sigma^2) \\
&\quad - \frac{1}{2\sigma^2}\left(\sum_{j=1}^{n}(x_{1,j}-\mu_1)^2 + \cdots + \sum_{j=1}^{n}(x_{m,j}-\mu_m)^2 \right)
\end{aligned}
$$

Setting derivatives equal to 0 for a fixed i value we get

$$
\frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu_i} = \frac{1}{\sigma^2}\sum_{j=1}^{n}(x_{i,j}-\mu_i) = 0 \quad \frac{\partial \log L(\boldsymbol{\theta})}{\partial(\sigma^2)} = -\frac{(n+m)}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-\mu_i)^2 = 0
$$

$$
\implies \hat{\mu_i} = \frac{1}{n}\sum_{j=1}^{n} x_{i,j} = (\bar{X}_n)_i \qquad \implies (\hat{\sigma^2}) = \frac{1}{n+m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j} - (\bar{X}_n)_i)^2
$$

To form the Hessian matrix to determine if these are indeed local maxima, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m, \theta_{m+1})^T = (\mu_1, \dots, \mu_m, \sigma^2)^T$ we first calculate the necessary derivatives.

$$
\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu_k \partial \mu_i} = -\frac{n}{\sigma^2}\delta_{ik} \qquad \frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial(\sigma^2)\partial \mu_i} = -\frac{1}{\sigma^4}\sum_{j=1}^{n}(x_{i,j}-\mu_i)
$$

$$
\therefore \left.\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu_k \partial \mu_i}\right|_{\boldsymbol{\theta}=\hat{\theta}} = -\frac{n}{(\hat{\sigma^2})}\delta_{ik} \qquad \therefore \left.\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial(\sigma^2)\partial \mu_i}\right|_{\boldsymbol{\theta}=\hat{\theta}} = 0
$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial(\sigma^2)^2} = \frac{(n+m)}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-\mu_i)^2$$

$$\therefore \left.\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial(\sigma^2)^2}\right|_{\boldsymbol{\theta}=\hat{\theta}} = \frac{(n+m)}{2(\hat{\sigma^2})^2} - \frac{(n+m)(\hat{\sigma^2})}{(\hat{\sigma^2})^3}$$

$$= -\frac{(n+m)}{2(\hat{\sigma^2})^2}$$

Which leads us to the following Hessian matrix

$$H = \begin{pmatrix} -\frac{n}{(\hat{\sigma^2})} & 0 & \cdots & \cdots & 0 \\ 0 & -\frac{n}{(\hat{\sigma^2})} & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \cdots & -\frac{n}{(\hat{\sigma^2})} & 0 \\ 0 & \cdots & \cdots & 0 & -\frac{(n+m)}{2(\hat{\sigma^2})^2} \end{pmatrix}$$

A diagonal matrix is negative definite if and only if all of its entries are negative. Clearly, since $n, m > 0$ and $(\hat{\sigma^2}) > 0$, we see that all entries of the diagonal matrix H are indeed negative and hence H is negative definite as required. Hence the MLE estimate of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = (\hat{\mu_1}, \ldots, \hat{\mu_m}, \hat{\sigma^2})^T = \left((\bar{X}_n)_1, \ldots, (\bar{X}_n)_m, \frac{1}{n+m}\sum_{i=1}^{m}\sum_{j=1}^{n}(x_{i,j}-(\bar{X}_n)_i)^2\right)$$

$\square$

Q5. MLE of shifted exponential distribution

Let $X_1, \dots, X_n$ be a random sample from a population with pdf

$$f(x|\theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & \text{otherwise} \end{cases} = e^{-(x-\theta)}\mathbb{1}(x \geq \theta)$$

We wish to maximise the likelihood function $L(\theta) = f(x_1, \dots, x_n|\theta)$. Since the X_i's are independent (as they are drawn from a random sample from a population), this is the product of their individual pdf's. Since n is finite, we can also map each $x_i \mapsto x_{(k)}$, its corresponding order statistic.

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} f(x_i|\theta) \\ &= \prod_{i=1}^{n} e^{-(x_i-\theta)}\mathbb{1}(x_i \geq \theta) \\ &= \prod_{k=1}^{n} e^{-(x_{(k)}-\theta)}\mathbb{1}(x_{(k)} \geq \theta) \\ &= e^{-\sum_{k=1}^{n}(x_{(k)}-\theta)}\mathbb{1}(x_{(1)} \geq \theta)\dots\mathbb{1}(x_{(n)} \geq \theta) \end{aligned}$$

Since the $x_{(i)}$'s are ordered, we see that $\mathbb{1}(x_{(1)} \geq \theta)\dots\mathbb{1}(x_{(n)} \geq \theta) = \mathbb{1}(x_{(1)} \geq \theta)$. Hence:

$$L(\theta) = e^{n\theta}e^{-n\bar{X}_n}\mathbb{1}(x_{(1)} \geq \theta)$$

Since $L(\theta)$ is positive and monotonically increasing in θ for $\theta \leq x_{(1)}$, we see that $L(\theta)$ is maximised at $\theta = x_{(1)} = \min(X_1, \dots, X_n)$. Hence the MLE of θ is
$\hat{\theta} = \min(X_1, \dots, X_n)$. $\square$

Q6. Comparison of estimators for mean of Normal

Let $X_i \sim N(\mu, \sigma_i^2)$, where σ_i^2 are known and positive for $i = 1, \ldots, n$ and $X_1, \ldots, X_n$ are independent. Let $\hat{\mu} = \frac{\sum_{i=1}^n (X_i/\sigma_i^2)}{\sum_{i=1}^n (1/\sigma_i^2)}$ be the MLE of μ.

Part a)

Since σ_i^2 are known, we can treat both it and $\phi = \sum_{i=1}^n (1/\sigma_i^2)$ as a fixed scalar allowing us to move it outside of the $\mathbb{E}$ brackets. Hence,

$$\begin{aligned}
\mathbb{E}[\hat{\mu}] &= \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i/\sigma_i^2)}{\phi}\right] \\
&= \frac{1}{\phi}\mathbb{E}\left[\sum_{i=1}^n (X_i/\sigma_i^2)\right] \\
&= \frac{1}{\phi}\sum_{i=1}^n \mathbb{E}\left[\frac{X_i}{\sigma_i^2}\right] \\
&= \frac{1}{\phi}\sum_{i=1}^n \frac{1}{\sigma_i^2}\mathbb{E}[X_i] \\
&= \frac{1}{\phi}\phi\mu = \mu
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\hat{\mu}^2] &= \mathbb{E}\left[\left(\frac{\sum_{i=1}^n (X_i/\sigma_i^2)}{\phi}\right)^2\right] \\
&= \frac{1}{\phi^2}\mathbb{E}\left[\left(\sum_{i=1}^n (X_i/\sigma_i^2)\right)^2\right] \\
&= \frac{1}{\phi^2}\mathbb{E}\left[\sum_{i=1}^n \frac{X_i^2}{(\sigma_i^2)^2} + 2\sum_{j=1}^n\sum_{k=1}^{j-1}\frac{X_jX_k}{\sigma_j^2\sigma_k^2}\right] \\
&= \frac{1}{\phi^2}\left(\sum_{i=1}^n \frac{\mathbb{E}[X_i^2]}{(\sigma_i^2)^2} + 2\sum_{j=1}^n\sum_{k=1}^{j-1}\frac{\mathbb{E}[X_jX_k]}{\sigma_j^2\sigma_k^2}\right) \\
&= \frac{1}{\phi^2}\left(\sum_{i=1}^n \frac{\sigma_i^2 + \mu^2}{(\sigma_i^2)^2} + 2\sum_{j=1}^n\sum_{k=1}^{j-1}\frac{\mu^2}{\sigma_j^2\sigma_k^2}\right) \\
&= \frac{1}{\phi^2}\left(\sum_{i=1}^n \frac{1}{\sigma_i^2} + \mu^2\left(\sum_{i=1}^n \frac{1}{(\sigma_i^2)^2} + 2\sum_{j=1}^n\sum_{k=1}^{j-1}\frac{1}{\sigma_j^2\sigma_k^2}\right)\right) \\
&= \frac{1}{\phi^2}\left(\phi + \mu^2\left(\sum_{i=1}^n \frac{1}{\sigma_i^2}\right)^2\right) \\
&= \frac{1}{\phi^2}(\phi + \mu^2\phi^2) = \frac{1}{\phi} + \mu^2
\end{aligned}$$

$$\therefore \mathrm{Var}(\hat{\mu}) = \mathbb{E}[\hat{\mu}^2] - \mathbb{E}[\hat{\mu}]^2 = \frac{1}{\phi} + \mu^2 - \mu^2 = \frac{1}{\phi}$$

Where we appeal to the fact that $\mathbb{E}[X_jX_k] = \mathbb{E}[X_j]\mathbb{E}[X_k] = \mu^2$ for $j \neq k$ since the X_i's are independent. Also, we use the fact that $\mathbb{E}[X_i^2] = \sigma_i^2 + \mu^2$. We notice that since $\mathbb{E}[\hat{\mu}] = \mu$ we have $\mathrm{Bias}_\mu(\hat{\mu}) = 0$ so $\hat{\mu}$ is an unbiased estimator of μ.

Part b)

First we do some trivial calculations for $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Since the X_i's are independent, $\text{Var}(\sum X_i) = \sum \text{Var}(X_i)$.

$$\begin{aligned}\mathbb{E}[\bar{X}_n] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] \\ &= \frac{1}{n}n\mu = \mu\end{aligned} \qquad \begin{aligned}\text{Var}(\bar{X}_n) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n \text{Var}(X_i) \\ &= \sum_{i=1}^n \frac{\sigma_i^2}{n^2}\end{aligned}$$

Clearly again $\bar{X}_n$ is an unbiased estimator of μ, meaning we can compare the relative efficiency of our two unbiased estimators $\hat{\mu}$ and $\bar{X}_n$ since $\text{MSE}(\hat{\mu}) = \text{Var}(\hat{\mu})$ and $\text{MSE}(\bar{X}_n) = \text{Var}(\bar{X}_n)$.

$$RE_\mu(\hat{\mu}, \bar{X}_n) = \frac{\text{Var}(\bar{X}_n)}{\text{Var}(\hat{\mu})} = \sum_{i=1}^n \frac{\sigma_i^2}{n^2}\sum_{j=1}^n \frac{1}{\sigma_j^2} = \left(\sum_{i=1}^n \frac{\sigma_i^2}{n}\right)\left(\sum_{j=1}^n \frac{1}{n}\frac{1}{\sigma_j^2}\right)$$

We then appeal to the Chebyshev sum inequality which states that for sequences a_i and b_j such that $a_1 \leq \cdots \leq a_n$ and $b_1 \geq \cdots \geq b_n$ then

$$\left(\sum_{i=1}^n \frac{a_i}{n}\right)\left(\sum_{j=1}^n \frac{b_j}{n}\right) \geq \frac{1}{n}\sum_{i=1}^n a_i b_i$$

Without loss of generality, rearrange the sequence of fixed σ_i^2's (i.e. consider this a_i) so that they are ordered, hence $\sigma_1^2 \leq \cdots \leq \sigma_n^2$. Hence the sequence $b_j = \frac{1}{\sigma_j^2}$ satisfies $b_1 \geq \cdots \geq b_n$. Thus, we can conclude that

$$\begin{aligned}RE_\mu(\hat{\mu}, \bar{X}_n) &= \left(\sum_{i=1}^n \frac{\sigma_i^2}{n}\right)\left(\sum_{j=1}^n \frac{1}{n}\frac{1}{\sigma_j^2}\right) \\ &\geq \frac{1}{n}\sum_{i=1}^n \sigma_i^2 \frac{1}{\sigma_i^2} = \frac{1}{n}n = 1\end{aligned}$$

Therefore, since $RE_\mu(\hat{\mu}, \bar{X}_n) \geq 1$ (i.e. $\text{Var}(\bar{X}_n) \geq \text{Var}(\hat{\mu})$), we can conclude that $\hat{\mu}$ is a better estimator of μ than $\bar{X}_n$. $\square$

Q7. Location-scale and Exponential Family of transformed Gamma random variable

Let X be a random variable such that $X \sim \text{Gamma}(\gamma, \alpha)$ (shape-scale parameterisation) with pdf

$$f_X(x) = \frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-\frac{1}{\gamma}x} \mathbb{1}(x > 0)$$

Here we have that α is known and γ is unknown. Let $Y = \sigma \log(X)$. Then we can establish the cdf of Y:

$$\begin{aligned}
F_Y(y) = P(Y \leq y) &= P(\sigma \log(X) \leq y) \\
&= P(\log(X) \leq \frac{y}{\sigma}) \\
&= P(X \leq e^{\frac{y}{\sigma}}) \\
&= F_X(e^{\frac{y}{\sigma}}) \\
&= \int_0^{e^{\frac{y}{\sigma}}} \frac{1}{\Gamma(\alpha)\gamma^\alpha} x^{\alpha-1} e^{-\frac{1}{\gamma}x} dx
\end{aligned}$$

Hence, we can find the pdf of Y:

$$\begin{aligned}
f_Y(y) = \frac{d}{dy} F_Y(y) &= \frac{d}{dy} F_X(e^{\frac{y}{\sigma}}) \\
&= \frac{1}{\sigma} e^{\frac{y}{\sigma}} f_X(e^{\frac{y}{\sigma}}) \\
&= \frac{1}{\sigma\Gamma(\alpha)\gamma^\alpha} e^{\frac{y}{\sigma}} (e^{\frac{y}{\sigma}})^{\alpha-1} e^{-\frac{1}{\gamma} e^{\frac{y}{\sigma}}} \mathbb{1}(e^{\frac{y}{\sigma}} > 0) \\
&= \frac{1}{\sigma\Gamma(\alpha)\gamma^\alpha} e^{(\frac{\alpha}{\sigma} y - \frac{1}{\gamma} e^{\frac{y}{\sigma}})} \\
&= \frac{1}{\sigma\Gamma(\alpha)\gamma^\alpha} e^{\left(\frac{\alpha}{\sigma} y - e^{\left(\frac{y - \sigma \log \gamma}{\sigma}\right)}\right)}
\end{aligned}$$

where the support set of Y is $y \in (-\infty, \infty)$.

Part a)

Let $\sigma > 0$ be unknown. To show Y is in a location-scale family, we want to show that for $\mu \in (-\infty, \infty)$, $\beta > 0$, we can write $f(y) = \frac{1}{\beta} g(\frac{y-\mu}{\beta})$ (i.e. $g(y) = \beta f(\beta y + \mu)$) for a well defined pdf $g(y)$. Let $\beta = \sigma$ and $\mu = \sigma \log \gamma$. Then:

$$\begin{aligned}
g(y) = \beta f(\beta y + \mu) &= \sigma \frac{1}{\sigma\Gamma(\alpha)\gamma^\alpha} e^{\left(\frac{\alpha}{\sigma}(\sigma y + \sigma \log \gamma) - e^{\left(\frac{\sigma y + \sigma \log \gamma - \sigma \log \gamma}{\sigma}\right)}\right)} \\
&= \frac{e^{\alpha \log \gamma}}{\Gamma(\alpha)\gamma^\alpha} e^{(\alpha y - e^y)} \\
&= \frac{1}{\Gamma(\alpha)} e^{\alpha y} e^{-e^y}
\end{aligned}$$

We can now verify that $g(y)$ is a pdf by first noticing that $g(y) \geq 0 \;\; \forall y \in (-\infty, \infty)$, and then ensuring that $\int_{-\infty}^{\infty} g(y)dy = 1$, where we make the substitution $u = e^y$

$$\begin{aligned}
\int_{-\infty}^{\infty} g(y)dy &= \int_{-\infty}^{\infty} \frac{1}{\Gamma(\alpha)} e^{\alpha y} e^{-e^y} dy \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} u^{\alpha} e^{-u} u^{-1} du \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} u^{\alpha-1} e^{-u} du \\
&= \frac{1}{\Gamma(\alpha)} \Gamma(\alpha) = 1
\end{aligned}$$

Therefore, we can see that $g(y)$ is a well defined pdf. Hence, this verifies that under these conditions, Y is in a location-scale family. $\square$

Part b)

Let $\sigma > 0$ be known. Y is in an exponential family if we can write

$$f(y|\boldsymbol{\theta}) = c(\boldsymbol{\theta})h(y) \exp\left\{\sum_{i=1}^{k} w_i(\boldsymbol{\theta})t_i(y)\right\}$$

where $c(\boldsymbol{\theta}) \geq 0$, $w_i(\boldsymbol{\theta})$, $h(y) \geq 0$ are all real valued functions, where $\boldsymbol{\theta} = (\alpha, \gamma, \sigma)$.

Since σ is known, we can replace $z = y/\sigma$. Then:

$$f(z|\boldsymbol{\theta}) = \frac{1}{\sigma\Gamma(\alpha)\gamma^{\alpha}} e^{\left(\alpha z - \frac{1}{\gamma} e^z\right)}$$

So we can see this satisfies our requirements with:

$$\begin{aligned}
c(\boldsymbol{\theta}) &= \frac{1}{\sigma\Gamma(\alpha)\gamma^{\alpha}} \\
h(z) &= 1 \\
(w_1(\boldsymbol{\theta}), w_2(\boldsymbol{\theta})) &= \left(\alpha, -\frac{1}{\gamma}\right) \\
(t_1(z), t_2(z)) &= (z, e^z)
\end{aligned}$$

Thus under these conditions, Y is in an exponential family. $\square$

Q8. Sufficient statistic for $\frac{1}{x^2}$ distribution

Let $X_1, \dots, X_n$ be a random sample from a population with pdf (with $\theta > 0$)

$$f(x|\theta) = \begin{cases} \frac{\theta}{x^2} & x \geq \theta \\ 0 & \text{otherwise} \end{cases} = \frac{\theta}{x^2}\mathbb{1}(x \geq \theta)$$

We can also calculate (for $z > \theta$, 1 otherwise)

$$\begin{aligned} P(X > z) &= \int_z^\infty \frac{\theta}{x^2} dx \\ &= \left[-\frac{\theta}{x}\right]_z^\infty = \frac{\theta}{z} \end{aligned}$$

Part a)

We can attempt to find the method of moments estimator for θ, however we will soon establish that these moments do not exist. We can calculate the moments of X quite easily for $n \in \mathbb{N}_{\geq 1}$

$$\begin{aligned} \mathbb{E}[X^n] &= \int_\theta^\infty \theta x^{n-2} dx \\ &= \begin{cases} [\theta \log(x)]_\theta^\infty & n = 1 \\ [\theta \frac{1}{n-1} x^{n-1}]_\theta^\infty & n = 2, 3, \dots \end{cases} \\ &= \infty \ \text{ for } \ n \in \mathbb{N}_{\geq 1} \end{aligned}$$

Thus since the moments of X don't exist, we cannot calculate the method of moments estimator for θ. [sad :( ]

Part b)

The likelihood function for X is

$$\begin{aligned} L(\theta) = f(\mathbf{x}|\theta) &= \prod_{i=1}^n \frac{\theta}{x_i^2}\mathbb{1}(x_i \geq \theta) \\ &= \theta^n \mathbb{1}(x_{(1)} \geq \theta) \prod_{i=1}^n \frac{1}{x_i^2} \end{aligned}$$

Since $L(\theta)$ is positive and monotonically increasing in θ for $\theta \leq x_{(1)}$ (given that $\theta > 0$), we see that $L(\theta)$ is maximised at $\theta = x_{(1)} = \min(X_1, \dots, X_n)$. Hence the MLE of θ is $\hat{\theta} = \min(X_1, \dots, X_n)$.

We can then calculate

$$\begin{aligned} P(\hat{\theta} > z) &= P(\min(X_1, \dots, X_n) > z) \\ &= P(X_1 > z, \dots, X_n > z) \\ &= P(X_1 > z) \dots P(X_n > z) \\ &= \left(\frac{\theta}{z}\right)^n \end{aligned}$$

Thus the cdf of $\hat{\theta}$ is

$$F_{\hat{\theta}}(z) = 1 - P(\hat{\theta} > z) = 1 - \left(\frac{\theta}{z}\right)^n$$

and so the pdf is

$$f_{\hat{\theta}}(z) = \frac{d}{dz} F_{\hat{\theta}}(z) = \frac{n\theta^n}{z^{n+1}}$$

Part c)

We wish to find a sufficient statistic for θ. We return to the joint pdf

$$f(\mathbf{x}|\theta) = \theta^n \mathbb{1}(x_{(1)} \geq \theta) \prod_{i=1}^{n} \frac{1}{x_i^2}$$

We claim that $T(\mathbf{X}) = x_{(1)}$ is a sufficient statistic for θ. This is clear to see since we can write

$$f(\mathbf{x}|\theta) = \underbrace{\theta^n \mathbb{1}(x_{(1)} \geq \theta)}_{g(T(\mathbf{x})|\theta)} \underbrace{\prod_{i=1}^{n} \frac{1}{x_i^2}}_{h(\mathbf{x})}$$

Thus, by the factorisation theorem, we see that $T(\mathbf{X}) = x_{(1)}$ is a sufficient statistic for θ. $\square$

Q9. Sufficient Statistic for Multivariate Normal

Let $\mathbf{x}_1, \ldots \mathbf{x}_n = \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ be a random sample from a two-dimensional multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}\right)$ where $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_{11}, \sigma_{12}, \sigma_{22} \in \mathbb{R}^+$, where all parameters are unknown and $\det(\boldsymbol{\Sigma}) > 0$. To find a sufficient statistic for $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_{11}, \sigma_{12}, \sigma_{22})$, we first consider the joint pdf

$$\begin{aligned} f(\mathbf{x}|\boldsymbol{\theta}) = f(\mathbf{x}_1, \ldots, \mathbf{x}_n|\boldsymbol{\theta}) &= \prod_{i=1}^{n} \frac{\exp\left[-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\right]}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}} \\ &= \left(\frac{1}{(2\pi)^k \det(\boldsymbol{\Sigma})}\right)^{\frac{n}{2}} \prod_{i=1}^{n} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\right] \\ &= \left(\frac{1}{(2\pi)^k \det(\boldsymbol{\Sigma})}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu})\right] \end{aligned} \tag{9.1}$$

We then expand the the inside of the exponent as follows, where we write the statistic $\mathbf{T}_1(\mathbf{x}_i) = \bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$ and we make use of the fact that $(\boldsymbol{\Sigma}^{-1})^T = (\boldsymbol{\Sigma}^T)^{-1} = \boldsymbol{\Sigma}^{-1}$, and $(ABC)^T = C^T B^T A^T$

$$\begin{aligned} \sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \boldsymbol{\mu}) &= \sum_{i=1}^{n}((\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \boldsymbol{\mu}))^T \boldsymbol{\Sigma}^{-1} ((\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \boldsymbol{\mu})) \\ &= \sum_{i=1}^{n} \left[(\mathbf{x}_i - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right] \\ &\quad + \sum_{i=1}^{n} \left[(\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + (\mathbf{x}_i - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right] \\ &= \sum_{i=1}^{n} \left[(\mathbf{x}_i - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right] \\ &\quad + \sum_{i=1}^{n} \left[2(\mathbf{x}_i - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\right] \end{aligned} \tag{9.2}$$

However, we then notice that

$$\begin{aligned} \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) &= \sum_{i=1}^{n} \left(\mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}} - \bar{\mathbf{x}} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}} + \bar{\mathbf{x}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - \mathbf{x}_i^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}\right) \\ &= n\bar{\mathbf{x}} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}} - n\bar{\mathbf{x}} \boldsymbol{\Sigma}^{-1} \bar{\mathbf{x}} + n\bar{\mathbf{x}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} - n\bar{\mathbf{x}} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \\ &= 0 \end{aligned}$$

Hence, the last term in (9.2) vanishes.

Thus, we can now write (9.1) as

$$
\begin{aligned}
f(\mathbf{x}|\boldsymbol{\theta}) &= \left(\frac{1}{(2\pi)^k \det(\boldsymbol{\Sigma})}\right)^{\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right] \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})\right] \\
&= \left(\frac{1}{(2\pi)^k \det(\boldsymbol{\Sigma})}\right)^{\frac{n}{2}} \exp\left[-\frac{n}{2}(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right] \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})\right]
\end{aligned} \tag{9.3}
$$

Now, to deal with the quantity in the right-hand exponential, we notice that $(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})$ is a 1×1 matrix, hence we can use a trick with the trace (for which we know it obeys the cyclicity property, i.e. $\mathrm{tr}(ABC)=\mathrm{tr}(BCA)=\mathrm{tr}(CAB)$ [and clearly $\mathrm{tr}(A+B)=\mathrm{tr}(A)+\mathrm{tr}(B)$]) to turn this term into

$$
\begin{aligned}
\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) &= \mathrm{tr}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) \\
&= \sum_{i=1}^{n}\mathrm{tr}(\mathbf{x}_i-\bar{\mathbf{x}})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}}) \\
&= \sum_{i=1}^{n}\mathrm{tr}\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T \\
&= \mathrm{tr}\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T
\end{aligned}
$$

We can now write down the statistic

$$
\mathbf{T}_2(\mathbf{x}_i) = \hat{\boldsymbol{\Sigma}} = \sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})^T
$$

Then we can rewrite (9.3) as

$$
f(\mathbf{x}|\boldsymbol{\theta}) = \underbrace{\left(\frac{1}{(2\pi)^k \det(\boldsymbol{\Sigma})}\right)^{\frac{n}{2}} \exp\left[-\frac{n}{2}(\bar{\mathbf{x}}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})\right] \exp\left[-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\Sigma}}\right]}_{g((\mathbf{T}_1(\mathbf{x}_i),\mathbf{T}_2(\mathbf{x}_i))|\boldsymbol{\theta})} \cdot \underbrace{1}_{h(\mathbf{x}_i)}
$$

Thus, by the factorisation theorem, we have found sufficient statistics for the multivariate normal distribution, namely $\bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}}$. It is worth pointing out that due to the more sophisticated matrix calculations we have used throughout that this has method has found sufficient statistics for a multivariate normal distribution of any N size. It is easy and tedious to express our final function in terms of the $\boldsymbol{\theta}$ provided by the question - we shall leave this as an exercise for the reader. □

Q10. Incompleteness of bound statistics for a uniform distribution

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Uniform}(\theta, \theta+1)$, $\theta \in \mathbb{R}$. We wish to show that the minimal sufficient statistic $T = (X_{(1)}, X_{(n)})$ is *not* complete for $A = \{f(\mathbf{x}|\theta) : \theta \in \mathbb{R}\}$, that is, there exists a function g such that $\mathbb{E}_\theta[g(T)] = 0 \nRightarrow P_\theta(g(T) = 0) = 1$

We appeal to the range statistic $R(T) = X_{(n)} - X_{(1)}$. From Example 6.2.17 of Casella and Berger, we know that $R(T)$ is an ancillary statistic - that is, the distribution of R, which is $h(r|\theta) = n(n-1)r^{n-2}(1-r)\mathbb{1}(0 < r < 1)$, does not depend on the parameter θ. This means that $\mathbb{E}_\theta[R(T)] = k$ for some $k \in \mathbb{R}$, which does not depend on θ. Thus, we choose our g to be $g(T) = X_{(n)} - X_{(1)} - k$. Then clearly $\mathbb{E}_\theta[g(T)] = 0$. However, $P_\theta(g(T) = 0) = P_\theta(R(T) = k) = 0 \neq 1$ since $h(r|\theta)$ is a continuous distribution. Therefore we conclude that T is *not* a complete statistic for A. $\square$

Q11. Minimal sufficiency for a scaled-shifted exponential

Let $X_1, \ldots, X_n$ be a random sample from a population with pdf

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}e^{-\frac{x-\theta}{\theta}} & x \geq \theta \\ 0 & \text{otherwise} \end{cases} = \frac{1}{\theta}e^{-\frac{x-\theta}{\theta}}\mathbb{1}(x \geq \theta)$$

where $\theta > 0$.

Part a)

We first consider the joint pdf

$$\begin{aligned} f(\mathbf{x}|\theta) &= \prod_{i=1}^{n} \frac{1}{\theta}e^{-\frac{x_i-\theta}{\theta}}\mathbb{1}(x_i \geq \theta) \\ &= \frac{1}{\theta^n}\mathbb{1}(x_{(1)} \geq \theta)e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}e^n \\ &= \underbrace{\frac{1}{\theta^n}\mathbb{1}(x_{(1)} \geq \theta)e^{-\frac{n\bar{x}}{\theta}}}_{g(T(\boldsymbol{x})|\theta)}\underbrace{e^n}_{h(\boldsymbol{x})} \end{aligned}$$

We claim that $T = (x_{(1)}, \bar{x})$ is a minimal sufficient statistic for θ. By the factorisation theorem, it is clear that T is sufficient for θ. We want to show that the ratio $f(x|\theta)/f(y|\theta)$ is constant as a function of θ if and only if $T(x) = T(y)$ to prove that it is minimal sufficient. That $T(x) = T(y)$ implies the ratio is constant is trivial to show. Suppose that the ratio is constant, say K. Then

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{\theta^{-n}\mathbb{1}(x_{(1)} \geq \theta)e^{-\frac{n\bar{x}}{\theta}}e^n}{\theta^{-n}\mathbb{1}(y_{(1)} \geq \theta)e^{-\frac{n\bar{y}}{\theta}}e^n} = \frac{\mathbb{1}(x_{(1)} \geq \theta)}{\mathbb{1}(y_{(1)} \geq \theta)}e^{-\frac{n}{\theta}(\bar{x}-\bar{y})} = K$$

We first observe that since K is independent of θ, this implies $\lim_{x_{(1)},y(1)\to\theta} \frac{\mathbb{1}(x_{(1)}\geq\theta)}{\mathbb{1}(y_{(1)}\geq\theta)}$ must be 1. Thus, $\mathbb{1}(x_{(1)} \geq \theta) = \mathbb{1}(y_{(1)} \geq \theta)$ and hence $x_{(1)} = y_{(1)}$.

This then suggests that

$$e^{-\frac{n}{\theta}(\bar{x}-\bar{y})} = K$$

But since this must be true *for all* θ, this implies that $\bar{x} - \bar{y} = 0$ and hence $\bar{x} = \bar{y}$. Hence, we see that $f(x|\theta)/f(y|\theta)$ is constant as a function of θ if and only if $T(x) = T(y)$ and so T is a minimal sufficient statistic for θ. □

Part b)

It appears natural to believe that $X \sim f(x|\theta)$ is in an exponential family. However, we know from lectures that if X is a random variable from an exponential family, then the support set of X does *not* depend on on the parameter θ. Clearly, the support set of $f(x|\theta)$ depends on θ since $f(x|\theta) = 0$ if $x \leq \theta$. Therefore, we conclude that X is *not* in an exponential family. □

Q12. UMVUE of mean of a Normal

Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(\mu, 1), \mu \in \mathbb{R}$.

Part a)

We want to calculate the UMVUE of μ^2 and calculate its variance. We will take it as granted that $T_1 = \bar{X}_n$ is a complete and sufficient statistic for μ. Clearly $T_1 \overset{i.i.d}{\sim} N(\mu, 1/n)$. We can then consider

$$\begin{aligned}
&\mathbb{E}_\mu[T_1^2] = \mu^2 + 1/n \\
\implies& \mu^2 = \mathbb{E}_\mu[T_1^2] - 1/n \\
\implies& \mu^2 = \mathbb{E}_\mu[T_1^2 - 1/n]
\end{aligned}$$

Clearly then, if we let $T_2 = T_1^2 - 1/n$ then $\text{Bias}(T_2) = 0$. By the Lehmann-Scheffé Theorem, since T_1 is a complete and sufficient statistic and $T_2 = T_2(T_1)$ is an unbiased estimator of μ^2, then T_2 is is the UMVUE of μ^2.

We can then calculate the variance, where we use standard moment of normal results.

$$\begin{aligned}
\mathbb{E}_\mu[T_2^2] &= \mathbb{E}_\mu[(T_1^2 - 1/n)^2] \\
&= \mathbb{E}_\mu\left[T_1^4 - \frac{2}{n}T_1^2 + \frac{1}{n^2}\right] \\
&= \mathbb{E}_\mu[T_1^4] - \frac{2}{n}\mathbb{E}_\mu[T_1^2] + \frac{1}{n^2} \\
&= \mu^4 + \frac{6}{n}\mu^2 + \frac{3}{n^2} - \frac{2}{n}(\mu^2 + 1/n) + \frac{1}{n^2} \\
&= \mu^4 + \frac{4}{n}\mu^2 + \frac{2}{n^2}
\end{aligned}$$

$$\therefore \text{Var}(T_2) = \mathbb{E}_\mu[T_2^2] - \mathbb{E}_\mu[T_2]^2 = \mu^4 + \frac{4}{n}\mu^2 + \frac{2}{n^2} - \mu^4 = \frac{4\mu^2}{n} + \frac{2}{n^2}$$

Part b)

Since $X_1, \dots, X_n$ are drawn from a normal distribution, it is clear that $f(x|\theta)$, T_2, $\gamma(\mu) = \mu^2$ all satisfy the necessary conditions to use the Cramér-Rao Inequality (supposing $I_n(\theta)$ is finite which we will show below). We then calculate the necessary quantities

$$\begin{aligned}
I_n(\mu) = nI_1(\mu) &= -n\mathbb{E}_\mu\left[\frac{\partial^2}{\partial\mu^2}\log f(X_1|\mu)\right] \qquad\qquad \gamma'(\mu) = 2\mu \\
&= -n\mathbb{E}_\mu[-1] = n
\end{aligned}$$

Which means that $\text{CRLB}(T_2) = \frac{(\gamma'(\mu))^2}{I_n(\mu)} = \frac{4\mu^2}{n}$. Hence, as expected from the Cramér-Rao Inequality, we have

$$\text{Var}(T_2) = \frac{4\mu^2}{n} + \frac{2}{n^2} \geq \frac{4\mu^2}{n} = \text{CRLB}(T_2)$$

□

Q13. UMVUE of p^2 of Bernoulli

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Bernoulli}(p),\ p \in (0,1)$ and require that $n > 2$. We wish to find the UMVUE of $\gamma(p) = p^2$.

We will take for granted that $T = \sum_{i=1}^n X_i$ is a complete and sufficient statistic for p. We are interested in calculating $p^2 = P(X_1 = 1, X_2 = 1)$. We can then define an unbiased estimator for $\gamma(p)$ as

$$T_0 = \begin{cases} 1 & \text{if } X_1, X_2 = 1 \\ 0 & \text{otherwise} \end{cases}$$

By construction, we have $\mathbb{E}[T_0] = P(X_1 = 1, X_2 = 1) = p^2$ so T_0 is unbiased for $\gamma(p)$. Now we can define $T_1 = \mathbb{E}[T_0|T]$. Then

$$\begin{aligned}
\mathbb{E}[T_0|T = t] &= \mathbb{E}\left[T_0 \middle| \sum_{i=1}^n X_i = t\right] \\
&= P\left(X_1 = 1, X_2 = 1 \middle| \sum_{i=1}^n X_i = t\right) \\
&= \frac{P\left(X_1 = 1, X_2 = 1, \sum_{i=1}^n X_i = t\right)}{P\left(\sum_{i=1}^n X_i = t\right)} \\
&= \frac{P\left(X_1 = 1, X_2 = 1, \sum_{i=3}^n X_i = t-2\right)}{P\left(\sum_{i=1}^n X_i = t\right)} \\
&= \frac{P(X_1 = 1)P(X_2 = 1)P(\sum_{i=3}^n X_i = t-2)}{P\left(\sum_{i=1}^n X_i = t\right)} \mathbb{1}(t \geq 2) \\
&= \frac{p^2 \binom{n-2}{t-2} p^{t-2}(1-p)^{n-t}}{\binom{n}{t} p^t (1-p)^{n-t}} \mathbb{1}(t \geq 2) \\
&= \frac{\binom{n-2}{t-2}}{\binom{n}{t}} \mathbb{1}(t \geq 2) \\
&= \frac{(n-2)!\ t!}{(t-2)!\ n!} \mathbb{1}(t \geq 2)
\end{aligned}$$

Thus we see that in defining

$$T_1 = \mathbb{E}\left[T_0 \middle| \sum_{i=1}^n X_i\right] = \frac{(n-2)!(\sum_{i=1}^n X_i)!}{(\sum_{i=1}^n X_i - 2)!n!} \mathbb{1}\left(\sum_{i=1}^n X_i \geq 2\right)$$

by the Rao-Blackwell theorem we know that T_1 is unbiased, and by Lehmann-Scheffé we know that it is the UMVUE for p^2. $\square$