Problem 1

Modelling service times is one area of research in queuing theory. We are given a data set with service times (in minutes) for 174 customers. The distribution of service times is modelled using the one parameter Weibull distribution, which has the pdf

$$f(x) = \frac{2x}{\theta} e^{-x^2/\theta} I_{(0,\infty)}(x). \tag{1}$$

c) The method of moments estimator $\hat{\theta}_1$ of θ is $\hat{\theta}_1 = \frac{4(\bar{X})^2}{\pi}$. The reasoning is as follows. The Weibull distribution with parameters k, λ has the pdf

$$f(y) = \frac{k}{\lambda}\left(\frac{y}{\lambda}\right)^{k-1} e^{-(y/\lambda)^k} I_{(0,\infty)}(y).$$

The expected value of a Weibull distributed random variable Y is $\mathrm{E}Y = \lambda\Gamma(1+\frac{1}{k})$. Note that setting $\lambda = \sqrt{\theta}$ and $k = 2$ gives the pdf for the one parameter Weibull distribution (1). Thus, we have that

$$\mathrm{E}X = \sqrt{\theta}\Gamma\left(1+\frac{1}{2}\right) = \sqrt{\theta}\frac{\sqrt{\pi}}{2}.$$

Solving this equation for θ yields $\theta = \frac{4(\mathrm{E}X)^2}{\pi}$. Thus, the method of moments estimator is $\hat{\theta}_1 = \frac{4(\bar{X})^2}{\pi}$.

d) The asymptotic distribution of $\sqrt{n}(\hat{\theta}-\theta)$ is normal. This can be inferred directly from proposition 2 in the notes for Unit 2 (page 15). To make this situation fit into the proposition, define $h(\mu_1) = \theta = \frac{4(\mu_1^2)}{\pi}$. Since h is continuous, then $\hat{\theta} = h(m_{1,n}) = \frac{4(\bar{X}_n)^2}{\pi}$ converges to θ almost surely (which follows from the strong law of large numbers and the continuity property). Moreover, since h is differentiable in μ_1, the limiting distribution of $\hat{\theta}$ is normal, i.e.

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{h,\theta}).$$

Note that the random variable X fulfills the property that $\mathrm{E}(X^2) < \infty$, which is an assumption made in the proposition. This result is obtained in the next part of the problem. Next we can use the delta method to find the variance $\sigma^2_{h,\theta}$. First of all, as a consequence of the central limit theorem, we have that

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{2}\sqrt{\theta\pi}\right) \xrightarrow{\mathcal{D}} N\left(0, \theta\left(1-\frac{\pi}{4}\right)\right),$$

where the variance of X has been calculated in part e). Knowing this, the delta method gives

$$\sqrt{n}\left(\frac{4(\bar{X})^2}{\pi} - \frac{4\left(\frac{1}{2}\sqrt{\theta\pi}\right)^2}{\pi}\right) \xrightarrow{\mathcal{D}} N\left(0, \left(h'\left(\frac{1}{2}\sqrt{\theta\pi}\right)\right)^2 \theta\left(1-\frac{\pi}{4}\right)\right)$$
$$\implies \sqrt{n}(\hat{\theta}-\theta) \xrightarrow{\mathcal{D}} N\left(0, \left(\frac{8}{\pi}\left(\frac{1}{2}\sqrt{\theta\pi}\right)\right)^2 \theta\left(1-\frac{\pi}{4}\right)\right)$$
$$\implies \sqrt{n}(\hat{\theta}-\theta) \xrightarrow{\mathcal{D}} N\left(0, 16\theta^2\left(\frac{1}{\pi}-\frac{1}{4}\right)\right),$$

which means that $\sigma^2_{h,\theta} \approx 16\theta^2\left(\frac{1}{\pi}-\frac{1}{4}\right)$. This fact will be used to argue in part f) also.

e) The maximum likelihood estimator of θ is $\hat{\theta}_2 = \sum_{i=1}^n X_i^2 = \bar{X}^2$. The reasoning is as follows. The likelihood function is

$$L(\theta;x) = \prod_{i=1}^n \frac{2x_i}{\theta} e^{-x_i^2/\theta} I_{(0,\infty)}(x_i) = \left(\frac{2}{\theta}\right)^n \prod_{i=1}^n x_i \cdot e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^2} \prod_{i=1}^n I_{(0,\infty)}(x_i).$$

The log-likelihood function is

$$l(\theta;x) = \log L(\theta;x) = n\log 2 - n\log\theta + \sum_{i=1}^n \log x_i - \frac{1}{\theta}\sum_{i=1}^n x_i^2,$$

for $X_{(1)} > 0$. The first derivative of the log-likelihood function is

$$\frac{\partial}{\partial\theta} l(\theta;x) = -\frac{n}{\theta} + \frac{\sum_{i=1}^n x_i^2}{\theta^2},$$

which, by equating to zero yields a stationary point at $\theta^* = \frac{1}{n}\sum_{i=1}^n x_i^2 = \bar{x}^2$. The second derivative of the log-likelihood function is

$$\frac{\partial^2}{\partial\theta^2} l(\theta;x) = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n x_i^2}{\theta^3},$$

which, when inserting θ^* gives

$$\frac{\partial^2}{\partial\theta^2} l(\theta^*;x) = \frac{n}{(\bar{x}^2)^2} - \frac{2n}{(\bar{x}^2)^2} = -\frac{n}{(\bar{x}^2)^2} < 0,$$

which shows that θ^* is a maximum of $l(\theta;x)$, and indeed that $\hat{\theta}_2 = \bar{X}^2$. Next we need to show that this is also a method of moments estimator. First, note that the variance of X is

$$\text{Var}X = \theta\left(\Gamma\left(1+\frac{2}{2}\right) - \frac{1}{\theta}(\text{E}X)^2\right) = \theta\left(1-\frac{\pi}{4}\right),$$

since it has a Weibull distribution.

We solve the following system of equations

$$\begin{aligned} m_1 &= \text{E}X = \sqrt{\theta}\frac{\sqrt{\pi}}{2}, \\ m_2 - m_1^2 &= \text{Var}X = \theta\left(1-\frac{\pi}{4}\right), \end{aligned}$$

where m_i, $i = 1, 2$ are the first two moments of X. The system has the solution

$$m_2 = \theta\frac{\pi}{4} + \theta\left(1-\frac{\pi}{4}\right) = \theta,$$

which means that the method of moment estimator is $\hat{\theta}_{MME} = \hat{m}_2 = \widehat{\text{E}X^2} = \bar{X}^2 = \hat{\theta}_2$.

f) Which of these estimators is preferable? At first glance I would prefer $\hat{\theta}_2 = \bar{X}^2$ because it is calculated based on the sufficient (and complete) statistic for this one parameter Weibull distribution. This is the case because this distribution belongs to the exponential family.

Note that since both are method of moments estimators, they are both consistent. What about efficiency?

As already noted, the distribution of X belongs to the exponential family, which means that the MLE is the UMVUE and coincides with the estimator whose variance attains the Cramér-Rao Lower Bound ($CRLB$). Since $\hat{\theta}_2$ is the MLE in this case, I would still argue that this is preferable, since it is consistent, but also asymptotically efficient. It can be shown that $\hat{\theta}_1$ is not asymptotically efficient, which means that it is at a disadvantage in this regard. In the following, non-efficiency of $\hat{\theta}_1$ will be shown for completeness of this argument. First of all, we calculate the expected value of $\bar{X}^2$

$$\begin{aligned} \text{E}\bar{X}^2 &= \text{Var}\bar{X} + (\text{E}\bar{X})^2 \\ &\overset{i.i.d.}{=} \frac{1}{n^2}\sum \text{Var}X + \left(\frac{1}{n}\sum \text{E}X\right)^2 \\ &= \frac{1}{n}\text{Var}X + (\text{E}X)^2 \\ &= \frac{1}{n}\theta(1-\frac{\pi}{4}) + \frac{\theta\pi}{4}. \end{aligned}$$

Thus, we get that

$$\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} l(\theta;x)\right) = \frac{n}{\theta^2} - \frac{2n\mathrm{E}\bar{X}^2}{\theta^3} = \frac{n - 2 + \frac{\pi}{2} - \frac{\pi n}{2}}{\theta^2}.$$

This means that

$$CRLB = \frac{1}{-\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} l(\theta;x)\right)} = \left(\frac{2 - \frac{\pi}{2} + n(\frac{\pi}{2} - 1)}{\theta^2}\right)^{-1}.$$

This leads to

$$\lim_{n\to\infty} \frac{\mathrm{Var}(\hat{\theta}_1)}{CRLB} = \lim_{n\to\infty} \left(\frac{2 - \frac{\pi}{2} + n(\frac{\pi}{2} - 1)}{\theta^2}\right)\left(16\theta^2\left(\frac{1}{\pi} - \frac{1}{4}\right)\right) \longrightarrow \infty,$$

which means that $\hat{\theta}_1$ is not asymptotically efficient.

Problem 3

Suppose that $X_1, \ldots, X_n$ are i.i.d. Normal random variables with mean θ and variance θ^2 where $\theta > 0$. Define

$$\hat{\theta}_n = \bar{X}_n \left(1 + \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2 - n\bar{X}_n^2}{3\sum_{i=1}^n (X_i - \bar{X}_n)^2}\right).$$

a) Show that $\hat{\theta}_n \xrightarrow{\mathcal{P}} \theta$ as $n \longrightarrow \infty$.

Proof. The definition states that $\hat{\theta}_n$ converges in probability to θ if, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \geq \epsilon) = 0.$$

Looking at $P(|\hat{\theta}_n - \theta| \geq \epsilon) = P(|\hat{\theta}_n - \theta|^2 \geq \epsilon^2)$, Chebyshev's inequality yields

$$P(|\hat{\theta}_n - \theta|^2 \geq \epsilon^2) \leq \frac{\mathrm{E}[|\hat{\theta}_n - \theta|^2]}{\epsilon^2} = \frac{1}{\epsilon^2}(\mathrm{Var}_\theta[\hat{\theta}_n] + (\mathrm{B}_\theta(\hat{\theta}_n))^2),$$

where the last equality is attributed to the bias-variance decomposition of the mean squared error. This is a rather mechanical way of proving consistency of the estimator. Instead of using this, let us use the known fact that a method of moments estimator is consistent. Define $\hat{\theta}_n$ as a function of the moments, as done in the following

$$\begin{aligned}
\hat{\theta}_n &= \bar{X}_n \left(1 + \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2 - n\bar{X}_n^2}{3\sum_{i=1}^n (X_i - \bar{X}_n)^2}\right) = \bar{X}_n \left(1 + \frac{\sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) - n\bar{X}_n^2}{3\sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + (\bar{X}_n)^2)}\right) \\
&= \bar{X}_n \left(1 + \frac{\sum_{i=1}^n X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2 - n\bar{X}_n^2}{3(\sum_{i=1}^n X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2)}\right) = m_{1.n}\left(1 + \frac{nm_{2,n} - 2nm_{1,n}^2}{3nm_{2,n} - 3nm_{1,n}^2}\right) \\
&= m_{1.n}\left(1 + \frac{1}{3}\frac{m_{2,n} - 2m_{1,n}^2}{m_{2,n} - m_{1,n}^2}\right) =: h(m_{1,n}, m_{2,n}),
\end{aligned}$$

where the raw sample moments $m_{1,n} = \frac{1}{n}\sum_{i=1}^n X_i = \bar{X}_n$ and $m_{2,n} = \frac{1}{n}\sum_{i=1}^n X_i^2$ have been defined. A large sample property of the method of moments estimator is that

$$h(m_{1,n}, m_{2,n}) \xrightarrow{a.s.} h(\mu_1, \mu_2),$$

if $\mathrm{E}(X^4) < \infty$ and h is a continuous function of the moments. For this normal distribution, it can be shown that $\mathrm{E}(X^4) = 10\theta^4 < \infty$. Moreover, the function h is a continuous function of the moments because

$$m_{2,n} - m_{1,n}^2 = \frac{(n-1)S_n^2}{n} = \frac{\theta^2}{n}\frac{(n-1)S_n^2}{\theta^2}.$$

Since $\frac{(n-1)S_n^2}{\theta^2} \sim \chi_{n-1}^2$, $Y := \frac{n}{\theta^2}(m_{2,n} - m_{1,n}^2)$ has a continuous distribution which will not take the value 0, since $P(Y = 0) = 0$ in the continuous case. Notice that

$$h(\mu_1, \mu_2) = \mu_1\left(1 + \frac{1}{3}\frac{\mu_2 - \mu_1^2 - \mu_1^2}{\mu_2 - \mu_1^2}\right) = \theta\left(1 + \frac{1}{3}\frac{\theta^2 - \theta^2}{\theta^2}\right) = \theta.$$

Thus, we have shown that $\hat{\theta}_n \xrightarrow{\mathcal{P}} \theta$ as $n \longrightarrow \infty$.

□

b) The Cramér-Rao Lower Bound ($CRLB$) for unbiased estimators of θ, assuming all regularity conditions are satisfied, is

$$CRLB = \frac{1}{-\mathrm{E}(\frac{\partial^2}{\partial\theta^2}\log L(\theta; x))},$$

where $L(\theta; x)$ is the likelihood function. Thus,

$$L(\theta; x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\theta} e^{-\frac{1}{2\theta^2}(x_i - \theta)^2} = (2\pi\theta^2)^{-n/2} e^{-\frac{1}{2\theta^2}\sum_{i=1}^n (x_i - \theta)^2},$$

which yields

$$\log L(\theta; x) = -\frac{n}{2}\log 2\pi\theta^2 - \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i - \theta)^2.$$

The first derivative of the log-likelihood is

$$\frac{\partial}{\partial\theta}\log L(\theta; x) = -\frac{n}{\theta} + \frac{1}{\theta^3}\sum_{i=1}^{n}(x_i - \theta)^2 + \frac{1}{\theta^2}\sum_{i=1}^{n}(x_i - \theta),$$

and the second derivative is

$$\begin{aligned}
\frac{\partial^2}{\partial\theta^2}\log L(\theta; x) &= \frac{n}{\theta^2} + \left(-\frac{3}{\theta^4}\sum_{i=1}^{n}(x_i - \theta)^2 - \frac{2}{\theta^3}\sum_{i=1}^{n}(x_i - \theta)\right) + \left(-\frac{2}{\theta^3}\sum_{i=1}^{n}(x_i - \theta) - \frac{n}{\theta^2}\right) \\
&= \frac{n}{\theta^2} - \frac{3}{\theta^4}\sum_{i=1}^{n}(x_i - \theta)^2 - \frac{4}{\theta^3}\sum_{i=1}^{n}(x_i - \theta) - \frac{n}{\theta^2} \\
&= -\frac{3\sum_{i=1}^{n}(x_i - \theta)^2 + 4\theta\sum_{i=1}^{n}(x_i - \theta)}{\theta^4}
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\log L(\theta; x)\right) &= -\frac{1}{\theta^4}\left(3\sum_{i=1}^{n}\mathrm{E}((x_i - \theta)^2) + 4\theta\sum_{i=1}^{n}\mathrm{E}(x_i - \theta)\right) \\
&= -\frac{1}{\theta^4}\left(3\sum_{i=1}^{n}(\mathrm{E}(x_i^2) + \theta^2 - 2\theta\mathrm{E}(x_i)) + 4\theta\sum_{i=1}^{n}(\mathrm{E}(x_i) - \theta)\right) \\
&= -\frac{1}{\theta^4}\left(3\sum_{i=1}^{n}(\mathrm{Var}(x_i) + \mathrm{E}(x_i)^2 + \theta^2 - 2\theta\mathrm{E}(x_i))\right) \\
&= -\frac{1}{\theta^4}\left(3\sum_{i=1}^{n}(\theta^2 + \theta^2 + \theta^2 - 2\theta^2)\right) \\
&= -\frac{3n\theta^2}{\theta^4} \\
&= -\frac{3n}{\theta^2}.
\end{aligned}$$

Finally, this means that

$$CRLB = \frac{1}{-\mathrm{E}(\frac{\partial^2}{\partial\theta^2}\log L(\theta; x))} = \frac{\theta^2}{3n}.$$

c) Following the argument in a), if h is differentiable in (μ_1, μ_2), the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ is $N(0, \sigma^2_{h,\theta})$.

The estimator $\hat{\theta}_n$ of θ is asymptoptically efficient if

$$\lim_{n\to\infty} \frac{\text{Var}_\theta(\hat{\theta}_n)}{CRLB} = 1.$$

But what is the value of $\text{Var}_\theta(\hat{\theta}_n) = \sigma^2_{h,\theta}$ in this case? Using the delta method to calculate this variance yields $(h'(\theta))^2\theta^2 = \theta^2$, which means that

$$\lim_{n\to\infty} \frac{\text{Var}_\theta(\hat{\theta}_n)}{CRLB} = \lim_{n\to\infty} \frac{\theta^2}{\frac{\theta^2}{3n}} = \lim_{n\to\infty} 3n \longrightarrow \infty.$$

Thus, $\hat{\theta}_n$ is not asymptotically efficient.