Student Name: **Luke Nguyen**
Student ID: **D5850A**

Statistical Methods and Data Analysis (EN.625.603)
Problem Set 5

Question 5.2.15
The exponential pdf is a measure of lifetimes of devices that do not age. However, the exponential pdf is a special case of the *Weibull distribution*, which can measure time to failure of devices where the probability of failure increases as time does. A Weibull random variable Y has pdf $f_Y(y;\alpha,\beta) = \alpha\beta y^{\beta-1}e^{-\alpha y^{\beta}}, 0 \leq y, 0 < \alpha, 0 < \beta$.
Find the maximum likelihood estimator for α assuming that β is known.
Solution
Applying **Definition 5.2.1** as follows:

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^{n} f_Y(y_i;\alpha) \\ &= \prod_{i=1}^{n} \alpha\beta y_i^{\beta-1}e^{-\alpha y_i^{\beta}} \\ &= \alpha^n\beta^n \prod_{i=1}^{n} y_i^{\beta-1}e^{-\alpha y_i^{\beta}} \\ \implies \ln[L(\alpha)] &= n\ln\alpha + n\ln\beta + \sum_{i=1}^{n}(\beta-1)\ln y_i - \alpha\sum_{i=1}^{n} y_i^{\beta} \\ \implies \frac{\partial \ln[L(\alpha)]}{\partial\alpha} &= \frac{n}{\alpha} - \sum_{i=1}^{n} y_i^{\beta} = 0 \\ \implies \hat{\alpha} &= \boldsymbol{\frac{n}{\sum_{i=1}^{n} y_i^{\beta}}} \end{aligned}$$

Question 5.4.4
A sample size $n = 16$ is drawn from a normal distribution where $\sigma = 10$ but μ is unknown. If $\mu = 20$, what is the probability that the estimator $\hat{\mu} = \bar{Y}$ will lie between 19.0 and 21.0?
Solution
Applying **Collary 4.3.1**, the probability that the estimator $\hat{\mu} = \bar{Y}$ will lie between 19.0 and 21.0 can be calculated as follows

$$\begin{aligned} P(19.0 \leq \bar{Y} \leq 21.0) &= P(\bar{Y} \leq 21.0) - P(\bar{Y} \leq 19.0) \\ &= P\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq \frac{21.0-\mu}{\sigma/\sqrt{n}}\right) - P\left(\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \leq \frac{19.0-\mu}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{21.0-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{19.0-\mu}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{21.0-20}{10/\sqrt{16}}\right) - \Phi\left(\frac{19.0-20}{10/\sqrt{16}}\right) \end{aligned}$$

$$
\begin{aligned}
&= \Phi(0.4) - \Phi(-0.4) \\
&= 0.6554 - 0.3446 \\
&= \mathbf{0.3108}
\end{aligned}
$$

Question 5.4.16
Is the maximum likelihood estimator for σ^2 in a normal pdf, where both μ and σ^2 are unknown, asymptotically unbiased?
Solution
Suppose that $Y_1, Y_2, \ldots, Y_n$ are a random sample from a normal pdf with mean μ and variance σ^2. Using the result from **Example 5.2.5**, we know that the maximum likelihood estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

With the same terminologies and concepts, we find the expected value of the estimator for σ^2 as follows

$$
\begin{aligned}
E[\hat{\sigma}^2] &= E\left[\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}(Y_i - \bar{Y})^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}(Y_i^2 - 2Y_i\bar{Y} + \bar{Y}^2)\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}Y_i^2 - \sum_{i=1}^{n}2Y_i\bar{Y} + \sum_{i=1}^{n}\bar{Y}^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}Y_i^2 - 2\bar{Y}\sum_{i=1}^{n}Y_i + n\bar{Y}^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}Y_i^2 - n\bar{Y}^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}E\left[Y_i^2\right] - E\left[\bar{Y}^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - \left(\sigma^2 + \frac{\mu^2}{n}\right) \\
&= \frac{n-1}{n}\sigma^2 \qquad (1)
\end{aligned}
$$

Also,

$$
\begin{aligned}
\lim_{n\to\infty} E[\hat{\sigma}^2] &= \lim_{n\to\infty}\frac{n-1}{n}\sigma^2 \\
&= \sigma^2 \qquad (2)
\end{aligned}
$$

From (1) and (2), we can conclude that the maximum likelihood estimator for σ^2 is asymptotically unbiased by applying **Definition 5.4.1**.

Question 5.5.2
Let $X_1, X_2, ..., X_n$ be random sample of size n from the Poisson distribution, $p_k(k;\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}, k = 0, 1, 2, ...$ Show that $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$ is an efficient estimator for λ.
Solution
First, we have

$$\begin{aligned}
\text{Var}(\hat{\lambda}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \\
&= \frac{1}{n^2}\text{Var}(\lambda) \\
&= \frac{1}{n^2}\sum_{i=1}^n \lambda \\
&= \frac{\lambda}{n} \qquad (1)
\end{aligned}$$

As illustrated in **Example 5.5.1**, we start with

$$\begin{aligned}
\ln p_{X_i}(X_i;\lambda) &= -\lambda + X_i \ln\lambda - \ln X_i! \\
\frac{\partial \ln p_{X_i}(X_i;\lambda)}{\partial\lambda} &= -1 + \frac{X_i}{\lambda} \\
\frac{\partial^2 \ln p_{X_i}(X_i;\lambda)}{\partial\lambda^2} &= -\frac{X_i}{\lambda^2}
\end{aligned}$$

Using the above calculation and applying **Theorem 5.5.1**, we calculate The Cramer-Rao lower bound as follows

$$\begin{aligned}
\text{Var}(\hat{\lambda}) &\geq \left\{-nE\left[\frac{\partial^2 \ln p_X(X;\lambda)}{\partial\lambda^2}\right]\right\}^{-1} \\
&\geq \left\{-nE\left[\frac{\partial^2 \ln\left(\frac{e^{-\lambda}\lambda^k}{k!}\right)}{\partial\lambda^2}\right]\right\}^{-1} \\
&\geq \left\{-nE\left[-\frac{X_i}{\lambda^2}\right]\right\}^{-1} \\
&\geq \left\{-n\left[-\frac{\lambda}{\lambda^2}\right]\right\}^{-1} \\
&\geq \frac{\lambda}{n} \qquad (2)
\end{aligned}$$

From (1), (2), and **Definition 5.5.2**, we can conclude that $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$ is an efficient estimator for λ.

Question 5.6.1
Let $X_1, X_2, ..., X_n$ be a random sample of size n from the geometric distribution, $p_k(k;p) = p(1-p)^{k-1}, k = 1, 2, ...$ Show that $\hat{p} = \sum_{i=1}^n X_i$ is sufficient for p.

Solution
We check the likelihood function as follows

$$
\begin{aligned}
L(p) &= \prod_{i=1}^{n} p(1-p)^{k_i-1} \\
&= p^n(1-p)^{\sum_{i=1}^{n} k_i - n} \\
&= p_{\hat{p}}(h(k_1, ..., k_n); p)
\end{aligned}
$$

Which implies that $\hat{p} = \sum_{i=1}^{n} X_i$ is sufficient for p by **Definition 5.6.1**.

Question 5.7.1
How large a sample must be taken from a normal pdf where $E(Y) = 18$ in order to guarantee that $\hat{\mu}_n = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ has a 90% probability of lying somewhere in the interval $[16, 20]$? Assume that $\sigma = 5.0$.
Solution
We can calculate the value of n as follows

$$
\begin{aligned}
P\left(\frac{16-\mu}{\sigma/\sqrt{n}} \le \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \le \frac{20-\mu}{\sigma/\sqrt{n}}\right) &= 0.90 \\
P\left(\frac{16-18}{5/\sqrt{n}} \le \frac{\bar{Y}-18}{5/\sqrt{n}} \le \frac{20-18}{5/\sqrt{n}}\right) &= 0.90 \\
P\left(\frac{-2}{5/\sqrt{n}} \le \frac{\bar{Y}-18}{5/\sqrt{n}} \le \frac{2}{5/\sqrt{n}}\right) &= 0.90 \\
P\left(\frac{-2}{5/\sqrt{n}} \le Z \le \frac{2}{5/\sqrt{n}}\right) &= 0.90 \\
P\left(Z \le \frac{2}{5/\sqrt{n}}\right) - P\left(Z \le \frac{-2}{5/\sqrt{n}}\right) &= 0.90 \\
\Phi\left(\frac{2}{5/\sqrt{n}}\right) - \Phi\left(\frac{-2}{5/\sqrt{n}}\right) &= 0.90 \\
\frac{2}{5/\sqrt{n}} &\approx 1.65 \\
n &\approx \mathbf{17.02}
\end{aligned}
$$