1. Let $X_1, \ldots, X_{n_1}$ be a random sample of size $n_1$ from $N(\mu_1, \sigma^2)$.

   (a) Find a constant $c$ such that $c\sum_{i=1}^{n_1-1}(X_{i+1} - X_i)^2$ is an unbiased estimator of $\sigma^2$.

   **Solution**: We have

$$\begin{aligned} E\left\{c\sum_{i=1}^{n_1-1}(X_{i+1} - X_i)^2\right\} &= c\sum_{i=1}^{n_1-1} E(X_{i+1}^2 - 2X_{i+1}X_i + X_i^2) \\ &= c\sum_{i=1}^{n_1-1}(\sigma^2 + \mu^2 - 2\mu^2 + \sigma^2 + \mu^2) \\ &= 2c(n_1 - 1)\sigma^2. \end{aligned}$$

   To make the estimator unbiased, one can let

$$c = \frac{1}{2(n_1 - 1)}.$$

   (b) Let $Y_1, \ldots, Y_{n_2}$ be another random sample of size $n_2$ from $N(\mu_2, \sigma^2)$. Assuming $X$ and $Y$ are mutually independent, show that $S_p^2 = aS_1^2 + (1-a)S_2^2$ is an unbiased estimator of $\sigma^2$, where

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2 \quad \text{and} \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2,$$

   with a constant $a \in [0, 1]$.

   **Solution**: We know that $(n_1 - 1)S_1^2/\sigma^2 \sim \chi^2_{n_1-1}$, $(n_2 - 1)S_2^2/\sigma^2 \sim \chi^2_{n_2-1}$, and $E(S_1^2) = E(S_2^2) = \sigma^2$. Hence,

$$E(S_p^2) = aE(S_1^2) + (1-a)E(S_2^2) = a\sigma^2 + (1-a)\sigma^2 = \sigma^2,$$

   regardless what the value $a$ is.

   (c) Find $a$ such that $\text{var}(S_p^2)$ is minimized and the estimator becomes

$$S_p^2 = \frac{1}{n_1 + n_2 - 2}\left\{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i - \bar{Y})^2\right\}.$$

   [Hint]: Show that $\text{var}(S_p^2) = g(a)\sigma^4$ and find $a$ such that $g'(a) = 0$ and $g''(a) > 0$.

**Solution**: We have

$$\mathrm{var}(S_p^2) = a^2 \frac{2\sigma^4}{n_1 - 1} + (1-a)^2 \frac{2\sigma^4}{n_2 - 1}.$$

Let

$$g(a) = \frac{a^2}{n_1 - 1} + \frac{(1-a)^2}{n_2 - 1}.$$

One can derive $a = (n_1 - 1)/(n_1 + n_2 - 2)$ by letting $g'(a) = 0$. The resulting estimator is

$$S_p^2 = \frac{1}{n_1 + n_2 - 2} \left\{ \sum_{i=1}^{n_1} (X_i - \bar{X})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2 \right\}.$$

2. For women in a certain high-risk population, suppose that the number of lifetime events of domestic violence involving emergency room treatments is assumed to have a Poisson distribution

$$f_X(x|\theta) = \theta^x e^{-\theta}/x!, \quad x = 0, 1, \ldots, \quad \theta > 0.$$

Let $X_1, \ldots, X_n$ be a random sample chosen for the high-risk population.

(a) Find the maximum likelihood estimator (MLE) of the parameter $\tau(\theta)$, where

$$\tau(\theta) = P(X_1 = 0) = e^{-\theta}.$$

**Solution**: The MLE of $\theta$ is $\bar{X}$. By the invariance property of MLE, the MLE of $\tau(\theta)$ is $\tau(\theta) = e^{-\bar{X}}$.

(b) Find Crámer-Rao Lower Bound (CRLB) for every unbiased estimator of the parameter $\tau(\theta)$.

**Solution**: One have $\tau(\theta) = e^{-\theta}$ and $d\tau(\theta)/d\theta = -e^{-\theta}$. One also have

$$E\left\{ -\frac{\partial^2}{\partial \theta^2} \ell(\theta|x) \right\} = E(X\theta^{-2}) = \theta^{-1}.$$

Therefore, the CRLB is $\theta e^{-2\theta}/n$.

(c) Let $Y = \sum_{i=1}^n X_i$. Show that $T = (1 - 1/n)^Y$ is an unbiased estimator of $\tau(\theta)$ using the fact that $Y \sim \text{Poisson}(n\theta)$. Derive the variance of $T$ and compare the variance to the CRLB.

**Solution**: We have

$$
\begin{aligned}
E(T) &= E\left\{(1-1/n)^Y\right\} \\
&= \sum_{y=0}^{\infty}(1-1/n)^y \frac{(n\theta)^y}{y!} \\
&= e^{-n\theta}\sum_{y=0}^{\infty}\frac{(n\theta-\theta)^y}{y!} \\
&= e^{-n\theta}e^{n\theta-\theta} \\
&= e^{-\theta}.
\end{aligned}
$$

We can conclude that $\tau(\hat{\theta})$ is an unbiased estimator of $\tau(\theta)$. In addition, we have

$$
\begin{aligned}
E(T^2) &= E\{(1-1/n)^{2Y}\} \\
&= \sum_{y=0}^{\infty}(1-1/n)^{2y}\frac{(n\theta)^y e^{-n\theta}}{y!} \\
&= \sum_{y=0}^{\infty}(1+1/n^2-2/n)^y\frac{(n\theta)^y e^{-n\theta}}{y!} \\
&= \sum_{y=0}^{\infty}(n\theta+\theta/n-2\theta)^y\frac{e^{-(n\theta+\theta/n-2\theta)}}{y!}e^{\theta/n-2\theta} \\
&= e^{\theta/n-2\theta}.
\end{aligned}
$$

Therefore, $\mathrm{var}(T) = E(T^2) - \{E(T)\}^2 = e^{\theta/n-2\theta} - e^{-2\theta} = e^{-2\theta}(e^{\theta/n}-1)$. We know that the CRLB is $e^{-2\theta}\theta/n$, which is smaller than $e^{-2\theta}(e^{\theta/n}-1)$ since $e^{\theta/n}-1 = \theta/n + \sum_{i=2}^{\infty}(\theta/n)^i > \theta/n$. The variance does not reach CRLB.

(d) Show that $T$ is the uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.

**Solution**: Since $T$ is an unbiased of $\tau(\theta)$ and a function of the complete and sufficient statistic $\sum_{i=1}^{n} X_i$, we can conclude $T$ is the UMVUE.

3. Assume that the distribution of wages (in thousands of dollars) in a large U.S. city follows a Pareto distribution with pdf

$$
f_Y(y) = \theta\gamma^\theta y^{-(\theta+1)}, \quad 0<\gamma<y<\infty, \quad 0<\theta<\infty,
$$

where $\gamma$ is a unknown parameter but $\theta$ is assumed a known constant. Let $Y_1, \cdots, Y_n$ be a random sample from $f_Y(y)$. To test the null hypothesis $H_0 : \gamma \leq \gamma_0$ versus the alternative hypothesis $H_1 : \gamma > \gamma_0$, one intend to develop a likelihood ratio test (LRT) to conclude whether the minimum wage is smaller than some value $\gamma_0$.

(a) Derive the likelihood ratio test statistic $\lambda(\boldsymbol{y})$ and show that the rejection region $R = \{\boldsymbol{y} : \lambda(\boldsymbol{y}) \leq c\}$ is equivalent to $R^* = \{\boldsymbol{y} : y_{(1)} \geq c^*\}$, where $y_{(1)}$ is the minimum order statistic.

**Solution**: The likelihood ratio test statistic is

$$\lambda(\boldsymbol{y}) = \frac{\sup_{\gamma \in \Theta_0} L(\gamma|\boldsymbol{y})}{\sup_{\gamma \in \Theta} L(\gamma|\boldsymbol{y})},$$

where

$$L(\gamma|\boldsymbol{y}) = \theta^n \gamma^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)} I(y_{(1)} > \gamma).$$

Maximized over the overall parameter space, the denominator of $\lambda(\boldsymbol{y})$ is

$$\sup_{\gamma \in \Theta} L(\gamma|\boldsymbol{y}) = L(y_{(1)}|y_1, \ldots, y_n) = \theta^n y_{(1)}^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)}.$$

Maximized over the null parameter space, the numerator of $\lambda(\boldsymbol{y})$ is

$$\sup_{\gamma \in \Theta_0} L(\gamma|\boldsymbol{y}) = \begin{cases} \theta^n y_{(1)}^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)} & \text{if } \ y_{(1)} \leq \gamma_0, \\ \theta^n \gamma_0^{n\theta} \prod_{i=1}^{n} y_i^{-(\theta+1)} & \text{if } \ y_{(1)} > \gamma_0. \end{cases}$$

The LRT statistic becomes

$$\lambda(\boldsymbol{y}) = \begin{cases} 1 & \text{if } \ y_{(1)} \leq \gamma_0, \\ \gamma_0^{n\theta} y_{(1)}^{-n\theta} & \text{if } \ y_{(1)} > \gamma_0. \end{cases}$$

By drawing the graph of $\lambda(\boldsymbol{y})$ versus $y_{(1)}$, one can find the critical region $R = \{\boldsymbol{y} : \lambda(\boldsymbol{y}) \leq c\}$ is equivalent to $R^* = \{\boldsymbol{y} : y_{(1)} \geq c^*\}$.