1. Let $X$ be a $B(n, p)$ (binomial) random variable and $Y$ be a $B(n-1, p)$ random variable.

(a) Show that $E[X^k] = npE[(Y + 1)^{k-1}]$.

(b) Find $k$ such that $P(X = k)$ attains its maximum.

解答:

(a) 

$$E[X^k] = np \sum_{i=1}^{n} i^{k-1} \binom{n-1}{i-1} p^{i-1}(1-p)^{n-i}$$

$$= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j(1-p)^{n-1-j}$$

$$= npE[(Y + 1)^{k-1}]$$

(b) 

$$\frac{P(X = k)}{P(X = k - 1)} = \frac{(n - k + 1)p}{k(1-p)} \geq 1 \Leftrightarrow k \leq (n + 1)p.$$ 

$\therefore k = [(n + 1)p]$.

2. Suppose $X, Y$ are random variables with joint density

$$f_{X,Y}(x, y) = \frac{1}{2\pi \sqrt{x(y-x)}} e^{-y/2}, \quad 0 < x < y.$$ 

(a) Find the distribution of $Y$. [Hint: use the substitution $x = ys$.]

(b) Find the conditional distribution of $X$ given $Y = y$.

(c) Compute $E[X|Y = 1]$ and $\text{Var}(X|Y = 1)$.

解答:

(a) The p.d.f. of $Y$ is

$$f_Y(y) = \frac{1}{2} e^{-y/2}, \quad y > 0.$$ 

(b) The p.d.f of the conditional distribution is

$$f(x|y) = \frac{1}{\pi \sqrt{y-x}} e^{-y/2}, \quad 0 < x < y.$$
3. Let $X_1, X_2, \ldots, X_n$ be independent and identically with $N(\theta, 1)$ distribution. Find the best unbiased estimator of $\theta^2$. Calculate its variance, and show that it is greater than the Cramer-Rao Lower Bound.

解答:

$\sum_i X_i$ is a complete sufficient statistic for $\theta$ when $X_i \sim N(\theta, 1)$. $\bar{X}^2 - 1/n$ is a function of $\sum_i X_i$. Therefore, by Theorem 7.3.23, $\bar{X}^2 - 1/n$ is the unique best unbiased estimator of its expectation.

$$E\left(\bar{X}^2 - \frac{1}{n}\right) = \text{Var}\bar{X} + (E\bar{X})^2 - \frac{1}{n} = \frac{1}{n} + \theta^2 - \frac{1}{n} = \theta^2.$$  

Therefore, $\bar{X}^2 - 1/n$ is the UMVUE of $\theta^2$. We will calculate

$$\text{Var}(\bar{X}^2 - 1/n) = \text{Var}(\bar{X}^2) = E(\bar{X}^4) - [E(\bar{X}^2)]^2,$$

but first we derive some general formulas that will also be useful in later exercises. Let $Y \sim N(\theta, \sigma^2)$. Then here are formulas for $EY^4$ and $\text{Var}Y^2$.

$$EY^4 = E[Y^3(Y - \theta + \theta)] = EY^3(Y - \theta) + EY^3\theta = EY^3(Y - \theta) + \theta EY^3.$$  

$$EY^3(Y - \theta) = \sigma^2 E(3Y^2) = \sigma^2(3\sigma^2 + \theta^2) = 3\sigma^4 + 3\theta^2\sigma^2. \quad \text{(Stein's Lemma)}$$  

$$\theta EY^3 = \theta(3\theta \sigma^2 + \theta^3) = 3\theta^2 \sigma^2 + \theta^4. \quad \text{(Example 3.6.6)}$$  

$$\text{Var}Y^2 = 3\sigma^4 + 6\theta^2 \sigma^2 + \theta^4 - (\sigma^2 + \theta^2)^2 = 2\sigma^4 + 4\theta^2 \sigma^2.$$  

Thus,

$$\text{Var}\left(\bar{X}^2 - \frac{1}{n}\right) = \text{Var}\bar{X}^2 = \frac{2}{n^2} + 4\theta^2 \frac{1}{n^2} > \frac{4\theta^2}{n}.$$  

To calculate the Cramer-Rao lower bound, we have

$$E_\theta\left(\begin{array}{c} \partial^2 \log f(X|\theta) \\ \partial \theta^2 \end{array}\right) = E_\theta\left(\begin{array}{c} \partial^2 \log \frac{1}{\sqrt{2\pi}} e^{-(X-\theta)^2/2} \\ \partial \theta^2 \end{array}\right)$$  

$$= E_\theta\left(\begin{array}{c} \partial^2 \log(2\pi)^{-1/2} - \frac{1}{2}(X-\theta)^2 \\ \partial \theta^2 \end{array}\right) = E_\theta\left(\frac{\partial}{\partial \theta}(X-\theta)\right) = -1,$$

and $\tau(\theta) = \theta^2$, $[\tau'(\theta)]^2 = (2\theta)^2 = 4\theta^2$ so the Cramer-Rao Lower Bound for estimating $\theta^2$ is

$$\frac{[\tau'(\theta)]^2}{-nE_\theta(\frac{\partial^2}{\partial \theta^2} \log f(X|\theta))} = \frac{4\theta^2}{n}.$$  

Thus, the UMVUE of $\theta^2$ does not attain the Cramer-Rao bound. (However, the ratio of the variance and the lower bound $\to 1$ as $n \to \infty$.)

4. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be random samples drawn from $X$ and $Y$ respectively where $X$ has distribution function $F$ and $Y$ has distribution function $G$. The testing hypothesis is

$$H_0 : F = G \quad \text{against} \quad H_1 : F \neq G.$$  

Let $Z_{(1)} \leq Z_{(2)} \leq \cdots \leq Z_{(n+m)}$ be the order statistics of the combined samples. Define $W = \sum_{k=1}^{n+m} kI_k$ where $I_k = 1$ if $Z_{(k)}$ is an $X$ and $I_k = 0$ if $Z_{(k)}$ is an $Y$. 

(c) $E[X|Y = 1] = 1/2, \quad \text{Var}(X|Y = 1) = 1/8.$
(a) Compute the mean and variance of the random variable \( W \) under \( H_0 \).

(b) Suppose in an experimental project, Hospital A applies its treatment, say treatment A, to 10 cancer patients and they survive 3.1, 2.8, 3.2, 5.1, 4.2, 2.1, 1.5, 3.6, 4.3, 4.7 years respectively. At the same time Hospital B applies its treatment, say treatment B, to 12 cancer patients and they survive 1.2, 2.4, 3.5, 2.8, 2.3, 4.4, 8.2, 2.5, 1.1, 2.7, 1.8, 0.6 years respectively. Use the result in part (a) to make a judgement about these two treatments. Type I error is 0.05.

解答:

(a) \( W = \sum_{i=1}^{n+m} iI_i \). Thus \( E[W] = \sum_{i=1}^{n+m} iE[I_i] \).

\[
I_i = \begin{cases} 
1 & \text{if } Z_{(i)} \text{ is an } X \\
0 & \text{otherwise}
\end{cases}
\]

\[
E[I_i] = \frac{n}{n+m} \quad \text{and} \quad E[I_i I_j] = \frac{n(n+1)}{2(n+m)}.
\]

\[
E[W] = \frac{n}{n+m} \sum_{i=1}^{n+m} i = \frac{n(n+m+1)}{2}.
\]

\[
\text{Var}(W) = E[W^2] - E^2[W]
\]

\[
E[W^2] = E \left[ \sum_{i=1}^{n+m} i^2 I_i + \sum_{i\neq j} ijI_iI_j \right]
\]

\[
= \sum_{i=1}^{n+m} i^2 E[I_i] + \sum_{i\neq j} ij E[I_iI_j]
\]

where

\[
E[I_i] = \frac{n}{n+m}
\]

\[
E[I_i I_j] = \frac{n}{n+m} \frac{n-1}{n+m-1}.
\]

\[
E[W^2] = \sum_{i=1}^{n+m} i^2 I_i + \sum_{i\neq j} ij I_iI_j
\]

\[
\text{Var}(W) = \frac{mn(n+m+1)}{12}
\]

(b) \( X \) has score \( W = 13 + 11.5 + \cdots + 20 = 140.5 \)

\[
E[W] = (1 + 2 + \cdots + 22) \cdot \frac{10}{22} = 115 \quad \text{under } H_0
\]

\[
\text{Var}(W) = \frac{mn(n+m+1)}{12} \bigg|_{n=10, m=12} = 230
\]

\[
\sqrt{\text{Var}(W)} = \sqrt{230} \approx 15.4
\]

\[
\therefore 95\% \text{ C.I. is } (115 - 1.96(14.4), 115 + 1.96(14.4)). \text{ Since 140.5 is located inside the confidence interval, therefore these two treatments are not significantly different.}
\]