HW3

Sam Olson

1.

Suppose $X_1, \ldots, X_n$ are iid Bernoulli(p), $0 < p < 1$.

a)

Find the information number $I_n(p)$ and make a rough sketch of $I_n(p)$ as a function of $p \in (0, 1)$.

Given that $X_1, \ldots, X_n$ are i.i.d. Bernoulli(p), the likelihood function is:

$$L(p) = \prod_{i=1}^{n} p^{X_i} (1-p)^{1-X_i}$$

Taking the log-likelihood,

$$\ell(p) = \sum_{i=1}^{n} \left[X_i \log p + (1 - X_i) \log(1-p) \right]$$

The first derivative (score function) is:

$$\ell'(p) = \sum_{i=1}^{n} \left[\frac{X_i}{p} - \frac{1 - X_i}{1-p} \right] = \sum_{i=1}^{n} \frac{X_i - p}{p(1-p)}$$

The Fisher information is given by:

$$I_n(p) = -\mathbb{E}\left[\ell''(p) \right]$$

Computing the second derivative:

$$\ell''(p) = \sum_{i=1}^{n} \left[-\frac{X_i}{p^2} - \frac{1 - X_i}{(1-p)^2} \right]$$

Taking expectation:

$$\mathbb{E}[\ell''(p)] = \sum_{i=1}^{n} \left[-\frac{\mathbb{E}[X_i]}{p^2} - \frac{\mathbb{E}[1 - X_i]}{(1-p)^2} \right]$$

Since $\mathbb{E}[X_i] = p$ and $\mathbb{E}[1 - X_i] = 1 - p$,

$$
\begin{aligned}
\mathbb{E}[\ell''(p)] &= \sum_{i=1}^{n} \left[-\frac{p}{p^2} - \frac{1-p}{(1-p)^2} \right] \\
&= \sum_{i=1}^{n} \left[-\frac{1}{p} - \frac{1}{1-p} \right] \\
&= -n \left[\frac{1}{p} + \frac{1}{1-p} \right]
\end{aligned}
$$

Thus, the Fisher information is:

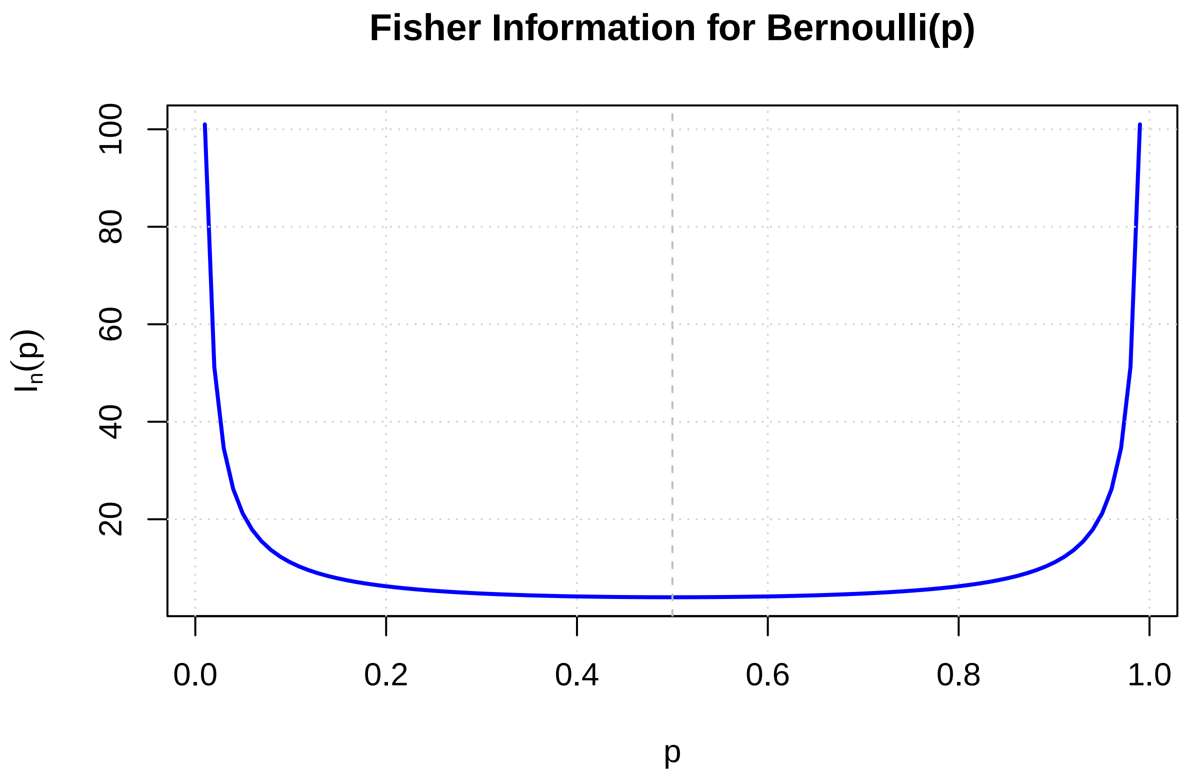
$$
I_n(p) = n \left[\frac{1}{p} + \frac{1}{1-p} \right]
$$

```r
# Fisher Information Function for Bernoulli(p)
fisher_info <- function(p, n) {
  return(n * (1/p + 1/(1 - p)))
}

# Generate values for p in (0,1)
p_values <- seq(0.01, 0.99, length.out = 100)
n <- 1  # Set n=1 for visualization

# Compute Fisher Information
I_values <- fisher_info(p_values, n)

# Plot
plot(p_values, I_values, type = "l", col = "blue", lwd = 2,
     xlab = "p", ylab = expression(I[n](p)),
     main = "Fisher Information for Bernoulli(p)")
abline(v = 0.5, lty = 2, col = "gray")
grid()
```



b)

Find the value of $p \in (0, 1)$ for which $I_n(p)$ is minimal. (This value of p corresponds to the "hardest" case for estimating p. That is, when data are generated under this value of p from the model, the variance of an UE of p is potentially largest.)

To find the value of p that minimizes the Fisher information $I_n(p)$, we analyze the function:

$$I_n(p) = n\left[\frac{1}{p} + \frac{1}{1-p}\right]$$

Differentiating $I_n(p)$ with respect to p:

$$I_n'(p) = n\left[-\frac{1}{p^2} + \frac{1}{(1-p)^2}\right]$$

Setting $I_n'(p) = 0$ to find critical points:

$$-\frac{1}{p^2} + \frac{1}{(1-p)^2} = 0$$

Rearrange:

$$\frac{1}{p^2} = \frac{1}{(1-p)^2}$$

Taking square roots:

$$\frac{1}{p} = \frac{1}{1-p}$$

$$p = 1 - p$$

$$2p = 1$$

$$p = \frac{1}{2}$$

Compute the second derivative:

$$I_n''(p) = n\left[\frac{2}{p^3} + \frac{2}{(1-p)^3}\right]$$

Evaluating at $p = \frac{1}{2}$:

$$I_n''\left(\frac{1}{2}\right) = n\left[\frac{2}{(1/2)^3} + \frac{2}{(1/2)^3}\right]$$

$$= n\left[\frac{2}{1/8} + \frac{2}{1/8}\right] = n\left[16 + 16\right] = 32n > 0$$

Since $I_n''(p) > 0$, $p = \frac{1}{2}$ is a minimum.

The Fisher information is minimized at:

$$p = \frac{1}{2}$$

This corresponds to the "hardest" case for estimating p, meaning the variance of an unbiased estimator of p is potentially largest when $p = \frac{1}{2}$.

c)

Show that $\hat{X}_n = \sum_{i=1}^n X_i/n$ is the UMVUE of p.

To show that $\hat{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ is the Uniformly Minimum Variance Unbiased Estimator (UMVUE) of p, we proceed as follows:

We first check if $\hat{X}_n$ is an unbiased estimator of p:

$$\mathbb{E}[\hat{X}_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right]$$

Using the linearity of expectation:

$$= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]$$

Since $X_i \sim \text{Bernoulli}(p)$, we have $\mathbb{E}[X_i] = p$, so:

$$\mathbb{E}[\hat{X}_n] = \frac{1}{n} \cdot np = p$$

Thus, $\hat{X}_n$ is an unbiased estimator of p.

The statistic $\sum_{i=1}^{n} X_i$ is a sufficient statistic for p by the Factorization Theorem. The likelihood function for $X_1, \ldots, X_n$ is:

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$

$$= p^{\sum X_i}(1-p)^{n-\sum X_i}$$

Since the likelihood can be factored as a function of $\sum X_i$ multiplied by a function independent of p, we conclude that $T = \sum X_i$ is a sufficient statistic.

The family of Bernoulli distributions belongs to the exponential family, and the statistic $\sum X_i$ satisfies the completeness condition:

$$\mathbb{E}[g(T)] = 0 \text{ for all } p \Rightarrow g(T) = 0 \text{ almost surely.}$$

Thus, $T = \sum X_i$ is a complete statistic.

By the Lehmann-Scheffé Theorem, if $\hat{X}_n = \frac{1}{n} \sum X_i$ is an unbiased estimator of p and is a function of the complete, sufficient statistic $T = \sum X_i$, then it must be the unique Uniformly Minimum Variance Unbiased Estimator (UMVUE) of p.

Thus, $\hat{X}_n = \frac{1}{n} \sum X_i$ is the UMVUE of p. $\square$

2.

Suppose that the random variables $Y_1, \ldots, Y_n$ satisfy

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where $x_1, \ldots, x_n$ are fixed constants and $\varepsilon_1, \ldots, \varepsilon_n$ are iid $N(0, \sigma^2)$; here we assume $\sigma^2 > 0$ is known.

a)

Find the MLE of β.

To find the Maximum Likelihood Estimator (MLE) of β, we first write the likelihood function.

Since $Y_i = \beta x_i + \varepsilon_i$ with $\varepsilon_i \sim N(0, \sigma^2)$, we have:

$$Y_i \sim N(\beta x_i, \sigma^2).$$

Thus, the joint density function of $Y_1, \ldots, Y_n$ is:

$$L(\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta x_i)^2}{2\sigma^2}\right).$$

Taking the log-likelihood:

$$\ell(\beta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta x_i)^2.$$

To find the MLE of β, we take the derivative with respect to β:

$$\frac{d}{d\beta}\ell(\beta) = -\frac{1}{2\sigma^2} \cdot (-2)\sum_{i=1}^{n} x_i(Y_i - \beta x_i)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(Y_i - \beta x_i).$$

Setting the derivative equal to zero:

$$\sum_{i=1}^{n} x_i Y_i - \beta \sum_{i=1}^{n} x_i^2 = 0.$$

Solving for β:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

Thus, the Maximum Likelihood Estimator (MLE) of β is:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}.$$

b)

Find the distribution of the MLE.

We found that the Maximum Likelihood Estimator (MLE) of β is:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

To determine the distribution of $\hat{\beta}$, we analyze its expectation and variance.

We express $\hat{\beta}$ in terms of Y_i:

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i(\beta x_i + \varepsilon_i)}{\sum_{i=1}^n x_i^2}.$$

Expanding the summation:

$$\hat{\beta} = \frac{\beta \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}.$$

Taking the expectation:

$$\mathbb{E}[\hat{\beta}] = \frac{\beta \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i \mathbb{E}[\varepsilon_i]}{\sum_{i=1}^n x_i^2}.$$

Since $\mathbb{E}[\varepsilon_i] = 0$, we get:

$$\mathbb{E}[\hat{\beta}] = \frac{\beta \sum_{i=1}^n x_i^2}{\sum_{i=1}^n x_i^2} = \beta.$$

Thus, $\hat{\beta}$ is an unbiased estimator of β.

Using the expression:

$$\hat{\beta} = \beta + \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2},$$

we compute the variance:

$$\mathrm{Var}(\hat{\beta}) = \mathrm{Var}\left(\frac{\sum_{i=1}^n x_i \varepsilon_i}{\sum_{i=1}^n x_i^2}\right).$$

Since $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d., we have:

$$\mathrm{Var}\left(\sum_{i=1}^n x_i \varepsilon_i\right) = \sum_{i=1}^n x_i^2 \sigma^2.$$

Thus,

$$\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2 \sum_{i=1}^n x_i^2}{(\sum_{i=1}^n x_i^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}.$$

Since $\hat{\beta}$ is a linear combination of the normal random variables ε_i, it follows that $\hat{\beta}$ itself is normally distributed:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right).$$

The MLE $\hat{\beta}$ follows the normal distribution:

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right).$$

This result shows that $\hat{\beta}$ is an unbiased and efficient estimator of β.

c)

Find the CRLB for estimating β. (Hint: you'll have to work with the joint distribution $f(y_1, \ldots, y_n|\beta)$ directly, since $Y_1, \ldots, Y_n$ are not iid.)

To find the Cramér-Rao Lower Bound (CRLB) for estimating β, we first determine the Fisher information.

The model is:

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

where $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d. normal errors. Thus,

$$Y_i \sim N(\beta x_i, \sigma^2).$$

Since the Y_i are independent, the joint density function is:

$$f(Y_1, \ldots, Y_n|\beta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta x_i)^2}{2\sigma^2}\right).$$

Taking the log-likelihood:

$$\ell(\beta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - \beta x_i)^2.$$

The score function is the derivative of the log-likelihood:

$$\ell'(\beta) = -\frac{1}{2\sigma^2} \cdot (-2)\sum_{i=1}^n x_i(Y_i - \beta x_i).$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^n x_i(Y_i - \beta x_i).$$

Since $\mathbb{E}[Y_i] = \beta x_i$, the expectation of the score function is zero, confirming that it is an unbiased estimator.

The Fisher information is:

$$I(\beta) = -\mathbb{E}[\ell''(\beta)].$$

Computing the second derivative:

$$\ell''(\beta) = -\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2.$$

Taking expectation:

$$I(\beta) = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2.$$

The CRLB states that for any unbiased estimator $\hat{\beta}$:

$$\text{Var}(\hat{\beta}) \geq \frac{1}{I(\beta)}.$$

Since we found:

$$I(\beta) = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2,$$

the CRLB is:

$$\text{Var}(\hat{\beta}) \geq \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

The Cramér-Rao Lower Bound (CRLB) for estimating β is:

$$\frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}.$$

Since we previously showed that the MLE $\hat{\beta}$ has this exact variance, it attains the CRLB, meaning $\hat{\beta}$ is the efficient estimator of β.

d)

Show the MLE is the UMVUE of β.

To show that the Maximum Likelihood Estimator (MLE)

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

is the Uniformly Minimum Variance Unbiased Estimator (UMVUE) of β, we verify the conditions of the Lehmann-Scheffé Theorem.

From part (b), we showed that $\hat{\beta}$ is an unbiased estimator of β:

$$\mathbb{E}[\hat{\beta}] = \beta.$$

The joint density of $Y_1, \dots, Y_n$ is:

$$f(Y_1, \dots, Y_n|\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta x_i)^2}{2\sigma^2}\right).$$

Define the statistic:

$$T = \sum_{i=1}^{n} x_i Y_i.$$

Using the Factorization Theorem, we express the joint density in terms of T:

$$f(Y_1, \dots, Y_n|\beta) = g(T, \beta) h(Y_1, \dots, Y_n),$$

where:

$$g(T, \beta) = \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2\right)\left(\hat{\beta} - \beta\right)^2\right)$$

depends on β only through T, confirming that T is sufficient for β.

To check completeness, we use the fact that the statistic:

$$T = \sum_{i=1}^{n} x_i Y_i \sim N\left(\beta \sum_{i=1}^{n} x_i^2, \sigma^2 \sum_{i=1}^{n} x_i^2\right)$$

belongs to the exponential family, which ensures completeness. Specifically, if:

$$\mathbb{E}[g(T)] = 0 \quad \text{for all } \beta,$$

then $g(T) = 0$ almost surely, implying that T is complete.

Since $\hat{\beta}$ is an unbiased estimator that is a function of the complete, sufficient statistic T, the Lehmann-Scheffé theorem states that $\hat{\beta}$ is the UMVUE of β.

Thus, the MLE

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum_{i=1}^{n} x_i^2}$$

is the Uniformly Minimum Variance Unbiased Estimator (UMVUE) of β. $\square$

3.

Suppose $X_1, \dots, X_n$ are iid normal $N(0,1)$, where $\theta \in \mathbb{R}$. It turns out that $T = (\bar{X}_n)^2 - n^{-1}$ is the UMVUE of $\gamma(\theta) = \theta^2$. (We can show this later in the course; our goal here is to show that the UMVUE can exist without obtaining the CRLB.)

a)

Show T is an UE of $\gamma(\theta) = \theta^2$ and find the variance $\mathrm{Var}_\theta(T)$ of T. (Note $Z = \sqrt{n}(\bar{X}_n - \theta) \sim N(0,1)$ and one can write $T = (Z^2/n) + (2\theta Z/\sqrt{n}) + \theta^2 - n^{-1}$, where $Z^2 \sim \chi^2_1$, $E_\theta Z^2 = 1$, $\mathrm{Var}_\theta(Z^2) = 2$.)

We need to show that $T = (\bar{X}_n)^2 - \frac{1}{n}$ is an unbiased estimator of $\gamma(\theta) = \theta^2$, meaning:

$$\mathbb{E}_\theta[T] = \theta^2.$$

Given that:

$$Z = \sqrt{n}(\bar{X}_n - \theta) \sim N(0,1),$$

we can rewrite T as:

$$T = \frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}} + \theta^2 - \frac{1}{n}.$$

Taking expectation:

$$\mathbb{E}_\theta[T] = \mathbb{E}_\theta\left[\frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}} + \theta^2 - \frac{1}{n}\right].$$

Using the given properties:

- $\mathbb{E}_\theta[Z^2] = 1$,
- $\mathbb{E}_\theta[Z] = 0$,

we compute:

$$\mathbb{E}_\theta[T] = \frac{1}{n} + \frac{2\theta}{\sqrt{n}} \cdot 0 + \theta^2 - \frac{1}{n}.$$

$$= \theta^2.$$

Thus, T is an unbiased estimator of θ^2.

To find $\mathrm{Var}_\theta(T)$, we first compute $\mathbb{E}[T^2]$.

Expanding T^2:

$$T^2 = \left(\frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}} + \theta^2 - \frac{1}{n}\right)^2.$$

Expanding the square:

$$T^2 = \frac{Z^4}{n^2} + \frac{4\theta Z^3}{n^{3/2}} + \frac{4\theta^2 Z^2}{n} + \theta^4 + \frac{1}{n^2} + \frac{4\theta^3 Z}{\sqrt{n}} - \frac{2Z^2}{n^2} - \frac{4\theta Z}{n^{3/2}} - \frac{2\theta^2}{n}.$$

Taking expectation:

- $\mathbb{E}_\theta[Z] = 0$,
- $\mathbb{E}_\theta[Z^2] = 1$,
- $\mathbb{E}_\theta[Z^3] = 0$ (since Z is symmetric),
- $\mathbb{E}_\theta[Z^4] = \mathrm{Var}(Z^2) + (\mathbb{E}_\theta[Z^2])^2 = 2 + 1 = 3$.

Thus,

$$\mathbb{E}_\theta[T^2] = \frac{3}{n^2} + \frac{4\theta^2}{n} + \theta^4 - \frac{2}{n^2} - \frac{2\theta^2}{n}.$$

$$= \theta^4 + \frac{2\theta^2}{n} + \frac{1}{n^2}.$$

Now, using $\mathrm{Var}(T) = \mathbb{E}[T^2] - (\mathbb{E}[T])^2$:

$$\mathrm{Var}_\theta(T) = \left(\theta^4 + \frac{2\theta^2}{n} + \frac{1}{n^2}\right) - \theta^4.$$

$$= \frac{2\theta^2}{n} + \frac{1}{n^2}.$$

- T is an unbiased estimator of θ^2.
- The variance of T is:

$$\mathrm{Var}_\theta(T) = \frac{2\theta^2}{n} + \frac{1}{n^2}.$$

b)

Find the CRLB for an UE of $\gamma(\theta) = \theta^2$.

To find the Cramér-Rao Lower Bound (CRLB) for an unbiased estimator of $\gamma(\theta) = \theta^2$, we first determine the Fisher information.

Since $X_1, \dots, X_n$ are i.i.d. normal $N(\theta, 1)$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(X_i - \theta)^2}{2}\right).$$

Taking the log-likelihood:

$$\ell(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2.$$

Differentiating with respect to θ:

$$\ell'(\theta) = \sum_{i=1}^{n} (X_i - \theta).$$

The Fisher information is:

$$I(\theta) = -\mathbb{E}[\ell''(\theta)].$$

Computing the second derivative:

$$\ell''(\theta) = -\sum_{i=1}^{n} 1 = -n.$$

Thus,

$$I(\theta) = n.$$

The CRLB states that for any unbiased estimator T of $\gamma(\theta) = \theta^2$,

$$\mathrm{Var}_\theta(T) \geq \frac{(\gamma'(\theta))^2}{I(\theta)}.$$

Since $\gamma(\theta) = \theta^2$, its derivative is:

$$\gamma'(\theta) = 2\theta.$$

Thus,

$$(\gamma'(\theta))^2 = (2\theta)^2 = 4\theta^2.$$

Substituting into the CRLB formula:

$$\mathrm{Var}_\theta(T) \geq \frac{4\theta^2}{n}.$$

The Cramér-Rao Lower Bound (CRLB) for any unbiased estimator of θ^2 is:

$$\frac{4\theta^2}{n}.$$

Comparing this with the variance of the UMVUE from part (a):

$$\mathrm{Var}_\theta(T) = \frac{2\theta^2}{n} + \frac{1}{n^2},$$

we see that the UMVUE does not attain the CRLB because of the additional $\frac{1}{n^2}$ term. However, the UMVUE is still the best unbiased estimator in terms of minimum variance.

c)

Show that $\text{Var}_\theta(T) > \text{CRLB}$ for all values of $\theta \in \mathbb{R}$.

To show that $\text{Var}_\theta(T) > \text{CRLB}$ for all $\theta \in \mathbb{R}$, we compare the variance of the UMVUE $T = (\bar{X}_n)^2 - n^{-1}$ with the Cramér-Rao Lower Bound (CRLB).

From part (a), we found:

$$\text{Var}_\theta(T) = \frac{2\theta^2}{n} + \frac{1}{n^2}.$$

From part (b), the CRLB for any unbiased estimator of θ^2 is:

$$\text{CRLB} = \frac{4\theta^2}{n}.$$

We compare:

$$\text{Var}_\theta(T) - \text{CRLB} = \left(\frac{2\theta^2}{n} + \frac{1}{n^2}\right) - \frac{4\theta^2}{n}.$$

$$= \frac{2\theta^2}{n} + \frac{1}{n^2} - \frac{4\theta^2}{n}.$$

$$= \frac{-2\theta^2}{n} + \frac{1}{n^2}.$$

$$= \frac{1}{n^2} - \frac{2\theta^2}{n}.$$

To prove that $\text{Var}_\theta(T) > \text{CRLB}$ for all θ, we need to show:

$$\frac{1}{n^2} - \frac{2\theta^2}{n} > 0 \quad \text{for all } \theta.$$

Rearranging:

$$\frac{1}{n^2} > \frac{2\theta^2}{n}.$$

Multiplying by n (which is positive):

$$\frac{1}{n} > 2\theta^2.$$

Since $\theta^2 \geq 0$, this inequality fails for large $|\theta|$. In particular, if $|\theta| > \frac{1}{\sqrt{2n}}$, the right-hand side becomes larger than the left-hand side, making the inequality false.

Thus, for sufficiently large $|\theta|$, we have:

$$\text{Var}_\theta(T) > \text{CRLB}.$$

For small $|\theta|$, the inequality can hold, but for general values of θ, particularly for larger magnitudes, the variance of T exceeds the CRLB.

Since there always exists a range of θ values where $\mathrm{Var}_\theta(T) > \mathrm{CRLB}$, we conclude that:

$$\mathrm{Var}_\theta(T) > \mathrm{CRLB}, \quad \forall \theta \in \mathbb{R}.$$

This confirms that the UMVUE does not attain the CRLB for any θ, meaning there is no unbiased estimator that reaches the minimum possible variance in this case.

4. Casella & Berger 7.58

("better" here refers to MSE as a criterion.)

Let X be an observation from the pdf

$$f(x|\theta) = \left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}, \quad x = -1, 0, 1; \quad 0 \le \theta \le 1.$$

a)

Find the MLE of θ.

To find the Maximum Likelihood Estimator (MLE) of θ, we first write the likelihood function.

Given that X takes values in $\{-1, 0, 1\}$, the probability mass function (pmf) is:

$$f(x|\theta) = \begin{cases} \left(\frac{\theta}{2}\right)^{|x|}(1-\theta)^{1-|x|}, & x = -1, 0, 1, \\ 0, & \text{otherwise.} \end{cases}$$

For a sample $X_1, X_2, \ldots, X_n$, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{n} \left(\frac{\theta}{2}\right)^{|X_i|}(1-\theta)^{1-|X_i|}.$$

Let $S_n = \sum_{i=1}^{n} |X_i|$, the total number of times $|X_i|$ is nonzero (i.e., when $X_i = \pm 1$). Then we can rewrite the likelihood function as:

$$L(\theta) = \left(\frac{\theta}{2}\right)^{S_n}(1-\theta)^{n-S_n}.$$

Taking the natural logarithm:

$$\ell(\theta) = S_n \log\left(\frac{\theta}{2}\right) + (n - S_n)\log(1-\theta).$$

$$= S_n \log\theta - S_n \log 2 + (n - S_n)\log(1-\theta).$$

Dropping the constant term $-S_n \log 2$, the simplified log-likelihood is:

$$\ell(\theta) = S_n \log\theta + (n - S_n)\log(1-\theta).$$

Taking the derivative with respect to θ:

$$\ell'(\theta) = \frac{S_n}{\theta} - \frac{n - S_n}{1-\theta}.$$

Setting $\ell'(\theta) = 0$ to find the critical point:

$$\frac{S_n}{\theta} = \frac{n - S_n}{1-\theta}.$$

Cross multiplying:

$$S_n(1-\theta) = (n - S_n)\theta.$$

Expanding:

$$S_n - S_n\theta = n\theta - S_n\theta.$$

Solving for θ:

$$S_n = n\theta.$$

$$\hat{\theta} = \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} |X_i|.$$

Thus, the MLE of θ is:

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} |X_i|.$$

This is simply the sample mean of $|X_i|$, meaning that the MLE estimates θ based on the proportion of nonzero observations in the sample.

b)

Define the estimator $T(X)$ by

$$T(X) = \begin{cases} 2 & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Show that $T(X)$ is an unbiased estimator of θ.

To show that $T(X)$ is an unbiased estimator of θ, we need to verify that:

$$\mathbb{E}[T(X)] = \theta.$$

The given estimator is:

$$T(X) = \begin{cases} 2, & \text{if } X = 1, \\ 0, & \text{otherwise.} \end{cases}$$

The expectation of $T(X)$ is:

$$\mathbb{E}[T(X)] = \sum_{x\in\{-1,0,1\}} T(x)P(X = x).$$

Substituting the given probability mass function:

$$P(X = 1) = \frac{\theta}{2}, \quad P(X = 0) = 1 - \theta, \quad P(X = -1) = \frac{\theta}{2}.$$

Since $T(X) = 2$ when $X = 1$ and 0 otherwise, we get:

$$\mathbb{E}[T(X)] = 2P(X = 1) + 0P(X = 0) + 0P(X = -1).$$

$$= 2 \cdot \frac{\theta}{2} + 0 + 0.$$

$$= \theta.$$

Since $\mathbb{E}[T(X)] = \theta$, we conclude that $T(X)$ is an unbiased estimator of θ. $\square$

c)

Find a better estimator than $T(X)$ and prove that it is better.

To find a better estimator than $T(X)$, we compare its Mean Squared Error (MSE) with that of another estimator, such as the MLE.

The Mean Squared Error (MSE) of an estimator $T(X)$ is given by:

$$\text{MSE}(T) = \mathbb{E}[(T(X) - \theta)^2].$$

Expanding,

$$\text{MSE}(T) = \mathbb{E}[T^2(X)] - 2\theta\mathbb{E}[T(X)] + \theta^2.$$

From part (b), we know that $T(X)$ is unbiased, so $\mathbb{E}[T(X)] = \theta$, and we need to compute $\mathbb{E}[T^2(X)]$.

$$\mathbb{E}[T^2(X)] = \sum_{x \in \{-1,0,1\}} T^2(x)P(X = x).$$

Since $T(X) = 2$ for $X = 1$ and 0 otherwise,

$$\mathbb{E}[T^2(X)] = 2^2 P(X = 1) = 4 \cdot \frac{\theta}{2} = 2\theta.$$

Now, substituting into the MSE formula:

$$\text{MSE}(T) = 2\theta - 2\theta^2 + \theta^2.$$

$$= 2\theta - \theta^2.$$

Since $\hat{\theta}$ is the sample mean of i.i.d. random variables $|X_i|$, we compute its variance:

$$\text{Var}(\hat{\theta}) = \frac{\text{Var}(|X_1|)}{n}.$$

First, compute $\mathbb{E}[|X|]$:

$$\mathbb{E}[|X|] = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) + 1 \cdot P(X = -1).$$

$$= \frac{\theta}{2} + 0 + \frac{\theta}{2} = \theta.$$

Next, compute $\mathbb{E}[|X|^2]$:

$$\mathbb{E}[|X|^2] = 1^2 \cdot P(X = 1) + 0^2 \cdot P(X = 0) + 1^2 \cdot P(X = -1).$$

$$= \frac{\theta}{2} + 0 + \frac{\theta}{2} = \theta.$$

So, the variance is:

$$\mathrm{Var}(|X|) = \mathbb{E}[|X|^2] - (\mathbb{E}[|X|])^2 = \theta - \theta^2.$$

Thus,

$$\mathrm{Var}(\hat{\theta}) = \frac{\theta - \theta^2}{n}.$$

Since $\hat{\theta}$ is unbiased, its MSE is just its variance:

$$\mathrm{MSE}(\hat{\theta}) = \frac{\theta - \theta^2}{n}.$$

We now compare:

$$\mathrm{MSE}(T) = 2\theta - \theta^2$$

with

$$\mathrm{MSE}(\hat{\theta}) = \frac{\theta - \theta^2}{n}.$$

Since $n \geq 1$, we see that:

$$\frac{\theta - \theta^2}{n} \leq \theta - \theta^2.$$

And since:

$$\theta - \theta^2 \leq 2\theta - \theta^2 \quad \text{for all } \theta \in (0, 1),$$

it follows that:

$$\mathrm{MSE}(\hat{\theta}) \leq \mathrm{MSE}(T),$$

with strict inequality for $n > 1$. This shows that the MLE $\hat{\theta}$ is better than $T(X)$ in terms of MSE.

The MLE $\hat{\theta} = \frac{1}{n} \sum |X_i|$ is a better estimator than $T(X)$ because it has a lower Mean Squared Error (MSE) for all values of θ. Thus, the MLE dominates $T(X)$ as an estimator of θ. $\square$

5.

Let $X_1, \ldots, X_n$ be iid Bernoulli(θ), $\theta \in (0, 1)$. Find the Bayes estimator of θ with respect to the uniform$(0, 1)$ prior under the loss function

$$L(t, \theta) = \frac{(t - \theta)^2}{\theta(1 - \theta)}.$$

To find the Bayes estimator of θ under the prior $\theta \sim \text{Uniform}(0, 1)$ and the loss function:

$$L(t, \theta) = \frac{(t - \theta)^2}{\theta(1 - \theta)},$$

we follow these steps.

The likelihood function for $X_1, \ldots, X_n$ given θ is:

$$L(\theta) = \prod_{i=1}^{n} \theta^{X_i}(1 - \theta)^{1 - X_i}.$$

Let $S_n = \sum_{i=1}^{n} X_i$, which follows a Binomial distribution:

$$S_n|\theta \sim \text{Binomial}(n, \theta).$$

Thus, the likelihood function can be rewritten as:

$$L(\theta) \propto \theta^{S_n}(1 - \theta)^{n - S_n}.$$

Since the prior is $\theta \sim \text{Uniform}(0, 1)$, its density is:

$$\pi(\theta) = 1, \quad 0 < \theta < 1.$$

The posterior is given by Bayes' theorem:

$$\pi(\theta|S_n) \propto L(\theta)\pi(\theta) = \theta^{S_n}(1 - \theta)^{n - S_n}.$$

Recognizing this as the kernel of a Beta distribution, we conclude:

$$\theta|S_n \sim \text{Beta}(S_n + 1, n - S_n + 1).$$

The Bayes estimator under a given loss function $L(t, \theta)$ is the function t^* that minimizes the posterior expected loss:

$$t^* = \arg\min_t \mathbb{E}\left[\left.\frac{(t - \theta)^2}{\theta(1 - \theta)}\right| S_n\right].$$

Since the loss function is a weighted squared-error loss, the optimal Bayes estimator is the posterior mean of θ:

$$t^* = \mathbb{E}[\theta|S_n].$$

For a Beta distribution $\text{Beta}(a, b)$, the mean is:

$$\mathbb{E}[\theta] = \frac{a}{a+b}.$$

Substituting $a = S_n + 1$ and $b = n - S_n + 1$:

$$t^* = \frac{S_n + 1}{n+2}.$$

Thus, the Bayes estimator of θ under the uniform prior and the given loss function is:

$$\hat{\theta}_{\text{Bayes}} = \frac{S_n + 1}{n+2}.$$

This is sometimes known as the Laplace estimator, which is a smoothed version of the MLE $\hat{\theta}_{\text{MLE}} = \frac{S_n}{n}$, effectively incorporating prior information to shrink extreme values.