

# HW3

Sam Olson

1.

Suppose  $X_1, \dots, X_n$  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  $X_1, \dots, X_n$  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,

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

The first derivative is:

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

The Fisher information is:

$$I_n(p) = -E[\log(L(p))'']$$

Computing the second derivative:

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

Taking expectation:

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

Given we know the distribution of the random variables, we know  $E[X_i] = p$  and  $E[1 - X_i] = 1 - p$ . This allows us to simplify the expression:

$$-E[\log(L(p))''] = -\sum_{i=1}^n \left[ -\frac{p}{p^2} - \frac{1-p}{(1-p)^2} \right] = -\sum_{i=1}^n \left[ -\frac{1}{p} - \frac{1}{1-p} \right] = n \left[ \frac{1}{p} + \frac{1}{1-p} \right]$$

Simplifying

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

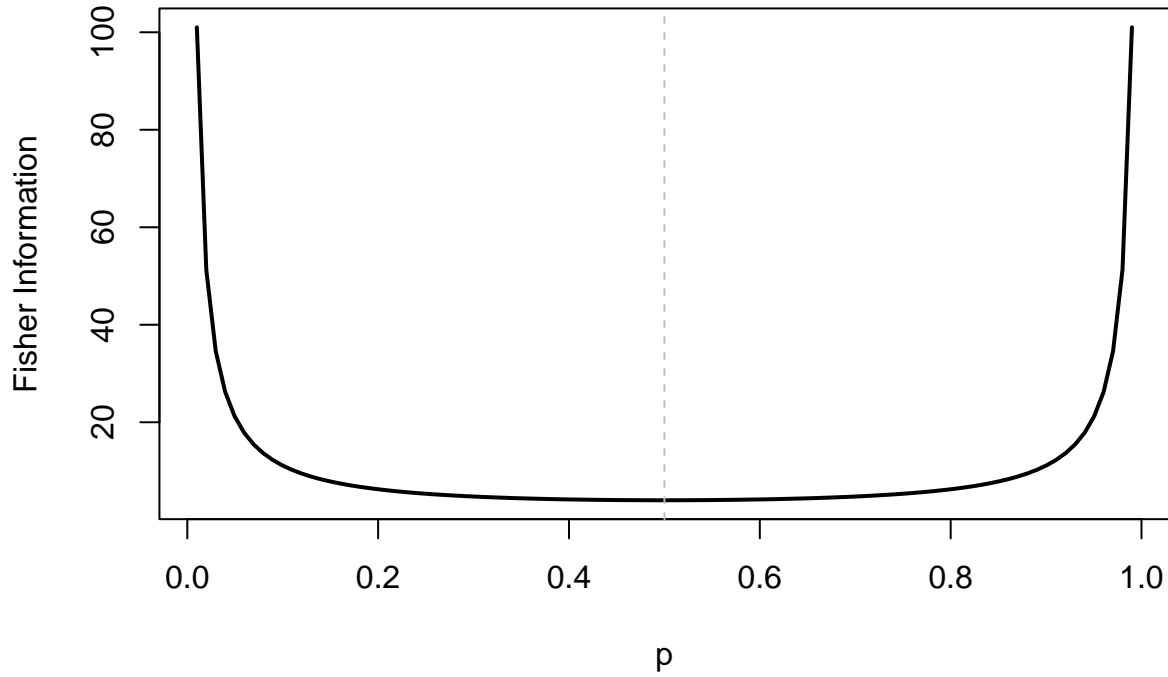
**Sketch**

```
# functional form
fisher_info <- function(p, n) {
  return(n * (1/p + 1/(1 - p)))
}

# setup
p_values <- seq(0.01, 0.99, length.out = 100)
n <- 1
I_values <- fisher_info(p_values, n)

# plot
plot(x = p_values,
     y = I_values,
     type = "l",
     col = "black", lwd = 2,
     xlab = "p", ylab = "Fisher Information",
     main = "Fisher Information for Bernoulli(p)")
abline(v = 0.5, lty = 2, col = "gray")
```

## Fisher Information for Bernoulli(p)



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 use the functional form of the Fisher Information:

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

Differentiating  $I_n(p)$  with respect to  $p$ , and setting equal to zero:

$$I_n(p)' = n \left[ -\frac{1}{p^2} + \frac{1}{(1-p)^2} \right] = -\frac{1}{p^2} + \frac{1}{(1-p)^2} = 0$$

This gives us the expression:

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

Taking square roots:

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

To ensure this is a maximum, we also check whether the second derivative is positive (since we are minimizing and not maximizing) at  $\frac{1}{2}$ :

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

$$I_n \left( \frac{1}{2} \right)'' = n \left[ \frac{2}{(1/2)^3} + \frac{2}{(1/2)^3} \right] = n \left[ \frac{2}{1/8} + \frac{2}{1/8} \right] = n [16 + 16] = 32n > 0$$

So this is in fact a minimum, hence the Fisher information is minimized at:

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

c)

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

Note to self: Uniformly Minimum Variance Unbiased Estimator (UMVUE)

We start by checking if  $\hat{X}_n$  is an unbiased estimator of  $p$ :

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

$$Bias(\bar{X}_n) = E[\hat{X}_n] - E[X] = p - p = 0$$

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

Now as far as the “Uniformly Minimum Variance” part of the question:

Note again the Fisher Information formula we’ve found:

$$I_n(p) = \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} = \frac{n}{p(1-p)}$$

By the definition, the Cramér-Rao Lower Bound, for any unbiased estimator  $T$  of  $p$ :

$$\text{Var}_p(T) \geq \frac{(\gamma'(p))^2}{I_n(p)}$$

Here, we are estimating  $\gamma(p) = p$ , so  $\gamma'(p) = 1$ . Therefore:

$$\text{Var}_p(T) \geq \frac{1^2}{I_n(p)} = \frac{p(1-p)}{n}$$

We compute the variance of  $\hat{X}_n = S_n/n$ :

$$E[\hat{X}_n] = E \left[ \frac{S_n}{n} \right] = \frac{1}{n} E[S_n] = \frac{np}{n} = p$$

$$\text{Var}(\hat{X}_n) = \text{Var} \left( \frac{S_n}{n} \right) = \frac{1}{n^2} \text{Var}(S_n)$$

Since  $S_n \sim \text{Binomial}(n, p)$ , we know:

$$\text{Var}(S_n) = np(1 - p)$$

Thus:

$$\text{Var}(\hat{X}_n) = \frac{np(1 - p)}{n^2} = \frac{p(1 - p)}{n}$$

Comparing with the CRLB:

$$\text{Var}(\hat{X}_n) = \frac{p(1 - p)}{n} = \frac{1}{I_n(p)}$$

Since  $\hat{X}_n$  is unbiased and attains the CRLB, it is the UMVUE.

## 2.

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

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

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

a)

Find the MLE of  $\beta$ .

To find the Maximum Likelihood Estimator (MLE) of  $\beta$ , 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  $Y_1, \dots, Y_n$  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:

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

To find the MLE of  $\beta$ , we take the derivative with respect to  $\beta$  and set to zero:

$$\frac{d}{d\beta} \log(L(\beta)) = -\frac{1}{2\sigma^2} \cdot (-2) \sum_{i=1}^n x_i (Y_i - \beta x_i) = \frac{1}{\sigma^2} \sum_{i=1}^n x_i (Y_i - \beta x_i) \rightarrow \sum_{i=1}^n x_i Y_i - \beta \sum_{i=1}^n x_i^2 = 0$$

Solving for  $\beta$ , we get our MLE of  $\beta$  as::

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

To ensure this is a maximum, we take the second derivative at the MLE and see if it is negative:

$$\log(L(\beta))'' = -\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2 < 0$$

So this is in fact the maximum.

b)

Find the distribution of the MLE.

From part a), the MLE of  $\beta$  is:

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

To determine the distribution of  $\hat{\beta}$ , determine its expectation and variance, noting that since  $\hat{\beta}$  is a linear combination of the normal random variables  $\varepsilon_i$ , it follows that  $\hat{\beta}$  itself is normally distributed.

That being said, given  $Y_i = \beta x_i + \varepsilon_i$ , we may write:

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

Taking the expectation, noting our data is treated as “fixed”, we may write:

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

Noting  $E[\varepsilon_i] = 0$

Because  $E[\hat{\beta}] = \beta$ , it has zero bias and  $\hat{\beta}$  is an unbiased estimator of  $\beta$ . Not needed for the distribution, but will need this note for later.

Let us then analyze the variance. We start again with definitions:

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

Simplifying:

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

We thus conclude:

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

c)

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

To find the CRLB, we first calculate the Fisher information.

Note the joint density:

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

Taking the log-likelihood:

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

We take the derivative:

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

The Fisher information is then:

$$I_n(\beta) = -E[\log(L(\beta))''] = -E\left[-\frac{1}{\sigma^2} \sum_{i=1}^n x_i^2\right] = \frac{1}{\sigma^2} \sum_{i=1}^n x_i^2$$

We then have what we need to calculate the CRLB using the information we've gathered.

The CRLB is:

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

d)

Show the MLE is the UMVUE of  $\beta$ .

Now we just need to compare the variance of our MLE of  $\beta$  to the value calculated in part c). To that end:

We have already calculated the expectation of  $\hat{\beta}_{MLE}$ , which is  $\beta$ , so via Bias calculation:

$$\text{Bias}(\hat{\beta}_{MLE}) = E[\hat{\beta}_{MLE}] - \beta = \beta - \beta = 0$$

Hence it is unbiased. We then just need to determine if our MLE attains the CRLB. If so, then the MLE is the UMVUE.

Recall the variance of the MLE:

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

And the CRLB:

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

These are one and the same! So we do indeed satisfy:

$$\text{Var}(\hat{\beta}_{MLE}) = CRLB$$

Such that the MLE is the UMVUE.



### 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  $\text{Var}_\theta(T)$  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  $Z^2 \sim \chi_1^2$ ,  $E_\theta Z^2 = 1$ ,  $\text{Var}_\theta(Z^2) = 2$ .)

Given:

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

we can rewrite  $T$  in terms of  $Z$ , specifically:

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

Taking expectation:

$$E_\theta[T] = E_\theta \left[ \frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}} + \theta^2 - \frac{1}{n} \right] = \frac{1}{n} + \frac{2\theta}{\sqrt{n}}(0) + \theta^2 - \frac{1}{n} = \theta^2$$

Thus,  $T$  is an unbiased estimator of  $\theta^2$ .

We then must calculate the variance of  $T$ , to that end, we find  $E[T^2]$ :

As defined:

$$T^2 = \left( \frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}} + \theta^2 - \frac{1}{n} \right)^2 = \frac{Z^4}{n^2} + \frac{4\theta Z^3}{n^{3/2}} + \frac{4\theta^2 Z^2}{n} + \theta^4 + \frac{1}{n^2} + \frac{4\theta^3 Z}{\sqrt{n}} - \frac{2Z^2}{n^2} - \frac{4\theta Z}{n^{3/2}} - \frac{2\theta^2}{n}$$

Though that's quite a lot, we can actually simplify it quite a bit when taking expectation, noting the distribution of  $Z$  aids in these calculations

(Note:  $E_\theta[Z] = 0$ ,  $E_\theta[Z^2] = 1$ ,  $E_\theta[Z^3] = 0$ , and  $E_\theta[Z^4] = \text{Var}(Z^2) + (E_\theta[Z^2])^2 = 2 + 1 = 3$ .)

Thus,

$$E_\theta[T^2] = \frac{3}{n^2} + \frac{4\theta^2}{n} + \theta^4 - \frac{2}{n^2} - \frac{2\theta^2}{n} = \theta^4 + \frac{2\theta^2}{n} + \frac{1}{n^2}$$

Now we can calculate the variance:

$$\text{Var}_\theta(T) = \text{Var}\left(\frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}} + \theta^2 - \frac{1}{n}\right) = \text{Var}\left(\frac{Z^2}{n} + \frac{2\theta Z}{\sqrt{n}}\right) = \frac{1}{n^2} \text{Var}(Z^2) + \frac{4\theta^2}{n} \text{Var}(Z) + \frac{2\theta}{n^{3/2}} \text{Cov}(Z^2, Z)$$

Simplifying:

$$\text{Var}_\theta(T) = \frac{2}{n^2} + \frac{4\theta^2}{n}$$

**b)**

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

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:

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

Differentiating with respect to  $\theta$ , and getting expectation to derive the Fisher Information:

$$\log(L(\theta))' = \sum_{i=1}^n (X_i - \theta) \rightarrow I_n(\theta) = -E[\log(L(\theta))'] = -E\left[-\sum_{i=1}^n 1\right] = -(-n) = n$$

The CRLB by definition is given by:

$$\frac{(\gamma'(\theta))^2}{I_n(\theta)}$$

We just need now to calculate the numerator. To that end, note that  $\gamma(\theta) = \theta^2$ , making its derivative:

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

Thus the CRLB is:

$$\frac{(\gamma'(\theta))^2}{n} = \frac{(2\theta)^2}{n} = \frac{4\theta^2}{n}$$

**c)**

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

We compare the variance we calculated from part a) with the CRLB from part b). To that end, note:

From part a):

$$\text{Var}_\theta(T) = \frac{2}{n^2} + \frac{4\theta^2}{n}$$

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

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

Comparing these two quantities directly, their difference is given by:

$$\text{Var}_\theta(T) - \text{CRLB} = \frac{2}{n^2} + \frac{4\theta^2}{n} - \frac{4\theta^2}{n} = \frac{2}{n^2} > 0$$

So, for  $n > 0$ , and  $\forall \theta \in \mathbb{R}$ , it holds that:

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

## 4. Casella & Berger 7.58

(“better” here refers to MSE as a criterion.)

Let  $X$  be an observation from the pdf

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

a)

Find the MLE of  $\theta$ .

Given that  $X$  takes values in  $\{-1, 0, 1\}$ , it is discrete, so we note the pmf:

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

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

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

Let  $S_n = \sum_{i=1}^n |X_i|$ . We may then rewrite the likelihood function as:

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

Using our log-likelihood technique:

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

We find the maximum the typical route, i.e., taking the derivative with respect to  $\theta$  and setting equal to zero:

$$\log(L(\theta))' = \frac{S_n}{\theta} - \frac{n-S_n}{1-\theta} = 0 \rightarrow \frac{S_n}{\theta} = \frac{n-S_n}{1-\theta}$$

After some simplifying:

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

And we arrive at our “MLE” (in quotes because there’s our second check to account for):

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

To double check, we take the second derivative (at the MLE) and see if it is negative:

$$\log(L(\theta))'' = -\frac{S_n}{\theta^2} - \frac{n - S_n}{(1 - \theta)^2}$$

$$\log(L(\hat{\theta}))'' = -\frac{S_n n^2}{S_n^2} - \frac{(n - S_n)n^2}{(n - S_n)^2} = -\frac{n^2}{S_n} - \frac{n^2}{n - S_n} < 0$$

Noting:  $S_n > 0$  and  $n - S_n > 0$

So yes, this is our maximum and our MLE!

Jetzt zock' ich Fortnite und trink' Cola! Yippee!

**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  $\theta$ .

To test for bias, we find the the expectation of  $T(X)$ :

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

Using the pmf from part a), each possible outcome/observation of  $X$  has its associated probability given by:

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

So we need to do the more “manual” calculation of expectation:

$$E[T(X)] = 2P(X = 1) + 0P(X = 0) + 0P(X = -1)$$

Since  $T(X) = 2$  when  $X = 1$  and 0 otherwise from the initial definition of  $T$ .

Thus, we calculate:

$$E[T(X)] = 2 \cdot \frac{\theta}{2} + 0 + 0 = \theta$$

So, via Bias calculation, we know  $T(X)$  is an unbiased estimator of  $\theta$  because  $E[T(X)] = \theta$ .

**c)**

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

By “better” we are making note of the “hint” to compare MSE, and “better” corresponding to smaller MSE compared to  $T(X)$ .

By definition, the MSE of the estimator  $T(X)$  is:

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

From part b), we know that  $T(X)$  is unbiased, so the unknown quantity in the above expression is  $E[T^2(X)]$ . Solving for that:

$$E[T^2(X)] = \sum_{x \in \{-1, 0, 1\}} T^2(x)P(X = x) = 2^2 P(X = 1) = 4 \cdot \frac{\theta}{2} = 2\theta$$

Returning to the MSE, our goal is to then find a better (smaller) MSE than  $T(X)$ , which is:

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

Our first guess will be to use the sample mean, the MLE from part a):

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

To find the relevant quantities to find its MSE, we note/derive expectation for a discrete random variable:  $E[|X|]$ :

$$E[|X|] = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) + 1 \cdot P(X = -1) = \frac{\theta}{2} + 0 + \frac{\theta}{2} = \theta$$

Next,  $E[|X|^2]$ :

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

So, the variance is:

$$\text{Var}(|X|) = E[|X|^2] - (E[|X|])^2 = \theta - \theta^2 \rightarrow \text{Var}(\hat{\theta}) = \frac{\theta - \theta^2}{n}$$

Since  $\hat{\theta}$  is unbiased, i.e.

$$E[\hat{\theta}] = E\left[\frac{S_n}{n}\right] = E\left[\frac{1}{n} \sum_{i=1}^n |X_i|\right] = \frac{1}{n} E\left[\sum_{i=1}^n |X_i|\right] = \frac{n\theta}{n} = \theta$$

The MSE of  $\hat{\theta}$  is:

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

We now comparing the two estimators:

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

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

Since  $n > 0$ :

$$\text{MSE}(T) - \text{MSE}(\hat{\theta}) = 2\theta - \theta^2 - \left( \frac{\theta - \theta^2}{n} \right) = \frac{2n\theta - n\theta^2 - \theta + \theta^2}{n}$$

for  $n = 1$ :

$$\text{MSE}(T) - \text{MSE}(\hat{\theta}) = \theta(2 - \theta) - \frac{\theta(1 - \theta)}{n} = \theta(2 - \theta) - \theta(1 - \theta) = \theta \rightarrow \text{MSE}(\hat{\theta}) < \text{MSE}(T)$$

for  $n > 1$ , noting  $0 \leq \theta \leq 1 \rightarrow 0 \leq (1 - \theta) \leq 1$  and  $\theta(2 - \theta) > \theta(1 - \theta)$ :

$$\text{MSE}(T) - \text{MSE}(\hat{\theta}) = \theta(2 - \theta) - \frac{\theta(1 - \theta)}{n} = \frac{n\theta(2 - \theta) - \theta(1 - \theta)}{n} > 0 \rightarrow \text{MSE}(\hat{\theta}) < \text{MSE}(T)$$

Taken together:

$$\text{MSE}(\hat{\theta}) < \text{MSE}(T)$$

So 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 for all values of  $\theta$  (while also being unbiased!)

## 5.

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

The likelihood function is:

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

Let  $S_n = \sum_{i=1}^n X_i$ , we may then write the above expression as:

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

Noting the prior,  $\theta \sim \text{Uniform}(0, 1)$ , the posterior is:

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

Since Beta is a known distribution, we may write the posterior density:

$$f_{\theta|S_n}(\theta) = \frac{\theta^{S_n} (1 - \theta)^{n-S_n}}{B(S_n + 1, n - S_n + 1)}$$

By definition, our Bayes estimate minimizes the posterior expected loss:

$$E_{\theta|S_n} L(h(X), \theta) = \int_0^1 \frac{(h(X) - \theta)^2}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta$$

Using the expressions detailed previously:

$$E_{\theta|S_n} L(h(X), \theta) = \int_0^1 \frac{h(X)^2 - 2h(X)\theta + \theta^2}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta$$

For simplicity, we can break us this evaluation:

$$E_{\theta|S_n} L(h(X), \theta) = h(X)^2 \int_0^1 \frac{1}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta - 2h(X) \int_0^1 \frac{\theta}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta + \int_0^1 \frac{\theta^2}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta$$

(1):

$$E_{\theta|S_n} \left[ \frac{1}{\theta(1 - \theta)} \right] = \int_0^1 \frac{1}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta = \frac{n(n-1)}{S_n(n - S_n)}$$

(2):

$$E_{\theta|S_n} \left[ \frac{\theta}{\theta(1 - \theta)} \right] = \int_0^1 \frac{\theta}{\theta(1 - \theta)} f_{\theta|S_n}(\theta) d\theta = \frac{(n-1)}{(n - S_n)}$$

(3):



$$E_{\theta|S_n} \left[ \frac{\theta^2}{\theta(1-\theta)} \right] = \int_0^1 \frac{\theta^2}{\theta(1-\theta)} f_{\theta|S_n}(\theta) d\theta = \frac{(S_n+1)(n+2)}{(S_n+1)(n-S_n+1)}$$

Combining the results of (1) through (3) gives us:

$$E_{\theta|S_n} L(h(X), \theta) = h(X)^2 \frac{n(n-1)}{S_n(n-S_n)} - 2h(X) \frac{(n-1)}{(n-S_n)} + \frac{(S_n+1)(n+2)}{(S_n+1)(n-S_n+1)}$$

Similar to the MLE method, we derive and set equal to zero:

$$\frac{d}{dh(X)} E_{\theta|S_n} L(h(X), \theta) = 2h(X) \frac{n(n-1)}{S_n(n-S_n)} - 2 \frac{(n-1)}{(n-S_n)} = 0 \rightarrow 2h(X) \frac{n(n-1)}{S_n(n-S_n)} = 2 \frac{(n-1)}{(n-S_n)}$$

Isolating  $h(X)$ :

$$h(X) = \frac{\frac{(n-1)}{(n-S_n)}}{\frac{n(n-1)}{S_n(n-S_n)}} = \frac{S_n}{n}$$

Givins us our Bayes estimator of  $\theta$ :

$$\hat{\theta}_{\text{Bayes}} = \frac{S_n}{n} = \bar{X}_n$$

Extra check time!

To confirm this is a minimum, we compute the second derivative, just like the MLE method.

Evaluating for the Bayes estimator gives us:

$$\frac{d^2}{dh(X)^2} E_{\theta|S_n} L(h(X), \theta) = 2 \frac{n(n-1)}{S_n(n-S_n)} > 0$$

So our Bayes estimator,  $\hat{\theta}_{\text{Bayes}} = \bar{X}_n$  does in fact minimize!