

# Statistical Inference and Estimation Techniques

Mohamad Lakkis

Novemeber 2024

## Bootstrap Part: Exercise 2 from Wasserman Chapter 8

We care about finding 3 kinds of CI for  $\theta$

In the code we will estimate the Coverage of these 3 confidence intervals (i.e. their bounds) and also we will estimate  $P(\theta \in CI)$

Before proceeding into getting these confidence intervals, let's talk about  $\theta = T(F)$ , which is the True Skwewness of the population. The true skewness is defined as:

$$\theta = \frac{E[(X - \mu)^3]}{\sigma^3} \quad (1)$$

Where  $X = e^Y, Y \sim N(0, 1)$ ,  $\mu$  is the true mean of  $X$ , and  $\sigma$  is the standard deviation of  $X$ .

Now we need to actually calculate this  $\theta$ , I used two methods to get it (one is exact, and the other one using Monte Carlo Approximation (in the code)).

I will start with the theoretical one.

### **Exact Calculation:**

Let us focus on the numerator of the equation (1), We have:

$$\begin{aligned} E[(X - \mu)^3] &= E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] = E[X^3] - 3\mu E[X^2] + 3\mu^2 E[X] - \mu^3 \\ &= E[X^3] - 3E[X^2]E[X] + 3E[X]^3 - E[X]^3 = E[X^3] - 3E[X^2]E[X] + 2E[X]^3 \end{aligned}$$

Now we need to calculate  $E[X^3]$ ,  $E[X^2]$ , and  $E[X]$ .

$$E[X] = E[e^Y] = M_Y(t)|_{t=0}$$

But we know that the MGF of a normal distribution is:

$$M_Y(t) = E[e^{tY}] = e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}, \text{ so for } N(0, 1) \text{ we get } M_Y(t) = e^{\frac{1}{2}t^2},$$

And now getting back to  $E[X] : E[X] = e^{\frac{1}{2}}$

Similarly, we get  $E[X^2] = E[e^{2Y}] = M_Y(t)|_{t=2} = e^2$   
and also  $E[X^3] = E[e^{3Y}] = M_Y(t)|_{t=3} = e^{9/2}$

$$E[(X - \mu)^3] = e^{9/2} - 3e^2 e^{1/2} + 2e^{3/2} = e^{9/2} - 3e^{5/2} + 2e^{3/2} \approx 62.43$$

Now let's look at the denominator of the equation (1), we have:  $\sigma^3$ , to get this we will first calculate the  $Var(X)$ :

$$Var(X) = E[X^2] - E[X]^2 = e^2 - e = e(e - 1)$$

So,  $\sigma^3 = (Var[X])^{\frac{3}{2}} = [e(e - 1)]^{\frac{3}{2}} \approx 10.1$  Thus,

$$\theta = \frac{62.43}{10.1} \approx 6.18$$

### Monte Carlo Approximation:

We can see from the result of the code that indeed the approximation is very close to the exact value.

```
1 import numpy as np
2 from scipy.stats import skew
3 n = 100000 # number of samples per set
4 N = 1000 # number of sets
5 Y = np.random.normal(0, 1, (N, n))
6 X = np.exp(Y)
7 skewness_values = skew(X, axis=1)
8 print("Skewness from Monte Carlo Simulation is: ", skewness_values.mean())
```

### Let's now move to the confidence intervals:

Note that the Skewness in each bootstrap sample is calculated as:

$$\hat{\theta}^* = T(\hat{F}_n) = \frac{\int (x - \hat{\mu})^3 d\hat{F}_n(x)}{\hat{\sigma}^3} = \frac{1}{n} \sum_{i=1}^n \frac{(X_i^* - \bar{X}_n^*)^3}{\hat{\sigma}^3}$$

, where  $\hat{\mu} = \bar{X}_n^*$ , and  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i^* - \bar{X}_n^*)^2$  This is the formula that we are using in the code ( the function called: `calc_skew_sample` )

### Normal confidence Interval:

The normal confidence interval is calculated as:

$$\hat{\theta} \pm z_{\alpha/2} \hat{\sigma}_{boot}$$

### Pivotal Confidence Interval:

The pivotal confidence interval is calculated as:

$$C_n = \left( 2\hat{\theta} - \hat{\theta}_{1-\alpha/2}^*, 2\hat{\theta} - \hat{\theta}_{\alpha/2}^* \right).$$

### Percentile Confidence Interval:

$$C_n = \left( \hat{\theta}_{\alpha/2}^*, \hat{\theta}_{1-\alpha/2}^* \right).$$

## Results

We have noticed from the results that with this small n (n=50), none of these CI are good.

My analysis is that with this small n, the bootstrap samples are not good enough to give us a good estimate of the CI of the true skewness.

This can be seen from the low probability of the true skewness being in the CI in all three cases.

```
The coverage probability for Normal Confidence Interval is: 0.129
The coverage probability for Pivotal Confidence Interval is: 0.18
The coverage probability for Percentile Confidence Interval is: 0.028
Now the estimation of Normal CI is: (1.3128672022080914, 4.046375807463007)
Now the estimation of Pivotal CI is: (1.8181126517904476, 4.311218031795157)
Now the estimation of Percentile CI is: (1.048024977875942, 3.541130357880651)
```

Figure 1: The code of this is named: `Exercise_2_Reguar.py`

**Note:** I have also tried to increase n and see if  $P(\theta \in CI)$  becomes larger, and indeed it is the case. Which is logical since we will be getting a better estimate of the sample skew in each bootsrap sample (since n is larger). And

of course a direct consequence of increasing  $n$  is that  $\hat{F}_n$  will become a better estimate of the true distribution of the data  $F$

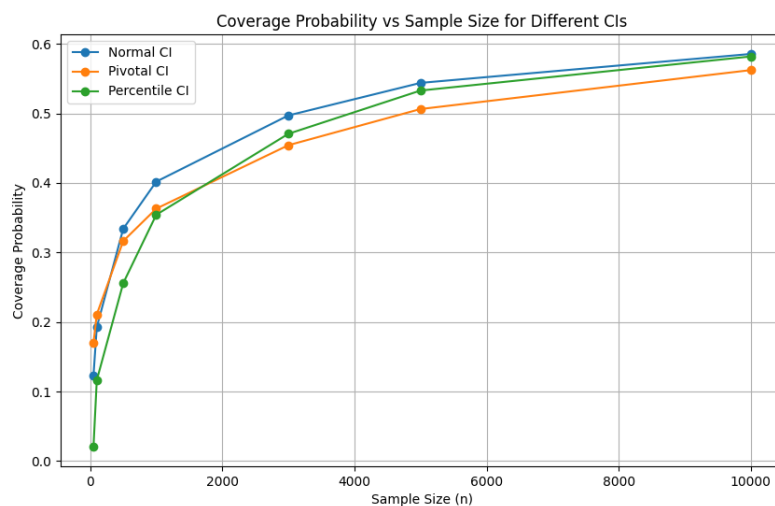


Figure 2: The Code of this is named: Exercise\_2\_many\_n.py

## Bootstrap Part: Exercise 6 from Wasserman Chapter 8

```

=====PART A=====
True Theta: 148.4131591025766
Estimated Normal CI: (131.48675115141884, 184.18861906179166)

```

Figure 3: The code of this is named: Exercise\_6\_Wasserman.py

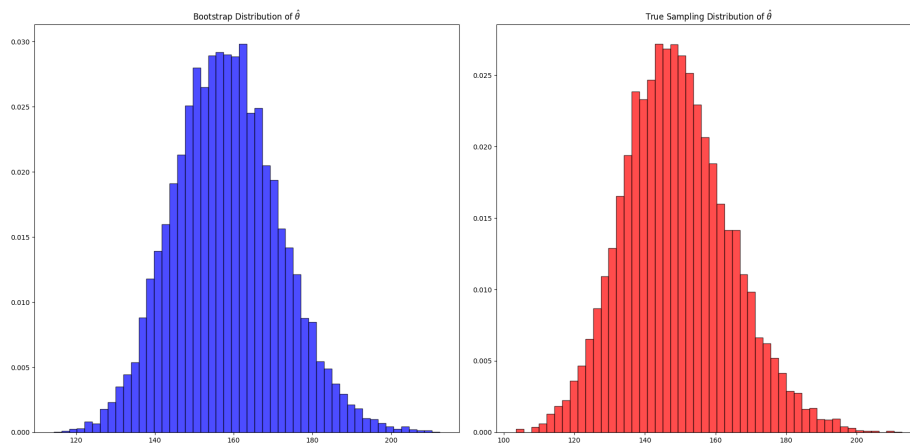


Figure 4: The code of this is named: Exercise\_6\_Wasserman.py

## Results Meaning:

We can see that both distributions are very similar, which shows us that in this example the bootstrap replications give us a good estimate of the distribution of  $\hat{\theta}$ .

## Bootstrap Part: Exercise 8 from Wasserman Chapter 8

**Goal:** Compute  $Var_{\hat{F}_n}(T_n^*)$ , where  $T_n^* = \bar{X}_n^*$

Now notice that since this is the variance with respect to the empirical distribution, this means that we are now assuming that  $X_i^* \sim \hat{F}_n$ . Reminder  $\hat{F}_n$  is the empirical distribution means that this is a discrete distribution, where we have a point mass at each  $X_j$  ( $P(X_i^* = X_j) = \frac{1}{n} \quad \forall j$ ) where  $X_i^* \sim \hat{F}_n$ ,

Note that in this exercise when I write  $Var$  this is equivalent to the variance with respect to  $\hat{F}_n$  (i.e.  $Var_{\hat{F}_n} \equiv Var$ )

$$Var[T[X^*]] = Var\left[\frac{1}{n} \sum_{i=1}^n X_i^*\right] = \frac{1}{n} Var[X_i^*]$$

Let us now compute  $Var[X_i^*]$ :

$$Var[X_i^*] = E[(X_i^* - E[X_i^*])^2]$$

We have that  $E[X_i^*] = \sum_{j=1}^n P(X_i^* = X_j) = \frac{1}{n} \sum_{j=1}^n X_j = \bar{X}_n$ , and so plugging this in

$$Var[X_i^*] = E[(X_i^* - \bar{X}_n)^2] = \sum_{j=1}^n P(X_i^* = X_j)(X_i^* - \bar{X}_n)^2 = \frac{1}{n} \sum_{j=1}^n [X_i^* - \bar{X}_n]^2$$

And so,

$$Var[X_i^*] = \hat{\alpha}_2$$

With  $\hat{\alpha}_2 = \frac{1}{n} \sum_{j=1}^n [X_i^* - \bar{X}_n]^2$  (the sample variance of the bootstrap sample) Now going back to the initial variance, and plugging this in we get:

$$Var[T(X^*)] = \frac{\hat{\alpha}_2}{n}$$

**Note:** So what we did here is that we have computed the exact value of this quantity, but usually especially with harder statistics  $T_n$ , this quantity might be hard to get so we use the bootstrap to estimate it. In other words we approximate it as so:  $Var_{\hat{F}_n}(T_n^*) \approx Var_{boot}(T_n^*)$ .

Lastly notice that this shows that sometimes we can escape from approximating this quantity from bootstrap (if it is easy to compute it like in this problem), but the main approx (that is  $F$  by  $\hat{F}_n$ ) is inevitable in this kind of setup!

## Problem 1 in section [Score function, Fisher Information, MLE estimations, Properties of the MLE]

**Showing that**  $X = \theta \sqrt{-2 \log(U)}$

In order to show the following result we will proceed with the inverse sampling method, which states that if  $U \sim U(0, 1)$ , then  $F^{-1}(U) \sim F$  (where  $F$  is the CDF of the random variable that we are interested in). So let's get the CDF of this distribution; **Side Note: We can actually integrate this**

pdf and indeed we get 1

$$\begin{aligned} F(x) &= \int_0^x \frac{t}{\theta^2} e^{-\frac{t^2}{2\theta^2}} dt, \quad \theta > 0 \\ &= \int_0^x \frac{1}{\theta^2} t e^{-\frac{t^2}{2\theta^2}} dt = \left[ -e^{-\frac{t^2}{2\theta^2}} \right]_0^x = 1 - e^{-\frac{x^2}{2\theta^2}} \end{aligned}$$

Now we need to find the inverse of this function, which is:

$$U = 1 - e^{-\frac{x^2}{2\theta^2}} \Rightarrow e^{-\frac{x^2}{2\theta^2}} = 1 - U \Rightarrow -\frac{x^2}{2\theta^2} = \log(1 - U) \Rightarrow x^2 = -2\theta^2 \log(1 - U)$$

But now notice that  $1 - U$  can be changed to  $U$  since  $1 - U \sim U(0, 1)$  (it is just the same from the symmetry of Uniform), so we get:

$$X = \theta \sqrt{-2\log(U)}$$

**Showing that  $X^2 \sim \text{Exp}(\frac{1}{2\theta^2})$**

This is straight forward to prove especially since  $X$  is 1-dimensional, we will proceed as in notes 1.

Let  $Y = g(X) = X^2$ . We can see that the support of  $Y$  is  $y > 0$ .

We just showed that  $F_X(x) = 1 - e^{-\frac{x^2}{2\theta^2}}$

Let us now get the CDF of  $Y$ :

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) \\ &= F_X(\sqrt{y}) = 1 - e^{-\frac{y}{2\theta^2}} \end{aligned}$$

Notice that is the the CDF of an  $\text{Exp}(\lambda) = 1 - e^{-\lambda x}$ , with  $\lambda = \frac{1}{2\theta^2}$

Thus,  $Y = X^2 \sim \text{Exp}(\frac{1}{2\theta^2})$

**MLE of  $\theta^2$  (NOTE: I Will denote by  $\hat{\theta}_{MLE}^2$ ) (This doesn't mean that I will square the MLE!!!), I know it is the same from the equivariant property but just to avoid confusion**

Suppose we have an i.i.d sample  $X_1, X_2, \dots, X_n$  from the distribution  $F$  from above. The likelihood function is:

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

taking the log of this function we get:

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

Now in order to get the MLE of  $\theta^2$  we need to take the derivative of this function with respect to  $\theta^2$  and set it to zero.

$$\frac{d\ell(\theta^2)}{d\theta^2} = -\frac{n}{\theta^2} + \frac{1}{2(\theta^2)^2} \sum_{i=1}^n X_i^2 = 0$$

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

And so we get:

$$\hat{\theta}_{MLE}^2 = \frac{1}{2n} \sum_{i=1}^n X_i^2$$

**Proving that  $\hat{\theta}_{MLE}^2 \sim \Gamma(\alpha = n, \beta = \frac{\theta^2}{n})$**

First we know that  $Exp(\frac{1}{2\theta^2}) \equiv \Gamma(\alpha = 1, \beta = 2\theta^2)$ , so now let us find the distribution of  $S_n = \sum_{i=1}^n X_i^2$

This is a straight forward result since we know that this also follows a Gamma distribution since all  $X_i^2$  are i.i.d, and so  $S_n \sim \Gamma(\alpha = n, \beta = 2\theta^2)$ , now using the properties of the Gamma distribution scaling this by  $\frac{1}{2n}$ , will only affect  $\beta$  and Thus:

$$\hat{\theta}_{MLE}^2 = \frac{1}{2n} S_n \sim \Gamma(\alpha = n, \beta = \frac{2\theta^2}{2n}) \equiv \Gamma(\alpha = n, \beta = \frac{\theta^2}{n})$$

**Proving that  $\hat{\theta}_{MLE}^2$  is unbiased**

**IMPORTANT NOTE** I did this part (i.e. proving unbiased and getting variance and MSE) before the new given, I will use these new given to verify my results from before

**OLD:**



$$E(\hat{\theta}_{MLE}^2) = E\left(\frac{1}{2n} \sum_{i=1}^n X_i^2\right) = \frac{1}{2n} E\left(\sum_{i=1}^n X_i^2\right)$$

But now since  $X_i$  are independent,

$$= \frac{1}{2n} \sum_{i=1}^n E(X_i^2)$$

And since they are identically distributed,  $E(X_i^2)$  is the same for all  $i$ , so we get:

$$\frac{1}{2n} n E(X_1^2) = \frac{1}{2} E(X_1^2)$$

Now let's calculate  $E[X_1^2]$ :

$$E[X_1^2] = \int_0^\infty x^2 \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} dx = \int_0^\infty \frac{x^3}{\theta^2} e^{-\frac{x^2}{2\theta^2}} dx$$

We will make the substitution  $u = -\frac{x^2}{2\theta^2}$ :

$$\begin{aligned} x &= \theta\sqrt{-2u}, & dx &= \frac{\theta\sqrt{-2u}}{2u} du \\ E[X_1^2] &= \frac{1}{\theta^2} \int_A^B \theta^3(-2u)\sqrt{-2u}e^u \frac{\theta\sqrt{-2u}}{2u} du \\ &= 2\theta^2 \int_A^B u e^u du \\ &= 2\theta^2 [(e^u u - e^u)|_A^B] \\ &= 2\theta^2 [e^{-\frac{x^2}{2\theta^2}} (-\frac{x^2}{2\theta^2}) - e^{-\frac{x^2}{2\theta^2}}] \Big|_0^\infty \\ &\quad - 2\theta^2 [0 - 1] = 2\theta^2 \end{aligned}$$

Now going back to the expectation of  $\hat{\theta}_{MLE}^2$ :

$$E[\hat{\theta}_{MLE}^2] = \frac{1}{2} E[X_1^2] = \frac{1}{2} 2\theta^2 = \theta^2$$

Thus  $Bias(\hat{\theta}_{MLE}) = E[\hat{\theta}_{MLE}^2] - \theta^2 = \theta^2 - \theta^2 = 0$ , and so we just proved that the MLE estimator  $\hat{\theta}_{MLE}$  is an unbiased estimator of  $\theta^2$ .

**Getting**  $Var(\hat{\theta}_{MLE}^2)$

$$Var(\hat{\theta}_{MLE}^2) = Var\left(\frac{1}{2n} \sum_{i=1}^n X_i^2\right) = \frac{1}{4n^2} \sum_{i=1}^n Var(X_i^2) = \frac{1}{4n^2} n Var(X_1^2) = \frac{1}{4n} Var(X_1^2)$$

Now we need to compute  $Var(X_1^2)$ ;

$$Var(X_1^2) = E[X_1^4] - E[X_1^2]^2$$

But notice that the second term is already computed in the previous section  $E[X_1^2] = 2\theta^2$ , and so we have  $E[X_1^2]^2 = 4\theta^4$  so we just need to compute  $E[X_1^4]$ .

$$Var(X_1^2) = E[X_1^4] - 4\theta^4$$

$$E[X_1^4] = \int_0^\infty x^4 \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}} dx = \int_0^\infty \frac{x^5}{\theta^2} e^{-\frac{x^2}{2\theta^2}} dx$$

We proceed similarly to the previous section, we make the substitution  $u = \frac{x^2}{2\theta^2}$ :

$$E[X_1^4] = \frac{1}{\theta^2} \int_A^B \theta^5 (\sqrt{2u})^5 e^{-u} \frac{\theta \sqrt{2u}}{2u} du$$

$$= 4\theta^4 \int_A^B u^2 e^{-u} du$$

By integration by parts we get:

$$4\theta^4 [e^{-u}(-u^2 - 2u - 2)]|_A^B = -4\theta^4 [e^{-\frac{x^2}{2\theta^2}} (\frac{x^4}{4\theta^4} + \frac{x^2}{\theta^2} + 2)]|_0^\infty$$

$$= -4\theta^4 [0 - 2] = 8\theta^4$$

And so we have found:

$$E[X_1^4] = 8\theta^4$$

And going back to the variance of  $X_1^2$ :

$$Var(X_1^2) = E[X_1^4] - E[X_1^2]^2 = 8\theta^4 - 4\theta^4 = 4\theta^4$$

And Thus we get finally plug in this value in the Variance of  $\hat{\theta}_{MLE}^2$ :

$$Var(\hat{\theta}_{MLE}^2) = \frac{1}{4n} Var(X_1^2) = \frac{1}{4n} 4\theta^4 = \frac{\theta^4}{n}$$

**Getting the MSE of  $\hat{\theta}_{MLE}^2$ :**

$$MSE(\hat{\theta}_{MLE}^2) = Var(\hat{\theta}_{MLE}^2) + Bias(\hat{\theta}_{MLE}^2)^2 = \frac{\theta^4}{n}$$

Since we have shown from before that the bias of this estimator is zero.

**New Given(Much easier)**

**Proving Unbiased**

Since we know the distribution of  $\hat{\theta}_{MLE}^2 \sim \Gamma(n, \frac{\theta^2}{n})$ , we know that  $E[\hat{\theta}_{MLE}^2] = n \frac{\theta^2}{n} = \theta^2$ .

Thus we just proved that  $\hat{\theta}_{MLE}^2$  is an Unbiased estimator of  $\theta^2$

**Getting  $Var(\hat{\theta}_{MLE}^2)$**

Also very direct since we know the distribution of  $\hat{\theta}_{MLE}^2$ , and so:

$$Var[\hat{\theta}_{MLE}^2] = n \left(\frac{\theta^2}{n}\right)^2 = \frac{\theta^4}{n}$$

Which verifies our result from before!

**Getting the MSE of  $\hat{\theta}_{MLE}^2$ :**

$$MSE(\hat{\theta}_{MLE}^2) = Var(\hat{\theta}_{MLE}^2) + Bias(\hat{\theta}_{MLE}^2)^2 = \frac{\theta^4}{n}$$

Since we have shown from before that the bias of this estimator is zero.

## Cramér-Rao Bound (CR Bound) and UMVUE

First Let us find the CR bound

$$var(\hat{\theta}_{MLE}^2) \geq \frac{\left(\frac{d}{d\theta}b(\theta^2) + \frac{d}{d\theta}\theta^2\right)^2}{nI(\theta)}$$

Where  $b(\theta^2) = Bias(\hat{\theta}_{MLE}^2)$ , but notice we know the bias =0, and we have  $var(\hat{\theta}_{MLE}^2)$ , and so:

$$\frac{\theta^4}{n} \geq \frac{4\theta^2}{nI_X(\theta)}$$

So let us now find the fisher information of  $\theta$ :

$$I_X(\theta) = -E\left[\frac{\partial^2 \ell(\theta)}{\partial \theta^2}\right]$$

Where  $\ell(\theta) = \log(X) - 2\log(\theta) - \frac{1}{2\theta^2}X^2$ , the log-likelihood function of  $\theta$  for ONE OBSERVATION.

Taking the second derivative of  $\ell(\theta)$  w.r.t. to  $\theta$  we find:

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = \frac{2}{\theta^2} - \frac{3}{\theta^4}X^2$$

Now plugging this into fisher information:

$$I_X(\theta) = -\frac{2}{\theta^2} + \frac{3}{\theta^4}E[X^2]$$

But we know that  $X^2 \sim \exp(\frac{1}{2\theta^2})$ , and so  $E[X^2] = 2\theta^2$

$$I_X(\theta) = -\frac{2}{\theta^2} + \frac{6}{\theta^2} = \frac{4}{\theta^2}$$

And so we get:

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

Finally going back to the CR bound:

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

$$\frac{\theta^4}{n} \geq \frac{\theta^4}{n}$$

Which is True,  $\forall \theta$ !

And so Indeed it verifies the CR Bound.

In fact it is the best unbiased estimator of  $\theta^2$  (when working with MSE) since it achieves the CR bound. So we have reached the minimum value possible of the MSE, under the unbiasedness constraint. Which means that the  $\hat{\theta}_{MLE}^2$  is an **Efficient** and **UMVUE** (i.e. uniformly minimum variance unbiased estimator) estimator of  $\theta^2$ .

## Asymptotic Properties of the $\hat{\theta}_{MLE}^2$

### Asymptotic Normality

We know:  $\hat{\theta}_{MLE}^2 = \frac{1}{2} \bar{X}^2$ .

We will use CLT to prove this, it is straight forward (unformally "we will take  $X^2$  as our variable instead of  $X$ ", meaning we will apply the CLT to  $\bar{X}^2$ ), we know from the CLT that:

$$\frac{(\bar{X}^2 - E[X^2])}{\sqrt{V[\bar{X}^2]}} \xrightarrow{d} N(0, 1)$$

Now we also know that  $V[\bar{X}^2] = \frac{1}{n} 4\theta^4$ , and  $E[X^2] = 2\theta^2$ , plugging these in we get:

$$\frac{\sqrt{n}(\bar{X}^2 - 2\theta^2)}{2\theta^2} \xrightarrow{d} N(0, 1)$$

$$\sqrt{n}(\bar{X}^2 - 2\theta^2) \xrightarrow{d} 2\theta^2 N(0, 1) \equiv N(0, 4\theta^4)$$

Multiply by a constant ( $\frac{1}{2}$ ) (which gets applied to both sides)

$$\sqrt{n}(\hat{\theta}_{MLE}^2 - \theta^2) \xrightarrow{d} N(0, \theta^4)$$

And so we just proved that the  $\hat{\theta}_{MLE}^2$ , is asymptotically normal (corrected by  $\sqrt{n}$ ).

### Asymptotic Optimality

To do this manually we need to show two things: Asymptotic Normality (already proved) and Asymptotic UMVUE (Meaning asymptotically unbiased and has the minimum variance among all asymptotically unbiased estimator of  $\theta^2$ ). So more formally we need to show that:

**1. Asymptotically Normal:**(Already Proven)

**2. Asymptotically Unbiased:**

This is a direct result from proving asymptotic normality, because the mean of the limiting distribution is 0 (and so as discussed in the notes) this shows Asymptotic unbiasedness

Notice that when we talk about asymptotic bias of the estimator, we are talking about the bias of the limiting distribution of the estimator. AND NOT

$$\lim_{n \rightarrow \infty} [E(\hat{\theta}_{MLE}^2)] = \theta^2$$

This is as discussed in the notes, is unbiasedness in the limit!!

**3. Asymptotically Minimum Variance(i.e. equal to the CR Bound):**

Similarly here, we need to prove, that the following inequality is a straight equality:

$$v(\theta) \geq \frac{\frac{dg(\theta)}{d\theta}}{I(\theta)}$$

Where g here is the square function since we are estimating  $\theta^2$ . We know from above that the asymptotic variance is this:

$$v(\theta) = \theta^4$$

Now let us get the RHS:

$$\frac{\frac{dg(\theta)}{d\theta}}{I(\theta)} = \frac{\theta^2}{4} \left[ \frac{dg(\theta)}{d\theta} \right]^2 = \frac{\theta^2}{4} \left[ \frac{d\theta^2}{d\theta} \right]^2 = \frac{\theta^2}{4} [2\theta]^2 = \theta^4$$

Which is exactly equal to the asymptotic variance, and so we have shown that the MLE of  $\theta^2$  is asymptotically optimal.

Also notice that when we talk about the asymptotic variance of the estimator, it is the variance of the limiting distribution, NOT the variance in the limit

$$\lim_{n \rightarrow \infty} [Var(\hat{\theta}_{MLE}^2)]$$

## Consistency

We care to show that  $\hat{\theta}_{MLE}^2 \xrightarrow{P} \theta^2$

By the Definition of Convergence in Prob. We know we need to prove:

$$P(|\hat{\theta}_{MLE}^2 - \theta^2| > \epsilon) \xrightarrow{P} 0, \quad \forall \epsilon > 0$$

Now I will use the notation of  $Y = X^2$ , and so we have:

$$P(|\frac{1}{2}\bar{Y}_n - \theta^2| > \epsilon) = P(|\bar{Y}_n - 2\theta^2| > 2\epsilon)$$

But Notice that  $E(\bar{Y}_n) = 2\theta^2$ , and so by the Weak Law of Large Numbers (WLLN) we know that:

$$\bar{Y}_n \xrightarrow{P} 2\theta^2$$

Which means that:

$$P(|\bar{Y}_n - 2\theta^2| > 2\epsilon) \xrightarrow{P} 0$$

And so we have shown:

$$P(|\hat{\theta}_{MLE}^2 - \theta^2| > \epsilon) \xrightarrow{P} 0, \quad \forall \epsilon > 0$$

Thus,  $\hat{\theta}_{MLE}^2 \xrightarrow{P} \theta^2$

## MLE of $\theta$

We will proceed like in the previous section, but now we will find the MLE of  $\theta$  instead of  $\theta^2$ .

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

Now taking the derivative and setting it to zero:

$$\frac{d\ell(\theta)}{d\theta} = -\frac{2n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^n X_i^2 = 0$$

And so we get:

$$\hat{\theta}_{MLE} = \sqrt{\frac{1}{2n} \sum_{i=1}^n X_i^2}$$

We can notice that this is the square root of the MLE of  $\theta^2$ : ( $\hat{\theta}_{MLE}^2$ ) that we have found before. (This is the Equivariant Property of the MLE)

$$\hat{\theta}_{MLE} = \sqrt{\hat{\theta}_{MLE}^2}$$

### Finding the Distribution of $\hat{\theta}_{MLE}$

For the sake of simple notations I will take  $U = \hat{\theta}_{MLE}^2$  and  $V = \hat{\theta}_{MLE}$ , so now our goal is to find the distribution of  $V$ .

We will proceed by the method of transformation.

Re-stating what we will do, we have the pdf of r.v.  $U$ , we care now to find the pdf of a transformation of this r.v.  $V = \sqrt{U}$ .

We know that  $U \sim \Gamma(n, \frac{\theta^2}{n})$

$$f_U(u) = \frac{n^n}{\Gamma(n)\theta^{2n}} u^{n-1} e^{-\frac{nu}{\theta^2}}, \quad u > 0$$

In this case I won't use the same method that I used in the start of the previous section, because here it is easier to work with pdf rather than cdf.

This method tell us that if  $V = g(U)$  and  $U = h(V)$ , then:

$$f_V(v) = f_U(h(v)) \left| \frac{dh(v)}{dv} \right|$$

In our case  $h(v) = u = v^2$ , and so  $f_V(v) = f_U(v^2) \left| \frac{d(v^2)}{dv} \right| = f_U(v^2) 2v$

$$f_V(v) = f_U(v^2) 2v = \frac{n^n}{\Gamma(n)\theta^{2n}} (v^2)^{n-1} e^{-\frac{n(v^2)}{\theta^2}} 2v$$

Thus, we have found the distrbition of  $\hat{\theta}_{MLE}$ , which is:

$$f_V(v) = \frac{2n^n}{\Gamma(n)\theta^{2n}} v^{2n-1} e^{-\frac{nv^2}{\theta^2}}, \quad v > 0$$

### Verifying the equation of the moments

We have the PDF of  $\hat{\theta}_{MLE}$ :

$$f_{\hat{\theta}_{MLE}}(v) = \frac{2n^n}{\Gamma(n)\theta^{2n}} v^{2n-1} e^{-\frac{nv^2}{\theta^2}}, \quad v > 0.$$

And so by definition the moment function is:

$$E(\hat{\theta}_{MLE}^r) = \int_0^\infty v^r f_{\hat{\theta}_{MLE}}(v) dv.$$
$$E(\hat{\theta}_{MLE}^r) = \frac{2n^n}{\Gamma(n)\theta^{2n}} \int_0^\infty v^{r+2n-1} e^{-\frac{nv^2}{\theta^2}} dv.$$



We now solve this similar to the previous section: take substitution  $u = \frac{nv^2}{\theta^2}$ ,  
 $\rightarrow v = \frac{\theta}{\sqrt{n}}\sqrt{u} \rightarrow dv = \frac{\theta}{2\sqrt{nu}} du$ .

Plugging this into our integral and getting out of the integral the constants:

$$E(\hat{\theta}_{MLE}^r) = \frac{2n^n}{\Gamma(n)\theta^{2n}} \cdot \left(\frac{\theta}{\sqrt{n}}\right)^{r+2n-1} \cdot \frac{\theta}{2\sqrt{n}} \int_0^\infty u^{(r+2n-2)/2} e^{-u} du.$$

Now just simplifying the constant terms (lots of cancelation, to make it look nicer and easier to work with)

$$E(\hat{\theta}_{MLE}^r) = \frac{n^{-r/2}\theta^r}{\Gamma(n)} \int_0^\infty u^{(r+2n-2)/2} e^{-u} du.$$

$$E(\hat{\theta}_{MLE}^r) = \frac{n^{-r/2}\theta^r}{\Gamma(n)} \int_0^\infty u^{(\frac{r+2n-2}{2})} e^{-u} du.$$

I noticed that it is similar to a gamma function, so I will try to make it explicitly look like it: First we know that the Gamma Function is:

$$\int_0^\infty u^{\alpha-1} e^{-u} du = \Gamma(\alpha), \quad \text{Equation(1)}$$

Let us define the exponent of  $u$  as:

$$\alpha - 1 = \frac{r + 2n - 2}{2} \rightarrow \alpha = \frac{r + 2n}{2}.$$

Now, the integral becomes:

$$E(\hat{\theta}_{MLE}^r) = \frac{n^{-r/2}\theta^r}{\Gamma(n)} \int_0^\infty u^{\alpha-1} e^{-u} du.$$

where  $\alpha = \frac{r+2n}{2}$ , we obtain:

$$E(\hat{\theta}_{MLE}^r) = \theta^r n^{-r/2} \frac{\Gamma\left(\frac{r+2n}{2}\right)}{\Gamma(n)}.$$

And Thus, we have shown that the moments of the MLE of  $\theta$  are:

$$E(\hat{\theta}_{MLE}^r) = \left(\frac{\theta}{\sqrt{n}}\right)^r \frac{\Gamma\left(n + \frac{r}{2}\right)}{\Gamma(n)}.$$

### Getting the Bias of $\hat{\theta}_{MLE}$

Now our work is very easy since we have  $E(\hat{\theta}_{MLE}^r)$ , we can easily get the bias of this estimator.

$$E[\hat{\theta}_{MLE}] = E[\hat{\theta}_{MLE}^1] = \frac{\theta}{\sqrt{n}} \frac{\Gamma(\frac{1}{2} + n)}{\Gamma(n)}$$

$$Bias(\hat{\theta}_{MLE}) = E[\hat{\theta}_{MLE}] - \theta = \frac{\theta}{\sqrt{n}} \frac{\Gamma(\frac{1}{2} + n)}{\Gamma(n)} - \theta$$

Just a quick note, notice that here the  $MLE$  is biased for finite  $n$ . (Note: that is okay since the MLE requires Consistency asymptotically, and so it is okay to have bias for finite  $n$ ) we will show later that indeed this  $MLE$  is consistent

### Getting the Variance of $\hat{\theta}_{MLE}$

$$Var(\hat{\theta}_{MLE}) = E[\hat{\theta}_{MLE}^2] - E[\hat{\theta}_{MLE}]^2 = \theta^2 - \frac{\theta^2}{n} \left( \frac{\Gamma(\frac{1}{2} + n)}{\Gamma(n)} \right)^2$$

For simplicity I will denote the term  $\frac{\Gamma(\frac{1}{2} + n)}{\Gamma(n)}$  as  $A$ , and so we get:

$$Var(\hat{\theta}_{MLE}) = \theta^2 - \frac{\theta^2}{n} A^2$$

### Getting the MSE of $\hat{\theta}_{MLE}$

$$MSE(\hat{\theta}_{MLE}) = Var(\hat{\theta}_{MLE}) + Bias(\hat{\theta}_{MLE})^2 = \theta^2 - \frac{\theta^2}{n} A^2 + \left( \frac{\theta}{\sqrt{n}} A - \theta \right)^2$$

$$MSE(\hat{\theta}_{MLE}) = \theta^2 \left[ 2 - \frac{2A}{\sqrt{n}} \right]$$

### CR Bound

We know from before that

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

Now we need to check if:

$$Var(\hat{\theta}_{MLE}) \geq \frac{\theta^2}{4n} (1 + \frac{d}{d\theta} Bias(\hat{\theta}_{MLE}))^2$$

We have:

$$Var(\hat{\theta}_{MLE}) = \theta^2 - \frac{\theta^2}{n} A^2 = \theta^2 (1 - \frac{A^2}{n})$$

And

$$CR = \frac{\theta^2}{4n} (1 + \frac{d}{d\theta} Bias(\hat{\theta}_{MLE}))^2 = \frac{\theta^2}{4n} (1 + \frac{d}{d\theta} (\frac{\theta}{\sqrt{n}} A - \theta))^2$$

$$CR = \frac{\theta^2}{4n} (\frac{1}{\sqrt{n}} A)^2$$

$$CR = \frac{A^2 \theta^2}{4n^2}$$

And so we need to check if:

$$\theta^2 (1 - \frac{A^2}{n}) \geq \frac{A^2 \theta^2}{4n^2}$$

Which reduces to:

$$(1 - \frac{A^2}{n}) \geq \frac{A^2}{4n^2}$$

Let us compare them via simulation:

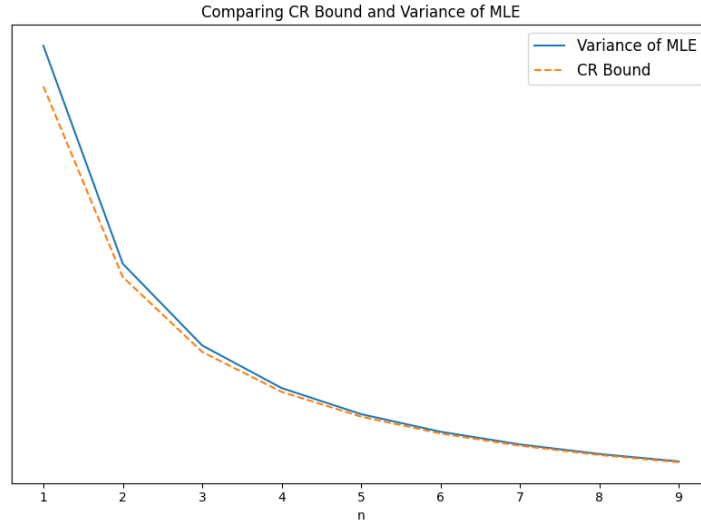


Figure 5: Code in Comparing\_CR\_Bound\_VAR\_MLE\_THETA.py

We can see how for  $\forall n > 0$  this inequality holds. So this is a verification of the CR Bound. Furthermore we can see how as  $n$  increases the bound becomes tighter and tighter, meaning that the variance of the MLE is getting closer to the CR Bound. (also a note from this graph is that we can see that when increasing  $n$  the variance of the MLE becomes smaller as expected since we are using a large sample from the data )

And so to answer the question it doesn't reach the CR bound exactly ( i.e. becoming an equality ), but we expect that as  $n$  gets bigger they will converge as shown in the graph above.

## Asymptotic Properties of the $\hat{\theta}_{MLE}$ (First Directly (i.e. not relying on $\hat{\theta}_{MLE}^2$ ))

Note that in this section I relied more on simulations rather than actual proof, honestly working with Gamma functions is hard, especially when dealing with limits.

### Asymptotic Normality

To show this formally we need to prove:

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, v(\theta)) \quad , \forall \theta$$

Where  $v(\theta)$  is the asymptotic variance of the limiting distribution.

And we know from notes3 page 21, that under regularity conditions of the MLE, the asymptotic variance of the limiting distribution is equal to the inverse of the information number (i.e.  $v(\theta) = \frac{1}{I(\theta)}$ ), and so we need to show that:

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, \frac{1}{I(\theta)})$$

Side note (from this we know that the asymptotic variance of  $\hat{\theta}_{MLE} = \frac{1}{I(\theta)}$ ) (when adjusted by  $\sqrt{n}$ )

Now like I mentioned before, I will rely more on simulations rather than actual proof.

So to prove this let me first discuss the set-up of the simulation:

We will sample from the distribution of  $X^2$  (i.e.  $Exp(\lambda = \frac{1}{2\theta^2})$ ),  $n$  (a large

$n$  to mimic the  $\infty$ ) i.i.d observations and then calculate the MLE of  $\theta$  using the formula obtained before  $\hat{\theta}_{MLE} = \sqrt{\frac{1}{2n} \sum_{i=1}^n X_i^2}$  and then repeat this process many times. Then we will have many values of  $\hat{\theta}_{MLE}$  and each one of them we will multiply it by its  $(\sqrt{n})$  and subtract  $\theta$  which we will then draw a histogram and compare it to  $N(0, \frac{1}{I(\theta)}) \equiv N(0, \frac{\theta^2}{4})$  (since we already know the value of  $I(\theta)$ )

Note that I will do this for one value of  $\theta$ , but the same logic applies if we wanted to choose many  $\theta$ , because the MLE should converge  $\forall \theta$ , I am just doing for one  $\theta$  just to illustrate the concept.

From the graph below we can see how the histogram of the adjusted MLE of  $\theta$  is very close to the normal distribution with the mean 0 and variance  $\frac{\theta^2}{4}$ , which verifies the asymptotic normality of the MLE of  $\theta$ . (Note this simulation was done with  $\theta = 1$ ). And in the code as we increase  $n$  they become to look a like even more, which verifies the convergence in distribution!!

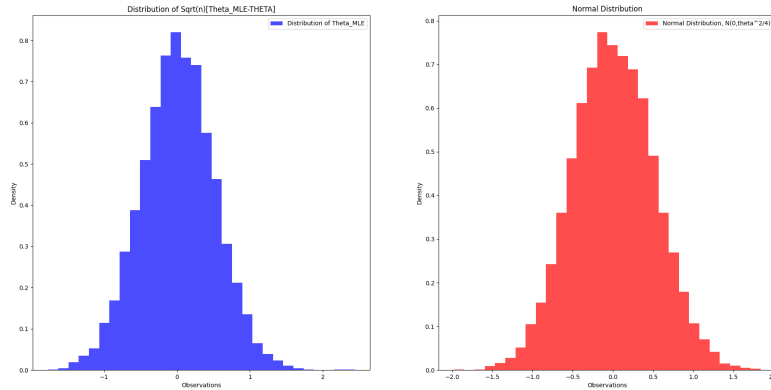


Figure 6: Code in distribution\_of\_theta\_MLE.py

## Asymptotic Optimality (Efficient)

To do this manually we need to show two things: Asymptotic Normality (already proved) and Asymptotic UMVUE (Meaning asymptotically unbiased and has the minimum variance among all asymptotically unbiased estimator of  $\theta$ ). So more formally we need to show that:

1. **Asymptotically Normal:** (Already Proven)
2. **Asymptotically Unbiased:**

This is a direct result from proving asymptotic normality, because the mean of the limiting distribution is 0 (and so as discussed in the notes) this shows Asymptotic unbiasedness

Notice that when we talk about asymptotic bias of the estimator, we are talking about the bias of the limiting distribution of the estimator. AND NOT

$$\lim_{n \rightarrow \infty} [E(\hat{\theta}_{MLE})] = \theta$$

This is as discussed in the notes, is unbiasedness in the limit!!

### **3. Asymptotically Minimum Variance(i.e. equal to the CR Bound):**

Similarly here, we need to prove, that the following inequality is a straight equality:

$$v(\theta) \geq \frac{\frac{dg(\theta)}{d\theta}}{I(\theta)}$$

Where g here is the identity function since we are estimating  $\theta$  and not a function of it. We know from above that the asymptotic variance adjusted is this:

$$v(\theta) = \frac{\theta^2}{4}$$

Which is exactly equal to the RHS of the inequality, and so we have shown that the MLE of  $\theta$  is asymptotically optimal.

Also notice that when we talk about the asymptotic variance of the estimator, it is the variance of the limiting distribution, NOT the variance in the limit

$$\lim_{n \rightarrow \infty} [Var(\hat{\theta}_{MLE})]$$

### **Consistency**

We need to show that:

$$\hat{\theta}_n \xrightarrow{p} \theta, \forall \theta$$

Note: here I am writing  $\hat{\theta}_n$  to make it explicit that this is convergence of the sequence of MLEs to the True parameter  $\theta$ .

We have  $\hat{\theta}_n = \sqrt{\frac{1}{2n} \sum_{i=1}^n X_i^2}$  We know that  $E[X_i^2] = 2\theta^2$ , and from it we know that  $E[\frac{1}{n} \sum_{i=1}^n X_i^2] = 2\theta^2$ .

Now by the WLLN we have that:

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{d} E[\frac{1}{n} \sum_{i=1}^n X_i^2] = 2\theta^2$$

Now notice that  $2\theta^2$  is just a constant and so we can use from theorem 5.4 from Wasserman's book:

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} 2\theta^2$$

Now also using properties in theorem 5.5 (property f) we have that, if  $g$  is a continuous function (here I will take  $g(x) = \sqrt{\frac{x}{2}}$ ) which is continuous on  $x > 0$ :

$$g\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) \xrightarrow{P} g(2\theta^2)$$

$$\sqrt{\frac{1}{2} \frac{1}{n} \sum_{i=1}^n X_i^2} \xrightarrow{P} \theta$$

And so:

$$\hat{\theta}_n \xrightarrow{P} \theta$$

Thus we have shown that the MLE of  $\theta$  is consistent.

Notice how we can also verify this using simulation, where we see that as we increase  $n$  the probability of  $\hat{\theta}_{MLE}$  deviating from the True  $\theta$  tends to 0.

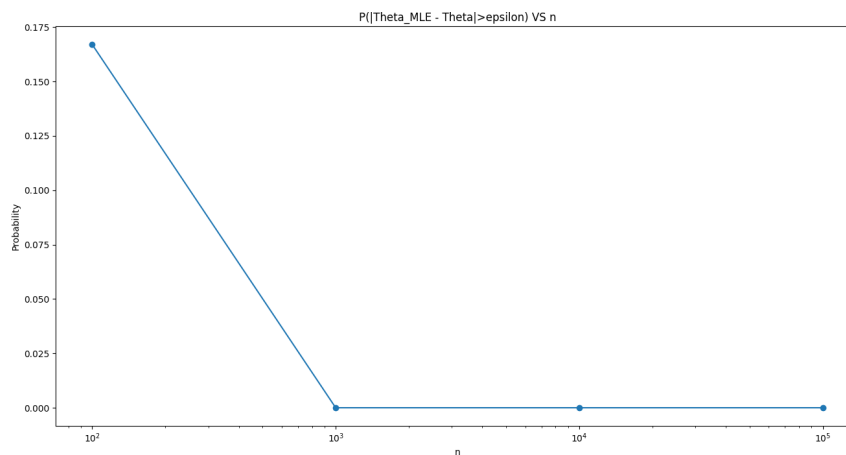


Figure 7: Code in theta\_MLE\_Consistent.py

## Asymptotic Properties of the $\hat{\theta}_{MLE}$ using the Delta Method (i.e. Getting it from $\hat{\theta}_{MLE}^2$ )

### Asymptotic Normality

For simplicity I will denote  $\hat{\theta}_{MLE}^2$  as  $W$  and  $\hat{\theta}_{MLE}$  as  $V$ .

We know that

$$\frac{\sqrt{n}(W - \theta^2)}{\theta^2} \xrightarrow{d} N(0, 1)$$

As asked we will use the delta method to find the limiting distribution of  $V$ .  
Take  $g(x) = \sqrt{x}$ , which is differentiable and  $g'(\theta^2) \neq 0$ . (which is required by the delta method)

Now we have that  $g(W) = \sqrt{W} = V$ .

And  $g(\theta^2) = \theta$ . And

$$\frac{dg(W)}{dW} = \frac{1}{2\sqrt{W}}$$

Which evaluated at  $\theta^2$  is:

$$g'(\theta^2) = \frac{1}{2\sqrt{\theta^2}} = \frac{1}{2\theta}$$

And so by the delta method we have that:

$$\frac{\sqrt{n}(g(W) - g(\mu))}{|g'(\mu)|\sigma} \xrightarrow{d} N(0, 1)$$

With  $\sigma = \sqrt{\text{Var}(W)} = \sqrt{\frac{\theta^4}{n}} = \frac{\theta^2}{\sqrt{n}}$

And  $\mu = E(W) = \theta^2$

And so we get by plugging in the values:

$$\frac{\sqrt{n}(V - \theta)}{\theta^2 \frac{1}{2\theta}} \xrightarrow{d} N(0, 1)$$

$$\frac{\sqrt{n}(V - \theta)}{\frac{\theta}{2}} \xrightarrow{d} N(0, 1)$$

$$\sqrt{n}(V - \theta) \xrightarrow{d} N(0, \frac{\theta^4}{4})$$

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, \frac{\theta^4}{4})$$

Which verifies the asymptotic normality of the MLE of  $\theta$ .



## Asymptotic Optimality

This is as discussed earlier a direct result from the above.

Side note:

Now I will just show how did we obtain the CR bound for  $\theta^2$  in the first section

I will now obtain  $I(\theta^2)$  from  $I(\theta) = \frac{4}{\theta^2}$  (that we calculated in the previous sections). I am doing this just to illustrate the power of the formula (7) in the notes which states:

$$I(g(W))\left[\frac{dg(W)}{W}\right]^2 = I(W)$$

And so in this case

$$I(\theta)\left[\frac{1}{4\theta^2}\right] = I(\theta^2)$$
$$\frac{1}{\theta^4} = I(\theta^2)$$

So now we obtained the fisher information of  $\theta^2$ , which we can use now to obtain the CR Bound of the variance of  $\hat{\theta}_{MLE}^2$ . And that is  $\frac{1}{nI(\theta^2)}$ , notice that unlike before here we immediately put  $\frac{1}{nI(\theta^2)}$ , because this is the information with respect to  $\theta^2$  and NOT  $\theta$ .

Which we can see that indeed it is equal to the CR bound that we previously found for  $\hat{\theta}_{MLE}^2$ .

This is a nice way to show from where we got the derivative in the numerator in the first method i.e. :

$$\frac{(\frac{d}{d\theta}\theta^2)^2}{nI(\theta)}$$

So to summarize:

$$\frac{(\frac{d}{d\theta}\theta^2)^2}{nI(\theta)} \equiv \frac{1}{nI(\theta^2)}$$

## Consistency

This is also the same as in the previous section.

## Method of Moments Estimator of $\theta$

### Finding the MME of $\theta$

Notice here it is not specified what exact moment we should take, but we will assume it is the first moment (so MOM) is generally different for each moment.

Another note here we are taking the MOM with the distribution of  $X$  and not  $X^2$ , since "our data" comes from  $X$  and we would like to know an estimate of  $\theta$  based on these data.

But in theory we could use the MOM of  $X^2$  and then take the square root of it to get the MOM of  $\theta$ . (But we don't have "data" on it)

So the method of moment estimator of  $\theta$  is:

$$E[X] = \hat{\alpha}_1$$

Where  $\hat{\alpha}_1$  is the sample mean.

Now let us compute  $E[X]$ :

$$E[X] = \int_0^\infty \frac{x^2}{\theta^2} e^{-\frac{x^2}{2\theta^2}} dx$$

Make substitution:  $u = \frac{x^2}{2\theta^2}$ ;  $x = \theta\sqrt{2u}$  and we get  $dx = \frac{\theta}{\sqrt{2}\sqrt{u}} du$ , plugging them into the integral, and taking the constant out of the integral we get:

$$= \sqrt{2}\theta \int_A^B u^{\frac{1}{2}} e^{-u} du$$

Now let us compute the bound of the new variable  $u$ ,  $A = 0$ ;  $B = \infty$  (got them by plugging 0 and  $\infty$  into the substitution)

$$= \sqrt{2}\theta \int_0^\infty u^{\frac{1}{2}} e^{-u} du$$

Notice how this is in the form of the gamma function (see equation 1), with  $\alpha = 1 + 0.5 = \frac{3}{2}$ , and so we obtain,

$$= \sqrt{2}\theta \Gamma\left(\frac{3}{2}\right) = \sqrt{2}\theta \frac{\sqrt{\pi}}{2} = \theta \sqrt{\frac{\pi}{2}}$$

Thus:

$$E[X] = \theta \sqrt{\frac{\pi}{2}}$$

Going back to the equation of moments:

$$\theta\sqrt{\frac{\pi}{2}} = \hat{\alpha}_1$$

$$\theta\sqrt{\frac{\pi}{2}} = \frac{1}{n} \sum_{i=1}^n X_i$$

And so we get:

$$\hat{\theta}_{MM} = \frac{1}{n} \sqrt{\frac{2}{\pi}} \sum_{i=1}^n X_i$$

**Showing that  $\hat{\theta}_{MM}$  is Unbiased  $\forall n$**

$$E[\hat{\theta}_{MM}] = E\left[\frac{1}{n} \sqrt{\frac{2}{\pi}} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sqrt{\frac{2}{\pi}} \sum_{i=1}^n E[X_i] = \sqrt{\frac{2}{\pi}} E[X] = \sqrt{\frac{2}{\pi}} \theta \sqrt{\frac{\pi}{2}} = \theta$$

Thus we have shown that the MOM Estimator is unbiased !

**Is it UMVUE?**

You can't directly conclude that it is UMVUE only because it is unbiased we need to get the variance and compare to the bound, the bound in this case since we are estimating and this estimator is unbiased we get

$$CR = \frac{\theta^2}{4n}$$

Let us compute  $Var[\hat{\theta}_{MM}]$  and compare it to CR:

$$Var[\hat{\theta}_{MM}] = E[X^2] - E[X]^2 = 2\theta^2 - \frac{\pi}{2}\theta^2 = \theta^2(2 - \frac{\pi}{2})$$

And so:

$$\theta^2(2 - \frac{\pi}{2}) \geq \frac{\theta^2}{4n}$$

We can see that they this is not a straight equality, and furthermore the variance of the MOM does NOT depend on n, so it is not decreasing as n increase. But as discussed in the notes, not being able to reach the bound does NOT mean that it is NOT UMVUE, but here I am suspecting that it is not since it is not depending on n (i.e. not decreasing with n), but I wasn't able to prove if it is actually UMVUE or not, I just know that it doesn't reach the bound!

### Comparing $\hat{\theta}_{MLE}$ and $\hat{\theta}_{MM}$ in terms of MSE

$$MSE(\hat{\theta}_{MM}) = \theta^2(2 - \frac{\pi}{2})$$

$$MSE(\hat{\theta}_{MLE}) = \theta^2[2 - \frac{2A}{\sqrt{n}}]$$

Where  $A = \frac{\Gamma(n+\frac{1}{2})}{\Gamma(n)}$

First we can notice how the MSE of the method of moments is constant, it does NOT depend on n. Also we know that  $\frac{A}{\sqrt{n}} \approx 1$  as we take n to be larger, from this figure:

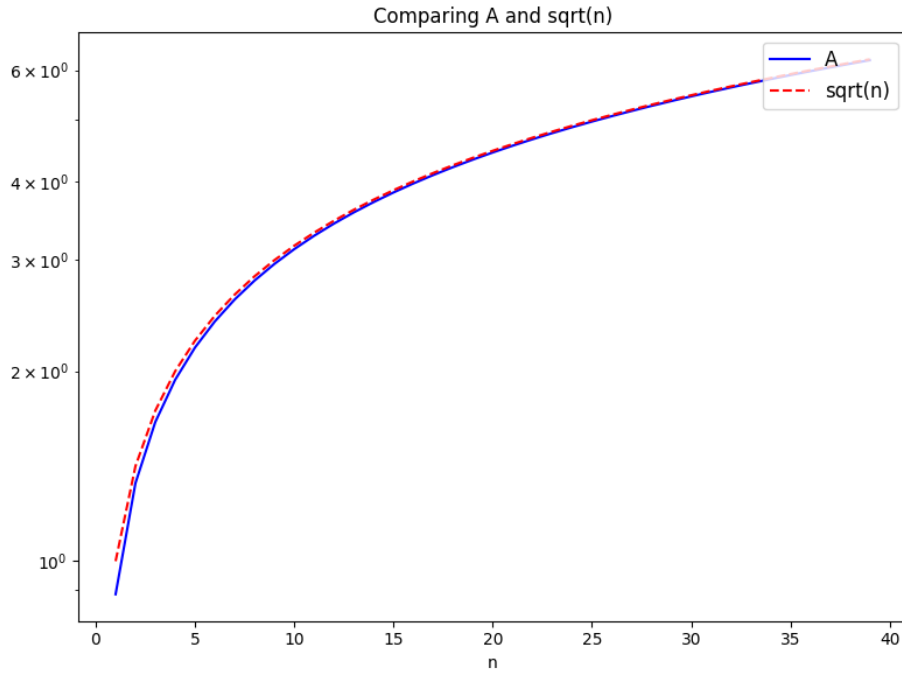


Figure 8: Code in A\_vs\_sqrt.py

And so we see that the  $MSE(\hat{\theta}_{MLE}) \rightarrow 0$  and the  $MSE(\hat{\theta}_{MM}) = \theta^2[2 - \frac{\pi}{2}]$ , which is a positive constant  $\neq 0$ , and so MSE of MLE is better. (we expect that the MSE of the MLE will decrease up to certain level and then it will plateau)

We will verify this by simulation.

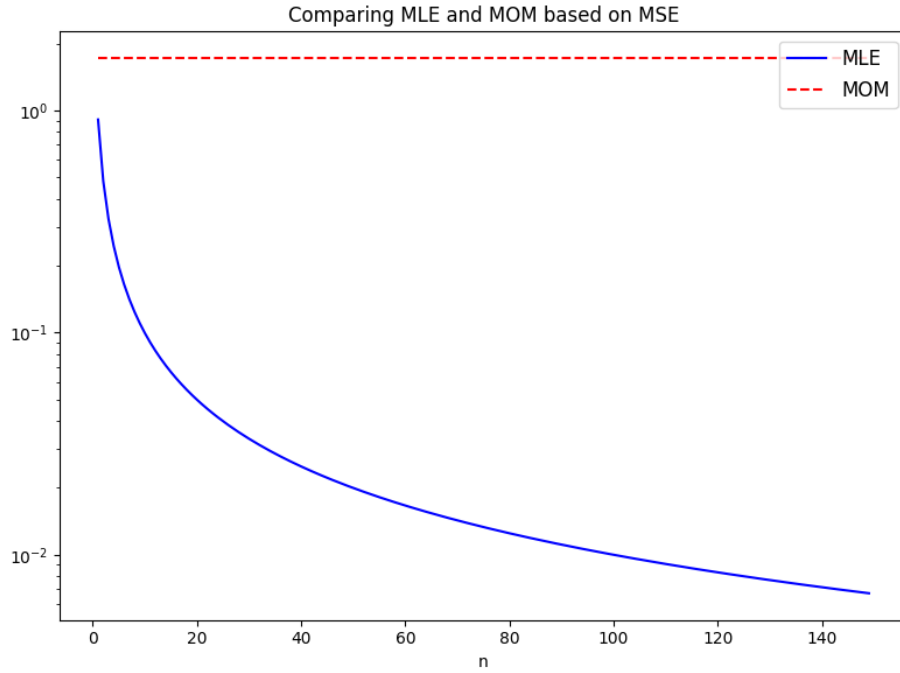


Figure 9: Code in MSE\_MM\_MLE.py

## Problem 2 in the section [Score function, Fisher Information, MLE estimations, Properties of the MLE]

### Part 1

We have  $X_1, \dots, X_n$ , i.i.d  $X_i \sim Pois(\lambda)$

**Computation of the  $\hat{\lambda}_{MLE}$**

We have  $f(x, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$ , the pdf of  $Pois(\lambda)$

$$L(\lambda) = \prod_{i=1}^n f(x_i, \lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}$$

Taking the log of the likelihood function:

$$\ell(\lambda) = \sum_{i=1}^n x_i \log(\lambda) - \sum_{i=1}^n \log(x_i!) - n\lambda$$

Taking the derivative and setting it to zero:

$$\frac{d\ell(\lambda)}{d\lambda} = \sum_{i=1}^n \frac{x_i}{\lambda} - n = 0$$

Which give us:

$$\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}_n$$

**Finding the asymptotic distribution of  $\hat{\lambda}_{MLE}$**

We know from the CLT that:

$$\frac{\bar{X}_n - E[X]}{\sqrt{V(\bar{X}_n)}} \xrightarrow{d} N(0, 1)$$

We know  $E[X] = \lambda$  and  $V[\bar{X}_n] = \frac{1}{n}\lambda$ , and so we get:

$$\frac{\bar{X}_n - \lambda}{\sqrt{\frac{\lambda}{n}}} \xrightarrow{d} N(0, 1)$$

$$\sqrt{n}(\bar{X}_n - \lambda) \xrightarrow{d} N(0, \lambda)$$

And so what we have is that:

- $\hat{\lambda}_{MLE}$  is asymptotically normal
- $\hat{\lambda}_{MLE}$  is asymptotically Unbiased
- The asymptotic variance of  $\hat{\lambda}_{MLE}$  is  $\lambda$

**Comparing The asymptotic variance of  $\hat{\lambda}_{MLE}$  and limit of the sequence of variances  $Var(\hat{\lambda}_{MLE})$**

$$\lim_{n \rightarrow \infty} [Var(\hat{\lambda}_{MLE})] = \lim_{n \rightarrow \infty} [\frac{1}{n}\lambda] = 0$$

Which we can see that it is different than the asymptotic variance.  
But now if you we notice that:

$$nVar(\hat{\lambda}_{MLE}) = \lambda$$

Now let us find the  $I(\lambda)$ :

$$I(\lambda) = -E[\frac{d^2}{d\lambda^2} \log(f(X, \lambda))] = -E[\frac{d^2}{d\lambda^2} \log(\frac{\lambda^x}{x!} e^{-\lambda})]$$

$$I(\lambda) = -E[-\frac{x}{\lambda^2}] = E[\frac{x}{\lambda^2}] = \frac{1}{\lambda^2} E[X] = \frac{1}{\lambda}$$

So we can see (which is a verification of what we talked about in class), that

$$\lim_{n \rightarrow \infty} [nVar(\hat{\lambda}_{MLE})] = \frac{1}{I(\lambda)}$$

We discussed in class that this is usually the case, and not the that the asymptotic variance is equal to the limit of the sequence of variances.

Rather a factor (in this case  $n$ ) multiplied by the variance of the MLE converges to the Fisher Information (the asymptotic variance).

Lastly we can see how the MLE of  $\lambda$  in the case of poisson is asymptotically Efficient since it reached the CR bound.

**Finding the MLE of  $\theta = \frac{1}{\lambda}$**

We know from the equivariance property that the MLE of  $\theta$  is the function of the MLE of  $\lambda$  (i.e.  $\hat{\theta}_{MLE} = g(\hat{\lambda}_{MLE})$ )

And so in particular we have:

$$\hat{\theta}_{MLE} = \frac{1}{\hat{\lambda}_{MLE}} = \frac{1}{\bar{X}_n} = \frac{n}{\sum_{i=1}^n X_i}$$

### Finding the asymptotic distribution of $\hat{\theta}_{MLE}$

We know that  $\hat{\theta}_{MLE} = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\hat{\lambda}_{MLE}}$ , and so we can use the delta method to find the limiting distribution of  $\hat{\theta}_{MLE}$ , from the limiting distribution of  $\hat{\lambda}_{MLE}$ .

We know:

$$\sqrt{n}(\hat{\lambda}_{MLE} - \lambda) \xrightarrow{d} N(0, \lambda)$$

Now take  $g(x) = \frac{1}{x}$ , which is differentiable and  $g'(\lambda) \neq 0$  in particular  $g'(x) = -\frac{1}{x^2}$  and  $|g'(\lambda)| = \frac{1}{\lambda^2} = \theta^2$  and  $g(\lambda) = \frac{1}{\lambda} = \theta$ .

And so by the delta method we have that:

$$\frac{\sqrt{n}(g(\hat{\lambda}_{MLE}) - g(\lambda))}{|g'(\lambda)|} \xrightarrow{d} N(0, \lambda)$$

And so we have that:

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{d} N(0, \theta^3)$$

And so we have that the asymptotic distribution of  $\hat{\theta}_{MLE}$  is  $N(0, \theta^3)$ .

### Asymptotic Variance of $\hat{\theta}_{MLE}$

We have that the asymptotic variance of  $\hat{\theta}_{MLE}$  is  $\theta^3$ .

### Variance in the limit of the sequence of variances (i.e. $\lim_{n \rightarrow \infty} Var(\hat{\theta}_{MLE})$ )

$$\lim_{n \rightarrow \infty} [Var(\hat{\theta}_{MLE})] = \lim_{n \rightarrow \infty} [Var(\frac{n}{\sum_{i=1}^n X_i})] = \lim_{n \rightarrow \infty} [n^2 Var(\frac{1}{\sum_{i=1}^n X_i})]$$

I suspect that either:

- $\lim_{n \rightarrow \infty} var(\hat{\theta}_{MLE}) \rightarrow \theta^3$
- Or
- $\lim_{n \rightarrow \infty} n * var(\hat{\theta}_{MLE}) \rightarrow \theta^3$
- Why  $\theta^3$ , since it is the asymptotic variance, which also we can obtain from:
- $I(\frac{1}{\lambda})[\frac{d\frac{1}{\lambda}}{d\lambda}]^2 = I(\lambda)$ , which give us that  $I(\theta) = (\frac{1}{\theta})^3$ .



- And so like discussed in class we suspect that the sequences of variances of the MLE of  $\theta$  will converge to the Fisher Information of  $\theta$ , (maybe adjusted by a factor  $n$ , like it is in this case .)
- And so from the simulation results we can see how like the previous case  $n * Var[\hat{\theta}_{MLE}] \rightarrow \theta^3$ , and NOT  $Var[\hat{\theta}_{MLE}] \rightarrow \theta^3$ . Since  $Var[\hat{\theta}_{MLE}] \rightarrow 0$ , as  $n \rightarrow \infty$
- Which verifies the discussion made in class! That in most cases it is the sequences of  $n * Var$  that converges to the asymptotic variance, and not the sequence of variance itself.

```
n * var(theta_MLE) = 0.007993001131192076  theta^3 = 0.008000000000000002
var(theta_MLE) = 7.993001131192076e-08  theta^3 = 0.008000000000000002
```

Figure 10: Code in variance\_in\_limit.py

```
1 import numpy as np
2 lambda_true = 5
3 theta_true = 1 / lambda_true
4 n_simulations = 10000
5 theta_mle_simulations = []
6 n = 100000
7 for _ in range(n_simulations):
8     X_samples = np.random.poisson(lambda_true, n)
9     lambda_mle = np.mean(X_samples)
10    theta_mle = 1 / lambda_mle
11    theta_mle_simulations.append(theta_mle)
12 VAR = np.var(theta_mle_simulations)
13 print("n * var(theta_MLE) = ", n * VAR, " theta^3 = ", theta_true**3)
14 print("var(theta_MLE) = ", VAR, " theta^3 = ", theta_true**3)
```

Figure 11: Code of the result above

## Part 2

We have now a different set-up;

$X_1 \sim Pois(\mu_1)$  and  $X_2 \sim Pois(\mu_2)$  (two independent Poisson random variables)

with  $\mu_1 = \log(\alpha)$  and  $\mu_2 = \log(\alpha + \beta)$   
 We are asked to find the MLE of  $\alpha$  and  $\beta$ .

### Finding the MLE of $\alpha$ and $\beta$

$$L(\alpha, \beta) = f(x_1, \mu_1)f(x_2, \mu_2) = \frac{e^{-\mu_1}\mu_1^{x_1}}{x_1!} \frac{e^{-\mu_2}\mu_2^{x_2}}{x_2!}$$

Taking the log of the likelihood function:

$$\ell(\alpha, \beta) = x_1 \log(\mu_1) - \mu_1 - \log(x_1!) + x_2 \log(\mu_2) - \mu_2 - \log(x_2!)$$

Plugging the corresponding  $\mu_1$  and  $\mu_2$ :

$$\ell(\alpha, \beta) = x_1 \log(\log(\alpha)) - \log(\alpha) - \log(x_1!) + x_2 \log(\log(\alpha + \beta)) - \log(\alpha + \beta) - \log(x_2!)$$

Now computing the partial derivatives and setting them to zero:

$$\frac{\partial \ell(\alpha, \beta)}{\partial \alpha} = -\frac{1}{\alpha} - \frac{x_1}{\alpha \log(\alpha)} - \frac{1}{\alpha + \beta} - \frac{x_2}{(\alpha + \beta) \log(\alpha + \beta)} = 0$$

Now the second one:

$$\frac{\partial \ell(\alpha, \beta)}{\partial \beta} = -\frac{1}{\alpha + \beta} - \frac{x_2}{(\alpha + \beta) \log(\alpha + \beta)} = 0$$

From the second equation we find:

$$\alpha + \beta = e^{x_2}$$

Plugging this into the first equation we get:

$$\alpha = e^{x_1}$$

And:

$$\beta = e^{x_2} - e^{x_1}$$

Thus:

$$\hat{\alpha}_{MLE} = e^{x_1} \quad \text{and} \quad \hat{\beta}_{MLE} = e^{x_2} - e^{x_1}$$

### Problem 3: Section[Score function, Fisher Information, MLE estimations, Properties of the MLE]

Goal: Show that the MLE of  $P(X_1 \leq u) = F(u)$  is:

$$\hat{F}_{MLE}(u) = \Phi(u - \bar{X}_n)$$

We will be using the equivariant property of the MLE.

Let us first find the MLE of  $\mu$  on an i.i.d sample of size  $n$ , coming from this distribution.

$$L(\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(X_i - \mu)^2}{2}}$$

Taking the log of the likelihood function:

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

now taking the derivative and setting it equal to 0:

$$\frac{d\ell(\mu)}{d\mu} = \sum_{i=1}^n (X_i - \mu) = 0$$

$$\hat{\mu}_{MLE} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}_n$$

Now we will try expressing the function  $F(u)$  in terms of  $\mu$  if we were able to do this then we can use the equivariant property of the MLE to find the MLE of  $F(u)$ .

We know that:

$$F(u) = P(X_1 \leq u) = P(Z \leq \frac{u - \mu}{\sigma}) = \Phi(\frac{u - \mu}{\sigma})$$

But we know  $\sigma = 1$ , and so

$$F(u) = \Phi(u - \mu)$$

And so we can use the equivariant property of the MLE to find the MLE of  $F(u)$ , from the MLE of  $\mu$ :

$$\hat{F}_{MLE}(u) = \Phi(u - \bar{X}_n)$$

## Non Parametric Estimation

Showing that non-parametric estimator  $\hat{F}_{np}$  of  $F(u)$  based on the sample  $X_1, \dots, X_n$  is:

$$\hat{F}_{np}(u) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq u)$$

This is "a" non-parametric estimator of  $F(u)$  which comes from the definition of the empirical distribution function.

### Unbiasedness of $\hat{F}_{np}$

$$E\left[\frac{1}{n} \sum_{i=1}^n I(X_i \leq u)\right] = \frac{1}{n} \sum_{i=1}^n E[I(X_i \leq u)] = \frac{1}{n} \sum_{i=1}^n P(X_i \leq u) = \frac{1}{n} n F(u) = F(u)$$

And so we have shown that  $\hat{F}_{np}$  is unbiased.

### Showing $\hat{F}_{np}(u)$ is invariant under permutations of $X_i$

We want to show that the estimator  $\hat{F}_{NP}(u)$  is invariant under any permutation of the sample.

Let  $\pi$  be any permutation of the indices  $\{1, 2, \dots, n\}$ . The permuted sample (that we had) is:

$$X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)}.$$

The estimator  $\hat{F}_{NP}(u)$  for the permuted sample is:

$$\hat{F}_{NP}^{\text{perm}}(u) = \frac{1}{n} \sum_{i=1}^n I(X_{\pi(i)} < u).$$

Since the indicator function  $I(X_{\pi(i)} < u)$  is 1 if  $X_{\pi(i)} < u$  and 0 otherwise, and we see that the sum of these indicators doesn't change if we change the order of these indicators, and so:

$$\hat{F}_{NP}^{\text{perm}}(u) = \frac{1}{n} \sum_{i=1}^n I(X_{\pi(i)} < u) = \frac{1}{n} \sum_{i=1}^n I(X_i < u) = \hat{F}_{NP}(u).$$

Thus, the non-parametric estimator  $\hat{F}_{NP}(u)$  is invariant under any permutation of the sample. And from this we see how the non parametric estimator depends only on the number of observations less than  $u$ , not on the order in which the observations appear!!!

### Asymptotic Distribution of $\hat{F}_{np}(u)$

This follows from the CLT, take  $\hat{F}_{NP}(u) = \frac{1}{n} \sum_{i=1}^n I(X_i < u) = \bar{Y}_n$ ,  $\forall u$ , So in other words we are now taking the average of the bernoulli random variables  $I(X_i < u)$  "as our new random variables", so by the CLT we have that:

$$\sqrt{n}(\bar{Y}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Now let us get  $\mu$  and  $\sigma^2$  (note that  $\sigma^2 = \text{Var}[I(X_i < u)]$  and not for the variance of the  $\bar{Y}_n$ , since we already did that to get the  $\sqrt{n}$  in the numerator) (and of course  $\mu$  is the same for both (since i.i.d)), we know that  $E[I(X_i < u)] = P(X_i < u) = F(u)$ , and  $\text{Var}[I(X_i < u)] = F(u)(1 - F(u))$ , and so we have that:

$$\sqrt{n}(\bar{Y}_n - F(u)) \xrightarrow{d} N(0, F(u)(1 - F(u)))$$

Thus:

$$\sqrt{n}(\hat{F}_{np}(u) - F(u)) \xrightarrow{d} N(0, F(u)(1 - F(u)))$$

And so we just proved that the non-parametric estimator is asymptotically normal, and unbiased with an asymptotic variance equal to  $F(u)(1 - F(u))$ .

## Comparing $\hat{F}_{MLE}(u)$ with $\hat{F}_{NP}(u)$ in terms of asymptotic efficiency

### Evaluating the asymptotic efficiency of the MLE estimate

To evaluate this we need first to get the limiting distribution of  $\hat{F}_{MLE}(u)$ . Let us first start by getting the asymptotic distribution of  $\bar{X}_n$ . Which is a straight result from the CLT:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, 1)$$

Now we will use the delta method to find the limiting distribution of  $\hat{F}_{MLE}(u)$ , we know that  $\hat{F}_{MLE}(u) = \Phi(u - \bar{X}_n)$ , and so we can take  $g(x) = \Phi(u - x)$ ,

which is differentiable and  $g'(\mu) \neq 0$  (see why below), and so we have that:

$$\frac{\sqrt{n}(g(\bar{X}_n) - g(\mu))}{|g'(\mu)|} \xrightarrow{d} N(0, 1)$$

- $g'(x) = \varphi(u - x)$
- $g'(\mu) = \varphi(u - \mu)$       We know that  $\varphi(x) > 0, \quad \forall x$
- $g(\mu) = \Phi(u - \mu)$
- $g(\bar{X}_n) = \Phi(u - \bar{X}_n) = \hat{F}_{MLE}(u)$

And so by plugging these values we get:

$$\frac{\sqrt{n}(\hat{F}_{MLE}(u) - \Phi(u - \mu))}{\varphi(u - \mu)} \xrightarrow{d} N(0, 1)$$

And so we have that:

$$\sqrt{n}(\hat{F}_{MLE}(u) - \Phi(u - \mu)) \xrightarrow{d} N(0, \varphi^2(u - \mu))$$

And so we have found that the asymptotic variance of  $\hat{F}_{MLE}(u)$  is  $\varphi^2(u - \mu)$ , and so we can see that the asymptotic efficiency of the MLE is:

$$e_{asympt}(\hat{F}_{MLE}(u)) = \frac{\frac{1}{I(F(u))}}{\text{asymptotic Variance}} = \frac{\frac{1}{I(F(u))}}{\varphi^2(u - \mu)} \quad (**)$$

**Now we need to get  $\frac{1}{I(F(u))}$**

We know that:

$$\frac{1}{I(F(u))} = \frac{(\frac{dF(u)}{d\mu})^2}{I(\mu)} = \frac{(\frac{d}{d\mu}\Phi(u - \mu))^2}{I(\mu)}$$

Note that the second equality follows from part 1

Let us first get  $I(\mu)$

$$I(\mu) = -E\left[\frac{d^2}{d\mu^2} \log(f(X, \mu))\right] = -E\left[\frac{d^2}{d\mu^2} \log\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2}}\right)\right] = 1$$

Now let us get  $\frac{d}{d\mu}\Phi(u - \mu)$ : To do that we need to use the chain rule:

$$\frac{d}{d\mu}\Phi(u - \mu) = \frac{d}{d(u - \mu)}\Phi(u - \mu) \frac{d(u - \mu)}{d\mu} = -\varphi(u - \mu)$$

And so we get by pluggin these quantities back:

$$\frac{1}{I(F(u))} = \frac{(-\varphi(u - \mu))^2}{1} = \varphi^2(u - \mu)$$

Now finally going back to (\*\*),

$$e_{asympt}(\hat{F}_{MLE}(u)) = \frac{\frac{1}{I(F(u))}}{\text{asymptotic Variance}} = \frac{\varphi^2(u - \mu)}{\varphi^2(u - \mu)} = 1$$

So we have shown that  $\hat{F}_{MLE}(u)$  is asymptotically efficient!!!

### **Evaluating the asymptotic efficiency of the non-parametric estimate**

We know that the Asymptotic Variance of the NP Estimate is (from the asymptotic distribution of  $\hat{F}_{NP}(u)$ ):

$$F(u)(1 - F(u))$$

And We have already computed  $\frac{1}{I(F(u))}$ , and so we get:

$$e_{asympt}(\hat{F}_{NP}(u)) = \frac{\frac{1}{I(F(u))}}{\text{Asymptotic Variance}} = \frac{\varphi^2(u - \mu)}{F(u)(1 - F(u))}$$

We can see how this factor is not equal to 1. (and we know it is less than 1) from the CR bound (i.e. we know that Asymptotic variance  $\geq \frac{1}{I(F(u))}$ ) So clearly we can see how the *MLE* estimate is better than *NP* estimate, in terms of asymptotic efficiency!

## Problem 1: Section [Sufficiency, Exponential Families, Rao-Blackwell, Construction of UMVUE]

**Proving that the order statistics  $Y = X_{(1)}, \dots, X_{(n)}$  is sufficient for  $\theta$  (i.e. for the family of distributions of  $X$ , as discussed in the notes)**

Just to make it clear in terms of notation: the pdf of each  $X_i$  is denoted by  $m$ . (which I won't use here, but to make explicit differentiation between joint and the marginal pdfs)

In general we have many ways to prove Sufficiency, but here I will use the third definition in the notes, that the ratio of densities, is a constant as function of  $\theta$ .

We first have the joint density of  $X = X_1, \dots, X_n$ , given by:

$$f(x|\theta), \quad \text{where } x = x_1, \dots, x_n$$

Now let us understand the joint density of the order statistic:

We know that a statistic partition the sample space(of size  $n$ ) into level sets:

$$B_y = \{x : Y = y\}, \quad \text{where } y = x_{(1)}, \dots, x_{(n)}$$

This give us that an observation of  $Y$ , comes from one of these level sets, in other words, if  $Y = y$ (I know in continous case this will not happen, but here just to illstrute a point!), then we know it came from the set  $B_y$ .

Now let us actually understand this set  $B_y = \{x : Y = y\}$ , the question to ask is that how many observations of  $X$  are in this set?

Very simple answer " $n!$ " since we can have any permutation of the  $n$  observations of  $X$  to get the same value of  $Y$ . But notice that the observations  $X_i$  are i.i.d so any permutation of them won't affect the joint distribution  $f(x|\theta)$

And so we get that the joint density of  $Y$  is:

$$g(y|\theta) = n!f(x|\theta)$$

In other words: the factor  $n!$  because there are  $n!$  permutations of  $x = x_1, \dots, x_n$  that results in the same  $y$ , each contributing with the same joint density function  $g(x|\theta)$



Thus we can see that the ratio of the joint density of  $X$  to the joint density of  $Y$  is:

$$\frac{f(x|\theta)}{g(y|\theta)} = \frac{1}{n!}$$

Which is a constant as a function of  $\theta$ , and so we have shown that the order statistics  $Y = X_{(1)}, \dots, X_{(n)}$  is sufficient for  $\theta$ .

Based on this analysis we can that yes indeed  $Y$  is a reduction from the original data  $X$ , since it removes information about the order of the observations, and only keeps the information about the values of the observations. which is sufficient (as shown previously) to estimate  $\theta$

## Problem 2: Section [Sufficiency, Exponential Families, Rao-Blackwell, Construction of UMVUE]

### Part 1: Proving that the density function of problem 1 section 1 is a member of the exponential family

The density function is given by:

$$f(x, \theta) = \frac{x}{\theta^2} e^{-\frac{x^2}{2\theta^2}}$$

where  $x > 0$  and  $\theta > 0$ .

Reminder of the canonical form of the exponential family:

$$f(x, \theta) = h(x) e^{<T(x), \eta(\theta)> - A(\eta)}$$

Where  $\eta$  is the natural parameter, and  $T(x)$  is the sufficient statistic.

Let us first rewrite the density function in the form of the exponential family:

$$f(x, \theta) = x e^{<\frac{x^2}{2}; -\frac{1}{\theta^2}> - \log(\theta^2)}$$

And so we can identify the sufficient statistic  $T(x) = \frac{x^2}{2}$ , and the natural parameter  $\eta = -\frac{1}{\theta^2}$ , and the function  $h(x) = x$ , and  $A(\eta) = \log(\theta^2) = \log(\frac{-1}{\eta})$ . And so we have shown that the density function is a member of the exponential family.

To also check that this distribution is in the exponential family, if we consider the joint distribution of i.i.d observations we will get that the sufficient

statistic is indeed their sum of the statistic applied to each observation ! (i.e.  $T_{joint}(X) = \sum_{i=1}^n T(X_i)$ ), which comes directly from the multiplication of each exponential term, (multiplication since i.i.d)!

## Part 2: Finding the minimal sufficient statistic

In the first part of the question I got that  $T_{joint}(X)$  is "a" sufficient statistic, for this exponential family. I will now verify that is indeed a minimal sufficient statistic.

Reminder of the definition of a minimal sufficient statistic:

Let  $f_X(x|\theta)$  be the joint pdf/pmf of a sample  $X = (X_1, \dots, X_n)$  and  $T = T(X_1, \dots, X_n)$  be a statistic for which for all realizations  $x, x_*$  from the distribution

$$\frac{f(x|\theta)}{f(x_*|\theta)} \text{ does not depend on } \theta \iff T(x) = T(x_*)$$

then  $T$  is minimal and sufficient for  $\theta$ .

We need to prove the two directions:

But before let us get the joint density function of the sample  $X = X_1, \dots, X_n$ :

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{x_i}{\theta^2} e^{-\frac{x_i^2}{2\theta^2}} = \left(\frac{1}{\theta^2}\right)^n \left(\prod_{i=1}^n x_i\right) e^{-\frac{1}{2\theta^2} \sum_{i=1}^n x_i^2}$$

And so:

$$f(x|\theta) = \left(\frac{1}{\theta^2}\right)^n \left(\prod_{i=1}^n x_i\right) e^{-\frac{1}{2\theta^2} \sum_{i=1}^n T(x_i)}$$

**Direction1: Assuming  $T_{joint}(x) = T_{joint}(x_*)$  and proving that the ratio of the densities does not depend on  $\theta$**

$$\frac{f(x|\theta)}{f(x_*|\theta)} = \frac{\left(\frac{1}{\theta^2}\right)^n \left(\prod_{i=1}^n x_i\right) e^{-\frac{1}{2\theta^2} \sum_{i=1}^n T(x_i)}}{\left(\frac{1}{\theta^2}\right)^n \left(\prod_{i=1}^n x_{*i}\right) e^{-\frac{1}{2\theta^2} \sum_{i=1}^n T(x_{*i})}} = \frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n x_{*i}}$$

Which is independent of  $\theta$  and so we have shown the first implication!

**Direction2:** Assuming that the ratio of the densities does not depend on  $\theta$  and proving that  $T_{joint}(x) = T_{joint}(x_*)$

We know from the assumption that:

$$\frac{f(x|\theta)}{f(x_*|\theta)} = \frac{\prod_{i=1}^n x_i}{\prod_{i=1}^n x_{*i}}$$

Which is the only way they are independent of  $\theta$ , and so for this to happen we require that the exponential terms  $e^{-\frac{1}{\theta^2} \sum_{i=1}^n T(x_{*i})} = e^{-\frac{1}{\theta^2} \sum_{i=1}^n T(x_i)}$

Which give us the condition that:

$$e^{\frac{1}{\theta^2} (\sum_{i=1}^n T(x_i) - \sum_{i=1}^n T(x_{*i}))} = 1$$

Which is only possible if:

$$\sum_{i=1}^n T(x_i) - \sum_{i=1}^n T(x_{*i}) = 0$$

And so we have shown the second implication!

Thus we have shown that the statistic  $T_{joint}(X)$  is minimal sufficient for  $\theta$  (which is also minimal sufficient for a function of  $\theta$  as discussed in the notes).

## Finding the $\hat{\theta}_{UMVUE}^2$

Here we can get it from the remark in the notes which is straight-forward, let me start by this method:

We know that for a sufficient statistic (in this case  $T_{joint}(X)$ ), of a 1-d exponential family, the  $T_{joint}(X)$  is the UMVUE of its mean (i.e.  $E(T_{joint}(X))$ ) since it achieves the inequality bound (we will verify this later).

Let us now get  $E(T_{joint}(X))$  to see  $T_{joint}(X)$  is the UMVUE of what function of  $\theta$  exactly:

$$E(T_{joint}(X)) = E\left[\sum_{i=1}^n T(X_i)\right] = nE[T(X_1)]$$

But we know from the properties of the exponential family that:

$$E(T(X_1)) = \frac{\partial A(\eta)}{\partial \eta} = \frac{\partial \frac{-1}{\eta}}{\partial \eta} = \frac{\partial \log(\frac{-1}{\eta})}{\partial \frac{-1}{\eta}} \frac{\partial \frac{-1}{\eta}}{\partial \eta} = -\frac{1}{\eta} = \theta^2$$

And so plugging this back in we get:

$$E(T_{joint}(X)) = n\theta^2$$

For now we know that  $T_{joint}(X)$  is a *UMVUE* estimate for  $n\theta^2$

Let us adjust this Sufficient statistic to get the *UMVUE* estimate for  $\theta^2$ :

$$T_{adj}(X) = \frac{T_{joint}(X)}{n} = \frac{1}{2n} \sum_{i=1}^n X_i^2$$

We can verify that this statistic is indeed sufficient (also minimal sufficient, easily verified like we did previously) for  $\theta^2$  (since we just scaled it by  $n$  (which is a constant))

And so we get:

$$E[T_{adj}(X)] = \theta^2$$

Thus we have found the *UMVUE* estimate for  $\theta^2$  to be  $T_{adj}(X) = \frac{1}{2n} \sum_{i=1}^n X_i^2$ .

## Finding the $\hat{\theta}_{MLE}^2$

Now to find the MLE of  $\theta^2$ , we will also use the properties of the exponential family:

$$\ell(\eta|X) = \eta T_{joint}(X) - nA(\eta) = \frac{\eta}{2} \sum_{i=1}^n x_i^2 - n \log\left(\frac{-1}{\eta}\right)$$

Taking the derivative and setting equal to 0:

$$\frac{d\ell(\eta|X)}{d\eta} = \frac{1}{2} \sum_{i=1}^n x_i^2 + \frac{n}{\eta} = 0$$

$$\hat{\eta}_{MLE} = -\frac{2n}{\sum_{i=1}^n x_i^2}$$

We know that  $\eta = -\frac{1}{\theta^2}$ , and from the equivariant property we know that:

$$\hat{\theta}_{MLE}^2 = -\frac{1}{\hat{\eta}_{MLE}} = \frac{1}{2n} \sum_{i=1}^n x_i^2$$

And so we have just found that  $\hat{\theta}_{MLE}^2 = \hat{\theta}_{UMVUE}^2 = \frac{1}{2n} \sum_{i=1}^n x_i^2$ .  
Which is consistent with what we have found in problem 1 !!!!

## Finding $I(\theta^2)$

We know from the properties of the exponential family that:

$$I(\eta) = \frac{\partial^2 A(\eta)}{\partial \eta^2} = \frac{1}{\eta^2}$$

Let us now use the delta method to find  $I(\theta^2)$ , we know that  $\theta^2 = -\frac{1}{\eta}$ , so take  $g(\eta) = -\frac{1}{\eta} = \theta^2$  and so;  $g'(\eta) = \frac{1}{\eta^2}$ , we know:

$$I(g(\eta)) \left[ \frac{dg(\eta)}{d\eta} \right]^2 = I(\eta)$$

$$I(\theta^2) \left[ \frac{1}{\eta^2} \right]^2 = I(\eta)$$

$$I(\theta^2) = \eta^2 = \frac{1}{\theta^4}$$

Now let us verify this by comparing with the result in **\*\*1\*\***;

Now notice that in one we directly got  $I(\theta)$ , so let us check if we use the delta method now on  $I(\theta^2)$ , we get the same  $I(\theta)$  as before:

Now similarly we take here  $f(x) = \sqrt{x}$ , and so we find:

$$I(\theta) \left[ \frac{df(\theta^2)}{d\theta^2} \right] = I(\theta^2)$$

And so we indeed find that:

$$I(\theta) = \frac{4}{\theta^2},$$

Which is consistent with problem **\*\*1\*\***.

## Verifying that indeed the *UMVUE* estimate reach the **CR** bound

We know that the *UMVUE* estimate for  $\theta^2$  is the  $T_{adj}(X) = \frac{1}{2n} \sum_{i=1}^n X_i^2$ , and we know that the *CR* bound is given by:

$$Var(T_{adj}(X)) \geq \frac{1}{I(\theta^2)}$$

let us compute the variance of  $T_{adj}(X)$ :

$$Var(T_{adj}(X)) = \frac{1}{4n^2} Var\left(\sum_{i=1}^n X_i^2\right) = \frac{1}{4n^2} n Var(X_i^2) = \frac{1}{4n} Var(X_i^2) = \frac{1}{4n} 4\theta^4 = \frac{\theta^4}{n}$$

And thus, we just verified that the *UMVUE* estimate reaches the CR bound.  
And of course by construction we know that it is unbiased !

### **Problem 3: Section [Sufficiency, Exponential Families, Rao-Blackwell, Construction of UMVUE]**

**Finding the method of moment estimator  $\hat{\theta}_{MM}$  for  $\theta$**

We know that the method of moment estimator is given by:

$$E[X] = \frac{1}{n} \sum_{i=1}^n x_i$$

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

$$\hat{\theta}_{MM} = \frac{1}{n} \sum_{i=1}^n x_i - \frac{1}{2}$$

**Computing the MSE of the MM estimator of  $\theta$**

**It is unbiased**,  $E[\hat{\theta}_{MM}] = \frac{2\theta+1}{2} - \frac{1}{2} = \theta$

Now let us compute the variance of the MM estimator:

$$Var(\hat{\theta}_{MM}) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2}\right) = \frac{1}{n^2} n Var[X_i] = \frac{1}{n} Var[X_i] = \frac{1}{n} \frac{(\theta + 1 - \theta)^2}{12} = \frac{1}{12n}$$

Thus:

$$MSE[\hat{\theta}_{MM}] = Var(\hat{\theta}_{MM}) + Bias^2[\hat{\theta}_{MM}] = \frac{1}{12n}$$

## Applying Rao-Blackwell to the MM estimator

To do this we first need a sufficient statistic for  $\theta$ .

Looking at the parameter  $\theta$  one statistic that comes to mind is  $T(X) = (\min(X), \max(X)) = (\min\{X_1, \dots, X_n\}, \max\{X_1, \dots, X_n\})$ .

**Verification that  $T(X)$  is a sufficient statistic for  $\theta$**

I will use the factorization theorem to show that  $T(X)$  is sufficient for  $\theta$ .

Recall that the factorization theorem states that a statistic  $T(X)$  is sufficient for  $\theta$  if the joint density of the sample  $X$  can be written as:

$$f(x|\theta) = g(T(x), \theta)h(x)$$

In this case we know that the joint density of the i.i.d sample  $X = X_1, \dots, X_n$  is given by:

$$f(x|\theta) = \prod_{i=1}^n f(x_i|\theta) = I\{\theta \leq x_1 \leq \theta + 1, \theta \leq x_2 \leq \theta + 1, \dots, \theta \leq x_n \leq \theta + 1\}$$

Where  $I$  is the indicator function.

Notice how the condition now can be written as:

$$f(x|\theta) = I\{\theta \leq \min(x) \leq \max(x) \leq \theta + 1\}$$

Thus we have written  $f(x|\theta) = g(T(x), \theta)h(x)$ , where  $h(x) = 1$  and  $g(T(x), \theta) = I\{\theta \leq \min(x) \leq \max(x) \leq \theta + 1\}$

Thus we just proved that  $T(X)$  is sufficient for  $\theta$ .

**Applying Rao-Blackwell to the MM estimator using  $T(X)$**

$$\begin{aligned}\hat{\theta}_{RB} &= E[\hat{\theta}_{MM}|T(X)] = E\left[\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{2} | T(X)\right] \\ &= -\frac{1}{2} + \frac{1}{n} E[x_{(1)} + x_{(n)} + \sum_{i=2}^{n-1} x_{(i)} | x_{(1)}, x_{(n)}] \\ &= -\frac{1}{2} + \frac{\max(x) + \min(x)}{n} + \frac{1}{n} E\left[\sum_{i=2}^{n-1} x_{(i)} | \max(x), \min(x)\right]\end{aligned}$$

So we can see how now knowing the max, min we know that the rest  $x_{(i)} \sim U[\min(x), \max(x)]$ , and so we get:

$$= -\frac{1}{2} + \frac{\max(x) + \min(x)}{n} + \frac{1}{n}(n-2)\frac{1}{2}(\min(x) + \max(x))$$

Thus:

$$\hat{\theta}_{RB} = \frac{\min(x) + \max(x) - 1}{2}$$

### Getting the Expectation and variance of $\min(X)$ and $\max(X)$

We should expect that this new estimator is still unbiased, this is because we have used the RB theorem, which reduces("or keep the same") the variance of the estimator, without interfering with the bias of the estimator. In this case the bias = 0, so we expect that this is the case for this estimator

#### The Proof:

In order to do this we need to know the distr. of each one.

I will start with **Max**, the min follows in a very similar way:

#### Max(X)

We have that the CDF of each  $X_i$  is:

$$F_{X_I}(x_i) = P(X_I \leq x_i) = x_i - \theta$$

Now cdf of  $\max(X)$ :

$$F_{\max(X)}(m_x) = P(X_1 \leq m_x, X_2 \leq m_x, \dots, X_n \leq m_x) = (m_x - \theta)^n$$

when  $\theta \leq m_x \leq \theta + 1$

$$f_{\max(X)}(m_x) = \frac{d}{dm_x} F_{\max(X)}(m_x) = n(m_x - \theta)^{n-1}$$

Now let us get the expectation of  $\max(X)$ :

$$E[\max(X)] = \int_{\theta}^{\theta+1} m_x n(m_x - \theta)^{n-1} dm_x = n \int_{\theta}^{\theta+1} m_x (m_x - \theta)^{n-1} dm_x$$



Very easy integral to compute: just use the substitution  $u = m_x - \theta$  and so  $du = dm_x$ , and also  $m_x = u + \theta$ , the new bounds are 0 and 1, obtained them by plugging the old bounds in the substitution, and so we get:

$$= n \int_0^1 (u+\theta)u^{n-1}du = n \int_0^1 (u^n + \theta u^{n-1})du = n \left[ \frac{u^{n+1}}{n+1} + \theta \frac{u^n}{n} \right]_0^1 = n \left[ \frac{1}{n+1} + \frac{\theta}{n} \right] = \frac{n}{n+1} + \theta$$

And so we found that:

$$E[\max(X)] = \frac{n}{n+1} + \theta$$

Now let us get the variance of  $\max(X)$ :

$$\text{Var}[\max(X)] = E[\max(X)^2] - E[\max(X)]^2$$

The second term:

$$E[\max(X)]^2 = \left( \frac{n}{n+1} + \theta \right)^2 = \frac{n^2}{(n+1)^2} + \frac{2n\theta}{n+1} + \theta^2$$

Let us now compute the first term:

$$E[\max(X)^2] = \int_{\theta}^{\theta+1} m_x^2 n (m_x - \theta)^{n-1} dm_x = n \int_{\theta}^{\theta+1} m_x^2 (m_x - \theta)^{n-1} dm_x$$

Again very easy integral to compute, just use the substitution  $u = m_x - \theta$  and so  $du = dm_x$ , and also  $m_x = u + \theta$ , the new bounds are 0 and 1, obtained them by plugging the old bounds in the substitution, and so we get:

$$\begin{aligned} &= n \int_0^1 (u+\theta)^2 u^{n-1} du = n \int_0^1 (u^n + 2\theta u^{n-1} + \theta^2 u^{n-2}) du = n \left[ \frac{u^{n+2}}{n+2} + 2\theta \frac{u^{n+1}}{n+1} + \theta^2 \frac{u^n}{n} \right]_0^1 \\ &= n \left[ \frac{1}{n+2} + 2\theta \frac{1}{n+1} + \theta^2 \frac{1}{n} \right] \end{aligned}$$

And so we also found:

$$E[\max(X)^2] = \frac{n}{n+2} + \frac{2\theta n}{n+1} + \theta^2$$

And so we get the variance of  $\max(X)$ :

$$\text{Var}[\max(X)] = \frac{n}{n+2} + \frac{2\theta n}{n+1} + \theta^2 - \frac{n^2}{(n+1)^2} - \frac{2n\theta}{n+1} - \theta^2$$

Thus we get the variance of  $\max(X)$ :

$$\text{Var}[\max(X)] = \frac{n}{(n+1)^2(n+2)}$$

$\min(X)$

We will now compute the expectation and variance of the minimum of the random variables  $X_1, X_2, \dots, X_n$  using a similar approach.

The CDF of  $\min(X)$  is:

$$F_{\min(X)}(m_x) = P(X_1 \geq m_x, X_2 \geq m_x, \dots, X_n \geq m_x) = (1 - (m_x - \theta))^n = (1 - (m_x - \theta))^n$$

for  $\theta \leq m_x \leq \theta + 1$ .

Differentiating to get the PDF:

$$f_{\min(X)}(m_x) = \frac{d}{dm_x} F_{\min(X)}(m_x) = -n(1 - (m_x - \theta))^{n-1}$$

Now let us get the expectation of  $\min(X)$ .

The expectation of  $\min(X)$  is:

$$E[\min(X)] = \int_{\theta}^{\theta+1} m_x \cdot n(1 - (m_x - \theta))^{n-1} dm_x = n \int_{\theta}^{\theta+1} m_x (1 - (m_x - \theta))^{n-1} dm_x$$

We use the substitution  $u = 1 - (m_x - \theta)$ , so  $du = -dm_x$  and  $m_x = \theta + 1 - u$ .

The new bounds are from 1 to 0, giving us:

$$\begin{aligned} &= n \int_1^0 (\theta + 1 - u) u^{n-1} (-du) = n \int_0^1 (\theta + 1 - u) u^{n-1} du \\ &= n \int_0^1 (u^{n-1}(\theta + 1) - u^n) du = n \left[ (\theta + 1) \frac{u^n}{n} - \frac{u^{n+1}}{n+1} \right]_0^1 \\ &= n \left[ \frac{\theta + 1}{n} - \frac{1}{n+1} \right] = \frac{\theta + 1}{n+1} \end{aligned}$$

Thus, we have:

$$E[\min(X)] = \frac{1}{n+1} + \theta$$

To find the variance of  $\min(X)$ , we first compute  $E[\min(X)^2]$ :

$$E[\min(X)^2] = \int_{\theta}^{\theta+1} m_x^2 \cdot n(1 - (m_x - \theta))^{n-1} dm_x = n \int_{\theta}^{\theta+1} m_x^2 (1 - (m_x - \theta))^{n-1} dm_x$$

Using the same substitution  $u = 1 - (m_x - \theta)$  with  $m_x = \theta + 1 - u$ , the bounds convert from 1 to 0:

$$= n \int_0^1 (\theta + 1 - u)^2 u^{n-1} du$$

Expanding  $(\theta + 1 - u)^2$  gives:

$$\begin{aligned}
&= n \int_0^1 ((\theta + 1)^2 u^{n-1} - 2(\theta + 1)u^n + u^{n+1}) du \\
&= n \left[ (\theta + 1)^2 \frac{u^n}{n} - 2(\theta + 1) \frac{u^{n+1}}{n+1} + \frac{u^{n+2}}{n+2} \right]_0^1 \\
&= n \left[ \frac{(\theta + 1)^2}{n} - \frac{2(\theta + 1)}{n+1} + \frac{1}{n+2} \right]
\end{aligned}$$

and so:

$$E[\min(X)^2] = \frac{n(\theta + 1)^2}{n} - \frac{2n(\theta + 1)}{n+1} + \frac{n}{n+2}$$

Then the variance of  $\min(X)$  is:

$$\text{Var}[\min(X)] = E[\min(X)^2] - (E[\min(X)])^2$$

Substituting  $E[\min(X)] = \frac{1}{n+1} + \theta$ :

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

Thus, we get:

$$\text{Var}[\min(X)] = \frac{n}{(n+1)^2(n+2)}$$

### Getting the MSE of the RB estimator

Now we can get the expectation of the RB estimator:

$$E[\hat{\theta}_{RB}] = E\left[\frac{\min(x) + \max(x) - 1}{2}\right] = \frac{E[\min(x)] + E[\max(x)] - 1}{2} = \frac{\frac{1}{n+1} + \theta + \frac{n}{n+1} + \theta - 1}{2}$$

Thus we get that indeed this estimator is unbiased.

$$E[\hat{\theta}_{RB}] = \theta$$

Now let us get the variance of the RB estimator:

$$\text{Var}[\hat{\theta}_{RB}] = \text{Var}\left[\frac{\min(x) + \max(x) - 1}{2}\right] = \frac{1}{4} \text{Var}[\min(x) + \max(x)]$$

$$= \frac{1}{4}(Var[\min(x)] + Var[\max(x)])$$

$$= \frac{1}{4}\left(\frac{n}{(n+1)^2(n+2)} + \frac{n}{(n+1)^2(n+2)}\right) = \frac{n}{2(n+1)^2(n+2)}$$

And so we have found that:

$$Var[\hat{\theta}_{RB}] = \frac{n}{2(n+1)^2(n+2)}$$

And so we have found that the MSE of the RB estimator is:

$$MSE[\hat{\theta}_{RB}] = \frac{n}{2(n+1)^2(n+2)}$$

### **Comparing $MSE[\hat{\theta}_{MM}]$ and $MSE[\hat{\theta}_{RB}]$**

We have found that the MSE of the MM estimator is:

$$MSE[\hat{\theta}_{MM}] = \frac{1}{12n}$$

And the MSE of the RB estimator is:

$$MSE[\hat{\theta}_{RB}] = \frac{n}{2(n+1)^2(n+2)}$$

Now let us compare them, we suspect that the RB estimator to have a lower variance, let us see if that is the case:

$$MSE[\hat{\theta}_{MM}] \geq MSE[\hat{\theta}_{RB}]$$

$$\frac{1}{12n} \geq \frac{n}{2(n+1)^2(n+2)}$$

$$n^3 - 2n^2 + 5n + 2 \geq 0$$

This polynomial has ONE real root  $n_1 \approx -0.34$ . Let us evaluate the sign of this polynomial at the left of this root (since we don't care about the right part since  $n$  is the number of observations, and so it should be a positive integer):

$$n = 1$$

$$1 - 2 + 5 + 2 = 6 > 0$$

Which implies that  $g(n) = n^3 - 2n^2 + 5n + 2 \geq 0$ ,  $\forall n \geq n_1 \approx -0.34$ , and so the inequality holds for  $n \in [-0.34, \infty)$ , and in our case we are interested in  $n \in \mathbb{N}$ , and so we have shown that the RB estimator has a lower MSE than the MM estimator  $\forall n$  (and of course  $\forall \theta \in \Theta$ ).

## Problem 4: Section [Sufficiency, Exponential Families, Rao-Blackwell, Construction of UMVUE]

Need to show:  $var[\hat{\psi}_{UMVUE}] > \frac{1}{I(\psi)}$  (strict inequality)

From now on I will denote  $\hat{\psi}_{UMVUE}$  as  $\hat{\psi}$

### Getting $I(\psi)$

In order to get this we need to get  $I(\mu)$ , since  $\psi = e^{t\mu}$  (i.e. a function of  $\mu$ ) and then we use the delta method to get  $I(\psi)$

$$I(\mu) = -E\left[\frac{d^2}{d\mu^2} \log(f(X, \mu))\right] = -E\left[\frac{d^2}{d\mu^2} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X-\mu)^2}{2\sigma^2}}\right)\right] = \frac{1}{\sigma^2}$$

Now we use the delta method to get  $I(\psi)$ :

Take  $g(\mu) = \psi = e^{t\mu}$ , where  $t \neq 0$ , and so  $g'(\mu) = te^{t\mu}$ , and so we get:

$$I(\mu) = I(\psi) \left[\frac{dg(\mu)}{d\mu}\right]^2$$

$$\frac{1}{\sigma^2} = I(\psi) t^2 e^{2t\mu}$$

and so we get:

$$I(\psi) = \frac{e^{-2t\mu}}{t^2 \sigma^2}$$

### Finding the UMVUE estimator, $\hat{\psi}$

We know from the notes "proposition" page 33, that in order to find an UMVUE of  $g(\theta)$  (in this case  $g(\mu)$ ) we need to solve this with respect to the function  $h$ :

$$E_{\theta}[h(T(X))] = g(\theta), \quad \forall \theta \in \Theta$$

where  $T(X)$  is a complete and sufficient statistic for  $\theta$ .

So in order to start looking for this function  $h$ , we need to find a complete and sufficient statistic for  $\mu$ , when  $\sigma^2$  is known.

And indeed we know one  $T(X) = \sum_{i=1}^n X_i$  (from the notes it is indeed complete and sufficient for  $\mu$ ).

Now let us start looking for the function  $h$ :

The first thing that we can try to do is to take  $h(\sum_i X_i) = e^{t \sum_i X_i}$ .

$$E[h(\sum_i X_i)] = E[e^{t \sum_i X_i}] = E[e^{tX_1} e^{tX_2} \dots e^{tX_n}] = E[e^{tX_1}] E[e^{tX_2}] \dots E[e^{tX_n}] = \prod_{i=1}^n M_{X_i}(t)$$

But we also know that the moment generating function of a normal distribution is (from homework 1, I showed this and also from notes 1):

$$M_{X_i}(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

And so we get:

$$E[h(\sum_i X_i)] = \prod_{i=1}^n e^{\mu t + \frac{\sigma^2 t^2}{2}} = e^{n\mu t} e^{\frac{n\sigma^2 t^2}{2}}$$

Now remember that we are trying to find  $h$  such that  $E[h(\sum_i X_i)] = e^{t\mu}$ , so let us adjust to remove the  $n$  in  $n\mu t$

We also know from notes 1 that the MGF of:

$$M_{bX}(t) = M_X(bt)$$

where here  $b = \frac{1}{n}$ , and  $X = \sum_i X_i$ , and so by taking  $h(T(X)) = e^{t \frac{1}{n} \sum_i X_i}$ , we get the expectation to be:

$$E[h(T(X))] = E[e^{t \frac{1}{n} \sum_i X_i}] = M_X(\frac{1}{n}t) = e^{t\mu} e^{\frac{\sigma^2 t^2}{2n}}$$

Now lastly let us adjust this by a constant factor  $e^{-\frac{\sigma^2 t^2}{2n}}$ , i.e. taking  $h(T(X)) = e^{-\frac{\sigma^2 t^2}{2n}} e^{t \frac{1}{n} \sum_i X_i}$  and so we get the expectation to be:

$$E[h(T(X))] = E[e^{-\frac{\sigma^2 t^2}{2n}} e^{t \frac{1}{n} \sum_i X_i}] = e^{-\frac{\sigma^2 t^2}{2n}} E[e^{t \frac{1}{n} \sum_i X_i}] = e^{t\mu}$$

notice the second equality is because the term  $e^{-\frac{\sigma^2 t^2}{2n}}$  is constant !!! Thus we have found the UMVUE estimator to be:

$$\hat{\psi} = e^{-\frac{\sigma^2 t^2}{2n}} e^{t \frac{1}{n} \sum_i X_i}$$

### Finding the variance of $\hat{\psi}$

We know that the variance of  $\hat{\psi}$  is given by:

$$\text{Var}[\hat{\psi}] = E[\hat{\psi}^2] - E[\hat{\psi}]^2$$

Let us first get the second term:

$$(E[\hat{\psi}])^2 = e^{2t\mu}$$

Now let us get the first term:

$$E[\hat{\psi}^2] = E[e^{-\frac{\sigma^2 t^2}{n}} e^{2t \frac{1}{n} \sum_i X_i}] = e^{-\frac{\sigma^2 t^2}{n}} E[e^{2t \frac{1}{n} \sum_i X_i}] = e^{-\frac{\sigma^2 t^2}{n}} e^{2t\mu} e^{\frac{2\sigma^2 t^2}{n}}$$

The third equality is also using the fact that:  $M_{2[\frac{1}{n} \sum_i X_i]}(t) = M_{\frac{1}{n} \sum_i X_i}(2t)$ , and the second MGF we already computed it above (so here just plug  $2t$  instead of  $t$ ).

And so we get that:

$$E[\hat{\psi}^2] = e^{2t\mu} e^{\frac{\sigma^2 t^2}{n}}$$

Thus we get the variance of  $\hat{\psi}$ :

$$\text{Var}[\hat{\psi}] = e^{2t\mu} e^{\frac{\sigma^2 t^2}{n}} - e^{2t\mu} = e^{2t\mu} (e^{\frac{\sigma^2 t^2}{n}} - 1)$$

### Showing that $\text{var}[\hat{\psi}] > \frac{1}{nI(\psi)}$ (strict inequality)

We know that  $I(\psi) = \frac{e^{-2t\mu}}{t^2 \sigma^2}$ , and so we get:

$$\frac{1}{nI(\psi)} = \frac{t^2 \sigma^2}{n e^{-2t\mu}}$$

Let us now compare the two, we know from the CR bound that:

$$e^{2t\mu} (e^{\frac{\sigma^2 t^2}{n}} - 1) \geq \frac{t^2 \sigma^2}{n e^{-2t\mu}}$$

$$e^{\frac{\sigma^2 t^2}{n}} \geq \frac{t^2 \sigma^2}{n} + 1$$

Now to compare these two sides, I will write the maclaurin expansion of  $e^x$  (or taylor about zero):

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

here  $x = \frac{\sigma^2 t^2}{n}$

$$e^{\frac{\sigma^2 t^2}{n}} = 1 + \frac{\sigma^2 t^2}{n} + \frac{[\frac{\sigma^2 t^2}{n}]^2}{2!} + \frac{[\frac{\sigma^2 t^2}{n}]^3}{3!} + \dots$$

And now notice how the first two terms in this series are exactly equal to the right hand side. And since the rest of the terms are positive (strictly positive, since  $\frac{\sigma^2 t^2}{n} > 0$ ), we can conclude that the inequality is strict.

**Thus, we have shown that  $\text{var}[\hat{\psi}] > \frac{1}{nI(\psi)}$  (strict inequality) .**

Side Note: Notice how when  $n \rightarrow \infty$  the inequality becomes an equality, but for finite  $n$  it is strict.

**Finding the efficiency of  $\hat{\psi}$**

$$e(\hat{\psi}) = \frac{1}{nI(\psi)\text{Var}[\hat{\psi}]} = \frac{1}{nI(\psi)e^{2t\mu}(e^{\frac{\sigma^2 t^2}{n}} - 1)} = \frac{t^2 \sigma^2}{n[e^{\frac{\sigma^2 t^2}{n}} - 1]}$$

$$e(\hat{\psi}) \approx \frac{t^2 \sigma^2}{t^2 \sigma^2 + \frac{t^4 \sigma^4}{2n} + \frac{t^6 \sigma^6}{n^2 6} + \dots}$$

Now we can see how as discussed before asymptotically the efficiency of the UMVUE estimator is 1, but for finite  $n$  it is strictly less than 1.

## Problem 5: Section [Sufficiency, Exponential Families, Rao-Blackwell, Construction of UMVUE]

**Part 1: Getting the UMVUE estimator for  $F(u) = P(X \leq u)$**

We have that our data are i.i.d from  $\sim N(\mu, 1)$ , we care to find the UMVUE estimator for  $F(u) = P(X \leq u)$ , for any fixed  $u$

**Using The Lehmann-Scheffe Theorem**

For Lehmann-scheffe theorem we require to have an Unbiased estimator for  $F(u)$ , and a complete and sufficient statistic for  $\mu$  (which are given in the question), and we remain to only find the Expectation of the Unbiased estimator given the complete sufficient statistic.



The unbiased estimator for  $F(u)$  is given by:

$$E[I(X < u)|\bar{X}_n] = 1.P(X < u|\bar{X}_n) + 0.P(X \geq u|\bar{X}_n) = P(X < u|\bar{X}_n) \quad (**)$$

And so to solve the problem we need to find the conditional distribution of  $X|\bar{X}_n$ .

And as we know in order to do this we need to get the the joint distribution of  $X$  and  $\bar{X}_n$ , and from it get the conditional. So let's do it:

**Step 1: In order to get the joint, we need the marginal of  $\bar{X}_n$ , let's get it:**

We know from notes one that the sum of independent normal r.v.s is also normal with  $N(\sum_i a_i \mu_i, \sum_i a_i^2 \sigma_i^2)$ , and since we have i.i.d  $X_i \sim N(\mu, 1)$ , so  $\mu_i = \mu$  and  $a_i = \frac{1}{n}$  and  $\sigma^2 = 1$ , we get:

$$\bar{X}_n \sim N(\mu, \frac{1}{n})$$

**Step 2: Getting the Joint dist.**  $Y = (X, \bar{X}_n)$

We know  $E[Y] = (\mu, \mu)$  and that:

$$\Sigma = \begin{bmatrix} Var[X] & Cov[X, \bar{X}_n] \\ Cov[\bar{X}_n, X] & Var[\bar{X}_n] \end{bmatrix}$$

Let us now get  $Cov[\bar{X}_n, X] = Cov[X, \bar{X}_n]$ : ( $X$  is on of the  $X_i$ 's say  $X_j$ )

$$Cov[\bar{X}_n, X] = Cov[\frac{1}{n} \sum_{i=1}^n X_i, X_j]$$

Now notice that we can take the sum and the  $\frac{1}{n}$  out of the covariance (properties of linear combinations of the covariance function):

$$= \frac{1}{n} \sum_{i=1}^n Cov(X_i, X_j) = \frac{1}{n} Cov(X_j, X_j) + \frac{1}{n} \sum_{i \neq j} Cov(X_i, X_j)$$

Notice that the second term is  $= 0$ , from independence of the  $X_i$ 's, and the first term is just the variance of  $X_j$  which is 1, and so we get:

$$Cov[\bar{X}_n, X] = \frac{1}{n}$$

And so we get the covariance matrix:

$$\Sigma = \begin{bmatrix} 1 & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} \end{bmatrix}$$

**Step 3: Getting the conditional distribution of  $X|\bar{X}_n$**

We know from the properties of the bivariate normal distribution ( $Y = (\bar{X}_n, X)$ ) that the conditional distribution of  $X|\bar{X}_n$  is normal (notes 1) with the  $\mu, \sigma^2$  given from the wasserman book page 40, thm 2.44 part 2:

$$E[X|\bar{X}_n] = \mu + Cov[X, \bar{X}_n] \frac{1}{Var[\bar{X}_n]} (\bar{X}_n - \mu)$$

And so we get:

$$E[X|\bar{X}_n] = \mu + \frac{1}{n} n (\bar{X}_n - \mu) = \bar{X}_n$$

And we know also that the variance of  $X|\bar{X}_n$  is:

$$Var[X|\bar{X}_n] = Var[X] - Cov[X, \bar{X}_n]^2 Var[\bar{X}_n]^{-1}$$

And so we get:

$$Var[X|\bar{X}_n] = 1 - \frac{1}{n^2} n = 1 - \frac{1}{n} = \frac{n-1}{n}$$

Thus we have that  $X|\bar{X}_n \sim N(\bar{X}_n, \frac{n-1}{n})$

**Getting the UMVUE estimator**

Going back to (\*\*), we have that the UMVUE estimator is:

$$\hat{F}_{UMVUE}(u) = P(X < u|\bar{X}_n) = P\left(\frac{X|\bar{X}_n - E[X|\bar{X}_n]}{Var[X|\bar{X}_n]^{0.5}} < \frac{u - E[X|\bar{X}_n]}{Var[X|\bar{X}_n]^{0.5}}\right)$$

But notice that we know that  $X|\bar{X}_n$  is normal, and this scaled form is the form of  $Z \sim N(0, 1)$  to get:

$$\hat{F}_{UMVUE}(u) = P\left(Z < \sqrt{\frac{n}{n-1}}(u - \bar{X}_n)\right) = \Phi\left(\sqrt{\frac{n}{n-1}}(u - \bar{X}_n)\right)$$

## Part 2: Simulation and Comparison

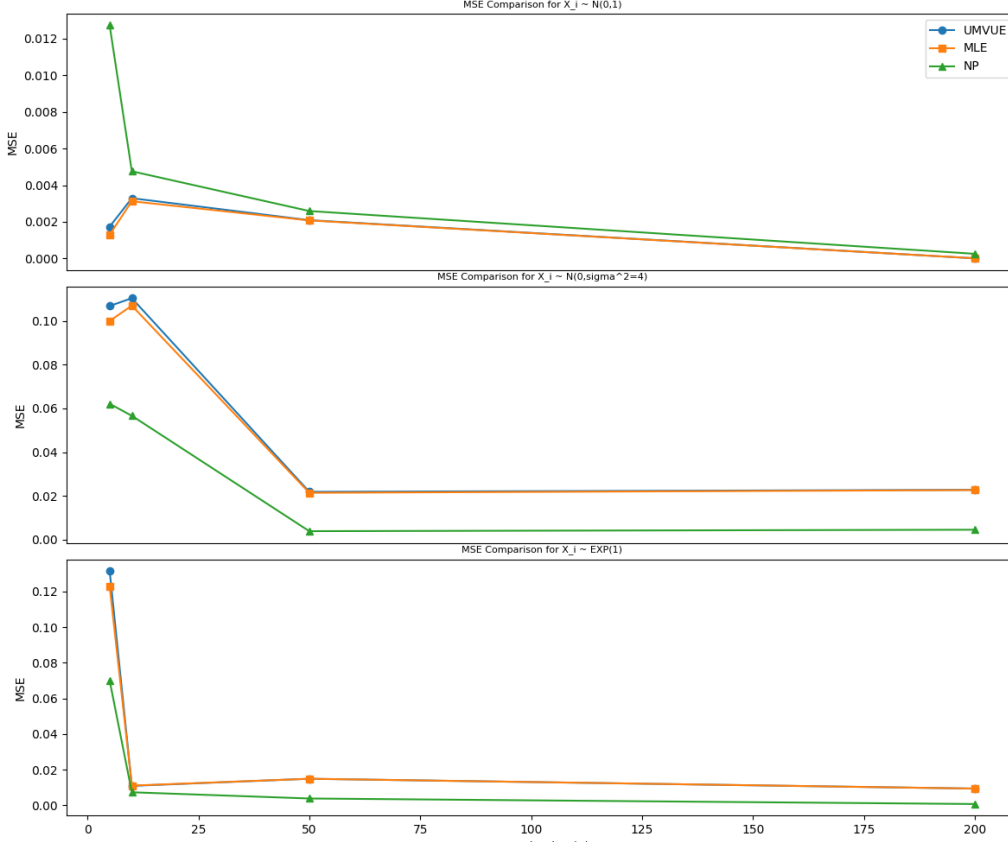


Figure 12: Code available in: Comparing\_CDFS\_USING\_MSE.py

### Comments on the plot

- We can see how in the first case (i.e.  $X_i \sim N(0,1)$ ), the MLE and UMVUE estimators do a very good job in estimating the true cdf, which is expected, since both of these estimators, assumed that the data  $X_i \sim N(\mu,1)$ . And this can be concluded implicitly from their formulas,  $\hat{F}_{MLE}(u) = \Phi(u - \bar{X}_n)$ , since the  $\bar{X}_n$  will converge to the true mean  $\mu$ , and so  $\Phi(u - \bar{X}_n) \approx \Phi(u - \mu)$  for large  $n$ , and the same logic applies to the UMVUE estimator. So as expected when we keep on increasing  $n$ , we are seeing  $MSE \rightarrow 0$

- We can see from the three graphs how the MLE estimator becomes more and more close as  $n$  increases, to the UMVUE estimator(which verifies the property of the MLE, asymptotically efficient)
- From the three figures we can see how when we increase  $n$ , the empirical cdf gets more and more accurate( since it becomes "closer" to the true cdf), and so the MSE decreases.
- But now notice how in the second case, when we change the variance  $\sigma^2 = 4$ , the non-paramateric estimator does a better job in estimating the true cdf. Since both the MLE and UMVUE assumed the data to have a variance of "1"(not the case in 2) and are normal(not the case in 3).
- Now looking at the second and third plot we can see how when we don't know the true distirbution of the data (or we are not sure about some parameters "in this case  $\sigma^2$ " or wrong) the non-parametric estimator does a better job in estimating the true cdf.

## Problem 6: Section [Sufficiency, Exponential Families, Rao-Blackwell, Construction of UMVUE]

We can get the  $\hat{\eta}_{MLE}$  very easily by solving the equation:

$$\sum_{i=1}^n T(X_i) = n \frac{\partial A(\eta)}{\partial \eta}$$

Which is equivalent to solving the dervative of the log-likelihood function with respect to  $\eta$  equal to zero.

So we get:

$$\hat{\eta}_{MLE} = g^{-1}\left(\frac{1}{n} \sum_{i=1}^n T(X_i)\right)$$

Where  $g^{-1}$  is the inverse of the function  $g(\eta) = \frac{\partial A(\eta)}{\partial \eta}$

TO be able to write this we need to make sure that  $g^{-1}$  exists, but in this case as we were told in the question, assume that  $\hat{\eta}_{MLE}$  exists.

### CR bound

We know that the CR bound is given by:

$$Var[\hat{\eta}_{MLE}] \geq \frac{\left\{ \frac{dE[\hat{\eta}_{MLE}]}{d\eta} \right\}^2}{nI(\eta)}$$

Which in general is not the case since, unless specific conditions are met.

### Finding the function $\tau = g(\eta)$ whose MLE reaches the bound

Very straightforward, since we know from the one dimensional case the  $T(X)$  is UMVUE of its mean  $E[T]$  (which we will now also verify), so let's take:  $\tau = E[T]$ , and so we find that  $\hat{\tau}_{MLE} = \bar{T}$  (from the equivariant property of the MLE to be able to conclude this).

### Verifying it is unbiased

$$E[\hat{\tau}_{MLE}] = \frac{1}{n} n E[T] = E[T] = \tau$$

Thus, we have shown that  $\hat{\tau}_{MLE}$  is unbiased.

### Verifying that it reaches the CR bound, and thus the UMVUE (since also unbiased)

We know that the CR bound is given by:

$$Var[\hat{\tau}_{MLE}] \geq \frac{1}{nI(\tau)}, \quad \text{The numerator is 1 since unbiased}$$

Let us first get  $I(\tau)$ :

$$I(\eta) = I(\tau) \left\{ \frac{\partial E[T]}{\partial \eta} \right\}^2 = I(\tau) \left\{ \frac{\partial \tau}{\partial \eta} \right\}^2$$

Now, let us look at  $\frac{\partial \tau}{\partial \eta}$

$$\frac{\partial \tau}{\partial \eta} = \frac{\partial E[T]}{\partial \eta} = \frac{\partial^2 A(\eta)}{\partial \eta^2} = I(\eta)$$

And so going back:

$$I(\eta) = I(\tau)I(\eta)^2$$

And so:

$$I(\tau) = \frac{1}{I(\eta)} = \frac{1}{Var[T]} \quad (\text{From Notes, since here } m = 1)$$

Now let us get  $VAR[\hat{\tau}_{MLE}]$ :

$$VAR[\hat{\tau}_{MLE}] = \frac{Var[T]}{n}$$

Going back to the bound:

$$Var[\hat{\tau}_{MLE}] \geq \frac{1}{nI(\tau)} = \frac{1}{n \frac{1}{Var[T]}} = \frac{Var[T]}{n}$$

Which is exactly equal to  $VAR[\hat{\tau}_{MLE}]$ , thus we have just shown that the MLE of  $\tau$  reaches the minimum possible variance!!!

## Last Section: From Exam 2023

### Part 1: Minimal sufficient statistic for $X = (X_1, \dots, X_n)$

To get this I will first prove that the  $Exp(\lambda)$  distribution is in the exponential family.

The pdf of the  $Exp(\lambda)$  is:

$$f(x; \lambda) = \lambda e^{-\lambda x} = e^{-\lambda x + \log(\lambda)}$$

Which give us that:

- $h(X) = 1$
- $\theta = \lambda$

- $T(X) = X$
- $\eta(\theta) = -\lambda$
- $A(\eta) = -\log(\lambda) = -\log(-\eta)$

And so for the joint density of  $X = (X_1, \dots, X_n)$  we get:

$$T(X) = \sum_{i=1}^n X_i$$

from notes, or just expand the density for one for multiple  $X$ , which is then the product (from independence) of the densities of each  $X_i$ .

And so we have that  $T(X)$  is a sufficient statistic for  $\lambda$ .

Let us now Prove that  $T(X)$  is minimal sufficient.

We will proceed as we did in previous sections:

**First Implications: Assume  $T(x) = T(x^*)$ , prove that ratio of densities is independent of  $\lambda$**

Ratio is:

$$\frac{f(x; \lambda)}{f(x^*; \lambda)} = \frac{e^{-\lambda T(x) + n \log(\lambda)}}{e^{-\lambda T(x^*) + n \log(\lambda)}} = e^{-\lambda(T(x) - T(x^*))} = e^{-\lambda(T(x) - T(x^*))} = 1$$

**Second Implications: Assume Ratio is independent of  $\lambda$**

From the form above we see that the ratio is independent of  $\lambda$  gives us that

$$T(x) = T(x^*).$$

Thus  $T(X) = \sum_{i=1}^n X_i$  is minimal sufficient.

## Part 2: Getting the UMVUE estimator for $\lambda$

**We know that  $T$  is complete**

And so now our job is to find:

$$E_\lambda[h(T(X))] = \lambda$$

And so we need to find the function  $h$  that satisfies this, once we do then  $h(T)$  is our UMVUE. (From proposition in the notes)

We know that  $T \sim \Gamma(\alpha = n, \beta = \frac{1}{\lambda})$  (sum of exponentials!)

So let us first try to take  $h(T) = T$  (identity function), we see that  $E_\lambda[T] = \frac{n}{\lambda}$ .

So one approach that I will try is to take  $h(T) = \frac{1}{T}$ , let us compute the expectation of this:

$$E\left[\frac{1}{T}\right] = \int_0^\infty \frac{1}{t} \frac{1}{\left(\frac{1}{\lambda}\right)^n \Gamma(n)} t^{n-1} e^{-t\lambda} dt = \frac{\lambda^n}{(n-1)!} \int_0^\infty t^{n-2} e^{-t\lambda} dt$$

Remember the Gamma Function is:

$$\int_0^\infty u^{\alpha-1} e^{-u} du = \Gamma(\alpha), \quad \text{Equation(1)}$$

So let us adjust the integral to look like the gamma function

Take  $u = \frac{t}{\lambda}$ , and so  $dt = \frac{du}{\lambda}$ , the bound stay the same "just plug in the substitution":

$$\frac{\lambda^n}{(n-1)!} \frac{1}{\lambda^{n-1}} \int_0^\infty u^{n-2} e^{-u} du$$

and now this integral is the gamma function with  $\alpha = n - 1$ , and so the integral evaluates to  $(n-2)!$ , we get after simplifications:

$$E\left[\frac{1}{T}\right] = \frac{\lambda}{(n-1)}$$

And so by looking at this we can scale this current function by  $(n-1)$  (which is a constant so it doesn't affect the expectation) to get the UMVUE estimator:

$$\hat{\lambda}_{UMVUE} = (n-1) \frac{1}{T} = \frac{n-1}{\sum_{i=1}^n X_i}$$

### Part 3: CR bound

We know that the CR bound is given by:

$$\text{Var}[\hat{\lambda}_{UMVUE}] \geq \frac{1}{nI(\lambda)}$$

Let us check that this is a straight inequality (i.e.  $\text{var}[\hat{\lambda}_{UMVUE}] > \frac{1}{nI(\lambda)}$ ).

**Getting  $I(\lambda)$ :**

Very straight-forward:

$$I(\lambda) = -E\left[\frac{d^2}{d\lambda} \log(f(X_i; \lambda))\right] = \frac{1}{\lambda^2}$$



So the bound is:

$$Var[\hat{\lambda}_{UMVUE}] \geq \frac{\lambda^2}{n}$$

**Getting the variance of the UMVUE estimator:**

$$Var[\hat{\lambda}_{UMVUE}] = E[\hat{\lambda}_{UMVUE}^2] - E[\hat{\lambda}_{UMVUE}]^2$$

The second term we know  $E[\hat{\lambda}_{UMVUE}]^2 = \lambda^2$

The first term is:

$$E[\hat{\lambda}_{UMVUE}^2] = E[(n-1)^2 \frac{1}{T^2}] = (n-1)^2 E[\frac{1}{T^2}]$$

And by proceeding as before we find  $E[\frac{1}{T^2}] = \frac{\lambda^2}{(n-1)(n-2)}$  So the first term becomes:

$$E[\hat{\lambda}_{UMVUE}^2] = (n-1) \frac{\lambda^2}{(n-2)}$$

Thus the variance is:

$$Var[\hat{\lambda}_{UMVUE}] = (n-1) \frac{\lambda^2}{(n-2)} - \lambda^2 = \frac{\lambda^2}{n-2}$$

Comparing the two we get:

$$\frac{\lambda^2}{n-2} > \frac{\lambda^2}{n}$$

Which is true for  $n \notin (0, 2)$ , and we have that  $n = 50 > 2$ , thus we have shown that indeed this UMVUE does NOT achieve the CR bound.

**Efficiency of the UMVUE estimator**

$$e(\hat{\lambda}_{UMVUE}) = \frac{1}{nI(\lambda)Var[\hat{\lambda}_{UMVUE}]} = \frac{1}{n \frac{1}{\lambda^2} \frac{\lambda^2}{n-2}} = \frac{n-2}{n} = \frac{48}{50} = 0.96$$

**Part 4: Arguing why it does not achieve the CR bound but still does NOT contradicts what we saw in the slides**

We know from the \*remark page 35\* that the sufficient statistic  $T$  (when  $m=1$ , i.e. one dimensional exponential family) is the unbiased estimator of its mean ( $E[T(X)]$ ), which in this case turns out to be  $\frac{n}{\lambda}$  and NOT  $\lambda$ , so this doesn't contradict what we just got, since,  $Var(T[X])$  does indeed reach

the information bound but for  $I(\frac{n}{\lambda})$ (as expected from the statement), let us verify this fact:

**Step 1: Get  $I(\frac{n}{\lambda}) = I(n\mu)$**

$$I(n\mu)\{\frac{d\frac{n}{\lambda}}{d\lambda}\}^2 = I(\lambda)$$

But we know from before that  $I(\lambda) = \frac{1}{\lambda^2} = \mu^2$

$$I(n\mu) = \frac{1}{n^2\mu^2}$$

**Step 2: Let us now get the variance of  $T(X)$**

Very easy since we know that  $T(X) \sim \Gamma(\alpha = n, \beta = \mu = \frac{1}{\lambda})$ :

$$Var[T(X)] = n\mu^2$$

**Step 3: Now let us compare  $Var[T(X)]$  with  $\frac{1}{nI(n\mu)}$**

We know from the CR bound:

$$Var[T(X)] \geq \frac{1}{nI(n\mu)}$$

$$n\mu^2 \geq n\mu^2$$

Thus this is a straight inequality( $Var[T(X)] = \frac{1}{nI(n\mu)}$ ).

One more thing  $E[T(X)] = \frac{n}{\lambda} = n\mu$ .

And so that  $T(X)$  is UMVUE estimator of  $n\mu$  (which verifies what we saw in the slides).

## Part 5: Effects of stopping before reaching n=50

Yes, they can stop before reaching  $n = 50$ .

**BUT THEY SHOULD BE AWARE OF COUPLE OF THINGS:**

The sufficient statistic should now include another term and that is the number of points they stopped at, since the number of points they stopped at is a function of the data, and so it should be included in the sufficient statistic. The second part of the sufficient statistic will remain the same (but with a different  $n$ ),  $T(X) = \sum_{i=1}^k X_i$ , where  $k$  is the number of points they stopped at.

But we should keep in mind that the efficiency of the  $\hat{\lambda}_{UMVUE}$ , will become lower:

$$e(\hat{\lambda}_{UMVUE}) = \frac{k-2}{k} < \frac{n-2}{n}, \text{ for } k < n$$

Here assuming we stopped at  $k > 2$ (for what was discussed above!)

Of course the remaining properties will remain the same, but instead of  $n$  we will have  $k$ .

So they can stop, but they should be aware that the efficiency of the estimator will decrease.

## Part 6: Getting the distribution of $W_i = X_i | X_i \geq \tau$

Very simple since we know:

$$f_{X_i|X_i > \tau}(w) = \frac{f_{X_i}(x_i)}{P(X_i > \tau)}$$

since we want to rescale the density of  $f_{X_i}$  to only accounts for the values that are greater than  $\tau$ .

And so we get:

$$f_{X_i|X_i > \tau}(w) = \frac{\lambda e^{-\lambda w}}{e^{-\lambda \tau}} = \lambda e^{-\lambda(w-\tau)}$$

when  $w > \tau$  and 0 otherwise, so in otherwords:

$$f_{X_i|X_i > \tau}(w) = I(w > \tau) \lambda e^{-\lambda(w-\tau)}$$

We can clearly see that this is NOT an exponential, and this is because of the term  $I(w > \tau)$ , which makes it impossible to break it down into a term depending on  $w$  (which we would take in  $h(w)$ ) and a term depending on  $\tau$ (which we would have taken in  $A(\eta)$ ).

## Part 7: Finding a minimal sufficient statistic for $(\lambda, \tau)$ —Family of $W$

Let us first find the joint density of  $W = (W_1, ..W_k)$ , we know that they independent and so the joint density of independent r.v.s is the product of their densities.

And so we get:

$$f(W; \lambda, \tau) = \prod_{i=1}^n f_{X_i|X_i > \tau}(w_i) = \prod_{i=1}^k I(w_i > \tau) \lambda e^{-\lambda(w_i-\tau)}$$

$$\begin{aligned}
&= I(w_1 > \tau)I(w_2 > \tau) \dots I(w_k > \tau) \lambda^k e^{-\lambda \sum_{i=1}^k (w_i - \tau)} \\
&= I(\min\{w_1, \dots, w_n\} > \tau) \lambda^n e^{-\lambda \sum_{i=1}^k w_i + k\lambda\tau}
\end{aligned}$$

So we found:

$$f(W; \lambda, \tau) = I(w_{(1)} > \tau) \lambda^k e^{-\lambda \sum_{i=1}^k w_i + k\lambda\tau}$$

So one sufficient statistic would be  $T(W) = (\min\{W\}, \sum_{i=1}^k W_i)$ , since we can use the factorization theorem to prove that indeed it is sufficient, by taking  $h(W) = 1$ .

### Proving it is minimal sufficient

**Goal Prove:**  $\frac{f(w|\tau, \lambda)}{f(w^*|\tau, \lambda)}$  Does not depend on  $(\tau, \lambda) \iff T(w) = T(w^*)$

**First Implications: Assume**  $T(w) = T(w^*)$ , **prove that ratio of densities is independent of**  $(\lambda, \tau)$

$$\frac{f(w|\tau, \lambda)}{f(w^*|\tau, \lambda)} = \frac{I(w_{(1)} > \tau) \lambda^k e^{-\lambda \sum_{i=1}^k w_i + k\lambda\tau}}{I(w_{(1)}^* > \tau) \lambda^k e^{-\lambda \sum_{i=1}^k w_i^* + k\lambda\tau}} = \frac{I(w_{(1)} > \tau) e^{-\lambda \sum_{i=1}^k w_i}}{I(w_{(1)}^* > \tau) e^{-\lambda \sum_{i=1}^k w_i^*}}$$

But notice that we know (from our assumption) that  $T(w) = T(w^*)$ , and so we get that  $w_{(1)} = w_{(1)}^*$  and  $\sum_{i=1}^k w_i = \sum_{i=1}^k w_i^*$ , and so the ratio is indeed independent of  $(\lambda, \tau)$ .

**Second Implications: Assume Ratio is independent of**  $(\lambda, \tau)$

Also very straight-forward:

From the ratio:

$$\frac{I(w_{(1)} > \tau) e^{-\lambda \sum_{i=1}^k w_i}}{I(w_{(1)}^* > \tau) e^{-\lambda \sum_{i=1}^k w_i^*}}$$

for it to be independent of  $(\lambda, \tau)$ , we need to have that  $w_{(1)} = w_{(1)}^*$  and  $\sum_{i=1}^k w_i = \sum_{i=1}^k w_i^*$ , and so we get that  $T(w) = T(w^*)$ .

## Part 8: Getting the UMVUE estimator for $\lambda$

Since we have that our previous statistic is complete and sufficient, we can use the Lehmann-Scheffe theorem to get the UMVUE estimator for  $\lambda$ .

Our goal: Find a function  $h$  such that  $E_\lambda[h(T(W))] = \lambda$ ,

very simple we proceed as we did in Part 2 of this problem.

We start by taking  $h(T(W)) = \sum_{i=1}^n (w_i - \tau)$  (but since  $\tau$  is unknown here I will put instead of  $\tau$ ,  $w_1$ ) So now I will try and find the expectation of this function:

$$E\left[\sum_{i=1}^n (w_i - w_{(1)})\right]$$

To help us in this case since we don't know the distribution of  $w_{(1)}$ , we will proceed with the hint, that  $(w_{(i)} - w_{(i-1)})(n - i + 1) \text{ i.i.d} \sim \text{Exp}(\lambda)$ , so we need a way to express  $h(T) = \sum_{i=1}^n (w_i - w_{(1)})$  in terms of the hint. Let's go step by step:

$$h = \sum_{i=1}^n (w_i - w_{(1)}) = \sum_{i=1}^n (w_{(i)} - w_{(1)}) = \sum_{i=2}^n (w_{(i)} - w_{(1)})$$

Now let us look at  $w_{(i)} - w_{(1)}$

$$w_{(i)} - w_{(1)} = \sum_{k=2}^i \sigma_k$$

with  $\sigma_k = w_{(k)} - w_{(k-1)}$ .

Let us now un-wrap the double sum, in terms of the  $\sigma_k$ 's:

$$\sum_{i=2}^n (w_{(i)} - w_{(1)}) = \sum_{i=2}^n \sum_{k=2}^i \sigma_k = \sigma_2 + (\sigma_2 + \sigma_3) + (\sigma_2 + \sigma_3 + \sigma_4) + (\sigma_2 + \sigma_3 + \sigma_4 + \sigma_5) + \dots + \sigma_n$$

So we can see that each  $\sigma_k$  appears  $n - k + 1$  times, and so we can write the sum as:

$$\sum_{i=2}^n (w_{(i)} - w_{(1)}) = \sum_{k=2}^n \sigma_k (n - k + 1) = \sum_{k=2}^n (w_{(k)} - w_{(k-1)})(n - k + 1)$$

And so now we have:

$$h = \sum_{k=2}^n (w_{(k)} - w_{(k-1)})(n - k + 1)$$

From the hint we know that  $(w_{(k)} - w_{(k-1)})(n - k + 1) \text{ i.i.d} \sim \text{Exp}(\lambda)$ , and so we can write the expectation of  $h$  as:

$$E[h] = \frac{n-1}{\lambda}$$

But this is not what we want, let us now find another function  $h$ , taking  $h = \frac{1}{\sum_{i=1}^n (w_i - w_{(1)})} = \frac{1}{\sum_{i=2}^n (w_i - w_{(1)})}$  WE know from part 2 that  $E[\frac{1}{\sum_i Y_i}]$ , where  $Y_i$  are n i.i.d  $\sim Exp(\lambda)$  is  $\frac{\lambda}{n-1}$ , applying here to our "n" which is "n-1" to get:

$$E[h] = \frac{\lambda}{n-2}$$

And thus just by scaling this function by  $(n-2)$  we get the UMVUE estimator for  $\lambda$ :

$$\hat{\lambda}_{UMVUE} = \frac{n-2}{\sum_{i=1}^n (w_i - w_{(1)})}$$

## Part 9: Pivot CI for $\tau$

So as a starting point we will start by finding the distribution of  $W_{(1)}$ , since we will proceed as the hint with the statistic found previously. Very easy and very similar to what we did in the Uniform distribution case:

$$\begin{aligned} F_{W_{(1)}}(w) &= P(W_{(1)} \leq w) = 1 - P(W_{(1)} > w) = 1 - P(W_1 > w, W_2 > w, \dots, W_n > w) \\ &= 1 - P(W_1 > w)P(W_2 > w) \dots P(W_n > w) = 1 - (1 - F(w))^n \end{aligned}$$

and so we can find the CDF of  $W$ :

$$F_W(w) = 1 - e^{w-\tau}$$

Plugging back in we find:

$$F_{W_{(1)}}(w) = 1 - e^{-n\lambda(w-\tau)}, \quad \text{for } w > \tau$$

Now we can see that this is shifted exponential  $Exp(n\lambda) + \tau$ . Now our goal is to make this independent of  $\tau$ , so to make it easier for us I will remove  $\tau$  from  $W_{(1)}$  by taking  $W_{(1)} - \tau$ , and so we get its distribution to be  $Exp(n\lambda)$ . Which is good:

$V = W_{(1)} - \tau \sim Exp(n\lambda)$ , and so we can use this "V" as our pivot.

We know  $F_V(v) = 1 - e^{-\lambda n v}$

$$P(a \leq V \leq b) \geq 1 - \alpha$$

Now we know that  $V > 0$ , and so

$$P(0 \leq V \leq b) = 1 - e^{-\lambda n b} \geq 1 - \alpha$$

TO get this  $b$  I will set them equal to each other:

$$1 - e^{-\lambda nb} = 1 - \alpha \rightarrow b = -\frac{\log \alpha}{n\lambda}$$

And so going back:

$$P(0 \leq w_{(1)} - \tau \leq -\frac{\log \alpha}{n\lambda}) \geq 1 - \alpha$$

And so we get:

$$P\left(w_{(1)} + \frac{\log \alpha}{n\lambda} \leq \tau \leq w_{(1)}\right) \geq 1 - \alpha$$

And so we get the pivot CI for  $\tau$ :

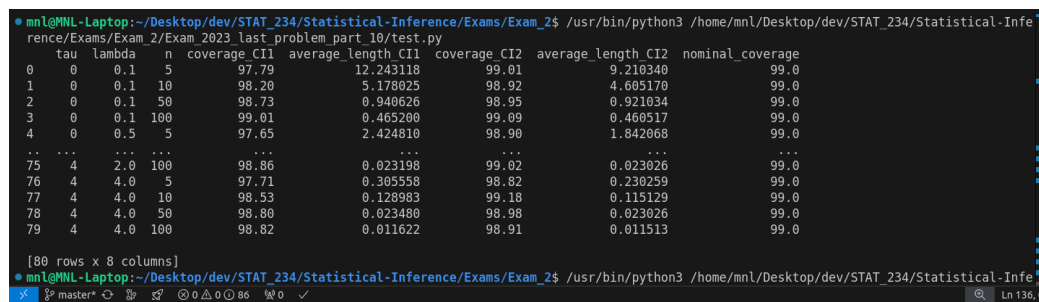
$$\left[w_{(1)} + \frac{\log \alpha}{n\lambda}, w_{(1)}\right]$$

But here notice that we don't know  $\lambda$ , so we plug instead of it its estimate  $\hat{\lambda}_{UMVUE}$  obtained in part 8.

And so this won't give us an exact CI, with coverage of  $1 - \alpha$ , but it will give us an approximate CI.

## Part 10: Simulation of confidence interval

### CI 1 VS CI 2 VS Normal Coverage (Taking $\alpha = 0.05$ )



```

mnl@mnl-Laptop: ~/Desktop/dev/STAT_234/Statistical-Inference/Exams/Exam_2$ /usr/bin/python3 /home/mnl/Desktop/dev/STAT_234/Statistical-Inference/Exams/Exam_2/Exam_2023_last_problem_part_10/test.py
  tau  lambda    n coverage_C1 average_length_C1 coverage_C2 average_length_C2 nominal_coverage
0      0      0.1     5      97.79      12.243118      99.01      9.210340             99.0
1      0      0.1    10      98.20       5.178025      98.92      4.605170             99.0
2      0      0.1    50      98.73      0.940626      98.95      0.921034             99.0
3      0      0.1   100      99.01      0.465200      99.09      0.460517             99.0
4      0      0.5     5      97.65      2.424810      98.90      1.842068             99.0
..    ...    ...    ...    ...    ...    ...    ...
75     4      2.0   100      98.86      0.023198      99.02      0.023026             99.0
76     4      4.0     5      97.71      0.305558      98.82      0.230259             99.0
77     4      4.0    10      98.53      0.128983      99.18      0.115129             99.0
78     4      4.0    50      98.80      0.023480      98.98      0.023026             99.0
79     4      4.0   100      98.82      0.011622      98.91      0.011513             99.0

[80 rows x 8 columns]
mnl@mnl-Laptop: ~/Desktop/dev/STAT_234/Statistical-Inference/Exams/Exam_2$ /usr/bin/python3 /home/mnl/Desktop/dev/STAT_234/Statistical-Inference/Exams/Exam_2/Exam_2023_last_problem_part_10/test.py

```

Figure 13: Quick Look at the table

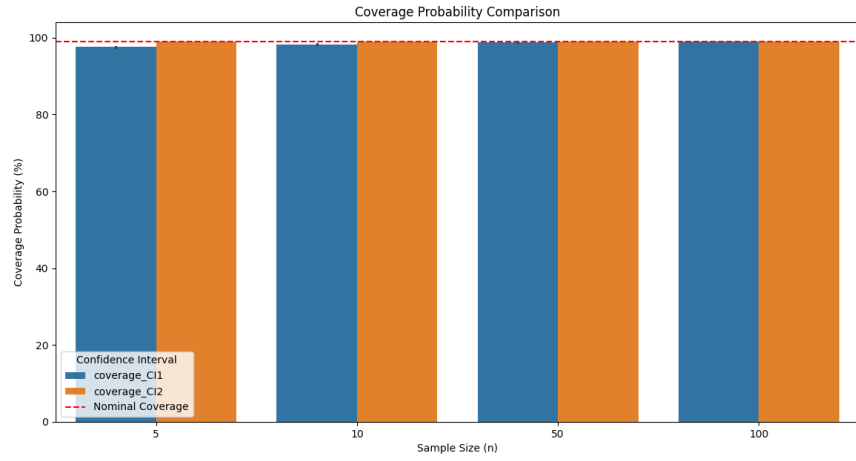


Figure 14: Comparison of the coverage probabilities of the two CIs

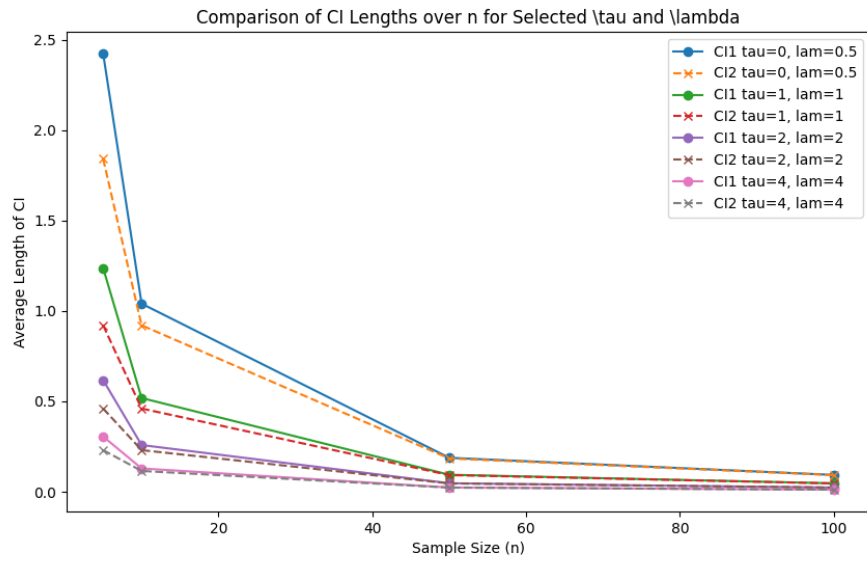


Figure 15: Comparison of the lengths of the two CIs



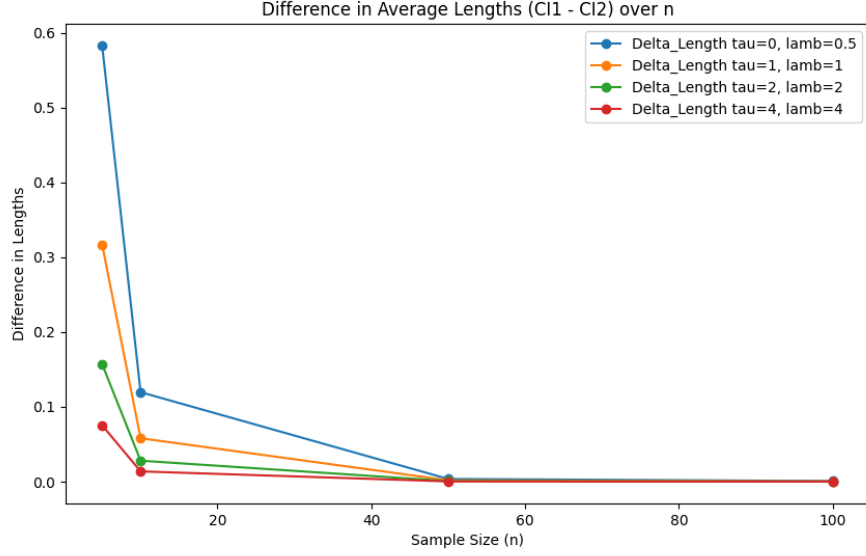


Figure 16: Comparison of the difference in lengths of the two CIs

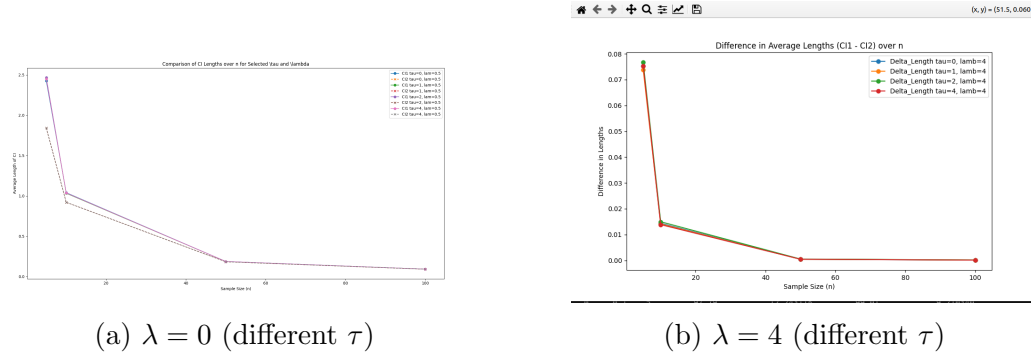


Figure 17: Effect of  $\tau$  (as we can see  $\tau$  DO NOT AFFECT the length of the interval)

## Comments on the plots

**From Figure 14:** We can clearly see how the  $CI_2$  is an exact CI, with coverage of  $1 - \alpha$ , and how the  $CI_1$  is an approximate CI, but we can see how the coverage of the  $CI_1$  is very close to the coverage of the  $CI_2$ , especially

as  $n$  increases !

**From Figure 15:** We can see how indeed the  $CI_2$  has a smaller length than the  $CI_1$ , and this is expected since the  $CI_2$  is an exact CI, and so it should have a smaller length than the approximate CI. But similar to above we see how as  $n$  increases there the  $CI_1$  length becomes very close to the length of  $CI_2$ . This can be easily see explicitly from **Figure 16**

**Figure 17** We can see from figure 17, how  $\tau$  do NOT have any effect on the length of the interval, and it is the value of  $\lambda$  that has the effect on the length of the interval. Which can be see from the difference between figures \*15\* and \*17\* where we can see in 15 there is difference in each scenario, which from 17 we can see that this is the effect of  $\lambda$  (and  $n$  of course) and NOT  $\tau$ .

**Figure 17** WE can see from the difference between (a) and (b) how as  $\lambda$  increases the length of the interval decreases, and this is expected since the length of the interval is inversely proportional to  $\lambda$ . (same for  $n$ )

**From all** We can see how the *UMVUE* estimator when used instead of  $\lambda$  in the CI, gives us a very good approximation of the exact CI, especially as  $n$  increases.

## PART 10: Bootstrap Confidence Interval, from the physic-slab.txt Data

```
the W_min values = 2.738

CI1 (Using UMVUE estimate): (np.float64(1.800168536658994), np.float64(2.738))
Bootstrap Pivotal Interval: (np.float64(1.8679999999999999), np.float64(2.738))
Bootstrap Normal Interval: (np.float64(2.232239724015083), np.float64(3.243760275984917))
Bootstrap Percentile Interval: (np.float64(2.738), np.float64(3.608))
```

Figure 18: Code available in: part\_2.py

And so we can see how  $CI_1$  and the Bootstrap Pivotal Interval are appropriate and reliable for estimating  $\tau$  from the shifted exponential distribution. Especially since the pivotal interval (in bootstrap) is somewhat similar to the  $CI_1$ . We can see how the normal and percentile:

DO NOT RESPECT THE CONDITION  $\tau \leq W_{min}$  We saw this from the upper bound of the confidence interval ! ( which give us that these two intervals are NOT valid)