# ST117 E3

Homework Lab Group 003 Pod E

## This submission was created by:


1. Name and WARWICK ID: DANIEL GUO 5645242 Question A, B
2. Name and WARWICK ID: ZHIJIAN LIN 5655296 Question A, B
3. Name and WARWICK ID: QINLING SI 5614637 Question C
4. Name and WARWICK ID: TOM O'CONNELL 5628105 Question C


## Question A

### 1. Estimation of Exponential Distribution Parameter

Let $X_1, X_2, ...X_n$ be i.i.d. random variables following an exponential distribution with rate $\lambda$. We estimate $\theta = 1/\lambda$ using two estimators:

1. $\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$
2. $\hat{\theta}_2 = \frac{1}{n+1}\sum_{i=1}^{n} X_i$

**Bias Calculation:**

Since $X_i \sim Exp(\lambda)$, the expectation of each $X_i$ is:

$E[X_i] = \theta$ it is unbiased

Therefore, the bias of $\hat{\theta}_1$ is:

$$E[\hat{\theta}_1] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = \frac{1}{n}n\theta = \theta$$

Therefore, estimator $\hat{\theta}_1$ is unbiased.

The bias of $\hat{\theta}_2$ is:

$$E[\hat{\theta}_2] = E[\frac{1}{n+1}\sum_{i=1}^{n} X_i] = \frac{1}{n+1}E[\sum_{i=1}^{n} X_i] = \frac{n}{n+1}\theta$$

Therefore it is biased and the bias is:

$$E[\hat{\theta}_2] - \theta = \frac{n}{n+1}\theta - \theta = -\frac{\theta}{n+1}$$

Underestimation by $\frac{\theta}{n+1}$.

**Variance Calculation:**

The variance of $X_i$ is:

$$Var(X_i) = \frac{1}{\lambda^2} = \theta^2$$

And since each $X_i$ is independent, the variance sum will be:

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

Therefore, the variance of $\hat{\theta}_1$ is (using the variance formula:

$$Var(\hat{\theta}_1) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}n\theta^2 = \frac{\theta^2}{n}$$

and the variance of $\hat{\theta}_2$ is:

$$Var(\hat{\theta}_2) = Var(\frac{1}{n+1}\sum_{i=1}^{n} X_i) = \frac{n\theta^2}{(n+1)^2}$$

**Mean Squared Error (MSE) Calculations:**

Formula: $MSE(\hat{\theta}) = Var(\hat{\theta} + (Bias(\hat{\theta}))^2$

Therefore, since $\hat{\theta}_1$ is unbiased, its MSE is:

$$MSE(\hat{\theta}_1) = Var(\hat{\theta}_1) = \frac{\theta^2}{n}$$

And $\hat{\theta}_2$ MSE is:

$$\begin{aligned} MSE(\hat{\theta}_2) &= Var(\hat{\theta}_2) + (Bias(\hat{\theta}_2))^2 \\ &= \frac{n\theta^2}{(n+1)^2} + (-\frac{\theta}{n+1})^2 = \frac{n\theta^2 + \theta^2}{(n+1)^2} \\ &= \frac{\theta^2(n+1)}{(n+1)^2} = \frac{\theta^2}{n+1} \end{aligned}$$

## 2.

**Bias:**

$\hat{\theta}_1$ is unbiased while $\hat{\theta}_2$ is underestimated.

**Variance:**

$\hat{\theta}_2$ has slightly lower variance than $\hat{\theta}_1$ since $\frac{n\theta^2}{(n+1)^2} \leq \frac{\theta^2}{n}$ for all n.

**MSE:**

$\hat{\theta}_2$ has slightly lower variance than $\hat{\theta}_1$ since $\frac{\theta^2}{n+1} \leq \frac{\theta^2}{n}$ for all n.

**Conclusion**

Estimator $\hat{\theta}_1$ is preferred if unbiasedness is the highest priority.

Estimator $\hat{\theta}_2$ has lower MSE and variance, so it could perform better in terms of overall error and lower variance, despite its small bias. So that $\hat{\theta}_2$ would be preferred when n is small, since as the sample space increases, the variance and MSE and the differences between them converge to 0.

# Question B

## 1.

Assume the waiting times between thefts follow a geometric distribution, which models the no. trials til the first theft. The PMF of a geometric random variable X with parameter p is:

$P(X = k) = (1 - p)^(k - 1)p$ for $k = 1, 2, 3, 4....$

The expectation of a geometric random variable is: $E[X] = \frac{1}{p}$

**Method of Moments Estimator:**

Equate the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ to the expectation:

$$\hat{p}_M oM = \frac{1}{\overline{X}} = \frac{n}{\sum_{i=1}^{n} X_i}$$

**Maximum Likelihood Estimator:**

The likelihood function is:

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

The log-likelihood is:

$$Log(L(p)) = \sum_{i_1}^{n}((X_i - 1)log(1 - p) + log(p))$$

Differentiate with respect to p:

$$\frac{\partial}{\partial p}(Log(L(p))) = \sum_{i_1}^{n}(\frac{1}{p} - \frac{X_i - 1}{1 - p})$$

Set equals 0:

$$0 = \sum_{i_1}^{n}(\frac{1}{p} - \frac{X_i - 1}{1 - p})$$

rearrange:

$$\sum_{i_1}^{n}\frac{X_i - 1}{1 - p} = \sum_{i_1}^{n}\frac{1}{p}$$

Divide by n:

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

Rearrgange:

$$\overline{X} - 1 = \frac{1 - p}{p}$$
$$\overline{X} - 1 = \frac{1}{p} - 1$$
$$\overline{X} = \frac{1}{p}$$
$$\hat{p}_M LE = \frac{1}{\overline{X}}$$

Therefore, they are the same.

### 2.

We know:

$$E[\overline{X}] = \frac{1}{p}$$

and

$$E[\hat{p}] = E[\frac{1}{\overline{X}}]$$

so,

$$E[\frac{1}{\overline{X}}] \geq \frac{1}{E[\overline{X}]} = \frac{1}{\frac{1}{p}} = p$$

Therefore, it is an overestimation as:

$$E[\hat{p}] \geq p$$

### 3.

```
#Define the data given
thefts_vector <- c(1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 
# Find where thefts occurred
theft_days <- which(thefts_vector == 1)
# Compute waiting time between each thefts
waiting_time <- diff(theft_days)
#Mean of waiting times
mean_X <- mean(waiting_time)
#calculate parameters
p_hat <- 1 / mean_X
#print
cat("The Methods of Moments Estimator and the Maximum Likelihood Estimator for :", p_hat, "\n")
```

```
## The Methods of Moments Estimator and the Maximum Likelihood Estimator for : 0.4146341
```

## Question C

**1.**

**2.**

**3.**

**4.**

**5.**

Typically, solutions to an exercise contain the following components:

Some text explaining how you approach the task…

Theoretical calculations (if needed), including assumptions and rationales

```r
# definitions of functions
# commented R commands
```

Figures (if applicable)

Some text explaining what has been achieved, interpretations, and answers to the questions in the description of the task.