CS 6313.001

Statistical Methods for Data Science

By

Prof. Min Chen

MINI PROJECT-1

DUO GROUP-41

1.Amit Kumar	AXK210047
2.Vedant Paresh Shah	VXS200021

Contribution of each group member:

Both worked together and finished the questions as instructed. First went through all the details required, followed the class Note and textbook, learned R then wrote down the scripts. Both of us worked efficiently to complete the required project and finished it on time.

Question #1:

(a.) Analytically compute the probability that the lifetime of the satellite exceeds 15 years using Density Function.

Let XA = lifetime of block A

XB = lifetime of block B

T = lifetime of the satellite

P (T > 15) = 1 - P (T <= 15)

= 1 - F (T <= 15)

$$= 1 - \int_{0}^{15} (fT(t)) \qquad \text{\# Here fT(t) is CDF}$$

$$= 1 - \int_{0}^{15} (0.2exp\text{^}(-0.1t) - 0.2exp\text{^}(-0.2t))\, dt$$

$$= 1 - \int_{0}^{15} [\ 0.2(((exp(-0.1t) / (-\ 0.1)) - ((exp(-0.2t) / (-\ 0.2))]\ dt$$

$$= 1 - \int_{0}^{15} [-2 * \exp^{(- 0.1 t)} + \exp^{(- 0.2 t)}] \, dt$$

$$= 1 - [((-2*\exp^{(-0.1*15)}) + \exp^{(-0.2*15)} - (-2*\exp^{(-0.1*0)} + \exp^{(-0.2*0)}))]$$

$$= 1 - [-2*\exp^{(-1.5)} + \exp^{(- 3)} + 2*\exp^{(0)} - \exp^{(0)}]$$

$$= 1 - [\exp^{(-3)} - 2*\exp^{(-1.5)} + 1]$$

$$= 1 - 0.049787 + 0.446260 - 1$$

$$\boxed{\mathbf{P (T > 15) = 0.3964}}$$

The probability density function of T where T > 15 years is 0.3964 .

(b.) Using Monte Carlo Simulation compute E[T] and P(T>15)

I) Simulate a Draw of XA, XB, and T

For exponential distribution of XA and XB we can use rexp(n, rate) function where n is the number of observations and rate is the value of lambda(λ) or vector of rates.

```
```{r}
#Question 1(b):1
#Simulate a Draw of XA, XB, and T

SLt <- max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10))
print(SLt)

```
```

Output:

SLt	11.74429

The Lifetime of Satellite(T) after 1 draw of XA and XB is T = 11.74 Years.

II) Repeat Above Step 10,000 Times

Using the same code as above in replicate() function where we will run line of code for 10,000 times we will get 10,000 values of T.

```{r}
#Question 1(b):2
#Repeat Above Step 10,000 Times

SLt10K <- replicate(10000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
print(SLt10K)

```

Output:

SLt10k	Num [1:10000] 1.9044500 19.0761546 10.6661079 1.0545865 1.9179961 5.8887402 ………….

The Lifetime of Satellite(T) after 10,000 draws of XA and XB will be stored in SLt10k.

III) Plot Histogram of T with 10,000 draws and superimpose a curve of density function

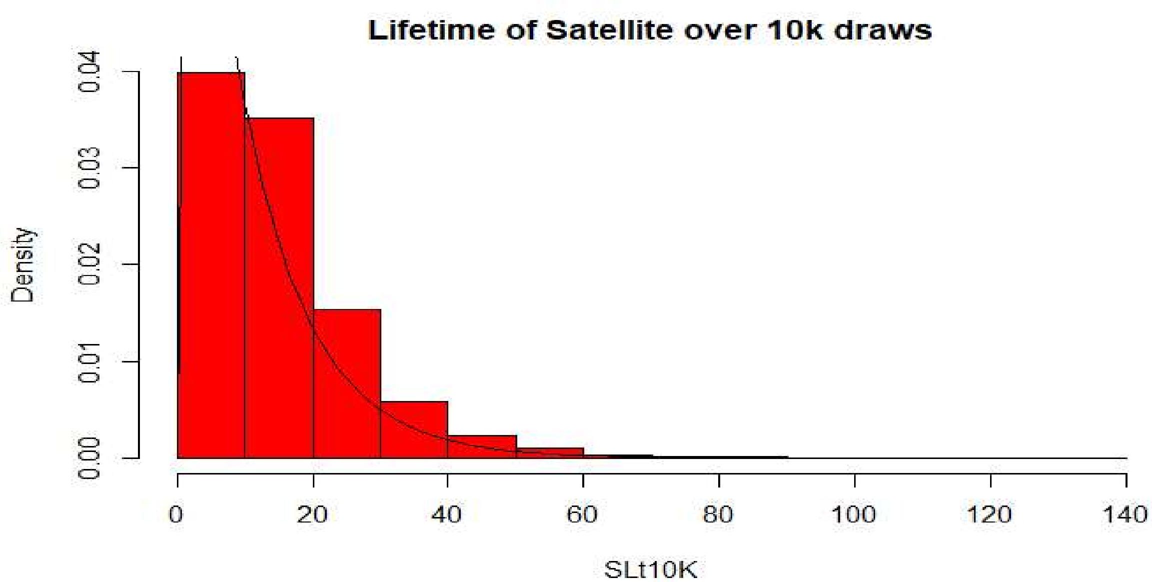
Using above values of T, we can draw a histogram of T and superimpose the curve by using the density function dexp(n,rate) where n is the number of observations and rate is the value of lambda(λ) or vector of rates.

```
```{r}
#Question 1(b):3
#Plot Histogram of T with 10,000 draws and superimpose a curve of density function

SLt10K <- replicate(10000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
hist(SLt10K,prob=TRUE,col= "red",main="Lifetime of Satellite over 10k draws")
curve(dexp(x,0.1),add=TRUE)

```
```

Output:



IV) Find E[T] using above 10,000 values of T

Using the 10,000 values of T we can find the mean of T using the mean() function.

```
```{r}
#Question 1(b):4
#Find E[T] using above 10,000 values of T

SLt10K <- replicate(10000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
Avg <- mean(SLt10K)
print(Avg)

```
```

Output:

Avg	14.91229

The Mean of the Lifetime of Satellite using the Monte Carlo Simulation is E[SLt10k]=14.91 Years and given Mean of Lifetime of Satellite is E[T]=15 Years both values are Comparable to each other. The difference in both values is very small.

V) Estimating that the Probability of T where T lasts more than 15 years

Here we will calculate the Probability of T where T>15.

```{r}
#Question 1(b):5
#Estimating that the Probability of T where T lasts more than 15 years

SLt10K <- replicate(10000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
Tn <- SLt10K > 15
TotalSum <- sum(Tn)
ProbT <- TotalSum / 10000
ProbT
```

Output:

ProbT	0.3922

The Probability of the Lifetime of the Satellite using the Monte Carlo Simulation where the Lifetime of Satellite (T) is greater than 15 in 10000 Draws is ProbT[SLt10k] = 0.3922 whereas the Probability we calculated in Question 1(a.) was P(T>15) = 0.3964 which is comparable. The difference in both values is very small.

VI) Estimating E[T] and P(T>15) 4 more times over 10,000 draws

By using the code given in Question 1(b.) we will calculate E[T] and P(T>15) 4 more times over 10,000 draws.

```{r}
#Question 1(b):6
#Estimating E[T] and P(T>15) 4 more times over 10,000 draws

SLt10K <- replicate(10000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
Avg <- mean(SLt10K)
Avg
Tn <- SLt10K > 15
TotalSum <- sum(Tn)
ProbT <- TotalSum / 10000
ProbT

```

Output:

Test #	Mean (E[SLt10k])	Probability (ProbT(SLt10k))
Test -1	14.91229	0.3922
Test -2	14.87832	0.3888
Test -3	15.12651	0.4031
Test -4	15.13226	0.4029
Test -5	15.12285	0.4019

The Value of E[T] has a very minor variation to the value 15. These values are very close to original value given E[T] = 15. Also, the value P(T>15) has minor variation from the original value calculated by us in Question 1(a.) where P(T>15) = 0.3964.

(c.) Estimating E[T] and P(T>15) 5 times over 1,000 and 100,000 draws.

1. Estimating E[T] and P(T>15) Over 1,000 Draws

By using the code given in Question 1(b.) we will calculate E[T] and P(T>15) over 1,000 draws.

```
```{r}
#Question 1(c)
#Estimating E[T] and P(T>15) Over 1,000 Draws

SLt1K <- replicate(1000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
Avg <- mean(SLt1K)
Avg
Tn <- SLt1K > 15
TotalSum <- sum(Tn)
ProbT <- TotalSum / 1000
ProbT

```
```

Output:

Test #	Mean (E[SLt1k])	Probability (ProbT(SLt1k))
Test -1	14.95364	0.415
Test -2	14.94149	0.387
Test -3	14.56947	0.372
Test -4	15.49855	0.41
Test -5	15.05182	0.401

2. Estimating E[T] and P(T>15) Over 100,000 Draws

By using the code given in Question 1(b.) we will calculate E[T] and P(T>15) over 100,000 draws.

```
```{r}
#Question 1(c)
#Estimating E[T] and P(T>15) over 100,000 Draws

SLt100K <- replicate(100000, max(rexp(n = 1, rate = 1/10), rexp(n = 1, rate = 1/10)))
Avg <- mean(SLt100K)
Avg
Tn <- SLt100K > 15
TotalSum <- sum(Tn)
ProbT <- TotalSum / 100000
ProbT

```
```

Output:

Test #	Mean (E[SLt100k])	Probability (ProbT(SLt100k))
Test -1	15.02656	0.39825
Test -2	14.98768	0.39561
Test -3	14.92633	0.39455
Test -4	15.02739	0.397
Test -5	15.01898	0.39647

It is clear from running these Monte Carlo Simulations that as we increase the size of the draws the values of the E[T] and P(T>15) have less variation in them. It can be observed that we get a more accurate value and less variation in value of both E[T] and P(T>15) when we draw 100,000 times than when we draw 1,000 times.

Question #2: Monte Carlo approach estimate the value of π based on 10, 000 replications.

Monte Carlo approach to estimate the value of π

To calculate the value of pi, we will have to find the probability of a point falling inside the circle under the square = Area of square of side 1/Area of circle of radius 1 = pi/4

To estimate the value of pi using the Monte Carlo approach, we will have to generate 10,000 random numbers between 0 and 1 for both x and y coordinates. And then we will have to check if the number falls within the range of the circle or not.

````
```{r}
#Question 2
#Monte Carlo approach to estimate the value of π
counter <- 10000
counter
x <- runif(counter, min = 0, max = 1)
print(x)
y <- runif(counter, min = 0, max = 1)
print(y)

inside.circle <- (x - 0.5)^2 + (y - 0.5)^2 <= 0.5^2
print(inside.circle)

pi.value <- (sum(inside.circle)/counter)*4
print(pi.value)

```
````

Output:

[1] 3.136

Counter	10000
x	[1] 0.8707107916 0.1914565384 0.0962700143 [4] 0.2683930718 0.7195527819 0.2095570192...........
y	[1] 0.570991406 0.160455309 0.954698283 [4] 0.033608761 0.945261253 0.939486394................
Inside.circle	[1] TRUE TRUE FALSE FALSE TRUE FALSE TRUE [8] TRUE TRUE TRUE TRUE TRUE FALSE FALSE ….....
pi.value	[1] 3.136

The value of pi generated using Monte Carlo simulation over 10000 iterations is 3.136 which is close to 3.1415 i.e., the true value of pi. As we increase the counter, the value of the calculated pi will be closer to the actual value.