CS 6313 STATISTICAL METHODS FOR DATA SCIENCE

Mini Project 1

SEPTEMBER 10, 2021

AFRAA ALSHAMMARI
PRAVALIKA DOMAL

Contribution of each member:

Both worked together to finish and submit this mini project. Starting with learning R and conducted the needed statistics scripts to solve Q1 and Q2. Afraa did the documentation and report the experiments and Pravalika worked on analytical solution scripts for finding accuracy of scripts.

1. (10 points) Consider Exercise 4.11 from the textbook. In this exercise, let X_A be the lifetime of block A, X_B be the lifetime of block B, and T be the lifetime of the satellite. The lifetimes are in years. It is given that X_A and X_B follow independent exponential distributions with mean 10 years. One can follow the solution of Exercise 4.6 to show that the probability density function of T is

$$f_T(t) = \begin{cases} 0.2\exp(-0.1t) - 0.2\exp(-0.2t), & 0 \le t < \infty, \\ 0, & \text{otherwise,} \end{cases}$$

and $E(T) = 15$ years.

Section A:

λ= 1/µ

λ= 1/10 =**0.1** years^-1

where t is the random variable representing the lifetime of the satellite.

According to the Probability Distribution Function (PDF) which states that Fx(x)=P(X<+x)

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

= 1-F (t<=15)

$$=1 - \int_0^{15} (0.2\exp^{-0.1*t} - 0.2\exp^{-0.2*t})$$

$$=1 - \left(0.2 ((\exp^{-0.1*t} / -0.1) - (\exp^{-0.2*t} / -0.2))\right|_0^{15}$$

$$= 1 - (-2\exp^{-0.1*t} + \exp^{-0.2*t})\Big|_0^{15}$$

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

$$= 1 - ((-2\exp^{-1.5} + \exp^{-3}) - (-2\exp^{0} + \exp^{0}))$$ # any value to the power 0=1

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

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

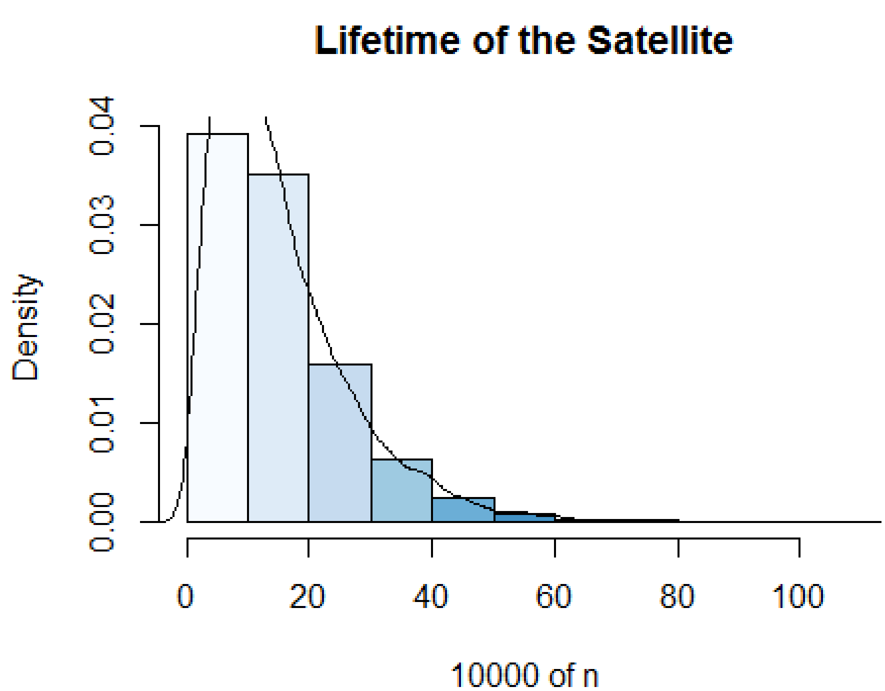
=1 – (0.04978706836 - 0.44626032029 + 1)

=1 - 0.60352674807

=**0.39647325192**

Section B:

```
#Process 1


#(i) Simulate one draw of the block lifetimes XA,XB and the satellite lifetime T.
(i) print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 7.250919

#(ii) Repeat the previous step 10,000 times and save these draws.
(ii) V=replicate(10000,max(rexp(n=1,0.1),rexp(n=1,0.1)))

#(iii) Make a histogram of the of T using 'hist'  the density function.Then use curve function to dr
aw the density.
hist(V,probability = TRUE, col =blues9, xlab ="10000 of n", main = "Lifetime of the Satellite")
x<-V
lines(density(x))
```



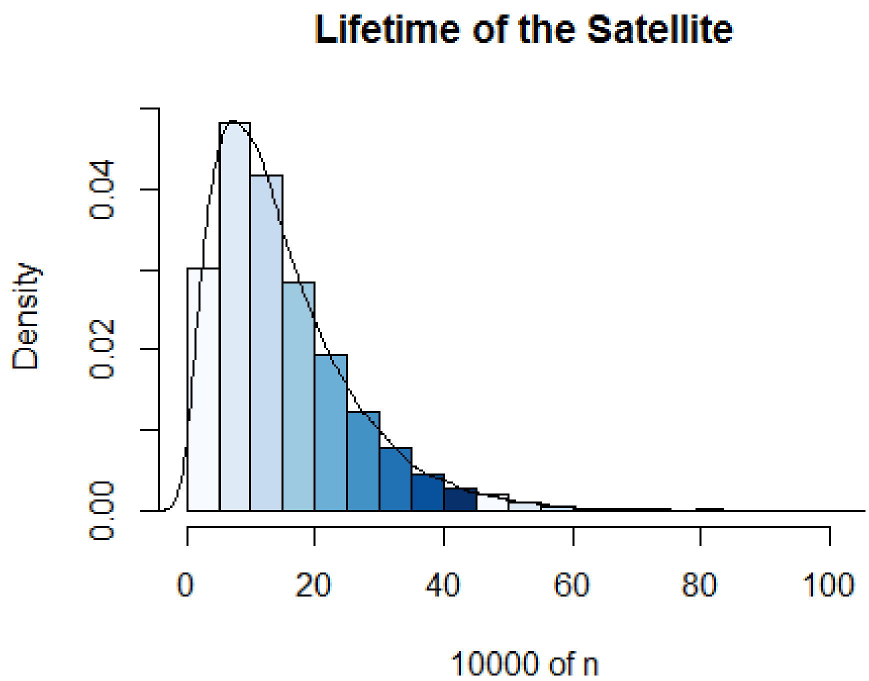
```r
#(iv)Use the saved draws to estimate E(T).
mean(V)

## [1] 15.18305
```

In this part of the question, the value of the given E(t) from the question which equals to 15, and the obtained from the Monti Carlo approach that is equal to 15.18305 has an insignificant difference between them indicating the accuracy of the calculations.

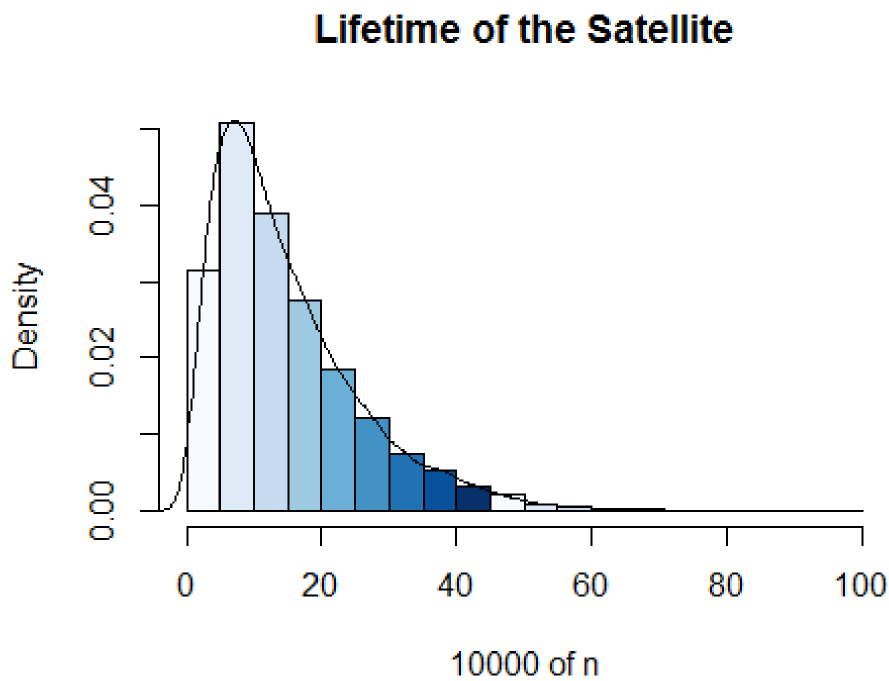
```r
#(v)Use the saved draws to estimate the probability
# Here P(t)>=15 is needed, therefor epex will be used and n will start from 15
1-pexp(15,1/mean(V))

## [1] 0.3723415
```

In this part of the question a comparation between the value obtained in part (a) which equals to 0.39647325192, and the obtained from the Monti Carlo approach that is equal to 0.3723415. The difference between these values is less than 0.2 which according to the statistically significant is and indictor for a rejection of the null hypothesis is in order. Therefore, the hypothesis of the lifetime of the satellite exceeds 15 is accepted.

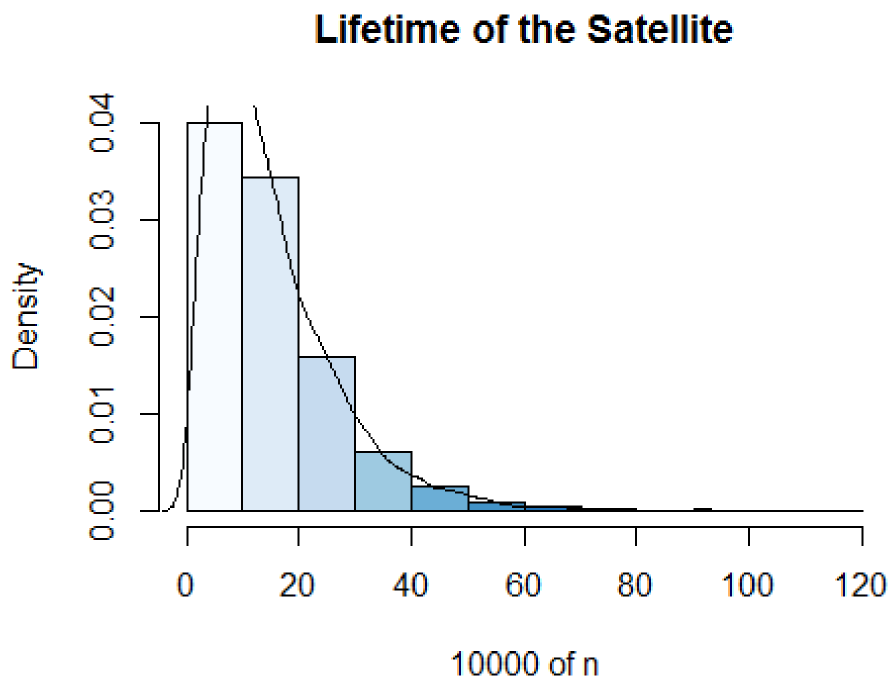
```r
#(vi)Repeat the above process of obtaining E(T) and the probability 4 more time
# Process 2
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 22.34487

(ii)V=replicate(10000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="10000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```



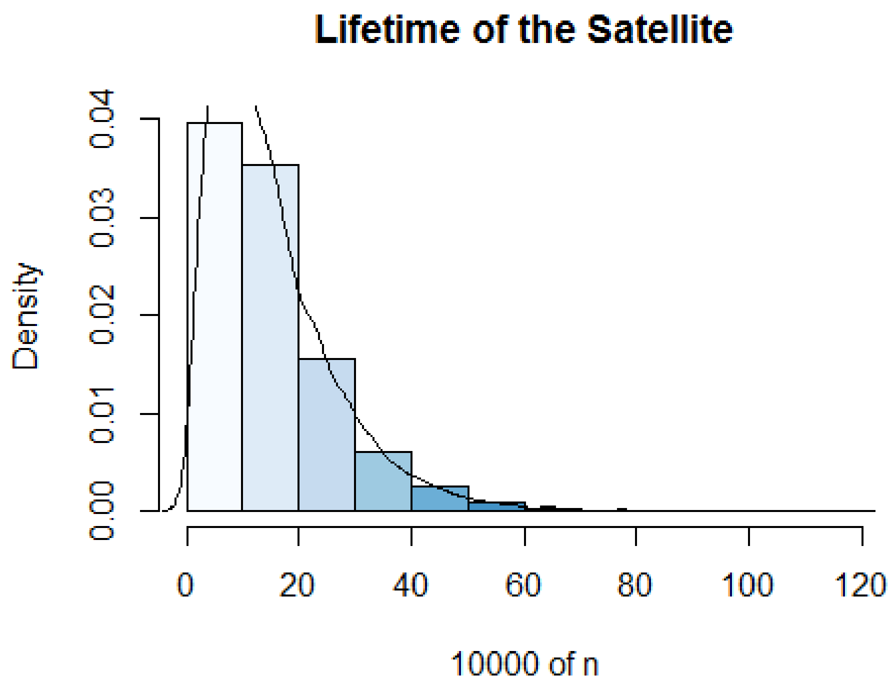

```
mean(V)

## [1] 15.08155

1-pexp(15,1/mean(V))

## [1] 0.369874



# Process 3
print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 2.19438

V=replicate(10000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
hist(V,probability = TRUE, col =blues9, xlab ="10000 of n", main = "Lifetime of the Satellite")
x<-V
lines(density(x))
```




```
(iv) mean (V)

## [1] 14.97882

(v) 1-pexp(15,1/mean(V))

## [1] 0.3673597



# Process 4
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 7.721175

(ii)V=replicate(10000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="10000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```




```
(iv)mean(V)

## [1] 15.0691

(v)1-pexp(15,1/mean(V))

## [1] 0.3695704



# Process 5
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 17.74189

(ii)V=replicate(10000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="10000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```




```
(iv)mean(V)

## [1] 14.92103

(v)1-pexp(15,1/mean(V))

## [1] 0.3659375
```

Results: To compare the result of the 5 process Table 1 was created.

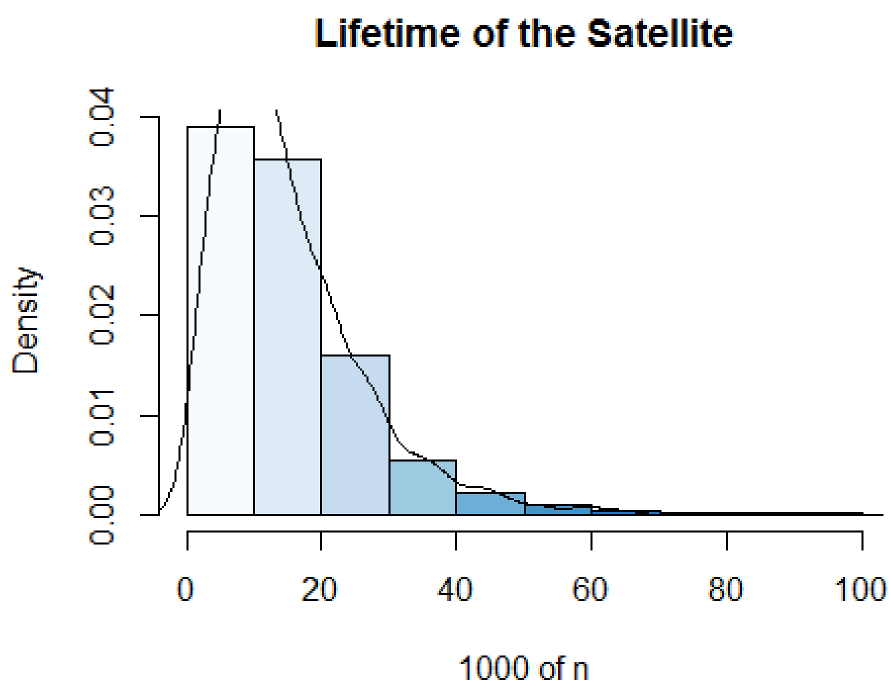
#	E(T)	P (>15)
1	15.01108	0.3681512
2	14.71731	0.3608807
3	14.9477	0.3665945
4	14.79901	0.3629169
5	14.78544	0.3625794

Table 1. Result of the 5 Processes with n=10000

The result of both E (T) and P (>15) that are generated are in consistence with given value of E (T) and the analytically calculated value of P (>15). In addition, All the probabilities values are supporting the statement in part (v) of the question in accepting the hypothesis of the question stating that the lifetime of the satellite exceeds 15 and rejecting the null hypothesis stating otherwise (<=15).

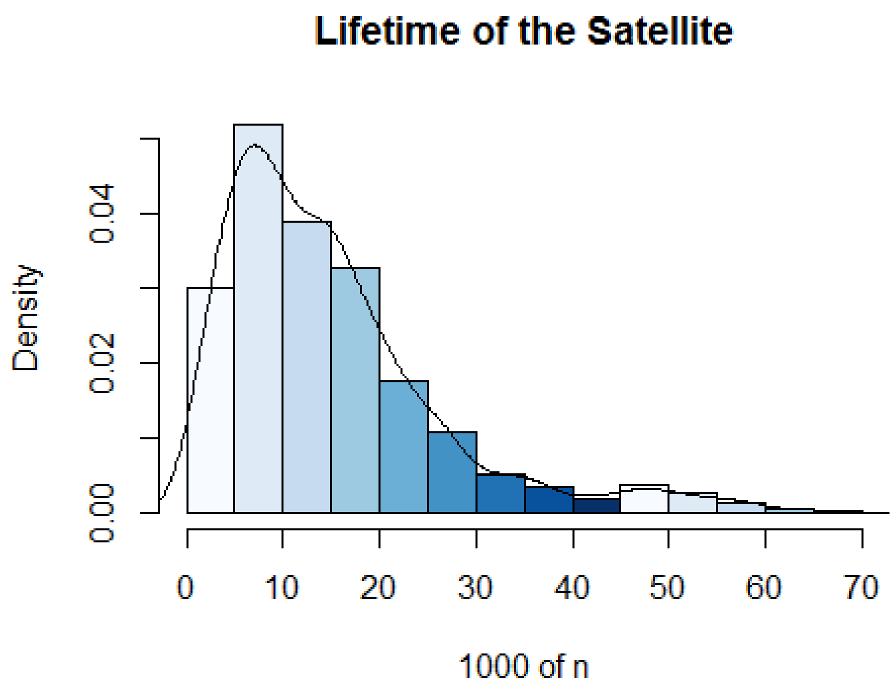
Section C:

```
# N= 1000
# Process 1
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 11.44768

(ii)V=replicate(1000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="1000 of n", main = "Lifetime of the Satellite")
     x<-V
    lines(density(x))
```



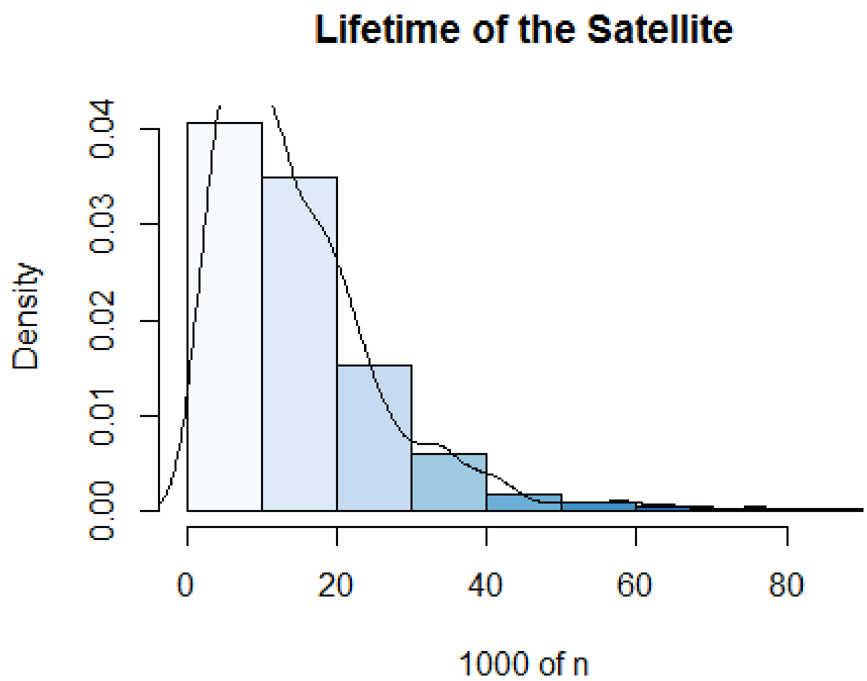

```
(iv)mean(V)

## [1] 15.18499

(v)1-pexp(15,1/mean(V))

## [1] 0.3723886



# Process 2
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 10.3407
```

```
(ii)V=replicate(1000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="1000 of n", main = "Lifetime of the Satellite")
  x<-V
  lines(density(x))
```

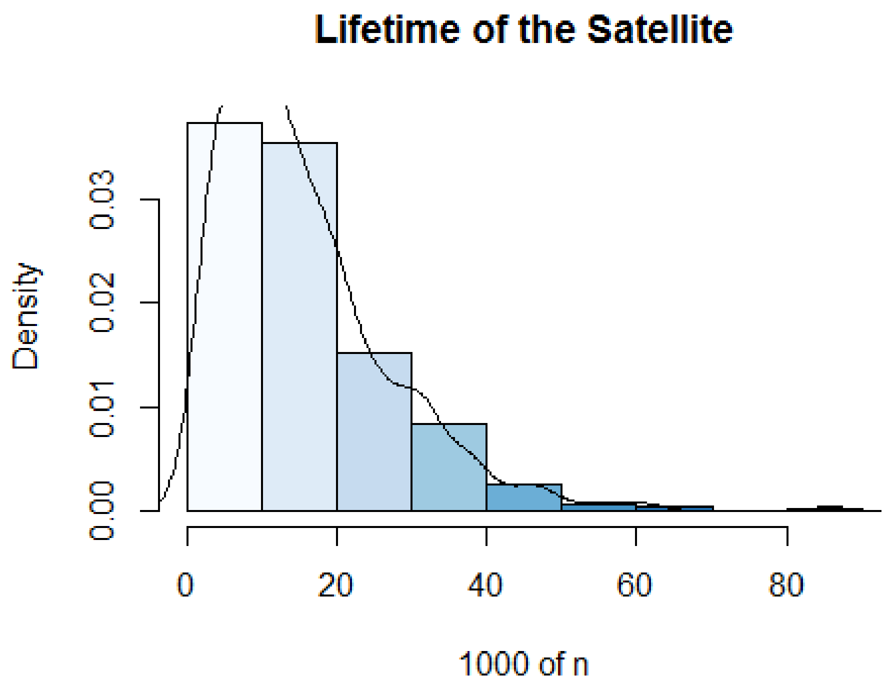



```
(iv)mean(V)

## [1] 15.02864

(v)1-pexp(15,1/mean(V))

## [1] 0.3685811
```

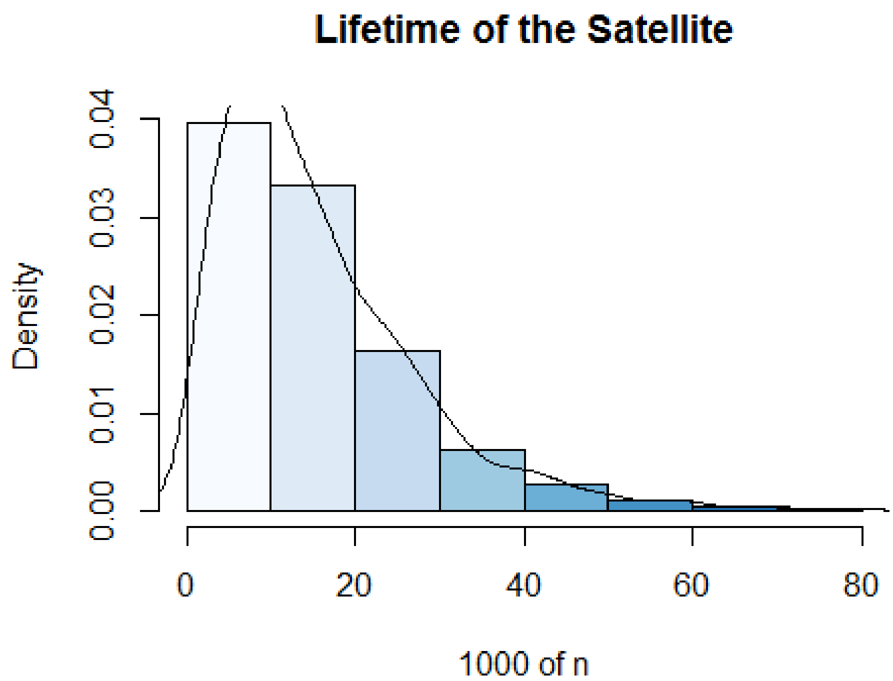
```
# Process 3
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 12.46927

(ii)V=replicate(1000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="1000 of n", main = "Lifetime of the Satellite")
  x<-V
  lines(density(x))
```



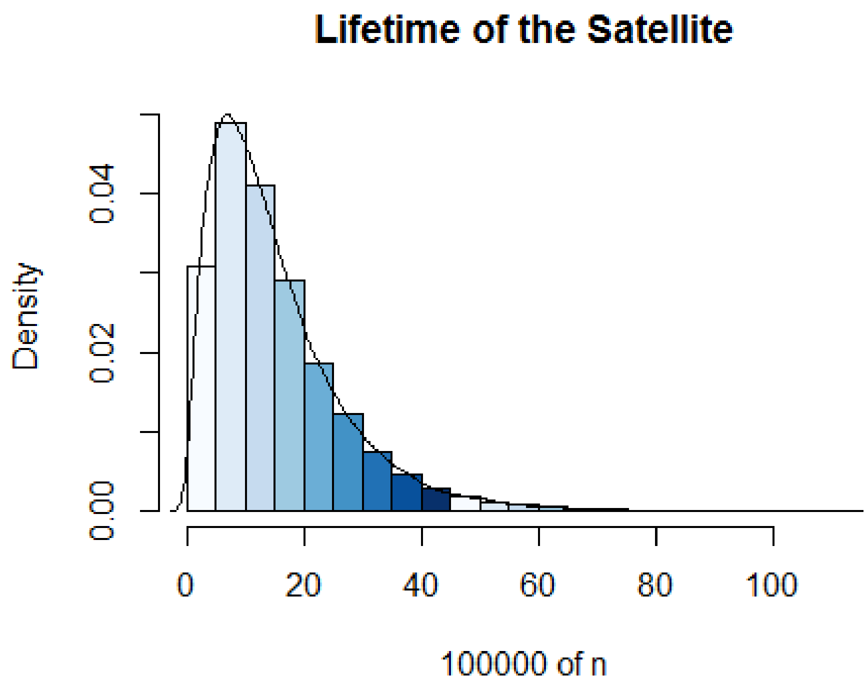

```
(iv)mean(V)

## [1] 14.77815

(v)1-pexp(15,1/mean(V))

## [1] 0.362398



# Process 4
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 24.40523

(ii)V=replicate(1000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="1000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```



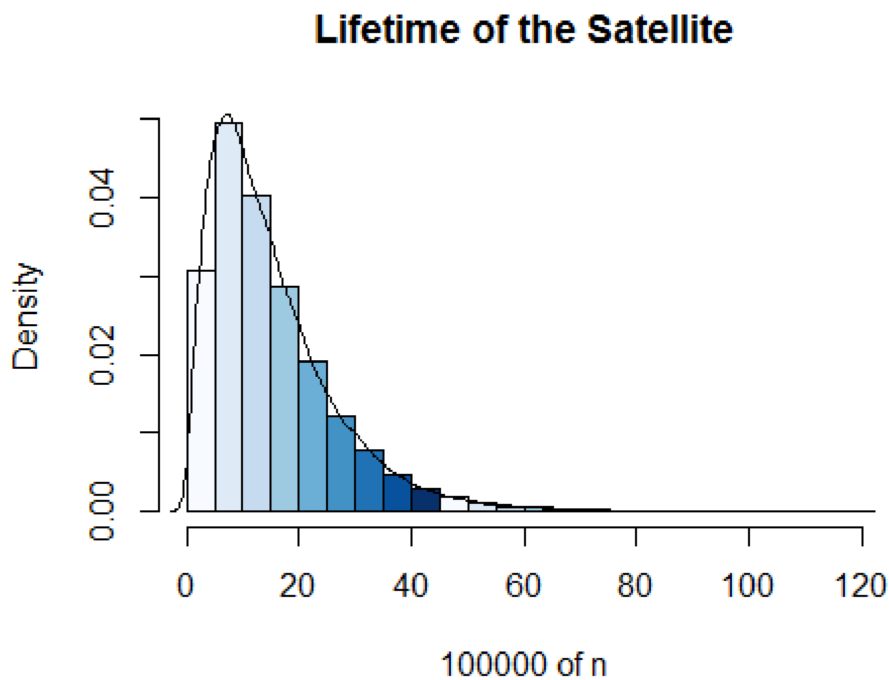

```
(iv)mean(V)

## [1] 15.55194

(v)1-pexp(15,1/mean(V))

## [1] 0.38117



# Process 5
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 9.389427

(ii)V=replicate(1000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="1000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```



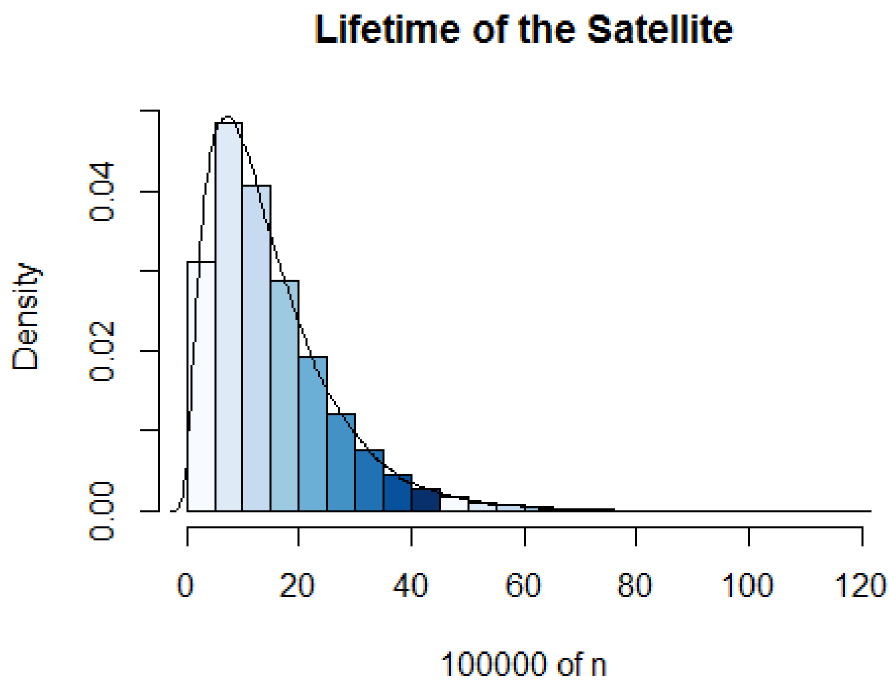

```
(iv)mean(V)

## [1] 15.47992

(v)1-pexp(15,1/mean(V))

## [1] 0.3794633



(D)N=100000

# Process 1
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 22.28447

(ii)V=replicate(100000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="100000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```



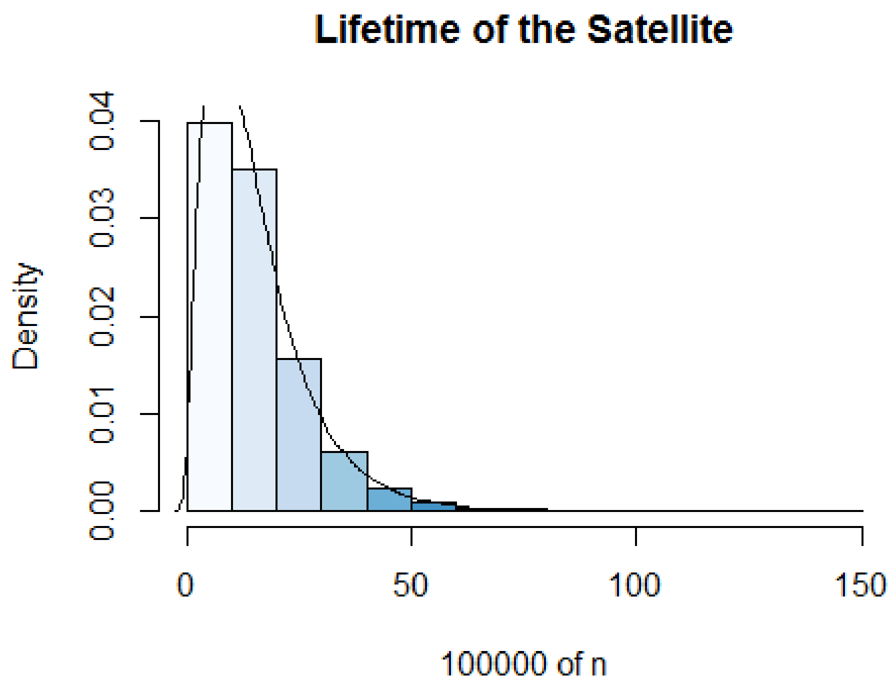

```
(iv)mean(V)

## [1] 14.98837

(v)1-pexp(15,1/mean(V))

## [1] 0.3675942



# Process 2
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 10.24337

(ii)V=replicate(100000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="100000 of n", main = "Lifetime of the Satellite")
    x<-V
   lines(density(x))
```



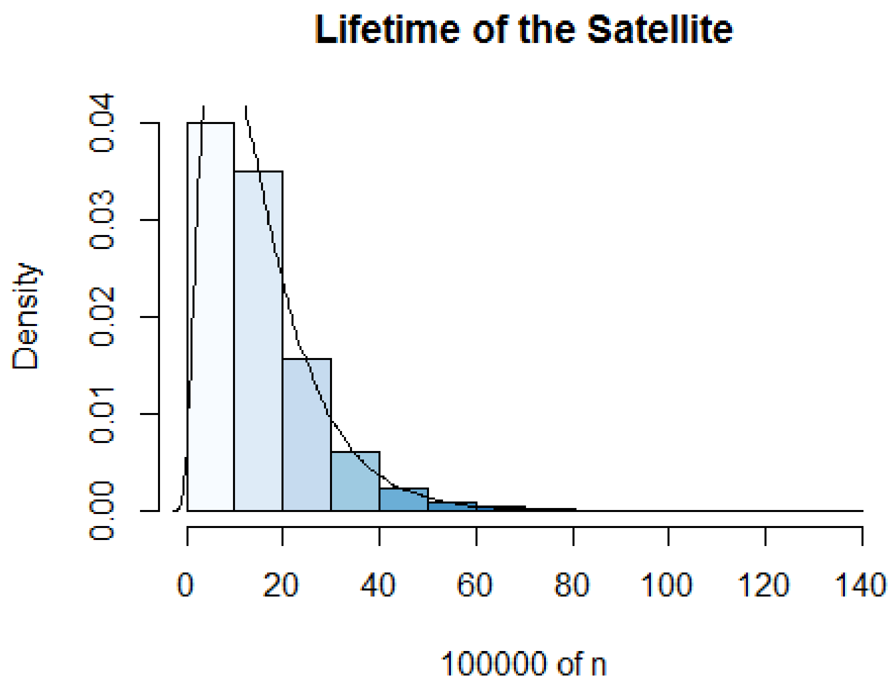

```
(iv)mean(V)

## [1] 14.97791

(v)1-pexp(15,1/mean(V))

## [1] 0.3673373



# Process 3
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 25.01517

(ii)V=replicate(100000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="100000 of n", main = "Lifetime of the Satellite")
     x<-V
    lines(density(x))
```




```
(iv)mean(V)

## [1] 15.03014

(v)1-pexp(15,1/mean(V))

## [1] 0.3686179



# Process 4
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 15.12155

(ii)V=replicate(100000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="100000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```




```
(iv)mean(V)

## [1] 15.00976

(v)1-pexp(15,1/mean(V))

## [1] 0.3681186

# Process 5
(i)print (max(rexp(n=1,0.1),rexp(n=1,0.1)))

## [1] 7.485639

(ii)V=replicate(100000,max(rexp(n=1,0.1),rexp(n=1,0.1)))
(iii)hist(V,probability = TRUE, col =blues9, xlab ="100000 of n", main = "Lifetime of the Satellite")
    x<-V
    lines(density(x))
```




```
(iv)mean(V)

## [1] 14.97396

(v)1-pexp(15,1/mean(V))
```

```
## [1] 0.36
```

Results:

#	E(T)	P (>15)
1	15.01108	0.3681512
2	14.71731	0.3608807
3	14.9477	0.3665945
4	14.79901	0.3629169
5	14.78544	0.3625794

Table 1. Result of the 5 Processes with n=10000

#	E(T)	P (>15)
1	14.67756	0.3598859
2	14.73066	0.3612141

3	14.42983	0.3536268
4	14.87843	0.3648858
5	15.10984	0.3705635

Table 2. Result of the 5 Processes with n=1000

#	E(T)	P (>15)
1	14.982	0.3674377
2	14.99936	0.3678638
3	15.01062	0.3681398
4	15.01571	0.3682646
5	15.00791	0.3680733

Table 3. Result of the 5 Processes with n=100000

The results in the above 3 tables, the relation between the number of the sample (n) and the reduction of the variance and accuracy of the results for both E (T) and P (>15) are in a positive correlation. This is visible by comparing the result of Table 2 with both tables 1 and 3 respectively. Concluding the findings statistically demonstrate the Central Limit theorem as the variance get reduced while the number of elements in the sample increases and the distribution is towards a normal distribution.

2. (10 points) Use a Monte Carlo approach estimate the value of π based on 10,000 replications. [**Ignorable hint**: First, get a relation between π and the probability that a randomly selected point in a unit square with coordinates — (0, 0), (0, 1), (1, 0), and (1, 1) — falls in a circle with center (0.5, 0.5) inscribed in the square. Then, estimate this probability, and go from there.]

(i) P(E) is probability that we need to find out. Prove that P(E) = π/4.

P(E) = area of the circle/area of the square = π * ((0.5)^2) / (1 ^ 2) = π * 25/ 100 = π / 4

(ii) Monte Carlo simulation can be used to estimate the probability by randomly creating a large number of points (say, N) inside the square and calculating the proportion the number of points that lie within the circle Use N = 10,000

```
n <- 10000

x <- runif(n,min=0,max=1)                  # generates x coordinates

y <- runif(n,min=0,max=1)                  # generates y coordinates

ci <- (x-0.5) ^2 + (y-0.5)^2 <= 0.5^2      # find points the fall in circle

val.pi<-(sum (ci)/n)*4                     # estimated value of pi

val.pi
```

```
##[1] 3.1476
```

3.1476 is the generated value of pi, which is near to 3.14159, which is the actual value of pi used in mathematical computations. We can get more accurate value by larger value of N.