CS 6313.002 - Statistical Methods For Data Science

MINI PROJECT - #1

Group - 4

Group members :

- Sirisha Satish
- Sravan Kumar Guduru

Contribution :

- We both had discussed the approaches to both the questions and have divided the project equally.
- Sirisha completed half of the questions and the other half was completed by Sravan.
- After simulating the code in R, we analyzed every question and generated the project report together.
- Both of us have completed our work efficiently.

1. **a)**

Let $\mathbf{X_A} \rightarrow$ Lifetime of block A
Let $\mathbf{X_B} \rightarrow$ Lifetime of block B

From the given information, we need to the probability of the satellite if its lifetime exceeds 15 years → T>15

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

= 1 - $\int_0^{15} f(t)$ (i.e, over the values 0 to 15)

Where, $f(t) = 0.2\, exp(-\ 0.1t)\ -\ 0.2\, exp(-\ 0.2t)$

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

= 1 – $[(-\ 0.2\, exp(-\ 0.1t))/0.1\ -\ (-\ 0.2\, exp(-\ 0.2t)/0.2)]_0^{15}$

= 1 – $[-\ 2\, exp(-\ 0.1t)\ +\ 1\, exp(-\ 0.2t)]_0^{15}$

= 1 –
$[(-\ 2\, exp(-\ 0.1 * 15)\ +\ exp(-\ 0.2 * 15))\ -\ (-\ 2\, exp(-\ 0.1 * 0)\ +\ exp(-\ 0.2 * 0))\ ]$

$= 1 - [- 2\, exp(- 1.5) \;+\; exp(- 3) \;+\; 2\, exp(0) \;-\; 1\, exp(0)]$

$= 1 - [exp(- 3) \;-\; 2\, exp(- 1.5) \;+\; 1]$

$= 1 - [0.04978706 \;-\; 2\,(0.22313016) \;+\; 1]$

$= 1 - [1.04978706 \;-\; 2\,(0.22313016)]$

$= 1 - 0.603526$

$= 0.39647326$

b) i) Simulating a Draw of X_A, X_B and T

```
#1. b) Creating a function to compute the #probability that the lifetime of a
#satellite exceeds 15 years


calc_lifetime <- function(t) {
  return(0.2*exp(-0.1*t)-0.2*exp(-0.2*t))
}
#when t = 30
calc_lifetime(30)
```

```
> #1. b) Creating a function to compute the
> #probability that the lifetime of a satellite
> #exceeds 15 years
>
> calc_lifetime <- function(t) {
+   return(0.2*exp(-0.1*t)-0.2*exp(-0.2*t))
+ }
> #when t = 30
> calc_lifetime(30)
[1] 0.009461663
```

ii) repeating steps 10000 times

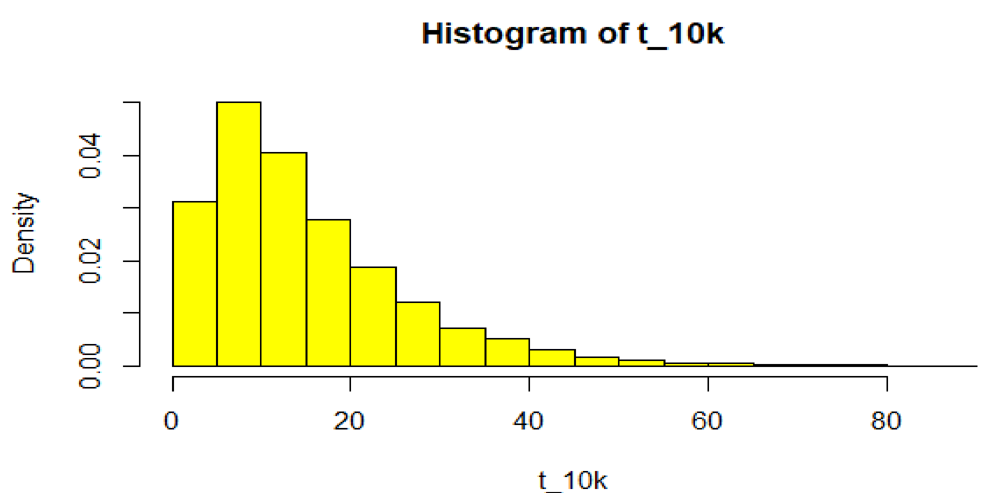
```
#b)ii)repeating the steps 10,000 times by simulating 10,000 draws
#from the distribution of T
#using replicate function
t_10k = replicate(10000,max(rexp(n=1, rate =0.1),rexp(n=1, rate =0.1)))
t_10k
```

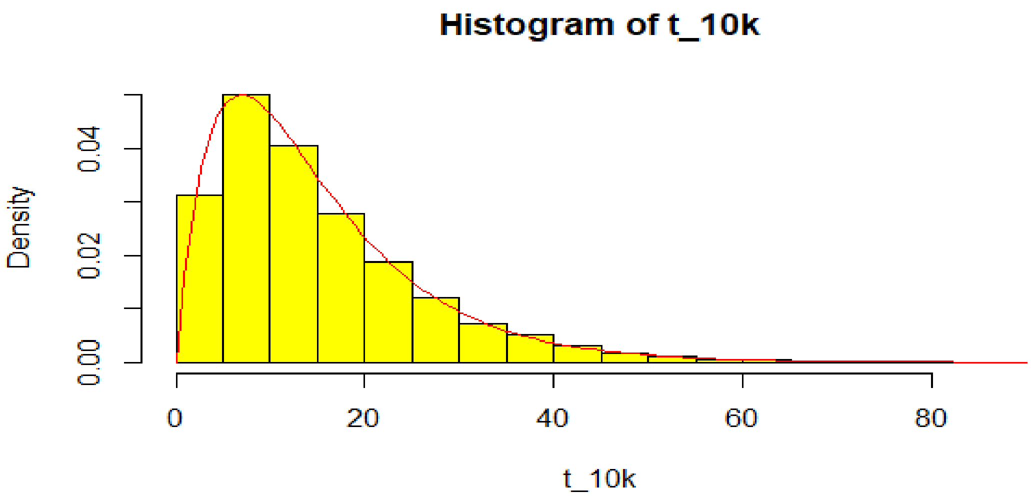
```
[442] 23.7719592 20.8512800 21.8356419 18.1088314 14.6890447 22.5011758  6.6569407
[449] 12.0731970 61.0587119  7.7759029 29.0135719  8.0959493  3.0153370  0.5812671
[456] 17.7363379 10.1696690 11.4770965 12.1645843  3.5317464  9.1151040  2.1152801
[463] 11.4885169  2.7030776 17.1893167  6.7700307  1.5207690 18.3657145 14.9517604
[470]  6.8002154 27.1177191  4.4526984  8.3801720 25.3422651 12.5486624 23.2494782
[477]  3.9434232 10.7830788  3.2139470 11.1751008  5.7765310  3.1249962  8.2043148
[484]  2.6537260  6.3451973  9.0653712  8.9100348 11.5082224  8.6682050 13.4962128
[491] 22.0756748 10.4673367 17.7842286  6.6863077  5.7317595  4.1660835  3.1802046
[498]  2.4097818 22.8046851 11.3958146 23.8809136  5.3762718  2.3609511  8.0917170
[505] 21.5459774  7.8532771 13.5902628 18.7965688  5.3143386  9.1302186 40.6172416
[512]  7.6026915  4.0342955 42.8207226 20.0944368 10.3409016 10.3333661  1.5360070
[519] 23.1125402 12.0699281  8.3665409 10.2815434 15.2880525  6.1968220  9.9736586
[526]  0.5317402  8.2091810  1.0261189 17.1092998 16.3288208 13.2367314  4.7669686
[533] 10.5176253  5.1768614 34.9261710 25.9796279  8.0005702  5.1315336 18.0238300
[540]  0.7848313 67.5903581 20.7980610 19.0751494 20.1656432 15.6213973 12.9603259
```

iii) Plotting a histogram of the 10000 draws using hist function.

```
#b)iii) Plotting a histogram to superimpose the density function mentioned above
#using hist (without curve)
hist(t_10k, col="yellow",prob=TRUE)

#superimposing it with a 'curve'
curve(calc_lifetime, col='red',add=TRUE)
```





iv) Calculating the expected value of T i.e, the mean by mean()-

```
#b)iv) estimating the expected value(mean)
mean(t_10k)
```

```
> mean(t_10k)
[1] 14.94749
>
```

The analytical calculation of probability density function gives expected values as 15 which is approximately close to the Monte Carlo simulation i.e., 14.94749.

v) Estimating the probability that the satellite lasts more than 15 years.

```
#b)v)estimating the probability that the satellite lasts more than 15 years.
1-pexp(15, rate=1/mean(t_10k))
```

```
> 1-pexp(15, rate=1/mean(t_10k))
[1] 0.3665893
>
```

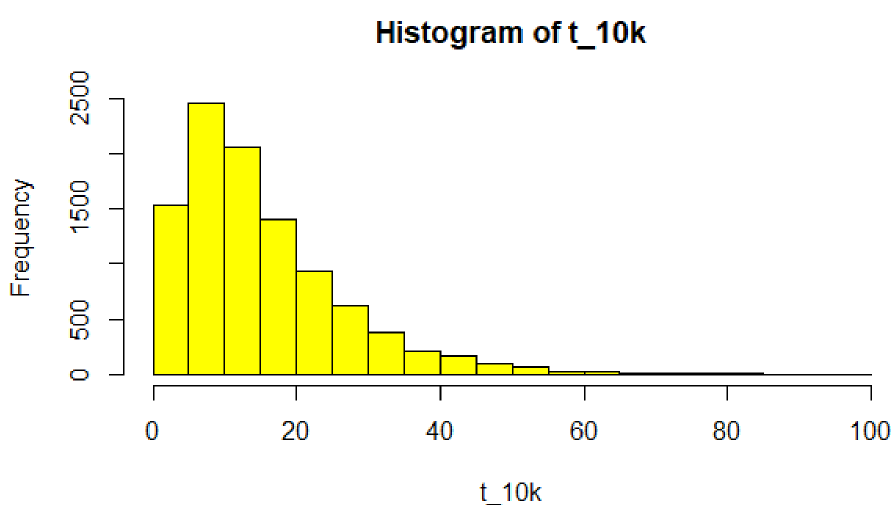
The probability calculated in (a) is 0.39647326 and the probability calculated here is 0.3665893 which is slightly different because the sample size is 1000 random variables and the differences in the mean.

vi) Estimating of the probability four more times

Test 1

```
#b)vi) estimating the probability 4 more times
#TEST 1
t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_10k,col="yellow")
1-pexp(15,rate=1/mean(t_10k))
mean(t_10k)
```

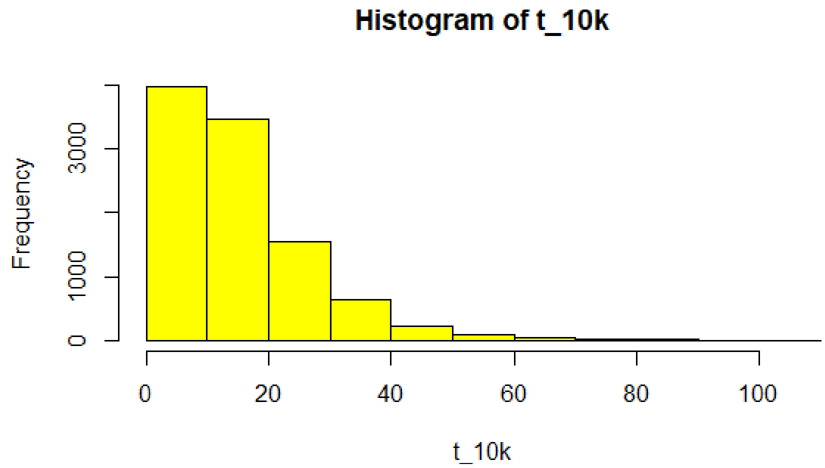
```
> #b)vi) estimating the probability 4 more times
> #TEST 1
> t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_10k,col="yellow")
> 1-pexp(15,rate=1/mean(t_10k))
[1] 0.3710054
> mean(t_10k)
[1] 15.128
```



Test 2

```
#TEST 2
t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_10k,col="yellow")
1-pexp(15,rate=1/mean(t_10k))
mean(t_10k)
```

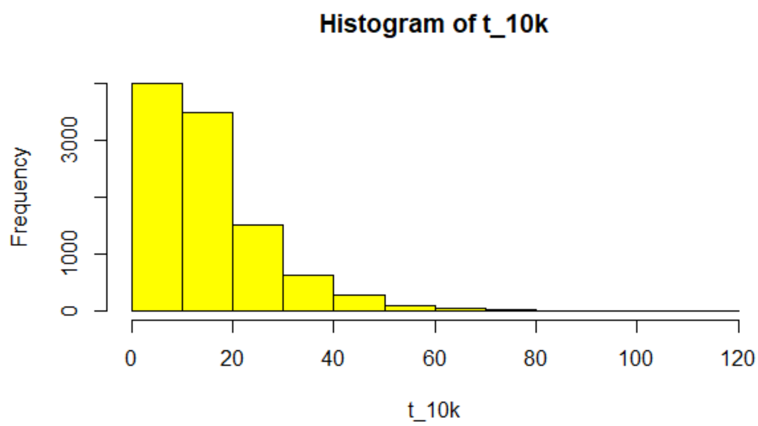
```
> #TEST 2
> t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_10k,col="yellow")
> 1-pexp(15,rate=1/mean(t_10k))
[1] 0.3704014
> mean(t_10k)
[1] 15.10319
```



Test 3

```
#TEST3
t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_10k,col="yellow")
1-pexp(15,rate=1/mean(t_10k))
mean(t_10k)
```

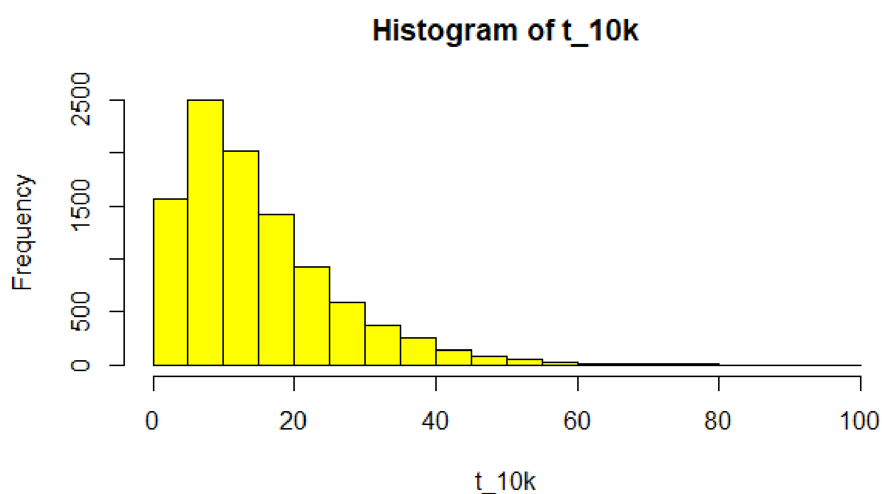
```
> #TEST3
> t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_10k,col="yellow")
> 1-pexp(15,rate=1/mean(t_10k))
[1] 0.3706542
> mean(t_10k)
[1] 15.11357
>
```



Test 4

```r
#TEST 4
t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_10k,col="yellow")
1-pexp(15,rate=1/mean(t_10k))
mean(t_10k)
```

```
#TEST 4
t_10k = replicate(10000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_10k,col="yellow")
1-pexp(15,rate=1/mean(t_10k))
] 0.3669412
mean(t_10k)
] 14.96179
```



From the above simulations we can conclude that the value mean E(T) is closer to 15 and the probability P(T>15) is approximately equal to the value we got by calculating analytically in question 1 part A.

c) Repeat part (vi) 5 times with 1,000 and 100,000 draws

Sample size = 1000

```
t_1k = replicate(1000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_1k,col="yellow")
1-pexp(15,rate=1/mean(t_1k))
mean(t_1k)
```

1st:

```
> t_1k = replicate(1000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_1k,col="yellow")
> 1-pexp(15,rate=1/mean(t_1k))
[1] 0.3601519
> mean(t_1k)
[1] 14.68818
>
```

2nd:

```
> t_1k = replicate(1000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_1k,col="yellow")
> 1-pexp(15,rate=1/mean(t_1k))
[1] 0.3635296
> mean(t_1k)
[1] 14.82368
>
```

3rd:

```
> t_1k = replicate(1000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_1k,col="yellow")
> 1-pexp(15,rate=1/mean(t_1k))
[1] 0.376586
> mean(t_1k)
[1] 15.35927
>
```

4th:

```
> t_1k = replicate(1000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_1k,col="yellow")
> 1-pexp(15,rate=1/mean(t_1k))
[1] 0.3778796
> mean(t_1k)
[1] 15.41339
>
```

5th:

```
> t_1k = replicate(1000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_1k,col="yellow")
> 1-pexp(15,rate=1/mean(t_1k))
[1] 0.3715394
> mean(t_1k)
[1] 15.14998
>
```

Test	**E(T)**	**P(T>15)**
1	14.68818	0.3601519
2	14.82368	0.365296
3	15.35927	0.376586
4	15.41339	0.3778796
5	15.14998	0.3715394

Sample size 100000:

```r
t_100k = replicate(100000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
hist(t_100k,col="yellow")
1-pexp(15,rate=1/mean(t_100k))
mean(t_100k)
```

1st :

```
> t_100k = replicate(100000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_100k,col="yellow")
> 1-pexp(15,rate=1/mean(t_100k))
[1] 0.3695234
> mean(t_100k)
[1] 15.06718
>
```

2nd:

```
> t_100k = replicate(100000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_100k,col="yellow")
> 1-pexp(15,rate=1/mean(t_100k))
[1] 0.3680012
> mean(t_100k)
[1] 15.00497
>
```

3rd:

```
> t_100k = replicate(100000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_100k,col="yellow")
> 1-pexp(15,rate=1/mean(t_100k))
[1] 0.368234
> mean(t_100k)
[1] 15.01446
>
```

4th:

```
> t_100k = replicate(100000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_100k,col="yellow")
> 1-pexp(15,rate=1/mean(t_100k))
[1] 0.3668219
> mean(t_100k)
[1] 14.95694
>
```

5th:

```
> t_100k = replicate(100000,max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> hist(t_100k,col="yellow")
> 1-pexp(15,rate=1/mean(t_100k))
[1] 0.3684889
> mean(t_100k)
[1] 15.02487
>
```

Test	E(T)	P(T>15)
1	15.06718	0.3695234
2	15.00497	0.3680012
3	15.01446	0.368234
4	14.95694	0.3668219
5	15.02487	0.3684889

Observation:

- As sample sizes increases from 1000 to 100000 change in variation of E(T) and P(T>15) is reduced.

2. (Area of circle) / (Area of square) = pi / 4

```r
pi_func <- function() {
  #The runif() function generates random deviates of the uniform distribution and is written as
  #runif(n, min = 0, max = 1) can be used to generate points inside unit square
  # 10000 values are generated for the x-coordinate
  x = runif(10000)
  # 10000 values are generated for the y-coordinate
  y = runif(10000)
  #We now find the distance of the point from the center (1/2, 1/2)
  r = sqrt((x - 0.5)^2  + (y - 0.5)^2)
  # We find the number of points that fall within the circle
  sum(r<=0.5)/10000*4
}
pi_func()
```

```
> pi_func <- function() {
+   #The runif() function generates random deviates of the uniform distribution and is written as
+   #runif(n, min = 0, max = 1) can be used to generate points inside unit square
+   # 10000 values are generated for the x-coordinate
+   x = runif(10000)
+   # 10000 values are generated for the y-coordinate
+   y = runif(10000)
+   #We now find the distance of the point from the center (1/2, 1/2)
+   r = sqrt((x - 0.5)^2  + (y - 0.5)^2)
+   # We find the number of points that fall within the circle
+   sum(r<=0.5)/10000*4
+ }
> pi_func()
[1] 3.162
```

Original value of pi =3.141
Our estimate of pi = 3.162