CS 6313 - Statistics for Data Science
Mini Project 1
Group 29 - Niharika Rajaram, Tejas Prakash Bobhate

Contribution of each member:
Niharika Rajaram:
- Solved all the problems.
- Came up with a multiline solution to solve (1b).
- Studied the theory of Monte Carlo simulation to understand and implement question 2.

Tejas Prakash Bobhate:
- Solved all the problems.
- Came up with a single line solution to solve (1b).
- Plotted the graph for inscribing square and circle in question 2.

Both of us solved all the questions and mentioned the best solution we had in the report.
Niharika filled in the question 1 of section 1 while Tejas worked on the question 2 of section 1 and the section 2 of the mini project report.

Section 1

Solutions:

(1a)

classmate

Date

Page

1(a) Given the probability density function of T,

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

With this density, we can infer the cumulative density function :

$$F_T(t) = \int_0^x f_T(t)\,dt \qquad , \quad (t>0)$$

Probability that the lifetime of the sattelite exceeds 15 years $P(T>15)$ can be given by,

$$P(T>15) = 1 - P(T \le 15) = 1 - F(T \le 15)$$

$$\Rightarrow P(T>15) = 1 - \int_0^{15} (0.2e^{-0.1t} - 0.2e^{-0.2t})\,dt$$

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

$$= 1 - 0.2 * \left[(-1/0.1)e^{-0.1t} - (-1/0.2)e^{-0.2t}\right]_0^{15}$$

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

$$= 1 - [(-0.44626032029 + 0.04978706836) - (-2+1)]$$

$$P(T>15) = 0.3965$$

(1b)

i.

One draw of satellite lifetime T using Xa and Xb is the maximum of Xa,Xb (Note : Xa and Xb start simultaneously, if any one fails the other will continue working, therefore we consider the maximum of Xa and Xb).

lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}

```
> Xa = rexp(n=1,0.1)
> Xa
[1] 2.237966
> Xb = rexp(n=1,0.1)
> Xb
[1] 17.30918
> T=max(Xa,Xb)
> T
[1] 17.30918
```

This can be written in a single line as

```
T=max(rexp(n=1,0.1),rexp(n=1,0.1))
```

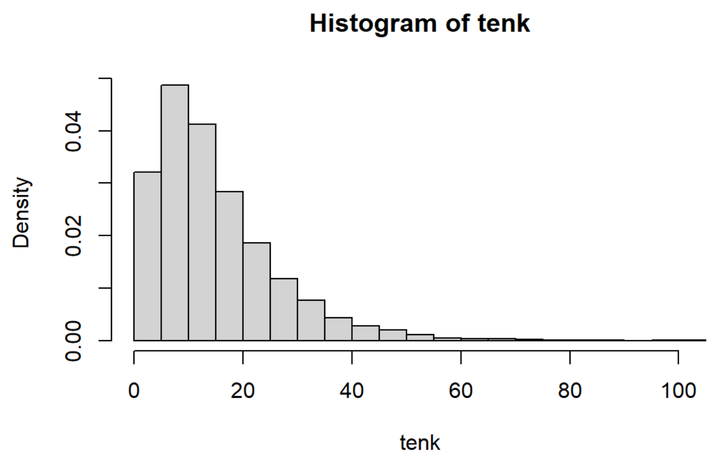
ii.

Replicating the previous step to draw the value of satellite lifetime 10,000 times.

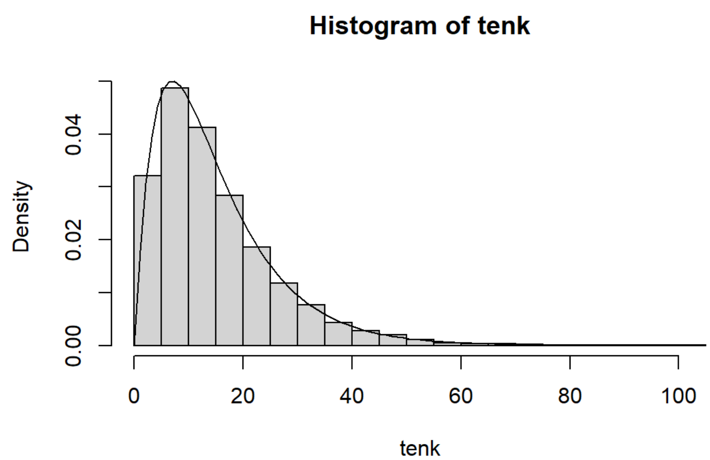
```
> tenk <- replicate(10000, max(rexp(n=1,rate=0.1),rexp(n=1,rate=0.1)))
> tenk
   [1] 19.4367924 15.8697999 10.3446516 10.0337986  3.3529360 41.1985554
   [7] 13.7868050  2.3718996 38.5511751 17.1988011 25.0480609  2.5996389
  [13] 49.2195492  9.8828872 11.1096467  6.3422155 18.4496954  2.9470498
  [19]  2.4549250 12.0854180 16.1787385 12.0608026 12.3706334  9.8948469
  [25] 11.0207032 44.0295148 13.7761128 10.2222789 34.5869603 11.3560056
  [31] 51.2589990 10.4306831 19.0505382 12.7323433 18.9195191 30.9101722
  [37] 37.6860142  7.3716178  4.2106867 10.9780473 35.8034687 15.3174771
```

iii.

hist(tenk, probability = TRUE)



curve(lt, add = TRUE)



We see in the above plot that the probability density function curve overlaps and matches the histogram. This proves that the Monte Carlo estimation is providing the correct results.
iv.

```
> mean(tenk)
[1] 14.86095
```

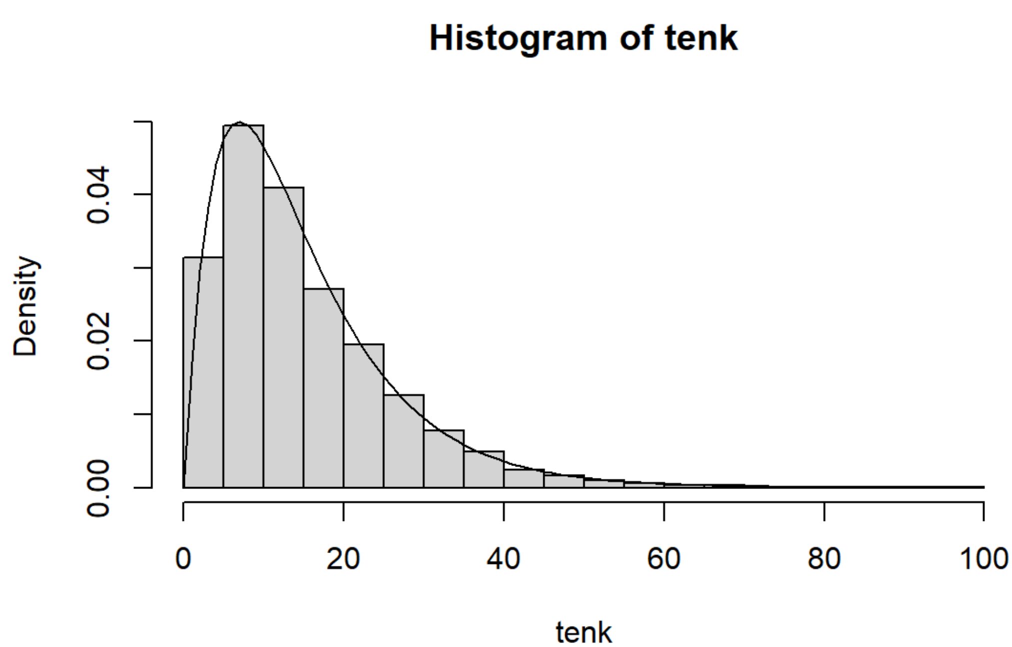
v.

```
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.3896
```

vi.

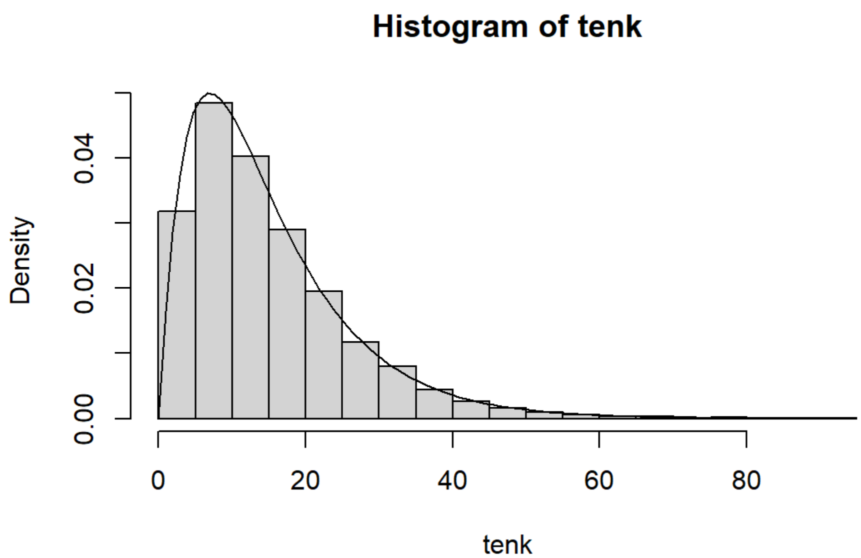
Test 2

```
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.82199
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.3909
```



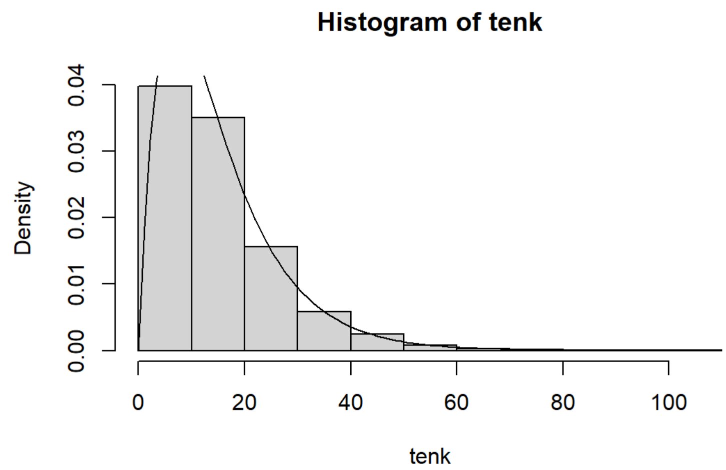
Test 3

```
> # picking 10000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.93469
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.3974
>
```



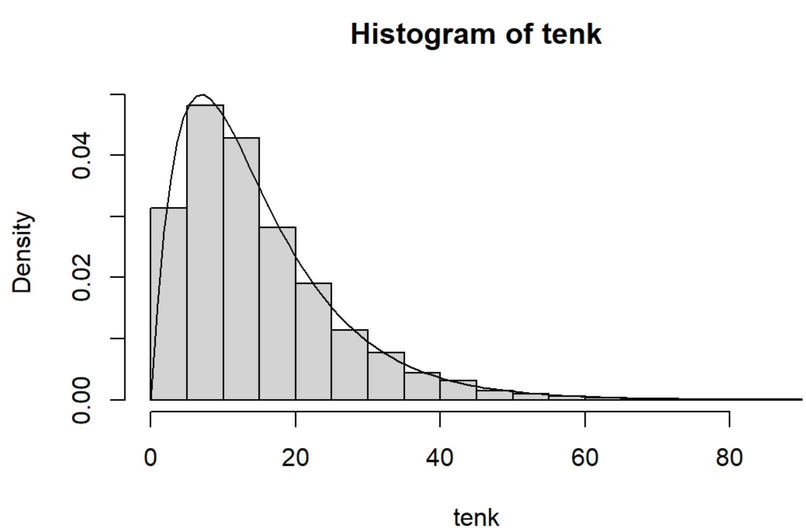
Test 4

```
> # picking 10000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 15.01152
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.3919
>
```



Test 5

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 10000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.81709
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.3878
```



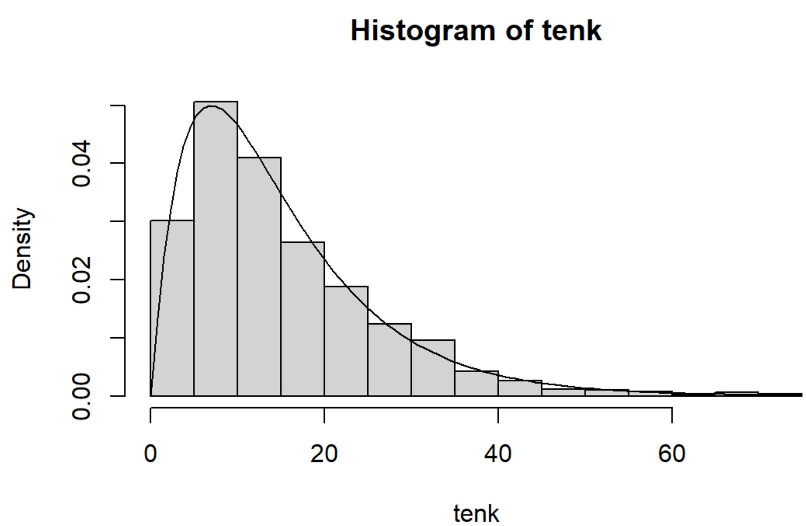
Test	E(T)	P(T > 15)
1	14.86095	0.3896
2	14.82199	0.3909
3	14.93469	0.3974
4	15.01152	0.3919
5	14.81709	0.3878

Average E(T) = 14.8892
Average P(T > 15) = 0.39152

(1c)
Sample size = 1000

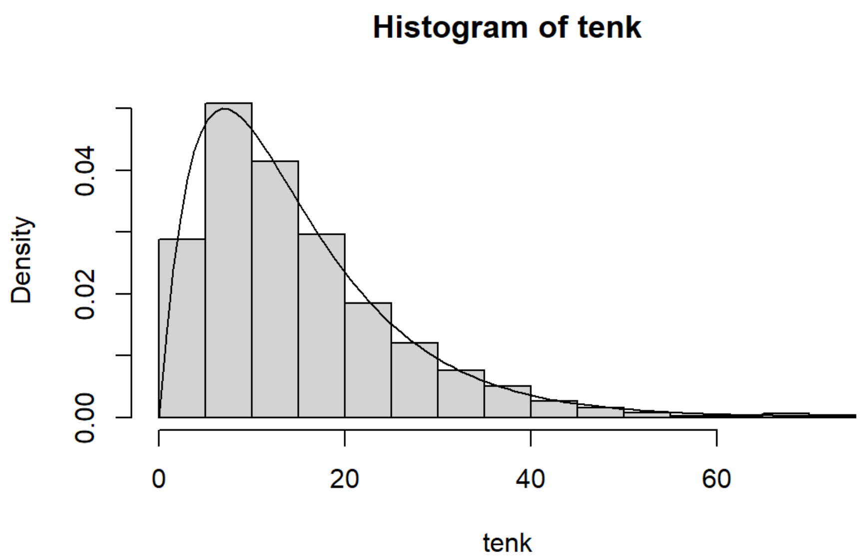
Task 1

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 1000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 15.05642
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.391
```



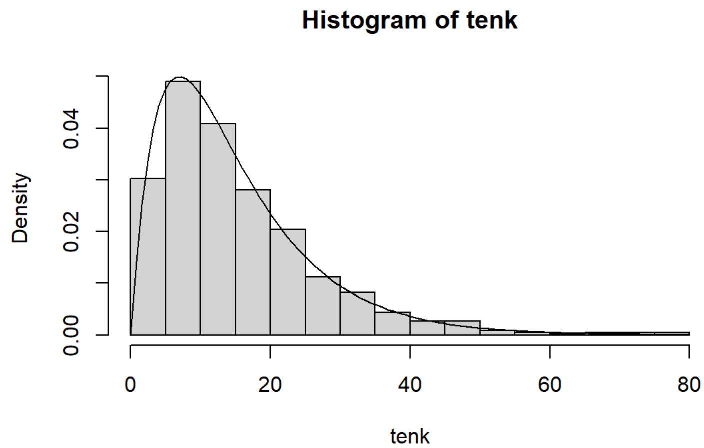
Task 2

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 1000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.96947
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.395
```



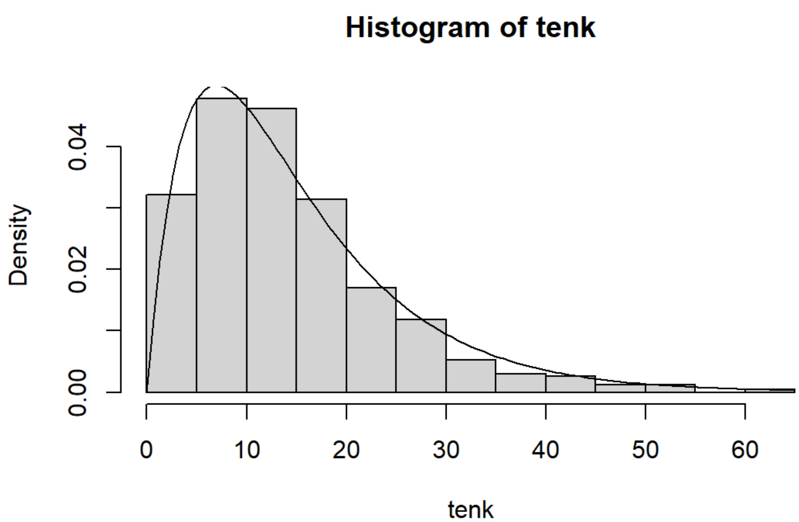
Task 3

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 1000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 15.24657
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.4
```



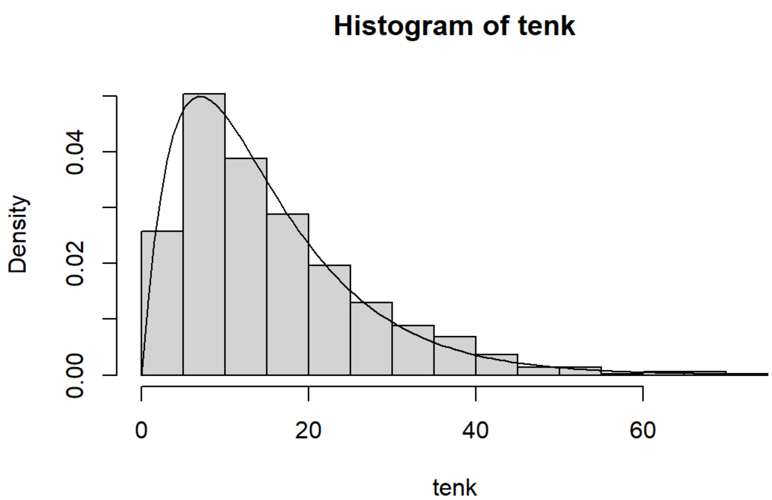
Task 4

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 1000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.0465
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.369
```



Task 5

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 1000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 15.89178
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.425
>
```



Test	E(T)	P(T > 15)
1	15.05642	0.391
2	14.96947	0.395
3	15.24657	0.4
4	14.0465	0.369
5	15.89178	0.425

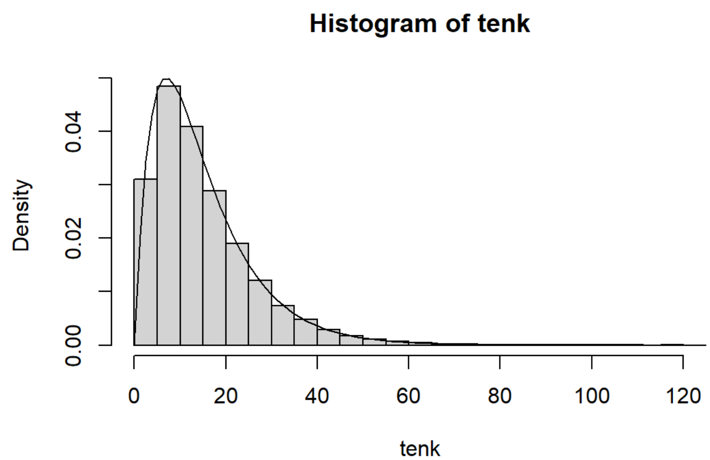
Average E(T) = 15.042148

Average P(T > 15) = 0.396

Sample size: 100000

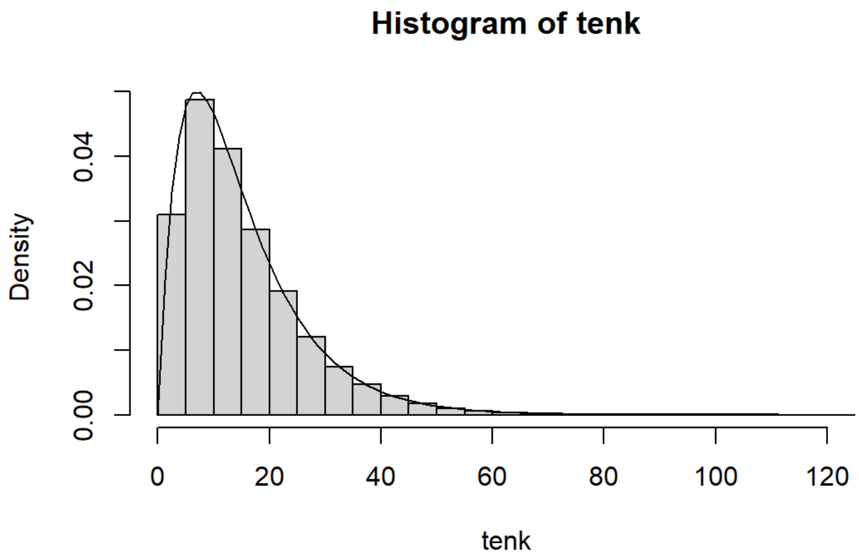
Task 1

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 100000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 15.03663
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.39781
```



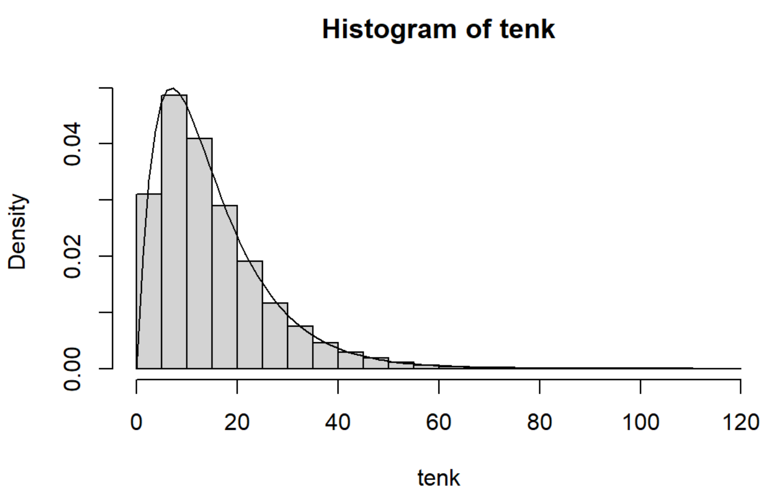
Task 2

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 100000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.97876
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.39572
```



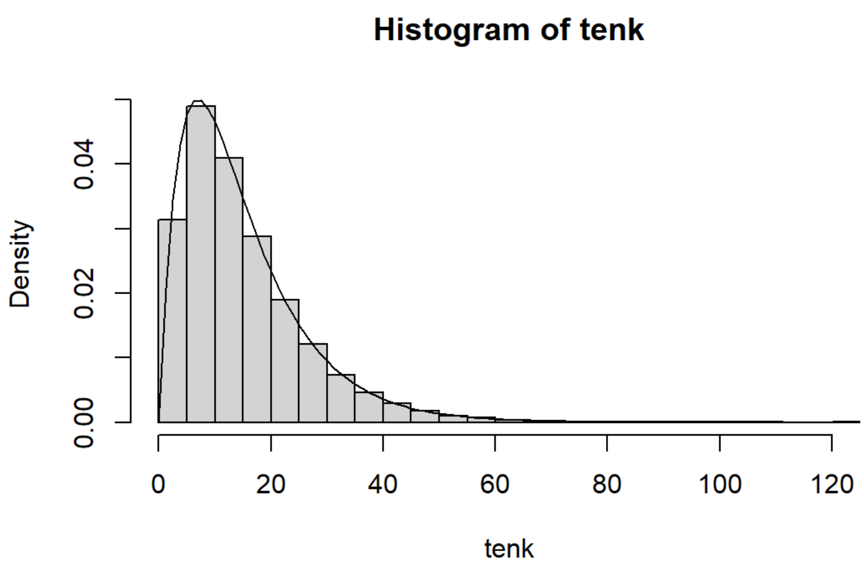
Task 3

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 100000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 15.00218
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.39692
```



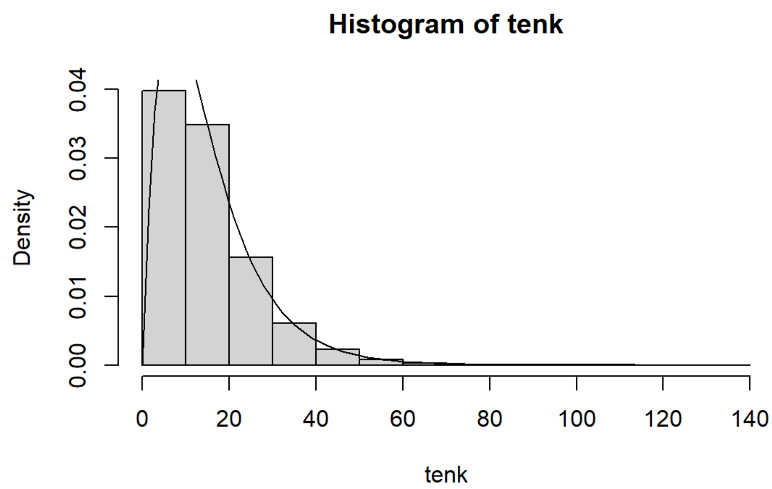
Task 4

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 100000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.93474
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.39446
```



Task 5

```
> lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
> # picking 100000 random points
> tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
> # plotting a histogram of the randomly generated points
> hist(tenk, probability=TRUE)
> # adding a curve to the histogram
> curve(lt, add=TRUE)
> # E(T)
> mean(tenk)
[1] 14.98687
> # probability of lifetime of sattelite being more than 15 years
> sum(tenk > 15)/size
[1] 0.39717
```



Test	E(T)	P(T > 15)
1	15.03663	0.39781
2	14.97876	0.39572
3	15.00218	0.39692
4	14.93474	0.39446
5	14.99687	0.39717

Average E(T) = 14.989836
Average P(T > 15) = 0.396416

Conclusion:

Runs	E(T)	P(T > 15)
1000	15.042148	0.396
10000	14.8892	0.39152
100000	14.989836	0.396416

From the above table, we see that, with the increase in the number of draws, the E(T) and P(T >15) gets closer to the analytical answers of 15 and 0.3965 respectively.

(2)
The square of unit length with vertices at (0,0), (0,1), (1,0), (1,1). Inside this, we will inscribe a circle of radius 0.5 units.

Area of square of unit length = 1^2

Area of inscribed circle of radius 0.5 unit = $\pi * (0.5)^2$

The probability that the point will lie inside the circle = $\frac{Area\ of\ inscribed\ circle\ of\ radius\ 0.5\ unit}{Area\ of\ square\ of\ unit\ length}$

P(x inside circle) = $\frac{\pi * (0.5)^2}{1}$

Considering this, if we can determine the probability of point lying inside the circle, then the value of π can be approximated using $\pi = 4 * P(x\ inside\ the\ circle)$

We will randomly generate 10000 points inside the square and check how many of these are satisfying the circle. This count will give us the probability of the point lying inside the circle.

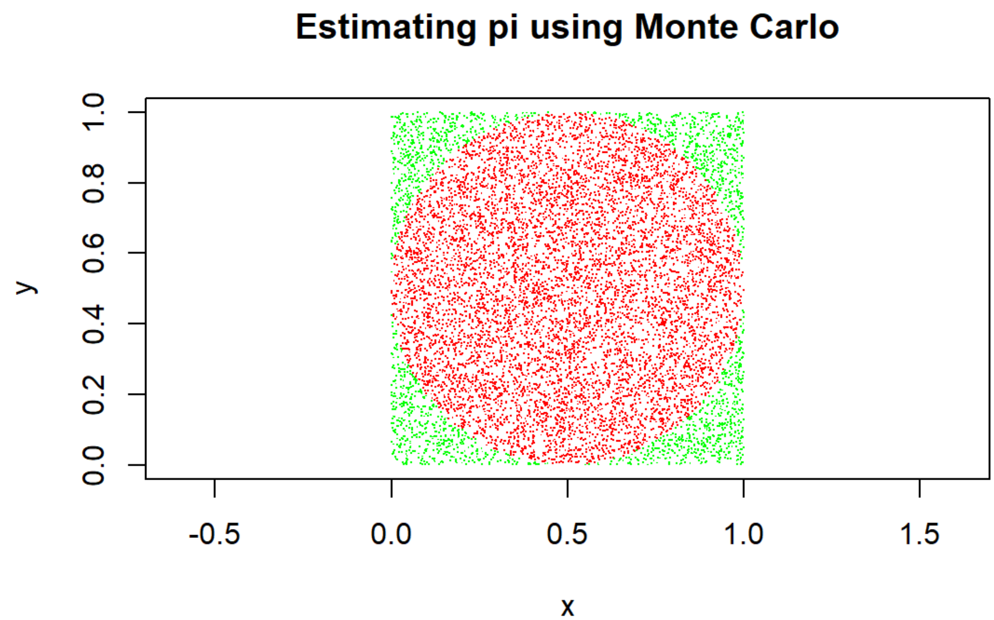
```
# setting up 10000 runs
runs <- 10000

square_x = runif(runs, min=0, max=1)
square_y = runif(runs, min=0, max=1)

circle_points = (square_x-0.5)^2 + (square_y-0.5)^2 <= (0.5)^2

pi = 4 * sum(circle_points) / runs

print(pi)

plot(square_x, square_y,
     pch='.',
     asp=1,
     xlab='x',
     ylab='y',
     col=ifelse(circle_points, "red", "green"),
     main = "Estimating pi using Monte Carlo"
     )
```

The value of pi as estimated by Monte Carlo estimation approach is,

```
> print(pi)
[1] 3.1592
```

which is close to the actual value of pi = 3.1415

The plot of the inscribed square and the circle is:



Section 2

Question 1 - Code

Note: The number of draws can be changed in the size variable as needed.

```
# sample size
size=100000
# defining the probability density function
lt<-function(x){(0.2*exp(-0.1*x)-0.2*exp(-0.2*x))}
# picking 100000 random points
tenk <- replicate(size, max(rexp(n=1, rate=0.1), rexp(n=1, rate = 0.1)))
# plotting a histogram of the randomly generated points
hist(tenk, probability=TRUE)
# adding a curve to the histogram
curve(lt, add=TRUE)
# E(T)
mean(tenk)
# probability of lifetime of sattelite being more than 15 years
sum(tenk > 15)/size
```

Question 2 - Code

```r
# setting up 10000 runs
runs <- 10000

# defining coordinates of the square
# points are randomly generated within this region for uniform distribution
square_x = runif(runs, min=0, max=1)
square_y = runif(runs, min=0, max=1)

# defining the condition which all the points inside the circle should satisfy
circle_points = (square_x-0.5)^2 + (square_y-0.5)^2 <= (0.5)^2

# as per explanation of pi
pi = 4 * sum(circle_points) / runs

print(pi)

# plot the square and circle
plot(square_x, square_y,
     pch='.',
     asp=1,
     xlab='x',
     ylab='y',
     col=ifelse(circle_points, "red", "green"),
     main = "Estimating pi using Monte Carlo"
     )
```