Statistical Methods for Data Science

Mini Project 1

Mini Project Group No # 11
Names of group members: Parikh Parashar, Prajapati Aakash

Contribution of each group member:
Both worked together to finish this mini project. Aakash did documentation and done all tabular calculation. Parashar did R code. We collaborated and worked on each task parallelly.

---

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

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

and $E(T) = 15$ years.

(a) Use the above density function to analytically compute the probability that the lifetime of the satellite exceeds 15 years.

Q.1

a. We are given;

$E(X_A) = E(X_B) = 1/\lambda = 1/10 = 0.10$

and $E(T) = 15$.

now, $P(T>15) = 1 - P(T \leq 15)$

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

$$= 1 - \left(0.2 \left[\frac{e^{-0.1t}}{-0.1}\right]_0^{15} - 0.2 \left[\frac{e^{-0.2t}}{-0.2}\right]_0^{15}\right.$$

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

$$= 1 + 2e^{-1.5} - 1 + e^{-3}$$

$$= 2e^{-1.5} + e^{-3}$$

$$\boxed{\therefore\ P(T>15) = 0.39648}$$

(b) Use the following steps to take a Monte Carlo approach to compute $E(T)$ and $P(T > 15)$.

i. Simulate one draw of the block lifetimes X_A and X_B. Use these draws to simulate one draw of the satellite lifetime T.

ii. Repeat the previous step 10,000 times. This will give you 10,000 draws from the distribution of T. Try to avoid 'for' loop. Use 'replicate' function instead. Save these draws for reuse in later steps. [**Bonus**: 1 bonus point for not taking more than 1 line of code for steps (i) and (ii).]

iii. Make a histogram of the draws of T using 'hist' function. Superimpose the density function given above. Try using 'curve' function for drawing the density. Note what you see.

iv. Use the saved draws to estimate $E(T)$. Compare your answer with the exact answer given above.

v. Use the saved draws to estimate the probability that the satellite lasts morethan 15 years. Compare with the exact answer computed in part (a).

vi. Repeat the above process of obtaining an estimate of $E(T)$ and an estimate of the probability four more times. Note what you see.

Solution:

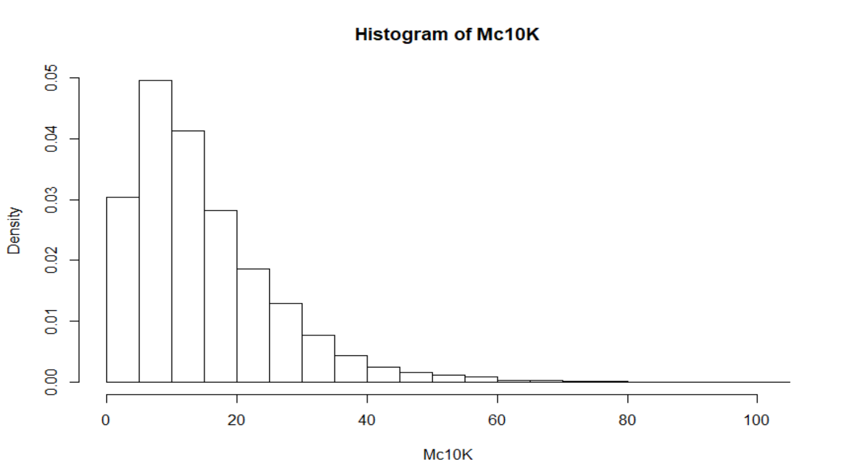
i & ii) Randomly generate 10000 draws of X_A and X_B from the distribution T:

```
> Mc10K = replicate(10000,max(rexp(1,0.1),rexp(1,0.1)))
```

Values	
Mc10K	num [1:10000] 16.22 5.57 6.2 4.03 1.77 ...

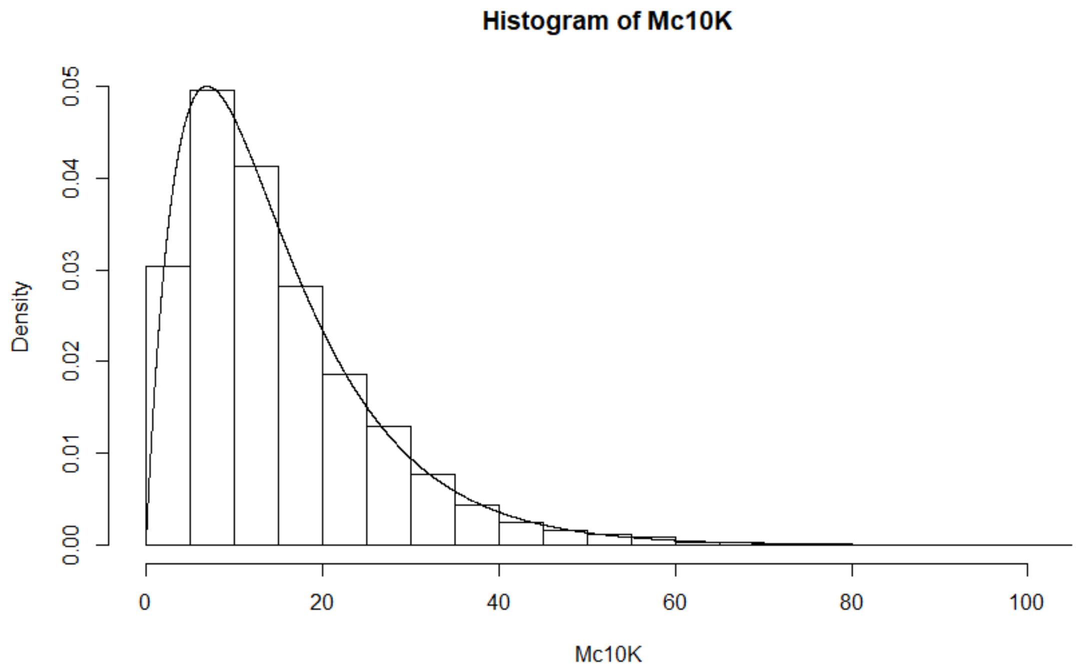
iii) Create histogram of the draw of T:

```
> hist(Mc10K,freq=FALSE)
>
```



Used 'curve' function for drawing the density:

```
> curve((0.2 * exp(-0.1 * x) - 0.2 * exp(-0.2 * x)), from=0, to=80, n=10000, add=TRUE)
>
```



iv) Estimating the E(T) using randomly generated draws

```
> meanMc10k = mean(Mc10k)
>
```

meanMc10k	14.9325098850204

v) To estimate the probability of T when the lifetime T exceeds 15 years using the draws

```
> predictedMc10k = mean(abs(Mc10k>15))
>
```

predictedMc10k	0.3973

vi) We can conclude from the table that value of expectation and probability are depended on the random number generated using the function. And expected values are equal to the theoretically calculated values with minor errors.

Draw 2

```
> Mc10k = replicate(10000,max(rexp(1,0.1),rexp(1,0.1)))
> mean(Mc10k)
[1] 14.96821
> mean(abs(Mc10k>15))
[1] 0.3952
> 
```

Draw 3

```
> Mc10k = replicate(10000,max(rexp(1,0.1),rexp(1,0.1)))
> mean(Mc10k)
[1] 14.89335
> mean(abs(Mc10k>15))
[1] 0.3934
> 
```

Draw 4

```
> Mc10k = replicate(10000,max(rexp(1,0.1),rexp(1,0.1)))
> mean(Mc10k)
[1] 15.06208
> mean(abs(Mc10k>15))
[1] 0.3975
> 
```

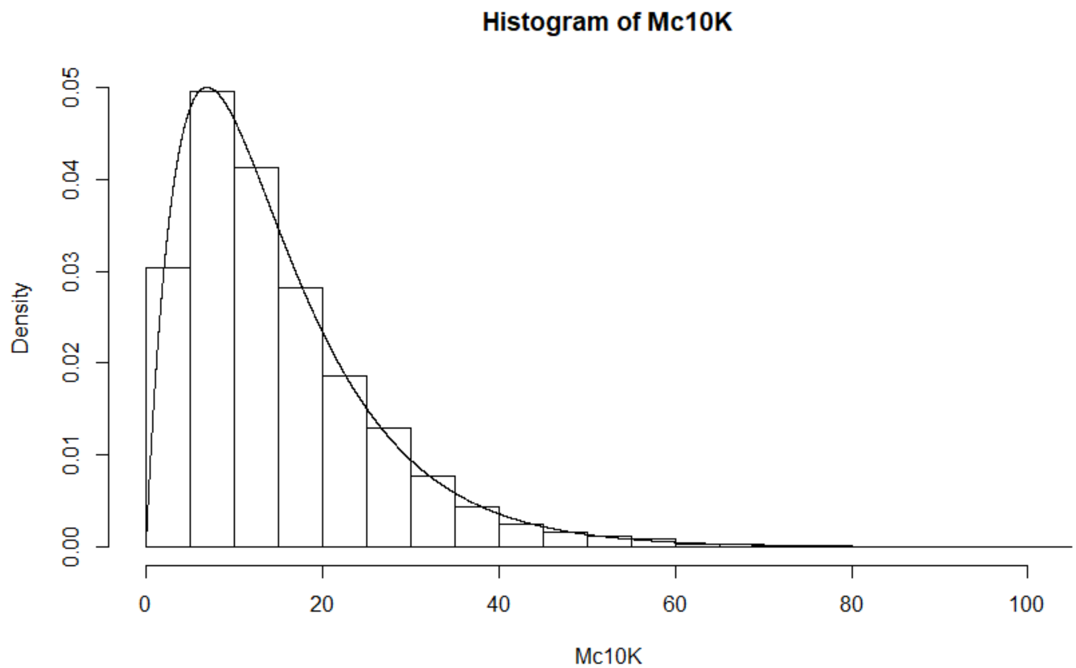
Draw 5

```
> Mc10k = replicate(10000,max(rexp(1,0.1),rexp(1,0.1)))
> mean(Mc10k)
[1] 15.06069
> mean(abs(Mc10k>15))
[1] 0.395
>
```

From the below table we can conclude following:

1. Average of E(T) = 14.77 with average error = ±0.07704 for 10000 random numbers
2. Average of P(T>15) = 0.3792 with average error = ±0.002591 for 10000 random numbers

Draws #	E(T)	Error in E(T)	P(T>15)	Error in P(T>15)
1	15.2275	±0.227	0.39	±0.00648
2	14.4934	±0.506	0.376	±0.020
3	14.6360	±0.3639	0.381	±0.015
4	14.76869	±0.2313	0.361	±0.035
5	14.78782	±0.2121	0.388	±0.0084



(c) Repeat part (vi) five times using 1,000 and 100,000 Monte Carlo replications instead of 10,000. Make a table of results. Comment on what you see and provide an explanation

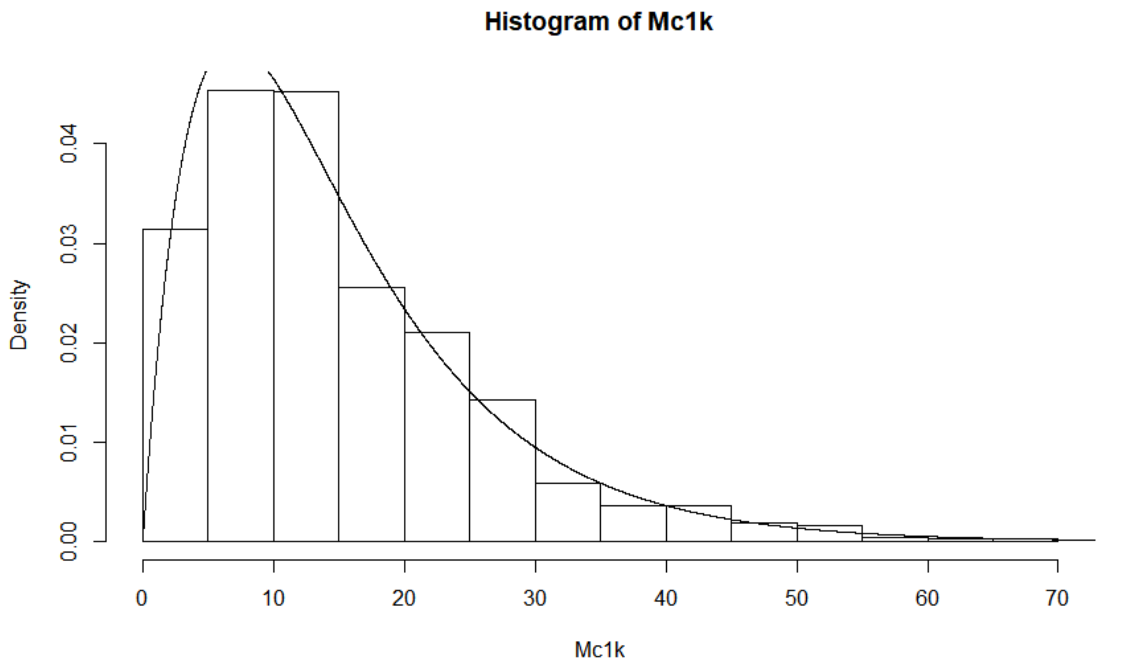
Using 1000 random numbers:

Draws #	E(T)	Error in E(T)	P(T>15)	Error in P(T>15)
1	14.9325	±0.068	0.3973	±0.00082
2	14.968	±0.032	0.3952	±0.00128
3	14.89335	±0.106	0.3934	±0.00308
4	15.06208	±0.062	0.3975	±0.00102
5	15.06069	±0.060	0.395	±0.00148

From the above table from 1000 random variable we can conclude following:

1. Average of E(T) = 14.98 with average error = ±0.084 for 1000 random numbers

2. Average of P(T>15) = 0.39568 with average error = ±0.000208 for 1000 random numbers

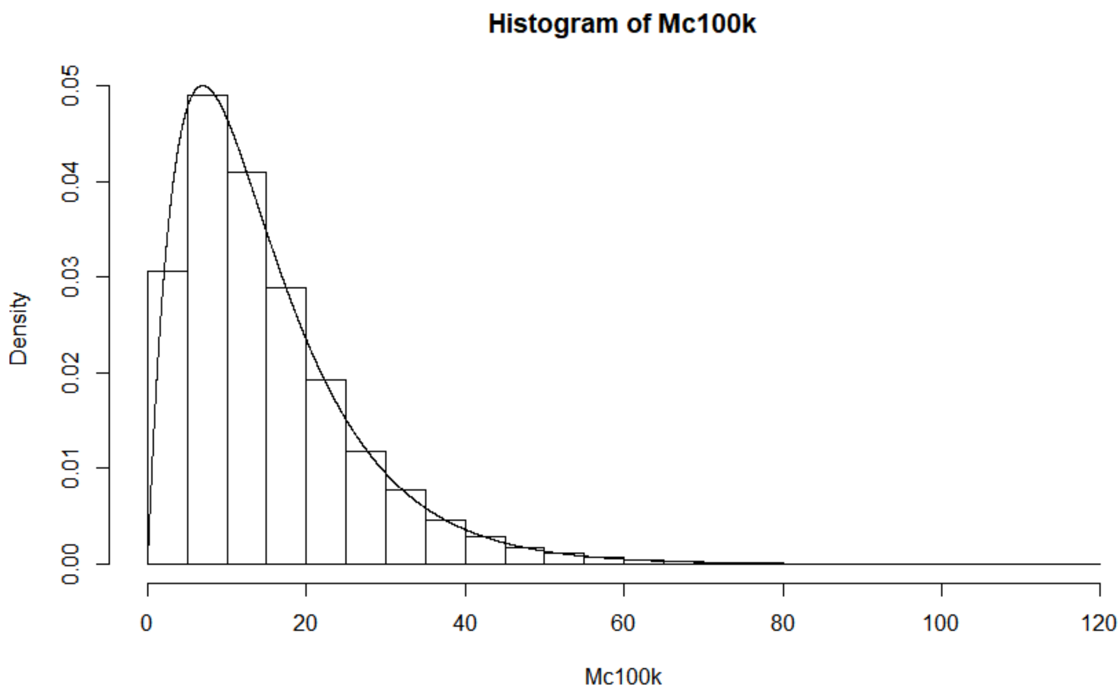


Using 100000 random numbers:

Draws #	E(T)	Error in E(T)	P(T>15)	Error in P(T>15)
1	15.0098	0.0098	0.397	0.0010
2	14.960	0.0390	0.3944	0.0019
3	14.9796	0.0203	0.39712	0.00064
4	14.9864	0.01397	0.397	0.00091
5	15.0019	0.00197	0.396	0.00018

From the above table from 100000 random variable we can conclude following:

1. Average of E(T) = 14.988 with average error = ±0.0123 for 100000 random numbers
2. Average of P(T>15) = 0.39568 with average error = ±0.00016 for 100000 random numbers

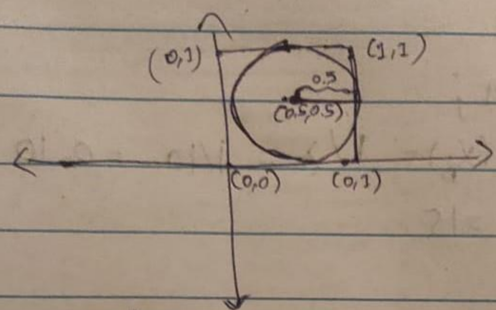


From all the tables above we can conclude that average values are more accurate when size of random variables is high.

Therefor, we can say that Monte Carlo follows the Laws of Large numbers.

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.]

2.



circle ÷ $x^2+y^2=1$ unit circle.

probability of a point falling in the circle cover the space of a square = $\frac{\text{Area of a circle}}{\text{Area of square}} = \frac{\pi}{4}$

So now we generate 10,000 values of x,y from 0 to 1. from that we can calculate π value.

R code:

```
#randomly generated 10000 x,y coordinates
x = runif(10000, min=0, max = 1)
y = runif(10000, min=0, max = 1)
#circle equatiion
circle <-(x-0.5)^2+(y-0.5)^2 <= 0.5^2
#average of circle 10000 ittration value
pi = (sum(circle)/10000)*4
#pi value
pi
```

Test #	Value of Pi	Error in Pi
1	3.1476	0.0076
2	3.1348	0.0052
3	3.1748	0.0348
4	3.1424	0.0024
5	3.1376	0.0024

Average Pi=3.1468

Average Error=±0.01048

So we got average Pi value 3.1468 which is close to actual value 3.14159.