STATISTICAL METHODS FOR DATA SCIENCE CS6313-001 FALL 2019

Mini Project #1

Participant #1 : Reetish Chand Guntakal Patil (RXG190006)

1. (10 points) Consider Exercise 4.11 from the textbook. In this exercise, let XA be the

lifetime of block A, XB be the lifetime of block B, and T be the lifetime of the satellite.

The lifetimes are in years. It is given that XA and XB 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

fT (t) =

(

0:2 exp(􀀀0:1t) 􀀀 0:2 exp(􀀀0:2t); 0 \_ t < 1;

0; otherwise;

and E(T) = 15 years.

1

(a) Use the above density function to analytically compute the probability that the

lifetime of the satellite exceeds 15 years.

(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 XA and XB. 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 more

than 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.

(c) Repeat part (vi) \_ve 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.

# Solution (b)

1. Simulate one draw of the block lifetimes XA and XB. Use these draws to simulate one draw of the satellite lifetime T.

# Ans:

Given, E(X)=1 ie mean is 1. So, 𝛌 = 1/E(X) => 1/10=0.1. Thus. 𝛌 = 0.1

As the given distribution follows an exponential distribution use rexp() in R to simulate one draw of the block lifetimes of XA and XB. Let **a** be the draw of XA and **b** be the draw of XB.

Here

*randA = rexp(1, 0.1)*

which gives us the value a = 4.247287

*randB = rexp(1, 0.1)*

Which gives us the value b = 5.998225

Now find the maximum value of both the blocks and name it as **maxValue**

*maxValue = max(randA , randB)*

Which gives us maxValue = 5.998225

1. Repeat the previous step 10,000 times.

# Ans:

This gives 10,000 draws from the distribution of T. Use replicate() in R to repeat the process 10,000. In this case **for** loop can be used but replicate simplifies the code.

*list\_nDraws=replicate(10000,max(rexp(1,0.1),rexp(1,0.1)))*

Now list\_nDraws contains 10000 random variables following an exponential distribution.

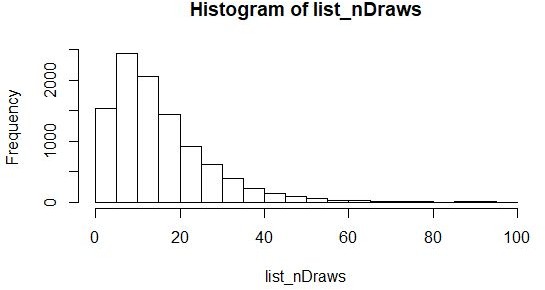
1. Make a histogram of the draws of T using ‘hist’ function. Superimpose the density function given above. Try using ‘curve’ function for drawing the density. Note what you see.

# Ans:

To generate the histogram for the above list named as **list\_nDraws** we use the ‘hist’ function in R.

*hist(list\_nDraws, probability = TRUE)*

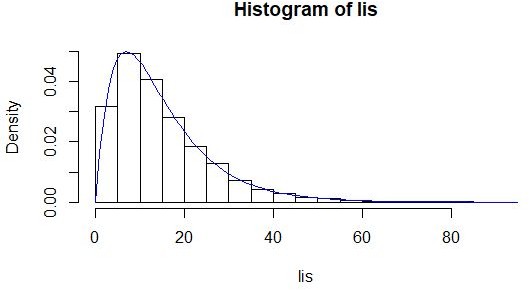
Which gives us the following



Now use the curve function to get an exponential curve over the histogram since the function defined in the question is exponentially distributed. The following R code is used to superimpose the curve and the histogram.

*curve((0.2\*exp(-0.1\*x)) - (0.2\*exp(-0.2\*x)), add = TRUE, col= "BLUE")*

The curve generated by the above command is as follows



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

# Ans:

E(T) is the mean. To find the mean of 10000 draws mean() is used in R.

*mean\_Value = mean(list\_nDraws)*

For the current draw

*mean\_Value = 15.00929*.

Observation: In the problem the given E(T) = 15. The derived E(T) value is 15.00929. Thus, the derived E(T) varies by ± 0.2 units.

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

# Ans:

In order to estimate P(T>15), count of all the values in list\_nDraws greater than 15 are required. This is achieved by :

*count=length(which(list\_nDraws >15))*

Probability is given by the count calculated in the previous step divided by the sample size ie 10000.

*prob\_15 = count/10000*

For the current draw

*count = 3964*

*prob\_15 = 0.3964*

**Observation:** The probability got above is almost the same as the probability calculated from the exponential function given in the question which is

P ( T > 15 ) = 0.3965

So the difference between both values is 0.0001 which is almost negligible so it proves that the probability is the same in both cases.

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

# Ans:

In order to repeat the entire process again, a user defined function with the required steps has been created in R as follows() :

*find\_prob<-function(n,mean,lam){ list\_nDraws=replicate(10000,max(rexp(1,0.1),rexp(1,0.1))) hist(list\_nDraws,probability = TRUE)*

*curve((0.2\*exp(-0.1\*x))-(0.2\*exp(-0.2\*x)),add=TRUE,col="BLUE") mean\_Value =mean(list\_nDraws) count=length(which(list\_nDraws >15))*

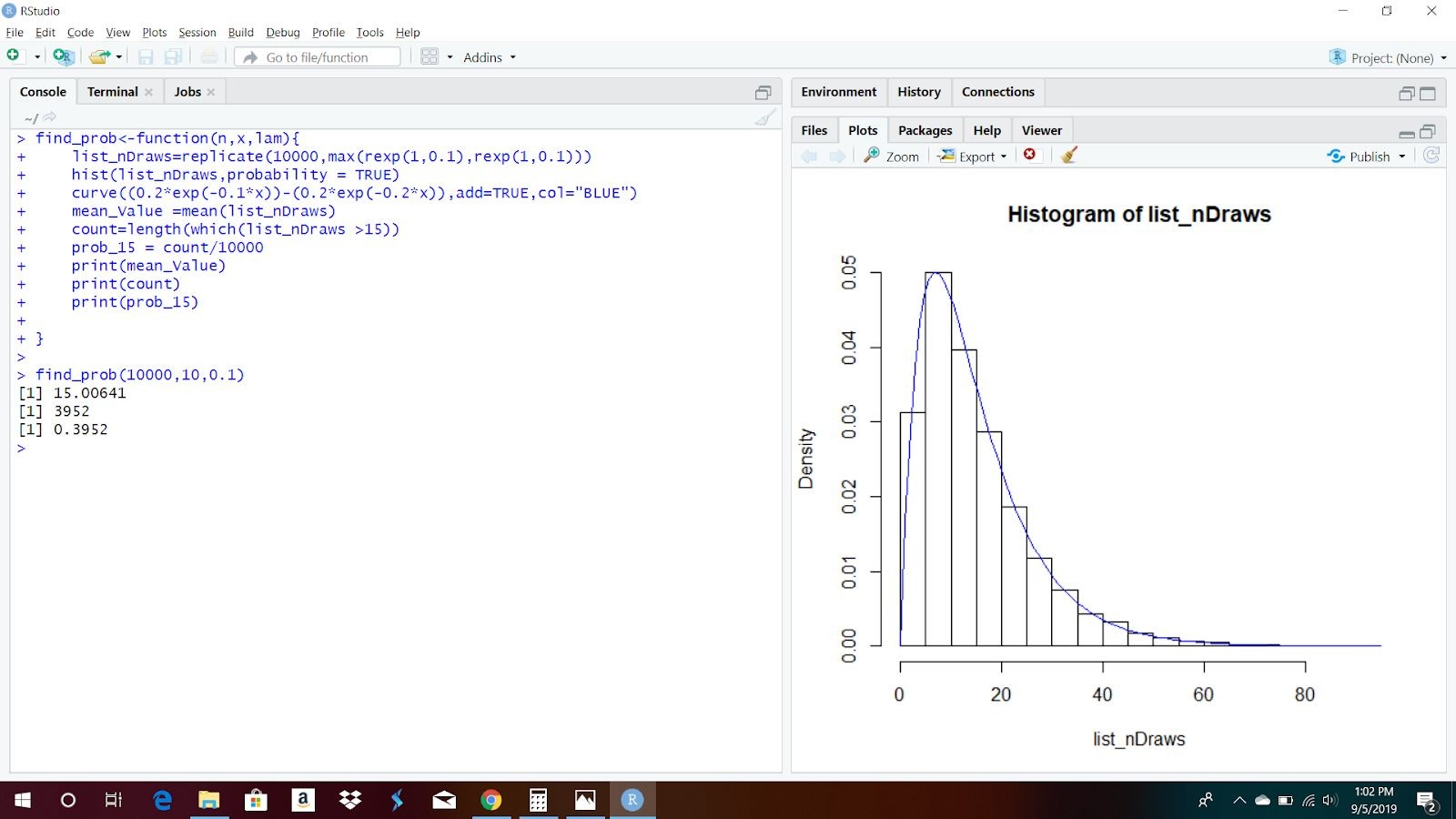
*prob\_15 = count/10000 print(“E(T) : “,mean\_Value)*

*print(“ count of variables >15: ”,count)*

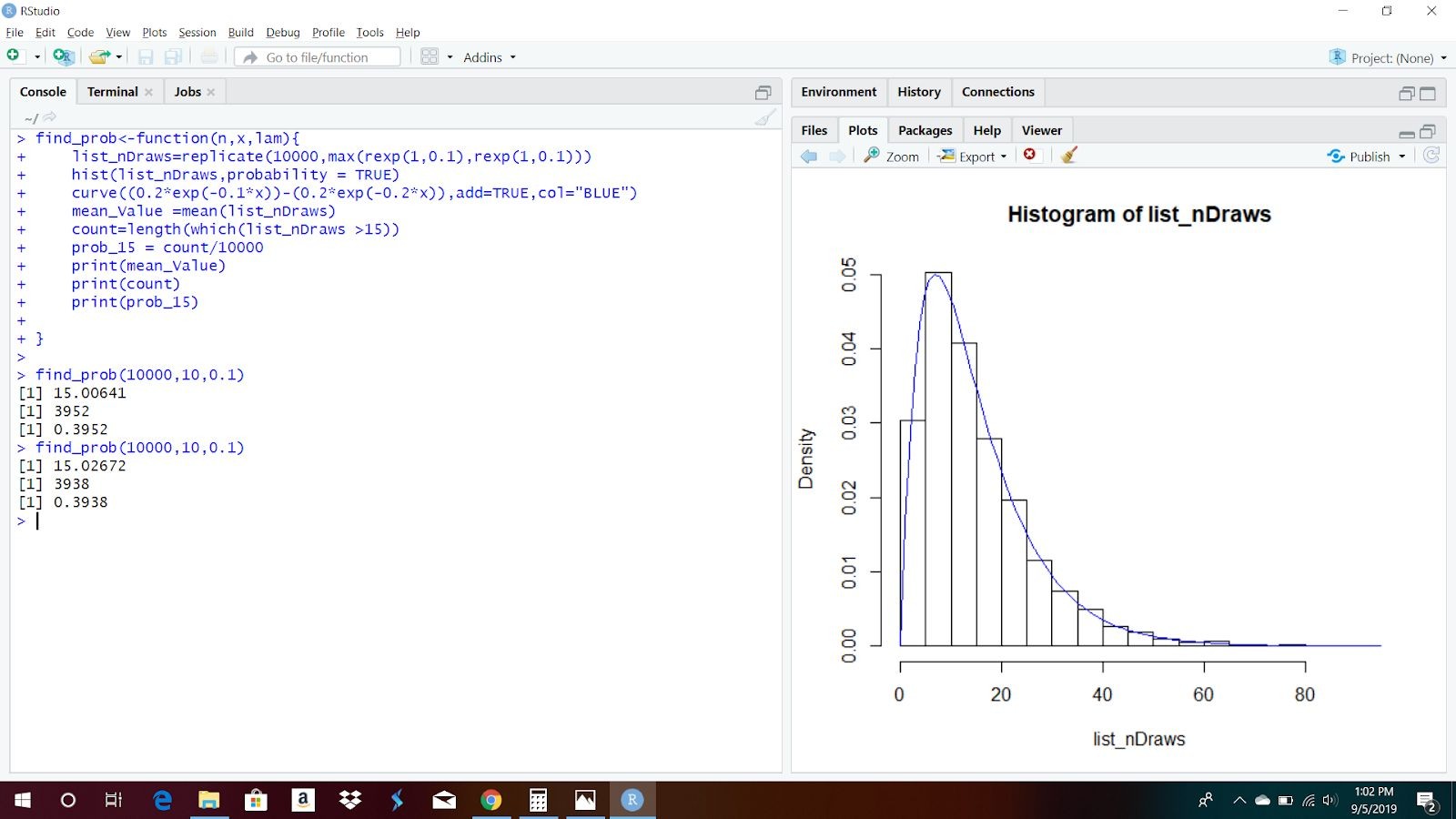
*print(“Probability: “,prob\_15) }*

This function is called 4 times by passing the parameters sample size ie 10000, E(X)= 10, 𝛌=0.1

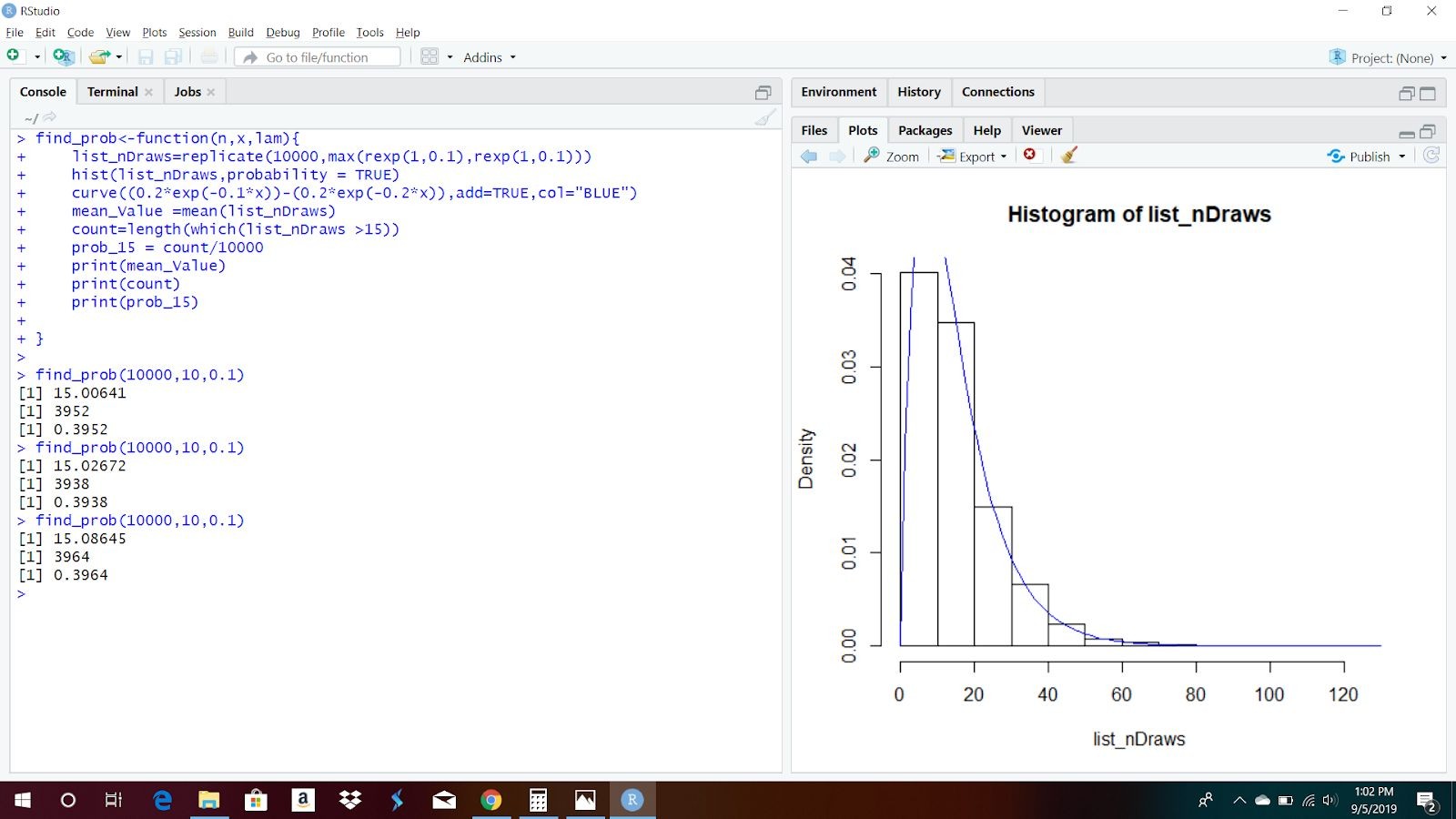
# Trial 1



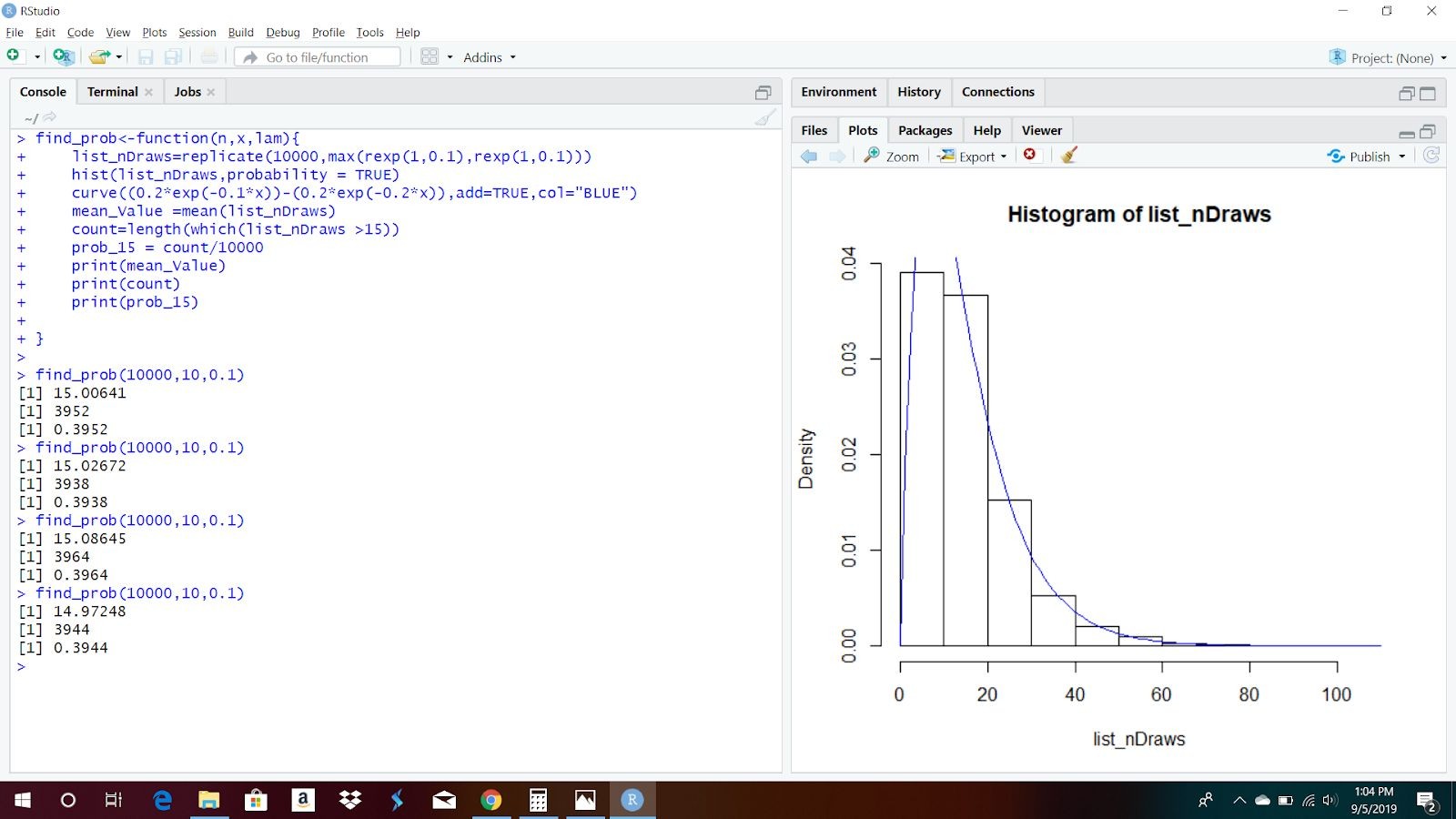
**Trial 2**



**Trial 3**



**Trial 4**



The table of mean and probabilities calculated above is as follows

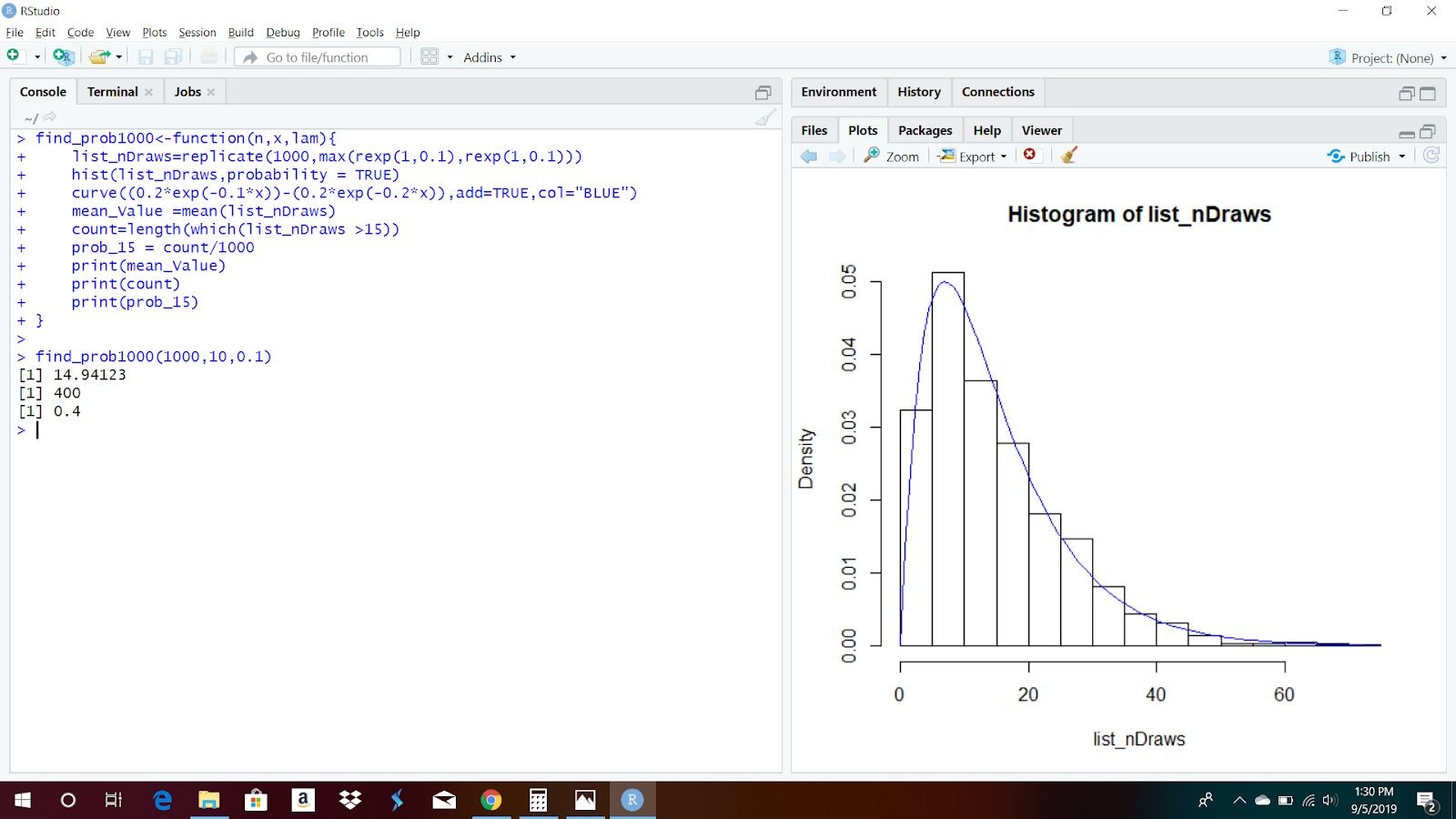
|  |  |  |  |
| --- | --- | --- | --- |
| **Sample Size** | **Mean Value** | **Count > 15** | **P ( T>15 )** |
| 10000 | 15.00641 | 3952 | 0.3952 |
| 10000 | 15.02672 | 3938 | 0.3938 |
| 10000 | 15.08645 | 3964 | 0.3964 |
| 10000 | 14.97248 | 3944 | 0.3944 |

1. (c). Repeat part (vi) ve 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

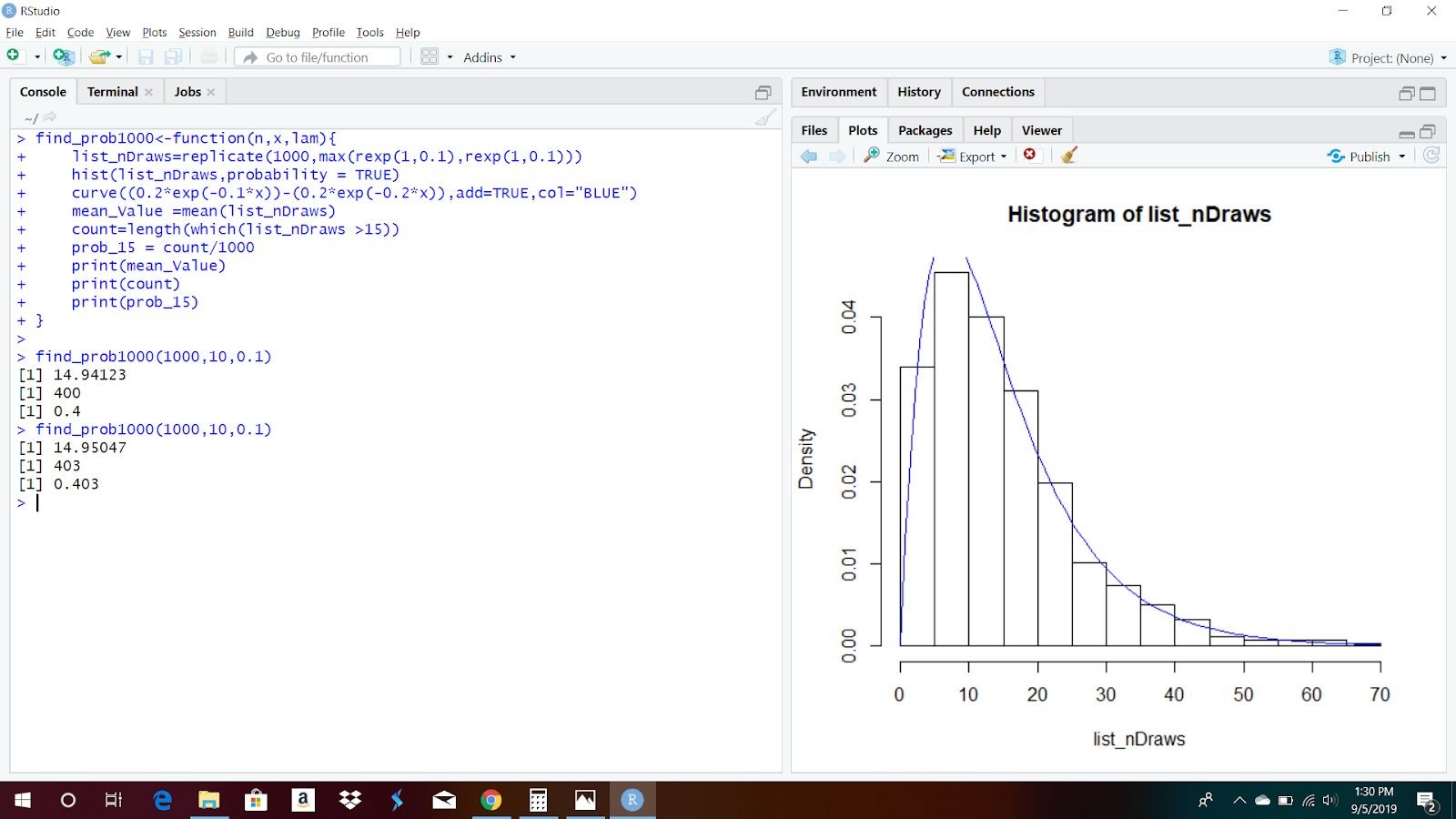
# Ans:

Case 1 - When n = 1,000

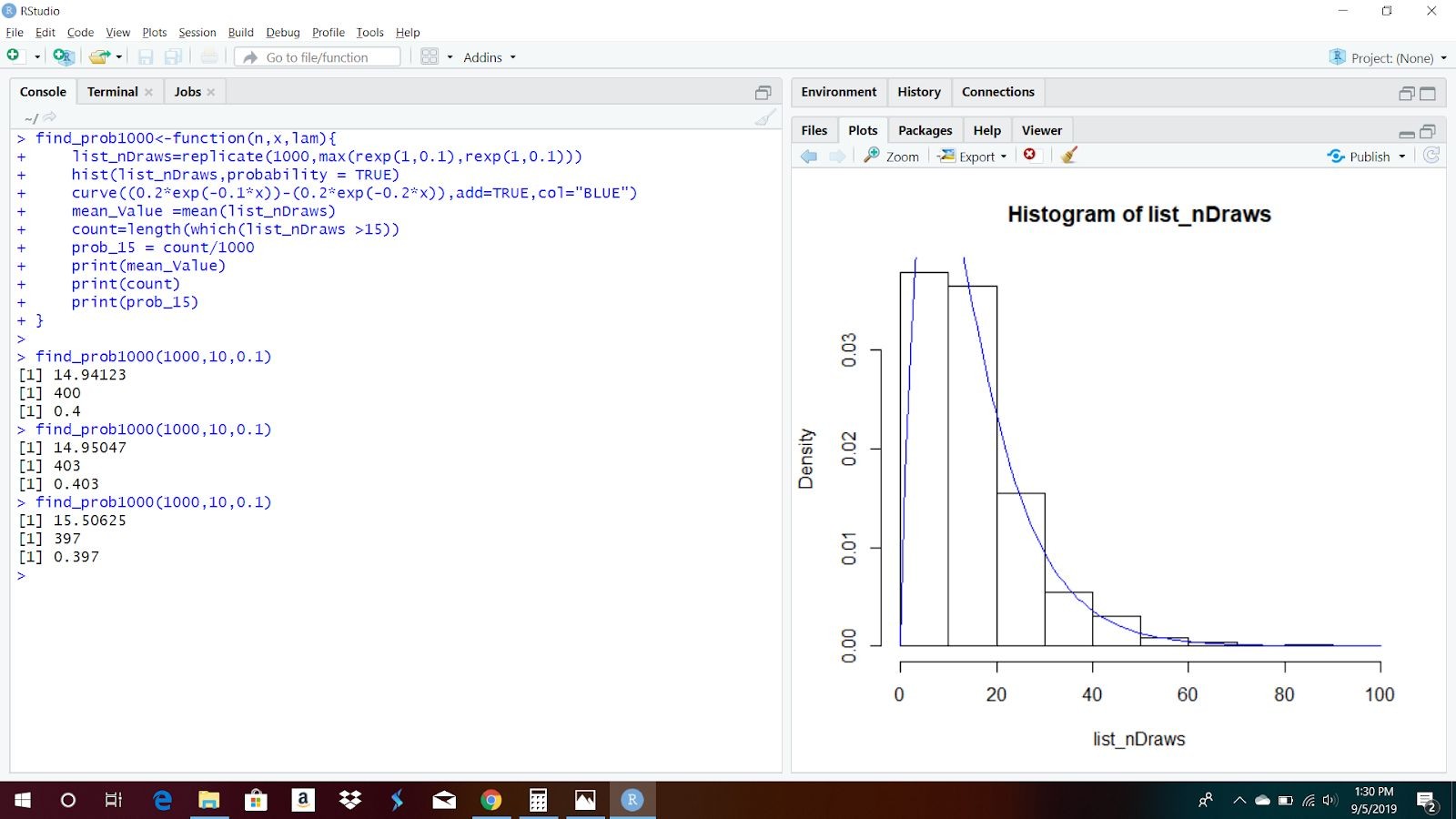
# Trial 1



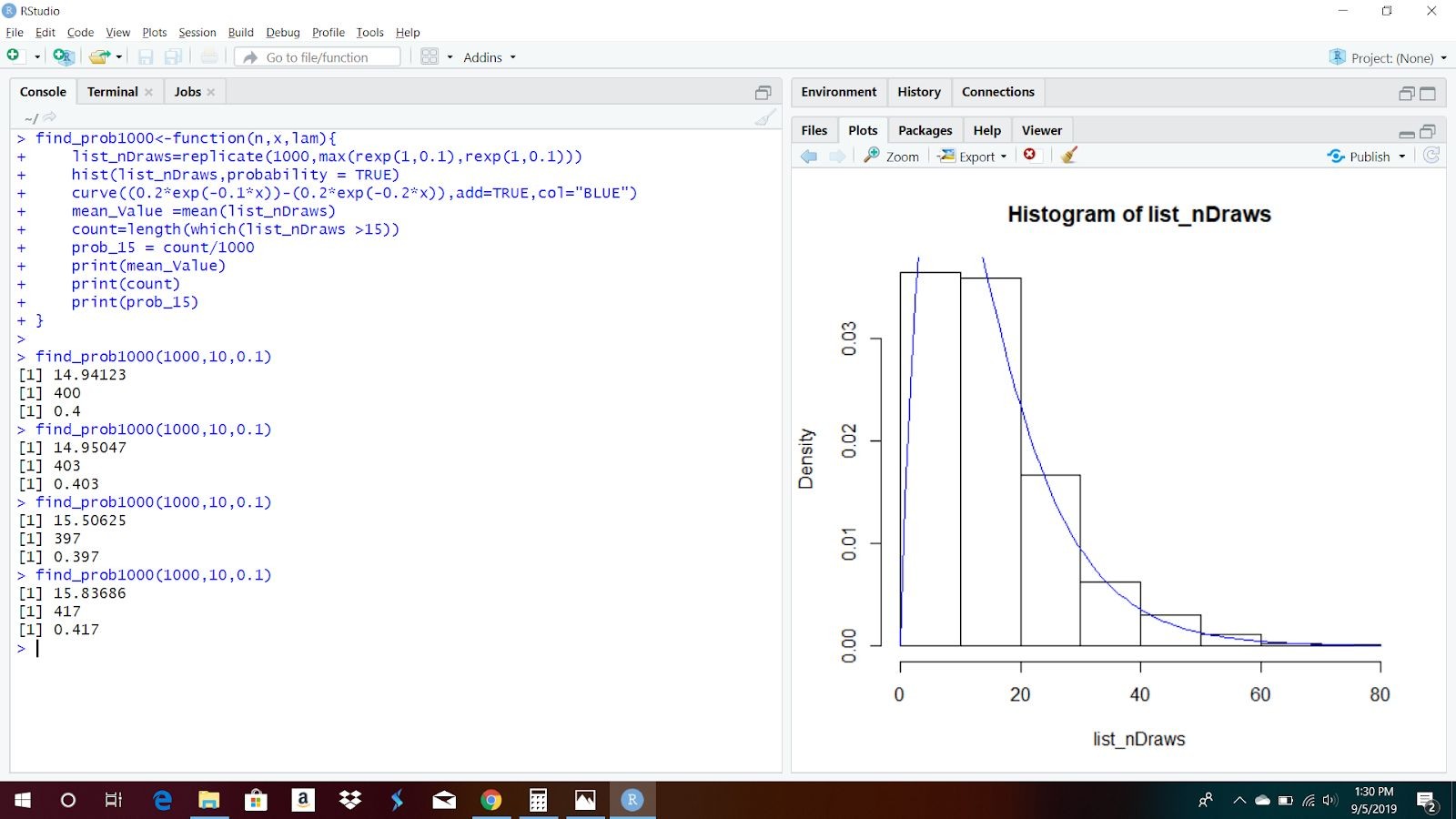
**Trial 2**



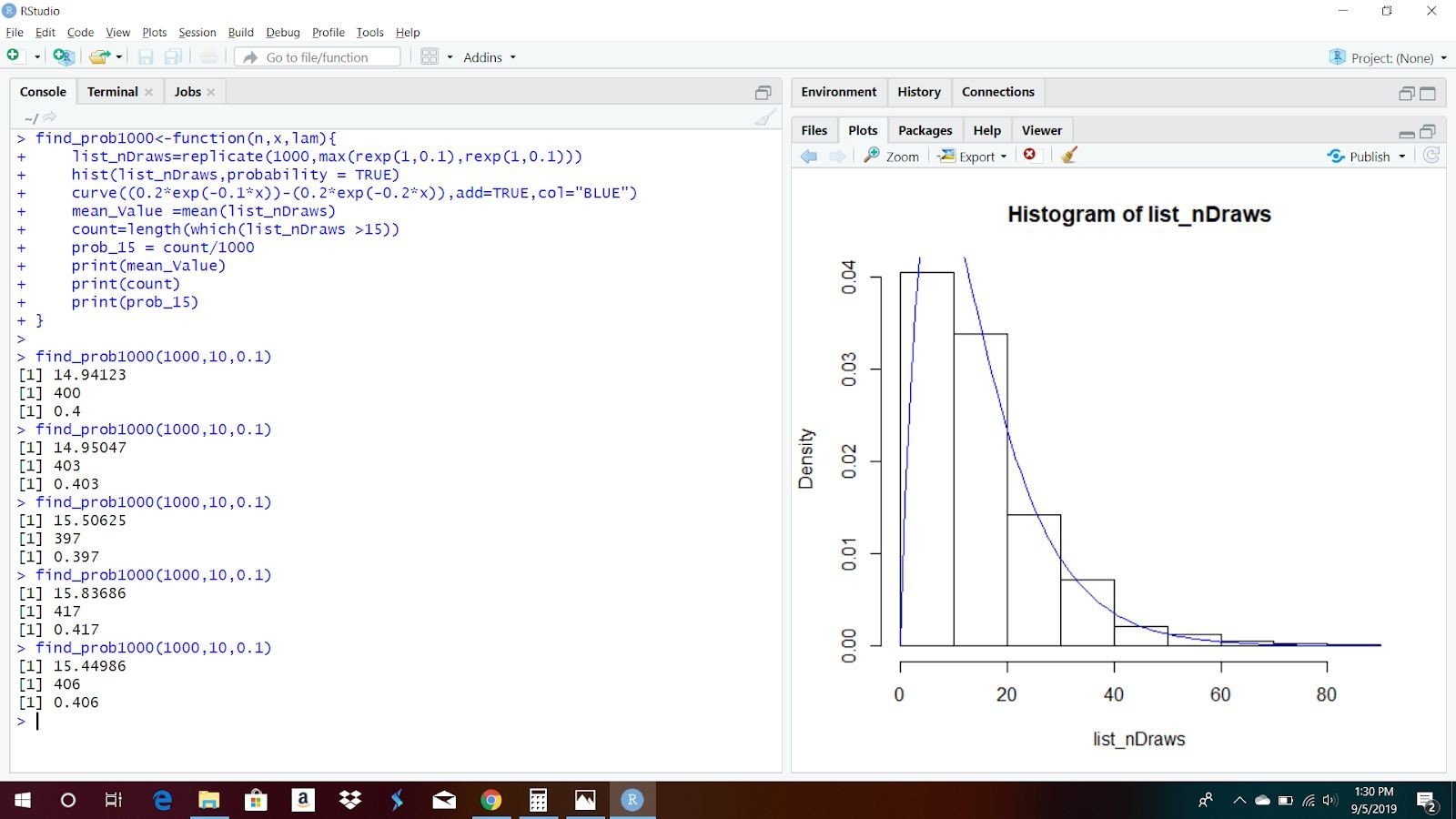
**Trial 3**



**Trial 4**



**Trial 5**

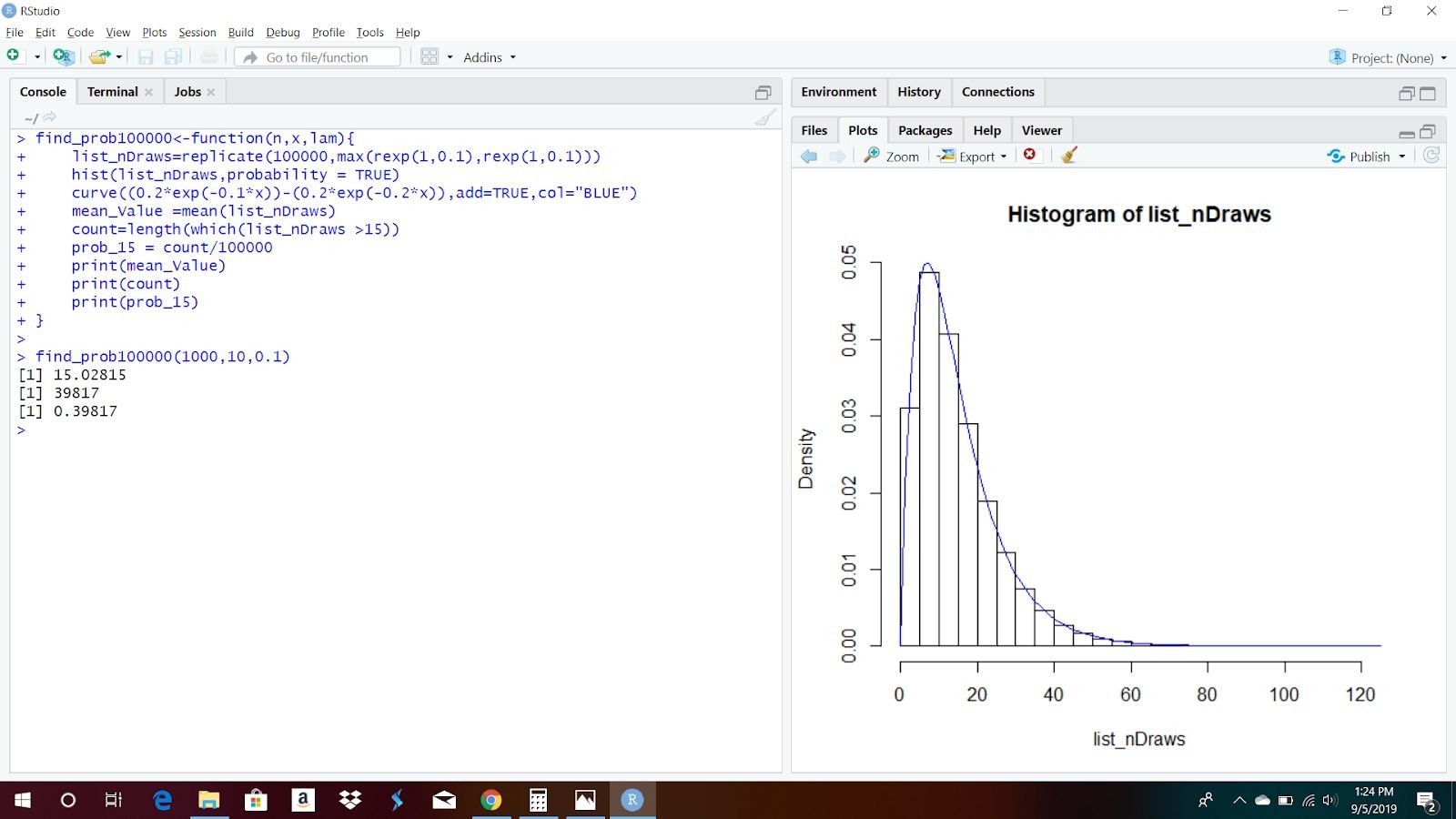


The table of results in the 5 trials are given below

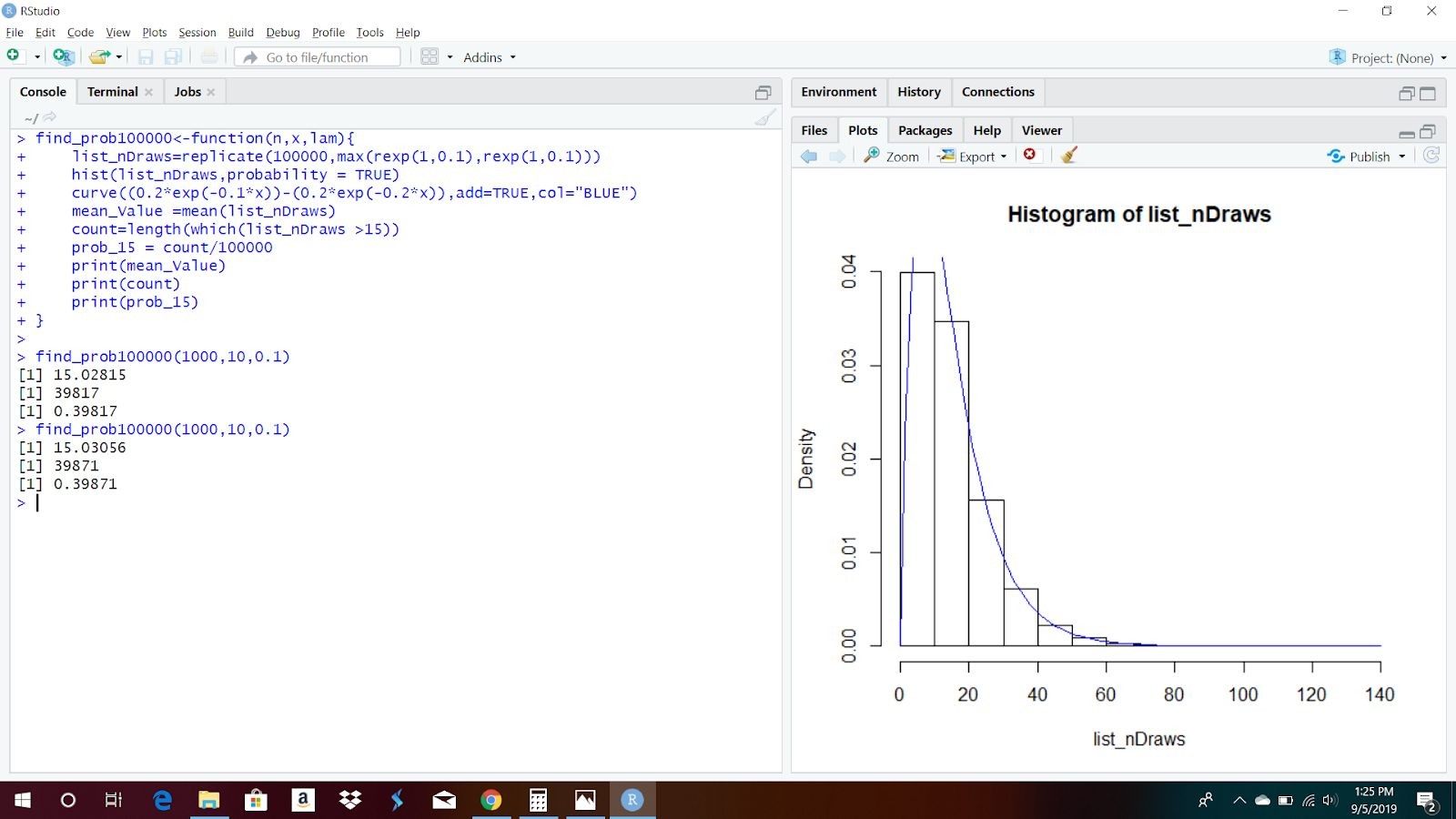
|  |  |  |  |
| --- | --- | --- | --- |
| **Sample Size** | **Mean Value** | **Count > 15** | **P ( T>15 )** |
| 1000 | 14.94123 | 400 | 0.4 |
| 1000 | 14.95047 | 403 | 0.403 |
| 1000 | 15.50625 | 397 | 0.397 |
| 1000 | 15.83686 | 417 | 0.417 |
| 1000 | 15.44986 | 406 | 0.406 |

Case 2 - When n = 100,000

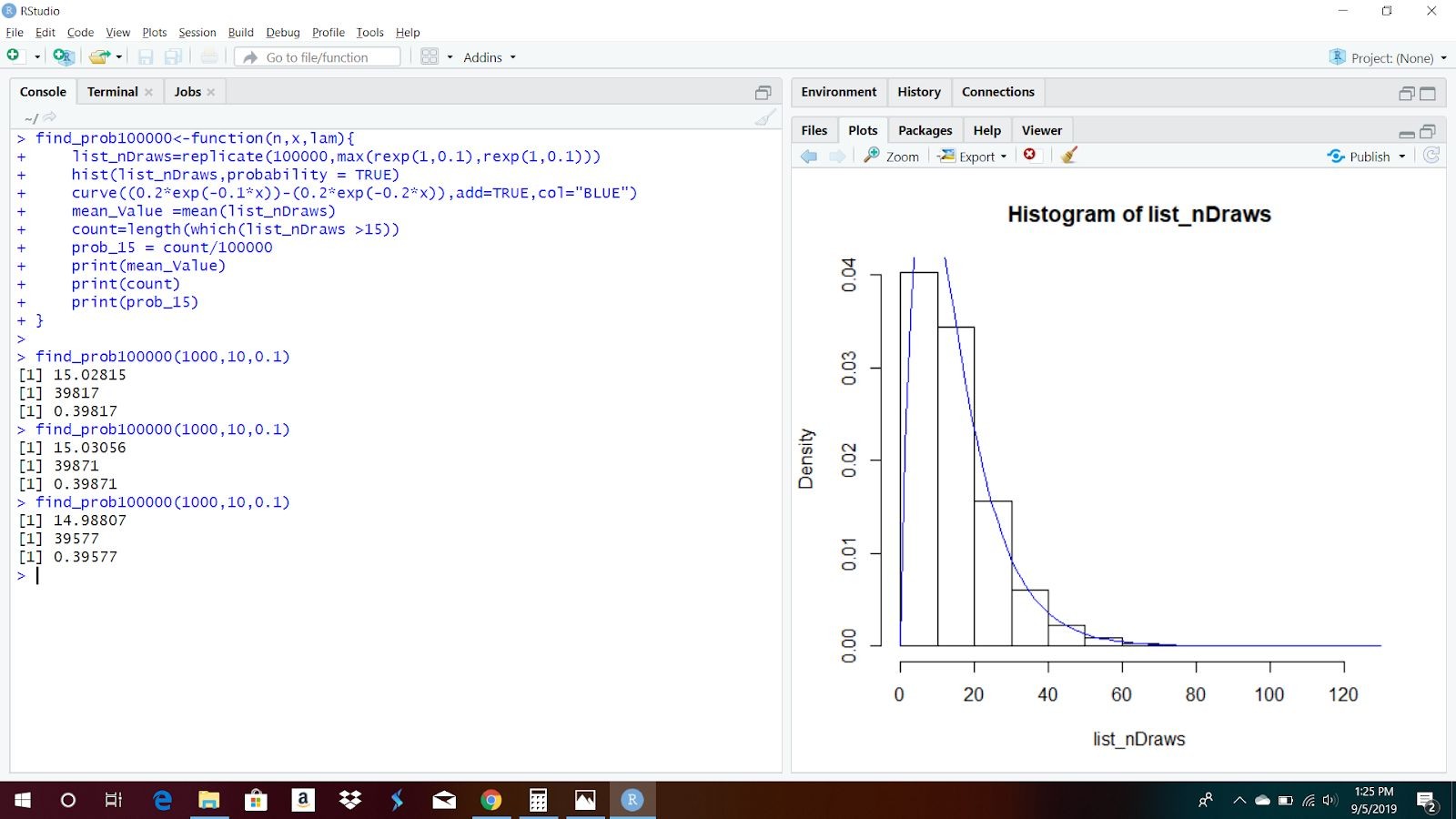
# Trial 1



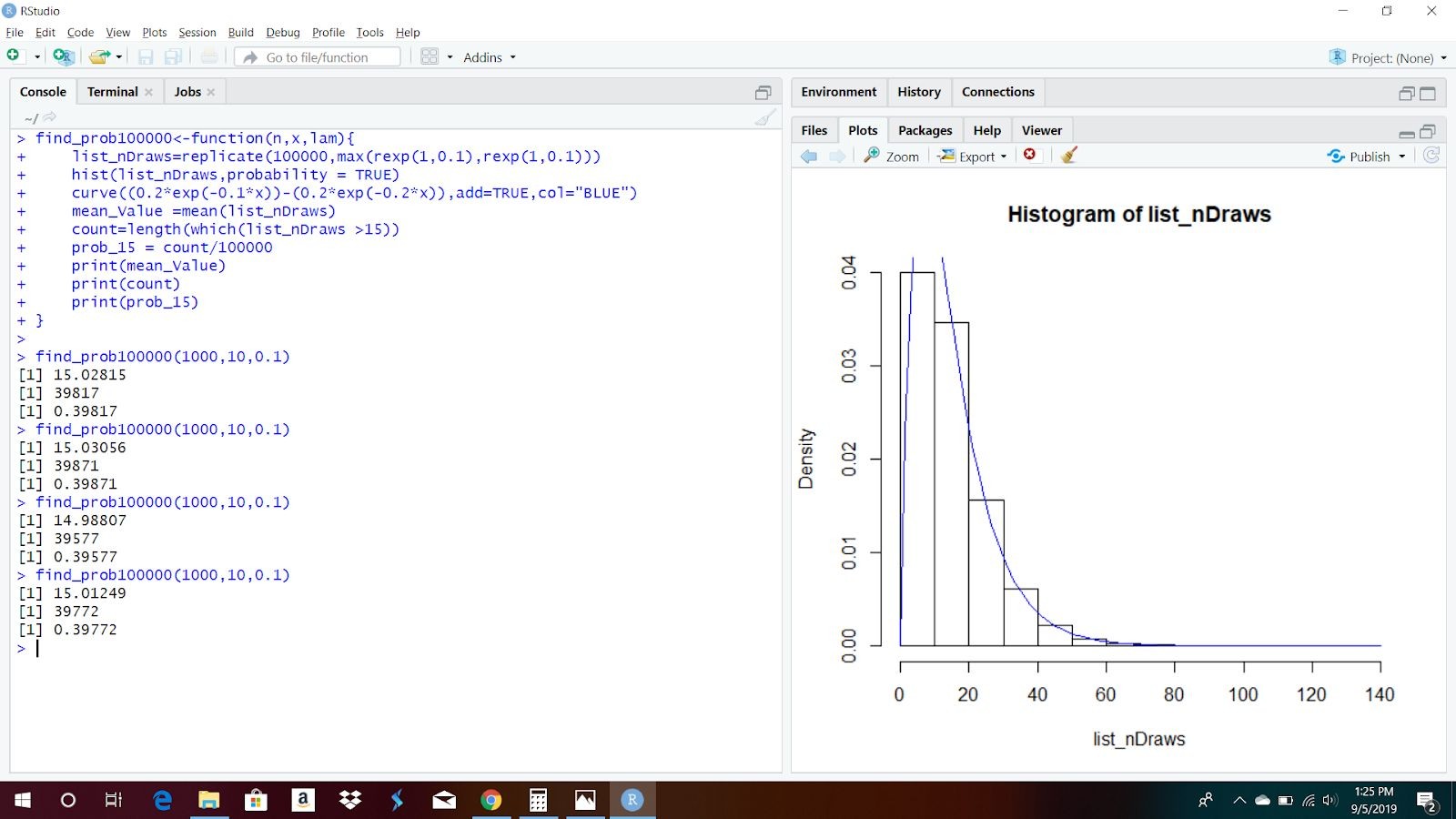
**Trial 2**



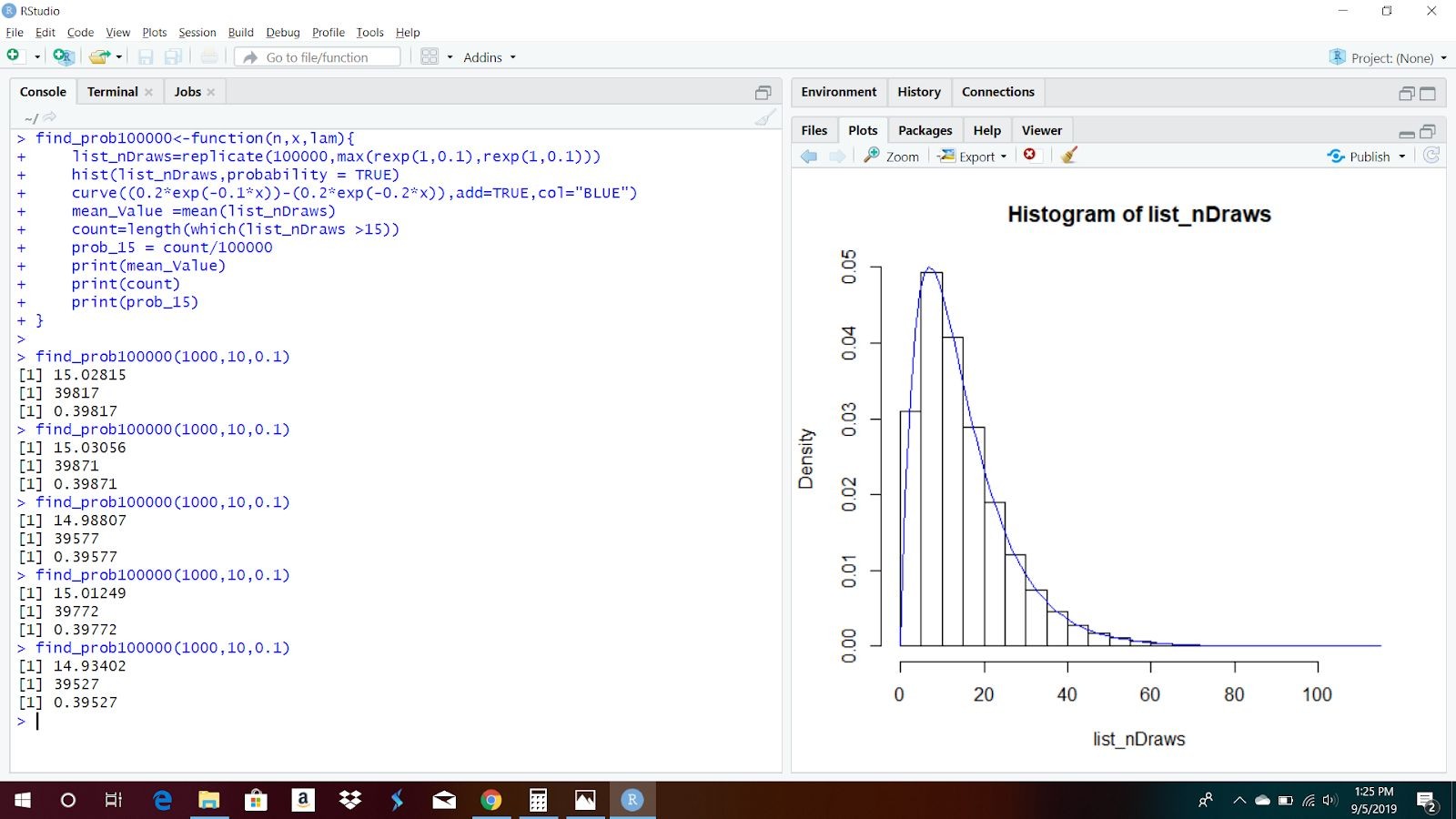
**Trial 3**



**Trial 4**



**Trial 5**



The table of results containing all the calculated values are as follows

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample Size** | **Mean Value** | **Count > 15** | **P ( T>15 )** |
| 1000 | 15.02815 | 39817 | 0.39817 |
| 1000 | 15.03056 | 39871 | 0.39871 |
| 1000 | 14.98807 | 39577 | 0.39577 |
| 1000 | 15.01249 | 39772 | 0.39772 |
| 1000 | 14.93402 | 39527 | 0.39527 |

# Observations:

The table shows that the observations in both the cases when there are 1000 and 100000 cases generated, the expected value and the probability of life of satellite greater than 15 years are varying by a maximum of +0.03 or -0.07 times the given values calculated above,

E ( T ) = 15 and P ( T > 15) = 0.4 (approx.)

2. (10 points) Use a Monte Carlo approach estimate the value of \_ based on 10; 000

replications.

**Solution 2.** Use a Monte Carlo approach to estimate the value of π based on 10,000 replications.

# Ans:

Method used :Consider a circle inscribed in a square. Let the side of square be 1 and radius of circle be 0.5. The value of π is estimated by finding the probability of a point within the circle and within the square.

P(X) = probability of a point within the circle & within the square P(X) = Area of circle / Area of square

P(X) = 𝝅r2 / s2

P(X) = 𝝅(0.5)2 / (1)2

P(X) = 𝝅\*0.25/1 = 𝝅/4 Therefore, P(X) = 𝝅/4

We can estimate the probability P(X) using R as follows:

* We use 2 random variables randX, randY which follow a standard uniform distribution ( uniform distribution can be used in any case). To generate a uniform random variable runif() is used.

*randX =runif(10000,min=0,max=1) randY =runif(10000,min=0,max=1)*

* If a point is inside the circle then it is inside the square too. So, we need to check if a point is inside the circle by using the formulae :

D = **√**(x-coordinate)2 + (y-coordinate)2

The D value should be less than the diameter of the circle ie 2\*0.5 = 1. So if the D value is less than 1 then the point is in the circle.

To do the above in R :

*randPoint=sqrt((randX ^2)+(randX ^2))*

* P(X) is calculated by the sum of all the points in randPoint which are less than 1 divided by the sample size ie 10000. To do this in R :

*in\_points=sum(randPoint<1) prob=iin\_pts/count print(prob)*

Thus P(X) is achieved. P(X) = 0.78690

* To estimate the 𝝅 value we need to calculate 4\*P(X):

= 4 \* 0.78690

=3.1476

Thus the estimated value of 𝝅 is 3.1476

This entire process is repeated 3 times to find the value of 𝝅 using the following function: findPi<-function(){

randX=runif(10000,min=0,max=1) randY=runif(10000,min=0,max=1) randPoint=sqrt((randA^2)+(randB^2)) *in\_points*=sum(randPoint<1) print(4\*in\_pts/10000)

}

|  |  |  |
| --- | --- | --- |
| **Trial No.** | **P(X)** | 𝝅 **value** |
| 1 | 0.78690 | 3.1476 |
| 2 | 0.78670 | 3.1468 |
| 3 | 0.78683 | 3.1475 |

*Thus, the estimated value of* 𝝅 **is 3.147**