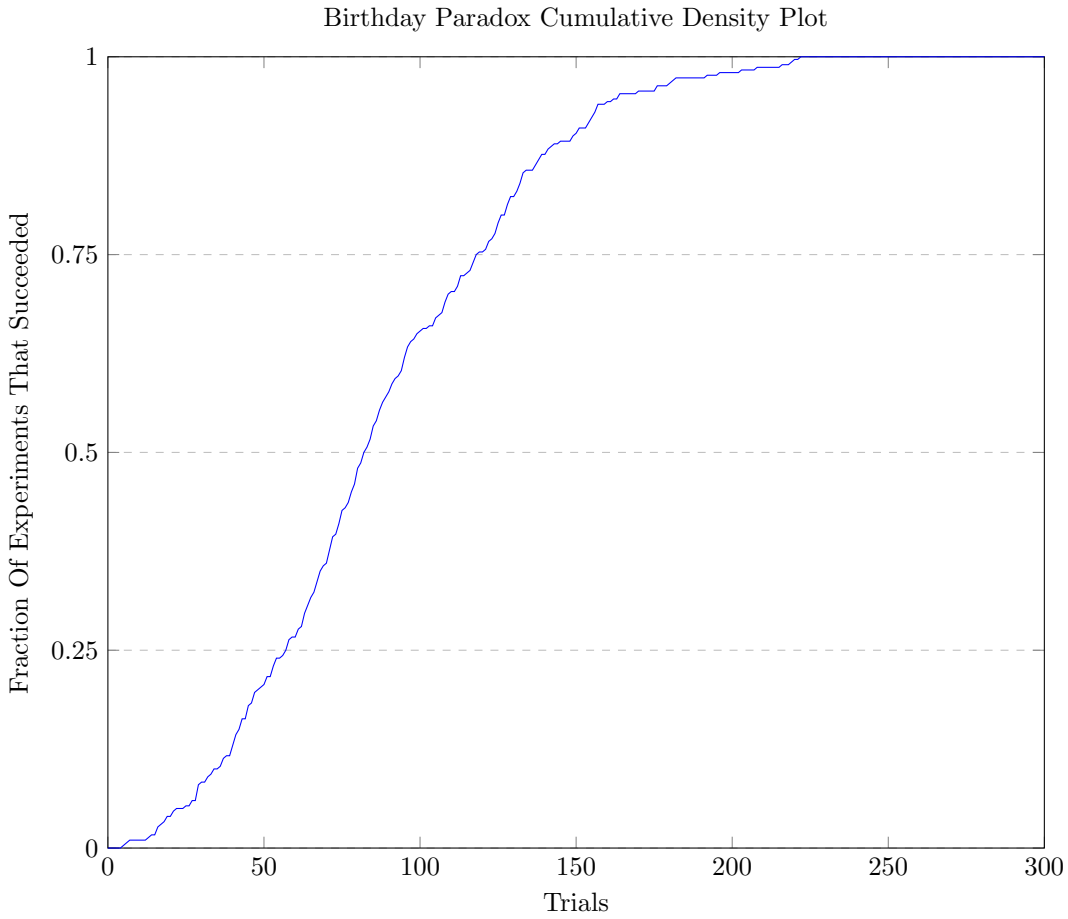
Birthday Paradox

A: (5 points) Question: Generate random numbers in the domain [n] until two have the same value. How many random trials did this take? We will use k to represent this value.

Answer: it took about 87 random trails to do the requested.

B: (10 points) Question: Repeat the experiment m = 300 times, and record for each time how many random trials this took. Plot this data as a cumulative density plot where the x-axis records the number of trials required k, and the y-axis records the fraction of experiments that succeeded (a collision) after k trials. The plot should show a curve that starts at a y value of 0, and increases as k increases, and eventually reaches a y value of 1.

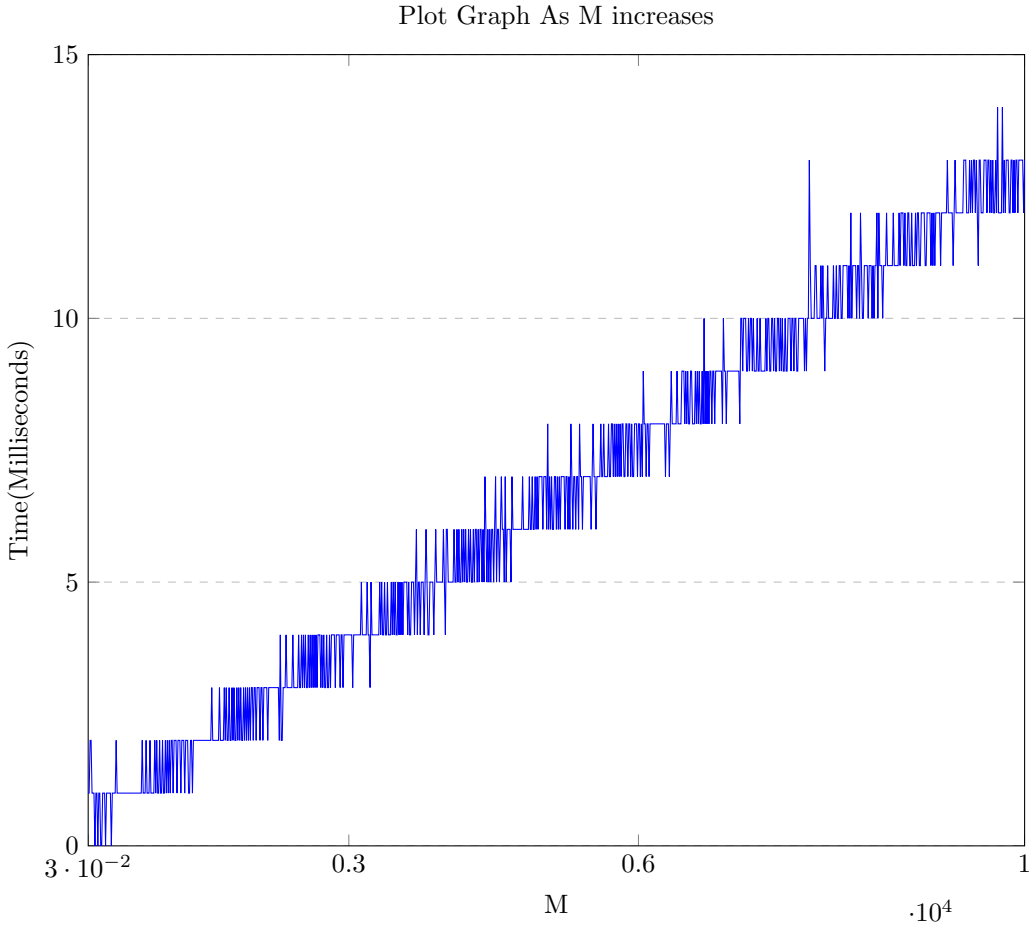


C: (10 points) Question: Empirically estimate the expected number of k random trials in order to have a collision. That is, add up all values k, and divide by m.

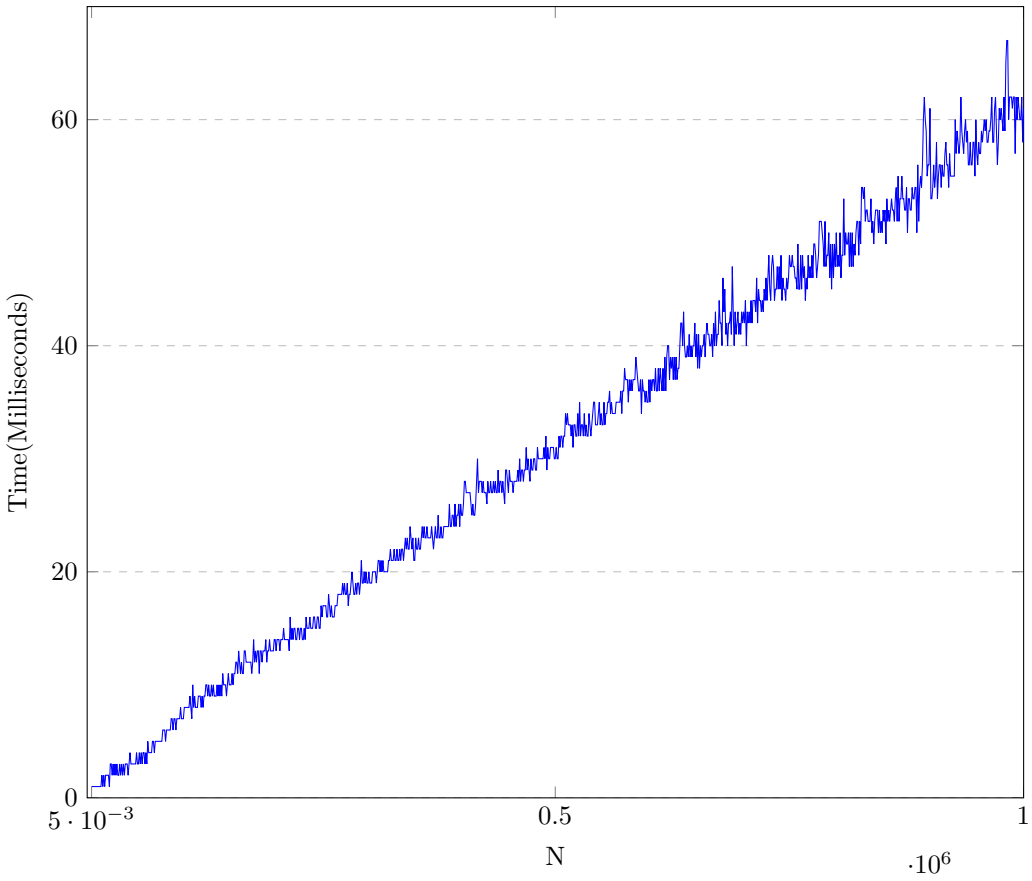
Answer: After performing the requested calculations I got 85 as the expected value

D: (10 points) Question: Describe how you implemented this experiment and how long it took for m = 300 trials. Show a plot of the run time as you gradually increase the parameters n and m. (For at least 3 fixed values of m between 300 and 10,000, plot the time as a function of n.) You should be able to reach values of n = 1,000,000 and m = 10,000.

Answer: Implementation was done in java. I created an array that holds the amount of each number was randomly selected. Once a number appeared twice we are done gathering the information. Next we organize the information.



Plot Graph As N increases

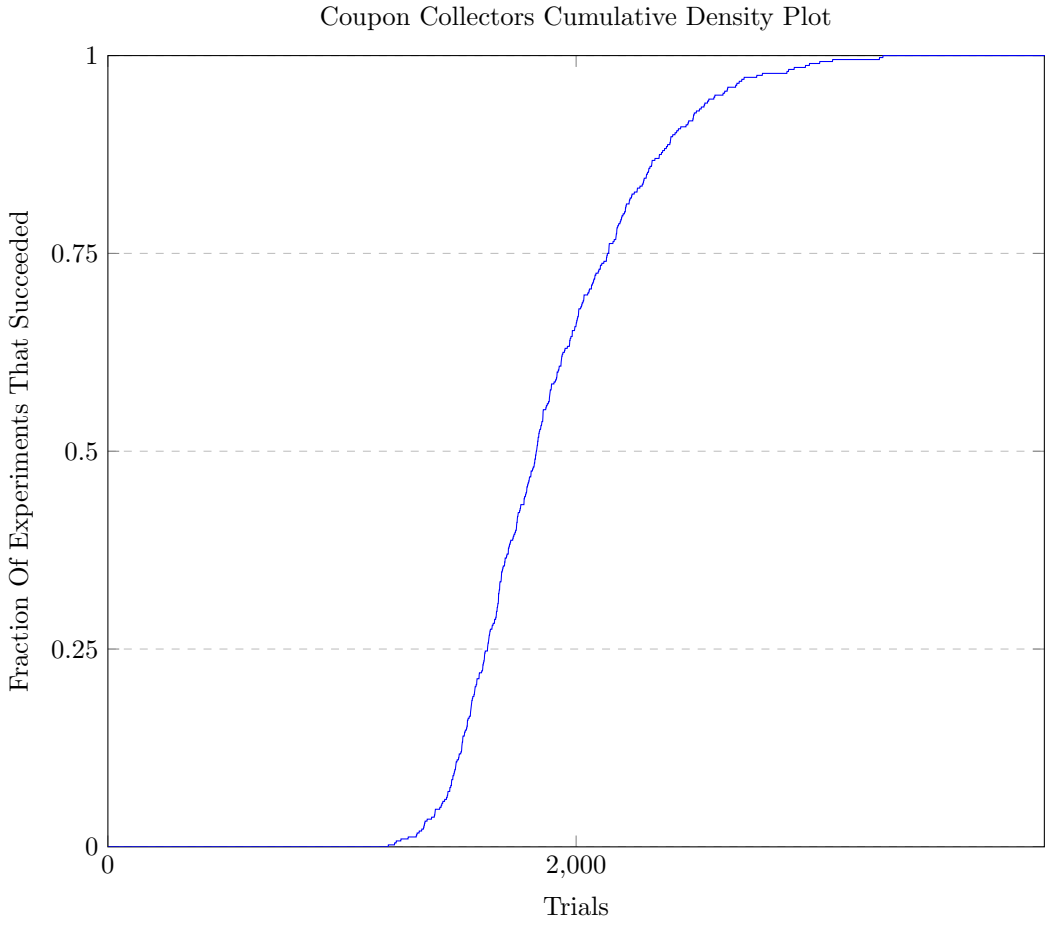


Coupon Collectors

A: (5 points) Question: Generate random numbers in the domain n until every value i exist in n has had one random number equal to i. How many random trials did this take? We will use k to represent this value.

Answer: it took about 1845 random trails to do the requested.

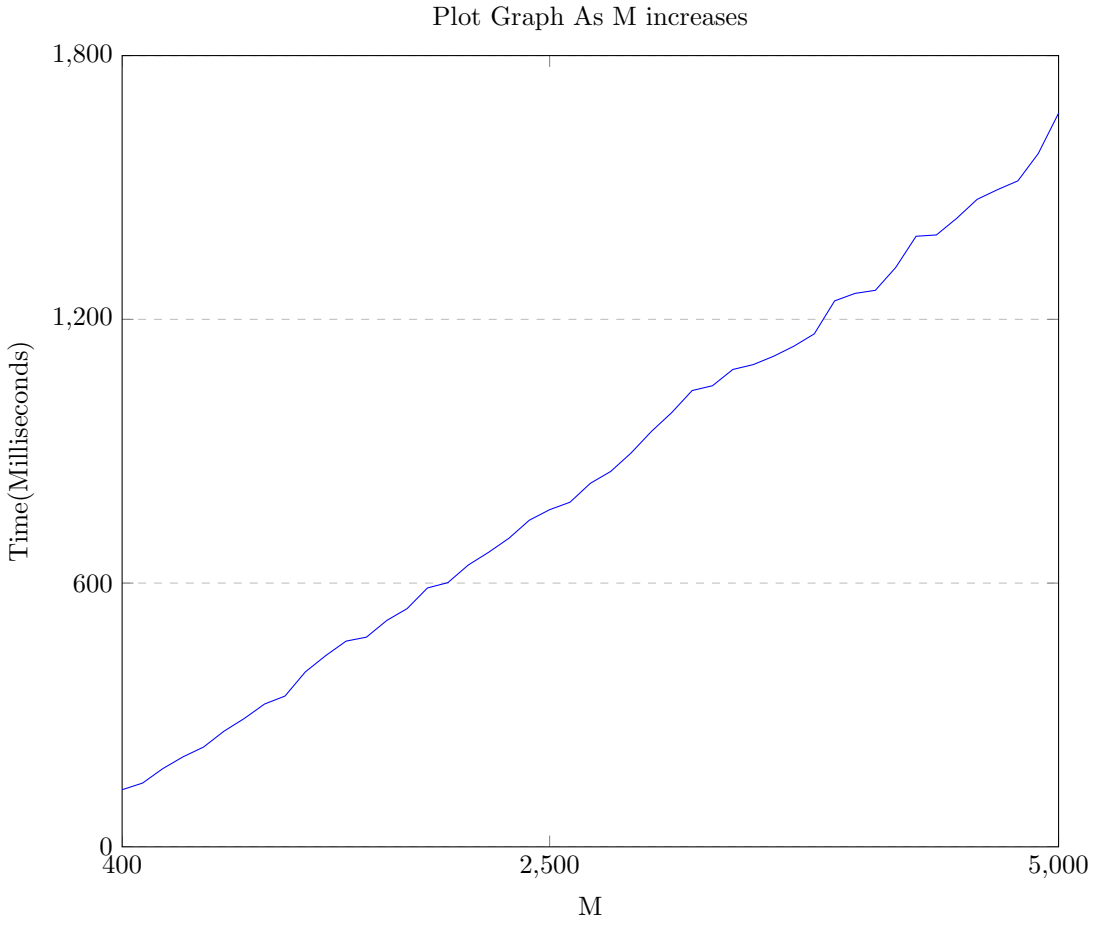
B: (10 points) Question: Repeat step A for m = 400 times, and for each repetition record the value k of how many random trials we required to collect all values i exist in n. Make a cumulative density plot as in 1.B.

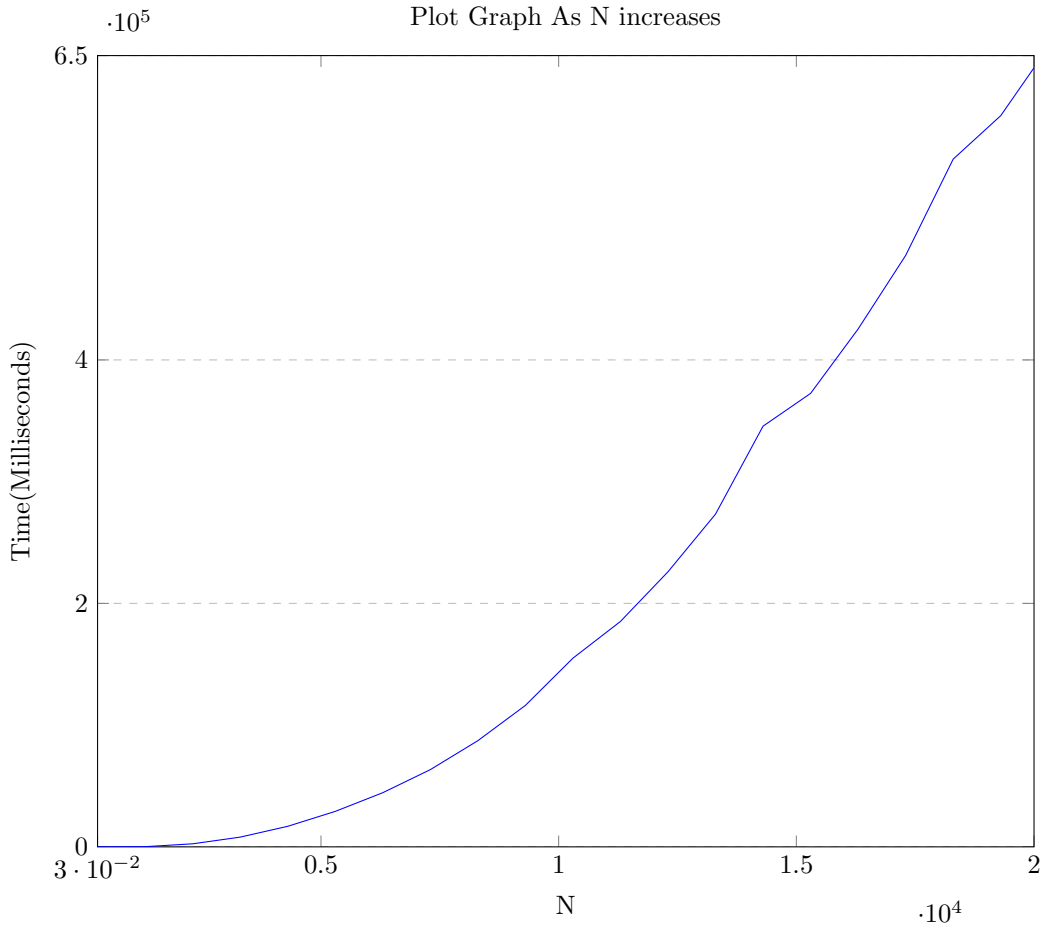


C: (10 points) Use the above results to calculate the empirical expected value of k.

Answer: After performing the requested calculations I got 1880 as the expected value

D: (10 points) Describe how you implemented this experiment and how long it took for n = 300 and m = 400 trials. Show a plot of the run time as you gradually increase the parameters n and m. (For at least 3 fixed values of m between 400 and 5,000, plot the time as a function of n.) You should be able to reach n = 20,000 and m = 5,000.


Plot Graph As M increases
1,800
1,200
600
0
Time(Milliseconds)
400
2,500
5,000
M



Comparing Experiments to Analysis

A: (15 points) Question: Calculate analytically (using formulas from the notes in L2) the number of random trials needed so there is a collision with probability at least 0.5 when the domain size is n = 5000. There are a few formulas stated with varying degree of accuracy, you may use any of these – the more accurate formula, the more sure you may be that your experimental part is verified, or is not (and thus you need to fix something). "Show your work, including describing which formula you used." How does this compare to your results from Q1.C?

1 - $\prod_{i=1}^{k-1}(\frac{n-i}{n}) = \frac{1}{2}, n = 5,000, solve for k$

To solve this I wrote a simple java program
for (int k = 2; k ¡ 100; k++)

```
int n = 5000;
double top1 = n - 1;
double mult = top1 / n;

for (int i = 2; i ¡ k; i++)
double top = n - i;
double fraction = top / n;
mult = mult * fraction;

double answer = 1 - mult;
System.out.println("Answer " + k + ": " + answer);
```

Output
84: 0.49559435841072497
85: 0.5039674920611069
86: 0.5123008381944804

$1 - \prod_{i=1}^{k-1}(\frac{n-i}{n}) = \frac{1}{2}, k \approx 85$

Answer: The expected value I calculated was 85, The better answer is 84, so my first calculation was pretty accurate

B: (15 points) Question: Calculate analytically (using formulas from the notes in L2) the expected number of ran- dom trials before all elements are witnessed in a domain of size n = 300? Again, there are a few formulas you may use – the more accurate, the more confidence you might have in your experimental part. "Show your work, including describing which formula you used." How does this compare to your results from Q2.C?

$k = T = nH_n \approx n(\gamma + ln(n)), \gamma = .577, n = 300$
$300(.577 + ln(300)) = 300(.577 + 5.704) = 1884.42$

Answer: The expected value I calculated was 1880, The better answer is 1884, so my first calculation was pretty accurate