Assignment 4 Report

Problem Statement:

Your task this time is to verify the *Birthday Problem* and the *Coupon Collector's* Problem. As you recall, these are both relevant to understanding how a hash table, especially one using *separate chaining*, will behave. But scientists, including engineers, shouldn't always just take things for granted. So, you will run experiments to validate the following two expressions:

$$C_1(m) \sim \sqrt{\pi m / 2}$$

$$B_0(m) \sim m \ln m$$

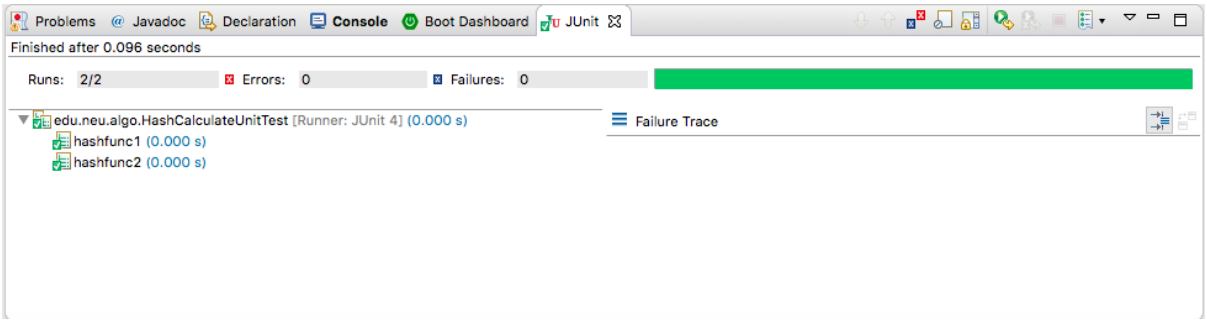
where

m is the number of bins/slots; and
$C_1(m)$ is the (average) number of hashes/throws before the first collision in encountered; and
$B_0(m)$ is the (average) number of hashes/throws before all bins/slots are filled (i.e. there are zero empty bins).

Analysis:

Hash function is calculated by random integer modulo bins/slot.

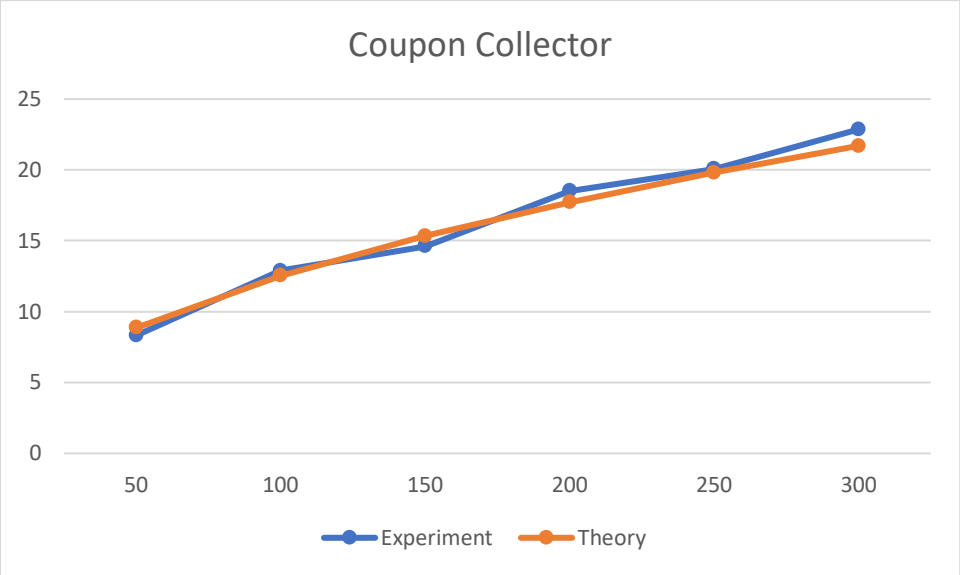
Snapshot of unit test case passed:



For Coupon

Number of run: 50

Coupon Collector bin/slot	Experiment	Theory
50	8.32	8.86
100	12.9	12.53
150	14.62	15.34
200	18.52	17.72
250	20.06	19.81
300	22.86	21.7



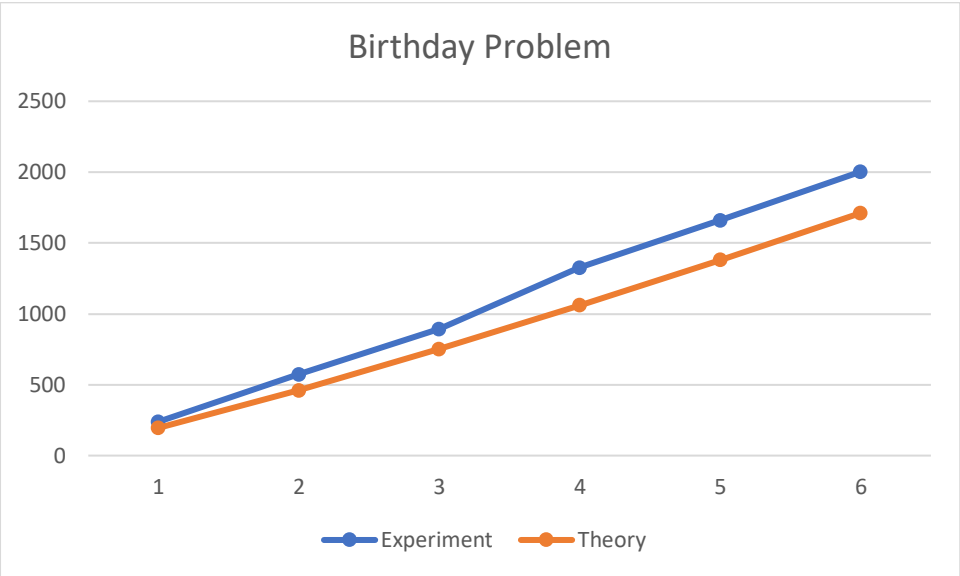
Conclusion

From above graph it is evident that experiment value also comes to ~ sqrt(PI * m / 2).

For Birthday

Number of run: 50

Birthday bin/slot	Experiment	Theory
50	239.32	195.6
100	572.9	460.51
150	893.62	751.59
200	1326.52	1059.66
250	1660.06	1380.36
300	2002.86	1711.13



Conclusion

From above graph, it is evident that experiment value also comes to ~ m*In(m).

Final Conclusion:

Experiment validate the two expressions provided for Coupon Collector and Birthday Problem.