D2 : Experiments and stacks

A)

N=1000:

```
1) How many cards in collection? 1000
   Total numbers generated in the process: 6731

2) How many cards in collection? 1000
   Total numbers generated in the process: 7246

3) How many cards in collection? 1000
   Total numbers generated in the process: 7508
```

N=10000:

```
1) How many cards in collection? 10000
   Total numbers generated in the process: 101029

2) How many cards in collection? 10000
   Total numbers generated in the process: 102684

3) How many cards in collection? 10000
   Total numbers generated in the process: 108012
```

B)

```
N=1000:
How many cards in collection? 1000
How many times to run experiment? 4

Test number 1
Total numbers generated in the process: 7527

Test number 2
Total numbers generated in the process: 9943

Test number 3
Total numbers generated in the process: 9997

Test number 4
Total numbers generated in the process: 7799

Mean: 8816.0
Standard deviation: 1157.6592762985144
```

```
N=10000:
How many cards in collection? 10000
How many times to run experiment? 4


Test number 1
```

```
Total numbers generated in the process: 91810

Test number 2
Total numbers generated in the process: 94179

Test number 3
Total numbers generated in the process: 92689

Test number 4
Total numbers generated in the process: 81698

Mean: 90094.0
Standard deviation: 4920.8414422738715
```

C)

N=1000, T=10 :

```
Mean: 7062.0
Standard deviation: 1610.9211960862642
```

N=1000, T=100 :

```
Mean: 7508.0
Standard deviation: 1207.2518254283154
```

N=1000, T=1000 :

```
Mean: 7509.0
Standard deviation: 1274.6372774244444
```

N=1000, T=10000 :

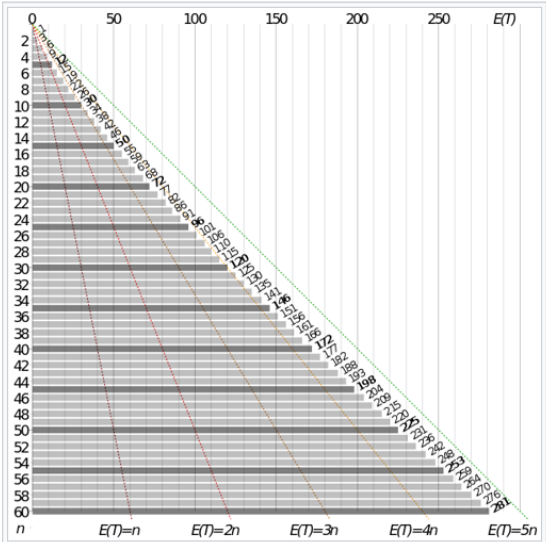
```
Mean: 7490.0
Standard deviation: 1279.8029371743135
```

Javadoc : Sjá html-skránna „CouponCollectorStats“

D)

Quote and picture 1 below was taken from the website source *https://en.wikipedia.org/wiki/Coupon_collector_%27s_problem* on the 27.August 2019:

> *„ The mathematical analysis of the problem reveals that the expected number of trials needed grows as $\Theta(n\log(n))$. For example, when n = 50 it takes about 225 trials on average to collect all 50 coupons.“*



Graph of number of coupons, *n* vs the expected number of tries (i.e., time) needed to collect them all, *E* (*T*)

Picture 1. Number of coupons (n) versus the expected number of tries

As shown on the graph here above, the more coupons that there is to be collected, the number of tries increases substantially for collection all the different coupons (as to be expected).

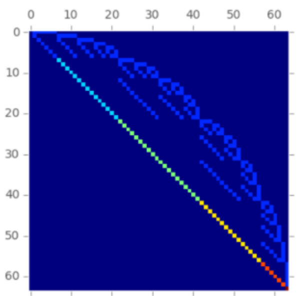
With tests run on the coupons to be collected N=1000 and the tests run for each sequence vary (10, 100, 1000, 10000), we can see that the results give a varying mean of results. This can be expected, as the coupons generated are all random, and therefore the mean of tests vary up and down randomly.

If we however look in greater details into the effect it has of changing the number of tests run (on the same sample size N=1000), we see that with increasing number of tests run (T), the standard deviation increases (shown in section C here above).

This gives us the information on how much the members of the group (the coupons to be collected) differ from the mean value. So with increasing number of tests, the difference between the mean value and the random coupons collected increases, even though the mean of tests does not increase or decrease in the same linear way.

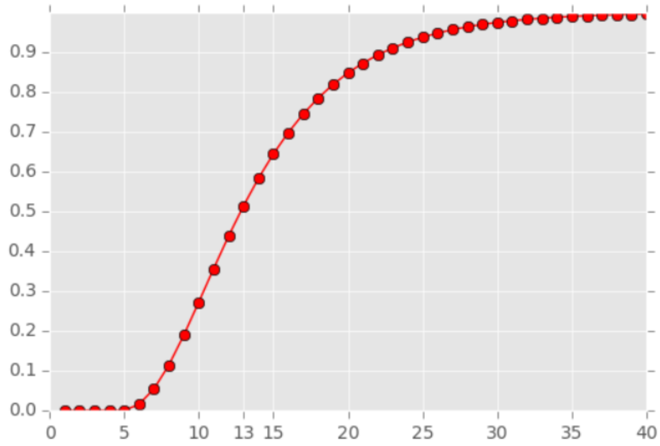
According to a study made on dice roles[1], which is similar to the coupon collector problem here, the experiment was to get every side of the dice facing upwards (at least) one time. When all sides have faced upwards (at least) once, the experiments stops and the number of throws of the dice is valuated.

This was done with various combinations of different lengths of dice roll sets, and set up transition probabilities based on all the possible previous state – see picture 2 below[1].



Picture 2. Results of the amount of dice roll sets versus transition probabilities

In this same experiment, the expected number of steps to the final state was also taken under consideration, as well as the variance in the number of steps. The expected number of dice rolls was 14,7 with a variance of about 39. Then, going from 1 roll to 40 rolls gives the following graph – see below picture 3[1].



Picture 3. Cumulative distribution function (CDF)

According to the study[1], the CDF on picture 3 above will go on to infinity, but was stopped at 40 as it was close enough to 1. What this tells us is, that for the first coupon (or first roll of the dice) the expected outcome of getting a value that we have not had before is equal to one. And for the last coupon (or last number on the dice), it could be a very long time before we finally get this last coupon (number).

The final conclusion of this particular study[1] showed that roughly 95% of the times the dice will be rolled 28 times or fewer. Witch gets to the conclusion that approximately 5% of the time, it will take 29 times or more to roll the dice and getting all six sides facing upwards.

The same conclusion can be made with the experiments done in part A to C in this assignment. As coupons get collected, the next coupon will get harder and harder to find. Therefore, the mean of the experiments made does not vary substantially (from 7062 tries to 7509 tries, when N=1000). Meaning that somewhere in between these two numbers is highly likely to find the expected number of tries it takes to collect all 1000 coupons.

1) https://phaethonprime.wordpress.com/2015/09/08/non-uniform-coupon-collectorsthe-problem/ used on August 28, 2019.