MA3K7 Week 13 Rubric

Meadow Thomas (2110165)

Aim: To investigate the process of randomly selecting two string ends, from a bowl of strings, and tying them together.

1 Entry

NOTATION Some notation I will be using throughout this investigation;

- N will denote the number of strings in the bowl.
- $2N$ will denote the number of string ends in the bowl.
- Loop size $= n$, is a loop created using n different strings. (A loop of size 1 is a loop formed by tying together two ends of the same string.)
- $P_N(allsize1)$, the probability all loops formed are of size 1 for a given N.
- $E_N(nloops)$, the expected number of loops of size n formed for N strings in the bowl.
- $P_N(singleloop)$, the probability of forming a single loop of size N.

I KNOW

- All strings are equal length.
- The probability of selecting any string end is equal.
- We can create loops of size ranging 1 to N.
- The maximum number of size 1 loops that can be formed is N, this occurs when all of the strings are tied to themselves.
- Continuing the process of tying two ends together N times results in all strings being part of a closed loop of some size.

Maximum and Minimum loops;

- The maximum number of loops we can make is N, this is done by each piece of string being tied to itself creating N loops of size 1.
- The minimum number of loops we can make is 1, this is done by each piece of string being ties to another creating a loop of size N.

Formula I may need;

- to calculate the expected value,

  $$\sum_{i=1}^{N} x_i p_i$$

  Where x_i is the outcome and p_i is the probability of that outcome.

ASSUMPTION Throughout the investigation some assumptions I have made include,

- All the strings are identical, same , size and texture.
- All the string ends have an equal probability of being selected, so the selection is completely random.

I WANT The questions I aim to answer by the end of this investigation;

- What is the probability of selecting two ends of the same string?/What is the probability of creating a loop of size 1?
- What is the expected number of loops formed for a given N?
- How will the number of strings in the bowl impact the expected number of loops formed?
- What does the distribution of loop sizes look like?

SPECIALISE Before we start the attack phase of this investigation we will look at the simplest case, N=1. This is where there is only one string in the bowl.

- $N = 1$, number of string ends is $2N = 2$.
- $P(allsize1) = 1$.
- $E_1(1loops) = 1$.

2 Attack

To begin we will try and simplify the investigation slightly by focussing on the expected number of loops of any size formed for a given N.

Throughout this part of this investigation we will focus on proving our first conjecture,
CONJECTURE As N increases, $E_N(loops)$ increases logarithmically.

I WANT We need to come with a formula for $E_N(loops)$.
STUCK Unsure of how to find this formula we start by investigating small values of N to see if a pattern arises.
N=1 was completed in the entry phase, where we found $E_1(loops) = 1$.
SPECIALISE N=2,
There are 4 string ends in the bowl.
Number the string ends,

$$1_a, 1_b, 2_a, 2_b$$

First, we start by selecting a string end at random from the bowl. It doesn't matter what string end we select first so, without loss of generality, assume we select string end 1_a.

- First pair Selection.

  To create a loop on the first pair selection we must now select the string end from the same string as 1_a, that being 1_b. Resulting in a loop of size 1 being formed.

  The probability of creating a loop on the first pair selection is,

  $$\frac{1}{3}$$

  And the expected number of loops formed on the first pairing is,

  $$(\frac{2}{3} \times 0) + (\frac{1}{3} \times 1) = \frac{1}{3}$$

- Second pair selection.

STUCK Is the probability of creating a loop on the second pairing dependent on the outcome of the first pairing?

After the first pairing we will either be left with 2 string ends from the same string or 2 string ends from different strings remaining in the bowl.

AHA! However, both of these pairings will result in the formation of a loop.

The first case is obvious. For the second case this would mean that we tied the two strings in the bowl together creating a single string of size 2, so we are actually left with 2 string ends of the same string its just a longer string.

Hence, the probability of creating a loop on the second pair selection is,

$$1$$

And the expected number of loops formed on the second pairing is,

$$(0 \times 0) + (1 \times 1) = 1$$

Therefore, the probability of creating 2 loops for N=2 is,

$$\frac{1}{3} \times 1 = \frac{1}{3}$$

And the expected number of loops formed for N=2 is,

$$E_2(loops) = \frac{1}{3} + 1 = \frac{4}{3}$$

STUCK Can we generalise this case to get a formula for the expected number of loops formed for any value of N?

Let N be the number of strings in the bowl.
First, we select a string end at random.
After this selection there are $2N - 1$ string ends remaining in the bowl, of which only 1 will result in the formation of a loop, that being the other end of the same string, as we saw in the case for N=2.
Thus the probability of creating a loop on the first pair selection is,

$$\frac{1}{2N-1}.$$

And the Expected number of loops on the first pairing can be calculated by,

$$(0 \times \frac{2N-2}{2N-1}) + (1 \times \frac{1}{2N-1}) = \frac{1}{2N-1}$$

There is now 2 cases, the same as we saw for N=2,

- Case 1

  We have created a loop.

  Through creating the loop we have effectively removed a string from the bowl since its ends can no longer be picked.

  Hence, there are $N - 1$ strings remaining in the bowl.

- Case 2

  We haven't created a loop.

  In this case we have instead combined 2 strings into 1, since the new longer string we created took 4 string ends and changed them in to 2 string ends.

  This is equivalent to removing a string.

AHA! Hence, we get the same result as the first case. There are $N-1$ strings remaining in the bowl.

AHA! We have shown that each pairing is independent of the previous pairing for all values of N. So, if we continue in this way, on the second pairing there are $N-1$ strings in the bag and so by the same logic,
the probability of creating a loop on the second pair selection is,

$$\frac{1}{2(N-1)-1} = \frac{1}{2N-3}.$$

And the Expected number loops formed on the second pairing is,

$$(0 \times \frac{2N-4}{2N-3}) + (1 \times \frac{1}{2N-3}) = \frac{1}{2N-3}$$

We can repeat this process until we get to the last pair selection, where there are only 2 string ends remaining in the bowl.

STUCK Will the last 2 string ends always result in a loop?
As we saw for N=2 the string ends are guaranteed to make a loop. We can generalise this for all N, since by tying these ends together we have removed all string ends from the bowl and so there can only be loops remaining. Hence the probability of creating a loop on the last pairing will always be 1.

AHA! Therefore, we get a formula for the expected number of loops of any size formed for a given N,

$$E_N(loops) = \sum_{r=1}^{N} \frac{1}{2r-1}$$

Following from this I used python to plot a graph of the Expected number of loops formed against N, the number of strings in the bowl.

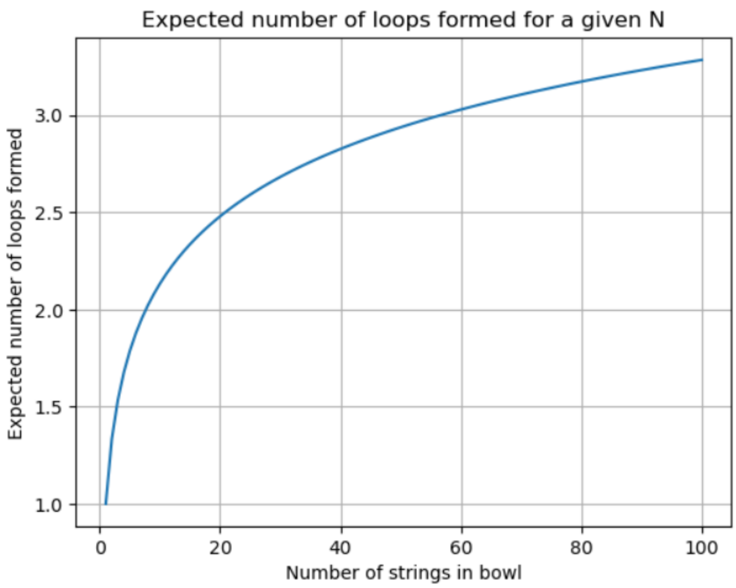

Figure 1: Expected number of loops formed for a given N.

From this graph and from our formula for $E_N(loops)$ we can see that our first conjecture is true. As N increases the expected number of loops formed increases logarithmically.

We will now investigate further in to the specific sizes of the loops formed.

We will look at the formation of Size 1 loops and size N loops in particular.

I WANT We will aim to answer these questions;

- What is the probability that all loops formed are size 1 for a given N?
- What is the probability of creating a single loop (a size N loop)?

From these questions we get our second conjecture,

CONJECTURE $P_N(allsize1) \to 0$ as $N \to \infty$

To answer these questions we will again start by looking at small values of N.
We have already seen the answer to the first question for N=1 and N=2, since if we create 2 loops for N=2 they must both be size 1 loops.

SPECIALISE Lets now look at N=3,
There are now 6 string ends in the bowl and we will label them in the same way as before.

$$1_a, 1_b,\ 2_a,\ 2_b,\ 3_a, 3_b$$

Without loss of generality suppose we select 1_a as our first string end.
So to create a loop of size 1 on the first pairing we must tie it to 1_b the probability of this happening is 1/5.

$$1_a, {\color{red}1_b},\ 2_a,\ 2_b,\ 3_a, 3_b$$

Suppose that we have successfully created a size 1 loop using 1_a and 1_b, we are now left with 4 string ends and we are in the exact situation for $N = 2$.
We know that the probability of all loops formed being of size 1 for N=2, that is $\frac{1}{3}$.

AHA! So we can calculate the probability that all loops formed are size 1 for N=3 by,

$$P_3 = (allsize1) = \frac{1}{5} \times \frac{1}{3} = \frac{1}{15}$$

N	$P_N(allsize1)$
1	1
2	$\frac{1}{3}$
3	$\frac{1}{15}$

STUCK How can we extend this to find a formula for all values of N?

If we want all loops being formed to be size 1 we must have a loop being formed on every pairing. From this we can see that to calculate the probability of all loops formed being size 1, we must multiply the probability's of forming a size 1 loop on every pairing.

AHA! Which gives us the formula,

$${\color{red}P_N(allsize1) = \prod_{r=1}^{N} \frac{1}{2N-1}}$$

I then used python to produce the probabilities for the first 10 values of N and plotted a graph for $N = 1, .., 10$.
From our equation and the graph we see that our conjecture was correct and

$$P_N(allsize1) \to 0 \text{ as } N \to \infty$$

Figure 2 shows the printed values for x=N and y=$P_N(allsize1)$

```
x = 1, y = 1
x = 2, y = 1/3
x = 3, y = 1/15
x = 4, y = 1/105
x = 5, y = 1/945
x = 6, y = 1/10395
x = 7, y = 1/135135
x = 8, y = 1/2027025
x = 9, y = 1/34459425
x = 10, y = 1/654729075
```

Figure 2: First 10 values of N

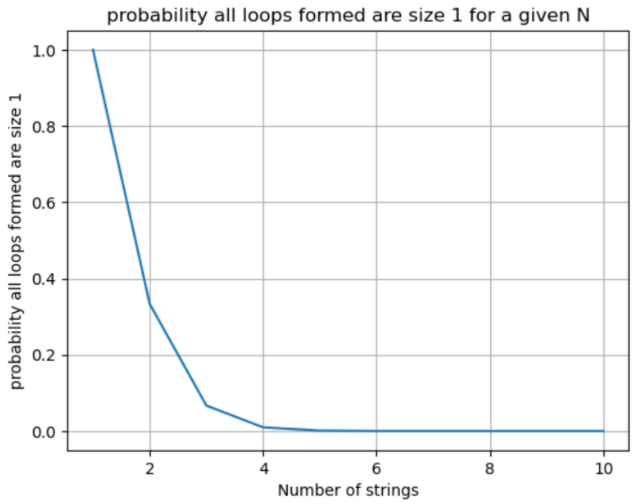

Figure 3: Probability that all loops are of size 1.

We will now look at the probability of forming a single loop for a given N, that is the probability of creating a size N loop.

This leads to our third Conjecture,

CONJECTURE The probability of forming a single loop, of size N, will decrease exponentially as N increases.

SPECIALISE For N=1,

$$P_1(singleloop) = 1$$

Because we cannot form more than one loop as there are only 2 string ends.

SPECIALISE For N=2, As before we will label our string ends,

$$1_a, 1_b,\ 2_a,\ 2_b$$

Suppose without loss of generality that we select 1_a.

- First pair selection $\rightarrow$ In order to create a single loop we do not want to create a loop on the first pairing. Therefore, we cannot select 1_b as this will create a size 1 loop. Hence, we can select either 2_a or 2_b as this will create a string of length 2 instead of a loop.

  Therefore, for the first pair selection the probability of not creating a loop is $\frac{2}{3}$.

- Second pair selection $\rightarrow$ After this we are left with 2 string ends, these will be 1_b and either 2_a or 2_b depending on the first pairing.

  Either of these pairings will result in a loop of size 2 being formed.

  This is because for N=2 we can either have 2 loops of size 1 or 1 loop of size 2. Since we have made it so the first pairing is not a size 1 loop, it must mean our pairings will result in a loop of size 2.

Therefore,

$$P_2(singleloop) = 2/3$$

From this case we can see,

$$P_2(singleloop) = 1 - \text{probability of forming a loop at the second pairing.}$$
$$P_2(singleloop) = 1 - \frac{1}{3}$$

Stuck Is it true that

$$P_N(singleloop) = 1 - \text{probability of forming a loop at Nth pairing ?}$$

To see if this is true for all N we will look at one more case,

Specialise let N=3, As before we label our string ends,

$$1_a, 1_b,\ 2_a,\ 2_b,\ 3_a, 3_b$$

Assume without loss of generality that we select 1_a,

- First pair selection $\rightarrow$ We cannot select 1_b since this will create a loop, so the probability of not creating a loop is $\frac{4}{5}$.

  without loss of generality assume our first pair is 1_a and 2_a. It doesn't matter which of the 4 other string ends we choose since in all cases we create a new string of length 2 and a string of length 1.

- Second pair selection $\rightarrow$ Now we are in the case of N=2, as we have 2 strings. We know the probability of creating a single loop to be $\frac{2}{3}$.

In order for us to create a single loop for N=3, we must not make a loop on the first pair selection, not make a loop on the second pair selection and then finally make a loop on the third pair selection.
Therefore,

$$P_3(singleloop) = \tfrac{4}{5} \times \tfrac{2}{3} \times 1 = \tfrac{8}{15}$$

Stuck But $\frac{8}{15} \neq 1 - \frac{1}{15}$
So, it is not true that

$$P_N(singleloop) = 1 - \text{probability of forming a loop at Nth pairing?}$$

Why is this not true?

Aha! This is because calculating the probability of forming a single loop is dependent on all the pairings of string ends, bar the very last, not forming a loop. To create a single loop we require a certain outcome at each of the previous pairings, whereas to create a loop with a string end pairing it doesn't matter what the previous out come was.
In other words we want to make a string of size N and then on the last pairing it will become a loop of size N.

The probability of not creating a loop at each pairing can be calculated by,

$$\tfrac{2N-2}{2N-1}$$

Therefore, we get an equation to calculate the probability of creating a single loop for N strings in the bowl,

$$P_N(singleloop) = \textstyle\prod_{r=1}^{N} \tfrac{2N-2}{2N-1}$$

Stuck This equation doesn't work! when using python to plot the graph of probability of forming a single loop against N we get the graph in figure 4.

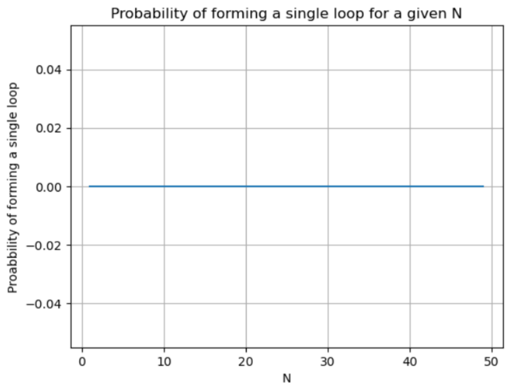

Figure 4: Probability of forming a single loop of size N.

STUCK Why does our equation not work?

AHA The reason is because forming a loop and forming a single loop for N=1 is the exact same thing, so when we subtract the probabilities we get 0. Therefore, when we substitute 1 into our equation we get 0, which is then multiplying everything in the equation.
To solve this issue we can start at $r = 2$ as we know that for $N = 1$ the $P_1(singleloop) = 1$ and since we are multiplying the probabilities it doesn't make a difference to just remove this case. Therefore, we get an equation to calculate the probability of creating a single loop for N strings in the bowl,

$$P_N(singleloop) = \prod_{r=1}^{N} \frac{2N-2}{2N-1}$$

Following from this I used python to plot a graph for the probability of forming a single loop of size N using the correct formula and this is shown in figure 5.

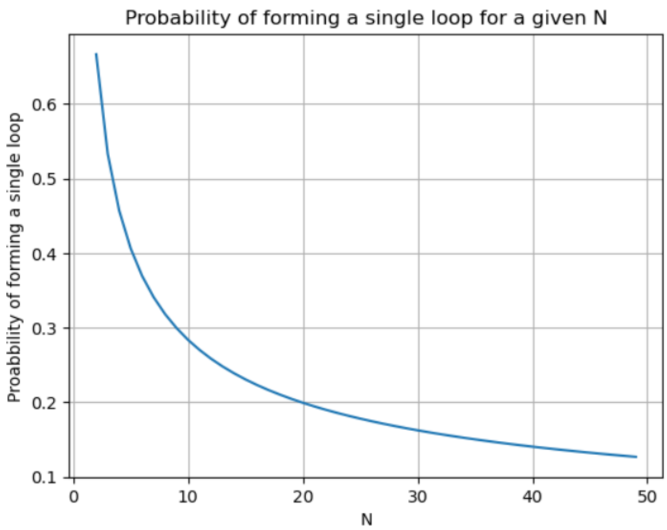

Figure 5: Probability of forming a single loop of size N.

AHA! From the graph and our equation we see that our conjecture was correct and the probability of forming a single loop of size N decreases exponentially as N increases.

3 Review

Check

I have investigated the outcomes of randomly selecting 2 string ends from a bowl of N strings and tying them together. Throughout the investigation I have done all calculations by hand and checked using python. As well as using python to provide observational evidence for my conjectures by plotting the graphs for the equations.
One assumption that I made for this investigation was that the probability of selecting any string end was the same. By making this assumption it allowed me to calculate the probability of creating particular loop sizes, as well as the expected number of loops formed. Changing this assumption is something that I will explore in the extend section of this investigation.
Some consequences of my conjectures include,

- As the number of strings in the bowl increases the expected number of loops formed increases logarithmically.
- The probability that all of the loops formed are size one tends to zero as the number of strings in the bowl tends to infinity.
- The probability of forming a single loop, of size equal to the number of strings in the bowl, decreases exponentially as the number of strings in the bowl increases.

Through proving these conjectures I believe I have answered the question. I have investigated mainly into the number of strings formed as well as the sizes of the strings formed.

Reflect

Throughout this investigation I have often tried to simplify the question by looking closely at specific cases, such as, N=2 or N=3. This helped to guide my investigation when i was stuck and didn't know where to start. This method was also extremely useful when trying to come up with equations. Working with the smaller values of N meant that I could carry out all of the cases by hand, allowing me to spot patterns in the results, which I then extended my to larger values of N.
Simplifying the question in this way highlighted many crucial ideas, such as,

- each pairing being independent of the previous pairing, when calculating the expected number of loops formed for a given N.
- To calculate the probability that all loops formed are size 1, we must multiply the probabilities of forming a loop at each pairing.

This investigation was very open ended and there are many different ways to take it. By breaking down the investigation into smaller more manageable questions it became a lot easier to work through.
One of the main ideas of this investigation was trying to understand what happened as the number of strings in the bowl increased. All of my conjectures are linked to the idea of hat impact the number of strings in the bowl increasing has on my results.

To answer the questions I asked in my entry phase;

- What is the probability of selecting two ends of the same string?/What is the probability of creating a loop of size 1?

$$P_N(singleloop) = \prod_{r=1}^{N} \frac{2N-2}{2N-1}$$

- What is the expected number of loops formed for a given N?

$$E_N(loops) = \sum_{r=1}^{N} \frac{1}{2r-1}$$

- How will the number of strings in the bowl impact the expected number of loops formed?

  As N increases the expected number of loops formed increases logarithmically.

EXTEND How could this investigation be extended?

Firstly, to extend this investigation I looked at adjusting the assumption that 'all the string ends have an equal probability of being selected'.
What if there was a biased end selection, favouring ends from the same string?

I want to investigate how adding a bias to the experiment would change the expected number of loops formed for a given N.

Lets suppose that instead of all of the string ends having an equal probability of being selected, selecting a string end from the same string as your first selection now has a probability double that of any other string end.
Will the expected number of loops formed be double for a given N?

I WANT What is the expected number of loops of any size formed for a given N?

Base Case, N=1.
If N=1, there are only 2 string ends and so the probability of creating a loop on the first pairing is 1.

$$E_1(loops) = 1$$

SPECIALISE Lets now look at the case for N=2,
lets label the strings as before,

$$1_a, 1_b,\ 2_a,\ 2_b$$

Suppose without loss of generality our first selection is 1_a.
Here is where our calculation differs to before.
As now we still have 3 options to choose from with still only 1 resulting in the formation of a loop.

$$1_b,\ 2_a,\ 2_b$$

But the probability of selecting 1_b is no longer $\frac{1}{3}$ but it is instead $\frac{2}{4} = \frac{1}{2}$. This is because the probability of selecting 1_b is double that of the other 2, so we can think of it as having another 1_b in the bowl.
Therefore, we can calculate the expected number of loops formed when N=2,

$$E_2(loops) = (0 \times \tfrac{1}{2}) + (1 \times \tfrac{1}{2}) + 1 = \tfrac{3}{2}$$

We can see that the method for calculating the expected number of loops formed for a given N is exactly the same as before with only small altercations.

Therefore we can take the equation that we had for the expected number of loops formed and adjust it accordingly.
Original equation,

$$E_N(loops) = \sum_{r=1}^{N} \frac{1}{2r-1}$$

STUCK How do we need to adjust it?

This equation sums all the probabilities of selecting a string end that will make a loop. Previously this was always 1 string end out of the string ends remaining in the bowl. However, since the probability of selecting a string end from the same string is now double that of the others, it is like we are adding another string end to the bowl that can result in a loop being formed. So, to adjust our formula we will now have 2 as the numerator and 2r as the denominator.

AHA! New equation ,

$$E_N(loops) = \sum_{r=1}^{N} \frac{2}{2r} = \sum_{r=1}^{N} \frac{1}{r}$$

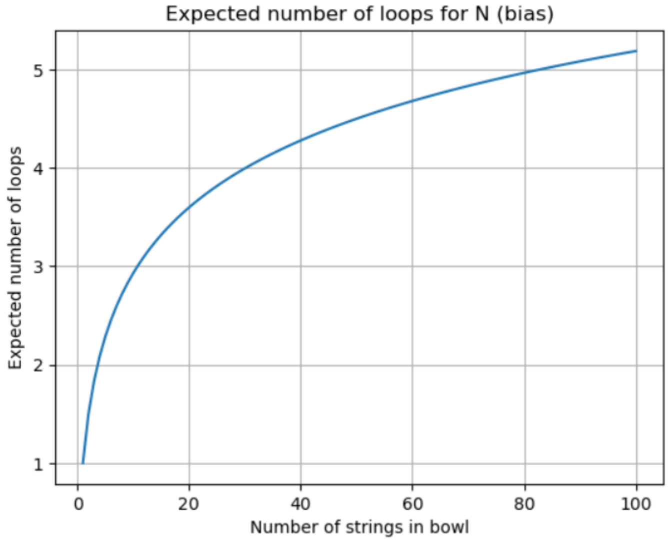

Figure 6: Probability of forming a single loop of size N.

From the table we can see that the results are very similar to the original experiment, however the expected values are slightly higher.
At the start of this extension I asked the question, will the expected values be double the original expected values?

STUCK The answer is no, we can compare the expected number of loops formed when N=100,

- Unbiased $E_{100}(loops) = 3.2843421893016322$
- Biased $E_{100}(loops) = 5.187377517639621$

AHA! This is because we didn't directly double the probability of creating a loop, we said the probability was double that of the others. These statements are very different.

If we look at the case of N=3,

- The unbiased probability of forming a loop on the first pairing is $\frac{1}{5}$
- The biased probability of forming a loop on the first pairing is $\frac{1}{3}$

$$2 \times \frac{1}{5} \neq \frac{1}{3}$$

We can extend this further by adjusting the bias more.
Here are the graphs for the expected number of loops formed for a given N, when the probability of selecting a string end of the same string already selected is,

- 3 times that of the others
- 4 times that of the others
- 0.5 times that of the other

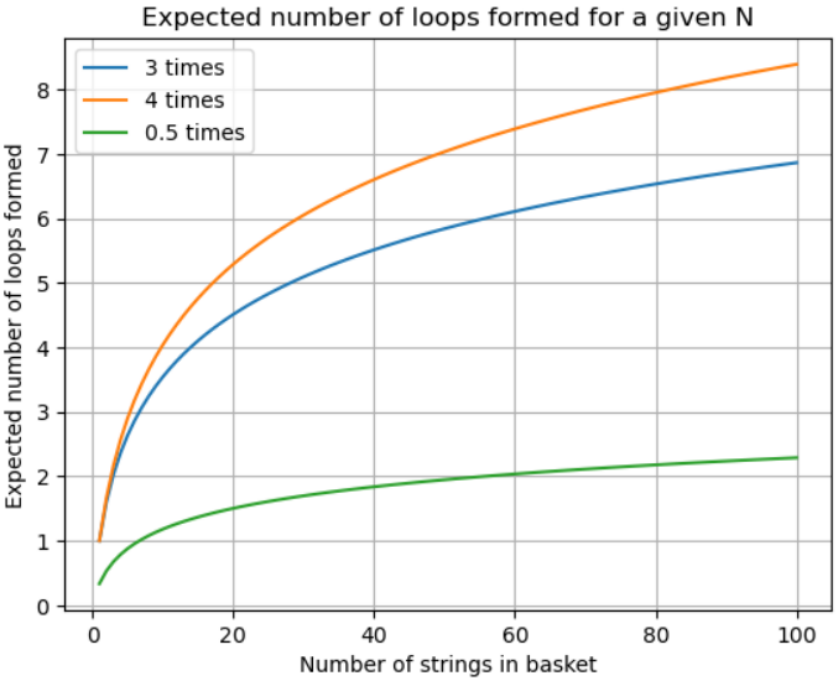

Figure 7: Expected number of loops formed for different probabilities.