Stochastic Loop Formation

Harshitha Mutyala

Contents

1 Entry **2**

1.1 Problem set up . . . 2

1.2 Definitions . . . 2

1.3 Objectives . . . 2

1.4 Example . . . 3

2 Attack **4**

2.1 Probability of forming a loop of length i in k tying sessions . . . 4

2.2 Expected number of loops . . . 7

2.3 Expected number of loops with length 1 when $k = N$ . . . 11

3 Review **13**

3.1 Expected number of loops with restrictions . . . 14

1 Entry

1.1 Problem set up

To begin with, there are N strings all jumbled up in a bowl, which means that there are $2N$ string ends in the bowl. I then blindly and randomly choose two string ends and tie them together, reducing the number of total string ends in the bowl by 2. I repeat this process k times, where $1 \leq k \leq N$, until a number of string loops have been formed, and not necessarily until I have tied all string ends together. A loop is formed when two ends of a string, or a sequence of connected strings, are tied together. At the end of the process, we are guaranteed at least one loop. The possible final objects in the bowl are: a single string, a long string, a loop made of only one string, and a loop made of multiple strings. These are illustrated in Figure 1.

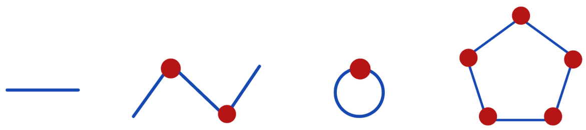

Figure 1: Possible string shapes.

1.2 Definitions

We now provide some definitions that we will use throughout the rest of the project:

- Let us refer to the nth time we tie the string as the nth tying session.
- Define a chain of strings to be multiple strings attached to each other when it is not yet a loop.
- Define the length of the loop or chain to be the number of strings that make up the loop or chain.
- Let C_n be the number of string ends before the nth tying session.
- Let A_n be a random variable representing the length of the loop formed in the nth tying session. If no loop has formed, then the length of the loop is given as 0.
- Let Z_n be a random variable representing the number of loops in the bowl after n tying sessions.

1.3 Objectives

The main goals for this project are:

- What is the expected number of string loops formed in k tying sessions when we start with N strings?
- What is the probability of forming a loop of length $k = N$?
- What is the probability of forming a loop of length i in k tying sessions?
- What is the expected number of loops with length 1?

1.4 Example

I now attempt to introduce myself to the problem by looking at the case where we only have 5 strings in the bowl and we tie all string ends together. Initially we have 10 string ends, and at the nth tying session we have $C_n = 10 - 2(n-1)$ string ends remaining as we tie two string ends together each tying session. At first, I can pick from any of the 10 string ends, but to form a loop I need to pick the other end of that string. So I only have 1 choice for the second string end from the remaining 9 string ends in the bowl.

Notice that picking the end of a string of length 1 is just as likely as picking the end of a string with length greater than 1. So I will go about this problem by ignoring the loops once they have formed, and thinking that chains of any length are just lone strings. So in the beginning of each tying session, we can choose any string end and the other end of that same string to form a loop. So we can choose from C_n string ends for the first string end and 1 string end for the second string. This means that in the nth tying session, the probability of tying a loop is:

$$\frac{C_n}{C_n} \times \frac{1}{C_n - 1} = \frac{1}{C_n - 1} = \frac{1}{10 - 2(n-1) - 1}.$$

This means that the expected number of loops at the end of the 5 tying sessions is:

$$\mathbb{E}[Z_5] = \sum_{n=1}^{5} \frac{1}{10 - 2(n-1) - 1} \times 1 = \frac{563}{315}.$$

2 Attack

2.1 Probability of forming a loop of length i in k tying sessions

We need to find the probability of forming a loop of length i in k tying sessions, when we start with N strings. Let us first look at the case when $i = k = N$. Since we must tie up all string ends, all the strings will be involved in loops so we do not have to worry about forming no loops. Notice that this is the probability of forming no loops in the first $N-1$ tying sessions, and then forming a loop in the Nth tying session.

Recall that A_n is the random variable representing the length of the loops formed in the nth tying session. So we see that $\mathbb{P}[A_n = 0]$ represents the probability of no loops forming in a tying session. Similarly, $\mathbb{P}[A_N = N]$ represents the probability of us forming a loop with a chain of length N. This means that the probability that we need to find is:

$$A = \Big(\prod_{n=1}^{N-1} \mathbb{P}[A_n = 0]\Big)\mathbb{P}[A_N = N | A_1 = \ldots = A_{N-1} = 0].$$

Our next step is to find the probability distribution for A_n. Let us see what happens in the first two tying sessions:

- First tying session:

  $\mathbb{P}[A_1 = 1]$ is the probability of making a loop in the first tying session. $\mathbb{P}[A_1 = 0] = 1 - \mathbb{P}[A_1 = 1]$ is the probability of making no loops in the first tying sessions. Clearly we cannot make a loop with more than length 1 in the first tying session. So this gives us:

  $$\mathbb{P}[A_1 = 1] = \frac{1}{C_1 - 1}, \quad \mathbb{P}[A_1 = 0] = 1 - \frac{1}{C_1 - 1}, \quad \mathbb{P}[A_1 \geq 2] = 0.$$

- Second tying session:

  $\mathbb{P}[A_2 = 1] = \mathbb{P}[A_2 = 1 | A_1 = 1]\mathbb{P}[A_1 = 1] + \mathbb{P}[A_2 = 1 | A_1 = 0]\mathbb{P}[A_1 = 0]$, where in the second term we have to avoid making a loop of the chain of length 2. $\mathbb{P}[A_2 = 2] = \mathbb{P}[A_2 = 2 | A_1 = 0]\mathbb{P}[A_1 = 0]$ since we only get a chain of length 2 if no loops were formed in the first tying session. Once again, $\mathbb{P}[A_2 = 0]$ is the probability that no loops were formed. All of this gives us:

  $$\begin{aligned}
  \mathbb{P}[A_2 = 1] &= \frac{1}{C_2 - 1} \times \mathbb{P}[A_1 = 1] + \frac{C_2 - 2}{C_2} \times \frac{1}{C_2 - 3} \times \mathbb{P}[A_1 = 0] \\
  \mathbb{P}[A_2 = 2] &= \frac{2}{C_2} \times \frac{1}{C_2 - 1} \times \mathbb{P}[A_1 = 0] \\
  \mathbb{P}[A_2 = 0] &= 1 - \mathbb{P}[A_2 = 1] - \mathbb{P}[A_2 = 2] = 1 - \frac{1}{C_2 - 1} \\
  \mathbb{P}[A_2 \geq 3] &= 0.
  \end{aligned}$$

This approach seems too complicated for $\mathbb{P}[A_n \neq 0]$, and it does not seem like I would get a closed-form generalized formula at the end. However, there seems to be a pattern for $\mathbb{P}[A_n = 0]$.

Conjecture 1: $\mathbb{P}[A_n = 0] = 1 - \frac{1}{C_n - 1}$ for $1 \leq n \leq N - 1$.

We try and calculate the number of ways to not form a loop in the first $N - 1$ tying sessions.

As I did in the specialization in the entry stage, I will ignore loops once they have formed, and I will treat string ends of longer chains like any other strings.

At first, we can choose any string end so we have $2N$ options. For the second string end, we can't choose what we have just chosen, and we also can't choose the other end of the string that we just chose. If we did, then we would form a loop, leading to a contradiction. So there are $2N-2$ ways to choose the second string. Using the same logic for the second tying session, there are $2N-2$ and $2N-4$ ways to choose the first and second string ends respectively. So in each tying session, for the first choice we have the option to choose from any of the string ends, and for the second choice, we have two fewer string ends to choose from. This is the same as C_n and C_n-2 in the nth tying session. We see that:

$$\begin{aligned}\mathbb{P}[A_n=0] &= \frac{C_n}{C_n}\times\frac{C_n-2}{C_n-1}\\ &= 1-\frac{1}{C_n-1}.\end{aligned}$$

Since this is true for all $1\leq n\leq N-1$, we have proved our claim.

Note that if we don't form any loops in the first $N-1$ tying sessions, then we end up with a chain of length N. This means that there are only 2 string ends left before the Nth tying session, so they have to be tied together to form a loop of length N. Conversely, I can only form a loop of length N in the Nth tying session if I get no loops in all of the previous tying sessions. So it is clear that

$$\mathbb{P}[A_N=N|A_1=0,\ldots,A_{N-1}=0]=1.$$

Since each tying session we lose 2 string ends to choose from, we have that $C_n=2N-2(n-1)$. We can now calculate A to be:

$$\begin{aligned}A &= \prod_{n=1}^{N-1}\mathbb{P}[A_n=0]\times 1\\ &= \prod_{n=1}^{N-1}\left(1-\frac{1}{C_n-1}\right)\\ &= \prod_{n=1}^{N-1}\left(1-\frac{1}{2N-2(n-1)-1}\right)\\ &= \prod_{t=1}^{N-1}\frac{2t}{2t+1} \text{ by using the substitution } t=N-n.\end{aligned}$$

We now need to find a way to calculate the probability of forming a loop of length i, without relying on all string ends being tied up. Consider representing each string as a pair containing the labels of the string ends. So N strings can be labelled as $(1,2),\ldots,(2N-1,2N)$. In the first tying sessions, we need to choose two different numbers, say a and b, from the set $\{1,\ldots,2N\}$. If we pick from the same pair, then we have a loop so we remove the elements of the loop pair from $\{1,\ldots,2N\}$. If we pick from two different pairs, then merge the two pairs to make a list with 4 elements. For example, if a is in the pair (a,a') and b is in the pair (b,b'), then merging them would give (a,a',b,b').

We repeat this process, this time choosing numbers from $\{1,\ldots,2N\}\setminus\{a,b\}$. We do this k times, keeping track of the number of elements in loops. Then the length of the loop is clearly half the number of elements in a loop list, since there are 2 string ends involved in a knot.

Perhaps I can implement this reasoning in a recurrence relation, but then I wouldn't know how to merge the lists and remove used string ends from the set of all available string ends. I was not sure where to go from here so I made 10000 simulations to represent this data using the same logic. I did this by using Python to create a function which gave me the simulated probabilities of loops of length $1 \leq i \leq k$ occurring in k tying sessions. I then plotted Figure 2, where we have $k = N = 100$ tying sessions.

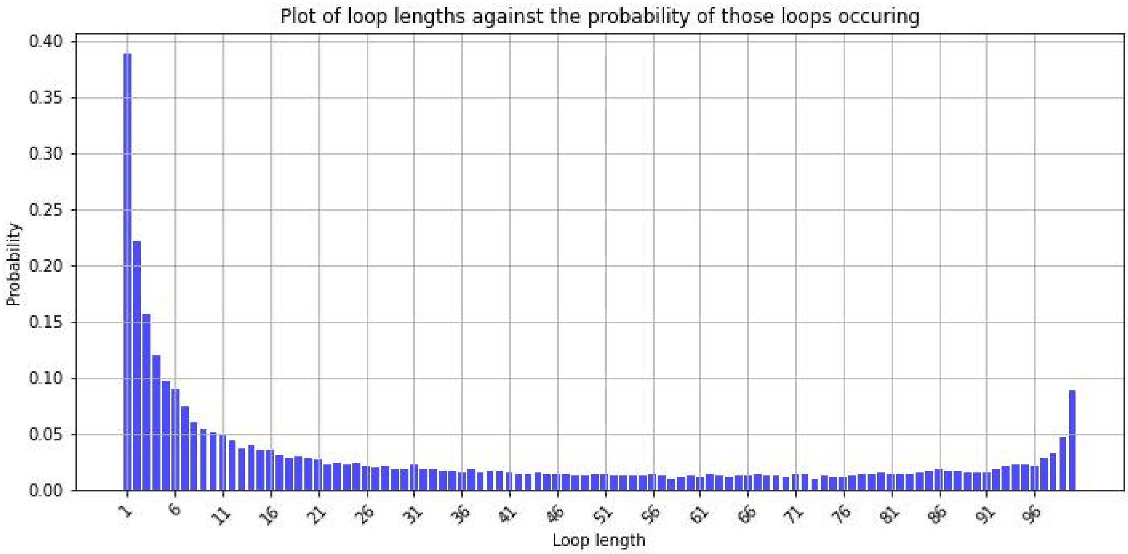

Figure 2: Plot of loop lengths against simulated probabilities of those loops occurring in $N = 100$ tying sessions.

Using the formula given above, I got $A \approx 0.0887$. Using the function used to plot Figure 2, I got the following:

```
Probability of a loop of length 100 occurring: 0.0889.
```

This evidences the validity of Figure 2.

2.2 Expected number of loops

We are going to try to use the specialization from the entry phase to find the expected number of loops for k tying sessions when we start with N strings. By looking at the result of the specialization example, I will try to conjecture the formula. Recall that Z_k is the random variable representing the number of loops in the bowl after the kth tying session.

Conjecture 2: $\mathbb{E}[Z_k] = \sum_{n=1}^{k} \frac{1}{2N-2(n-1)-1}$ when $1 \leq k \leq N$.

Since we ended up with a summation formula at the end of the the specialization, I will try to generalize the formula using recursion. In each tying session, we either create a loop or extend a chain. As seen in the specialization, the probability of creating a loop in the nth tying session is

$$\frac{C_n}{C_n} \times \frac{1}{C_n - 1} = \frac{1}{C_n - 1}.$$

This means that the probability of extending a chain, or equivalently not forming a loop, in a tying session is given by

$$1 - \frac{1}{C_n - 1} = \frac{C_n - 2}{C_n - 1}.$$

Then we see that

$$\mathbb{E}[Z_k] = \frac{1}{C_k - 1}(1 + \mathbb{E}[Z_{k-1}]) + \frac{C_k - 2}{C_k - 1}(\mathbb{E}[Z_{k-1}]).$$

We have that $C_n = 2N - 2(n-1)$, so the expected number of loops in k tying sessions when we start with N strings is given by:

$$\begin{aligned}\mathbb{E}[Z_k] &= \frac{1}{2N - 2(k-1) - 1}(1 + \mathbb{E}[Z_{k-1}]) + \frac{2N - 2(k-1) - 2}{2N - 2(k-1) - 1}(\mathbb{E}[Z_{k-1}]) \\ &= \mathbb{E}[Z_{k-1}] + \frac{1}{2N - 2k + 1} \\ &= \sum_{n=1}^{k} \frac{1}{2N - 2n + 1} \quad \text{as required.} \end{aligned} \tag{1}$$

We see that $\sum_{n=1}^{5} \frac{1}{2\times 5 - 2n + 1} = \frac{563}{315}$, which was the same thing we got for our specialization in the entry stage.

The equality (1) can be visualized in Figure 3, where we have 100 strings to start with.

We can see in Figure 3 that the expected number of loops rapidly increases when k gets closer to 100 as we have fewer string ends to choose from, so the probability of making a loop is higher. This is consistent with our intuition.

However, the question says that at the end of us tying the strings, we are left with some loops in the bowl. This means that there is at least 1 loop at the end of our tying sessions. So the

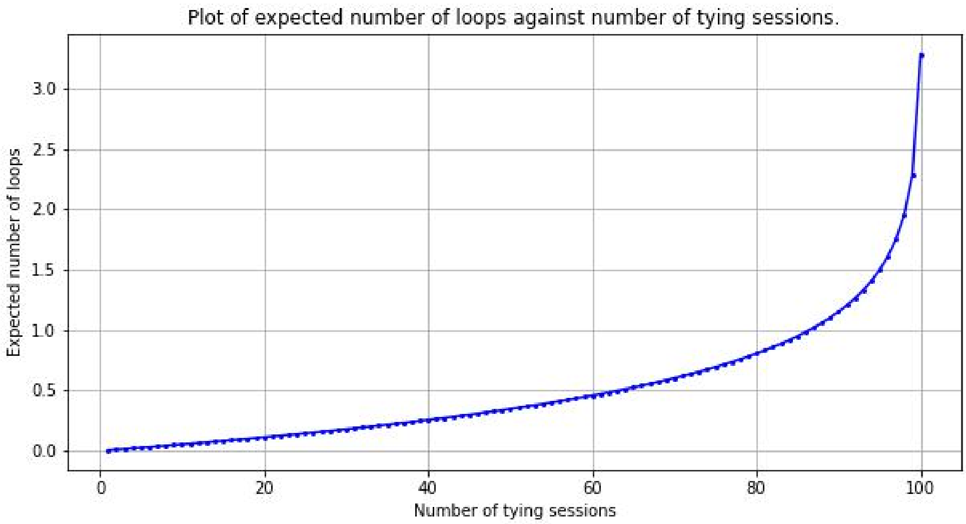

Figure 3: Plot of number of tying sessions against expected number of loops.

formula I have given above only works when $k = N$, as all the strings are a part of a loop since all string ends are tied up. So our question then becomes what is the expected number of loops given that we have at least one loop. We now try to find a probability distribution for the number of loops, which we do by finding the probability of having L loops in k tying sessions when we start with N strings.

I start off by noting that the number of ways to arrange L loops in k tying sessions is given by $\binom{k}{L}$. I am not sure where to go from here so let me try to find the probability of getting all L loops in the first L tying sessions. This means that we end up with L loops of length 1 in the first L tying sessions, and no more loops for the rest of the $k - L$ tying sessions. We have seen before that the probability of a loop during a tying session is $\frac{1}{C_n - 1}$. So the probability that we are looking for is:

$$\Big(\prod_{n=1}^{L} \frac{1}{C_n - 1}\Big) \prod_{n=L+1}^{k} \Big(1 - \frac{1}{C_n - 1}\Big).$$

This is one of the cases that we need to account for when finding the probability distribution for the number of loops. The rest of the cases depend on in which tying sessions the L loops are made. So let us construct a set which contains all possible positions in which we can have the L loops:

$$X = \{P = \{x_1, \dots, x_L\} : x_i \in \{1, \dots, k\} \text{ and } x_i \neq x_j \text{ when } i \neq j\}.$$

We also need a set which gives us the tying sessions that won't make a loop:

$$Y_P = \{1, \dots, k\} \setminus P.$$

Then we can clearly see that the probability of having L loops in k tying sessions when we start off with N strings is given by:

$$\mathbb{P}[Z_k = L] = \sum_{P \in X} \Big(\prod_{n \in P} \frac{1}{C_n - 1}\Big) \prod_{m \in Y_P} \Big(1 - \frac{1}{C_m - 1}\Big).$$

I see now that this formula doesn't work when $L = 0$ as X would equal $\{\emptyset\}$. This would mean that the probability of forming no loops is 0, which is not true when $k < N$. We can create a separate formula for when $L = 0$ by recalling the probability of forming no loops in the nth tying session. So we get:

$$\mathbb{P}[Z_k = 0] = \prod_{n=1}^{k} \Big(1 - \frac{1}{C_n - 1}\Big).$$

This is depicted in Figure 4, where we look at the probability distribution of the number of loops when we have 20 tying sessions and we start with 20 strings.

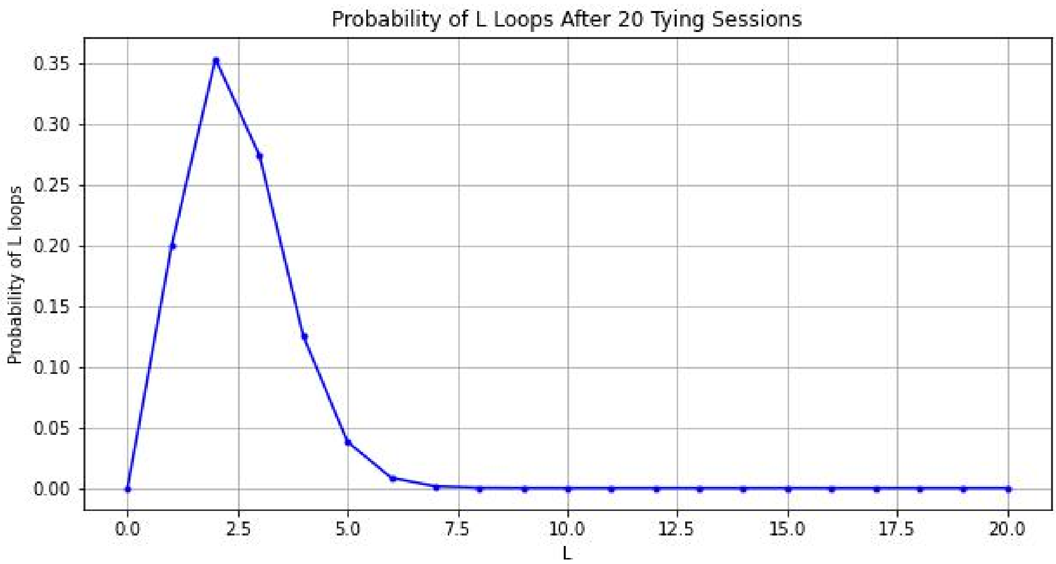

Figure 4: Plot of L against probability of L loops occurring after 10 tying sessions.

It is easy to see that $\mathbb{P}[Z_k \geq 1] = 1 - \mathbb{P}[Z_k = 0]$. So we know that

$$\begin{aligned}\mathbb{P}[Z_k = L | Z_k \geq 1] &= \frac{\mathbb{P}[Z_k = L]}{1 - \mathbb{P}[Z_k = 0]} \text{ when } L \geq 1 \\ \mathbb{P}[Z_k = 0 | Z_k \geq 1] &= 0.\end{aligned}$$

We can now use these probabilities to calculate the expected number of loops given the condition that we must have at least 1 loop by the end of the process.

$$\mathbb{E}[Z_k | Z_k \geq 1] = \sum_{L=1}^{k} L \times \frac{\mathbb{P}[Z_k = L]}{1 - \mathbb{P}[Z_k = 0]}. \tag{2}$$

This is too burdensome to calculate by hand so I created a Python function which does this for me. I found that it is quicker to do this when I use simulated probabilities, and I was able to see visualizations for when I started off with a higher number of strings than 20. So I used Python to run 1000 simulations of the process and got simulated probabilities for the number of loops made in k tying sessions. This can be seen in Figure 5.

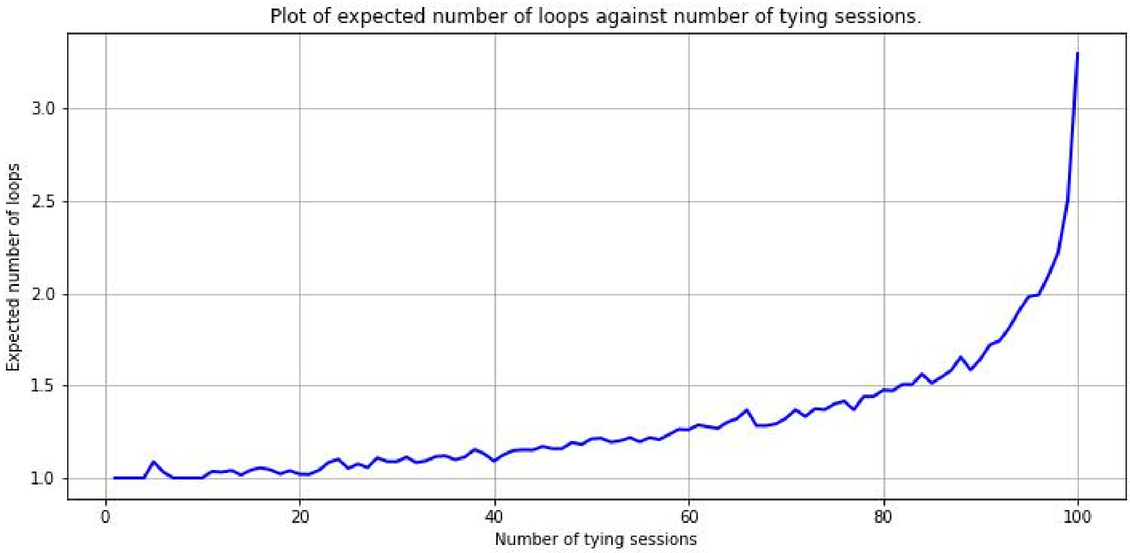

Figure 5: Plot of expected number of loops against number of tying sessions given that we have at least one loop by the end of the process.

We start off with 100 strings like we did in Figure 3. We can see that the curve in Figure 3 is smoother since I did 10000 simulations, whereas I only used 1000 simulations in Figure 5 for faster computations. However, both curves have similar shapes, with the expected values on Figure 5 being higher of course. This evidences the fact that equation (2) is correct.

2.3 Expected number of loops with length 1 when $k = N$

Notice that since $k = N$, we don't have to worry about forming no loops. Once again, we label the N strings as $(1, 2), \ldots, (2N-1, 2N)$. Let us start by trying to find the probability that the string represented by $(1, 2)$ is tied to itself.

Throughout the process of picking the $2N$ strings, notice that the order in which we pick the strings is a random permutation of the set $\{1, \ldots, 2N\}$. Suppose we wrote out these permuted numbers in a straight line. Then clearly the labels of string ends that are tied together would be next to each other. I was not sure how to progress from here, so I used Python to give me all of the permutations of $\{1, \ldots, 4\}$ where 1 and 2 are next to each other.

```
(1, 2, 3, 4)  (3, 1, 2, 4)  (3, 4, 1, 2)  (1, 2, 4, 3)
(4, 1, 2, 3)  (4, 3, 1, 2)  (2, 1, 3, 4)  (3, 2, 1, 4)
(3, 4, 2, 1)  (2, 1, 4, 3)  (4, 2, 1, 3)  (4, 3, 2, 1).
```

Notice that in `(3, 1, 2, 4)`, the string (1,2) does not form a loop. Instead, we form a loop of length 2. This means that not only do 1 and 2 have to be next to each other, but the first appearing number must be on an odd position. Also it doesn't matter which order 1 and 2 are in as long as they are in the right positions. This means that only 8 out of the $4! = 24$ total permutations result in (1,2) forming a loop. Doing the same thing for $N = 3$, we see that (1,2) becomes a loop in only 144 out of the total 720 permutations possible. We see that

$$\frac{2}{2} = \frac{1}{2 \times 1 - 1}, \quad \frac{8}{24} = \frac{1}{2 \times 2 - 1}, \quad \text{and} \quad \frac{144}{720} = \frac{1}{2 \times 3 - 1},$$

which leads to the following conjecture.

Conjecture 3: The probability that the string (1,2) will form a loop in N tying sessions is given by $\frac{1}{2N-1}$.

Consider a random permutation of the set $\{1, \ldots 2N\}$, and let us represent this by $x = ((x_1, x_2), \ldots, (x_{2N-1}, x_{2N}))$. Think of x_i as the ith string end being chosen in the game, so $x_j = 1$ for some $1 \leq j \leq 2N$. This means that if the string (1,2) was to form its own loop of length 1, then 2 has to be in the same pair as x_j in x. There are $2N - 1$ spaces for 2 to be in and only one of those spaces would lead to (1,2) becoming a loop, so there is a $\frac{1}{2N-1}$ probability of (1,2) becoming a loop.

Notice that any string of length 1 has the same probability of becoming a loop of length 1. Let B be a random variable representing the number of loops of length 1 after all N tying sessions. Then

$$\begin{aligned} \mathbb{E}[B] &= \sum_{n=1}^{N} \frac{1}{2N-1} \times 1 \\ &= \frac{N}{2N-1}. \end{aligned} \tag{3}$$

I used Python to create 10000 simulations of the process and compared the expected values to the theoretical values that can be found from the formula above.

```
(N=10) Theoretical: 0.5263157894736842, Simulated: 0.5131
(N=50) Theoretical: 0.5050505050505051, Simulated: 0.5039
(N=75) Theoretical: 0.5033557046979866, Simulated: 0.499
(N=100) Theoretical: 0.5025125628140703, Simulated: 0.5002
```

We can see that the simulated expected values are getting closer to 0.5, which agrees with the fact that

$$\lim_{N\to\infty} \mathbb{E}[B] = \frac{1}{2}.$$

3 Review

- I found the probability of making a loop of length N, but I did not get a closed formula for the more general probability distribution. I came up with a method to make a simulation of this instead, and checked the simulation with the probability of making a loop of length N. I obtained equations (2) and (3) for the expected number of loops, which I then checked with the help of simulations.

- The specialization I did in the entry stage helped me to grasp the problem at hand much better and gave me a starting point with my line of reasoning. Python really helped me with checking my answers and computing solutions that I could not find closed-form expressions for. The simulations helped me visualize data in Python that otherwise would have taken a lot of time to compute.

- I see now that I could have simply proved Conjecture 1 by noticing that the probability of tying a loop in a tying session is $\frac{1}{C_n - 1}$, so clearly the probability of not tying a loop is $1 - \frac{1}{C_n - 1}$.

- As an extension, we can look at what happens when there are restrictions placed on which string ends can be tied together. Suppose we assign each string a top and a bottom. Then we can only make top-bottom and bottom-top knots. If we pick two bottom string ends (or two top string ends) then we don't make a knot in that tying session. I am now going to try to find the expected number of loops for this process in k tying sessions, where $1 \leq k \leq K$.

3.1 Expected number of loops with restrictions

We can see that in each tying session, we can make a loop, extend a chain or make no connection between strings. We lose two string ends if a knot is made and we lose none if a knot is not made. At any point between tying sessions, we have an equal number of tops of strings and bottoms of strings.

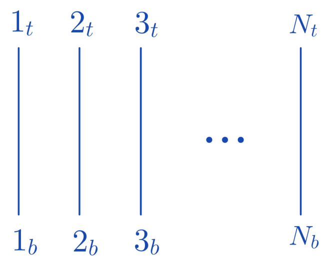

Figure 6: We have an equal number of top and bottom strings.

All this means that in each tying session:

- Probability of a loop: $\frac{C_n}{C_n} \times \frac{1}{C_n - 1} = \frac{1}{C_n - 1}$
- Probability of chain extension: $\frac{C_n}{C_n} \times \frac{C_n/2}{C_n - 1} = \frac{C_n/2}{C_n - 1}$
- Probability of no knot: $\frac{C_n}{C_n} \times \frac{(C_n/2)-1}{C_n - 1} = \frac{(C_n/2)-1}{C_n - 1}$.

But notice that C_n does not depend on just n, and it depends on the previous tying sessions as well.

Notice that $\mathbb{E}[Z_k] = \mathbb{P}[$ we make a loop in the kth tying session $] + \mathbb{E}[Z_{k-1}]$. Let S_k be a random variable representing the number of knots made in k tying sessions. Let us first try to see what happens when $k = 1$ and $k = 2$.

- When $k = 1$:

  Clearly $\mathbb{E}[Z_1]$ is just the probability of forming a loop in the first tying session, so it is equal to $\frac{1}{C_1 - 1} = \frac{1}{2N - 1}$.
- When k=2:

  The probability of making a loop in the second tying session depends on how many untied string ends are in the bowl. This depends on whether or not a knot was made in the first tying session. So we have

$$\begin{aligned}
\mathbb{E}[Z_2] &= \mathbb{P}[S_1 = 1]\mathbb{P}[\text{ loop in 2nd tying session } |S_1 = 1] \\
&\quad + \mathbb{P}[S_1 = 0]\mathbb{P}[\text{ loop in 2nd tying session } |S_1 = 0] + \mathbb{E}[Z_1] \\
&= \left(\frac{C_1}{C_1} \times \frac{C_1/2}{C_1 - 1}\right) \times \left(\frac{C_1 - 2}{C_1 - 2} \times \frac{1}{C_1 - 3}\right) + \left(\frac{C_1}{C_1} \times \frac{C_1/2 - 1}{C_1 - 1}\right) \times \left(\frac{1}{C_1 - 1}\right) + \mathbb{E}[Z_1]
\end{aligned}$$

I now propose a conjecture using the specialization:

Conjecture 4: $\mathbb{E}[Z_k] = \sum_{i=0}^{k-1} \mathbb{P}[S_{k-1} = i]\mathbb{P}[\text{ loop in } k\text{th tying session } |S_{k-1} = i] + \mathbb{E}[Z_{k-1}]$.

My answer for $\mathbb{E}[Z_2]$ makes me realize that if we have all possible combinations of a knot occurring or not occurring in k tying sessions, then we can find the probability of a loop forming given those combinations, and multiply it by the probability of the combination occurring. Let $Q_k = \{Q = D_1 \cap \ldots \cap D_k : D_j \in \{R_j, F_j\}$ and $1 \leq j \leq k\}$, where R_j is the event that a knot is formed in the jth tying session, and F_j is the event that a knot is not formed in the jth tying session. Q_{k-1} gives all the possible combinations of knots being formed or not formed in $k-1$ tying sessions.

We need to find the probability of a loop being formed in the kth tying session given all possible combinations. So

$$\mathbb{E}[Z_k] = \sum_{Q \in Q_{k-1}} \mathbb{P}[Q]\mathbb{P}[\text{ loop in the } k\text{th tying session } |Q] + \mathbb{E}[Z_{k-1}]$$

But actually, if we have formed i knots in $k-1$ tying sessions, then the number of string ends left in the bowl is $2N-2i$. So the probability of a loop forming in the kth tying session is $\frac{2N-2i}{2N-2i} \times \frac{1}{2N-2i-1}$. We see that i depends only on the number of knots made. So instead of looking at each $Q \in Q_{k-1}$, we can instead look at the number of tying session for which $D_j = R_j$, where $1 \leq j \leq k-1$. This is clearly the random variable represented by S_{k-1}.

The minimum number of knots that could be made is 0 and the maximum would be k. So we get:

$$\mathbb{E}[Z_k] = \sum_{i=0}^{k-1} \mathbb{P}[S_{k-1} = i]\mathbb{P}[\text{ loop in the } k\text{th tying session } |S_{k-1} = i] + \mathbb{E}[Z_{k-1}]$$

as required. Using that fact that $\mathbb{P}[\text{ loop in the } k\text{th tying session } |S_{k-1} = i] = \frac{1}{2N-2i-1}$, and summing over all $\mathbb{E}[Z_j]$s, we get

$$\begin{aligned}\mathbb{E}[Z_k] &= \sum_{j=1}^{k}\sum_{i=0}^{j-1} \mathbb{P}[S_{k-1} = i]\mathbb{P}[\text{ loop in the } k\text{th tying session } |S_{k-1} = i] \\ &= \sum_{j=1}^{k}\sum_{i=0}^{j-1} \mathbb{P}[S_{k-1} = i]\frac{1}{2N-2i-1}.\end{aligned}$$

We now need to find $\mathbb{P}[S_k = i]$, where $0 \leq i \leq k-1$. Let me now attempt to find a recursion formula for $\mathbb{P}[S_k = i]$, so let $i \geq 1$ for now.

There are two situations that could've occurred in the $k-1$th tying session: either we tied a knot or we didn't tie a knot. To get i successful knots in k tying sessions, we either had $i-1$ knots in the $k-1$th tying session and made a knot kth tying session, or we had i knots in the $k-1$th tying session and we didn't make a knot in the kth tying session.

$$\begin{aligned}\mathbb{P}[S_k = i] = &\ \mathbb{P}[S_{k-1} = i]\mathbb{P}[\text{no knot in } k\text{th tying session}|S_{k-1} = i] \\ &+ \mathbb{P}[S_{k-1} = i-1]\mathbb{P}[\text{knot in } k\text{th tying session}|S_{k-1} = i-1]\end{aligned}$$

If we have made s knots, then there are $2N-2s$ string ends left in the bowl. To make a knot, I have to first pick any string end, then I need to pick the other kind of string end.

These string ends make up $N - s$ of the total $2N - 2s - 1$ string ends left in the bowl. So $\mathbb{P}[\text{knot in } k\text{th tying session}|S_{k-1} = s] = \frac{N-s}{2N-2s-1}$. Clearly we can't make any knots before we have attempted to tie any strings together, so $\mathbb{P}[S_0 = i] = 0$ for all i. If we make no knots in k tying sessions, then the number of string ends won't change in those sessions. All of this combined gives us:

$$\begin{aligned}
\mathbb{P}[S_k = i] &= \mathbb{P}[S_{k-1} = i]\Big(1 - \frac{N-i}{2N-2i-1}\Big) + \mathbb{P}[S_{k-1} = i-1]\Big(\frac{N-(i-1)}{2N-2(i-1)-1}\Big) \\
&= \mathbb{P}[S_{k-1} = i]\Big(\frac{N-i-1}{2N-2i-1}\Big) + \mathbb{P}[S_{k-1} = i-1]\Big(\frac{N-i+1}{2N-2i+1}\Big) \\
\mathbb{P}[S_0 = i] &= 0 \\
\mathbb{P}[S_k = 0] &= \prod_{n=1}^{k}\Big(1 - \frac{N}{2N-1}\Big) = \Big(\frac{N-1}{2N-1}\Big)^k.
\end{aligned}$$

I am not sure on how to find a closed expression for $\mathbb{P}[S_k = i]$ that won't be messy, so I will use the recursion and

$$\mathbb{E}[Z_k] = \sum_{j=1}^{k}\sum_{i=0}^{j-1}\mathbb{P}[S_{k-1} = i]\frac{1}{2N-2i-1}$$

to code a graph in Python depicting my results. I will compare these results with $\mathbb{E}[Z_k]$ values that I found from 10000 simulations in Python.

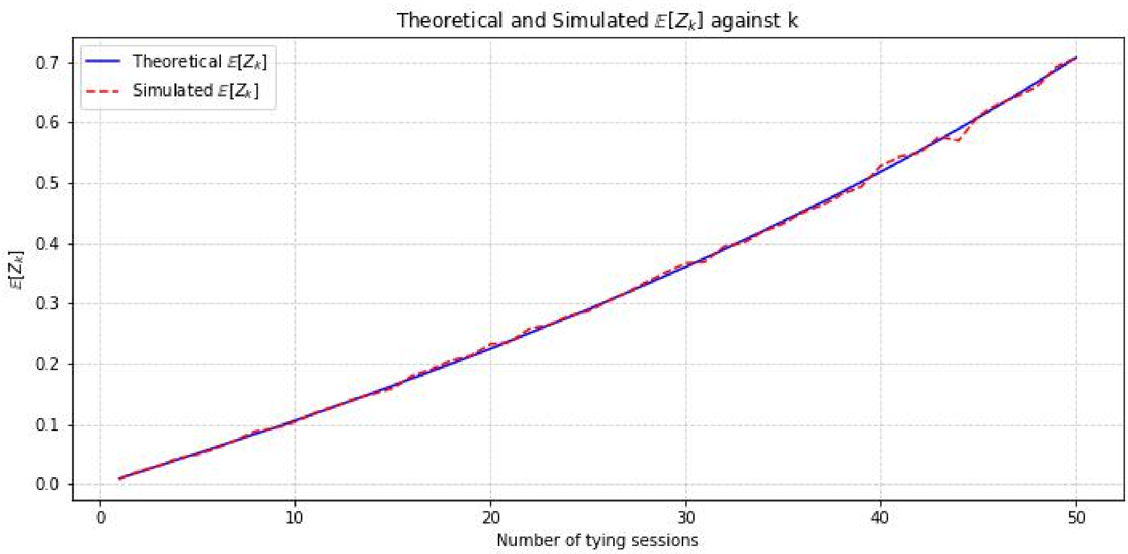

Figure 7: Plot of number of tying sessions against theoretical and simulated $\mathbb{E}[Z_k]$ values.

Figure 7 evidences the validity of my formula for $\mathbb{E}[Z_k]$.