8-2018

The Expected Number of Patterns in a Random Generated Permutation on [n] = {1,2,...,n}

Evelyn Fokuoh
East Tennessee State University

The Expected Number of Patterns in a Random Generated Permutation on $[n] = \{1, 2, \ldots, n\}$

A thesis

presented to

the faculty of the Department of Mathematics

East Tennessee State University

In partial fulfillment

of the requirements for the degree

Master of Science in Mathematical Sciences

---

by

Evelyn Fokuoh

August 2018

Anant Godbole, Ph.D., Chair

Teresa Haynes, Ph.D.

Nicole Lewis, Ph.D.

Keywords: Subsequence, random permutation, subadditivity, Fekete's Lemma

ABSTRACT

The Expected Number of Patterns in a Random Generated Permutation on

$[n] = \{1, 2, \ldots, n\}$

by

Evelyn Fokuoh

Previous work by Flaxman (2004) and Biers-Ariel et al. (2018) focused on the number of distinct words embedded in a string of *words* of length n. In this thesis, we will extend this work to *permutations*, focusing on the maximum number of distinct permutations contained in a permutation on $[n] = \{1, 2, \ldots, n\}$ and on the expected number of distinct permutations contained in a random permutation on $[n]$. We further considered the problem where repetition of subsequences are as a result of the occurrence of (Type A and/or Type B) replications. Our method of enumerating the Type A replications causes double counting and as a result causes the count of the number of distinct sequences to go down.

ACKNOWLEDGMENTS

I probably might not have the right words to express how grateful I am to my advisor, Dr. Anant Godbole for his help. I would not have been able to make it without his efforts, directions and intense commitment to this work. Not forgetting the support from my committee members. I would like to also thank Abdul Nasah, my friend who was a great source of help to me in making this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

A major problem of interest in combinatorics is the number of distinct subsequences that exist as subsequences of a given string. The number of distinct subsequence equals 2^n- the number of repetitions. For example, in the word 10110, the embedded subsequences are {}, 1, 0, 10, 01, 00, 11, 101, 100, 110, 111, 1011, 1010, 1110, 0110, and 10110, where throughout this thesis we use the notation {} for empty set. There are thus 16 *distinct* subsequences among the $2^5 = 32$ subsequences of 10110. Similarly the permutation 1324 contains (using order-isomorphic representations we can, e.g., rename both 13 and 12 as 12) the sub-permutations {}, 1, 12, 21, 132, 123, 213, and 1324. The rest of the subpermutations are repetitions. Flaxman et al. (2004) focused on determining the maximizing string in the case of binary words. This turned out to be the alternating string. In other words, for the alphabet $\Sigma = \{1, 2\}$, the sequence $(1, 2, 1, 2, \ldots)$, gives the maximum number of distinct subsequences, which turn out to be

$$Fib(n+3) - 1 \sim ((1+\sqrt{5})/2)^{n+3}/\sqrt{5},$$

which is asymptotic to $(1.62)^n$. This result was generalized for a finite alphabet Σ of size d. Biers-Ariel et al. (2018) defined T_n as a fixed binary string of length n, t_i as the i^{th} letter of T_n, S_n as a random binary string of length n and $\phi(T_n)$ as the number of distinct subsequences of T_n. The authors of this paper focused on $E[\phi(S_n)]$ when $\Pr[s_i = 1] = \alpha \in [0, 1]$. The minimum value of $E(\phi(S_n)$ on a fixed length-n string on the the alphabet $\{0, 1\}$, is trivial. It occurs when $\alpha = 0$ or $\alpha = 1$, and in this case $E(\phi(S_n)) = n + 1$. If $n = 2, \alpha = 1/2$, the four possibilities for T_n

are 11, 10, 01, and 00. These contain 3, 4, 4, and 3 subsequences respectively, so $E(S_n) = (3 + 4 + 4 + 3)/4 = 3.5$. A key result in Biers-Ariel et al. (2018), which enables one to count $\phi(T_n)$ recursively, is the following: Given T_n, let l be the greatest number less than n such that $t_l = t_n$, and if no such number exists, let $l = 0$. Then, $\nu(T_n)$, the number of distinct new subsequences created at the nth stage in a string T_n, is

$$\nu(T_n) = \begin{cases} n & \text{if } l = 0 \\ \sum_{i=l}^{n-1} \nu(T_i) & \text{if } l > 0 \end{cases}$$

As an example, we consider the word $\{10010\}$ of length $n = 5$. We have $\phi(T_5) = 1 + \nu(T_1) + \nu(T_2) + \nu(T_3) + \nu(T_4) + \nu(T_5) = 1 + 1 + 2 + 2 + 5 + 7 = 18$. Biers-Ariel et al.(2018) then used the above expression and linearity of expectation to find $E(\phi(S_n))$ for each $\alpha \in (0, 1)$. Later in the article, the authors focused on random strings from an alphabet of size d, where the letter $j \in [d]$ is independently selected with probability α_j. In other words, $\Pr[s_i = j] = \alpha_j$ for all $i \in [n], j \in [d]$. (Note that $\sum_{j=1}^{d} \alpha_j = 1$). They gave a recursion to count $\nu(T_n)$ as in the case of $d = 2$. They also considered the case where the letters were generated according to a Markov chain.

To give a specific example of a result in Biers-Ariel et al. (2018), suppose we are considering the case of binary words with unequal probabilities, i.e., $\Pr[s_i = 1] = \alpha \in [0, 1]; \Pr[s_i = 0] = 1 - \alpha$ for all $1 \leq i \leq n$. Then we have

$$E[\phi(S_n)] = \begin{cases} n+1 & \text{if } \alpha = 0, 1 \\ \frac{\left(1-2\sqrt{\alpha(1-\alpha)}\right)\left(1-\left(1-\sqrt{\alpha(1-\alpha)}\right)^n\right)+\left(1+2\sqrt{\alpha(1-\alpha)}\right)\left(\left(1+\sqrt{\alpha(1-\alpha)}\right)^n-1\right)}{2\sqrt{\alpha(1-\alpha)}} & \text{if } \alpha \neq 0, 1 \end{cases}$$

Thus, $E[\phi(S_n)] = 2(\frac{3}{2})^n - 1$ when $\alpha = 0.5$, as proved by Flaxman et al (2004).

The main open problem stated in Biers Ariel et al. (2018) is extending the above work to permutations. Both the referees of their paper were enthusiastic about this direction.

In this thesis, $\phi(\pi_n)$ is the number of distinct permutations contained in a permutation π_n on $[n]$. π_n may be fixed or random. In the random case, we can consider $E(\phi(\pi_n))$. $\nu(j)$ is the number of new permutations created by the jth entry of the sequence $(\pi(1), \ldots, \pi(n))$ for $j = (1, 2, \ldots, n)$. Finally, $\psi(k, \pi)$ is the number of repeats of permutations of length k. Some of the basic definitions of interest are the Type A repeats which are repeats in which two subpermutations have the same pattern and the values are identical except for two values which have the same rank in the pattern, e.g., subpermutations 2146 and 2145 of length $k = 4$. Both generate 2134 patterns. If n is odd, then the singleton $\{n\}$ is a type C duplicate. Any duplicate which is neither a Type A nor a Type C is a type B duplicate.

Our two main directions will be the following:

(A) To determine the maximum number of embedded permutations as done by Flaxman et al. (2004) for words, since the minimum number of embedded permutations is trivial. (The permutation $\pi_n = 123...n$ contains the permutations $\{\}$, $1, 12, \ldots, 123\ldots n$, or $n+1$ permutations).

(B) In the random case to make progress in determining $E(\phi(\pi_n))$.

These two topics will be presented in Sections 3 and 4 respectively, after some

preliminary results are presented in Section 2. Other issues such as the variance of $\phi(\pi_n)$ or the distribution of $\phi(\pi_n)$ are very difficult and we will not consider these. Trying to find a recursion that counts the number of new permutations created at the jth position (this was the approach taken by Biers-Ariel et al. 2018) will be an option; this will hopefully give an exact expression for $E(\phi(\pi_n))$. Another option might be to show that $E(\phi(\pi_n))/a^n \to 1$ for some a (using subadditivity and Fekete's Lemma). We will explore all these options in the coming pages.

2 PRELIMINARY RESULTS

In this part of the thesis we determine $(\mathbf{E}[\phi(n)])^{1/n}$ for $n = \{1, 2, 3, 4, 5, 6, 7, 8\}$. $\nu(j)$ is defined to be the number of new permutations created by the jth entry of the sequence $(\pi(1), \ldots, \pi(n))$ for $j = (1, 2, \ldots, n)$ and $\phi(n)$ is the number of distinct permutations contained in the sequence $(\pi(1), \pi(2), .., \pi(n))$. Hence $\phi(n) = \sum_{j=1}^{n} \nu(j)+1$. This was the approach taken by Biers-Ariel et al. (2018) for words, but after several attempts at finding analogs in the case of permutations, we abandoned this line of investigation. For small values of n, we were able to use R (a free software for statistical computing) to write an Rcode. This is available in the Appendix. We derived histograms for $n = \{1, 2, 3, 4, 5, 6, 7, 8\}$. The R on the x-axis defines the number of embedded sequences for each value of n while the y-axis defines the rate at which each embedded sequence occur (count). For instance in Figure 1, the possible permutation for $n = 1$ is 1. The number of embedded subsequence when $n = 1$ is $\{\{\}, 1\}$. These two appear just once. For Figure 1, the mean of the embedded sequences when $n = 1$ is 2. From our calculation, we had the variance of these sequences to be 0, the expected number of distinct permutations $(\mathbf{E}[\phi(1)])^{1/1}$ was 2 and the value of the constant C was 1 (Mean/2^n). For Figure 2, with $n = 2$, we can generate two possible permutations namely 12 and 21. The number of embedded sequences for 12 is $\{\{\}, 1, 12\}$ and for 21 is $\{\{\}, 2, 21\}$. However, by order isormorphism the 2 in 21 turns to be 1. Since we have three embedded sequences in both 12 and 21, it means the rate(frequency) at which we see the count of 3 is 2 times. These embedded sequences give us a mean value of 3, and a variance of 0.222. The $(\mathbf{E}[\phi(2)])^{1/2}$ is 1.732 and the constant C is 0.75.

Now for Figure 3 , where $n = 3$, the possible permutations are $123, 132, 213, 213, 231, 321$. There are 4 embedded sequences including the empty set in 123 and 321 but the rest contains 5 embedded sequences. Thus the rate at which we see the count of 4 is 2 and the rate ate which we see the count of 5 is 4. The mean value of these embedded sequences when $n = 3$ is 4.667 , the variance is 0.222, $(\mathbf{E}[\phi(3)])^{1/3}$ is 1.671 and the value of the constant C is 0.5833. This idea is used to explain the nature of the histograms, mean etc for Figures 4 through Figure 8. Detailed information on the embedded sequences for $n = \{1, 2, ..., 8\}$ can be found in the Rcode under Appendix.

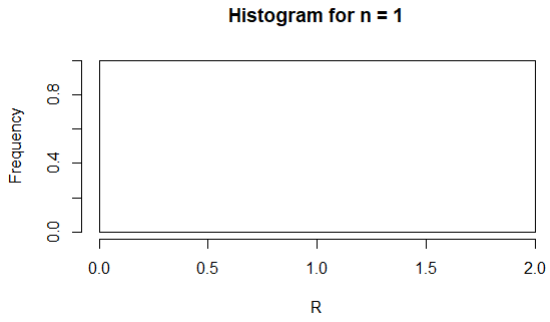

Figure 1: A Histogram showing the value of $(\mathbf{E}[\phi(n)]$ at different values of R when $n = 1$.

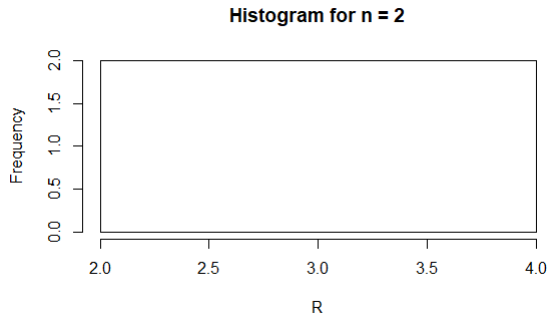

Figure 2: A Histogram showing the value of $(\mathbf{E}[\phi(n)]$ at different values of R when $n = 2$.

The mean for the embedded sequences for Figure 2 where $n = 2$ is 3 and the variance is 0. The $(\mathbf{E}[\phi(n)])^{1/n}$ and the value of the constant C are 1.732, 0.75 respectively.

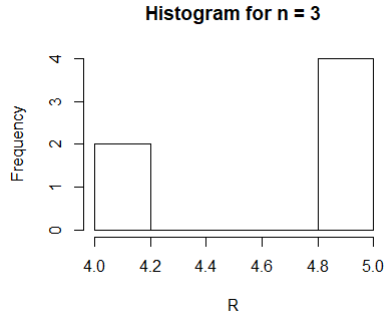

Figure 3: A Histogram showing the value of $(\mathbf{E}[\phi(n)]$ at different values of R when $n = 3$.

For the Figure 3 , the mean of the 6 embedded sequences is 4.667 and the variance $= 0.222$. The $(\mathbf{E}[\phi(n)])^{1/n} = 1.671$ and $C = 4.667/2^3 = 0.5833$.

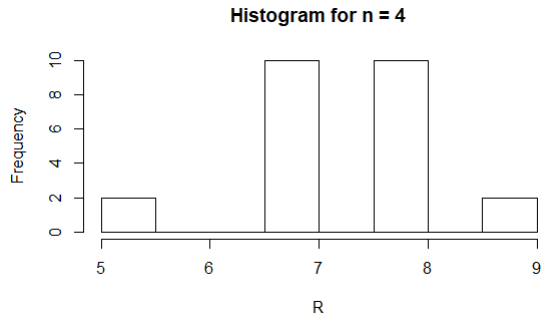

Figure 4: A Histogram showing the value of ($\mathbf{E}[\phi(n)]$ at different values of R when $n = 4$.

When $n = 4$, the mean of the 24 embedded sequences is 7.417 and the variance = 0.9097. The $(\mathbf{E}[\phi(n)])^{1/n} = 1.6502$ and $C = 7.417/2^4 = 0.4636$.

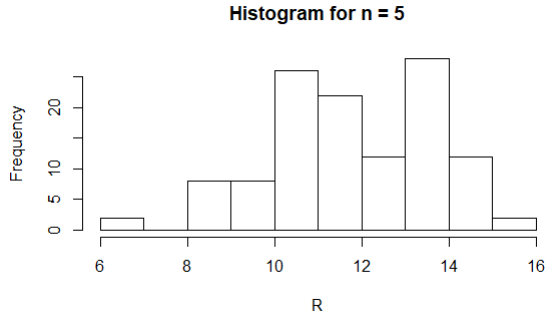

Figure 5: A Histogram showing the value of ($\mathbf{E}[\phi(n)]$ at different values of R when $n = 5$.

For Figure 5, the mean of the 120 embedded sequences is 12.2833 and the variance = 4.1969. The $(\mathbf{E}[\phi(n)])^{1/n} = 1.6514$ and $C = 12.283/2^5 = 0.3838$.

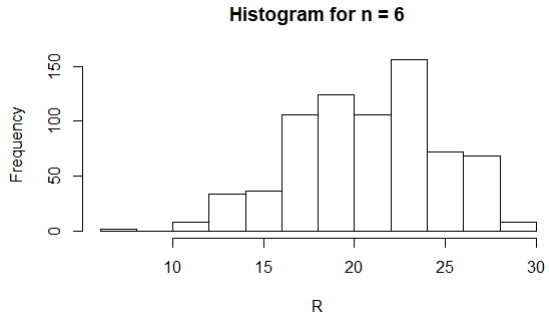

Figure 6: A Histogram showing the value of $(\mathbf{E}[\phi(n)]$ at different values of R when $n = 6$.

The mean of the 720 embedded sequences in Fig. 6 is 21.1444 with a variance of 25.6347. $(\mathbf{E}[\phi(n)])^{1/n} = 1.6629$ and $C = 21.144/2^6 = 0.3304$.

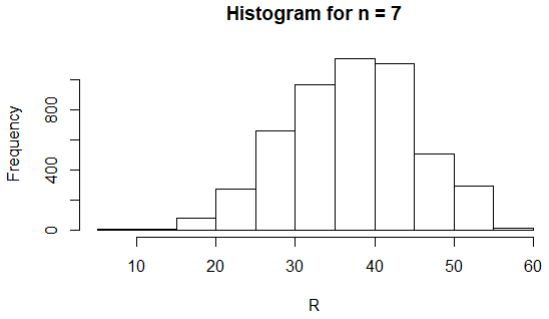

Figure 7: A Histogram showing the value of $(\mathbf{E}[\phi(n)]$ at different values of R when $n = 7$.

We generate 5040 embedded sequences when $n = 7$ with a mean of 37.60397 and variance of 76.0793. The $(\mathbf{E}[\phi(n)])^{1/n} = 1.6789$ and $C = 37.604/2^7 = 0.2938$.

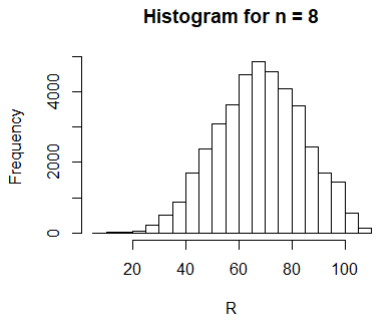

Figure 8: A Histogram showing the value of $(\mathbf{E}[\phi(n)]$ at different values of R when $n = 8$.

The mean for the embedded sequences is 68.63661 with a variance of 201.3242. The

$$(\mathbf{E}[\phi(n)])^{1/n} = 1.6966 \text{ and the constant } C = 68.637/2^8 = 0.2681$$

We observe that the $(\mathbf{E}[\phi(n)])$ was 2 for $n = 1$, then it decreased to 1.732 for $n = 2$. Then it reduced to 1.671 for $n = 3$. However after $n = 4$, it started to increase and finally to 1.6966 for $n = 8$. The value of the constant C on the other hand was reducing for values of n from 1 through 8. Our interest in deriving the values of the constant C was because we believe $(\mathbf{E}[\phi(n)]) = C_n * 2^n$. Hence, for reasons that we will give later, we believe that the following holds:

Conjecture 2.1. *Since* $(\mathbf{E}[\phi(n)])$ *is increasing after* $n = 4$ *and approaching a constant number for values of* $n = \{1, 2, .., 8\}$*, then*

$$\lim_{n\to\infty} (\mathbf{E}[\phi(n)])^{1/n} = 2.$$

2.1 Possible use of Fekete's Lemma(Subadditivity)

Suppose we can show that that for each $\pi \in S_{n+m}$, where S_n is the set of all permutations on $[n]$, we have that $\phi(\pi_{m+n}) \leq \phi(\pi_n)\phi(\pi_{\{n+1,\ldots,n+m\}})$, then we would have made great progress. This was what occurred in Biers-Ariel et al (2018); they exploited subadditivity to prove the existence of a limit for $(\mathbf{E}(\phi_n))^{1/n}$.

Theorem 2.2. Fekete's Lemma*: If a_n is a real sequence for which*

$$a_{n+m} \leq a_n + a_m, \quad (n, m = 1, 2, \ldots)$$

then

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_n \frac{a_n}{n},$$

where the existence of the limit is the key conclusion.

As an example, we can consider $\pi = 24315$ $n = 3$ and $m = 2$. Then, $\phi(\pi_{n+m}) = 14$ due to the embedded subpermutations

$$\{\}, 1, 12, 21, 123, 132, 231, 213, 321, 3214, 2314, 1324, 2431, 24315;$$

$\phi(\pi_n) = 5$ due to

$$\{\}, 1, 12, 21, 132;$$

and $\phi(\pi_{\{n+1,\ldots,n+m\}}) = 3$ due to

$$\{\}, 1, 12.$$

The inequality holds. If this was to be true for any Π, then we could say that in general

$$\mathbf{E}[\phi_{n+m}] \leq \mathbf{E}(\phi_n.\phi_{\{n+1,\ldots,n+m\}}).$$

Since this is on disjoint intervals, these values will be independent of each other. If we assume that $n \leq m$ then we would have

$$\begin{aligned} \mathbf{E}(\phi_{n+m}) &\leq \mathbf{E}(\phi_n.\phi_{\{n+1,\ldots,n+m\}}) \\ &= \mathbf{E}(\phi_n)\mathbf{E}(\phi_{\{n+1,\ldots,n+m\}}) \\ &= \mathbf{E}(\phi_n)\mathbf{E}(\phi_m) \qquad (1) \\ Then, \log \mathbf{E}(\phi_{n+m}) &\leq \log \mathbf{E}(\phi_n) + \log \mathbf{E}(\phi_m) \qquad (2) \end{aligned}$$

Letting $a_n = \log \mathbf{E}(\phi_n)$, this would lead to the conclusion that

$$a_{n+m} \leq a_n + a_m,$$

and thus, via Fekete's Lemma, that

$$\lim(\mathbf{E}(\phi_n))^{1/n} = D.$$

This is because we would have had

$$\frac{a_n}{n} = \frac{1}{n} \log \mathbf{E}(\phi_n) \to C,$$

and thus

$$\log \mathbf{E}(\phi_n)^{\frac{1}{n}} \to C,$$

which would imply that

$$\mathbf{E}(\phi_n)^{\frac{1}{n}} \to e^C := D.$$

This was however a failed attempt. We actually found examples where the inequality didn't hold. In an attempt to fix this, we examined the Erdős-deBruijn generalization of Fekete's lemma.

2.2 Erdős-Debruijn Generalisation of Fekete's Lemma

Theorem 2.3. Fekete's Lemma*: If we have* $\phi(n+m) \leq \phi(m) + \phi(n) + a(m+n)$, *where* $\sum_{n=1}^{\infty} \frac{a_n}{n^2} < \infty$, *then* $\lim_{n\to\infty} \frac{a_n}{n}$ *exists.*

We consider as examples the cases where $\phi_{m+n} \leq \phi_n + \phi_m + n + m$, and $\phi_{m+n} \leq \phi_n + \phi_m + \sqrt{n+m}$. Theorem 3 could not be used in the first case but could be used in the second case, since the series $\sum \frac{1}{n}$ diverges but the series $\sum \frac{1}{n^{3/2}}$ converges. Our efforts to find a "buffer" function that led to Theorem 3 being valid were unsuccessful as well.

2.3 Trying to use a Recursion Similar to that in Biers-Ariel et al (2018)

We have that

$$\phi(\pi_n) = \sum_{j=1}^{n} \nu(j)$$

where $\nu(j)$ is the number of new permutations counted by the jth entry. We can calculate $\nu(j)$ as follows: If $\nu(j)$ is the smallest term, increase all the previous entries by 1 and label the jth entry as 1. Retain all those sequences that have not occurred before. If we consider as an example the permutation 42681, the possible

subpermutations are:

$$\begin{aligned} &j = 1; \nu(j) = 1; 1 \\ &j = 2; \nu(j) = 1; 21 \\ &j = 3; \nu(j) = 2; 12, 213 \\ &j = 4; \nu(j) = 2; 23, 2134 \\ &j = 5; \nu(j) = 5; 321, 231, 3241, 32451, 2341. \end{aligned} \tag{3}$$

We can observe that the last entry $j = 5$ was obtained by increasing all previous entries we have obtained by 1 and finally ending with a term 1.

If $\nu(j)$ is the largest term encountered, we augment each previous term by the next largest value and retain all those sequences that have not occurred before. We take a look at the permutation 42689:

$$\begin{aligned} &j = 1; \nu(j) = 1; 1 \\ &j = 2; \nu(j) = 1; 21 \\ &j = 3; \nu(j) = 2; 12, 213 \\ &j = 4; \nu(j) = 2; 123, 2134 \\ &j = 5; \nu(j) = 2; 1234, 21345. \end{aligned} \tag{4}$$

However, this process is very difficult to track when terms other than the largest or smallest appear at the jth position. We thus abandoned our efforts to adopt a

method similar to that in Biers-Ariel et al. (2018), which was used so successfuly for words.

2.4 Heuristic Reason why we think $\mathbf{E}(\phi_n)^{1/n} \longrightarrow 2$

We see from the histograms that $\mathbf{E}(\phi_n)^{1/n}$ is decreasing up till $n = 5$ and increasing after that, up till $n = 8$. Will it approach 2 as $n \to \infty$? Here we argue that it might be possible, and the rest of the thesis is devoted to gathering evidence that the limit might equal 2.

Now for large k, we observed that $k! \gg \binom{n}{k}$. To obtain the repeats of length k for large k, we are essentially asking how many repeats are possible in a small set of $\binom{n}{k}$ positions, where there exist $k!$ possible permutations. Intuition suggests that this would be a small number. This thesis will be devoted to classifying and understanding these repeats.

Let the number of repeats of permutations of length k be denoted by $\psi(k, \pi)$. **Suppose** that for a large k, say $k \geq n/2$, we have at most $\binom{n}{k}/2$ repeats, so that $\max \psi(k, \pi) \leq \binom{n}{k}/2$. Then the total number of repeats is

$$\sum_{k=1}^{n/2} \max \psi(k, \pi) + \sum_{n/2+1}^{n} \max \psi(k, \pi)$$

which is bounded above by

$$\sum_{k=1}^{n/2} \binom{n}{k} + \sum_{n/2+1}^{n} \binom{n}{k}/2,$$

which simplifies to

$$\frac{1}{2} * 2^n + \frac{1}{2} * 2^n * \frac{1}{2} = \frac{3}{4} * 2^n.$$

Therefore, the number of distinct permutations would be at least $2^n - \frac{3}{4} * 2^n = \frac{1}{4} * 2^n$, and hence we would have that

$$\mathbf{E}(\phi_n)^{1/n} \geq \frac{1}{4^{1/n}} * 2 \to 2.$$

2.5 Tpes of Repeats

In the permutation $n = 2146537$, let $k = 4$. Then we have $k! = 24$ and $\binom{n}{k} = 35$. The successive terms are $(1, 2)$ and $(5, 6)$. The possible subpermutations of length 4 from the above permutation $n = 2146537$ are obtained in the Table 1 below. From Table 1, the colored pattern under the reduced column indicate some of the Type A repeat. In this table, if we consider case 1 and case 2, thus the subpermutations 2146 and 2145. These two subpermutations generate the same reduce pattern 2134. This is because 6 in the subpermutation 2146 and 5 in subpermutation 2145 turn to have the same rank (which is 4) in the reduced pattern. Again from case 4 and case 7, the subpermutations 2147 and 2167 both generate a reduce pattern of 2134. This is because 4 in the subpermutation 2147 and 6 in the subpermutation 2167 have the same rank of 3 in the reduced pattern. The rest of the Type A repeats are found in the Table 1 below and those that can be paired are colored with the same color.

Table 1: Types of repeats.

CASE	SUBPERMUTATION	REDUCTION
1	2146	2134
2	2145	2134
3	2143	2143
4	2147	2134
5	2165	2143
6	2163	2143
7	2167	2134
8	2153	2143
9	2157	2134
10	2137	2134
11	2465	1243
12	2463	1342
13	2467	1234
14	2453	1342
15	2457	1234
16	2437	1324
17	2653	1432
18	2657	1324
19	2637	1324
20	2537	1324
21	1465	1243
22	1463	1342
23	1465	1234
24	1453	1342
25	1457	1234
26	1437	1324
27	1653	1432
28	1657	1324
29	1637	1324
30	1537	1324
31	4653	2431
32	4657	1324
33	4637	2314
34	4537	2314
35	6537	3214

The only type C present in this permutation is 7. We start by analyzing type A repeats with successions. A succession is defined as an occurrence where $\pi(i+1) = \pi(i) \pm 1$.

There are however other types of events that cause duplication other than successions. In this thesis, we focus on what we call Type A duplicates, which are defined more precisely (than done before) later.

2.6 Calculation of the Expected Number of Patterns in a Random Permutation

We start by calculating some values of $[\mathbf{E}(\phi(n)]^{1/n}$ by hand and by using the R code found in the Appendix. The expected number of patterns for $n = 2, 3, 4$ are in the tables below:

From the Table 2, we derive two possible permutations of length 2. This is found in the first column. The number of patterns is 3 as explained for Figure 2 for each of these permutations. The probability that each of these permutations occur is 0.5. The $[\mathbf{E}(\phi(2)]^{1/2} = (6/2)^{0.5} = 1.7321$. Hence, the expected number of patterns for $n = 2$ is 1.7321.

Table 2: Expected number of patterns for n=2.

π_n	number of patterns (ϕ_n)	Probability
12	{},1,12 =3	1/2=0.5
21	{},1,21 =3	0.5

For $n = 3$, the column 1 of Table 3 gives us information on all the 6 possible permutations for $n = 3$. This is similar to the number of permutations we derived for Figure 3. The $[\mathbf{E}(\phi(3)]^{1/3} = (28/6)^{1/3} = 1.671$. Since we have 6 permutations in all, the probability that each of these permutations occur is 0.167.

Table 3: Expected number of patterns for n=3.

π_n	number of patterns (ϕ_n)	Probability
123	{},1,12 ,123= 4	1/6=0.167
132	{},1,12,21,132 = 5	0.167
213	{},1,21,12,213 = 5	0.167
231	{},1,12,21,231=5	0.167
312	{},1,21,12,312 =5	0.167
321	{},1,21,321=4	0.167

For $n = 4$, we have a total of 24 possible permutations of length 4 each in the Table 4. $[\mathbf{E}(\phi(4)]^{1/4} = (178/24)^{1/4} = 1.6502$.

Table 4: Expected number of patterns for n=4.

π_n	number of patterns (ϕ_n)	Contribution to Expected Value
1234	{},1,12 ,123,1234= 5	5*1/24=5/24
1243	{},1,12,21,123,132,1243 = 7	7/24
1324	{},1,12,21,123,132,213,1324 = 8	8/24
1342	{},1,12,21,123,132,231,1342 =8	8/24
1423	{},1,12,21,312,132,123,1423 =8	8/24
1432	{},1,12,21,132,321,1432=7	7/24
⋮	⋮	⋮
4312	{},1,12,21,312,321,4312 =7	7/24
4321	{},1,21,321,4321=5	5/24

The probability that each of these permutations in Table 4 will occur is 1/24. However, the contribution of 5 to the expected value is 5/24.

The corresponding values for $[\mathbf{E}(\phi(n)]^{1/n}$ for $n = 5, 6$ and 7 are :

$$[\mathbf{E}(\phi(5)]^{1/5} = (1,474/120)^{1/5} = 1.6514.$$

$$[\mathbf{E}(\phi(6)]^{1/6} = (15,224/720)^{1/6} = 1.66289.$$

$$[\mathbf{E}(\phi(7)]^{1/7} = (189,524/5040)^{1/7} = 1.6789.$$

We see that the minimum of $[\mathbf{E}(\phi(n)]^{1/n}$ is at $n = 4$, and we observe an increasing trend from there on.

3 UPPER AND LOWER BOUNDS ON $\phi(n)$

It is clear that there are at most $k!$ distinct permutations of length k. But only $\binom{n}{k}$ of these may actually be present. For small values of k, $k! < \binom{n}{k}$, and this inequality flips for large values of k. As an example we consider $n = 7$. The values of $k!$ are $(1, 2, 6, 24, 120, 720, 5040)$, and the corresponding values for $\binom{n}{k}$ are $(7, 21, 35, 35, 21, 7, 1)$. In any case, we have the bound

$$\phi(n) \leq \sum_{k=0}^{n} \min\left(\binom{n}{k}, k!\right).$$

We believe (and will argue below) that $\sum_{k=0}^{n} \min(\binom{n}{k}, k!)$ is not a bad bound. To explain this we take a look at $\sum_{k=0}^{n} \min(\binom{n}{k}, k!)$. We start by analyzing the relationship between $\binom{n}{k}$ and $k!$:

By Stirling's approximation

$$k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k,$$

so on taking the limit we obtain

$$\lim_{k\to\infty} \frac{k!}{\sqrt{2\pi k}(\frac{k}{e})^k} = 1.$$

For any k, however, we have the bounds

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \leq k! \leq \sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\frac{1}{12k}}.$$

The other term is $\binom{n}{k}$. If k is not of magnitude $O(n)$, (and thus $k \neq \frac{n}{3}$, for example), then we can approximate

$$\frac{(n-k)^k}{k!} \leq \binom{n}{k} \leq \frac{n^k}{k!},$$

and thus

$$\binom{n}{k} \sim \frac{n^k}{\sqrt{2\pi k}(\frac{k}{e})^k}.$$

Since

$$k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k,$$

on equating the two we obtain

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k = \frac{n^k}{\sqrt{2\pi k}(\frac{k}{e})^k},$$

or

$$2\pi k\left(\frac{k^2}{ne^2}\right)^k = 1.$$

In order to get a close estimate for the solution, we ignore the linear factor $2\pi k$, and obtain

$$\frac{k^2}{ne^2} = 1,$$

or

$$k = e\sqrt{n}.$$

Albert et al (3) noted the bound $\sum \min\left(\binom{n}{k}, k!\right)$, but dismissed it without further analysis. Let us summarize the history of the *lower* bound on the maximum number of permutations $\max \phi(n)$ contained in an n permutation:

Coleman (6) proved that

$$\max \phi(n) \geq 2^{n-2\sqrt{n}+1}$$

establishing that

$$(\max \phi(n))^{1/n} \to 2$$

as $n \to \infty$, but leaving open the question of whether

$$\frac{\max \phi(n)}{2^n} \to 1.$$

Albert's et al's lower bound on the maximum number of embedded distinct permutations was larger than what Coleman had. Specifically, they proved that

$$\max \phi(n) \geq 2^n(1 - \frac{6\sqrt{n}}{2^{\frac{\sqrt{n}}{2}}}) = 2^n(1 - \epsilon), \tag{5}$$

thus showing that

$$\frac{\max \phi(n)}{2^n} \to 1.$$

Miller (2008) proved a better lower bound than in (5). Specifically, she proved that

$$\max \phi(n) \geq 2^n - O(n^2 2^{n-\sqrt{2n}}), \tag{6}$$

which she matched with the upper bound of

$$\max \phi(n) \leq 2^n - \Theta(n 2^{n-\sqrt{2n}}). \tag{7}$$

We delved through the paper to uncover the value of the constant in (7) and found that it was enormous. With $l = \sqrt{2n}$, (7) can be fleshed out as

$$\max \phi(n) \leq 2^n - (2^n(\frac{nl - 27l - \frac{27n}{2} - 27}{80(l+3)2^l}) - (2^{16}.n^4.2^{(n-l)})). \tag{8}$$

In the Table 5 below, we compare our upper bound $\sum \min(\binom{n}{k}, k!)$ to Miller's upper bound. We observed clearly the magnitude of the constant in (8) leads to the "trivial" bound, doing much better for values of n at least till 50.

Table 5: Comparison of our upper bound with Miller's upper bound.

n	Our upper bound	Miller's upper bound
1	2	49,182.218
2	3	1,048,580.26
3	5	2,654,216.42
4	9	37,791,894.67
5	16	146,409,613.6
10	540	$3.02059167*10^{10}$
100	1267650591324628...	$4.594862602*10^{10}$
200	160693804425899027...	$1.60693806*10^{68}$

Next, we study the asymptotics of the $\sum \min(\binom{n}{k}, k!)$ bound, and will uncover that while not as good as (8), does satisfy several desirable criteria. We have already seen that $k! = \binom{n}{k}$ around $k = e\sqrt{n}(1 + o(1)) = A$. We thus truncate as follows:

$$\begin{aligned}
\sum \min\left(\binom{n}{k}, k!\right) &= \sum_{0}^{A} k! + \sum_{A+1}^{n} \binom{n}{k} \\
&= \sum_{0}^{A} k! + \sum_{0}^{n} \binom{n}{k} - \sum_{0}^{A} \binom{n}{k} \\
&= \sum_{0}^{n} \binom{n}{k} - \sum_{0}^{A} \left(\binom{n}{k} - k!\right) \\
&= 2^n - \sum_{0}^{A} \left(\binom{n}{k} - k!\right) \\
&= 2^n - \sum_{0}^{A} \binom{n}{k} + \sum_{0}^{A} k!. \qquad (9)
\end{aligned}$$

We need to find a good upper bounds on the two sums in (9). First we have

$$\begin{aligned}
\sum_0^A \binom{n}{k} &= 2^n \sum_0^A \binom{n}{k} \frac{1}{2^k} \frac{1}{2^{n-k}} \\
&= 2^n \Pr(\mathcal{B}i(n, \frac{1}{2}) \leq A) \\
&\geq 2^n \Pr(\mathcal{B}i(n, \frac{1}{2}) = A)(1+\epsilon) \qquad (10) \\
&= \binom{n}{A}(1+\epsilon), \qquad (11)
\end{aligned}$$

where we used, in (11), Proposition A.2.5 in Barbour et al (1992). Since $A \sim e\sqrt{n}$, we see that

$$\binom{n}{A} \sim \frac{n^A}{(\sqrt{n}e)!} \sim (\sqrt{n}e^2)^{\sqrt{n}e}.$$

Plugging this into (11), we see that

$$\begin{aligned}
\sum \min\left(\binom{n}{k}, k!\right) &\geq 2^n - (\sqrt{n}e^2)^{\sqrt{n}e} + \sum_0^A k! \\
&\sim 2^n - (\sqrt{n}e^2)^{\sqrt{n}e} + A \cdot A! \\
&= 2^n - 2^{c\sqrt{n}\ln n} + A \cdot A! \\
&\sim 2^n - 2^{c\sqrt{n}\ln n} + C\sqrt{n} \cdot (\sqrt{n})^{e\sqrt{n}} \\
&= 2^n - 2^{c\sqrt{n}\ln n}(1+o(1)). \qquad (12)
\end{aligned}$$

Thus it turns out that asymptotically, Miller gets a better upper bound which is $2^n - (n2^{n-\sqrt{2}n})$. This is because

$$2^n - (n2^{n-\sqrt{2}n}) \leq 2^n - 2^{c\sqrt{n}\ln n}$$

since $2^{c\sqrt{n}\ln n} \leq n2^{n-\sqrt{2}n}$. As seen however, our bound does far better for small values of n.

4 ANALYSIS

For any n, k, most of the duplicates are created by choosing one entry in a pair that has the same rank (all other terms are the same). We call such duplicates as *Type A duplicates.*

4.1 Interesting Examples

If we consider as an example the permutation 132, the possible patterns are in the Table 6 below,

Table 6: Patterns in the permutation 132.

1	12	21	132
1	12	32	132
2	13	null	null
3	null	null	null

Here the pair 1, 2 creates a duplicate since both have a rank of 1. In the second column, a duplicate is also generated from the pair 12 and 13. This is because the 2 and 3 in these terms have the same rank. When we pair these up the duplicate created is due to (2, 3). This is true for all permutations on $n = 2$ from which we derive 2 possible permutation 12 and 21 in Tables 7 and 8 respectively. The Tables 9 through 14 gives us all possible patterns of the permutations on $n = 3$. Since n is odd, all the 3rd rows of the 1st column gives us type C repeats since they cannot be paired.

Table 7: Patterns in the permutation 12.

1	12
1	12
2	null

Table 8: Patterns in the permutation 21.

1	21
1	21
2	null

Table 9: Patterns in the permutation 123.

1	12	123
1	12	123
2	13	null
3	23	null

Table 10: Patterns in the permutation 132.

1	12	21	132
1	13	32	132
2	12	null	null
3	null	null	null

Table 11: Patterns in the permutation 213.

1	12	21	213
1	13	21	213
2	23	null	null
3	null	null	null

Table 12: Patterns in the permutation 231.

1	12	21	231
1	23	21	231
2	null	31	null
3	null	null	null

Table 13: Patterns in the permutation 312.

1	12	21	312
1	12	31	312
2	null	32	null
3	null	null	null

Table 14: Patterns in the permutation 321.

1	21	321
1	21	132
2	3l	null
3	32	null

We now extend this to a larger n, using say 4132.

1. The old subsequence of size 3, before the 4 was added, was 132. In this term, the only length 3 permutation is 132.

2. The subsequences of size 2 from 132 are $13, 12, 32$. After adding 4, we get $413, 412, 432$ as additional pattern of length 3.

Table 15: Patterns in the permutation 4132.

1	12	21	312	321	132	4132
1	13	41	413	432	132	4132
2	12	43	412	null	null	null
3	null	42	null	null	null	null
4	null	32	null	null	null	null

From the Table 15 above, it can be observed that subsequences of size 1 become subsequences of size 2 when 4 is added; and subsequences of size 2 become subsequences of size 3. Hence the original pair of length 2 in the 2nd column turns to be a pair of length 3 in the 4th column. We can observe that subsequences of size 3 are derived from old subsequences of size 3 and old subsequences of size 2 plus the element n. Hence in general, the subsequences of size k is obtained from old subsequences of size k plus old subsequences of size $k-1$ + a new term, namely n. Thus by Pascal's identity, we have

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}.$$

If we consider the permutation $n = 41325$, the subsequences of size 3 we generate are $413, 412, 432, 132, 135, 125, 415, 435, 425, 325$. The old subsequences of size 3 generated from 4132 are $413, 412, 432, 132$. Again, the old subsequences of size 2 that we generate from 4132 are $13, 12, 41, 43, 42, 32$. When we add the new term 5, each of these length of 2 permutations changes to $135, 125, 415, 435, 425$ and 325. Hence the subsequences of length 3 from 41325 is the sum of the old subsequences of length 3 generated from 4132 and the old subsequences of length 2 from 4132 with 5 added to each pattern. This holds for larger values of n.

We have defined type A duplicates as those that occur when two terms of a permutation can be paired because they have the same rank in a permutation of length k and thus yield the same pattern. Type B duplicates are those that remain after all the type A and type C duplicates have been enumerated. (Type C duplicates may occur as a result of the singletons for odd values of n.) We take a look at an example:

Table 16: Patterns in the permutation 41352.

1	12	21	123	132	213	231	312	321	3124	4132	3142	3241	1342	41352
1	13	41	135	132	415	452	413	432	4135	4132	4152	4352	1352	41352
2	15	43	null	152	435	352	412	null	null	null	null	null	null	null
3	12	42	null	null	null	null	null	null	null	null	null	null	null	null
4	45	32	null	null	null	null	null	null	null	null	null	null	null	null
5	35	52	null	null	null	null	null	null	null	null	null	null	null	null

In the first column of the above table (Table 16), we have duplicates generated by (1,2) and (3,4). All these form a rank of 1 and hence they can be paired. In the second column, we generate a duplicate of generated by the pair (3, 5) from (13, 15). This is a duplicate of type A since 3 and 5 have a same rank in these two permutations. The total number of permutations generated including $\{\}$ is $2^n = 32$. After removing the pairings we have $32 - 2(10) = 12$. Here we have 10 paired patterns that form a Type A and 12 remaining distinct permutations. The 5 in the first column is neither a Type A nor a Type B so it forms a Type C duplicate. We have 15 distinct permutations, but on removing the duplicates we see that we destroy some permutations, leading to only 12 remaining distinct permutations. From the above permutation, 5 is larger

than all the other terms so the same pairing remains. Old subsets of size 2 are created by 12 and 13. New subsets of size 2 are created by adding 5 to all the old subsets of size 1. Old subsets of size 3 are 132,432,412,413. The new subsets generated are 415,425,435,135,125,325. By pairing, we realize that 425 and 435 can be paired because 2 and 3 are of the same rank. 413 and 412 are also pairs with 2 and 3 having the same rank. 135 and 125 can also be paired. We realize the permutation 432 cannot be paired.

Now the role played by Type B repeats is best seen in the permutation 456123. There are exactly two 123 patterns and they have no elements in common. Here is another example:

The permutation 312645 contain the subsets 315,325,164,165,364,365. By pairing 315 and 325 we generate a 213 pattern because 1 and 2 turn to have the same rank. The permutation 164 and 165 also generate 132 pattern.164 and 165 are pairs because 4 and 5 have the same rank.Here again, there are two 312 patterns namely 312 and 645 which do not have any elements in common.

We clearly need to better understand Type B duplicates. All repeats are either a type A or type B, except for the trivial case of up to one Type C duplicate. We therefore have $\mathbf{E}(\phi(\mathbf{n})) = 2^n - \mathbf{E}(\psi(\mathbf{n}))$, where $\mathbf{E}(\phi(\mathbf{n}))$ and $\mathbf{E}(\psi(\mathbf{n}))$ are the expected number of distinct patterns and duplicates respectively.

Since all repeats are of type A or type B, and the singletons can be considered as a type C, our goal is to get an upper bound on $\mathbf{E}(\psi(\mathbf{n}))$.Where

$$\begin{aligned}
\mathbf{E}(\psi(\mathbf{n})) &\leq \mathbf{E}(\psi_{\mathbf{A}}(\mathbf{n})) + \mathbf{E}(\psi_{\mathbf{B}}(\mathbf{n})) + 1 \\
&= \sum_{b,c} \mathbf{E}(\psi_{\mathbf{A},\mathbf{b},\mathbf{c}}(\mathbf{n})) + \sum \mathbf{E}(\psi_{\mathbf{B}}(\mathbf{n})) + 1 \\
&= \sum_{k} \sum_{b,c} \mathbf{E}(\psi_{\mathbf{A},\mathbf{b},\mathbf{c}}(\mathbf{n},\mathbf{k})) + \sum \mathbf{E}(\psi_{\mathbf{B}}(\mathbf{n})) + 1 \\
&= \sum_{1} + \sum_{2} + 1. \qquad (13)
\end{aligned}$$

Here we let,

1. $\mathbf{E}(\psi_{\mathbf{A}}(\mathbf{n}))$ is the number of repeats caused by Type A and $\mathbf{E}(\psi_{\mathbf{B}}(\mathbf{n}))$ is the number of repeats caused by Type B.

2. $\sum_{b,c} \mathbf{E}(\psi_{\mathbf{A},\mathbf{b},\mathbf{c}}(\mathbf{n})$ is the number of repeats of type A caused by b and c where b and c are the two different fixed points and $\sum \mathbf{E}(\psi_{\mathbf{B}}(\mathbf{n}))$ is the number of repeats caused by Type B situations.

3. $\sum_k \sum_{b,c} \mathbf{E}(\psi_{\mathbf{A},\mathbf{b},\mathbf{c}}(\mathbf{n},\mathbf{k}))$ is the number of repeats caused by b and c of length k.

4. $\sum_1$ denotes the duplicates of type A and $\sum_2$ denotes the duplicates of type B.

An upper bound on $\sum_1$ is given in the following main theorem.

Theorem 4.1.

$$\begin{aligned}
\mathbf{E}\left(\sum_{\mathbf{1}}\right) &= \sum_{k=1}^{n} \sum_{t=1}^{n-1} \sum_{r=0}^{n-2} \sum_{s=0}^{\min(t-1,r)} \frac{2(n-t)(n-r-1)}{n!} \times \\
&\quad \binom{r}{s} [(t-1)\dots(t-1-(s-1))] \times \\
&\quad [(n-r-2)(n-r-2-1)\dots(n-r-2-(t-2-s))] \times \\
&\quad (n-2-t+1)! \binom{n-r-t-1+s}{k-1} \qquad (14)
\end{aligned}$$

where

1. *k is the number of repeats of type A of size k with the least being 1 and the maximum as n.*

2. *t is the numerical difference between b and c and the least difference we can have is 1.*

3. *r is the position difference between b and c and the least is 0.*

4. *s the number of terms that are in value between b and c and are physically between b and c.*

5. *2 is the way we can permute the two numbers(b and c or c and b).*

6. *$(n-t)$ is the number of ways we can have the numerical difference being t.*

7. *$(n-r-1)$ is the number of ways we can place the two numbers.*

8. *$n!$ defines the total number of permutation and it is needed since we are dealing with expected value.*

9. *$\binom{r}{s}$ defines the positions of the s numbers, out of the r numbers, that are in value between b and c.*

10. *$[(t-1)\dots(t-1-(s-1))]$ is the number of ways we can choose the s numbers.*

11. *$[(n-r-2)(n-r-2-1)\dots(n-r-2-(t-2-s))]$ is the number of ways we can permute the remaining numbers that lie outside b and c.*

12. $(n-2-t+1)!$ *is the number of ways we can permute the remaining numbers after taking out b and c and the numbers that lies within this interval.*

13. $\binom{n-r-t-1+s}{k-1}$ *is the number of ways to choose either b or c, and k-1 other legal numbers, so as to generate a type A duplicate.*

We will see that

1. $\sum_1$ yields good values (i.e. smaller than $\binom{n}{k}$) for small examples, but there are exceptions.

2. $\sum_1$ is almost exact for $k = n - 1$.

3. $\sum_1$ may cause double counting thus causing the number distinct subsequences to be underestimated.

Considering small values of n, (say $n = 4$ and $k = 1, 2, 3$) we see on the next page the comparison of the formula in Theorem 4.1 with the value of $\binom{n}{k}$ in the Table 17. We can observe from the first row of the Table 17 that our formula gave us a value greater than $\binom{n}{k}$ when k was 1. However, when k was 2, our formula gave us a smaller value than $\binom{n}{k}$. We see that we do not always produce a number that is smaller than $\binom{n}{k}$. We therefore believe that further research work needs to be carried out so that we can determine the true nature of the above Theorem.

Table 17: Comparison of the Main theorem with $\binom{n}{k}$.

k	Formula	$\binom{n}{k}$
1	5.224	4
2	4.833	6
3	1.5	4

Derivation of the number of distinct subsequences: For example, taken

$$\pi = 41325, k = 3,$$

we obtain from Table 18:

Table 18: Distinct subsequences in a permutation.

312	213	123	132	321
413	415	135	132	432
412	435	125	null	null
null	425	null	null	null
null	325	null	null	null

From table 18, given $2^n - \sum_1$ (where $\sum_1$ is repeats of Type A), we have the total left to be 2. In reality, there are 5 distinct subsequences of length 3 which are under the permutation 312, 213, 123, 132 and 321. However, after removing pairs and taking the type A duplicates out we are left with only 2 distinct categories (132 and 321). Because of double counting the total number of duplicates is 8 (these are under categories 312, 213 and 123). The established upper bound might not be a good upper bound because we might have a duplication caused by a and c and the same duplication caused by b and c.

4.2 Special Cases of Theorem 4

4.2.1 Case 1, Successions with $k = n - 1$

When $k = n-1$ we can only have successions. No other values of r, t are possible, e.g. $t \geq 2$. We can only have $t = 1, r = 0$. The formula reduces to

$$\begin{aligned} & \frac{2(n-1)(n-1)(n-2)!}{n!}\binom{n-2}{n-2} \\ = & \frac{2(n-1)(n-1)(n-2)!}{n(n-1)(n-2)!} \\ = & \frac{2(n-1)}{n} \\ \approx & \ 2 \\ \leq & \ 2\binom{n}{k} = \binom{n}{n-1} = \binom{n}{1} = n. \end{aligned} \tag{15}$$

This gives the expected number of successions to be 2, as per the result in Allison et al (2013). The formula is thus exact as stated earlier.

4.2.2 Case 1 - Successions with $r = 0$

By way of a motivating example, assuming 5 and 8 are in the positions 15 and 16, if we pick 5, then we will not consider the numbers 6,7,8. In this case, we will be left with 16 choose 11 possibilities for the other numbers.

We define $t = |c - b| = 1$, $(n-1)$ = number of starting positions and $2(n-1)(n-2)!/n!$ = number of ways of arranging the remaining numbers. The formula thus yields

$$
\begin{aligned}
& \frac{2(n-1)(n-1)(n-2)!}{n!}\binom{n-2}{k-1} \\
= \; & \frac{2(n-1)(n-1)(n-2)!}{n(n-1)(n-2)!}\binom{n-2}{k-1} \\
= \; & \frac{2(n-1)}{n}\binom{n-2}{k-1} \\
= \; & \frac{2(n-1)}{n}\frac{(n-2)!}{(k-1)![(n-2)-(k-1)]!} \\
= \; & \frac{2n(n-1)(n-2)!k(n-k)}{n^2k(k-1)!(n-k-1)!(n-k)} \\
= \; & \frac{2n!k(n-k)}{n^2k!(n-k)!} \\
= \; & 2\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)\left(\frac{n!}{(n-k)!k!}\right) \\
= \; & 2\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)\binom{n}{k}. \qquad (16)
\end{aligned}
$$

4.2.3 Case 2 - Near Successions with $r = 0$

First consider $t = 2, r = 0$, where the two numbers are next to each other but differ by 2. The formula gives

$$\begin{aligned}
& \frac{2(n-1)(n-2)(n-2)!}{n!}\binom{n-3}{k-1} \\
= \; & \frac{2(n-2)}{n}\binom{n-3}{k-1} \\
= \; & \frac{2(n-2)}{n}\binom{n-3}{k-1} \\
= \; & \frac{2(n-2)}{n}\frac{k}{k}\frac{(n-3)!n(n-1)(n-k)(n-k-1)}{(k-1)!n(n-1)(n-k-2)(n-k)(n-k-1)} \\
= \; & \frac{2kn!(n-k)(n-k-1)}{nk!n(n-1)(n-k)!} \\
= \; & \frac{2k}{n^2}\frac{n-k}{n}\frac{(n-k-1)}{n-1}\frac{n!}{(n-k)!k!} \\
= \; & \frac{2k}{n^2}\frac{n-k}{n}\frac{(n-k-1)}{n-1}\binom{n}{k} \\
\leq \; & \frac{2k}{n}\left(\frac{n-k}{n}\right)^2\binom{n}{k}. \qquad (17)
\end{aligned}$$

Now for $t = 3$ and $r = 0$, (for example 5,8). These numbers are next to each other but they differ by 3.

$$\begin{aligned}
& \frac{2(n-1)(n-3)(n-2)!}{n!}\binom{n-4}{k-1} \\
= \; & \frac{2(n-3)}{n}\binom{n-4}{k-1} \\
= \; & \frac{2(n-3)(n-4)!(n-2)(n-1)n(n-k-2)(n-k-1)(n-k)k}{n.k(k-1)!(n-k-3)!(n-k-2)(n-k-1)(n-k-1)(n-k)(n-2)(n-1)n} \\
= \; & \frac{2.k.n!(n-k-2)(n-k-1)(n-k)}{k!(n-k)!(n-2)(n-1)n^2} \\
= \; & \frac{2k}{n}\left(\frac{n-k}{n}\right)\left(\frac{n-k-1}{n-1}\right)\left(\frac{n-k-2}{n-2}\right) \\
\leq \; & 2\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)^3\binom{n}{k}. \qquad (18)
\end{aligned}$$

In general, the total contribution of the $r = 0$ case can be bounded by

$$
\begin{aligned}
&= \left[2\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)\binom{n}{k} + 2\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)^2\binom{n}{k} + 2\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)^3\binom{n}{k} + \ldots \right] \\
&= 2\left(\frac{k}{n}\right)\left(1-\frac{k}{n}\right)\binom{n}{k}\left[1 + \left(1-\frac{k}{n}\right)^2 + \left(1-\frac{k}{n}\right)^3 + \ldots\right] \\
&\leq 2\left(\frac{k}{n}\right)\left(1-\frac{k}{n}\right)\frac{1}{1-\left(1-\frac{k}{n}\right)}\binom{n}{k}, \qquad (19)
\end{aligned}
$$

yielding

Proposition 4.2. *The total contribution of duplicates by adjacent terms is* $\leq 2\left(1-\frac{k}{n}\right)\binom{n}{k}$

We combine these contributions for $k \geq n/2$ to yield

$$
\begin{aligned}
2\sum_{k=n/2+1}^{n}\left(1-\frac{k}{n}\right)\binom{n}{k} &= 2\sum_{k=n/2}^{n}\binom{n}{k} - 2\sum_{k=n/2+1}^{n}\frac{k}{n}\binom{n}{k} \\
&= 2.\frac{1}{2}.2^n - 2\sum_{k=n/2+1}^{n}\frac{k}{n}\frac{n!}{(n-k)!k!} \\
&= 2^n - 2\sum_{k=n/2+1}^{n}\binom{n-1}{k-1} = 2^n - 2\sum_{j=n/2}^{n-1}\binom{n-1}{j} \\
&= 2^{n-1}.
\end{aligned}
\qquad (20)
$$

4.2.4 Other Cases: Non-successions with $r = 1$

In this next step, we consider the case $r = 1$ and $t = 1$. Thus the position difference for the two numbers b, c is 1 and the numerical difference is 1. An example

of such numbers are $5x6$. The formula yields

$$\begin{aligned}
& \frac{2(n-1)(n-2)(n-2)!}{n!}\binom{n-3}{k-1} \\
= \; & \frac{2(n-2)}{n}\binom{n-3}{k-1} \\
= \; & \frac{2(n-2)(n-3)!(n-1)n(n-k-1)(n-k)k}{n.k(k-1)!(n-2-k)!n(n-1)(n-k-1)(n-k)} \\
= \; & 2\binom{n}{k}\frac{k}{n}\frac{(n-k-1)(n-k)}{n(n-1)} \\
\leq \; & 2\binom{n}{k}\frac{k}{n}\frac{(n-k)}{n}\frac{(n-k)}{n}.
\end{aligned} \tag{21}$$

Now we consider $r = 1, t = 2$. As an example, we consider the numbers $5x7$, and note we can either have the number 6 between them or outside. When 6 lies outside of these two numbers, we will have another number to lie between this interval. This gives us two conditions(the case number in-value (eg 6) lies between them and the case when another number lies between them but we have to rule both the 6 and that number out since they can cause a repetition. This gives, via Theorem 4:

$$\begin{aligned}
& \frac{2(n-2)(n-2)}{n!}\left[(n-3)!\binom{n-3}{k-1} + ((n-2)! - (n-3)!)\binom{n-4}{k-1}\right] \\
= \; & \frac{2(n-2)(n-2)}{n!}\left[(n-3)!\binom{n-3}{k-1} + (n-2)!\binom{n-4}{k-1} - (n-3)!\binom{n-4}{k-1}\right] \\
= \; & \frac{2(n-2)(n-2)}{n!}\left[(n-2)!\binom{n-4}{k-1} + (n-3)!\left(\binom{n-3}{k-1} - \binom{n-4}{k-1}\right)\right] \\
= \; & \frac{2(n-2)(n-2)}{n!}\left[(n-2)!\binom{n-4}{k-1} + (n-3)!\binom{n-4}{k-2}\right] \\
= \; & \frac{2(n-2)(n-2)(n-2)!}{n!}\binom{n-4}{k-1} + \frac{2(n-2)(n-2)(n-3)!}{n!}\binom{n-4}{k-2}
\end{aligned} \tag{22}$$

For the first part, we have

$$\begin{aligned}
& \frac{2(n-2)(n-2)(n-2)!}{n(n-1)(n-2)!}\binom{n-4}{k-1} \\
= & \frac{2(n-2)(n-2)(n-4)!}{n(n-1)(n-k-3)!(k-1)!} \\
= & \frac{2(n-2)(n-2)(n-4)!kn(n-1)(n-3)(n-k-2)(n-k-1)(n-k)}{n(n-1)(n-k-3)!(k-1)!k} \\
= & \frac{2n!(n-2)(n-k-2)(n-k-1)k(n-k)}{k!n(n-1)(n-k)!(n-3)} \\
= & 2\binom{n}{k}\frac{n-k}{n} * \frac{n-k-1}{n-1} * \frac{(n-k-2)(n-2)}{(n-1)(n-3)}\frac{k}{n}
\end{aligned} \tag{23}$$

Since $(n-k-2)(n-2) \leq (n-k-1)(n-3)$, (23) reduces to

$$\frac{k}{n}\left(\frac{n-k}{n}\right)^2\binom{n}{k}.$$

For the 2nd part, note that

$$\begin{aligned}
& \frac{2(n-2)(n-2)(n-3)!}{n!}\binom{n-4}{k-2} \\
= & \frac{2(n-2)(n-2)(n-3)!}{n!}\binom{n-4}{k-2} \\
= & \frac{2(n-2)(n-2)(n-3)!n(n-1)(n-3)(n-k-1)(n-k)k(k-1)}{n(n-1)(n-k-1)!(n-k)!} \\
= & 2\frac{k}{n} * \frac{k-1}{n-1} * \frac{n-k}{n} * \frac{n-k-1}{n-1}\frac{1}{n-3}\binom{n}{k} \\
\leq & 2\left(\frac{k}{n}\right)^2 * \left(\frac{n-k}{n}\right)^2 * \frac{1}{n-3}\binom{n}{k}.
\end{aligned} \tag{24}$$

5 CONCLUSION

In conclusion, we see that many of the earlier terms in the formula of Theorem 4 can be analyzed so as to be smaller that $\binom{n}{k}$. In future work, we need to complete this analysis and make further progress towards proving our conjecture.

BIBLIOGRAPHY

[1] Yonah Biers-Ariel, Anant Godbole, and Elizabeth Kelley *Expected Number of Distinct Subsequences in Randomly Generated Binary Strings*. Discrete Mathematics and Theoretical Computer Science 19(2), 10 pages, (2018).

[2] Flaxman, Abraham, Aram W. Harrow, and Gregory B. Sorkin. *Strings with Maximally Many Distinct Subsequences and Substrings*. Electronic Journal of Combinatorics 11.1 (2004).

[3] M. H. Albert, Micah Coleman, Ryan Flynn, and Imre Leader. *Permutations containing many Patterns*. Annals of Combinatorics 11, 265–270, (2007)

[4] Collins, Michael. The Number of Distinct Subsequences of a Random Binary String. Unpublished. `https://arxiv.org/pdf/1310.7288.pdf`.

[5] Allison Miller. *Asymptotic bounds for permutations containing many different patterns*. Journal of Combinatorial Theory Series A, 116, 92–108, (2008).

[6] Micah, Coleman. *An answer to a question by Wilf on packing distinct patterns in a permutation*. Micah Coleman. Electron. J. Combin., 11(1):Note 8, 4 pp. (electronic), 2004.

[7] Allison, T., Godbole, A., Hawley, K., and Kay, B. *Covering n- permutations with (n + 1)-permutations*. Electronic Journal of Combinatorics 20, Paper P6 (2013).

APPENDIX

$n = 3$

```
> result$count <- number
> result
  perm       elements count
1  123     1, 12, 123     3
2  132 1, 12, 21, 132     4
3  213 1, 21, 12, 213     4
4  312 1, 21, 12, 231     4
5  231 1, 12, 21, 312     4
6  321     1, 21, 321     3
>
>
>
> R=((result$count)+1);R# Number of Embedded Sequences for each possible sequence of n
[1] 4 5 5 5 5 4
> table (R )# Yields the frequency table of the number of Embedded Sequences for each possible sequence of n
R
4 5
2 4
>
> S=sum(R);S # Total number of Embedded Sequences
[1] 28
> A=length(R);A
[1] 6
>
> prob.dist<-prop.table(table(R))
> prob.dist
R
        4         5
0.3333333 0.6666667
>
> E<-(S/A);E
[1] 4.666667
>
> V<-sum(unique((result$count)+1)^2*prob.dist)-E^2;V
[1] 0.2222222
>
```

$n = 4$

```
> result
   perm                                elements count
1  1234                        1, 12, 123, 1234     4
2  1243              1, 12, 21, 123, 132, 1243     6
3  1324        1, 12, 21, 132, 123, 213, 1324     7
4  1423        1, 12, 21, 132, 123, 231, 1342     7
5  1342        1, 12, 21, 123, 132, 312, 1423     7
6  1432              1, 12, 21, 132, 321, 1432     6
7  2134              1, 21, 12, 213, 123, 2134     6
8  2143              1, 21, 12, 213, 132, 2143     6
9  3124        1, 21, 12, 231, 213, 123, 2314     7
10 4123              1, 21, 12, 231, 123, 2341     6
11 3142 1, 21, 12, 213, 231, 312, 132, 2413     8
12 4132        1, 21, 12, 231, 321, 132, 2431     7
13 2314        1, 12, 21, 312, 123, 213, 3124     7
14 2413 1, 12, 21, 312, 132, 213, 231, 3142     8
15 3214              1, 21, 12, 321, 213, 3214     6
16 4213        1, 21, 12, 321, 231, 213, 3241     7
17 3412              1, 12, 21, 312, 231, 3412     6
18 4312              1, 21, 12, 321, 231, 3421     6
19 2341              1, 12, 21, 123, 312, 4123     6
20 2431        1, 12, 21, 132, 312, 321, 4132     7
21 3241        1, 21, 12, 213, 321, 312, 4213     7
22 4231        1, 21, 12, 231, 321, 312, 4231     7
23 3421              1, 12, 21, 312, 321, 4312     6
24 4321                    1, 21, 321, 4321     4
>
> R=((result$count)+1);R# Number of Embedded Sequences for each possible sequence of n
 [1] 5 7 8 8 8 7 7 7 8 7 9 8 8 9 7 8 7 7 7 8 8 8 7 5
> table (R )# Yields the frequency table of the number of Embedded Sequences for each possible sequence of n
R
 5  7  8  9
 2 10 10  2
>
> S=sum(R);S # Total number of Embedded Sequences
[1] 178
> A=length(R);A
[1] 24
>
> prob.dist<-prop.table(table(R))
> prob.dist
R
        5         7         8         9
0.08333333 0.41666667 0.41666667 0.08333333
>
>
> E<-(S/A)
> E
[1] 7.416667
>
>
> V<-sum(unique((result$count)+1)^2*prob.dist)-E^2
> V
[1] 0.9097222
```

$n = 5$

```
> result
     perm                                                                elements count
1   12345                                                      1, 12, 123, 1234, 12345     5
2   12354                                            1, 12, 21, 123, 132, 1234, 1243, 12354     8
3   12435                                 1, 12, 21, 123, 132, 213, 1243, 1234, 1324, 12435    10
4   12534                                 1, 12, 21, 123, 132, 231, 1243, 1234, 1342, 12453    10
5   12453                                 1, 12, 21, 123, 132, 312, 1234, 1243, 1423, 12534    10
6   12543                                       1, 12, 21, 123, 132, 321, 1243, 1432, 12543     9
7   13245                                 1, 12, 21, 132, 123, 213, 1324, 1234, 2134, 13245    10
8   13254                                 1, 12, 21, 132, 123, 213, 1324, 1243, 2143, 13254    10
9   14235                      1, 12, 21, 132, 123, 231, 213, 1342, 1324, 1234, 2314, 13425    12
10  15234                                 1, 12, 21, 132, 123, 231, 1342, 1234, 2341, 13452    10
11  14253           1, 12, 21, 132, 123, 213, 231, 312, 1324, 1342, 1423, 1243, 2413, 13524    14
12  15243                      1, 12, 21, 132, 123, 231, 321, 1342, 1432, 1243, 2431, 13542    12
13  13425                      1, 12, 21, 123, 132, 312, 213, 1423, 1234, 1324, 3124, 14235    12
14  13524           1, 12, 21, 123, 132, 312, 213, 231, 1423, 1243, 1324, 1342, 3142, 14253    14
15  14325                           1, 12, 21, 132, 123, 321, 213, 1432, 1324, 3214, 14325    11
16  15324                 1, 12, 21, 132, 123, 321, 231, 213, 1432, 1342, 1324, 3241, 14352    13
17  14523                           1, 12, 21, 123, 132, 312, 231, 1423, 1342, 3412, 14523    11
18  15423                           1, 12, 21, 132, 123, 321, 231, 1432, 1342, 3421, 14532    11
19  13452                                1, 12, 21, 123, 132, 312, 1234, 1423, 4123, 15234    10
20  13542                      1, 12, 21, 123, 132, 312, 321, 1243, 1423, 1432, 4132, 15243    12
21  14352                 1, 12, 21, 132, 123, 213, 321, 312, 1324, 1432, 1423, 4213, 15324    13
22  15342                 1, 12, 21, 132, 123, 231, 321, 312, 1342, 1432, 1423, 4231, 15342    13
23  14532                           1, 12, 21, 123, 132, 312, 321, 1423, 1432, 4312, 15423    11
24  15432                                            1, 12, 21, 132, 321, 1432, 4321, 15432     8
25  21345                                            1, 21, 12, 213, 123, 2134, 1234, 21345     8
26  21354                                1, 21, 12, 213, 123, 132, 2134, 2143, 1243, 21354    10
27  21435                                1, 21, 12, 213, 132, 123, 2143, 2134, 1324, 21435    10
28  21534                           1, 21, 12, 213, 132, 123, 231, 2143, 2134, 1342, 21453    11
29  21453                           1, 21, 12, 213, 123, 132, 312, 2134, 2143, 1423, 21534    11
30  21543                                      1, 21, 12, 213, 132, 321, 2143, 1432, 21543     9
31  31245                                1, 21, 12, 231, 213, 123, 2314, 2134, 1234, 23145    10
32  31254                           1, 21, 12, 231, 213, 132, 123, 2314, 2143, 1243, 23154    11
33  41235                                1, 21, 12, 231, 213, 123, 2341, 2314, 1234, 23415    10
```

33	41235					1,	21,	12,	231,	213,	123,	2341,	2314,	1234,	23415	10
34	51234							1,	21,	12,	231,	123,	2341,	1234,	23451	8
35	41253			1,	21,	12,	231,	213,	312,	123,	132,	2314,	2341,	2413,	1243,	23514	13
36	51243					1,	21,	12,	231,	321,	123,	132,	2341,	2431,	1243,	23541	11
37	31425		1,	21,	12,	213,	231,	312,	123,	132,	2413,	2134,	2314,	3124,	1324,	24135	14
38	31524		1,	21,	12,	213,	231,	312,	132,	123,	2413,	2143,	2314,	3142,	1342,	24153	14
39	41325			1,	21,	12,	231,	213,	321,	132,	123,	2431,	2314,	3214,	1324,	24315	13
40	51324			1,	21,	12,	231,	321,	132,	123,	213,	2431,	2341,	3241,	1324,	24351	13
41	41523			1,	21,	12,	213,	231,	312,	132,	123,	2413,	2341,	3412,	1342,	24513	13
42	51423				1,	21,	12,	231,	321,	132,	123,	2431,	2341,	3421,	1342,	24531	12
43	31452			1,	21,	12,	213,	231,	123,	312,	132,	2134,	2413,	4123,	1423,	25134	13
44	31542			1,	21,	12,	213,	231,	132,	312,	321,	2143,	2413,	4132,	1432,	25143	13
45	41352	1,	21,	12,	231,	213,	321,	312,	123,	132,	2314,	2431,	2413,	4213,	1423,	25314	15
46	51342			1,	21,	12,	231,	321,	123,	132,	312,	2341,	2431,	4231,	1423,	25341	13
47	41532			1,	21,	12,	213,	231,	312,	321,	132,	2413,	2431,	4312,	1432,	25413	13
48	51432						1,	21,	12,	231,	321,	132,	2431,	4321,	1432,	25431	10
49	23145						1,	12,	21,	312,	123,	213,	3124,	1234,	2134,	31245	10
50	23154					1,	12,	21,	312,	123,	213,	132,	3124,	1243,	2143,	31254	11
51	24135		1,	12,	21,	312,	132,	123,	213,	231,	3142,	3124,	1324,	2134,	2314,	31425	14
52	25134			1,	12,	21,	312,	132,	213,	123,	231,	3142,	1342,	2134,	2341,	31452	13
53	24153		1,	12,	21,	312,	123,	132,	213,	231,	3124,	3142,	1423,	2143,	2413,	31524	14
54	25143			1,	12,	21,	312,	132,	213,	231,	321,	3142,	1432,	2143,	2431,	31542	13
55	32145								1,	21,	12,	321,	213,	123,	3214,	2134,	32145	9
56	32154								1,	21,	12,	321,	213,	132,	3214,	2143,	32154	9
57	42135					1,	21,	12,	321,	231,	213,	123,	3241,	3214,	2314,	2134,	32415	12
58	52134						1,	21,	12,	321,	231,	213,	123,	3241,	2341,	2134,	32451	11
59	42153			1,	21,	12,	321,	213,	231,	312,	132,	3214,	3241,	2413,	2143,	32514	13
60	52143						1,	21,	12,	321,	231,	213,	132,	3241,	2431,	2143,	32541	11
61	34125						1,	12,	21,	312,	123,	231,	213,	3412,	3124,	2314,	34125	11
62	35124			1,	12,	21,	312,	132,	231,	213,	123,	3412,	3142,	2314,	2341,	34152	13
63	43125						1,	21,	12,	321,	213,	231,	123,	3421,	3214,	2314,	34215	11
64	53124					1,	21,	12,	321,	231,	213,	123,	3421,	3241,	2341,	2314,	34251	12
65	45123								1,	12,	21,	312,	231,	123,	3412,	2341,	34512	9
66	54123								1,	21,	12,	321,	231,	123,	3421,	2341,	34521	9
67	34152			1,	12,	21,	312,	123,	213,	231,	132,	3124,	3412,	4123,	2413,	35124	13
68	35142		1,	12,	21,	312,	132,	213,	231,	321,	3142,	3412,	4132,	2413,	2431,	35142	14
68	35142		1,	12,	21,	312,	132,	213,	231,	321,	3142,	3412,	4132,	2413,	2431,	35142	14
69	43152				1,	21,	12,	321,	213,	231,	312,	132,	3214,	3421,	4213,	2413,	35214	13
70	53142		1,	21,	12,	321,	231,	213,	312,	132,	3241,	3421,	4231,	2431,	2413,	35241	14
71	45132						1,	12,	21,	312,	231,	321,	132,	3412,	4312,	2431,	35412	11
72	54132								1,	21,	12,	321,	231,	132,	3421,	4321,	2431,	35421	10
73	23415								1,	12,	21,	123,	312,	213,	4123,	1234,	3124,	41235	10
74	23514				1,	12,	21,	123,	312,	132,	213,	231,	4123,	1243,	3124,	3142,	41253	13
75	24315				1,	12,	21,	132,	312,	123,	213,	321,	4132,	1324,	3124,	3214,	41325	13
76	25314	1,	12,	21,	132,	312,	123,	213,	321,	231,	4132,	1342,	3142,	3124,	3241,	41352	15
77	24513				1,	12,	21,	123,	312,	132,	213,	231,	4123,	1423,	3142,	3412,	41523	13
78	25413				1,	12,	21,	132,	312,	213,	321,	231,	4132,	1432,	3142,	3421,	41532	13
79	32415					1,	21,	12,	213,	321,	312,	123,	4213,	2134,	3214,	3124,	42135	12
80	32514				1,	21,	12,	213,	321,	312,	132,	231,	4213,	2143,	3214,	3142,	42153	13
81	42315				1,	21,	12,	231,	321,	213,	312,	123,	4231,	2314,	3214,	3124,	42315	13
82	52314				1,	21,	12,	231,	321,	312,	123,	213,	4231,	2341,	3241,	3124,	42351	13
83	42513		1,	21,	12,	213,	321,	231,	312,	132,	4213,	2413,	3241,	3412,	3142,	42513	14
84	52413		1,	21,	12,	231,	321,	312,	132,	213,	4231,	2431,	3241,	3421,	3142,	42531	14
85	34215						1,	12,	21,	312,	123,	321,	213,	4312,	3124,	3214,	43125	11
86	35214				1,	12,	21,	312,	132,	321,	213,	231,	4312,	3142,	3214,	3241,	43152	13
87	43215										1,	21,	12,	321,	213,	4321,	3214,	43215	8
88	53214								1,	21,	12,	321,	231,	213,	4321,	3241,	3214,	43251	10
89	45213						1,	12,	21,	312,	321,	231,	213,	4312,	3412,	3241,	43512	11
90	54213								1,	21,	12,	321,	231,	213,	4321,	3421,	3241,	43521	10
91	34512									1,	12,	21,	123,	312,	231,	4123,	3412,	45123	9
92	35412						1,	12,	21,	132,	312,	231,	321,	4132,	3412,	3421,	45132	11
93	43512						1,	21,	12,	213,	321,	312,	231,	4213,	3421,	3412,	45213	11
94	53412								1,	21,	12,	231,	321,	312,	4231,	3421,	3412,	45231	10
95	45312								1,	12,	21,	312,	321,	231,	4312,	3412,	3421,	45312	10
96	54312											1,	21,	12,	321,	231,	4321,	3421,	45321	8
97	23451											1,	12,	21,	123,	312,	1234,	4123,	51234	8
98	23541						1,	12,	21,	123,	312,	132,	321,	1243,	4123,	4132,	51243	11
99	24351				1,	12,	21,	132,	123,	312,	213,	321,	1324,	4132,	4123,	4213,	51324	13
100	25341				1,	12,	21,	132,	312,	123,	231,	321,	1342,	4132,	4123,	4231,	51342	13
101	24531					1,	12,	21,	123,	132,	312,	321,	1423,	4123,	4132,	4312,	51423	12
102	25431								1,	12,	21,	132,	312,	321,	1432,	4132,	4321,	51432	10
103	32451							1,	21,	12,	213,	321,	123,	312,	2134,	4213,	4123,	52134	11

```
103 32451                       1, 21, 12, 213, 321, 123, 312, 2134, 4213, 4123, 52134    11
104 32541                       1, 21, 12, 213, 321, 132, 312, 2143, 4213, 4132, 52143    11
105 42351            1, 21, 12, 231, 213, 321, 312, 123, 2314, 4231, 4213, 4123, 52314    13
106 52341                       1, 21, 12, 231, 321, 123, 312, 2341, 4231, 4123, 52341    11
107 42531      1, 21, 12, 213, 231, 321, 312, 132, 2413, 4213, 4231, 4312, 4132, 52413    14
108 52431                 1, 21, 12, 231, 321, 132, 312, 2431, 4231, 4321, 4132, 52431    12
109 34251                 1, 12, 21, 312, 123, 213, 321, 3124, 4312, 4123, 4213, 53124    12
110 35241      1, 12, 21, 312, 132, 213, 321, 231, 3142, 4312, 4132, 4213, 4231, 53142    14
111 43251                            1, 21, 12, 321, 213, 312, 3214, 4321, 4213, 53214    10
112 53241                 1, 21, 12, 321, 231, 213, 312, 3241, 4321, 4231, 4213, 53241    12
113 45231                            1, 12, 21, 312, 231, 321, 3412, 4312, 4231, 53412    10
114 54231                            1, 21, 12, 321, 231, 312, 3421, 4321, 4231, 53421    10
115 34521                                  1, 12, 21, 123, 312, 321, 4123, 4312, 54123     9
116 35421                            1, 12, 21, 132, 312, 321, 4132, 4312, 4321, 54132    10
117 43521                            1, 21, 12, 213, 321, 312, 4213, 4321, 4312, 54213    10
118 53421                            1, 21, 12, 231, 321, 312, 4231, 4321, 4312, 54231    10
119 45321                                       1, 12, 21, 312, 321, 4312, 4321, 54312     8
120 54321                                                  1, 21, 321, 4321, 54321     5
>
> R=((result$count)+1);R# Number of Embedded Sequences for each possible sequence of n
  [1]  6  9 11 11 11 10 11 11 13 11 15 13 13 15 12 14 12 12 11 13 14 14 12  9  9 11 11 12 12 10 11 12 11  9 14
 12 15 15 14
 [40] 14 14 13 14 14 16 14 14 11 11 12 15 14 15 14 10 10 13 12 14 12 12 14 12 13 10 10 14 15 14 15 12 11 11 14
 14 16 14 14
 [79] 13 14 14 14 15 15 12 14  9 11 12 11 10 12 12 11 11  9  9 12 14 14 13 11 12 12 14 12 15 13 13 15 11 13 11
 11 10 11 11
[118] 11  9  6
> table (R )# Yields the frequency table of the number of Embedded Sequences for each possible sequence of n
R
 6  9 10 11 12 13 14 15 16
 2  8  8 26 22 12 28 12  2
>
> S=sum(R);S # Total number of Embedded Sequences
[1] 1474
> A=length(R);A
[1] 120
>
> prob.dist<-prop.table(table(R))
> prob.dist
R
         6          9         10         11         12         13         14         15         16
0.01666667 0.06666667 0.06666667 0.21666667 0.18333333 0.10000000 0.23333333 0.10000000 0.01666667
>
>
> E<-(S/A)
> E
[1] 12.28333
>
>
> V<-sum(unique((result$count)+1)^2*prob.dist)-E^2
> V
[1] -4.196944
>
```

$n = 6$

```
> table (R )# Yields the frequency table of the number of Embedded Sequences for each possible sequence of n
R
 7 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
 2  8 14 20 18 18 58 48 52 72 70 36 80 76 66  6 44 24  8
>
> S=sum(R);S # Total number of Embedded Sequences
[1] 15224
> A=length(R);A
[1] 720
>
> prob.dist<-prop.table(table(R))
> prob.dist
R
          7          11          13          14          15          16          17          18          19
       20
0.002777778 0.011111111 0.019444444 0.027777778 0.025000000 0.025000000 0.080555556 0.066666667 0.072222222 0.
100000000
         21          22          23          24          25          26          27          28          29
0.097222222 0.050000000 0.111111111 0.105555556 0.091666667 0.008333333 0.061111111 0.033333333 0.011111111
>
>
> E<-(S/A)
> E
[1] 21.14444
>
>
> V<-sum(unique((result$count)+1)^2*prob.dist)-E^2
> V
[1] 25.63469
>
```

$n = 7$

```
> S=sum(R);S # Total number of Embedded Sequences
[1] 189524
> A=length(R);A
[1] 5040
>
> prob.dist<-prop.table(table(R))
> prob.dist
R
          8           13           16           17           19           20           21           22
    23
0.0003968254 0.0015873016 0.0023809524 0.0055555556 0.0051587302 0.0019841270 0.0015873016 0.0130952381 0.0150
793651
          24           25           26           27           28           29           30           31
    32
0.0142857143 0.0103174603 0.0202380952 0.0174603175 0.0206349206 0.0333333333 0.0388888889 0.0285714286 0.0373
015873
          33           34           35           36           37           38           39           40
    41
0.0365079365 0.0380952381 0.0507936508 0.0412698413 0.0575396825 0.0329365079 0.0515873016 0.0428571429 0.0333
333333
          42           43           44           45           46           47           48           49
    50
0.0460317460 0.0730158730 0.0301587302 0.0373015873 0.0230158730 0.0230158730 0.0238095238 0.0230158730 0.0079
365079
          51           52           53           54           55           56
0.0261904762 0.0142857143 0.0079365079 0.0047619048 0.0047619048 0.0019841270
>
> E<-(S/A);E
[1] 37.60397
>
> V<-sum(unique((result$count)+1)^2*prob.dist)-E^2;V
[1] 76.07927
```

$n = 8$

```
> table (R )# Yields the frequency table of the number of Embedded Sequences for each possible sequence of n
R
   9   15   19   20   21   23   24   26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
  41   42
   2    8   12   20   14   26   20    8   48   36   72   48  108   96   72   68  146   88  134  228  224  188
 426  238
  43   44   45   46   47   48   49   50   51   52   53   54   55   56   57   58   59   60   61   62   63   64
  65   66
 266  336  424  258  416  504  548  640  450  386  908  592  756  624  808  736  944  504  990  896  856  856
 892 1056
  67   68   69   70   71   72   73   74   75   76   77   78   79   80   81   82   83   84   85   86   87   88
  89   90
1136  900  960  796  860  880 1014  856  968  816  712  888  916  746 1000  746  716  604  540  604  608  452
 434  348
  91   92   93   94   95   96   97   98   99  100  101  102  103  104  105  106  107  108  109  110
 356  248  468  306  326  458  340  170  220  248  144  156   96  108   64   72   16   32    8    8
>
> S=sum(R);S # Total number of Embedded Sequences
[1] 2767428
> A=length(R);A
[1] 40320
> E<-(S/A);E
[1] 68.63661
>
> V<-sum(unique((result$count)+1)^2*prob.dist)-E^2;V
[1] 201.3242
>
```

VITA

EVELYN FOKUOH

Education:	B.S. Mathematics with Economics, University of Cape Coast, Cape Coast, Ghana 2013 M.S. Mathematical Sciences, East Tennessee State University Johnson city, Tennessee 2018
Professional Experience:	Graduate Teaching/Research Assistant, East Tennessee State University, Johnson city,TN 2016-2018