Generalised inversion frequency distribution

Author:
Tong, Sin Keong

Publication Date:
2019

DOI:
https://doi.org/10.26190/unsworks/22089

GENERALISED INVERSION FREQUENCY DISTRIBUTION

Tong Sin Keong

School of Mathematics and Statistics
UNSW Sydney

December 2019

Submitted in partial fulfilment of the requirements of the degree of
Master of Mathematics

Thesis/Dissertation Sheet

Surname/Family Name	:	**Tong**
Given Name/s	:	**Sin Keong**
Abbreviation for degree as give in the University calendar	:	**M. Sc.**
Faculty	:	**Science**
School	:	**Mathematics and Statistics**
Thesis Title	:	GENERALISED INVERSION FREQUENCY DISTRIBUTION

Abstract 350 words maximum: (PLEASE TYPE)

The thesis is a study of the distribution of inversion counts for the permutations of multisets by a four-tier architecture of integers, partitions, multisets and the permutations of the multisets. It introduces two insertion methods to link the hierarchical and peer to peer relationships between these entities. It centers around the generating function for the inversion count distribution for the permutation of the multisets. The main result is a recursive function for the parent/child relationship between the permutations of multisets.

The thesis also studies the link between the coefficients of the generating polynomial and the Ferrers diagram and also delivers an integer partition formula as a special case of the closed form. It also analyses the conformance of natural and computer generated sequences with the expected distribution of partition and inversion counts.

INCLUSION OF PUBLICATIONS STATEMENT

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in their thesis in lieu of a Chapter if:

- The candidate contributed greater than 50% of the content in the publication and is the "primary author", ie. the candidate was responsible primarily for the planning, execution and preparation of the work for publication
- The candidate has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not:

☒ This thesis contains no publications, either published or submitted for publication
(if this box is checked, you may delete all the material on page 2)

☐ Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement
(if this box is checked, you may delete all the material on page 2)

☐ This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below

CANDIDATE'S DECLARATION

I declare that:

- I have complied with the UNSW Thesis Examination Procedure
- where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.

Candidate's Name	**Signature**	**Date (dd/mm/yy)**

POSTGRADUATE COORDINATOR'S DECLARATION *To only be filled in where publications are used in lieu of Chapters* I declare that: • the information below is accurate • where listed publication(s) have been used in lieu of Chapter(s), their use complies with the UNSW Thesis Examination Procedure • the minimum requirements for the format of the thesis have been met.		
PGC's Name	**PGC's Signature**	**Date (dd/mm/yy)**

For each publication incorporated into the thesis in lieu of a Chapter, provide all of the requested details and signatures required

Details of publication #1: *Full title:* *Authors:* *Journal or book name:* *Volume/page numbers:* *Date accepted/ published:*						
Status	*Published*		*Accepted and In press*		*In progress (submitted)*	
The Candidate's Contribution to the Work Insert text describing how the candidate has contributed to the work						
Location of the work in the thesis and/or how the work is incorporated in the thesis: Insert text						
PRIMARY SUPERVISOR'S DECLARATION I declare that: • the information above is accurate • this has been discussed with the PGC and it is agreed that this publication can be included in this thesis in lieu of a Chapter • All of the co-authors of the publication have reviewed the above information and have agreed to its veracity by signing a 'Co-Author Authorisation' form.						
Primary Supervisor's name		**Primary Supervisor's signature**			**Date (dd/mm/yy)**	

Add additional boxes if required

COPYRIGHT STATEMENT

Signed ……………………………………………..........................

Date ……………………………………………...........................

AUTHENTICITY STATEMENT

‘I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis.’

Signed ……………………………………………..........................

Date ……………………………………………...........................

ORIGINALITY STATEMENT

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Signed ……………………………………………..............

Date ……………………………………………..............

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

- Reproduce it and provide a copy to another member of the University; and/or,
- Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the purpose of future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct, and am aware of any potential plagiarism penalties which may apply.

By signing this declaration I am agreeing to the statements and conditions above.

Signed: ______________________ Date: __________

Acknowledgements

To my father who taught me to give more than I received and motivated me to look beyond the horizon.
To my high school maths teacher Brother McKay, for his kindness and for providing me with the nourishment for seeking the truth embodied in mathematics.
To my supervisor, Dr Thomas Britz who has elevated me to a higher plane through his abstract and critical thinking. His guidance and reviews provided me with the enlightenment and sustenance for completion of the thesis.
I am deeply indebted to Professor Catherine Greenhill for providing the support in editing and streamlining the ideas in the thesis. She guided me to alternative proofs and also helped me to discover the links to works in related area.
To Dr Kooguan Choo who spent many hours proofreading the initial article and shaping it into a mathematical document. He exemplifies the classical Chinese scholastic values and selflessness.
To Dr Diana Combe for her incisive assessment of the abstract design of the thesis and invaluable feedback.
To Professor Alexander Molev for his illuminating critics of the initial version of the document.
To Dr Les Gordon who has been my companion in the journey through mathematics. I am also grateful to him for providing insights deserving showers of gold coins and feedback on the thesis.

Margaret, this is dedicated to the times when our paths merged.

Abstract

The thesis is a study of the distribution of inversion counts for the permutations of multisets by a four-tier architecture of integers, partitions, multisets and the permutations of the multisets. It introduces two insertion methods to link the hierarchical and peer to peer relationships between these entities. It centers around the generating function for the inversion count distribution for the permutation of the multisets. The main result is a recursive function for the parent/child relationship between the permutations of multisets.

The secondary result is a rediscovery of the closed form expression for the generating function as a product of Gaussian binomial coefficients, also known as q-nomials. For a partition $n = n_1 + n_2 + \cdots + n_k$, the inversion count distribution is given by the coefficients of the polynomial

$$P(n_1, n_2, \ldots, n_k) = \frac{G(n_1 + n_2 + \cdots + n_k)}{G(n_1)G(n_2)\cdots G(n_k)}$$
$$\text{where} \quad G(n) = (x^n - 1)(x^{n-1} - 1)\ldots(x - 1).$$

The thesis also studies the link between the coefficients of the generating polynomial and the Ferrers diagram and also delivers an integer partition formula as a special case of the closed form. It also analyses the conformance of natural and computer generated sequences with the expected distribution of partition and inversion counts.

Contents

Chapter 1

Concepts

1.1 Inversion

Five cards each with a digit from 0 to 4 are laid on the table from left to right, as illustrated below:

$$4 \quad 1 \quad 0 \quad 4 \quad 2$$

Take the leftmost card (here with value 4) and record the number I_1 of cards to its right of lower value. In the example above, $I_1 = 3$ since 1, 0 and 2 are smaller than 4. Repeat this count for the remaining cards, recording $I_2 = 1, I_3 = 0$ and $I_4 = 1$. Now, let $I = I_1 + I_2 + I_3 + I_4 = 5$. This is the *inversion count* $I = \mathrm{Inv}(s)$ for the sequence $s = 41042$. It can be shown by Lemma 2.2 to follow that the maximum inversion count for a sequence of 5 numbers of which 4 are distinct is 9. In this thesis, we consider inversion count for various types of sequences, some random and others not. We address the natural question: "What is the probability distribution of I, given some random sequence?". Once such a probability distribution is known, it can be used to analyse the digits of classical irrational numbers such as π and e to determine whether they conform to expected distributions. We can also use inversion count frequency distribution to analyse the efficiency of sorting algorithms and also measure the randomness of quasi-random sequences generated by computers.

1.2 Inversion distribution

The objective of the thesis is to deliver the expected distribution of inversion count for the permutations of the elements of a multiset (See Section 2.1.2) This section presents some of terminologies and related concepts at a high level using the digits 0 to 9. It will be assumed that $d_i \leq d_j$ for $1 \leq i < j \leq 5$.

Let $S = \{00000, 00001, \ldots, 99999\}$ be the set of 5-digits numbers. For $s \in S$, let $Inv(s)$ be the inversion count of its digits. The multiplicities of the digits of a 5 digit number naturally induces 7 partitions of 5, these being: '1-1-1-1-1', '2-1-1-1', '2-2-1', '3-2', '3-1-1', '4-1', '5'. Let $S_1 \subset S$ be those numbers where the digits are distinct ('1-1-1-1-1'), so that $|S_1| = 10 \times 9 \times 8 \times 7 \times 6 = 30,240$. For $s_1 \in S_1$, $0 \leq Inv(s_1) \leq 10$. For instance, $Inv(25689) = 0$, $Inv(94310) = 10$. We are interested in the relative frequency of the Inversion Frequency Distribution (IFD) in $I_F(S_1)$. The result is tabulated in the first row of Table 1.1 below. Here, $I_F(S_1) = (f_0, f_1, \ldots, f_{10})$ where $f_0 = 1, f_1 = 4, f_2 = 9, \ldots, f_{10} = 1$. Note that the sum of the row is $120 = 5!$ which is the factor to obtain actual frequencies. S_1 corresponds to the integer partition $5 = 1 + 1 + 1 + 1 + 1$ and is denoted as (1,1,1,1,1).

Next, we turn the attention to those 5-digit numbers S_2 which have 4 distinct digits with one repeated digit ('2-1-1-1') (e.g., 03074) with $|S_2| = 10 \times \frac{9\times8\times7}{3!} \times \frac{5!}{2} = 50,400$. For $s_2 \in S_2$, $0 \leq Inv(s_2) \leq 9$. For instance, $Inv(02256) = 0$, $Inv(77641) = 9$. S_2 is denoted as $(2,1,1,1)$ and the relative frequency is given by the second row in Table 1.1. Let $\sigma(S)$ denote the permutations for the set S. For instance, $\sigma(\{1,2,2,3,4\})$ contains the 30 permutations of 1,2,2,3,4. The natural question that arises is whether $I_F(\{d_1, d_2, d_2, d_3, d_4\}) = I_F(\{d_1, d_1, d_2, d_3, d_4\})$? A formal proof is given by the Family Partition Theorem (Theorem 6.3) which provides a direct proof that the IFD is invariant of the ranking of the elements.

Table 1.1 also extends the calculations for the partitions (1-2-2), (1-1-3), (2-3), (1-4), (5). Although the table can be computed, the numbers quickly get out of control for large number of digits. The final row is the weighted sum of the rows by multiplying IFD by the size of the dataset.

Table 1.1: Inversion distribution table for $n = 5$

	Inversion Frequency Distribution (IFD)											
Partition	0	1	2	3	4	5	6	7	8	9	10	Count
1-1-1-1-1	1	4	9	15	20	22	20	15	9	4	1	30,240
1-1-1-2	1	3	6	9	11	11	9	6	3	1		50,400
1-2-2	1	2	4	5	6	5	4	2	1			10,800
1-1-3	1	2	3	4	4	3	2	1				7,200
2-3	1	1	2	2	2	1	1					900
1-4	1	1	1	1	1							450
5	1											10
Total	2002	5148	10098	14850	18150	17754	14850	9900	5148	1848	252	100,000

An important objective of the thesis is to construct IFDs at the partition level. Observe also that in Table 1.1, IFD for a partition is symmetrical about the median position. For instance, for the row '1-4', $f_0 = f_4, f_1 = f_3$. However, it is not symmetrical for the column total. The thesis analyses the partition and inversion count distribution for the digits of irrational numbers and computer generated numbers in Chapter 9.

1.3 Generating function for the symmetric group

The inversion count distribution for each partition is associated with a generating function. The IFD for the partition '1-1-1-1-1' is (1, 4, 9, 15, 20, 22, 20, 15, 9, 4, 1) and this is represented by the generating polynomial where the coefficient of x^k corresponds to the frequency count of the inversion count k:

$$P(1-1-1-1-1) = 1+4x+9x^2+15x^3+20x^4+22x^5+20x^6+15x^7+9x^8+4x^9+x^{10}\,. \tag{1.1}$$

Note that the sum of coefficients of P(1-1-1-1-1) is $5! = 120$.

1.4 Generating function for the partitions with repeating elements

For the partition '1-1-1-2', the permutations of the sets $S_2 = \{d_1, d_2, d_3, d_4, R\}$ where $0 \leq d_1 \leq d_2 \leq d_3 \leq d_4$ and $R = d_1, d_2, d_3, d_4$ spans all the 5-digits number with one repeating digit. There are $\frac{10\times9\times8\times7}{3!} \times 4 = 840$ permutations of the elements of S_2. The collection of S_2 is defined as the *partition family* for '1-1-1-2'. For fixed values of d_1, d_2, d_3, R, the $\frac{5!}{2} = 60$ permutations can be split into inversion counts of 0 to 9. It is natural to ask whether the IFD (1,3,6,9,11,11,9,6,3,1) accounts for the permutations of both the sets $S_2' = \{3, 3, 5, 7, 9\}, S_2'' = \{0, 2, 6, 6, 8\}$? The Family Partition Theorem (Theorem 6.3) provides a direct proof that the IFD is invariant of the choice of elements. The Closed Form Theorem (Theorem 8.6) also provides an indirect proof.

Chapter 2

Notation, terminology and preliminary results

2.1 Notation

This section defines the notation and provides examples of how they are used.

2.1.1 Set

Let $S = \{a, b, c, \ldots\}$ be a set with total order $a \preceq b \preceq c \preceq \cdots$.

In the thesis, the set S will be the set of the first 10^n positive integers, where $n \in \mathbb{Z}^+$, represented as strings of length n:

$$\overbrace{0 \cdots 0}^{n}, \ldots, \overbrace{9 \cdots 9}^{n} .$$

In Chapter 9, the values $n = 6, 7, 8, 9$ will be used. For the purpose of this thesis, the ordering $\preceq$ is simply the usual integer order $a \leq b \leq c \leq \cdots$.

2.1.2 Multiset

A *multiset* S is a collection of elements in which elements may be repeated. In the thesis, the elements are formed by the concatenation of digits. The distinct elements of a multiset will be denoted as $\{e_1, e_2, \ldots, e_k\}$. Unless otherwise stated, it will be assumed that $e_i < e_j$ when $i < j$.

Associated with each element e_i is the *multiplicity* n_i which is the number of times the element is repeated in S. Given the elements $e_1, e_2, \ldots, e_k$, each multiset in this thesis can be represented simply as $S = [n_1, n_2, \ldots, n_k]$ where the multiplicities are associated with each of the elements $e_1, e_2, \ldots, e_k$, respectively. For instance, the set $S = \{a, a, a, b, c, c\}$ can be represented as $[3, 1, 2]$. It may also be denoted as $\{a^3bc^2\}$.

2.1.3 Rank of multiset $R(S)$

For a multiset $S = [n_1, n_2, \ldots, n_k]$, the *rank* $R(S) = k$ is the number of distinct elements in S.

2.1.4 Permutation of multiset $\sigma(S)$

The *permutation set* formed by the elements of S is defined as $\sigma(S)$. The elements of $\sigma(S)$ are called *sequences*. For instance if $S = \{a, b, b, c\}$, then

$$\sigma(S) = \{abbc, abcb, acbb, babc, bacb, bbac, bbca, bcab, bcba, cabb, cbab, cbba\} . \quad (2.1)$$

2.1.5 *Inversion count Inv*(s)

The *inversion count* $\mathrm{inv}(s)$ of any sequence $s = s_1, \ldots, s_n$ of elements of s is the number of pairs of elements in s which are out of order:

$$\mathrm{Inv}(s) = \left|\{(i,j) \ : \ 1 \le i < j \le N\,,\ s_i > s_j\}\right|.$$

For a multiset, $m(S)$ denotes the maximum inversion count of the permutations of S. In (2.1), the element *cbba* has inversion count 5 and $m(S) = 5$. Lemma 2.2 expresses $m(S)$ for each multiset $S = [n_1, n_2, \ldots, n_k]$.

2.1.6 *Inversion frequency distribution* $I_F(S)$

The *Inversion Frequency Distribution (IFD)* of multiset S is the $(m(S)+1)$-tuple

$$I_F(S) = (f_0, f_1, \ldots, f_{m(S)})$$

where, for $0 \le i \le m(S)$, the number f_i is the number of sequences in $\sigma(S)$ with inversion count i:

$$f_i = \left|\{s \in \sigma(S) \ : \ \mathrm{Inv}(s) = i\}\right|.$$

Table 2.1: Inversion frequency distribution for $\sigma[2,1,1]$

Inv(s)	Permutation
0	aabc
1	aacb, abac
2	abca, acab, baac
3	acba, baca, caab
4	bcaa, caba
5	cbaa

From Table 2.1, the number of sequences with inversion counts $0,1,2,3,4,5$ are $1,2,3,3,2,1$, respectively. Thus,

$$f_0 = 1\,,\ f_1 = 2\,,\ f_2 = 3\,,\ f_3 = 2\,,\ f_4 = 2\,,\ f_5 = 1\,, \quad m(S) = 5\,, \quad I_F(S) = (1,2,3,3,2,1)\,.$$

2.1.7 *Generating polynomial* $P(S)$

The *generating polynomial* $P(S)$ is a representation of $I_F(S) = (f_0, f_1, \ldots, f_{m(S)})$ in polynomial form:

$$P(S) = \sum_{i=0}^{m(S)} f_i\, x^i\,.$$

For the previous example where $S = [2,1,1]$ and $\sigma(S) = (1,2,3,3,2,1)$,

$$P(S) = 1 + 2x + 3x^2 + 3x^3 + 2x^4 + x^5\,.$$

It has an important role in the closed form expression for, as well as the recursive calculations of, the inversion count frequency distribution; see Theorem 7.2. The generating polynomial also acts an operator in the parent/child relationship between partitions; see Examples 4.2 and 4.3.

As we are concerned only with the coefficients of the polynomial, it will be assumed that $x \neq 1$.

2.1.8 *Cayley's notation*

Cayley's notation was used by P.A. MacMahon [10]. He remarked: "This notation is exceeding illuminating, and is a striking example of mathematics that has gained by an appropriate notation". We will however use the modified notation $G(n)$ to avoid notational ambiguities later in the thesis.

$$G(n) = (x^n - 1)(x^{n-1} - 1) \cdots (x - 1), n \in \mathcal{Z}^+ .$$

2.1.9 *Partition family*

The *partition family* $\mathcal{F}(n_1, n_2, \ldots, n_k)$ is the collection of permutation sets on any permutation of the multiplicities. For instance,

$$\mathcal{F}(1,2,3) = \{\sigma[1,2,3], \sigma[1,3,2], \sigma[2,1,3], \sigma[2,3,1], \sigma[3,1,2], \sigma[3,2,1]\} .$$

The partition family establishes a one-to-many relationship between the positive integer partitions of an integer n and the multisets with multiplicities given by the permutations of the summands of the partition.

2.1.10 *Partial integer partition count $A(n, p, m)$*

Let $A(n, p, m)$ be the number of partitions of a positive integer n into p parts each of size at most m. For instance, $A(6, 3, 4)$ is the number of partitions of 6 into 3 natural numbers, each of which is less than or equal 4, namely

$$\begin{aligned}
6 &= 4 + 2 + 0 \, , \\
6 &= 4 + 1 + 1 \, , \\
6 &= 3 + 3 + 0 \, , \\
6 &= 3 + 2 + 1 \, , \\
6 &= 2 + 2 + 2 \, .
\end{aligned}$$

There are five such partitions, so $A(6, 3, 4) = 5$. The number $A(n, p, m)$ is an extension of the Euler partition of the integer n into m parts [1]. The coefficients of $P(S)$ can be expressed in terms of $A(n, p, m)$; see Corollary 5.5.

Next, we develop two results about the properties of the permutations of a multiset. These results will enable us to further the study of inversion count distribution by the insertion method of the next chapter.

2.2 Supporting lemmas

We will first establish a well-known result for the cardinality for the permutation set $\sigma(S)$.

Lemma 2.1. Let $S = [n_1, n_2, \ldots, n_k]$ be a multiset with $n = \sum_{i=1}^{k} n_i$ elements. Then

$$|\sigma(S)| = \frac{n!}{n_1! n_2! \cdots n_k!}\,.$$

Proof. Map the multiset S to a set S^* so that if $e \in S$ is repeated r times, then the elements $\overbrace{e, e, \ldots, e}^{r} \in S$ are mapped to $e^{(1)}, e^{(2)}, \ldots, e^{(r)} \in S^*$. The number of such mappings is $n!$. The positions of $e \in S^*$ can be permuted in $r!$ ways to form the same permutation in $\sigma(S)$. By applying the multiplicative principle of counting, the proof is now complete. $\square$

Recall that $m(S)$ is the maximum inversion count of the sequences in the permutation set $\sigma(S)$. The next lemma establishes the value of $m(S)$ in terms of the multiplicities of the elements of S.

Lemma 2.2. Let $S = [n_1, n_2, \ldots, n_k]$ be a multiset with k distinct elements. Then

$$m(S) = \begin{cases} 0 & , \quad \text{if } k = 1 \\ \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j & , \quad \text{if } k > 1 \end{cases}$$

Proof 1: It is clear that if $k = 1$, then $m(S) = 0$.
For $k > 1$, the maximum inversion $m(S)$ can be obtained by arranging the elements of $s \in \sigma(S)$ in reverse order which corresponds to the element $s = s_1 s_2 \cdots s_n$, where $s_i \le s_j, 1 \le i \le j \le |S|$. Now, s consists of $\binom{|S|}{2}$ pairs and since each group of identical elements has zero inversion count, the values $\binom{n_i}{2}$ must be subtracted from the maximum possible inversion count. As $|S| = n_1 + n_2 + \cdots + n_k$, we have

$$\begin{aligned}
m(S) &= \binom{|S|}{2} - \sum_{i=1}^{k} \binom{n_i}{2} \\
&= \binom{n_1 + n_2 + \cdots + n_k}{2} - \sum_{i=1}^{k} \binom{n_i}{2} \\
2m(S) &= \sum_{i=1}^{k} n_i \times \left(\sum_{i=1}^{k} n_i - 1 \right) - \sum_{i=1}^{k} n_i(n_i - 1) \\
&= \left[\sum_{i=1}^{k} n_i \right]^2 - \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} n_i(n_i - 1) \\
&= \sum_{i=1}^{k} n_i^2 + 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j - \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} n_i^2 + \sum_{i=1}^{k} n_i \\
&= 2 \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j\,.
\end{aligned}$$

$\square$

Proof 2: For $k > 1$, the maximum inversion $m(S)$ can be obtained by arranging the elements of $s \in \sigma(S)$ in reverse order with $s = \underbrace{e_k \cdots e_k}_{n_k} \underbrace{e_{k-1} \cdots e_{k-1}}_{n_{k-1}} \cdots \underbrace{e_1 \cdots e_1}_{n_1}$.

For $2 \leq i \leq k$, the element e_i is followed by $n_{i-1} + \cdots + n_1$ elements of lower ranking. We have

$$\begin{aligned} Inv(s) &= n_k(n_{k-1} + \cdots + n_1) + n_{k-1}(n_{k-2} + \cdots + n_1) + \cdots + n_2(n_1) \\ &= n_1(n_2 + \cdots + n_k) + n_2(n_3 + \cdots + n_k) + \cdots + n_{k-1}n_k \\ &= \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} n_i n_j \,. \end{aligned}$$

$\square$

Chapter 3

Overview of the inversion distribution

3.1 Development history and overview

In 1750, G. Cramer [6] noticed, for a $n \times n$ matrix $A = (a_{i,j})$, the relationship between the sign of determinant $\det A$ and the parity of the inversion count:

$$\det(A) = \sum_{\pi \in \sigma(S_n)} (-1)^{\mathrm{Inv}(\pi)} \prod_{i=1}^{n} a_{i,\pi_i}\,, \quad \text{where } \pi = (\pi_1, \pi_2, \ldots, \pi_n)\,.$$

The first-known work on inversion distribution for the symmetric group was published by O. Rodrigues [14] in 1839, although it is generally attributed to Muir [13] in 1899.

$$\begin{aligned} P(\overbrace{1, 1, \ldots, 1}^{k}) &= \frac{\displaystyle\prod_{i=1}^{k}(x^i - 1)}{(x-1)^k} \\ &= \frac{G(k)}{\underbrace{G(1)\cdots G(1)}_{k}} \qquad \text{See notation in (2.1.8)} \end{aligned} \tag{3.1}$$

For a multiset $S = [n_1, n_2, \ldots, n_k]$, the objective of the thesis is to develop the generating function for the inversion distribution of the sequences created by the permutations of the elements of S. The generating function is given by

$$P(n_1, n_2, \ldots, n_k) = \frac{G(n_1 + n_2 + \cdots + n_k)}{G(n_1)G(n_2)\cdots G(n_k)}\,. \tag{3.2}$$

By setting $n_i = 1$ for $1 \le i \le k$, equation (3.2) reduces to equation (3.1). To add to the words of P.A. MacMahon in Section 2.1.8, (3.2) is truly remarkable in that $G(n)$ can be considered as an object and is described as “the q-analogue of $n!$” by R.P. Stanley [16]. Note that (3.2) can also assume the role of coefficients of a multinomial expansion. In 1913, P.A. MacMahon [11] published an article on the distribution of greater index, which is later named major index in his honor, for the multiset with three distinct elements ($k = 3$) is given by equation (3.2). He went on to prove that the distribution of major index for the permutations of a multiset is identical to the inversion distribution [10]. As the technique naturally extends to the general case, it was recognised by R.P. Stanley [16] as a complete solution. The proof utilises the recursive parent/child relationship in the partition structure which is formally proven in Theorem 7.2 for the general case.

In 1967, L. Carlitz [3] independently provided a combinatoric proof for the general case. The proof relies on the inversion distribution satisfying recursive relations of the permutation by algebraic expressions. It delves into the parent/child and peer to peer relationships between the permutations. R.P. Stanley [16, p.64] describes this type of proof as "semi-combinatorial" where the proof is a verification rather than a direct proof.

By using the Euler Pentagonal Theorem [1], D.E. Knuth [9] provided a beautiful combinatorial closed form expression for the inversion count distribution of the symmetric group. However, this form is of little computational value despite of its beauty.

Let $I_n(k)$ denote the number of elements with inversion count k in S_n:

$$I_n(k) = \sum_{j \geq 1} (-1)^j \left[\binom{n+k-u_j-1}{k-u_j} + \binom{n+k-u_j-j-1}{k-u_j-j} \right] \tag{3.3}$$

where $n \geq k \geq u_j + j$ and

$$u_j = \frac{3j^2 - j}{2} \,.$$

Knuth also outlined ideas on obtaining the closed form for the permutation of multisets by considering the mapping of inversions with the cycles of permutations. R.P. Stanley [16] provided two '"semi-combinatorial" proofs. The first proof is based on decomposition properties of the inversion distribution of a multiset. The second proof is a mapping of permutation cycles.

In summary, the distribution of inversion for multisets expressed as q-nomial form in (3.2) has been established by the combinations of the different methods listed below:

- By the link between major index and permutation of a multiset.
- By decomposition of permutation of multiset into components.
- By recursive relationships between the permutations of a multiset.
- By mapping of permutation cycles in a multiset.

3.2 Thesis overview

In my early University days, I came across three women sorting the 60,000 enrolment forms in a basketball court over two or three weeks. Their method was to segregate the forms into alphabet piles around the court, sort the piles separately and then consolidate the piles into a single pile. The initial curiosity inspired me to try to measure the efficiency of the method. As the sort process untangles pairs of out of order, it led to the development of a model for measuring the expected number of pairs out of order. The outcome, given in this thesis, is a study of inversion count distribution in order to define a mechanism for measuring how far a sequence deviates from the sorted state.

The thesis develops the hierarchical relationship between integer partitions and permutations of multisets. The many-to-many parent/child relationships between the permutations of multisets are expressed by insertion of elements.

For $n \in \mathbb{Z}$, the integer partitions of n can be formed by inserting an element into integer partitions of $n-1$. A partition $P = (n_1, n_2, \ldots, n_k)$ where $n_i \in \mathbb{Z}$, $1 \leq i \leq k$ and $n_1 + n_2 + \cdots + n_k = n$ is the child of partitions $P_i = (n'_1, n'_2, \ldots, n'_k)$, $1 \leq i \leq k$ where

$$n'_m = \begin{cases} n_m & , m \neq i \\ n_m - 1 & , m = i \,. \end{cases}$$

Therefore, a child partition with k distinct elements has k parent partitions. A partition with k distinct elements is the parent of $k+1$ partitions. This is illustrated in Figure 3.1 below.

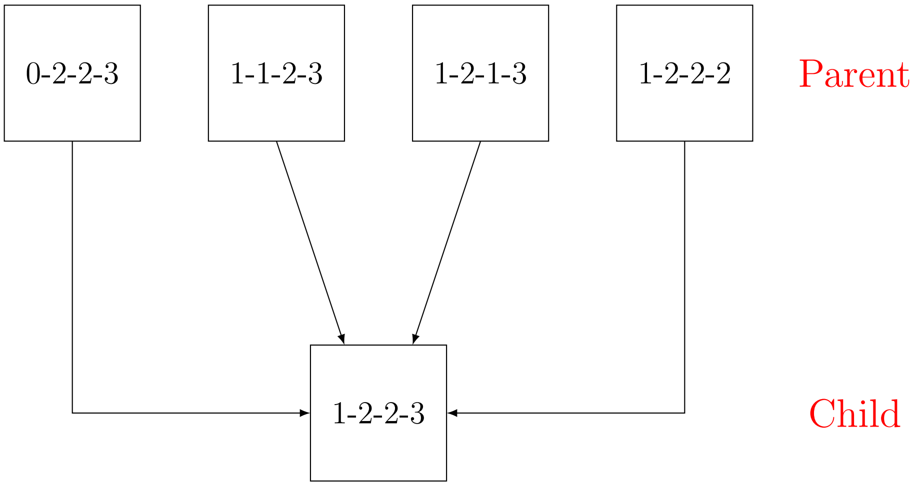


Figure 3.1: Hierarchy of partition

In Figure 3.1 above, let $S = \{a, b, b, c, c, d, d, d\}$. The parents for the permutations of S are the permutations of

$$\begin{aligned} S_1 &= \{b, b, c, c, d, d, d\}\,, \\ S_2 &= \{a, b, c, c, d, d, d\}\,, \\ S_3 &= \{a, b, b, c, d, d, d\}\,, \\ S_4 &= \{a, b, b, c, c, d, d\}\,. \end{aligned}$$

The thesis develops methods for calculating the inversion count for the permutations of a multiset when one or more copies of new element is inserted. It develops decomposition techniques for the permutation of multiset by insertion processes. The insertion process can also be linked to Ferrers diagrams which leads to a generating polynomial for integer partitions as a special case of the closed form of the distribution of inversion count in Theorem 8.6.

The two types of insertions explored are the insertion of a single element into first or last position of a sequence and also the insertion of multiple copies of a new element into a sequence represented by the upper diagonal of a hypercube.

For the permutations of a multiset, the inversion count frequency distribution is represented by a generating polynomial. The thesis derives the generating polynomial for two distinct elements in q-nomial form and uses it to form the building blocks for a closed form expression of the inversion count distribution.

The inversion count distribution of the integer partitioning provides a link to Ferrers diagrams. Lemma 8.10 established a generating polynomial for the integer partition function $p(n)$ in terms of the coefficients of the polynomial $P([n, n])$.

Chapter 4

The insertion process

For a multiset S, the elements s of the permutation set $\sigma(S)$ are referred to as sequences. This chapter demonstrates the process of building sequences by insertion of elements to the sequences in a parent/children hierarchy. The objective is to provide a methodology to calculate the inversion count distribution for the permutations of a multiset.

The *insertion position* k of an element into a sequence of length n is counted from left to right starting at zero, where $0 \le k \le n$.

$$\begin{array}{ccccccccc} \uparrow & \boxed{s_1} & \uparrow & \boxed{s_2} & \uparrow & \cdots & \uparrow & \boxed{s_n} & \uparrow \\ 0 & & 1 & & 2 & & n-1 & & n \end{array}$$

Insertion positions for a sequence

4.1 Insertion of single copy of a new element

The following example demonstrates the insertion process and its relationship to the inversion count distribution.

Example 4.1. Let $T = \{b, b, d\}$. Then sequences in $\sigma(T) = \{bbd, bdb, dbb\}$ have inversion counts 0,1,2, respectively. Therefore, $I_F(T) = (1, 1, 1)$. Let $S = \{b, b, c, d\}$. We will form $\sigma(S)$ and $I_F(S)$ by inserting the element c into positions 0, 1, 2, 3 of each element of $\sigma(T)$ as in Table 4.1 below. The notations for the table are:

- IP - Insertion Position.
- $I(s)$, $I(t)$ - Inversion count for the sequence $s \in \sigma(S)$, $t \in \sigma(T)$.

Table 4.1: Inversion count distribution by insertion

t	$I(t)$	IP	s	$I(s)$	IP	s	$I(s)$	IP	s	$I(s)$	IP	s	$I(s)$
bbd	0	0	cbbd	2	1	bcbd	1	2	bbcd	0	3	bbdc	1
bdb	1	0	cbdb	3	1	bcdb	2	2	bdcb	3	3	bdbc	2
dbb	2	0	cdbb	4	1	dcbb	5	2	dbcb	4	3	dbbc	3

By combining the columns $I(s)$ in Table 4.1, we have

$$I_F(T) = I_F([2, 1]) \quad = (1, 1, 1) \tag{4.1a}$$

$$I_F(S) = I_F([2, 1, 1]) = (1, 2, 3, 3, 2, 1)\,. \tag{4.1b}$$

4.2 Insertion of multiple copies of a new element

The next two examples demonstrate the insertion process for the construction of $I_F(S)$ as a sum of its parent partitions. The examples are simplified so that the inserting elements are either of the highest or lowest ranking. It will be seen in the Partition Family Theorem (Theorem 6.6) that $I_F(S)$ is invariant of the ranking of the inserting element.

Example 4.2. Let us calculate $I_F(S), I_F(S'), I_F(S'')$ for

$$S = \{b^2cd\}\,, \quad S' = \{ab^2cd\}\,, \quad \text{and} \quad S'' = \{b^2cde\}\,.$$

Note that S is the parent of S' and S''. From Equation (4.1b), $I_F(S) = (1, 2, 3, 3, 2, 1)$. By Lemma 2.2, the maximum inversion count for $\sigma(S')$ is 9 and therefore $I_F(S')$ is a 10-tuple. To calculate $I_F(S')$, note that S' is obtained from S by adding the element a. Since the letter a is of lower ranking than b, c and d, insertion into position k results in a permutation $s' \in \sigma(S')$ with insertion count k greater than that of s. This inserts in $I_F(S')$ k zeros to the leftmost coordinates, then inserts $4-k$ zeros to the rightmost coordinates, where $0 \le k \le 4$. Therefore,

$$\begin{array}{lll}
I_F(S') & = (1,2,3,3,2,1,0,0,0,0) & \text{Insertion into position 0} \\
& + (0,1,2,3,3,2,1,0,0,0) & \text{Insertion into position 1} \\
& + (0,0,1,2,3,3,2,1,0,0) & \text{Insertion into position 2} \\
& + (0,0,0,1,2,3,3,2,1,0) & \text{Insertion into position 3} \\
& + (0,0,0,0,1,2,3,3,2,1) & \text{Insertion into position 4} \\
& = (1,3,6,9,11,11,9,6,3,1)\,. &
\end{array}$$

$$\begin{array}{lll}
\text{Therefore,} \quad P(S') & = 1 + 3x + 6x^2 + 9x^3 + 11x^4 + 11x^5 + 9x^6 + 6x^7 + 3x^8 + x^9 \\
& = (1 + 2x + 3x^2 + 3x^3 + 2x^4 + x^5)(1 + x + x^2 + x^3 + x^4) \\
& = P(S)(1 + x + x^2 + x^3 + x^4)\,.
\end{array}$$

Note that $P(S')$ is formed by multiplying the $P(S)$ by the polynomial matching the insertion, namely $1 + x + x^2 + x^3 + x^4$.

Now, S'' is formed by inserting the element e into S. Since e is of higher ranking than b, c, d, inserting e into position k of $s \in \sigma(S)$ results in a permutation $s'' \in \sigma(S'')$ with inversion count $4-k$ greater than that of s. This inserts in $I_F(S'')$ $4-k$ zeros to the leftmost coordinates, then inserts k zeros to the rightmost coordinates. Therefore,

$$\begin{array}{lll}
I_F(S'') & = (0,0,0,0,1,2,3,3,2,1) & \text{Insertion into position 0} \\
& + (0,0,0,1,2,3,3,2,1,0) & \text{Insertion into position 1} \\
& + (0,0,1,2,3,3,2,1,0,0) & \text{Insertion into position 2} \\
& + (0,1,2,3,3,2,1,0,0,0) & \text{Insertion into position 3} \\
& + (1,2,3,3,2,1,0,0,0,0) & \text{Insertion into position 4} \\
& = (1,3,6,9,11,11,9,6,3,1)\,. &
\end{array}$$

$$\begin{array}{lll}
\text{Therefore,} \quad P(S'') & = 1 + 3x + 6x^2 + 9x^3 + 11x^4 + 11x^5 + 9x^6 + 6x^7 + 3x^8 + x^9 \\
& = (1 + 2x + 3x^2 + 3x^3 + 2x^4 + x^5)(1 + x + x^2 + x^3 + x^4) \\
& = P(S)(1 + x + x^2 + x^3 + x^4)\,.
\end{array}$$

Therefore,

$$\sigma(S'') = \sigma(S') \qquad \text{and} \qquad P(S') = P(S'')\,.$$

The next example demonstrates the insertion of multiple copies of an element and also provides a geometric interpretation.

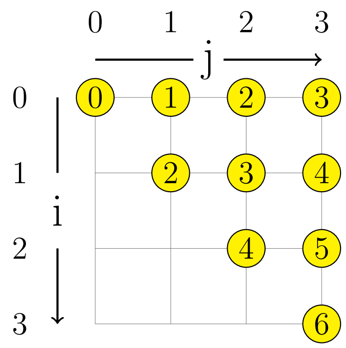
Example 4.3. We will calculate $P(S)$, $P(S')$, and $P(S'')$ for

$$S = \{bcd\}\,, \quad S' = \{a^2bcd\}\,, \quad \text{and} \quad S'' = \{a^3bcd\}\,.$$

The elements of $\sigma(S)$ are $\{bcd, bdc, cbd, cdb, dbc, dcb\}$ with inversion counts 0,1,1,2,2,3, respectively. Therefore, $I_F(S) = (1,2,2,1)$ and $P(S) = 1 + 2x + 2x^2 + x^3$.

We will use $T = \{abcd\}$ as an intermediate set to explain the insertion process. Each item $s' \in \sigma(S')$ is formed by inserting 2 copies of a into positions i and j of $s \in \sigma(S)$, where $0 \leq i \leq j \leq 3$. Insertion of the first copy of element a into position i of $s \in \sigma(S)$ forms $t \in \sigma(T)$ where $\mathrm{Inv}(t) = \mathrm{Inv}(s) + i$. Insertion of the second copy of a into position j in $s \in \sigma(S)$ forms $s' \in \sigma(S')$. Now, $\mathrm{Inv}(s') = \mathrm{Inv}(t) + j$ since the position of the first copy of a does not affect the increase in inversion count of the second copy of a. Therefore, $\mathrm{Inv}(s') = \mathrm{Inv}(t) + j = \mathrm{Inv}(s) + i + j$. Insertion of the two copies of a into position i, j shifts $I_F(S)$ to the right by $i + j$.

Figure 4.1: Insertion of 2 copies of a into elements of $\sigma(b, c, d)$



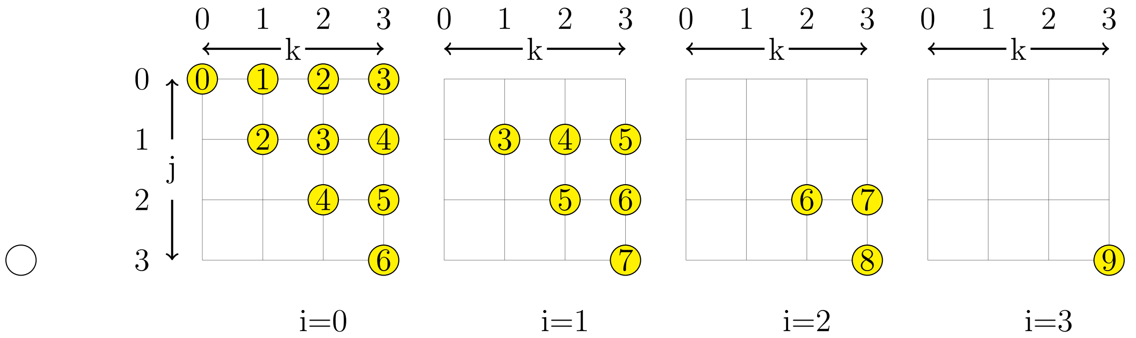
In Figure 4.1, the horizontal axis (reading downwards) is i and the vertical axis (reading across) is j. The circled value is $i + j$. Notice that the insertions correspond to the upper diagonal of the square. For instance, inserting a into position 1 and 3 of bdc (inversion count 1) gives $badca$ (inversion count 5) and $5 = 1 + (1 + 3)$. Since increase of inversion by $k \geq 0$ has the effect of multiplying by x^k, the insertion of two copies of a can be treated as multiplying $P(S)$ by the operator $1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6$.

$$\begin{aligned} P(S') &= (1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6)P(S) \\ &= (1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6)(1 + 2x + 2x^2 + x^3) \\ &= 1 + 3x + 6x^2 + 9x^3 + 11x^4 + 11x^5 + 9x^6 + 6x^7 + 3x^8 + x^9 \\ I_F(S') &= (1, 3, 6, 9, 11, 11, 9, 6, 3, 1)\,. \end{aligned} \tag{4.2}$$

Note that in the factor in the RHS of first line of (4.2), the coefficient of $x^i, 1 <= i \leq 6$ corresponds to the number of circles with value i in Figure 4.1.

Next, $s'' \in S''$ is formed by inserting 3 copies of a into positions i, j, k of $s \in S$, where $0 \leq i \leq j \leq k \leq 3$. Insertion into positions i, j, k shifts $(1, 2, 2, 1)$, the inversion count frequency of S, to the right by $i + j + k$.

Figure 4.2: Insertion of a, a, a into $\sigma(b, c, d)$



The triplets (i, j, k) form an upper diagonal of a 3 dimensional cube. Thus

$$
\begin{aligned}
P(S'') &= (1 + x + 2x^2 + 3x^3 + 3x^4 + 3x^5 + 3x^6 + 2x^7 + x^8 + x^9)P(S) \\
&= (1 + x + 2x^2 + 3x^3 + 3x^4 + 3x^5 + 3x^6 + 2x^7 + x^8 + x^9)(1 + 2x + 2x^2 + x^3) \\
&= 1 + 3x + 6x^2 + 10x^3 + 14x^4 + 17x^5 + 18x^6 + 17x^7 + 14x^8 \\
&\quad + 10x^9 + 6x^{10} + 3x^{11} + x^{12} \\
I_F(S'') &= (1, 3, 6, 10, 14, 17, 18, 17, 14, 10, 6, 3, 1)
\end{aligned}
$$

In the examples above, we have limited the insertion element to be either of lowest or highest ranking, relative to the elements in set S.

Chapter 5

Preliminary results

5.1 The generating polynomial $P(S)$

The coefficients of the generating polynomial $P(S)$ represent the inversion frequency distribution of the permutations $\sigma(S)$ of the multiset S. This polynomial provides the algebraic tool for the insertion process as demonstrated in Examples 4.2 and 4.3. In this chapter, we will further develop the properties of this generating polynomial. In particular, the exact form for $P(S)$ where S consists of two distinct elements ($R(S) = 2$) is established.

The reader would have noticed that the coordinates f_k are symmetrical under reflection: $(f_0, f_1, \ldots, f_{m-1}, f_m) = (f_m, f_{m-1}, \ldots, f_2, f_1)$. The following lemma provides a formal proof of this fact.

Lemma 5.1. Let S be a multiset with $I_F(S) = \{f_0, f_1, \ldots, f_{m(S)}\}$. Then $f_j = f_{m(S)-j}$ for $j = 0, 1, \ldots, m(S)$.

Proof. Reflect $s = s_1 s_2 \ldots s_{|S|} \in \sigma(S)$ about its median position to form $s' = s'_1 s'_2 \ldots s'_{|S|}$ so that $s_i = s'_{|S|-i+1}$, $1 \leq i \leq |S|$.

For a pair $(i, j), 1 \leq i < j \leq |S|$, there are three cases to consider:

1. $s_i = s_j$. The pair does not contribute to the inversion count.
2. $s_i > s_j$. The pair is included in $\mathrm{Inv}(s)$.
3. $s_i < s_j$. The pair is included in $\mathrm{Inv}(s')$ since $s'_{|S|-j+1} > s'_{|S|-i+1}$.

Therefore, $\mathrm{Inv}(s') + \mathrm{Inv}(s) = m(S)$, by reflection. For every sequence $s \in \sigma(S)$ where $\mathrm{Inv}(s) = m$, there exists a unique sequence in $s' \in \sigma(S)$ where $\mathrm{Inv}(s') = m(S) - m$. Thus for $0 \leq k \leq m(S)$,

$$f_k = |s \in \sigma(S) : \mathrm{Inv}(s) = k| = |s' \in \sigma(S) : \mathrm{Inv}(s') = m(S) - k| = f_{m(S)-k}\,.$$

□

The following corollary is used for analysing the inversion count mean and median of datasets in Chapter 9.

Corollary 5.2. For a multiset S, the mean $\overline{X}$ of the inversion frequency distribution $I_F(S) = \{f_0, f_1, \ldots, f_{m(S)}\}$, is equal to the median value M; indeed,

$$\overline{X} = M = \frac{m(S)}{2}\,.$$

Proof. Let $T = \sum_{i=0}^{m(S)} f_i$; then $\overline{X} = \frac{1}{T} \sum_{i=0}^{m(S)} i f_i$.

There are two cases to consider:

Case 1: $m(S)$ is even, and so $m(S) = 2M$, where M is the median.

$$\begin{aligned}
\overline{X} &= \frac{1}{T}\left(\sum_{i=0}^{M-1} i f_i + M f_M + \sum_{j=M+1}^{2M} j f_j\right) \\
&= \frac{1}{T}\left(\sum_{i=0}^{M-1} i f_i + M f_M + \sum_{i=0}^{M-1} (2M-i) f_i\right) && \text{(by Lemma 5.1)} \\
&= \frac{1}{T}\left(2M \sum_{i=0}^{M-1} f_i + M f_M\right) \\
&= \frac{M}{T}\left(\sum_{i=0}^{M-1} f_i + f_M + \sum_{i=0}^{M} f_i\right) \\
&= \frac{M}{T}\left(\sum_{i=0}^{M-1} f_i + f_M + \sum_{i=M+1}^{2M} f_i\right) && \text{(by Lemma 5.1)} \\
&= \frac{M}{T} \sum_{i=0}^{2M} f_i \\
&= M\,.
\end{aligned}$$

Case 2: $m(S)$ is odd, and so $m(S) = 2M' + 1$ where M' is the median.

$$\begin{aligned}
\overline{X} &= \frac{1}{T}\left(\sum_{i=0}^{M'} i f_i + \sum_{j=M'+1}^{2M'+1} j f_j\right) \\
&= \frac{1}{T}\left(\sum_{i=0}^{M'} i f_i + \sum_{i=0}^{M'} (2M'+1-i) f_i\right) \\
&= \frac{2M'+1}{T} \sum_{i=0}^{M'} f_i \\
&= \frac{2M'+1}{2T}\left(\sum_{i=0}^{M'} f_i + \sum_{i=0}^{M'} f_i\right) \\
&= \frac{2M'+1}{2T}\left(\sum_{i=0}^{M'} f_i + \sum_{i=M'+1}^{2M'+1} f_i\right) && \text{(by Lemma 5.1)} \\
&= \frac{2M'+1}{2} = \frac{m(s)}{2}
\end{aligned}$$

□

Lemma 5.3 formalises the calculations of the insertion process described in Examples 4.2 and 4.3.

Lemma 5.3. Let S be a multiset with k distinct elements and let $S' = S \cup e^m$ where $e < \min(S)$ or $e > \max(S)$. Then

$$P(S') = P(S) \sum_{i=0}^{m|S|} A(i, |S|, m) x^i$$

where $A(i, |S|, m)$ is the number of partitions of the integer i into m parts each of size at most $|S|$.

Proof. Suppose that $e < \min(S)$ and insert m identical copies of e into S. Let c_i denote the number of copies of e inserted into position i, where $0 \le i \le |S|$ and $0 \le c_i \le m$. The insertion positions can be represented as a $(|S|+1)$ -tuple $\boldsymbol{c} = (c_0, c_1, \ldots, c_{|S|})$.

Since $e \notin S$, the elements in S can be regarded as being identical in ranking for the insertion process . In the formation of $\sigma(S')$, each $(|S|+1)$-tuple $(c_0, c_1, \ldots, c_{|S|})$ increases the inversion count of $s \in \sigma(S)$ by

$$K = \sum_{j=0}^{|S|} j c_j \quad \text{where} \quad 0 \le K \le m|S| \,. \tag{5.1}$$

The maximum value of $K = m|S|$ is obtained by inserting all the m copies of e into position $|S|$. The application of $(c_0, c_1, \ldots, c_{|S|})$ to $\sigma(S)$ inserts K zeros to the left of $I_F(S)$ and appends $M(S) - K$ zeros to the right of $I_F(S)$ to form $I_F(S')$.

For a fixed value of K, the count of tuples $(c_0, c_1, \ldots, c_{|S|})$ satisfying Equation (5.1) is given by $A(K, |S|, m)$. Now group the tuples by their value of K, the group increases the inversion count for each $s \in S$ by K. This represents multiplying the coefficient of each term in $P(S)$ by x^K. The lemma now follows by the definition of coefficients of $P(S')$.

By similar argument, the lemma is also true if $e > \max(S)$. ☐

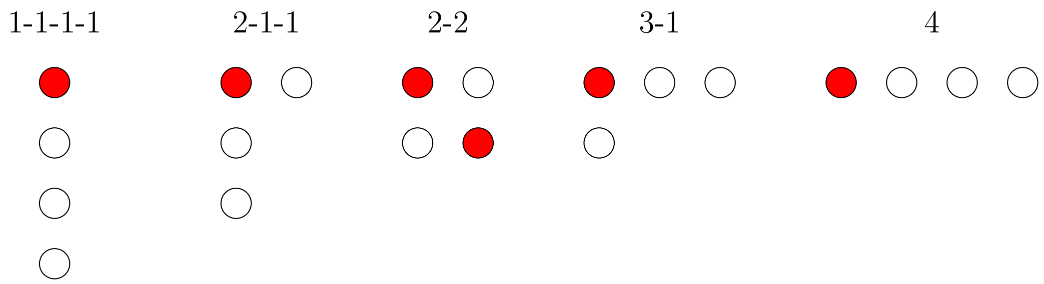
5.2 Ferrers diagram

A Ferrers diagram [2] is a representation of an integer partition n

$$n = n_1 + n_2 + \cdots + n_k \,, \quad n_1 \ge n_2 \ge \cdots \ge n_k \ge 0 \,, \quad n_1, n_2, \ldots, n_k \in \mathbb{Z} \,.$$

Figure 5.1 below shows the partitions of the integer 4. The circles in the south-east diagonal are marked as red. The *conjugate* of the Ferrers diagram is obtained by reflecting along this diagonal. The conjugate pairs are (1,1,1,1) and (4), (2,1,1) and (1,3), and (2,2) and (2,2). The partition (2,2) maps to itself is termed as self conjugate. By considering the reflection image along the diagonal, it is clear that each Ferrers diagram has a unique conjugate.

Figure 5.1: Ferrers diagram for $n = 4$



The insertion of n copies of an element (see Example 4.3) into the permutations of a multiset can be represented by the partitions of a Ferrers diagram. In this section, we will examine this link which leads to a generating polynomial for integer partition in Lemma 5.4. Each partition counted by $A(n, p, m)$ can be represented as a Ferrers diagram for n with the restriction that the number of summands is no more than p and the maximum value of each summand is m. To illustrate, Table 5.1 below demonstrates the relationship between $A(10, 7, 5)$ and $A(10, 5, 7)$ using the correspondence between the conjugate pairs in the Ferrers diagram.

Table 5.1: A(10,7,5) and A(10,5,7)

A(10,5,7)	A(10,7,5)	A(10,5,7)	A(10,7,5)	A(10,5,7)	A(10,7,5)	A(10,5,7)	A(10,7,5)
7-3	2-2-2-1-1-1-1	7-2-1	3-2-1-1-1-1-1	7-1-1-1	4-1-1-1-1-1-1	6-4	2-2-2-2-1-1
6-3-1	3-2-2-1-1-1	6-2-2	3-3-1-1-1-1	6-2-1-1	4-2-1-1-1-1	6-1-1-1-1	5-1-1-1-1-1
5-5	2-2-2-2-2	5-4-1	3-2-2-2-1	5-3-2	3-3-2-1-1	5-3-1-1	4-2-2-1-1
5-2-2-1	4-3-1-1-1	5-2-1-1-1	5-2-1-1-1	4-4-2	3-3-2-2	4-4-1-1	4-2-2-2
4-3-3	3-3-3-1	4-3-2-1	4-3-2-1	4-3-1-1-1	5-2-2-1	4-2-2-2	4-4-1-1
4-2-2-1-1	5-3-1-1	3-3-3-1	4-3-3	3-3-2-2	4-4-2	3-3-2-1-1	5-3-2
3-2-2-2-1	5-4-1	2-2-2-2-2	5-5				

Lemma 5.4.

$$A(n, p, m) = A(n, m, p)$$

Proof. $A(n, p, m)$ is the number of integer partitions of n into p blocks of size at most m. Each such integer partitions can be represented as a Ferrers diagram that represents an integer partition of n into m blocks of size at most p. This is a bijection. □

Corollary 5.5 below is the special case of Lemma 5.3 in which the multiset S consists of two distinct elements.

Corollary 5.5. Let $S' = [n_1, n_2]$. Then

$$P(S') = \sum_{i=0}^{n_1 n_2} A(i, n_1, n_2)x^i = \sum_{i=0}^{n_1 n_2} A(i, n_2, n_1)x^i \,. \tag{5.2}$$

Proof. Let $S = [n_1]$ and $S' = [n_1, n_2]$. Since $P(S') = 1$, apply Lemmas 5.3 and 5.4, we have

$$\begin{aligned} P(S') &= P(S) \sum_{i=0}^{n_1 n_2} A(i, n_1, n_2) x^i \\ &= \sum_{i=0}^{n_1 n_2} A(i, n_1, n_2) x^i \\ &= \sum_{i=0}^{n_1 n_2} A(i, n_2, n_1) x^i \,. \end{aligned}$$

$\square$

Chapter 6

Partition family theorem

This chapter presents the Partition Family Theorem (Theorem 6.6) below which states that multisets belonging to the same partition family have the same inversion count frequency distribution.

6.1 Partition family with two distinct elements

The following lemma forms the base case for a proof by induction of the Partition Family Theorem. It also lays the groundwork for calculating the generating polynomial $P(S)$ by recursion as well as deriving its closed form.

Lemma 6.1. Let $S = [n_1, n_2]$ and $S' = [n_2, n_1]$, where n_1, n_2 are positive integers. Then $I_F(S) = I_F(S')$ and $P(S) = P(S')$.

Proof 1: Write $I_F(S) = (f_0, f_1, \ldots, f_M)$ and $I_F(S') = (f'_0, f'_1, \ldots, f'_{M'})$.

By Lemma 2.2, we have $M = M'$. Denote $M = m(S) = m(S')$. Then, Corollary 5.5 implies that, for $0 \leq i \leq M$,

$$f_i = A(i, n_1, n_2) = A(i, n_2, n_1) = f'_i \,.$$

It follows that $I_F(S) = I_F(S')$ and $P(S) = P(S')$. □

Proof 2: This is an elementary proof based on the inversion count of the elements in a sequence. Let $S = \{a^{n_1} b^{n_2}\}$ and $S' = \{a^{n_2} b^{n_1}\}$ and set $n = n_1 + n_2$. Then for an element $s \in \sigma(S)$, we form the unique element $s' \in \sigma(S')$ by the two following operations (A) and (B).

(A) Reflect s to form the permutation $s*$.

(B) Replace the elements a by b and b by a in $s*$ to form $s' \in \sigma(S')$.

Let x be in position k_1 and y be in position k_2 in s. Consider the two cases:

Case 1: If $x = y$, then the pair aa is transformed into bb and vice versa, and, therefore, the contribution of the inversion count of this pair to the inversion count does not change.

Case 2: If $x \neq y$, then consider the three subcases (a), (b), (c) below, in which $\overset{M}{|}$ is the median position in each diagram, and (A) and (B) corresponds the reflection and swap operations as described above. Note that if n is even, then the median position is not occupied by an element.

(a) x or y is at the median position: Suppose that x is at the median; then
$(..\overset{M}{x}..y..) \overset{(A)}{\longmapsto} (..y..\overset{M}{x}..) \overset{(B)}{\longmapsto} (..x..\overset{M}{y}..)$.

(b) k_1, k_2 are on the same side of the median position: Then
$(..x..y..\overset{M}{|}..) \overset{(A)}{\longmapsto} (..\overset{M}{|}..y..x..) \overset{(B)}{\longmapsto} (..\overset{M}{|}..x..y..)$

(c) k_1, k_2 are on the opposite sides of the median position: Then
$(..x..\overset{M}{|}..y..) \overset{(A)}{\longmapsto} (..y..\overset{M}{|}..x..) \overset{(B)}{\longmapsto} ..x..\overset{M}{|}..y..)$

In all three cases, the contribution of the pair x, y to the inversion count does not change. Since the mapping $S \longmapsto S'$ is a bijection and since it preserves the inversion count, it follows that $I_F(S) = I_F(S')$ and $P(S) = P(S')$. $\square$

The next corollary is an important result for the iterative process in the calculation of $I_F(S)$. A generalised result for $P(m, n)$ is given later by Lemma 8.5.

Corollary 6.2. For any positive integer n,

$$P(n, 1) = P(1, n) = 1 + x + \cdots + x^n = \frac{x^{n+1} - 1}{x - 1}. \quad x \neq 1$$

Proof. Let $S = \{a^n\}$ and $S' = \{a^n b\}$. Note that $\sigma(S)$ has only one element $s = \overbrace{aa \ldots a}^{n}$ and $P(S) = 1$. Then each element $s' \in \sigma(S')$ is formed by inserting a copy of b into position i of s, where $0 \leq i \leq n$:

$$s = \overbrace{a \ldots a}^{i} b \overbrace{a \ldots a}^{n-i}.$$

By inserting b into position i, each element to the right of b increases the inversion count of s by 1. Since $\sigma(S')$ consists of $n + 1$ elements: one with element b in position i for each $0 \leq i \leq n$. Therefore,

$$P(n, 1) = 1 + x + \cdots + x^n.$$

Hence by Lemma 6.1, $P(1, n) = 1 + x + \cdots + x^n$. $\square$

Corollary 6.3 below proves that the insertion of a lower or higher ranking element of any multiplicity to two multisets with the same inversion count frequency distribution yield multisets with the same inversion count frequency distribution. It will be used in the proof of the Partition Family Theorem to follow.

Corollary 6.3. Let $S = [n_1, n_2, \ldots, n_k]$ and $S' = [n'_1, n'_2, \ldots, n'_k]$ where $\{n_1, n_2, \ldots, n_k\}$ and $\{n'_1, n'_2, \ldots, n'_k\}$ are permutations of each other , with $I_F(S) = I_F(S')$. Let e_L, e_H be elements such that $e_L < \min(s), \min(s')$ and $e_H > \max(s), \max(s')$, for all $s \in S$, $s' \in S'$ and $m \in \mathbb{Z}^+$. Then

$$\begin{array}{lllllll} I_F(S \cup e_L^m) & = & I_F(S' \cup e_L^m) & = & I_F(S \cup e_H^m) & = & I_F(S' \cup e_H^m) \\ P(S \cup e_L^m) & = & P(S' \cup e_L^m) & = & P(S \cup e_H^m) & = & P(S' \cup e_H^m). \end{array}$$

Proof. Let $n = \sum_i n_i = \sum_i n'_i$. By definition, $I_F(S) = I_F(S')$ if and only $P(S) = P(S')$. Combining this with Lemma 5.3 gives

$$\begin{aligned} & P(S \cup e_L^m) = P(S) \sum_{i=0}^{nm} A(i,m,n)x^i = P(S') \sum_{i=0}^{nm} A(i,m,n)x^i = P(S' \cup e_L^m) \\ \text{and} \quad & P(S \cup e_H^m) = P(S) \sum_{i=0}^{nm} A(i,m,n)x^i = P(S') \sum_{i=0}^{nm} A(i,m,n)x^i = P(S' \cup e_H^m) \end{aligned} \tag{6.1}$$

By Corollary 5.5,

$$P(S \cup e_L^m) = P(S \cup e_H^m) \tag{6.2}$$
$$P(S' \cup e_L^m) = P(S' \cup e_H^m)\,. \tag{6.3}$$

By Equation (6.1) and (6.2), we have

$$P(S \cup e_L^m) = P(S' \cup e_L^m) = P(S \cup e_H^m) = P(S' \cup e_H^m)\,.$$

□

6.2 Two sort processes α and β

For a sequence s of length n where the elements may be repeated, the α-sort and β- sort processes are defined as follows:

- The α-sort arranges the first $n-1$ elements in ascending order, while position n in the sequence does not move.
- The β-sort arranges the last $n-1$ elements in ascending order, while position 1 in the sequence does not move.

These sort processes will be used for the proof in Theorem 6.6 and can be combined together to sort a sequence of length n as shown below:

Example 6.4.

(A) (9,1,8,3,1,8,1,6) $\xrightarrow{\alpha}$ (1,1,1,3,8,8,9,6)
$\xrightarrow{\beta}$ (1,1,1,3,6,8,8,9)

(B) (9,1,8,3,1,8,1,0) $\xrightarrow{\alpha}$ (1,1,1,3,8,8,9,0)
$\xrightarrow{\beta}$ (1,0,1,1,3,8,8,9)
$\xrightarrow{\alpha}$ (0,1,1,1,3,8,8,9)

In the case of (B) where the lowest ranking element is in the last position, an extra α-sort operation is required to complete the sort.

The following lemma is self-evident and is stated without proof.

Lemma 6.5. Let $s = s_1 s_2 \ldots s_n$, $n \geq 3$ be a sequence. Then the following operations will sort the elements of s into non-descending order.

1. If $s_n > s_i$ for some $1 \leq i \leq n$, , then the sort operations $\alpha\beta$ arrange the elements of s into non-descending order.
2. If $s_n \leq s_i$, for all $1 \leq i \leq n$, , then the sort operations $\alpha\beta\alpha$ arrange the elements of s into non-descending order.

6.3 Partition Family Theorem

The theorem below asserts that the multisets belonging to the same partition family have the same inversion count frequency distribution.

Theorem 6.6. Let $N = [n_1, n_2, \ldots, n_k]$ and $M = [m_1, m_2, \ldots, m_k]$, where $(n_1, n_2, \ldots, n_k)$ and $(m_1, m_2, \ldots, m_k)$ are permutations of each other. Then

$$I_F(N) = I_F(M) \quad and \quad P(N) = P(M)\,.$$

Proof. The proof is by induction on the number of distinct elements k. Note that it is valid to assume that the multisets M and N span over the same set of elements $\{e_1, e_2, \ldots, e_k\}$ where $e_1 \leq e_2 \leq \cdots \leq e_k$. The case in which $k = 2$, that is, when $N = [n_1, n_2]$ and $M = [n_2, n_1]$, is given by Lemma 6.1.

Assume that the theorem holds for all multisets with $k - 1$ distinct elements with $k \geq 3$ and consider the multisets M and N as in the theorem. Define $Z = [z_1, z_2, \ldots, z_k]$ to be the permutation of the multiplicities of M and N in non-descending order. We will now prove that $I_F(M) = I_F(Z) = I_F(N)$.

We will first prove that the inversion count frequency is invariant under application of the α-sort and of the β-sort. That is, if $Y = [y_1, y_2, \ldots, y_{k-1}, y_k]$ is formed by applying the α-sort and β-sort to $M = [m_1, m_2, \ldots, m_{k-1}, m_k]$, then $I_F(M) = I_F(Y)$.

Denote $M' = [m_1, m_2, \ldots, m_{k-1}]$. Arrange M' in nondecreasing order to form $Y' = [y_1, y_2, \ldots, y_{k-1}]$. By the induction hypothesis, $I_F(M') = I_F(Y')$. It follows from Corollary 6.3 that:

$$I_F(M) = I_F\big(M' \cup e_k^{m_k}\big) = I_F\big(Y' \cup e_k^{y_k}\big) = I_F(Y)\,. \tag{6.4}$$

Denote $M'' = [m_2, m_3, \ldots, m_k]$ Arrange M'' to form Y'' in nondecreasing order to form $Y'' = [y_2, y_3, \ldots, y_k]$. By the induction hypothesis, $I_F(M'') = I_F(Y'')$. Then it follows from Corollary 6.3 that:

$$I_F(M) = I_F\big(e_1^{m_1} \cup M''\big) = I_F\big(e_1^{y_1} \cup Y''\big) = I_F(Y)\,. \tag{6.5}$$

By Lemma 6.5, the application of $\alpha\beta\alpha$-sort to the multiset M results in the sorted multiset Z. By Equations (6.4) and (6.5), we have $I_F(M) = I_F(Z)$. By the same arguments above for N, we also have $I_F(N) = I_F(Z)$. Therefore, $I_F(N) = I_F(M)$, and induction concludes the proof. $\square$

For the purpose of calculating inversion distribution, a very important corollary of Theorem 6.6 is that when inserting an element e of multiplicity n into a multiset S to form S', it is valid to assume that it is either of higher ranking or lower ranking than all the elements in S. In practice, it is easier to form a new sequence by inserting the lowest order element using Lemma 5.3.

Corollary 6.7. Let S be a multiset and write $S' = S \cup e_\alpha^n$ and $S'' = S \cup e_\beta^n$ where $e_\alpha, e_\beta \notin S$. Then

$$I_F(S') = I_F(S'') \quad \text{and} \quad P(S') = P(S'')\,.$$

Proof. This result has been established in Corollary 6.3 where e_α, e_β are both either of higher ranking or of lower ranking than the elements of S. Since the multisets S' and S'' belong to the same partition family, then by Theorem 6.6, $I_F(S') = I_F(S'')$. By definition, it follows that $P(S') = P(S'')$. □

Chapter 7

Parent-child relationship between partition families

Lemma 5.3 demonstrates a process of constructing $\sigma(S)$ incrementally by inserting many copies of a new element of either the lowest or highest ranking. This section introduces a method of insertion into the leading position of a sequence which encapsulates the parent/child relationship between partitions of length n and those of $n+1$.

Example 7.1. We show how to derive the equality

$$P(1,1,3,2) = P(1,3,2) + xP(1,3,2) + x^2P(1,1,2,2) + x^5P(1,1,3,1)\,.$$

Let $S = \{a,b,c,c,c,d,d\}$. Then the parents of $\sigma(S)$ are

$$\begin{aligned}
\sigma(S_a) &= \sigma(b,c,c,c,d,d)\\
\sigma(S_b) &= \sigma(a,c,c,c,d,d)\\
\sigma(S_c) &= \sigma(a,b,c,c,d,d)\\
\sigma(S_d) &= \sigma(a,b,c,c,c,d)\,.
\end{aligned}$$

The mapping $\mathcal{F} : \underset{\alpha\in S}{\cup}\sigma(S_\alpha) \to \sigma(S)$ which appends α to the beginning of the sequences in $\sigma(S_\alpha)$ is surjective since if $s \in \sigma(S)$, the element in the leading position in s is a,b,c,d. It is also an injective mapping since each parent element in $s_\alpha \in \sigma(S_\alpha)$ is mapped to a unique element in $\sigma(S)$. Therefore, $\mathcal{F}$ is a bijective mapping.

Let $K(\alpha)$ denote the number of elements in S of lower ranking than α. The insertion of α into position 0 of a sequence in $\sigma(S_\alpha)$ forms $s \in \sigma(S)$ where $\mathrm{Inv}(s) = \mathrm{Inv}(s_\alpha) + K(\alpha)$. Since $K(a) = 0$, $K(b) = 1$, $K(c) = 2$, and $K(d) = 5$, the equality now follows from the definition of $P(S)$.

In fact, we have the following general result:

Theorem 7.2. Recursive expression for generating polynomial $P(S)$
Let $S = [n_1, n_2, \ldots, n_k]$ and, for each $0 \le m \le k$, define $S_m = [n'_{m,1}, \ldots, n'_{m,k}]$ where

$$n'_{m,i} = \begin{cases} n_i & , \quad i \neq m \\ n_i - 1 & , \quad i = m\,. \end{cases} \tag{7.1}$$

Then $P(S) = \sum_{m=0}^{k} x^{c_m} P(S_m)$ where

$$c_m = \begin{cases} 0 & , \quad m = 0 \\ \sum_{j=1}^{m-1} n'_{m,j} & , \quad m > 0\,. \end{cases} \tag{7.2}$$

Proof. Let $\{e_1, \ldots, e_k\}$ be the distinct elements of S and assume that $e_i < e_j$ when $i < j$. Let $s = s_1 s_2 \ldots s_{|S|} \in \sigma(S)$, and suppose that $s_1 = e_m$, where $1 \leq m \leq k$. Then $s = e_m s'$ where $s' = s_2 \ldots s_{|S|} \in \sigma(S_m)$, and S_m is defined by (7.1).

Define $\mathcal{F}_m : \sigma(S_m) \to \sigma(S)$ to be the mapping which appends e_m to position 1 of the sequences in $\sigma(S_m)$ and let

$$\mathcal{F} = \bigcup_{1 \leq m \leq k} \mathcal{F}_m\,.$$

This mapping is surjective since the position 1 of each element $s \in \sigma(S)$ is equal to e_m for some m where $1 \leq m \leq k$ and the remaining positions satisfies $s' = s_2 \ldots s_{|S|} \in \sigma(S_m)$. If $\mathcal{F}(u) = \mathcal{F}(v)$, where $u = u_1 u_2 \ldots u_{|S|}$ and $v = v_1 v_2 \ldots v_{|S|}$, then $u_1 = v_1$, and $u_2 \ldots u_{|S|} = v_2 \ldots v_{|S|}$. Therefore, $u = v$ and $\mathcal{F}$ is injective. It follows that $\mathcal{F}$ is a bijective mapping.

Let $s = e_m s'_m$ where $s'_m \in S_m$,

$$\begin{aligned} Inv(s) &= Inv(s'_m) + \text{Number of elements in } s'_m \text{ of higher ranking than } e_m \\ &= Inv(s'_m) + \sum_{j=1}^{m-1} n'_{m,j} \\ &= Inv(s'_m) + c_m\,. \qquad \text{[By (7.2)]} \end{aligned}$$

By the definition of $I_F(S)$ in Section 2.1.6, the above expression can be written as

$$I_F(S) = \sum_{m=0}^{k} I_F(S_m) + c_m\,.$$

It follows from the definition of $P(S)$ that

$$P(S) = \sum_{m=0}^{k} x^{c_m} P(S_m)\,.$$

$\square$

Using Theorem 7.2, the inversion count frequency distributions are calculated iteratively for the digits $\underbrace{0 \cdots 0}_{n}, \ldots, \underbrace{9 \cdots 9}_{n}$ where $n = 5, 6, 7, 8, 9, 10$; see Tables 1.1, A1–A7.

Chapter 8

A closed form expression for the generating function $P(S)$

8.1 A closed form expression for the generating function $P(S)$

This chapter delivers a closed form expression for the generating function $P(S)$ where $S = [n_1, n_2, \ldots, n_k]$. To this end, Theorem 6.6 may be applied and allows us to assume that $n_i \leq n_j$ and $e_i < e_j$ for $1 \leq i < j \leq k$. Recall that the ranking $R(S)$ of S is the number of distinct elements in S. The Decomposition Lemma (Lemma 8.2) decomposes a generating polynomial for a multiset of ranking k as a product of $k-1$ generating polynomials composed of two elements in the form $P(n, m)$. The formula for generating polynomial where $R(S) = 2$ (Lemma 8.5) provides a closed form for $P(n, m)$. The two lemmas combine together to provide a closed form expression for $P(S)$; see Theorem 8.6.

Example 8.1. Show that $P(1, 2, 3) = P(1, 5)P(2, 3)$.
Let $S = \{a, b, b, c, c, c\}$, $T = \{a, x, x, x, x, x\}$, $U = \{b, b, c, c, c\}$.

A sequence $s \in \sigma(S)$ is obtained from a unique element $u \in \sigma(U)$ by an insertion position for a into u. Since b and c are of higher ranking than a, these elements can be considered as having an identical higher ranking for the purpose of the insertion. u is constructed by the insertion of 2 copies of b into 3 copies of c. Therefore,

$$\begin{aligned} Inv(s) &= Inv(u) + \text{insertion position of } a \text{ into } u \\ &= Inv(u) + \text{insertion position of } a \text{ into 5 } x\text{'s.} \end{aligned} \tag{8.1}$$

By using (8.1) to sum over all the sequences $s \in \sigma(S)$ and by the definition of P(S), it follows that $P(S) = P(T)P(U)$.

The example can also be verified algebraically.
By corollary 6.2, $P(T) = 1 + x + x^2 + x^3 + x^4 + x^5$.
By Table A1, $P(U) = 1 + x + 2x^2 + 2x^3 + 2x^4 + x^5 + x^6$ and
$P(S) = 1+2x+4x^2+6x^3+8x^4+9x^5+9x^6+8x^7+6x^8+4x^9+2x^{10}+x^{11} = P(T)P(U)$.

Lemma 8.2. (**Decomposition Lemma**)
Let $n_1, n_2, \ldots, n_k \in \mathbb{Z}^+$, where $k \geq 3$. Then

$$P(n_1, n_2, \ldots, n_k) = \prod_{i=1}^{k-1} P(n_i, n_{i+1} + \cdots + n_k)\,. \tag{8.2}$$

Proof. By Theorem 6.6, we can assume elements satisfy $e_i < e_j, 1 \leq i \leq< k$ The coefficients of $P(n_1, n_2)$ form the inversion count frequency distribution for two distinct elements e_1, e_2 with multiplicity of n_1, n_2 respectively.

Another way of looking at $P(n_1, n_2)$ is to consider the inversion count frequency distribution resulting from inserting n_1 copies of the element e_1 into n_2 copies of the element e_2. The result is a permutation in $\sigma[n_1, n_2]$. Each such sequence is of length $n_1 + n_2$.

Now, $P(n_1, n_2, n_3)$ arises from the insertion of n_1 copies of the element e_1 into elements of $\sigma[n_2, n_3]$. Since e_1 is of lower ranking than e_2, e_3, these elements can be thought as having identical higher ranking for insertion purposes.

$$P(n_1, n_2, n_3) = P(n_1, n_2 + n_3)P(n_2, n_3)\,. \tag{8.3}$$

Assume that (8.2) holds for $m = k$. Any permutation of the multiset $[n_1, n_2, \ldots, n_{k+1}]$ is formed by inserting n_1 copies of e_1 into some permutation $s \in \sigma(n_2 + n_3 + \cdots + n_{k+1})$. Therefore,

$$\begin{aligned} P(n_1, n_2, \ldots, n_{k+1}) &= P(n_1, n_2 + \cdots + n_{k+1})\; P(n_2, \ldots, n_{k+1}) \\ &= P(n_1, n_2 + \cdots + n_{k+1}) \prod_{i=2}^{k} P(n_i, n_{i+1} + \cdots + n_{k+1})\,. \end{aligned}$$

By induction,

$$P(n_1, n_2, \ldots, n_{k+1}) = \prod_{i=1}^{k} P(n_i, n_{i+1} + \cdots + n_{k+1})\,.$$

□

The following example demonstrates an application of Lemma 8.2.

Example 8.3.

$$\begin{aligned} P(1, 2, 2, 3, 4) &= P(1, 2 + 2 + 3 + 4)\; P(2, 2 + 3 + 4)\; P(2, 3 + 4)\; P(3, 4) \\ &= P(1, 11)\; P(2, 9)\; P(2, 7)\; P(3, 4) \\ &= P(11, 1)\; P(9, 2)\; P(7, 2)\; P(4, 3)\,. \end{aligned}$$

The next example demonstrates a technique for calculating $P(n, m)$.

Example 8.4. Show that

$$P(2, 2) = \frac{(x^4 - 1)(x^3 - 1)}{(x^2 - 1)(x - 1)}\,.$$

Any sequence $s \in \sigma[2, 3]$ can be considered to be the permutation of 2 copies of e_1 and 3 copies of e_2.

1. If $s = e_1 s'$, then $\mathrm{Inv}(s) = \mathrm{Inv}(s')$, where $s' \in \sigma[1, 3]$.
2. If $s = e_2 s'$, then $\mathrm{Inv}(s) = \mathrm{Inv}(s') + 2$, where $s' \in \sigma[2, 2]$.

By partitioning according to the element in the first position of $s \in \sigma[2, 3]$, we have

$$P(2, 3) = P(1, 3) + x^2 P(2, 2)\,. \tag{8.4}$$

Similarly, by considering $s \in \sigma[3,2]$, s is the permutation of the 3 copies of e_1 and 2 copies of e_2, so

$$P(3,2) = P(2,2) + x^3 P(1,3)\,. \tag{8.5}$$

By Lemma 6.1, $P(2,3) = P(3,2)$ and $P(1,3) = P(3,1)$. Also, by Corollary 6.2

$$P(3,1) = 1 + x + x^2 + x^3 = \frac{x^4-1}{x-1}\,. \tag{8.6}$$

By equating the right hand sides of (8.4) and (8.5) and using (8.6),

$$\begin{aligned} P(2,2) &= \frac{(x^3-1)P(3,1)}{x^2-1} \\ &= \frac{(x^4-1)(x^3-1)}{(x^2-1)(x-1)}\,. \end{aligned}$$

The following lemma provides a closed form expression for the generating polynomial $P(n,m)$.

Lemma 8.5. Closed form expression for $P(S)$ where $R(S) = 2$
For $n, m \in \mathbb{Z}^+$,

$$P(n,m) = \frac{G(n+m)}{G(n)G(m)} \qquad \text{where} \qquad G(n) = \prod_{i=1}^{n}(x^i - 1)\,. \tag{8.7}$$

Proof. Let $S = \{e_1^{n+1}e_2^m\}$, $S' = \{e_1^m e_2^{n+1}\}$ where $e_1 < e_2$. By Lemma 6.1, $I_F(S) = I_F(S')$.

Case (1): Let $s \in \sigma(\{e_1^{n+1}e_2^m\})$. There are two subcases for the element in position 1 of s.

A. If $s = e_1 s'$, where $s' \in \sigma[n,m]$, then $\mathrm{Inv}(s) = \mathrm{Inv}(s')$.
B. If $s = e_2 s'$, where $s' \in \sigma[n+1, m-1]$, then $\mathrm{Inv}(s) = \mathrm{Inv}(s') + n + 1$.

By Theorem 7.2, we have

$$P(n+1,m) = P(n,m) + x^{n+1}P(n+1,m-1)\,. \tag{8.8}$$

Case (2): Let $s \in \sigma(\{e_2^{n+1}e_1^m\})$. By a similar argument to the previous case, we have

$$P(n+1,m) = P(n+1,m-1) + x^m P(n,m)\,. \tag{8.9}$$

By equating the right hand sides of (8.8) and (8.9), we see that

$$P(n,m) = \frac{x^{n+1}-1}{x^m-1}\,P(n+1,m-1)\,. \tag{8.10}$$

By applying (8.10) repeatedly, we have

$$
\begin{aligned}
P(n,m) &= \frac{(x^{n+1}-1)(x^{n+2}-1)}{(x^m-1)(x^{m-1}-1)}\, P(n+2,m-2) \\
&= \frac{(x^{n+1}-1)(x^{n+2}-1)(x^{n+3}-1)}{(x^m-1)(x^{m-1}-1)(x^{m-2}-1)}\, P(n+3,m-3) \\
&\;\;\vdots \\
&= \prod_{i=1}^{m} \frac{(x^{n+i}-1)}{(x^{m+1-i}-1)}\, P(n+m,0) \\
&= \prod_{i=1}^{m} \frac{(x^{n+i}-1)}{(x^{m+1-i}-1)} \\
&= \frac{\displaystyle\prod_{i=1}^{m}(x^{n+i}-1)}{\displaystyle\prod_{i=1}^{m}(x^{i}-1)} \\
&= \frac{\displaystyle\prod_{i=1}^{m}(x^{n+i}-1)}{\displaystyle\prod_{i=1}^{m}(x^{i}-1)} \times \frac{\displaystyle\prod_{i=1}^{n}(x^{i}-1)}{\displaystyle\prod_{i=1}^{n}(x^{i}-1)} \\
&= \frac{\displaystyle\prod_{i=1}^{n+m}(x^{i}-1)}{\displaystyle\prod_{i=1}^{m}(x^{i}-1)\prod_{i=1}^{n}(x^{i}-1)} \\
&= \frac{G(n+m)}{G(n)G(m)}\,.
\end{aligned}
$$

This completes the proof. $\square$

We now finally consider the main result of the thesis, namely a closed form expression for the generating polynomial $P(n_1, n_2, \ldots, n_k)$.

Theorem 8.6. A closed form expression for $P(n_1, n_2, \ldots, n_k)$
For $n_1, n_2, \ldots, n_k \in \mathbb{Z}^+$,

$$P(n_1, n_2, \ldots, n_k) = \frac{G(n_1 + n_2 + \cdots + n_k)}{G(n_1)G(n_2)\cdots G(n_k)}\,.$$

Proof. By Lemma 8.2 and Lemma 8.5,

$$\begin{aligned}
& P(n_1, n_2, \dots, n_k) \\
= \; & P(n_1, n_2 + \cdots + n_k) P(n_2, n_3 + \cdots + n_k) \cdots P(n_{k-1}, n_k) \\
= \; & \frac{G(n_1 + \cdots + n_k)}{G(n_1) G(n_2 + \cdots + n_k)} \frac{G(n_2 + \cdots + n_k)}{G(n_2) G(n_3 + \cdots + n_k)} \cdots \frac{G(n_{k-1} + n_k)}{G(n_{k-1}) G(n_k)} \\
= \; & \frac{G(n_1 + n_2 + \cdots + n_k)}{G(n_1) G(n_2) \cdots G(n_k)} .
\end{aligned}$$

□

Note that the Partition Family Theorem (Theorem 6.6) was not used in the proof of Theorem 8.6. Furthermore, the symmetry of $P(n_1, n_2, \dots, n_k)$ provides an alternative proof to Theorem 6.6.

Example 8.7. Let us calculate $P(1, 2, 3, 4)$:

$$\begin{aligned}
P(1,2,3,4) &= P(1, 2+3+4)\, P(2, 3+4)\, P(3,4) \quad \text{by Lemma 8.2} \\
&= P(1,9)\, P(2,7)\, P(3,4) \\
&= P(9,1)\, P(7,2)\, P(4,3) \\
&= \frac{(x^{10}-1)}{(x-1)} \times \frac{(x^8-1)(x^9-1)}{(x^2-1)(x-1)} \times \frac{(x^5-1)(x^6-1)(x^7-1)}{(x^3-1)(x^2-1)(x-1)} .
\end{aligned}$$

Therefore, cancelling and multiplying gives

$$\begin{aligned}
P(1,2,3,4) = {} & 1 + 3x + 8x^2 + 17x^3 + 33x^4 + 57x^5 + 93x^6 + 141x^7 + 204x^8 + 280x^9 \\
& + 369x^{10} + 466x^{11} + 568x^{12} + 667x^{13} + 758x^{14} + 833x^{15} + 887x^{16} + 915x^{17} \\
& + 915x^{18} + 887x^{19} + 833x^{20} + 758x^{21} + 667x^{22} + 568x^{23} + 466x^{24} + 369x^{25} \\
& + 280x^{26} + 204x^{27} + 141x^{28} + 93x^{29} + 57x^{30} + 33x^{31} + 17x^{32} + 8x^{33} \\
& + 3x^{34} + x^{35} .
\end{aligned}$$

The coefficients of $P(1, 2, 3, 4)$ above agree with the row entry '4-3-2-1' in Table A7 which is calculated using Theorem 6.6.

The original theorem by Muir [13] in 1899 for the permutation group S_n can be recovered from Theorem 8.6 by setting $n_i = 1$ for $i = 1, 2, \dots, k$.

Corollary 8.8. (Muir)

$$P(\overbrace{1, 1, \dots, 1}^{k}) = \frac{1}{(x-1)^k} \prod_{i=1}^{k} (x^i - 1) .$$

8.2 Integer partition polynomial

We will next establish the relationship between the coefficients of the generating polynomial for $P([n, n])$ and $p(n)$, the integer partition of n. Recall that $A(n, p, m)$ is number of partitions of n into p parts of size at most m. It also corresponds to the number of insertions of m elements into a sequence of length p where the sum of inversion count positions is equal to n. The ideas were demonstrated in Examples 4.2–4.3 and formalised in Lemma 5.3.

Applying Corollary 5.5 with $n_1 = n_2 = n$ gives

$$P([n,n]) = \sum_{i=0}^{n^2} A(i,n,n)x^i \,.$$

For $0 \leq i \leq n$, the insertion positions therefore corresponds to the integer partitions of i.

Example 8.9. Let $S = [6,6]$. By Theorem 8.6,

$$P(S) = \frac{G(12)}{G(6)\,G(6)} = 1 + x + 2x^2 + 3x^3 + 5x^4 + 7x^5 + 11x^6 + \cdots \,.$$

The sum of coefficients of $x^0, x^1, \ldots, x^6$ correspond to $p(1), p(2), \ldots, p(6)$, the number of integer partitions of $1, 2, \ldots, 6$, respectively (see A000041 in the OEIS [15]). In Table 8.1, the 11 partitions of the integer 6 are mapped to insertion positions as shown.

	Insertion Position						
Partition	0	1	2	3	4	5	6
6	5	0	0	0	0	0	1
5+1	4	1	0	0	0	1	0
4+2	4	0	1	0	1	0	0
4+1+1	3	2	0	0	1	0	0
3+3	4	0	2	0	0	0	0
3+2+1	3	1	1	1	0	0	0
3+1+1+1	2	1	0	1	0	0	0
2+2+2	3	0	3	0	0	0	0
2+2+1+1	2	2	2	0	0	0	0
2+1+1+1+1	1	4	1	0	0	0	0
1+1+1+1+1+1	0	6	0	0	0	0	0

Table 8.1: Partition $\leftrightarrow$ Insertion Position

Table 8.1 represents the 11 different possible ways of inserting 6 copies of e_1 into the sequence $e_2e_2e_2e_2e_2e_2$. The possible insertion positions are $0, 1, \ldots, 6$. For instance, the partition 3+2+1 corresponds insertion one copy of e_1 into each of position 3, 2, 1. The remaining three copies are inserted into position 0.

Lemma 8.10. Let $P(S) = \sum_{k=0}^{n^2} f_k\, x^k = \frac{G(2n)}{G(n)G(n)}$ be the generating polynomial of $S = [n,n]$. Then $f_k = p(k)$ for each $0 \leq k \leq n$.

Proof. Let $S = \{a^n\, b^n\}$. Then each sequence in $\sigma(S)$ is constructed by inserting n copies of the element a into n copies of b into positions $i = 0, 1, \ldots, n$. Let $q_i \geq 0$ denote the number of copies of a inserted into position i. Then

$$q_0 + q_1 + \cdots + q_n = n \,.$$

Let $I(k)$ be the set of $(n+1)$-tuples where the insertion results in an increase of k in the inversion count. Note that the elements of $I(K)$ are not necessarily in order. Then

$$I(k) = \{(q_0, q_1, \ldots, q_n) \,|\, q_0 + q_1 + \cdots + q_n = n,\ \sum_{i=0}^{n} i\, q_i = k\}\,.$$

Let $J(k)$ be the set of integer partitions of k over m summands. Thus,

$$J(k) = \{(p_1, p_2, \ldots, p_m) \,|\, \sum_{i=1}^{m} p_i = k,\ p_1, p_2, \ldots, p_m \in \mathbb{Z}^+\}\,.$$

For each partition $\pi = (p_1, \ldots, p_m)$ of k, let ℓ_i be the number of times that integer i occurs in the partition π. Then

$$\ell_i = \sum_{j=1}^{m} \delta(i, p_j)\,,$$

where $\delta(i, p_j)$ is the Kronecker delta. Define $\ell_0 = m - \sum_{i=1}^{m} \ell_i$ and note that $l_0 \geq 0$. Also, define

$$L(k) = \{(\ell_0, \ell_1, \ldots, \ell_m) \,|\, \ell_0 + \ell_1 + \cdots + \ell_m = k,\ \sum_{i=0}^{m} i\, \ell_i = k\}\,.$$

Let $\ell = (\ell_0, \ell_1, \ldots, \ell_m) \in L(k)$ and let $p = (p_1, p_2, \ldots, p_n)$ be a n-tuple which is initially filled with zeros. Let $\mathcal{M}_k : \ell \to p$ be the mapping that sequentially replaces each set of ℓ_i leftmost zero coordinates in p by ℓ_i copies of i, for $i = 1, 2, \ldots, m$. The resulting object $\mathcal{M}_k = p$ represents the insertion positions of n copies of b into n copies of a such that the inversion count of the resultant sequence is k. Therefore, $c \in I(k)$ and so $\mathcal{M}_k$ is a mapping from $L(k) \to J(k)$. By this construction, $\mathcal{M}_k$ is injective.

For a given partition $p = (p_1, p_2, \ldots, p_n)$ of S where $p_1 + p_2 + \cdots + p_m = k$, let ℓ_i be the number of coordinates in p with the value i where $1 \leq i \leq m$ and let $\ell_0 = k - \sum_{i=1}^{m} \ell_i$. Since $\ell = (\ell_1, \ell_2, \ldots, \ell_m) \in L(k)$, $\mathcal{M}_k$ is surjective. Furthermore, we have $L(k) = I(k)$. By the definition of the generating polynomial $P(S)$, $f_k = |I(k)| = |J(k)|$ for $0 \leq k \leq n$. The proof is now complete. $\square$

Example 8.11. By proving and using an extension of the Euler Pentagonal Theorem [1]

$$p(n) = p(n-1) + p(n-2) - p(n-5) - p(n-7) + \cdots\,.$$

P.A. MacMahon [12] calculated by hand the values $p(1), \ldots, p(200)$, which took an estimated 20,000 operations. By applying Lemma 8.10 with $n = 200$, we calculated

$$p(200) = 3,972,999,029,388$$

which took Matlab 2.4 seconds on a P7 Pentium Processor. This value has historical significance since it was used to verify the Hardy-Ramanujan Asymptotic

Formula [8] for integer partitions:

$$p(n) \approx \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{2n}{3}}}.$$

Chapter 9

Analysis of the distribution of sequences

9.1 Datasets

The purpose of this section is to analyse the fit of natural and computer generated sequences with the expected inversion frequency distribution at the integer partitions level. The element sets are 10^n digits of the numbers $\overbrace{0\ldots0}^{n}$ to $\overbrace{9\ldots9}^{n}$ for $n = 6, 7, 8, 9, 10$ extracted from consecutive digits of the datasets. Ten datasets including six natural sequences and four computer generated sequences are used for the thesis.

1. The 5,000,000th Fibonacci number with 1,044,938 digits created by a Python application. Denote F_n as the n^{th} Fibonacci number. Now $F_n = F_{n-1} + F_{n-2}$, $n \geq 3$, so the final $k \geq 1$ digits of F_n forms a cycle whenever a pair of consecutive terms have the same values. Since there are 10^{2k} choices for the consecutive pair, then it follows that the digits of the Fibonacci numbers must form cycles. The cycle lengths for $n = 1, 2$ are 60, 300 respectively. For $n \geq 3$, the cycle is 1.5×10^n [17]. It is an interesting study to determine if the cycle of digits affects the partition and inversion frequency distributions.
2. The first 5 million digits of $\sqrt{2}$ created by a python application.
3. Dataset is created by approximately the first 2 million digits of e [7].
4. The largest known prime at the time of writing, GIMPS prime $2^{74,207,281} - 1$ with 22,338,618 digits [8].
5. Dataset of $1,437,849$ digits created by a Python application for 300000! with the trailing zeros stripped off.
6. Dataset formed by the first billion digits of π.
7. Dataset of one billion digits created by Microsoft VBA (Visual Basic for Application).
   (i) The dataset MS_A consisting of approximately 10^9 digit using a Visual Basic for Application Version 1640.
   (ii) It was discovered that rnd() call to return 9 digit numbers has a loop of 100,663,295 irrespective of the seeding. The dataset MS_B contains all the digits of a single cycle in MS_A.
8. The dataset MS_C of 10^9 digits created by Visual C# 2012.
9. A dataset with 10^9 digits created by a Python 3.5 application. The Python engine is based on entropic values of the environmental variables.
10. A dataset with 10^9 digits using the function randi() in MATLAB R2017b by concatenating 10 digit numbers.

Tables A24–A34 provide a breakdown of the mean inversion counts and the frequency of the partition counts for each of the datasets. The application for the data extraction and calculations is documented in Appendix A2.

For a dataset of size N, sequences of n consecutive digits are extracted from the datasets as a sliding window where $n = [\log_{10} N]$ digits. For example, the dataset for $\sqrt{2}$, $n = [\log_{10} 5000000] = 6$, the first number is obtained from positions 1 to 6, the second number from positions 2 to 7, and the 49999995^{th} number from positions 4999995 to 5000000.

The tests conducted in the thesis are the distribution of partition and the inversion count for the datasets. The partition distribution is categorical data and therefore normal distribution analysis cannot be applied to it. Preliminary study of inversion distribution using Kurtosis count [18] suggests that the distribution is asymptotically normal for large values of n. The Pearson's χ^2 testing [7] is chosen because it does not assume normality although it does assume finite variances and finite covariance which is the case for the datasets. It is applicable to categorical data which can be classified into mutually exclusive classes where the probability of each class is known. For instance, in the gaming industry, it can be used to test loaded dice, slot machine randomness and the gravitational tilt of roulette tables. The three χ^2 tests conducted in this chapter are:

A Apply χ^2 test of the actual partition probability in Table A24 to Table A34 with the expected partition distribution for the datasets in Table A8 and Table A9.

B Apply χ^2 test to the actual partition mean of inversion count in Table A24 to Table A34 for each partition of the datasets with the expected inversion count mean.

C Apply χ^2 test to the IFD for the dataset partitions with the calculated distributions in (Tables A10 to Table A23).

The following legends are used for the tables in this chapter.

1. χ^2 - Pearson coefficient
2. DF - Degrees of freedom.
3. CV - Critical Value (χ^2 value for 0.95)

9.2 χ^2 test for the partition probability

The purpose of this section is to establish for a given value of n, the conformance of the datasets to the expected partition probabilities. Tables A8–A9 tabulate the probabilities that each partition occurs if we assume that each digit is chosen uniformly at random for $n = 6, 7, 8, 9, 10$. In applying the χ^2 test, the degree of freedom is $p(n) - 1$ where $p(n)$ is the integer partition of n. The level of significance is $\alpha = 0.95$. The Pearson correlation coefficient is

$$\chi^2 = \sum_{i=1}^{I_{p(n)}} \frac{(E_i - X_i)^2}{E_i}\,, \tag{9.1}$$

where

$I_{p(n)}$ = Number of integer partitions of n.
E_i = Probability of partition i × size of dataset.
(See Tables A8–A9)
X_i = Actual population of the partition i in the dataset.

The null hypothesis H_0 is that the population spread of the partition for the dataset is consistent with the expected population spread. The alternative hypothesis H_1 is that the population spread of the partition for the dataset is not consistent with the expected population spread.

Table 9.1: χ^2 test for the distribution of partitions for the datasets

Dataset	Size	Digits	$I_{p(n)}$	χ^2	Conclusion
$F_{5000000}$	1044930	6	11	0.73	H_0
300000!	1437846	6	11	0.97	H_1
$\sqrt{2}$	4999995	6	11	0.07	H_0
e	2000063	6	11	0.26	H_0
M_{49}	22338612	7	15	0.95	H_1
π	999999992	9	30	0.62	H_0
MSC_A	1000004008	9	30	1.00	H_1
MSC_B	100663295	8	22	0.59	H_0
MSC_C	1083333411	9	30	1.00	H_1
Python	999995552	9	30	0.99	H_1
MATLAB	999999991	9	30	1.00	H_1

It is evident that the partition distributions for all the computer generated sequences for the datasets do not satisfy the expected distributions.

9.3 χ^2 test for the inversion count mean of the partition

For each dataset, the expected mean of the inversion count for each partition is compared with the actual value. Corollary 5.2 proved that the expected inversion mean for a partition is equal to the median.

For a given dataset, the χ^2 test is applied over the partitions. In Tables A27–A34, the mean value of the inversion count at the partition level is calculated and are listed alongside of the expected mean. The χ^2 test is applied to the partitions of the datasets.

In applying the χ^2 test, the degree of freedom is $p(n) - 1$ where $p(n)$ is the integer partition of n. The level of significance is $\alpha = 0.95$. The Pearson correlation coefficient is

$$\chi^2 = \sum_{i=1}^{I_{p(n)}} \frac{P_i(E_i - X_i)^2}{E_i}, \tag{9.2}$$

where

$I_{p(n)}$ = Number of integer partitions of n
E_i = Expected mean value of the inversion count for the partition
E_i = Probability of partition i × size of dataset (see Tables A8–A9)
X_i = Actual mean of the partition i (see Tables A24–A34)
P_i = Population of partition i.

The null hypothesis H_0 is that the mean of the inversion count of the partition for the dataset is consistent with the expected population mean. The alternative hypothesis H_1 is that the mean of the inversion count of the partition for the dataset is not consistent with the expected population mean.

Table 9.2: Pearson test for inversion count mean by partition

Dataset	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
F_{500000}	10	12.62	0.75	H_0
300000!	10	17.03	0.93	H_0
$\mathbf{M}_{49}$	15	12.62	0.99	$\mathbf{H}_1$
$\sqrt{2}$	10	12.26	0.73	H_0
e	10	7.70	0.34	H_0
π	29	62.23	0.99	$\mathbf{H}_1$
$\mathbf{MS}_A$	29	990.55	1.00	$\mathbf{H}_1$
$\mathbf{MS}_B$	21	505.86	1.00	$\mathbf{H}_1$
$\mathbf{MS}_C$	29	44.91	0.97	$\mathbf{H}_1$
Python	29	43.72	0.96	$\mathbf{H}_1$
MATLAB	29	380.01	1.00	$\mathbf{H}_1$

9.4 χ^2 test of the IFD for the datasets

In this section, we will analyse the fit of the inversion count distribution between the calculated values in Tables A1–A7 and that of the ten datasets in Tables A10–A23. The χ^2 test is on the spread of the inversion count for each partition. For instance, the dataset $F_{5000000}$ contains 1044930 digits. The number of digits extracted from the dataset $n = [\log_{10} 1044930] = 6$ which has 11 partitions. For each partitions in Table A10, χ^2 is calculated from the spread of inversion count. For the partition $[n_1, n_2, \ldots, n_k]$, the Pearson correlation coefficient is

$$\chi^2 = \sum_{i=0}^{M} \frac{(E_i - X_i)^2}{E_i}\,, \tag{9.3}$$

where

M	=	Maximum inversion count for partition $(n_1, n_2, \ldots, n_k)$ (see Lemma 2.2)
E_i	=	Probability of inversion count i × size of partition (see Tables A1–A5)
X_i	=	Count of inversion count i in the partition $(n_1, n_2, \ldots, n_k)$ for the dataset.

The null hypothesis H_0 is that the population of the spread of the inversion count for the a partition of the dataset is consistent with the spread of the expected population. The alternative hypothesis H_1 is that the population of the spread of the inversion count for the a partition of the dataset is not consistent with the spread of the expected population.

The detailed analysis for the datasets in Table 9.3 below can be found in Tables A10–A23 in the appendices.

Table 9.3: χ^2 inversion count analysis for the partitions of datasets

Dataset	Size	No. digits	Partitions Tested	Partitions Failed
$F_{5000000}$	1044930	6	11	1
300000!	1437846	6	11	0
$\sqrt{2}$	4999995	6	11	0
e	2000063	6	11	0
M_{49}	22338612	7	15	1
π	999999995	6	11	1
π	999999994	7	15	1
π	999999993	8	22	3
π	999999992	9	30	3
$\mathbf{MS}_A$	1000004008	9	30	30
$\mathbf{MS}_B$	100663288	9	22	5
$\mathbf{MS}_C$	1083333411	9	30	4
Python	999995552	9	30	3
MATLAB	999999991	9	30	6

9.5 Summary of distributional tests

1. e and $\sqrt{2}$ passed all three tests.
2. $F_{5000000}$ passed two of the three tests.
3. M_{49}, π and 30000! passed one of the three tests.
4. Three partitions for π failed the χ^2 test for the inversion count distribution for 9 consecutive digits ($n = 9$). As a result, tests were also conducted for $n = 6, 7, 8$ to determine the parent/child relationship for the failing partitions.
   (A) For $n = 6$, the dataset passed 10 out 11 tests.
   The partition (3,2,1) failed the χ^2 test.
   (B) For $n = 7$, the dataset passed 14 out 15 tests.
   The partition (4,3) failed the χ^2 test.
   (C) For $n = 8$, the dataset passed 19 out 22 tests.
   The partitions (1,1,1,1,1,1,1,1), (3-2-1-1-1), (4,4) failed the χ^2 test.
   (D) For $n = 9$, the dataset passed 27 out 30 tests.
   The partitions (3-3-2-1), (4-2-2-1), (4-4-1) failed the χ^2 test.

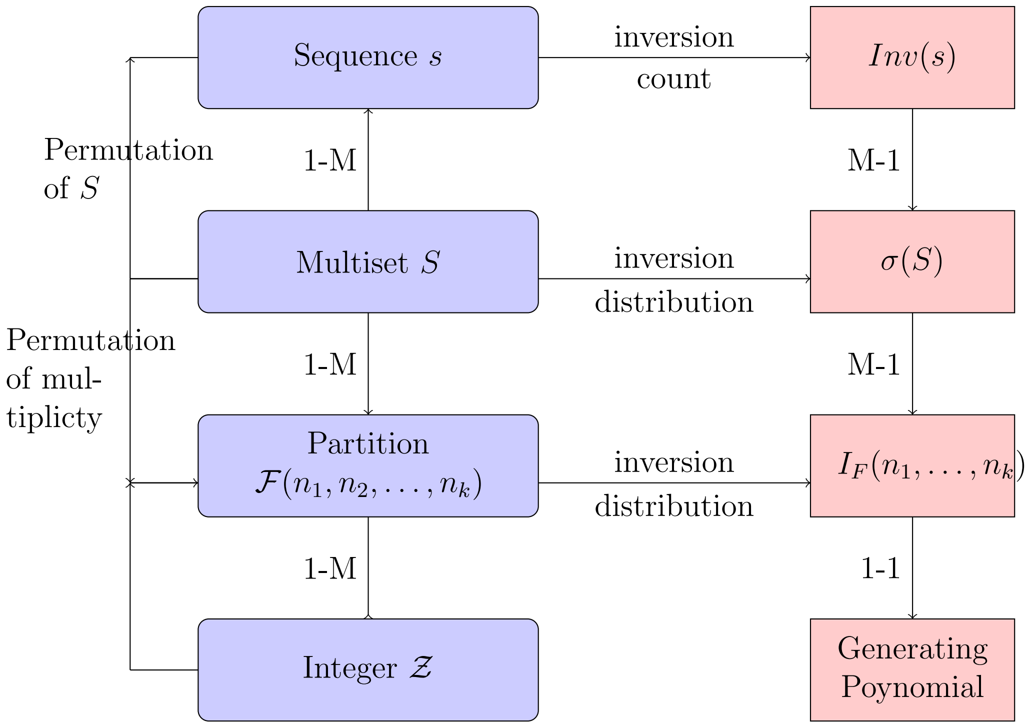
   Note the parent/child relationships between the partitions (3-2-1) and (3-3-2-1) and between (4-3), (4-4), (4-4-1).
5. All four computer generated sequences failed all three tests against the expected values.
6. The digits of dataset MS_A contain repeated sets of the 100,663,295 digits and failed all the partition tests. Let k be the repetition factor for MS_A. In Equation (9.2), by substituting E_i and X_i by kE_i and kX_i, respectively, χ^2 increases by a factor of k. Thus, the repetition of data resulted in all the partitions failing the χ^2 test.

Chapter 10

Conclusion

The first part of the thesis is a study of expected inversion distribution sequences by insertion techniques. It provides an elementary and self-contained approach to the structure of the permutations of multisets and the relationships. This approach makes this structure clearer and more accessible for readers than previous approaches such as Stanley's "semi-combinatorial" proof [16, p. 64]. The hierarchical structure of partitions and their relationships is summarised in the Entity Relationship diagram below.

Figure 10.1: Entity Relationship Diagram



The closed form for the generating function is created by:

- Four tiers structure of integer $\rightarrow$ partition $\rightarrow$ multiset $\rightarrow$ sequence.
- Permutation of multiset by the ordering of elements.
- Permutation of multiset by the multiplicities.
- Insertion method as upper diagonal of hypercube.
- Insertion method into leading position of a sequence.
- Expansion of generating function as products of generating polynomial with two distinct elements. (Polynomial of rank 2)

- Closed form for generating polynomial $P(n_1, n_2, \ldots, n_k)$.

The insertion method provided a tool for linking the Ferrer diagram in with integer partition and as a result, a generating polynomial for integer partition was delivered.

These are potential areas for further research:

1. Find an asymptotic function for $P(n_1, n_2, \ldots, n_k)$. Preliminary studies indicate that such a function is asymptotic normal. For the symmetric group $S_n = P(\overbrace{1, 1, \ldots, 1}^{n})$, Conger and Viswanath [4] gave the approximation function by a probabilistic approach:
$$\left| P\left(\frac{\operatorname{Inv}(\pi) - \frac{1}{2}\binom{n}{2}}{\sqrt{n(n-1)(2n+5)/12}} \le x \right) - \phi(x) \right| \le \frac{C}{\sqrt{n}}\,.$$
   (A) π is an element of the permutation group S_n.
   (B) $\operatorname{Inv}(\pi)$ is the inversion count of the permutation.
   (C) $\phi(x)$ is the standardised normal function.
   (D) $P()$ is the probability function.
2. Establish the asymptotic function for an integer n by summing all the partitions by the probability of the partition. This could be a very difficult task. Preliminary study indicates that the function is slightly skewed to the left. The recommended approach is a probabilistic rather than deriving an exact function.
3. Of lesser practical importance but higher in academic pursuit is a combinatorial closed form for $P(n_1, n_2, \ldots, n_k)$ generalising Knuth's pentagonal expansion in (3.3).
4. Another method of measuring inversion count that is more pertinent to computing science is to define the inversion count as the sum distance between pairs of order. Sort algorithms such as the Bubble and Merge sorts [5] compares (near) adjacent pairs and progressively reduce the distance between pairs out of order on each pass.
5. The thesis assumes that each element has equal probability of being selected. While this is applicable to digits of natural and computer generated sequence, in the real world, the model needs to be adjusted by the probabilities of elements being selected.
6. A partial sort is the ranking of top k items from a set of size of n. For instance, an internet search may retrieve 10 million items but it is likely the user will only want to see the first 100. The efficiency of a sort algorithm is determined by the number of comparisions $C(n, k)$. It is a rich topic of practical importance.

The second part of the thesis is the analysis of inversion frequencies and partition distributions were applied to computer generated (MATLIB, Python and Microsoft VBA and C++) and natural sequences ($\sqrt{2}$, e, π, $M49$, $n!$ and Fibonacci numbers).

The conclusions are:

- The natural sequences conform better than the computer generated sequences in the expected values of partition and inversion frequency distributions.

- There are some issues with the randomness of the first billion digits of π. It could be an interesting study to increase the size of the database to determine if the partitions failing the tests spur negative child patterns.
- The Microsoft randomiser for Visual Basic for Application produces repeated patterns irrespective of seeding.

Appendices

A1 Supporting data tables

Table A1: IFD table for $n = 6$

	Proba			IFD															
Partition	-bility	Mean	SD	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1-1-1-1-1-1	0.15	7.5	2.66	1	5	14	29	49	71	90	101	101	90	71	49	29	14	5	1
1-1-1-1-2	0.45	7	2.61	1	4	10	19	30	41	49	52	49	41	30	19	10	4	1	
1-1-2-2	0.23	6.5	2.57	1	3	7	12	18	23	26	26	23	18	12	7	3	1		
2-2-2	0.01	6	2.52	1	2	5	7	11	12	14	12	11	7	5	2	1			
1-1-1-3	0.1	6	2.48	1	3	6	10	14	17	18	17	14	10	6	3	1			
1-2-3	0.04	5.5	2.43	1	2	4	6	8	9	9	8	6	4	2	1				
3-3	0	4.5	2.29	1	1	2	3	3	3	3	2	1	1						
1-1-4	0.01	4.5	2.22	1	2	3	4	5	5	4	3	2	1						
2-4	0	4	2.16	1	1	2	2	3	2	2	1	1							
1-5	0	2.5	1.71	1	1	1	1	1	1										
6	0	0	0	1															

In Tables A2–A7, $k/M-k$ denotes the calculated inversion count frequency for f_k and $f_{m(S)-k}$ (see Lemma 5.1).

Table A2: Expected IFD for $n = 7$

			Std.	Max	IFD $k/M-k$										
Partition	Probabilty	Mean	Dev	Inv	0	1	2	3	4	5	6	7	8	9	10
1-1-1-1-1-1-1	0.06048	10.5	10.5	21	1	6	20	49	98	169	259	359	455	531	573
2-1-1-1-1-1	0.31752	10	10	20	1	5	15	34	64	105	154	205	250	281	292
2-2-1-1-1	0.31752	9.5	9.5	19	1	4	11	23	41	64	90	115	135	146	
2-2-2-1	0.05292	9	9	18	1	3	8	15	26	38	52	63	72	74	
3-1-1-1-1	0.10584	9	9	18	1	4	10	20	34	51	69	85	96	100	
3-2-1-1	0.10584	8.5	8.5	17	1	3	7	13	21	30	39	46	50		
3-2-2	0.00756	8	8	16	1	2	5	8	13	17	22	24	26		
3-3-1	0.00504	7.5	7.5	15	1	2	4	7	10	13	16	17			
4-1-1-1	0.01764	7.5	7.5	15	1	3	6	10	15	20	24	26			
4-2-1	0.00756	7	7	14	1	2	4	6	9	11	13	13			
4-3	0.000315	6	6	12	1	1	2	3	4	4	5				
5-1-1	0.001512	5.5	5.5	11	1	2	3	4	5	6					
5-2	0.000189	5	5	10	1	1	2	2	3	3					
6-1	0.000063	3	3	6	1	1	1	1							
7	0.000001	0	0	0	1										

Table A3: Expected IFD for $n = 8$

			Std.	Max	IFD $k/M - k$													
Partition	Prob.	Mean	Dev	Inv	0	1	2	3	4	5	6	7	8	9	10	12	13	14
1-1-1-1-1-1-1-1	0.018144	14	4.04	28	1	7	27	76	174	343	602	961	1415	1940	2493	3450	3736	3836
2-1-1-1-1-1	0.169344	13.5	4.01	27	1	6	21	55	119	224	378	583	832	1108	1385	1818	1918	
2-2-1-1-1-1	0.31752	13	3.98	26	1	5	16	39	80	144	234	349	483	625	760	946	972	
2-2-2-1-1	0.127008	12.5	3.95	25	1	4	12	27	53	91	143	206	277	348	412	486		
2-2-2-2	0.005292	12	3.92	24	1	3	9	18	35	56	87	119	158	190	222	248		
3-1-1-1-1-1	0.084672	12.5	3.93	25	1	5	15	35	69	120	189	274	369	465	551	651		
3-2-1-1-1	0.169344	12	3.89	24	1	4	11	24	45	75	114	160	209	256	295	330		
3-2-2-1	0.042336	11.5	3.86	23	1	3	8	16	29	46	68	92	117	139	156			
3-3-1-1	0.014112	11	3.81	22	1	3	7	14	24	37	53	70	86	100	109			
3-3-2	0.002016	10.5	3.77	21	1	2	5	9	15	22	31	39	47	53	56			
4-1-1-1-1	0.021168	11	3.76	22	1	4	10	20	35	55	79	105	130	151	165			
4-2-1-1	0.021168	10.5	3.73	21	1	3	7	13	22	33	46	59	71	80	85			
4-2-2	0.001512	10	3.7	20	1	2	5	8	14	19	27	32	39	41	44			
4-3-1	0.002016	9.5	3.64	19	1	2	4	7	11	15	20	24	27	29				
4-2-2-1	0.0000315	8	3.46	16	1	1	2	3	5	5	7	7	8					
5-1-1-1	0.0028224	9	3.49	18	1	3	6	10	15	21	27	32	35	36				
5-2-1	0.0012096	8.5	3.45	17	1	2	4	6	9	12	15	17	18					
5-3	0.0000504	7.5	3.35	15	1	1	2	3	4	5	6	6						
6-1-1	0.0002016	6.5	3.04	13	1	2	3	4	5	6	7							
6-2	0.0000252	6	3	12	1	1	2	2	3	3	4							
7-1	0.0000072	3.5	2.29	7	1	1	1	1										
8	0.0000001	0	0	9	1													

Table A4: Expected IFD for $n = 9$, Part A

			Std.	Max	Inversion count Frequency $k/M - k$								
Partition	Probabilty	Mean	SD	Inv	0	1	2	3	4	5	6	7	8
1-1-1-1-1-1-1-1-1	0.0036288	18	4.8	36	1	8	35	111	285	628	1230	2191	3606
2-1-1-1-1-1-1-1	0.0653184	17.5	4.77	35	1	7	28	83	202	426	804	1387	2219
2-2-1-1-1-1-1	0.2286144	17	4.74	34	1	6	22	61	141	285	519	868	1351
2-2-2-1-1-1	0.190512	16.5	4.72	33	1	5	17	44	97	188	331	537	814
2-2-2-2-1	0.0285768	16	4.69	32	1	4	13	31	66	122	209	328	486
3-1-1-1-1-1-1	0.0508032	16.5	4.7	33	1	6	21	56	125	245	434	708	1077
3-2-1-1-1-1	0.190512	16	4.67	32	1	5	16	40	85	160	274	434	643
3-2-2-1-1	0.1143072	15.5	4.65	31	1	4	12	28	57	103	171	263	380
3-2-2-2	0.0063504	15	4.62	30	1	3	9	19	38	65	106	157	223
3-3-1-1-1	0.0254016	15	4.6	30	1	4	11	25	49	86	139	209	295
3-3-2-1	0.0127008	14.5	4.57	29	1	3	8	17	32	54	85	124	171
3-3-3	0.0002016	13.5	4.5	27	1	2	5	10	17	27	41	56	74
4-1-1-1-1-1	0.0190512	15	4.56	30	1	5	15	35	70	125	204	309	439
4-2-1-1-1	0.0381024	14.5	4.54	29	1	4	11	24	46	79	125	184	255
4-2-2-1	0.0095256	14	4.51	28	1	3	8	16	30	49	76	108	147
4-3-1-1	0.0063504	13.5	4.46	27	1	3	7	14	25	40	60	84	111
4-3-2	0.0009072	13	4.43	26	1	2	5	9	16	24	36	48	63
4-4-2	0.0002268	12	4.32	24	1	2	4	7	12	17	24	31	39
5-1-1-1-1	0.0038102	13	4.34	26	1	4	10	20	35	56	83	115	150
5-2-1-1	0.0038102	12.5	4.31	25	1	3	7	13	22	34	49	66	84
5-2-2	0.0002722	12	4.28	24	1	2	5	8	14	20	29	37	47
5-3-1	0.0003629	11.5	4.23	23	1	2	4	7	11	16	22	28	34
5-4	0.0000113	10	4.08	20	1	1	2	3	5	6	8	9	11
6-1-1-1	0.0004234	10.5	3.99	21	1	3	6	10	15	21	28	35	41
6-2-1	0.0001814	10	3.96	20	1	2	4	6	9	12	16	19	22
6-3	0.0000076	9	3.87	18	1	1	2	3	4	5	7	7	8
7-1-1	0.0000259	7.5	3.45	15	1	2	3	4	5	6	7	8	8
7-2	0.0000032	7	3.42	14	1	1	2	2	3	3	4	4	4
8-1	0.0000008	4	2.58	8	1	1	1	1	1	1	1	1	1
9	0	0	0	0	1								

Table A5: Expected IFD for $n = 9$, Part B

			Std.	Max	Inversion count Frequency $k/M-k$									
Partition	Probabilty	Mean	SD	Inv	9	10	11	12	13	14	15	16	17	18
1-1-1-1-1-1-1-1-1	0.0036288	18	4.8	36	5545	8031	11021	14395	17957	21450	24584	27073	28675	29228
2-1-1-1-1-1-1-1	0.0653184	17.5	4.77	35	3326	4705	6316	8079	9878	11572	13012	14061	14614	14614
2-2-1-1-1-1-1	0.2286144	17	4.74	34	1975	2730	3586	4493	5385	6187	6825	7236	7378	7236
2-2-2-1-1-1	0.190512	16.5	4.72	33	1161	1569	2017	2476	2909	3278	3547	3689	3689	3547
2-2-2-2-1	0.0285768	16	4.69	32	675	894	1123	1353	1556	1722	1825	1864	1825	1722
3-1-1-1-1-1-1	0.0508032	16.5	4.7	33	1541	2087	2688	3304	3886	4382	4744	4935	4935	4744
3-2-1-1-1-1	0.190512	16	4.67	32	898	1189	1499	1805	2081	2301	2443	2492	2443	2301
3-2-2-1-1	0.1143072	15.5	4.65	31	518	671	828	977	1104	1197	1246	1246	1197	1104
3-2-2-2	0.0063504	15	4.62	30	295	376	452	525	579	618	628	618	579	525
3-3-1-1-1	0.0254016	15	4.6	30	394	500	605	700	776	825	842	825	776	700
3-3-2-1	0.0127008	14.5	4.57	29	223	277	328	372	404	421	421	404	372	328
3-3-3	0.0002016	13.5	4.5	27	93	110	125	137	142	142	137	125	110	93
4-1-1-1-1	0.0190512	15	4.56	30	589	750	910	1055	1171	1246	1272	1246	1171	1055
4-2-1-1-1	0.0381024	14.5	4.54	29	334	416	494	561	610	636	636	610	561	494
4-2-2-1	0.0095256	14	4.51	28	187	229	265	296	314	322	314	296	265	229
4-3-1-1	0.0063504	13.5	4.46	27	139	166	189	206	215	215	206	189	166	139
4-3-2	0.0009072	13	4.43	26	76	90	99	107	108	107	99	90	76	63
4-4-2	0.0002268	12	4.32	24	45	51	54	56	54	51	45	39	31	24
5-1-1-1-1	0.0038102	13	4.34	26	185	217	243	260	266	260	243	217	185	150
5-2-1-1	0.0038102	12.5	4.31	25	101	116	127	133	133	127	116	101	84	66
5-2-2	0.0002722	12	4.28	24	54	62	65	68	65	62	54	47	37	29
5-3-1	0.0003629	11.5	4.23	23	39	43	45	45	43	39	34	28	22	16
5-4	0.0000113	10	4.08	20	11	12	11	11	9	8	6	5	3	2
6-1-1-1	0.0004234	10.5	3.99	21	45	47	47	45	41	35	28	21	15	10
6-2-1	0.0001814	10	3.96	20	23	24	23	22	19	16	12	9	6	4
6-3	0.0000076	9	3.87	18	8	8	7	7	5	4	3	2	1	1
7-1-1	0.0000259	7.5	3.45	15	7	6	5	4	3	2	1			
7-2	0.0000032	7	3.42	14	3	3	2	2	1	1				
8-1	0.0000008	4	2.58	8										
9	0	0	0	0										

Table A6: Expected IFD for $n = 10$ Part A

			Std.	Max	Inversion count frequency $k/M-k$											
Partition	Prob.	Mean	Dev	Inv	0	1	2	3	4	5	6	7	8	9	10	11
1-1-1-1-1-1-1-1-1-1	0.00036288	22.5	5.59	45	1	9	44	155	440	1068	2298	4489	8095	13640	21670	32683
2-1-1-1-1-1-1-1-1	0.0163296	22	5.568	44	1	8	36	119	321	747	1551	2938	5157	8483	13187	19496
2-2-1-1-1-1-1-1	0.1143072	21.5	5.545	43	1	7	29	90	231	516	1035	1903	3254	5229	7958	11538
2-2-2-1-1-1-1	0.190512	21	5.523	42	1	6	23	67	164	352	683	1220	2034	3195	4763	6775
2-2-2-2-1-1	0.071442	20.5	5.5	41	1	5	18	49	115	237	446	774	1260	1935	2828	3947
2-2-2-2-2	0.00285768	20	5.477	40	1	4	14	35	80	157	289	485	775	1160	1668	2279
3-1-1-1-1-1-1-1	0.0217728	21	5.508	42	1	7	28	84	209	454	888	1596	2673	4214	6300	8982
3-2-1-1-1-1-1	0.1524096	20.5	5.485	41	1	6	22	62	147	307	581	1015	1658	2556	3744	5238
3-2-2-1-1-1	0.190512	20	5.462	40	1	5	17	45	102	205	376	639	1019	1537	2207	3031
3-2-2-2-1	0.0381024	19.5	5.439	39	1	4	13	32	70	135	241	398	621	916	1291	1740
3-3-1-1-1-1	0.031752	19.5	5.424	39	1	5	16	41	90	176	315	524	819	1213	1712	2313
3-3-2-1-1	0.0381024	19	5.401	38	1	4	12	29	61	115	200	324	495	718	994	1319
3-3-2-2	0.0031752	18.5	5.377	37	1	3	9	20	41	74	126	198	297	421	573	746
3-3-3-1	0.0014112	18	5.339	36	1	3	8	18	35	62	103	159	233	326	435	558
4-1-1-1-1-1-1	0.0127008	19.5	5.393	39	1	6	21	56	126	251	455	764	1203	1792	2541	3446
4-2-1-1-1-1	0.047628	19	5.37	38	1	5	16	40	86	165	290	474	729	1063	1478	1968
4-2-2-1-1	0.0285768	18.5	5.346	37	1	4	12	28	58	107	183	291	438	625	853	1115
4-2-2-2	0.0015876	18	5.323	36	1	3	9	19	39	68	115	176	262	363	490	625
4-3-1-1-1	0.0127008	18	5.307	36	1	4	11	25	50	90	150	234	345	484	649	835
4-3-2-1	0.0063504	17.5	5.284	35	1	3	8	17	33	57	93	141	204	280	369	466
4-3-3	0.0001512	16.5	5.22	33	1	2	5	10	18	29	46	66	92	122	155	189
4-4-1-1	0.0007938	16.5	5.188	33	1	3	7	14	26	43	67	98	137	182	232	284
4-4-2	0.0001134	16	5.164	32	1	2	5	9	17	26	41	57	80	102	130	154
5-1-1-1-1-1	0.00381024	17.5	5.204	35	1	5	15	35	70	126	209	324	474	659	875	1114
5-2-1-1-1	0.00762048	17	5.18	34	1	4	11	24	46	80	129	195	279	380	495	619
5-2-2-1	0.00190512	16.5	5.156	33	1	3	8	16	30	50	79	116	163	217	278	341
5-3-1-1	0.00127008	16	5.115	32	1	3	7	14	25	41	63	91	125	164	206	249
5-3-2	0.00018144	15.5	5.091	31	1	2	5	9	16	25	38	53	72	92	114	135
5-4-1	0.00009072	14.5	4.992	29	1	2	4	7	12	18	26	35	46	57	68	78
5-5	0.000001134	12.5	4.787	25	1	1	2	3	5	7	9	11	14	16	18	19
6-1-1-1-1	0.00063504	15	4.916	30	1	4	10	20	35	56	84	119	160	205	251	295
6-2-1-1	0.00063504	14.5	4.89	29	1	3	7	13	22	34	50	69	91	114	137	158
6-2-2	0.00004536	14	4.865	28	1	2	5	8	14	20	30	39	52	62	75	83
6-3-1	0.00006048	13.5	4.822	27	1	2	4	7	11	16	23	30	38	46	53	59
6-4	0.00000189	12	4.69	24	1	1	2	3	5	6	9	10	13	14	16	16
7-1-1-1	0.00006048	12	4.491	24	1	3	6	10	15	21	28	36	44	51	56	59
7-2-1	0.00002592	11.5	4.463	23	1	2	4	6	9	12	16	20	24	27	29	30
7-3	0.00000108	10.5	4.387	21	1	1	2	3	4	5	7	8	9	10	10	10
8-1-1	0.00000324	8.5	3.862	17	1	2	3	4	5	6	7	8	9	9	8	7
8-2	0.000000405	8	3.83	16	1	1	2	2	3	3	4	4	5	4	4	3
9-1	0.00000009	4.5	2.872	9	1	1	1	1	1	1	1	1	1	1		
10	0.000000001	0	0	10	1											

Table A7: Expected IFD for $n = 10$ Part B

	Inversion count Frequency $k/M-k$												
Partition	12	13	14	15	16	17	18	19	20	21	21	22	23
1-1-1-1-1-1-1-1-1-1	32683	47043	64889	86054	110010	135853	162337	187959	211089	230131	243694	250749	250749
2-1-1-1-1-1-1-1-1	19496	27547	37342	48712	61298	74555	87782	100177	110912	119219	124475	126274	124475
2-2-1-1-1-1-1-1	11538	16009	21333	27379	33919	40636	47146	53031	57881	61338	63137	63137	61338
2-2-2-1-1-1-1	6775	9234	12099	15280	18639	21997	25149	27882	29999	31339	31798	31339	29999
2-2-2-2-1-1	3947	5287	6812	8468	10171	11826	13323	14559	15440	15899	15899	15440	14559
2-2-2-2-2	2279	3008	3804	4664	5507	6319	7004	7555	7885	8014	7885	7555	7004
3-1-1-1-1-1-1-1	8982	12265	16095	20352	24851	29352	33579	37246	40087	41886	42502	41886	40087
3-2-1-1-1-1-1	5238	7027	9068	11284	13567	15785	17794	19452	20635	21251	21251	20635	19452
3-2-2-1-1-1	3031	3996	5072	6212	7355	8430	9364	10088	10547	10704	10547	10088	9364
3-2-2-2-1	1740	2256	2816	3396	3959	4471	4893	5195	5352	5352	5195	4893	4471
3-3-1-1-1-1	2313	3002	3753	4529	5285	5971	6538	6943	7154	7154	6943	6538	5971
3-3-2-1-1	1319	1683	2070	2459	2826	3145	3393	3550	3604	3550	3393	3145	2826
3-3-2-2	746	937	1133	1326	1500	1645	1748	1802	1802	1748	1645	1500	1326
3-3-3-1	558	690	822	947	1057	1141	1195	1214	1195	1141	1057	947	822
4-1-1-1-1-1-1	3446	4486	5622	6798	7945	8987	9849	10465	10786	10786	10465	9849	8987
4-2-1-1-1-1	1968	2518	3104	3694	4251	4736	5113	5352	5434	5352	5113	4736	4251
4-2-2-1-1	1115	1403	1701	1993	2258	2478	2635	2717	2717	2635	2478	2258	1993
4-2-2-2	625	778	923	1070	1188	1290	1345	1372	1345	1290	1188	1070	923
4-3-1-1-1	835	1034	1235	1425	1591	1720	1802	1830	1802	1720	1591	1425	1235
4-3-2-1	466	568	667	758	833	887	915	915	887	833	758	667	568
4-3-3	189	224	254	280	299	308	308	299	280	254	224	189	155
4-4-1-1	284	336	383	422	450	465	465	450	422	383	336	284	232
4-4-2	154	182	201	221	229	236	229	221	201	182	154	130	102
5-1-1-1-1-1	1114	1364	1610	1835	2022	2156	2226	2226	2156	2022	1835	1610	1364
5-2-1-1-1	619	745	865	970	1052	1104	1122	1104	1052	970	865	745	619
5-2-2-1	341	404	461	509	543	561	561	543	509	461	404	341	278
5-3-1-1	249	290	326	354	372	378	372	354	326	290	249	206	164
5-3-2	135	155	171	183	189	189	183	171	155	135	114	92	72
5-4-1	78	87	93	96	96	93	87	78	68	57	46	35	26
5-5	19	20	20	19	18	16	14	11	9	7	5	3	2
6-1-1-1-1	295	334	365	385	392	385	365	334	295	251	205	160	119
6-2-1-1	158	176	189	196	196	189	176	158	137	114	91	69	50
6-2-2	83	93	96	100	96	93	83	75	62	52	39	30	20
6-3-1	59	64	66	66	64	59	53	46	38	30	23	16	11
6-4	16	18	16	16	14	13	10	9	6	5	3	2	1
7-1-1-1	59	60	59	56	51	44	36	28	21	15	10	6	3
7-2-1	30	30	29	27	24	20	16	12	9	6	4	2	1
7-3	10	10	9	8	7	5	4	3	2	1	1		
8-1-1	7	6	5	4	3	2	1						
8-2	3	3	2	2	1	1							
9-1													
10													

Table A8: Partition probability $n = 6, 7, 8$

n = 6		n = 7		n = 8	
Partition	Probability	Partition	Probability	Partition	Probability
1-1-1-1-1-1	0.1512	1-1-1-1-1-1-1	0.06048	1-1-1-1-1-1-1-1	0.018144
1-1-1-1-2	0.4536	2-1-1-1-1-1	0.31752	2-1-1-1-1-1	0.169344
1-1-2-2	0.2268	2-2-1-1-1	0.31752	2-2-1-1-1-1	0.31752
2-2-2	0.0108	2-2-2-1	0.05292	2-2-2-1-1	0.127008
1-1-1-3	0.1008	3-1-1-1-1	0.10584	2-2-2-2	0.005292
1-2-3	0.0432	3-2-1-1	0.10584	3-1-1-1-1-1	0.084672
3-3	0.0009	3-2-2	0.00756	3-2-1-1-1	0.169344
1-1-4	0.0108	3-3-1	0.00504	3-2-2-1	0.042336
2-4	0.00135	4-1-1-1	0.01764	3-3-1-1	0.014112
1-5	0.00054	4-2-1	0.00756	3-3-2	0.002016
6	0.00001	4-3	0.000315	4-1-1-1-1	0.021168
		5-1-1	0.001512	4-2-1-1	0.021168
		5-2	0.000189	4-2-2	0.001512
		6-1	0.000063	4-3-1	0.002016
		7	0.000001	4-4	0.0000315
				5-1-1-1	0.0028224
				5-2-1	0.0012096
				5-3	0.0000504
				6-1-1	0.0002016
				6-2	0.0000252
				7-1	0.0000072
				8	0.0000001

Table A9: Partition probability $n = 9, 10$

n = 9		n = 10	
Partition	Probability	Partition	Probability
1-1-1-1-1-1-1-1-1	0.0036288	1-1-1-1-1-1-1-1-1-1	0.00036288
2-1-1-1-1-1-1-1	0.0653184	2-1-1-1-1-1-1-1-1	0.0163296
2-2-1-1-1-1-1	0.2286144	2-2-1-1-1-1-1-1	0.1143072
2-2-2-1-1-1	0.190512	2-2-2-1-1-1-1	0.190512
2-2-2-2-1	0.0285768	2-2-2-2-1-1	0.071442
3-1-1-1-1-1-1	0.0508032	2-2-2-2-2	0.00285768
3-2-1-1-1-1	0.190512	3-1-1-1-1-1-1-1	0.0217728
3-2-2-1-1	0.1143072	3-2-1-1-1-1-1	0.1524096
3-2-2-2	0.0063504	3-2-2-1-1-1	0.190512
3-3-1-1-1	0.0254016	3-2-2-2-1	0.0381024
3-3-2-1	0.0127008	3-3-1-1-1-1	0.031752
3-3-3	0.0002016	3-3-2-1-1	0.0381024
4-1-1-1-1-1	0.0190512	3-3-2-2	0.0031752
4-2-1-1-1	0.0381024	3-3-3-1	0.0014112
4-2-2-1	0.0095256	4-1-1-1-1-1-1	0.0127008
4-3-1-1	0.0063504	4-2-1-1-1-1	0.047628
4-3-2	0.0009072	4-2-2-1-1	0.0285768
4-4-1	0.0002268	4-2-2-2	0.0015876
5-1-1-1-1	0.00381024	4-3-1-1-1	0.0127008
5-2-1-1	0.00381024	4-3-2-1	0.0063504
5-2-2	0.00027216	4-4-3	0.0001512
5-3-1	0.00036288	4-4-1-1	0.0007938
5-4	0.00001134	4-4-2	0.0001134
6-1-1-1	0.00042336	5-1-1-1-1-1	0.00381024
6-2-1	0.00018144	5-2-1-1-1	0.00762048
6-3	0.00000756	5-2-2-1	0.00190512
7-1-1	0.00002592	5-3-1-1	0.00127008
7-2	0.00000324	5-3-2	0.00018144
8-1	0.00000081	5-4-1	0.00009072
9	0.00000001	5-5	0.000001134
		6-1-1-1-1	0.00063504
		6-2-1-1	0.00063504
		6-2-2	0.00004536
		6-3-1	0.00006048
		6-4	0.00000189
		7-1-1-1	0.00006048
		7-2-1	0.00002592
		7-3	0.00000108
		8-1-1	0.00000324
		8-2	0.000000405
		9-1	0.00000009
		10	0.000000001

The following legends apply to Tables A10–A23:

- χ^2 - Pearson coefficient
- DF - Degrees of freedom.
- CV - Critical Value (χ^2 value for 0.95).

Table A10: χ^2 test for IFD - $F_{5000000}$

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1	15	11.32	0.27	H_0
2-1-1-1-1	14	15.88	0.68	H_0
2-2-1-1	13	19.86	0.90	H_0
2-2-2	12	15.25	0.77	H_0
3-1-1-1	12	18.85	0.91	H_0
3-2-1	11	11.47	0.60	H_0
3-3	9	8.23	0.49	H_0
4-1-1	9	20.34	0.98	H_1
4-2	8	5.18	0.26	H_0
5-1	5	3.78	0.42	H_0
6	0	-	-	-

Table A11: χ^2 test for IFD - 300000!

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1	15	13.12	0.48	H_0
2-1-1-1-1	14	16.53	0.78	H_0
2-2-1-1	13	11.84	0.54	H_0
2-2-2	12	15.87	0.85	H_0
3-1-1-1	12	12.08	0.64	H_0
3-2-1	11	16.83	0.92	H_0
3-3	9	7.50	0.52	H_0
4-1-1	9	7.23	0.49	H_0
4-2	8	8.92	0.74	H_0
5-1	5	4.62	0.67	H_0
6	0	-	-	-

Table A12: χ^2 test for IFD - $\sqrt{2}$

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1	15	17.98	0.74	H_0
2-1-1-1-1	14	22.23	0.93	H_0
2-2-1-1	13	17.20	0.81	H_0
2-2-2	12	16.2	0.82	H_0
3-1-1-1	12	8.79	0.28	H_0
3-2-1	11	15.73	0.85	H_0
3-3	9	13.12	0.84	H_0
4-1-1	9	6.68	0.33	H_0
4-2	8	12.95	0.89	H_0
5-1	5	10.02	0.93	H_0
6	-	-	-	-

Table A13: χ^2 test for IFD - e

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1	15	7.24	0.05	H_0
2-1-1-1-1	14	22.93	0.94	H_0
2-2-1-1	13	10.08	0.31	H_0
2-2-2	12	17.67	0.87	H_0
3-1-1-1	12	15.41	0.78	H_0
3-2-1	11	11.67	0.61	H_0
3-3	9	12.36	0.81	H_0
4-1-1	9	13.42	0.86	H_0
4-2	8	11.16	0.81	H_0
5-1	5	5.26	0.62	H_0
6	0	-	-	-

Table A14: χ^2 test for IFD - $M49$

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1	21	20.66	0.52	H_0
2-1-1-1-1-1	20	22.65	0.69	H_0
2-2-1-1-1	19	29.73	0.95	H_1
2-2-2-1	18	20.42	0.69	H_0
3-1-1-1-1	18	22.34	0.78	H_0
3-2-1-1	17	10.38	0.11	H_0
3-2-2	16	21.89	0.85	H_0
3-3-1	15	17.03	0.68	H_0
4-1-1-1	15	6.78	0.04	H_0
4-2-1	14	21.42	0.91	H_0
4-3	12	8.7	0.27	H_0
5-1-1	11	8.28	0.31	H_0
5-2	10	10.5	0.6	H_0
6-1	6	4.91	0.45	H_0
7	0	0	-	-

Table A15: χ^2 test for IFD - π, (6 digits)

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1	15	11.32	0.27	H_0
2-1-1-1-1	14	15.88	0.68	H_0
2-2-1-1	13	19.86	0.90	H_0
2-2-2	12	15.25	0.77	H_0
3-1-1-1	12	18.85	0.91	H_0
3-2-1	11	11.47	0.60	H_0
3-3	9	8.23	0.49	H_0
4-1-1	9	20.34	0.98	H_1
4-2	8	5.18	0.26	H_0
5-1	5	3.78	0.42	H_0
6	0	0	-	-

Table A16: χ^2 test for IFD - π (7 digits)

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1	21	23.5	0.68	H_0
2-1-1-1-1-1	20	7.13	0.004	H_0
2-2-1-1-1	19	17.81	0.47	H_0
2-2-2-1	18	23.36	0.82	H_0
3-1-1-1-1	18	23.41	0.83	H_0
3-2-1-1	17	19.51	0.7	H_0
3-2-2	16	6.81	0.02	H_0
3-3-1	15	8.51	0.1	H_0
4-1-1-1	15	11.57	0.29	H_0
4-2-1	14	9.63	0.21	H_0
4-3	12	26.32	0.99	H_1
5-1-1	11	7.12	0.21	H_0
5-2	10	12.86	0.77	H_0
6-1	6	2.72	0.16	H_0
7	0	0	-	-

Table A17: χ^2 test for IFD - π (8 digits)

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1	28	42.12	0.96	H_1
2-1-1-1-1-1-1	27	20.68	0.2	H_0
2-2-1-1-1-1	26	21.48	0.28	H_0
2-2-2-1-1	25	18.93	0.2	H_0
2-2-2-2	24	17.93	0.19	H_0
3-1-1-1-1-1	25	18.5	0.18	H_0
3-2-1-1-1	24	40.85	0.98	H_1
3-2-2-1	23	34.32	0.94	H_0
3-3-1-1	22	33.25	0.94	H_0
3-3-2	21	24.49	0.73	H_0
4-1-1-1-1	22	18.56	0.33	H_0
4-2-1-1	21	24.67	0.74	H_0
4-2-2	20	29.91	0.93	H_0
4-3-1	19	20.57	0.64	H_0
4-4	16	33.47	0.994	H_1
5-1-1-1	18	20.51	0.7	H_0
5-2-1	17	18.93	0.67	H_0
5-3	15	22.97	0.92	H_0
6-1-1	13	9.32	0.59	H_0
6-2	12	12.5	0.59	H_0
7-1	7	4.45	0.27	H_0
8	0	0	-	-

Table A18: χ^2 test for IFD - π (9 digits)

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1-1	36	43.09	0.81	H_0
2-1-1-1-1-1-1-1	35	31.88	0.38	H_0
2-2-1-1-1-1-1	34	45.35	0.91	H_0
2-2-2-1-1-1	33	29.68	0.37	H_0
2-2-2-2-1	32	23.6	0.14	H_0
3-1-1-1-1-1-1	33	44.82	0.92	H_0
3-2-1-1-1-1	32	28.19	0.34	H_0
3-2-2-1-1	31	32.65	0.61	H_0
3-2-2-2	30	37.49	0.84	H_0
3-3-1-1-1	30	23.76	0.22	H_0
3-3-2-1	29	49.51	0.99	H_1
3-3-3	27	34.87	0.86	H_0
4-1-1-1-1-1	30	25.78	0.31	H_0
4-2-1-1-1	29	17.98	0.06	H_0
4-2-2-1	28	53.2	0.997	H_1
4-3-1-1	27	19.86	0.16	H_0
4-3-2	26	30.99	0.77	H_0
4-4-1	24	38.7	0.97	H_1
5-1-1-1-1	26	30.18	0.74	H_0
5-2-1-1	25	14.55	0.05	H_0
5-2-2	24	30.12	0.82	H_0
5-3-1	23	21.34	0.44	H_0
5-4	20	21.12	0.61	H_0
6-1-1-1	21	13.48	0.11	H_0
6-2-1	20	9.28	0.02	H_0
6-3	18	14.76	0.32	H_0
7-1-1	15	11.66	0.3	H_0
7-2	14	15.36	0.65	H_0
8-1	8	10.48	0.77	H_0
9	0	0	-	-

Table A19: χ^2 test for IFD - MS_A

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1-1	36	354.39	1	H_1
2-1-1-1-1-1-1-1	35	367.51	1	H_1
2-2-1-1-1-1-1	34	458.18	1	H_1
2-2-2-1-1-1	33	310.72	1	H_1
2-2-2-2-1	32	524.71	1	H_1
3-1-1-1-1-1-1	33	491.42	1	H_1
3-2-1-1-1-1	32	358.36	1	H_1
3-2-2-1-1	31	367.18	1	H_1
3-2-2-2	30	256.49	1	H_1
3-3-1-1-1	30	399.2	1	H_1
3-3-2-1	29	369.68	1	H_1
3-3-3	27	136.22	1	H_1
4-1-1-1-1-1	30	619.35	1	H_1
4-2-1-1-1	29	433.64	1	H_1
4-2-2-1	28	456.2	1	H_1
4-3-1-1	27	373.38	1	H_1
4-3-2	26	176.54	1	H_1
4-4-1	24	295.41	1	H_1
5-1-1-1-1	26	296.49	1	H_1
5-2-1-1	25	269.75	1	H_1
5-2-2	24	126.19	1	H_1
5-3-1	23	649.51	1	H_1
5-4	20	136.74	1	H_1
6-1-1-1	21	149.32	1	H_1
6-2-1	20	336.75	1	H_1
6-3	18	182.06	1	H_1
7-1-1	15	193.83	1	H_1
7-2	14	108.32	1	H_1
8-1	8	86.41	1	H_1
9		0	-	-

Table A20: χ^2 test for IFD - MS_B

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1	28	24.42	0.34	H_0
2-1-1-1-1-1-1	27	22.67	0.3	H_0
2-2-1-1-1-1	26	31.21	0.78	H_0
2-2-2-1-1	25	30.59	0.80	H_0
2-2-2-2	24	42.88	0.99	H_1
3-1-1-1-1-1	25	19.42	0.22	H_0
3-2-1-1-1	24	36.12	0.95	H_1
3-2-2-1	23	35.56	0.95	H_1
3-3-1-1	22	20.27	0.43	H_0
3-3-2	21	19.69	0.46	H_0
4-1-1-1-1	22	19.64	0.40	H_0
4-2-1-1	21	35.37	0.97	H_1
4-2-2	20	19.96	0.54	H_0
4-3-1	19	21.78	0.71	H_0
4-4	16	10.87	0.18	H_0
5-1-1-1	18	14.35	0.29	H_0
5-2-1	17	39.86	1.00	H_1
5-3	15	20.46	0.85	H_0
6-1-1	13	16.92	0.61	H_0
6-2	12	10.97	0.47	H_0
7-1	7	10	0.81	H_0
8	-	-	-	n/a

Table A21: χ^2 test for IFD - MS_C

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1-1-1	45	27.27	0.02	H_0
2-1-1-1-1-1-1-1-1	44	38.25	0.28	H_0
2-2-1-1-1-1-1-1	43	56.05	0.91	H_0
2-2-2-1-1-1-1	42	33.94	0.19	H_0
2-2-2-2-1-1	41	39.18	0.45	H_0
2-2-2-2-2	40	66.62	1.00	H_1
3-1-1-1-1-1-1-1	42	51.07	0.84	H_0
3-2-1-1-1-1-1	41	33.35	0.2	H_0
3-2-2-1-1-1	40	45.99	0.76	H_0
3-2-2-2-1	39	30.43	0.17	H_0
3-3-1-1-1-1	39	38.21	0.5	H_0
3-3-2-1-1	38	37.04	0.49	H_0
3-3-2-2	37	48.25	0.9	H_0
3-3-3-1	36	41.85	0.77	H_0
4-1-1-1-1-1-1	39	44.75	0.76	H_0
4-2-1-1-1-1	38	59.18	0.99	H_1
4-2-2-1-1	37	54.1	0.97	H_1
4-2-2-2	36	42.76	0.8	H_0
4-3-1-1-1	36	28.99	0.21	H_0
4-3-2-1	35	36.34	0.59	H_0
4-3-3	33	43.4	0.89	H_0
4-4-1-1	33	22.19	0.08	H_0
4-4-2	32	34.3	0.64	H_0
5-1-1-1-1-1	35	37.92	0.66	H_0
5-2-1-1-1	34	26.67	0.19	H_0
5-2-2-1	33	41.32	0.85	H_0
5-3-1-1	32	23.15	0.13	H_0
5-3-2	31	37.48	0.8	H_0
5-4-1	29	41.57	0.94	H_0
5-5	25	23.38	0.45	H_0
6-1-1-1-1	30	17.26	0.03	H_0
6-2-1-1	29	48.24	0.99	H_1
6-2-2	28	25.41	0.39	H_0
6-3-1	27	30.53	0.71	H_0
6-4	24	27.05	0.7	H_0
7-1-1-1	24	28.81	0.77	H_0
7-2-1	23	21.32	0.44	H_0
7-3	21	17.75	0.34	H_0
8-1-1	17	14.35	0.36	H_0
8-2	16	22.21	0.86	H_0
9-1	9	9.35	0.59	H_0
10	0	0	0	-

Table A22: χ^2 test for IFD - Python

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1-1	36	23.94	0.06	H_0
2-1-1-1-1-1-1-1	35	35.56	0.56	H_0
2-2-1-1-1-1-1	34	38.71	0.74	H_0
2-2-2-1-1-1	33	46.59	0.94	H_0
2-2-2-2-1	32	31.13	0.49	H_0
3-1-1-1-1-1-1	33	40.29	0.82	H_0
3-2-1-1-1-1	32	32.94	0.58	H_0
3-2-2-1-1	31	41.74	0.91	H_0
3-2-2-2	30	29.46	0.51	H_0
3-3-1-1-1	30	39.96	0.89	H_0
3-3-2-1	29	34.35	0.77	H_0
3-3-3	27	21.95	0.26	H_0
4-1-1-1-1-1	30	52.45	0.993	H_1
4-2-1-1-1	29	38.91	0.9	H_0
4-2-2-1	28	29.67	0.62	H_0
4-3-1-1	27	28.03	0.59	H_0
4-3-2	26	17.34	0.1	H_0
4-4-1	24	13.42	0.04	H_0
5-1-1-1-1	26	33.2	0.84	H_0
5-2-1-1	25	31.78	0.84	H_0
5-2-2	24	17.21	0.16	H_0
5-3-1	23	19.53	0.33	H_0
5-4	20	25.47	0.82	H_0
6-1-1-1	21	16.18	0.24	H_0
6-2-1	20	13.24	0.13	H_0
6-3	18	20.96	0.82	H_0
7-1-1	15	23.71	0.93	H_0
7-2	14	24.71	0.96	H_1
8-1	8	17.24	0.97	H_1
9	0	0	-	-

Table A23: χ^2 test for IFD - MATLAB

Partition	DF	χ^2	$P(\chi^2 < CV)$	Hypothesis
1-1-1-1-1-1-1-1-1-1	45	34.03	0.12	H_0
2-1-1-1-1-1-1-1-1	44	43.71	0.52	H_0
2-2-1-1-1-1-1-1	43	39.61	0.38	H_0
2-2-2-1-1-1-1	42	62.75	0.98	H_1
2-2-2-2-1-1	41	54.51	0.92	H_0
2-2-2-2-2	40	36.15	0.36	H_0
3-1-1-1-1-1-1-1	42	23.92	0.01	H_0
3-2-1-1-1-1-1	41	45.48	0.71	H_0
3-2-2-1-1-1	40	51.91	0.9	H_0
3-2-2-2-1	39	47.97	0.85	H_0
3-3-1-1-1-1	39	42.34	0.67	H_0
3-3-2-1-1	38	44.86	0.79	H_0
3-3-2-2	37	41.56	0.72	H_0
3-3-3-1	36	57.81	0.99	H_1
4-1-1-1-1-1-1	39	49.21	0.87	H_0
4-2-1-1-1-1	38	47.11	0.85	H_0
4-2-2-1-1	37	36.49	0.51	H_0
4-2-2-2	36	52.12	0.96	H_1
4-3-1-1-1	36	52.61	0.96	H_1
4-3-2-1	35	48.39	0.93	H_0
4-3-3	33	31.87	0.48	H_0
4-4-1-1	33	33.19	0.54	H_0
4-4-2	32	27.02	0.28	H_0
5-1-1-1-1-1	35	35.95	0.57	H_0
5-2-1-1-1	34	42.65	0.85	H_0
5-2-2-1	33	44.95	0.92	H_0
5-3-1-1	32	28.31	0.35	H_0
5-3-2	31	56.98	1.00	H_1
5-4-1	29	31.49	0.66	H_0
5-5	25	17.1	0.12	H_0
6-1-1-1-1	30	39.35	0.88	H_0
6-2-1-1	29	21.01	0.14	H_0
6-2-2	28	13.43	0.009	H_0
6-3-1	27	22.6	0.29	H_0
6-4	24	27.63	0.72	H_0
7-1-1-1	24	14.77	0.07	H_0
7-2-1	23	31.13	0.88	H_0
7-3	21	20.99	0.54	H_0
8-1-1	17	16.51	0.51	H_0
8-2	16	7.07	0.03	H_0
9-1	9	6.86	0.35	H_0
10	-	0	0	-

Table A24: Frequency & Mean - $F_{5000000}$

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-1-1-1-1	157845	157992	7.497	7.5
2-1-1-1-1	473267	473975	6.991	7.0
2-2-1-1	237526	236988	6.505	6.5
2-2-2	11239	11285	6.014	6.0
3-1-1-1	105260	105328	6.009	6.0
3–2-1	45329	45141	5.502	5.5
3-3	965	940	4.491	4.5
4-1-1	11573	11285	4.542	4.5
4-2	1375	1411	4.01	4.0
5-1	540	564	2.511	2.5
6	11	10	0	0.0

Table A25: Frequency & Mean - $\sqrt{2}$

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-1-1-1-1	755316	756000	7.4973	7.5
2-1-1-1-1	2268728	2268000	6.9965	7.0
2-2-1-1	1133971	1134000	6.5031	6.5
2-2-2	54005	54000	6.0030	6.0
3-1-1-1	504235	504000	6.0056	6.0
3-2-1	215732	216000	5.4904	5.5
3-3	4508	4500	4.4361	4.5
4-1-1	53910	54000	4.4954	4.5
4-2	6796	6750	4.0511	4.0
5-1	2757	2700	2.5455	2.5
6	42	50	0	0.0

Table A26: Frequency & Mean - e

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-1-1-1-1	302232	302410	7.503	7.5
2-1-1-1-1	906770	907229	6.999	7.0
2-2-1-1	454604	453614	6.496	6.5
2-2-2	21671	21601	5.975	6.0
3-1-1-1	201437	201606	6.004	6.0
3-2-1	86199	86403	5.500	5.5
3-3	1777	1800	4.521	4.5
4-1-1	21678	21601	4.507	4.5
4-2	2621	2700	3.939	4.0
5-1	1054	1080	2.516	2.5
6	20	20	0	0.0

Table A27: Frequency & Mean - 300000!

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-1-1-1-1	217312	217402	7.508	7.5
2-1-1-1-1	651799	652207	7.000	7.0
2-2-1-1	327479	326103	6.508	6.5
2-2-2	15605	15529	5.956	6.0
3-1-1-1	144571	144935	5.997	6.0
3-2-1	61941	62115	5.492	5.5
3-3	1218	1294	4.555	4.5
4-1-1	15223	15529	4.467	4.5
4-2	1898	1941	3.981	4.0
5-1	785	776	2.434	2.5
6	15	14	0.000	0.0

Table A28: Frequency & Mean - M_{49}

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-1-1-1-1-1	1350810	1351038	10.493	10.5
2-1-1-1-1-1	7088692	7092949	9.999	10.0
2-2-1-1-1	7093093	7092949	9.502	9.5
2-2-2-1	1183903	1182158	8.998	9.0
3-1-1-1-1	2362096	2364316	8.999	9.0
3-2-1-1	2368793	2364316	8.499	8.5
3-2-2	168837	168880	8.004	8.0
3-3-1	113159	112586	7.517	7.5
4-1-1-1	393483	394053	7.498	7.5
4-2-1	169285	168880	7.014	7.0
4-3	7158	7037	6.055	6.0
5-1-1	33627	33776	5.509	5.5
5-2	4229	4222	5.106	5.0
6-1	1425	1407	3.060	3.0
7	20	22	0.000	0.0

Table A29: Frequency & Mean - MS_B

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-1-1-1-1-1-1	1826309	1826435	14.0021	14.0
2-1-1-1-1-1-1	17047001	17046724	13.5000	13.5
2-2-1-1-1-1	31957432	31962607	12.9999	13.0
2-2-2-1-1	12786336	12785043	12.4984	12.5
2-2-2-2	532923	532710	11.997	12 .0
3-1-1-1-1-1	8530313	8523362	12.5002	12.5
3-2-1-1-1	17043916	17046724	12.0016	12.0
3-2-2-1	4263358	4261681	11.4998	11.5
3-3-1-1	1418065	1420560	11.0015	11.0
3-3-2	203167	202937	10.5006	10.5
4-1-1-1-1	2130759	2130840	10.9977	11.0
4-2-1-1	2131307	2130840	10.4998	10.5
4-2-2	152403	152203	10.0035	10.0
4-3-1	202730	202937	9.5100	9.5
4-4	3255	3171	7.9239	8.0
5-1-1-1	283386	284112	8.9945	9.0
5-2-1	121895	121762	8.4935	8.5
5-3	5135	5073	7.4869	7.5
6-1-1	20254	20294	6.4863	6.5
6-2	2561	2537	5.9879	6.0
7-1	776	725	3.4026	3.5
8	7	10	0.0000	0.0

Table A30: Frequency & Mean - π

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-$\cdots$-1	3631181	3628800	17.9978	18.0
2-1-$\cdots$-1	65342549	65318400	17.4988	17.5
2-2-1-1-1-1-1	228597082	228614399	17.0002	17.0
2-2-2-1-1-1	190481668	190511999	16.4999	16.5
2-2-2-2-1	28572040	28576800	15.9989	16.0
3-1-1-1-1-1-1	50796060	50803200	16.4999	16.5
3-2-1-1-1-1	190520840	190511999	16.0001	16.0
3-2-2-1-1	114318703	114307199	15.4997	15.5
3-2-2-2	6347093	6350400	14.9956	15.0
3-3-1-1-1	25403542	25401600	14.9998	15.0
3-3-2-1	12703937	12700800	14.4986	14.5
3-3-3	201998	201600	13.5097	13.5
4-1-1-1-1-1	19053904	19051200	15.0011	15.0
4-2-1-1-1	38108708	38102400	14.5003	14.5
4-2-2-1	9526839	9525600	14.0002	14.0
4-3-1-1	6349876	6350400	13.5005	13.5
4-3-2	908296	907200	13.0093	13.0
4-4-1	226376	226800	12.0227	12.0
5-1-1-1-1	3810585	3810240	13.0015	13.0
5-2-1-1	3810333	3810240	12.4995	12.5
5-2-2	272577	272160	11.9883	12.0
5-3-1	362606	362880	11.5088	11.5
5-4	11297	11340	10.0264	10.0
6-1-1-1	423197	423360	10.4980	10.5
6-2-1	181028	181440	10.0091	10.0
6-3	7604	7560	9.0178	9.0
7-1-1	26049	25920	7.5101	7.5
7-2	3199	3240	7.0672	7.0
8-1	816	810	4.0723	4.0
9	9	10	0.0000	0.0

Table A31: Frequency & Mean - MS_A

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-$\cdots$-1	3624242	3628800	18.0023	18.0
2-1-$\cdots$-1	65323120	65318400	17.4971	17.5
2-2-1-1-1-1-1	228612721	228614399	17.0006	17.0
2-2-2-1-1-1	190445434	190511999	16.5000	16.5
2-2-2-2-1	28596300	28576800	16.0027	16.0
3-1-1-1-1-1-1	50829685	50803200	16.5015	16.5
3-2-1-1-1-1	190545488	190511999	16.0006	16.0
3-2-2-1-1	114327458	114307199	15.4981	15.5
3-2-2-2	6347260	6350400	14.9971	15.0
3-3-1-1-1	25361239	25401600	15.0022	15.0
3-3-2-1	12708686	12700800	14.4917	14.5
3-3-3	201322	201600	13.5094	13.5
4-1-1-1-1-1	19073225	19051200	15.0018	15.0
4-2-1-1-1	38089234	38102400	14.5032	14.5
4-2-2-1	9530887	9525600	13.9942	14.0
4-3-1-1	6349032	6350400	13.4931	13.5
4-3-2	908251	907200	13.0086	13.0
4-4-1	228248	226800	12.0479	12.0
5-1-1-1-1	3802180	3810240	12.9897	13.0
5-2-1-1	3806335	3810240	12.499	12.5
5-2-2	272081	272160	11.9993	12.0
5-3-1	363662	362880	11.6328	11.5
5-4	11292	11340	10.0620	10.0
6-1-1-1	421890	423360	10.4769	10.5
6-2-1	182172	181440	9.9802	10.0
6-3	7754	7560	9.2063	9.0
7-1-1	26377	25920	7.4747	7.5
7-2	3476	3240	6.8728	7.0
8-1	943	810	3.5567	4.0
9	0	10	0.0000	0.0

Table A32: Frequency & Mean - MS_C

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-. . .-1	3629234	3628800	18.0029	18.0
2-1-. . .-1	65358048	65318405	17.5007	17.5
2-2-1-1-1-1-1	228637484	228614416	17.0000	17.0
2-2-2-1-1-1	190489960	190512014	16.5003	16.5
2-2-2-2-1	28562402	28576802	16.0000	16.0
3-1-1-1-1-1-1	50808374	50803204	16.5003	16.5
3-2-1-1-1-1	190505627	190512014	15.9998	16.0
3-2-2-1-1	114274857	114307208	15.5008	15.5
3-2-2-2	6355236	6350400	14.9969	15.0
3-3-1-1-1	25399196	25401602	14.9997	15.0
3-3-2-1	12695837	12700801	14.5008	14.5
3-3-3	201381	201600	13.5008	13.5
4-1-1-1-1-1	19058919	19051201	14.9988	15.0
4-2-1-1-1	38100158	38102403	14.5008	14.5
4-2-2-1	9525188	9525601	14.0022	14.0
4-3-1-1	6349315	6350400	13.4973	13.5
4-3-2	908382	907200	12.9963	13.0
4-4-1	226956	226800	11.9868	12.0
5-1-1-1-1	3813847	3810240	13.0008	13.0
5-2-1-1	3812086	3810240	12.5034	12.5
5-2-2	271819	272160	12.0034	12.0
5-3-1	361635	362880	11.5001	11.5
5-4	11443	11340	9.9692	10.0
6-1-1-1	423357	423360	10.5103	10.5
6-2-1	181381	181440	10.0119	10.0
6-3	7660	7560	9.0604	9.0
7-1-1	26027	25920	7.516	7.5
7-2	3403	3240	6.9668	7.0
8-1	843	810	4.0107	4.0
9	17	10	0.0000	0.0

Table A33: Frequency & Mean Python

	Frequency		Mean	
Partition	Actual	Exp.	Actual	Exp.
1-1-. . .-1	3623408	3628784	18.0013	18.0
2-1-. . .-1	65291573	65318109	17.5004	17.5
2-2-1-1-1-1-1	228597060	228613383	17.0001	17.0
2-2-2-1-1-1	190505257	190511153	16.4999	16.5
2-2-2-2-1	28573940	28576673	16.0010	16.0
3-1-1-1-1-1-1	50817196	50802974	16.4994	16.5
3-2-1-1-1-1	190521632	190511153	16.0002	16.0
3-2-2-1-1	114323924	114306692	15.5000	15.5
3-2-2-2	6348302	6350372	14.9997	15.0
3-3-1-1-1	25409279	25401487	14.9994	15.0
3-3-2-1	12705258	12700744	14.4999	14.5
3-3-3	201573	201599	13.486	13.5
4-1-1-1-1-1	19045989	19051115	15.0005	15.0
4-2-1-1-1	38111257	38102231	14.4982	14.5
4-2-2-1	9529693	9525558	14.0000	14.0
4-3-1-1	6351445	6350372	13.4991	13.5
4-3-2	907074	907196	13.0004	13.0
4-4-1	227118	226799	12.0133	12.0
5-1-1-1-1	3807904	3810223	12.9994	13.0
5-2-1-1	3808146	3810223	12.499	12.5
5-2-2	272548	272159	12.0192	12.0
5-3-1	363815	362878	11.5037	11.5
5-4	11238	11340	10.0500	10.0
6-1-1-1	421816	423358	10.5092	10.5
6-2-1	181475	181439	9.9999	10.0
6-3	7614	7560	8.9980	9.0
7-1-1	25871	25920	7.5368	7.5
7-2	3261	3240	6.9197	7.0
8-1	869	810	4.0299	4.0
9	17	10	0.000	0.0

Table A34: Frequency & Mean - MATLAB

	Frequency		Mean	
Partition	Actual	Expected	Actual	Excepted
1-1-1-1-1-1-1-1-1	3626925	3628800	18.0013	18.0
2-1-1-1-1-1-1-1	65318282	65318399	17.5004	17.5
2-2-1-1-1-1-1	228635599	228614398	17.0002	17.0
2-2-2-1-1-1	190560331	190511998	16.5002	16.5
2-2-2-2-1	28576352	28576800	15.9996	16.0
3-1-1-1-1-1-1	50792445	50803200	16.4993	16.5
3-2-1-1-1-1	190499720	190511998	15.9999	16.0
3-2-2-1-1	114296958	114307199	15.5002	15.5
3-2-2-2	6349861	6350400	14.9969	15.0
3-3-1-1-1	25400616	25401600	15.0024	15.0
3-3-2-1	12700452	12700800	14.5013	14.5
3-3-3	201260	201600	13.5214	13.5
4-1-1-1-1-1	19039559	19051200	15.0002	15.0
4-2-1-1-1	38093464	38102400	14.5007	14.5
4-2-2-1	9524275	9525600	13.9968	14.0
4-3-1-1	6345456	6350400	13.5028	13.5
4-3-2	905797	907200	12.9915	13.0
4-4-1	227286	226800	11.9957	12.0
5-1-1-1-1	3807643	3810240	12.996	13.0
5-2-1-1	3808746	3810240	12.5007	12.5
5-2-2	272936	272160	12.0092	12.0
5-3-1	363065	362880	11.4901	11.5
5-4	11324	11340	10.0155	10.0
6-1-1-1	422990	423360	10.5054	10.5
6-2-1	181276	181440	10.0063	10.0
6-3	7430	7560	8.9828	9.0
7-1-1	25874	25920	7.4908	7.5
7-2	3247	3240	6.9587	7.0
8-1	817	810	4.0392	4.0
9	6	10	0.0000	0.0

A2 Extraction application

This application is written in Visual Basic for Application. It extracts d consecutive digits from a dataset. The variable `numDigits` stipulates the number of digits. The data may span over multiple records and is controlled by the variable `singleRecord`. In some cases, the input record is preceded and ended by a quote and it is necessary to stipulate `skip_quote` = `True`. The application takes care of numbers which span across the consecutive records. The output is written directly to a XLS file and are controlled by the variables `colMax`, `rowNo`, `colNo`, `HeaderCol`.

The application uses product of prime numbers to identify the partitions. To run the application, select $d = 2, 3, \ldots, 10$ where d is the number of consecutive digits to be selected from the dataset (See Section 9.1). Replace the variables `cFactord`, `cPatternd` which are commented out and rename them as `cFactor` and `cPattern` respectively. For a given value of d, `cPatternd` stores the possible partitions of d. Internally, the application identifies a partition by the product of prime factors by mapping the frequency count to a prime number. Thus the prime number assignments $p(f)$ for frequencies f are:

$$p(1) = 2\,,\ p(2) = 3\,,\ \ldots\,,\ p(3) = 5\,,\ p(4) = 7\,,\ p(10) = 29\,.$$

Thus, the partition [1,2,2,4] is represented as $p(1) \times p(2)^2 \times p(7) = 126$. This representation allows the application to detect the patterns in any permutations.

```
Sub Process_fileM()

'   This subroutine calculates the inversion distribution
    from an input file.
'   The file may have multiple records in which case there
    are carry over digits from the previous record.
'   The gap option specifies the number of digits to skip (0
     meaning the next digit)
'   The quote option is used to ignore the first and last
    digit of the input record.

Dim sTime, eTime, prime, cFactor, pattern As Variant
Dim iCount, dCount, dSize, duration As Long
Dim textline, invString, dString As String
Dim digits(10), numDigits As Integer
Dim length As Double
Dim inversionCount As Integer
Dim invCount(100) As Double
Dim inversion(42, 45) As Double
prime = Array(0, 2, 3, 5, 7, 11, 13, 17, 19, 23, 29)
'cFactor2 = Array(4, 3)
'cFactor3 = Array(8, 6, 5)
'cFactor4 = Array(16, 12, 9, 10, 7) '5
'cFactor5 = Array(32, 24, 18, 20, 15, 14, 11)   '7
```

```
'cFactor6 = Array(64, 48, 36, 27, 40, 30, 25, 28, 21, 22,
   13)  '11
'cFactor7 = Array(128, 96, 72, 54, 80, 60, 45, 50, 56, 42,
   35, 44, 33, 26, 17)  '15
'cFactor8 = Array(256, 192, 144, 108, 81, 160, 120, 90,
   100, 75, 112, 84, 63, 70, 49, 88, 66, 55, 52, 39, 34,
   19)  '22
'cFactor9 = Array(512, 384, 288, 216, 162, 320, 240, 180,
   135, 200, 150, 125, 224, 168, 126, 140, 105, 98, 176,
   132, 99, 110, 77, 104, 78, 65, 68, 51, 38, 23) '30
'cFactor10 = Array(1024, 768, 576, 432, 324, 243, 640, 480,
    360, 270, 400, 300, 225, 250, 448, 336, 252, 189, 280,
   210, 175, 196, 147, 352, 264, 198, 220, 165, 154, 121,
   208, 156, 117, 130, 91, 136, 102, 85, 76, 57, 46, 29)
   '42
'pattern2 = Array(”’1−1”, ‘‘’2−0”, ‘‘”)
'pattern3 = Array(”’1−1−1”, ‘‘’2−1”, ‘‘’3−0”, ‘‘”)
'pattern4 = Array(”’1−1−1−1”, ‘‘’2−1−1”, ‘‘’2−2”, ‘‘’3−1”,
   ‘‘’4”)
'pattern5 = Array(”’1−1−1−1−1”, ‘‘’2−1−1−1”, ‘‘’2−2−1”,
   ‘‘’3−1−1”, ‘‘’3−2”, ‘‘’4−1”, ‘‘’5”)
'pattern6 = Array(”’1−1−1−1−1−1”, ‘‘’2−1−1−1−1”,
   ‘‘’2−2−1−1”, ‘‘’2−2−2”, ‘‘’3−1−1−1”, ‘‘’3−2−1”, ‘‘’3−3”,
    ‘‘’4−1−1”, ‘‘’4−2”, ‘‘’5−1”, ‘‘’6”) '11
'pattern7 = Array(”’1−1−1−1−1−1−1”, ‘‘’2−1−1−1−1−1”,
   ‘‘’2−2−1−1−1”, ‘‘’2−2−2−1”, ‘‘’3−1−1−1−1”, ‘‘’3−2−1−1”,
   ‘‘’3−2−2”, ‘‘’3−3−1”, ‘‘’4−1−1−1”, ‘‘’4−2−1”, ‘‘’4−3”,
   ‘‘’5−1−1”, ‘‘’5−2”, ‘‘’6−1”, ‘‘’7”)
'pattern8 = Array(”’1−1−1−1−1−1−1−1”, ‘‘’2−1−1−1−1−1−1”,
   ‘‘’2−2−1−1−1−1”, ‘‘’2−2−2−1−1”, ‘‘’2−2−2−2”,
   ‘‘’3−1−1−1−1−1”, ‘‘’3−2−1−1−1”, ‘‘’3−2−2−1”,
   ‘‘’3−3−1−1”, ‘‘’3−3−2”, ‘‘’4−1−1−1−1”, ‘‘’4−2−1−1”,
   ‘‘’4−2−2”, ‘‘’4−3−1”, ‘‘’4−4”, ‘‘’5−1−1−1”, ‘‘’5−2−1”,
   ‘‘’5−3”, ‘‘’6−1−1”, ‘‘’6−2”, ‘‘’7−1”, ‘‘8”)
'pattern9 = Array(”’1−1−1−1−1−1−1−1−1”,
   ‘‘’2−1−1−1−1−1−1−1”, ‘‘’2−2−1−1−1−1−1”, ‘‘’2−2−2−1−1−1”,
    ‘‘’2−2−2−2−1”, ‘‘’3−1−1−1−1−1−1”, ‘‘’3−2−1−1−1−1”,
   ‘‘’3−2−2−1−1”, ‘‘’3−2−2−2”, ‘‘’3−3−1−1−1”, ‘‘’3−3−2−1”,
   ‘‘’3−3−3”, ‘‘’4−1−1−1−1−1”, ‘‘’4−2−1−1−1”, ‘‘’4−2−2−1”,
   ‘‘’4−3−1−1”, ‘‘’4−3−2”, ‘‘’4−4−1”, ‘‘’5−1−1−1−1”,
   ‘‘’5−2−1−1”, ‘‘’5−2−2”, ‘‘’5−3−1”, ‘‘’5−4”, ‘‘’6−1−1−1”,
    ‘‘’6−2−1”, ‘‘’6−3”, ‘‘’7−1−1”, ‘‘’7−2”, ‘‘’8−1”, ‘‘’9”)
'pattern10 = Array(”’1−1−1−1−1−1−1−1−1−1”,
   ‘‘’2−1−1−1−1−1−1−1−1”, ‘‘’2−2−1−1−1−1−1−1”,
   ‘‘’2−2−2−1−1−1−1”, ‘‘’2−2−2−2−1−1”, ‘‘’2−2−2−2−2”,
   ‘‘’3−1−1−1−1−1−1−1”, ‘‘’3−2−1−1−1−1−1”, ‘‘’3−2−2−1−1−1”,
    ‘‘’3−2−2−2−1”, ‘‘’3−3−1−1−1−1”, ‘‘’3−3−2−1−1”,
```

```
    ''3-3-2-2", ''3-3-3-1", ''4-1-1-1-1-1-1",
    ''4-2-1-1-1-1", ''4-2-2-1-1", ''4-2-2-2",
    ''4-3-1-1-1", ''4-3-2-1", ''4-3-3", ''4-4-1-1",
    ''4-4-2", ''5-1-1-1-1-1", ''5-2-1-1-1", ''5-2-2-1",
    ''5-3-1-1", ''5-3-2", ''5-4-1", ''5-5",
    ''6-1-1-1-1", ''6-2-1-1", ''6-2-2", ''6-3-1",
    ''6-4", ''7-1-1-1", ''7-2-1", ''7-3", ''8-1-1",
    ''8-2", ''9-1", ''10", ''")

Dim Idx(1024) As Integer
Dim endFlag As Boolean
    sTime = Now()
    iCount = 0
    numDigits = 9
    Gap = 1
    rData = ''"
    colMax = 36
    rowNo = 121
    colNo = 9
    HeaderCol = 1
    skip_quote = False
    numpatterns = 30
    singleRecord = False
    For i = 0 To numpatterns - 1
            factor = cFactor(i)
             Idx(factor) = i
    Next
    infile = ''Input file name"
    Open infile For Input As #1
    endFlag = False
    Do While Not EOF(1) And endFlag = False
     Line Input #1, pData
     If skip_quote = True Then
           pData = Mid(pData, 2, Len(pData) - 2)
     End If
     pLength = Len(pData)
     If singleRecord = True Then
          pData = pData & Left(pData, numDigits - 1)
     End If
     pData = rData & pData
     If Gap > 1 Then
          cycle = Int(Len(pData) / Gap)
     Else
          cycle = Len(pData) + 1 - numDigits
     End If
     rData = Right(pData, numDigits - 1)
     For i = 1 To cycle
```

```
            factor = 1
            For j = 0 To 10
                digits(j) = 0
            Next
            invString = Mid(pData, Gap * (i - 1) + 1,
               numDigits)
            For j = 1 To numDigits
                iData = Mid(invString, j, 1)
                indexPos = Int(iData)
                digits(indexPos) = digits(indexPos) + 1
            Next
            For j = 0 To 10
                If digits(j) > 0 Then
                    index = digits(j)
                    factor = factor * prime(index)
                 End If
             Next
             vCount = 0
             For j = 1 To numDigits - 1
                 lChar = Mid(invString, j, 1)
                 For k = j + 1 To numDigits
                     rChar = Mid(invString, k, 1)
                     If (lChar > rChar) Then
                         vCount = vCount + 1
                     End If
                 Next
            Next
            iPos = Idx(factor)
            inversion(iPos,vCount) = inversion(iPos,vCount) +
                1
            dCount = dCount + 1
            iCount = iCount + 1
            endFlag = True
            Next
        Loop
Close #1
For i = 0 To numpatterns - 1
    Cells(rowNo + i, HeaderCol) = pattern(i)
    For j = 0 To colMax
        If inversion(i, j) > 0 Then
            Cells(rowNo + i, colNo + j) = inversion(i,j)
            sCount = sCount + inversion(i,j)
        End If
    Next
Next
Close #1
End Sub
```

References

[1] G.E. Andrews, Euler's pentagonal number theorem, *Math. Mag.* **56** (1983), 279–284.

[2] M. Bóna, *A Walk Through Combinatorics*, World Scientific Publishing Co., Inc., River Edge, NJ, 2002.

[3] L. Carlitz, Enumeration of symmetric arrays, *Duke Math. J.* **33** (1966), 771–782.

[4] M. Conger and D. Viswanath, Normal approximations for descents and inversions of permutations of multisets, *J. Theoret. Probab.* **20** (2007), 309–325.

[5] T.H. Cormen, *Introduction To Algorithms*, NetLibrary, Inc. 2nd ed., MIT Press, Cambridge, MA, 2001.

[6] G. Strang, *Essays In Linear Algebra*, Wellesley-Cambridge Press, Wellesley, MA, 2012.

[7] P.E. Greenwood and M.S. Nikulin, *A Guide To Chi-Squared Testing*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1996.

[8] G.H. Hardy and S. Ramanujan, Asymptotic formulae in combinatory analysis, *Proc. London Math. Soc.* **17** (1918), 75–115.

[9] D.E. Knuth, *The Art Of Computer Programming. Vol. 3*, second edition, Addison-Wesley, Reading, MA, 1998.

[10] P.A. MacMahon, Two applications of general theorems in combinatory analysis: (1) to the theory of inversions of permutations; (2) to the ascertainment of the numbers of terms in the development of a determinant which has amongst its elements an arbitrary number of zeros, *Proc. London Math. Soc. (2)* **15** (1916), 314–321.

[11] P.A. MacMahon, The indices of permutations and the derivation therefrom of functions of a single variable associated with the permutations of any assemblage of objects, *Amer. J. Math.* **35** (1913), 281–322.

[12] P.A. MacMahon, The parity of $p(n)$, the number of partitions of n, when $n \leq 1000$, *J. London Math. Soc.* **1** (1926), 225–226.

[13] T. Muir, On a simple term of a determinant, *Proc. Royal Soc. Edinburgh* **21** (1898–9), 331–377.

[14] O. Rodrigues, Note sur les Inversions, ou derangements produits das les permutations, *Journal de Mathematiques* **4** (1839), 236–240.

[15] *The On-Line Encyclopedia of Integer Sequences (OEIS)*, `https://oeis.org/`, last retrieved on 19-03-2019.

[16] R.P. Stanley, *Enumerative Combinatorics, Vol. 1*, Cambridge University Press, Cambridge, MA, 1997.

[17] S. Washburn, The periods of the digits of the Fibonacci numbers, *Pi Mu Epsilon J.* **10** (1997), 540–546.

[18] Charles Zaiontz, http://www.real-statistics.com/tests-normality-and-symmetry/analysis-skewness-kurtosis/, last retrieved on 05-08-2019.