EXAMENSARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Parikh Matrices and Permutation Statistics

av

Christopher Yamba

2006 - No 8

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 10691 STOCKHOLM

Parikh Matrices and Permutation Statistics

Christopher Yamba

Examensarbete i matematik 20 poäng, fördjupningskurs

Handledare: Mats Oldin och Yishao Zhou

2006

Abstract

In this paper I introduce a permutation statistic, the *Parikh matrix statistic,* which counts the number of permutations with a given Parikh matrix and prove this is equidistributed with the descent statistic, when we consider permutations in S_n. First I will introduce a generalization of the classical Parikh vector and show a matrix completeness for Parikh matrices of words in S_n. When we consider S_n, this Parikh matrix completeness implies that the Parikh matrix statistic is equivalent to what can be defined as a Parikh vector statistic of order 2. In order to prove the equidistribution between the Parikh matrix statistic and the descent set statistic I make use of results from R.P. Stanley on descent set statistics. I define analogous inclusive and exclusive Parikh vector statistics. The desired result follows from the Principle of Inclusion- Exclusion.

Contents

Notations

$P(S)$	the power set of S
$P_M(S)$	the set of all multisets with the elements from S
$[n]$	the set $\{1,...,n\}$
$\left(\!\binom{n}{k}\!\right)$	the number of k-combinations of $[n]$ with repetions
$S_n(M)$	the set of all permutations of length n of elements from a multiset M
$D(\pi)$	the descent set of π
$\alpha(S)$	the inclusive descent set statistic
$\beta(S)$	the exclusive descent set statistic
Σ	an alphabet
Σ^*	the set of all words with letters from the alphabet Σ
$\lvert w\rvert_u$	the number of occurrences of the word u as a scattered subword in w
$a_{i,j}$	the word $a_{i,j} = a_i a_{i+1} \cdots a_j$ of consecutive letters from an ordered alphabet $\Sigma = \{a_1 < a_2 < \cdots < a_n\}$
M_{k+1}	the set of upper triangular $(k+1)\times(k+1)$-matrices $(m_{i,j})_{1\le i,j\le k+1}$
$\psi_{M_k} : \Sigma^* \to M_{k+1}$	the Parikh matrix mapping given an alphabet of k letters
$\psi_j : \Sigma^* \to \mathbb{N}^{k-j+1}$	the Parikh vector mapping of order j
$\vartheta_k(v)$	the inclusive Parikh vector statistic of order j
$\rho_j(v)$	the exclusive Parikh vector statistic of order j
$\vartheta_{M_k}((m_{i,j})_{1\le i,j\le k+1})$	the inclusive Parikh matrix statistic over an alphabet of k letters
$\rho_{M_k}((m_{i,j})_{1\le i,j\le k+1})$	the exclusive Parikh matrix statistic over an alphabet of k letters
$D_{\phi_B}(\pi)$	the descent vector of π
$\beta_{\phi_B}(S)$	the descent vector statistic
$i(\pi)$	the number of inversions in a permutation π

1 Introduction

The Parikh vector mapping is an important tool in the theory of formal languages, introduced 1966 by R.J. Parikh in [6]. Given an ordered alphabet with *k* letters $\Sigma_k = \{a_1 < \cdots < a_k\}$ and a word $w \in \Sigma^*$, where Σ^* denotes the set of all words with letters from the alphabet Σ_k, the Parikh vector mapping is a morphism $\psi : \Sigma^* \to \mathbb{N}^k$, defined by $\psi(w) = (|w|_{a_1}, ..., |w|_{a_k})$, where $|w|_{a_i}$ denotes the number of occurrences of the letter a_i in the word w and $\mathbb{N}$ denotes the set of nonnegative integers. One important result concerning the Parikh vector mapping, found in [6], is that the image by the Parikh vector mapping of a context-free language is always a semilinear set.

The Parikh vector contains *numerical* properties of a word expressed as a vector. For example, consider the ordered 3-letter alphabet $\Sigma_3 = \{a < b < c\}$ and the word $w = abccbac$. Then we get the Parikh vector

$\psi(w) = \psi(abccbac) = (|w|_a, |w|_b, |w|_c) = (|abccbac|_a, |abccbac|_b, |abccbac|_c) = (2, 2, 3) \in \mathbb{N}^3$.

The Parikh vector mapping is not injective since two words can have the same Parikh vector. Thus, much information of the words is lost in this mapping and therefore the natural continuation of this topic was the development of an extension of the Parikh vector mapping. This new tool, the *Parikh matrix mapping*, contains more information about a word than the Parikh vector mapping. The Parikh matrix mapping was introduced several decades later in [1], quite recently 2001. It contains information about the number of occurrences of certain subwords. Here subwords means scattered subwords defined in the following way:

Definition 1.1 A word u is a subword of a word w if there exist words $x_1, ..., x_n$ and $y_0, ..., y_n$, some possibly empty, such that $u = x_1 \cdots x_n$ and $w = y_0 x_1 y_1 \cdots x_n y_n$.

Let $|w|_u$ denote the number of occurrences of u as a scattered subword of w. For example, take the word $w = abccbac \in \Sigma^*$ and consider the subword $u = abc$. Then $|w|_u$ means the number of occurrences of $u = abc$ as a scattered subword of $w = abccbac$. We have following 4 occurrences $\begin{matrix} 1\ \underline{a}\underline{b}\underline{c}cbac \\ 2\ \underline{a}\underline{b}c\underline{c}bac \\ 3\ \underline{a}\underline{b}ccba\underline{c} \\ 4\ \underline{a}bcc\underline{b}a\underline{c} \end{matrix}$. Thus $|w|_u = 4$. When working with Parikh matrices we will repeatedly encounter subwords denoted $a_{i,j}$, where i, j, $i \le j$, denotes the indices of the letters position in the given alphabet. So in other words we are dealing with subwords consisting of consecutive letters from an alphabet.

For example we might be in the ordered alphabet $\Sigma = \{a_1 < \cdots < a_5\}$. The notation $a_{2,4}$ means that we are considering the word $a_{2,4} = a_2 a_3 a_4$. If we let $w = a_2 a_2 a_1 a_3 a_4$ then we have that $|w|_{a_{2,4}}$ denotes the number of occurrences of $a_{2,4} = a_2 a_3 a_4$ as a scattered subword in $w = a_2 a_2 a_1 a_3 a_4$. We have the following 2 occurrences $\begin{matrix} 1\ w = \underline{a_2} a_2 a_1 \underline{a_3}\,\underline{a_4} \\ 2\ w = a_2 \underline{a_2} a_1 \underline{a_3}\,\underline{a_4} \end{matrix}$. Thus $|w|_{a_{2,4}} = 2$.

The Parikh Matrix mapping is a morphism $\psi_{M_k} : \Sigma^* \to M_{k+1}$ where M_{k+1} is a collection of $(k+1)$ - dimensional upper triangular matrices with nonnegative integral entries and unit diagonal. The classical Parikh vector will appear as *the second diagonal* (the diagonal above the main diagonal) in the Parikh matrix. The other entries above the second diagonal contain information about the order of the letters (in the word in examined) and are of the form $|w|_{a_{i,j}}$, $i < j$. The main diagonal entries are 1's and 0's below it. Given the alphabet $\Sigma_k = \{a_1 < \cdots < a_k\}$, we have $\psi_{M_k} : \Sigma^* \to M_{k+1}$, defined by

$$\psi_{M_k}(w) = \begin{bmatrix} 1 & |w|_{a_{1,1}} & |w|_{a_{1,2}} & \dots & |w|_{a_{1,k}} \\ 0 & 1 & |w|_{a_{2,2}} & \dots & |w|_{a_{2,k}} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & |w|_{a_{k-1,k-1}} & |w|_{a_{k-1,k}} \\ \dots & \dots & \dots & 1 & |w|_{a_{k,k}} \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix}.$$

(I should point out that this definition is equivalent to Definition 2.8 on page 12).

In this thesis I will examine words in S_k, the set of all permutations of the elements of the set $[k] = \{1, \dots, k\}$. The Parikh matrices we deal with in this paper will have entries of the form

$|w|_{i\cdots j}$, $i \le j$. Consider the classical Parikh vector mapping of $\psi(w) = (|w|_1, |w|_2, \ldots, |w|_k)$, for $w \in S_k$.

Since the letters of words belonging to S_k occur exactly once, (given the alphabet $\sum_k = \{1 < \cdots < k\}$) we have that all the words of S_k have the same Parikh vector, namely $\psi(w) = (\underbrace{1,\ldots,1}_{k})$. So, given $\sum_5 = \{1 < 2 < 3 < 4 < 5\}$ what is the Parikh matrix of $w = 13245 \in S_5$?

First of all note that since we are working with S_5, the value of $|w|_{i\cdots j}$ is either 0 or 1. In general the Parikh matrices of words from S_k will have entries from $\mathbb{Z}_2$.

In our case we have $w = 13245 \in S_5$ and

$$\psi_{M_5}(w) = \psi(13245) = \begin{bmatrix} 1 & |w|_1 & |w|_{12} & |w|_{123} & |w|_{1234} & |w|_{12345} \\ 0 & 1 & |w|_2 & |w|_{23} & |w|_{234} & |w|_{2345} \\ 0 & 0 & 1 & |w|_3 & |w|_{34} & |w|_{345} \\ 0 & 0 & 0 & 1 & |w|_4 & |w|_{45} \\ 0 & 0 & 0 & 0 & 1 & |w|_5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

The Parikh matrix mapping is not injective. This can easily be seen. If we consider the case studied in this thesis, that is, $\psi_{M_k} : S_k \to M_{k+1}$, we have a surjective mapping from a set with $k!$ elements to a set with less than $k!$ elements, namely 2^{k-1} elements. This might not be obvious to the reader at this moment, but it will be seen later, when we show that the Parikh matrix of a word in S_k is completely determined by its third diagonal. The third diagonal can be seen as a binary string of length $k-1$ and we will show that it is the same as the Parikh vector of second order.

It can easily be realized if one considers for example the subword 234 can only exist if the word 23 and 34 exists. Generalize this idea and the result is clear. This is proved in the Parikh Matrix Completeness Theorem. So as a result we will focus on this third diagonal. It is very special in this thesis and will be referred to as the *Parikh vector of second order*.

Let's examine the entries of the above Parikh matrix a bit closer. The main (first) diagonal is, by the Parikh matrix Theorem (**Theorem 2.10**), $m_{i,i} = 1$, $\forall\ 1 \le \mathrm{i} \le \mathrm{k+1}$. So we may think of it as the unit vector $(\underbrace{1,...,1}_{k+1})$. The second diagonal $(|w|_1,...,|w|_k)$ is just the classical Parikh vector (**Corollary 2.11**).

Let $w = 13245 \in S_5$. Then in this case as we established earlier we have $\psi(w) = (|w|_1,...,|w|_5) = (\underbrace{1,...,1}_{5})$. Now it seems natural to continue this procedure, investigating the entries by going through the diagonals in increasing order. Let us continue.

The third diagonal of the Parikh matrix in this case is $(|w|_{12}, |w|_{23}, |w|_{34}, |w|_{45}) = (1,0,1,1)$.

The general case would be the *j+1* diagonal corresponds to $(|w|_{1 \cdot 2 \cdots j}, |w|_{23 \cdots (j+1)}, ..., |w|_{(k-j+1) \cdots k})$.

Now, why not call this a *Parikh vector of j:th order*, if we let the classical Parikh vector be the *Parikh vector of first order*. In the same way we have **Corollary 2.11**, we can establish a Corollary such that the third diagonal corresponds to the Parikh vector of second order. This follows directly from the Parikh matrix theorem (**Theorem 2.10**) in the same way we get **Corollary 2.11**. We will get a generalized Parikh vector of order i ,where $1 \le i \le k$ and our attention will be directed toward $i = 2$, the *Parikh vector of second order*.

Two words with the same (classical) Parikh vector will in many cases have Parikh matrices that are different.

I should also mention that Parikh mapping is a morphism from $(\Sigma^*, \bullet, \lambda)$ (where $\bullet$ denotes concatenation) to $(\mathbb{N}^k, +, (\underbrace{0,...,0}_{k}))$ and the Parikh matrix mapping is a morphism from $(\Sigma^*, \bullet, \lambda)$ to M_k consisting of the set of $(k+1)$-dimensional upper triangular matrices as defined above together with the operation multiplication of matrices and with as its unit the unit matrix of dimension $k+1$, $(M_{k+1}, \times, (1)_{k+1 \times k+1})$.

I have now provided some background of Parikh matrices. Now I will provide some background on permutation statistics.

What is permutation statistics?

Permutation statistics is a branch of enumerative combinatorics dealing with enumeration of permutations with respect to properties such as the *descent set* of a permutation.

Definition 1.2

The descent set of a permutation $\pi = \pi_1 \cdots \pi_k \in S_k$ is denoted $D(\pi)$ and is defined by $D(\pi) = \{i : \pi_i > \pi_{i+1}\} \subseteq [k-1]$.

Sometimes it is useful to think of it as a mapping $D : S_k \to P([k-1])$, where $P([k-1])$ denotes the powerset of $[k-1]$, that is the set of all subsets of $[k-1] = \{1,...,k-1\}$. The permutation statistic that counts the number of permutations with a given descent set is called the *descent (set) statistic* and has been extensively studied.

The descent set statistic is denoted by $\beta(S) = \#\{\pi \in S_k : D(\pi) = S\}$, $S \subseteq [k-1]$. In this thesis we will consider descent set statistic of words that belong to S_k. Consider the powerset of $[k-1]$, $P([k-1])$. It will be very useful to represent each subset $S \subseteq [k-1]$ as a binary string corresponding to *S*. For example if we consider $S = \{1,3,4\} \subseteq [4] = \{1,2,3,4\}$, may represent this by letting position 1, 3, and 4 in the binary string of length 4 correspond to a 1 and the rest correspond to 0. Formally this is a mapping ϕ_B, where *B* stands for binary, defined as follows:

Definition 1.3 Let $P([k-1])$ denote the powerset of $[k-1]$. Define $\phi_B : P([k-1]) \to \mathbb{Z}_2^{k-1}$, by $\phi_B(S) = v$, such that $v = v_1 \cdots v_k$ and $v_i = \begin{cases} v_i = 0 \text{ if } i \in S \\ v_i = 1 \text{ if } i \notin S \end{cases}$.

This is easily seen to be a bijection.

Having this way of representing sets we also have a way of representing descent sets as binary strings. I sometimes denote descent sets of permutations as $D_{\phi_B}(\pi)$, where the set is mapped to its binary string. (or alternatively $\phi_B(D(\pi))$). For example we have $D_{\phi_B}(13245) = 1011$. (Note that $\psi_2(13245) = 1011$).

By this point you might have begun to suspect what I'm aiming for. I want to investigate possible connections between Parikh vectors of second order of words in S_k (they are binary strings of length $(k-1)$) and descent sets of words in S_k. I will do this in this by introducing a new permutation statistic by thinking of the Parikh matrix mapping as the property in question. To be more specific, I am interested in counting the number of permutations mapping to a given Parikh matrix , and as a consequence of the Parikh Matrix Completeness

(**Theorem 3.5**). This turns out to be the same as determining the number of permutations mapping to a certain Parikh vector of second order, since every Parikh matrix of a word in S_k is completely determined by its third diagonal, which is the diagonal corresponding to the Parikh vector of second order. The number of permutations mapping to a given Parikh vector *v* of second order is denoted by $\rho_2(v)$. The number of permutations mapping to a given *descent vector* v is denoted by β_{ϕ_B} (see the section of **Notations** below).

Later I will prove that $\rho_2(v) = \beta_{\phi_B}(v)$ (see **Theorem 6.12**). In other words that the Parikh statistic of order 2 is equidistributed with the descent statistic. This may be considered the main result of this thesis and possibly a new contribution connecting the theory of formal languages with the permutation statistics. This opens up a vast number of interesting enumerative problems to explore further.

2 Parikh Matrices Basic Definitions and Theorems

We have been informally introduced to what Parikh matrices represents, namely it contains information of the number of certain subwords (of consecutive letters) of a word, given an ordered alphabet. In this section I will more formally go through the theory as found in [1], before going further and introducing my own results concerning Parikh statistics and descent statistics. First of all we need some definitions.

Definition 2.1 [1] An ordered alphabet is a totally ordered finite set denoted by $\Sigma = \{a_1 < ... < a_k\}$.

In this thesis, working in S_k, we will mostly consider the ordered alphabet where the letters are a finite set of integers $\Sigma = \{1 < ... < k\}$.

Definition 2.2 A *finite multiset* is a finite set where repetitions of elements are taken into consideration: Let $N = \{1, 2, ..., k\}$ be a set. A finite multiset with respect to *N* of cardinality *n* is denoted $M = \{1^{b_1}, ..., k^{b_k}\}$, $\#M = \sum_{i=1}^{k} b_i = n$, where b_i denotes the number of repetitions of element i.

We will consider words *w* of length *n* that are permutations of elements of multisets M. This is denoted by $w \in S_n(M)$.

Example 2.3 Let $N = \{1, ..., 5\}$ and consider the multiset $M = \{1^2, 2^1, 3^3, 4^0, 5^1\}$. If we consider words $w \in S_n(M)$, we are considering words that are permutations of length $\#M = 2+1+3+0+1 = 7$. The number of such words is equal to the multinomial coefficient $\binom{7}{2,1,3,0,1} = \frac{7!}{2!1!3!0!1!} = 280$. One such word is $w = 3125331 \in S_5(M)$.

Definition 2.4 [1] Let $\Sigma = \{a_1 < \cdots < a_k\}$ be an ordered alphabet. The *Parikh vector mapping* is a morhpism $\psi_1 : \Sigma^* \to \mathbb{N}^k$ defined by $\psi_1(w) = (|w|_{a_1}, ..., |w|_{a_k})$, where $|w|_{a_i}$ denotes the number of occurences of the letter a_i in w.

Given $w_1, w_2 \in \Sigma^*$, we have $\psi_1(w_1 * w_2) = \psi_1(w_1) + \psi(w_2)$, where $*$ denotes concatenation and Σ^* denotes the set of all words with letters from Σ.

Example 2.5 Consider the ordered alphabet $\Sigma_7 = \{1 < 2 < \cdots < 7\}$ and the words $w_1 = 1234 \in \Sigma^*$ and $w_2 = 5367 \in \Sigma^*$. Then we have $w_1 * w_2 = 12345367$ and

$$\psi(w_1 * w_2) = (|w_1 * w_2|_1, |w_1 * w_2|_2, |w_1 * w_2|_3, |w_1 * w_2|_4, |w_1 * w_2|_5, |w_1 * w_2|_6, |w_1 * w_2|_7) = (1,1,2,1,1,1,1)$$

$$\psi(w_1) = (|w_1|_1 |w_1|_2 |w_1|_3 |w_1|_4 |w_1|_5 |w_1|_6 |w_1|_7) = (1,1,1,1,0,0,0)$$

$$\psi(w_2) = (|w_2|_1 |w_2|_2 |w_2|_3 |w_2|_4 |w_2|_5 |w_2|_6 |w_2|_7) = (0,0,1,0,1,1,1)$$

So we have $\psi(w_1 * w_2) = (1,1,2,1,1,1,1) = (1,1,1,1,0,0,0) + (0,0,1,0,1,1,1) = \psi(w_1) + \psi(w_2)$.

In this thesis, working in S_k, we have that the Parikh vector mapping $\psi_1(w) = (|w|_{a_1}, ..., |w|_{a_k})$ is $\psi_1(w) = (|w|_1, ..., |w|_k) = (\underbrace{1,...,1}_{k})$, since each letter of the *k*-letter alphabet in a permutation in S_k of length *k* occurs exactly once.

The following definition concerns how the subwords are notated.

Definition 2.6 [1] Consider the alphabet $\Sigma = \{a_1 < ... < a_k\}$**.** Then the $a_{i,j}$ denotes the word $a_{i,j} = a_i a_{i+1} \cdots a_j$ of consecutive letters from the alphabet. In particular $a_{i,i} = a_i$.

As we will work with words belonging to S_k, the above subwords will have the form $i(i+1)\cdots j$, where i and j are letters from the ordered alphabet $\Sigma_k = \{1 < \cdots < k\}$.

Definition 2.7 [1] The *number of occurrences* of $a_{i,j}$ as a scattered subword of *w* is denoted by $|w|_{a_{i,j}} \in \mathbb{N}$.

Again, in our case S_k, we will have $|w|_{i,j} \in \mathbb{Z}_2$, and since each subword in a permutation can occur at most once.

We have informally defined the Parikh matrix mapping but we will do it more formally here.

Definition 2.8 [1] Let $\Sigma = \{a_1 < ... < a_k\}$ be the an ordered alphabet.
The *Parikh matrix mapping* is a morphism $\psi_{M_k} : \Sigma^* \to M_{k+1}$, where M_{k+1} denotes the set of upper triangular $(k+1)\times(k+1)$ matrices $(m_{i,j})_{1\le i,j\le k+1}$, such that $m_{i,j} \in \mathbb{N}$.
ψ_{M_k} is defined on Σ by $\psi_{M_k}(a_l) = (m_{i,j})_{1\le i,j\le(k+1)}$, where we set $m_{i,i} = 1$, $m_{l,l+1} = 1$ for $1 \le i \le k+1$ and for all other entries in $\psi_{M_k}(a_l)$ we set $m_{i,j} = 0$.

M_{k+1} has a unit which is the identity matrix I_k. For $w \in \Sigma^*$ with $w = a_{i_1} \cdots a_{i_s}$, $a_{i_j} \in \Sigma$, define $\psi_{M_k}(w) = \psi_{M_k}(a_{i_1}) \cdots \psi_{M_k}(a_{i_s})$ where multiplication is matrix multiplication.

We have $\psi_{M_k}(a_r a_t) = \psi_{M_k}(a_r)\psi_{M_k}(a_t)$, $a_r, a_t \in \Sigma$.

Example 2.9 Let Σ_3 be the ordered alphabet $\{1 < 2 < 3\}$ and assume that $w = 22113$. Then $\psi_{M_3}(w)$ is a 4×4 triangular matrix and can be computed as follows:

$\psi_{M_3}(22113) = \psi_{M_3}(2)\psi_{M_3}(2)\psi_{M_3}(1)\psi_{M_3}(1)\psi_{M_3}(3) =$

$$\begin{bmatrix}1&0&0&0\\0&1&1&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&1&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&1&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&1\\0&0&0&1\end{bmatrix}=$$

$$\begin{bmatrix}1&2&0&0\\0&1&2&2\\0&0&1&1\\0&0&0&1\end{bmatrix}.$$

However as we mentioned earlier although two words have the same Parikh vector they can have different Parikh matrices. For instance, take $u = 11223$ given the same alphabet as above. $u = 11223$ has the same Parikh vector as 22113 but it can easily be verified that

$\psi_{M_3}(11223) = (1)\psi_{M_3}(1)\psi_{M_3}(2)\psi_{M_3}(2)\psi_{M_3}\psi_{M_3}(3) =$

$$\begin{bmatrix}1&2&4&4\\0&1&2&2\\0&0&1&1\\0&0&0&1\end{bmatrix} \neq \begin{bmatrix}1&2&0&0\\0&1&2&2\\0&0&1&1\\0&0&0&1\end{bmatrix} = \psi_{M_3}(22113).$$

Theorem 2.10 (Parikh matrix mapping) [1] Let $\Sigma = \{a_1 < \cdots < a_k\}$ be the an ordered alphabet with $k \geq 1$ and $w \in \Sigma^*$, then the matrix $\psi_{M_k}(w) = (m_{i,j})_{1\leq i,j\leq k+1}$ has the following properties:

1) $m_{i.j} = 0$, $\forall\ 1 \leq j < i \leq k+1$,

2) $m_{i,i} = 1$, $\forall\ 1 \leq \mathrm{i} \leq \mathrm{k}+1$

3) $m_{i,j+1} = |w|_{a_{i,j}}$, $\forall\ 1 \leq \mathrm{i} \leq \mathrm{j} \leq \mathrm{k}+1$.

Proof [1] Clearly property 1) and 2) follows directly from the definition. Let's prove property 3) by induction. Assume that $|w| = n$. If $n \le 1$ the assertion is clearly true. Assume that 3) holds for all words of length *n* and let *w* be of length $n+1$. $w = ua_i$, where $|u| = n$ and $a_i \in \Sigma$ with $1 \le i \le k$. It follows that $\psi_{M_k}(w) = \psi_{M_k}(ua_i) = \psi_{M_k}(u)\psi_{M_k}(a_i)$.

Assume that $\psi_{M_k}(u) = \begin{bmatrix} 1 & m_{1,2} & \dots & \dots & m_{1,k+1} \\ 0 & 1 & \dots & \dots & m_{2,k+1} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & m_{k,k+1} \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix}$. By the induction hypothesis $\psi_{M_k}(u)$ has property 3). From **Definition 2.8** we have that $\psi_{M_k}(a_i) = \begin{bmatrix} 1 & 0 & \dots & & & 0 \\ \dots & & & & & \dots. \\ 0 & \dots & 1 & 1 & \dots & 0 \\ & & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots. \\ 1 & 0 & \dots & & & 1 \end{bmatrix}$.

All the elements in the matrix $\psi_{M_k}(a_i)$ are 0, except the main diagonal, that is all 1:s and entry $m_{i,i+1} = 1$. So we have $\psi_{M_k}(w) = \psi_{M_k}(ua_i) = \psi_{M_k}(u)\psi_{M_k}(a_i) =$

$$\begin{bmatrix} 1 & m_{1,2} & \dots & \dots & m1,k+1 \\ 0 & 1 & \dots & \dots & m_{2,k+1} \\ \dots & \dots & & \dots & \dots. \\ & & & \dots & \\ & & & \dots & m_{k,k+1} \\ 0 & 0 & \dots & \dots & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & & & 0 \\ \dots & & & & & \dots. \\ 0 & \dots & 1 & 1 & \dots & 0 \\ & & & & & \\ \dots & \dots & \dots & \dots & \dots & \dots. \\ 1 & 0 & \dots & & & 1 \end{bmatrix} = N.$$

The resulting matrix *N* has the following property: $n_{j,i+1} = m_{j,i} + m_{j,i+1}$ for all $1 \le j \le i$ and for all indices, $n_{p,q} = m_{p,q}$. □

Given the above theorem we can easily get the following corollary, since this follows directly from property 3) $m_{i,j+1} = |w|_{a_{i,j}}$, $\forall\ 1 \le \mathrm{i} \le \mathrm{j} \le \mathrm{k}+1$, in the above theorem:

Corollary 2.11 [1] Let $\Sigma = \{a_1, \dots, a_k\}_<$ be an ordered alphabet. The Parikh matrix $\psi_{M_k}(w) = (m)_{1 \le i,j \le k+1}$ has as its *second diagonal,* the Parikh vector of *w*. That is $(m_{1,2}, m_{2,3}, \dots, m_{k,k+1}) = (|w|_{a_1}, \dots, |w|_{a_k}) = \psi(w)$. □

Example 2.12 Given the alphabet $\Sigma = \{1 < 2 < 3\}$ and $w = 22113$ we have

$\psi_{M_3}(22113) = \begin{bmatrix} 1 & \underline{2} & 0 & 0 \\ 0 & 1 & \underline{2} & 2 \\ 0 & 0 & 1 & \underline{1} \\ 0 & 0 & 0 & 1 \end{bmatrix}$. We can see that the Parikh vector $\psi(22113) = (2,2,1)$ equals the second diagonal of the matrix.

Now I will proceed and present some of the new results on Parikh matrices in connection with permutations statistics that I have developed.

3 The Generalized Parikh Mapping

This section is the part where I start develop the Parikh matrix subject further to establish new connections with existing permutation statistics. Some of the ideas have been mentioned informally in the introduction. Here I will attempt to give a more formal foundation for these ideas. I am going to show a *matrix completeness* (**Theorem 3.5**) for Parikh matrices of words in S_n. We shall see that the Parikh matrix of words in S_n is completely determined by its third diagonal. Therefore it is natural to generalize **Corollary 2.11** so that it connects each *j+1:th* diagonal with a certain Parikh vector of order j in the following way:

Definition 3.1 (Generalization of the Parikh Vector Mapping) Let $\Sigma = \{a_1,...,a_k\}_<$ be an ordered alphabet. Define the *Parikh vector mapping of order j* as the mapping $\psi_j : \Sigma^* \to \mathbb{Z}^{k-j+1}$, defined by $\psi_j(w) = (|w|_{a_{1,j}}, |w|_{a_{2,j+1}}, ..., |w|_{a_{k-j+1,k}})$.

In particular $j = 1$ gives the classical Parikh vector mapping. If we set $a_{i,i} = a_i$, then $\psi_1(w) = (|w|_{a_{1,1}}, |w|_{a_{2,2}}, ..., |w|_{a_{k,k}}) = (|w|_{a_1}, |w|_{a_2}, ..., |w|_{a_k})$, corresponding to the second diagonal.

Analogically, which is easy to verify by using **Theorem 2.10**, we get a generalization of **Corollary 2.11** as follows:

Corollary 3.2 Let $\Sigma = \{a_1, ..., a_k\}_<$ be an ordered alphabet. The Parikh matrix $\psi_{M_k}(w) = (m)_{1 \le i, j \le k}$ has as its *j*:th diagonal, the Parikh vector of order *j* of the word *w*. That is $(m_{1,j+1}, m_{2,j+2}, ..., m_{k-j+1,k}) = \psi_j(w) = (|w|_{a_{1,j}}, |w|_{a_{2,j+1}}, ..., |w|_{a_{k-j+1,k}})$.

Proof Again this follows directly from property 3) in **Theorem 2.10**, same as for **Corollary 2.11** and by the definition of the generalized Parikh vector mapping. □

Example 3.3 (Generalized Parikh vector and the Parikh matrix diagonals). Let $\Sigma_5 = \{1 < 2 < 3 < 4 < 5\}$, $M = \{1, 2, 3, 4^2, 5^2\}$ and consider the word $w = 1235445 \in S_7(M)$.

Then we have $\psi_{M_5}(w) = \begin{bmatrix} 1 & |w|_1 & |w|_{12} & |w|_{123} & |w|_{1234} & |w|_{12345} \\ 0 & 1 & |w|_2 & |w|_{23} & |w|_{234} & |w|_{2345} \\ 0 & 0 & 1 & |w|_3 & |w|_{34} & |w|_{345} \\ 0 & 0 & 0 & 1 & |w|_4 & |w|_{45} \\ 0 & 0 & 0 & 0 & 1 & |w|_5 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

By **Corollary 3.2** we have the following:

The (trivial) Parikh vector of order $j = 0$ corresponds to the main diagonal. That is $\psi_0(w) = (1,1,1,1,1,1)$ corresponds to the underlined main (first) diagonal in

$$\begin{bmatrix} \underline{1} & 1 & 1 & 1 & 2 & 2 \\ 0 & \underline{1} & 1 & 1 & 2 & 2 \\ 0 & 0 & \underline{1} & 1 & 2 & 2 \\ 0 & 0 & 0 & \underline{1} & 2 & 2 \\ 0 & 0 & 0 & 0 & \underline{1} & 2 \\ 0 & 0 & 0 & 0 & 0 & \underline{1} \end{bmatrix}.$$

The (classical) Parikh vector of order $j = 1$ corresponds to the second diagonal. That is

$\psi_1(w) = (|w|_1, |w|_2, |w|_3, |w|_4, |w|_5) = (1,1,1,2,2)$ corresponds to the underlined second diagonal

in $\begin{bmatrix} 1 & \underline{1} & 1 & 1 & 2 & 2 \\ 0 & 1 & \underline{1} & 1 & 2 & 2 \\ 0 & 0 & 1 & \underline{1} & 2 & 2 \\ 0 & 0 & 0 & 1 & \underline{2} & 2 \\ 0 & 0 & 0 & 0 & 1 & \underline{2} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

The Parikh vector of second order ($j = 2$) corresponds to the third diagonal. That is

$\psi_2(w) = (|w|_{12}, |w|_{23}, |w|_{34}, |w|_{45}) = (1,1,2,2)$ corresponds to the underlined third diagonal in

$\begin{bmatrix} 1 & 1 & \underline{1} & 1 & 2 & 2 \\ 0 & 1 & 1 & \underline{1} & 2 & 2 \\ 0 & 0 & 1 & 1 & \underline{2} & 2 \\ 0 & 0 & 0 & 1 & 2 & \underline{2} \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$. etc

The following result will be useful when proving the Matrix Completeness Theorem (**Theorem 3.5**) for words in S_k. The Matrix Completeness Theorem simplifies the task of finding the connection between descent statistic and Parikh statistics because we can consider binary strings in both cases instead of binary strings and binary matrices.

Proposition 3.4 Consider the Parikh matrix mapping on the set of permutations of the elements of [*k*]. Define $\psi_{M_k} : S_k \to M_{k+1}$, by

$$\psi_{M_k}(\pi) = \begin{bmatrix} 1 & 1 & |\pi|_{12} & |\pi|_{123} & |\pi|_{1234} & \cdots & |\pi|_{123\cdots(k-1)} & |\pi|_{123\cdots k} \\ 0 & 1 & 1 & |\pi|_{23} & |\pi|_{234} & \cdots & \cdots & |\pi|_{23\cdots k} \\ 0 & 0 & 1 & 1 & |\pi|_{34} & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & |\pi|_{(k-2)(k-1)} & |\pi|_{(k-2)(k-1)k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & |\pi|_{(k-1)k} \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix} = (m_{i,j})_{1 \le i, j \le k+1}$$

Then for $1 \le i < j \le k$, we have.

1) If $|\pi|_{i\cdots(j-1)}=1$ and $|\pi|_{(i+1)\cdots j}=1$, then $|\pi|_{i\cdots j}=1$.

2) If $|\pi|_{i\cdots(j-1)}=0$ and $|\pi|_{(i+1)\cdots j}=1$, then $|\pi|_{i\cdots j}=0$.

3) If $|\pi|_{i\cdots(j-1)}=1$ and $|\pi|_{(i+1)\cdots j}=0$, then $|\pi|_{i\cdots j}=0$.

4) If $|\pi|_{i\cdots(j-1)}=0$ and $|\pi|_{(i+1)\cdots j}=0$, then $|\pi|_{i\cdots j}=0$.

Proof Let $\pi\in S_k$. Then the entries of $\psi_{M_k}(\pi)$ are exclusively 0's and 1's, since for each word $\pi\in S_k$ every letter of $\Sigma=\{1<\cdots<k\}$ occur exactly once and by consequence the subwords of $\pi\in S_k$ can occur at most once, thus $|\pi|_{i\cdots j}$ is either 0 or 1. We will use this property of S_k in this proof.

1) For $1\le i<j\le k$, assume that $|\pi|_{i(i+1)\cdots(j-1)}=1$ and $|\pi|_{(i+1)\cdots(j-1)(j)}=1$. This means that the words $i(i+1)\cdots(j-1)$ and $(i+1)\cdots(j-1)(j)$ both exist as subwords of π and therefore, since we are working in S_k, $i(i+1)\cdots(j)$ must also exist as a subword of π. Thus $|\pi|_{i\cdots j}=1$.

2) For $1\le i<j\le k$, assume that $|\pi|_{i(i+1)\cdots(j-1)}=0$ and $|\pi|_{(i+1)\cdots(j-1)j}=1$**.** This means that the word $i(i+1)\cdots(j-1)$ does not exist as a subword of π. Then $i(i+1)\cdots(j-1)j$ cannot exist since $i(i+1)\cdots(j-1)$ is a subword of $i(i+1)\cdots(j-1)j$. Thus $|\pi|_{i\cdots j}=0$.

3) For $1\le i<j\le k$, assume that $|\pi|_{i\cdots(j-1)}=1$ and $|\pi|_{(i+1)\cdots j}=0$. Then by a similar reasoning as in 2) we get that $|\pi|_{i\cdots j}=0$.

4) For $1\le i<j\le k$, assume that both $|\pi|_{i\cdots(j-1)}=0$ and $|\pi|_{(i+1)\cdots j}=0$. Then by the same reasoning as in 2) we get that $|\pi|_{i\cdots j}=0$. □

In conclusion the only way to get $|\pi|_{i\cdots j}=1$ is if both when $|\pi|_{i\cdots(j-1)}=1$ and $|\pi|_{(i+1)\cdots j}=1$ otherwise $|\pi|_{i\cdots j}=0$. We can think of this as a rule when computing the Parikh matrix $\psi_{M_k}(w)$ of a word $\pi\in S_k$. All entries of the main diagonal in a Parikh matrix are 1's, by the Parikh matrix Theorem (**Theorem 2.10**).

As we mentioned earlier about words $\pi\in S_k$, all the entries of the second diagonal of the Parikh matrix (the classical Parikh vector, **Corollary 2.11**) also consists exclusively of 1's. By **Corollary 3.2**, the third diagonal is the Parikh vector of second order

$\psi_2(\pi) = (|\pi|_{12}, |\pi|_{23}, ..., |\pi|_{k(k-1)})$. The entries $|\pi|_{i(i+1)}$ of this vector depends on whether or not $i(i+1)$ exist as a subword of π.

If $i(i+1)$ exist as a subword of π then $|\pi|_{i(i+1)} = 1$. If $i(i+1)$ does not exist as a subword of π then $|\pi|_{i(i+1)} = 0$.

By the above proposition we can automatically compute the entries in the fourth diagonal from the third diagonal and the fifth diagonal from the fourth diagonal etc, until we have completed the computation of the matrix.

Now having established this result we have the tools for proving the Matrix completeness.

Theorem 3.5 (Parikh Matrix Completeness Theorem) Consider the Parikh vector of second order $\psi_2 : S_k \to \mathbb{Z}_2^{k-1}$, $\psi_2(\pi) = (|\pi|_{12}, ..., |\pi|_{(k-1)k})$, for $\pi \in S_k$. Then the Parikh matrix mapping $\psi_{M_k} : S_k \to M_{k+1}$, $\psi_{M_k}(\pi) = (m)_{1 \le i, j \le j}$ is completely determined by $\psi_2(\pi)$.

Proof By the Parikh matrix Theorem (**Theorem 2.10**) we have

$$\psi_{M_k}(\pi) = \begin{bmatrix} 1 & 1 & |\pi|_{12} & |\pi|_{123} & |\pi|_{1234} & \cdots & |\pi|_{123\cdots(k-1)} & |\pi|_{123\cdots k} \\ 0 & 1 & 1 & |\pi|_{23} & |\pi|_{234} & \cdots & \cdots & |\pi|_{23\cdots k} \\ 0 & 0 & 1 & 1 & |\pi|_{34} & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & |\pi|_{(k-2)(k-1)} & |\pi|_{(k-2)(k-1)k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & |\pi|_{(k-1)k} \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

All the entries below the main diagonal are always 0 (by **Theorem 2.10**). All the entries in the main diagonal are always 1 (by **Theorem 2.10**). Since we are working in S_k, all the entries of the second diagonal are always 1 (by **Corollary 2.11** (classic vector)). By **Corollary 3.2** we have that the third diagonal corresponds to the Parikh vector of second order so the Parikh vector of second order determines the third diagonal in the Parikh matrix. By **Proposition 3.4** we have, for $1 \le i < j \le k$, that the values of $|\pi|_{i\cdots(j-1)}$ and $|\pi|_{(i+1)\cdots j}$ determines the value of $|\pi|_{i\cdots j}$. Therefore the Parikh vector of second order $(|\pi|_{12}, |\pi|_{23}, ..., |\pi|_{(k-2)(k-1)}, |\pi|_{(k-1)k})$

determines $(|\pi|_{123}, |\pi|_{234}, ..., |\pi|_{(k-3)(k-2)(k-1)}, |\pi|_{(k-2)(k-1)k})$ which is the Parikh vector of third order. By **Corollary 3.2** this corresponds to the fourth diagonal of the Parikh matrix. In the same way we get by **Proposition 3.4**, that the fifth diagonal follows from the third diagonal in the same way. We get all the diagonals by the same reasoning. In conclusion the Parikh vector of second order is enough information to compute all the entries of the Parikh matrix, when we consider words that belong to S_k. □

The Parikh matrix completeness for permutations in S_k significantly speeds up the computations of Parikh matrices.

To demonstrate this, here is a simple example

Example 3.6 Let $\sum_5 = \{1 < 2 < 3 < 4 < 5\}$ and consider the words in S_5. Consider the following 3 permutations $\pi_1 = 51324$, $\pi_2 = 12534$, $\pi_3 = 25134$, $\pi_4 = 12345$, $\pi_5 = 54321$. Then we have:

1) $\psi_2(\pi_1) = (|\pi_1|_{12}, |\pi_1|_{23}, |\pi_1|_{34}, |\pi_1|_{45}) = (1,0,1,0)$
2) $\psi_2(\pi_2) = (|\pi_2|_{12}, |\pi_2|_{23}, |\pi_2|_{34}, |\pi_2|_{45}) = (1,1,1,0)$
3) $\psi_2(\pi_3) = (|\pi_3|_{12}, |\pi_3|_{23}, |\pi_3|_{34}, |\pi_3|_{45}) = (0,1,1,0)$
4) $\psi_2(\pi_4) = (|\pi_4|_{12}, |\pi_4|_{23}, |\pi_4|_{34}, |\pi_4|_{45}) = (1,1,1,1)$
5) $\psi_2(\pi_5) = (|\pi_5|_{12}, |\pi_5|_{23}, |\pi_5|_{34}, |\pi_5|_{45}) = (0,0,0,0)$

Let us just compute the diagonals resulting from these (third diagonals) Parikh vectors of second order as described earlier. To make it easier to visually follow the computations I have removed the all entries below the third diagonal since they are the same for all Parikh matrices over words in S_k.

Then we get:

1) $\psi_{M_5}(\pi_1) = \begin{bmatrix} & & 1 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 \\ & & & & 1 & 0 \\ & & & & & 0 \\ & & & & & \\ & & & & & \end{bmatrix}$

2) $\psi_{M_5}(\pi_2) = \begin{bmatrix} & & 1 & 1 & 1 & 0 \\ & & & 1 & 1 & 0 \\ & & & & 1 & 0 \\ & & & & & 0 \\ & & & & & \\ & & & & & \end{bmatrix}$

3) $\psi_{M_5}(\pi_3) = \begin{bmatrix} & & 0 & 0 & 0 & 0 \\ & & & 1 & 1 & 0 \\ & & & & 1 & 0 \\ & & & & & 0 \\ & & & & & \\ & & & & & \end{bmatrix}$

4) $\psi_{M_5}(\pi_4) = \begin{bmatrix} & & 1 & 1 & 1 & 1 \\ & & & 1 & 1 & 1 \\ & & & & 1 & 1 \\ & & & & & 1 \\ & & & & & \\ & & & & & \end{bmatrix}$

5) $\psi_{M_5}(\pi_5) = \begin{bmatrix} & & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 \\ & & & & 0 & 0 \\ & & & & & 0 \\ & & & & & \\ & & & & & \end{bmatrix}$

It is also visually more apparent if viewed as arithmetical triangles where and bottom row corresponds to the third diagonal and the next row to the fourth diagonal etc. Entries follow the rules from **Proposition 3.4** . Then we get

$$
1)\;\begin{array}{c} 0 \\ 0\quad 0 \\ 0\quad 0\quad 0 \\ 1\quad 0\quad 1\quad 0 \end{array}
\qquad
2)\;\begin{array}{c} 0 \\ 1\quad 0 \\ 1\quad 1\quad 0 \\ 1\quad 1\quad 1\quad 0 \end{array}
\qquad
3)\;\begin{array}{c} 0 \\ 0\quad 0 \\ 0\quad 0\quad 0 \\ 0\quad 1\quad 1\quad 0 \end{array}
\qquad
4)\;\begin{array}{c} 1 \\ 1\quad 1 \\ 1\quad 1\quad 1 \\ 1\quad 1\quad 1\quad 1 \end{array}
\qquad
5)\;\begin{array}{c} 0 \\ 0\quad 0 \\ 0\quad 0\quad 0 \\ 0\quad 0\quad 0\quad 0 \end{array}
$$

These arithmetical triangles have the form $\begin{bmatrix} \psi_5(\pi) \\ \psi_4(\pi) \\ \psi_3(\pi) \\ \psi_2(\pi) \end{bmatrix}$, by **Corollary 3.2.**

A consequence of this result is that if we are able to show something for Parikh vectors $\psi_2(\pi)$, $\pi \in S_k$, it can also be used for Parikh matrices $\psi_{M_k}(\pi)$, $\pi \in S_k$. For example if we are counting the number of permutations mapping to a given Parikh vector we are counting what is also equivalent to the number of permutations mapping to a given Parikh matrix.

Similar to the descent set statistic, I will define an analogous permutation statistic for the Parikh mappings. Consider the Parikh vector of order *j*. I will define a permutation statistic by counting the number of permutations that maps onto a specific vector of order 2. By **Theorem 3.5**, the Parikh vector of second order of a word $w \in S_n$ determines the Parikh matrix of this word. As a consequence, counting the number of permutations mapping to a specific Parikh vector of order 2 is equivalent to counting the number of permutations mapping to a given Parikh matrix.

Now let's revisit the descent sets.

4 Descent Statistic

Permutation statistics is an ongoing branch of enumerative combinatorics and is currently under a lot of development. Computing the descent set statistic has been done by several. The methods I have chosen to present in this paper can be found in Stanley's enumerative combinatorics and is based on the Principle of Inclusion-Exclusion. First let's go through the permutation statistics of interest in this thesis, the descent set statistic. We have already been introduced to it informally in the introduction. Now I will go through material in a more formal manner.

Definition 4.1 Let $\pi = \pi_1\pi_2 \cdots \pi_k = S_k$. Define the *descent set mapping* by $D : S_k \to P([k-1])$, where $D(\pi) = \{i : \pi_i > \pi_{i+1}\} = S \subseteq [k-1]$.

The following function will be useful because it maps each descent set of a permutation to a corresponding binary string and thus we can compare it to the binary Parikh vector of a permutation.

Definition 4.2 Let $\phi_B : P([k-1]) \to \mathbb{Z}_2^{k-1}$ be a function defined by

$\phi_B(S) = v = (v_1, v_2, ..., v_{k-1}) \in \mathbb{Z}_2^{k-1}$, such that $v_i = \begin{cases} 0 \text{ if } i \in S \\ 1 \text{ if } i \notin S \end{cases}$, where $S \subseteq [k-1]$.

It's easy to see that ϕ_B in the above definition is a bijection.

We can consider the composition $\phi_B \circ D : S_k \xrightarrow{D} P[k-1] \xrightarrow{\phi_B} \mathbb{Z}_2^{k-1}$, where $\pi \mapsto v$. We may call the vector *v* mapped by this composition function the *descent vector* v of π and denote this by $D_{\phi_B}(\pi) = v$.

Example 4.3 Consider the permutation $\pi = \pi_1\pi_2\pi_3\pi_4\pi_5\pi_6 = 245631 \in S_6$, then $D(\pi) = \{4,5\}$, since we have the following descents $\pi_4 = 6 > 3 = \pi_5$ and $\pi_5 = 3 > 1 = \pi_6$ and all other are *ascents* ($\pi_i < \pi_{i+1}$). The corresponding descent vector is $D_{\phi_B}(\pi) = \phi_B(\{4,5\}) = 11100$.

So, how many permutations in S_6 has got the descent set $D(\pi) = \{4,5\}$?
How is this computed?
In finding a formula for the *descent set statistic* we shall find the Principle of Inclusion-Exclusion very useful. I want to use binary strings as representatives for descent sets, since I will show the connection to Parikh vectors of second order later in the thesis. Therefore I will use a function ϕ_B between sets and binary strings.

First let's define the *inclusive descent statistic*. This is an important construction in this thesis. Because we can use the Principle of Inclusion-Exclusion on it to establish the exclusive descent statistic formula. Later we will also define an analogue for Parikh vectors of second order and in the proof of the **Main Theorem 6.2** we will eventually arrive at this formula.

Definition 4.5 [3] Given $S \subseteq [n-1]$ let $\alpha(S)$ denote the number of permutations $\pi \in S_n$ whose descent set is contained in *S*. That is $\alpha(S) = \#\{\pi \in S_n : D(\pi) \subseteq S\}$.

Proposition 4.6 [3] Let $S = \{s_1, ..., s_k\}_< \subseteq [n-1]$. Then $\alpha(S) = \binom{n}{s_1, s_2 - s_1, s_3 - s_2 ..., n - s_k}$.

Proof ([3] see page 22, proposition 1.3.11) Consider the permutation $\pi = a_1 a_2 \cdots a_n$ with $D(\pi) \subseteq S$, we may first choose $a_1 < a_2 < ... < a_{S_1}$ in $\binom{n}{s_1}$ ways. Proceed by choosing $a_{S_1+1} < a_{S_1+2} < ... < a_{S_2}$ in $\binom{n-s_1}{s_2-s_1}$ ways etc. From this we get

$\alpha(S) = \binom{n}{s_1}\binom{n-s_1}{s_2-s_1}\binom{n-s}{s_3-s_2}\cdots\binom{n-s_k}{n-s_k} = \binom{n}{s_1, s_2 - s_1, ..., n - s_k}$ as desired.

Now let's define the *exclusive descent statistic*.

Definition 4.7 [3] Given $S \subseteq [n-1]$**,** the *exclusive descent statistic* is defined as $\beta(S) = \#\{\pi \in S_n : D(\pi) = S\}$**.**

It is clear that $\alpha(S) = \sum_{T \subseteq S} \beta(T)$. This "holds the key" for application of the Principle of Inclusion-Exclusion to get a formula for $\beta(S)$.

I will not go through this here but from Stanleys enumerative combinatorics [3] (by application of the Principle of Inclusion-Exclusion) we have the formula:

(4.8-9) $\beta(S)=\sum_{T\subseteq S}(-1)^{|S-T|}\alpha(T)=\sum_{1\le i_1<i_2<\ldots<i_j\le k}(-1)^{k-j}\binom{n}{s_{i_1},s_{i_2}-s_{i_1},\ldots,n-s_{i_j}}$, where

$\alpha(S)=\binom{n}{s_1,s_2-s_1,\ldots,n-s_k}$, $S=\{s_1<s_2<\ldots<s_k\}\subseteq[n-1]$ and $|S-T|=\#\{S\setminus T\}$

Example 4.10 Given $n-1=7$ and $S=\{1,5\}$. We get

$S=\{s_1=1,s_2=5\}\subseteq\{1,2,\ldots,7\}$. $\beta(S)=\sum_{T\subseteq S}(-1)^{|S-T|}\alpha(T)=$

$$=(-1)^{2-0}\binom{8}{0}+(-1)^{2-1}\left(\binom{8}{1,8-1}+\binom{8}{5,8-5}\right)+(-1)^{2-2}\binom{8}{1,5-1,8-5}=217 .$$

In Stanleys enumerative combinatorics [3], page 69, the descent statistic is also written in an alternative form as a determinant. It can be shown that

$$\beta_n(S)=\sum_{1\le i_1<i_2<\ldots<i_j\le k}(-1)^{k-j}\binom{n}{s_{i_1},s_{i_2}-s_{i_1},\ldots,n-s_{i_j}}=n!\det[1/(s_{j+1}-s_i)].$$

We have

$$\beta_n(S)=n!\begin{vmatrix}\frac{1}{(s_1-s_0)!} & \frac{1}{(s_2-s_0)!} & \cdots & \frac{1}{(s_{k+1}-s_0)!}\\ \frac{1}{(s_1-s_1)!} & \frac{1}{(s_2-s_1)!} & \cdots & \frac{1}{(s_{k+1}-s_1)!}\\ \cdots & \cdots & \frac{1}{(s_{j+1}-s_i)!} & \cdots\\ \frac{1}{(s_1-s_k)!} & \frac{1}{(s_2-s_k)!} & \cdots & \frac{1}{(s_{k+1}-s_k)!}\end{vmatrix}.$$

So we can express the computation of the sum in **Example 4.10** as

$$\beta_n(S)=8!\begin{vmatrix}\frac{1}{1!} & \frac{1}{5!} & \frac{1}{8!}\\ 1 & \frac{1}{4!} & \frac{1}{7!}\\ 0 & 1 & \frac{1}{3!}\end{vmatrix}=217 .$$

5 Comparing the Parikh Vector Mapping to the Descent Vector Mapping

This section is simply dedicated to providing some computations showing of the distribution of the Parikh Statistic and the Descent statistic. The computations indicate that they seem to be equidistributed. We do this by computing the Parikh vector mapping of second order and then for the same value of *n* = 5 we compute the Descent vector mapping (by using the function ϕ_B mapping descent sets to their corresponding binary strings)

Computations 5.1 Consider the Parikh vector mapping of second order on S_5. We have $\psi_2 : S_5 \to \mathbb{Z}_2^4$, where $\psi_2(\pi) = (|\pi|_{12}, |\pi|_{23}, |\pi|_{34}, |\pi|_{45}) = v$

(For convenience I will notate *v* as a word. For example 1110 instead of $(1,1,1,0)$).

$\mathbb{Z}_2^4$ S_5

1111 – 12345

0000 - 54321

0001 – 45321,43521,43251,43215

1110 – 12354,12534,15234,51234

1000 – 54312,54132,51432,15432

0111 – 21345,23145,23415,23451

1100 - 12543,51243,51423,54123,15243,15423

0011 – 34521,43215,32415,32145,34251,32451

1101 – 12435,12453,14235,14253,14523,41235,41253,41523,45123

0010 – 53421,35421,53241,35241,32541,53214,35214,32514,32154

1011 – 13245,31245,13425,31425,34125,13452,31452,34152,34512

0100 – 54231,54213,52431,52413,52143,25431,25413,25143,21543

1001 – 41325,41325,41532,43125,43152,43512,45132,45312,14325,14352,14532

0110 – 52314,52314,23514,52134,25134,21534,23154,21354,52341,25341,23541

1010 - 13254,13524,13542,31254,31524,31542,35124,35142,35412,51324,51342,
53124,53142,53412,15324,15342

0101- 45231,42531,24531,45213,42513,24513,42153,24153,21453,42315,24315
42135,24135,21435,42351,24351

Computation 5.2 Consider $\rho_2(v) = \#\{\pi \in S_5 : \psi_2(\pi) = v \in \mathbb{Z}_2^4\}$.

$\rho_2(1111) = 1$

$\rho_2(0000) = 1$

$\rho_2(1110) = 4$

$\rho_2(0001) = 4$

$\rho_2(0111) = 4$

$\rho_2(1000) = 4$

$\rho_2(1100) = 6$

$\rho_2(0011) = 6$

$\rho_2(0010) = 9$

$\rho_2(1101) = 9$

$\rho_2(0100) = 9$

$\rho_2(1011) = 9$

$\rho_2(1001) = 11$

$\rho_2(0110) = 11$

$\rho_2(1010) = 16$

$\rho_2(0101) = 16$

Let's check the descent set distribution. Given $\phi_B \circ D : S_k \xrightarrow{D} P([k-1]) \xrightarrow{\phi_B} \mathbb{Z}_2^{k-1}$, where $\pi \mapsto S \mapsto v$. For $k = 5$, we have

Computation 5.3 (Descent vector mapping, *n* = 5)

$\mathbb{Z}_2^4$ S_5

1111 – 12345

0000 – 54321

1110 – 12354,12453,13452,23451

0001 – 54312,54213,53214,43215

0111 – 21345,31245,41235,51234

1000 – 45321,35421,25431,15432

1100 – 34521,14532,23541,24531,12543,13542

0011 – 32145,52134,43125,42135,54123,53124

0010 – 32154,42153,43152,43251,52143,53142,53241,54132,54231

1101 – 34512,24513,23514,23415,14523,13524,13425,12534,12435

0100 – 21543,31542,41532,51432,32541,42531,52431,43521,53421

1011 – 45123,35124,25134,15234,34125,24135,14235,23145,13245

1001 – 34215,35214,35412,45213,45312,14325,15324,15423,25314,25413,24315

0110 – 51243,41253,21453,31254,21354,52341,42351,32451,41352,31452,51342

0101 – 53412,41325,41523,42315,42513,43512,51324,51423,52314,52413,21435,
21534,31425,31524,32415,32514

1010 - 13254,25341,25143,24351,24153,23154,15342,15243,14352,14253,45231,
45132,35241,35142,34251,34152

Computation 5.4 Consider $\beta_{\phi_B}(v) = \#\{\pi \in S_5 : D_{\phi_B}(\pi) = v \in \mathbb{Z}_2^4\}$.

$\beta_{\phi_B}(1111) = 1$

$\beta_{\phi_B}(0000) = 1$

$\beta_{\phi_B}(1110) = 4$

$\beta_{\phi_B}(0001) = 4$

$\beta_{\phi_B}(0111) = 4$

$\beta_{\phi_B}(1000) = 4$

$\beta_{\phi_B}(1100) = 6$

$\beta_{\phi_B}(0011) = 6$

$\beta_{\phi_B}(0010) = 9$

$\beta_{\phi_B}(1101) = 9$

$\beta_{\phi_B}(0100) = 9$

$\beta_{\phi_B}(1011) = 9$

$\beta_{\phi_B}(1001) = 11$

$\beta_{\phi_B}(0110) = 11$

$\beta_{\phi_B}(1010) = 16$

$\beta_{\phi_B}(0101) = 16$

We see that for S_5 and $v \in \mathbb{Z}_2^{5-1}$, $\rho_2(v) = \beta_{\phi_B}(v)$. In the main theorem I will prove that this is true for all $v \in \mathbb{Z}_2^{k-1}$ when we are considering ψ_2 and D_{ϕ_B} over S_k.

6 Yamba's Theorem

We have seen that for $v \in \mathbb{Z}_2^{5-1}$, $\rho_2(v) = \beta_{\phi_B}(v)$. I conjecture that this is true for all $v \in \mathbb{Z}_2^{k-1}$ and for all $k \geq 2$ when we are considering ψ_2 and D_{ϕ_B} over S_k. The theorem can be considered as the main result of this paper.

The proof of the *exclusive descent set statistic* formula is based on the Principle of Inclusion-Exclusion. There are several methods of computing this statistic but I have chosen the methods found in [3]. (See **Proposition 4.6** and **Theorem 4.11**).
It is achieved by first defining the descent set inclusively in terms of subsets of a descent set and inverting the formula by identifying that it holds the setting of the Principle of Inclusion-Exclusion. To get the analogue thinking for the Parikh vectors of second order I will make use

of subvectors (defined in **Definition 6.1**) instead of subsets. The bijection ϕ_B, between subsets and binary vectors motivates the use of this. This is useful later when we define *the inclusive descent set statistic*.

Out of simplicity we will denote the binary vectors as binary words.

Definition 6.1 Consider the set of vectors of $\mathbb{Z}_2^{k-1}$. Let $u = u_1 \cdots u_{k-1} \in \mathbb{Z}_2^{k-1}$ and $v = v_1 \cdots v_{k-1} \in \mathbb{Z}_2^{k-1}$. Define a relation $\leq$ on the set $\mathbb{Z}_2^{k-1}$ by $u \leq v$ if $u_1 \leq v_1, ..., u_{k-1} \leq v_{k-1}$. If $u \leq v$ we may call *u* a *sub vector* of *v* and *v* a *super vector* of *u*.

Example 6.2 Let $\Sigma = \{1 < 2 < 3 < 4 < 5\}$ and consider $w = 13452 \in S_5$ and $z = 13524 \in S_5$. Then we have $\psi_2(w) = (|w|_{12}, |w|_{23}, |w|_{34}, |w|_{45}) = u = (u_1, u_2, u_3, u_4) = (1,0,1,1)$ and $\psi_2(z) = (|z|_{12}, |z|_{23}, |z|_{34}, |z|_{45}) = v = (v_1, v_2, v_3, v_4) = (1,0,1,0)$. We have

$v_1 = 1 \leq 1 = u_1$

$v_2 = 0 \leq 0 = u_2$

$v_3 = 1 \leq 1 = u_3$

$v_4 = 0 \leq 1 = u_4$

Thus $v \leq u$, so we have that *v* is a sub vector of *u* and *u* is a super vector of *v*.

Descent statistic is originally formulated in terms of sets but one should keep in mind the bijection ϕ_B to be aware of the connection between the descent statistic and the Parikh statistic.

Definition 6.8 Let $\Sigma = \{a_1 < \cdots < a_k\}$ and $v \in \mathbb{Z}_2^{k-1}$. The *inclusive Parikh vector statistic of second order* is defined by $\vartheta_2(v) = \#\{w \in S_k : \psi_2(w) \geq v\}$.

Definition 6.9 Let $\Sigma = \{1 < \cdots < k\}$, and $v \in \mathbb{Z}_2^{k-1}$.

The *exclusive Parikh vector statistic of second order* is defined by

$\rho_2(v) = \#\{w \in S_k : \psi_2(w) = v \in \mathbb{Z}_2^{k-1}\}$

We will now define the Parikh matrix statistic $\rho_{M_k}((m_{i,j})_{k+1})$, counting the number of permutations with a given Parikh matrix $(m_{i,j})_{k+1}$. By the Parikh Matrix Completeness Theorem (**Theorem 3.5**), we get that $\rho_{M_k}((m_{i,j})_{k+1}) = \rho_2(v)$, where v is a Parikh vector of second order corresponding to the third diagonal in $(m_{i,j})_{k+1}$ (**Corollary 3.2**).

Definition 6.11 Let $\sum = \{1 < \cdots < k\}$, consider $\psi_{M_k}(S_k)$ and let $(m)_{1\le i,j\le k+1} \in \psi_{M_k}(S_k)$. The *Parikh matrix statistic* is defined by

$\rho_{M_k}((m_{i,j})_{k+1}) = \#\{w \in S_k : \psi_{M_k}(w) = (m)_{1\le i,j\le k+1} \in \psi_{M_k}(S_k)\}$

Now I will show that the defined Parikh vector statistic is in fact equidistributed with the descent set statistic, when we consider words in S_k. The proof given for this is inspired by Stanley's proof of the inclusive descent statistic. It can be simplified by observing the similarity of reasoning. Choosing a permutation with a given "inclusive descent" corresponds to choosing its elements of $[n]$ satisfying a given ascent pattern. In the Parikh case we have a given subword pattern and choose among $[n]$ incides in the same way as in the descent case. To compare the similarity in reasoning see [3], proposition 1.3.11.

Main Theorem 6.12

Let ϕ_B be the function from **Definition 4.3**. For $S \subseteq [k-1]$ and $v = \phi_B(S) \in \mathbb{Z}_2^{k-1}$ consider $\rho_2(v) = \#\{w \in S_k : \psi_2(w) = v \in \mathbb{Z}_2^{k-1}\}$ and $\beta_{\phi_B}(v) = \#\{w \in S_k : D_\phi(w) = v \in \mathbb{Z}_2^{k-1}\}$.

Then $\rho_2(v) = \beta_{\phi_B}(v)$. (The Parikh vector statistic of second order is equidistributed to the descent statistic when we consider words in S_k). The proof follows the same ideas as in [3], page 22.

Proof Consider $\phi_B : P([k-1]) \to \mathbb{N}_2^{k-1}$, $\phi_B(S) = v$, $S \subseteq [k-1]$,

Let $\vartheta_2(v)$ be the inclusive Parikh statistic $\vartheta_2(v) = \#\{\pi \in S_k : \psi_2(\pi) \ge v \in \mathbb{Z}_2^{k-1}\}$ and $\rho_2(v)$ be the exclusive Parikh statistic $\rho_2(v) = \#\{\gamma \in S_k : \psi_2(\gamma) = v \in \mathbb{Z}_2^{k-1}\}$.

Clearly $\vartheta_2(v)=\sum_{u\geq v}\rho_2(u)$ (in the same way $\alpha(S)=\sum_{T\subseteq S}\beta(T)$, [3], page 22), so if we can find a formula for $\vartheta_2(v)$, we can invert it, in analogue with $\alpha(S)=\sum_{T\subseteq S}\beta(T)\Leftrightarrow\beta(S)=\sum_{T\subseteq S}(-1)^{|S-T|}\alpha(T)$ to get the desired result.

Let $v=v_1\cdots v_{k-1}\in\mathbb{Z}_2^{k-1}$, such that $\phi_B^{-1}(v)=S=\{s_1,...,s_j\}\subseteq[k-1]$. To create a $\pi\in S_k$ that has the Parikh vector of second order $\psi_2(\pi)\geq v$, first let's take a closer look at v.

We have, for $1\leq i\leq j$, $v=v_1\cdots v_{s_1}\cdots v_{s_2}\cdots v_{s_j}$ where $v_{s_i}=0$ and $v_l=1$ for all other $s_i\neq l$.

This means that

$$\psi_2(\pi)=\left(\underbrace{...,|\pi|_{s_1(s_1+1)},...,|\pi|_{s_2(s_2+1)},...,|\pi|_{s_j(s_j+1)},...}_{k-1}\right)\geq\underbrace{(...,\underset{s_1}{0},1,...,1,\underset{s_2}{0},1,...,1,\underset{s_j}{0},1,...)}_{k-1}=v.$$

Now to create a $\pi\in S_k$ that has the Parikh vector of second order $\psi_2(\pi)\geq v$.

First place the subword $12\cdots(s_1-1)s_1$ among *k* available indices of π. This can be done in $\binom{k}{s_1}$ ways. Then place $(s_1+1)\cdots(s_2-1)s_2$ among the $k-s_1$ positions.

This can be done in $\binom{k-s_1}{s_2-s_1}$ ways. If we continue this process we see that

$$\vartheta_2(v)=\binom{k}{s_1}\binom{k-s_1}{s_2-s_1}\binom{k-s_2}{s_3-s_2}\cdots\binom{k-s_j}{k-s_j}=\binom{k}{s_1,s_2-s_1,s_3-s_2,...,k-s_j}.$$

Now recall that we have $\vartheta_2(v)=\sum_{u\geq v}\rho_2(u)$. Hence by the Principle of Inclusion-Exclusion (the formula) and **Theorem 4.11** we have

$$\rho_2(v)=\sum_{1\leq i_1<i_2<\cdots<i_t\leq j}(-1)^{j-t}\binom{k}{s_{i_1},s_{i_2}-s_{i_1},...,k-s_{i_t}}=\beta(S)=\beta_{\phi_B}(v).\ \square$$

Let $S=\{s_1,...,s_j\}\subseteq[k-1]$. Then to choose a permutation $\pi=a_1\cdots a_k\in S_k$, where $D(\pi)\subseteq S$. We can do this by first choosing the elements a_i from $[k]$ $a_1<\cdots<a_{s_1}$. This can be done in $\binom{k}{s_1}$ ways and this corresponds to placing the subword $1\cdot2\cdots(s_1-1)s_1$ among *k* available positions.

I claim that $D_\phi(\pi) = v = \psi_2(\pi^{-1})$. But first let's define what the inverse permutation π^{-1} is .

Definition 6.13 Given a permutation $\pi = a_1 a_2 \cdots a_k = \begin{pmatrix} 1 & 2 & \dots & k \\ a_1 & a_2 & \dots & a_k \end{pmatrix} \in S_k$

The *inverse permutation* is defined as $\pi^{-1} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \\ 1 & 2 & \dots & k \end{pmatrix} \in S_k$.

There is a one-to-one correspondence between a permutation and its inverse permutation. To get a permutation of the same form as $\pi = a_1 a_2 \cdots a_k$ the indices a_i in π^{-1} are placed in increasing order we get $\pi^{-1} = b_1 b_2 \cdots b_k \in S_k$.

Example 6.14 Consider $\pi = 13452 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 5 & 2 \end{pmatrix} \in S_5$. Then

$$\pi^{-1} = \begin{pmatrix} 1 & 3 & 4 & 5 & 2 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 2 & 3 & 4 \end{pmatrix} = 15234 \in S_5$$

Theorem 6.15 Let $\pi \in S_k$, let $\Sigma = \{1 < \cdots < k\}$ be an ordered alphabet and let

$D_{\phi_B}(\pi) = v \in \mathbb{Z}_2^{k-1}$.

Then for $k \ge 2$ exist we have that $D_\phi(\pi) = v = \psi_2(\pi^{-1})$.

Proof 1

Consider $\pi = a_1 \cdots a_k \in S_k$ such that $D(\pi) = S = \{s_1, s_2, \dots, s_j\} \subseteq [k-1]$. This means

$a_1 < a_2 < \cdots < a_{s_1} > a_{s_1+1} < \cdots < a_{s_2} > a_{s_2+1} < \cdots < a_{s_j} > a_{s_j+1} < \cdots < a_{s_k}$. Let's use an alternative notation for π:

$$\pi = \begin{pmatrix} 1 & 2 & \dots & s_1 & s_1+1 & \dots & s_2 & s_2+1 & \dots & s_j & s_j+1 & \dots & k \\ a_1 & <a_2< & \dots & <a_{s_1}> & a_{s_1+1}< & \dots & <a_{s_2}> & a_{s_2+1}< & \dots & <a_{s_j}> & a_{s_j+1}< & \dots & <a_{s_k+1} \\ & 1 \quad 1 & & 1 \quad 0 & 1 & & 1 \quad 0 & 1 & & 1 \quad 0 & 1 & & 1 \end{pmatrix}$$

. With standard notation we have

$$\pi = \begin{pmatrix} 1 & 2 & \dots & s_1 & s_1+1 & \dots & s_2 & s_2+1 & \dots & s_j & s_j+1 & \dots & k \\ a_1 & a_2 & \dots & a_{s_1} & a_{s_1+1} & \dots & a_{s_2} & a_{s_2+1} & \dots & a_{s_j} & a_{s_j+1} & \dots & a_{s_k+1} \end{pmatrix}.$$

The first row is indices of π, the second row is the letters a_i of $\pi = a_1 \cdots a_k \in S_k$ with inequalities. The third row is the descent vector $D_{\phi_B}(\pi) = \phi_B(S) = v$.

If we let indices change places then we get (with an alternative notation)

$$\pi^{-1} = \begin{pmatrix} a_1 & < a_2 < & \ldots & < a_{s_1} > & a_{s_1+1} < & \ldots & < a_{s2} > & a_{s_2+1} < & \ldots & < a_{s_j} > & a_{s_j+1} < & \ldots & < a_{s_k+1} \\ 1 & 2 & \ldots & s_1 & s_1+1 & \ldots & s_2 & s_2+1 & \ldots & s_j & s_j+1 & \ldots & k \\ & & & & & & & & & & & & \end{pmatrix}$$

. With standard notation we have

$$\pi^{-1} = \begin{pmatrix} a_1 & a_2 & \ldots & a_{s_1} & a_{s_1+1} & \ldots & a_{s_2} & a_{s_2+1} & \ldots & a_{s_j} & a_{s_j+1} & \ldots & a_{s_k+1} \\ 1 & 2 & \ldots & s_1 & s_1+1 & \ldots & s_2 & s_2+1 & \ldots & s_j & s_j+1 & \ldots & s_k \\ & & & & & & & & & & & & \end{pmatrix}.$$

Now $1 \cdot 2 \cdots s_1$ must exist as a subword of π^{-1} since indices $a_1 < a_2 < \cdots < a_{s_1}$.

$s_1(s_1+1)$ does not exist as a subword of π^{-1} since indices $a_{s_1} > a_{s_1+1}$ means the letter s_1+1 is to the left of s_1 in π^{-1}. Therefore we have $\left(\left|\pi^{-1}\right|_{12}, \left|\pi^{-1}\right|_{23}, \ldots, \left|\pi^{-1}\right|_{s_1(s_1+1)}, \ldots\right) = \left(1,1,\ldots,\underbrace{0}_{s_1},\ldots\right)$.

If we continue the same reasoning for s_i , $1 \le i \le j$ we will attain all the entries of what is the Parikh vector of second order.

We have $\psi_2(\pi^{-1}) = \left(\left|\pi^{-1}\right|_{12}, \left|\pi^{-1}\right|_{23}, \ldots, \left|\pi^{-1}\right|_{s_1(s_1+1)}, \ldots\right) = 11 \cdots 1\underbrace{0}_{s_1}1 \cdots 1\underbrace{0}_{s_2}1 \cdots 1\underbrace{0}_{s_j}1 \cdots 11 = D_{\phi_B}(\pi)$ □

I will now use this result for an alternative proof of **Theorem 6.14**.

Proof 2 of Theorem 6.12

Since there a one-to-one correspondence between the set of permutations and their inverse permutations and by **Theorem 6.15,** $\psi_2(\pi^{-1}) = D_{\phi_B}(\pi)$, we have for $v \in \mathbb{Z}_2^{k-1}$ that $\rho_2(v) = \#\{\gamma \in S_k : \psi_2(\gamma) = v\} = \#\{\gamma^{-1} \in S_k : D(\gamma^{-1}) = v\} = \beta_{\phi_B}(v)$. For each permutation $\gamma \in S_k$, where $\psi_2(\gamma) = v$, there is exactly one permutation $\gamma^{-1} \in S_k$, where $D(\gamma^{-1}) = v$.

So $\beta_{\phi_B}(v) = \rho_2(v)$. □

Corollary 6.16 Let $\Sigma = \{1 < 2 < \cdots < k\}$, let $v \in \mathbb{Z}_2^{k-1}$ and let $(m_{i,j})_{k+1} \in M_k$ be a Parikh matrix of a word belonging to S_k, where *v* corresponds to the third diagonal. Then $\rho_{M_k}((m_{i,j})_{k+1}) = \beta_{\phi_B}(v)$.

Proof

As mentioned earlier by the Parikh Matrix Completeness Theorem (**Theorem 3.5**), we get that $\rho_{M_k}((m_{i,j})_{k+1}) = \rho_2(v)$, where v is a Parikh vector of second order corresponding to the third diagonal in $(m_{i,j})_{k+1}$ (**Corollary 3.2**). By the **Main Theorem** $\rho_2(v) = \beta_{\phi_B}(v)$ so it follows that $\rho_{M_k}((m_{i,j})_{k+1}) = \rho_2(v) = \beta_{\phi_B}(v)$. □

7 $\rho_2(v)$ for binary words

In this section I will present some results in computing the number of binary word mapping to a given parikh vector of second order.

Definition 7.1 Let (b_i, b_j) denote the inversion of the letters b_i and b_j in the word $w = b_1 b_2 \cdots b_n \in S_n(M)$, (where M is a finite multiset), if $i < j$ and $b_i > b_j$.

Let $i(w)$ denote the number of inversions of $w = b_1 b_2 \cdots b_n \in S_n(M)$.

Definition 7.2 The number of words with k inversions is denoted $inv(w) = \#\{w \in S_n(M) : i(w) = k\}$, where $0 \le k \le n-1$.

A generating function for this would have the form $\sum_{w \in S_n(M)} q^{i(w)}$.

Let's say we want to count $\#\{w \in S_n(M) : \psi_2(w) = (|w|_{12}) = (k)\}$, where $0 \le k \le n-1$ and $M = \{1^a, 2^b\}$, $\#M = a + b = n$.

can be attained by observing that counting the number of subwords 21 in w. That is, $|w|_{21}$ corresponds to counting the number of inversions in a permutation and the fact that we are dealing with binary words. Also, since we are working with binary words, symmetry gives us that $\psi_2(\overline{w}) = i(w)$, where $\overline{w}$ is the *(binary) complement word* (for example $w = 1122121 \Rightarrow \overline{w} = 2211212$). And since we have, by

Proposition 1.3.17 ([3] page 26), that for $M=\{1^{a_1},2^{a_2}...,m^{a_m}\}$, with $\#M=n$

$$\Rightarrow \sum_{w\in S_n(M)} q^{i(w)} = \begin{bmatrix} n \\ a_1,a_2,...,a_m \end{bmatrix}_q$$. We get in particular

$M=\{1^{a_1},2^{a_2}\}$, with $\#M=n \Rightarrow \sum_{w\in S_n(M)} q^k = \sum_{w\in S_n(M)} q^{i(w)} = \begin{bmatrix} n \\ a_1,a_2 \end{bmatrix}_q$, where $k=\psi_2(\overline{w})$.

Similar observations have also been made by D. Foata and G.-N. Han in [4], pages 24-31.

Definition 7.3 (Binary words) [4] Let $BW(N,n)$ denote the set of all words of length $(N+n)$ having exactly N letters equal to 1 and n letters equal to 0. If $x=x_1x_2\cdots x_{N+n}$ is such a word, the inversion number $inv(x)$, is defined as the number of scattered subwords, 10 of the word x.

Example 7.4 We can also write the number of 1's that appear to the left of each letter equal to 0. For the word $x=100101001$ we have
$i(x)=1+1+2+3+3=10$.

In [4] it is shown that $\sum_{x\in BW(N,n)} q^{i(x)} = \begin{bmatrix} N+n \\ n \end{bmatrix}_q$. I will not go through this here in more detail (it can also be found in [3], proposition 1.3.17, page 26), but since we have shown that $i(x)=\psi_2(\overline{x})$ we can make use of this result to get a generating function for Parikh vector statistic of second order on binary words.

Proposition 7.12 [4]

$\begin{bmatrix} N+n \\ n \end{bmatrix}_q = \sum_{x\in BW(N,n)} q^{i(x)}$, where $\begin{bmatrix} N+n \\ n \end{bmatrix}_q$ is the $q-$binomial coefficient)

As a consequence we get $\begin{bmatrix} N+n \\ n \end{bmatrix}_q = \sum_{x\in BW(N,n)} q^{i(x)} = \sum_{\overline{x}\in BW(N,n)} q^k$, where $k=\psi_2(\overline{x})$

8 Summary, Conclusions and Suggestion of Further Research

Theorem 6.12 and **Theorem 6.15** are the main result of this thesis with consequences for Formal language theory (and future Parikh matrix theory) and permutation statistics. The result connects formal languages with permutation statistics in such a way that the results established for descent set statistics (and inversions statistics) is transferred to Parikh matrix theory and thus motivates future research concerning, for example, lattice theory of Parikh matrices and the permutation statistics of various Parikh vector mappings (generalized Parikh vectors). For instance a generalized formula for the Parikh vector statistic $\rho_2(v) = \#\{w \in S_n(M) : \psi_2(w) = v \in \mathbb{N}^{k-1}\}$, where $\sum = \{1,...,k\}_<$, $M = \{1^{n_1},...,k^{n_k}\}$, such that $\#M = \sum_{i=1}^{k} n_i = n$ would be desirable. It would also be interesting to study algorithms for listing the words with a given Parikh vector/matrix. The next step would be to investigate the Parikh vector statistic ρ_2 for arbitrary words $w \in S_n(M)$. The general case, $\rho_2(v) = \#\{w \in S_n(M) : \psi_2(w) = v \in \mathbb{N}^{k-1}\}$, gets more complicated.

References

[1] Alexandru Mateescu, Arto Salomaa, Kai Salomaa, Sheng Yu, *On an Extension of the Parikh Mapping,* TUCS Technical Report No 364, 2003.

[2] Alexandru Mateescu, Arto Salomaa, Sheng Yu, *Subword Histories and Parikh Matrices*, Turku Centre of Computer Science, TUCS Technical Report No 442, 2004.

[3] Richard P. Stanley, *Enumerative Combinatorics Volume 1,* Cambridge University Press,1997.

[4] M.Lothaire, *Combinatorics on words*, Cambridge University Press, 2002, http://www-igm.univ-mlv.fr/%7Eberstel/Lothaire/AlgCWContents.html, 2006.

[5] D. Foata, Guo-Niu Han, *The q-Series in Combinatorics; Permutation Statistics (preliminary version)*, Aug 17, 2004.

[6] Parikh. R.J., *On Context-Free Languages*, J.Assoc. Comp. Mach., 13 (1966) 570-581