

## ENUMERATING SET PARTITIONS BY THE NUMBER OF POSITIONS BETWEEN ADJACENT OCCURRENCES OF A LETTER

*Toufik Mansour, Mark Shattuck, Stephan Wagner*

A *partition*  $\Pi$  of the set  $[n] = \{1, 2, \dots, n\}$  is a collection  $\{B_1, \dots, B_k\}$  of nonempty disjoint subsets of  $[n]$  (called *blocks*) whose union equals  $[n]$ . Suppose that the subsets  $B_i$  are listed in increasing order of their minimal elements and  $\pi = \pi_1 \pi_2 \cdots \pi_n$  denotes the canonical sequential form of a partition of  $[n]$  in which  $i \in B_{\pi_i}$  for each  $i$ . In this paper, we study the generating functions corresponding to statistics on the set of partitions of  $[n]$  with  $k$  blocks which record the total number of positions of  $\pi$  between adjacent occurrences of a letter. Among our results are explicit formulas for the total value of the statistics over all the partitions in question, for which we provide both algebraic and combinatorial proofs. In addition, we supply asymptotic estimates of these formulas, the proofs of which entail approximating the size of certain sums involving the Stirling numbers. Finally, we obtain comparable results for statistics on partitions which record the total number of positions of  $\pi$  of the same letter lying between two letters which are strictly larger.

### 1. INTRODUCTION

A *partition*  $\Pi$  of the set  $[n] = \{1, 2, \dots, n\}$  is a collection  $\{B_1, B_2, \dots, B_k\}$  of nonempty disjoint subsets of  $[n]$  whose union equals  $[n]$ . The elements of a partition are called *blocks*. We assume that  $B_1, B_2, \dots, B_k$  are listed in increasing order of their minimal elements, that is,  $\min B_1 < \min B_2 < \cdots < \min B_k$ . The set of all partitions of  $[n]$  with  $k$  blocks is denoted by  $P(n, k)$  and the set of all partitions of  $[n]$  by  $P(n)$ . The cardinality of  $P(n, k)$  is the well-known Stirling number of the

---

2010 Mathematics Subject Classification. 05A18, 05A16.

Keywords and Phrases. Set partition, Stirling number, partition statistics.

second kind [11], which is usually denoted by  $S_{n,k}$ . The cardinality of  $P(n)$  is given by  $B_n = \sum_{k=0}^n S_{n,k}$ , the  $n^{\text{th}}$  Bell number, which satisfies the recurrence relation

$$(1) \quad B_n = \sum_{j=0}^{n-1} \binom{n-1}{j} B_j, \quad n \geq 1,$$

with initial value  $B_0 = 1$ . Any partition  $\Pi$  can be written in the *canonical sequential form*  $\pi = \pi_1\pi_2 \cdots \pi_n$ , where  $i \in B_{\pi_i}$  for each  $i$  (see, e.g., [4, 8]). From now on, we identify each partition with its canonical sequential form. For example, if  $\Pi = \{1, 4\}, \{2, 5, 7\}, \{3\}, \{6\}$  is a partition of  $[7]$ , then its canonical sequential form is  $\pi = 1231242$  and in such a case we write  $\Pi = \pi$ .

In this paper, we consider statistics on  $P(n, k)$  which record the total number of positions between adjacent occurrences of a letter as well as statistics which record the total number of positions of the same letter lying between two letters which are strictly larger. For other statistics on finite set partitions, see, e.g., [2, 5, 9, 12]. To aid in our analysis, we will represent a partition  $\pi = \pi_1\pi_2 \cdots \pi_n$  of  $[n]$  with exactly  $k$  blocks as a set of points  $(\pi_i, i)$ ,  $i = 1, 2, \dots, n$ , in the lattice  $\mathbb{Z}^2$ . We will call this picture a *graph representation*. For example, the graph representation of the partition  $\pi = 1231242 \in P(7, 4)$  is given below in Figure 1.

In fact, a set  $A$  of points in the first quarter of the lattice  $\mathbb{Z}^2$  is a graph representation for a member of  $P(n, k)$  if  $A$  contains only points of the form  $(j, i)$  such that  $j \leq i$ ,  $j = 1, 2, \dots, k$  and  $i = 1, 2, \dots, n$ , with at least one point on each vertical line and no two points on the same horizontal line.

In this paper, we study the ordinary generating functions corresponding to the aforementioned statistics on  $P(n, k)$ . Several of our results are readily expressed in terms of Stirling numbers of the second kind and Bell numbers. Among the results are explicit formulas for the total value of statistics on  $P(n, k)$  and  $P(n)$ , for which we provide both algebraic and combinatorial proofs. In addition, we also provide asymptotic estimates of these formulas, the proofs of which entail finding the approximate size of certain sums involving the Stirling numbers.

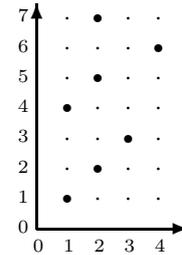


Figure 1.

## 2. MAIN RESULTS

Having drawn the graph representation of a partition  $\pi$  of  $[n]$ , we say that the two points  $(j, i)$  and  $(j, i')$  lying on the vertical line  $x = j$  have  *$j$ -distance  $m$*  if there are  $m$  points in the interior of the subset of the first quadrant of  $\mathbb{Z}^2$  bounded by the line segment between  $(j, i)$  and  $(j, i')$  and the horizontal lines emanating in the positive direction from these points. For instance, the points  $(2, 2)$  and  $(2, 5)$  have distance 1; see Figure 1. We denote the total sum of the

$j$ -distances between any two adjacent points lying on the line  $x = j$  in the graph representation of the partition  $\pi$  by  $d_j(\pi)$ . For example, if  $\pi = 1231242$ , then  $d_1(\pi) = d_2(\pi) = 2$  and  $d_3(\pi) = d_4(\pi) = 0$ ; see Figure 1. Define  $dis(\pi) = \sum_{j \geq 1} d_j(\pi)$  for any partition  $\pi$  of  $[n]$ . In our example, Figure 1, we have  $dis(\pi) = 4$ . In other words, the statistics  $d_j$  (and  $dis$ ) can be defined directly on the canonical sequential form of the partition as follows. Let  $\pi = \pi_1 \cdots \pi_n$  be any partition of  $[n]$  with exactly  $k$  blocks. Two elements  $\pi_i, \pi_{i'}$  have  $j$ -distance  $m$  if  $\pi_i = \pi_{i'} = j$  and  $|\{\pi_s \mid \pi_s > j, i < s < i'\}| = m$ .

Putting this in subword terminology, the statistic  $d_j$  records the number of occurrences of the generalized pattern 1-2-1 in which the 1's correspond to adjacent occurrences of the letter  $j$ , which implies that the  $dis$  statistic records the total number of occurrences of the pattern 1-2-1 in which the 1's correspond to adjacent occurrences of the same letter. For information on the enumeration problem of generalized patterns, see, e.g., [1] (for strings), [7] (permutations), [6] (compositions), or [5] (partitions). The statistics  $d_j$  and  $dis$  also have an interpretation directly in terms of sets. Suppose  $\Pi = \{B_1, B_2, \dots, B_k\} \in P(n, k)$ , where the blocks are arranged by increasing size of minimal elements, and let  $B_j = \{b_1, b_2, \dots, b_r\}$ , with  $b_1 < b_2 < \dots < b_r$ . For each consecutive pair  $b_i, b_{i+1}$ ,  $1 \leq i \leq r - 1$ , consider the number of elements  $c$  in  $[n]$  occurring in blocks of  $\Pi$  to the right of  $B_j$ , with  $b_i < c < b_{i+1}$ . Doing this for each pair  $b_i, b_{i+1}$ , and adding the resulting numbers, yields  $d_j$  and repeating this for all blocks  $B_j$ , and summing over  $j$ , yields  $dis$ .

Let  $F_n(r; q_1, q_2, \dots)$  be the joint generating function for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the statistics  $d_1, d_2, \dots$ ; that is,

$$F_n(r; q_1, q_2, \dots) = \sum_{k \geq 0} \sum_{\pi} r^k \prod_{j \geq 1} q_j^{d_j(\pi)},$$

where the internal sum is over all partitions of  $[n]$  with exactly  $k$  blocks. Suppose that there are  $j + 1$  occurrences of the letter 1 within a member of  $P(n, k)$ , their positions being denoted by  $1, i_1, i_2, \dots, i_j$ . From the definitions, we obtain the recurrence relation

$$(2) \quad F_n(r; q_1, q_2, \dots) = rF_{n-1}(r; q_2, q_3, \dots) + r \sum_{j=1}^{n-1} \left( F_{n-1-j}(r; q_2, q_3, \dots) \sum_{2 \leq i_1 < i_2 < \dots < i_j \leq n} q_1^{i_j - j - 1} \right),$$

with the initial condition  $F_0(r; q_1, q_2, \dots) = 1$ . We must first simplify the rightmost sum in (2) above.

**Lemma 2.1.** Let  $a_{n,j} = \sum_{2 \leq i_1 < i_2 < \dots < i_j \leq n} x^{i_j - 1}$  for all  $0 < j < n$ . Then

$$a_{n,j} = \frac{x^j}{(1-x)^j} - \frac{x^n}{(1-x)^j} \sum_{i=0}^{j-1} \binom{n-1-j+i}{i} (1-x)^i.$$

**Proof.** Direct calculations lead to

$$\begin{aligned} a_{n,j} &= \sum_{2 \leq i_1 < i_2 < \dots < i_{j-1} \leq n-1} \sum_{i_j = i_{j-1} + 1}^n x^{i_j - 1} \\ &= \frac{x}{1-x} \sum_{2 \leq i_1 < i_2 < \dots < i_{j-1} \leq n-1} x^{i_{j-1} - 1} - \frac{x^n}{1-x} \sum_{2 \leq i_1 < i_2 < \dots < i_{j-1} \leq n-1} 1 \\ &= \frac{x}{1-x} a_{n-1, j-1} - \frac{x^n}{1-x} \binom{n-2}{j-1}. \end{aligned}$$

Using the initial condition  $a_{n,1} = \frac{x-x^n}{1-x}$  and iterating the above recurrence, we obtain

$$a_{n,j} = \frac{x^j}{(1-x)^j} - \frac{x^n}{(1-x)^j} \sum_{i=0}^{j-1} \binom{n-1-j+i}{i} (1-x)^i,$$

as claimed. □

Combining Lemma 2.1 and (2) yields

$$(3) \quad F_n(r; q_1, q_2, \dots) = r F_{n-1}(r; q_2, q_3, \dots)$$

$$+ r \sum_{j=1}^{n-1} F_{n-1-j}(r; q_2, q_3, \dots) \left( \frac{1}{(1-q_1)^j} - \sum_{i=0}^{j-1} \binom{n-1-j+i}{i} \frac{q_1^{n-j}}{(1-q_1)^{j-i}} \right),$$

with the initial condition  $F_0(r; q_1, q_2, \dots) = 1$ .

Define  $F(t, r; q_1, q_2, \dots) = \sum_{n \geq 0} F_n(r; q_1, q_2, \dots) t^n$ . Multiplying (3) by  $t^n$  and summing over all  $n \geq 1$ , we get

$$(4) \quad F(t, r; q_1, q_2, \dots) = 1 + \frac{rt}{1-t/(1-q_1)} F(t, r; q_2, q_3, \dots)$$

$$- \frac{rt^2 q_1}{1-q_1} \sum_{n \geq 0} t^n \left[ \sum_{j=0}^n F_{n-j}(r; q_2, q_3, \dots) \left( \sum_{i=0}^j \binom{n-j+i}{i} \frac{q_1^{n-j}}{(1-q_1)^{j-i}} \right) \right]$$

$$\begin{aligned} &= 1 + \frac{rt}{1-t/(1-q_1)} F(t, r; q_2, q_3, \dots) \\ &\quad - \frac{rt^2 q_1}{1-q_1} \sum_{n \geq 0} t^n \left[ \sum_{m=0}^n F_m(r; q_2, q_3, \dots) \left( \sum_{i=0}^{n-m} \binom{m+i}{i} \frac{q_1^m}{(1-q_1)^{n-m-i}} \right) \right] \end{aligned}$$

$$\begin{aligned}
 &= 1 + \frac{rt}{1 - t/(1 - q_1)} F(t, r; q_2, q_3, \dots) \\
 &\quad - \frac{rt^2 q_1}{1 - q_1} \sum_{k \geq 0} \left[ t^k q_1^k F_k(r; q_2, q_3, \dots) \sum_{n \geq k} \frac{t^{n-k}}{(1 - q_1)^{n-k}} \sum_{i=0}^{n-k} \binom{k+i}{i} (1 - q_1)^i \right] \\
 &= 1 + \frac{rt}{1 - t/(1 - q_1)} F(t, r; q_2, q_3, \dots) \\
 &\quad - \frac{rt^2 q_1}{1 - q_1} \sum_{k \geq 0} \left[ t^k q_1^k F_k(r; q_2, q_3, \dots) \sum_{n \geq 0} \frac{t^n}{(1 - q_1)^n} \sum_{i=0}^n \binom{k+i}{i} (1 - q_1)^i \right].
 \end{aligned}$$

Using the fact that  $\sum_{n \geq 0} \frac{t^n}{(1 - q_1)^n} \sum_{i=0}^n \binom{m+i}{i} (1 - q_1)^i = \frac{1}{(1 - t)^{m+1} (1 - t/(1 - q_1))}$ , we obtain the following result.

**Theorem 2.2.** *The generating function  $F(t, r; q_1, q_2, \dots)$  satisfies*

$$\begin{aligned}
 F(t, r; q_1, q_2, \dots) &= 1 + \frac{rt(1 - q_1)}{1 - q_1 - t} F(t, r; q_2, q_3, \dots) \\
 &\quad - \frac{rt^2 q_1}{(1 - t)(1 - q_1 - t)} F\left(\frac{tq_1}{1 - t}, r; q_2, q_3, \dots\right).
 \end{aligned}$$

**2.1. The statistic *dis***

Let  $F(t, r; q) = F(t, r; q, q, \dots)$ . Theorem 2.2 gives

$$F(t, r; q) = 1 + \frac{rt(1 - q)}{1 - q - t} F(t, r; q) - \frac{rt^2 q}{(1 - t)(1 - q - t)} F\left(\frac{tq}{1 - t}, r; q\right),$$

which is equivalent to

$$F(t, r; q) = \frac{1 - q - t}{1 - q - t - rt(1 - q)} - \frac{rt^2 q}{(1 - t)(1 - q - t - rt(1 - q))} F\left(\frac{tq}{1 - t}, r; q\right).$$

Applying this recurrence relation an infinite number of times, we obtain the following result.

**Theorem 2.3.** *The generating function  $F(t, r; q)$  for the number of partitions of  $[n]$  according to the statistic *dis* is given by*

$$\sum_{j \geq 0} \frac{(-1)^j r^j t^{2j} q^{j^2} (1 - q)^j (1 - q - t)}{\prod_{i=1}^j (1 - q - t(1 - q^i)) \prod_{i=0}^j (1 - q - t(1 + rq^i - rq^{i+1}))}.$$

REMARK. Theorem 2.3 for  $q = 1$  implies

$$\begin{aligned}
 F(t, r; 1) &= \sum_{j \geq 0} \lim_{q \rightarrow 1} \left( \frac{(-1)^j r^j t^{2j} q^{j^2} (1-q)^j (1-q-t)}{\prod_{i=1}^j (1-q-t(1-q^i)) \prod_{i=0}^j (1-q-t(1+rq^i-rq^{i+1}))} \right) \\
 &= \sum_{j \geq 0} r^j t^j \lim_{q \rightarrow 1} \frac{1}{\prod_{i=1}^j \left( 1-t \lim_{q \rightarrow 1} \frac{1-q^i}{1-q} \right)} = \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=1}^j (1-it)},
 \end{aligned}$$

which gives us the familiar fact (see, e.g., p. 34 of [10]) that the generating function for the number partitions of  $[n]$  with exactly  $k$  blocks is given by

$$\frac{t^k}{\prod_{i=1}^k (1-it)} = \sum_{n \geq k} S_{n,k} t^n,$$

where  $S_{n,k}$  is the Stirling number of the second kind.

When  $q = 0$  in Theorem 2.3, we get

$$F(t, r; 0) = \frac{1-t}{1-t-rt} = \sum_{j \geq 0} \left( \frac{rt}{1-t} \right)^j,$$

which implies that there are  $\binom{n-1}{k-1}$  members of  $P(n, k)$  with zero *dis* if  $n \geq k \geq 1$  and hence  $2^{n-1}$  members of  $P(n)$  with zero *dis* if  $n \geq 1$ . This also follows directly from the definitions since partitions  $\pi = \pi_1 \pi_2 \cdots \pi_n$  for which  $dis(\pi) = 0$  must be non-decreasing.

**Corollary 2.4.** *The generating function for the total dis over all the partitions of  $[n]$  is given by*

$$t \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=0}^j (1-it)} \left( \sum_{i=2}^j \frac{\binom{i}{2}}{1-it} \right).$$

**Proof.** Let  $F'(t, r; 1) = \frac{d}{dq} F(t, r; q) |_{q=1}$ . Theorem 2.3 implies

$$F'(t, r; 1) = \sum_{j \geq 0} \lim_{q \rightarrow 1} \frac{(-1)^j r^j t^{2j} q^{j^2} (1-q)^j (1-q-t)}{\prod_{i=1}^j (1-q-t(1-q^i)) \prod_{i=0}^j (1-q-t(1+rq^i-rq^{i+1}))} V_j(t; q),$$

where

$$V_j(t, r; q) = \frac{j^2}{q} - \frac{1}{1-q-t} - \frac{j - \sum_{i=1}^j \frac{1-iq^{i-1}t}{1-t\frac{1-q^i}{1-q}}}{1-q} + \sum_{i=0}^j \frac{1+rq^{i-1}t(i-(i+1)q)}{1-q-t(1+rq^i-rq^{i+1})}.$$

By the preceding remark, we get

$$(5) \quad F'(t, r; 1) = \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=0}^j (1-it)} \lim_{q \rightarrow 1} V_j(t, r; q).$$

On the other hand,

$$\begin{aligned} \lim_{q \rightarrow 1} V_j(t, r; q) &= j^2 + r(j+1) - \frac{j}{t} - \lim_{q \rightarrow 1} \frac{j - \sum_{i=1}^j \frac{1-iq^{i-1}t}{1-t\frac{1-q^i}{1-q}}}{1-q} \\ &= j^2 + r(j+1) - \frac{j}{t} - \sum_{i=1}^j \lim_{q \rightarrow 1} \frac{d}{dq} \left( \frac{1-iq^{i-1}t}{1-t\frac{1-q^i}{1-q}} \right) \\ &= j^2 + r(j+1) - \frac{j}{t} + \sum_{i=1}^j \frac{\binom{i}{2}t}{1-it}. \end{aligned}$$

Hence, by (5),

$$\begin{aligned} F'(t, r; 1) &= \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=0}^j (1-it)} \left( j^2 + r(j+1) - \frac{j}{t} + t \sum_{i=1}^j \frac{\binom{i}{2}}{1-it} \right) \\ &= \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=0}^j (1-it)} \left( j^2 + r(j+1) - \frac{r(j+1)}{1-(j+1)t} + t \sum_{i=1}^j \frac{\binom{i}{2}}{1-it} \right) \\ &= \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=0}^j (1-it)} \left( j^2 - \frac{r(j+1)^2 t}{1-(j+1)t} + t \sum_{i=1}^j \frac{\binom{i}{2}}{1-it} \right) \\ &= t \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=0}^j (1-it)} \left( \sum_{i=1}^j \frac{\binom{i}{2}}{1-it} \right), \end{aligned}$$

as claimed. □

The above corollary, together with the fact that

$$\frac{t^k}{\prod_{i=1}^k (1-it)} = \sum_{n \geq k} S_{n,k} t^n,$$

yields the following result.

**Corollary 2.5.** *The total dis over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$\sum_{i=2}^k \binom{i}{2} \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k},$$

for all  $n \geq k \geq 1$ .

This formula can be simplified; for this and later purposes, we convert the ordinary generating function to an exponential generating function, which has several advantages. This can be achieved by means of a partial fraction decomposition: one has

$$\begin{aligned} t \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=1}^j (1-it)} \binom{j}{2} \frac{1}{1-it} &= \sum_{j \geq 0} \frac{r^j}{\prod_{i=1}^j (t^{-1}-i)} \binom{j}{2} \frac{1}{t^{-1}-i} \\ &= \sum_{j \geq 1} \left( \frac{a_j(r)}{(t^{-1}-j)^2} + \frac{b_j(r)}{t^{-1}-j} \right) \end{aligned}$$

for certain  $a_j(r)$  and  $b_j(r)$  that depend only on  $r$ . In order to determine these values, we need the expansion around  $u = t^{-1} = m$  for fixed  $m$ : for  $j \geq m$ , one has

$$\begin{aligned} &\frac{r^j}{\prod_{i=1}^j (t^{-1}-i)} \binom{j}{2} \frac{1}{t^{-1}-i} \\ &= \frac{r^j}{u-m} \prod_{\substack{i=1 \\ i \neq m}}^j (u-i)^{-1} \left( \frac{\binom{m}{2}}{u-m} + \sum_{\substack{i=1 \\ i \neq m}}^j \frac{\binom{i}{2}}{m-i} + O(u-m) \right) \\ &= \frac{r^j}{u-m} \prod_{\substack{i=1 \\ i \neq m}}^j (m-i)^{-1} \cdot \prod_{\substack{i=1 \\ i \neq m}}^j \left( 1 + \frac{u-m}{m-i} \right)^{-1} \cdot \left( \frac{\binom{m}{2}}{u-m} + \sum_{\substack{i=1 \\ i \neq m}}^j \frac{\binom{i}{2}}{m-i} + O(u-m) \right) \end{aligned}$$

$$\begin{aligned}
&= \frac{r^j (-1)^{j-m}}{(m-1)!(j-m)!(u-m)} \cdot \left( 1 - \sum_{\substack{i=1 \\ i \neq m}}^j \frac{u-m}{m-i} + O((u-m)^2) \right) \\
&\quad \cdot \left( \frac{\binom{m}{2}}{u-m} + \sum_{\substack{i=1 \\ i \neq m}}^j \frac{\binom{i}{2}}{m-i} + O(u-m) \right) \\
&= \frac{r^j (-1)^{j-m}}{(m-1)!(j-m)!(u-m)} \cdot \left( \frac{\binom{m}{2}}{u-m} + \sum_{\substack{i=1 \\ i \neq m}}^j \frac{\binom{i}{2} - \binom{m}{2}}{m-i} + O(u-m) \right) \\
&= \frac{r^j (-1)^{j-m}}{(m-1)!(j-m)!(u-m)} \cdot \left( \frac{\binom{m}{2}}{u-m} - \frac{j^2 - j + 2 + 2mj - 4m}{4} + O(u-m) \right) \\
&= \frac{r^j (-1)^{j-m} \binom{m}{2}}{(m-1)!(j-m)!(u-m)^2} - \frac{r^j (-1)^{j-m} (j^2 - j + 2 + 2mj - 4m)}{4(m-1)!(j-m)!(u-m)} + O(1).
\end{aligned}$$

Hence, we have

$$a_m(r) = \sum_{j \geq m} \frac{r^j (-1)^{j-m} \binom{m}{2}}{(m-1)!(j-m)!} = \frac{r^m \binom{m}{2}}{(m-1)!} \sum_{k \geq 0} \frac{(-r)^k}{k!} = \frac{r^m \binom{m}{2} e^{-r}}{(m-1)!},$$

and, similarly,

$$b_m(r) = -\frac{r^m e^{-r} (r^2 - 4mr + 3m^2 - 5m + 2)}{4(m-1)!}.$$

Now we can pass on easily from the ordinary generating function

$$\sum_{j \geq 1} \left( \frac{a_j(r)}{(t^{-1} - j)^2} + \frac{b_j(r)}{t^{-1} - j} \right) = \sum_{j \geq 1} a_j(r) \sum_{k \geq 0} (k+1) j^k t^{k+2} + \sum_{j \geq 1} b_j(r) \sum_{k \geq 0} j^k t^{k+1}$$

to the exponential generating function

$$\begin{aligned}
&\sum_{j \geq 1} a_j(r) \sum_{k \geq 0} \frac{(k+1) j^k t^{k+2}}{(k+2)!} + \sum_{j \geq 1} b_j(r) \sum_{k \geq 0} \frac{j^k t^{k+1}}{(k+1)!} \\
&= \sum_{j \geq 1} a_j(r) \frac{e^{jt} (jt - 1) + 1}{j^2} + \sum_{j \geq 1} b_j(r) \frac{e^{jt} - 1}{j}.
\end{aligned}$$

This can be simplified to obtain

$$\begin{aligned}
 & \sum_{j \geq 1} a_j(r) \frac{e^{jt}(jt-1)+1}{j^2} + \sum_{j \geq 1} b_j(r) \frac{e^{jt}-1}{j} \\
 &= \sum_{j \geq 1} \frac{r^j \binom{j}{2} e^{-r}}{(j-1)!} \cdot \frac{e^{jt}(jt-1)+1}{j^2} - \sum_{j \geq 1} \frac{r^j e^{-r}(r^2-4jr+3j^2-5j+2)}{4(j-1)!} \cdot \frac{e^{jt}-1}{j} \\
 &= t \sum_{j \geq 2} \frac{r^j e^{-r} e^{jt}}{2(j-2)!} - \sum_{j \geq 1} \frac{(j-1)r^j e^{-r}(e^{jt}-1)}{2j!} \\
 &\quad - \sum_{j \geq 1} \frac{(r^2-4jr+3j^2-5j+2)r^j e^{-r}(e^{jt}-1)}{4j!} \\
 &= \frac{tr^2 e^{2t}}{2} e^{r(e^t-1)} - \sum_{j \geq 1} \frac{(r^2-4jr+3j^2-3j)r^j e^{-r}(e^{jt}-1)}{4j!} \\
 &= \frac{tr^2 e^{2t}}{2} e^{r(e^t-1)} - \sum_{j \geq 0} \frac{r^{j+2} e^{-r}(e^{jt}-1)}{4j!} \\
 &\quad + \sum_{j \geq 1} \frac{r^{j+1} e^{-r}(e^{jt}-1)}{(j-1)!} - \sum_{j \geq 2} \frac{3r^j e^{-r}(e^{jt}-1)}{4(j-2)!} \\
 &= \frac{tr^2 e^{2t}}{2} e^{r(e^t-1)} - \frac{r^2}{4} (e^{r(e^t-1)} - 1) + r^2 (e^t e^{r(e^t-1)} - 1) - \frac{3r^2}{4} (e^{2t} e^{r(e^t-1)} - 1) \\
 &= \frac{r^2}{4} e^{r(e^t-1)} ((2t-3)e^{2t} + 4e^t - 1).
 \end{aligned}$$

Extracting coefficients implies that the total *dis* over  $P(n, k)$  is given by

$$\frac{k(k-1)(2n-3k)}{4} S_{n-1,k} + \frac{(k-1)(4n-5k+2)}{4} S_{n-1,k-1} + \frac{n-k+1}{2} S_{n-1,k-2}.$$

Simplifying this, using the recurrence  $S_{n,k} = kS_{n-1,k} + S_{n-1,k-1}$  several times, yields the following result.

**Theorem 2.6.** *The total dis over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$\frac{(n+1-k)S_{n+1,k} - \left[ \binom{k}{2} + n \right] S_{n,k}}{2},$$

for all  $n \geq k \geq 1$ .

It is well known that  $S_{n,k} = \frac{k^n}{k!} + O((k-1)^n)$  for fixed  $k$ ; hence one has the following immediate corollary.

**Corollary 2.7.** *For fixed  $k$ , the total dis over all the partitions of  $[n]$  with exactly  $k$  blocks is asymptotically*

$$\frac{(k-1)k^n}{4k!}(2n-3k) + O(n(k-1)^n),$$

and the average dis over all such partitions is therefore asymptotically

$$\frac{(k-1)(2n-3k)}{4} + O\left(n\left(\frac{k-1}{k}\right)^n\right).$$

For the total over all partitions of  $[n]$ , without restrictions as to the number of blocks, one simply has to set  $r = 1$  to find the exponential generating function

$$\frac{1}{4}e^{e^t-1}((2t-3)e^{2t} + 4e^t - 1),$$

and again extracting coefficients yields the following result.

**Theorem 2.8.** *The total dis over all the partitions of  $[n]$  is given by*

$$\frac{2n+7}{4}B_{n+1} - \frac{3}{4}B_{n+2} - \frac{2n+1}{4}B_n.$$

Since  $\frac{B_{n+1}}{B_n} = \frac{n}{\log n} + O\left(\frac{n \log \log n}{\log^2 n}\right)$  (see [3]), one obtains an asymptotic formula for the mean.

**Corollary 2.9.** *The average dis over all the partitions of  $[n]$  is asymptotically*

$$\frac{n^2}{2 \log n} + O\left(\frac{n^2 \log \log n}{\log^2 n}\right).$$

### 2.2. The statistic $m$ -distance

Our aim in this subsection is to study the generating function  $G(t, r; q) = F(t, r; q, 1, 1, \dots)$ , which is the generating function for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the statistic 1-distance. Theorem 2.2 gives

$$G(t, r; q) = 1 + \frac{rt(1-q)}{1-q-t}F(t, r; 1) - \frac{rt^2q}{(1-t)(1-q-t)}F\left(\frac{tq}{1-t}, r; 1\right).$$

By the  $q = 1$  case, we get

$$G(t, r; q) = 1 + \frac{rt(1-q)}{1-q-t} \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=1}^j (1-it)} - \frac{rt^2q}{1-q-t} \sum_{j \geq 0} \frac{r^j t^j q^j}{\prod_{i=0}^j (1-t(1+iq))},$$

which implies the following result.

**Theorem 2.10.** *The generating function for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the statistic 1-distance is given by*

$$G(t, r; q) = 1 + \frac{rt}{1 - q - t} \sum_{j \geq 0} r^j t^j \left[ \frac{1 - q}{\prod_{i=1}^j (1 - it)} - \frac{tq^{j+1}}{\prod_{i=0}^j (1 - t(1 + iq))} \right].$$

For instance, the generating function for the number of partitions  $\pi$  of  $[n]$  with exactly  $k$  blocks and  $d_1(\pi) = 0$  is given by

$$G(t, r; 0) = 1 + \frac{rt}{1 - t} \sum_{j \geq 0} \frac{r^j t^j}{\prod_{i=1}^j (1 - it)},$$

which implies that the number of partitions of  $[n]$  with exactly  $k$  blocks and having zero  $d_1$  is given by  $\sum_{j=k-1}^{n-1} S_{j,k-1}$  if  $n \geq k \geq 1$ . Thus, the number of partitions of  $[n]$  with zero  $d_1$  is given by  $\sum_{j=0}^{n-1} B_j$  if  $n \geq 1$ , where  $B_j$  is the  $j^{th}$  Bell number. This is also apparent from the definitions since members of  $P(n, k)$  with zero  $d_1$  must start with  $n - j$  1's for some  $j$ ,  $k - 1 \leq j \leq n - 1$ , and have no other occurrences of 1.

Differentiating the generating function  $G(t, r; q)$  with respect to  $q$ , setting  $q = 1$ , and finding the coefficient of  $r^k$  yields

$$(6) \quad [r^k] \left( \frac{d}{dq} G(t, r; q) \Big|_{q=1} \right) = \frac{t^k}{\prod_{i=0}^k (1 - it)} \sum_{i=1}^k \frac{(i - 1)t}{1 - it},$$

which implies the following result.

**Corollary 2.11.** *The total  $d_1$  over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$\sum_{i=2}^k \left( (i - 1) \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right),$$

for all  $n \geq k \geq 1$ .

Corollary 2.11 may be extended to find the total  $d_m$  over all the partitions of  $[n]$  with exactly  $k$  blocks as follows. At first, delete all the letters  $1, 2, \dots, m - 1$  from the partition  $\pi$  and denote the resulting partition by  $\pi'$ . Then counting  $d_1(\pi')$  over all possible partitions  $\pi'$  yields the following result.

**Corollary 2.12.** *For all  $n > k \geq 1$ , the total  $d_m$  over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$\sum_{i_1=0}^{n-j_1} \sum_{i_2=0}^{n-j_2} \cdots \sum_{i_{m-1}=0}^{n-j_{m-1}} \left[ \prod_{s=1}^{m-1} \binom{n-j_s}{i_s} \sum_{i=2}^{k+1-m} \binom{i-1}{j} \sum_{j=k+1-m}^{n-j_m} i^{n-j_m-j} S_{j,k+1-m} \right],$$

where  $j_s = i_1 + \cdots + i_{s-1} + s$ .

As in the case of the statistic  $dis$ , it is possible to derive a simpler formula for  $d_1$  as well as its asymptotic behavior by means of exponential generating functions. In a totally analogous manner, one obtains the partial fraction decomposition

$$\sum_{k \geq 1} \frac{r^k t^k}{\prod_{i=1}^k (1-it)} \sum_{i=1}^k \frac{(i-1)t}{1-it} = \sum_{j \geq 2} \frac{r^j e^{-r}}{(j-2)!(t^{-1}-j)^2} - \sum_{j \geq 1} \frac{r^j e^{-r} (j-r-1)}{(j-1)!(t^{-1}-j)},$$

and as an immediate consequence the exponential generating function

$$\sum_{j \geq 2} \frac{r^j e^{-r}}{(j-2)!} \cdot \frac{e^{jt}(jt-1)+1}{j^2} - \sum_{j \geq 1} \frac{r^j e^{-r} (j-r-1)}{(j-1)!} \cdot \frac{e^{jt}-1}{j}.$$

Since the sums do not evaluate directly to elementary functions, we differentiate with respect to  $t$  (which merely means a shift of coefficients) to obtain, after some simplifications,

$$\sum_{j \geq 2} \frac{r^j e^{-r}}{(j-2)!} \cdot t e^{jt} - \sum_{j \geq 1} \frac{r^j e^{-r} (j-r-1)}{(j-1)!} \cdot e^{jt} = r^2 e^t (te^t - e^t + 1) e^{r(e^t-1)}.$$

Extracting coefficients implies that the total  $d_1$  over  $P(n, k)$  is given by

$$k(k-1)(n-k-1)S_{n-2,k} + (k-1)(2n-2k-1)S_{n-2,k-1} + (n-k)S_{n-2,k-2},$$

which may be simplified using the recurrence for  $S(n, k)$  to obtain the following result.

**Theorem 2.13.** *The total  $d_1$  over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$(n-k)S_{n,k} - (n-1)S_{n-1,k},$$

for all  $n \geq k \geq 1$ .

Furthermore, one has the following asymptotic result which parallels Corollary 2.7.

**Corollary 2.14.** *For fixed  $k$ , the total  $d_1$  over all the partitions of  $[n]$  with exactly  $k$  blocks is asymptotically*

$$\frac{(k-1)(n-k-1)k^{n-1}}{k!} + O(n(k-1)^n),$$

and the average  $d_1$  over all such partitions is therefore asymptotically

$$\frac{(k-1)(n-k-1)}{k} + O\left(n\left(\frac{k-1}{k}\right)^n\right).$$

From the exponential generating function

$$e^t(te^t - e^t + 1)e^{e^t-1}$$

that results when  $r = 1$ , one also obtains the following theorem for the total  $d_1$  over all partitions of  $[n]$  which parallels Theorem 2.8.

**Theorem 2.15.** *The total  $d_1$  over all the partitions of  $[n]$  is given by*

$$(n+1)B_n - B_{n+1} - (n-1)B_{n-1}.$$

**Corollary 2.16.** *The average  $d_1$  over all the partitions of  $[n]$  is asymptotically  $n + O(n/\log n)$ .*

### 2.3. Combinatorial proofs

In this section, we provide direct combinatorial proofs of Corollaries 2.5 and 2.11 and Theorems 2.13 and 2.15. To do so, we first consider some additional statistics on  $P(n, k)$ . Let  $\pi = \pi_1\pi_2 \cdots \pi_n \in P(n, k)$ . Define the statistic  $\ell$  by  $\ell(\pi) := t - 1$ , where  $t$  denotes the position in  $\pi$  corresponding to the rightmost 1,  $1 \leq t \leq n$ . The total  $\ell$  on  $P(n, k)$  is easily computed.

**Lemma 2.17.** *The total  $\ell$  over all the partitions of  $[n]$  with exactly  $k$  blocks is given by  $(n-k)S_{n,k}$  for all  $n \geq k \geq 1$ .*

**Proof.** Assume  $k \geq 2$  and induct on  $n \geq k$ , the  $n = k$  case clear. If  $n > k$  and  $\lambda = \lambda_1\lambda_2 \cdots \lambda_n \in P(n, k)$ , then considering the cases (i)  $\lambda_n = 1$ ; (ii)  $\lambda_n > 1$ , with  $\lambda_1\lambda_2 \cdots \lambda_{n-1} \in P(n-1, k)$ ; or (iii)  $\lambda_n = k$ , with  $\lambda_1\lambda_2 \cdots \lambda_{n-1} \in P(n-1, k-1)$ , implies that the total  $\ell$  value over  $P(n, k)$  is given by

$$\begin{aligned} & (n-1)S_{n-1,k} + (k-1)(n-1-k)S_{n-1,k} + (n-k)S_{n-1,k-1} \\ & = (n-k)(S_{n-1,k-1} + kS_{n-1,k}) = (n-k)S_{n,k}, \end{aligned}$$

which completes the induction.  $\square$

We define two more statistics as follows. If  $\pi = \pi_1\pi_2 \cdots \pi_n \in P(n, k)$ , then let  $\alpha(\pi)$  denote the number of distinct elements of  $\{2, 3, \dots, k\}$  which occur to the left of the rightmost 1 in  $\pi$  and let  $\beta(\pi)$  denote the number of positions to the right of the rightmost 1 in  $\pi$  and not corresponding to the initial occurrence of a letter. For example, if  $\pi = 1232134252 \in P(10, 5)$ , then  $\alpha(\pi) = 2$  and  $\beta(\pi) = 3$ .

**Lemma 2.18.** *The total  $\alpha$  over all the partitions of  $[n]$  with exactly  $k$  blocks is equal to the total  $\beta$  for all  $n \geq k \geq 1$ .*

**Proof.** If  $a(\pi)$  denotes the number of distinct elements of  $[k]$  occurring to the left of (and including) the rightmost 1 in  $\pi \in P(n, k)$ , then  $\alpha(\pi) = a(\pi) - 1$  and  $\beta(\pi) = (n - k) - (\ell(\pi) + 1 - a(\pi))$ , from the definitions. Thus, the total  $\alpha$  over  $P(n, k)$  equals the total  $\beta$  if and only if

$$\sum_{\pi \in P(n, k)} (a(\pi) - 1) = \sum_{\pi \in P(n, k)} (n - k - (\ell(\pi) + 1 - a(\pi))),$$

which reduces to  $\sum_{\pi \in P(n, k)} \ell(\pi) = (n - k)S_{n, k}$ , the statement of the previous lemma.  $\square$

We now provide a direct proof of Corollary 2.11 as follows. First let  $\gamma(\pi)$  denote the total number of positions in  $\pi \in P(n, k)$  corresponding to letters greater than 1 and not corresponding to an initial occurrence of a letter. Then the total  $\gamma$  over  $P(n, k)$  is given by

$$\sum_{i=2}^k \left( (i - 1) \sum_{j=k}^{n-1} i^{n-1-j} S_{j, k} \right).$$

To see this, suppose  $i$  and  $j$  are given, where  $2 \leq i \leq k \leq j \leq n - 1$ , and consider those members of  $P(n, k)$  which may be expressed uniquely as

$$(7) \quad \pi = \pi'ixy,$$

where  $\pi'$  is a partition with  $i - 1$  blocks,  $x$  is a non-empty word in the alphabet  $[i]$  of length  $n - j$  whose last letter is greater than 1, and  $y$  is a possibly empty word. Note that there are  $(i - 1)i^{n-1-j}$  choices for the word  $x$  and  $S_{j, k}$  choices for the remaining letters  $\pi'iy$  which together constitute a partition of a  $j$ -element set into  $k$  blocks. Thus, the total  $\gamma$  over  $P(n, k)$  may be obtained by finding the number of partitions which may be expressed as in (7) for each  $i$  and  $j$  and then summing over all possible values of  $i$  and  $j$ , which yields the expression above.

Furthermore, from the definitions and Lemma 2.18, we may write

$$t(\gamma) = t(d_1) - t(\alpha) + t(\beta) = t(d_1),$$

where  $t$  of a statistic denotes its total taken over  $P(n, k)$ , which implies that the total  $d_1$  over  $P(n, k)$  is as in Corollary 2.11.  $\square$

This proof may be extended to explain Corollary 2.5 as follows. We first generalize the prior statistics. For each  $i \in [k - 1]$ , let  $\alpha_i(\pi)$  denote the number of distinct elements of the set  $\{i + 1, i + 2, \dots, k\}$  which occur to the left of the rightmost  $i$  in  $\pi = \pi_1\pi_2 \cdots \pi_n \in P(n, k)$  and let  $\beta_i(\pi)$  denote the number of positions to the right of the rightmost  $i$  in  $\pi$  that do not correspond to an initial occurrence of a letter but are larger than  $i$ . Note that  $\alpha_i$  and  $\beta_i$  are  $\alpha$  and  $\beta$  when  $i = 1$ . For each  $i$ , note that the total  $\alpha_i$  over  $P(n, k)$  equals the total  $\beta_i$ , by Lemma 2.18, upon deleting all occurrence of  $1, 2, \dots, i - 1$  within members of  $P(n, k)$  and

then considering  $\alpha_1$  and  $\beta_1$  on the resulting partitions, much like as in the proof of Corollary 2.12 above.

Given  $i \in [k - 1]$  and  $\pi \in P(n, k)$ , let  $\gamma_i(\pi)$  denote the number of positions of  $\pi$  greater than  $i$  and not corresponding to an initial occurrence of a letter. Then

$$(8) \quad \sum_{i=1}^{k-1} t(\gamma_i) = \sum_{i=2}^k \left( \binom{i}{2} \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right).$$

To see this, first fix  $i$  and  $j$ , where  $2 \leq i \leq k \leq j \leq n - 1$ , and consider the set of all ordered pairs  $(\pi, a)$ , where  $\pi = \pi'ixy$  is a member of  $P(n, k)$  which may be expressed as in (7) above and  $a$  is a positive integer strictly less than the last letter of  $x$ . Then the right side of (8) gives the total number of ordered pairs as  $i$  and  $j$  vary over all possible values. By definition and similar reasoning,  $t(\gamma_i)$  equals the total number of ordered pairs  $(\pi, a)$  where  $a = i$  for each  $i \in [k - 1]$ , which gives (8).

From the definitions and the fact that  $t(\alpha_i) = t(\beta_i)$ , we may write for each  $i \in [k - 1]$ ,

$$t(\gamma_i) = t(d_i) - t(\alpha_i) + t(\beta_i) = t(d_i),$$

which implies

$$t(dis) = \sum_{i=1}^{k-1} t(d_i) = \sum_{i=1}^{k-1} t(\gamma_i),$$

from which Corollary 2.5 follows from (8). □

Finally, this proof can be extended to explain Theorems 2.13 and 2.15 as follows. Given  $\pi = \pi_1\pi_2 \cdots \pi_n \in P(n, k)$ , let  $b(\pi)$  denote the number of positions  $i > 1$  such that  $\pi_i = 1$ . Then the total  $b$  over all members of  $P(n, k)$  is  $(n - 1)S_{n-1,k}$ , as there are  $S_{n-1,k}$  occurrences of 1 in the  $i^{th}$  position, taken over  $P(n, k)$ , for each  $i, 2 \leq i \leq n$ . Since there are clearly  $(n - k)S_{n,k}$  positions in all of the members of  $P(n, k)$  which do not correspond to initial occurrences of letters, we see from the definition that

$$t(\gamma) = (n - k)S_{n,k} - (n - 1)S_{n-1,k},$$

upon subtracting all non-initial 1's. However, we have  $t(\gamma) = t(d_1)$ , from the proof of Corollary 2.11 above. Summing the expression in Theorem 2.13 over  $k$  yields

Theorem 2.15, upon noting  $\sum_{k=1}^n kS(n, k) = B_{n+1} - B_n$ . □

REMARK. We may give another expression for the total  $d_1$  over  $P(n, k)$  as follows. Suppose that  $\pi \in P(n, k)$  is to contain exactly  $i + 1$  1's for some  $i, 0 \leq i \leq n - k$ , and that exactly  $j$  letters greater than 1 come after the final occurrence of 1 for some  $j, 0 \leq j \leq n - i - 1$ . If  $j < n - 1 - i$ , then we may write  $\pi = 1\alpha 1\beta$ , where  $\alpha$  is non-empty and contains  $i - 1$  1's and  $n - 1 - i - j$  letters greater than 1 and  $\beta$  is a possibly empty word having length  $j$ . There are then  $\binom{n - 2 - j}{i - 1} S_{n-1-i,k-1}$  such partitions, each having a  $d_1$  value of  $n - 1 - i - j$ .

Summing over all possible  $i$  and  $j$  implies that the total  $d_1$  over  $P(n, k)$  is given by

$$\begin{aligned} \sum_{i=0}^{n-k} S_{n-1-i, k-1} \sum_{j=0}^{n-2-i} \binom{n-2-j}{i-1} (n-1-i-j) &= \sum_{i=0}^{n-k} S_{n-1-i, k-1} \sum_{j=1}^{n-1-i} \binom{i-1+j}{j} j \\ &= \sum_{i=1}^{n-k} i S_{n-1-i, k-1} \binom{n-1}{i+1}. \end{aligned}$$

Equating this with the expression in Theorem 2.13 for the total  $d_1$  over  $P(n, k)$  yields the following recurrence for Stirling numbers:

$$(n-k)S_{n,k} - (n-1)S_{n-1,k} = \sum_{i=1}^{n-k} i S_{n-1-i, k-1} \binom{n-1}{i+1}, \quad n \geq k \geq 1.$$

### 3. OTHER STATISTICS

Suppose  $\pi = \pi_1\pi_2 \cdots \pi_n$  is represented graphically as in Figure 1 above. We will call a point  $P$  in the graph of  $\pi$  *internal* if there exist points in the graph both above and below it lying to its right. In the partition pictured in Figure 1, only the points (1, 4) and (2, 5) are internal. Note that a point  $(\pi_i, i)$  belonging to the graph of a partition  $\pi = \pi_1\pi_2 \cdots \pi_n$  is internal if and only if there exist  $j$  and  $k$  with  $j < i < k$  such that  $\pi_i < \min\{\pi_j, \pi_k\}$ . We will call the corresponding letter  $\pi_i$  in  $\pi$  an *internal letter*. Given  $\pi \in P(n, k)$  and  $m \in [k-1]$ , let  $int_m(\pi)$  record the number of internal points of  $\pi$  whose  $x$ -coordinate is  $m$ . Then  $int(\pi) := \sum_{m=1}^{k-1} int_m(\pi)$  records the total number of internal points of  $\pi$ . For example, if  $\pi = 123214331431 \in P(12, 4)$ , then  $int_1(\pi) = 2$ ,  $int_2(\pi) = 1$ ,  $int_3(\pi) = 2$ , and  $int(\pi) = 5$ .

Let  $F_n(r; q_1, q_2, \dots)$  be the joint generating function for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the statistics  $int_1, int_2, \dots$ ; that is,

$$F_n(r; q_1, q_2, \dots) = \sum_{k \geq 0} \sum_{\pi} r^k \prod_{j \geq 1} q_j^{int_j(\pi)},$$

where the internal sum is over all partitions of  $[n]$  with exactly  $k$  blocks.

If  $n \geq 2$ , suppose that there are to be a total of  $j+1$  1's in a partition of  $[n]$ , where  $1 \leq j \leq n-1$ . If  $j = n-1$ , then there is only one possible partition, so assume  $j \leq n-2$ . Suppose that  $j-k+1$  of the 1's appear before the first 2 and  $i$  of the 1's are internal (so the partition ends in a run of exactly  $k-i$  1's). Then the  $i$  1's are to be distributed in  $n-j-2$  positions in  $\binom{n-j-3+i}{i}$  possible ways. Thus, the total  $int_1$  weight contribution is

$$\sum_{k=0}^j \sum_{i=0}^k q_1^i \binom{n-j-3+i}{i} = \sum_{i=0}^j q_1^i \binom{n-j-3+i}{i} (j-i+1).$$

If  $n \geq 2$ , then we may write

$$(9) \quad F_n(r; q_1, q_2, \dots) = r + rF_{n-1}(r; q_2, q_3, \dots) + r \sum_{j=1}^{n-2} F_{n-1-j}(r; q_2, q_3, \dots) \sum_{i=0}^j q_1^i \binom{n-j-3+i}{i} (j-i+1).$$

Define  $F(t, r; q_1, q_2, \dots) = \sum_{n \geq 0} F_n(r; q_1, q_2, \dots) t^n$ . Multiplying both sides of (9) by  $t^n$  and summing over  $n \geq 2$  yields

$$(10) \quad F(t, r; q_1, q_2, \dots) = 1 + \frac{rt^2}{1-t} + rtF(t, r; q_2, q_3, \dots) + r \sum_{n \geq 3} t^n \left( \sum_{j=1}^{n-2} F_{n-1-j}(r; q_2, q_3, \dots) \sum_{i=0}^j q_1^i \binom{n-j-3+i}{i} (j-i+1) \right).$$

Now

$$(11) \quad r \sum_{n \geq 3} t^n \left( \sum_{j=1}^{n-2} F_{n-1-j}(r; q_2, q_3, \dots) \sum_{i=0}^j q_1^i \binom{n-3-j+i}{i} (j+1-i) \right) = rt^3 \sum_{n \geq 0} t^n \left( \sum_{j=0}^n F_{n+1-j}(r; q_2, q_3, \dots) \sum_{i=0}^{j+1} q_1^i \binom{n-1-j+i}{i} (j+2-i) \right) = rt^3 \sum_{n \geq 0} t^n \left( \sum_{j=1}^{n+1} F_j(r; q_2, q_3, \dots) \sum_{i=0}^{n-j+2} q_1^i \binom{j-2+i}{i} (n+3-j-i) \right) = rt^2 \sum_{j \geq 1} t^j F_j(r; q_2, q_3, \dots) \left( \sum_{n \geq 0} t^n \sum_{i=0}^{n+1} q_1^i \binom{j-2+i}{i} (n+2-i) \right).$$

Observe that

$$\begin{aligned} \sum_{n \geq 0} t^n \sum_{i=0}^{n+1} q_1^i \binom{j-2+i}{i} (n+2-i) &= \sum_{n \geq 0} t^n q_1^{n+1} \binom{j-1+n}{n+1} + \sum_{i \geq 0} t^i q_1^i \binom{j-2+i}{i} \sum_{n \geq 0} t^n (n+2) \\ &= \frac{1}{t} \left( \frac{1}{(1-tq_1)^{j-1}} - 1 \right) + \left( \frac{1}{(1-tq_1)^{j-1}} \right) \left( \frac{2-t}{(1-t)^2} \right), \end{aligned}$$

so that (10) and (11) imply

$$\begin{aligned} F(t, r; q_1, q_2, \dots) &= 1 + \frac{rt^2}{1-t} + rtF(t, r; q_2, q_3, \dots) \\ &\quad - rt(F(t, r; q_2, q_3, \dots) - 1) + \frac{rt(1-tq_1)}{(1-t)^2} \left( F\left(\frac{t}{1-tq_1}, r; q_2, q_3, \dots\right) - 1 \right), \end{aligned}$$

which yields the following recurrence for the generating function

$$H(t, r; q_1, q_2, \dots) := F(t, r; q_1, q_2, \dots) - 1.$$

**Theorem 3.1.** *The generating function  $H(t, r; q_1, q_2, \dots)$  satisfies*

$$H(t, r; q_1, q_2, \dots) = \frac{rt}{1-t} + \frac{rt(1-tq_1)}{(1-t)^2} H\left(\frac{t}{1-tq_1}, r; q_2, q_3, \dots\right).$$

REMARK. Theorem 3.1 yields  $F(t, 1; 0, 0, \dots) = \frac{1-2t}{1-3t+t^2}$ , which implies that the number of partitions of  $[n]$  with zero *int* value is  $F_{2n-1}$  if  $n \geq 1$ , where  $F_m$  is the Fibonacci number defined by  $F_m = F_{m-1} + F_{m-2}$  if  $m \geq 2$ , with  $F_0 = 0, F_1 = 1$ .

**3.1. The statistic *int***

Let  $H(t, r; q) := H(t, r; q, q, \dots)$  and  $F(t, r; q) := H(t, r; q, q, \dots) + 1$ . Theorem 3.1 gives

$$H(t, r; q) = \frac{rt}{1-t} + \frac{rt(1-tq)}{(1-t)^2} H\left(\frac{t}{1-tq}, r; q\right),$$

and applying this recurrence relation an infinite number of times, we obtain the following result.

**Theorem 3.2.** *The generating function  $F(t, r; q)$  for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the statistic *int* is given by*

$$1 + \sum_{j \geq 1} \frac{r^j t^j}{1 - [(j-1)q + 1]t} \prod_{i=1}^{j-1} \frac{1 - itq}{(1 - [(i-1)q + 1]t)^2}.$$

Setting  $q = 0$  in Theorem 3.2 implies

$$F(t, r; 0) = 1 + \sum_{j \geq 1} \frac{r^j t^j}{(1-t)^{2j-1}},$$

which implies that the number of partitions of  $[n]$  with exactly  $k$  blocks and having zero *int* is given by  $\binom{n+k-2}{n-k}$  if  $n \geq k \geq 1$ . This also follows directly from the definitions. Note that a member  $\pi$  of  $P(n, k)$  for which  $int(\pi) = 0$  must be of the form  $\pi = \alpha_1 \cdots \alpha_{k-1} \alpha_k \beta_{k-1} \beta_{k-2} \cdots \beta_1$ , where  $\alpha_i$  is a non-empty string of the letter  $i$  for each  $i \in [k]$  and  $\beta_i$  is a (possibly empty) string of the letter  $i$  for each  $i \in [k-1]$ . Conditioning on the value of  $m := \sum_{i=1}^k |\alpha_i|$  implies that the members of  $P(n, k)$  with zero *int* have cardinality

$$\sum_{m=k}^n \binom{m-1}{k-1} \binom{n-m+k-2}{k-2} = \binom{n+k-2}{2k-2}.$$

Note that  $\sum_{k=1}^n \binom{n+k-2}{2k-2} = \sum_{k=0}^{n-1} \binom{2n-2-k}{k} = F_{2n-1}$ , in accordance with the last remark.

Differentiating the generating function  $F(t, r; q)$  with respect to  $q$ , setting  $q = 1$ , and finding the coefficient of  $r^k$  yields

$$(12) \quad [r^k] \left( \frac{d}{dq} F(t, r; q) \Big|_{q=1} \right) = \frac{t^{k+1}}{\prod_{i=1}^k (1-it)} \left( \frac{1}{1-kt} + \sum_{i=1}^k \frac{i-2}{1-it} \right),$$

which implies the following result.

**Corollary 3.3.** *The total int over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$\sum_{j=k}^{n-1} k^{n-1-j} S_{j,k} + \sum_{i=1}^k \left( (i-2) \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right),$$

for all  $n \geq k \geq 1$ .

Comparing this to Corollary 2.11, we also obtain the following trivial consequence.

**Corollary 3.4.** *The total int over all the partitions of  $[n]$  with exactly  $k$  blocks is always less or equal to the total  $d_1$  over all such partitions.*

Substituting  $q = -1$  into  $F(t, r; q)$  and collecting the coefficient of  $r^k$  yields

$$(13) \quad [r^k](F(t, r; -1)) = \frac{(-1)^{k-1} t^3 (1 + (k-1)t)}{(1-t)^2} \left( \frac{(-t)^{k-3}}{\prod_{i=1}^{k-3} (1+it)} \right), \quad k \geq 3,$$

which implies the following result.

**Corollary 3.5.** *The partitions of  $[n]$  having exactly  $k$  blocks and even int value differ in number with those having odd int value by*

$$(n-2) \cdot \delta_{k,3} + (-1)^{k-1} \sum_{i=3}^{n-1} (-1)^{n-i} (i-2) [S_{n-i, k-3} - (k-1) S_{n-1-i, k-3}],$$

for all  $n \geq k \geq 3$ .

We now provide a combinatorial explanation for the expression in Corollary 3.3 for the total int over  $P(n, k)$ , rewritten slightly as

$$\sum_{i=2}^k \left( (i-1) \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right) - \sum_{i=1}^{k-1} \left( \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right).$$

Given  $i$  and  $j$ , where  $2 \leq i \leq k \leq j \leq n-1$ , consider those members of  $P(n, k)$  which may be decomposed exactly as in (7) above except that now the last letter of the word  $x$  must be less than  $i$  (instead of greater than 1). Note that, once again, there are  $(i-1)i^{n-1-j}$  choices for the word  $x$  and  $S_{j,k}$  choices for the remaining letters  $\pi'iy$ . Call a letter  $\pi_i$  in a partition  $\pi = \pi_1\pi_2 \cdots \pi_n$  *secondary* if there exists a letter to its left which is larger. The total number of secondary letters in all the members of  $P(n, k)$  may be obtained by finding the number of partitions which may be expressed as in (7) for each  $i$  and  $j$  and then summing over all possible values of  $i$  and  $j$ , which gives the first part of the expression above.

From the total number of secondary letters, we must subtract the total number of secondary letters which are not internal, i.e., those letters less than  $k$  for which there is no strictly larger letter occurring to the right. Given  $i$  and  $j$ , where  $1 \leq i \leq k-1 < j \leq n-1$ , consider all members of  $P(n, k)$  which may be decomposed as

$$(14) \quad \pi = \pi' i \rho,$$

where  $\pi'$  is a member of  $P(j, k)$  and  $\rho$  is a word of length  $n-1-j$  in the alphabet  $[i]$ . Finding the number of partitions which may be expressed as in (14) for each  $i$  and  $j$  and then summing over all possible values yields the total number of letters within members of  $P(n, k)$  which are secondary but not internal. This gives the second sum in the expression above, which we subtract to obtain the total *int*.  $\square$

While it seems that there are no nice formulas for the total *int* analogous to Theorem 2.6 and Theorem 2.8 (the partial fraction approach leads to very complicated exponential generating functions), one can at least again determine the asymptotic behavior. For fixed  $k$ , note that the generating function is rational, with a dominant double pole at  $\frac{1}{k}$ . By means of the classical singularity analysis [3], we get the expansion

$$\frac{t^{k+1}}{\prod_{i=1}^k (1-it)} \left( \frac{1}{1-kt} + \sum_{i=1}^k \frac{i-2}{1-it} \right) = \frac{k-1}{k \cdot k!(1-tk)^2} + \frac{3-k-k^2-kH_k}{k \cdot k!(1-tk)} + O(1),$$

where  $H_k$  denotes the  $k$ th harmonic number, which yields

**Corollary 3.6.** *For fixed  $k$ , the total *int* over all the partitions of  $[n]$  with exactly  $k$  blocks is asymptotically*

$$\frac{k^{n-1}}{k!} ((k-1)n + 2 - k^2 - kH_k) + O(n(k-1)^n),$$

and the average *int* over all such partitions is therefore asymptotically

$$\frac{(k-1)n}{k} + \frac{2}{k} - k - H_k + O\left(n \left(\frac{k-1}{k}\right)^n\right).$$

For the total over all partitions, regardless of the number of blocks, one has to proceed differently: in view of Corollary 3.4, the total  $int$  is at most the total  $d_1$ , which is  $(n+1)B_n - B_{n+1} - (n-1)B_{n-1}$ , by Theorem 2.15. On the other hand, it is greater than

$$\sum_{k=1}^n \sum_{i=1}^k \left( (i-2) \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right),$$

by Corollary 3.3. The partial fractions technique can then be applied to this expression to yield the exponential generating function  $e^t(e^t-1)(t-1)e^{e^t-1}$ , as in the proof of Theorem 2.15 above. Extracting coefficients yields the following theorem.

**Theorem 3.7.** *The total  $int$  over all partitions of  $[n]$  satisfies*

$$(n+1)B_n - B_{n+1} - 2(n-1)B_{n-1} \leq t(int) \leq (n+1)B_n - B_{n+1} - (n-1)B_{n-1},$$

and so the average  $int$  is  $\frac{t(int)}{B_n} = n + O(n/\log n)$ .

**Proof.** Simply note that the difference between the two sides of the inequality is  $(n-1)B_{n-1} = O(B_n \log n)$ . Therefore,

$$\frac{t(int)}{B_n} = \frac{(n+1)B_n - B_{n+1} - (n-1)B_{n-1}}{B_n} + O(\log n) = n + O(n/\log n). \quad \square$$

### 3.2. The statistic $int_1$

Setting  $q_2 = q_3 = \dots = 1$  in Theorem 3.1 and using the  $q = 1$  case implies that the generating function  $G(t, r; q) := F(t, r; q, 1, 1, \dots)$  for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the  $int_1$  statistic is given by

**Theorem 3.8.** *The generating function for the number of partitions of  $[n]$  with exactly  $k$  blocks according to the statistic  $int_1$  is given by*

$$G(t, r; q) = 1 + \frac{rt}{1-t} + \frac{rt(1-tq)}{(1-t)^2} \sum_{j \geq 1} r^j t^j \prod_{i=1}^j \frac{1}{1-(q+i)t}.$$

Setting  $q = 0$  in Theorem 3.8 implies that the number of elements of  $P(n, k)$  having zero  $int_1$  is given by  $\sum_{j=k-1}^{n-1} (n-j)S_{j,k-1}$  if  $n \geq k \geq 2$ , and hence the number

of elements of  $P(n)$  having zero  $int_1$  is given by  $1 + \sum_{j=1}^{n-1} (n-j)B_j$  if  $n \geq 2$ . This

also follows directly from the definitions since within a member of  $P(n)$  having zero  $int_1$ , the 1's can only occur as runs at the very beginning or at the very end.

Differentiating the generating function  $G(t, r; q)$  with respect to  $q$ , setting  $q = 1$ , and finding the coefficient of  $r^k$  yields

$$(15) \quad [r^k] \left( \frac{d}{dq} G(t, r; q) \Big|_{q=1} \right) = \frac{t^{k+1}}{\prod_{i=1}^k (1-it)} \left( -\frac{1}{1-t} + \sum_{i=1}^{k-1} \frac{1}{1-(i+1)t} \right), \quad k \geq 2,$$

which implies the following result.

**Corollary 3.9.** *The total  $int_1$  over all the partitions of  $[n]$  with exactly  $k$  blocks is given by*

$$\sum_{i=2}^k \left( \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right) - \sum_{j=k}^{n-1} S_{j,k},$$

for all  $n \geq k \geq 1$ .

The combinatorial proof above for Corollary 3.3 applies to Corollary 3.9 as well. There are  $i^{n-1-j}$  choices for the word  $x$ , which now must end in a 1. Thus, the first sum counts the total number of secondary 1's over  $P(n, k)$ . From this, we subtract the total number of secondary 1's which are not internal, which is given by  $\sum_{j=k}^{n-1} S_{j,k}$  (to see this, let  $i = 1$  in (14)). □

Substituting  $q = -1$  into  $G(t, r; q)$  and collecting the coefficient of  $r^k$  yields

$$(16) \quad [r^k](G(t, r; -1)) = \frac{t^2(1+t)}{(1-t)^2} \left( \frac{t^{k-2}}{\prod_{i=1}^{k-2} (1-it)} \right), \quad k \geq 2,$$

which implies the following result.

**Corollary 3.10.** *The partitions of  $[n]$  having exactly  $k$  blocks and even  $int_1$  value differ in number with those having odd  $int_1$  value by*

$$(n-1) \cdot \delta_{k,2} + \sum_{i=3}^n (i-2)[S_{n-i,k-2} + S_{n+1-i,k-2}],$$

for all  $n \geq k \geq 2$ .

Finally, let us determine the asymptotic behavior for the total  $int_1$ . For fixed  $k$ , we can again expand around the dominant pole  $t = \frac{1}{k}$ :

$$\begin{aligned} & \frac{t^{k+1}}{\prod_{i=1}^k (1-it)} \left( -\frac{1}{1-t} + \sum_{i=1}^{k-1} \frac{1}{1-(i+1)t} \right) \\ &= \frac{1}{k \cdot k!(1-kt)^2} - \frac{2(2k-1)}{(k-1)k \cdot k!(1-kt)} + O(1), \end{aligned}$$

so that one obtains the following result.

**Corollary 3.11.** *For fixed  $k$ , the total  $int_1$  over all the partitions of  $[n]$  with exactly  $k$  blocks is asymptotically*

$$\frac{k^{n-1}}{k!} \left( n - \frac{3k-1}{k-1} \right) + O(n(k-1)^n),$$

and the average  $int_1$  over all such partitions is therefore asymptotically

$$\frac{n}{k} - \frac{3k-1}{k(k-1)} + O\left( n \left( \frac{k-1}{k} \right)^n \right).$$

For the total  $int_1$  over all partitions of  $[n]$ , one has to rely on estimates again. First observe

$$\sum_{i=2}^k \left( \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right) - \sum_{j=k}^{n-1} S_{j,k} = \sum_{i=1}^k \left( \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right) - 2 \sum_{j=k}^{n-1} S_{j,k}$$

and

$$\sum_{k=1}^n \sum_{j=k}^{n-1} S_{j,k} = \sum_{j=1}^{n-1} \sum_{k=1}^j S_{j,k} = \sum_{j=1}^{n-1} B_j < \sum_{j=0}^{n-1} \binom{n-1}{j} B_j = B_n.$$

Furthermore, one finds, as in the proofs of Theorem 2.8 and Theorem 2.15 above,

$$\sum_{k=1}^n \sum_{i=1}^k \left( \sum_{j=k}^{n-1} i^{n-1-j} S_{j,k} \right) = (n-1)B_{n-1}.$$

The following theorem is now immediate.

**Theorem 3.12.** *The total  $int_1$  over all partitions of  $[n]$  satisfies*

$$(n-1)B_{n-1} - 2B_n \leq t(int_1) \leq (n-1)B_{n-1},$$

and so the average  $int_1$  is  $\frac{t(int_1)}{B_n} = \log n + O(\log \log n)$ .

**Acknowledgment.** The third author (S.W.) was supported by the National Research Foundation of South Africa under grant number 70560.

**REFERENCES**

1. A. BURSTEIN, T. MANSOUR: *Counting occurrences of some subword patterns*. *Discr. Math. Theor. Comput. Sci.*, **6** (1) (2003), 1–12.
2. J. CIGLER: *A new  $q$ -analog of Stirling numbers*. *Sitzungsberichte*, **201** (1992), 97–109.
3. P. FLAJOLET, R. SEDGEWICK: *Analytic Combinatorics*. Cambridge University Press, Cambridge, 2009.
4. M. KLAZAR: *On abab-free and abba-free set partitions*. *European J. Combin.*, **17** (1996), 53–68.

5. T. MANSOUR, A. MUNAGI: *Enumeration of partitions by long rises, levels and descents*. J. Integer Seq., **12** (2009), Article 09.1.8.
6. T. MANSOUR, B. SIRHAN: *Counting  $\ell$ -letter subwords in compositions*. Discr. Math. Theor. Comput. Sci., **8** (2006), 285–298.
7. T. MANSOUR, A. VAINSHTEIN: *Counting occurrences of 132 in a permutation*. Adv. Appl. Math., **28** (2) (2002), 185–195.
8. S. MILNE: *A  $q$ -analog of restricted growth functions, Dobinski's equality, and Charlier polynomials*. Trans. Amer. Math. Soc., **245** (1978), 89–118.
9. B. SAGAN: *A maj statistic for set partitions*. European J. Combin., **12** (1991), 69–79.
10. R. STANLEY: *Enumerative Combinatorics*, Vol. **1**. Cambridge University Press, Cambridge, 1996.
11. D. STANTON, D. WHITE: *Constructive Combinatorics*. Springer, New York, 1986.
12. C. WAGNER: *Partition statistics and  $q$ -Bell numbers ( $q = -1$ )*. J. Integer Seq., **7** (2004), Article 04.1.1.

Department of Mathematics,  
University of Haifa, 31905 Haifa,  
Israel

E-mail: [toufik@math.haifa.ac.il](mailto:toufik@math.haifa.ac.il)

Department of Mathematics,  
University of Tennessee,  
Knoxville, TN 37996, USA

E-mail: [shattuck@math.utk.edu](mailto:shattuck@math.utk.edu)

Department of Mathematical Sciences,  
Stellenbosch University,  
Matieland 7602,  
South Africa

E-mail: [swagner@sun.ac.za](mailto:swagner@sun.ac.za)

(Received December 3, 2009)

(Revised April 25, 2010)