Notes on Number Theory and Discrete Mathematics
Print ISSN 1310–5132, Online ISSN 2367–8275
Vol. 22, 2016, No. 1, 42–54

Set partitions and m-excedances

Toufik Mansour[1] and Mark Shattuck[2]

[1] Department of Mathematics, University of Haifa
3498838 Haifa, Israel
e-mail: tmansour@univ.haifa.ac.il

[2] Department of Mathematics, University of Haifa
3498838 Haifa, Israel
e-mail: maarkons@excite.com

Received: 16 June 2015 **Accepted:** 10 August 2015

Abstract: In this paper, we consider a generalized m-excedance statistic on set partitions which is analogous to the usual excedance statistic on permutations when $m = 0$. We study it from a probabilistic perspective in which set partitions are regarded as geometrically distributed words satisfying a so-called restricted growth property. We derive a general set of recurrences satisfied by the relevant generating functions and, in the $m = 0$ and $m = 1$ cases, find an explicit formula for the distribution of the statistic recording the number of m-excedances in partitions having a fixed number of blocks. When $m = 0$, one can also determine the joint distribution for the number of excedances with the major index statistic on set partitions. Further recurrences may be given in this case as well as formulas for the mean number of excedances and the variance.
Keywords: Set partition, Geometric random variable, Excedance, q-generalization.
AMS Classification: 05A15, 05A18, 60C05.

1 Introduction

A *partition* of the set $[n] = \{1, 2, \dots, n\}$ is a collection of non-empty, pairwise disjoint subsets, called *blocks*, whose union is $[n]$. (If $n = 0$, then there is a single empty partition of $[0] = \varnothing$ which has no blocks.) A partition Π having exactly k blocks will be called a k-partition. Let $\mathcal{P}_{n,k}$ denote the set of all k-partitions of $[n]$; recall that $|\mathcal{P}_{n,k}| = S(n, k)$, the Stirling number of the second kind (see, e.g., [16, A008277]). A partition $\Pi \in \mathcal{P}_{n,k}$ is said to be in *standard form* if it is written as $\Pi = B_1/B_2/\cdots/B_k$, where $\min(B_1) < \min(B_2) < \cdots < \min(B_k)$.

A statistic on a finite discrete structure is simply a function from the collection of all finite combinatorial objects of the given structure to the non-negative integers. P. A. MacMahon perhaps began the modern study of permutation statistics (see [7, 8]) when he enumerated permutations according to four fundamental statistics, namely, the inversion index, the major index, the descent number, and the excedance number. These statistics have been generalized in different ways and joint distributions of various combinations of them have been studied (see [15] and references contained therein). Furthermore, analogues of these statistics have been considered for other discrete structures, including set partitions. See, for example, Sagan [14] and Goyt [6], where the inversion index, major index, and descent number are studied for set partitions; we also refer the reader to the related paper by Wachs and White [18] and to the text [9, Section 5.7]. Here, we wish to introduce and study an analogue of the excedance statistic for set partitions. Such a statistic does not seem to have been previously considered.

Recall that an *excedance* within a permutation $\sigma = \sigma_1\sigma_2\cdots$ is an index i such that $\sigma_i > i$. The excedance statistic (and its joint distribution with other statistics) has been an object of study on the set of all permutations as well as various restricted subsets; see, e.g., [3, 5, 11]. We define excedances for set partitions as follows. Given $\Pi = B_1/B_2/\cdots/B_k \in \mathcal{P}_{n,k}$ in standard form and a non-negative integer m, we will say that $j \in [n]$ is an *m-excedance* if $j \in B_i$ and $j > i + m$. We will often refer to a 0-excedance simply as an *excedance*. For example, if $\Pi = 1,4,5/2,3,9/6,7/8 \in \mathcal{P}_{9,4}$ and $m = 3$, then there are four 3-excedances in Π, as witnessed by the elements 5, 9, 7, and 8.

We consider here the m-excedance statistic on set partitions in a probabilistic setting in which the letters of a sequence assume positive integral values according to a certain distribution. Let $0 < p < 1$ and $q = 1 - p$. A *geometric* random variable X with parameter p (which we will denote here by geom(p)) is one that assumes positive integral values according to the rule

$$P(X = i) = pq^{i-1}, \qquad i = 1, 2, \ldots.$$

A *geometrically distributed word* $w = w_1w_2\cdots w_n$ with parameter p is one whose letters are generated by a sequence of independent and identically distributed geom(p) random variables. Statistics on geometrically distributed words have been an object of recent study; see, e.g., [1, 2, 4].

Recall that a partition $\Pi = B_1/B_2/\cdots/B_k \in \mathcal{P}_{n,k}$ in standard form may also be represented, equivalently, by the *canonical sequential form* $\pi = \pi_1\pi_2\cdots\pi_n$ wherein $i \in B_{\pi_i}$ for $1 \le i \le n$ (see, e.g., [17]). For example, the partition $\Pi \in \mathcal{P}_{9,4}$ above has the canonical sequential form $\pi = 122113342$. Note that $\pi \in \mathcal{P}_{n,k}$ has the *restricted growth property* (see, e.g., [12]), meaning π is a function from $[n]$ onto $[k]$ such that the first occurrence of i comes earlier than the first occurrence of j for all $1 \le i < j \le k$. One might consider the probability $P_{n,k}$ that a geometrically distributed word of length n contains precisely the letters in $[k]$ and possesses the restricted growth property, i.e., that it corresponds to a canonical sequential form of a member of $\mathcal{P}_{n,k}$. Such a probability has been previously computed [10, Theorem 3.1]:

$$P_{n,k} = \frac{p^n}{[k]_q!}\sum_{j=0}^{k}(-1)^j q^{\binom{j}{2}}\binom{k}{j}_q [k-j]_q^n, \qquad 1 \le k \le n. \tag{1}$$

Note that in computing the probability $P_{n,k}$ above and in what follows, one may regard the event of an element $i \in [n]$ belonging to block B_j within a partition in standard form as having probability pq^{j-1}.

Formula (1) has the property that when p and q are viewed as indeterminates, the resulting expression for $P_{n,k}$ can be written as a polynomial in p and q with non-negative integer coefficients in which each separate monomial term is an encoding of some member of $\mathcal{P}_{n,k}$ (hence $P_{n,k}$ reduces to $S(n,k)$ when $p = q = 1$). Moreover, the $p = 1$ case of (1) is equivalent to a certain q-analogue of $S(n,k)$ that has been previously studied (see, e.g., [14] and [19, Theorem 5.3]). Here, we will consider a refinement of $P_{n,k}$ in terms of excedances which has similar dual probabilistic and combinatorial interpretations.

Let Y be a geometrically distributed word of length n with parameter p. Let $\mathcal{P}_{n,k,r}$ denote the subset of $\mathcal{P}_{n,k}$ whose members contain exactly r m-excedances and let $P_{n,k,r}$ denote the probability that the random variable Y belongs to $\mathcal{P}_{n,k,r}$. Given $1 \le k \le n$, let $a(n,k;z) = \sum_{r=0}^{n-1} P_{n,k,r} z^r$. In the next section, we find a general recurrence for all m satisfied by the generating function of the polynomial sequence $a(n,k;z)$. In the third section, we focus on the $m = 0$ and $m = 1$ cases and compute explicitly the generating functions in these cases and provide additional recurrence formulas. An expression for the total number of excedances within all of the members of $\mathcal{P}_{n,k}$ is also given. In the case $m = 0$, extracting the coefficient that marks word length in the generating function yields an explicit formula for $a(n,k;z)$ (see Theorem 3.1 below). Viewed combinatorially when $p = 1$ and q is an indeterminate, this formula is seen also to be equivalent to the joint distribution of the major index and excedance statistics on $\mathcal{P}_{n,k}$. See [15, Theorem 1.1] for a comparable result in terms of the major index and excedance statistics on permutations.

We will make use of the following notational conventions. Empty sums assume the value 0, and empty products the value 1. If m and n are positive integers, then $[m,n] = \{m, m+1, \ldots, n\}$ if $m \le n$, with $[m,n] = \varnothing$ if $m > n$. The polynomials $[n]_q$ and $[n]_q!$ are defined for a positive integer n as $[n]_q = 1+q+\cdots+q^{n-1}$ and $[n]_q! = \prod_{i=1}^{n} [i]_q$, with $[0]_q = 0$ and $[0]_q! = 1$. Finally, for $n \ge 0$, the q-binomial coefficient $\binom{n}{k}_q$ is defined as $\frac{[n]_q!}{[k]_q![n-k]_q!}$ if $0 \le k \le n$ and as 0, otherwise.

2 Formulas for m-excedances

Given $1 \le k \le n$ and $m \ge 0$, let $a(n,k;z) = a_m(n,k;z)$ denote the generating function for the probability that a geometrically distributed word of length n belongs to $\mathcal{P}_{n,k}$ and contains r m-excedances. For example, if $n = 4$, $k = 3$, and $m = 0$, then we have

$$a(4;3;z) = p^4(q^3z^3 + q^3(1+q)z^2 + q^3(1+q+q^2)z).$$

We have the following recurrence relations satisfied by $a(n,k;z)$.

Lemma 2.1. *If* $k = 1$*, then* $a(n,1;z) = p^n z^{n-m-1}$ *for* $n \ge m+2$ *and* $a(n,1;z) = p^n$ *for* $1 \le n \le m+1$*. If* $k \ge 2$*, then*

$$a(n,k;z) = pq^{k-1}za(n-1,k-1;z) + p[k]_q za(n-1,k;z), \quad n \ge k+m+1, \tag{2}$$

and

$$a(n,k;z) = pq^{k-1}a(n-1,k-1;z) + p([n-m-1]_q z + q^{n-m-1}[k-(n-m-1)]_q)a(n-1,k;z), \tag{3}$$

for $m+2 \le n \le k+m$, *with* $a(n,k;z) = P_{n,k}$ *if* $n \le m+1$.

Proof. The formulas for $k = 1$ follow from the definitions, so assume $k \ge 2$. First note that if $n \le m+1$, then it is not possible for any member of $[n]$ to form an m-excedance, whence $a(n,k;z) = a(n,k;1) = P(n,k)$ in this case. To show (3), we consider the block containing the element n. If n occupies a block by itself, then n cannot form an m-excedance since $n \le k+m$, and thus the weight of such members of $\mathcal{P}_{n,k}$ is given in this case by $pq^{k-1}a(n-1,k-1;z)$. On the other hand, if n does not occupy its own block, then the contribution towards the total weight is $pz[n-m-1]_q a(n-1,k;z)$ if n belongs to the i-th block (from the left) for some $i \in [n-m-1]$ and $pq^{n-m-1}[k-(n-m-1)]_q a(n-1,k;z)$ if n belongs to the i-th block for some $i \in [n-m,k]$, which implies (3). Finally, if $n \ge k+m+1$, then n always contributes an m-excedance regardless of its position, which implies a weight of $pq^{k-1}za(n-1,k-1;z)$ for members of $\mathcal{P}_{n,k}$ in which n belongs to a block by itself and a weight of $p[k]_q za(n-1,k;z)$ for the rest of the members of $\mathcal{P}_{n,k}$. This gives (2) and completes the proof. □

We have the following recurrence relation satisfied by the polynomials $a(n,k;z)$.

Lemma 2.2. *If* $k \ge 2$ *and* $m \ge 0$, *then*

$$a(k+m,k;z) = p^{k+m}q^{\binom{k}{2}} + \sum_{i=2}^{k} p^{k-i+1}q^{\binom{k}{2}-\binom{i}{2}}([i-1]_q z + q^{i-1})a(i+m-1,i;z). \tag{4}$$

Proof. By (3) with $n = k+m$, we have

$$\begin{aligned} a(k-i+m,k-i;z) &- pq^{k-i-1}a(k-i+m-1,k-i-1;z) \\ &= p([k-i-1]_q z + q^{k-i-1}[1]_q)a(k-i+m-1,k-i;z), \end{aligned}$$

which implies

$$\begin{aligned} &\sum_{i=0}^{k-2} p^{i+1}q^{(k-1)+\cdots+(k-i)}([k-i-1]_q z + q^{k-i-1})a(k-i+m-1,k-i;z) \\ &= \sum_{i=0}^{k-2} p^i q^{(k-1)+\cdots+(k-i)}(a(k-i+m,k-i;z) - pq^{k-i-1}a(k-i+m-1,k-i-1;z)) \\ &= \sum_{i=0}^{k-2} p^i q^{\binom{k}{2}-\binom{k-i}{2}}a(k-i+m,k-i;z) - \sum_{i=1}^{k-1} p^i q^{\binom{k}{2}-\binom{k-i}{2}}a(k-i+m,k-i;z) \\ &= a(k+m,k;z) - p^{k-1}q^{\binom{k}{2}}a(m+1,1;z). \end{aligned}$$

Thus,

$$a(k+m,k;z) = p^{k+m}q^{\binom{k}{2}} + \sum_{i=2}^{k} p^{k-i+1}q^{\binom{k}{2}-\binom{i}{2}}([i-1]_q z + q^{i-1})a(i+m-1,i;z),$$

as claimed. □

By Lemma 2.2 and induction on m, we obtain the following explicit formula for $a(k+m, k; z)$.

Theorem 2.3. *If $k \geq 2$ and $m \geq 0$, then*

$$a(k+m, k; z) = p^{k+m} q^{\binom{k}{2}} (1 + d_1 + \cdots + d_m), \tag{5}$$

where

$$d_j = \sum_{i_m=2}^{k} \sum_{i_{m-1}=2}^{i_m} \cdots \sum_{i_{m+1-j}=2}^{i_{m+2-j}} u_{i_m} u_{i_{m-1}} \cdots u_{i_{m+1-j}}$$

and $u_b = [b-1]_q z + q^{b-1}$.

Define $A_k(t, z) = \sum_{n \geq k+m+1} a(n, k; z) t^n$. The generating function $A_k(t, z)$ satisfies the following recurrence.

Theorem 2.4. *If $k \geq 2$ and $m \geq 0$, then*

$$A_k(t, z) = \frac{pq^{k-1} zt}{1 - p[k]_q zt} A_{k-1}(t, z) + \frac{p[k]_q z t^{k+m+1}}{1 - p[k]_q zt} a(k+m, k; z), \tag{6}$$

with $A_1(t, z) = \frac{(pt)^{m+2} z}{1 - pzt}$, where $a(k+m, k; z)$ is given by Theorem 2.3.

Proof. By the initial values when $k = 1$, we have $A_1(t, z) = \frac{(pt)^{m+2} z}{1 - pzt}$. Multiplying both sides of (2) by t^n, and summing over $n \geq k + m + 1$, gives

$$A_k(t, z) = pq^{k-1} zt A_{k-1}(t, z) + p[k]_q zt (A_k(t, z) + a(k+m, k; z) t^{k+m}),$$

which implies

$$A_k(t, z) = \frac{pq^{k-1} zt}{1 - p[k]_q zt} A_{k-1}(t, z) + \frac{p[k]_q z t^{k+m+1}}{1 - p[k]_q zt} a(k+m, k; z).$$

□

Using (6) and programming (if necessary), one may compute $A_k(t, z)$ when k and m are given, and extracting the coefficient of t^n from the resulting formula yields $a(n, k; z)$. Note that

$$\frac{[z^i] a(n, k; z)}{a(n, k; 1)}$$

gives the (conditional) probability that a geometrically distributed member of $\mathcal{P}_{n,k}$ belongs to $\mathcal{P}_{n,k,i}$. Observe that this probability can be positive only if $\max\{0, n-k-m\} \leq i \leq n-m-1$. We illustrate these values below for a given n, k and m and several p.

3 Cases $m = 0$ and $m = 1$

Note that for any given m, the recurrences in (6) may be solved up to k in determining $A_k(t, z)$. In this section, we find formulas for $A_k(t, z)$ for all $k \geq 1$ in the cases when $m = 0$ and $m = 1$. Extracting the coefficient of x^n then gives explicit expressions for the polynomials $a(n, k; z)$ in these cases. When $m = 0$, we are able to provide further recurrences satisfied by $a(n, k; z)$.

Table 1: Values of $\frac{[z^i]a(n,k;z)}{a(n,k;1)}$ for $n = 10$, $k = 6$, $m = 3$ and $i = 1, 2, \ldots, 6$.

$p\backslash i$	1	2	3	4	5	6
0.125	.039523862	.27061578	.41912956	.21403574	.05076996	.005925096
0.250	.027542704	.23035554	.41485439	.24660750	.07057252	.010067335
0.375	.019690081	.19707889	.40283629	.27219189	.09219174	.016011103
0.500	.014759958	.17165881	.38730827	.28930842	.11318550	.023779039
0.625	.011739323	.15345823	.37195956	.29861487	.13118609	.033041933
0.750	.009890500	.14100071	.35895153	.30222981	.14473280	.043194647
0.875	.008723362	.13263367	.34886712	.30260013	.15363133	.053544390

3.1 The case $m = 0$

Let $A_k(t,z) = \sum_{n\geq k+1} a(n,k;z)t^n$, where $a(n,k;z)$ denotes in this subsection the $m = 0$ case. By (6) when $m = 0$, we have

$$A_k(t,z) = \frac{pq^{k-1}zt}{1-p[k]_q zt}A_{k-1}(t,z) + \frac{p[k]_q zt^{k+1}}{1-p[k]_q zt}a(k,k;z), \qquad k \geq 2,$$

with $A_1(t,z) = \frac{p^2zt^2}{1-pzt}$. Note that by the definitions (or Theorem 2.3), we have $a(k,k;z) = p^k q^{\binom{k}{2}}$, which implies

$$A_k(t,z) = \frac{pq^{k-1}zt}{1-p[k]_q zt}A_{k-1}(t,z) + \frac{[k]_q p^{k+1} q^{\binom{k}{2}} zt^{k+1}}{1-p[k]_q zt}, \qquad k \geq 2. \tag{7}$$

Iteration of a recurrence of the form $A_k(t,z) = u_k A_{k-1}(t,z) + v_k$ implies

$$A_k(t,z) = A_1(t,z)\prod_{j=2}^{k} u_j + \sum_{j=2}^{k} v_j \prod_{i=j+1}^{k} u_i.$$

Applying this formula to the our particular case, namely (7), gives

$$\begin{aligned} A_k(t,z) &= \frac{p^{k+1}q^{\binom{k}{2}}z^k t^{k+1}}{\prod_{j=1}^{k}(1-p[j]_q zt)} + p^{k+1}q^{\binom{k}{2}}t^{k+1}\sum_{j=2}^{k}\frac{[j]_q z^{k-j+1}}{\prod_{i=j}^{k}(1-p[i]_q zt)} \\ &= p^{k+1}q^{\binom{k}{2}}t^{k+1}\sum_{j=1}^{k}\frac{[j]_q z^{k-j+1}}{\prod_{i=j}^{k}(1-p[i]_q zt)}. \end{aligned}$$

Now, by the partial fraction decomposition, one can write $A_k(t,z) = \sum_{j=1}^{k} w_j/(1-p[j]_q zt)$, where

$$\begin{aligned} w_\ell &= \lim_{t\to\frac{1}{pz[\ell]_q}} (1-p[\ell]_q zt)A_k(t,z) \\ &= \sum_{j=1}^{\ell} \frac{(-1)^{k-\ell} q^{\binom{k-\ell}{2}+\binom{j}{2}}[j]_q}{[\ell-j]_q![k-\ell]_q![\ell]_q^{j+1} z^j}. \end{aligned}$$

Hence,

$$A_k(t,z) = \sum_{\ell=1}^{k} \frac{\sum_{j=1}^{\ell} \frac{(-1)^{k-\ell} q^{\binom{k-\ell}{2}+\binom{j}{2}} [j]_q}{[\ell-j]_q! [k-\ell]_q! [\ell]_q^{j+1} z^j}}{1 - p[\ell]_q z t}.$$

Comparing the coefficient of t^n on both sides of this last equation implies the following result.

Theorem 3.1. *Let $n > k \geq 1$. Then the generating function $a(n,k;z)$ with $m = 0$ is given by*

$$a(n,k;z) = \frac{p^n}{[k]_q!} \sum_{\ell=1}^{k} (-1)^{k-\ell} q^{\binom{k-\ell}{2}} \binom{k}{\ell}_q \left(\sum_{j=1}^{\ell} q^{\binom{j}{2}} [j]_q! [\ell]_q^{n-j} \binom{\ell-1}{j-1}_q z^{n-j} \right). \tag{8}$$

Taking $z = 1$ in (8) retrieves the previous formula for $P_{n,k}$, which was shown in [10] in a different way.

Corollary 3.2. *If $n > k \geq 1$, then*

$$P_{n,k} = \frac{p^n}{[k]_q!} \sum_{\ell=1}^{k} (-1)^{k-\ell} q^{\binom{k-\ell}{2}} [\ell]_q^n \binom{k}{\ell}_q. \tag{9}$$

Proof. Letting $z = 1$ in (8) gives

$$a(n,k;1) = \frac{p^n}{[k]_q!} \sum_{\ell=1}^{k} (-1)^{k-\ell} q^{\binom{k-\ell}{2}} [\ell]_q^n \binom{k}{\ell}_q \left(\sum_{j=1}^{\ell} q^{\binom{j}{2}} [j]_q! [\ell]_q^{-j} \binom{\ell-1}{j-1}_q \right).$$

Thus, to complete the proof, it suffices to show the identity

$$\sum_{j=1}^{\ell} q^{\binom{j}{2}} [j]_q! [\ell]_q^{-j} \binom{\ell-1}{j-1}_q = 1,$$

which may be rewritten as

$$\sum_{j=1}^{\ell} \frac{[j]_q}{[\ell]_q} \prod_{i=1}^{j-1} \left(1 - \frac{[i]_q}{[\ell]_q}\right) = 1, \qquad \ell \geq 1, \tag{10}$$

upon noting $1 - \frac{[i]_q}{[\ell]_q} = \frac{q^i [\ell-i]_q}{[\ell]_q}$ for $1 \leq i \leq \ell - 1$.

To show (10), we argue probabilistically as follows. Suppose that we start with a single white ball and $\ell - 1$ distinguishable black balls (labeled $1, 2, \ldots, \ell-1$) in an urn. Assign the event of drawing the r-th black ball from the urn a probability of $\frac{q^r}{[\ell]_q}$ for $1 \leq r \leq \ell - 1$, while the probability of drawing the white ball is $\frac{1}{[\ell]_q}$. One first draws a ball from the urn, and if it is not white, changes the color of the black ball labeled 1 to white (note that the probability of selecting this particular ball on subsequent draws stays the same, only the color of it has changed). Then one draws a ball once again from the urn and, if it is not white, changes the black ball labeled 2 to white. We repeat this process wherein one subsequently changes the color of the r-th black ball to white for $r \geq 1$ until one draws a white ball. Note the process must terminate after at most ℓ draws since after $\ell - 1$ draws all of the balls in the urn would be white. The j-th term in the sum on the left-hand side of (10) is seen to give the probability that a white ball is first encountered on the j-th draw. Summing over all j then gives 1, which completes the proof of (10). $\square$

Extracting the coefficient of z^r in (8), we obtain the following result.

Corollary 3.3. *The (conditional) probability that a geometrically distributed member of $\mathcal{P}_{n,k}$ contains exactly r excedances, where $n-k \le r \le n-1$, is given by*

$$\frac{p^n q^{\binom{n-r}{2}}[n-r]_q!}{[k-1]_q! P_{n,k}} \binom{k-1}{n-r-1}_q \sum_{\ell=0}^{k-1} (-1)^\ell q^{\binom{\ell}{2}} [k-\ell]_q^{r-1} \binom{k+r-n}{\ell}_q .$$

It is possible also to give further recurrences satisfied by the polynomials $a(n,k;z)$.

Proposition 3.4. *If $n > k \ge 1$ and $m = 0$, then*

$$a(n,k;z) = \sum_{j=0}^{k-1} (pz)^{j+1} q^{\binom{k}{2}-\binom{k-j}{2}} [k-j]_q a(n-j-1, k-j; z). \qquad (11)$$

Proof. We provide a combinatorial proof as follows. Consider the largest element, $n-j$, not belonging to its own block, where $0 \le j \le k-1$. Then the elements of $[n-j-1]$ comprise a member of $\mathcal{P}_{n-j-1,k-j}$, and thus there are $k-j$ choices regarding the block to contain $n-j$. Since $n > k$, the element $n-j$ always forms an excedance regardless of its position and hence it contributes $p[k-j]_q z$ towards the weight. The elements of $[n-j+1, n]$ then each occupy their own block and all of them form excedances as well; thus, they contribute $(pz)^j q^{(k-j)+\cdots+(k-1)} = (pz)^j q^{\binom{k}{2}-\binom{k-j}{2}}$. Summing over all j then gives (11). □

Proposition 3.5. *If $n \ge k \ge 1$ and $m = 0$, then*

$$a(n,k;z) = p^n q^{\binom{k}{2}} ([k]_q z)^{n-k} + \sum_{i=k}^{n-1} p^{n-i} q^{k-1} [k]_q^{n-i-1} z^{n-i} a(i, k-1; z). \qquad (12)$$

Proof. We consider the position of the first occurrence of the letter k within a geometrically distributed member of $\mathcal{P}_{n,k}$. If the first occurrence of k is in position k, then the weight in that case is $p^n q^{\binom{k}{2}}([k]_q z)^{n-k}$ since the first k letters contribute $p^k q^{\binom{k}{2}}$, while the final $n-k$ contribute $(p[k]_q z)^{n-k}$ towards the weight. So suppose that the first occurrence of k is at position $i+1$ for some $i \in [k, n-1]$. In this case, the first i letters have weight $a(i, k-1; z)$ and the final $n-i-1$ letters have weight $(p[k]_q z)^{n-i-1}$, with the first occurrence of k contributing $pq^{k-1}z$. Summing over all possible i then gives (12). □

Remark. The foregoing two propositions may be generalized using similar reasoning to any m, though more cases are needed in defining the recurrences. For example, formula (11) can be generalized for $k \ge 2$ to

$$a(n,k;z) = \sum_{j=0}^{k-1} (pz)^{j+1} q^{\binom{k}{2}-\binom{k-j}{2}} [k-j]_q a(n-j-1, k-j; z), \qquad n \ge k+m+1,$$

with

$$\begin{aligned} a(n,k;z) = &\sum_{j=0}^{n-m-2} p^{j+1} q^{\binom{k}{2}-\binom{k-j}{2}} ([n-m-j-1]_q(z-1) + [k-j]_q) a(n-j-1, k-j; z) \\ &+ \sum_{j=n-m-1}^{k-1} p^{j+1} q^{\binom{k}{2}-\binom{k-j}{2}} [k-j]_q a(n-j-1, k-j; z), \qquad m+2 \le n \le k+m, \end{aligned}$$

and $a(n, k; z) = P_{n,k}$ if $n \le m + 1$. Similarly, formula (12) may be extended to any m upon considering cases based on whether or not $n \ge k + m + 1$ and whether or not $k > m$.

Let X be the random variable that records the number of excedances in a geometric word of length n. Let F be the event that a geometric word of length n belongs to $\mathcal{P}_{n,k}$. Then the conditional mean $\mu := E[X|F]$ is the average value of the number of excedances among members of $\mathcal{P}_{n,k}$. Let $\sigma^2 := \text{Var}[X|F] = E[(X - \mu)^2|F]$ denote the conditional variance. Then there are the following explicit formulas for μ and σ^2.

Theorem 3.6. *If $n \ge k \ge 1$, then*

$$\mu = n - \frac{p^n}{P_{n,k}} \sum_{r=1}^{k} \sum_{i=1}^{r} \frac{(-1)^{k-r} q^{\binom{k-r}{2} + \binom{i}{2}} [r]_q^{n-i}}{[k-r]_q! [r-i]_q!} \tag{13}$$

and

$$\sigma^2 = \frac{2p^n}{P_{n,k}} \sum_{r=1}^{k} \sum_{i=1}^{r} \frac{(-1)^{k-r} q^{\binom{k-r}{2} + \binom{i}{2}} i [r]_q^{n-i}}{[k-r]_q! [r-i]_q!} - (n - \mu) - (n - \mu)^2. \tag{14}$$

Proof. Let $Y = n - X$. Note that the random variable Y records the length of the longest initial run of the form $12 \cdots i$ within a geometrically distributed word of length n. Then we have the conditional probability formula

$$P(Y \ge i|F) = \frac{p^i q^{\binom{i}{2}}}{P_{n,k}} \sum_{j=0}^{n-i} p^j q^{i(n-i-j)} [i]_q^j \binom{n-i}{j} P_{n-i-j,k-i}, \qquad 1 \le i \le k. \tag{15}$$

To show (15), note first that the intersection of events $\{Y \ge i\} \cap F$ occurs if and only if the first i letters of a geometric word form the sequence $12 \cdots i$, with the rest of the letters of the word that do not belong to $[i]$ constituting a partition of $k - i$ blocks. Observe that the probability that the first i letters occur as $12 \cdots i$ is $p^i q^{\binom{i}{2}}$. To compute the probability of the intersection of events above, suppose further that exactly j of the final $n - i$ positions of the word correspond to members of $[i]$. The probability of this occurring is given by $p^j q^{i(n-i-j)} [i]_q^j \binom{n-i}{j} P_{n-i-j,k-i}$. To see this, we choose j of the final $n - i$ letters of a geometric word of length n to correspond to members of $[i]$. Once this selection is made, the probability is $p^j [i]_q^j$ that these positions do in fact contain members of $[i]$, while the probability that the remaining $n - i - j$ positions of the word form a partition of $k - i$ blocks on the letters in $[i + 1, k]$ is $q^{i(n-i-j)} P_{n-i-j,k-i}$, by independence (note that the $q^{i(n-i-j)}$ factor accounts for each letter in this partition being at least $i + 1$). Summing over all j yields (15).

By a well-known formula for the expected value [13, p. 180] and (15), we then have

$$E[Y|F] = \sum_{i=1}^{k} P(Y \ge i|F) = \frac{1}{P_{n,k}} \sum_{i=1}^{k} \sum_{j=0}^{n-i} p^{i+j} q^{\binom{i}{2} + i(n-i-j)} [i]_q^j \binom{n-i}{j} P_{n-i-j,k-i}.$$

Applying formula (1) to the $P_{n-i-j,k-i}$ factor in the last formula, interchanging summation, and using the binomial theorem yields

$$E[Y|F] = \frac{p^n}{P_{n,k}} \sum_{r=1}^{k} \sum_{i=1}^{r} \frac{(-1)^{k-r} q^{\binom{k-r}{2} + \binom{i}{2}} [r]_q^{n-i}}{[k-r]_q! [r-i]_q!}.$$

Noting that $\mu = n - E[Y|F]$ gives (13). Recalling the fact (see [13, p. 180]) that

$$E[Z^2] = 2\sum_{i\geq 1} iP(Z \geq i) - E[Z],$$

for a non-negative integer-valued random variable Z, and observing Var$[X|F]$ = Var$[Y|F]$, implies formula (14). □

Regarding the variables p and q as indeterminates in (13), and taking $p = q = 1$, implies the following result.

Corollary 3.7. *If $n \geq k \geq 1$, then the total number of excedances within all of the members of $\mathcal{P}_{n,k}$ is given by*

$$nS(n,k) - \sum_{r=1}^{k}\sum_{i=1}^{r} \frac{(-1)^{k-r} r^{n-i}}{(k-r)!(r-i)!}.$$

Remark. For k fixed and n large, note that $nS(n,k) \sim \frac{nk^n}{k!}$, whereas the sum $\sum_{r=1}^{k}\sum_{i=1}^{r} \frac{(-1)^{k-r} r^{n-i}}{(k-r)!(r-i)!}$ behaves like $\sum_{i=1}^{k} \frac{k^n}{(k-i)!k^i}$, which is negligible compared to $nS(n,k)$. Similar observations applied to formulas (13) and (14) imply that the conditional mean and variance are approximately n and $b_k - a_k(1 + a_k)$, where

$$a_k = \sum_{i=1}^{k} \frac{q^{\binom{i}{2}}[k]_q!}{[k-i]_q![k]_q^i} \quad \text{and} \quad b_k = \sum_{i=1}^{k} \frac{2q^{\binom{i}{2}} i[k]_q!}{[k-i]_q![k]_q^i}.$$

Suppose $\Pi = B_1/B_2/\cdots/B_k \in \mathcal{P}_{n,k}$ is in standard form. Given $1 \leq i \leq k-1$, let d_i denote the number of elements $b \in B_i$ such that $b > \min(B_{i+1})$. Then the *major index* [14] of Π is defined as

$$\text{maj}(\Pi) = 1d_1 + 2d_2 + \cdots + (k-1)d_{k-1}.$$

For example, for the partition $\Pi \in \mathcal{P}_{9,4}$ given in the introduction, we have maj$(\Pi) = 2+2+0 = 4$. Let exc denote the statistic recording the number of excedances. Then we have the following result.

Theorem 3.8. *If $n > k \geq 1$, then*

$$\sum_{\Pi\in\mathcal{P}_{n,k}} q^{maj(\Pi)} z^{exc(\Pi)} = \frac{1}{[k]_q!}\sum_{\ell=1}^{k}(-1)^{k-\ell} q^{\binom{k-\ell}{2}-\binom{k}{2}} \binom{k}{\ell}_q \left(\sum_{j=1}^{\ell} q^{\binom{j}{2}}[j]_q![\ell]_q^{n-j}\binom{\ell-1}{j-1}_q z^{n-j}\right). \tag{16}$$

Proof. Consider the statistic maj$^*(\Pi) = \binom{k}{2}$ + maj(Π) on $\mathcal{P}_{n,k}$. Note that maj* may be obtained from maj be adding $j-1$ whenever the first element of the block B_j is encountered for $1 \leq j \leq k$ as one creates a member of $\mathcal{P}_{n,k}$ by adding elements one-by-one starting with 1. Let

$$b(n,k;z) = \sum_{\Pi\in\mathcal{P}_{n,k}} q^{\text{maj}^*(\Pi)} z^{\text{exc}(\Pi)}.$$

Considering whether or not the element n occupies its own block within a member of $\mathcal{P}_{n,k}$ implies

$$b(n,k;z) = q^{k-1}zb(n-1,k-1;z) + [k]_q zb(n-1,k;z), \qquad n > k \geq 2,$$

with $b(k,k;z) = q^{\binom{k}{2}}$ for $k \geq 1$ and $b(n,1;z) = z^{n-1}$ for $n \geq 2$. Upon comparison with Lemma 2.1, we see that $b(n,k;z)$ satisfies the same recurrence and initial conditions as the $m = 0$ case of $a(n,k;z)$ when $p = 1$ and q is regarded as an indeterminate. Taking $p = 1$ in (8), and dividing both sides by $q^{\binom{k}{2}}$, then gives (16). □

Remark. Given a partition Π, let $\text{init}(\Pi)$ denote the length of the longest initial increasing run of letters in the canonical sequential form of Π. Then replacing z by $1/z$ in (16), and multiplying both sides by z^n, gives the joint distribution of the maj and init statistics on $\mathcal{P}_{n,k}$.

3.2 The case $m = 1$

Taking $m = 1$ in Theorem 2.4 gives

$$A_k(t,z) = \frac{pq^{k-1}zt}{1-p[k]_q zt}A_{k-1}(t,z) + \frac{p[k]_q zt^{k+2}}{1-p[k]_q zt}a(k+1,k;z), \qquad k \geq 2,$$

with $A_1(t,z) = \frac{p^3zt^3}{1-pzt}$. By Theorem 2.3 with $m = 1$, we have

$$a(k+1,k;z) = p^{k+1}q^{\binom{k}{2}}\left(1 + \sum_{i=1}^{k-1}([i]_q z + q^i)\right), \qquad k \geq 2.$$

This implies $A_k(t,z) = u_k A_{k-1}(t,z) + v_k$, where

$$u_k = \frac{pq^{k-1}zt}{1-p[k]_q zt}, \quad v_k = \frac{p^{k+2}q^{\binom{k}{2}}[k]_q zt^{k+2}}{1-p[k]_q zt}\left(1 + \sum_{i=1}^{k-1}([i]_q z + q^i)\right).$$

Therefore,

$$A_k(t,z) = A_1(t,z)\prod_{j=2}^{k} u_j + \sum_{j=2}^{k} v_j \prod_{i=j+1}^{k} u_i,$$

which implies

$$A_k(t,z) = \frac{p^{k+2}q^{\binom{k}{2}}z^k t^{k+2}}{\prod_{j=1}^{k}(1-p[j]_q zt)} + p^{k+2}q^{\binom{k}{2}}zt^{k+2}\sum_{j=2}^{k}\frac{[j]_q z^{k-j}w_j}{\prod_{i=j}^{k}(1-p[i]_q zt)},$$

where $w_j = 1 + q[j-1]_q + z\sum_{i=1}^{j-1}[i]_q$. Hence,

$$A_k(t,z) = p^{k+2}q^{\binom{k}{2}}t^{k+2}\sum_{j=1}^{k}\frac{[j]_q z^{k-j+1}w_j}{\prod_{i=j}^{k}(1-p[i]_q zt)}.$$

By the partial fraction decomposition, we have $A_k(t,z) = \sum_{\ell=1}^{k}\frac{a_\ell}{1-p[\ell]_q zt}$, where

$$\begin{aligned}
a_\ell &= \frac{q^{\binom{k}{2}}}{[\ell]_q^{k+2}z^{k+2}}\sum_{j=1}^{\ell}\frac{[j]_q z^{k-j+1}w_j}{\prod_{i=j}^{\ell-1}(1-[i]_q/[\ell]_q)\prod_{i=\ell+1}^{k}(1-[i]_q/[\ell]_q)} \\
&= \frac{(-1)^{k-\ell}q^{\binom{k-\ell}{2}}}{[k-\ell]_q!}\sum_{j=1}^{\ell}\frac{q^{\binom{j}{2}}[j]_q w_j}{[\ell]_q^{j+2}[\ell-j]_q!z^{j+1}}.
\end{aligned}$$

Hence, we can state the following result.

Theorem 3.9. *If* $1 \leq k < n-1$ *and* $m = 1$*, then*

$$a(n,k;z) = p^n \sum_{\ell=1}^{k} \left[\frac{(-1)^{k-\ell} q^{\binom{k-\ell}{2}}}{[k-\ell]_q!} \sum_{j=1}^{\ell} \frac{q^{\binom{j}{2}} [j]_q \left(1 + q[j-1]_q + z \sum_{i=1}^{j-1} [i]_q\right)}{[\ell-j]_q!} [\ell]_q^{n-j-2} z^{n-j-1} \right]. \tag{17}$$

We remark that analogous formulas for $a(n,k;z)$ may be computed for other small m in a similar fashion.

Acknowledgement

We wish to thank Stephan Wagner for the versions of formulas (13) and (14) presented which make clear their asymptotic behavior.

References

[1] Archibald, M., & Knopfmacher A. (2009) The average position of the d-th maximum in a sample of geometric random variables, *Statist. Probab. Lett.*, 79, 864–872.

[2] Archibald, M., & Knopfmacher A. (2014) The largest missing value in a sample of geometric random variables, *Combin. Probab. Comput.*, 23(5), 670–685.

[3] Baril, J.-L. (2013) Statistics-preserving bijections between classical and cyclic permutations, *Inform. Process. Lett.*, 113(1–2), 17–22.

[4] Brennan, C., & Knopfmacher, A. (2013) Descent variation of samples of geometric random variables, *Discrete Math. Theor. Comput. Sci.*, 15:2, 1–12.

[5] Elizalde, S. (2011-2) Fixed points and excedances in restricted permutations, *Electron. J. Combin.*, 18(2), #P29.

[6] Goyt, A. (2008) Avoidance of partitions of a three element set, *Adv. in Appl. Math.*, 41, 95–114.

[7] MacMahon, P. A. (1913) The indices of permutations and derivation therefrom of functions of a single variable associated with the permutations of any assemblage of objects, *Amer. J. Math.*, 35(3), 281–322.

[8] MacMahon, P. A. (1960) *Combinatory Analysis*, Vol. 2, Cambridge University Press, 1915–1916. Reprinted by Chelsea, New York.

[9] Mansour, T. (2012) *Combinatorics of Set Partitions*, CRC Press, Boca Raton.

[10] Mansour, T., & Shattuck, M. (2012) Set partitions as geometric words, *Australas. J. Combin.*, 53, 31–39.

[11] Mantaci, R. & Rakotondrajao, F. (2003) Exceedingly deranging!, *Adv. in Appl. Math.*, 30(1–2), 177–188.

[12] Milne, S. (1978) A q-analog of restricted growth functions, Dobinski's equality, and Charlier polynomials, *Trans. Amer. Math. Soc.*, 245, 89–118.

[13] Ross, S. (2010) *A First Course in Probability*, 8-th edition, Prentice Hall, Upper Saddle River, N.J.

[14] Sagan, B. E. (1991) A maj statistic for set partitions, *European J. Combin.*, 12, 69–79.

[15] Shareshian, J. & Wachs, M. L. (2010) Eulerian quasisymmetric functions, *Adv. Math.*, 225, 2921–2966.

[16] Sloane, N. J. (2010) The On-line Encyclopedia of Integer Sequences, `http://oeis.org`.

[17] Stanton, D. & White, D. (1986) *Constructive Combinatorics*, Springer, New York.

[18] Wachs, M. & White, D. (1991) p, q-Stirling numbers and set partition statistics, *J. Combin. Theory Ser. A*, 6, 27–46.

[19] Wagner, C. G. (1996) Generalized Stirling and Lah numbers, *Discrete Math.*, 160, 199–218.