Records in set partitions

Arnold Knopfmacher*
The John Knopfmacher Centre for Applicable Analysis and Number Theory
Department of Mathematics, University of the Witwatersrand
P. O. Wits, 2050, Johannesburg, South Africa
arnold.knopfmacher@wits.ac.za

Toufik Mansour
Department of Mathematics, University of Haifa
31905 Haifa, Israel
toufik@math.haifa.ac.il

Stephan Wagner†
Department of Mathematical Sciences, Stellenbosch University
7602 Stellenbosch, South Africa.
swagner@sun.ac.za

Submitted: April 23, 2010; Accepted: July 13, 2010; Published: Aug 9, 2010
Mathematics Subject Classification: Primary: 05A18; Secondary: 05A15, 05A16

Abstract

A partition of $[n] = \{1, 2, \ldots, n\}$ is a decomposition of $[n]$ into nonempty subsets called blocks. We will make use of the canonical representation of a partition as a word over a finite alphabet, known as a restricted growth function. An element a_i in such a word π is a strong (weak) *record* if $a_i > a_j$ ($a_i \geqslant a_j$) for all $j = 1, 2, \ldots, i-1$. Furthermore, the position of this record is i. We derive generating functions for the total number of strong (weak) records in all words corresponding to partitions of $[n]$, as well as for the sum of the positions of the records. In addition we find the asymptotic mean values and variances for the number, and for the sum of positions, of strong (weak) records in all partitions of $[n]$.

*This material is based upon work supported by the National Research Foundation under grant number 2053740

†This material is based upon work supported by the National Research Foundation under grant number 70560

1 Introduction

Let $\pi = a_1 a_2 \cdots a_n$ be any permutation of length n. An element a_i in π is a *record* if $a_i > a_j$ for all $j = 1, 2, \ldots, i-1$. Furthermore, the position of this record is i. The number of records in permutations was first studied by Rényi [11], see also [5]. A survey of results on this topic can be found in [3]. Recently, Myers and Wilf [7] extended the study of records to multiset permutations and words. In the literature records are also referred to as *left–to–right maxima* or *outstanding elements*. In particular the study of records has applications to observations of extreme weather problems, test of randomness, determination of minimal failure, and stresses of electronic components. The recent paper by Kortchemski [6] defines a new statistic srec, where $\mathsf{srec}(\pi)$ is the sum over the positions of all records in π. For instance, the permutation $\pi = 451632$ has 3 records $4, 5, 6$ and $\mathsf{srec}(\pi) = 1 + 2 + 4 = 7$.

A word over an alphabet A, a set of positive integers, is defined as any ordered sequence of possibly repeated elements of A. Recently, Prodinger [10] studied the statistic srec for words over the alphabet $\mathbb{N} = \{1, 2, 3, \ldots\}$, equipped with geometric probabilities $p, pq, pq^2, \ldots$ with $p + q = 1$. In the case of words there are two versions: A *strong record* in a word $a = a_1 \cdots a_n$ is an element a_i such that $a_i > a_j$ for all $j = 1, 2, \ldots, i-1$ (that is, it must be strictly larger than elements to the left) and a *weak record* is an element a_i such that $a_i \geqslant a_j$ for all $j = 1, 2, \ldots, i-1$ (must be only larger or equal to elements to the left). Furthermore, the position i is called the position of the strong record (weak record). In [10], Prodinger found the expected value of the sum of the positions of strong records, in random geometrically distributed words of length n. Previously, Prodinger [9] also studied the number of strong and weak records in samples of geometrically distributed random variables. Records have recently also been studied for compositions in [4], where a composition $\sigma = \sigma_1 \sigma_2 \ldots \sigma_m$ of n is an ordered collection of positive integers whose sum is n.

A *partition* of $[n] = \{1, 2, \ldots, n\}$ is a decomposition of $[n]$ into nonempty subsets called blocks. A partition with k blocks is called a k-partition and denoted by $B_1|B_2|\ldots|B_k$ where the blocks are listed in standard order, that is, $min(B_1) < \cdots < min(B_k)$. We will also make use of the representation of a partition as a word over a finite alphabet. That is, we represent the partition $\pi = B_1|B_2|\ldots|B_k$ in the canonical word form $\pi = \pi_1, \pi_2, \ldots, \pi_n$ such that $j \in B_{\pi_j}$, for $1 \leqslant j \leqslant n$. For example, 1231242 is the canonical word corresponding to the partition $\{1, 4\}, \{2, 5, 7\}, \{3\}, \{6\}$ of [7]. A word over the alphabet $[k]$ represents a partition of $[n]$ with k blocks if and only if each number from the set $[k]$ appears at least once in π and for each i, j with $1 \leqslant i < j \leqslant k$ the first occurrence of i precedes the first occurrence of j. Words satisfying these properties are also known as *restricted growth functions*.

It is natural with respect to such words to consider once again record values (left-to-right maxima). In particular, strong records of a set partition π (in word form) correspond to the well studied statistic *number of blocks* in the partition π; thus the number of set partitions of n with k strong records is the *Stirling partition number* (Stirling number of the second kind), henceforth denoted $S(n, k)$.

However weak records have not previously been considered. In addition we will consider the statistic *sum of positions of records* in π. In this paper we find generating functions for these parameters as well as mean values and variances and their asymptotic behaviour as $n \to \infty$.

The number of partitions of $[n]$ is called the Bell number, with exponential generating function

$$\sum_{n\geqslant 0} B_n \frac{x^n}{n!} = e^{e^x-1}.$$

Asymptotically we have as $n \to \infty$

$$B_n \sim n! \frac{e^{e^r-1}}{r^n\sqrt{2\pi r(r+1)e^r}} \tag{1}$$

where r is the positive root of $re^r = n+1$, from which we get

$$r \equiv r(n) = \log n - \log\log n + O\left(\frac{\log\log n}{\log n}\right).$$

For the number of strong records in set partitions we merely quote the known results concerning the number of blocks, as found, for example in [2]. Thus the mean number of strong reords over all partitions of $[n]$ is $B_{n+1}/B_n - 1$ and asymptotically we find that the mean and variance are respectively

$$\frac{n}{\log n} \text{ and } \frac{n}{\log^2 n}.$$

2 Weak records in set partitions

We want to determine the distribution of the number of weak records in words (restricted growth functions) that correspond to set partitions of $[n]$. It turns out that the excess of the number of weak records over strong records is comparatively small. Hence it is more interesting to study *additional weak records*, that is, weak records that are not also strong records.

Since such words can be decomposed as

$$\pi = 1(1)^*2(12)^*\cdots k(12\cdots k)^* \tag{2}$$

for some k, where $(Q)^*$ denotes an arbitrary word over an alphabet Q including the empty word, we have the generating function

$$\sum_{k\geqslant 1} \frac{u^k x^k}{\prod_{j=1}^k (1-(j+v-1)x)},$$

where u marks the number of strong records and v marks additional weak records. We expand it into partial fractions:

$$\sum_{k\geqslant 1} \frac{u^k x^k}{\prod_{j=1}^k (1-(j+v-1)x)} = \sum_{k\geqslant 1} \frac{u^k}{\prod_{j=1}^k (1/x-(j+v-1))} = \sum_{m\geqslant 1} \frac{a_m}{1/x-(m+v-1)}.$$

The coefficient a_m can be found by multiplying by $\frac{1}{x}-(m+v-1)$ and setting $x=\frac{1}{m+v-1}$:

$$\begin{aligned}
a_m &= \sum_{k\geqslant 1} \frac{u^k(1/x-(m+v-1))}{\prod_{j=1}^{k}(1/x-(j+v-1))}\Big|_{x=1/(m+v-1)} \\
&= \sum_{k\geqslant m} \frac{u^k}{\prod_{\substack{j=1\\ j\neq m}}^{k}(1/x-(j+v-1))}\Big|_{x=1/(m+v-1)} \\
&= \sum_{k\geqslant m} u^k \prod_{j=1}^{m-1}(m-j)^{-1} \prod_{j=m+1}^{k}(m-j)^{-1} \\
&= \frac{1}{(m-1)!}\sum_{k\geqslant m}\frac{(-1)^{k-m}u^k}{(k-m)!} = \frac{u^m e^{-u}}{(m-1)!}.
\end{aligned}$$

Therefore,

$$\sum_{k\geqslant 1}\frac{u^k x^k}{\prod_{j=1}^{k}(1-(j+v-1)x)} = \sum_{m\geqslant 1}\frac{u^m e^{-u}x}{(m-1)!(1-(m+v-1)x)}.$$

Now we expand $\frac{x}{1-(m+v-1)x}$ into a geometric series:

$$\sum_{k\geqslant 1}\frac{u^k x^k}{\prod_{j=1}^{k}(1-(j+v-1)x)} = \sum_{m\geqslant 1}\frac{u^m e^{-u}}{(m+v-1)(m-1)!}\sum_{n\geqslant 1}((m+v-1)x)^n.$$

Since we would like to work with the exponential generating function rather than the ordinary generating function, we introduce a factor $\frac{1}{n!}$:

$$\begin{aligned}
&\sum_{m\geqslant 1}\frac{u^m e^{-u}}{(m+v-1)(m-1)!}\sum_{n\geqslant 1}\frac{((m+v-1)x)^n}{n!} \\
&\qquad = \sum_{m\geqslant 1}\frac{u^m e^{-u}}{(m+v-1)(m-1)!}\left(e^{(m+v-1)x}-1\right)
\end{aligned}$$

In order to obtain an elementary function, we differentiate with respect to x (since we are dealing with exponential generating functions now, this merely means a shift of coefficients) to obtain

$$\sum_{m\geqslant 1}\frac{u^m e^{-u}e^{(m+v-1)x}}{(m-1)!} = ue^{ue^x+vx-u}. \tag{3}$$

Note that $u=v=1$ yields e^{e^x+x-1}, which is indeed the derivative of the generating function of the Bell numbers. The equation (3) can also be interpreted in another way: u marks the number of blocks in a set partition, while v marks the number of elements (other than 1) in the first block. Indeed, ue^{vx} generates a single block (to which the element 1 is added), while $e^{u(e^x-1)}$ generates an arbitrary number of additional blocks.

There is also a simple bijection that shows this identity: in a word that corresponds to a set partition, replace every 1 between the first occurrence of p and the first occurrence of $p+1$ (if any) by p and vice versa. Then the additional weak records are exactly mapped to elements of the first block.

The generating function immediately leads to explicit formulae for the number of set partitions with a prescribed number of strong and additional weak records:

Theorem 2.1. *The number of partitions of $[n]$ with exactly k strong and ℓ additional weak records is*

$$\binom{n-1}{\ell} S(n-1-\ell, k-1).$$

The total number of partitions of $[n]$ with ℓ additional weak records is

$$\binom{n-1}{\ell} B_{n-1-\ell}.$$

From (3), we see that the generating function for the total number of additional weak records is

$$\frac{d}{dv} u e^{ue^x+vx-u}\Big|_{v=1} = uxe^x e^{u(e^x-1)}. \tag{4}$$

This generating function also arises in another interesting context. Let v mark singleton blocks and u all blocks, then the appropriate bivariate generating function is $e^{u(e^x-1-x+vx)}$. Since the number of singleton blocks equal to $\{1\}$ in partitions of $[n]$ is equal to the number of partitions of $[n-1]$ we see that the generating function for the number of singleton blocks in all partitions of $[n]$, excluding any singleton blocks equal to $\{1\}$ is

$$\frac{d}{dv}\frac{d}{dx} e^{u(e^x-1-x+vx)}\Big|_{v=1} - ue^{u(e^x-1)} = u^2 xe^x e^{u(e^x-1)}. \tag{5}$$

(Once again, differentiation with respect to x merely means a shift of coefficients.)

We now give a bijection between the number of elements (other than 1) in the first block and the number of singleton blocks in all partitions of $[n]$, excluding any singleton blocks equal to $\{1\}$. Let $B_1|B_2|B_3|...|B_k$ be a partition with k blocks such that B_{i_j} contains only one element for $j = 1, 2, 3, \ldots s$. We construct a new partition of $[n]$ by doing the following: All the elements in $B_{i_1} \cup ... \cup B_{i_s}$ together with 1 define a block (the first block) and each element of B_1 other than the element 1 defines a singleton block. The rest of the blocks remain unchanged. This mapping is an involution on the set of partitions of $[n]$ and thus it is a bijection. For example, $1|234 \to 1|234$, $12|34 \to 1|2|34$ and $123|4 \to 14|2|3$.

Comparing the generating functions (4) and (5), we notice that actually a stronger result holds: since they differ only by a factor of u, we can deduce that the total number of elements that are in the same block as the element 1 in all partitions of $[n]$ into k blocks is exactly the total number of singletons, excluding those of the form $\{1\}$, in all partitions of $[n]$ into $k+1$ blocks. Let us show this interesting relation by means of a bijection as well. We consider set partitions in which one of the elements that are in

the same block as 1 (excluding 1 itself) is marked. Clearly the number of such marked partitions is exactly the total number of elements that are in the same block as 1. For each such marked partition, remove the marked element from its block and make it form a singleton block; for example, $12\underline{4}|3 \to 12|3|\underline{4}$. The length clearly increases by 1, and the result of the mapping (which can obviously be reversed) is a set partition in which one of the singleton blocks (other than $\{1\}$) is marked. The total number of such marked set partitions is clearly the total number of singleton blocks other than $\{1\}$ over all partitions, which completes the bijective proof of our assertion.

Let us finally turn to asymptotic results on the number of additional weak records:

Theorem 2.2. *The distribution of the number of additional weak records in a random partition of $[n]$ is asymptotically Gaussian, with mean*

$$\frac{(n-1)B_{n-1}}{B_n} = \log n - \log\log n + O\left(\frac{\log\log n}{\log n}\right)$$

and variance

$$\frac{(n-1)(n-2)B_{n-2} + (n-1)B_{n-1}}{B_n} - \frac{(n-1)^2 B_{n-1}^2}{B_n^2} = \log n - \log\log n + O\left(\frac{\log\log n}{\log n}\right).$$

Proof. The explicit formulae for mean and variance follow easily by differentiating the generating function with respect to v and extracting coefficients. For instance, the total number of additional weak records over all partitions of $[n]$ is

$$(n-1)![x^{n-1}]\frac{d}{dv}e^{e^x+vx-1}\Big|_{v=1} = (n-1)![x^{n-1}]xe^{e^x+x-1} = (n-1)B_{n-1},$$

which yields the formula for the mean. The computation of the variance is similar. In order to obtain asymptotic estimates for the moments as well as the limiting distribution, we need an extension of (1):

$$B_{n+h} = B_n \cdot \frac{(n+h)!}{n!r^h}\left(1 + O\left(\frac{\log n}{n}\right)\right), \tag{6}$$

uniformly for $h = O(\log n)$, where r is the positive root of $re^r = n+1$ again. See [1] for an even stronger form that includes further terms in the asymptotic expansion.

It is not difficult to derive from this formula that both mean and variance are asymptotically equal to $r + O\left(\frac{\log^2 n}{n}\right)$, from which the stated formulae follow. For the limiting distribution, consider the probability that there are exactly ℓ additional weak records, which equals, for $\ell = O(\log n)$,

$$\binom{n-1}{\ell}\frac{B_{n-1-l}}{B_n} = \binom{n-1}{\ell}\cdot\frac{(n-1-l)!r^{\ell+1}}{n!}\left(1 + O\left(\frac{\log n}{n}\right)\right)$$

$$= \frac{r^{\ell+1}}{n\ell!}\left(1+O\left(\frac{\log n}{n}\right)\right).$$

Now set $\ell = r + t\sqrt{r}$ and apply Stirling's formula to obtain, for $t = o(r^{1/6})$,

$$\begin{aligned}
&\binom{n-1}{\ell}\frac{B_{n-1-l}}{B_n}\\
&= \frac{r}{n}\cdot\exp\left(\ell+\ell\log r-\ell\log\ell-\frac{1}{2}\log(2\pi\ell)\right)\left(1+O\left(\frac{\log n}{n}\right)\right)\\
&= \frac{r}{n}\cdot\exp\left(\ell\left(1-\log(1+tr^{-1/2})\right)-\frac{1}{2}\log(2\pi r)+O(r^{-1/2})\right)\\
&= \frac{r}{n}\cdot\exp\left((r+t\sqrt{r})\left(1-tr^{-1/2}+\frac{t^2}{2}r^{-1}\right)-\frac{1}{2}\log(2\pi r)+O((t^3+1)r^{-1/2})\right)\\
&= \frac{r}{n}\cdot\exp\left(r-\frac{t^2}{2}-\frac{1}{2}\log(2\pi r)+O((t^3+1)r^{-1/2})\right)\\
&= \frac{1}{\sqrt{2\pi r}}\exp\left(-\frac{t^2}{2}+o(1)\right),
\end{aligned}$$

which completes the proof of the theorem. □

3 Sum of positions of records in set partitions

We will first study this parameter from the point of view of restricted growth functions. Let $\mathsf{sumrec}(\pi)$ be the sum of the positions of the strong records in π. For instance, if $\pi = 121231314321$ then $\mathsf{sumrec}(\pi) = 1+2+5+9 = 17$. Let $f_k(x,q)$ be the generating function for the number of partitions of $[n]$ with exactly k blocks according to the statistic sumrec, that is,

$$f_k(x,q) = \sum_{n\geqslant k}\sum_{\pi\in P(n,k)} x^n q^{\mathsf{sumrec}(\pi)}$$

where $P(n,k)$ denotes the set of all partitions of $[n]$ with exactly k blocks. Making use of the decomposition (2) again, we obtain

$$f_k(x,q) = \frac{xq}{1-kx}f_{k-1}(xq,q)$$

with the initial condition $f_1(x,q) = \frac{xq}{1-x}$. Applying this recurrence relation we have

$$f_k(x,q) = \prod_{j=1}^{k}\frac{xq^{k+1-j}}{1-jq^{k-j}x}. \tag{7}$$

Thus the generating function for set partitions into k subsets is given by

$$f_k(x,q) = x^k q^{k(k+1)/2}\prod_{j=1}^{k}(1-jq^{k-j}x)^{-1}, \tag{8}$$

In order to determine the mean of the sumrec parameter, we differentiate with respect to q and set $q = 1$ to obtain

$$x^k \prod_{j=1}^{k} (1-jx)^{-1} \left(\frac{k(k+1)}{2} + \sum_{j=1}^{k} \frac{j(k-j)x}{1-jx} \right). \tag{9}$$

Substitute $x^{-1} = y$ and rewrite this as

$$\prod_{j=1}^{k} (y-j)^{-1} \left(\frac{k(k+1)}{2} + \sum_{j=1}^{k} \frac{j(k-j)}{y-j} \right). \tag{10}$$

The partial fraction decomposition has the form

$$\sum_{m=1}^{k} \left(\frac{a_{k,m}}{(y-m)^2} + \frac{b_{k,m}}{y-m} \right).$$

In order to determine the coefficients $a_{k,m}$ and $b_{k,m}$, we consider the expansion of (10) at $y = m$, which is given by

$$\begin{aligned}
&(y-m)^{-1} \cdot \prod_{\substack{j=1 \\ j\neq m}}^{k} (y-m+m-j)^{-1} \cdot \left(\frac{k(k+1)}{2} + \frac{m(k-m)}{y-m} + \sum_{\substack{j=1 \\ j\neq m}}^{k} \frac{j(k-j)}{y-j} \right) \\
&\quad = (y-m)^{-1} \cdot \prod_{\substack{j=1 \\ j\neq m}}^{k} \left((m-j)^{-1} \left(1 + \frac{y-m}{m-j} \right)^{-1} \right) \\
&\qquad \cdot \left(\frac{k(k+1)}{2} + \frac{m(k-m)}{y-m} + \sum_{\substack{j=1 \\ j\neq m}}^{k} \frac{j(k-j)}{y-j} \right) \\
&\quad = (y-m)^{-1} \cdot \frac{(-1)^{k-m}}{(m-1)!(k-m)!} \cdot \prod_{\substack{j=1 \\ j\neq m}}^{k} \left(1 - \frac{y-m}{m-j} + O\left((y-m)^2\right) \right) \\
&\qquad \cdot \left(\frac{k(k+1)}{2} + \frac{m(k-m)}{y-m} + \sum_{\substack{j=1 \\ j\neq m}}^{k} \frac{j(k-j)}{m-j} + O\left(y-m\right) \right) \\
&\quad = (y-m)^{-1} \cdot \frac{(-1)^{k-m}}{(m-1)!(k-m)!} \cdot \left(1 - \sum_{\substack{j=1 \\ j\neq m}}^{k} \frac{y-m}{m-j} + O\left((y-m)^2\right) \right)
\end{aligned}$$

$$\cdot \left(\frac{k(k+1)}{2} + \frac{m(k-m)}{y-m} + \sum_{\substack{j=1 \\ j\neq m}}^{k} \frac{j(k-j)}{m-j} + O\left(y-m\right) \right)$$

$$= (y-m)^{-1} \cdot \frac{(-1)^{k-m}}{(m-1)!(k-m)!}$$

$$\cdot \left(\frac{m(k-m)}{y-m} + \frac{k(k+1)}{2} + \sum_{\substack{j=1 \\ j\neq m}}^{k} \frac{j(k-j)-m(k-m)}{m-j} + O\left(y-m\right) \right)$$

$$= (y-m)^{-1} \cdot \frac{(-1)^{k-m}}{(m-1)!(k-m)!}$$

$$\cdot \left(\frac{m(k-m)}{y-m} + \frac{k(k+1)}{2} + \sum_{\substack{j=1 \\ j\neq m}}^{k} (m+j-k) + O\left(y-m\right) \right)$$

$$= (y-m)^{-1} \cdot \frac{(-1)^{k-m}}{(m-1)!(k-m)!}$$

$$\cdot \left(\frac{m(k-m)}{y-m} + (m+2)k - 2m + O\left(y-m\right) \right).$$

This shows that

$$a_{k,m} = \frac{m(k-m)(-1)^{k-m}}{(m-1)!(k-m)!} \qquad \text{and} \qquad b_{k,m} = \frac{((m+2)k-2m)(-1)^{k-m}}{(m-1)!(k-m)!}.$$

Passing to the exponential generating function, we have to replace

$$\frac{1}{y-m} = \frac{x}{1-mx} = \sum_{\ell=0}^{\infty} m^{\ell} x^{\ell+1}$$

by

$$\frac{e^{mx}-1}{m} = \sum_{\ell=0}^{\infty} \frac{m^{\ell} x^{\ell+1}}{(\ell+1)!}$$

and similarly $\frac{1}{(y-m)^2}$ by $\frac{e^{mx}(mx-1)+1}{m^2}$. Furthermore, we sum over all k to obtain the bivariate generating function

$$\sum_{k=1}^{\infty} u^k \sum_{m=1}^{k} \frac{m(k-m)(-1)^{k-m}}{(m-1)!(k-m)!} \cdot \frac{e^{mx}(mx-1)+1}{m^2}$$

$$+ \sum_{k=1}^{\infty} u^k \sum_{m=1}^{k} \frac{((m+2)k-2m)(-1)^{k-m}}{(m-1)!(k-m)!} \cdot \frac{e^{mx}-1}{m}.$$

Interchanging the order of summation, we can simplify this as follows:

$$\begin{aligned}
&\sum_{m=1}^{\infty} \frac{e^{mx}(mx-1)+1}{m!} \sum_{k=m}^{\infty} \frac{(-1)^{k-m}(k-m)u^k}{(k-m)!} \\
&\qquad + \sum_{m=1}^{\infty} \frac{e^{mx}-1}{m!} \sum_{k=m}^{\infty} \frac{(-1)^{k-m}(km+2k-2m)u^k}{(k-m)!} \\
&= \sum_{m=1}^{\infty} \frac{e^{mx}(mx-1)+1}{m!} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell}\ell u^{\ell+m}}{\ell!} \\
&\qquad + \sum_{m=1}^{\infty} \frac{e^{mx}-1}{m!} \sum_{\ell=0}^{\infty} \frac{(-1)^{\ell}(m^2+m\ell+2\ell)u^{\ell+m}}{\ell!} \\
&= -\sum_{m=1}^{\infty} \frac{e^{mx}(mx-1)+1}{m!} \cdot u^{m+1}e^{-u} \\
&\qquad + \sum_{m=1}^{\infty} \frac{e^{mx}-1}{m!} \left(m^2u^me^{-u} - (m+2)u^{m+1}e^{-u}\right) \\
&= ue^{u(e^x-1)} \left(ue^x(e^x-x-1) + e^x - 1\right).
\end{aligned}$$

Note that this equals

$$\frac{d}{dx} ue^{u(e^x-1)}(e^x-x-1),$$

and since differentiation of exponential generating functions only means a shift of coefficients, we find that the total of sumrec, summed over all set partitions of n into k subsets, is also exactly the number of non-singletons in all set partitions of $n+1$ into k subsets: note that the exponential generating function for the latter is given by

$$\frac{d}{dv} e^{uv(e^x-x-1)+ux}\Big|_{v=1} = ue^{u(e^x-1)}(e^x-x-1). \tag{11}$$

Let us also provide a purely combinatorial proof of this fact. Consider a set partition of $[n]$ with one marked block; assume that the record associated with this block (in other words, the smallest number in the block) is ℓ. Increase all elements of this block by 1, and add a new element r between 1 and ℓ to the block (which is possible, since all other elements are now $\geqslant \ell+1$). All elements in other blocks are increased by 1 if they are $\geqslant r$, otherwise they remain the same. As a result, we obtain a set partition of $[n+1]$ with the same number of blocks and a marked block that cannot be a singleton block (since we added an element to a nonempty block).

This procedure maps a set partition of $[n]$ with a marked block whose associated record is ℓ to ℓ distinct set partitions of $[n+1]$ with the same number of blocks and a marked non-singleton block. For instance, $15|\underline{24}|3$ is mapped to $26|\underline{135}|4$ and $16|\underline{235}|4$. The process is clearly reversible, and thus provides us with the desired combinatorial proof of the above fact.

4 Another approach to the sumrec parameter

We now give an alternative way to deal with the parameter sumrec. It turns out to be simpler than that of the previous section, however, we believe that the preceding method will be more widely applicable to other parameters of restricted growth functions. Let $s(n,k)$ be defined by

$$s(n,k) := \sum_{\pi \in P(n,k)} q^{\mathsf{sumrec}(\pi)}.$$

A set partition of $[n]$ is obtained by adding n to a set partition of $[n-1]$; if n is added as a singleton, then sumrec changes by n; otherwise, it remains unchanged. Therefore, one has

$$s(n,k) = ks(n-1,k) + q^n s(n-1,k-1).$$

For $q = 1$, one obtains the recursion for the Stirling numbers of the second kind, as it should be. It is now easy to prove— one possible way is induction—the formula

$$\begin{aligned}\mathsf{sr}(n,k) &= \frac{d}{dq} s(n,k)\Big|_{q=1} \\ &= \sum_{\pi \in P(n,k)} \mathsf{sumrec}(\pi) \\ &= kS(n+1,k) - (n+1)S(n,k-1).\end{aligned}$$

This result can also be derived from the generating function (11) of the previous section.

Let $S(q,u,x)$ be the trivariate generating function

$$S(q,u,x) = \sum_{n \geqslant 0} \sum_{k \geqslant 0} \frac{u^k x^n}{n!} s(n,k).$$

Then the recursion for $s(n,k)$ translates to the equation

$$\frac{\partial}{\partial x} S(q,u,x) = u \frac{\partial}{\partial u} S(q,u,x) + quS(q,u,qx),$$

which can probably not be solved explicitly. However, one can determine explicit expressions (albeit complicated) for the derivatives with respect to q at $q = 1$, since this amounts to solving linear partial differential equations of the form

$$S_x(u,x) = uS_u(u,x) + uS(u,x) + f(u,x),$$

which can be done by standard techniques (see [8]). Making use of computer algebra, one obtains

$$\frac{\partial}{\partial q} S(q,u,x)\Big|_{q=1} = ue^{u(e^x-1)} \left(ue^x(e^x - x - 1) + e^x - 1\right)$$

and

$$\frac{\partial^2}{\partial q^2} S(q,u,x)\Big|_{q=1} + \frac{\partial}{\partial q} S(q,u,x)\Big|_{q=1}$$

$$= \frac{u}{2} e^{u(e^x-1)} \Big(2u^3 e^{4x} - u^2 e^{3x} \big(4u(x+1) - 13\big) \\ + 2ue^{2x} \big(u^2(x+1)^2 - u(x^2+8x+10) + 8\big) \\ + e^x \big(u^2(2x^2+10x+7) - 2u(x^2+7x+9) + 2\big) + 2(u-1)\Big).$$

Setting $u = 1$ as well, one finds the mean and variance of sumrec in terms of the Bell numbers B_n:

Theorem 4.1. *The mean and variance of* sumrec*, taken over all partitions of* $[n]$*, are given by*

$$\mathsf{mean} = \frac{B_{n+2} - B_{n+1} - (n+1)B_n}{B_n},$$
$$\mathsf{variance} = \frac{B_{n+4} - \frac{3}{2}B_{n+3} - (2n+\frac{7}{2})B_{n+2} - \frac{1}{2}B_{n+1}}{B_n} - \mathsf{mean}^2.$$

Another application of (6) now shows that the mean is asymptotically $\sim \frac{n^2}{\log^2 n}$, while the variance is $\sim \frac{n^3}{2\log^3 n}$. It follows that the parameter sumrec is concentrated around its mean.

5 Sum of positions of additional weak records in set partitions

Let us finally discuss the parameter sumwrec, defined as the sum of positions of all additional weak records. While it would be possible to follow the approach used in Section 3, we restrict ourselves to a discussion analogous to the preceding section. As n is added to a set partition of $[n-1]$, it becomes a new weak record if and only if it is added to the last block. If we define

$$w(n,k) := \sum_{\pi \in P(n,k)} q^{\mathsf{sumwrec}(\pi)}$$

in analogy to $s(n,k)$, then this yields

$$w(n,k) = (k - 1 + q^n)w(n-1,k) + w(n-1,k-1),$$

from which one can prove easily by induction that

$$\mathsf{swr}(n,k) = \frac{d}{dq} w(n,k)\Big|_{q=1} \\ = \sum_{\substack{\pi \text{ set partition of} \\ [n] \text{ into } k \text{ blocks}}} \mathsf{sumwrec}(\pi) \\ = \frac{(n-1)(n+2)}{2} S(n-1,k).$$

This also implies that the mean of sumwrec over all partitions of $[n]$ is

$$\frac{(n-1)(n+2)B_{n-1}}{2B_n}.$$

Again, we can introduce a trivariate generating function:

$$W(q,u,x) = \sum_{n\geqslant 0}\sum_{k\geqslant 0}\frac{u^k x^n}{n!}w(n,k).$$

Then the recursion for $w(n,k)$ translates to the equation

$$\frac{\partial}{\partial x}W(q,u,x) = u\frac{\partial}{\partial u}W(q,u,x) + qW(q,u,qx) + (u-1)W(q,u,x).$$

Applying the same ideas as in the previous section, we can derive explicit expressions for derivatives with respect to q at $q=1$, so that we also find the variance:

Theorem 5.1. *The mean and variance of* sumwrec*, taken over all partitions of* $[n]$*, are given by*

$$\begin{aligned}\mathsf{mean} &= \frac{(n+2)(n-1)B_{n-1}}{2B_n},\\ \mathsf{variance} &= \frac{(n-1)(2n^2+5n+6)B_{n-1}}{6B_n} + \frac{(n-1)(3n^3+5n^2-10n-24)B_{n-2}}{12B_n} - \mathsf{mean}^2.\end{aligned}$$

Asymptotically, mean and variance are $\sim \frac{n}{2}\log n$ and $\sim \frac{n^2}{3}\log n$, so that the distribution of sumwrec is also concentrated around its mean.

References

[1] E. R. Canfield, Engel's inequality for Bell numbers, *J. Combin. Theory Ser. A* **72** (1995) 184–187.

[2] P. Flajolet and R. Sedgewick, Analytic Combinatorics, Cambridge University Press 2009. (Available at *http://algo.inria.fr/flajolet/Publications/books.html*.)

[3] N. Glick, Breaking records and breaking boards, *Amer. Math. Monthly* **85** (1978) 2–26.

[4] A. Knopfmacher and T. Mansour, Record statistics in a random composition. Submitted.

[5] D. E. Knuth, *The Art of Computer Programming*, volume 1: Fundamental Algorithms, Addison-Wesley, 1968, Third edition, 1997.

[6] I. Kortchemski, Asymptotic behavior of permutation records. arXiv: 0804.044v1, pages: 12, 2008.

[7] A. Myers and H. Wilf, Left-to-right maxima in words and multiset permutations, *Israel J. Math.*, to appear.

[8] A. D. Polyanin, V. F. Zaitsev and A. Moussiaux, *Handbook of first order partial differential equations*, volume 1 of *Differential and Integral Equations and Their Applications*, Taylor & Francis Ltd., London, 2002.

[9] H. Prodinger, Combinatorics of geometrically distributed random variables: Left-to-right maxima, *Discr. Math.* **153** (1996) 253–270.

[10] H. Prodinger, Records in geometrically distributed words: sum of positions, *Appl. Anal. Discr. Math.* **2** (2008) 234–240.

[11] A. Rényi, Théorie des éléments saillants d'une suite d'observations, *Ann. Fac. Sci. Univ. Clermont-Ferrand* **8** (1962) 7–13.