Record statistics in integer compositions

Arnold Knopfmacher, Toufik Mansour

FPSAC 2009, Hagenberg, Austria DMTCS proc. **AK**, 2009, 527–536

Record statistics in integer compositions

Arnold Knopfmacher[1†] and Toufik Mansour[2]

[1]*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
[2]*Department of Mathematics, University of Haifa, 31905 Haifa, Israel*
toufik@math.haifa.ac.il

A *composition* $\sigma = a_1 a_2 \ldots a_m$ of n is an ordered collection of positive integers whose sum is n. An element a_i in σ is a strong (weak) *record* if $a_i > a_j$ ($a_i \geq a_j$) 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 compositions of n, as well as for the sum of the positions of the records in all compositions of n, where the parts a_i belong to a fixed subset A of the natural numbers. In particular when $A = \mathbb{N}$, we find the asymptotic mean values for the number, and for the sum of positions, of records in compositions of n.

Keywords: Composition, Record, Left-to-right maxima, Generating function, Mellin transforms, Asymptotic estimates

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 was first studied by Rényi [13], compare also [7]. A survey of results on this topic can be found in [2]. In the literature records are also referred to as a *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 [8] defines a new statistic $srec$, where $srec(\pi)$ is the sum over the positions of all records in π. For instance, the permutation $\pi = 451632$ has 3 records $4, 5, 6$ and $srec(\pi) = 1 + 2 + 4 = 7$.

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 [12] 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 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, must be strictly larger than elements to the left) and *weak record* is an element $a_i \geq a_j$ 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). We denote the sum of

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

the positions of all strong (respectively, weak) records in a word a by $ssrec$ (respectively, $wsrec$). In [12], Prodinger found the expected value of the sum of the positions of strong records, in random geometrically distributed words of length n. Previously, Prodinger [10] also studied the number of strong and weak records, in samples of geometrically distributed random variables. He also studied further properties of such records in papers such as [11] and references therein.

A *composition* $\sigma = \sigma_1\sigma_2\ldots\sigma_m$ of n is an ordered collection of positive integers whose sum is n. Thus a composition σ of n with parts in A is a restricted word over the alphabet A. We denote the set of all compositions of n with m parts in A by $C_A(n,m)$. It is well known that the number of compositions of $n \geq 1$ with m parts in $\mathbb{N}$ is given by $\binom{n-1}{m-1}$ and that the total number of compositions of n is 2^{n-1}.

In this paper we find generating functions for these parameters, number of strong records, number of weak records, sum of positions of strong records, and sum positions of weak records in a random composition of n with parts in $A = [d] := \{1,2,\ldots,d\}$ or $A = \mathbb{N}$. We also study the mean values of these parameters as $n \to \infty$ in the case $A = \mathbb{N}$ by means of rational function asymptotics and Mellin transforms. Details of some of the lengthier proofs will be left to the full version of the paper. We remark that in [5], an asymptotic correspondence is established between compositions of n and samples of geometric variable of parameter $p = 1/2$ and length $n/2$. By exploiting this correspondence, and using the already established results of Prodinger for samples of geometric random variables, alternative derivations of our asymptotic results can be obtained.

1 Number strong records and weak records

Let $NSR_A(z,y,q)$ and $NWR_A(z,y,q)$ be the generating function for the number of compositions of n with m parts in A according to the number of strong and weak records, respectively, that is,

$$NSR_A(z,y,q) = \sum_{n,m\geq 0}\sum_{\sigma\in C_A(n,m)} z^n y^m q^{nsr(\sigma)},$$
$$NWR_A(z,y,q) = \sum_{n,m\geq 0}\sum_{\sigma\in C_A(n,m)} z^n y^m q^{nwr(\sigma)},$$

where $nsr(\sigma)$ and $nwr(\sigma)$ is the number of strong and weak records in composition σ, respectively. In this section we find an explicit formulas for those generating functions.

Theorem 1.1 *The generating function $NSR_{[d]}(z,y,q)$ is given by*

$$NSR_{[d]}(z,y,q) = \prod_{j=1}^{d}\left(1 + \frac{z^j yq}{1 - y\sum_{i=1}^{j} z^i}\right).$$

Proof: We denote the number of occurrences of the part d in the composition $\sigma \in C_{[d]}(n,m)$ by $\ell(\sigma)$. Now let us write equation for the generating function $NSR_{[d]}(z,y,q)$. The contribution of the case $\ell(\sigma) = 0$ is given by $NSR_{[d-1]}(z,y,q)$. Assume $\ell(\sigma) > 0$, then σ can be decomposed as $\sigma' d\sigma''$, where σ' is a composition with parts in $[d-1]$ and σ'' is a composition with parts in $[d]$. Thus, the contribution of the case $\ell(\sigma) > 0$ equals $z^d yqNSR_{[d-1]}(z,y,q)NSR[d](z,y,1)$. Therefore,

$$NSR_{[d]}(z,y,q) = NSR_{[d-1]}(z,y,q) + z^d yqNSR_{[d-1]}(z,y,q)NSR_{[d]}(z,y,1).$$

For $q=1$ and by induction we have that

$$NSR_{[d]}(z,y,1)=\frac{1}{1-y\sum_{j=1}^{d}z^j}.$$

Hence,

$$NSR_{[d]}(z,y,q)=\prod_{j=1}^{d}\left(1+\frac{z^jyq}{1-y\sum_{i=1}^{j}z^i}\right),$$

as claimed. □ Theorem 1.1 with $q=1$ gives that the generating function for the number of compositions of n with m parts in $[d]$ is given by $NSR_{[d]}(z,y,1)=\frac{1}{1-y\sum_{i=1}^{d}z^j}$, see for example [4].

Also from Theorem 1.1 we get that

$$\begin{aligned}\frac{\partial}{\partial q}NSR_{[d]}(z,y,1)&=\prod_{j=1}^{d}\left(1+\frac{z^jy}{1-y\sum_{i=1}^{j}z^i}\right)\left(\sum_{j=1}^{d}\frac{z^jy}{1-y\sum_{i=1}^{j-1}z^i}\right)\\&=\frac{1}{1-y\sum_{i=1}^{d}z^i}\left(\sum_{j=1}^{d}\frac{z^jy}{1-y\sum_{i=1}^{j-1}z^i}\right).\end{aligned}$$

Hence, the generating function for the number strong records in all compositions of n with parts in $\mathbb{N}$ is given by

$$f(z):=\frac{1}{1-\sum_{i\geq1}z^i}\sum_{j\geq1}\frac{z^j}{1-\sum_{i=1}^{j-1}z^i}=\frac{1-z}{1-2z}\sum_{j\geq1}\frac{z^j}{1-\sum_{i=1}^{j-1}z^i}.$$

Theorem 1.2 *The average number E_n^s of strong left-to-right maxima in the context of compositions of n has the asymptotic expansion*

$$E_n^s=\frac{1}{2}\left[\log_2 n-\frac{1}{2}+\frac{\gamma}{L}-\delta\left(\log_2 n\right)\right]+o(1).$$

Here and in the rest of the paper, $L=\log 2$; γ is Euler's constant and $\delta(x)$ is a periodic function of period 1 and mean 0 and small amplitude, which is given by the Fourier series

$$\delta(x)=\frac{1}{L}\sum_{k\neq0}\Gamma\left(-\chi_k\right)e^{2k\pi ix}.$$

The complex numbers χ_k are given by $\chi_k=2k\pi i/L$.

Proof: Firstly by summing the finite geometric series and using partial fraction decomposition,

$$f(z)=\frac{z-z^2}{1-2z}+(1-z)^2\sum_{k\geq2}\left[\frac{1}{1-2z}-\frac{1}{1-2z+z^k}\right].$$

Hence the average number E_n^s of strong left-to-right maxima in compositions of n satisfies

$$E_n^s = \frac{1}{2^{n-1}}[z^n]f(z) = \frac{1}{2} + \frac{1}{2^{n-1}}[z^n](1-z)^2 \sum_{k=2}^{n} \left[\frac{1}{1-2z} - \frac{1}{1-2z+z^k} \right].$$

Let ρ_k be the smallest positive root of the denominator polynomial $1-2z+z^k$ that lies between $1/2$ and 1. An application of the principle of the argument or Rouche's Theorem shows such a root to exist with all other roots of modulus greater than $3/4$. By dominant pole analysis,

$$q_{n,k} := [z^n]\frac{(1-z)^2}{1-2z+z^k} = c_k \rho_k^{-n} + O\left(\left(\frac{4}{3}\right)^n\right) \quad \text{with} \quad c_k = \frac{(1-\rho_k)^2}{\rho_k(2-k\rho_k^{k-1})},$$

for large n but fixed k. The denominator polynomial $1-2z+z^k$ behaves like a perturbation of $1-2z$ near $z=1/2$. By "bootstrapping" we find that

$$\rho_k = \frac{1}{2} + 2^{-k-1} + O(k2^{-2k})$$

and hence $c_k = \frac{1}{4} + O(k2^{-k})$. The use of this approximation can be justified for a wide range of values of k and n (see for example [3] or [6]).

Let us now restrict our attention to those k for which $n^{-3} \leq 2^{-k} \leq \frac{\log n}{n}$. For such k we can show that

$$q_{n,k} = 2^{n-2}\left(\exp\left(-\frac{n}{2^k}\right) + O\left(\frac{\log^3 n}{n}\right)\right). \tag{1}$$

Turning next to smaller values of $k \geq 2$, that is, k such that $2^{-k} > \frac{\log n}{n}$, we find that now the coefficients $q_{n,k}$ are relatively small, since for such k, $q_{n,k} = O(\frac{2^n}{n})$ as $n \to \infty$. Finally we must consider larger values of $k \leq n$ that is, k for which $n^{-3} > 2^{-k}$, or equivalently, $k \geq 3\log_2 n$. In this range we find that

$$q_{n,k} = 2^{n-2}\left(\exp\left(-\frac{n}{2^k}\right) + O\left(\frac{1}{n^2}\right)\right). \tag{2}$$

Then combining the estimates for $q_{n,k}$ over the range $2 \leq k \leq n$ above,

$$\begin{aligned} E_n^s - \frac{1}{2} &= \frac{1}{2}\sum_{k=2}^{n}\left(1 - \frac{q_{n,k}}{2^{n-2}}\right) \\ &\sim \frac{1}{2}\sum_{k\geq 0}\left(1 - \exp\left(-\frac{n}{2^k}\right)\right) - 1, \end{aligned}$$

as the additional tail sum $\sum_{k>n}\left(1-\exp\left(-\frac{n}{2^k}\right)\right)$ is exponentially small. It remains to estimate

$$h(n) := \sum_{k\geq 0}\left(1 - \exp\left(-\frac{n}{2^k}\right)\right),$$

as $n \to \infty$. For this we use Mellin transforms and find (see [1, Appendix B.7, equation (48)])

$$h(n) = \log_2 n + \frac{1}{2} + \frac{\gamma}{L} - \delta\left(\log_2 n\right) + O(1/n). \tag{3}$$

The asymptotic estimate for E_n^s follows. $\square$

Remarks Asymptotically we find that the expected number of strict left-to-right maxima is half the expected size of the *largest part* in a random composition of n (see [9]). Also, as mentioned in the introduction, the asymptotic correspondence established in [5] would allow one to use the results of Prodinger [10] in the case $p = 1/2$, to give an alternative proof of Theorem 1.2.

A similar approach to that of Theorem 1.3 leads to

Theorem 1.3 *The generating function $NWR_{[d]}(z, y, q)$ is given by*

$$NWR_{[d]}(z, y, q) = \prod_{j=1}^{d} \frac{1}{1 - \frac{z^j yq}{1-y\sum_{i=1}^{j-1} z^i}}.$$

The generating function for the total number of weak records in compositions over $\mathbb{N}$ is then

$$g(z) := \left.\frac{\partial NWR_{[\mathbb{N}]}(z, 1, q)}{\partial q}\right|_{q=1} = \frac{1-z}{1-2z} \sum_{k\geq 1} \frac{z^k}{1 - \sum_{i=1}^{k} z^i}$$
$$= \frac{(1-z)^2}{z} \sum_{k\geq 2} \left[\frac{1}{1-2z} - \frac{1}{1-2z+z^k}\right].$$

Theorem 1.4 *The average number E_n^w of weak left-to-right maxima in the context of compositions of n has the asymptotic expansion*

$$E_n^w = \log_2 n - \frac{3}{2} + \frac{\gamma}{L} - \delta\left(\log_2 n\right) + o(1).$$

Proof: The average number E_n^w of weak left-to-right maxima in compositions of n satisfies

$$E_n^w = \frac{1}{2^{n-1}}[z^n]g(z) = \frac{1}{2^{n-1}}[z^{n+1}](1-z)^2 \sum_{k=2}^{n} \left[\frac{1}{1-2z} - \frac{1}{1-2z+z^k}\right].$$

Then using the $q_{n,k}$ notation in the proof of Theorem 1.2

$$E_n^w = \sum_{k=2}^{n+1} \left(1 - \frac{q_{n+1,k}}{2^{n-1}}\right) = 2E_{n+1}^s - 1.$$

The asymptotic estimate then follows from that of Theorem 1.2. $\square$

2 The statistics $ssrec$ and $wsrec$ on the set of compositions

Let $NSR_A(z, y, q)$ and $NWR_A(z, y, q)$ be the generating function for the number of compositions of n with m parts in A according to the statistic $ssrec$ and $wsrec$, respectively, that is,

$$PSR_A(z, y, q) = \sum_{n,m\geq 0} \sum_{\sigma\in C_A(n,m)} z^n y^m q^{ssrec(\sigma)},$$
$$PWR_A(z, y, q) = \sum_{n,m\geq 0} \sum_{\sigma\in C_A(n,m)} z^n y^m q^{wsrec(\sigma)}.$$

Theorem 2.1 *The generating function $PSR_{[d]}(z,y,q)$ is given by*

$$1+\sum_{k=1}^{d} q^k \left(\sum_{d\geq j_1>j_2>\cdots>j_k\geq 1} \prod_{i=1}^{k} \frac{z^{j_i} y q^{i-1}}{1-yq^{i-1}\sum_{\ell=1}^{j_i} z^\ell} \right).$$

Proof: We denote the number of occurrences of the part d in the composition $\sigma \in C_{[d]}(n,m)$ by $\ell(\sigma)$. Decomposing according to $\ell(\sigma)=0$ and $\ell(\sigma)>0$ leads to

$$PSR_{[d]}(z,y,q) = PSR_{[d-1]}(z,y,q) + z^d yq PSR_{[d-1]}(z,qy,q) PSR_{[d]}(z,y,1). \tag{4}$$

For $q=1$, $PSR_{[d]}(z,y,1)=\frac{1}{1-y\sum_{j=1}^{d} z^j}$. Hence,

$$\begin{aligned} PSR_{[d]}(z,y,q) &= PSR_{[d-1]}(z,y,q) + \frac{z^d yq}{1-y\sum_{i=1}^{d} z^i} PSR_{[d-1]}(z,qy,q) \\ &= PSR_{[d-2]}(z,y,q) + \sum_{j=d-1}^{d} \frac{z^j yq}{1-y\sum_{i=1}^{j} z^i} PSR_{[j-1]}(z,qy,q) \\ &\vdots \\ &= 1+\sum_{j=1}^{d} \frac{z^j yq}{1-y\sum_{i=1}^{j} z^i} PSR_{[j-1]}(z,qy,q). \end{aligned}$$

Iterating the above recurrence relation d times we get the desired result. □ From this we derive

Corollary 2.2 *The generating function $v_d(z)=\frac{\partial}{\partial q} PSR_{[d]}(z,1,q)\mid_{q=1}$ is given by*

$$\frac{z}{1-\sum_{j=1}^{d} z^j} \sum_{j=0}^{d-1} \frac{z^j}{\left(1-\sum_{i=1}^{j} z^i\right)^2}.$$

The above corollary gives that the generating function for the number of compositions of n according to the total of the statistic $ssrec$ is given by

$$v(z) := \frac{z(1-z)}{1-2z} \sum_{j\geq 0} \frac{z^j}{\left(1-\sum_{i=1}^{j} z^i\right)^2}.$$

The rather lengthy proof of the asymptotic behaviour of the coefficients of $v(z)$ will be left for the journal version of the paper. We obtain

Theorem 2.3 *The average sum of the positions of the strong records e_n^s in compositions of n has the asymptotic expansion*

$$e_n^s = \frac{n}{4\log 2}\big(1+\delta_2(\log_2 n)\big) + o(n),$$

where $\delta_2(x)$ is a periodic function of period 1, mean zero and small amplitude, which is given by the Fourier series

$$\delta_2(x) = \sum_{k\neq 0} \chi_k \Gamma\left(-1-\chi_k\right) e^{2k\pi i x}.$$

With reference again to [5], Theorem 2.3 is seen to correspond to the $p = 1/2$ case of the results of Prodinger [12].

The corresponding results for $PWR_{[d]}(z, y, q)$ are as follows.

Theorem 2.4 *The generating function $PWR_{[d]}(z, y, q)$ satisfies the following recurrence relation*

$$PWR_{[d]}(z,y,q) = PWR_{[d-1]}(z,y,q) + \frac{z^d yq}{1 - y\sum_{i=1}^{d-1} z^i} PWR_{[d]}(z,qy,q).$$

Corollary 2.5 *The generating function $w_d(z) = \frac{\partial}{\partial q} PWR_{[d]}(z,1,q) \mid_{q=1}$ is given by*

$$w_d(z) = \frac{1}{1-\sum_{j=1}^{d} z^j} \sum_{j\geq 1} \frac{z^j}{\left(1-\sum_{i=1}^{j} z^i\right)^2}.$$

The above corollary gives that the generating function for the number of compositions of n according to the total of the statistic $swrec$ is given by

$$w(z) := \frac{1-z}{1-2z} \sum_{j\geq 1} \frac{z^j}{\left(1-\sum_{i=1}^{j} z^i\right)^2}.$$

Theorem 2.6 *The average sum of the positions of the weak records e_n^w in compositions of n has the asymptotic expansion*

$$e_n^w = \frac{n}{2\log 2}\left(1+\delta_2\left(\log_2 n\right)\right) + o(n),$$

where $\delta_2(x)$ is the same periodic function that occured in Theorem 2.3.

Proof: The generating functions $v(z)$ and $w(z)$ are related as follows,

$$v(z) = \frac{z(1-z)}{1-2z} + zw(z).$$

From this we see that

$$[z^{n+1}]v(z) = 2^{n-1} + [z^n]w(z).$$

So that $e_n^w = 2e_{n+1}^s - 1$. The result then follows from Theorem 2.3. □

Now, our aim is to present a combinatorial explanation for the fact that the number (sum) of the positions of weak records in all compositions of n plus 2^{n-1} equals the number (sum) of the positions of strong records in all compositions of $n+1$, for $n \geq 1$. In order to do that we need the following notations.

Let $sw_{n,r}$ (respectively, $ss_{n,r}$) be the sum of r-th power of the positions of weak (respectively, strong) records in all the compositions of n, namely,

$$sw_{n,r} = \sum_{\sigma\in C_n} \sum_{\sigma_i \text{ is a weak record of } \sigma} i^r,$$
$$ss_{n,r} = \sum_{\sigma\in C_n} \sum_{\sigma_i \text{ is a strong record of } \sigma} i^r,$$
$$sw'_{n,r} = \sum_{\sigma\in C_n(A)} \sum_{\sigma_i \text{ is a weak record of } \sigma, i>1} i^r,$$
$$ss'_{n,r} = \sum_{\sigma\in C_n(A)} \sum_{\sigma_i \text{ is a strong record of } \sigma, i>1} i^r,$$

where $C_n = \cup_{m=1}^n C_{n,m}$ is the set of all compositions of n. From the definitions, each first letter is a weak (strong) record. Therefore,

$$sw_{n,r} = |C_n| + sw'_{n,r} \text{ and } ss_{n,r} = |C_n| + ss'_{n,r}, \tag{5}$$

where $|C_n| = 2^{n-1}$ is the number of compositions of n.

Theorem 2.7 *For all $n \geq 1$,*

$$ss_{n+1,r} = sw_{n,r} + 2^{n-1}.$$

Proof: It is not hard to see that $\sigma_1 \cdots \sigma_m$ is a composition of n and σ_i, $i > 1$, is a weak record if and only if $\sigma_1 \cdots \sigma_{i-1}(\sigma_i + 1)\sigma_{i+1} \cdots \sigma_m$ is a composition of n and $\sigma_i + 1$, $i > 1$, is a strong record. Therefore, the multiset of all positions i, $i > 1$, of the weak records in all compositions of n is the same multiset as all positions i, $i > 1$, of the strong records in all compositions on $n + 1$. In other words, $ss'_{n+1,r} = sw'_{n,r}$ for all n and r.

Hence, by (5) we have

$$ss_{n+1,r} = 2^n + ss'_{n+1,r} = 2^n + sw'_{n,r} = 2^{n-1} + 2^{n-1} + sw'_{n,r} = 2^{n-1} + sw_{n,r},$$

as requested. □

References

[1] P. Flajolet and R. Sedgewick, Analytic Combinatorics, Cambridge University Press 2008. (Web edition available at *http://algo.inria.fr/flajolet/Publications/books.html*.)

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

[3] X. Gourdon and H. Prodinger, A generating function approach to random subgraphs of the n-cycle, *Discr. Math.* **169** (1997) 227–232.

[4] S. Heubach and T. Mansour, Compositions of n with parts in a set, *Congr. Numer.* **168** (2004) 127–143.

[5] P. Hitczenko, and G. Louchard, Distinctness of compositions of an integer: a probabilistic analysis, *Random Structures Algorithms* **19** (2001), 407–437.

[6] D. E. Knuth, The average time for carry propagation, *Indagationes Math.* **40** (1978) 238–242.

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

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

[9] A. M. Odlyzko and B. Richmond, On the compositions of an integer, *Lecture Notes in Math.* **829** (1980) 199–210.

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

[11] H. Prodinger, Combinatorics of geometrically distributed random variables: value and position of large left-to-right maxima. *Discrete Math.* **254** (2002), 459–471.

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

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

[14] W. Szpankowski, Average Case Analysis of Algorithms on Sequences, Wiley, 2001.