Discrete Mathematics 338 (2015) 236–247

Contents lists available at ScienceDirect

Discrete Mathematics

journal homepage: www.elsevier.com/locate/disc

Longest run of equal parts in a random integer composition

Ayla Gafni *

Algorithms Project, INRIA-Rocquencourt, F78153 Le Chesnay, France

ARTICLE INFO

Article history:
Received 28 May 2014
Received in revised form 24 August 2014
Accepted 4 October 2014
Available online 23 October 2014

Keywords:
Asymptotic combinatorics
Enumerative combinatorics
Probabilistic combinatorics
Integer compositions
Residue analysis

ABSTRACT

This paper examines a problem in enumerative and asymptotic combinatorics involving the classical structure of integer compositions. What is sought is an analysis on average and in distribution of the length of the longest run of consecutive equal parts in a composition of size n. The problem was posed by Herbert Wilf at the Analysis of Algorithms conference in July 2009 (see arXiv:0906.5196).

1. Introduction

A *composition* of an integer n is a sequence $(x_1, \ldots, x_m)$ of positive integers such that

$$n = x_1 + \cdots + x_m, \quad \text{and} \quad x_i \geq 1.$$

The x_i are called the *parts* and n is the *size* of the composition. We wish to know the *length of the longest run of equal parts* (which we denote by the random variable L) in a random composition of size n. For instance, the composition

$$3, 2, 1, 4, 4, 4, 4, 4, 7, 3, 5, 5, 4, 2$$

has $L = 5$. A composition with $L = 1$ is known as a Carlitz composition. The characteristics of Carlitz compositions and their generating function $C^{\langle 2 \rangle}(z)$ (see Proposition 1) are studied in great detail in [8,10].

Integer compositions have received a lot of attention in recent years. In [2], Brennan and Knopfmacher study the distribution of the number of large ascents in an integer composition. Falah and Mansour complete the study of ascents in [3] by examining small ascents. In the context of ascents, descents, and levels, finding the size of the longest run of equal parts amounts to finding the longest string of consecutive levels in the composition. The longest run in a random word of length n has been studied in depth by Grabner et al. in [6]. The problem of analyzing the longest run in an integer composition is similar, but is complicated by the fact that we are working with an infinite "alphabet" (i.e., the natural numbers) and our "letters" are weighted by their numerical value.

The solution to the longest run problem can be broken down into four main parts. In Section 2, we find a family of generating functions for integer compositions that keeps track of the longest run of equal parts. In Section 3, we analyze the generating functions using singularity analysis to find an asymptotic estimate of the number of compositions of size n with no run of length k. In Section 4, we use that estimate to describe the probability distribution of the random variable L, and in Section 5, we calculate the mean and variance of the distribution. The analysis here has some similarities to the analytic

* Correspondence to: Pennsylvania State University, Department of Mathematics, 109 McAllister Bldg, 16802 University Park, PA, United States.
E-mail address: gafni@math.psu.edu.

http://dx.doi.org/10.1016/j.disc.2014.10.003

treatment of compositions in [1,8,10], and the methods and notation used in this paper are detailed in the book *Analytic Combinatorics* by Flajolet and Sedgewick [5].

Throughout this paper, we will use log to denote the natural logarithm, and lg to denote the base 2 logarithm. We will also use $\log^m(x)$ and $\lg^m(x)$ to denote an iterated logarithm. That is, $\log^3(x) = \log\log\log(x)$.

2. Enumerative aspects of compositions

The enumeration of integer compositions is easily solved using basic combinatorics. We can create a graphical model of a composition by representing the integers in unary using small discs ("•") and drawing bars between some of the balls. The following is an example using the composition $2+3+1+1+3=10$:

$$\bullet\bullet \mid \bullet\bullet\bullet \mid \bullet \mid \bullet \mid \bullet\bullet\bullet\,.$$

Using this "balls-and-bars" model, we see that the number of compositions of the integer n is $C_n = 2^{n-1}$, since a composition can be viewed as the placement of separation bars at a subset of the $n-1$ spaces between the balls.

We can also find the enumeration of compositions with the symbolic method [5, p. 40]. If the integers are represented in unary, then the combinatorial class of positive integers ($\mathcal{I}$) can be thought of as a sequence of atoms ($\mathcal{Z}$) so that

$$\mathcal{I} = \textsc{Seq}_{\geq 1}(\mathcal{Z}) \Longrightarrow I(z) = \frac{z}{1-z}.$$

Since an integer composition is simply a sequence of positive integers, we can easily derive the generating function for the class $\mathcal{C}$ of compositions from the specification

$$\mathcal{C} = \textsc{Seq}\ (\mathcal{I}) \Longrightarrow C(z) = \frac{1}{1-I(z)} = \frac{1}{1-\frac{z}{1-z}} = \frac{1-z}{1-2z}.$$

Throughout this note, we let $[z^n]f(z)$ be the coefficient of z^n in the expansion of $f(z)$ at 0:

$$[z^n]\sum_n f_n z^n = f_n.$$

We find that our result using the symbolic method is consistent with the above combinatorial argument, since

$$[z^n]C(z) = [z^n]\frac{1}{1-2z} - [z^n]\frac{z}{1-2z} = 2^n - 2^{n-1} = 2^{n-1}.$$

Now that we have a generating function for *all* integer compositions, we need another generating function for compositions, which keeps track of the longest run of equal parts. We begin by examining Smirnov words, i.e., words over an m-ary alphabet such that no letter occurs twice in a row. Words over the m-ary alphabet $\{a_1, \ldots, a_m\}$ can be represented by the multivariate generating function

$$W(x_1, \ldots, x_m) = \frac{1}{1-(x_1+\cdots+x_m)},$$

where x_j marks the number of times the letter a_j occurs in a word. That is, the expression $[x_1^{n_1}, \ldots, x_m^{n_m}]W(x_1, \ldots, x_m)$ denotes the number of words in which the letter a_1 occurs n_1 times, a_2 occurs n_2 times, and so on.

Similarly, let $S(y_1, \ldots, y_m)$ be the generating function for Smirnov words, where y_j marks the number of times the letter a_j occurs in a word. Now, given a Smirnov word, one can obtain *any* word by replacing a_j with a nonempty sequence of a_j (i.e., $a_j \times \textsc{Seq}\ (a_j)$). In terms of generating functions, this translates to

$$W(x_1, \ldots, x_m) = S\left(\frac{x_1}{1-x_1}, \ldots, \frac{x_m}{1-x_m}\right).$$

We use this to find the generating function for Smirnov words in terms of the generating function for all words:

$$S(y_1, \ldots, y_m) = W\left(\frac{y_1}{1+y_1}, \ldots, \frac{y_m}{1+y_m}\right),$$

so that we have

$$S(y_1, \ldots, y_m) = \left(1 - \sum_{j=1}^{m} \frac{y_j}{1+y_j}\right)^{-1}.$$

We would like to generalize $S(y_1, \ldots, y_m)$ to the generating function $S^{\langle k\rangle}(y_1, \ldots, y_m)$ for words over an m-ary alphabet such that no letter occurs k times in a row. We can obtain this via the substitution

$$y_j \longmapsto \sum_{i=1}^{k-1} y_j^i = \frac{y_j - y_j^k}{1-y_j},$$