Discrete Mathematics 311 (2011) 723–731

Contents lists available at ScienceDirect

Discrete Mathematics

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

The largest missing value in a composition of an integer

Margaret Archibald [a,*], Arnold Knopfmacher [b]

[a] *Laboratory of Foundational Aspects of Computer Science, Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa*

[b] *The John Knopfmacher Centre for Applicable Analysis and Number Theory, University of the Witwatersrand, P.O. Wits, 2050 Johannesburg, South Africa*

ARTICLE INFO

Article history:
Received 14 June 2010
Received in revised form 30 November 2010
Accepted 12 January 2011
Available online 16 February 2011

Keywords:
Compositions
Generating functions
Geometric random variables
Rice's method
Largest missing value

ABSTRACT

In this paper we find, asymptotically, the mean and variance for the largest missing value (part size) in a composition of an integer n. We go on to show that the probability that the largest missing value and the largest part of a composition differ by one is relatively high and we find the mean for the average largest value in compositions that have this property. The average largest value of compositions with at least one non-zero missing value is also found, and used to calculate how many distinct values exceed the largest missing value on average.

1. Introduction

A composition of n is an ordered partition of the integer n — i.e., it is a sequence of positive integer values whose sum is n, where the order of the values is important. For convenience we denote the composition $a_1 + a_2 + \cdots + a_k = n$ by the word $a_1 a_2 \cdots a_k$ The compositions of 4 are 4, 31, 13, 22, 211, 121, 112, and 1111.

Recently the subject of gaps and missing values in compositions, as well as in the related topic of geometric random variables, has attracted a lot of interest. See [3–7].

In this paper we are interested in the largest value which does not occur in a composition of n. That is, the largest part size smaller than the maximum part that is not present in the composition. It will be an element of the set $\{0, 1, \ldots, n-1\}$, since we take the largest missing value of a complete composition (a composition with no missing values) to be 0. The compositions of 4 which are complete are 211, 121, 112, and 1111. The compositions of 4 that are non-complete are 4, 31, 13, and 22, which have non-zero largest missing values 3, 2, 2 and 1 respectively. Complete compositions have been studied in [4].

As an example, the composition 1131,617,221,131 of 30 with largest part 7 has two missing values, since both 4 and 5 do not occur in the sample, so 5 is the largest missing value in this composition. As the largest value is 7, the number of values larger than the largest missing value is 2.

If the largest value which is missing is exactly one smaller than the largest part, then we say that the composition has the LMV property (the 'largest missing value' property). For example, the composition 312,311,621 of 20 has largest part 6 and largest missing value 5 and thus has the LMV property. We will show asymptotically that over 80% of non-complete compositions have this property.

* Corresponding author. Tel.: +27 21 650 3206; fax: +27 86 634 0327.
E-mail addresses: margaret.archibald@uct.ac.za, zigarch@gmail.com (M. Archibald), arnold.knopfmacher@wits.ac.za (A. Knopfmacher).

0012-365X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.disc.2011.01.012

The proofs in this paper stem from the link between compositions and samples of geometric random variables with parameter $p = \frac{1}{2}$. The relationship which allows us to reduce this problem to one for geometric random variables is explained fully in [4–6], and uses a probabilistic argument by applying the uniform probability measure to the set of all compositions of n.

We are now able to use analytic tools on geometric samples to produce asymptotic results for compositions of n as $n \to \infty$. Let the sample of geometric random variables be given by $(\Gamma_1, \Gamma_2, \ldots, \Gamma_n)$ where $\mathbb{P}\{\Gamma_j = i\} = 2^{-i}$, for $1 \le j \le n$ and $i \in \mathbb{N}$. Then we want to find the largest value which does not appear in the sequence $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$ as long as there is at least one larger value which does appear. Analogously to the case of compositions, if the geometric sample is 'complete' (has all values from 1 up to the maximum occurring in the sample) the largest missing value is taken to be 0, and a sample has the LMV property if the largest value and the largest missing value differ by exactly 1.

In all cases, we use $\delta_*(x)$ to be a periodic function with a mean of 0, period 1 and small amplitude.

Theorem 1. *The average largest missing value in a non-complete composition is asymptotic to (as $n \to \infty$)*

$$\log_2 n + \frac{\gamma}{L} - \frac{1}{2} + \delta_E(\log_2 n), \tag{1}$$

where

$$\delta_E(x) := -\frac{1}{L} \sum_{k \neq 0} \Gamma(-\chi_k) e^{2k\pi i x}.$$

Theorem 2. *The variance for the largest missing value in a non-complete composition satisfies (as $n \to \infty$)*

$$\mathbb{V}(\lambda_n^{(g1)}) \sim 1 + \frac{2}{L} \sum_{h \ge 1} \frac{(-1)^{h-1}}{h(2^h - 1)} + \delta(\log_2 n), \tag{2}$$

where

$$\delta(x) := -\frac{2}{L} \sum_{l \ge 1} \frac{(-1)^l}{l!(2^l - 1)} \sum_{k \neq 0} \Gamma(-\chi_k + l) e^{2k\pi i x}.$$

The layout of the paper is as follows: In Section 3 we discuss the largest missing value and prove Theorems 1 and 2. We look at probabilities in Section 4, and then go on to discuss the average largest value of LMV compositions in Section 5. We conclude with a discussion on the number of values larger than the largest missing value in Section 6. Finally, in Section 7, we give a comparison of the mean largest parts of different classes of compositions and refer to ongoing work concerning the case of geometric random variables with $p \neq \frac{1}{2}$.

In the following table, the columns are labelled A to G and the meaning of each column is:

$A =$ the number of non-complete compositions of n

$B =$ the proportion of non-complete compositions of n

$C =$ the number of compositions of n with the LMV property

$D =$ the proportion of compositions of n with the LMV property

$E =$ the sum of the largest missing values

$F =$ the average largest missing value for non-complete compositions

$G =$ the average difference between the maximum value and the largest missing value for non-complete compositions.

2. Notation

In this paper we use the following notation:

$L := \log 2$

$\chi_k := \dfrac{2k\pi i}{L}$, where $k \in \mathbb{Z}$ and $k \neq 0$

$H_k := \displaystyle\sum_{i=1}^{k} \frac{1}{i}$ (denotes the kth harmonic number)

$\gamma := 0.577\ldots$ (denotes Euler's constant)

GEOM $\left(\dfrac{1}{2}\right)$ denotes a geometric random variable of parameter $p = \dfrac{1}{2}$.

3. Geometric samples with largest missing value *k*

In order to derive asymptotic estimates, it will be convenient to adopt a probabilistic viewpoint. That is, rather than think of the proportion of compositions with a given property we will equip the set of all compositions of n with the uniform probability measure and will be interested in the probability that a randomly chosen composition of n has this property. In that setting, compositions of n are closely related to the special case for geometric random variables when $p = 1/2$, as shown in [4–6]. It follows from the methods of [4–6] that the results obtained in this paper for compositions of n are asymptotically the same as those for geometric samples of length $n/2$.

In this section we find the expectation and variance of the largest value which is absent from the composition, assuming there is at least one larger value which does occur. We present two theorems which are proved by the use of generating functions, residue calculus and asymptotic analysis.

If the largest missing value in a sample of $n\mathrm{GEOM}(\frac{1}{2})$ variables is k and the largest value to occur in the sample is j then j must have a minimum value of $k+1$, and all of the values from $k+1$ to j must occur. We define the EGF (exponential generating function) of these samples as follows. Let $C(z)$ be the EGF for complete samples. From [4], we know that

$$\mathbb{P}\left(\mathrm{GEOM}\left(\frac{1}{2}\right)\text{ sample of length } n \text{ is complete}\right) = \begin{cases} \dfrac{1}{2}, & \text{for } n \geq 1, \\ 1, & \text{for } n = 0. \end{cases}$$

Hence $C(z) = \frac{1}{2}\mathrm{e}^z + \frac{1}{2}$. We use this to find the EGF of samples whose largest missing value is k. We can use $\prod_{i=1}^{k-1}\mathrm{e}^{\frac{z}{2^i}}$ to represent all the values smaller than k, which are allowed to occur any number of times (including 0), and $C(\frac{z}{2^k}) - 1$ to represent the values larger than k. That is, we want to have a (non-empty) complete $\mathrm{GEOM}(\frac{1}{2})$ sample, but where the first or smallest value is now not 1 but $k+1$, and hence we shift the $C(z)$ to the right by writing $C(\frac{z}{2^k})$. Thus for $p = \frac{1}{2}$ the EGF of samples where the largest missing value is k is

$$\begin{aligned} S_k(z) &:= \prod_{i=1}^{k-1}\mathrm{e}^{\frac{z}{2^i}}\left(C\left(\frac{z}{2^k}\right) - 1\right) = \prod_{i=1}^{k-1}\mathrm{e}^{\frac{z}{2^i}}\left(\frac{1}{2}\mathrm{e}^{\frac{z}{2^k}} + \frac{1}{2} - 1\right) \\ &= \frac{1}{2}(\mathrm{e}^{\frac{z}{2^k}} - 1)\prod_{i=1}^{k-1}\mathrm{e}^{\frac{z}{2^i}} = \frac{1}{2}(\mathrm{e}^{\frac{z}{2^k}} - 1)\mathrm{e}^{z\sum_{i=1}^{k-1}\frac{1}{2^i}} \\ &= \frac{1}{2}(\mathrm{e}^{z-\frac{z}{2^k}} - \mathrm{e}^{z-\frac{z}{2^{k-1}}}). \end{aligned}$$

3.1. *Proof of Theorem 1 — expected value*

We define

$$S(z) := \sum_{k\geq 1} kS_k(z), \tag{3}$$

so the average value for the largest missing value in a $\mathrm{GEOM}(\frac{1}{2})$ sample is

$$\begin{aligned} n![z^n]\sum_{k\geq 1} k\frac{1}{2}(\mathrm{e}^{z-\frac{z}{2^k}} - \mathrm{e}^{z-\frac{z}{2^{k-1}}}) &= \frac{1}{2}\sum_{k\geq 1} k\left(\left(1-\frac{1}{2^k}\right)^n - \left(1 - \frac{1}{2^{k-1}}\right)^n\right) \\ &= \frac{1}{2}\sum_{r=1}^{n}\binom{n}{r}(-1)^r(1-2^r)\sum_{k\geq 1} k\frac{1}{2^{kr}} \\ &= \frac{1}{2}\sum_{r=1}^{n}\binom{n}{r}(-1)^r\frac{2^r}{1-2^r}. \end{aligned}$$

This expression can be approximated using 'Rice's method'. This technique is briefly explained in the following lemma (see [2,10,12]).

Lemma 1. *Let $\mathcal{C}$ be a curve surrounding the points* $1, 2, \ldots, n$ *in the complex plane, and let $f(z)$ be analytic inside $\mathcal{C}$. Then*

$$\sum_{k=1}^{n}\binom{n}{k}(-1)^k f(k) = -\frac{1}{2\pi\mathrm{i}}\int_C [n; z]f(z)\mathrm{d}z,$$

where the kernel is given by

$$[n;z] = \frac{(-1)^{n-1}n!}{z(z-1)\cdots(z-n)} = \frac{\Gamma(n+1)\Gamma(-z)}{\Gamma(n+1-z)}. \tag{4}$$

By extending the contour of integration, it turns out that under suitable growth conditions (see [2]) the asymptotic expansion of our alternating sum is

$$\sum \operatorname{Res}([n;z]f(z)) + \textit{smaller order terms},$$

where the sum is taken over all poles different from $1, \ldots, n$.

The function we use in this case is $f(z) := \frac{2^z}{1-2^z}$, and hence there is a double pole at $z = 0$ (the kernel also has a pole here) and simple poles at $z = \chi_k = \frac{2k\pi \mathrm{i}}{\log 2}$, where $k \in \mathbb{Z}$, $k \neq 0$.

For the double pole at $z = 0$,

$$f(z) = \frac{2^z}{1-2^z} \sim \frac{-1}{zL}\left(1 + \frac{zL}{2}\right),$$

and

$$[n;z] \sim -\frac{1}{z}(1 + zH_n).$$

This gives a residue of

$$\frac{1}{2} + \frac{H_n}{L} = \frac{1}{2} + \frac{1}{L}\log n + \frac{\gamma}{L}.$$

Because of the 1/2 outside our sum, the main (non-oscillating) term from Lemma 1 is thus

$$\frac{1}{2}\left(\log_2 n + \frac{\gamma}{L} + \frac{1}{2}\right).$$

For the simple poles at $z = \chi_k$,

$$f(z) = \frac{2^{z+\chi_k}}{1-2^{z+\chi_k}} \sim \frac{-1}{zL},$$

since $\mathrm{e}^{\chi_k} = 1$. The kernel expands to

$$[n;z] \sim n^{\chi_k}\Gamma(-\chi_k)$$

and so the small fluctuating terms which always arise from the poles at $z = \chi_k$ in problems of this type are

$$-\frac{1}{2L}\sum_{k\neq 0}\Gamma(-\chi_k)n^{\chi_k}.$$

Consequently the expected value for the largest missing value for all geometric samples is

$$\frac{1}{2}\left(\log_2 n + \frac{\gamma}{L} + \frac{1}{2} - \frac{1}{L}\sum_{k\neq 0}\Gamma(-\chi_k)\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\right). \tag{5}$$

However, this includes the case of complete samples for which the largest missing value was defined to be zero. We want the average value in the case of non-complete samples (we insist on there being at least one missing value), hence we must divide by the probability that the sample is not complete, namely $\frac{1}{2}$, (see [4]).

Since compositions are asymptotically equivalent to samples of length $\frac{n}{2}$, the $\log_2 n$ in (5) becomes $\log_2 \frac{n}{2}$, so that for non-complete compositions we have

$$\log_2 n + \frac{\gamma}{L} - \frac{1}{2} - \frac{1}{L}\sum_{k\neq 0}\Gamma(-\chi_k)\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}, \tag{6}$$

as in Theorem 1.

Remark. There are other techniques besides Rice's method that can be used for the problems considered in this paper. See, for example, papers by Louchard and Prodinger such as [8,7].

3.2. Proof of Theorem 2 — variance

Now we consider the function with a k^2 replacing the k in $T(z)$ (see (3)), and use a similar method to find the coefficients of $n![z^n]$. This will give us the second moment, and the difference between this and the expected value squared gives the variance. We define

$$V(z) := \sum_{k\geq 1} k^2 S_k(z),$$

so that

$$\begin{aligned} n![z^n]V(z) &= n![z^n]\sum_{k\geq 1} k^2 S_k(z) \\ &= \frac{1}{2}\sum_{r=1}^{n}\binom{n}{r}(-1)^r(1-2^r)\sum_{k\geq 1}k^2\frac{1}{2^{kr}} \\ &= -\frac{1}{2}\sum_{r=1}^{n}\binom{n}{r}(-1)^r\frac{2^r(1+2^r)}{(1-2^r)^2}. \end{aligned} \tag{7}$$

As before, this is a candidate for Rice's method, but the new function ($f(z) = \frac{2^r(1+2^r)}{(1-2^r)^2}$) means that there are triple poles at $z = 0$ and double poles at $z = \chi_k$. The residue at $z = 0$ gives us the main term of the second moment, namely

$$\frac{1}{2}\left(\log_2^2 n + \log_2 n\left(1+\frac{2\gamma}{L}\right)+\frac{1}{3}+\frac{\gamma}{L}+\frac{\gamma^2}{L^2}+\frac{\pi^2}{6L^2}\right).$$

Dividing by the probability that the sample is not complete removes the factor of a half, giving

$$\log_2^2 n + \log_2 n\left(1+\frac{2\gamma}{L}\right)+\frac{1}{3}+\frac{\gamma}{L}+\frac{\gamma^2}{L^2}+\frac{\pi^2}{6L^2}. \tag{8}$$

Now we subtract the square of the expectation from (8). Squaring the fluctuations of the expectation means that we will have some $k = 0$ terms appearing (any k term multiplied by a $-k$ term will give this), which means a contribution to the main term. Prodinger has already computed this in [11], page 252, and from his results we have that the main term obtained from squaring $\frac{1}{L}\sum_{k\neq 0}\Gamma(-\chi_k)e^{2k\pi i\log_2 n}$ is

$$\frac{\pi^2}{6L^2}-\frac{11}{12}-\frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(2^h-1)}.$$

So the main term of the variance is

$$\begin{aligned} &\log_2^2 n + \log_2 n\left(1+\frac{2\gamma}{L}\right)+\frac{1}{3}+\frac{\gamma}{L}+\frac{\gamma^2}{L^2}+\frac{\pi^2}{6L^2} \\ &\quad -\left(\log_2^2 n+\frac{\gamma^2}{L^2}+\frac{1}{4}+2\log_2 n\frac{\gamma}{L}+2\frac{\gamma}{2L}+2\log_2 n\frac{1}{2}+\frac{\pi^2}{6L^2}-\frac{11}{12}-\frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(2^h-1)}\right) \end{aligned}$$

which simplifies to

$$1+\frac{2}{L}\sum_{h\geq 1}\frac{(-1)^{h-1}}{h(2^h-1)}.$$

For the fluctuations of the variance, we have the fluctuations of the second moment, from which we subtract all fluctuations which arise from squaring the expectation. The residue for the double pole in

$$\frac{2^r(1+2^r)}{(1-2^r)^2}\frac{\Gamma(n+1)\Gamma(-z)}{\Gamma(n+1-z)}$$

at $z = \chi_k$ is asymptotic to (the factor of a half cancels as before)

$$-n^{\chi_k}\frac{\Gamma(-\chi_k)}{L}-2n^{\chi_k}\frac{\Gamma(-\chi_k)}{L}\log_2 n+2n^{\chi_k}\frac{\Gamma'(-\chi_k)}{L^2}, \tag{9}$$

which is summed on all $k \neq 0$. We find the expression for the square of the fluctuations — without the $k = 0$ terms of the Fourier series in [11], page 255 — to be

$$\frac{2}{L}\sum_{l\geq 1}\frac{(-1)^l\Gamma(-\chi_k+l)}{l!(2^l-1)}+\frac{2}{L^2}\Gamma(-\chi_k)(\psi(-\chi_k)+\gamma)$$

Table 1
Table showing values for different properties of compositions of 1 to n.

n	A	B	C	D	E	F	G
1	0	0.	0	–	0	–	–
2	1	0.5	1	1.	1	1.	1.
3	1	0.25	1	1.	2	2.	1.
4	4	0.5	4	1.	8	2.	1.
5	8	0.5	6	0.75	18	2.25	1.25
6	14	0.4375	14	1.	39	2.78571	1.
7	31	0.484375	26	0.83871	89	2.87097	1.16129
8	63	0.492188	54	0.857143	195	3.09524	1.14286
9	129	0.503906	105	0.813953	415	3.21705	1.23256
10	248	0.484375	213	0.858871	865	3.4879	1.14113
11	509	0.49707	423	0.831041	1823	3.58153	1.19253
12	1011	0.493652	849	0.839763	3787	3.7458	1.17804
13	2044	0.499023	1697	0.830235	7869	3.8498	1.19716
14	4089	0.499146	3399	0.831255	16,230	3.96919	1.20249
15	8167	0.498474	6799	0.832497	33,352	4.08375	1.1942
16	16,360	0.499268	13,608	0.831785	68,422	4.18227	1.19976
17	32,725	0.499344	27,220	0.83178	140,013	4.27847	1.19798
18	65,482	0.499588	54,451	0.831541	285,991	4.36748	1.19996
19	131,017	0.49979	108,901	0.831197	583,235	4.4516	1.20023
20	262,176	0.500061	217,789	0.830698	1187,718	4.53023	1.20267

where $\psi(-\chi_k) = \frac{\Gamma'(-\chi_k)}{\Gamma(-\chi_k)}$ is the digamma function. Thus the fluctuations for the variance are

$$
\begin{aligned}
&-\sum_{k\neq 0}\frac{\Gamma(-\chi_k)}{L}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} - 2\sum_{k\neq 0}\frac{\Gamma(-\chi_k)}{L}\log_2 n\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} + 2\sum_{k\neq 0}\frac{\Gamma'(-\chi_k)}{L^2}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\\
&\quad-\left((-2)\left(\log_2 n + \frac{\gamma}{L} + \frac{1}{2}\right)\frac{1}{L}\sum_{k\neq 0}\Gamma(-\chi_k)\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} + \sum_{k\neq 0}\frac{2}{L}\sum_{l\geq 1}\frac{(-1)^l\Gamma(-\chi_k+l)}{l!(2^l-1)}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\right.\\
&\quad\left.+\sum_{k\neq 0}\frac{2}{L^2}\Gamma'(-\chi_k)\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} + \sum_{k\neq 0}\frac{2}{L^2}\Gamma(-\chi_k)\gamma\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\right)\\
&= -\sum_{k\neq 0}\frac{\Gamma(-\chi_k)}{L}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} - \sum_{k\neq 0}\frac{2\Gamma(-\chi_k)\log_2 n}{L}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} + \sum_{k\neq 0}\frac{2\Gamma'(-\chi_k)}{L^2}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\\
&\quad+\sum_{k\neq 0}\frac{2\log_2 n\Gamma(-\chi_k)}{L}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} + \sum_{k\neq 0}\frac{2\gamma\Gamma(-\chi_k)}{L^2}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} + \sum_{k\neq 0}\frac{\Gamma(-\chi_k)}{L}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\\
&\quad-\sum_{k\neq 0}\frac{2}{L}\sum_{l\geq 1}\frac{(-1)^l\Gamma(-\chi_k+l)}{l!(2^l-1)}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} - \sum_{k\neq 0}\frac{2\Gamma'(-\chi_k)}{L^2}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n} - \sum_{k\neq 0}\frac{2\gamma\Gamma(-\chi_k)}{L^2}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}\\
&= -\sum_{k\neq 0}\frac{2}{L}\sum_{l\geq 1}\frac{(-1)^l\Gamma(-\chi_k+l)}{l!(2^l-1)}\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}.
\end{aligned}
$$

The result in Theorem 2 follows.

4. Probability that a composition has the LMV property

We now ask a different question, and again relate it to geometric random variables with parameter $p = \frac{1}{2}$. We want to know the probability that the largest missing value is one less than the largest part.

Theorem 3. *The probability that a non-complete composition of n has the LMV property is asymptotically*

$$4 - 2\log_2 3 + \frac{2}{L}\sum_{k\neq 0}(3^{\chi_k} - 1)\Gamma(-\chi_k)\mathrm{e}^{2k\pi\mathrm{i}\log_2 n}.$$

This means that with a probability of approximately 0.83, the largest part and the largest missing value will be adjacent, assuming the largest missing value is non-zero. This result agrees well with Column D in Table 1.

4.1. Proof of Theorem 3

The generating function for all samples whose largest missing value is k and which have only one value larger than k is

$$P_k(z) = \prod_{i=1}^{k-1} \mathrm{e}^{\frac{z}{2^i}} (\mathrm{e}^{\frac{z}{2^{k+1}}} - 1) = \mathrm{e}^{z - \frac{3z}{2^{k+1}}} - \mathrm{e}^{z - \frac{z}{2^{k-1}}}. \tag{10}$$

Let

$$P(z) := \sum_{k\geq 1} P_k(z).$$

So the probability that there is only one value larger than the largest missing value is:

$$\begin{aligned} n![z^n]P(z) &= n![z^n] \sum_{k\geq 1} (\mathrm{e}^{z-\frac{3z}{2^{k+1}}} - \mathrm{e}^{z-\frac{z}{2^{k-1}}}) \\ &= \sum_{k\geq 1} \left(\left(1 - \frac{3}{2^{k+1}}\right)^n - \left(1 - \frac{1}{2^{k-1}}\right)^n \right) \\ &= \sum_{k\geq 1} \sum_{r=0}^{n} \binom{n}{r} (-1)^r \left(\frac{1}{2^{k+1}}\right)^r (3^r - 4^r) \\ &= \sum_{r=0}^{n} \binom{n}{r} (-1)^r \frac{3^r - 4^r}{2^r(2^r-1)} \end{aligned}$$

which can be approximated using Rice's method. We use the function $f(z) := \frac{3^z - 4^z}{2^z(2^z-1)}$, which has no poles at $z = 0$, so there is only a simple pole (from the kernel) at $z = 0$. We find the residue at $z = 0$ to be

$$\frac{\log\left(\frac{4}{3}\right)}{\log 2} = 2 - \log_2 3. \tag{11}$$

Now we look at the simple poles at $z = \chi_k$, from which we find the fluctuations in this case to be

$$\frac{1}{L} \sum_{k\neq 0} (3^{\chi_k} - 1)\Gamma(-\chi_k) \mathrm{e}^{2k\pi \mathrm{i} \log_2 n}.$$

Dividing by the probability $\frac{1}{2}$ that the sample is non-complete gives the result in Theorem 3.

5. The average largest value of compositions with the LMV property

In this section we find the average largest part size for compositions which have the LMV property. That is, if a non-complete composition has a largest part and a largest missing value differing by one, what is the average of the largest part? We find the average largest missing value in this case, and then add one to the result.

Theorem 4. *The average largest part for compositions which have the LMV property is asymptotically*

$$\log_2 n + \frac{\frac{\gamma}{L} + \frac{\log_2 3}{2} + \delta_b(\log_2 n)}{1 + \delta_a(\log_2 n)}$$

where

$$\delta_a(x) := \frac{1}{L(2 - \log_2 3)} \sum_{k\neq 0} (3^{\chi_k} - 1)\Gamma(-\chi_k) \mathrm{e}^{2k\pi \mathrm{i} x}$$

and

$$\delta_b(x) := \frac{1}{L^2(2 - \log_2 3)} \sum_{k\neq 0} (\Gamma(-\chi_k)(3^{\chi_k} \log(3/2) - L) - \Gamma'(-\chi_k)(3^{\chi_k} - 1)) \mathrm{e}^{2k\pi \mathrm{i} x}.$$

Remark. Ignoring the tiny fluctuations, the average largest value for compositions is known to be (see [9])

$$\log_2 n + \frac{\gamma}{L} - \frac{1}{2} \tag{12}$$

and for complete compositions is

$$\log_2 n - \frac{0.337875\ldots}{L^2} + \frac{\gamma}{L} - 1 \tag{13}$$

from [7] (subtract 1 for compositions) where the constant 0.337875 . . . is the value of a certain Dirichlet series at $s = 0$.

5.1. Proof of Theorem 4

The generating function for all GEOM($\frac{1}{2}$) samples whose largest missing value is k and whose largest part is $k+1$ is (for $p = 1/2$)

$$P_k(z) = (e^{z-\frac{3z}{2^{k+1}}} - e^{z-\frac{z}{2^{k-1}}}),$$

as in (10). If we define the function

$$R(z) := \sum_{k\geq 1} kP_k(z)$$

then the average largest missing value is given by

$$\begin{aligned} n![z^n]R(z) &= n![z^n]\sum_{k\geq 1} k(e^{z-\frac{3z}{2^{k+1}}} - e^{z-\frac{z}{2^{k-1}}}) \\ &= \sum_{k\geq 1} k\left(\left(1-\frac{3}{2^{k+1}}\right)^n - \left(1-\frac{1}{2^{k-1}}\right)^n\right) \\ &= \sum_{k\geq 1} k\sum_{r=0}^{n}\binom{n}{r}(-1)^r\left(\frac{1}{2^{k+1}}\right)^r(3^r-4^r) \\ &= \sum_{r=0}^{n}\binom{n}{r}(-1)^r\frac{3^r-4^r}{(2^r-1)^2} \end{aligned}$$

which is an expression to which Rice's method can be applied. By using the function $f(z) := \frac{(3^z-4^z)}{(2^z-1)^2}$, we have a simple pole at $z = 0$, and there is also a simple pole from the kernel at $z = 0$. So there are double poles at $z = 0$ and at $z = \chi_k$ for $k \neq 0$. The residue at $z = 0$ is

$$\begin{aligned} \frac{\log\left(\frac{4}{3}\right)}{L}\log_2 n + \frac{\log\left(\frac{4}{3}\right)(2\gamma+\log 3)}{2L^2} &= \frac{\log 4 - \log 3}{L}\log_2 n + \frac{\log 4(2\gamma+\log 3) - \log 3(2\gamma+\log 3)}{2L^2} \\ &= (2-\log_2 3)\log_2 n + \frac{2\gamma}{L} + \log_2 3 - \frac{\gamma\log_2 3}{L} - \frac{(\log_2 3)^2}{2}. \end{aligned} \tag{14}$$

For the fluctuations, we look at the residues at $z = \chi_k$ to get

$$\begin{aligned} &\frac{\log_2 n}{L}\sum_{k\neq 0}(3^{\chi_k}-1)\Gamma(-\chi_k)e^{2k\pi i\log_2 n} \\ &\quad + \sum_{k\neq 0}\Gamma(-\chi_k)\frac{3^{\chi_k}\log(3/2) - L}{L^2}e^{2k\pi i\log_2 n} - \sum_{k\neq 0}\Gamma'(-\chi_k)\frac{3^{\chi_k}-1}{L^2}e^{2k\pi i\log_2 n}. \end{aligned} \tag{15}$$

We divide by $\frac{1}{2}$ to consider only non-complete compositions and then divide by the probability that the sample has the LMV property (see Theorem 3) to get the average largest missing value. Finally we need to add 1 to get the average largest value in a GEOM($\frac{1}{2}$) sample, but we must also subtract 1 for the case of compositions.

6. Average number of values larger than the largest missing value

Here we want to find the average number of parts larger than the largest missing value in non-complete compositions of n. To do this we need only subtract the average largest missing value from the average largest part.

The average largest value for all compositions of n and for complete compositions of n was given in Eqs. (12) and (13). Since the asymptotic probability that a composition of n is complete is 1/2, we can thus obtain the average largest part (denoted by $\mathcal{L}$) for non-complete compositions of n, since the average largest part of all compositions of n is

$$(\mathcal{L}[\text{complete compositions of } n] + \mathcal{L}[\text{non-complete compositions of } n])/2.$$

Hence, ignoring fluctuations, we have

$$\begin{aligned}\mathcal{L}[\text{non-complete compositions of } n] &= 2\left(\log_2 n + \frac{\gamma}{L} - \frac{1}{2}\right) - \left(\log_2 n - \frac{0.337875\ldots}{L^2} + \frac{\gamma}{L} - 1\right)\\ &= \log_2 n + \frac{\gamma}{L} + \frac{0.337875\ldots}{L^2}. \end{aligned} \tag{16}$$

So to find the average number of values larger than the largest missing value, we subtract the main term for the average largest missing value for non-complete compositions of n (see Theorem 1) from the main term for the average largest part for non-complete compositions of n in (16) to get a constant of

$$\frac{0.337875\ldots}{L^2} + \frac{1}{2} \approx 1.20324.$$

This estimate agrees well with the data in Column G of Table 1.

7. Concluding remarks

7.1. Comparison of the mean largest parts

We can compare the constant values found in this paper with those in the general compositions and complete compositions cases as follows (we ignore the tiny fluctuations).

The average largest part for all compositions of n is (see Eq. (12))

$$\log_2 n + \frac{\gamma}{L} - \frac{1}{2} \approx \log_2 n + 0.3324\ldots.$$

The average largest part for complete compositions is (see Eq. (13))

$$\log_2 n - \frac{0.337875\ldots}{L^2} + \frac{\gamma}{L} - 1 \approx \log_2 n - 0.8708\ldots.$$

The average largest part for compositions with the LMV property is (see Theorem 4)

$$\log_2 n + \frac{\gamma}{L} + \frac{\log_2 3}{2} \approx \log_2 n + 1.624\ldots.$$

The average largest part of non-complete compositions is (see Eq. (16))

$$\log_2 n + \frac{\gamma}{L} + \frac{0.337875\ldots}{L^2} \approx \log_2 n + 1.5359\ldots.$$

7.2. GEOM(p) samples for $p \neq \frac{1}{2}$

It is important to note that the methods used in this paper only work for $p = \frac{1}{2}$. The authors are now investigating the case for $p \neq \frac{1}{2}$ in [1].

Acknowledgement

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

References

[1] M. Archibald, A. Knopfmacher, The last missing value in a sample of geometric random variables (in preparation).
[2] P. Flajolet, R. Sedgewick, Mellin transforms and asymptotics: finite differences and Rice's integrals, Theoret. Comput. Sci. 144 (1995) 101–124.
[3] W. Goh, P. Hitczenko, Gaps in samples of geometric random variables, Discrete Math. 307 (2007) 2871–2890.
[4] P. Hitczenko, A. Knopfmacher, Gap-free compositions and gap-free samples of geometric random variables, Discrete Math. 294 (2005) 225–239.
[5] P. Hitczenko, G. Louchard, Distinctness of compositions of an integer: a probabilistic analysis, Random Struct. Alg. 19 (2004) 407–437.
[6] P. Hitczenko, C.D. Savage, On the multiplicity of parts in a random composition of a large integer, SIAM J. Discrete Math. 18 (2005) 418–435.
[7] G. Louchard, H. Prodinger, On gaps and unoccupied urns in sequences of geometrically distributed random variables, Discrete Math. 308 (2008) 1538–1562.
[8] G. Louchard, H. Prodinger, Asymptotics of the moments of extreme-value related distribution functions, Algorithmica 46 (2006) 431–467.
[9] A. Odlyzko, B. Richmond, On the compositions of an integer, in: Combinatorial Mathematics, VII (Proc. Seventh Australian Conf., Univ. Newcastle, Newcastle, 1979), in: Lecture Notes in Math., vol. 829, Springer, Berlin, 1980, pp. 199–210.
[10] H. Prodinger, Combinatorics of geometrically distributed random variables: left-to-right maxima, Discrete Math. 153 (1996) 253–270.
[11] H. Prodinger, Periodic oscillations in the analysis of algorithms, J. Iran. Stat. Soc. (JIRSS) 3 (2004) 251–270.
[12] W. Szpankowski, Average Case Analysis of Algorithms on Sequences, John Wiley and Sons, New York, 2001.