Journal of Integer Sequences, Vol. 16 (2013),
Article 13.2.2

The Largest Missing Value in a Composition of an Integer and Some Allouche-Shallit-Type Identities

Guy Louchard
Université Libre de Bruxelles
Département d'Informatique, CP 212
Boulevard du Triomphe
B-1050 Bruxelles
Belgium
louchard@ulb.ac.be

Helmut Prodinger
University of Stellenbosch
Mathematics Department
7602 Stellenbosch
South Africa
hproding@sun.ac.za

Dedicated to Jean-Paul Allouche on the occasion of his 60th birthday.

Abstract

Archibald and Knopfmacher recently considered the largest missing value in a composition of an integer and established the mean and variance. Our alternative, probabilistic approach produces (in principle) *all* moments in an almost automatic way. In order to show that our forms match the ones given by Archibald and Knopfmacher, we have to derive some identities which are interesting on their own. We construct a one-parameter family of identities, and the first one is (equivalent to) the celebrated identity due to Allouche and Shallit. We finally provide a simple direct analysis of the LMV(-1) case: if the largest missing value is exactly one smaller than the largest value, we say that the sequence has the LMV(-1) property.

1 Introduction

Our attention was recently attracted by a paper by M. Archibald and A. Knopfmacher [1] about some asymptotic properties of a sequence of n geometric independent identically distributed random variables, with distribution pq^{j-1}, $q := 1-p$. In the particular case $p = 1/2$, (which is related to compositions of integers, see below), and using generating functions techniques, the authors study the asymptotic mean and variance of largest missing value (if it exists). If the **l**argest **m**issing **v**alue is exactly one smaller than the largest value, they say that the sequence has the LMV(-1) property. They compute the LMV(-1) probability, and the corresponding mean value.

Here, within our alternative and more powerful approach, we think about an urn model, with urns labelled $1, 2, \ldots$, the probability of each ball falling into urn j being given by pq^{j-1}. In this probabilistic urn model, we throw n balls, and have the following properties, which we collect below although the full relevance of them will only appear later in the text.

- We have asymptotic independence of urns, for all events related to urn j, $j = O(\log n)$. This is proved, by Poissonization-DePoissonization, in [11], [3] and [8] (in this paper for $p = 1/2$, but the proof is easily adapted). The error term is $O(n^{-C})$ where C is a positive constant.

- We obtain asymptotic distributions of the random variables (RV) of interest. The number of balls in urn j, $j = O(\log n)$ is now Poisson-distributed with parameter $(np/q)q^j$. The asymptotic distributions are related to Gumbel distribution functions or convergent series of such. The error term is $O(n^{-1})$.

- We have uniform integrability for the moments of our RV. To show that the limiting moments are equivalent to the moments of the limiting distributions, we need a suitable rate of convergence. This is related to a uniform integrability condition (see Loève [9, Section 11.4]). For the kind of limiting distributions we consider here, the rate of convergence is analyzed in detail in [10, 3]. The error term is $O(n^{-C})$.

- Asymptotic expressions for the moments are obtained by Mellin transforms. The error term is $O(n^{-C})$.

- $\Gamma(s)$ decreases exponentially in the direction $i\infty$ ([5]):
$$|\Gamma(\sigma + it)| \sim \sqrt{2\pi}|t|^{\sigma-1/2}e^{-\pi|t|/2}.$$
Also, we have a "slow increase property" for all other functions we encounter. So inverting the Mellin transforms is easily justified.

We proceed as follows: from the asymptotic properties of the urns, we obtain the asymptotic distributions of our RV of interest. Next we compute the Laplace transform $\phi(\alpha)$ from which we can derive the dominant part of probabilities and moments as well as the (tiny) periodic part in the form of a Fourier series.

If we compare the approach in this paper with other ones that appeared previously (related to similar problems), then we notice the following. Traditionally, one would stay with exact enumerations as long as possible, and only at a late stage move to asymptotics. Doing this, one would, in terms of asymptotics, carry many unimportant contributions around, which makes the computations quite heavy, especially when it comes to higher moments. Here, however, approximations are carried out as early as possible, and this allows for streamlined (and often automatic) computations of the higher moments.

The paper is organized as follows: in Section 2, we consider compositions of an integer. Section 3 is devoted to a first computation of the dominant part of the moments. In Section 4, we analyze the periodicities and we briefly consider the conditioning on the event that there exists a largest missing value. Section 5 contains the main result of this paper: we present some general Allouche-Shallit-type formulas[1] and the direct moments computations. Section 6 is devoted to the LMV(-1) case and Section 7 concludes the paper.

2 Compositions of an integer

Let us consider the composition of an integer υ with n parts, i.e., $\upsilon = \sum_{i=1}^{n} x_i$, where x_i are natural numbers. Considering all compositions as equiprobable, we know ([8]) that the number of parts n is asymptotically Gaussian, $\upsilon \to \infty$:

$$n \sim \mathcal{N}\left(\frac{\upsilon}{2}, \frac{\upsilon}{4}\right), \tag{1}$$

and that the part sizes are asymptotically identically distributed as GEOM(1/2) and *independent*. All distributions we analyze in this paper depend on $\log_2 n$. Hence, with (1), the same random variables related to υ are asymptotically equivalent by replacing $\log_2 n$ by $\log_2 \upsilon - 1$ ([8]).

We will consider the case $p = 1/2$ and use the following notations:

$$\begin{aligned}
n &:= \text{number of balls, } n \text{ large},\\
\mathcal{P}(\lambda, u) &:= e^{-\lambda}\lambda^u/u!, \text{ (Poisson distribution)},\\
\log &:= \log_2,\\
\eta &:= j - \log n,\\
L &:= \ln 2,\\
\tilde{\alpha} &:= \alpha/L,\\
\chi_l &:= \frac{2l\pi\mathbf{i}}{L}.
\end{aligned}$$

Let $\mathcal{B}$ be the event that there exists a largest missing value, i.e., a maximal empty urn before the last non-empty urn. Following previous results of the authors [10], let J be the position of this maximal empty urn before the last non-empty urn. We have (here and in

[1] We were kindly informed by Allouche that the first forms of the identities in question are due to Woods and Robbins [13, 12]. However, Allouche and Shallit greatly extended them and made them into what they are today, thanks to the great book [4]. Thus, also, since this is the birthday volume for Jean-Paul Allouche, we took the liberty to use the term Allouche-Shallit-type identities.

the sequel $\sim$ always denotes $\sim_{n\to\infty}$)

$$\mathbb{P}[J=j;\mathcal{B}] \sim \varphi(\eta),$$
$$\varphi(\eta) := \exp(-e^{-L\eta}) \sum_{r\geq 1} \prod_{i=1}^{r} \left[1-\exp(-e^{-L(\eta+i)})\right] \exp(-e^{-L(\eta+r)}).$$

Here is the explanation: we recall that we have i.i.d. geometrically distributed RVs, and that the urns are asymptotically independent, for all events related to urn j containing $O(1)$ balls. Also the number of balls in each such urn is now asymptotically Poisson-distributed with parameter npq^{j-1} in urn j. So the asymptotic number ℓ of balls in urn j is given by

$$\exp\left(-npq^{j-1}\right) \frac{\left(npq^{j-1}\right)^{\ell}}{\ell!},$$

and with $p=1/2$, $\eta = j-\log n$, this is equivalent to $\mathcal{P}\left(e^{-L\eta},\ell\right)$. Here, urn j is empty, we have $r\geq 1$ non-empty urns after urn j, and all urns after urn $j+r$ are empty. This leads to

$$\mathbb{P}[J=j;\mathcal{B}] \sim \mathcal{P}\big(e^{-L\eta},0\big) \sum_{r\geq 1} \prod_{i=1}^{r} \left[1-\mathcal{P}\big(e^{-L(\eta+i)},0\big)\right] \mathcal{P}\big(e^{-L(\eta+r)},0\big).$$

Following [10] again, we now compute the Laplace transform (with the change of variables $y=e^{-L\eta}$):

$$\begin{aligned}
\phi(\alpha) &= \int_{-\infty}^{\infty} e^{-\alpha\eta}\varphi(\eta)d\eta \\
&= \int_0^{\infty} y^{-\tilde{\alpha}} e^{-y} \sum_{r\geq 1} \prod_{i=1}^{r} \left[1-e^{-ye^{-Li}}\right] e^{-ye^{-Lr}} \frac{dy}{Ly} \qquad (2) \\
&= \int_0^{\infty} y^{-\tilde{\alpha}} e^{-y} \sum_{r\geq 1} \prod_{i=1}^{r} \left[1-e^{-y2^{-i}}\right] e^{-y2^{-r}} \frac{dy}{Ly}.
\end{aligned}$$

Again by [10], we have, with

$$F_0(s) := \left.\phi(\alpha)\right|_{\alpha=-Ls}, \qquad (3)$$
$$\mathbb{P}(\mathcal{B}) \sim \phi(0) + \sum_{l\neq 0} F_0(\chi_l) e^{-2l\pi\mathbf{i}\log n}. \qquad (4)$$

But, from [11] and [7], we know that $\mathbb{P}(\mathcal{B})=1/2$. So we should first independently confirm that

$$\phi(0) = \int_0^{\infty} e^{-y} \sum_{r\geq 1} \prod_{i=1}^{r} \left(1-e^{-y/2^i}\right) e^{-y/2^r} \frac{dy}{Ly} = \frac{1}{2}. \qquad (5)$$

To do that, set

$$f_0(y) := \sum_{r\geq 1} \prod_{i=1}^{r} \left(1-e^{-y/2^i}\right) e^{-y/2^r}.$$

Then

$$
\begin{aligned}
f_0(2y) &= \sum_{r\geq 1}\prod_{i=1}^{r}\left(1-e^{-y/2^{i-1}}\right)e^{-y/2^{r-1}} \\
&= \sum_{r\geq 0}\prod_{i=0}^{r}\left(1-e^{-y/2^{i}}\right)e^{-y/2^{r}} \\
&= (1-e^{-y})e^{-y} + \sum_{r\geq 1}\prod_{i=0}^{r}\left(1-e^{-y/2^{i}}\right)e^{-y/2^{r}} \\
&= (1-e^{-y})e^{-y} + (1-e^{-y})\sum_{r\geq 1}\prod_{i=1}^{r}\left(1-e^{-y/2^{i}}\right)e^{-y/2^{r}} \\
&= (1-e^{-y})e^{-y} + (1-e^{-y})f_0(y).
\end{aligned}
$$

Now

$$
\begin{aligned}
G_0(t) := \int_0^\infty e^{-ty} f_0(y)\frac{dy}{y} &= \int_0^\infty e^{-2ty} f_0(2y)\frac{dy}{y} \\
&= \int_0^\infty e^{-2ty}\Big[(1-e^{-y})e^{-y} + (1-e^{-y})f_0(y)\Big]\frac{dy}{y} \\
&= \ln(2t+2) - \ln(2t+1) + \int_0^\infty e^{-2ty}(1-e^{-y})f_0(y)\frac{dy}{y}.
\end{aligned}
$$

So

$$
G_0(t) = \ln(2t+2) - \ln(2t+1) + G_0(2t) - G_0(2t+1).
$$

If one iterates this formally, one gets

$$
G_0(1) = \sum_{n\geq 3}(-1)^{\nu(n-1)}\ln n.
$$

Here $\nu(k)$ is the number of ones in the binary expansion of the integer k. We will only need the properties $\nu(2k) = \nu(k)$ and $\nu(2k+1) = 1 + \nu(k)$.

This is not a convergent series, but it makes good sense as follows. We have

$$
\begin{aligned}
G_0(1) &= \sum_{n\geq 2}(-1)^{\nu(n)}\ln(n+1) \\
&= \sum_{j\geq 1}(-1)^{\nu(j)}[\ln(2j+1) - \ln(2j+2)] \\
&= \sum_{j\geq 0}(-1)^{\nu(j)}\ln\left[\frac{2j+1}{2j+2}\right] + \ln 2 \\
&= -\frac{1}{2}\ln 2 + \ln 2 = \frac{\ln 2}{2} = \frac{L}{2} \qquad (6)
\end{aligned}
$$

by the Allouche-Shallit (or Woods-Robbins) identity ([4, 2, 11]). This proves (5).

Now we must confirm that $F_0(\chi_l) = 0$ in order to prove that the second part of (4) is null. Set

$$f_\alpha(y) = \sum_{r\geq 1} \prod_{i=1}^{r} \left(1 - e^{-y/2^i}\right) e^{-y/2^r} y^{-\alpha}.$$

Then

$$\begin{aligned}
f_\alpha(2y) &= \sum_{r\geq 1} \prod_{i=1}^{r} \left(1 - e^{-y/2^{i-1}}\right) e^{-y/2^{r-1}} (2y)^{-\alpha} \\
&= \sum_{r\geq 0} \prod_{i=0}^{r} \left(1 - e^{-y/2^{i}}\right) e^{-y/2^{r}} (2y)^{-\alpha} \\
&= (1-e^{-y})e^{-y}(2y)^{-\alpha} + \sum_{r\geq 1} \prod_{i=0}^{r} \left(1 - e^{-y/2^{i}}\right) e^{-y/2^{r}} (2y)^{-\alpha} \\
&= (1-e^{-y})e^{-y}(2y)^{-\alpha} + (1-e^{-y}) \sum_{r\geq 1} \prod_{i=1}^{r} \left(1 - e^{-y/2^{i}}\right) e^{-y/2^{r}} (2y)^{-\alpha} \\
&= (1-e^{-y})e^{-y}(2y)^{-\alpha} + (1-e^{-y})2^{-\alpha} f_\alpha(y).
\end{aligned}$$

Now

$$\begin{aligned}
G_\alpha(t) &:= \int_0^\infty e^{-ty} f_\alpha(y) \frac{dy}{y} = \int_0^\infty e^{-2ty} f_\alpha(2y) \frac{dy}{y} \\
&= \int_0^\infty e^{-2ty} \left[(1-e^{-y})e^{-y}(2y)^{-\alpha} + (1-e^{-y}) f_\alpha(y) 2^{-\alpha}\right] \frac{dy}{y} \\
&= 2^{-\alpha}\Gamma(-\alpha)[(2t+1)^\alpha - (2t+2)^\alpha] + \int_0^\infty e^{-2ty}(1-e^{-y}) f_\alpha(y) 2^{-\alpha} \frac{dy}{y}.
\end{aligned}$$

So

$$G_\alpha(t) = \Gamma(-\alpha)2^{-\alpha}[(2t+1)^\alpha - (2t+2)^\alpha] + 2^{-\alpha}G_\alpha(2t) - 2^{-\alpha}G_\alpha(2t+1).$$

By iteration we find

$$\begin{aligned}
M(\alpha) := G_\alpha(1) &= -\Gamma(-\alpha) \sum_{k\geq 1} \sum_{j=0}^{2^k-1} (-1)^{\nu(2^k+j)} 2^{-\alpha k} (2^k+j+1)^\alpha \\
&= \Gamma(-\alpha) \sum_{k\geq 1} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} 2^{-\alpha k} (2^k+j+1)^\alpha, \qquad (7)
\end{aligned}$$

and

$$\phi(\alpha) = M(\tilde{\alpha})/L. \qquad (8)$$

Now, by (3) and setting $s = \chi_l$,

$$F_0(\chi_l) = \frac{1}{L}\Gamma(\chi_l)\sum_{k\ge 1}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}(2^k+j+1)^{-\chi_l}.$$

It is already proved in [11] that this is zero. (But it follows readily from the computations that appear in this paper.)

3 The moments

The purpose of this section is to obtain the dominant terms of the moments of J, the first two of which have been computed already in [1].

The paper [10] contains the following theorem, which we express in our notation

$$F_k(s) := \phi^{(k)}(\alpha)\big|_{\alpha=-Ls}.$$

Theorem 1. *Let J be the position of the maximal empty urn before the last non-empty urn. The first two moments of the parameter J are asymptotically given by*

$$\mathbb{E}(J-\log n;\mathcal{B}) \sim \phi'(0) + \sum_{l\ne 0} F_1(\chi_l)e^{-2l\pi\mathbf{i}\log n},$$
$$\mathbb{E}((J-\log n)^2;\mathcal{B}) \sim \phi''(0) + \sum_{l\ne 0} F_2(\chi_l)e^{-2l\pi\mathbf{i}\log n}.$$

□

Now, by (7) and (8),

$$\begin{aligned}\phi(\alpha) &= M(\tilde{\alpha})/L,\\ M(\alpha) &= \Gamma(-\alpha)H(\alpha), \quad \text{with}\\ H(\alpha) &:= \sum_{k\ge 1}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}2^{-\alpha k}(2^k+j+1)^{\alpha}.\end{aligned} \tag{9}$$

The following expansion is well known, where γ is the Euler's constant ([5]):

$$\Gamma(-\alpha) = -\frac{1}{\alpha} - \gamma - \left(\frac{\pi^2}{12}+\frac{\gamma^2}{2}\right)\alpha + \cdots.$$

Now we expand $H(\alpha)$:

$$
\begin{aligned}
H(\alpha) &= 0 + H'(0)\alpha + H''(0)\frac{\alpha^2}{2} + H'''(0)\frac{\alpha^3}{6} + \cdots\,, \\
H'(0) &= \sum_{k\geq 1}\Bigg[-Lk\sum_{j=0}^{2^k-1}(-1)^{\nu(j)} + L\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log(2^k+j+1)\Bigg], \\
H''(0) &= \sum_{k\geq 1}\Bigg[L^2k^2\sum_{j=0}^{2^k-1}(-1)^{\nu(j)} - 2L^2k\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log(2^k+j+1) \\
&\qquad\qquad + L^2\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log^2(2^k+j+1)\Bigg], \\
H'''(0) &= \sum_{k\geq 1}\Bigg[-L^3k^3\sum_{j=0}^{2^k-1}(-1)^{\nu(j)} + 3L^3k^2\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log(2^k+j+1) \\
&\qquad\qquad - 3L^3k\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log^2(2^k+j+1) + \sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log^3(2^k+j+1)\Bigg].
\end{aligned} \tag{10}
$$

Note that, as $\phi(0) = 1/2$, the first term in these equations is identically 0 and $H'(0) = -\frac{L}{2}$. Set

$$
\begin{aligned}
f(x) &= \sum_{k\geq 0} k\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log(2^k+j+x), \\
g(x) &= \sum_{k\geq 0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log(2^k+j+x).
\end{aligned}
$$

A similar computation as before shows that

$$
f(x) = f\Big(\frac{x}{2}\Big) - f\Big(\frac{x+1}{2}\Big) + g(x) - \log(1+x)
$$

and

$$
g(x) = \log(1+x) + g\Big(\frac{x}{2}\Big) - g\Big(\frac{x+1}{2}\Big).
$$

(Later we will make one general observation that produces all the necessary identities at once.) Therefore

$$
g(0) = g(0) - g(\tfrac{1}{2}) \quad\Rightarrow\quad g(\tfrac{1}{2}) = 0 \quad\Rightarrow\quad g(1) = 1 + g(\tfrac{1}{2}) - g(1) \quad\Rightarrow\quad g(1) = \tfrac{1}{2}.
$$

Likewise

$$
\begin{aligned}
&f(0) = f(0) - f(\tfrac{1}{2}) + g(0) \quad\Rightarrow\quad f(\tfrac{1}{2}) = g(0) \quad\Rightarrow \\
&f(1) = f(\tfrac{1}{2}) - f(1) + g(1) - 1 \quad\Rightarrow\quad f(1) = \tfrac{1}{2}g(0) + \tfrac{1}{4} - \tfrac{1}{2} = \tfrac{1}{2}g(0) - \tfrac{1}{4}.
\end{aligned}
$$

We don't know what $g(0)$ is, but it will cancel out later on. Furthermore, we note that the constant $g(0)$ is related to $N'(0)$ from a paper by Flajolet and Martin [6]; this observation is, however, irrelevant here. The notation refers to $N(s)$ which is the analytic continuation of

$$\sum_{j\ge 1}(-1)^{\nu(j)}/j^s.$$

Set

$$h(x)=\sum_{k\ge 0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log^2(2^k+j+x).$$

Then we derive in a similar fashion that

$$h(x)=\log^2(1+x)+2g\Big(\frac{x}{2}\Big)+h\Big(\frac{x}{2}\Big)-2g\Big(\frac{x+1}{2}\Big)-h\Big(\frac{x+1}{2}\Big).$$

Hence

$$h(0)=2g(0)+h(0)-2g(\tfrac12)-h(\tfrac12)\quad\Rightarrow\quad 0=2g(0)-h(\tfrac12)\quad\Rightarrow\quad h(\tfrac12)=2g(0),$$

and

$$h(1)=1+2g(\tfrac12)+h(\tfrac12)-2g(1)-h(1)\quad\Rightarrow\quad 2h(1)=1+2g(0)-1\quad\Rightarrow\quad h(1)=g(0).$$

Note that

$$f(1)-\tfrac12h(1)=\tfrac12g(0)-\tfrac14-\tfrac12g(0)=-\tfrac14,$$

and the quantity $g(0)$ cancels out. Writing it in extended form,

$$\sum_{k\ge1}k\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log(2^k+j+1)-\frac12\sum_{k\ge0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\log^2(2^k+j+1)=-\frac14.$$

Hence $H''(0)=-L^2/2$.

Let us turn to the third derivative. Omitting the details (introducing various auxiliary functions), we can prove that $H'''(0)=-L^3/2$. But we will soon show (in Section 5) how one can streamline these computations to get all moments in a mechanical way.

Now we substitute $\alpha=\tilde\alpha$ in $M(\alpha)/L$ and expand w.r.t. α. By (8) and (9), this immediately gives

$$\begin{aligned}\phi(0)&=\frac12,\\ \phi'(0)&=\frac12\left(\frac{\gamma}{L}+\frac12\right),\\ \phi''(0)&=\frac12\left(\frac13+\frac{\gamma}{L}+\frac{\gamma^2}{L^2}+\frac{\pi^2}{6L^2}\right).\end{aligned}$$

Note that this conforms to [1], as it should.

4 The periodicities

Let us analyze $F_1(\chi_l)$. To simplify the notation, we will simply write χ. We have

$$\phi(\alpha) = \frac{1}{L}\Gamma(-\alpha/L)\sum_{k\ge1}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}e^{-\alpha k}(2^k+j+1)^{\alpha/L},$$

$$\begin{aligned}\phi'(\alpha) = \frac{1}{L}\biggl[&-\frac{1}{L}\Gamma(-\alpha/L)\psi(-\alpha/L)\sum_{k\ge1}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}e^{-\alpha k}(2^k+j+1)^{\alpha/L}\\ &+\Gamma(-\alpha/L)\biggl[-\sum_{k\ge1}\sum_{j=0}^{2^k-1}k(-1)^{\nu(j)}e^{-\alpha k}(2^k+j+1)^{\alpha/L}\\ &+\sum_{k\ge1}\sum_{j=0}^{2^k-1}\log(2^k+j+1)(-1)^{\nu(j)}e^{-\alpha k}(2^k+j+1)^{\alpha/L}\biggr]\biggr],\end{aligned}$$

$$F_1(\chi) = -\frac{1}{L^2}\Gamma(\chi)\sum_{k\ge1}\sum_{j=0}^{2^k-1}(2^k+j+1)^{-\chi}\bigl(\psi(\chi)+kL-L\log(2^k+j+1)\bigr).$$

We need an auxiliary function

$$g(x) := \sum_{k\ge0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}(2^k+j+x)^{-\chi}.$$

As usual,

$$\begin{aligned}g(x) &= (1+x)^{-\chi} + \sum_{k\ge0}\sum_{j=0}^{2^{k+1}-1}(-1)^{\nu(j)}(2^{k+1}+j+x)^{-\chi}\\ &= (1+x)^{-\chi} + \sum_{k\ge0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\Bigl(2^k+j+\frac{x}{2}\Bigr)^{-\chi} - \sum_{k\ge0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\Bigl(2^k+j+\frac{x+1}{2}\Bigr)^{-\chi},\end{aligned}$$

and so

$$g(x) = (1+x)^{-\chi} + g\Bigl(\frac{x}{2}\Bigr) - g\Bigl(\frac{x+1}{2}\Bigr).$$

Therefore

$$\begin{aligned}g(0) = 1 + g(0) - g(\tfrac12) \quad &\Rightarrow \quad g(\tfrac12) = 1 \quad \Rightarrow \quad g(1) = 1 + g(\tfrac12) - g(1)\\ &\Rightarrow \quad 2g(1) = 2 \quad \Rightarrow \quad g(1) = 1.\end{aligned}$$

We need a further auxiliary function

$$f(x) := \sum_{k\ge0} k \sum_{j=0}^{2^k-1}(-1)^{\nu(j)}(2^k+j+x)^{-\chi}$$

and derive

$$f(x) = \sum_{k\ge 1}(k-1)\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}(2^k+j+x)^{-\chi} - (1+x)^{-\chi} + g(x)$$
$$= f\Big(\frac{x}{2}\Big) - f\Big(\frac{x+1}{2}\Big) - (1+x)^{-\chi} + g(x).$$

Therefore

$$f(0) = f(0) - f(\tfrac{1}{2}) - 1 + g(0) \quad\Rightarrow\quad f(\tfrac{1}{2}) = -1 + g(0) \quad\Rightarrow\quad f(1) = f(\tfrac{1}{2}) - f(1)$$
$$\Rightarrow\quad f(1) = -\tfrac{1}{2} + \tfrac{1}{2}g(0).$$

We need a third function

$$h(x) := \sum_{k\ge 0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}(2^k+j+x)^{-\chi}\log(2^k+j+x).$$

The usual procedure (that we suppress) leads to

$$h(x) = h\Big(\frac{x}{2}\Big) - h\Big(\frac{x+1}{2}\Big) + g\Big(\frac{x}{2}\Big) - g\Big(\frac{x+1}{2}\Big) + (1+x)^{-\chi}\log(1+x).$$

Therefore

$$h(0) = h(0) - h(\tfrac{1}{2}) + g(0) - g(\tfrac{1}{2}) \quad\Rightarrow\quad h(\tfrac{1}{2}) = g(0) - 1 \quad\Rightarrow$$
$$h(1) = h(\tfrac{1}{2}) - h(1) + g(\tfrac{1}{2}) - g(1) + 1 \quad\Rightarrow\quad h(1) = \tfrac{1}{2}g(0).$$

Now

$$F_1(\chi) = -\frac{1}{L^2}\Gamma(\chi)\Big[\psi(\chi)(g(1)-1) + Lf(1) - L(h(1)-1)\Big]$$
$$= -\frac{1}{L}\Gamma(\chi)\Big[-\frac{1}{2} + \frac{1}{2}g(0) - \frac{1}{2}g(0) + 1\Big] = -\frac{1}{2L}\Gamma(\chi).$$

Finally, we turn to $F_2(\chi)$. Omitting the details, we can prove that

$$F_2(\chi) = -\frac{\Gamma(\chi)}{2L} + \frac{\Gamma(\chi)\psi(\chi)}{L^2}.$$

Note that the periodicities conform to [1]. We will not pursue these computations as we will soon give a simple, direct way of getting all the expressions we need.

To obtain the conditioned moments, we divide by $\mathbb{P}(\mathcal{B}) = 1/2$. So

$$\mathbb{E}(J - \log n|\mathcal{B}) \sim 2\Big[\phi'(0) + \sum_{l\ne 0}F_1(\chi_l)e^{-2l\pi\mathbf{i}\log n}\Big],$$
$$\mathbb{E}((J - \log n)^2|\mathcal{B}) \sim 2\Big[\phi''(0) + \sum_{l\ne 0}F_2(\chi_l)e^{-2l\pi\mathbf{i}\log n}\Big]$$

and we have the conditioned variance

$$\begin{aligned}\mathbb{V}(J|\mathcal{B}) \sim 2\phi''(0) - 4(\phi'(0))^2 + 2\sum_{l\neq 0} F_2(\chi_l)e^{-2l\pi\mathbf{i}\log n}\\ - 8\phi'(0)\left[\sum_{l\neq 0} F_1(\chi_l)e^{-2l\pi\mathbf{i}\log n}\right] - 4\left[\sum_{l\neq 0} F_1(\chi_l)e^{-2l\pi\mathbf{i}\log n}\right]^2.\end{aligned}$$

This has been computed in [1] already.

5 General Allouche-Shallit-type formulas and direct moments computations

In this section, we first prove a very simple expression for $H^{(m)}(0)$ which translates to $\phi(\alpha)$. As a byproduct, we have a direct form for all moments and also a family of new interesting identities.

Set

$$A(a,b) := \sum_{k\geq 0} k^a \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log^b(2^k+j+1).$$

We have seen already, by (6), that

$$A(0,1) = \frac{1}{2},$$

which is the Allouche-Shallit identity, and

$$A(0,2) - 2A(1,2) = \frac{1}{2}.$$

Introducing more and more auxiliary functions, one is led to

$$A(0,3) - 3A(1,2) + 3A(2,2) = \frac{1}{2},$$
$$A(0,4) - 4A(1,3) + 6A(2,2) - 4A(3,1) = \frac{1}{2}.$$

An obvious pattern appears:

$$\sum_{l=0}^{m} \binom{m}{l} (-1)^{m-l} A(l, m-l) = \frac{1}{2}.$$

All these formulas follow from the *master formula* for $m \geq 0$:

Theorem 2.

$$\sum_{k\geq 0}\sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \Big(\log(2^k+j+1) - k\Big)^m = \frac{1}{2}.$$

Proof. Here is the simple proof. Set

$$f(x) := \sum_{k\geq 0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\Big(\log(2^k+j+x)-k\Big)^m.$$

Then

$$\begin{aligned}
f(x) &= \log^m(1+x) + \sum_{k\geq 0}\sum_{j=0}^{2^{k+1}-1}(-1)^{\nu(j)}\Big(\log(2^{k+1}+j+x)-k-1\Big)^m \\
&= \log^m(1+x) + \sum_{k\geq 0}\sum_{j=0}^{2^{k+1}-1}(-1)^{\nu(j)}\Big(\log(2^k+\frac{j+x}{2})-k\Big)^m \\
&= \log^m(1+x) + \sum_{k\geq 0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\Big(\log(2^k+\frac{2j+x}{2})-k\Big)^m \\
&\qquad - \sum_{k\geq 0}\sum_{j=0}^{2^k-1}(-1)^{\nu(j)}\Big(\log(2^k+\frac{2j+1+x}{2})-k\Big)^m \\
&= \log^m(1+x) + f\Big(\frac{x}{2}\Big) - f\Big(\frac{x+1}{2}\Big).
\end{aligned}$$

Therefore, by setting $x=0$,

$$f(\tfrac{1}{2}) = 0,$$

and by setting $x=1$,

$$f(1) = \frac{1}{2}.$$

■

The master theorem leads to (note that the summation on k is from $k=0$, but the summation in $H^{(m)}$ starts from $k=1$)

$$H^{(m)}(0) = -\frac{1}{2}L^m, \qquad m\geq 1.$$

This gives the following Laplace transforms:

Theorem 3.

$$M(\alpha) = \Gamma(-\alpha)H(\alpha) = -\frac{1}{2}\Gamma(-\alpha)(e^{L\alpha}-1),$$
$$\phi(\alpha) = M(\tilde{\alpha})/L = -\frac{1}{2L}\Gamma(-\tilde{\alpha})(e^{L\tilde{\alpha}}-1).$$

Now that we found that $\phi(\alpha)$ has such a simple explicit form, we recover immediately $\phi(0)$, $\phi'(0)$, $\phi''(0)$, and a new expression $\phi'''(0)$ related to the third moment:

$$\phi'''(0) = \frac{\zeta(3)}{L^3} + \frac{\pi^2\gamma}{4L^3} + \frac{\pi^2}{8L^2} + \frac{\gamma}{2L} + \frac{\gamma^3}{2L^3} + \frac{1}{8} + \frac{3\gamma^2}{4L^2}. \tag{11}$$

Similarly, we recover $F_1(\chi)$, $F_2(\chi)$, and a new expression

$$F_3(\chi) = \Gamma(\chi)\left[-\frac{3\psi(1,\chi)}{2L^3} - \frac{3\psi^2(\chi)}{2L^3} + \frac{3\psi(\chi)}{L^2} - \frac{1}{2L}\right]. \tag{12}$$

For instance, we can state the third moment:

Theorem 4.

$$\mathbb{E}\big((J-\log n)^3; \mathcal{B}\big) \sim \phi'''(0) + \sum_{l\neq 0} F_3(\chi_l) e^{-2l\pi\mathbf{i}\log n}$$

with (11) and (12).

All moments can in principle be computed automatically.

6 The LMV(-1) case

If the largest missing value is exactly one smaller than the largest value, we say, following Archibald and Knopfmacher [1], that the sequence has the LMV(-1) property. In this case, contrarily to the previous sections, our urn model leads almost immediately to all probabilistic properties: we have here a very simple form for $\phi(\alpha)$. Let $\mathcal{B}$ be now the event that the LMV(-1) property is satisfied, and J be the largest missing value exactly one smaller than the largest value. We easily derive (we just plug $r = 1$ into (2))

$$\begin{aligned}\phi(\alpha) &= \int_0^\infty y^{-\tilde{\alpha}} e^{-y}\left[1 - e^{-ye^{-L}}\right] e^{-ye^{-L}} \frac{dy}{Ly}\\ &= \frac{1}{L}\left(2^{-\tilde{\alpha}}3^{\tilde{\alpha}} - 2^{\tilde{\alpha}}\right)\Gamma(-\tilde{\alpha}).\end{aligned}$$

Hence, automatically,

$$\begin{aligned}\phi(0) &= 2 - \log 3,\\ \phi'(0) &= \frac{2\gamma}{L} + \log 3 - \frac{\gamma\log 3}{L} - \frac{\log^2 3}{2},\\ \phi''(0) &= -\frac{\log^3 3}{3} + \frac{2}{3} + \log^2 3 - \frac{\log 3\gamma^2}{L^2} + 2\frac{\gamma\log 3}{L} - \frac{\pi^2\log 3}{6L^2} - \log 3 + \frac{\pi^2}{3L^2} + 2\frac{\gamma^2}{L^2} - \frac{\gamma\log^2 3}{L}.\end{aligned}$$

The first two expressions are already given in [1]. The third one is new. All moments are obtained in the same way.

Also we easily derive

$$\begin{aligned}F_0(\chi_l) &= \frac{1}{L}(3^{-\chi_l} - 1)\Gamma(\chi_l),\\ F_1(\chi_l) &= -\frac{\Gamma(\chi_l)3^{-\chi_l}}{L^2}\left[L - L\log 3 + L3^{\chi_l} + \psi(\chi_l) - \psi(\chi_l)3^{\chi_l}\right],\\ F_2(\chi_l) &= -\frac{\Gamma(\chi_l)3^{-\chi_l}}{L^3}\Big[-L^2 + L^2 3^{\chi_l} + 2\log 3L^2 - 2\psi(\chi_l)L - 2\psi(\chi_l)L3^{\chi_l}\\ &\qquad - \log^2 3L^2 + 2\psi(\chi_l)\log 3L - \psi(1,\chi_l) + \psi(1,\chi_l)3^{\chi_l} - \psi^2(\chi_l) + \psi^2(\chi_l)3^{\chi_l}\Big].\end{aligned}$$

Again the first two expressions are already given in [1]. The third one is new. That described the moments, but we can also get the (asymptotic) probability that $\mathcal{B}$ holds:

$$\mathbb{P}(\mathcal{B}) \sim \phi(0) + \sum_{l \neq 0} F_0(\chi_l) e^{-2l\pi \mathbf{i} \log n}.$$

7 Conclusion

We succeeded in analyzing the instance $p = 1/2$, related to random compositions. The (semi)automatic computations of all moments led us to a one parameter extension of the celebrated identity due to Allouche and Shallit. Unfortunately, this analysis cannot be extended to the case $p \neq 1/2$, as no expansions that involve $\nu(k)$ are available. Of course we could use a general integral form similar to (2) and use the theorems we developed in [10], but it would not be really explicit. One could also use some recursion on the number of gaps and some inclusion-exclusion expressions, but it would not be simpler.

The instance $p = 1/2$ is privileged inasmuch as it leads to very beautiful mathematics, and this is just appropriate for the present birthday issue.

References

[1] M. Archibald and A. Knopfmacher. The largest missing value in a composition of an integer. *Discrete Math.* **311** (2011), 723–731.

[2] J.-P. Allouche, H. Cohen, M. Mendès France, and J. O. Shallit. De nouveaux curieux produits infinis. *Acta Arith.* **49** (1987), 141–153.

[3] G. Louchard, H. Prodinger, and M. D. Ward. The number of distinct values of some multiplicity in sequences of geometrically distributed random variables. *Discr. Math. and Theoret. Comp. Sci.* (2005), 231–256. 2005 International Conference on Analysis of Algorithms.

[4] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, Cambridge, 2003.

[5] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, 1965.

[6] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. *J. Comput. Syst. Sci.* **31** (1985), 182–209.

[7] P. Hitczenko and A. Knopfmacher. Gap-free samples of geometric random variables. *Discrete Math.* **294** (2005), 225–239.

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

[9] M. Loève. *Probability Theory*, 3rd edition. D. Van Nostrand, 1963.

[10] G. Louchard and H. Prodinger. Asymptotics of the moments of extreme-value related distribution functions. *Algorithmica* **46** (2006), 431–467. Long version: http://www.ulb.ac.be/di/mcs/louchard/moml.ps.

[11] G. Louchard and H. Prodinger. On gaps and unoccupied urns in sequences of geometrically distributed random variables. *Discrete Math.* **308-309** (2008), 1538–1562. Long version: http://www.ulb.ac.be/di/mcs/louchard/gaps18.ps.

[12] D. Robbins. Solution to Elementary problem E2692. *Amer. Math. Monthly* **86** (1979), 395–395.

[13] D. R. Woods. Elementary problem E2692. *Amer. Math. Monthly* **85** (1978), 48.

2010 *Mathematics Subject Classification*: Primary 05A16; Secondary 60C05.
Keywords: composition, largest missing value, asymptotic expansion, identity, sum-of-digits function.

(Concerned with sequence A000120.)

Received April 18 2012; revised version received October 25 2012. Published in *Journal of Integer Sequences*, March 2 2013.

Return to Journal of Integer Sequences home page.