

The largest missing value in a composition of an integer and some Allouche-Shallit-like identities

Guy Louchard*

Helmut Prodinger†

February 11, 2011

Abstract

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

1 Introduction

Our attention was recently attracted by a paper by M. Archibald and A. Knopfmacher [3] 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 largest missing value is exactly one smaller than the largest value, they say that the sequence has the LMV property. They compute the LMV probability, and the corresponding mean value.

Here, we think about an urn model, with urns labelled $1, 2, \dots$, 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:

- We have asymptotic independence of urns, for all events related to urn j , $j = \mathcal{O}(\log n)$. This is proved, by Poissonization-DePoissonization, in [9], [10] and [6] (in this paper for $p = 1/2$, but the proof is easily adapted). The error term is $\mathcal{O}(n^{-C})$ where C is a positive constant.
- We obtain asymptotic distributions of the random variables of interest. The number of balls in urn j , $j = \mathcal{O}(\log n)$ 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 $\mathcal{O}(n^{-1})$.
- We have uniform integrability for the moments of our random variables. 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 [7, Section 11.4]). For the kind of limiting distributions we consider here, the rate of convergence is analyzed in detail in [8] and [10]. The error term is $\mathcal{O}(n^{-C})$.
- Asymptotic expressions for the moments are obtained by Mellin transforms. The error term is $\mathcal{O}(n^{-C})$.

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

†University of Stellenbosch, Mathematics Department, 7602 Stellenbosch, South Africa, email: hproding@sun.ac.za

- $\Gamma(s)$ decreases exponentially in the direction $i\infty$:

$$|\Gamma(\sigma + it)| \sim \sqrt{2\pi}|t|^{\sigma-1/2}e^{-\pi|t|/2}.$$

Also, we have a “slow increase property” for all other functions we encounter. So inverting the Mellin transforms is easily justified.

If we compare the approach in this paper with other ones that appeared previously (related to similar problems), then we can 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.

Let us now consider the composition of an integer ν , i.e. $\nu = \sum_{i=1}^n x_i$, where x_i are natural numbers. Considering all compositions as equiprobable, we know (see [6]) that the number of parts n is asymptotically Gaussian, $\nu \rightarrow \infty$:

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

and that the part sizes are asymptotically identically distributed as $\text{GEOM}(1/2)$ and *independent*. All distributions we analyze in this paper depend on $\log_2 n$. Hence, with (1), the same random variables related to ν are asymptotically equivalent by replacing $\log_2 n$ by $\log_2 \nu - 1$ (see [6]).

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

$$\begin{aligned} n &:= \text{number of balls, } n \text{ large,} \\ \log &:= \log_2, \\ \eta &:= j - \log n, \\ L &:= \ln 2, \\ \tilde{\alpha} &:= \alpha/L, \\ \chi_l &:= \frac{2l\pi i}{L}. \end{aligned}$$

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 [8], let J be the position of this maximal empty urn before the last non-empty urn. We have (here and in the sequel \sim always denotes $\sim_{n \rightarrow \infty}$)

$$\mathbb{P}[J = j; \mathcal{B}] \sim \varphi(\eta),$$

$$\varphi(\eta) := \exp(-e^{-L\eta}) \sum_{r \geq 1} \prod_{i=1}^r \left[1 - \exp(-e^{-L(\eta+i)})\right] \exp(-e^{-L(\eta+r)}).$$

Here is the explanation: urn j is empty, we have $r \geq 1$ non-empty urns after urn j , and all urns after urn $j + r$ are empty. Following again [8], we compute

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

Again by [8], we have, with

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

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

But, from [9] and [5], we know that $\mathbb{P}(\mathcal{B}) = 1/2$. So we should show that

$$\int_0^\infty e^{-y} \sum_{r \geq 1} \prod_{i=1}^r (1 - e^{-y/2^i}) e^{-y/2^r} \frac{dy}{Ly} = \frac{1}{2}. \quad (3)$$

Set

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

Then

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

Now

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

So

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

If one iterates this formally, one gets

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

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

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

by the Allouche-Shalit identity (see [2, 1, 9]). This proves (3).

Now we must confirm that $F_0(\chi_l) = 0$. Set

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

Then

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

Now

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

So

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

Iterating:

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

and

$$\phi(\alpha) = M(\tilde{\alpha})/L. \quad (4)$$

Now

$$F_0(s) = \frac{1}{L}\Gamma(s) \sum_{k \geq 1} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} 2^{sk} (2^k + j + 1)^{-s},$$

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

It is already proved in [9] that this is zero. (But it follows readily from the computations that appear in this paper.)

2 The moments

The purpose of this section is to obtain the dominant terms of the moments, the first two of which have been computed already in [3].

The paper [8] tells us that

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

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

Now

$$\begin{aligned} \phi(\alpha) &= M(\tilde{\alpha})/L, \\ M(\alpha) &= \Gamma(-\alpha)H(\alpha), \end{aligned} \quad (5)$$

$$H(\alpha) := \sum_{k \geq 1} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} 2^{-\alpha k} (2^k + j + 1)^\alpha,$$

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

$$H(\alpha) = 0 + H'(0)\alpha + H''(0)\frac{\alpha^2}{2} + H'''(0)\frac{\alpha^3}{6} + \dots,$$

$$H'(0) = \sum_{k \geq 1} \left[-Lk \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} + L \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log(2^k + j + 1) \right],$$

$$H''(0) = \sum_{k \geq 1} \left[L^2 k^2 \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} - 2L^2 k \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log(2^k + j + 1) + L^2 \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log^2(2^k + j + 1) \right],$$

$$H'''(0) = \sum_{k \geq 1} \left[-L^3 k^3 \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} + 3L^3 k^2 \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log(2^k + j + 1) \right. \\ \left. - 3L^3 k \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log^2(2^k + j + 1) + \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log^3(2^k + j + 1) \right].$$

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

$$f(x) = \sum_{k \geq 0} k \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log_2(2^k + j + x),$$

$$g(x) = \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log_2(2^k + j + x).$$

A similar computation as before shows that

$$f(x) = f\left(\frac{x}{2}\right) - f\left(\frac{x+1}{2}\right) + g(x) - \log_2(1+x)$$

and

$$g(x) = \log_2(1+x) + g\left(\frac{x}{2}\right) - g\left(\frac{x+1}{2}\right).$$

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

$$g(0) = g(0) - g\left(\frac{1}{2}\right) \Rightarrow g\left(\frac{1}{2}\right) = 0 \Rightarrow g(1) = 1 + g\left(\frac{1}{2}\right) - g(1) \Rightarrow g(1) = \frac{1}{2}.$$

Likewise

$$f(0) = f(0) - f\left(\frac{1}{2}\right) + g(0) \Rightarrow f\left(\frac{1}{2}\right) = g(0) \Rightarrow$$

$$f(1) = f\left(\frac{1}{2}\right) - f(1) + g(1) - 1 \Rightarrow f(1) = \frac{1}{2}g(0) + \frac{1}{4} - \frac{1}{2} = \frac{1}{2}g(0) - \frac{1}{4}.$$

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 [4]. Set

$$h(x) = \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log_2^2(2^k + j + x).$$

Then we derive in a similar style that

$$h(x) = \log_2^2(1+x) + 2g\left(\frac{x}{2}\right) + h\left(\frac{x}{2}\right) - 2g\left(\frac{x+1}{2}\right) - h\left(\frac{x+1}{2}\right).$$

Hence

$$h(0) = 2g(0) + h(0) - 2g\left(\frac{1}{2}\right) - h\left(\frac{1}{2}\right) \Rightarrow 0 = 2g(0) - h\left(\frac{1}{2}\right) \Rightarrow h\left(\frac{1}{2}\right) = 2g(0),$$

and

$$h(1) = 1 + 2g\left(\frac{1}{2}\right) + h\left(\frac{1}{2}\right) - 2g(1) - h(1) \Rightarrow 2h(1) = 1 + 2g(0) - 1 \Rightarrow h(1) = g(0).$$

Note that

$$f(1) - \frac{1}{2}h(1) = \frac{1}{2}g(0) - \frac{1}{4} - \frac{1}{2}g(0) = -\frac{1}{4},$$

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

$$\sum_{k \geq 1} k \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log_2(2^k + j + 1) - \frac{1}{2} \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log_2^2(2^k + j + 1) = -\frac{1}{4}.$$

Hence $H''(0) = -\frac{2L^2}{4} = -\frac{L^2}{2}$.

Let us turn to the third derivative. Omitting the details (introducing various auxiliary functions), we can prove that $H'''(0) = -\frac{L^3}{2}$. But we will soon show how one can streamline these computations to get all moments.

Now we substitute $\alpha = \tilde{\alpha}$ in $M(\alpha)/L$ and expand w.r.t. α . By (4) and (5), this immediately gives

$$\begin{aligned}\phi(0) &= \frac{1}{2}, \\ \phi'(0) &= \frac{1}{2} \left(\frac{\gamma}{L} + \frac{1}{2} \right), \\ \phi''(0) &= \frac{1}{2} \left(\frac{1}{3} + \frac{\gamma}{L} + \frac{\gamma^2}{L^2} + \frac{\pi^2}{6L^2} \right).\end{aligned}$$

Note that this conforms to [3].

3 The periodicities

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

$$\begin{aligned}\phi(\alpha) &= \frac{1}{L} \Gamma(-\alpha/L) \sum_{k \geq 1} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} e^{-\alpha k} (2^k + j + 1)^{\alpha/L}, \\ \phi'(\alpha) &= \frac{1}{L} \left[-\frac{1}{L} \Gamma(-\alpha/L) \psi(-\alpha/L) \sum_{k \geq 1} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} e^{-\alpha k} (2^k + j + 1)^{\alpha/L} \right. \\ &\quad + \Gamma(-\alpha/L) \left[-\sum_{k \geq 1} \sum_{j=0}^{2^k-1} k (-1)^{\nu(j)} e^{-\alpha k} (2^k + j + 1)^{\alpha/L} \right. \\ &\quad \left. \left. + \sum_{k \geq 1} \sum_{j=0}^{2^k-1} \log(2^k + j + 1) (-1)^{\nu(j)} e^{-\alpha k} (2^k + j + 1)^{\alpha/L} \right] \right], \\ F_1(\chi) &= -\frac{1}{L^2} \Gamma(\chi) \sum_{k \geq 1} \sum_{j=0}^{2^k-1} (2^k + j + 1)^{-\chi} (\psi(\chi) + kL - L \log(2^k + j + 1)).\end{aligned}$$

We need an auxiliary function

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

As usual,

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

and so

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

Therefore

$$g(0) = 1 + g(0) - g\left(\frac{1}{2}\right) \Rightarrow g\left(\frac{1}{2}\right) = 1 \Rightarrow g(1) = 1 + g\left(\frac{1}{2}\right) - g(1) \Rightarrow 2g(1) = 2 \Rightarrow g(1) = 1.$$

We need a further auxiliary function

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

and derive

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

Therefore

$$f(0) = f(0) - f\left(\frac{1}{2}\right) - 1 + g(0) \Rightarrow f\left(\frac{1}{2}\right) = -1 + g(0) \Rightarrow f(1) = f\left(\frac{1}{2}\right) - f(1) \Rightarrow f(1) = -\frac{1}{2} + \frac{1}{2}g(0).$$

We need a third function

$$h(x) := \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} (2^k + j + x)^{-\chi} \log_2(2^k + j + x).$$

The usual procedure (that we suppress) leads to

$$h(x) = h\left(\frac{x}{2}\right) - h\left(\frac{x+1}{2}\right) + g\left(\frac{x}{2}\right) - g\left(\frac{x+1}{2}\right) + (1+x)^{-\chi} \log_2(1+x).$$

Therefore

$$\begin{aligned} h(0) &= h(0) - h\left(\frac{1}{2}\right) + g(0) - g\left(\frac{1}{2}\right) \Rightarrow h\left(\frac{1}{2}\right) = g(0) - 1 \Rightarrow \\ h(1) &= h\left(\frac{1}{2}\right) - h(1) + g\left(\frac{1}{2}\right) - g(1) + 1 \Rightarrow h(1) = \frac{1}{2}g(0). \end{aligned}$$

Now

$$\begin{aligned} F_1(\chi) &= -\frac{1}{L^2} \Gamma(\chi) \left[\psi(\chi)(g(1) - 1) + Lf(1) - L(h(1) - 1) \right] \\ &= -\frac{1}{L} \Gamma(\chi) \left[-\frac{1}{2} + \frac{1}{2}g(0) - \frac{1}{2}g(0) + 1 \right] = -\frac{1}{2L} \Gamma(\chi). \end{aligned}$$

Finally, we must 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 [3]. We will not pursue these computations as we will soon give a simple, direct way of getting all the expressions we need.

4 Conditioning on \mathcal{B}

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

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

and we have the conditioned variance

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

This has been computed in [3] already.

5 General Allouche/Shallit-like 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

$$A(a, b) := \sum_{k \geq 0} k^a \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \log_2^b(2^k + j + 1).$$

We have seen already 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

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

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 5.1

$$\sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \left(\log_2(2^k + j + 1) - k \right)^m = \frac{1}{2}.$$

Here is the simple proof. Set

$$f(x) := \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \left(\log_2(2^k + j + x) - k \right)^m.$$

Then

$$\begin{aligned}
f(x) &= \log_2^m(1+x) + \sum_{k \geq 0} \sum_{j=0}^{2^{k+1}-1} (-1)^{\nu(j)} \left(\log_2(2^{k+1} + j + x) - k - 1 \right)^m \\
&= \log_2^m(1+x) + \sum_{k \geq 0} \sum_{j=0}^{2^{k+1}-1} (-1)^{\nu(j)} \left(\log_2\left(2^k + \frac{j+x}{2}\right) - k \right)^m \\
&= \log_2^m(1+x) + \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \left(\log_2\left(2^k + \frac{2j+x}{2}\right) - k \right)^m \\
&\quad - \sum_{k \geq 0} \sum_{j=0}^{2^k-1} (-1)^{\nu(j)} \left(\log_2\left(2^k + \frac{2j+1+x}{2}\right) - k \right)^m \\
&= \log_2^m(1+x) + f\left(\frac{x}{2}\right) - f\left(\frac{x+1}{2}\right).
\end{aligned}$$

Therefore, by setting $x = 0$,

$$f\left(\frac{1}{2}\right) = 0,$$

and by setting $x = 1$,

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

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

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

This gives the following

Theorem 5.2

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

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:

$$\phi'''(0) = \frac{\zeta(3)}{L^3} + \frac{\pi^2\gamma}{4L^3} + \frac{\pi^2}{8L^2} + \frac{\gamma}{2L} + \frac{\gamma^3}{2L^3} + \frac{1}{8} + \frac{3\gamma^2}{4L^2}. \quad (6)$$

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

$$F_3(\chi) = \Gamma(\chi) \left[-\frac{3\psi(1, \chi)}{2L^3} - \frac{3\psi^2(\chi)}{2L^3} + \frac{3\psi(\chi)}{L^2} - \frac{1}{2L} \right]. \quad (7)$$

So we have

Theorem 5.3

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

with (6) and (7).

All moments can in principle be computed automatically.

6 The LMV case

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 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))

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

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 [3]. The third one is new. All moments are obtained in the same way.

Also we easily derive

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

Again the first two expressions are already given in [3]. The third one is new. That described the moments, but we can also get the (asymptotic) probability that \mathcal{B} holds:

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

7 Conclusion

We were successful to analyze 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.

References

- [1] J.-P. Allouche, H. Cohen, M. Mendès France, and J. O. Shallit. De nouveaux curieux produits infinis. *Acta Arithmetica*, 49(2):141–153, 1987.
- [2] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.
- [3] M. Archibald and A. Knopfmacher. The largest missing value in a composition of an integer. *Discrete Mathematics*, 2010. to appear.

- [4] P. Flajolet and G.N. Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31:182–209, 1985.
- [5] P. Hitczenko and A. Knopfmacher. Gap-free samples of geometric random variables. *Discrete Mathematics*, 294:225–239, 2005.
- [6] P. Hitczenko and G. Louchard. Distinctness of compositions of an integer: a probabilistic analysis. *Random Structures and Algorithms*, 19(3,4):407–437, 2001.
- [7] M. Loève. *Probability Theory*. D. Van Nostrand, 1963.
- [8] G. Louchard and H. Prodinger. Asymptotics of the moments of extreme-value related distribution functions. *Algorithmica*, 46:431–467, 2006. Long version: <http://www.ulb.ac.be/di/mcs/louchard/moml.ps>.
- [9] G. Louchard and H. Prodinger. On gaps and unoccupied urns in sequences of geometrically distributed random variables. *Discrete Mathematics*, 308,9:1538–1562, 2008. Long version: <http://www.ulb.ac.be/di/mcs/louchard/gaps18.ps>.
- [10] G. Louchard, H. Prodinger, and M.D. Ward. The number of distinct values of some multiplicity in sequences of geometrically distributed random variables. *Discrete Mathematics and Theoretical Computer Science*, AD:231–256, 2005. 2005 International Conference on Analysis of Algorithms.