Matrix compositions: a Probabilistic analysis

G. Louchard

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

Abstract

In this paper, we analyze large matrix compositions of size N: we are interested in the asymptotic properties of random variables such as number M of parts, the associated processes and limiting trajectories, leading to some Brownian motions. Also we compute the asymptotic distributions of the last part size LP, the largest part $\mathcal{M}_N$, the first empty size $\mathcal{E}_N$ and some measure of distinctness $\mathcal{D}_N$. We finally analyze Carlitz compositions, where two successive parts must be different.

1 Introduction

During the last Gascom meeting 2006 in Dijon, our attention was attracted by two papers by Munarini et al. [24] and Castiglione et al. [5], dealing in particular with matrix compositions. Let C be a composition of an integer N into k parts, $(\gamma_1, \ldots, \gamma_k)$, where γ_i is a (column) vector of dimension r (its components are non-negative integers, not all of them being 0), be defined as $N = \sum_{u=1}^{k} |\gamma_u|$. The size $|\gamma|$ of each part (column) γ is given by the sum of its components. For instance, with $N = 21, r = 2, k = 5$ we have the possible matrix composition:

$$\begin{bmatrix} 0 & 3 & 3 & 4 & 1 \\ 1 & 2 & 5 & 0 & 2 \end{bmatrix}$$

Classical (r=1) compositions have been the subject of considerable research. In particular, the probabilistic aspects of the compositions have been analyzed in [21] and [10], where many other references are given. In this paper, considering all compositions as equiprobable, we analyze the asymptotic ($N \to \infty$) properties of some parameters. We are interested in several random variables (RV) such as the number M of parts, the associated processes and limiting trajectories, leading to some Brownian motions. Also we compute the asymptotic distributions of the last part size LP, the largest part $\mathcal{M}_N$. Define the indicator variable $I_i := [\![\text{value } i \text{ appears among the } k \text{ sizes } |\gamma_1|, \ldots, |\gamma_k|]\!]$. We are interested in stochastic properties of the distinctness measured by $\mathcal{D}_N := \sum_i I_i$ and of the first empty size $\mathcal{E}_N$.

The paper is organized as follows: in Section 2, we present a first analysis of several RV, in Section 3 we derive asymptotic distributions for $\mathcal{D}_N$, $\mathcal{E}_N$, $\mathcal{M}_N$. Section 4 is devoted to Carlitz compositions, where two successive parts must be different. Section 5 presents a simulation algorithm and analyzes hitting times and maximum for Carlitz compositions thickness (sizes of parts). Section 6 concludes the paper. Our computations are of course inspired by [21] and [10], but some important modifications and adaptations will be necessary.

2 Probabilistic analysis of matrix compositions

ù In this section, we first analyze the number M of parts. Then we obtain the asymptotic distribution of the last part size LP. In the sequel, asymptotic always means “ when $N \to \infty$". Next we prove the asymptotic independence between two successive intermediate parts. Finally, we consider the associated processes and limiting trajectories.

2.1 Number of parts

The size of a part (column part) is defined as the sum of its components. Let $T(m,n)$ be the number of matrix compositions $C(n)$ (Compositions of n) with m parts and let $h_m(i,n)$ be the number of C with same characteristics and last part of size i. We shall mark n by z, m by w and i by θ. Let

$$f(z) := \frac{1}{(1-z)^r} - 1$$

denote the generating function (GF) of the total number of possible (non-empty) columns of size u:

$$f(z) = \sum_{u=1}^{\infty} z^u h_1(u,u).$$

We have

$$[z^n]f(z) = \binom{r+n-1}{r-1} =: P(n), \text{ say },$$

and

$$P(n) \sim \tilde{P}(n), n \to \infty,$$

with

$$\tilde{P}(n) := \frac{n^{r-1}}{(r-1)!}.$$

Denote by $g_j(\theta,z)$ the GF of $h_j(i,n)$. Hence

$$\phi(w,\theta,z) := \sum_{j=1}^{\infty} w^j g_j(\theta,z) = \sum_{j=1}^{\infty} w^j f(z)^{j-1} f(\theta z) = \frac{wf(\theta z)}{1-wf(z)}.$$

Set (these rather general notations will be more significant in the Carlitz case)

$$\begin{aligned} h(w,z) &:= 1 - wf(z), \\ h(z) &:= h(1,z) = 1 - f(z), \\ A_1(w,\theta,z) &:= wf(\theta z), \\ D_1(w,z) &:= \phi(w,1,z) = A_1(w,1,z)/[1 - A_1(w,1,z)]. \end{aligned}$$

$h(w,z)$ was already obtained in [5] (for $r=2$) and in [24].

Notice that

$$1 + D_1(w,z) = 1/h(w,z), \tag{2.1}$$

so

$$\phi(w,\theta,z) = \frac{A_1(w,\theta,z)}{h(w,z)}. \tag{2.2}$$

First set $\theta = 1$. When $w = 1$ we get the GF of the total number $T(\cdot,n)$ of $C(n)$: this is given by $D_1(1,z)$. The dominant singularity of $D_1(1,z)$ is given by the root z^* (with smallest modulus) of $h(1,z)$, i.e. $z^* = 1 - 2^{-1/r}$.

By singularity analysis, $T(\cdot,n)$ is asymptotically given by

$$T(\cdot,n) \sim -\frac{1}{z^* h_z(1,z^*)} \frac{1}{z^{*n}} = C_1/z^{*n}, n \to \infty, \tag{2.3}$$

with

$$C_1 = \frac{-1}{z^* h_z(1,z^*)} = \frac{1}{2m(2^{1/m}-1)}.$$

(F_z always means differentiation with respect to z and similarly for w). Equation (2.3) was already obtained in [24].

Let us continue our asymptotic analysis. By Bender's Theorems 1.3 and (3.2) in [2] we can get more: we obtain the asymptotic distribution of the number of parts M in $C(n)$, where we consider all possible $T(\cdot, n)$ as equiprobable. Let

$$\begin{aligned} r_1 &:= -h_w/h_z, \\ r_2 &:= -(r_1^2 h_{zz} + 2r_1 h_{zw} + h_w + h_{ww})/h_z. \end{aligned}$$

Setting $w = 1, z = z^*$, we first derive

$$\begin{aligned} \mu_1 &:= -r_1/z^*, \\ \sigma_1^2 &:= \mu_1^2 - r_2/z^*. \end{aligned}$$

Then Bender's results lead to the following theorem.

Theorem 4. *(Horizontal Distribution in $C(n)$). The number M of parts in a $C(n)$ of large given n is asymptotically Gaussian:*

$$\frac{M - n\mu_1}{\sqrt{n}\sigma_1} \overset{\mathcal{D}}{\sim} \mathcal{N}(0,1), \quad n \to \infty.$$

Also a local limit theorem holds:

$$T(m,n) \sim \frac{C_1}{z^{*n}} \frac{e^{-(m-n\mu_1)^2/(2n\sigma_1^2)}}{\sqrt{2\pi n}\sigma_1}, \quad n \to \infty, m - n\mu_1 = \mathcal{O}(\sqrt{n}).$$

Now if we fix m and consider n as variable (there are, of course, an infinite number of compositions for given m), we can obtain another asymptotic form for $T(m,n)$: the conditioned distribution is given by

$$[w^m z^n] D_1(w,z) = \frac{1}{z^{*n}} [w^m z^n] D_1(w, zz^*).$$

But, for $z = 1$, the dominant singularity of $D_1(w, zz^*)$ is $w^* = 1$. So with

$$\begin{aligned} C_2 &= \frac{-1}{h_w(1,z^*)}, \\ \mu_2 &:= 1/\mu_1, \\ \sigma_2^2 &= \sigma_1^2/\mu_1^3, \end{aligned} \tag{2.4}$$

we obtain the following theorem.

Theorem 5. *(Vertical Distribution in $C(n)$). For large given number of parts $m, T(m,n)$ is asymptotically given by*

$$T(m,n) \sim \frac{C_2}{z^{*n}} \frac{e^{-(n-m\mu_2)^2/(2m\sigma_2^2)}}{\sqrt{2\pi m}\sigma_2}, \quad m \to \infty, n - m\mu_2 = \mathcal{O}(\sqrt{m}).$$

Note that $C_1/C_2 = \mu_1$.

2.2 Distribution of the last part size in C

We want to analyze the asymptotic distribution of the size LP of the last part of a $C(n)$. Setting $w = 1$ in (2.2), we derive

$$[z^n]\phi(1,\theta,z) \sim -\frac{f(\theta z^*)}{z^{*n} z^* h_z(1,z^*)}, \quad n \to \infty,$$

uniformly for θ in some complex neighbourhood of the origin. This may be checked by the method of singularity analysis of Flajolet and Odlyzko, as used in Flajolet and Soria [8] or Hwang [11].

Normalizing by the total number of $C(n)$ in (2.3) this leads to the following asymptotic PGF for the last part size LP:

$$G(\theta) = f(\theta z^*).$$

Expanding, this leads to

$$\pi_1(i) := [\theta^i]G(\theta) = z^{*i}P(i),$$

with typical geometric behaviour. Of course $G(1) = 1$ as $f(z^*) = 1$ and

$$\mathbb{E}(LP) = \sum_{i=1}^{\infty} i\pi_1(i) = z^* f'(z^*) \equiv \mu_2,$$

as it should.

Once we have chosen ℓ with distribution $\pi_1(\ell)$, we chose each column configuration of the last part size with equidistribution $1/P(\ell)$.

We can now obtain more information from (2.2). The asymptotic distribution of the last part size in $C(n)$ of m parts is related to $[w^m z^n]\phi(w, \theta, z)$ and Bender's Theorems 1 and 3 lead to the GF:

$$T(\theta, m, n) \sim \frac{C_1}{z^{*n}} \frac{e^{-(m-n\mu_1)^2/(2n\sigma_1^2)}}{\sqrt{2\pi n}\sigma_1} G(\theta),$$

for $m - n\mu_1 = \mathcal{O}(\sqrt{n}\,)$, uniformly for θ in some complex neighbourhood of the origin. Again the uniformity can be checked by following Bender's analysis. Normalizing by $T(m, n)$ (see Theorem 2.1), this gives again $G(\theta)$. So we have proved the following result:

Theorem 6. *For $m - n\mu_1 = \mathcal{O}(\sqrt{n}\,)$, the asymptotic distribution of the size LP of the last part in $C(n)$ with m parts is given by $\pi_1(j) = z^{*j}P(j)$ for $n \to \infty$.*

We can even be more precise in the computation of $\mathbb{E}(M)$. Returning to the proof of Bender's theorem, we must compute from (2.2)

$$[\theta^k w^m]\varphi_1\varphi_2 \left[\frac{z^*}{r(w)}\right]^n,$$

with

$$\begin{aligned} \varphi_1(\theta, w) &= A_1(w, \theta, r(w)), \\ \varphi_2(w) &= -\frac{1}{r(w)h_z(w, r(w))}, \end{aligned}$$

and $r(w)$ is the root of (2.1) with $r(1) = z^*$. But according to Hwang [[11], [12, Theorem 2]], we know that the mean value of M for large n is given by $\mathbb{E}(M) \sim n\mu_1 + v'(0)$, where

$$v(s) := \log[\varphi_1(1, e^s)\varphi_2(e^s)/(\varphi_1(1,1)\varphi_2(1))],$$

i.e.

$$v'(0) = \frac{\varphi_{1,w}(1,1)}{\varphi_1(1,1)} + \frac{\varphi_{2,w}(1)}{\varphi_2(1)}.$$

But it is easy to check that $\varphi_{1,w}(1,1) = 0$ and

$$\begin{aligned} \varphi_{2,w}(1) &= [r_w h_z + z^* h_{z,z} r_w + z^* h_{z,w}]/[z^* h_z]^2, \\ \varphi_2(1) &= -1/(z^* h_z), \\ r_w &= -h_w/h_z. \end{aligned}$$

For further use, we can also compute $C_2(j)$ in Theorem 5, based on a first part of fixed size j. But by simple symmetry, this is given by $C_2\pi_1(j)$.

It is finally possible to derive a large deviation result for the number of parts, using $r(w)$: see details in [21].

2.3 Asymptotic independence of two successive intermediate parts

Let us now turn to two successive parts m_1, m_1+1, of size k, j, such that their distances from the first and the last part are of order $\mathcal{O}(n)$. Let $T(m,n,k)$ be the total number of $C(n)$ with m parts and last part size k, let $\#(m,n,m_1,k,j)$ be the total number of $C(n)$ with m parts, part m_1 of size k, part m_1+1 of size j, and set $m_2 := m - m_1$. We have, conditioning on j,

$$\#(m,n,m_1,k,j) = \sum_{n_1} T(m_1,n_1,k)T(m_2,n-n_1,\cdot,|j).$$

With Theorem 6 (vertical form) and Theorem 5, with $C_2(j) = C_2\pi_1(j)$, this leads, after normalization by $T(m,n)$, to the following distribution:

$$\begin{aligned}
&\frac{\#(m,n,m_1,k,j)}{T(m,n)} \\
&\sim \sum_{n_1} \frac{e^{-(n_1-m_1\mu_2)^2/(2m_1\sigma_2^2)}}{\sqrt{2\pi m_1}\sigma_2}\pi_1(k)\pi_1(j)\frac{e^{-(n-n_1-m_2\mu_2)^2/(2m_2\sigma_2^2)}}{\sqrt{2\pi m_2}\sigma_2} \Bigg/ \frac{e^{-(n-m\mu_2)/(2m\sigma_2^2)}}{\sqrt{2\pi m}\sigma_2} \\
&\sim \pi_1(k)\pi_1(j),
\end{aligned}$$

which shows the asymptotic independence. So we obtain the following theorem.

Theorem 7. *The asymptotic ($n \to \infty$) distribution of intermediate part size is given by $\pi_1(k)$, with mean μ_2. The parts are asymptotically independent.*

2.4 Associated processes and limiting trajectories

Let us now turn to trajectories. Let x_i be the size of the ith part and set $X(j) := \sum_1^j x_i$. We have $E[X(m)] \sim m\mu_2$ and we can check that $[\mathsf{var}[X(m)] \sim m\sigma_2^2$ as given by (2.4).

We can now apply the Functional Central Limit Theorem (see for instance Billingsley [3] p. 168 ff.) and we obtain the following result, where $B(t)$ is the standard Brownian Motion (BM).

Theorem 8.

$$\frac{X([Mt]) - \mu_2 Mt}{\sigma_2\sqrt{M}} \Rightarrow B(t), M \to \infty, t \in [0,1],$$

where $M :=$ number of parts of $C(n)$.

Let us now condition on $X(m) = n$. A realization of X for fixed n is given by Theorem 8, where we stop at a random time m such that $X(m) = n$. Proceeding as in [18] it is easy to check that this amounts to fix $M = n\mu_1$ in Theorem 8 (denominator) and we obtain the following result.

Theorem 9. *Conditioned on $X(m) = n$,*

$$\frac{X([mt]) - \mu_2 mt}{\sigma_2\sqrt{n\mu_1}} \Rightarrow B(t).$$

Let us now consider the oscillations of $X(\cdot)$ around its mean. Let us define the lower oscillation bound W_n^- and the range W_n:

$$W_n^- := -\inf_{i\in[0,m]}[X(i) - \mu_2 i]$$

,

$$W_n := \left\{\sup_{[0,m]} - \inf_{[0,m]}\right\}[X(i) - \mu_2 i],$$

where m is the number of parts in $C(n)$.

But now the lower bound W_n^-, normalized by $\sqrt{n\mu_1}\sigma_2$, is asymptotically given by

$$\inf_{[0,1]} B(t),$$

and the range W_n corresponds to the range

$$\left\{\sup_{[0,1]} - \inf_{[0,1]}\right\} B(t).$$

The densities of these RV are well known (see, for instance, Ito and McKean, [13]): the first one is given by

$$f_1(x) = \frac{2e^{-x^2/2}}{\sqrt{2\pi}},$$

the second one is given by

$$f_2(x) = \frac{8}{\sqrt{2\pi}} \sum_{k=1}^{\infty} k^2(-1)^{k-1} \exp[-k^2x^2/2],$$

or

$$f_2(x) = (\frac{2}{\pi})^{1/2} \frac{1}{x} L'(\frac{x}{2}),$$

where

$$\begin{aligned} L(z) &= 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1} \exp[-2k^2z^2] \\ &= (2\pi)^{1/2}\frac{1}{z}\sum_{k=1}^{\infty} \exp[-2(k-1)^2\pi^2/8z^2], \end{aligned}$$

by Jacobi θ−relations (see e.g. Whittaker and Watson [26]).

This immediately leads to the following result

Theorem 10. *For large n, and with $M = n\mu_1$, $W_n^-/(\sigma_2\sqrt{M})$ has the asymptotic density f_1, while $W_n/(\sigma_2\sqrt{M})$ has the asymptotic density f_2.*

In summary, we can see the C as a BM with some thickness (part size). The distribution of the thickness is characterized by Theorem 7.

3 Some asymptotic distributions

In this section, we derive asymptotic distributions for $\mathcal{D}_N$, $\mathcal{E}_N$, $\mathcal{M}_N$. Recall that we consider the composition of an integer N into k parts, $(\gamma_1, \ldots, \gamma_k)$, i.e. $N = \sum_1^k |\gamma_u|$, γ_u being the uth column. The size of each part (column) is given by the sum of its components.

3.1 Measure of distinctness

We define the indicator variable $I_i := [\![$ value i appears among these k sizes $]\!]$. Considering all compositions as equiprobable, we are interested in stochastic properties of the distinctness measured by $\mathcal{D}_N := \sum_i I_i$. Let us first fix the number of parts to m. $\mathcal{D}_N$ will then be denoted by $\mathcal{D}_m$ (and similarly for $\mathcal{E}, \mathcal{M}$).

We can now proceed as in Louchard and Prodinger [22, Section 4.8 and 5.9] and in Louchard, Prodinger and Ward [23], where it is shown that, for the kind of limiting distributions we consider here, we can replace Binomials by Poisson distributions and where the rate of convergence is analyzed in detail. This entails that the limiting moments are given by the moments of the limiting distribution (this is related to a uniform integrability condition, see Loève [17, Sec.11.4]).

In [23], it is also proved that the random variables I_i are asymptotically independent. The proof can be easily adapted to our case.

We have

$$\mathbb{E}(\mathcal{D}_m) \sim \sum_{i=1}^{\infty} \mathbb{E}(I_i) \sim \sum_{i=1}^{\infty} \left\{1 - e^{-m\pi_1(i)}\right\} =: G(m) \text{ say}.$$

This is a harmonic sum, so we define

$$\Lambda(s) := \sum_{j=1}^{\infty} \pi_1(j)^{-s}. \tag{3.1}$$

We have

$$\int_0^{\infty} (1 - e^{-x}) x^{s-1} dx = -\Gamma(s).$$

The Mellin transform of $G(m)$ is (for a good reference on Mellin transforms, see Flajolet et al. [6] or Szpankowski [25])

$$-\Gamma(s)\Lambda(s). \tag{3.2}$$

The fundamental strip of (3.2) is $s \in \langle -1, 0 \rangle$. We have

$$\Gamma(s) \sim \frac{1}{s} - \gamma + \dots.$$

When $r = 1$, we have $\Lambda(s) = \frac{2^s}{1-2^s}$ and

$$\Lambda(s) \sim -\frac{1}{\ln(2)s} - 1/2 + o(s),$$

and the other simple poles are given by $s = \frac{2\pi\infty l}{\ln(2)}, \quad l \in \mathbb{Z} \setminus \{0\}$.

We conjecture that, here,

$$\Lambda(s) \sim \frac{C_3}{s} + C_4 + \dots$$

for some constants C_3, C_4. and that $\Lambda(s)$ has only other simples poles at some real values $s = \xi_l$, $l \in \mathbb{Z} \setminus \{0\}$, with residues $C_{5,l}$.

For instance, for $r = 2$, Figure 1 gives $\frac{1}{\Lambda(s)}, \quad s \in (-50, 50)$. The poles appear to be at

$$-46.4, -40.8, -36.0, -30.7, -25.6, -20.4, -15.4, -10.1, -5.1, 0,$$

$$5.1, 10.1, 15.4, 20.4, 25.6, 30.7, 36.0, 40.8, 46.4.$$

For $r = 1$, we know that the poles are at

$$-46.05, -40.93, -35.81, -30.70, -25.58, -20.46, -15.35, -10.23, -5.11, 0,$$

$$5.11, 10.23, 15.35, 20.46, 25.58, 30.70, 35.81, 40.93, 46.05.$$

Also, we have a "slow increase property": the behaviour of $\Gamma(s)\Lambda(s)$ is similar to the one of $\Gamma(s)$ which decreases exponentially towards $\infty\infty$ (imaginary axis). This is necessary to allow moving the line of integration to the right (see details in [25]).

Using

$$G(m) = \frac{1}{2\pi\infty} \int_{\beta-\infty\infty}^{\beta+\infty\infty} -\Gamma(s)\Lambda(s) m^{-s} ds, \quad -1 < \beta < 0,$$

the asymptotic expression of $G(m)$ (for large m) is obtained by moving the line of integration to the right, for instance to the line $\Re(s) = C_6 > 0$, taking residues into account (with a negative sign). This gives, modulo our conjecture,

$$G(m) = -\text{Res}\left[-\Gamma(s)\Lambda(s)m^{-s}\right]\Big|_{s=0} - \sum_{l \neq 0} \text{Res}\left[-\Gamma(s)\Lambda(s)m^{-s}\right]\Big|_{s=\xi_l} + \mathcal{O}\left(m^{-C_6}\right),$$

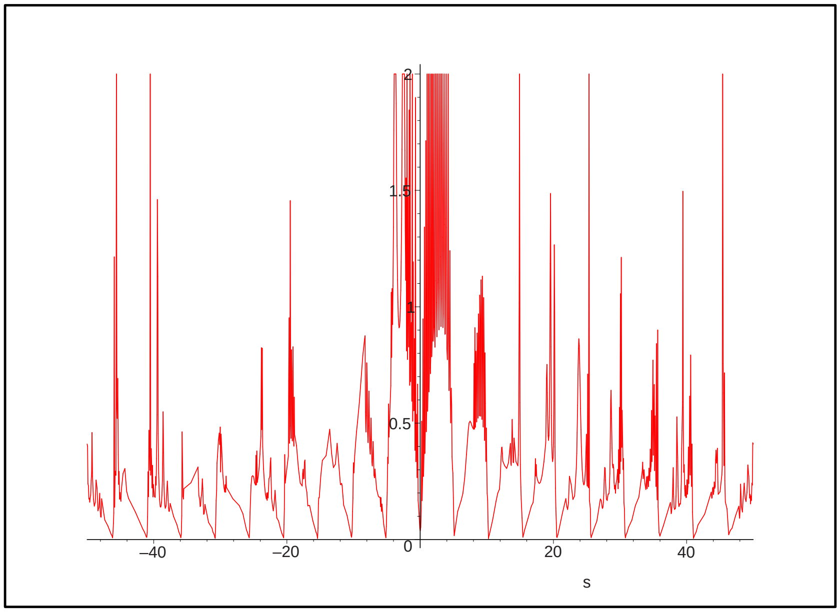

Figure 1: $\frac{1}{\Lambda(s)}, \quad s \in (-50, 50)$.

for $m \to \infty$.

This leads to

$$G(m) \sim G^*(m) + \sum_{l \neq 0} C_{5,l} \Gamma(\xi_l) e^{-\xi_l \ln m}, \quad m \to \infty, \tag{3.3}$$

with

$$G^*(m) := -C_3 \ln m - C_3 \gamma + C_4. \tag{3.4}$$

The higher centered moments of $\mathcal{D}_m$ can be obtained by analyzing

$$S_1(s) := \exp\{\ln(\tilde{G}_m(e^s)) - s\mathbb{E}(\mathcal{D}_m)\},$$

where $\tilde{G}_m(z)$ is the generating function of the probability distribution function of $\mathcal{D}_m$. Since, by asymptotic independence of I_i,

$$\begin{aligned} S_2(s) := \ln(\tilde{G}_m(e^s)) \quad &\sim \quad \sum_{j=1}^{\infty} \ln\left(1 + (e^s - 1)(1 - e^{m\pi_1(j)})\right) \\ &= \quad \sum_{i=1}^{\infty} \frac{(-1)^{i+1}}{i} (e^s - 1)^i \left\{ \sum_{j=1}^{\infty} (1 - e^{-m\pi_1(j)})^i \right\}, \end{aligned}$$

letting

$$V_i := \sum_{j=1}^{\infty} (1 - e^{-m\pi_1(j)})^i,$$

we see that

$$\begin{aligned} V_i &= \sum_{j=1}^{\infty} \left\{ \sum_{k=0}^{i} (-1)^k \binom{i}{k} e^{-km\pi_1(j)} \right\} \\ &= \sum_{j=1}^{\infty} \left\{ \sum_{k=0}^{i} (-1)^k \binom{i}{k} e^{-km\pi_1(j)} - \sum_{k=0}^{i} (-1)^k \binom{i}{k} \right\} \\ &= \sum_{j=1}^{\infty} \left\{ \sum_{k=0}^{i} (-1)^{k+1} \binom{i}{k} \left(1 - e^{-km\pi_1(j)}\right) \right\}. \end{aligned}$$

The last expression is a suitable form of an harmonic sum, which can again be asymptotically evaluated with Mellin transform.

One obtains

$$V_i \sim -C_3 \ln m - C_3\gamma + C_4 - C_3 \sum_{k=2}^{i} (-1)^{k+1} \binom{i}{k} \ln k + \beta_i(\ln m), \tag{3.5}$$

where $\beta_i(x)$ are periodic functions with mean 0. Hence,

$$S_2(s) = s[-C_3 \ln m - C_3\gamma + C_4] - C_3 \sum_{i=2}^{\infty} \frac{(-1)^{i+1}(e^s-1)^i}{i} B_i + \sum_{i=1}^{\infty} \frac{(-1)^{i+1}(e^s-1)^i}{i} \beta_i,$$

where $B_i = \sum_{k=2}^{i} (-1)^{k+1} \binom{i}{k} \ln k$. For instance,

$$\begin{aligned} B_1 &= 0, \\ B_2 &= -\ln 2, \\ B_3 &= -3\ln 2 + \ln 3, \\ B_4 &= -6\ln 2 + 4\ln 3 - \ln 4, \\ B_5 &= -10\ln 2 + 10\ln 3 - 5\ln 4 + \ln 5. \end{aligned}$$

In order to derive the constant term in the Fourier expansions (in $\ln m$), we consider

$$S_3(s) = \exp\left[\sum_{i=2}^{\infty} -C_3 \frac{(-1)^{i+1}(e^s-1)^i}{i} B_i\right]. \tag{3.6}$$

From this equation, we obtain

$$\begin{aligned} \tilde{\sigma}^2 := \mathsf{var}(\mathcal{D}_m) &\sim -C_3 \ln 2, \\ \tilde{\mu}_3 := \mu_3(\mathcal{D}_m) &\sim -C_3(-3\ln 2 + 2\ln 3), \\ \tilde{\mu}_4 := \mu_4(\mathcal{D}_m) &\sim 2C_3(-5\ln 2 + 6\ln 3 - 3\ln 4), \\ \tilde{\mu}_5 := \mu_5(\mathcal{D}_m) &\sim -C_3(-45\ln 2 + 70\ln 3 - 60\ln 4 + 24\ln 5). \end{aligned}$$

The neglected terms are made of periodic functions with small amplitude and of $\mathcal{O}(\frac{1}{m})$ contributions.

As $M \sim \mathcal{N}(N\mu_1, N\sigma_1^2)$, we must replace each $\ln m$ by $\ln N + \ln \mu_1$. The same modification is applicable in all next subsections.

3.2 The largest part

Set $L := -\ln(z^*), \log = \log_{1/z^*}$. For the maximum part size $\mathcal{M}_m$ we first compute

$$\tilde{\pi}_1(j) := \sum_{u \ge j} \pi_1(u) \sim \frac{\tilde{P}(j) z^{*j}}{1-z^*}, j \to \infty.$$

Next we have (j will only be used in the neighbourhood of $\log m + \mathcal{O}(1)$)

$$\mathbb{P}(\mathcal{M}_m \le j-1) \sim (1-\tilde{\pi}_1(j))^m \sim e^{-m\tilde{\pi}_1(j)} \left(1 - (m\tilde{\pi}_1(j))^2 \frac{1}{2m}\right).$$

Now, we proceed as in [19]. Set

$$\eta = j + 1 - \log m - (r-1)\log\log m + \log(r-1)! + \log(1-z^*).$$

Then, with integer j and $\eta = \mathcal{O}(1)$, the distribution is asymptotically given by the extreme-value (or Gumbel) distribution function $F(x) := e^{-e^{-x}}$: we obtain

$$\mathbb{P}[\mathcal{M}_m \le j] \sim e^{-e^{-L\eta}}. \tag{3.7}$$

The mean of this distribution is given by $\frac{\gamma}{L}$. From this and (3.7) we deduce, as in [19], that

$$\mathbb{E}(\mathcal{M}_m) \sim \log m + (r-1)\log\log m - 1 - \log(r-1)! - \log(1-z^*) + \frac{1}{2} + \frac{\gamma}{L} + \beta(\log m).$$

We have here another example of Gumbel-like distribution. The rate of convergence for this kind of distributions is fully analyzed in [22]; we will not give the details here. Let $P_n(k) := \sum_{i=0}^{k} \mathbb{P}[\mathcal{M}_m = i]$. We have $P_n(k) \sim F(\eta)$. Set

$$\nu = \frac{m(\log m)^{r-1} z^*}{(r-1)!(1-z^*)}.$$

Let m_i be the ith moment of $F(\eta)$, $\tilde{m}_1$ be the dominant part of the ith moment of $\mathcal{M}_m - \log\nu$, w_i be the corresponding periodic part, μ_i be the ith centered moment of $F(\eta)$, $\tilde{\mu}_1$ be the dominant part of the ith centered moment of $\mathcal{M}_m$, κ_i be the corresponding periodic part. We have, from (3.7),

$$\varphi(\alpha) = \int_{-\infty}^{\infty} F'(\eta)e^{\alpha\eta}d\eta = \Gamma(1-\tilde{\alpha}), \text{ with } \tilde{\alpha} := \alpha/L, \Re(\alpha) < L. \tag{3.8}$$

Now, we can compute (almost) mechanically all moments of $\mathcal{M}_m - \log\nu$ we need, from $\varphi(\alpha)$, using techniques and notations from [22]. This leads to

$$\begin{aligned}
m_1 &= \gamma/L,\\
m_2 &= (\pi^2/6+\gamma^2)/L^2,\\
m_3 &= (2\zeta(3)+\pi^2\gamma/2+\gamma^3)/L^3,\\
\tilde{m}_1 &= \gamma/L+1/2,\\
\tilde{m}_2 &= \gamma/L+1/3+(\pi^2/6+\gamma^2)/L^2,\\
\tilde{m}_3 &= 1/4+(2\zeta(3)+\pi^2\gamma/2+\gamma^3)/L^3+\gamma/L+(\pi^2/2+3\gamma^2)/(2L^2),\\
\sigma^2 &= \pi^2/(6L^2),\\
\mu_3 &= 2\zeta(3)/L^3,\\
\mu_4 &= 3\pi^4/(20L^4),\\
\tilde{\mu}_2 &= \pi^2/(6L^2)+1/12,\\
\tilde{\mu}_3 &= 2\zeta(3)/L^3.
\end{aligned}$$

Let us now turn to the fluctuating components, which are periodic functions of $\log\nu$. Note that, in contrast to the classical composition ($r=1$), the functions are periodic in $\log m + (r-1)\log\log m$. We derive

$$w_1 = -\sum_{l\neq 0}\Gamma(\chi_l)e^{-2l\pi i\log\nu}/L. \tag{3.9}$$

$$\kappa_2 = -2\gamma w_1/L - w_1^2 + 2\sum_{l\neq 0}\Gamma(\chi_l)\psi(\chi_l)e^{-2l\pi i\log\nu}/L^2,$$

where ψ is the digamma function. Next we obtain

$$\begin{aligned}
\kappa_3 \;=\; & -6(w_1+\gamma/L)\sum_{l\neq 0}\Gamma(\chi_l)\psi(\chi_l)e^{-2l\pi i\log\nu}/L^2 - 3\sum_{l\neq 0}\Gamma(\chi_l)\psi^2(\chi_l)e^{-2l\pi i\log\nu}/L^3\\
& -3\sum_{l\neq 0}\Gamma(\chi_l)\psi(1,\chi_l)e^{-2l\pi i\log\nu}/L^3 + 2w_1^3 + (3\gamma^2-\pi^2/2)w_1/L^2 + 6\gamma w_1^2/L,
\end{aligned}$$

where $\psi(1,x)$ is the trigamma function.

3.3 First empty part value

Another variable of interest, $\mathcal{E}_m$, is the first k such that $I_k = 0$, i.e. we are interested in the probability

$$\mathbb{P}(\mathcal{E}_m = k) = \mathbb{P}(I_i = 1, i = 1 \cdots k-1, I_k = 0).$$

This probability is asymptotically given by

$$\mathbb{P}(\mathcal{E}_m = k) \sim \prod_{i=1}^{k-1} (1 - e^{-m\pi_1(i)}) e^{-m\pi_1(k)}.$$

We set $\eta = k - \log m - (r-1)\log\log m + \log(r-1)!$. This equation leads asymptotically to

$$\varphi(\eta) = e^{-e^{-L\eta}} \prod_1^\infty [1 - e^{-e^{-L(\eta-i)}}]. \tag{3.10}$$

We see that $\mathcal{E}_m$ must have a very concentrated distribution.

3.4 Asymptotic distribution of $\mathcal{D}_m$

Now we turn to the explicit form of $f(x)$ such that $\mathbb{P}(\mathcal{D}_m = k) \sim f(\eta)$ for suitable η. Proceeding as in [10] and setting $k = \log m + (r-1)\log\log m - \log(r-1)! + \eta$ and $r_j = \log m + (r-1)\log\log m - \log(r-1)! + \eta + w_j$, we obtain the following result:

Theorem 11. *With k integer and $\eta = \mathcal{O}(1)$,*

$$\mathbb{P}(\mathcal{D}_m = k) \sim f(\eta) = \sum_{u=0}^\infty \varphi(\eta - u + 1) e^{-e^{-L(\eta+2-u+\log(1-z^*))}} \sum_{2-u \le w_1 < \ldots < w_u} \prod_{i=1}^u \frac{1 - e^{-e^{-L(\eta+w_i)}}}{e^{-e^{-L(\eta+w_i)}}}.$$

$\mathbb{P}(\mathcal{D}_m \le k) \sim F(\eta)$, with $F(\eta) := \sum_0^\infty f(\eta - i)$.

4 Probabilistic analysis of Carlitz Compositions

We remind that, in a Carlitz composition, two successive parts must be different In this section, we first analyze the number M of parts. Then we obtain the distribution of the last part size LP. Next we compute the correlation between two successive intermediate parts, leading to a Markov chain. Finally, we consider the associated processes and limiting trajectories. In this section we consider explicitly each column: $\infty := [i_1, \ldots, i_r]$, with size $i := \sum_1^r i_j$ and similar meaning for $\boldsymbol{\ell}, \boldsymbol{\theta}$. The weight $P(k)$ will play a crucial role.

4.1 Number of parts

Let $T(m,n)$ be the number of Carlitz Compositions (CC) of n with m parts and let $h_m(\infty, n)$ be the number of CC with same characteristics and last part ∞. We shall mark n by z, m by w and ∞ by $\boldsymbol{\theta}$. Let

$$f(\boldsymbol{\theta}, z) := \frac{1}{\prod_1^r (1 - \theta_i z)} - 1$$

be the GF of $h_1(\infty, u)$. The GF g_j of h_j satisfies

$$g_j(\boldsymbol{\theta}, z) = g_{j-1}(1, z) f(\boldsymbol{\theta}, z) - \sum_{\boldsymbol{\ell}} \sum_{k=1} h_{j-1}(\boldsymbol{\ell}, k) z^k \theta_1^{l_1} \ldots \theta_r^{l_r} z^l,$$

and $g_1(\boldsymbol{\theta}, z) = f(\boldsymbol{\theta}, z)$. (The first part can be of any size). Hence $\phi(w, \boldsymbol{\theta}, z) := \sum_1^\infty w^j g_j(\boldsymbol{\theta}, z)$ satisfies

$$\phi(w, \boldsymbol{\theta}, z) = w f(\boldsymbol{\theta}, z) + w\phi(w, \mathbf{1}, z) f(\boldsymbol{\theta}, z) - w\phi(w, z\boldsymbol{\theta}, z).$$

Iterating, this leads to

$$\phi(w,\boldsymbol{\theta},z) = A_1(w,\boldsymbol{\theta},z)[1+D_1(w,z)], \tag{4.1}$$

where

$$\begin{aligned} D_1(w,z) := \phi(w,\mathbf{1},z) &= A_1(w,\mathbf{1},z)/[1-A_1(w,\mathbf{1},z)], \\ A_1(w,\boldsymbol{\theta},z) &= \sum_{j=1}^{\infty}(-1)^{j+1}w^j f(\boldsymbol{\theta},z^j). \end{aligned} \tag{4.2}$$

$A_1(1,\mathbf{1},z)$ was already obtained in [24].

Notice that

$$1+D_1(w,z) = 1/h(w,z), \tag{4.3}$$

where

$$h(w,z) := 1 - A_1(w,\mathbf{1},z). \tag{4.4}$$

$A_1(w,\mathbf{1},z)$ can be simplified as follows

$$\begin{aligned} A_1(w,\mathbf{1},z) &= \sum_{j=1}^{\infty}(-1)^{j+1}w^j\left[\frac{1}{(1-z^j)^r}-1\right] \\ &= \sum_{j=1}^{\infty}(-1)^{j+1}w^j\sum_{k=0}^{\infty}\binom{k+r-1}{r-1}z^{jk} - \frac{w}{1+w} \\ &= \sum_{k=0}^{\infty}P(k)\frac{wz^k}{1+wz^k} - \frac{w}{1+w} \\ &= \sum_{k=1}^{\infty}P(k)\frac{wz^k}{1+wz^k}. \end{aligned}$$

This is also more convenient for numerical purposes.

First set $\boldsymbol{\theta} = \mathbf{1}$ in $A_1(w,\boldsymbol{\theta},z)$. When $w=1$ we get the GF of the total number $T(\cdot,n)$ of CC (n) (CC of n): this is given by $D_1(1,z)$. The dominant singularity of $D_1(1,z)$ is given by the root z^* (with smallest module) of $h(1,z)$, so

$$A_1(1,\mathbf{1},z^*) = \sum_{k=1}^{\infty}P(k)\frac{z^{*k}}{1+z^{*k}} = 1. \tag{4.5}$$

$h(1,z)$ is analytic for $|z|<1$. To be sure that z^* is the dominant singularity, we use the principle of the argument (see, for instance, Henrici [9]): the number of solutions of the equation $f(z)=0$ that lie inside a simple closed curve Γ, with $f(z)$ analytic inside and on Γ, is equal to the variation of the argument of $f(z)$ along Γ, a quantity also equal to the winding number of the transformed curve $f(\Gamma)$ around the origin. The argument was used in Flajolet and Prodinger [7] in a similar situation. Figure 2 represents $[\Re(h),\Im(h)]$ for $r=3, z=0.3\exp(it), t=0..2\pi$, where $h(1,z)$ is computed with fifty terms. The root is given by $z^* = 0.2238153681\ldots$.

The winding number is 1, so that h has only one root z^* for $|z|<0.3$. A similar analysis can be done for all r.

By singularity analysis, $T(\cdot,n)$ is asymptotically given by

$$T(\cdot,n) \sim -\frac{1}{z^*h_z(1,z^*)}\frac{1}{z^{*n}} = C_1/z^{*n}, n\to\infty, \tag{4.6}$$

with

$$C_1 = \frac{-1}{z^*h_z(1,z^*)}.$$

For $w\in[0,1]$, h is analytic for $|z|<1$ and for $w\in[1,\infty]$, h is analytic for $|z|<1/w$.

Theorems 4 and 5 are still valid, with suitable parameters $\mu_i,\sigma_i, i=1,2$.

Two useful relations can be derived:

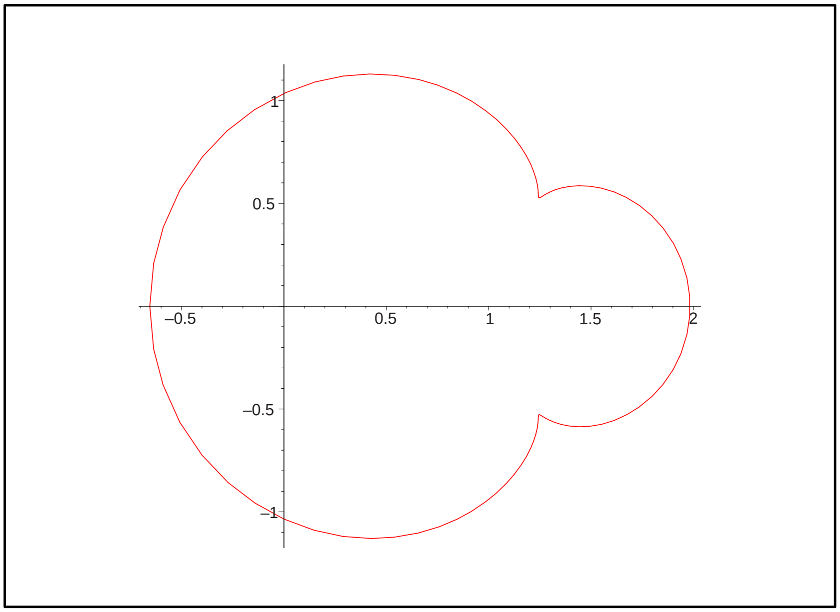

Figure 2: Winding number of h, $r = 3$.

$$h_w(1, z^*) = -\sum_{j=1}^{\infty} \frac{P(j) z^{*j}}{(1+z^{*j})^2},$$

$$z^* h_z(1, z^*) = -\sum_{j=1}^{\infty} \frac{P(j) j z^{*j}}{(1+z^{*j})^2}.$$

For further use, we have also computed $C_2(\mathbf{j})$ based on a first part $\mathbf{j}$. We first derive (conditioning on $\mathbf{j}$):

$$\phi(w, \boldsymbol{\theta}, z|\mathbf{j}) = A_2(w, \boldsymbol{\theta}, z|\mathbf{j}) + A_1(w, \boldsymbol{\theta}, z) D_2(w, z|\mathbf{j}), \tag{4.7}$$

where

$$\begin{aligned} D_2(w, z|\mathbf{j}) &:= \phi(w, \mathbf{1}, z|\mathbf{j}) = A_2(w, \mathbf{1}, z|\mathbf{j})/h(w, z), \\ A_2(w, \boldsymbol{\theta}, z|\mathbf{j}) &= \theta_1^{j_1} \dots \theta_r^{j_r} w z^j/(1 + w z^j), \end{aligned}$$

and $h(w, z)$ is given by (4.4).

This leads to

$$C_2(\mathbf{j}) = -\frac{A_2(1, \mathbf{1}, z^*|\mathbf{j})}{h_w(1, z^*)} = \frac{-z^{*j}}{1+z^{*j}} \frac{1}{h_w(1, z^*)}. \tag{4.8}$$

Note that $C_2(\mathbf{j})$ depends only on the size j, hence it will be denoted by $C_2(j)$. Note also that $\sum_j C_2(j) P(j) = C_2$, as it should.

4.2 Distribution of the last part size in CC

Following the lines of Section 2.2, we have

$$G(\boldsymbol{\theta}) = A_1(1, \boldsymbol{\theta}, z^*) = \sum_{j=1}^{\infty} (-1)^{j+1} f(\boldsymbol{\theta}, z^{*j}),$$

so

$$\pi_1(\boldsymbol{\ell}) := [\theta_1^{l_1} \dots [\theta_r^{l_r}] G(\boldsymbol{\theta}) = \sum_{j=1}^{\infty} (-1)^{j+1} z^{*lj} = \frac{z^{*l}}{1+z^{*l}}.$$

We note that $\pi_1(\boldsymbol{\ell})$ depends only on the size l. We will hence use the notation $\pi_1(\ell)$. Of course

$$\sum_{\ell=1}^{\infty} P(\ell)\pi_1(\ell) = 1.$$

4.3 Correlation between 2 successive intermediate parts

Let us now turn to two successive parts m_1, m_1+1, of columns values $\mathbf{k}, \mathbf{j}$, such that their distances from the first and the last part are of order $\mathcal{O}(n)$. Let $T(m,n,\mathbf{k})$ be the total number of $CC(n)$ with m parts and last part $\mathbf{k}$, let $\#(m,n,m_1,\mathbf{k},\mathbf{j})$ be the total number of $CC(n)$ with m parts, part m_1 of value $\mathbf{k}$, part m_1+1 of value $\mathbf{j}$, and we set $m_2 := m - m_1$. We have, conditioning on $\mathbf{j}$,

$$\#(m,n,m_1,\mathbf{k},\mathbf{j}) = \sum_{n_1} T(m_1,n_1,\mathbf{k})T(m_2,n-n_1,\cdot,|\mathbf{j}),$$

where $T(m,n,\cdot)$ denotes the the total number of $CC(n)$ with m parts. With the generalization of Theorem 6 (vertical form) and Theorem 5, with $C_2(\mathbf{j})$, this leads, after normalization by $T(m,n)$, to the following distribution, with $\mathbf{k} \neq \mathbf{j}$:

$$\begin{aligned}
&\frac{\#(m,n,m_1,\mathbf{k},\mathbf{j})}{T(m,n)} \\
&\sim \sum_{n_1} \frac{e^{-(n_1-m_1\mu_2)^2/(2m_1\sigma_2^2)}}{\sqrt{2\pi m_1}\sigma_2}\pi_1(\mathbf{k})\cdot \\
&\cdot C_2(\mathbf{j})\frac{e^{-(n-n_1-m_2\mu_2)^2/(2m_2\sigma_2^2)}}{\sqrt{2\pi m_2}\sigma_2} \Bigg/ \frac{e^{-(n-m\mu_2)/(2m\sigma_2^2)}}{\sqrt{2\pi m}\sigma_2} \\
&\sim \pi_1(\mathbf{k})C_2(\mathbf{j}).
\end{aligned} \tag{4.9}$$

Now we readily obtain the following distribution $\pi_2(\mathbf{k})$ for $\mathbf{k}$:

$$\begin{aligned}
\pi_2(\mathbf{k}) &= \sum_{\mathbf{j}\neq\mathbf{k}} \pi_1(\mathbf{k})C_2(\mathbf{j}) = \frac{z^{*k}}{1+z^{*k}}\frac{-1}{h_w(1,z^*)}\left[\sum_{j\neq k}\frac{z^{*j}}{1+z^{*j}}P(j) + \frac{z^{*k}}{1+z^{*k}}[P(k)-1]\right] \\
&= \frac{-1}{h_w(1,z^*)}\frac{z^{*k}}{(1+z^{*k})^2},
\end{aligned} \tag{4.10}$$

which depends only on the size k.

For further use, we compute, for $\ell \geq k$

$$\pi_2(\ell) = \pi_2^{(1)}(\ell)z^{*k} + \pi_2^{(2)}(\ell)z^{*2k} + \mathcal{O}(z^{*3k}), k \to \infty,$$

with

$$\pi_2^{(1)}(\ell) = -\frac{z^{*(l-k)}}{h_w(1,z^*)}, \pi_2^{(2)}(\ell) = \frac{2z^{*2(l-k)}}{h_w(1,z^*)}. \tag{4.11}$$

Note that, in further summations, $\pi_2(\ell)$ will be weighted by $P(\ell)$ and, for $\ell \geq k, k \gg 1$

$$P(\ell) \sim \tilde{P}(k)[1 + \mathcal{O}((\ell-k)/k)]. \tag{4.12}$$

We obtain $\sum_{k=1}^{\infty} P(k)\pi_2(k) = 1$ and $\sum_{k=1}^{\infty} kP(k)\pi_2(k) = \mu_2$ as it should. So we obtain the following theorem.

Theorem 12. *The asymptotic ($n \to \infty$) distribution of intermediate part size is given by $\pi_2(k)$, with mean μ_2.*

But we can get more from (4.9), which shows that the asymptotic joint distribution (in stationary distribution) of two successive intermediate part **k** and **j** is given by

$$\pi_1(k)C_2(j), \mathbf{j} \neq \mathbf{k}.$$

Normalizing by $\pi_2(k)$, this leads to the following Markov chain related to two successive intermediate parts:

$$\Pi(\mathbf{k}, \mathbf{j}) = \frac{z^{*j}(1+z^{*k})}{1+z^{*j}}, \mathbf{j} \neq \mathbf{k}. \tag{4.13}$$

Note that k may be equal to j. The chain is irreducible, recurrent positive and reversible. Of course the stationary distribution of Π is given by $\pi_2(k)$. We will use the notation $\Pi(k,j)$. Therefore, we obtain the following theorem

Theorem 13. *The asymptotic, $n \to \infty$, distribution of two successive intermediate parts in a CC (n) is given by a Markov chain $\Pi(i,j)$, with stationary distribution $\pi_2(k)$, and with mean μ_2.*

Let us now analyze the asymptotic behaviour of $\Pi(i,j)$. This will be used in Section 5.2. For fixed ℓ, we obtain

$$\Pi^+(\ell,k) := \sum_{j=k}^{\infty} \Pi(\ell,j)P(j) = (1+z^{*\ell})\frac{z^{*k}}{1-z^*}\tilde{P}(k)(1+\mathcal{O}(z^{*k}))(1+\mathcal{O}(1/k)), k \to \infty.$$

For further use, we set

$$\varphi_3(\ell) := \frac{1+z^{*\ell}}{1-z^*}. \tag{4.14}$$

4.4 Associated processes and limiting trajectories

Let us now turn to trajectories. Let x_i be the size of part i and set $X(j) := \sum_1^j x_i$. We have $E[X(m)] \sim m\mu_2$ and we must check that $\mathsf{var}[X(m)] \sim m\sigma_2^2$ as given by (2.4) with h given by (4.4). We first derive $m\sigma_X^2 := \mathsf{var}[X(m)] \sim m[S_2 - \mu_2^2] + 2m\sum_{k=1}^{\infty} C_k^x$ where

$$\begin{aligned} S_2 &:= \sum_{j=1}^{\infty} P(j)\pi_2(j)j^2, \\ C_k^x &:= \sum_{i\neq j} P(i)P(j)(i-\mu_2)\pi_2(i)\Pi^k(i,j)(j-\mu_2). \end{aligned}$$

The proof follows the one given in [21], with of course the introduction of $P(k)$ in suitable places. Theorems 8 and 9 are still valid.

5 Simulations and thickness

In this section, we present an algorithm for CC(n) simulation, then we analyze the CC thickness maximum.

5.1 Realizations

A simulated realization of a CC(n) proceeds as follows. Start with a part of size j at time $i=1$ given by $\pi_1(j)$. Proceed from a part of size k at time $i-1$ to a part of size j at time i by using a Markov chain defined by the probability matrix $\Pi(k,j)$ given by (4.13). Stop the chain as soon as $S := \sum_{i=1}^{m} j_i > n$. We choose each column of size l with equidistribution $1/P(l)$.

5.2 Hitting times and maximum for CC thickness

Let us call the part size “thickness". In this section, we derive the thickness hitting times (to high level) asymptotic distribution. This leads to an asymptotic density for the largest part. We will follow the lines of [21], but with some modifications essentially due to the presence of $P(k)$ in several expressions.

We consider the set x_i of RV describing the thickness of CC. Let us define the set $D := [k \cdot \cdot\infty), k \gg 1$. By (4.14), we see that the probability transition to D is $\mathcal{O}(\epsilon), \epsilon = z^{*k}$ and by standard properties (see Keilson [15], Aldous [1]) we know that the hitting time to D is such that (we drop k for ease of notation):

- $E_\ell[T_D] = \frac{C_7(k)}{\epsilon} + \psi(\ell) + \mathcal{O}(\epsilon)$ (Actually a Laurent series exists for ϵ sufficiently small),
- $\mathbb{P}_\ell[T_D \geq x] \sim e^{-x/E_\ell[T_D]}, x \to \infty$.

We should write $C_7(k)(\ell)$ but we will soon check that $C_7(k)$ is independent of ℓ. To compute $C_7(k)$ and $\psi(\ell)$, we use the classical relation:

$$
\begin{aligned}
E_\ell[T_D] &= 1 + \sum_{j \in D^C} \Pi[\ell, j] E_j[T_D] \text{ i.e.} && (5.1)\\
C_7(k) + \psi(\ell)\epsilon &= \epsilon + \sum_{j \in D^C} \Pi[\ell, j][C_7(k) + \psi(j)\epsilon] + \mathcal{O}(\epsilon^2) \\
&= \epsilon + \Pi C_7(k) - \epsilon\varphi_3(\ell)\tilde{P}(k)C_7(k) + \sum_{j=1}^{\infty} \Pi(\ell, j)\psi(j)\epsilon + \mathcal{O}(\epsilon^2), && (5.2)
\end{aligned}
$$

and φ_3 is given by (4.14). Refering to (4.12), we will neglect in the sequel the correction term $(\ell - k)/k$ and keep only the dominant terms.

Equation (5.1) leads to

$$1 = \sum_{\ell \in D} \pi_2(\ell) E_\ell(T_D).$$

(This is equivalent to a formula of Kac, see [14].)

Therefore, we obtain

$$1 = C_7(k)\tilde{P}(k) \sum_{\ell \in D} \pi_2^{(1)}(\ell). \tag{5.3}$$

The ε term leads to

$$0 = \tilde{P}(k) \sum_{\ell \in D} \pi_2^{(1)}(\ell)\psi(\ell) + C_7(k)\tilde{P}(k) \sum_{\ell \in D} \pi_2^{(2)}(\ell). \tag{5.4}$$

$\pi_2^{(1)}$ and $\pi_2^{(2)}$ are given by (4.11). This shows that $\psi(\ell) = \mathcal{O}(1/\tilde{P}(k))$ for $\ell \geq k$. Set tentatively $\psi(\ell) = \psi_1(\ell)/\tilde{P}(k)$.

Comparison of powers of ϵ in (5.2) leads to $[I - \Pi]C_7(k) = 0$, which confirms that $C_7(k)$ is independent of ℓ. We derive

$$\psi(\ell) = 1 - C_7(k)\tilde{P}(k)\varphi_3(\ell) + \sum_{j=1}^{\infty} \Pi(\ell, j)\psi(j),$$

or

$$[I - \Pi]\psi = \delta, \quad \text{with } \delta(\ell) := 1 - C_7(k)\tilde{P}(k)\varphi_3(\ell). \tag{5.5}$$

We must have $\pi_2 \delta = 0$, therefore

$$1 = C_7(k)\tilde{P}(k) \sum_{\ell=1}^{\infty} \pi_2(\ell)\varphi_3(\ell), C_7(k) = 1 \Bigg/ \left[\tilde{P}(k) \sum_{\ell=1}^{\infty} \pi_2(\ell)\varphi_3(\ell)\right],$$

which fixes

$$C_7(k) = -h_w(1 - z^*)/\tilde{P}(k).$$

(This is of course equivalent to (5.3)).

We denote by M_1 the Drazin inverse of $I-\Pi$. We refer to Campbell and Meyer [4] for a detailed definition and analysis of the Drazin inverse. We have

$$M_1 = \sum_{n\geq 0}(\Pi^n - 1\times\pi_2) = M_2 - 1\times\pi_2, \tag{5.6}$$

where $M_2 := [I-\Pi+1\times\pi_2]^{-1} = \sum_{n\geq 0}[\Pi-1\times\pi_2]^n$ is the potential used in Kemeny, Snell and Knapp [16]. Notice that $M_1\mathbf{1}=0$. Two expressions related to M_1 are proved in [10]:

$$\begin{aligned} M(j,k) &= -\frac{z^{*j}}{(1+z^{*j})}[\![j=k]\!] - \frac{z^{*j}z^{*k}}{(1+z^{*j})(1+z^{*k})^2 h_w} + \frac{z^{*k}}{(1+z^{*k})^3 h_w} \\ &\quad + C_8\frac{z^{*k}}{(1+z^{*k})^2} + [\![j=k]\!], \\ &= \frac{z^{*k}}{(1+z^{*k})^2}\left[-\frac{C_2}{1+z^{*k}} + 1 + C_9\right] + \mathcal{O}(z^{*k}z^{*j}), k=\mathcal{O}(1), j\gg 1, \end{aligned}$$

with

$$C_8 := -\frac{1}{h_w^2}\sum_{k=1}^{\infty}\frac{z^{*2k}}{(1+z^{*k)3}}, C_9 = C_8 - 1, h_w := h_w(1,z^*).$$

The solution of (5.5) is given by

$$\psi = M_1\delta + \pi_2\psi.$$

For $\ell\geq k$, we can compute the dominant term C_{10} of $[M_1\delta](\ell)$. This gives

$$[M_1\delta](\ell) \sim \frac{1}{1-z^*}\left[\sum_{u=1}^{\infty}\frac{z^{*u}}{(1+z^{*u})^2}P(u)(-C_2) + (1+C_9)\sum_{u=1}^{\infty}\frac{z^{*u}}{1+z^{*u}}P(u)\right][-C_7(k)\tilde{P}(k)],$$

so

$$C_{10} = -C_9 C_7(k)\tilde{P}(k)/(1-z^*).$$

To fix $C_{11} := \pi_2\psi$, we derive for $\ell\geq k$

$$C_{10} = -\pi_2\psi + \psi_1(\ell)/\tilde{P}(k),$$

so $C_{11} = -C_{10} + \psi_1(\ell)/\tilde{P}(k)$, which shows that ψ_1 is asymptotically independent of ℓ. To fix ψ_1, we first derive

$$\sum_{\ell\in D}\pi_2^{(1)}(\ell)\psi(\ell) = \sum_{\ell\in D}\pi_2^{(1)}(\ell)[M_1\delta](\ell) + C_{11}\sum_{\ell\in D}\pi_2^{(1)}(\ell),$$

and with (5.4),

$$C_{11} = \left[h_w(1-z^*)\frac{2}{h_w(1-z^{*2})} - \tilde{P}(k)\sum_{\ell\in D}\pi_2^{(1)}(\ell)[M_1\delta](\ell)\right]C_7(k),$$

or

$$C_{11} = \left[h_w(1-z^*)\frac{2}{h_w(1-z^{*2})} - C_{10}/C_7(k)\right]C_7(k),$$

and

$$\psi_1 = \frac{-2h_w(1-z^*)^2}{1-z^{*2}}.$$

We can summarize our results in the following form

Theorem 14. *The thickness hitting time $E_\ell[T_D]$ to $D := [k \cdot \cdot \infty], k \gg 1$ is given by*

$$E_\ell[T_D] \sim \frac{C_7(k)}{z^{*k}} + \psi(\ell) + \mathcal{O}(z^{*k}),$$

with $C_7(k) = -h_w(1-z^*)/\tilde{P}(k), \psi(\ell) = -[M_1 C_7(k)\tilde{P}(k)\varphi_3](\ell) + C_{11}$ (M_1 is the potential kernel given in (5.6)).

$$Pr_\ell[T_D \geq x] \sim e^{-xz^{*k}/C_7(k)}, x \to \infty.$$

Now we can proceed as in Section 3.2 for $\mathcal{M}_m$, we omit the details.

To analyze $\mathcal{E}_m$ and $\mathcal{D}_m$, we need an asymptotic independence of all RV I_i. We conjecture that it is true, but, even in the case of classical Carlitz compositions (see [10]), it was only proved an asymptotic independence of two by two RV I_i. If the conjecture is correct, the previous analysis is completely transposable.

6 Conclusion

In this paper, we have presented a stochastic analysis of several RV and processes related to matrix compositions. Many other research topics are possible like the number of distinct parts size of some multiplicity, as analyzed, for instance in [20] and [23]. Also other types of dependence are possible, like Carlitz compositions including 0 parts. We could have constraints on each column, leading to different forms for $f(z)$ and $f(\boldsymbol{\theta}, z)$. See [24] and [5], where many examples are given.

7 Acknowledgements.

The pertinent comments of the referee led to improvements in the presentation.

References

[1] D. Aldous. *Probability Approximations via the Poisson Clumping Heuristic.* Springer-Verlag, 1989.

[2] E.A. Bender. Central and local limit theorems applied to asymptotics enumeration. *Journal of Combinatorial Theory,* Series A, 15:91–111, 1973.

[3] P. Billingsley. *Convergence of Probability Measures.* Wiley, 1968.

[4] S. L. Campbell and C. D. Meyer. *Generalized Inverse of Linear Transformations.* Pitman, 1979.

[5] G. Castiglione, A. Frosini, E. Munarini, A. Restivo, and S. Rinaldi. Combinatorial aspects of L-convex polyominoes. Technical report, 2006.

[6] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[7] P. Flajolet and H. Prodinger. Level number sequences for trees. *Discrete Mathematics*, 65:149–156, 1987.

[8] P. Flajolet and M. Soria. General combinatorial schemes: Gaussian limit distribution and exponential tails. *Discrete Mathematics*, 114:159–180, 1993.

[9] P. Henrici. *Applied and Computational Complex Analysis.* Wiley, 1988.

[10] P. Hitczenko and G. Louchard. Distinctness of compositions of an integer: a probabilistic analysis. *Random Structures and Algorithms*, 19(3,4):407–437, 2001.

[11] H.K. Hwang. Théorèmes limites pour les structures aléatoires et les fonctions arithmétiques. 1994. Thèse, Ecole Polytechnique de Palaiseau.

[12] H.K. Hwang. On convergence rates in the central limit theorems for combinatorial structures. *European Journal of Combinatorics*, 19:329–343, 1998.

[13] K. Ito and H.P. McKean, Jr. *Diffusion Processes and their Sample Paths*. Springer, 1974.

[14] M. Kac. On the notion of recurrence in discrete stochastic processes. *Bulletin of the American Mathematical Society*, 53:1002–1019.

[15] J. Keilson. *Markov Chain Models-Rarity and Exponentiality*. Springer-Verlag, 1979.

[16] J. G. Kemeny, J. L. Snell, and A. W. Knapp. *Denumerable Markov Chains*. Van Nostrand, 1966.

[17] M. Loève. *Probability Theory*. D. Van Nostrand, 1963.

[18] G. Louchard. Probabilistic analysis of some (un)directed animals. *Theoretical Computer Science*, 159(1):65–79, 1996.

[19] G. Louchard. Probabilistic analysis of column-convex and directed diagonally-convex animals. II: Trajectories and shapes. *Random Structures and Algorithms*, 15:1–23, 1999.

[20] G. Louchard. The number of distinct part sizes of some multiplicity in compositions of an integer. a probabilistic analysis. *Discrete Mathematics and Theoretical Computer Science*, AC:155–170, 2003.

[21] G. Louchard and H. Prodinger. Probabilistic analysis of Carlitz compositions. *Discrete Mathematics and Theoretical Computer Science*, 5(1):71–96, 2002.

[22] 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.

[23] 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.

[24] E. Munarini, M. Poneti, and S. Rinaldi. Matrix compositions. In Formal Power series and Algebraic Combinatorics 2006, San Diego, 2006.

[25] W. Szpankowski, Average Case Analysis of Algorithms on Sequences. Wiley, New York, 2001.

[26] E.T. Whittaker, G.N. Watson. *A Course of Modern Analysis*. Cambridge University press (reprinted 1999), 1927.