Periodic Pólya urns
and an application to Young tableaux

Cyril Banderier[1][*]   Philippe Marchal[2][†]   Michael Wallner[3][‡]

February 7, 2018

[1],[3]: LIPN UMR-CNRS 7030, Université de Paris Nord, France.
[2]: LAGA, Institut Galilée, Université de Paris Nord, France.

Abstract

Pólya urns are urns where at each unit of time a ball is drawn and is replaced with some other balls according to its colour. We introduce a more general model: the replacement rule depends on the colour of the drawn ball *and* the value of the time $(\operatorname{mod} p)$. We exhibit a closed form for the corresponding generating functions (giving the exact composition of the urns at time n). We discuss some intriguing properties of the differential operators associated to the generating functions encoding the evolution of these urns: The initial *non-linear* partial differential equation indeed leads to *linear* differential equations and we prove that the moment generating functions are D-finite. When the time goes to infinity, we show that these *periodic Pólya urns* follow a rich variety of behaviours: their asymptotics fluctuations are described by a family of distributions, the generalized Gamma distributions, which can also be seen as powers of Gamma distributions. En passant, we establish some enumerative links with other combinatorial objects, and we give an application for a new result on the asympotics of Young tableaux: This approach allows us to prove that the law of the lower right corner in a triangular Young tableau follows asymptotically a product of generalized Gamma distributions.

Keywords: Pólya urn, Young tableau, analytic combinatorics, pumping moment, D-finite function, generalized Gamma distribution, Mittag-Leffler distribution

[*]http://lipn.fr/~banderier
[†]https://www.math.univ-paris13.fr/~marchal/
[‡]http://dmg.tuwien.ac.at/mwallner/

1   Periodic Pólya urns

Pólya urns were introduced in a simplified version by George Pólya and his PhD student Florian Eggenberger in [5, 6], with applications to disease spreading and conflagrations. They are still a widely used and powerful model, see e.g. Rivest's recent work on auditing elections [24], or the analysis of deanomization in Bitcoin's peer-to-peer network [7]. They are well-studied objects in combinatorial and probabilistic literature [1, 9, 19], and offer fascinatingly rich links with numerous objects like random recursive trees, m-ary search trees, branching random walks (see e.g. [2, 4, 13, 26]). In this paper we introduce a variation which offers new links with another important combinatorial structure: Young tableaux. We solve the enumeration aspects of this new model, derive the limit law for the evolution of the urn, and give some applications.

In the *Pólya urn model*, one starts with an urn with b_0 black balls and w_0 white balls at time 0. At every discrete time step one ball is drawn uniformly at random. After inspecting its colour it is returned to the urn. If the ball was black, a black balls and b white balls are added; if the ball was white, c black balls and d white balls are added (where $a, b, c, d \in \mathbb{N}$ are non-negative integers). This process can be described by the so-called *replacement matrix*:

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad a, b, c, d \in \mathbb{N}.$$

We call an urn and its associated replacement matrix *balanced* if $K := a + b = c + d$. In other words, in every step the same number K of balls is added to the urn. This results in a deterministic number of balls after n steps: $b_0 + w_0 + Kn$ balls.

Now, we introduce a more general model which has rich combinatorial, probabilistic, and analytic properties.

Definition 1.1. A *periodic Pólya urn* of period p with replacement matrices $M_1, M_2, \ldots, M_p$ is a variant of Pólya urn in which the replacement matrix M_k is used at steps $np + k$. Such a model is called *balanced* if each of its replacement matrices is balanced.

In this article, we illustrate the aforementioned rich properties on the following model.

Definition 1.2. We call a *Young Pólya urn* the periodic Pólya urn of period 2 with replacement matrix $M_1 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ for every odd step, and replacement matrix $M_2 := \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$ for every even step.

Let us describe the composition of the urn after n steps by pairs (b_n, w_n), starting with $b_0 = 1$ black ball and $w_0 = 1$ white ball shown in Figure 1. In the first step the matrix M_1 is used and gives the two compositions

$$(2, 1), \text{ and } (1, 2).$$

In the second step, matrix M_2 is used, in the third step, matrix M_1 is used again, in the fourth step, matrix M_2, etc. Thus, the possible compositions are

$$(3, 2), (2, 3), \text{ and } (1, 4), \text{ at time } 2, \text{ and}$$
$$(4, 2), (3, 3), (2, 4), \text{ and } (1, 5), \text{ at time } 3.$$

$$H_0 = xy$$

$$H_1 = x^2y + xy^2$$

$$H_2 = 2x^3y^2 + 2x^2y^3 + 2xy^4$$

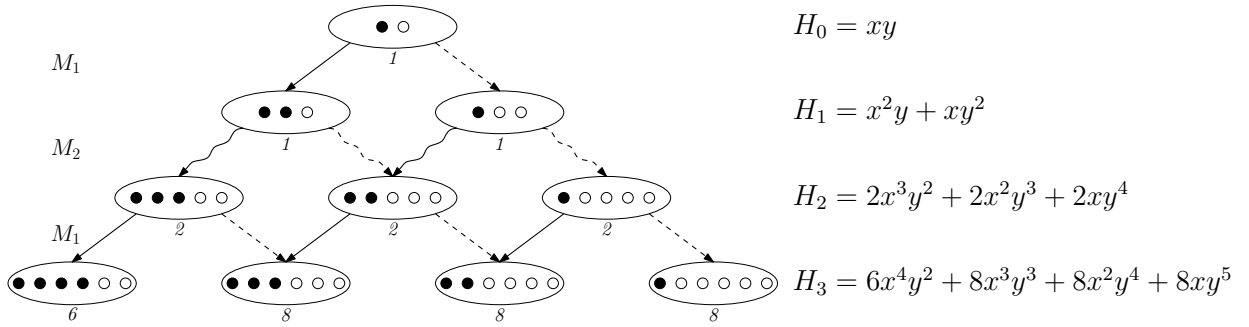$$H_3 = 6x^4y^2 + 8x^3y^3 + 8x^2y^4 + 8xy^5$$

Figure 1: The evolution of a Young Pólya urn with one initial black and one initial white ball. Black arrows mark that a black ball was drawn, dashed arrows mark that a white ball was drawn. Straight arrows indicate that the replacement matrix M_1 was used, curly arrows show that the replacement matrix M_2 was used. The number below each node is the number of possible transitions to reach such a composition. In this article we give a formula for H_n (which encodes all the possible compositions of the urn at time n) and their asymptotic behaviour.

In fact, each of these compositions is reached in different ways, and such a sequence of transitions is called a *history*. Each history comes with weight one. Implicitly, they induce a probability measure on the compositions at step n. So, let B_n and W_n be random variables for the number of black and white balls after n steps, respectively. As our model is balanced, $B_n + W_n$ is a deterministic process, reflecting the identity $b_n + w_n = b_0 + w_0 + n + \left\lfloor \frac{n}{2} \right\rfloor$. So we concentrate from now on our analysis on B_n and b_n.

For the classical model of a single balanced Pólya urn, the limit law of the random variable B_n is fully known: The possible limit laws include a rich variety of distributions. To name a few, let us mention the uniform distribution [8], the normal distribution [2], the Beta and Mittag-Leffler distributions [13]. Periodic Pólya urns (which include the classical model) lead to an even larger variety of distributions involving a product of *generalized Gamma distributions* [27].

Definition 1.3. The generalized Gamma distribution $\mathrm{GenGamma}(d, p)$ with parameters $d, p > 0$ is defined by the density function (having support $[0, +\infty]$)

$$f(x; d, p) := \frac{p \, x^{d-1} e^{-x^p}}{\Gamma\left(d/p\right)},$$

where Γ is the classical Gamma function $\Gamma(z) := \int_0^\infty t^{z-1} e^{-t} \, dt$.

Remark 1.4. Let $X \sim \Gamma(k)$ be a Gamma distributed random variable with parameter $k > 0$. Its density is equal to

$$f(x; k) = \frac{x^{k-1} e^{-x}}{\Gamma(k)}.$$

Then, for $q > 0$ the powers X^q are distributed like a generalized Gamma distribution with parameters $p = 1/q$ and $d = k/q$. In particular for $p = 1$ the generalized Gamma distribution is equal to the Gamma distribution.

Our main results are the enumeration result from Theorem 2.1, the application to Young tableaux in Theorem 4.1, and the following result (and its generalization in Theorem 3.1):

Theorem 1.5. The normalized random variable $\frac{2^{2/3}}{3}\frac{B_n}{n^{2/3}}$ of the number of black balls in a Young Pólya urn converges in law to a generalized Gamma distribution:

$$\frac{2^{2/3}}{3}\frac{B_n}{n^{2/3}} \quad \xrightarrow{\mathcal{L}} \quad \mathrm{GenGamma}\,(1,3)\,.$$

We give a proof of this results in Section 3. Let us first mention some articles where this distribution has already appeared before:

- in Janson [14], for the analysis of the area of the supremum process of the Brownian motion,
- in Peköz, Röllin, and Ross [22], as distributions of processes on walks, trees, urns, and preferential attachements in graphs (they also consider what they call a Pólya urn with immigration, which is a special case of a periodic Pólya urn),
- in Khodabin and Ahmadabadi [16] following a tradition to generalize special functions by adding parameters in order to capture several probability distributions, such as e.g. the normal, Rayleigh, and half-normal distribution, as well as the MeijerG function (see also the addendum of [14], mentioning a dozen of other generalizations of special functions).

In the next section we translate the evolution process into the language of generating functions by encoding the dynamics of this process into partial differential equations.

2   A functional equation for periodic Pólya urns

Let $h_{n,k,\ell}$ be the number of histories of a periodic Pólya urn after n steps with k black balls and ℓ white balls, with an initial composition of b_0 black balls and w_0 white balls, and with replacement matrices M_1 for the odd steps and M_2 for the even steps. We define the polynomials

$$H_n(x,y) := \sum_{k,\ell \geq 0} h_{n,k,\ell}x^k y^\ell.$$

Note that these are indeed polynomials as there are just a finite number of histories after n steps. We collect all these histories in the trivariate exponential generating function

$$H(x,y,z) := \sum_{n \geq 0} H_n(x,y)\frac{z^n}{n!}.$$

In particular, we get for the first 3 terms of $H(x,y,z)$ the expansion (compare Figure 1)

$$H(x,y,z) = xy + \left(xy^2 + x^2y\right)z + \left(2xy^4 + 2x^2y^3 + 2x^3y^2\right)\frac{z^2}{2} + \dots$$

Observe that the polynomials $H_n(x,y)$ are homogeneous due to balanced urn model.

Now it is our goal to derive a partial differential equation describing the evolution of the periodic Pólya urn model. For a comprehensive introduction into the method we refer to [8].

In order to capture the periodic behaviour we split the generating function $H(x, y, z)$ into odd and even steps. We define

$$H_e(x, y, z) := \sum_{n \geq 0} H_{2n}(x, y) \frac{z^{2n}}{(2n)!} \qquad \text{and} \qquad H_o(x, y, z) := \sum_{n \geq 0} H_{2n+1}(x, y) \frac{z^{2n+1}}{(2n + 1)!},$$

such that $H(x, y, z) = H_e(x, y, z) + H_o(x, y, z)$. Next, we associate to the replacement matrices M_1 and M_2 from Definition 1.2 the differential operators $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. We get

$$\mathcal{D}_1 := x^2 \partial_x + y^2 \partial_y \qquad \text{and} \qquad \mathcal{D}_2 := x^2 y \partial x + y^3 \partial_y,$$

where ∂_x and ∂_y are defined as the partial derivatives $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$, respectively. These model the evolution of the urn. For example, in the term $x^2 \partial_x$, the derivative ∂_x represents drawing a black ball and the multiplication with x^2 returning the black ball and an additional black ball into the urn. The other terms have analogous interpretations.

With these operators we are able to link odd and even steps with the following system

$$\begin{cases} \partial_z H_o(x, y, z) = \mathcal{D}_1 H_e(x, y, z), \\ \partial_z H_e(x, y, z) = \mathcal{D}_2 H_o(x, y, z). \end{cases} \tag{1}$$

Note that the derivative ∂_z models the evolution in time. This system of partial differential equations naturally corresponds to recurrences at the level of coefficients $h_{n,k,l}$, and *vice versa*. This philosophy is well explained in the *symbolic method* part of [10] and see also [8].

As a next step we want to eliminate the y variable in these equations. This is possible as the number of balls in each round and the number of black and white balls are connected due to the fact that we are dealing with balanced urns. First, as observed previously, one has

$$\text{number of balls after } n \text{ steps} = b_0 + w_0 + n + \left\lfloor \frac{n}{2} \right\rfloor. \tag{2}$$

Therefore, for any $x^k y^\ell z^n$ appearing in $H(x, y, z)$ with $b_0 = w_0 = 1$ we have

$$\begin{cases} k + \ell = 2 + \frac{3n}{2} & \text{if } n \text{ is even}, \\ k + \ell = 2 + \frac{3n}{2} - \frac{1}{2} & \text{if } n \text{ is odd}. \end{cases}$$

This translates directly into

$$\begin{cases} x \partial_x H_e(x, y, z) + y \partial_y H_e(x, y, z) = 2 H_e(x, y, z) + \frac{3}{2} z \partial_z H_e(x, y, z), \\ x \partial_x H_o(x, y, z) + y \partial_y H_o(x, y, z) = \frac{3}{2} H_o(x, y, z) + \frac{3}{2} z \partial_z H_o(x, y, z). \end{cases} \tag{3}$$

Finally, combining (1) and (3), we eliminate $\partial_y H_e$ and $\partial_y H_o$. After that it is legitimate to insert $y = 1$ as there appears no differentiation with respect to y anymore. As the urns are balanced, the exponents of y and x in each monomial are bound (see Equation (2)), so we are losing no information on the trivariate generating functions by setting $y = 1$. Hence, from now on we use the notation $H(x, z)$, $H_e(x, z)$, and $H_o(x, z)$ instead of $H(x, 1, z)$, $H_e(x, 1, z)$, and $H_o(x, 1, z)$, respectively. All of this leads to our first main enumeration theorem:

Theorem 2.1 (Linear differential equations and hypergeometric expressions for histories)**.** The generating functions describing the 2-periodic Young Pólya urn satisfy the following system:

$$\begin{cases} \partial_z H_e(x,z) = x(x-1)\partial_x H_o(x,z) + \dfrac{3}{2}z\partial_z H_o(x,z) + \dfrac{3}{2}H_o(x,z), \\ \partial_z H_o(x,z) = x(x-1)\partial_x H_e(x,z) + \dfrac{3}{2}z\partial_z H_e(x,z) + 2H_e(x,z). \end{cases} \tag{4}$$

Moreover, all these functions satisfy *linear* differential equations (they are *D-finite*, see e.g. [10, Appendix B.4] for more on this notion), which in return implies that H satisfies the equation $L.H(x,z) = 0$, where L is a differential operator of order 3 in ∂_z, and then one has the hypergeometric closed forms for $h_n := [z^n]H(x,z)$:

$$h_n = \begin{cases} 3^n \dfrac{\Gamma\left(\frac{n}{2}+1\right)\Gamma\left(\frac{n}{2}+\frac{2}{3}\right)}{\Gamma(2/3)} & \text{if } n \text{ is even,} \\ 3^n \dfrac{\Gamma\left(\frac{n}{2}+1/2\right)\Gamma\left(\frac{n}{2}+7/6\right)}{\Gamma(2/3)} & \text{if } n \text{ is odd.} \end{cases} \tag{5}$$

Alternatively, this sequence satisfies $h(n+2) = \frac{2}{3}h(n+1) + \frac{1}{4}(9\,n^2 + 21\,n + 12)h(n)$. This sequence is not found in the OEIS[1], we added it there, it is now `A293653`, and it starts like

$$1, 2, 6, 30, 180, 1440, 12960, 142560, 1710720, 23950080, 359251200, \dots$$

In the next section we will use Equations (4) to iteratively derive the moments of the distribution of black balls after n steps.

3   Moments of periodic Pólya urns

Let $m_r(n)$ be the r-th factorial moment of the distribution of black balls after n steps, i.e.

$$m_r(n) := \mathbb{E}\left(B_n(B_n - 1)\cdots(B_n - r + 1)\right).$$

Expressing them in terms of the generating function $H(x,z)$, it holds that

$$m_r(n) = \frac{[z^n]\left.\frac{\partial^r}{\partial x^r}H(x,z)\right|_{x=1}}{[z^n]H(1,z)}.$$

Splitting them into odd and even moments, we have access to these quantities via the partial differential equation (4). As a first step we compute $h_n := [z^n]H(1,z)$, the total number of histories after n steps. We substitute $x = 1$, which makes the equation independent of the derivative with respect to x. Then, the idea is to transform (4) into two independent differential equations for $H_e(1,z)$ and $H_o(1,z)$. This is achieved by differentiating the equations with respect to z and substituting the other one to eliminate $H_e(1,z)$ or $H_o(1,z)$, respectively. This decouples the system, but increases the degree of differentiation by 1. We get

$$\left(9z^2 - 4\right)\partial_z^2 H_e(1,z) + 39z\partial_z H_e(1,z) + 24H_e(1,z) = 0,$$
$$\left(9z^2 - 4\right)\partial_z^2 H_o(1,z) + 39z\partial_z H_o(1,z) + 21H_o(1,z) = 0.$$

In this case it is easy to extract the underlying recurrence relations and solve them explicitly.

---

[1]On-Line Encyclopedia of Integer Sequences, https://oeis.org.

This also leads to the closed forms (5) for h_n, from which it is easy to compute the asymptotic number of histories for $n \to \infty$. Interestingly, the first two asymptotic terms are the same for odd and even number of steps, only the third one differs. We get

$$h_n = n! \frac{\sqrt{\pi}}{2^{1/6}\Gamma\left(\frac{2}{3}\right)} \left(\frac{3}{2}\right)^n n^{1/6} \left(1 + O\left(\frac{1}{n}\right)\right).$$

As a next step we compute the mean. Therefore, we differentiate (4) once with respect to x, substitute $x = 1$, decouple the system, derive the recurrence relations of the coefficients, and solve them. Note again that the factor $(x - 1)$ prevents higher derivatives from appearing and is therefore crucial for this method. After normalization by h_n we get

$$m_1(n) = \begin{cases} \frac{3^{3/2}\Gamma\left(\frac{2}{3}\right)^2}{2\pi} \frac{\Gamma\left(\frac{n}{2}+\frac{4}{3}\right)}{\Gamma\left(\frac{n}{2}+\frac{2}{3}\right)} & \text{if } n \text{ is even,} \\ \frac{3^{3/2}\Gamma\left(\frac{2}{3}\right)^2}{4\pi} \frac{(n+1)\Gamma\left(\frac{n}{2}+\frac{5}{6}\right)}{\Gamma\left(\frac{n}{2}+\frac{7}{6}\right)} & \text{if } n \text{ is odd.} \end{cases}$$

For the asymptotic mean we discover again the same phenomenon that its first two terms are equal for odd and even n. We get

$$m_1(n) = \frac{3}{2^{2/3}} \frac{\Gamma\left(\frac{2}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} n^{2/3} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Repeating this process, we can derive also higher moments in a mechanical way. In general we get the closed form for the r-th factorial moment

$$m_r(n) = \frac{3^r}{2^{2r/3}} \frac{\Gamma\left(\frac{r}{3}+\frac{1}{3}\right)}{\Gamma\left(\frac{1}{3}\right)} n^{2r/3} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Therefore we see that the moments $\mathbb{E}\left(B_n^{*r}\right)$ of the rescaled random variable $B_n^* := \frac{2^{2/3}}{3} \frac{B_n}{n^{2/3}}$ converge for n to infinity to the limit

$$m_r := \frac{\Gamma\left(\frac{r}{3}+\frac{1}{3}\right)}{\Gamma\left(\frac{1}{3}\right)}. \tag{6}$$

Furthermore, by the asymptotic result there exist an $n_0 > 0$ and constants a_r and b_r independent of n such that $a_r < m_r < b_r$, for all $n \geq n_0$. Thus, by the limit theorem of Fréchet and Shohat [11][2] there exists a limit distribution to which our rescaled random variables B_n^* converge to. Furthermore, Carleman's condition

$$\sum_{r>0} m_r^{-1/(2r)} = +\infty,$$

---

[2]As a funny coincidence, Fréchet and Shohat mention in [11] that the generalized Gamma distribution with parameter $p \geq 1/2$ is uniquely characterized by its moments.

is satisfied, because $m_r^{-1/(2r)} = \left(\frac{3e}{r}\right)^{1/6}(1 + o(1))$ for large r. Therefore, the moment problem is determined (i.e. there exists a unique limit distribution) and the sequence of rescaled random variables $(B_n^*)_{n>0}$ converges in law to the unique limit distribution, and not only a subsequence.

It remains to characterize this limit distribution. Let $X \sim \mathrm{GenGamma}(d, p)$ be a generalized Gamma distributed random variable (as defined in Definition 1.3). It is a distribution determined by its moments, which are given by

$$\mathbb{E}(X^r) = \frac{\Gamma\left(\frac{d+r}{p}\right)}{\Gamma\left(\frac{d}{p}\right)}.$$

Thus, the structure of m_r in Formula (6) implies that B_n^* converges to $\mathrm{GenGamma}(1, 3)$, which completes the proof of Theorem 1.5. □

The same approach allows us to study the distribution of black balls for the urn with replacement matrices $M_1 = M_2 = \cdots = M_{p-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $M_p = \begin{pmatrix} 1 & \ell \\ 0 & 1+\ell \end{pmatrix}$. We call this model the Young Pólya urn of period p and parameter ℓ.

Theorem 3.1. The renormalized distribution of black balls in the Young Pólya urn of period p and parameter ℓ is asymptotically a distribution, which we call $\mathrm{ProdGenGamma}(p, \ell, b_0, w_0)$, defined as the following product of independent distributions:

$$\frac{p^\alpha}{p+\ell}\frac{B_n}{n^\alpha} \quad \xrightarrow{\mathcal{L}} \quad \mathrm{Beta}(b_0, w_0) \prod_{i=0}^{\ell-1} \mathrm{GenGamma}(b_0 + w_0 + p + i, p + \ell)$$

with $\alpha = p/(p + \ell)$, and where $\mathrm{Beta}(b_0, w_0)$ is as usual the law with support $[0, 1]$ and density $\frac{\Gamma(b_0+w_0)}{\Gamma(b_0)\Gamma(w_0)}x^{b_0-1}(1 - x)^{w_0-1}$.

Sketch. This follows from the following r-th (factorial) moment computation:

$$\mathbb{E}\left(B_n^r\right) = \frac{(p+\ell)^r}{p^{\alpha r}}\frac{\Gamma(b_0 + r)\Gamma(b_0 + w_0)}{\Gamma(b_0)\Gamma(b_0 + w_0 + r)}\prod_{i=0}^{\ell-1}\frac{\Gamma\left(\frac{b_0+w_0+p+r+i}{p+\ell}\right)}{\Gamma\left(\frac{b_0+w_0+p+i}{p+\ell}\right)}n^{\alpha r}\left(1 + O\left(\frac{1}{n}\right)\right),$$

which in turns characterizes the $\mathrm{ProdGenGamma}$ distribution. Here, Indeed, if for some independent random variables X, Y, Z, one has $\mathbb{E}(X^r) = \mathbb{E}(Y^r)\mathbb{E}(Z^r)$ (and if Y and Z are determined by their moments), then $X \stackrel{\mathcal{L}}{=} YZ$. □

This is consistent with our results on the Young Pólya urn introduced in Section 1. Indeed, there one has $w_0 = b_0 = 1, p = 2, \ell = 1$, and therefore the renormalized distribution of black balls $\frac{p^\alpha}{p+\ell}B_n/n^\alpha$ is asymptotically $\mathrm{Unif}(0, 1) \cdot \mathrm{GenGamma}(4, 3) = \mathrm{GenGamma}(1, 3)$.

We will now see what are the implications of this result on an apparently unrelated topic: Young tableaux.

4   Urns, trees, and Young tableaux

As predicted by Anatoly Vershik in [28], the 21st century should see a lot of challenges and advances on the links of probability theory with (algebraic) combinatorics. A key role is played here by Young tableaux, because of their ubiquity in representation theory [18]. Many results on their asymptotic shape have been collected, but very few results are known on their asymptotic content when the shape is fixed (see e.g. the works by Pittel and Romik [23, 25] and Marchal [21], who have studied the distribution of the values of the cells in random rectangular Young tableaux, while the case of Young tableaux with a more general shape seems to be very intricate). It is therefore pleasant that our work on periodic Pólya urns allows us to get advances on the case of a triangular shape, with any slope.

For any fixed integers $n, \ell, p \geq 1$, we introduce the quantity $N := p\ell n(n+1)/2$. We define a *triangular Young tableau* of slope $-\ell/p$ and of size N as a classical Young tableau with N cells with length $n\ell$ and height np such that the first p rows have length $n\ell$, the next p lines have length $(n-1)\ell$ and so on (see Figure 2). We now study what is the typical value of its lower right corner (with the French convention for drawing Young tableaux, see [18] but take however care that on page 2 therein, Macdonald advises readers preferring the French convention to "read this book upside down in a mirror"!).
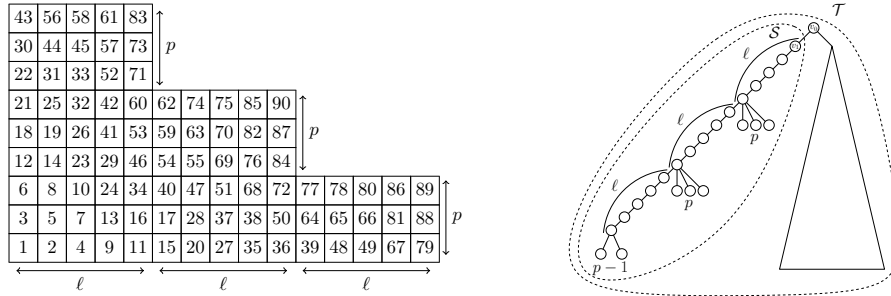

Figure 2: In this section, we see that there is a relation between Young tableaux with a given periodic shape, some trees and the periodic Pólya urns. The lower right corner of these Young tableaux is thus following the same generalized Gamma distribution we proved for urns.

It could be expected (e.g. via the Greene–Nijenhuis–Wilf algorithm for generating Young tableaux, see [12]) that the entries near the hypotenuse should be $N - o(N)$. Can we expect a more precise description of these $o(N)$ fluctuations? Our result on periodic urns enables us to exhibit the right critical exponent, and the limit law in the corner:

Theorem 4.1. Choose a uniform random triangular Young tableau $\mathcal{Y}$ of slope $-\ell/p$ and size $N := p\ell n(n+1)/2$ and put $\alpha = p/(p+\ell)$. Let X_n be the entry of the lower right. Then $(N - X_n)/n^{1+\alpha}$ converges in law to the distribution of the number of black balls in the periodic Pólya urn with replacement matrices $M_1 = \cdots = M_{p-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $M_p = \begin{pmatrix} 1 & \ell \\ 0 & 1+\ell \end{pmatrix}$, and with initial conditions $w_0 = \ell$, $b_0 = p$, i.e. we have the convergence in law, as n goes to infinity:

$$\frac{p^\alpha}{p+\ell} \frac{N - X_n}{n^{1+\alpha}} \xrightarrow{\mathcal{L}} \mathrm{ProdGenGamma}(p, \ell, b_0, w_0).$$

Sketch of proof. We first establish a link between Young tableaux and linear extensions of trees. Then we will be able to conclude via a link between these trees and periodic Pólya urns.

The right part of Figure 2 presents two trees (the "big" tree $\mathcal{T}$, which contains the "small" tree $\mathcal{S}$). More precisely, we define the rooted planar tree $\mathcal{S}$ as follows

- The left-most branch of $\mathcal{S}$ has $n\ell + 1$ vertices, which we call $v_1, v_2, \ldots, v_{n\ell+1}$, where v_1 is the root and $v_{n\ell+1}$ is the left-most leaf of the tree.

- For $2 \leq k \leq n-1$, the vertex $v_{k\ell}$ has $p+1$ children.

- The vertex $v_{n\ell}$ has $p-1$ children.

- All other vertices v_j (for $j < n\ell, j \neq k\ell$) have exactly one child.

Now, define $\mathcal{T}$ as the "big" tree obtained from the "small" tree $\mathcal{S}$ by adding a vertex v_0 as the father of v_1 and adding $N + 1 - n(p + \ell)$ children to v_0 (see Figure 2). Remark that the number of vertices of $\mathcal{T}$ is equal to $1 +$ the number of cells of $\mathcal{Y}$. Moreover, the hook length of each cell in the first row of $\mathcal{Y}$ is equal to the hook length of the corresponding vertex in the left-most branch of $\mathcal{S}$.

Let us now introduce a linear extension $E_{\mathcal{T}}$ of $\mathcal{T}$, i.e. a bijection from the set of vertices of $\mathcal{T}$ to $\{0, 1, \ldots, N\}$ such that $E_{\mathcal{T}}(u) < E_{\mathcal{T}}(u')$ whenever u is an ancestor of u'. A key result is the following: if $E_{\mathcal{T}}$ is a random uniform linear extension of $\mathcal{T}$, then X_n (the entry of the lower right corner in a random uniform Young tableau with shape $\mathcal{Y}$) has the same law as $E_{\mathcal{T}}(v_{n\ell})$:

$$X_n \overset{\mathcal{L}}{=} E_{\mathcal{T}}(v_{n\ell}). \tag{7}$$

What is more, recall that $\mathcal{T}$ was obtained from $\mathcal{S}$ by adding a root and some children to this root. Therefore, one can obtain a linear extension of the "big" tree $\mathcal{T}$ from a linear extension of the "small" tree $\mathcal{S}$ by a simple insertion procedure. This allows us to construct a random uniform linear extension $E_{\mathcal{T}}$ of $\mathcal{T}$ and a random uniform linear extension $E_{\mathcal{S}}$ of $\mathcal{S}$ such that

$$\left| \frac{2(p+\ell)}{n\ell p}(N - E_{\mathcal{T}}(v_{n\ell})) - (n\ell + p - E_{\mathcal{S}}(v_{n\ell})) \right| \to 0 \qquad \text{(in probability)}.$$

So, to summarize, we have now

$$E_{\mathcal{T}}(v_{n\ell}) \overset{\mathcal{L}}{=} E_{\mathcal{S}}(v_{n\ell}) + \text{deterministic quantity} + \text{smaller order error terms}. \tag{8}$$

The last step (which we just state here, see our long version for its full proof) is that

$$E_{\mathcal{S}}(v_{n\ell}) \overset{\mathcal{L}}{=} \text{distribution of periodic Pólya urn} + \text{deterministic quantity}. \tag{9}$$

Indeed, more precisely $N - E_{\mathcal{S}}(v_{n\ell})$ has the same law as the number of black balls in a periodic urn after $(n-1)p$ steps (an urn with period p, with adding parameter ℓ, and with initial conditions $w_0 = \ell$ and $b_0 = p$). Thus, our results on periodic urns from Section 3 and the conjunction of Equations (7), (8), and (9) gives the convergence in law for X_n which we wanted to prove. □

5   Conclusion and further work

In this article, we introduced Pólya urns with periodic replacements, and showed that they can be exactly solved with generating function techniques, and that the initial non-linear equation encoding their dynamics leads to linear (D-finite) moment generating functions, which we identify as a product of generalized Gamma distributions. Note that [17,20] involve the asymptotics of a related process (by grouping p units of time at once of our periodic Pólya urns). This related process is therefore "smoothing" the irregularities created by our periodic model, and allows us to connect with the usual famous key quantities for urns, such as the quotient of eigenvalues of the substitution matrix, etc). Our approach has the advantage to make a kind of Tauberian zoom on each unit of time, giving more asymptotic terms, and also exact enumeration.

In the full version of this work we will consider arbitrary periodic balanced urn models, and their relationship with Young tableaux. It remains a challenge to understand the asymptotic landscape of Young tableaux, even if it could be globally expected that they behave like a Gaussian free field, like for many other random surfaces [15]. As a first step, understanding the fluctuations and the universality of the critical exponents at the corner could help to get a more global picture. Note that our results on the lower right corner directly imply similar results on the upper right corner: just use our formulae by exchanging ℓ and p, i.e. for a slope corresponding to the complementary angle to 90°. Thus the critical exponent for the upper right corner is $2 - \alpha$. In fact, it is a nice surprise that there is even more structure: there is a *duality* between the limit laws X and X' of these two corners and we get the factorization as independent random variables (up to renormalization and slight modifications of the boundary conditions) $XX' \sim \Gamma(b_0)$. Similar factorizations of the exponential law, which is a particular case of Gamma distribution, have appeared recently in relation with functionals of Lévy processes, following [3].

Acknowledgments: Let us thank Cécile Mailler, Henning Sulzbach and Markus Kuba for kind exchanges on their work [17,20] and on related questions.

References

[1] Krishna B. Athreya and Peter E. Ney. *Branching processes*. Springer, 1972.

[2] Amitava Bagchi and Asim K. Pal. Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods*, 6(3):394–405, 1985.

[3] Jean Bertoin and Marc Yor. On subordinators, self-similar Markov processes and some factorizations of the exponential variable. *Electron. Comm. Probab.*, 6:95–106, 2001.

[4] Brigitte Chauvin, Cécile Mailler, and Nicolas Pouyanne. Smoothing equations for large Pólya urns. *J. Theoret. Probab.*, 28(3):923–957, 2015.

[5] Florian Eggenberger and George Pólya. Über die Statistik verketteter Vorgänge. *Z. Angew. Math. Mech.*, 3:279–290, 1923.

[6] Florian Eggenberger and George Pólya. Sur l'interprétation de certaines courbes de fréquence. *C. R. Acad. Sc.*, 187:870–872, 1928.

[7] Giulia Fanti and Pramod Viswanath. Deanonymization in the Bitcoin P2P Network. *Proceedings of the 31st Conference on Neural Information Processing Systems*, 2017.

[8] Philippe Flajolet, Philippe Dumas, and Vincent Puyhaubert. Some exactly solvable models of urn process theory. *Discrete Math. Theor. Comput. Sci. Proc.*, AG:59–118, 2006.

[9] Philippe Flajolet, Joaquim Gabarró, and Helmut Pekari. Analytic urns. *Ann. Probab.*, 33(3):1200–1233, 2005.

[10] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[11] Maurice Fréchet and James Shohat. A proof of the generalized second-limit theorem in the theory of probability. *Trans. Amer. Math. Soc.*, 33(2):533–543, 1931.

[12] Curtis Greene, Albert Nijenhuis, and Herbert S. Wilf. Another probabilistic method in the theory of Young tableaux. *J. Comb. Theory, Ser. A*, 37:127–135, 1984.

[13] Svante Janson. Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110(2):177–245, 2004.

[14] Svante Janson. Moments of Gamma type and the Brownian supremum process area. *Probab. Surv.*, 7:1–52, 2010. with addendum on pages 207–208.

[15] Richard Kenyon. Dominos and the Gaussian free field. *Ann. Probab.*, 29(3):1128–1137, 2001.

[16] Morteza Khodabin and Alireza Ahmadabadi. Some properties of generalized gamma distribution. *Math. Sci. Q. J.*, 4(1):9–27, 2010.

[17] Markus Kuba and Henning Sulzbach. On martingale tail sums in affine two-color urn models with multiple drawings. *J. Appl. Probab.*, 54(1):96–117, 2017.

[18] Ian G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Classic Texts in the Physical Sciences. The Clarendon Press, Oxford University Press, second edition, 2015.

[19] Hosam M. Mahmoud. *Pólya urn models*. CRC Press, 2009.

[20] Cécile Mailler. Describing the asymptotic behaviour of multicolour Pólya urns via smoothing systems analysis. *arXiv:1407.2879*, 2014.

[21] Philippe Marchal. Rectangular Young tableaux and the Jacobi ensemble. *Discrete Math. Theor. Comput. Sci. Proceedings*, BC:839–850, 2016.

[22] Erol A. Peköz, Adrian Röllin, and Nathan Ross. Generalized gamma approximation with rates for urns, walks and trees. *Ann. Probab.*, 44(3):1776–1816, 2016.

[23] Boris Pittel and Dan Romik. Limit shapes for random square Young tableaux. *Adv. Appl. Math.*, 38(2):164–209, 2007.

[24] Ronald L. Rivest. Tabulation audits: Explained and extended. *arXiv:1801.00528*, 2018.

[25] Dan Romik. *The surprising mathematics of longest increasing subsequences*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, New York, 2015.

[26] Robert T. Smythe and Hosam M. Mahmoud. A survey of recursive trees. *Theory Probab. Math. Statist.*, 51:1–29, 1994.

[27] Edney W. Stacy. A generalization of the gamma distribution. *Ann. Math. Statist.*, 33:1187–1192, 1962.

[28] Anatoly M. Vershik. Randomization of algebra and algebraization of probability. In *Mathematics unlimited—2001 and beyond.*, pages 1157–1166. Springer, 2001.