2005 International Conference on Analysis of Algorithms DMTCS proc. **AD**, 2005, 275–286

Classification of large Pólya-Eggenberger urns with regard to their asymptotics

Nicolas Pouyanne[1]

[1] *Département de mathématiques, LAMA UMR 8100 CNRS, Université de Versailles - Saint-Quentin*
45, avenue des Etats-Unis, 78035 Versailles cedex

This article deals with Pólya generalized urn models with constant balance in any dimension. It is based on the algebraic approach of Pouyanne (2005) and classifies urns having "large" eigenvalues in five classes, depending on their almost sure asymptotics. These classes are described in terms of the spectrum of the urn's replacement matrix and examples of each case are treated. We study the cases of so-called cyclic urns in any dimension and m-ary search trees for $m \geq 27$.

1 Introduction

We consider (generalized balanced) Pólya-Eggenberger urns with balls of s different types (or colours), s being any integer ≥ 2. Namely, under this model, the urn may contain balls of colours named $1, \ldots, s$ and evolves as a Markov process as follows. Its initial composition is described by a non random column-vector $U_1 = {}^t(U_{1,1}, \ldots, U_{1,s})$ whose k-th coordinate is the initial number of balls of colour k. One proceeds to successive draws of one ball at random in the urn, any ball being at any time equally likely drawn. At each draw, one inspects the colour of the drawn ball, places it back into the urn and adds other balls following invariably the same rule. This rule is given by the (non random) *replacement matrix*

$$R = (r_{i,j})_{1 \leq i,j \leq s},$$

the entry $r_{i,j}$ being the number of balls of colour j one adds if a ball of colour i has been drawn. In our model, the replacement matrix has nonnegative off-diagonal entries, but may have negative diagonal ones (that correspond better to prelevement than to addition of balls), submitted to the following assumptions:
1- (balance hypothesis) $\exists S \in \mathbb{Z}_{\geq 1},\ \forall k \in \{1, \ldots, s\}$,

$$\sum_{j=1}^{s} r_{k,j} = S; \tag{1}$$

2- (sufficient condition of tenability) $\forall k \in \{1, \ldots, s\}$,

$$r_{k,k} \geq 0 \ \text{ or } \ U_{1,k}\mathbb{Z} + \sum_{j=1}^{s} r_{j,k}\mathbb{Z} = r_{k,k}\mathbb{Z}. \tag{2}$$

The composition of the urn will be denoted by $U_n = {}^t(U_{n,1}, \ldots, U_{n,s})$, the number $U_{n,k}$ being the number of balls of colour k after $n-1$ draws. The subject of our study is the asymptotic behaviour of this random vector as n tends to infinity. Hypothesis 1- requires the total number of added balls at each draw to be always the same; this number will be denoted by S and called **balance** of the urn. If $|U_1|$ denotes the initial total number of balls, this implies that the urn contains $|U_1| + nS$ balls after the n-th draw. Arithmetical Hypothesis 2- is a classical sufficient condition for the process not to extinguish after a finite number of draws (Bagchi and Pal (1985), Gouet (1997), Flajolet et al. (2005), Pouyanne (2005)), as can be checked by an elementary induction. One can replace it by conditioning the whole asymptotic study to non extinction.

Pólya-Eggenberger urns have been studied by many authors since the original article Pólya (1930). Roughly speaking, employed methods have been direct probabilistic considerations, generating functions and partial differential equations, embedding in continuous time process and martingale arguments; one can refer to Flajolet et al. (2005) or Puyhaubert (2005) for good surveys on the subject.

1365–8050 © 2005 Discrete Mathematics and Theoretical Computer Science (DMTCS), Nancy, France

A Pólya-Eggenberger urn defined by U_1 and R being given, one can consider it as a random walk in $\mathbb{R}^s$ having U_1 as initial point, the increment $U_{n+1} - U_n$ at time n being at random one of R's rows, the probability of the k-th one to be chosen being equal to $U_{n,k}/(|U_1| + (n-1)S)$ (proportion of balls of colour k after the $(n-1)$-st draw).

Adopting this point of view, we *standardize* the process (or the urn) the following way. Our random vector of interest is $X_n = \frac{1}{S}U_n$. It can be interpreted as a Pólya-Eggenberger urn with balance 1 (*i.e.* with $S = 1$), the replacement matrix $\frac{1}{S}R$ having rational entries. The extension of this point of view to real-valued replacement matrix leads to the definition of what has been called **Pólya process** in Pouyanne (2005).

When 1 is simple eigenvalue of $\frac{1}{S}R$, the random vector U_n admits an almost sure non random drift. More precisely, there exists a non random vector v_1 such that U_n/n converges almost surely to Sv_1 as n tends to infinity. This vector v_1 is the only vector fixed by $\frac{1}{S}{}^tR$ whose coordinates' sum equals 1. When 1 is multiple eigenvalue of $\frac{1}{S}R$, then U_n/n converges almost surely to a random vector that follows a Dirichlet distribution (see below and Gouet (1997) for the almost sure asymptotics of U_n/n). The asymptotic behaviour of the difference $X_n - nv_1$ has been for a long time known to depend on the spectrum of R. We will say that a Pólya-Eggenberger urn with replacement matrix R and balance S is **small** when 1 is simple eigenvalue of $\frac{1}{S}R$ and when every other eigenvalue of $\frac{1}{S}R$ has a real part $\leq 1/2$. Otherwise, it will be said **large**.

When the urn is small, under some conditions of irreducibility, one can establish convergence in law of the normalisation $(X_n - nv_1)/\sqrt{n\log^\nu n}$ to a centered Gaussian vector, the integer ν depending only on the conjugacy class of R (Athreya and Karlin (1968), Janson (2004)). If one releases this irreducibility, considering for instance urns with triangular replacement matrix, convergence in distribution (to most often non normal laws) has been shown and moments have been computed in several cases in low dimension (see Janson (2005), Puyhaubert (2005)).

Our case of interest is the one of large urns. Almost sure convergence-like results on $X_n - nv_1$ have been established since the work of Athreya and Karlin (Athreya and Karlin (1968)) and refined in some more general cases by Janson (Janson (2004)) by means of embedding of the process in continuous time, but these results require still some irreducibility-type assumptions. In Pouyanne (2005), in any case of large urns, following a different method that stays in the discrete field, almost sure (and L^p for any $p \geq 1$) asymptotics is established and a way to compute the moments of limit random vectors is given.

In this paper, we classify large Pólya-Eggenberger urns with regard to their asymptotics, give some generic example of each case and some other new results about particular families of urns (general two-dimensional urn, cyclic urns, m-ary search trees).

2 Asymptotics of large Pólya-Eggenberger urns

Basic objects and notations are introduced in this section, following the method of Pouyanne (2005). Then we state the classification of large urns with regard to their asymptotics.

2.1 Notations and overview of the method

Let's consider an s-dimensional Pólya-Eggenberger urn $(U_n)_{n\geq 1}$ with balance S defined as in Section 1 by its initial composition U_1 and its replacement matrix R. Let τ_1 be the renormalized initial total number of balls, namely

$$\tau_1 = \frac{1}{S}|U_1| = |X_1|.$$

Let A be the $s \times s$ matrix with rational (or real if one admits the generalized definition of a Pólya process) entries defined as the transpose

$$A = \frac{1}{S}\,{}^tR.$$

We adopt notations and definitions of Pouyanne (2005). Let $w_1, \dots, w_s$ be the column-vectors of A and $x_1, \dots, x_s$ the generic coordinates of $\mathbb{R}^s$ or $\mathbb{C}^s$. With this notation, the *transition operator* Φ is given by the formula

$$\Phi(f)(x) = \sum_{k=1}^{s} x_k \big[f(x + w_k) - f(x)\big], \tag{3}$$

for any function $f : \mathbb{C}^s \to V$ (V is any vector space) and any $x = (x_1, \dots, x_s) \in \mathbb{C}^s$. This operator has a good decomposition on polynomials spaces given by so called reduced polynomials of the process; these polynomials are defined just below.

Let $(u_k)_{1\leq k\leq s}$ be a **Jordan basis** of the process, *i.e.* a basis of linear forms on $\mathbb{R}^s$ or $\mathbb{C}^s$ that satisfy
1- $u_1(x) = \sum_{1\leq k\leq s} x_k$ for all x;
2- $u_k \circ A = \lambda_k u_k + \varepsilon_k u_{k-1}$ for all $k \geq 2$, where the λ_k are complex numbers (necessarily eigenvalues of A) and where the ε_k are numbers in $\{0,1\}$ that satisfy $\lambda_k \neq \lambda_{k-1} \Longrightarrow \varepsilon_k = 0$.
This implies that the transpose of A has a Jordan normal form in this basis and that u_1 is fixed by A (balance assumption). A Jordan basis being chosen, we will denote by $(v_k)_{1\leq k\leq s}$ its dual basis of vectors of $\mathbb{R}^s$ or $\mathbb{C}^s$.

A subset $J \subseteq \{1,\dots,s\}$ is called **monogenic block of indices** when J has the form $J = \{m, m+1,\dots,m+r\}$ ($r \geq 0$, $m \geq 1$, $m+r \leq s$) with $\varepsilon_m = 0$, $\varepsilon_k = 1$ for every $k \in \{m+1,\dots,m+r\}$ and J is maximal for this property. In other words, J is monogenic when $\mathrm{Vect}\{u_j,\ j\in J\}$ is A-stable and when the matrix of the endomorphism of $\mathrm{Vect}\{u_j,\ j\in J\}$ induced by ${}^tA$ in the $(u_j)_j$ basis is one of the Jordan blocks of the Jordan normal form of A mentioned above. Any monogenic block of indices J is associated with a unique eigenvalue of A that will be denoted by

$$\lambda(J).$$

We denote by $\mathrm{Sp}(A)$ the set of eigenvalues of A and σ_2 the real number defined as

$$\sigma_2 = \begin{cases} 1 \text{ if } 1 \text{ is multiple eigenvalue of } A \\ \\ \max\{\Re\lambda,\ \lambda\in\mathrm{Sp}(A),\ \lambda\neq 1\} \text{ if } 1 \text{ is simple eigenvalue;} \end{cases}$$

hypotheses on A imply that $\sigma_2 < 1$ in the second case. The urn is called **large** when $1/2 < \sigma_2 \leq 1$. Otherwise it is called small. A monogenic block of indices J is called **principal block** when $\Re\lambda(J) = \sigma_2$ and J has maximal size among monogenic blocks that satisfy that property.

We denote by δ_k the k-th vector of the canonical basis of $\mathbb{Z}^s$. For any s-uple of nonnegative integers $\alpha = \sum_{k=1}^s \alpha_k\delta_k \in (\mathbb{Z}_{\geq 0})^s$, we use as usual the notations

$$\mathbf{u}^\alpha = \prod_{1\leq k\leq s} u_k^{\alpha_k} \quad \text{and} \quad \langle \alpha,\lambda\rangle = \sum_{1\leq k\leq s} \alpha_k\lambda_k$$

where λ_k denotes the eigenvalue associated with the linear form u_k and $\lambda = (\lambda_1,\dots,\lambda_s)$. Furthermore, the symbol $\alpha \leq \beta$ on s-uples of nonnegative integers will denote the **degree-antialphabetical order**, defined, if $|\alpha| = \sum_{1\leq k\leq s}\alpha_k$, by $\alpha = (\alpha_1,\dots,\alpha_s) < \beta = (\beta_1,\dots,\beta_s)$ when $\Big(|\alpha| < |\beta|\Big)$ or $\Big(|\alpha| = |\beta|$ and $\exists r \in \{1,\dots,s\}$ such that $\alpha_r < \beta_r$ and $\alpha_t = \beta_t$ for any $t > r\Big)$.

As shown in Pouyanne (2005), a Jordan basis being chosen, there exists a unique basis $(Q_\alpha)_{\alpha\in(\mathbb{Z}_{\geq 0})^s}$ of polynomials in s variables such that

1- $Q_0 = 1$ and $Q_{\delta_k} = u_k$ for all $k \in \{1,\dots,s\}$;

2- for all α, $Q_\alpha - \mathbf{u}^\alpha$ belongs to $\mathrm{Vect}\{Q_\beta,\ \beta < \alpha,\ \langle\beta,\lambda\rangle \neq \langle\alpha,\lambda\rangle\}$;

3- for all α, $\Phi(Q_\alpha) - \langle\alpha,\lambda\rangle Q_\alpha$ belongs to $\mathrm{Vect}\{Q_\beta,\ \beta<\alpha,\ \langle\beta,\lambda\rangle = \langle\alpha,\lambda\rangle\}$. The polynomial Q_α is named α-th **reduced polynomial** with regard to the choice of the Jordan basis $(u_k)_k$. The reduced polynomials can be recursively computed in any case; their unicity leads sometimes to a closed-form (for some triangular urns for instance). In any case, for any nonnegative integer p, the reduced polynomial $Q_{p\delta_1}$ is an eigenvector of Φ associated with the eigenvalue p and have the following closed-form

$$Q_{p\delta_1} = u_1(u_1+1)\dots(u_1+p-1).$$

These notations being adopted, one can state the general result on the asymptotics of large urns that is shown in Pouyanne (2005).

Theorem 1 *Take a* **large** *Pólya-Eggenberger urn. Fix a Jordan basis $(u_k)_{1\leq k\leq s}$ of linear forms of the process and $(v_k)_{1\leq k\leq s}$ its dual basis of vectors of $\mathbb{C}^s$; let $J_1,\dots,J_r$ be the principal blocks of indices of A and $\nu+1$ the common size of the J_k's ($r\geq 1$ and $\nu\geq 0$).*
1- **(Convergence of principal coordinates of the process)** *For any $k\in\{1,\dots,r\}$, the complex-valued process $u_{\min J_k}(X_n)/n^{\lambda(J_k)}$ converges to a random variable W_k as n tends to infinity almost surely and in any L^p, $p\geq 1$.*
*2- (***Random vector's asymptotics)**

$$X_n - nv_1 = \frac{1}{\nu!}\log^\nu n \sum_{1\leq k\leq r} n^{\lambda(J_k)} W_k\, v_{\max J_k} + o\left(n^{\sigma_2}\log^\nu n\right) \tag{4}$$

as n tends to infinity, the small o being almost sure and in L^p for every $p\geq 1$.
3- **(Joint moments of the limits)** *If one denotes by $(Q_\alpha)_{\alpha\in(\mathbb{Z}_{\geq 0})^s}$ the reduced polynomials of the process relative to the Jordan basis $(u_k)_{1\leq k\leq s}$, all joint moments of the random variables $W_1,\dots,W_r$ are given by: for all $\alpha_1,\dots,\alpha_r\in\mathbb{Z}_{\geq 0}$,*

$$E\left(\prod_{1\leq k\leq r} W_k^{\alpha_k}\right) = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+\langle\alpha,\lambda\rangle)} Q_\alpha(X_1)$$

where $\alpha=\sum_{1\leq k\leq r}\alpha_k\delta_{\min J_k}$.

2.2 Very first example

We give "slowly" one first example in dimensions 2.

Example 1 Consider the 2-colour urn defined by an initial condition U_1 and the replacement matrix

$$R=\begin{pmatrix} 15 & 5\\ 4 & 16\end{pmatrix}.$$

One has $S=20$ as balance and, with our notations, $A=\begin{pmatrix} 3/4 & 1/5\\ 1/4 & 4/5\end{pmatrix}$; the urn is large, its eigenvalues being 1 and $\sigma_2=11/20$. One can choose $u_1(x,y)=x+y$ and $u_2(x,y)=-\frac{5}{9}x+\frac{4}{9}y$ as Jordan basis, its dual basis being given by $v_1={}^t(4/9,5/9)$ and $v_2={}^t(-1,1)$. The only principal block of indices is $\{2\}$ in this case. The theorem asserts that

$$\frac{1}{n^{11/20}}\left(X_n-\frac{n}{9}\begin{pmatrix}4\\5\end{pmatrix}\right)\underset{n\to\infty}{\longrightarrow} W\begin{pmatrix}-1\\1\end{pmatrix}$$

almost surely and in $\mathrm{L}^{\geq 1}$, where W is a real-valued random variable. Computation of the first reduced polynomials that provide moments of W gives

$$\begin{aligned}
Q_{(0,2)} &= u_2^2+\tfrac{121}{162}u_1-\tfrac{11}{180}u_2=\tfrac{253}{324}x+\tfrac{583}{810}y+\tfrac{25}{81}x^2-\tfrac{40}{81}xy+\tfrac{16}{81}y^2,\\
Q_{(0,3)} &= u_2^3+\tfrac{121}{54}u_1u_2-\tfrac{11}{60}u_2^2+\tfrac{1331}{47385}u_1+\tfrac{1573}{1350}u_2,\\
Q_{(0,4)} &= u_2^4+\tfrac{121}{27}u_1u_2^2-\tfrac{11}{30}u_2^3+\tfrac{14641}{8748}u_1^2-\tfrac{14641}{37908}u_1u_2+\tfrac{16819}{3600}u_2^2+\tfrac{2298637}{1166400}u_1-\tfrac{4143403}{12636000}u_2.
\end{aligned} \tag{5}$$

For example, if one begins the process with one ball of each colour *i.e.* with $x_1=x_2=\frac{1}{20}$, then $EW=-\frac{1}{180}\frac{\Gamma(1/10)}{\Gamma(13/20)}\sim -0.038$, $EW^2=\frac{152}{2025}\frac{\Gamma(1/10)}{\Gamma(6/5)}\sim 0.777$ $EW^3=-\frac{133093}{12636000}\frac{\Gamma(1/10)}{\Gamma(7/4)}\sim -0.109$, *etc*. This is enough to show for instance that the distribution of W is not normal (the first three moments m_1,m_2,m_3 of a normal distribution satisfy the relation $2m_1^3-3m_1m_2+m_3=0$).

3 Classification; generic examples

A Pólya-Eggenberger urn will be called **real** when every eigenvalue of A associated with a principal block is real. Otherwise, it will be called **imaginary**. When all principal blocks of indices have size 1, the urn is called **principally semisimple**. Otherwise, it is called non principally semisimple. The expression semisimple is taken from linear algebra: an endomorphism is called semisimple when it admits a basis of eigenvectors over an extension of the ground field. The adverb principally refers to the restriction of A to the sum of its characteristic spaces that correspond to principal blocks of indices.

As can be directly seen from Theorem 1, a large real urn have an almost sure limit after substraction of the drift and suitable renormalization. On the contrary, an large imaginary urn gives rise to an oscillatory almost sure random phenomenon. Some authors (see for example Chern and Hwang (2001)) pointed out this fact claiming that no normalization that consists in dividing the difference $X_n - nv_1$ provides any limit law. Principal semisimplicity leads to asymptotics in the powers-of-n scale; non principal semisimplicity requires the addition of entire powers of $\log n$.

Theorem 1 leads to a classification of large Pólya-Eggenberger urns in five types depending on the form of their asymptotics. We summarize this classification in the following table. Subsection 3.1 deals with the particular case of so-called essentially Pólya urns. In Subsection 3.2, we give five examples related to the classification. These examples have generic virtues. Only the closed-forms of reduced polynomials that appear in some of these cases are due to the very particular forms of the replacement matrices and cannot straightforwardly be generalized to any urn.

In the table, "pss" means principally semisimple.

Large urn	**Almost sure and $\mathrm{L}^{\geq 1}$ asymptotics**	$W_1, \ldots, W_r$
Essentially Pólya	$\frac{X_n}{n} \underset{n\to\infty}{\longrightarrow} W_1 v_1 + \ldots + W_r v_r$	$(W_1, \ldots, W_r)$ Dirichlet
Real and pss	$\frac{X_n - nv_1}{n^{\sigma_2}} \underset{n\to\infty}{\longrightarrow} W_1 v_2 + \ldots W_r v_{r+1}$	Joint moments of W_k's
Real and not pss	$\frac{1}{\nu!}\frac{X_n - nv_1}{n^{\sigma_2}\log^{\nu} n} \underset{n\to\infty}{\longrightarrow} W_1 v_{\max J_1} + \ldots + W_r v_{\max J_r}$	Joint moments of W_k's
Imaginary and pss	$\frac{X_n - nv_1}{n^{\sigma_2}} = n^{i\Im\lambda(J_1)} W_1 v_1 + \ldots + n^{i\Im\lambda(J_r)} W_r v_r + o(1)$	Joint moments of W_k's
Imaginary and not pss	$\frac{1}{\nu!}\frac{X_n - nv_1}{n^{\sigma_2}\log^{\nu} n} = n^{i\Im\lambda(J_1)} W_1 v_{\max J_1} + \ldots + n^{i\Im\lambda(J_r)} W_r v_{\max J_r} + o(1)$	Joint moments of W_k's

Note that, in the imaginary case, the computation of joint moments of W_k's leads in particular to the computation of joint moments of $\Re W_k$'s, $\Im W_k$'s and $|W_k|^2$'s too.

3.1 Essentially Pólya urn

An urn will be called **essentially Pólya** when 1 is multiple eigenvalue of A, *i.e.* when $\sigma_2 = 1$. Let $r \geq 2$ be the multiplicity of 1 as eigenvalue of A. As shown in Gouet (1997) and Pouyanne (2005), the urn is necessarily semisimple. Using the so-called graph of the x_k's and w_k's (or the graph of the replacement matrix), one finds a basis $(u_1, \ldots, u_r)$ of linear forms fixed by A and a partition $I_1, \ldots I_r$ of $\{1, \ldots s\}$ such that for any $k \in \{1, \ldots r\}$, $u_k(w_j) = 1$ if $j \in I_k$ and $u_k(w_j) = 0$ if $j \notin I_k$. We denote by $(v_1, \ldots, v_r)$ its dual basis of $\ker(A - 1)$. For such a basis, $\sum_{k=1}^{s} x_k = \sum_{k=1}^{r} u_k$.

An adaptation of Theorem 1 presented in Pouyanne (2005) implies that there exist real random variables $W_1, \ldots, W_r$ such that

$$\frac{X_n}{n} \underset{n\to\infty}{\longrightarrow} W_1 v_1 + \ldots + W_r v_r$$

almost surely and in $\mathrm{L}^{\geq 1}$ where the random vector $(W_1, \ldots, W_r)$ has Dirichlet distribution with parameters $u_1(X_1), \ldots, u_r(X_1)$, whose density on the simplex $\{\xi_1 \geq 0, \ldots, \xi_r \geq 0, \sum_{k=1}^{r} \xi_k = 1\}$ of $\mathbb{R}^r$ is given by

$$(\xi_1, \ldots, \xi_r) \mapsto \Gamma(\tau_1) \prod_{k=1}^{r} \frac{\xi_k^{u_k(X_1)}}{\Gamma(u_k(X_1))}.$$

3.2 Examples

We give a list of five matrices R_k, $2 \leq k \leq 6$. For any k, we consider the 5-colour urn process defined by an initial condition and its replacement matrix R_k and deal with it in Example k-.

$$R_2 \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 & 0 \\ 2 & 0 & 0 & 0 & 2 \end{pmatrix}, R_3 = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 2 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 & 2 \end{pmatrix}, R_4 = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 & 0 \\ 1 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix},$$
$$R_5 = \begin{pmatrix} 6 & 2 & 0 & 0 & 0 \\ 0 & 6 & 2 & 0 & 0 \\ 2 & 0 & 6 & 0 & 0 \\ 3 & 0 & 0 & 5 & 0 \\ 4 & 0 & 0 & 0 & 4 \end{pmatrix}, R_6 = \begin{pmatrix} 6 & 1 & 0 & 0 & 1 \\ 0 & 6 & 1 & 0 & 1 \\ 1 & 0 & 6 & 0 & 1 \\ 0 & 1 & 1 & 6 & 0 \\ 1 & 0 & 0 & 1 & 6 \end{pmatrix}. \tag{6}$$

Example 2 Consider the 5-colour urn defined by an initial condition and its replacement matrix R_2. Even if A is not semisimple, the urn is real and principally semisimple, and admits a unique principal block of indices (of size 1). We choose $u_2 = x_2$. Then, for any choice of a Jordan basis $(u_1, \dots, u_5)$, one has $v_1 = {}^t(1,0,0,0,0)$ and $v_2 = {}^t(-1,1,0,0,0)$. Because of the particular triangular form of R_2 (zeros under the entry 3), one can in this case derive explicitely the reduced polynomials that intervene in the moments of W from their properties: $Q_{0,p,0,0,0} = x_2(x_2+3/4)\dots(x_2+3(p-1)/4)$ for any positive integer p. The almost sure asymptotics is given by

$$\frac{1}{n^{3/4}}(X_n - nv_1) \underset{n\to\infty}{\longrightarrow} Wv_2$$

where the moment generating function of W is, if one adopts the notation $\tau_2 = u_2(X_1)$,

$$E(\exp zW) = \sum_{p\geq 0} \frac{(3/4)^p}{p!} \frac{\Gamma(\tau_1)}{\Gamma(\tau_1 + 3p/4)} \frac{\Gamma(4\tau_2/3 + p)}{\Gamma(4\tau_2/3)} z^p. \tag{7}$$

Example 3 Consider the 5-colour urn defined by an initial condition and its replacement matrix R_3. The urn, real and principally semisimple, admits two principal blocks of indices (of size 1). We choose $u_2 = x_2$ and $u_3 = x_3$. Then, for any choice of a Jordan basis $(u_1, \dots, u_5)$, one has $v_1 = {}^t(1,0,0,0,0)$, $v_2 = {}^t(-1,1,0,0,0)$ and $v_3 = {}^t(-1,0,1,0,0)$. As in the preceding example, because of the particular form of the matrix, the reduced polynomials of interest can be computed: $Q_{0,p,q,0,0} = x_2(x_2+3/4)\dots(x_2+3(p-1)/4) \times x_3(x_3+3/4)\dots(x_3+3(q-1)/4)$ for any nonnegative integers p and q. The almost sure asymptotics is given by

$$\frac{1}{n^{3/4}}(X_n - nv_1) \underset{n\to\infty}{\longrightarrow} W_1v_2 + W_2v_3,$$

where the joint moments of the real random variables W_1 and W_2 are

$$EW_1^{p_1}W_2^{p_2} = (3/4)^{p_1+p_2} \frac{\Gamma(\tau_1)}{\Gamma(\tau_1 + 3(p_1+p_2)/4)} \frac{\Gamma(4\tau_2/3 + p_1)}{\Gamma(4\tau_2/3)} \frac{\Gamma(4\tau_3/3 + p_2)}{\Gamma(4\tau_3/3)} \tag{8}$$

for any nonnegative integers p_1 and p_2; in this formula, $\tau_2 = u_2(X_1)$ and $\tau_3 = u_3(X_1)$.

Example 4 Consider the 5-colour urn defined by an initial condition and its replacement matrix R_4, real and not principally semisimple. The eigenvalue $3/4$ of A has multiplicity 4, and A admits two principal blocks, of size 2. A natural Jordan basis is given by $u_2 = x_3$, $u_3 = 4x_2$, $u_4 = x_5$ and $u_5 = 4x_4$. With this choice, one has $v_1 = {}^t(1,0,0,0,0)$, $v_3 = \frac{1}{4}{}^t(-1,1,0,0,0)$ and $v_5 = \frac{1}{4}{}^t(-1,0,0,1,0)$. Suitable reduced polynomials admit a closed-form as in the preceding examples; the asymptotics is given by

$$\frac{1}{n^{3/4}\log n}(X_n - nv_1) \underset{n\to\infty}{\longrightarrow} W_1v_3 + W_2v_5$$

where the joint moments of the real random variables W_1 and W_2 are given by (8).

Example 5 Consider the 5-colour urn defined by an initial condition and its replacement matrix R_5. The matrix A is semisimple, and its eigenvalues are $1, 5/8, 5/8 \pm i\sqrt{3}/8$ and $1/2$ so that the urn is imaginary and principally semisimple. There are three principal blocks, of size one. A Jordan basis can be chosen so that $u_2 = x_4$, $u_3 = zx_1 - x_2 + \overline{z}x_3 + (1+\overline{z})x_4 + 2x_5$, and $u_4 = \overline{u_3}$ where $z = \exp(i\pi/3)$. Then $v_1 = \frac{1}{3}{}^t(1,1,1,0,0)$, $v_2 = \frac{1}{3}{}^t(-1,2,-4,3,0)$, $v_3 = \frac{1}{3}{}^t(\overline{z},1,z,0,0)$ and $v_4 = \overline{v_3}$. The almost sure asymptotics is then given by

$$\frac{1}{n^{5/8}}(X_n - nv_1) = W_1 v_2 + 2\Re\left(e^{i\log n\sqrt{3}/8} W_2 v_3\right) + o(1).$$

Because of the zeros on the fourth column of R_5, the reduced polynomials associated with the random variable W_1 are computable as already done in Example 2. Computation of the very first other reduced polynomials provides $Q_{0,1,1,0,0} = u_2u_3 + \frac{5}{16}(3 - i\sqrt{3})u_2$, $Q_{0,0,2,0,0} = u_3^2 + (3 + i\sqrt{3}/4)u_5 + \frac{1}{52}(-25 + 2i\sqrt{3})u_4 + \frac{1}{37}(66 - 19i\sqrt{3})u_2$, $Q_{0,0,1,1,0} = u_3u_4 + \frac{7}{4}u_5 + \frac{7}{5}u_2 + \frac{7}{4}u_1$. To avoid too much heaviness, we just give examples of joint moments when the initial composition of the urn consists in one ball of each colour (all of them being computed from the above reduced polynomials): $EW_1 = \frac{\Gamma(5/8)}{2\Gamma(1/4)} \sim 0.198$, $EW_2 = \frac{\Gamma(5/8)}{8\Gamma(5/4 + i\sqrt{3}/8)}(3 + \overline{z}) \sim 0.719 - 0.142i$, $E|W_2|^2 \sim 2.544$, $E(\Re W_2)^2 = \frac{1}{2}E|W_2|^2 + \frac{1}{2}\Re(EW_2^2) \sim 1.630$, $E(\Im W_2)^2 = \frac{1}{2}E|W_2|^2 - \frac{1}{2}\Re(EW_2^2) \sim 0.914$, *etc.* One can for example, with the help of symbolic computation (I did it with Maple) compute $Q_{0,0,2,2,0}$ and show that $E|W_2|^4 \sim 12.957$.

Example 6 Consider the 5-colour urn defined by an initial condition and its replacement matrix R_6. The double eigenvalues of A are λ and $\overline{\lambda}$, where $\lambda = (11 + i\sqrt{3})/16$, so that the urn is imaginary. It is not principally semisimple, having two principal blocks, of size two. If $(u_1, \ldots, u_5)$ is any Jordan basis with eigenforms $u_2 = x_1 + x_2 + x_3 - 2zx_4 - (1+\overline{z})x_5$ and $u_4 = \overline{u}_2$ (same z as before), then $v_1 = \frac{1}{49}{}^t(11,9,8,7,14)$, $v_3 = \frac{5+i\sqrt{3}}{336}{}^t(-z,-\overline{z},1,0,0)$ and $v_5 = \overline{v}_3$. The almost sure asymptotics is then given by

$$\frac{1}{n^{11/16}\log n}(X_n - nv_1) = 2\Re\left(n^{i\sqrt{3}/16} W_1 v_3\right) + o(1).$$

The expectation of W_1 is $EW_1 = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+\lambda)}u_2(X_1)$; when the initial composition of the urn consists in one ball of each colour, then $EW_1 = \frac{\Gamma(5/8)}{8\Gamma(21/16 + i\sqrt{3}/16)}\overline{z} \sim 0.103 - 0.173i$, $E|W_1|^2 \sim -1.394 - 2.468i$, $EW_1^2 \sim 0.453 - 0.219i$, *etc.*

4 Miscellaneous examples

4.1 General two-dimensional large urn

The general two-dimensional Pólya-Eggenberger urn process with balance 1 has a replacement matrix $R = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$ where a and b are nonnegative rational (real) numbers. The eigenvalues are 1 and $1-a-b$; the urn is large if and only if $a+b < 1/2$ and is always real and semisimple. If $a = b = 0$, the almost sure limit of X_n/n has a Dirichlet (or beta) distribution as already stated in Subsection 3.1; this case corresponds to the original Pólya urn (Pólya (1930)). When the urn is not principally Pólya, one can compute the general form of first moments of the renormalized process's limit.

Theorem 2 *Assume that a and b are two nonnegative real numbers such that $0 < a+b < 1/2$. Let $(X_n)_n$ be the large Pólya-Eggenberger urn process defined by the above replacement matrix R and the initial composition $X_1 = {}^t(x_1, y_1)$. We denote $v_1 = \frac{1}{a+b}{}^t(b,a)$, $v_2 = \frac{1}{a+b}{}^t(1,-1)$, $\tau_1 = x_1 + y_1$ and $\tau_2 = ax_1 - by_1$. Then, almost surely and in L^p for every $p \geq 1$,*

$$\frac{1}{n^{1-a-b}}\left(X_n - nv_1\right) \underset{n\to\infty}{\longrightarrow} Wv_2$$

where W is a real random variable that satisfies

$$EW = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+1-a-b)}\tau_2, \quad EW^2 = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+2-2a-2b)}\left(\tau_2^2 + (a-b)(1-a-b)\tau_2 + \frac{ab(1-a-b)^2}{1-2a-2b}\tau_1\right)$$

$$EW^3 = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+3-3a-3b)}\left(\tau_2^3 + \frac{3ab(1-a-b)^2}{1-2a-2b}\tau_1\tau_2 + 3(a-b)(1-a-b)\tau_2^2 \right.$$
$$\left. - \frac{(1-a-b)^2(4a^3-3a^2b-3ab^2+4b^3-2a^2+2ab-2b^2)}{1-2a-2b}\tau_2 + \frac{4ab(a-b)(1-a-b)^3}{2-3a-3b}\tau_1\right). \tag{9}$$

PROOF. The given basis is the dual of a Jordan one ($u_1(x,y) = x + y$ as usual, $u_2(x,y) = ax - by$). Because of Theorem 1, one just has to compute the corresponding first three reduced polynomials (they are given by the formulae; an incredule reader has just to verify that they are eigenvectors for the transition operator Φ). □

This theorem gives a generic answer to a natural question of S. Janson: is the limit law of W normal (notations of Theorem 2)?

Corollary 3 *The limit distribution (W) of a renormalized large two-dimensional Pólya-Eggenberger urn is generically not normal.*

PROOF. If a, b, τ_1, τ_2 are defined as above, the distribution of W is normal only if $2(EW)^3 - 3(EW)E(W^2) + E(W^3) = 0$, because the first three moments of a normal law satisfy this relation. This is the equation of an analytic hypersurface in the variables (a, b, τ_1, τ_2). □

4.2 Cyclic urns

If s is a positive integer, we call s-colour cyclic urn any Pólya-Eggenberger urn process defined by an initial composition X_1 and the replacement matrix

$$R = \begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & 0 & & \\ & & & \ddots & 1 \\ 1 & & & & 0 \end{pmatrix}. \tag{10}$$

Its colours are elements of $\mathbb{Z}/s\mathbb{Z}$, and a ball of colour $c+1$ is added in the urn when a ball of colour c is drawn. We will denote $e_s = \exp(2i\pi/s)$.

Theorem 4 *Let $(X_n)_n$ be an s-colour cyclic urn process, $s \geq 1$.*

1- $X_n/n \to v_1$ almost surely as n tends to infinity, where $v_1 = \frac{1}{s}{}^t(1, 1, \ldots, 1) \in \mathbb{R}^s$.

2- If $s \leq 5$, the urn is small and $(X_n - nv_1)/\sqrt{n}$ converges in distribution to a centered Gaussian vector with values in the hyperplane $\{x_1 + \ldots + x_s = 0\}$ of $\mathbb{R}^s$.

3- If $s = 6$, the urn is small and $(X_n - nv_1)/\sqrt{n \log n}$ converges in distribution to a centered Gaussian vector with values in the hyperplane $\{x_1 + \ldots + x_s = 0\}$ of $\mathbb{R}^s$.

4- Suppose $s \geq 7$. The urn is large, $(X_n - nv_1)/n^{\cos(2\pi/s)}$ is bounded almost surely and in $\mathrm{L}^{\geq 1}$ and has the almost sure asymptotics

$$\frac{1}{n^{\cos(2\pi/s)}}(X_n - nv_1) = 2\Re\left(n^{i\sin(2\pi/s)} W v_2\right) + o(1)$$

where $v_2 = \frac{1}{s}{}^t(1, e_s^{-1}, e_s^{-2}, \ldots, e_s^{1-s})$ and W is a complex random variable; if one denotes $u_2(x_1, \ldots, x_s) = x_1 + e_s x_2 + \ldots + e_s^{s-1} x_s$, $\tau_2 = u_2(X_1)$, $u_4(x_1, \ldots, x_s) = x_1 + e_s^2 x_2 + \ldots + e_s^{2(s-1)} x_s$ and $\tau_4 = u_4(X_1)$, the first moments of W are $EW = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+e_s)}\tau_2$,

$$EW^2 = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+2e_s)}\left(\tau_2^2 + \frac{e_s}{2-e_s}\tau_4\right) \text{ and } E|W|^2 = \frac{\Gamma(\tau_1)}{\Gamma(\tau_1+2\cos(2\pi/s))}\left(|\tau_2|^2 - \frac{1}{1-2\cos(2\pi/s)}\tau_1\right).$$

PROOF. The urn is irreducible, imaginary and principally semisimple, the spectrum of A consisting in all s-th roots of unity. The s linear forms defined by $u_\zeta = \sum_{k=0}^{s-1} \zeta^k x_k$ for any s-th root of unity ζ constitute a Jordan basis, the two eigenforms having $\sigma_2 = \cos(2\pi/s)$ as real part being u_{e_s} and its conjugate. Points

2- and *3-* are thus shown by Janson (2004). One just has to compute suitable reduced polynomials to complete the whole proof. We omit this computation in the text because of its length. One can perform it carefully with much patience or do it with the help of symbolic computation (I did both!). We just mention here the expression of $\Phi(u_{\zeta_1}u_{\zeta_2}\dots u_{\zeta_r})$ for any choice of s-th roots of unity $\zeta_1,\zeta_2,\dots,\zeta_r$, starting point of the work.

Notations: if f is any function $\mathbb{C}^s\to\mathbb{C}$ and σ any permutation of $\{1,\dots,s\}$, we denote $\sigma.f:(x_1,\dots,x_s)\mapsto f(x_{\sigma(1)},\dots,x_{\sigma(s)})$ (group action on such functions), $\mathrm{Stab}(f)$ the subgroup $\{\tau\in\mathfrak{S}_s,\ \tau.f=f\}$ and

$$\sum_{sym} f=\frac{1}{|\,\mathrm{Stab}(f)|}\sum_{\sigma\in\mathfrak{S}_s}\sigma.f.$$

When f is any vector-valued function defined on $\mathbb{C}^s$ and ζ any s-th root of unity, we denote, for any $x\in\mathbb{C}^s$,

$$\Phi_\zeta(f)(x)=\sum_{k=1}^{s}\zeta^k x_k\left[f(x+w_k)-f(x)\right]$$

where as usual w_k is the k-th column-vector of $A={}^tR$. Note that $\Phi_1=\Phi$, transition operator of the cyclic urn. This notations being adopted, one shows that $\Phi_\zeta(fu_{\zeta'})=\Phi_\zeta(f)u_{\zeta'}+\zeta' fu_{\zeta\zeta'}+\zeta'\Phi_{\zeta\zeta'}(f)$. In particular, $\Phi_\zeta(u_{\zeta'})=\zeta' u_{\zeta\zeta'}$. This is enough to show by induction on r that, for any choice of s-th roots of unity $\zeta_1,\dots,\zeta_r$,

$$\Phi_\zeta(u_{\zeta_1}\dots u_{\zeta_r})=\sum_{k=1}^{r}\sum_{sym}\zeta_1\dots\zeta_k u_{\zeta_1\dots\zeta_k\zeta}u_{\zeta_{k+1}}\dots u_{\zeta_r}.$$

This formula leads to the computation of the reduced polynomials. For example, if ζ_1 and ζ_2 are s-th roots of unity, if one denotes $\delta_{e_s^k}=\delta_{k+1}$ for any $k\in\{0,\dots,s-1\}$,

$$Q_{\delta_{\zeta_1}+\delta_{\zeta_2}}=u_{\zeta_1}u_{\zeta_2}+\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2-\zeta_1\zeta_2}u_{\zeta_1\zeta_2},\tag{11}$$

this formula being valid as soon as $\zeta_1\neq\overline{\zeta_2}$ or $\zeta_1\neq\exp(\pm 2i\pi/6)$; if (and only if) this condition is not satisfied, the above denominator vanishes and $Q_{\delta_{\zeta_1}+\delta_{\zeta_2}}=u_{\zeta_1}u_{\zeta_2}$ (but is not eigenvalue of the operator Φ any more). This formula is enough to compute the second order moments of *4-*. □

Remark 1 If the initial composition of the urn consists in only one ball of any colour, then $E|W|^2=(1+1/(2\cos(2\pi/s)-1))/\Gamma(1+2\cos(2\pi/s))$ tends to 1 as s tends to infinity. If the initial composition consists in one ball of each colour, then $E|W|^2=s!/(2\cos(2\pi/s)-1)/\Gamma(s+2\cos(2\pi/s))$ is equivalent to $1/s$ as s tends to infinity.

Remark 2 (Variance of $|W|^2$) The computation of the variance of $|W|^2$ for large cyclic urns lets the value $s=12$ appear as exceptional. Indeed, suppose that $s\geq 7$ and $s\neq 12$, and denote $c=\cos(2\pi/s)$. Then, if (for example) the initial composition of the urn consists in only one ball of any colour, then

$$E|W|^4=\frac{1}{\Gamma(4c^2-1)}\frac{8(8c^3-20c^2+c+2)}{(4c-1)(4c-5)(2c-1)^2}$$

and if this initial composition consists in one ball of each colour, then

$$E|W|^4=\frac{s!}{\Gamma(s+4c^2-2)}\frac{2(16sc^3+8c^3-20c^2-24sc^2+5sc+c+2)}{c(4c-1)(4c-5)(2c-1)^2}.$$

Suppose now that $s=12$. Then the above formulae of are not valid any more; if the initial composition of the urn consists in only one ball of any colour, then $E|W|^4=\frac{64c^3-156c^2+4c+17}{(4c-1)(4c-5)(2c-1)^2}=\frac{1901}{143}+\frac{1140}{143}\sqrt{3}$ and if this initial composition consists in one ball of each colour, then $E|W|^4=\frac{404c^3-620c^2+123c+4}{c(4c-1)(4c-5)(2c-1)^2}=\frac{307}{11}+\frac{467}{33}\sqrt{3}$.

The exceptional value $s=12$ can be pointed out for reduced polynomials of degree two that appear in the computation in the fourth order moments of W: formula (11) implies that $Q_{\delta_3+\delta_{s-3}}=|u_{e_s^2}|^2+u_1/(2\cos(4\pi/s)-1)$ if $s\neq 12$ and $Q_{\delta_3+\delta_{s-3}}=|u_{e_s^2}|^2$ if $s=12$. We do not say more about the computations that lead to these formulae.

Remark 3 Some authors would talk about "phase change" for $s = 7$ because of the type of asymptotics of the renormalized process. Some moments of order ≥ 2 have an exceptional behaviour for $s = 12$ as can be seen in Remark 2 (for a moment of order four). This appears as a technical reason in the computation of reduced polynomials, whose coefficients are fractions in e_s with cyclotomic factors at their denominators. Computations of moments of higher orders give rise to similar phenomena for various exceptional values of s. The natural question of the law of W (or of its real and imaginary parts, or even of its module's square) remains open.

4.3 m-ary search trees

An m-ary search tree can be seen as a Pólya-Eggenberger irreducible, semisimple and imaginary urn process with $m-1$ colours and $X_1 = {}^t(1,0,\dots,0)$ as initial composition (see Chauvin and Pouyanne (2004) for the replacement matrix). It is well known that this urn is large if and only if $m \geq 27$ (see Mahmoud (1992), Chern and Hwang (2001)). We assume that $m \geq 27$.

Eigenvalues of A are the roots of $\prod_{1\leq k\leq m-1}(z+k) - m!$. For any $\lambda \in \mathrm{Sp}(A)$ we denote

$$\gamma_k(\lambda) = \prod_{1\leq j\leq k-1}(1+\lambda/j) \text{ and } H_m(\lambda) = \sum_{1\leq k\leq m-1}(k+\lambda)^{-1}.$$

We choose a Jordan basis $(u_\lambda)_{\lambda\in\mathrm{Sp}(A)}$ and its dual basis $(v_\lambda)_\lambda$, derived from computations of the article with B. Chauvin: for any $\lambda \in \mathrm{Sp}(A)$, $u_\lambda = \sum_{1\leq k\leq m-1}\frac{1}{k}\gamma_k(\lambda)x_k$ and $v_\lambda = \frac{1}{H_m(\lambda)}\,{}^t(1/\gamma_2(\lambda),\dots,1/\gamma_m(\lambda))$.

We denote by λ_2 the non real eigenvalue of A having the largest real part (namely σ_2) and a positive imaginary part (named λ_2''). It follows from Chauvin and Pouyanne (2004) or from Theorem 1 that there exists a complex random variable W such that

$$\frac{1}{n^{\sigma_2}}\left(X_n - nv_1\right) = 2\Re\left(n^{i\lambda_2''}Wv_{\lambda_2}\right) + o(1).$$

Theorem 5 *With the convention* $\frac{\Gamma(\mu+1)}{\Gamma(\mu+k+1)} = \frac{(-1)^k}{k!\binom{m}{k}}$ *if* $\mu = -m-1$ *(when m is odd, $-m-1$ is an eigenvalue of A), the random variable W has the following first polynomial moments:* $EW = 1/\Gamma(1+\lambda_2)$,

$$EW^2 = \frac{1}{\Gamma(1+2\lambda_2)\Gamma(1+\lambda_2)^2}\left(1 + \sum_{\substack{\mu\in\mathrm{Sp}(A)\\ 1\leq k\leq m-1}}\frac{1}{(2\lambda_2-\mu)H_m(\mu)}\frac{1}{k!}\frac{\Gamma(\lambda_2+k)^2}{\Gamma(\lambda_2)^2}\frac{\Gamma(\mu+1)}{\Gamma(\mu+1+k)}\right),$$

$$\text{and } E|W^2| = \frac{1}{\Gamma(1+2\sigma_2)|\Gamma(1+\lambda_2)|^2}\left(1 + \sum_{\substack{\mu\in\mathrm{Sp}(A)\\ 1\leq k\leq m-1}}\frac{1}{(2\sigma_2-\mu)H_m(\mu)}\frac{1}{k!}\frac{|\Gamma(\lambda_2+k)|^2}{|\Gamma(\lambda_2)|^2}\frac{\Gamma(\mu+1)}{\Gamma(\mu+1+k)}\right). \tag{12}$$

PROOF. Let w_k be the k-th column-vector of A. Namely, $w_k = -k\delta_k + (k+1)\delta_{k+1}$ if $k \leq m-2$ and $w_{m-1} = -(m-1)\delta_{m-1} + m\delta_1$. One has $u_\lambda(w_k) = \frac{\lambda}{k}\gamma_k(\lambda)$ for any eigenvalue λ and for any $k \in \{1,\dots,m-1\}$. This leads to the computation of the linear form $\Phi(u_\lambda u_\mu) - (\lambda+\mu)u_\lambda u_\mu$, first in the x_k coordinates, then in the u_ν ones for any eigenvalues λ and μ:

$$\begin{aligned}\Phi(u_\lambda u_\mu) - (\lambda+\mu)u_\lambda u_\mu &= \textstyle\sum_{k=1}^{m-1} x_k\frac{\lambda}{k}\gamma_k(\lambda)\frac{\mu}{k}\gamma_k(\mu)\\ &= \textstyle\sum_{\nu\in\mathrm{Sp}(A)}\left[\sum_{1\leq k\leq m-1}\frac{\lambda\mu}{k^2}\frac{\gamma_k(\lambda)\gamma_k(\mu)}{\gamma_{k+1}(\nu)H_m(\nu)}\right]u_\nu.\end{aligned} \tag{13}$$

This leads to the result, the coefficient of u_ν in the expansion of the second order reduced polynomials corresponding to the indices λ and μ being the above coefficients in brackets [.] divided by $\lambda+\mu-\nu$. □

One can compute these moments for various values of $m \geq 27$. As example of application of this result, if $X_n^{(2)}$ denotes the number of nodes that contain one key after insertion of the $(n-1)$-st key in an m-ary search tree (see Chauvin and Pouyanne (2004)), then this random variable satisfies almost surely

$$X_n^{(2)} = \frac{2n}{3H_m(1)} + n^{\sigma_2}\rho^{(2)}\cos(\lambda_2''\log n + \psi^{(2)}) + o(n^{\sigma_2})$$

as n tends to infinity, where $\rho^{(2)}$ and $\psi^{(2)}$ are real random variables. We give numeric approximations of σ_2, $H_m(1)$ and $E(\rho^{(2)})^2$ for some values of m:

$m =$	27	28	29	30	40	50
$\sigma_2 \sim$	0.517	0.533	0.549	0.563	0.662	0.720
$\frac{2}{3H_m(1)} \sim$	0.231	0.228	0.225	0.223	0.203	0.191
$E(\rho^{(2)})^2 \sim$	44.06	43.32	42.62	41.96	36.95	33.66

References

K. Athreya and S. Karlin. Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Stat.*, 39:1801–1817, 1968.

A. Bagchi and A. Pal. Asymptotic normality in the generalized Polya-Eggenberger urn model, with an application to computer data structures. *SIAM J. Algebraic Discrete Methods*, 6:394–405, 1985.

B. Chauvin and N. Pouyanne. m-ary search trees when $m \geq 27$: A strong asymptotics for the space requirements. *Random Struct. Algorithms*, 24(2):133–154, 2004.

H.-H. Chern and H.-K. Hwang. Phase changes in random m-ary search trees and generalized quicksort. *Random Struct. Algorithms*, 19(3-4):316–358, 2001.

P. Flajolet, J. Gabarró, and H. Pekari. Analytic urns. *Annals of Probability*, 2005. To appear.

R. Gouet. Strong convergence of proportions in a multicolor Pólya urn. *J. Appl. Probab.*, 34(2):426–435, 1997.

S. Janson. Functional limit theorem for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.*, 110:177–245, 2004.

S. Janson. Limit theorems for triangular urn schemes. *Probability Theory and Related Fields*, 2005. To appear.

H. M. Mahmoud. *Evolution of random search trees.* John Wiley & Sons, 1992.

G. Pólya. Sur quelques points de la théorie des probabilités. *Ann. Inst. Poincaré*, 1:117–161, 1930.

N. Pouyanne. An algebraic approach of Pólya processes. *Submitted*, 2005. 40 pages, available at `http://www.math.uvsq.fr/~pouyanne/`.

V. Puyhaubert. Modèles d'urnes et phénomènes de seuils en combinatoire analytique. *Thèse de l'Ecole Polytechnique*, 2005. available at `http://algo.inria.fr/puyhaubert/`.