



**HAL**  
open science

## Fully Analyzing an Algebraic Polya Urn Model

Basile Morcrette

► **To cite this version:**

Basile Morcrette. Fully Analyzing an Algebraic Polya Urn Model. LATIN 2012 : 10th Latin American Theoretical INformatics Symposium, Apr 2012, Arequipa, Peru. pp.568-581, 10.1007/978-3-642-29344-3\_48 . hal-00675936

**HAL Id: hal-00675936**

**<https://hal.science/hal-00675936>**

Submitted on 21 Mar 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Fully Analyzing an Algebraic Pólya Urn Model

Basile Morcrette

ALGORITHMS Project, INRIA Paris-Rocquencourt, 78153 Le Chesnay (France)  
LIP6, Université Paris 6, 4 place Jussieu, 75005 Paris (France)  
`Basile.Morcrette@inria.fr`

**Abstract.** This paper introduces and analyzes a particular class of Pólya urns: balls are of two colors, can only be added (the urns are said to be *additive*) and at every step the same constant number of balls is added, thus only the color compositions varies (the urns are said to be *balanced*). These properties make this class of urns ideally suited for analysis from an “analytic combinatorics” point-of-view, following in the footsteps of Flajolet et al. [4]. Through an *algebraic* generating function to which we apply a multiple coalescing saddle-point method, we are able to give precise asymptotic results for the probability distribution of the composition of the urn, as well as local limit law and large deviation bounds.

**Keywords:** analytic combinatorics, Pólya urn models, multiple coalescing saddle-point method, Gaussian local limit law, large deviations.

*Dedicated to the memory of Philippe Flajolet.*

## 1 Introduction

A Pólya urn is an urn which contains balls of two colors (black and white), and which is coupled with an initial configuration and a set of evolution rules. A step then consists in randomly picking a ball from the urn, placing it back, and depending on its color, adding a fixed number of black and/or white balls. The question is: what does the urn look like after a large number of steps? This simple process has turned out to be extremely versatile, and has been used to model many different phenomena, such as population growth, epidemics, tree structures in computer science (BST,  $(a, b)$ -trees), electoral campaigns, etc.

This paper analyzes a class of balanced additive urns. *Balanced* urns are urns for which, at every step, the same constant number of balls is added. This property allows us to resort to a combinatorial treatment: enumerating all configurations using generating functions. Such an approach, introduced by Flajolet and his coauthors [4], is a departure from previous probabilistic methods. In *additive* urns, no ball is ever removed and the urn’s size is strictly increasing. We specifically consider a class of balanced additive urns which has *algebraic* generating functions.

Through the use of analytic combinatorics [6], we obtain precise probability results, including a Gaussian limit law with rate of convergence, a local limit law and large deviation bounds. Our analysis makes use of a multiple coalescing saddle-point method, which is not classical. As previously thought, analytic combinatorial methods can provide a wealth of valuable information that seems to be new. In addition, our results seem to reinforce the idea of a general theory for additive balanced urns.

This paper comes as the continuation of Flajolet’s work on urns [5] and [4]; these two papers were on subtractive and triangular urn models. Of course, this topic has been thoroughly addressed through probabilistic means. We can cite introductory books on urns [11] [13] and an article [1]; on the topic of limit distribution for urns, Janson’s papers are a reference [8] [9] [10] as is Smythe’s [16]; and on the topic of large additive urns, [3]. From an analytic point of view, Hwang et al.’s paper [7] considers “*diminishing*” urn models. And in direct relation to the present paper, we have obtained similar results [14] on a class of additive balanced urns which is linked to some family of  $k$ -trees [15].

In Sect. 2, we lay the ground work by introducing Pólya urn models as well as the main result of [4] on the isomorphism between balanced urns and differential systems. Then, in Sect. 3, we introduce our urn class, which can be viewed as a population growth model. We establish that the generating function is algebraic and provide our first results on mean and variance. Section 4 contains the main theorems of the paper regarding the limit distribution of our urn class. Finally we state the propositions and lemmas required by the proof of the main theorems, and give some extensive details on the part involving multiple coalescing saddle-points.

## 2 Urns and Differential Systems

In Pólya’s classical urn model, we have an urn containing balls of two different colors, black balls (**b** type) and white balls (**w** type). This system evolves with regards to particular rules (at each step, add and/or discard black and/or white balls), and these rules are specified by a  $2 \times 2$  matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad a, d \in \mathbb{Z}, \quad b, c \in \mathbb{Z}_{\geq 0}. \quad (1)$$

We start with an initial configuration  $(a_0, b_0)$ . At step 0, the urn contains  $a_0$  black balls and  $b_0$  white balls. The evolution between steps  $n$  and  $n + 1$  is now described. We uniformly draw a ball from the urn, we look at its color and *we put it back into the urn*. If the color is black, then we add  $a$  black balls and  $b$  white balls; if the color is white, we add  $c$  black balls and  $d$  white balls.

**Definition 1.** *The urn (1) is said to be balanced if the sums of its rows are constant, that is if  $a + b = c + d$ . This parameter is called the balance of the urn, and denoted by  $\sigma$ .*

*Remark 1.* At each step, we add  $\sigma$  balls in the urn. So, starting with  $a_0 + b_0$  balls in a balanced urn, we know that the total number of balls after  $n$  steps will be  $a_0 + b_0 + \sigma n$ . This balanced condition is the key requirement for using Flajolet-Dumas-Puyhaubert differential systems.

**Definition 2.** *An urn is additive if all the coefficients in its rule matrix (1) are strictly positive, i.e.,  $a, b, c, d > 0$ .*

*Remark 2.* No balls are ever removed from an additive urn. In our study, we focus on *balanced additive* urns.

The papers [5] and [4] introduce a new analytical and combinatorial approach to the study of these balanced urn models. The crucial starting point is an isomorphism theorem between the rules of an urn and a differential system. We briefly recall these results.

**Definition 3.** *A history of length  $n$  is a sequence of  $n$  steps, obtained by  $n$  successive draws from the urn. The exponential generating function of histories of the urn (1) is*

$$H(x, y, z; a_0, b_0) = \sum_{n,i,j} H_n(a_0, b_0; i, j) x^i y^j \frac{z^n}{n!}, \quad (2)$$

where  $H_n(a_0, b_0; i, j)$  is the number of histories of length  $n$  starting at the configuration  $(a_0, b_0)$ , and ending at a configuration  $(i, j)$ . We will shorten this to  $H(x, y, z)$  when there is no ambiguity on the initial configuration  $(a_0, b_0)$ .

**Theorem 1 (Flajolet–Dumas–Puyhaubert).** *We associate to the urn (1), with the balanced condition  $a + b = c + d$ , the following differential system (we denote  $t$ -differentiation by a point,  $\dot{X}(t) = \frac{d}{dt}X(t)$ ):*

$$\begin{cases} \dot{X} = X^{a+1} Y^b \\ \dot{Y} = X^c Y^{d+1} . \end{cases}$$

Let  $x_0$  and  $y_0$  be two complex variables such that  $x_0 y_0 \neq 0$ . Let  $X(x_0, y_0, t)$  and  $Y(x_0, y_0, t)$  be the solutions of the differential system with initial conditions  $X(x_0, y_0, 0) = x_0$ ,  $Y(x_0, y_0, 0) = y_0$ . Then the generating function of histories is given by

$$H(x_0, y_0, z) = X(x_0, y_0, z)^{a_0} Y(x_0, y_0, z)^{b_0}. \quad (3)$$

### 3 Preferential Growth Urns

As previously mentioned, [5] has signaled that analytic methods can be used in the treatment of urn models. Since then, a general theory has been described for several urn models: for two-color balanced subtractive models and triangular

models [4], as well as for some particular unbalanced models [7]. The extension of these methods for all additive urn models is an open problem. This paper presents a full asymptotic study for some restriction of the balanced additive models. More specifically, here we present a two-parameter urn model, whereas the entire class of balanced additive models is described with three parameters. Indeed, knowing the four matrix coefficients and the balance hypothesis, one parameter is redundant. Our class is studied through its histories generating function, which is in fact *algebraic* in our case.

**Definition 4.** *The class of preferential growth urns denoted by  $\mathcal{A}(\alpha, \beta)$  is defined by the two-color balanced matrix*

$$\begin{pmatrix} 2\alpha & \beta \\ \alpha & \alpha + \beta \end{pmatrix}, \text{ with } \alpha > 0, \beta > 0. \quad (4)$$

*Example 1.* A first example  $\mathcal{A}(1, 1)$  corresponds to the following evolution rules:

$$\begin{array}{l} \mathbf{b} \rightarrow \mathbf{b} \mathbf{b} \mathbf{b} \mathbf{w} \\ \mathbf{w} \rightarrow \mathbf{b} \mathbf{w} \mathbf{w} \mathbf{w} \end{array} \quad \text{corresponding to the urn } \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

If we pick a  $\mathbf{b}$  ball, we replace it by three  $\mathbf{b}$  (counting the ball we drew and are now replacing) and one  $\mathbf{w}$ . If we pick a  $\mathbf{w}$  ball, we replace it by one  $\mathbf{b}$  and three  $\mathbf{w}$ . In this particular example, we can see a model of population growth with two types of individuals ( $\mathbf{b}$  and  $\mathbf{w}$ ). Every individual has three children and two of them are of the same type as their parent. Every individual encourages its own type with the ratio 2/1. From this example, we choose to name our class *preferential growth urns*.

*Remark 3.* Here are some characteristics of this urn class. The balance of  $\mathcal{A}(\alpha, \beta)$  is  $\sigma := 2\alpha + \beta$ . The dissymmetry index is defined by  $p := \alpha - 2\alpha = \beta - (\alpha + \beta) = -\alpha$ . The two eigenvalues of the matrix are  $-p$  and  $\sigma$ . Besides, the ratio  $\rho$  of these two eigenvalues is  $\rho := \frac{-p}{\sigma} = \frac{\alpha}{2\alpha + \beta} \leq \frac{1}{2}$ . It is a small urn so, from the probabilistic analyses of Smythe [16] and Janson [8], we already have a Gaussian behavior for the limiting distribution of the two colored balls in the urn.

In this paper, thanks to analytic combinatorics, first we obtain concrete and precise asymptotic results with rate of convergence, local limit laws and probabilities of large deviations; second, the analytical proof deals with a multiple coalescing saddle-point method; and third, this is the first step towards a full study of additive balanced urns.

### 3.1 An Algebraic Generating Function

We start the urn process  $\mathcal{A}(\alpha, \beta)$  with no black balls  $\mathbf{b}$ , and one white ball  $\mathbf{w}$ . So, in the following,  $(a_0, b_0) = (0, 1)$ . The main result of this subsection is exhibiting the algebraic nature of the generating function of histories.

**Theorem 2.** *The bivariate generating function of histories  $H(x, z) := H(x, 1, z)$  is algebraic, and the following polynomial in  $y$  cancels it.<sup>1</sup>*

$$(z - A - B_x) y^{2\alpha+\beta} + B_x y^\alpha + A = 0, \quad (5)$$

$$\text{where } A := \frac{1}{2\alpha + \beta} = \sigma^{-1} \quad \text{and} \quad B_x := \frac{x^{-\alpha} - 1}{\alpha + \beta}.$$

*Proof.* The exponential generating function of histories is

$$H(x, y, z) = \sum_{n,i,j} H_n(i, j) x^i y^j \frac{z^n}{n!},$$

and the differential system associated to the urn is

$$\begin{cases} \dot{X} = X^{2\alpha+1} Y^\beta \\ \dot{Y} = X^\alpha Y^{\alpha+\beta+1}. \end{cases}$$

Let  $(X(x_0, y_0, t), Y(x_0, y_0, t))$  be the associated solution, then thanks to Theorem 1, and  $a_0 = 0$ ,  $b_0 = 1$ , we get  $H(x_0, y_0, z) = Y(x_0, y_0, z)$ . Rewriting the system,

$$\frac{\dot{X}}{X^{\alpha+1}} = X^\alpha Y^\beta = \frac{\dot{Y}}{Y^{\alpha+1}}, \quad \text{so,} \quad X^{-\alpha} - Y^{-\alpha} = x_0^{-\alpha} - y_0^{-\alpha}.$$

Then,

$$\frac{\dot{Y}}{Y^{\alpha+\beta+1}} (Y^{-\alpha} + x_0^{-\alpha} - y_0^{-\alpha}) = 1.$$

After a  $z$ -integration and naming the integration constant  $K(x_0, y_0)$ ,

$$\frac{1}{2\alpha + \beta} Y^{-(2\alpha+\beta)} + \frac{x_0^{-\alpha} - y_0^{-\alpha}}{\alpha + \beta} Y^{-(\alpha+\beta)} = -(z - K(x_0, y_0)).$$

We can set  $y_0 = 1$  (the variable is useless because of the balanced condition. Indeed, in  $H$ , the non-negative coefficients of  $x^i y^j z^n$  appear when  $i+j = a_0+b_0+\sigma n$ ), then we get the result. Thus, the generating function  $H(x, z) := H(x, 1, z)$  is algebraic, as solution of a polynomial of degree  $2\alpha + \beta$ .  $\square$

### 3.2 Mean and Variance

From the polynomial equation (5), we can directly extract some precise information concerning the expression of the mean and the variance of the urn composition. It is also possible to get all moments.

<sup>1</sup> In this paper, we only treat the case  $(a_0, b_0) = (0, 1)$ . The general case  $H = X^{a_0} Y^{b_0}$  derives directly from this study, because  $X$  and  $Y$  are bound:  $H(x, y, z) = (x^{-\alpha} - y^{-\alpha} + Y^{-\alpha})^{-a_0/\alpha} Y^{b_0}$ . We focus here on the function  $Y$ .

**Proposition 1.** *Let  $X_n$  be the random variable counting the number of black balls  $\mathbf{b}$  in the urn (4) at step  $n$ . Then we can express mean and variance asymptotically:<sup>2</sup>*

$$\mathbb{E}(X_n) = \frac{\alpha(2\alpha + \beta)}{\alpha + \beta}n + \frac{\alpha}{\alpha + \beta} \frac{\Gamma(\frac{1}{2\alpha + \beta})}{\Gamma(\frac{\alpha + 1}{2\alpha + \beta})} n^{\frac{\alpha}{2\alpha + \beta}} + \frac{\alpha}{\alpha + \beta} + O\left(n^{\frac{\alpha}{2\alpha + \beta} - 1}\right), \quad (6)$$

$$\mathbb{V}(X_n) = \frac{\alpha^3(2\alpha + \beta)}{(\alpha + \beta)^2}n + O\left(n^{\frac{\alpha + \beta}{2\alpha + \beta}}\right) \quad \text{where } \Gamma(x) := \int_0^\infty t^{x-1} e^{-t} dt. \quad (7)$$

*Proof.* The following techniques are described in [6]. We differentiate the equation (5) with regards to  $x$ , then we set  $x = 1$ ; we use the asymptotic expansion form and normalize by the asymptotic expansion of  $(1 - \sigma z)^{-1/\sigma} =: H(1, 1, z)$ . For variance, we differentiate twice and we use the same technique. It can be done for all moments. Besides, it is possible to express any term in the asymptotic expansion, thanks to Dr. Salvy's MAPLE package `gfun`.<sup>3</sup>  $\square$

## 4 Asymptotic Results

The aim of this section is to obtain properties on the limit distribution of the number of black balls in the urn. Our combinatorial point of view implies the study of  $z$ -coefficients of the generating function  $H(x, z)$ .

### 4.1 Probability Expression and the Case $x = 1$

**Lemma 1.** *The  $z^n$ -coefficient in the generating function of histories  $H(x, z)$  can be expressed by a contour integral (with a contour around the origin):*

$$[z^n]H(x, z) = \frac{\sigma^{n+1}}{2i\pi} \oint a_x(w) h_x(w)^{n+1} dw, \quad \text{where} \quad (8)$$

$$h_x(w) = \left[ 1 + \frac{\sigma(x^{-\alpha} - 1)}{\alpha + \beta} - (1 - w)^{\alpha + \beta} \left( \frac{\sigma(x^{-\alpha} - 1)}{\alpha + \beta} + (1 - w)^\alpha \right) \right]^{-1}, \quad (9)$$

$$a_x(w) = (1 - w)^{\alpha + \beta - 2} (x^{-\alpha} - 1 + (1 - w)^\alpha). \quad (10)$$

*Proof.* We start with the polynomial (5) which cancels  $y(x, z) := H(x, z)$ . We use a Cauchy formula to express the  $z^n$ -coefficient,

$$[z^n]H(x, z) = y_n(x) = \frac{1}{2i\pi} \oint \frac{y(x, z)}{z^{n+1}} dz. \quad (11)$$

<sup>2</sup> *Exact* expressions for mean and variance are also computable from this, see [6] (Prop.III.2, p. 158 and 728).

<sup>3</sup> available on <http://algo.inria.fr/libraries/>.

Then, by a Lagrange inversion<sup>4</sup> between  $z$  and  $y(z)$ , using the equation (5),

$$z = \sigma^{-1} + B_x - y^{-\sigma} (\sigma^{-1} + B_x y^\alpha) . \quad (12)$$

At this step, it is not yet possible to apply a saddle-point method, because the saddle-points are at  $\infty$ . Again, we change variables, and set  $w = 1 - y^{-1}$  to get the result.  $\square$

From this lemma, we have to evaluate this integral in order to have an expression of  $[z^n]H(x, z)$ . It will be necessary for describing the limit probability distribution. Indeed, our main goal is to understand the behavior of the probability generating function noted  $p_n(x)$ ,

$$p_n(x) = \sum_{i \geq 0} \mathbb{P}(X_n = i) x^i = \frac{[z^n]H(x, z)}{[z^n]H(1, z)} . \quad (13)$$

First, we focus on the evaluation of  $[z^n]H(1, z)$ . In the special case  $x = 1$ , we can solve explicitly the equation (5). We have  $H(1, z) = (1 - \sigma z)^{-1/\sigma}$ . Thus, we know with classical analytic combinatorics that  $y_n(1) \sim n^{-\frac{\sigma-1}{\sigma}} \Gamma(\sigma^{-1})^{-1} \sigma^n$ . For  $x \neq 1$ , we don't have access to the explicit solution, therefore we use a saddle-point method for the general case. In the Appendix A, the simple case  $x = 1$  is treated with this method. It is useful to guess the right normalization for the variables and the contour for the general case. For illustration, Fig. 1 shows the behavior of  $h_1(w)$  and how to choose the contour.

## 4.2 Limit Theorems

Here is the main result of the paper. It is described in the three following theorems which concerns the asymptotic distribution of the balls in the urn. Proposition 2 expresses the asymptotic expansion of the probability generating function  $p_n(x)$ . The refined saddle-point analysis is the core of the proof of this proposition. Then, the proofs of the theorems are based on this proposition and on Quasi-Power theorems from H.-K.Hwang, recalled in [6] (p. 645, 696, 700).

Let  $X_n$  be the random variable counting the number of black balls in the urn  $\mathcal{A}(\alpha, \beta)$  after  $n$  steps.

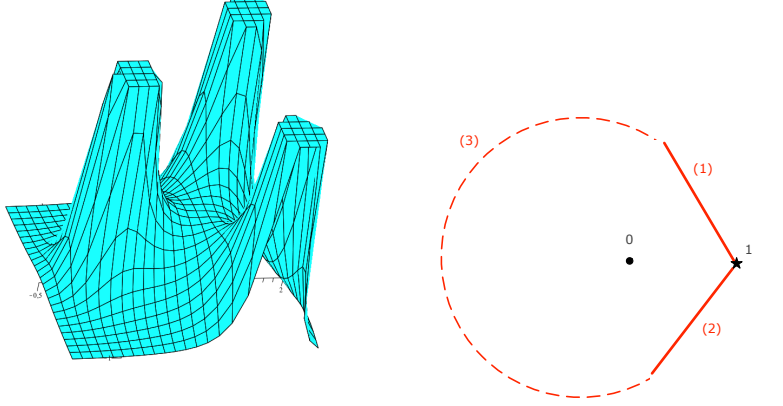
**Theorem 3.** (Gaussian limit law) *The random variable  $X_n$  has mean  $\mu_n$  and variance  $\nu_n^2$ , and the normalized random variable  $\frac{X_n - \mu_n}{\nu_n}$  converges in law to the standard normal law  $\mathcal{N}(0, 1)$ , with rate of convergence  $O\left(\frac{1}{\sqrt{n}}\right)$ ,*

$$\mu_n = \mathbb{E}(X_n) = \mu n + o(n) , \quad \nu_n^2 = \mathbb{V}(X_n) = \nu^2 n + o(n) ,$$

$$\mathbb{P} \left\{ \frac{X_n - \mu n}{\nu \sqrt{n}} \leq t \right\} = \Phi(t) + O\left(\frac{1}{\sqrt{n}}\right) , \quad \text{where } \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{v^2}{2}} dv ,$$

<sup>4</sup> For details on Lagrange inversion, see [6] (Appendix A.6, p. 732).





**Fig. 1.** *Left* - Double saddle-point for  $h_1(w)$  for the urn  $\mathcal{A}(1,1)$ . The three *peaks* correspond to zeros of the denominator of  $h_1(w)$  which are  $0$ ,  $1-j$  and  $1-j^2$ , where  $j$  is the third-root of unity  $\exp(2i\pi/3)$ . *Right* - Diagram from a top view of the three poles and the double saddle-point in  $1$ , with the three-parts contour  $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ .

with

$$\mu = \frac{\alpha(2\alpha + \beta)}{\alpha + \beta} \quad \text{and} \quad \nu^2 = \frac{\alpha^3(2\alpha + \beta)}{(\alpha + \beta)^2}.$$

**Theorem 4.** (*Local limit law*) We denote  $p_{n,k} = \mathbb{P}\{X_n = k\}$ . The distribution of  $X_n$  satisfies a local limit law of Gaussian type with rate of convergence  $O\left(\frac{1}{\sqrt{n}}\right)$ , i.e.

$$\sup_{t \in \mathbb{R}} \left| \nu \sqrt{n} p_{n, \lfloor \mu n + t\nu\sqrt{n} \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leq \frac{1}{\sqrt{n}}.$$

The last theorem concerns large deviation bounds.

**Definition 5.** (*Large deviations property*) Let  $(a_n)$  be a sequence tending to infinity. A sequence of random variables  $(X_n)$  with mean  $\mathbb{E}X_n \sim \mu a_n$ , satisfies a large deviation property, relative to the interval  $[t_0, t_1]$  containing  $\mu$ , if there exists a function  $W(t)$  such as  $W(t) > 0$  for  $t \neq \mu$ , and for  $n$  large enough,

$$\forall t, t_0 < t < \mu, \frac{1}{a_n} \log \mathbb{P}(X_n \leq ta_n) = -W(t) + o(1) \quad (\text{left tail}),$$

$$\forall t, \mu < t < t_1, \frac{1}{a_n} \log \mathbb{P}(X_n \geq ta_n) = -W(t) + o(1) \quad (\text{right tail}),$$

$W(t)$  is called the rate function, and  $a_n$  is the scale factor.

**Theorem 5.** (*Large deviations*) For  $\xi$  such as  $0 < \xi < 1$ , the sequence of random variables  $(X_n)$  satisfies a large deviation property relative to the interval  $[t_0, t_1]$ , with a scale factor  $n$ , and a rate  $W(t)$ , with  $x_0 = \xi$ ,  $x_1 = 2 - \xi$ ,

$$t_0 = x_0 \frac{\chi'(x_0)}{\chi(x_0)} \quad \text{and} \quad t_1 = x_1 \frac{\chi'(x_1)}{\chi(x_1)},$$

$$W(t) = - \min_{x \in [x_0, x_1]} \log \left( \frac{\chi(x)}{x^t} \right), \quad \text{where } \chi(x) = x^\mu e^{\frac{\nu^2}{2} \ln(x)^2}.$$

The main proposition in the proof of the theorems is the asymptotic expression of the probability generating function  $p_n(x)$ , obtained by the division of  $y_n(x)$  by  $y_n(1)$ , where  $y_n(x)$  denotes the  $z^n$ -coefficient in  $H(x, z)$ .

**Proposition 2.** For  $x = e^{iu/\sqrt{n}}$ , the expression of the probability generating function  $p_n(x)$  is, asymptotically when  $n \rightarrow \infty$ ,

$$p_n(x) = \exp \left( \mu i u \sqrt{n} - \frac{\nu^2}{2} u^2 \right) \left( 1 + O \left( u n^{-\frac{\beta}{2(2\alpha+\beta)}} \right) \right), \quad (14)$$

with

$$\mu = \frac{\alpha(2\alpha + \beta)}{\alpha + \beta} \quad \text{and} \quad \nu^2 = \frac{\alpha^3(2\alpha + \beta)}{(\alpha + \beta)^2}.$$

*Remark 4.* This expression (14) of  $p_n(x)$  is valid for  $x$  in a neighborhood of 1, with the radius  $n^{-1/2}$  (which will give the Gaussian limit law); for  $x$  in a real segment centered in 1 of length  $n^{-1/2}$  (which will give large deviation); for  $x$  in the unit circle,  $|x| = 1$  (which will give a local limit law).

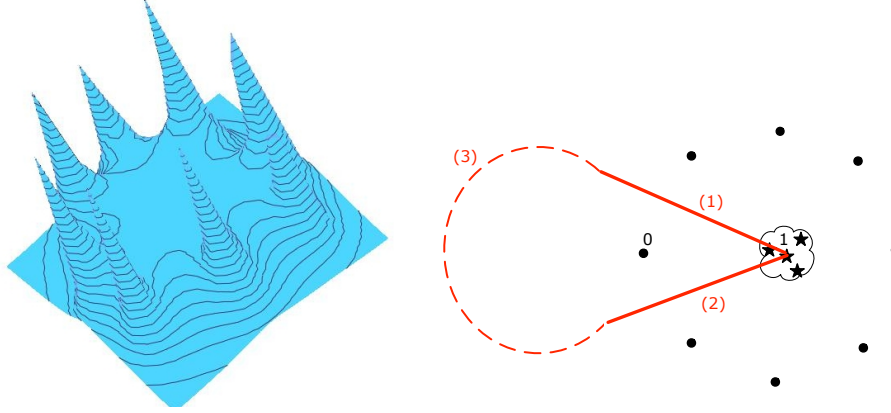
### 4.3 Proof of Proposition 2

**Proposition 3.** For all  $x$ ,  $h_x(w)$  has a simple pole in  $w = 0$ , and  $2\alpha + \beta - 1$  other simple poles. Saddle-points are in  $w = 1$  with multiplicity  $\alpha + \beta - 1$ , and in  $w = 1 - \gamma$ , with  $\gamma$  any  $\alpha$ th-root of  $(1 - x^{-\alpha})$ .

*Proof.* In order to find saddle-points, we look at zeros of the derivative of  $h_x$ ,

$$h'_x(w) = -\sigma h_x(w)^2 (1 - w)^{\alpha+\beta-1} (x^{-\alpha} - 1 + (1 - w)^\alpha).$$

*Remark 5.* In our case, when  $x \neq 1$ , we have  $\alpha + 1$  different saddle-points: one “huge” at  $w = 1$ , and  $\alpha$  secondary saddle-points depending on the  $x$  value. When  $x$  is in a neighborhood of 1, the  $\alpha$  saddle-points will merge at the point  $w = 1$ . We have to deal with  $\alpha + 1$  coalescing saddle-points. It differs from usual saddle-points method because it is no longer possible to treat each saddle-point independently. We need to consider all saddle-points together in order to find the asymptotic expansion.



**Fig. 2.** *Left* - Graph of  $|h_x(w)|$  for the urn  $\mathcal{A}(3,2)$ . There are  $\sigma = 8$  poles, and 4 saddle-points at point 1 and at  $1 - (1 - x^{-3})^{1/3}$ . *Right* - Multiple saddle-points for  $h_x(w)$ . The *peaks* (poles) corresponds to the zeros of the denominator of  $h_x(w)$ . The *stars* represent the different saddle-points. We illustrate *in red* the integration contour around 0.

We expect a Gaussian law (thanks to probabilistic results), so we use a renormalisation for  $x$  of order  $n^{-1/2}$ . For the renormalisation of  $w$ , we use the same scale as previously in the particular case  $x = 1$  (see Appendix A). So the normalizing factor is  $n^{-1/\sigma}$ .

The general idea for the choice of the contour is the same as in the previous simple case  $x = 1$ . We will use two segments beginning in the saddle-point area in  $w = 1$  and they have to follow the descents of the surface. This descent has to go between the pole 0 and the first closest pole. The segment will make an angle of  $\frac{\pi(\sigma-1)}{\sigma} = \pi - \frac{1}{2}\frac{2\pi}{\sigma}$ , as we can see on the right part of Fig. 2.

**Definition 6.** *The integration contour is described in three parts. We use the following parametrization for the segment  $\mathcal{C}_1$ . For  $t \in [0, n^2]$ ,*

$$w = 1 + \left(\frac{t}{n}\right)^{1/\sigma} e^{i\pi\frac{\sigma-1}{\sigma}}, \quad (15)$$

$$x = e^{iu/\sqrt{n}}. \quad (16)$$

*The same parametrization is used for the segment  $\mathcal{C}_2$ , with an angle of  $-\pi\frac{\sigma-1}{\sigma}$ . The circle part  $\mathcal{C}_3$  completes the contour.*

**Proposition 4.** *There is an asymptotic expansion in  $n$  for  $h_x(w)^n$ :*

$$h_x(w)^n = e^{\mu i u \sqrt{n} - \frac{\nu^2}{2} u^2} e^{-t} \exp\left(K t^{\frac{\alpha+\beta}{2(\alpha+\beta)}} u n^{\frac{-\beta}{2(\alpha+\beta)}} + O\left(n^{-1/2} (iu^3 + ut)\right)\right),$$

with

$$\mu = \frac{\alpha(2\alpha + \beta)}{\alpha + \beta}, \quad \nu^2 = \frac{\alpha^3(2\alpha + \beta)}{(\alpha + \beta)^2} \quad \text{and} \quad x = e^{iu/\sqrt{n}}.$$

*Remark 6.* Notice that this expansion is valid for all  $t$  and  $u$ . Besides, we obtain the same expansion using for  $x$  the expression  $x = 1 + iun^{-1/2} - u^2n^{-1}$ . Then, it will be possible to use this expansion for all  $|u| < 1$ , then we will have an expression for  $|x| < \frac{1}{\sqrt{n}}$  uniformly, which is a necessary hypothesis for Hwang's theorem on limit law. We will use  $x$  real near from 1 for large deviations theorem. The expression with  $x = e^{iu/\sqrt{n}}$  implies  $|x| = 1$  for the local limit law theorem.

**Lemma 2.** *[Neglect the tails] The  $\mathcal{C}_1$ -contribution is asymptotically equivalent on both intervals, for  $t \in [0, n^2]$  and for  $t \in [0, n^{\frac{1}{\sigma+1}}]$ . The error term is  $O\left(\exp\left(-n^{\frac{1}{\sigma+1}}\right)\right)$ , so it is exponentially negligible.*

*Proof.* On the path  $\mathcal{C}_1$ , the function  $h_x(t)$  is strictly decreasing for  $t > n^{\frac{1}{\sigma+1}}$ .  $\square$

**Lemma 3.** *[Central approximation] The  $\mathcal{C}_1$ -contribution can be written*

$$\frac{\sigma^n}{2i\pi} e^{i\pi/\sigma} n^{-\frac{\sigma-1}{\sigma}} \exp\left(\mu i u \sqrt{n} - \frac{\nu^2}{2} u^2\right) \int_0^{\frac{1}{\sigma+1}} t^{-1/\sigma} e^{-t} \left(1 + t^{\frac{\alpha+\beta}{\sigma}} O\left(un^{-\frac{\beta}{2\sigma}}\right)\right) dt.$$

*Proof.* We use Lemma 2 to restrict the interval. Then we can apply the asymptotic expansion of Proposition 4. Besides, we have to know an asymptotic expansion for  $a_x(w)$  on this path, and we have

$$a_x(t) \sim \left(\frac{t}{n}\right)^{(\sigma-2)/\sigma} e^{-i\pi(\sigma-2)/\sigma}.$$

**Lemma 4.** *[Complete the tails] The integral of  $t^{-1/\sigma} e^{-t}$  on segment  $[0, n^{\frac{1}{\sigma+1}}]$  is equivalent to the same integral on interval  $[0, \infty[$ .*

*The error term is  $O\left(\exp\left(-n^{\frac{1}{\sigma+1}}\right)\right)$ , so it is exponentially negligible.*

**Lemma 5.** *The  $\mathcal{C}_1$  and  $\mathcal{C}_2$  contribution can be written*

$$\frac{\sigma^n}{\Gamma\left(\frac{1}{\sigma}\right)} n^{\frac{1}{\sigma}-1} \exp\left(\mu i u \sqrt{n} - \frac{\nu^2}{2} u^2\right) \left(1 + O\left(un^{-\frac{\beta}{2\sigma}}\right)\right). \quad (17)$$

*Proof.* The work on segment  $\mathcal{C}_1$  is the same for segment  $\mathcal{C}_2$ . Adding the two contribution, we get a  $(e^{i\pi/\sigma} - e^{-i\pi/\sigma})$  factor. Thanks to Lemma 4, the integral part of calculus reduces to the simple Gamma factor  $\Gamma\left(\frac{\sigma-1}{\sigma}\right)$ .  $\square$

**Lemma 6.** *The circle part  $\mathcal{C}_3$  is exponentially negligible.*

Adding all the contributions, thanks to Lemmas 5 and 6, we obtain a global expression of  $y_n(x)$ ,

$$y_n(x) = \frac{\sigma^n}{\Gamma\left(\frac{1}{\sigma}\right)} n^{\frac{1}{\sigma}-1} \exp\left(\mu i u \sqrt{n} - \frac{\nu^2}{2} u^2\right) \left(1 + O\left(un^{-\frac{\beta}{2\sigma}}\right)\right). \quad (18)$$

We can express  $p_n(x)$  with  $x$  in a neighborhood of 1 with radius  $n^{-1/2}$ , and we obtain an expression suited to Quasi-Power Theorems,

$$p_n(x) = \left(x^\mu \exp\left(\frac{\nu^2}{2} \ln(x)^2\right)\right)^n \left(1 + O\left(n^{-\frac{\beta}{2\sigma}}\right)\right). \quad (19)$$

For the three main theorems, we use this expression which satisfies the hypothesis for applying H.-K.Hwang theorems: Quasi-Power Theorem, Quasi-Power Limit Theorem and Quasi-Power for large deviations (in [6], *p.645, 696, 700*).

## 5 Conclusion

Through the use of analytic combinatorics, we obtain precise probability results, including a Gaussian limit law with its rate of convergence, a local limit law and large deviation bounds. For the first time, this kind of probabilistic results on additive balanced urns are obtained from an analytic point of view. As previously thought, analytic combinatorial methods can provide a wealth of valuable information. In addition, our results seem to reinforce the idea of a general theory for additive balanced urns. In this way, the work on urns linked to  $k$ -trees in [14] use an other subclass of additive balanced urns, and similar probabilistic results were obtained. The natural next step is having a complete understanding of the whole class, following ideas of Flajolet et al. in [4].

**Acknowledgment.** I warmly thank my advisor and mentor, Philippe Flajolet, for guiding me and introducing me to research and especially to analytic combinatorics. This paper is dedicated to him. In addition, I'm grateful to Jérémie Lumbroso for his helpful remarks.

This work was supported by the ANR project 09 BLAN 0011 Boole and the ANR project 10 BLAN 0204 Magnum.

## References

1. BAGCHI, A., PAL, A. Asymptotic Normality in the Generalized Pólya-Eggenberger Urn Model, with an Application to Computer Data Structures. *SIAM Journal on Algebraic and Discrete Methods* **6**, No. 3, 394–405 (1985)
2. BANDERIER, C., FLAJOLET, P., SCHAEFFER, G., SORIA, M. Random Maps, Coalescing Saddles, Singularity Analysis, and Airy Phenomena. *Random Structures & Algorithms* **19**, No. 3/4, 194–246 (2001)

3. CHAUVIN, B., POUYANNE, N., SAHNOUN, R. Limit Distributions for Large Pólya Urns. *Annals of Applied Probability* **21**, No. 1, 1–32 (2011)
4. FLAJOLET, P., DUMAS, P., PUYHAUBERT, V. Some Exactly Solvable Models of Urn Process Theory. *Discrete Mathematics & Theoretical Computer Science Proceedings* **AG**, 59–118 (2006)
5. FLAJOLET, P., GABARRO, J., PEKARI, H. Analytic Urns. *Annals of Probability* **33**, No. 3, 1200–1233 (2005)
6. FLAJOLET, P., SEDGEWICK, R. *Analytic Combinatorics*. Cambridge University Press (2009)
7. HWANG, H.-K., KUBA, M., PANHOLZER, A. Analysis of Some Exactly Solvable Diminishing Urn Models. *In: 19th Formal Power Series and Algebraic Combinatorics*, Tianjin China (2007)
8. JANSON, S. Functional Limit Theorems for Multitype Branching Processes and Generalized Pólya Urns. *Stochastic Processes and their Applications* **110**, 177–245 (2004)
9. JANSON, S. Limit Theorems for Triangular Urn Schemes. *Probability Theory Related Fields* **134**, No. 3, 417–452 (2005)
10. JANSON, S. Plane Recursive Trees, Stirling Permutations and an Urn Model. *Discrete Mathematics & Theoretical Computer Science Proceedings* **AI**, 541–548 (2008)
11. JOHNSON, N. L., KOTZ, S. *Urn Models and Their Application*. John Wiley & Sons (1977)
12. MAHMOUD, H. Pólya Urn Models and Connections to Random Trees: A Review. *Journal of the Iranian Statistical Society* **2**, 53–114 (2003)
13. MAHMOUD, H. *Pólya Urn Models*. Chapman-Hall/CRC Press (2008)
14. MORCRETTE, B. Combinatoire analytique et modèles d’urnes, *Master’s thesis, MPRI - École Normale Supérieure de Cachan, INRIA Rocquencourt* (2010)
15. PANHOLZER, A., SEITZ, G. Ordered Increasing  $k$ -trees: Introduction and Analysis of a Preferential Attachment Network Model. *Discrete Mathematics & Theoretical Computer Science Proceedings* **AM**, 549–564 (2010)
16. SMYTHE, R. Central Limit Theorems for Urn Models. *Stochastic Processes and their Applications* **65**, 115–137 (1996)

## A Appendix - Saddle-point method when $x = 1$

We detail here the evaluation of the contour integral  $y_n(1)$  for the simple case  $x = 1$ ,

$$y_n(1) = \frac{\sigma^{n+1}}{2i\pi} \oint \frac{(1-w)^{\sigma-2} dw}{(1-(1-w)^\sigma)^{n+1}}. \quad (20)$$

Then,  $h_1(w) = (1 - (1-w)^\sigma)^{-1}$  and  $a_1(w) = (1-w)^{\sigma-2}$ . We have now to evaluate the integral. This is classical saddle-point method. The derivative  $h'_1$  becomes zero only for one value  $w = 1$ . There is only one saddle-point. We will use two segments starting at  $w = 1$  and descending along the surface, passing between the peaks (singular points of  $h_1$ ). The third part of the contour is a circle to join the two ends of segments. Figure 1 illustrates the three parts of the contour for the urn  $\mathcal{A}(1, 1)$ .

First we look at the parametrization of the two segments:

- $\mathcal{C}_1 : w = 1 + te^{i\pi\frac{\sigma-1}{\sigma}}$ , with  $t \in [0, L]$ ,
- $\mathcal{C}_2 : w = 1 + te^{-i\pi\frac{\sigma-1}{\sigma}}$ , with  $t \in [0, L]$ .

The two cases are similar. For  $\mathcal{C}_1$ , the integral rewrites

$$\frac{\sigma^{n+1}}{2i\pi} \left(-e^{-i\pi/\sigma}\right) \int_0^L -t^{\sigma-2} e^{2i\pi/\sigma} \left(\frac{1}{1+t^\sigma}\right)^{n+1} dt. \quad (21)$$

Adding the two contributions of the two paths  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , we get

$$\frac{\sigma^{n+1}}{\pi} \sin\left(\frac{\pi}{\sigma}\right) \int_0^L \frac{t^{\sigma-2} dt}{(1+t^\sigma)^{n+1}}. \quad (22)$$

The expansion  $h_1(t)^n = \exp(-nt^\sigma + O(nt^{2\sigma}))$  leads us to change variable with  $u = nt^\sigma$ . Thus, the integral rewrites

$$\int_0^L \frac{t^{\sigma-2} dt}{(1+t^\sigma)^{n+1}} = \frac{n^{-\frac{\sigma-1}{\sigma}}}{\sigma} \int_0^{nL^\sigma} u^{-1/\sigma} \exp(-u + O(u^2 n^{-1})) du. \quad (23)$$

We have to adjust the parameter  $L$  which is the length of each segment. In order to use the expansion approximation, some conditions impose themselves : first,  $nL^\sigma \rightarrow \infty$ , and second,  $nL^{2\sigma} \rightarrow 0$ . We decide to fix  $L \sim n^{\frac{-1}{\sigma+1}}$ . Thus,  $nL^\sigma = n^{\frac{1}{\sigma+1}}$  and  $nL^{2\sigma} = n^{-\frac{\sigma-1}{\sigma+1}}$ . In this case, we complete the tail of the Gaussian approximation,

$$\int_0^{n^{1/(\sigma+1)}} u^{-1/\sigma} e^{-u+O(u^2 n^{-1})} du \sim \int_0^{n^{1/(\sigma+1)}} u^{-1/\sigma} e^{-u} du + O(n^{-1}) \quad (24)$$

$$\sim \int_0^\infty u^{-1/\sigma} e^{-u} du + O(n^{-1}) \quad (25)$$

$$\sim \Gamma\left(\frac{\sigma-1}{\sigma}\right) + O(n^{-1}). \quad (26)$$

To conclude with this illustration example, we need to deal with the third part of the contour, which is the circle part. We show that this part is exponentially small. Indeed, it is possible to stretch the segments  $\mathcal{C}_1$  and  $\mathcal{C}_2$ . Thus, the radius of the circle part will grow to infinity and this integral part will be negligible. Besides, it is effectively possible to stretch the two segments because the function  $h_1$  is strictly decreasing along the segments. Its maximum is at the saddle-point. On  $[0, n^\sigma]$ , we have  $h_1'(t) = -\sigma t^{\sigma-1} (1+t^\sigma)^2 < 0$ . So, the evaluation of the integral on the segment  $[0, n^\sigma]$  is the same order as the evaluation on  $[0, n^{1/(\sigma+1)}]$ . Indeed, the majoration error is exponentially small. This error is about the order  $O(\exp(-n^{1/(\sigma+1)}))$ , which is negligible.

Gathering all parts, we obtain

$$y_n = \frac{\sigma^{n+1}}{\pi} \sin\left(\frac{\pi}{\sigma}\right) \frac{n^{-\frac{\sigma-1}{\sigma}}}{\sigma} \Gamma\left(\frac{\sigma-1}{\sigma}\right) (1 + O(n^{-1})) \quad (27)$$

$$= \frac{\sigma^n n^{1-1/\sigma}}{\Gamma(1/\sigma)} (1 + O(n^{-1})). \quad (28)$$