

# Fully Analyzing an Algebraic Pólya Urn Model

Basile Morcrette

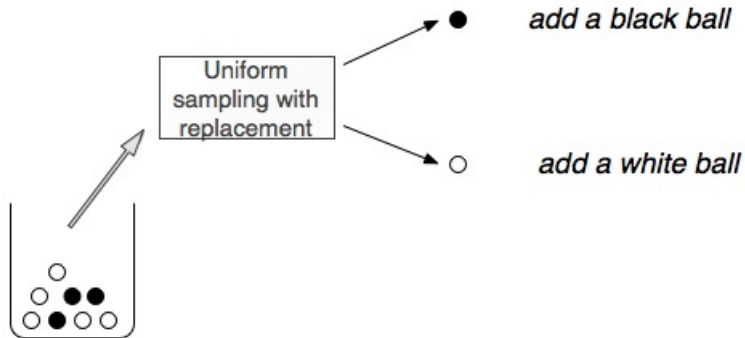
Algorithms project, INRIA Rocquencourt (*France*)  
*LIP6, UPMC*

LATIN, Arequipa (*Peru*)  
*April 19, 2012*

*Dedicated to Philippe Flajolet.*



# 1. Urns models



- ▶ an urn containing balls of two colours
- ▶ rules for urn evolution

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

# Balanced Pólya urns

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad \alpha, \delta \in \mathbb{Z}, \quad \beta, \gamma \in \mathbb{N}$$

**Balanced urn** :  $\boxed{\alpha + \beta = \gamma + \delta}$  (deterministic total number of balls)

A given initial configuration  $(a_0, b_0)$  :  
 $a_0$  balls  $\bullet$  (counted by  $x$ )  
 $b_0$  balls  $\circ$  (counted by  $y$ )

## Definition

**History of length  $n$**  : a sequence of  $n$  evolutions ( $n$  rules,  $n$  drawings)

$$H(x, y, z) = \sum_{n, a, b} H_{n, a, b} x^a y^b \frac{z^n}{n!}$$

$H_{n, a, b}$  : number of histories of length  $n$ , beginning in the configuration  $(a_0, b_0)$ , and ending in  $(a, b)$

## Combinatorics of histories - Example

We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .

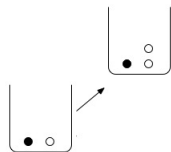


$$H(x, y, z) =$$

$xy$

## Combinatorics of histories - Example

We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .

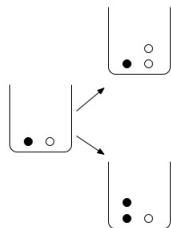


$$H(x, y, z) =$$

$xy$

## Combinatorics of histories - Example

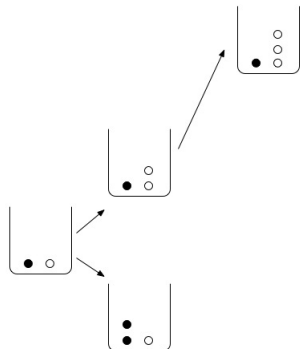
We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .



$$H(x, y, z) =$$
$$xy$$
$$+ (xy^2 + x^2y)z$$

## Combinatorics of histories - Example

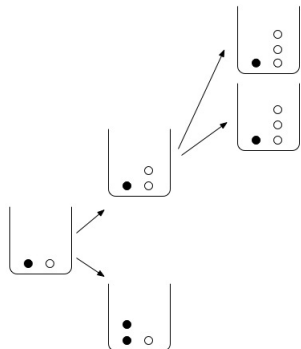
We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .



$$H(x, y, z) =$$
$$xy$$
$$+ (xy^2 + x^2y)z$$

## Combinatorics of histories - Example

We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .

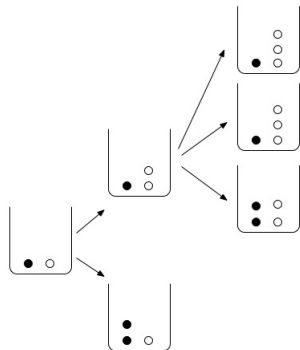


$$H(x, y, z) =$$
$$xy$$
$$+ (xy^2 + x^2y)z$$



## Combinatorics of histories - Example

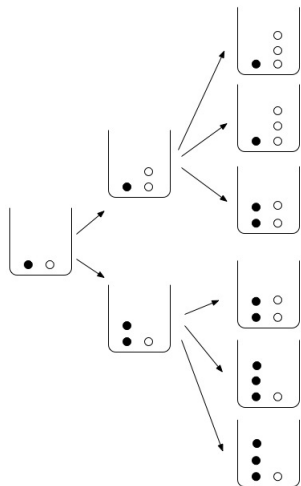
We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .



$$H(x, y, z) =$$
$$xy$$
$$+ (xy^2 + x^2y)z$$

## Combinatorics of histories - Example

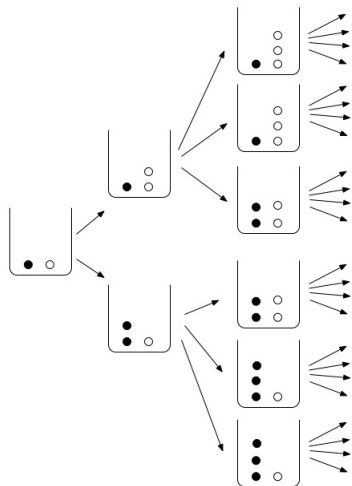
We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .



$$\begin{aligned}
 H(x, y, z) = & \\
 & xy \\
 + & (xy^2 + x^2y)z \\
 + & (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^2}{2}
 \end{aligned}$$

## Combinatorics of histories - Example

We consider this urn  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  with  $(a_0, b_0) = (1, 1)$ .



$$\begin{aligned}
 H(x, y, z) = & \\
 & xy \\
 + & (xy^2 + x^2y)z \\
 + & (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^2}{2} \\
 + & \dots
 \end{aligned}$$

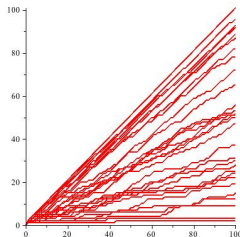
## Various behaviours

**Problem :** Understand the urn composition after  $n$  steps, and asymptotically when  $n$  tends to  $\infty$ .

# Various behaviours

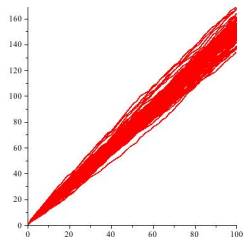
**Problem :** Understand the urn composition after  $n$  steps, and asymptotically when  $n$  tends to  $\infty$ .

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



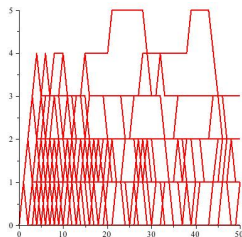
Pólya urn

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$



Preferential growth urn

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$



Triangular  $3 \times 3$  urn

# Probabilistic results

$$\text{Urn } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

$$\text{Ratio } \rho = \frac{\alpha - \gamma}{\alpha + \beta}$$

- ▶ **Small urns** :  $\rho \leq \frac{1}{2}$

**Gaussian** limit law [Smythe96] [Janson04]

- ▶ **Large urns** :  $\rho > \frac{1}{2}$

**Non gaussian** laws [Mahmoud] [Janson04]  
[Chauvin–Pouyanne–Sahnoun11]

## Tools :

- embedding in continuous time [Jan04] [ChPoSa11]
- martingales, central limit theorem

# Balanced urns and analysis

- ▶ First steps : [Flajolet–Gabarro–Pekari05], *Analytic urns*
- ▶ [Flajolet–Dumas–Puyhaubert06], on urns with negative coefficients, and triangular cases
- ▶ [Kuba–Panholzer–Hwang07], some unbalanced urns

Analytic approach : Isomorphism Theorem [FIDuPu06]

$$\text{Urn } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \text{ and } \begin{cases} (a_0, b_0) \\ \alpha + \beta = \gamma + \delta \end{cases} \implies \text{with } \begin{cases} H = X^{a_0} Y^{b_0} \\ \dot{X} = X^{\alpha+1} Y^{\beta} \\ \dot{Y} = X^{\gamma} Y^{\delta+1} \end{cases}$$

## Our goal

- ▶ balanced urns with positive coefficients
- ▶ two parameters class of urns study

## 2. Preferential growth urns

**Motivation** : characterization of **additive**  $2 \times 2$  urns (positive coefficients).

**Approach** : finding a class of urns with “nice” generating functions.

### Theorem [M12]

The balanced urns class  $\begin{pmatrix} 2\alpha & \beta \\ \alpha & \alpha + \beta \end{pmatrix}$ , with  $\alpha > 0, \beta \geq 0$ , has an **algebraic** bivariate generating function.

The histories GF  $H(x, 1, z)$  cancels the following polynomial in  $Y$

$$(z - a - b(x)) Y^{2\alpha+\beta} + b(x) Y^\alpha + a$$

$$\text{with } b(x) = \frac{x^{-\alpha} - 1}{\alpha + \beta} \text{ and } a = (2\alpha + \beta)^{-1}.$$



## Proof

$$\text{Differential system : } \begin{cases} \dot{X} = X^{2\alpha+1} Y^\beta \\ \dot{Y} = X^\alpha Y^{\alpha+\beta+1} \end{cases} \quad \dot{X} = \frac{\partial}{\partial z} X$$
$$\frac{\dot{X}}{X^{\alpha+1}} = \frac{\dot{Y}}{Y^{\alpha+1}} = X^\alpha Y^\beta$$

$$X^{-\alpha} - Y^{-\alpha} = x^{-\alpha} - y^{-\alpha}$$

$$\frac{\dot{Y}}{Y^{\alpha+\beta+1}} (Y^{-\alpha} + x^{-\alpha} - y^{-\alpha}) = 1$$

$$\frac{1}{2\alpha + \beta} Y^{-(2\alpha+\beta)} + \frac{x^{-\alpha} - y^{-\alpha}}{\alpha + \beta} Y^{-(\alpha+\beta)} = - \left( z - \frac{x^{-\alpha} - y^{-\alpha}}{\alpha + \beta} - \frac{1}{2\alpha + \beta} \right)$$

Balanced urn:  $a + b = a_0 + b_0 + n\sigma$ . We set  $y = 1$ .

$$\left( z - \frac{x^{-\alpha} - 1}{\alpha + \beta} - \frac{1}{2\alpha + \beta} \right) Y^{2\alpha+\beta} + \frac{x^{-\alpha} - 1}{\alpha + \beta} Y^\alpha + \frac{1}{2\alpha + \beta} = 0$$

## First observations

The balance  $\sigma = 2\alpha + \beta$       The ratio  $\rho = \frac{\alpha}{2\alpha + \beta} \leq \frac{1}{2}$

For  $x = 1$ , equation becomes :  $(z - \sigma^{-1})Y^\sigma + \sigma^{-1} = 0$

Thus, for  $(a_0, b_0) = (0, 1)$

$$H(1, 1, z) = (1 - \sigma z)^{-1/\sigma} \quad h_n \sim \frac{\sigma^n n^{1/\sigma-1}}{\Gamma(1/\sigma)}$$

### Proposition

Let  $X_n$  be the random variable counting the number of  $x$ -colored balls in the urn after  $n$  steps. Then,

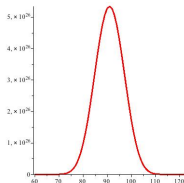
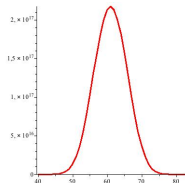
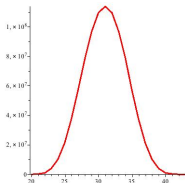
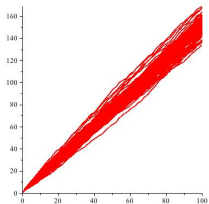
$$\mathbb{E}(X_n) = \frac{\alpha(2\alpha + \beta)}{\alpha + \beta} n + \frac{\alpha}{\alpha + \beta} \frac{\Gamma(\frac{1}{2\alpha + \beta})}{\Gamma(\frac{\alpha + 1}{2\alpha + \beta})} n^{\frac{\alpha}{2\alpha + \beta}} + \frac{\alpha}{\alpha + \beta} + O\left(n^{\frac{\alpha}{2\alpha + \beta} - 1}\right),$$

$$\mathbb{V}(X_n) = \frac{\alpha^3(2\alpha + \beta)}{(\alpha + \beta)^2} n + O\left(n^{\frac{\alpha + \beta}{2\alpha + \beta}}\right).$$

Example  $\alpha = 1, \beta = 1$

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \begin{array}{l} x \rightarrow xx y \\ y \rightarrow xy y \end{array}$$

Preferential growth



$$\left( z - \frac{x^{-1} - 1}{2} - \frac{1}{3} \right) Y^3 + \frac{x^{-1} - 1}{2} Y + \frac{1}{3} = 0$$

# Saddle-point method for $x=1$

$$\left(z - \frac{1}{3}\right) Y^3 + \frac{1}{3} = 0$$

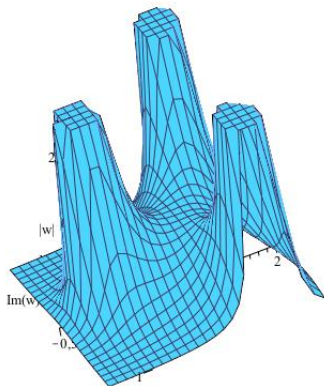
$$y_n = \frac{1}{2i\pi} \oint \frac{Y(z)}{z^{n+1}} dz$$

$$y_n = \frac{3^{n+1}}{2i\pi} \oint a(w) h(w)^{n+1} dw$$

$$\begin{cases} a(w) = 1 - w \\ h(w) = \frac{1}{w(w^2 - 3w + 3)} \end{cases}$$

$$h'(w) = \frac{-3(w-1)^2}{w^2(w^2 - 3w + 3)^2}$$

integrate with a right contour...



$$w \mapsto |h(w)|$$

3 poles

1 double saddle-point in  $w = 1$

## Saddle-point method for $x=1$ (next)

$$t \in [0..L]$$

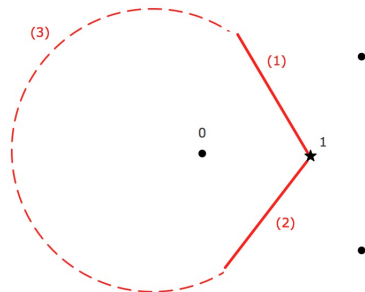
$$(1) w(t) = 1 + te^{i2\pi/3}$$

$$(2) w(t) = 1 + te^{-i2\pi/3}$$

$$h(w(t))^n = \exp(-n(t^3 + O(t^6)))$$

Choose  $L \dots nL^3 \rightarrow \infty$  and  $nL^6 \rightarrow 0$

We set  $L \sim n^{-1/4}$



$$\int_{(1)} + \int_{(2)} : \int_0^\infty ue^{-u^3} du \text{ and } \int_{(3)} \text{ exponentially small}$$

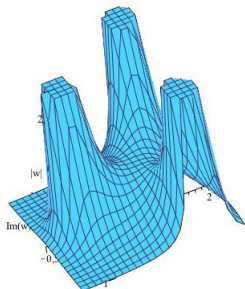
$$y_n = \frac{3^n}{\Gamma(1/3)} \left( n^{-2/3} + O\left(n^{-11/12}\right) \right)$$

## Saddle-point method for $x \neq 1$

$$\left(z - \frac{x^{-1} - 1}{2} - \frac{1}{3}\right) Y^3 + \frac{x^{-1} - 1}{2} Y + \frac{1}{3} = 0$$

$$y_n = \frac{3^{n+1}}{2i\pi} \oint a_x(w) h_x(w)^{n+1} dw$$

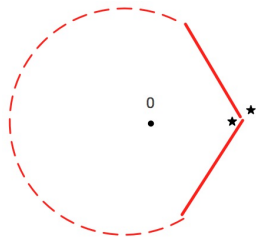
$$h'_x(1) = h'(x^{-1}) = 0$$



$w \mapsto |h_x(w)|$

3 poles

2 saddle-points in  $w = 1$  and  $w = x^{-1}$



$$x = 1 + \frac{\tilde{x}}{\sqrt{n}}, \quad |\tilde{x}| < 1$$

$$y_n(x) \sim \frac{3^n n^{-2/3}}{\Gamma(1/3)} \exp\left(\frac{3}{2}\sqrt{n}\tilde{x} - \frac{3}{8}\tilde{x}^2\right)$$

$$p_n(x) = \frac{y_n(x)}{y_n(1)} \sim \exp\left(\frac{3}{2}\sqrt{n}\tilde{x} - \frac{3}{8}\tilde{x}^2\right)$$

# Gaussian limit law

Let  $X_n$  be the random variable counting the number of  $\bullet$  balls in the urn after  $n$  steps.

## Theorem

$$\mathbb{P} \left\{ \frac{X_n - \frac{3}{2}n}{\sqrt{\frac{3n}{4}}} \leq t \right\} = \Phi(t) + O\left(\frac{1}{\sqrt{n}}\right).$$

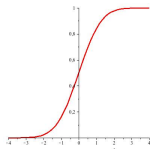
# Local limit law

## Theorem

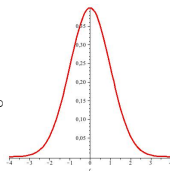
We set  $p_{n,k} = \mathbb{P}\{X_n = k\}$ . The  $X_n$  distribution satisfies a **local limit law** of **gaussian** type with rate of convergence  $O\left(\frac{1}{\sqrt{n}}\right)$ , i.e.

$$\sup_{t \in \mathbb{R}} \left| \frac{\sqrt{3n}}{2} p_{n, \lfloor 3n/2 + t\sqrt{3n}/2 \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \right| \leq \frac{1}{\sqrt{n}}.$$

$$\mathbb{P} \left\{ \frac{X_n - \frac{3}{2}n}{\sqrt{\frac{3n}{4}}} \leq t \right\} \xrightarrow{n \rightarrow \infty}$$



$$\frac{\sqrt{3n}}{2} \mathbb{P} \left\{ X_n = \left\lfloor \frac{3n}{2} + t \frac{\sqrt{3n}}{2} \right\rfloor \right\} \xrightarrow{n \rightarrow \infty}$$



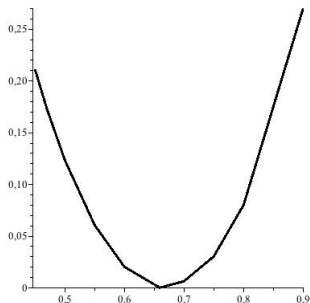


## Large deviations

- ▶ Exponentially small bound on the large deviation with regards to the mean : quantification on **rare events**

### Theorem

- ▶ si  $0.42 < t < 2/3$ ,  $\mathbb{P}(X_n \leq tn) \approx e^{-nW(t)}$  (left tail)
- ▶ si  $2/3 < t < 0.73$ ,  $\mathbb{P}(X_n \geq tn) \approx e^{-nW(t)}$  (right tail)



## General case

$$\left(z - \frac{x^{-\alpha} - 1}{\alpha + \beta} - \frac{1}{2\alpha + \beta}\right) Y^{2\alpha + \beta} + \frac{x^{-\alpha} - 1}{\alpha + \beta} Y^\alpha + \frac{1}{2\alpha + \beta} = 0$$

$$y_n(x) = \frac{\sigma^{n+1}}{2i\pi} \oint a_x(w) h_x(w)^{n+1} dw$$

$h_x(w) : \sigma = 2\alpha + \beta$  poles

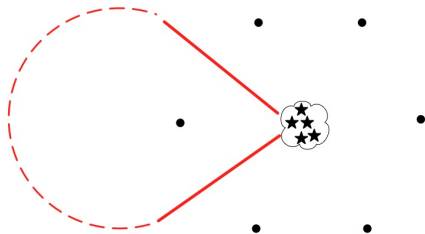
Saddle-point in **1**

with multiplicity  $\alpha + \beta - 1$

The other  $\alpha$  saddle-points

in  $1 - (1 - x^{-\alpha})^{1/\alpha}$

$x \sim 1 + O(n^{-1/2})$  and  $L \sim n^{-\frac{1}{\sigma+1}}$



$$y_n(x) \sim \frac{\sigma^n n^{\frac{1-\sigma}{\sigma}}}{\Gamma(1/\sigma)} \exp\left(\frac{\alpha\sigma}{\alpha + \beta} \sqrt{n\tilde{x}} - \frac{\alpha^3\sigma}{2(\alpha + \beta)^2} \tilde{x}^2\right)$$

$$p_n(x) \sim \exp\left(\frac{\alpha\sigma}{\alpha + \beta} \sqrt{n\tilde{x}} - \frac{\alpha^3\sigma}{2(\alpha + \beta)^2} \tilde{x}^2\right)$$

## Perspectives and conclusion

- ▶ first results on balanced additive small urns, with analytic methods
- ▶ precise results on limit laws (rate of convergence, local limit and large deviations)
- ▶ a general framework for the analysis of balanced algebraic urn model

$$\begin{pmatrix} (k+1)r & b \\ kr & b+r \end{pmatrix} \quad \text{with } r, b, k \geq 1.$$

## Perspectives and conclusion

- ▶ first results on balanced additive small urns, with analytic methods
- ▶ precise results on limit laws (rate of convergence, local limit and large deviations)
- ▶ a general framework for the analysis of balanced algebraic urn model

$$\begin{pmatrix} (k+1)r & b \\ kr & b+r \end{pmatrix} \quad \text{with } r, b, k \geq 1.$$

Thank you for your attention !