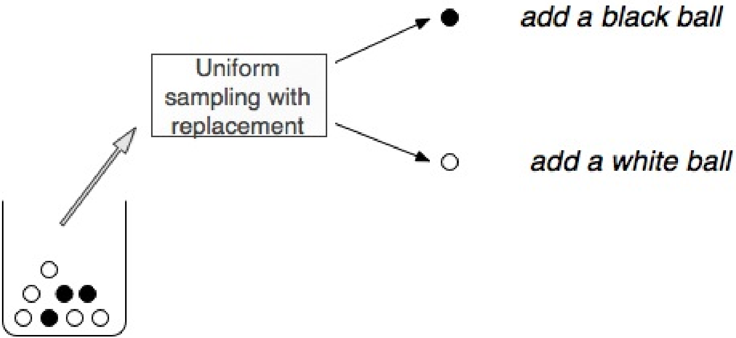
Analysis of an algebraic class of Pólya urn models

Basile Morcrette
(work with Philippe Flajolet)

Algorithms team-project, INRIA Rocquencourt.
LIP6, UPMC

AofA
13 June 2011

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)

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

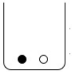
$H_{n,a,b}$: number of histories of length n, beginning in the configuration (a_0, b_0), and ending in (a, b)

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.

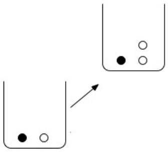
$$H(x, y, z) =$$

$$xyz$$



Combinatorics of histories - Example

We consider this urn $\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.

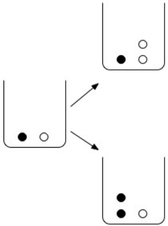


$$H(x, y, z) =$$

$$xyz$$

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.



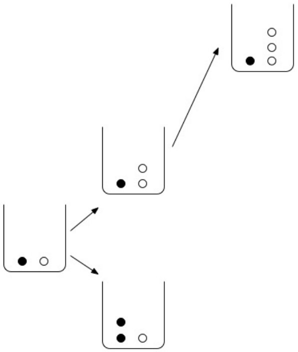
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.



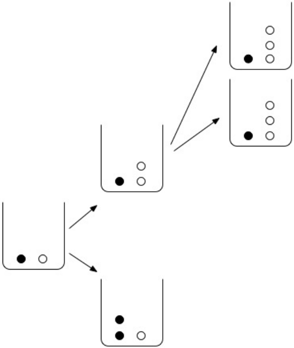
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.



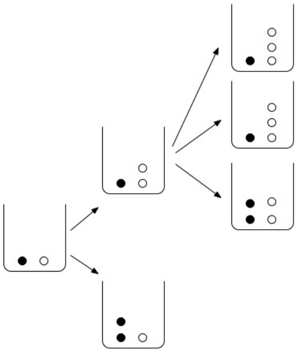
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.



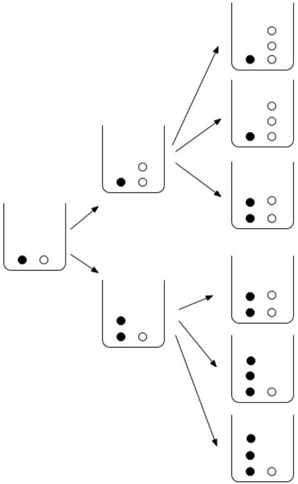
$$H(x, y, z) =$$

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.



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

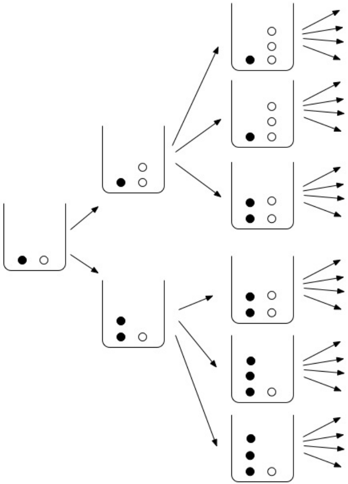
$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

$$+ \quad (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^3}{6}$$

Combinatorics of histories - Example

We consider this urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ with $(a_1, b_1) = (1, 1)$.



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

$$xyz$$

$$+ \quad (xy^2 + x^2y)\frac{z^2}{2}$$

$$+ \quad (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^3}{6}$$

$$+ \quad \dots$$

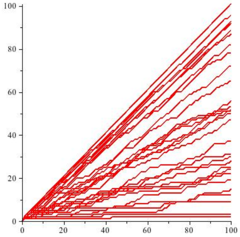
Various behaviours

Problem : Understand the urn composition after n steps, and asymptotically when n tends to ∞.

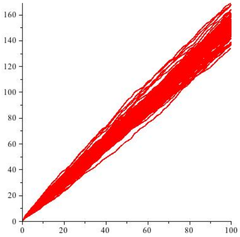
Various behaviours

Problem : Understand the urn composition after n steps, and asymptotically when n tends to ∞.

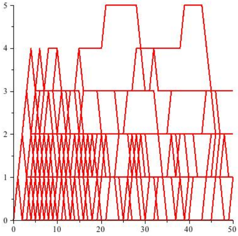
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & 0 \\ 0 & -1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$



Pólya urn



Preferential growing urn



Triangular 3×3 urn

Probabilistic results

Urn $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ Ratio $\rho = \dfrac{\alpha - \gamma}{\alpha + \beta}$

- Small urns : $\rho \leqslant \frac{1}{2}$

  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], unbalanced urns

Analytic approach : theorem [FlDuPu06]

$$\text{Urn } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \text{ and } \left\{ \begin{array}{c} (a_0, b_0) \\ \alpha+\beta=\gamma+\delta \end{array} \right. \Longrightarrow \begin{array}{c} H = X^{a_0}\, Y^{b_0} \\ \text{with } \left\{ \begin{array}{lcl} \dot{X} & = & X^{\alpha+1}\, Y^{\beta} \\ \dot{Y} & = & X^{\gamma}\, Y^{\delta+1} \end{array} \right. \end{array}$$

Our goal

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

2. Preferential growing urns

Motivation : caracterisation of additive 2×2 urns (positive coefficients).
Approach : finding a class of urns with "nice" generating functions.

Theorem [M.Fl11]

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

The histories GF $H(x, 1, z)$ annihils 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

Differential system : $\left\{ \begin{array}{rcl} \dot{X} & = & X^{2\alpha+1}\, Y^{\beta} \\ \dot{Y} & = & X^{\alpha}\, Y^{\alpha+\beta+1} \end{array} \right.$ $\qquad \dot{X} = \frac{\partial}{\partial z} X$

$$\frac{\dot{X}}{X^{\alpha+1}} = \frac{\dot{Y}}{Y^{\alpha+1}} = X^{\alpha} Y^{\beta}$$

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

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

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

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

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

First observations

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

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

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

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

Proposition

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

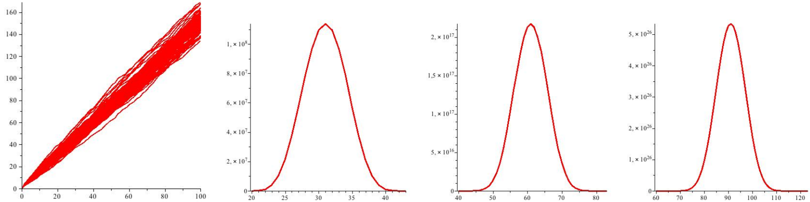
$$\mathbb{E}(X_n) = \frac{\alpha(2\alpha+\beta)}{\alpha+\beta} n + \frac{\alpha}{\alpha+\beta} \frac{\Gamma(\frac{1}{2\alpha+\beta})}{\Gamma(\frac{\alpha+1}{2\alpha+\beta})} n^{\frac{\alpha}{2\alpha+\beta}} + \frac{\alpha}{\alpha+\beta} + O\left(n^{\frac{\alpha}{2\alpha+\beta}-1}\right),$$

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

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

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \qquad \begin{array}{lcl} \mathtt{x} & \rightarrow & \mathtt{x\,x\,y} \\ \mathtt{y} & \rightarrow & \mathtt{x\,y\,y} \end{array}$$

Preferential growing



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

Saddle-point method for x=1

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

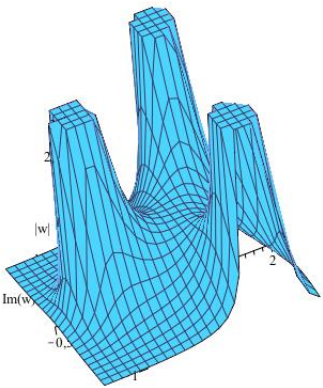
$$y_n = \frac{1}{2i\pi} \oint \frac{Y(z)}{z^{n+1}} \mathrm{d}z$$

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

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

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

integrate with a right contour...



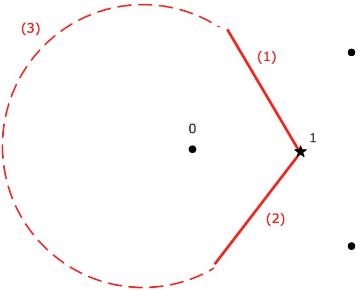
$w \mapsto |h(w)|$

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



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

Choose L... $nL^3 \to \infty$ and $nL^6 \to 0$
We set $L \sim n^{-1/4}$

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

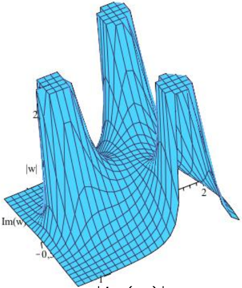
$$\boxed{y_n = \frac{3^n}{\Gamma(1/3)}\left(n^{-2/3} + O\left(n^{-11/12}\right)\right)}$$

Saddle-point method for $x \neq 1$

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

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

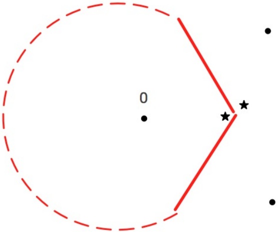
$$h'_x(1) = h'(x^{-1}) = 0$$



$w \mapsto |h_x(w)|$

3 poles

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



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

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

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

Gaussian limit law

Let X_n be the random variable couting the number of • balls in the urn after n steps.

Theorem

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

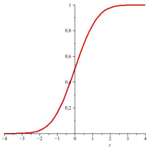
Local limit law

Theorem

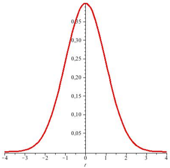
We set $p_{n,k} = \mathbb{P}\{X_n = k\}$. The X_n distribution satisfies a local limit law of gaussian type with speed of convergence $O\left(\frac{1}{\sqrt{n}}\right)$, i.e.

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

$$\mathbb{P}\left\{ \frac{X_n - \frac{3}{2}n}{\sqrt{\frac{3n}{4}}} \leqslant t \right\} \underset{n\to\infty}{\longrightarrow}$$



$$\frac{\sqrt{3n}}{2}\,\mathbb{P}\left\{X_n = \left\lfloor \frac{3n}{2} + t\frac{\sqrt{3n}}{2} \right\rfloor \right\} \underset{n\to\infty}{\longrightarrow}$$

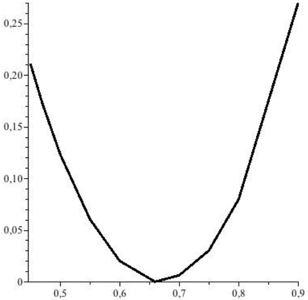


Large deviations

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

Theorem

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



General case

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

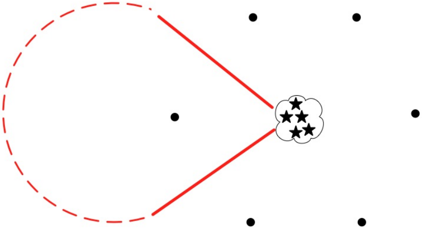
$$y_n(x) = \frac{\sigma^{n+1}}{2i\pi} \oint a_x(w) h_x(w)^{n+1} \mathrm{d}w$$

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

Saddle-point in 1
with multiplicity $\alpha+\beta-1$

The other α 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 additive small urns, with analytic methods
- precise results on limit laws (speed of convergence, local limit and large deviations)
- a door to unbalanced urns. Using PDE, and trying to make a full analysis

Perspectives and conclusion

- first results on additive small urns, with analytic methods
- precise results on limit laws (speed of convergence, local limit and large deviations)
- a door to unbalanced urns. Using PDE, and trying to make a full analysis

Thank you for your attention !

Unbalanced urns and PDE

Balanced case : $P_{n,a,b} = \dfrac{n![x^a y^b z^n]H(x,y,z)}{n![z^n]H(1,1,z)}$

Unbalanced case : All histories of length n are not equiprobable

If $g(x,y,z) = \sum_{n,a,b} P_{n,a,b} \dfrac{x^a y^b}{a+b} z^n$, then g satisfies the PDE :

$$\boxed{x(1 - zx^\alpha y^\beta)g_x + y(1 - zx^\gamma y^\delta)g_y = x^{a_0} y^{b_0}}$$

Then :

- change of variable
- elimination of one variable, for example y
- method of characteristics