Probabilistic Analytic Urns

Pólya Urns with Random Entries: an Analytic Combinatorics Approach

Basile Morcrette

Algorithms project, INRIA Rocquencourt
LIP6, UPMC

PCC, Warwick
August 2012

Dedicated to Philippe Flajolet.
Pólya Urn Model

- 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: \( \alpha + \beta = \gamma + \delta \) (deterministic total number of balls)

A given initial configuration \((b_0, w_0)\):
- \(b_0\) balls ● (counted by \(x\))
- \(w_0\) balls ○ (counted by \(y\))

**Definition**

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

\[
H(x, y, z) = \sum_{n,b,w} H_{n,b,w} x^b y^w z^n \frac{z^n}{n!}
\]

\(H_{n,b,w}\): number of histories of length \(n\), beginning in the configuration \((b_0, w_0)\), and ending in \((b, w)\)
Combinatorics of histories - Example

We consider this urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = \frac{xy}{24}
\]
Combinatorics of histories - Example

We consider this urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy
\]
Combinatorics of histories - Example

We consider this urn \[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\] with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z
\]
Combinatorics of histories - Example

We consider this urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z
\]
Combinatorics of histories - Example

We consider this urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z
\]
Combinatorics of histories - Example

We consider this urn \(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2 y)z
\]
Combinatorics of histories - Example

We consider this urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z + (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^2}{2}
\]
Combinatorics of histories - Example

We consider this urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) with \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z + (2xy^3 + 2x^2y^2 + 2x^3y)\frac{z^2}{2} + \ldots
\]
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 \( \infty \).

\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \quad \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix} \quad \begin{pmatrix}
0 & 1 & 0 \\
0 & -1 & 2 \\
0 & 0 & 1
\end{pmatrix}
\]

Pólya urn  
Friedman urn  
Triangular 3 × 3 urn
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 [FlDuPu06]

Urn \(\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\) and \(\{(b_0, w_0)\}\) where \(\alpha + \beta = \gamma + \delta\) \(\implies\) with

\[
\begin{align*}
H &= X^{b_0} Y^{w_0} \\
\dot{X} &= X^{\alpha+1} Y^{\beta} \\
\dot{Y} &= X^{\gamma} Y^{\delta+1}
\end{align*}
\]

### Our goal

- balanced urns with random coefficients
- obtain an *isomorphism theorem*
Balanced Pólya Urns with Random Entries

\[
\begin{pmatrix}
A & B \\
C & D
\end{pmatrix} \quad A, D \in \mathbb{Z}, \quad B, C \in \mathbb{N}
\]

\(A, B, C, D\) are random variables.

Balanced urn: \(A + B = C + D := \theta\) (deterministic total number of balls)

A given initial configuration \((b_0, w_0)\): \(b_0\) balls \(\bullet\), \(w_0\) balls \(\circ\).

Configuration \((B_n, W_n)\): number of black and white balls in the urn after \(n\) steps.

Total number of balls \(s_n := B_n + W_n\)

\[s_n = s_0 + \theta n = b_0 + w_0 + \theta n\]
An urn scheme is said to be tenable, if it is always possible to apply a rule, i.e., if it never reaches a deadlocked configuration.

\[
\begin{pmatrix}
-2 & 2 \\
2 & -2
\end{pmatrix}
\]

Starting configuration:

\((4,4) \rightarrow (2,6) \rightarrow (0,8) \rightarrow (2,6) \rightarrow \ldots\)

\((3,4) \rightarrow (1,6) \rightarrow \text{stuck : (}\)
Counting histories: Pólya example

Take the urn \( \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \) and \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z + (2x^3y + 2x^2y^2 + 2xy^3) \frac{z^2}{2} + \ldots
\]
Counting histories: Friedman example

Take the urn \( \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \) and \((b_0, w_0) = (1, 1)\).

\[
H(x, y, z) = xy + (xy^2 + x^2y)z + (xy^3 + 4x^2y^2 + x^3y)z^2 + \ldots
\]
Pólya–Friedman mixing model

\[
\left( \begin{array}{cc}
B_p & 1 - B_p \\
1 - B_p & B_p
\end{array} \right), \quad \text{with } B_p \sim \text{Bernoulli}(p)
\]

Starting configuration: \((b_0, w_0) = (1, 1)\).

\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\quad \text{with proba } p_1 := p
\]

\[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\quad \text{with proba } p_2 := 1 - p
\]

Pólya urn, with uniform limiting distribution

Friedman urn, with Gaussian limiting distribution
Pólya–Friedman mixing model

\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\text{ with proba } p_1
\]

\[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\text{ with proba } p_2
\]

\[p_1 = 0 \quad \text{Friedman}\]
\[p_1 = 0.4 \quad p_1 = 0.7 \quad p_1 = 1 \quad \text{Pólya}\]
Counting histories: Pólya–Friedman mixing model

\[
Q(x, y, z) = \begin{array}{c}
p^2 \quad xy \\
p(1-p) \quad (xy^2 + x^2y)z \\
\left(2p^2 + (1 - p)^2 + 3p(1 - p)\right) \quad xy^3 \\
\left(2p^2 + 4(1 - p)^2 + 6p(1 - p)\right) \quad x^2y^2 \\
\left(2p^2 + (1 - p)^2 + 3p(1 - p)\right) \quad x^3y \frac{z^2}{2} \\
+ \quad \ldots 
\end{array}
\]
Weighted Histories and Generating Function

**Definition**

\[ Q(x, y, z) = \sum_{b,w,n} Q_{n,b,w} x^b y^w z^n n! , \]

where \( Q_{n,b,w} \) is the weighted contribution for paths from the initial configuration \((b_0, w_0)\) to the configuration \((b, w)\) in \(n\) steps.

Total number of weighted histories:

\[ s_0 s_1 \ldots s_{n-1} = s_0 (s_0 + \theta)(s_0 + 2\theta) \ldots (s_0 + (n - 1)\theta) = \theta^n \frac{\Gamma(n + s_0/\theta)}{\Gamma(s_0/\theta)} \]

\[ Q(1, 1, z) = (1 - \theta z)^{-s_0/\theta} \]

\[ \mathbb{P}(B_n = b, W_n = w) = \frac{Q_{n,b,w}}{s_0 s_1 \ldots s_{n-1}} = \frac{[x^b y^w z^n]Q(x, y, z)}{[z^n]Q(1, 1, z)} \]
Isomorphism Theorem

For

\[
\begin{pmatrix}
B_p & 1 - B_p \\
1 - B_p & B_p
\end{pmatrix}, \quad \text{with } B_p \sim \text{Bernoulli}(p),
\]

the trivariate generating function of weighted histories \( Q \) satisfies

\[
Q(x, y, z) = X(z)^{b_0} \ Y(z)^{w_0}
\]

where

\[
\begin{cases}
X' &= p_1 \ X^2 + p_2 \ X \ Y \\
Y' &= p_1 \ Y^2 + p_2 \ X \ Y
\end{cases}
\]

and \( X(0) = x, \ Y(0) = y. \)
Isomorphism proof

\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} \text{ with proba } p_1 \quad \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix} \text{ with proba } p_2
\]

Differentiate = Pick a ball

\[
\partial_x[xx \ldots x] = (x x \ldots x) + (x x \ldots x) + \ldots + (x x \ldots x)
\]

\[
x \partial_x[xx \ldots x] = (x x \ldots x) + (x x \ldots x) + \ldots + (x x \ldots x)
\]

From step \(n\) to step \(n+1\):

\[
x^i y^j \xrightarrow{\mathcal{M}} i \left( p_1 x^{i+1} y^j + p_2 x^i y^{j+1} \right) + j \left( p_1 x^i y^{j+1} + p_2 x^{i+1} y^j \right)
\]

Then

\[
\mathcal{M} := \left( p_1 x^2 + p_2 xy \right) \partial_x + \left( p_1 y^2 + p_2 xy \right) \partial_y
\]
Isomorphism proof

Differentiate = Pick a ball

\[ \partial_x [xx \ldots x] = (xx \ldots x) + (x \bar{x} \ldots x) + \ldots + (xx \ldots \bar{x}) \]
\[ x \partial_x [xx \ldots x] = (xx \ldots x) + (xx \ldots x) + \ldots + (xx \ldots x) \]

From step \( n \) to step \( n + 1 \):

\[ x^i y^j \xrightarrow{m} i (p_1 x^{i+1} y^j + p_2 x^i y^{j+1}) + j (p_1 x^i y^{j+1} + p_2 x^{i+1} y^j) \]

Then

\[ M := (p_1 x^2 + p_2 xy) \partial_x + (p_1 y^2 + p_2 xy) \partial_y \]

\[ M^n [x^b_0 y^w_0] = \sum_{b, w} Q_{n,b,w} x^b y^w \]

\[ Q(x, y, z) = \sum_{n \geq 0} M^n [x^b_0 y^w_0] \frac{z^n}{n!} \]
Isomorphism proof

\[ Q(x, y, z) = \sum_{n \geq 0} \mathcal{M}^n [x^{b_0} y^{w_0}] \frac{z^n}{n!} \]

Let \((X(t), Y(t))\) be solution of

\[
\begin{align*}
X' &= p_1 X^2 + p_2 XY & X(0) &= x \\
Y' &= p_1 Y^2 + p_2 XY & Y(0) &= y
\end{align*}
\]

\[
\partial_t (X^i Y^j) = i X^{i-1} X' Y^j + j X^i Y^{j-1} Y' = i (p_1 X^{i+1} Y^j + p_2 X^i Y^{j+1}) + j (p_1 X^i Y^{j+1} + p_2 X^{i+1} Y^j)
\]

\[
\partial^n_t (X^i Y^j) = \mathcal{M}^n [x^i y^j] \quad x \rightarrow X \\
y \rightarrow Y
\]

\[ Q(X(t), Y(t), z) = \sum_{n \geq 0} \partial^n_t [X(t)^{b_0} Y(t)^{w_0}] \frac{z^n}{n!} = X(t + z)^{b_0} Y(t + z)^{w_0} \]

Then \(t = 0\), and it’s done! \[ Q(x, y, z) = X(z)^{b_0} Y(z)^{w_0} \]
Isomorphism Theorem

For any urn scheme \( (A, B, C, D) \) with initial conditions \((b_0, w_0)\), which is:

1. balanced positive (i.e. \( A + B = C + D = \theta \geq 0 \)),
2. tenable (i.e. no deadlock configuration at any step),

the trivariate generating function of weighted histories \( Q(x, y, z) \) satisfies:

\[
Q(x, y, z) = X(z)^{b_0} Y(z)^{w_0},
\]

where

\[
\begin{align*}
X' &= \sum_{k=-K}^{\theta} \pi_k X^{k+1} Y^{\theta-k} \\
Y' &= \sum_{k=-K}^{\theta} \tau_k X^{\theta-k} Y^{k+1}
\end{align*}
\]

with initial conditions \( X(0) = x, \ Y(0) = y \).
Application I: Coupon collector with delay

\[
\begin{pmatrix}
-B_p & B_p \\
0 & 0
\end{pmatrix}
\]

where \( B_p \) is a Bernoulli\((p)\) random variable, and \((b_0, w_0)\) initial condition.

\[
\begin{cases}
X' = (1 - p)X + pY; \\
Y' = Y,
\end{cases}
\]

\[Q(x, 1, z) = \left( e^z + (x - 1)e^{(1-p)z} \right)^{b_0} e^{w_0 z}.\]

The exact probability distribution of the number \( B_n \) of uncollected coupons (black balls) after \( n \) draws is given by

\[
P(B_n = b) = \sum_{j=b}^{b_0} (-1)^{j-b} \binom{b_0}{j} \binom{j}{b} \left( \frac{s_0 - pj}{s_0} \right)^n.\]
\textbf{Application II: a uniform urn}

\[
\begin{pmatrix}
\mathcal{U}_\theta & \theta - \mathcal{U}_\theta \\
\theta - \mathcal{U}_\theta & \mathcal{U}_\theta
\end{pmatrix},
\]

\(\mathcal{U}_\theta\) is a uniformly distributed random variable on the set \(\{0, 1, \ldots, \theta\}\).

\[
\begin{cases}
X' &= \sum_{k=0}^{\theta} \frac{1}{\theta + 1} X^{k+1} Y^{\theta-k} ; \\
Y' &= \sum_{k=0}^{\theta} \frac{1}{\theta + 1} X^{\theta-k} Y^{k+1} .
\end{cases}
\]

\[
Q(x, 1, z) = \frac{x^{b_0}}{(1 - \frac{\theta}{\theta + 1} (\sum_{l=0}^{\theta} x^l) z)^{b_0+w_0}} .
\]

The exact probability distribution of black balls is given by

\[
P(B_n = b) = \frac{T_{\theta,n,b-b_0}}{(\theta + 1)^n}, \quad \text{for } b \in \{b_0, b_0 + 1, \ldots, b_0 + \theta n\},
\]

where \(T_{\theta,n,k} := [x^k] (1 + x + \ldots + x^\theta)^n\).
Application III: More than 2 colors

$B_p$ a Bernoulli($p$), and $(b_0, r_0, g_0)$ the initial condition.

\[
\begin{pmatrix}
-1 & B_p & 1 - B_p \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{pmatrix},
\begin{cases}
X' = pY + (1 - p)H; \\
Y' = Y; \\
H' = H.
\end{cases}
\]

\[Q(x, y, h, z) = X^{b_0}(z) Y^{r_0}(z) H^{g_0}(z),\]

\[X(z) = (p y + (1 - p) h) (e^z - 1) + x;\]
\[Y(z) = y e^z;\]
\[H(z) = h e^z.\]

The distribution of red balls $R_n$ is

\[\mathbb{P}(R_n = r) = \left(\frac{p}{1 - p}\right)^r \sum_{j=0}^{b_0} (-1)^j \binom{j}{r} \binom{b_0}{j} (1 - p)^j \sum_{k=0}^{j} \binom{j}{k} (-1)^k \left(\frac{k}{s_0}\right)^n.\]
Conclusion

✓ general isomorphism theorem for balanced tenable urns with random entries, for two or more colors
✓ access to no trivial exact formulas for probability distribution of balls

✗ find asymptotics and limit distributions
✗ solve the differential system ?
✗ for the simple Pólya–Friedman mixing scheme, find the transition between Gaussian and uniform regime.
Advert: More about *Analytic Combinatorics*:


**French-British Workshop on Analytic Combinatorics**

in Oxford, from Sept. 5th to Sept. 7th