

Probabilistic *Analytic Urns*

Exactly Solvable Balanced Tenable Urns with Random Entries via the Analytic Methodology

Basile Morcrette¹ and Hosam M. Mahmoud²

¹ Algorithms project, INRIA Rocquencourt
¹ LIP6, UPMC

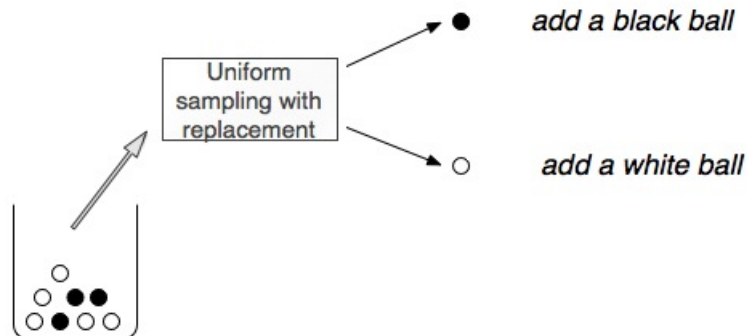
² George Washington University (USA)

AofA, Montreal
June 21, 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 with Random Entries

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

$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ are random variables.

Balanced urn: $\boxed{\mathcal{A} + \mathcal{B} = \mathcal{C} + \mathcal{D} := \theta}$
(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 $s_n := B_n + W_n$

$$s_n = s_0 + \theta n = b_0 + w_0 + \theta n$$

Tenability

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.

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

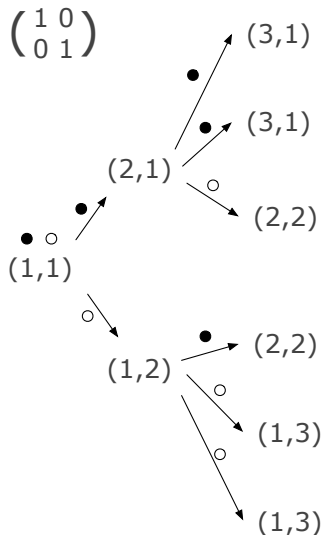
Starting configuration:

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

(3,4) \rightarrow (1,6) \rightarrow stuck :(

Counting histories: Pólya example

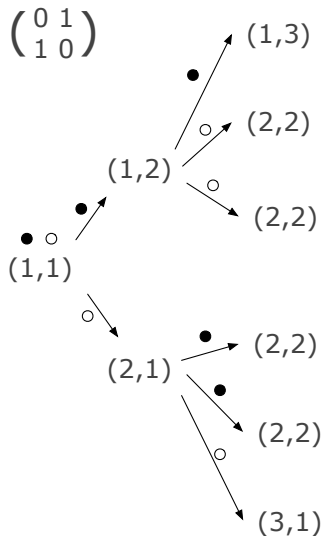
Take the urn $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $(b_0, w_0) = (1, 1)$.



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

Counting histories: Friedman example

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



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

$$xy$$

$$+ (xy^2 + x^2y)z$$

$$+ (xy^3 + 4x^2y^2 + x^3y)\frac{z^2}{2}$$

$$+ \dots$$

Pólya–Friedman mixing model

$$\begin{pmatrix} \mathcal{B}_p & 1 - \mathcal{B}_p \\ 1 - \mathcal{B}_p & \mathcal{B}_p \end{pmatrix}, \text{ with } \mathcal{B}_p \sim \text{Bernoulli}(p)$$

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

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

Pólya urn,
with *uniform* limiting distribution

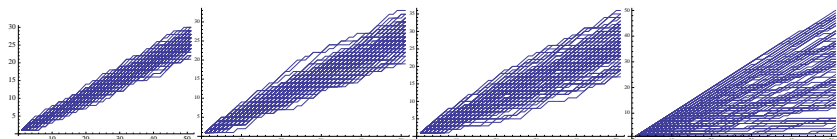
$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ with proba } p_2 := 1 - p$$

Friedman urn,
with *Gaussian* limiting distribution

Pólya–Friedman mixing model

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

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

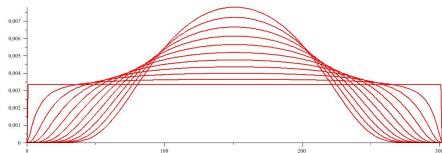
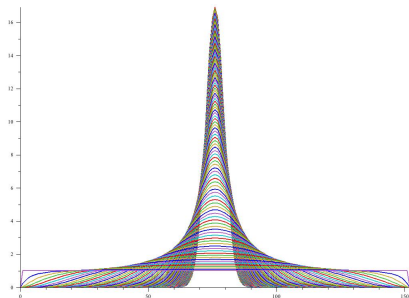


$p_1 = 0$
Friedman

$p_1 = 0.4$

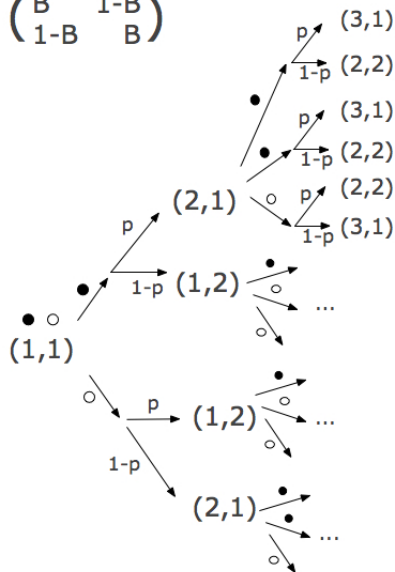
$p_1 = 0.7$

$p_1 = 1$
Pólya



Counting histories: Pólya–Friedman mixing model

$$\begin{pmatrix} B & 1-B \\ 1-B & B \end{pmatrix}$$



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

$$p^2 xy$$

$$+ p(1-p)(xy^2 + x^2y)z$$

$$+ ((2p^2 + (1-p)^2 + 3p(1-p)) xy^3$$

$$+ (2p^2 + 4(1-p)^2 + 6p(1-p)) x^2y^2$$

$$+ (2p^2 + (1-p)^2 + 3p(1-p)) x^3y) \frac{z^2}{2}$$

$$+ \dots$$

Weighted Histories and Generating Function

Definition

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

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:

$$s_0 s_1 \dots s_{n-1} = s_0 (s_0 + \theta) (s_0 + 2\theta) \dots (s_0 + (n-1)\theta) = \theta^n \frac{\Gamma(n + s_0/\theta)}{\Gamma(s_0/\theta)}$$

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

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

Isomorphism Theorem

For

$$\begin{pmatrix} \mathcal{B}_p & 1 - \mathcal{B}_p \\ 1 - \mathcal{B}_p & \mathcal{B}_p \end{pmatrix}, \text{ with } \mathcal{B}_p \sim \text{Bernoulli}(p),$$

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

$$\partial_x[xx \dots x] = (\cancel{x}x \dots x) + (x\cancel{x} \dots x) + \dots + (xx \dots \cancel{x})$$

$$x\partial_x[xx \dots x] = (\underline{x}x \dots x) + (x\underline{x} \dots x) + \dots + (xx \dots \underline{x})$$

From step n to step $n + 1$:

$$x^i y^j \xrightarrow{\mathfrak{M}} i (p_1 x^{i+1} y^j + p_2 x^i y^{j+1}) + j (p_1 x^i y^{j+1} + p_2 x^{i+1} y^j)$$

Then

$$\mathfrak{M} := (p_1 x^2 + p_2 xy) \partial_x + (p_1 y^2 + p_2 xy) \partial_y$$

Isomorphism proof

Differentiate = Pick a ball

$$\partial_x[xx \dots x] = (\cancel{x}x \dots x) + (x\cancel{x} \dots x) + \dots + (xx \dots \cancel{x})$$

$$x\partial_x[xx \dots x] = (\underline{x}x \dots x) + (x\underline{x} \dots x) + \dots + (xx \dots \underline{x})$$

From step n to step $n + 1$:

$$x^i y^j \xrightarrow{\mathfrak{M}} i (p_1 x^{i+1} y^j + p_2 x^i y^{j+1}) + j (p_1 x^i y^{j+1} + p_2 x^{i+1} y^j)$$

Then

$$\mathfrak{M} := (p_1 x^2 + p_2 xy) \partial_x + (p_1 y^2 + p_2 xy) \partial_y$$

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

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

Isomorphism proof

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

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

$$\begin{cases} X' = p_1 X^2 + p_2 XY & X(0) = x \\ Y' = p_1 Y^2 + p_2 XY & Y(0) = y \end{cases}$$

$$\begin{aligned} \partial_t (X^i Y^j) &= i X^{i-1} X' Y^j + j X^i Y^{j-1} Y' \\ &= i (p_1 X^{i+1} Y^j + p_2 X^i Y^{j+1}) + j (p_1 X^i Y^{j+1} + p_2 X^{i+1} Y^j) \end{aligned}$$

$$\boxed{\partial_t^n (X^i Y^j) = \mathfrak{M}^n [x^i y^j] \begin{array}{l} x \rightarrow X \\ y \rightarrow Y \end{array}}$$

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

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 $\begin{pmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{pmatrix}$ with initial conditions (b_0, w_0) , which is:

1. balanced positive (i.e. $\mathcal{A} + \mathcal{B} = \mathcal{C} + \mathcal{D} := \theta \geq 0$),
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

$$\begin{cases} X' &= \sum_{k=-K}^{\theta} \pi_k X^{k+1} Y^{\theta-k} & \pi_k := \mathbb{P}(\mathcal{A} = k) \\ Y' &= \sum_{k=-K}^{\theta} \tau_k X^{\theta-k} Y^{k+1} & \tau_k := \mathbb{P}(\mathcal{D} = k) \end{cases}$$

with initial conditions $X(0) = x$, $Y(0) = y$.

Application I: Coupon collector with delay

$$\begin{pmatrix} -\mathcal{B}_p & \mathcal{B}_p \\ 0 & 0 \end{pmatrix}$$

where \mathcal{B}_p is a Bernoulli(p) random variable, and (b_0, w_0) initial condition.

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

$$Q(x, 1, z) = (e^z + (x-1)e^{(1-p)z})^{b_0} e^{w_0 z}.$$

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

$$\mathbb{P}(B_n = b) = \sum_{j=b}^{b_0} (-1)^{j-b} \binom{b_0}{j} \binom{j}{b} \left(\frac{s_0 - pj}{s_0} \right)^n.$$

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}_θ is a uniformly distributed random variable on the set $\{0, 1, \dots, \theta\}$.

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

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

The exact probability distribution of black balls is given by

$$\mathbb{P}(B_n = b) = \frac{T_{\theta, n, b-b_0}}{(\theta+1)^n}, \quad \text{for } b \in \{b_0, b_0 + 1, \dots, b_0 + \theta n\},$$

where $T_{\theta, n, k} := [x^k] (1 + x + \dots + x^\theta)^n$.

Application III: More than 2 colors

\mathcal{B}_p a Bernoulli(p), and (b_0, r_0, g_0) the initial condition.

$$\begin{pmatrix} -1 & \mathcal{B}_p & 1 - \mathcal{B}_p \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{cases} X' = pY + (1-p)H; \\ Y' = Y; \\ H' = H. \end{cases}$$

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

$$X(z) = (py + (1-p)h)(e^z - 1) + x;$$

$$Y(z) = ye^z;$$

$$H(z) = he^z.$$

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.

