Algorithms Seminar 2002–2004,
F. Chyzak (ed.), INRIA, (2005), pp. 61–64.

Available online at the URL
`http://algo.inria.fr/seminars/`.

Analytic Urns of Triangular Form

Vincent Puyhaubert
Projet Algorithmes, Inria (France)

September 22, 2003

Summary by Julien Fayolle

Abstract

The urn model was first introduced as a mathematical object by Eggenberger and Pólya [2] in 1923. It has since become handy to model some problems arising from computer science, like for instance the coupon collector problem, hashing problems or even random tree generation. The original goal of Pólya and Eggenberger was related to public health and the spread of epidemics.

Until recently the only approach to urn problems was through the probabilistic toolkit. In [3] Flajolet, Gabarró, and Pekari first devised an approach using analytic combinatorics for certain families of urn (balanced ones). Puyhaubert [4] travels down this same path and uses the characteristics method for solving partial differential equations to obtain more precise results on the asymptotic behavior of the urns (moments, limit law).

1. Model

An urn contains balls of different types, say colors. Suppose first that we have two types of balls, say black ones and white ones. The given of the problem is the initial configuration of the urn, i.e., the number of black and of white balls at the origin of time and a 2×2 *replacement* matrix $M = (m_{i,j})_{i,j}$ with integral entries that codes the evolution rules of the urn.

The time is discrete and at time n a ball is picked (and then put back in the urn) uniformly at random from the urn. If a black (resp. white) ball is drawn from the urn, we add $m_{1,1}$ (resp. $m_{2,1}$) black and $m_{1,2}$ (resp. $m_{2,2}$) white balls in the urn. For this talk the replacement matrix has fixed entries (for instance they do not depend on the time) and the sum of the entries of any row is constant (we say the urn is *balanced*). This balance condition means the number of balls in the urn at time $n+1$ does not depend on the ball picked at time n. Our matrices are triangular (say upper-triangular, meaning we necessarily have $m_{2,1} = 0$). One last condition is *tenability*: if the entries of the matrix were negative one might end up trying to remove balls that no longer exist in the urn, therefore the entries are all positive integers.

As an example, we explain the urn model for the coupon collector problem. We have n different coupons to collect. We code coupons already collected as black balls in the urn and those not already collected as white balls. At time zero, there are just n white balls in the urn for no coupon has yet been found. Each time we pick a coupon (or ball) either it is one we already possess in our collection (black ball) and then the urn composition does not change, or it is a new coupon (white ball) and then we remove the white ball and add a black ball to mark we have one more coupon in our personal collection. We can model this behavior with the replacement matrix

$$M = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}$$

and the initial composition of n white balls.

2. Counting

This part actually initiates from [3] and their results on analytic urns. Throughout the rest of the summary we use the replacement matrix M with an initial composition U_0

$$M = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}, \quad U_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

to serve as an example of the methods.

We enumerate all possible compositions of an urn (knowing its initial composition and its replacement matrix) at time n with a polynomial $f_n(u,v)$ where u marks the number of black balls and v the number of white balls in the urn. A composition can be counted several times if there are several histories (picks of the distinguished balls) leading to this composition at time n.

For our example we have $f_0(u,v) = uv$, $f_1(u,v) = u^2v^2 + uv^3$, and $f_2(u,v) = 3uv^5 + 3u^2v^4 + 2u^3v^3$. Each history of length n (series of the n picks) has the same probability $2 \cdot 4 \cdot \ldots \cdot (2n)$.

From an urn composition u^rv^s at time n there are r possibilities of picking a black ball and this leads to a composition $u^{r+1}v^{s+1}$ at time $n+1$, and there are s possibilities of picking a white ball leading to a composition u^rv^{s+2} at time $n+1$. Hence from the u^rv^s composition at time n we have $ru^{r+1}v^{s+1} + su^rv^{s+2}$ possible compositions at time $n+1$. We introduce a differential operator Γ to code the evolution of the monomials u^rv^s:

$$\Gamma = uv \cdot u\frac{\partial}{\partial u} + u^0v^2 \cdot v\frac{\partial}{\partial v}. \tag{1}$$

This operator is linear so it extends to the whole polynomial allowing us to describe in a simple manner the possible evolutions of the urn composition: $f_{n+1}(u,v) = \Gamma f_n(u,v)$.

We introduce the main generating function for our study:

$$H(z,u,v) = \sum_{n\geq 0} f_n(u,v)\frac{z^n}{n!},$$

it satisfies the initial condition $H(0,u,v) = uv$ and the partial differential equation $\frac{\partial}{\partial z}H = \Gamma H$. We know there is one sole generating function satisfying these two conditions, and from it we can derive the moments of the random variable X_n counting the number of black balls in the urn and also its limit law.

Why focus uniquely on black balls? Because the balance condition on the urn means the population growth of the urn is deterministic: whatever the history of the urn the population increases by two at every replacement. It leads to $s + r = 2 + 2n = 2(n+1)$ balls in the urn at time n.

3. A Method for Solving PDEs

The partial differential equation

$$u^2v\frac{\partial H}{\partial u} + v^3\frac{\partial H}{\partial v} = \frac{\partial H}{\partial z} \tag{2}$$

can be solved using the characteristics method (see [1] for a thorough course on it). We first introduce the notion of first integral and some of its properties before applying them to our simple example.

Definition 1. Let $x'(t) = f(x)$ be an ordinary differential equation where x is a $\mathcal{C}^1$ function from an open I of $\mathbb{R}$ to an open U of $\mathbb{R}^n$ and f is a $\mathcal{C}^1$ function from U to $\mathbb{R}^n$. A $\mathcal{C}^1$ function ψ from U to $\mathbb{R}$ is said to be a first integral for the equation $x'(t) = f(x)$ if for any solution $x(t)$ to the differential equation $\psi(x(t))$ is constant.

The equation $x'(t) = f(x)$ is actually a system of n ODEs if we look at the coordinates $(x_i)_i$ and $(f_i)_i$ of the functions $x(t)$ and $f(x)$ on the canonical basis of $\mathbb{R}^n$. The benefit of using prime integrals is to avoid partial derivatives to deal solely with ODEs. The next proposition describes this link between ODEs and PDEs, its proof is quite straightforward.

Proposition 1. *A function ψ is a prime integral of $\frac{dx}{dt}(t) = f(x)$ if and only if for any point of U we have*

$$\sum_{i=1}^{n} f_i(x_1, \ldots, x_n) \frac{\partial \psi}{\partial x_i}(x_1, \ldots, x_n) = 0. \tag{3}$$

For a "good" differential system of n equation there exist $n-1$ prime integrals $(\psi_i)_i$ whose derivatives are linearly independent such that any prime integral can be expressed as $\phi(\psi_1, \ldots, \psi_{n-1})$ for any arbitrary $\mathcal{C}^1$ function ϕ.

For our guideline example we have to solve the PDE (2) which is of the type of (3). Given the proposition we only have to look for prime integrals of the system of three equations

$$\frac{du}{dt} = u^2 v, \qquad \frac{dv}{dt} = v^3, \qquad \text{and} \qquad \frac{dz}{dt} = -1. \tag{4}$$

We eliminate the dt from all three equations and obtain a system of two equations

$$\frac{dv}{v^3} = -dz, \qquad \text{and} \qquad \frac{du}{u^2} = \frac{dv}{v^2},$$

then we integrate and get two first integrals (checking it is easy) whose derivatives are linearly independent:

$$\psi_1(z, u, v) = z - \frac{1}{2v^2}, \qquad \text{and} \qquad \psi_2(z, u, v) = \frac{1}{v} - \frac{1}{u}.$$

The generating function for the urn is written in the form

$$H(z, u, v) = \phi\left(z - \frac{1}{2v^2}, \frac{1}{v} - \frac{1}{u}\right)$$

for some arbitrary function ϕ. We also have an initial condition $H(0, u, v) = uv$ that leads to

$$H(z, u, v) = uv(1 - 2v^2 z)^{-1/2}\left(1 - uv^{-1}(1 - (1 - 2v^2 z)^{1/2})\right)^{-1}.$$

4. Results

We now take a more general 2×2 triangular replacement matrix M and an initial composition (a_0, b_0) for the urn, thus with an initial population of $t_0 = a_0 + b_0$ balls.

$$M = \begin{pmatrix} a & b-a \\ 0 & b \end{pmatrix}$$

We solve the PDE

$$u^{a+1} v^{b-a} \frac{\partial H}{\partial u} + v^{b+1} \frac{\partial H}{\partial v} = \frac{\partial H}{\partial z}$$

associated to the urn using the characteristics method from the previous section. It leads to the generating function

$$H(z,u,v) = u^{a_0} v^{b_0} (1 - bv^b z)^{-b_0/b} (1 - u^a v^{-a} (1 - (1 - bv^b z)^{a/b}))^{-a_0/a}. \tag{5}$$

If we denote by X_n the random variable counting the number of black balls at time n, we obtain the probability of having $a_0 + ka$ black balls at time n and also a full expansion of its lth-order moment.

$$\mathbf{P}(X_n = a_0 + ka) = \frac{n!}{t_0 \cdots (t_0 + (n-1)s)} \binom{\frac{a_0}{a} + k - 1}{k} \sum_{i=0}^{k} (-1)^i \binom{k}{i} \binom{\frac{b_0 - ai}{b} + n - 1}{n} b^n \tag{6}$$

$$\mathbf{E}((X_n)^l) = a^l \frac{\Gamma(\frac{a_0 + la}{a})\Gamma(\frac{t_0}{b})}{\Gamma(\frac{a_0}{a})\Gamma(\frac{t_0 + la}{b})} n^{la/b} + O(n^{(l-1)a/b}). \tag{7}$$

We are also interested in the local limit law of the random variable X_n. The density can be expressed using Mittag–Leffler functions so that the limit law is a stable law lookalike. More precisely, for any positive x such that $xn^{a/b}$ is an integer we have the equivalent

$$\mathbf{P}(X_n = a_0 + axn^{a/b}) = \frac{1}{n^{a/b}} \frac{\Gamma(\frac{t_0}{b})}{\Gamma(\frac{a_0}{a})} x^{\frac{a_0}{a} - 1} \sum_{k \geq 0} \frac{(-1)^k}{\Gamma(\frac{b_0 - ka}{b})} \frac{x^k}{k!} + O\left(\frac{1}{n^{2a/b}}\right).$$

5. **Triangular Urns of Size 3**

We can easily adapt the generating function approach to urns filled with balls of three distinct colors. The matrix is then 3×3. We still ask for a triangular and balanced matrices. The methods are the same as for the two color case but they are original contributions from Puyhaubert since the probabilistic method needs additional constraints (irreducibility) to deal with 3×3 matrices.

Balanced 3×3 triangular replacement matrices (for balance 2 and 3) are classified according to the asymptotic growth of their three populations.

The generating function methodology leads to results for matrices of arbitrary size but the computations, especially to determine first integrals become heavier.

Bibliography

[1] Cartan (Henri). – *Calcul différentiel.* – Hermann, 1967.

[2] Eggenberger (Florian) and Pólya (George). – Über die Statistik verketteter Vorgänge. *Zeitschrift für angewandte Mathematik und Mechanik*, vol. 1, 1923, pp. 279–289.

[3] Flajolet (Philippe), Gabarró (Joaquim), and Pekari (Helmut). – Analytic urns. *Annals of Probability*, 2005.

[4] Puyhaubert (Vincent). – *Modèles d'urnes et phénomènes de seuil en combinatoire analytique.* – PhD thesis, École polytechnique, 2005.