

Pólya urn models: exercises

1 Exercises

Exercise 1.

When $R = \begin{pmatrix} 0 & 3 \\ 2 & 1 \end{pmatrix}$, code and count all histories of length 2 leading from $\begin{pmatrix} 2 \\ 0 \end{pmatrix}$ to $\begin{pmatrix} 4 \\ 4 \end{pmatrix}$.

Exercise 2.

Whenever $R = SI_2$, compute all numbers H_n , $n \geq 0$. [This urn is Pólya's original one in his article published in 1930.]

Exercise 3.

For any urn, if $N = \alpha + \beta$, show that the total number of histories of length n starting from $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ equals $N(N+S)(N+2S) \dots (N+(n-1)S) = n!S^n \binom{\frac{N}{S} + n - 1}{n}$.

Exercise 4. For any urn (*i.e.* for any R), show that

$$H\left(1, 1, z \mid \begin{matrix} u_0 \\ v_0 \end{matrix}\right) = \left(\frac{1}{1 - Sz}\right)^{\frac{u_0 + v_0}{S}}.$$

Exercise 5.

For the original urn ($R = SI_2$),

$$H\left(x, y, z \mid \begin{matrix} u_0 \\ v_0 \end{matrix}\right) = \frac{x^{u_0} y^{v_0}}{(1 - Szx^S)^{\frac{u_0}{S}} (1 - Szy^S)^{\frac{v_0}{S}}}.$$

Exercise 6.

All notations are those of Section 3.

6.1- (Linear functions)

Show that if V is a real vector space and if $f : \mathbb{R}^2 \rightarrow V$ is linear, then

$$\Phi(f) = f \circ A$$

where $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is defined by $A(v) = A \begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} := {}^t R \begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} v^{(1)} \\ v^{(2)} \end{pmatrix}$.

6.2- (Vector-valued martingale)

Denote $\tau := \alpha + \beta$. Show that the process $\left(\gamma_{n,\tau} ({}^t R)^{-1} (U_n)\right)_n$ is a martingale (with regard to the natural filtration) as soon as it is defined, *i.e.* as soon as all matrices $I_2 + \frac{1}{kS+\tau}R$, $k \in \mathbb{N}$ are invertible. Show that this martingale is not defined if, and only if $m \leq -1$ and S divides $m + \alpha + \beta$.

6.3- (Expectation)

Let u_1 and u_2 be the eigenforms defined in Section 1. Verify (or remember!) that $u_1 \circ A = Su_1$ and $u_2 \circ A = mu_2$. Show that for any $n \in \mathbb{N}$,

$$\mathbf{E}u_1(U_n) = n + \frac{\tau}{S}$$

and, when n tends to infinity,

$$\mathbf{E}u_2(U_n) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S} + \sigma\right)} \frac{b\alpha - c\beta}{S} n^\sigma \left(1 + O\left(\frac{1}{n}\right)\right).$$

When $R \neq SI_2$, using that $v = u_1(v)v_1 + u_2(v)v_2$ for any vector $v \in \mathbb{R}^2$, show that, when n tends to infinity,

$$\mathbf{E}U_n \sim nv_1$$

Example of 2-3 trees.

6.4- (Real-valued projected martingales)

Show that

$$\left(\frac{u_1(U_n)}{nS + \tau}\right)_n$$

is an almost surely bounded (thus convergent) martingale and compute its expectation. Show that

$$\left(\frac{u_2(U_n)}{\gamma_{n,\tau}(m)}\right)_n$$

is a martingale as well, as soon as $m \geq 0$ or $m + \tau$ is not a multiple of S .

6.5- (Second moments)

Denote by P and Q the 2-variable polynomials defined by

$$P(x, y) = u_1(x, y) \left(u_1(x, y) + 1\right) \quad \text{and} \quad Q(x, y) = \left(u_1(x, y) + \sigma\right) u_2(x, y).$$

Show that $\Phi(P) = 2SP$ and $\Phi(Q) = (S + m)Q$ and prove the asymptotics when n tends to infinity

$$\mathbf{E}P(U_n) = n^2 \left(1 + O\left(\frac{1}{n}\right)\right)$$

and

$$\mathbf{E}Q(U_n) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S} + \sigma\right)} \frac{b\alpha - c\beta}{S} n^{1+\sigma} \left(1 + O\left(\frac{1}{n}\right)\right)$$

(if one feels depressed, one can just show that $Q(U_n) \in O(n^{1+\sigma})$, :-)).

Suppose that $\sigma \neq 1/2$ and denote

$$R = u_2^2 - \frac{bc\sigma^2}{1-2\sigma}u_1 + (b-c)\sigma u_2.$$

Show that, in this case, $\Phi(R) = 2mR$ and that, when n tends to infinity,

$$\mathbf{E}R(U_n) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S} + 2\sigma\right)} R(\alpha, \beta) n^{2\sigma} \left(1 + O\left(\frac{1}{n}\right)\right).$$

Show that $(1, u_1, u_2, P, Q, R)$ is a basis of the vector space $\mathbb{R}_2[x, y]$ of polynomials of degree less than or equal to 2. Write x^2 , xy and y^2 in this basis and compute the asymptotics of the co-moment matrix $\mathbf{E}[U_n^t U_n]$ and of the covariance matrix $\mathbf{E}[(U_n - \mathbf{E}U_n)^t (U_n - \mathbf{E}U_n)]$ (one has to discuss whether $\sigma < 1/2$ or $\sigma > 1/2$).

Check what happens when $\sigma = 1/2$ and do the same job using $T = u_2^2 + \frac{2b-m}{2}u_2$ instead of R .

6.6- (For large urns, the second projected martingale is square-bounded)

Suppose that $\sigma > 1/2$. Expressing u_2^2 as a function of R , u_1 and u_2 , show that the martingale $\left(\frac{u_2(U_n)}{\gamma_{n,\tau}(m)}\right)_n$ is bounded in L^2 , thus convergent.

Exercise 7 (triangular urn).

Assume that $b = 0$, so that $R = \begin{pmatrix} S & 0 \\ S-m & m \end{pmatrix}$. Assume also that the initial number of black balls is non zero, *i.e.* that $\beta \neq 0$ (and check that $\beta = 0$ leads to a degenerate process). Let as above u_1 and u_2 be the linear forms

$$u_1(x, y) = \frac{x+y}{S} \quad \text{and} \quad u_2(x, y) = \frac{y}{S}.$$

For any $p \in \mathbb{N}^*$, let also A_p and B_p be the bivariate polynomials

$$A_p = u_1(u_1 + 1) \dots (u_1 + p - 1) = \frac{\Gamma(u_1 + p)}{\Gamma(u_1)}$$

and

$$B_p = u_2(u_2 + \sigma) \dots (u_2 + (p-1)\sigma) = \frac{\Gamma(u_2 + p\sigma)}{\Gamma(u_2)}.$$

Show that $\Phi(A_p) = pSA_p$ (as always, even if R is not triangular) and that $\Phi(B_p) = pmB_p$ for any $p \geq 1$. Deduce from this that, when n tends to infinity,

$$\mathbf{E}B_p(U_n) = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S} + p\sigma\right)} \frac{\Gamma\left(\frac{\beta}{S} + p\sigma\right)}{\Gamma\left(\frac{\beta}{S}\right)} n^{p\sigma} \left(1 + O\left(\frac{1}{n}\right)\right).$$

- Assume that $m \geq 1$.

Using the inversion formula

$$u_2^p = \sum_{k=1}^p (-\sigma)^{p-k} \left\{ \begin{matrix} p \\ k \end{matrix} \right\} B_k,$$

show that, for any $p \geq 1$,

$$\lim_{n \rightarrow \infty} \mathbf{E} \left(\frac{u_2(U_n)}{n^\sigma} \right)^p = \frac{\Gamma\left(\frac{\tau}{S}\right)}{\Gamma\left(\frac{\tau}{S} + p\sigma\right)} \frac{\Gamma\left(\frac{\beta}{S} + p\sigma\right)}{\Gamma\left(\frac{\beta}{S}\right)}. \quad (1)$$

so that the number of black balls $U_n^{(2)} = Su_2(U_n)$ converges in law to a random variable having the right side of Equality (1) as p -th moment (to make a complete proof of that fact, one has to check that a distribution having such a p -th moment is determined by its moments, which can be done by computing the asymptotics of (1) as p tends to infinity with the help of Stirling Formula). This law can be related to stable laws or to Mittag-Leffler ones.

- Assume that $m = 0$. Show that the process is deterministic (degenerate case).
- Assume that $m \leq -1$. Show that the number of black balls tends almost surely to zero (degenerate case again).

2 Hints

Exercise 1.

One possible solution: start from $W_0 = \mathbf{r}^2$. One can draw a tree of all possibilities: $W_1 \in \{\mathbf{r}\mathbf{b}^3\mathbf{r}, \mathbf{r}^2\mathbf{b}^3\}$, then $W_2 \in \{\mathbf{r}\mathbf{n}^6\mathbf{r}, \mathbf{r}^3\mathbf{n}^4\mathbf{r}, \mathbf{r}\mathbf{n}\mathbf{r}^2\mathbf{n}^3\mathbf{r}, \mathbf{r}\mathbf{n}^2\mathbf{r}^2\mathbf{n}^2\mathbf{r}, \mathbf{r}\mathbf{n}^3\mathbf{r}\mathbf{n}^3\}$ ou $W_2 \in \{\mathbf{r}\mathbf{b}^3\mathbf{r}\mathbf{b}^3, \mathbf{r}^2\mathbf{b}^6, \mathbf{r}^4\mathbf{b}^4, \mathbf{r}^2\mathbf{b}\mathbf{r}^2\mathbf{b}^3, \mathbf{r}^2\mathbf{b}^2\mathbf{r}^2\mathbf{b}^2\}$.

Amongst the ten histories of length 2, six of them lead to $\binom{4}{4}$ and four lead to $\binom{2}{6}$: starting from two red balls, the probability that the urn contains four red balls and four black ones after two drawings is $3/5$.

Beware: in the example, the configuration $\mathbf{r}\mathbf{n}^3\mathbf{r}\mathbf{n}^3$ is reached by two different histories. We count histories, not the different word that are potentially obtained.

Exercise 2

This is elementary enumerative combinatorics. Make the picture of a path in \mathbb{N}^2 and count the histories that follow each of these paths. For any $(p, q) \in \mathbb{N}^2$ such that $p + q = n$, one gets

$$\begin{aligned} H_n \left(\begin{matrix} \alpha & \alpha + pS \\ \beta & \beta + qS \end{matrix} \right) &= \binom{n}{p} \alpha(\alpha + S) \dots (\alpha + (p-1)S) \beta(\beta + S) \dots (\beta + (q-1)S) \\ &= n! S^n \binom{\frac{\alpha}{S} + p - 1}{p} \binom{\frac{\beta}{S} + q - 1}{q}; \end{aligned}$$

all others H_n vanish.

Exercise 5.

Computations on multivariate power series, based on the formula $\frac{1}{(1-X)^N} = \sum_{n \geq 0} \binom{N+n-1}{n} X^n$.

Exercise 6.

6.2- Apply **6.1-** to $f = \text{Id}$. This leads to $\mathbf{E}(U_{n+1}|U_n) = \left(I_2 + \frac{1}{nS+\tau}A\right)(U_n)$. This implies all the answers, because A is diagonalizable, with eigenvalues S and m .

6.3- An induction using **6.1-** leads to $\mathbf{E}u_1(U_n) = \gamma_{n,\tau}(S) \times u_1(U_0) = \frac{nS+\tau}{\tau} \times \frac{\tau}{S}$. For u_2 , apply Stirling Formula to $\mathbf{E}u_2(U_n) = \gamma_{n,\tau}(m) \times u_2(U_0)$ with a O -remainder. The third assertion is obtained by addition of asymptotic developments.

6.4- Using **6.1-** again, one gets $\mathbf{E}(u_1(U_{n+1})|U_n) = \left(1 + \frac{S}{nS+\tau}\right) \times u_1(U_n)$, so that $\mathbf{E}\left(\frac{u_1(U_{n+1})}{(n+1)S+\tau} | U_n\right) = \frac{u_1(U_n)}{nS+\tau}$, proving the martingale property. Same argument from $\mathbf{E}(u_2(U_{n+1})|U_n) = \left(1 + \frac{m}{nS+\tau}\right) \times u_2(U_n)$.

6.5- One gets $\Phi(P)$, $\Phi(Q)$ and $\Phi(R)$ by simple computation. Since $\Phi(P) = 2SP$, $\mathbf{E}P(U_n) = \gamma_{n,\tau}(2S) \times P(U_0)$ and the required asymptotics for $\mathbf{E}P(U_n)$ is obtained thanks to Stirling Formula. *Idem* for $\mathbf{E}Q(U_n)$ and $\mathbf{E}R(U_n)$. The remainder of the exercise is completely left to the reader.

6.6- $u_2^2 = R + \frac{bc\sigma^2}{1-2\sigma}u_1 - (b-c)\sigma u_2$, so that $\mathbf{E}u_2^2(U_n) = c_1n^{2\sigma}(1 + O(1/n)) + c_2n + c_3n^\sigma(1 + O(1/n))$ where c_1 , c_2 and c_3 are constants. Since $\sigma > 1/2$, the principal term is the one in $n^{2\sigma}$, proving that the martingale is square bounded (use Stirling formula again to get the asymptotics of $\gamma_{n,\tau}(m)^2$).