

Pólya Urn Models

Stefan Grosskinsky
Warwick Mathematics Institute

TU Delft, September 21, 2019

These notes and supplementary material are available on
<https://github.com/sgrosskinsky/urns>

Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

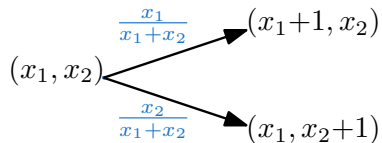
Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$

Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$



$$= \mathbb{P}[\underline{X}(n+1) = (x_1, x_2 + 1) | \underline{X}(n) = \underline{x}]$$

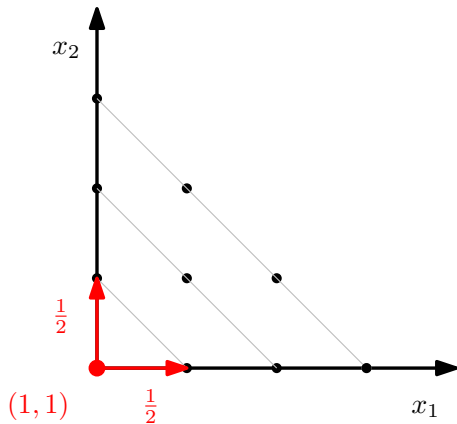
Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$

$$\begin{aligned} (x_1, x_2) &\begin{cases} \xrightarrow{\frac{x_1}{x_1+x_2}} (x_1+1, x_2) \\ \xrightarrow{\frac{x_2}{x_1+x_2}} (x_1, x_2+1) \end{cases} \\ &= \mathbb{P}[\underline{X}(n+1) = (x_1, x_2+1) | \underline{X}(n) = \underline{x}] \end{aligned}$$



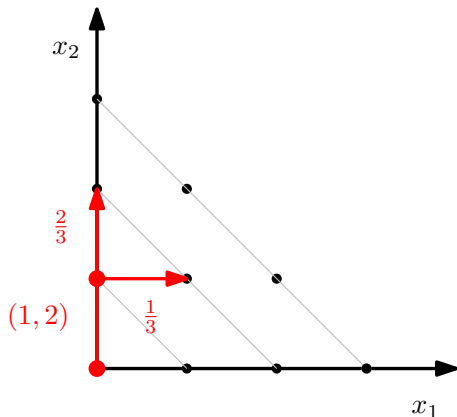
Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$

$$\begin{aligned} & \begin{array}{l} \xrightarrow{\frac{x_1}{x_1+x_2}} (x_1+1, x_2) \\ (x_1, x_2) \xrightarrow{\frac{x_2}{x_1+x_2}} (x_1, x_2+1) \end{array} \\ &= \mathbb{P}[\underline{X}(n+1) = (x_1, x_2+1) | \underline{X}(n) = \underline{x}] \end{aligned}$$



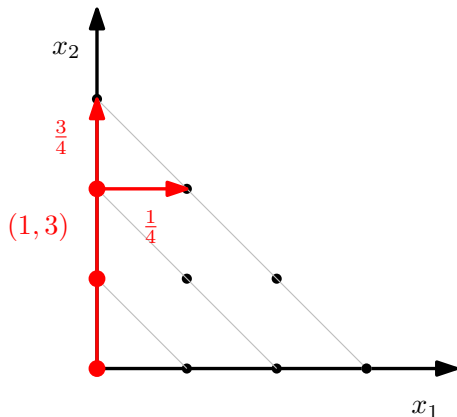
Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$

$$\begin{aligned} (x_1, x_2) &\begin{cases} \xrightarrow{\frac{x_1}{x_1+x_2}} (x_1+1, x_2) \\ \xrightarrow{\frac{x_2}{x_1+x_2}} (x_1, x_2+1) \end{cases} \\ &= \mathbb{P}[\underline{X}(n+1) = (x_1, x_2+1) | \underline{X}(n) = \underline{x}] \end{aligned}$$

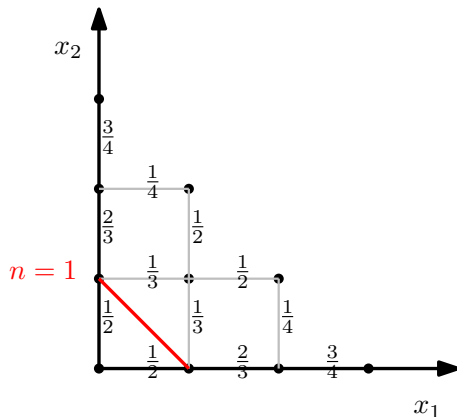
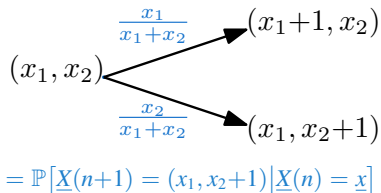


Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$

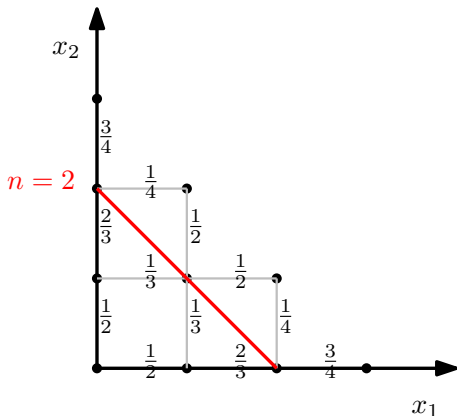
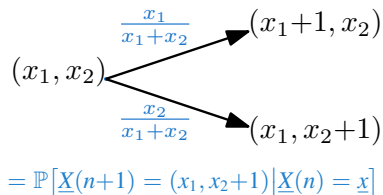


Pólya urn model

State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$

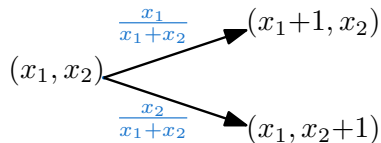


Pólya urn model

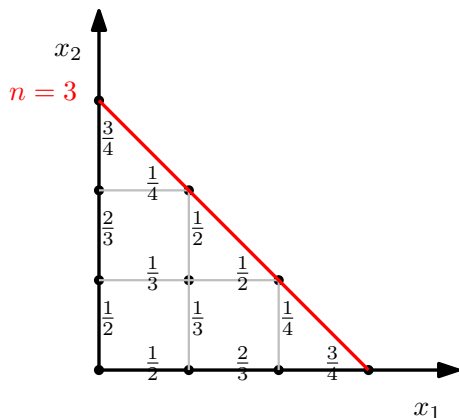
State space $\mathbb{N} \times \mathbb{N}$

Configurations $\underline{x} = (x_1, x_2) \in \mathbb{N}^2$

Random process $(\underline{X}(n) : n = 0, 1, \dots)$ with $\underline{X}(0) = (1, 1)$



$$= \mathbb{P}[\underline{X}(n+1) = (x_1, x_2+1) | \underline{X}(n) = \underline{x}]$$



Pólya urn model

Distribution at time n

For all $n \geq 0$ we have $\mathbb{P}[\underline{X}(n) = \underline{x}] = \frac{1}{n+1} \delta_{|\underline{x}|, n+2}$.

Induction. $\mathbb{P}[\underline{X}(n+1) = (k, n+3-k)] = \frac{1}{n+1} \left[\frac{k-1}{n+2} + \frac{n+2-k}{n+2} \right] = \frac{1}{n+2}$

Pólya urn model

Distribution at time n

For all $n \geq 0$ we have $\mathbb{P}[\underline{X}(n) = \underline{x}] = \frac{1}{n+1} \delta_{|\underline{x}|, n+2}$.

Induction. $\mathbb{P}[\underline{X}(n+1) = (k, n+3-k)] = \frac{1}{n+1} \left[\frac{k-1}{n+2} + \frac{n+2-k}{n+2} \right] = \frac{1}{n+2}$

Fraction of type i : $\xi_i(n) = \frac{X_i(n)}{|\underline{X}(n)|} = \frac{X_i(n)}{n+2}$

Pólya urn model

Distribution at time n

For all $n \geq 0$ we have $\mathbb{P}[\underline{X}(n) = \underline{x}] = \frac{1}{n+1} \delta_{|\underline{x}|, n+2}$.

Induction. $\mathbb{P}[\underline{X}(n+1) = (k, n+3-k)] = \frac{1}{n+1} \left[\frac{k-1}{n+2} + \frac{n+2-k}{n+2} \right] = \frac{1}{n+2}$

Fraction of type i : $\xi_i(n) = \frac{X_i(n)}{|\underline{X}(n)|} = \frac{X_i(n)}{n+2}$

Scaling limit

$(\xi_1, \xi_2)(n) \rightarrow (U, 1-U)$ in distribution as $n \rightarrow \infty$,

where U is a uniform random variable on $[0, 1]$.

Proof. $\mathbb{P}[\xi_1(n) \leq u] = \mathbb{P}[X_1(n) \leq u(n+2)] = \frac{1}{n+1} \lfloor u(n+2) \rfloor \rightarrow u, \quad u \in (0, 1).$

Extensions/Applications

[H.M. Mahmoud, Pólya urn models. CRC Press. 2008]

- $k > 2$ **types** of balls $\rightarrow k + n$ balls at time $n \geq 0$
- different **initial conditions** $\underline{a} \in \mathbb{N}^k \rightarrow$ **Dirichlet multinomial**

$$\mathbb{P}[\underline{X}(n) - \underline{a} = \underline{x}] \propto \prod_{i=1}^k \frac{\Gamma(x_i + a_i)}{x_i! \Gamma(a_i)} \delta_{|\underline{x}|, n}$$

- generalized **schemes** for adding balls
- a simple model for growth processes with **linear reinforcement**
can describe biological growth, evolution of market shares, ...

Extensions/Applications

[H.M. Mahmoud, Pólya urn models. CRC Press. 2008]

- $k > 2$ **types** of balls $\rightarrow k + n$ balls at time $n \geq 0$
- different **initial conditions** $\underline{a} \in \mathbb{N}^k \rightarrow$ **Dirichlet multinomial**

$$\mathbb{P}[\underline{X}(n) - \underline{a} = \underline{x}] \propto \prod_{i=1}^k \frac{\Gamma(x_i + a_i)}{x_i! \Gamma(a_i)} \delta_{|\underline{x}|, n}$$

- generalized **schemes** for adding balls
- a simple model for growth processes with **linear reinforcement** can describe biological growth, evolution of market shares, ...

Generalized Pólya Urn/Balls in boxes $(\underline{X}(n) : n \geq 0)$ on \mathbb{N}^k

$$\mathbb{P}[\underline{X}(n+1) = \underline{x} + \underline{e}_i | \underline{X}(n) = \underline{x}] = \frac{f_i x_i^\gamma}{\sum_{j=1}^k f_j x_j^\gamma}$$

with **fitness** $f_i > 0$ and **reinforcement parameter** $\gamma \geq 0$

[Oliviera, RS&A **34**(4): 454 (2009); Jiang et al., Proc. 2016 ACM SIGMETRICS, arxiv:1604.02097]

Generalized Pólya Urn Models

- **No reinforcement** $\gamma = 0, p_i = f_i / \sum_j f_j \rightarrow$ Multinomial

$$\mathbb{P}[\underline{X}(n) - \underline{X}(0) = \underline{x}] = \frac{n!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} \delta_{|\underline{x}|, n}$$

Generalized Pólya Urn Models

- **No reinforcement** $\gamma = 0, p_i = f_i / \sum_j f_j \rightarrow$ **Multinomial**

$$\mathbb{P}[\underline{X}(n) - \underline{X}(0) = \underline{x}] = \frac{n!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} \delta_{|\underline{x}|, n}$$

weak LLN and CLT

For all $i = 1, \dots, k$, as $n \rightarrow \infty$ in distribution

$$\xi_i(n) = \frac{X_i(n)}{|\underline{X}(n)|} \rightarrow p_i \quad \text{and} \quad \frac{X_i(n) - p_i n}{\sqrt{np_i(1 - p_i)}} \rightarrow \mathcal{N}(0, 1) .$$

Proof. $X_i(n) = X_i(0) + \sum_{t=1}^n Y_t$ with Y_1, Y_2, \dots i.i.d. $\text{Be}(p_i)$

Generalized Pólya Urn Models

- **No reinforcement** $\gamma = 0, p_i = f_i / \sum_j f_j \rightarrow$ **Multinomial**

$$\mathbb{P}[\underline{X}(n) - \underline{X}(0) = \underline{x}] = \frac{n!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n} \delta_{|\underline{x}|, n}$$

weak LLN and CLT

For all $i = 1, \dots, k$, as $n \rightarrow \infty$ in distribution

$$\xi_i(n) = \frac{X_i(n)}{|\underline{X}(n)|} \rightarrow p_i \quad \text{and} \quad \frac{X_i(n) - p_i n}{\sqrt{np_i(1 - p_i)}} \rightarrow \mathcal{N}(0, 1) .$$

Proof. $X_i(n) = X_i(0) + \sum_{t=1}^n Y_t$ with Y_1, Y_2, \dots i.i.d. $\text{Be}(p_i)$

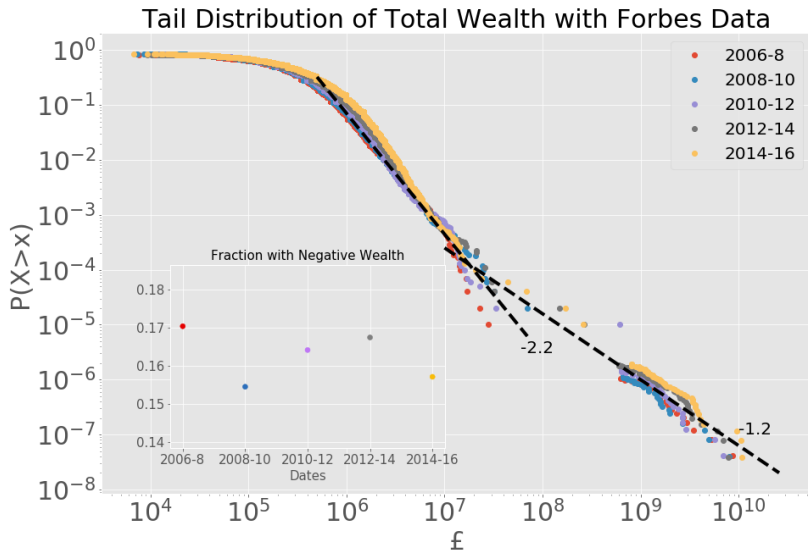
- **Superlinear reinforcement** $\gamma > 1, f_i = 1$

limiting behaviour

$$m(n) := \max_i X_i(n) \rightarrow \infty \quad \text{and} \quad n - m(n) \rightarrow Y < \infty$$

in distribution as $n \rightarrow \infty$.

Wealth distribution in UK



Wealth distribution in UK

