The Pólya Web

TDK dissertation

Author

Ákos Urbán

MSc student, BME

Supervisors

Dr Balázs Bárány

Department of Stochastics, BME

Dr Bálint Tóth

Alfréd Rényi Institute of Mathematics

Budapest University of Technology and Economics

October 2024

Contents

Acknowledgement

I would like to take this opportunity to express my gratitude to Bálint Tóth and Balázs Bárány for their help, advice and insights in more than a year of working together. In addition, for making probability theory my main interest during this time.

1 Preliminary

1.1 Introduction

This year marks the hundredth anniversary of Florian Eggenberger, György Pólya's student, defending his doctoral thesis on Pólya Urns in 1924 after writing an article on it in 1923 [5]. In the hundred years since then, almost everything that could be said on the subject has been said. Nevertheless, in this paper we will try to add a new touch to this rich subject. In doing so, we wish to pay tribute to our great predecessors and to the anniversary.

1.2 The Pólya Urn and the Dual Pólya Urn

The concept of Pólya Urn is well known. The main properties can be found in [9]. We give a short recap about the most important ones. The basic notions and theorems of probability theory which we use in this paper can be found in [4]. Consider an urn with some blue and red balls in it. We we draw a ball at random from the urn chosen uniformly and put back the ball with a new ball from the exact same color. This defines a Markov-chain in the following way.

Definition 1.2.1 (Pólya Urn)**.** *Let* $b_0, r_0 \in \mathbb{N}$ $(b_0 + r_0 > 0)$ *be the initial number of blue and red balls in the urn. We denote the number of blue and red balls after the nth draw with* B_n *and* R_n *with the following transition probalilities*

$$\mathbb{P}(B_0 = b_0, R_0 = r_0) = 1$$

and

$$\mathbb{P}(B_{n+1} = b+1,\ R_{n+1} = r \mid B_n = b,\ R_n = r) = \frac{b}{b+r},$$
$$\mathbb{P}(B_{n+1} = b,\ R_{n+1} = r+1 \mid B_n = b,\ R_n = r) = \frac{r}{b+r}.$$

This concept can be easily interpreted as a random walk on $\mathbb{N} \times \mathbb{N}$ (Figure 1). We call this the Pólya Walk.

We can consider the ratio of blue balls in the urn after n draws. This is the random variable in the form of

$$\xi_n = \frac{B_n}{B_n + R_n}. \tag{1}$$

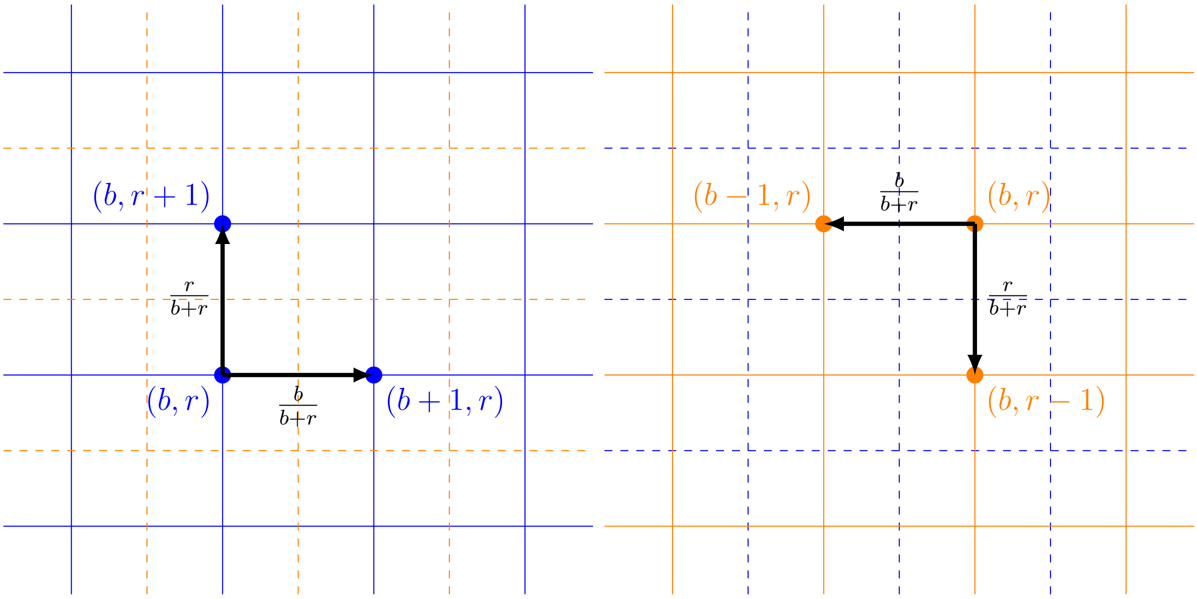

Figure 1: The Pólya Walk (left) and the Dual Pólya Walk (right) on $\mathbb{N} \times \mathbb{N}$.

It is known that ξ_n is a martingale and converges almost surely.

Theorem 1.2.2. *Let ξ_n be the sequence of random variables defined in Equation (1). Then we have the following*

$$\xi_n \overset{a.s.}{\to} \xi, \qquad as \quad n \to \infty,$$

where

$$\xi \sim BETA(b_0, r_0).$$

We use $\sim$ throughout the paper to indicate when two distributions are equal. However if we compare two series, it will mean their ratio tends to one. In this case ξ is a random variable with absolutely continuous distribution with probability density function

$$x \mapsto \frac{\Gamma(b_0 + r_0)}{\Gamma(b_0)\Gamma(r_0)} \cdot x^{b_0-1}(1-x)^{r_0-1} \cdot \mathbb{1}\left[0 \le x \le 1\right].$$

Now let us consider an urn with some blue and red balls in it. In this setting we remove a random ball from the urn chosen uniformly. In this case we can also define a similar Markov-chain.

Definition 1.2.3 (Dual Pólya Urn)**.** *Let $b_0, r_0 \in \mathbb{N}$ be the initial number of blue and red balls in the urn. We denote the number of blue and red balls after removing the nth ball by B_n and R_n with the following transition probabilities for*

$$\mathbb{P}(B_0 = b_0, R_0 = r_0) = 1$$

and for $b + r > 0$

$$\mathbb{P}(B_{n+1} = b - 1, R_{n+1} = r \,|\, B_n = b,\ R_n = r\,) = \frac{b}{b+r},$$
$$\mathbb{P}(B_n = b, R_{n+1} = r - 1 \,|\, B_n = b,\ R_n = r\,) = \frac{r}{b+r}.$$

Also for $b = r = 0$.

$$\mathbb{P}(B_{n+1} = 0, R_{n+1} = 0 \,|\, B_n = 0,\ R_n = 0\,) = 1.$$

We can also assign a random walk on $\mathbb{N} \times \mathbb{N}$ to the reverse Pólya urn (Figure 1). We call this the Dual Pólya Walk.

2 The Pólya Web

2.1 The Pólya Web and its dual

In this chapter we define a coupling of Pólya Walks on $\mathbb{N} \times \mathbb{N}$ in a way that two walks with different starting point stay independent until they meet. We call this construction the Pólya Web. This construction relates to the Random Walk Web introduced by B. Tóth and W. Werner in 1998 [8], but this time, we use Pólya Walks as primary constituents instead of simple symmetric random walks.

Definition 2.1.1 (The Pólya Web). *Let $X_{i,j}$ $(i, j \in \mathbb{N}$ with $i + j > 0)$ be indepentent random vector variables with the following distribution.*

$$\mathbb{P}(X_{i,j} = (1, 0)) = \frac{i}{i + j} \qquad \textit{and} \qquad \mathbb{P}(X_{i,j} = (0, 1)) = \frac{j}{i + j}.$$

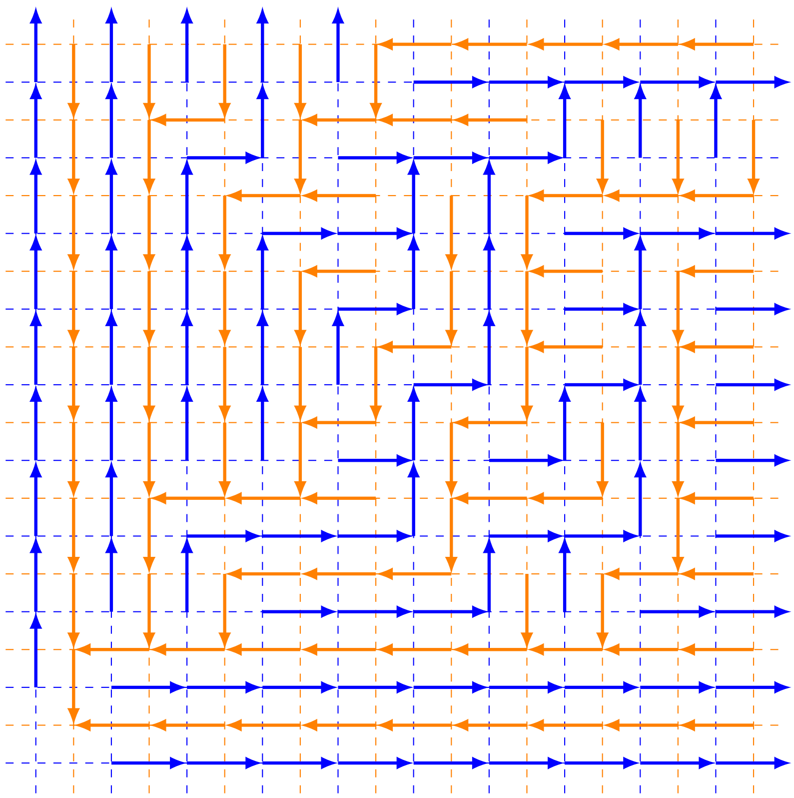

Figure 2: The Pólya Web (blue) and its dual (orange).

Definition 2.1.2 (Pólya Walk started from the pair (i, j)).

$$S^{(n)}_{i,j} := \begin{cases} (i, j) & \textit{if } n = i + j \\ S^{(n-1)}_{i,j} + X_{S^{(n-1)}_{i,j}} & \textit{if } i + j < n \end{cases}$$

Definition 2.1.3 (Dual Pólya Walk started from the pair (i,j))**.**

$$T_{i,j}^{(n)} := \begin{cases} (i,j) & \text{if } n = i+j \\ T_{i,j}^{(n+1)} - X_{T_{i,j}^{(n+1)}} & \text{if } 0 \leq i+j < n \end{cases}$$

Notice that the lower indices denote the starting point of the walk. The upper index denotes the total number of balls, thus it serves as a universal time to compare walks started from different points.

We can observe that the new definition coincides with our previous definitions of the Pólya Walks and Dual Pólya Walks.

Proposition 2.1.4. *The Pólya Walk defined above Definition 2.1.2 has the same distribution as the walk defined as a Markov chain in Definition 1.2.1.*

Proof. For $n = i+j$ we have

$$\mathbb{P}\Big(S_{i,j}^{(i+j)} = (i,j)\Big) = 1.$$

For $n > i+j$ let $(b,r) \in \mathbb{N} \times \mathbb{N}$. Then

$$\mathbb{P}\Big(S_{i,j}^{(n)} = (b+1,r) \,\Big|\, S_{i,j}^{(n-1)} = (b,r)\Big) = \mathbb{P}\Big(S_{i,j}^{(n-1)} + X_{S_{i,j}^{(n-1)}} = (b+1,r) \,\Big|\, S_{i,j}^{(n-1)} = (b,r)\Big)$$

$$= \mathbb{P}((b,r) + X_{b,r} = (b+1,r)) = \mathbb{P}(X_{b,r} = (1,0)) = \frac{b}{r+b},$$

and also

$$\mathbb{P}\Big(S_{i,j}^{(n)} = (b,r+1) \,\Big|\, S_{i,j}^{(n-1)} = (b,r)\Big) = \mathbb{P}\Big(S_{i,j}^{(n-1)} + X_{S_{i,j}^{(n-1)}} = (b,r+1) \,\Big|\, S_{i,j}^{(n-1)} = (b,r)\Big)$$

$$= \mathbb{P}((b,r) + X_{b,r} = (b,r+1)) = \mathbb{P}(X_{b,r} = (0,1)) = \frac{r}{r+b}.$$

■

Proposition 2.1.5. *The Dual Pólya Walk defined above Definition 2.1.3 has the same distribution as the walk defined as a Markov chain in Definition 1.2.3.*

Proof. For $n = i+j$ we have

$$\mathbb{P}\Big(T_{i,j}^{(i+j)} = (i,j)\Big) = 1.$$

For $n > i+j$ let $(b,r) \in \mathbb{N} \times \mathbb{N},\ b+r > 0$. Then

$$\mathbb{P}\Big(T_{i,j}^{(n)} = (b-1,r) \,\Big|\, T_{i,j}^{(n-1)} = (b,r)\Big) = \mathbb{P}\Big(T_{i,j}^{(n-1)} - X_{T_{i,j}^{(n-1)}} = (b-1,r) \,\Big|\, T_{i,j}^{(n-1)} = (b,r)\Big)$$

$$= \mathbb{P}((b,r) - X_{b,r} = (b-1,r)) = \mathbb{P}(X_{b,r} = (1,0)) = \frac{b}{r+b},$$

and also

$$\mathbb{P}\Big(T_{i,j}^{(n)} = (b,r-1) \,\Big|\, T_{i,j}^{(n-1)} = (b,r)\Big) = \mathbb{P}\Big(T_{i,j}^{(n-1)} - X_{T_{i,j}^{(n-1)}} = (b,r-1) \,\Big|\, T_{i,j}^{(n-1)} = (b,r)\Big)$$
$$= \mathbb{P}((b,r) - X_{b,r} = (b,r-1)) = \mathbb{P}(X_{b,r} = (0,1)) = \frac{r}{r+b}.$$

Finally for $b = r = 0$.

$$\mathbb{P}\Big(T_{i,j}^{(n)} = (0,0) \,\Big|\, T_{i,j}^{(n-1)} = (0,0)\Big) = \mathbb{P}\Big(T_{i,j}^{(n-1)} - X_{T_{i,j}^{(n-1)}} = (0,0) \,\Big|\, T_{i,j}^{(n-1)} = (0,0)\Big)$$
$$= \mathbb{P}((0,0) - X_{0,0} = (0,0)) = \mathbb{P}(X_{0,0} = (0,0)) = 1.$$

■

Throughout the paper we will use the following notations. For the number of blue balls we will use the notation $\mathrm{proj}_1 \circ S_{i,j}^{(n)}$, the projection onto the first coordinate. Likewise the number of red balls is denoted by $\mathrm{proj}_2 \circ S_{i,j}^{(n)}$. For the total number of balls we use the 1-norm on $\mathbb{R}^2$. Notice that

$$\mathrm{proj}_1 \circ S_{i,j}^{(n)} + \mathrm{proj}_2 \circ S_{i,j}^{(n)} = \left\|S_{i,j}^{(n)}\right\|_1 = n \tag{2}$$

where $\|\cdot\|_1$ denotes the 1-norm of a vector. We will also use these notations for the dual walk in the same way.

2.2 Orders

In this section, we define partial orderings in the sample space and in the plane so that they are consistent with certain properties of The Pólya Web.

Definition 2.2.1 (Order on the set $\{(0,1),(1,0)\}$)**.** *The* $\Subset$ *relation on the set* $\{(0,1),(1,0)\}$ *is defined in the following way*

$$(0,1) \Subset (0,1), \qquad (0,1) \Subset (1,0), \qquad (1,0) \Subset (1,0).$$

Proposition 2.2.2. *The relation in Definition 2.2.1 is an order on* $\{(0,1),(1,0)\}$*.*

Proof. Trivial. ■

From now on let us denote

$$\Omega = \{(0,1),(1,0)\}^{\mathbb{N}\times\mathbb{N}\setminus\{(0,0)\}}. \tag{3}$$

Definition 2.2.3 (Partial order on the set Ω)**.** *For any $\omega, \omega' \in \Omega$ we say that*

$$\omega \mathrel{\varepsilon\!\!-} \omega'$$

if and only if for any $i, j \in \mathbb{N}$ having $i + j > 0$

$$\omega_{i,j} \mathrel{\varepsilon\!\!-} \omega'_{i,j}.$$

Proposition 2.2.4. *The relation in Definition 2.2.3 is a partial order on Ω.*

Proof. Let us check the properties

- Reflexivity: Let $\omega \in \Omega$, then

  $$\omega_{i,j} \mathrel{\varepsilon\!\!-} \omega_{i,j},$$

  thus

  $$\omega \mathrel{\varepsilon\!\!-} \omega.$$

- Antisymmetry: Let $\omega, \omega' \in \Omega$ such that $\omega \mathrel{\varepsilon\!\!-} \omega'$ and $\omega' \mathrel{\varepsilon\!\!-} \omega$. Then

  $$\omega_{i,j} \mathrel{\varepsilon\!\!-} \omega'_{i,j} \quad \text{and} \quad \omega'_{i,j} \mathrel{\varepsilon\!\!-} \omega_{i,j},$$

  thus $\omega_{i,j} = \omega'_{i,j}$, which means

  $$\omega = \omega'.$$

- Transitivity: Let $\omega, \omega', \omega'' \in \Omega$ such that $\omega \mathrel{\varepsilon\!\!-} \omega'$ and $\omega' \mathrel{\varepsilon\!\!-} \omega''$. Then

  $$\omega_{i,j} \mathrel{\varepsilon\!\!-} \omega'_{i,j} \quad \text{and} \quad \omega'_{i,j} \mathrel{\varepsilon\!\!-} \omega''_{i,j},$$

  which implies

  $$\omega_{i,j} \mathrel{\varepsilon\!\!-} \omega''_{i,j},$$

  thus by definition

  $$\omega \mathrel{\varepsilon\!\!-} \omega''.$$

■

Definition 2.2.5 (Partial order on the plane)**.** *For $(i, j), (k, l) \in \mathbb{N} \times \mathbb{N}$ we say that*

$$(i, j) \succ (k, l)$$

if and only if

$$i \leq k \quad \textit{and} \quad j \geq l.$$

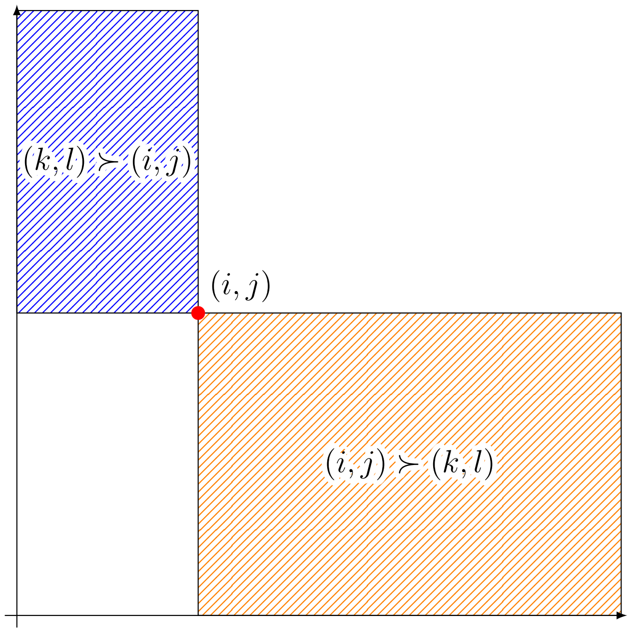

Figure 3: Partial order of the plane defined in Definition 2.2.5.

Proposition 2.2.6. *The relation defined in Definition 2.2.5 is a partial order on* $\mathbb{N} \times \mathbb{N}$*.*

Proof. We should check the properties one by one.

- Reflexivity: Let $(i,j) \in \mathbb{N} \times \mathbb{N}$. Then clearly

$$i \leq i \quad \text{and} \quad j \leq j,$$

thus by definition

$$(i,j) \succ (i,j).$$

- Antisymmetry: Suppose for $(i,j),(k,l) \in \mathbb{N} \times \mathbb{N}$ we have $(i,j) \succ (k,l)$ and $(i,j) \prec (k,l)$. Then by definition

$$i \leq k \text{ and } j \geq l \qquad \text{and} \qquad k \leq i \text{ and } l \geq j,$$

thus $i = k$ and $j = l$, which implies

$$(i,j) = (k,l).$$

- Transitivity: Suppose for $(i,j),(k,l),(m,n) \in \mathbb{N} \times \mathbb{N}$ we have $(i,j) \succ (k,l)$ and $(k,l) \succ (m,n)$. Then

$$i \leq k \text{ and } j \geq l \qquad \text{and} \qquad k \leq m \text{ and } l \geq n,$$

thus

$$i \leq m \quad \text{and} \quad j \geq n$$

which is by definition

$$(i,j) \succ (m,n).$$

■

We now point out a monotonic property of the previously defined ordering with respect to the Pólya Walk.

Proposition 2.2.7. *The random variable* $\mathrm{proj}_1 \circ S_{i,j}^{(n)}$ *is a decreasing, while* $\mathrm{proj}_2 \circ S_{i,j}^{(n)}$ *is an increasing function of* ω *with respect to the partial order on* Ω *defined in Definition 2.2.3 and the usual order on* $\mathbb{N}$.

Proof. We will only prove it for $\mathrm{proj}_1 \circ S_{i,j}^{(n)}$ by induction on $n \in \mathbb{N}$. The proof for $\mathrm{proj}_2 \circ S_{i,j}^{(n)}$ is completely similar. Let us suppose $\omega \succeq \omega'$.
For $n = i + j$ we have

$$\mathrm{proj}_1 \circ S_{i,j}^{(i+j)}(\omega) = i \leq i = \mathrm{proj}_1 \circ S_{i,j}^{(i+j)}(\omega').$$

Suppose for some n we have.

$$\mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega) \leq \mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega')$$

Then for $n+1$ let us consider the following two cases.

1. Suppose $S_{i,j}^{(n)}(\omega) = S_{i,j}^{(n)}(\omega')$. Then

$$\begin{aligned}
\mathrm{proj}_1 \circ S_{i,j}^{(n+1)}(\omega) &= \underbrace{\mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega)}_{=\mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega')} + \underbrace{\mathbb{1}\left[\omega_{S_{i,j}^{(n)}(\omega)} = (1,0)\right]}_{=\mathbb{1}\left[\omega_{S_{i,j}^{(n)}(\omega')} = (1,0)\right]} \\
&= \mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega') + \underbrace{\mathbb{1}\left[\omega_{S_{i,j}^{(n)}(\omega')} = (1,0)\right]}_{\leq \mathbb{1}\left[\omega'_{S_{i,j}^{(n)}(\omega')} = (1,0)\right]} \\
&\leq \mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega') + \mathbb{1}\left[\omega'_{S_{i,j}^{(n)}(\omega')} = (1,0)\right] = \mathrm{proj}_1 \circ S_{i,j}^{(n+1)}(\omega').
\end{aligned}$$

2. Suppose $S_{i,j}^{(n)}(\omega) \neq S_{i,j}^{(n)}(\omega')$. Then by the induction hypothesis

$$\mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega) < \mathrm{proj}_1 \circ S_{i,j}^{(n)}(\omega'),$$

thus

$$\begin{aligned}
&\text{proj}_1 \circ S_{i,j}^{(n+1)}(\omega) - \text{proj}_1 \circ S_{i,j}^{(n+1)}(\omega') \\
&= \underbrace{\text{proj}_1 \circ S_{i,j}^{(n)}(\omega) - \text{proj}_1 \circ S_{i,j}^{(n)}(\omega')}_{\leq -1} + \underbrace{\mathbb{1}\left[\omega_{S_{i,j}^{(n)}(\omega)} = (1,0)\right] - \mathbb{1}\left[\omega'_{S_{i,j}^{(n)}(\omega')} = (1,0)\right]}_{\leq 1} \\
&\leq 1 - 1 = 0.
\end{aligned}$$

■

Proposition 2.2.8. *The random variables* $\text{proj}_1 \circ S_{i,j}^{(n)}$ *and* $\text{proj}_2 \circ S_{i,j}^{(n)}$ *are an increasing function of* n *with respect to the the usual order on* $\mathbb{N}$*.*

Proof. We will only prove it for $\text{proj}_1 \circ S_{i,j}^{(n)}$. The proof for $\text{proj}_2 \circ S_{i,j}^{(n)}$ is analogous.

$$\text{proj}_1 \circ S_{i,j}^{(n+1)}(\omega) = \text{proj}_1 \circ S_{i,j}^{(n)}(\omega) + \underbrace{\mathbb{1}\left[\omega_{S_{i,j}^{(n)}(\omega)} = (1,0)\right]}_{\geq 0} \geq \text{proj}_1 \circ S_{i,j}^{(n)}(\omega)$$

■

2.3 Properties of the trajectories

An important observation is the fact that if two Pólya Walks meet at one point they will stay together for the rest of the time.

Lemma 2.3.1. *If there exists* $n \in \mathbb{N}$ *such that*

$$S_{i,j}^{(n)} = S_{k,l}^{(n)},$$

then for any $m \in \mathbb{N}$

$$S_{i,j}^{(n+m)} = S_{k,l}^{(n+m)}.$$

Proof. We will finish the proof using induction on $m \in \mathbb{N}$. For $m = 0$

$$S_{i,j}^{(n+0)} = \underbrace{S_{i,j}^{(n)} = S_{k,l}^{(n)}}_{\text{by our assumption}} = S_{k,l}^{(n+0)}.$$

Suppose it is true for some $m \in \mathbb{N}$ then

$$S_{i,j}^{(n+m+1)} = \underbrace{S_{i,j}^{(n+m)} + X_{S_{i,j}^{(n+m)}} = S_{k,l}^{(n+m)} + X_{S_{k,l}^{(n+m)}}}_{\text{since } S_{i,j}^{(n+m)} = S_{k,l}^{(n+m)} \text{ by the induction hypothesis}} = S_{k,l}^{(n+m+1)}.$$

■

The next lemma connects the order defined in Definition 2.2.5 with the paths of the trajectories of the random walks. Namely for any $(i, j) \succ (k, l)$ the trajectory of the walk started from (i, j) stays above the path started from (k, l). Thus the Pólya Walk preserves the order of the starting points (Figure 4).

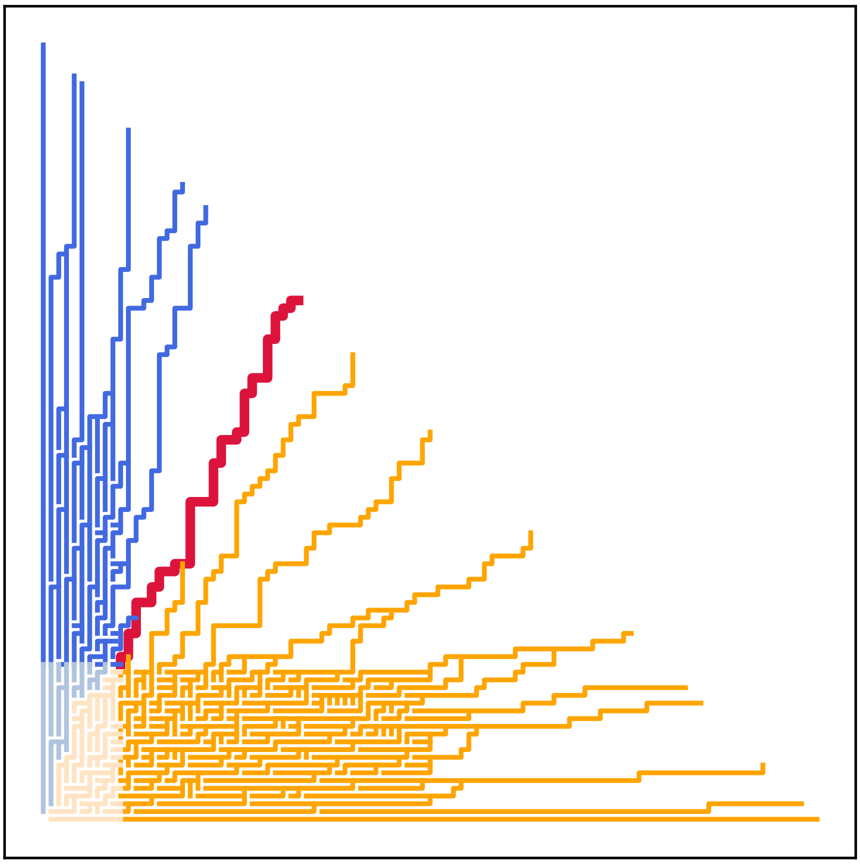

Figure 4: The trajectory of the random walk started from (i, j) marked by red. The trajectories started from $(i, j) \succ (k.l)$ marked dark orange, while $(k, l) \succ (i, j)$ marked dark blue. Notice that if we cannot compare a point to (i, j) (marked with light color) then the Pólya Walk started from it can have different outcomes with respect to the ordering.

Lemma 2.3.2. *For any*

$$(i, j) \succ (k, l)$$

and for any $n \in \mathbb{N}$ we have that

$$S_{i,j}^{(n)} \succ S_{k,l}^{(n)}.$$

Proof. We will prove it by induction on $n \in \mathbb{N}$.

First suppose $i+j \leq k+l$ and let $i+j \leq n \leq k+l$. Then

$$\text{proj}_1 \circ S_{i,j}^{(n)} = i + \underbrace{\sum_{m=i+j}^{n-1} \mathbb{1}\left[X_{S_{i,j}^{(m)}} = (1,0)\right]}_{\leq (k+l)-(i+j)} \leq k + \underbrace{(l-j)}_{\leq 0} \leq k = \text{proj}_1 \circ S_{k,l}^{(n)},$$

$$\text{proj}_2 \circ S_{i,j}^{(n)} = j + \underbrace{\sum_{m=i+j}^{n-1} \mathbb{1}\left[X_{S_{i,j}^{(m)}} = (0,1)\right]}_{\geq 0} \geq j \geq l = \text{proj}_2 \circ S_{k,l}^{(n)}.$$

Now suppose $i+j \geq k+l$ and let $k+l \leq n \leq i+j$. Then

$$\text{proj}_1 \circ S_{k,l}^{(n)} = k + \underbrace{\sum_{m=k+l}^{n-1} \mathbb{1}\left[X_{S_{k,l}^{(m)}} = (1,0)\right]}_{\geq 0} \geq k \geq i = \text{proj}_1 \circ S_{i,j}^{(n)}.$$

$$\text{proj}_2 \circ S_{k,l}^{(n)} = l + \underbrace{\sum_{m=k+l}^{n-1} \mathbb{1}\left[X_{S_{k,l}^{(m)}} = (0,1)\right]}_{\leq (i+j)-(k+l)} \leq j + \underbrace{(i-k)}_{\leq 0} \leq j = \text{proj}_1 \circ S_{i,j}^{(n)},$$

We concluded that for $n = \max\{i+j, k+l\}$ we have that

$$S_{i,j}^{(n)} \succ S_{k,l}^{(n)}.$$

Now suppose it is true for some $n \geq \max\{i+j, k+l\}$ we have that $S_{i,j}^{(n)} \succ S_{k,l}^{(n)}$. Consider the following two cases

1. $\text{proj}_1 \circ S_{i,j}^{(n)} < \text{proj}_1 \circ S_{k,l}^{(n)}$, then

$$\text{proj}_1 \circ S_{i,j}^{(n+1)} - \text{proj}_1 \circ S_{k,l}^{(n+1)}$$

$$= \underbrace{\text{proj}_1 \circ S_{i,j}^{(n)} - \text{proj}_1 \circ S_{k,l}^{(n)}}_{\leq -1} + \underbrace{\mathbb{1}\left[X_{S_{i,j}^{(n)}} = (1,0)\right] - \mathbb{1}\left[X_{S_{k,l}^{(n)}} = (1,0)\right]}_{\leq 1} \leq 0.$$

Then it also follows that

$$\text{proj}_2 \circ S_{i,j}^{(n+1)} = n+1 - \text{proj}_1 \circ S_{i,j}^{(n+1)} \geq n+1 - \text{proj}_1 \circ S_{k,l}^{(n+1)} = \text{proj}_2 \circ S_{k,l}^{(n+1)}$$

By definition it means

$$S_{i,j}^{(n+1)} \succ S_{k,l}^{(n+1)}.$$

2. $\text{proj}_1 \circ S_{i,j}^{(n)} = \text{proj}_1 \circ S_{k,l}^{(n)}$, then $\text{proj}_2 \circ S_{i,j}^{(n)} = \text{proj}_2 \circ S_{k,l}^{(n)}$, thus

$$S_{i,j}^{(n)} = S_{k,l}^{(n)},$$

thus by Lemma 2.3.1 we have that

$$S_{i,j}^{(n+1)} = S_{k,l}^{(n+1)},$$

which implies

$$S_{i,j}^{(n+1)} \succ S_{k,l}^{(n+1)}.$$

■

Lemma 2.3.3. *For any*

$$(i,j) \succ (k,l)$$

and for any $n \in \mathbb{N}$ we have that

$$T_{i,j}^{(n)} \succ T_{k,l}^{(n)}.$$

Proof. The proof is completely analogous to the proof of Lemma 2.3.2. ■

2.4 Connection between the trajectories

In this section we will show some important result corresponding to the paths of the Pólya Walks and the Dual Pólya Walks. The first lemma shows that the path of a Dual Pólya Walk on the Dual Web cannot cross the path of a Pólya Walk on the Pólya Web.

Lemma 2.4.1. *For any $1 \le m \le n$ and $i,j,k,l \in \mathbb{N}$ such that*

$$i + j = m,$$

$$k + l = n - 1.$$

(1) If $(k,l) \succ S_{i,j}^{(n)}$ then for any $m \le r \le n$ we have

$$T_{k,l}^{(r-1)} \succ S_{i,j}^{(r)}.$$

(2) If $S_{i,j}^{(n)} \succ (k,l)$ then for any $m \le r \le n$ we have

$$S_{i,j}^{(r)} \succ T_{k,l}^{(r-1)}.$$

Proof. We will prove (1), the proof of (2) is completely simmilar.

First for $r = n$ we have

$$T_{k,l}^{(r-1)} = T_{k,l}^{(n-1)} = \underbrace{(k,l) \succ S_{i,j}^{(n)}}_{\text{by assumption}} = S_{i,j}^{(r)}.$$

Suppose it is true for some $m < r \leq n$. Then consider the following two cases.

1. Suppose $T_{k,l}^{(r-1)} = S_{i,j}^{(r-1)}$. Then

$$T_{k,l}^{(r-2)} = T_{k,l}^{(r-1)} - X_{T_{k,l}^{(r-1)}} = T_{k,l}^{(r-1)} - X_{S_{i,j}^{(r-1)}} \succ S_{i,j}^{(r)} - X_{S_{i,j}^{(r-1)}} = S_{i,j}^{(r-1)}.$$

2. Suppose $T_{k,l}^{(r-1)} \neq S_{i,j}^{(r-1)}$. Then supposing

$$\text{proj}_1 \circ T_{k,l}^{(r-1)} > \text{proj}_1 \circ S_{k,l}^{(r-1)} \qquad \text{and} \qquad \text{proj}_2 \circ T_{k,l}^{(r-1)} < \text{proj}_2 \circ S_{k,l}^{(r-1)}$$

by the induction hypothesis leads to

$$\begin{aligned} \text{proj}_1 \circ S_{k,l}^{(r)} &\geq \text{proj}_1 \circ T_{k,l}^{(r-1)} > \text{proj}_1 \circ S_{k,l}^{(r-1)}, \\ \text{proj}_1 \circ S_{k,l}^{(r)} &\leq \text{proj}_2 \circ T_{k,l}^{(r-1)} < \text{proj}_2 \circ S_{k,l}^{(r-1)} \end{aligned}$$

which is a contradiction.

It follows that we have

$$\text{proj}_1 \circ T_{k,l}^{(r-1)} < \text{proj}_1 \circ S_{k,l}^{(r-1)} \qquad \text{and} \qquad \text{proj}_2 \circ T_{k,l}^{(r-1)} > \text{proj}_2 \circ S_{k,l}^{(r-1)}$$

which implies either

$$\begin{matrix} \text{proj}_1 \circ T_{k,l}^{(r-2)} < \text{proj}_1 \circ S_{k,l}^{(r-1)} & \quad \text{or} \quad & \text{proj}_1 \circ T_{k,l}^{(r-2)} < \text{proj}_1 \circ S_{k,l}^{(r-1)} \\ \text{proj}_2 \circ T_{k,l}^{(r-2)} > \text{proj}_2 \circ S_{k,l}^{(r-1)} & & \text{proj}_2 \circ T_{k,l}^{(r-2)} \geq \text{proj}_2 \circ S_{k,l}^{(r-1)}. \end{matrix}$$

In both cases we have

$$T_{k,l}^{(r-2)} \succ S_{i,j}^{(r-1)}.$$

■

The second lemma is the consequence of the first one. It shows that we can characterise the path of two neighbouring Pólya Walks with only one Dual Pólya Walk (Figure 5).

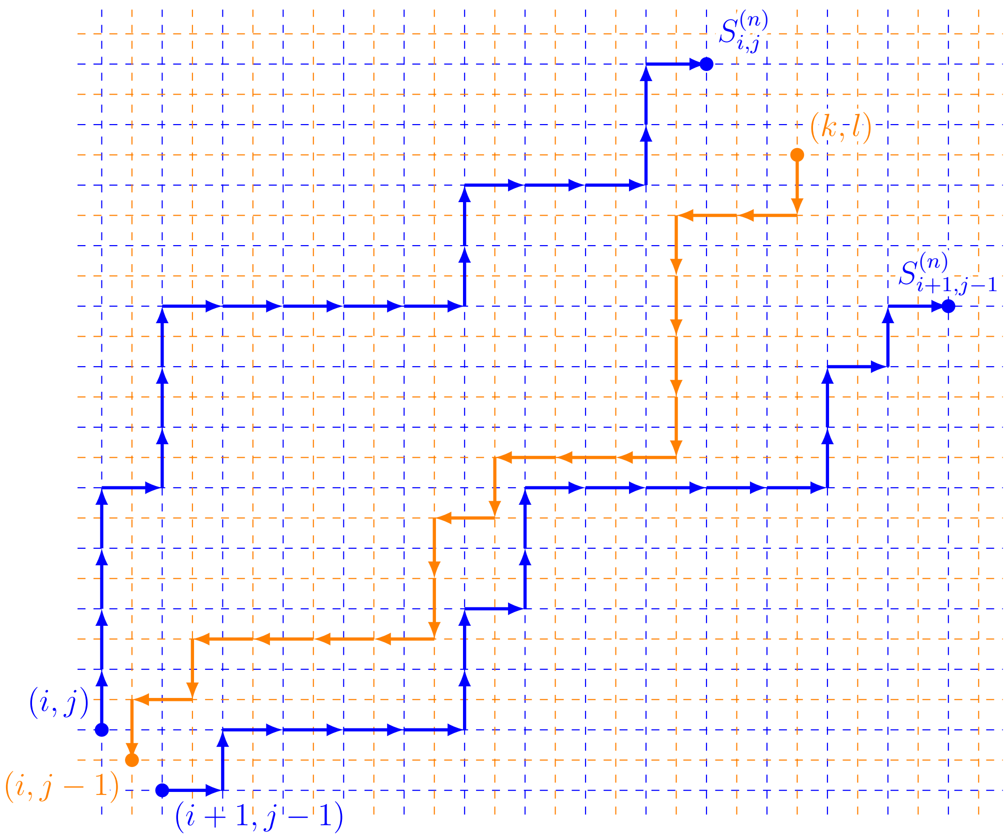

Figure 5: The characterization of two neighboring Pólya Walks (blue) with one Dual Pólya Walk (orange).

Lemma 2.4.2. *For any $m \leq n$ and $i+j=m, k+l=n-1$ we have that*

$$S_{i,j}^{(n)} \succ (k,l) \succ S_{i+1,j-1}^{(n)}$$

if and only if

$$T_{k,l}^{(m-1)} = (i,j-1).$$

Proof. Suppose $S_{i,j}^{(n)} \succ (k,l) \succ S_{i+1,j-1}^{(n)}$. Then Lemma 2.4.1 implies that

$$(i,j) = S_{i,j}^{(m)} \succ T_{k,l}^{(m-1)} \succ S_{i+1,j-1}^{(m)} = (i+1,j-1),$$

where

$$i \leq \text{proj}_1 \circ T_{k,l}^{(m-1)} \qquad \text{and} \qquad j-1 \leq \text{proj}_2 \circ T_{k,l}^{(m-1)}.$$

Notice that

$$\text{proj}_1 \circ T_{k,l}^{(m-1)} + \text{proj}_2 \circ T_{k,l}^{(m-1)} = m-1 = i+j-1,$$

thus

$$T_{k,l}^{(m-1)} = (i,j-1).$$

Suppose the opposite of $S_{i,j}^{(n)} \succ (k,l) \succ S_{i+1,j-1}^{(n)}$. Then two cases can happen by Lemma 2.3.2

$$S_{i,j}^{(n)} \succ S_{i+1,j-1}^{(n)} \succ (k,l) \qquad \text{or} \qquad (k,l) \succ S_{i,j}^{(n)} \succ S_{i+1,j-1}^{(n)}.$$

Then using Lemma 2.4.1

$$(i+1, j-1) \succ T_{k,l}^{(m-1)} \qquad \text{or} \qquad T_{k,l}^{(m-1)} \succ (i,j).$$

In both cases we have

$$T_{k,l}^{(m-1)} \neq (i, j-1).$$

■

2.5 The joint distribution of the limiting random variables

We have seen that the ratio of blue (or red) balls converges almost surely to a random variable with beta distribution. This motivates to introduce the same ratio for the Pólya Walk.

Definition 2.5.1 (Ratio of blue balls of the Pólya urn started form the pair $(i.j)$)**.**

$$\xi_{i,j}^{(n)} := \frac{\mathrm{proj}_1 \circ S_{i,j}^{(n)}}{\left\| S_{i,j}^{(n)} \right\|_1} = \frac{\mathrm{proj}_1 \circ S_{i,j}^{(n)}}{\mathrm{proj}_1 \circ S_{i,j}^{(n)} + \mathrm{proj}_2 \circ S_{i,j}^{(n)}} = \frac{\mathrm{proj}_1 \circ S_{i,j}^{(n)}}{n}.$$

Definition 2.5.2 (Ratio of red balls of the Pólya urn started form the pair $(i.j)$)**.**

$$\eta_{i,j}^{(n)} := \frac{\mathrm{proj}_2 \circ S_{i,j}^{(n)}}{\left\| S_{i,j}^{(n)} \right\|_1} = \frac{\mathrm{proj}_2 \circ S_{i,j}^{(n)}}{\mathrm{proj}_1 \circ S_{i,j}^{(n)} + \mathrm{proj}_2 \circ S_{i,j}^{(n)}} = \frac{\mathrm{proj}_2 \circ S_{i,j}^{(n)}}{n}.$$

A trivial observation is the following

$$\xi_{i,j}^{(n)} + \eta_{i,j}^{(n)} = 1.$$

Corollary 2.5.3.

$$\xi_{i,j}^{(n)} \overset{a.s.}{\to} \xi_{i,j} \sim BETA(i,j) \quad \text{and} \quad \eta_{i,j}^{(n)} \overset{a.s.}{\to} \eta_{i,j} \sim BETA(j,i), \qquad \text{as} \quad n \to \infty.$$

Proof. Applying Theorem 1.2.2 our proof is finished. ■

Now we prove some important properties of the joint distribution of the limiting beta variables. First we show that the joint distribution coincides with the ordering of the plane defined previously in Definition 2.2.5.

Lemma 2.5.4. *For any*

$$(i,j) \succ (k,l)$$

we have that

$$\frac{i}{i+j} \leq \frac{k}{k+l}.$$

Proof.

$$\frac{i}{i+j} - \frac{k}{k+l} = \underbrace{\frac{i \cdot l - k \cdot j}{(i+j)(k+l)} \leq \frac{k \cdot l - k \cdot l}{(i+j)(k+l)}}_{\text{since } i \leq k \text{ and } j \geq l} = 0.$$

∎

Lemma 2.5.5. *For any*

$$(i,j) \succ (k,l)$$

and for any $n \in \mathbb{N}$ we have that

$$\xi_{i,j}^{(n)} \leq \xi_{k,l}^{(n)} \qquad \text{and} \qquad \eta_{i,j}^{(n)} \geq \eta_{k,l}^{(n)}.$$

Proof. For any $n \in \mathbb{N}$ using Lemma 2.3.2 we have

$$\left(\mathrm{proj}_1 \circ S_{i,j}^{(n)},\ \mathrm{proj}_2 \circ S_{i,j}^{(n)}\right) = S_{i,j}^{(n)} \succ S_{k,l}^{(n)} = \left(\mathrm{proj}_1 \circ S_{k,l}^{(n)},\ \mathrm{proj}_2 \circ S_{k,l}^{(n)}\right),$$

then using Lemma 2.5.4

$$\xi_{i,j}^{(n)} = \frac{\mathrm{proj}_1 \circ S_{i,j}^{(n)}}{\mathrm{proj}_1 \circ S_{i,j}^{(n)} + \mathrm{proj}_2 \circ S_{i,j}^{(n)}} \leq \frac{\mathrm{proj}_1 \circ S_{k,l}^{(n)}}{\mathrm{proj}_1 \circ S_{k,l}^{(n)} + \mathrm{proj}_2 \circ S_{k,l}^{(n)}} = \xi_{k,l}^{(n)}.$$

Then it follows that

$$\eta_{i,j}^{(n)} = 1 - \xi_{i,j}^{(n)} \geq 1 - \xi_{k,l}^{(n)} = \eta_{k,l}^{(n)}.$$

∎

Proposition 2.5.6. *For any*

$$(i,j) \succ (k,l)$$

we have that

$$\mathbb{P}(\xi_{i,j} \leq \xi_{k,l}) = \mathbb{P}(\eta_{i,j} \geq \eta_{k,l}) = 1.$$

Proof.

$$\underbrace{1 = \mathbb{P}\left(\xi_{i,j}^{(n)} \leq \xi_{k,l}^{(n)}\right)}_{\text{by Lemma 2.5.5}} = \underbrace{\mathbb{P}\left(\lim_{n\to\infty} \xi_{i,j}^{(n)} \leq \lim_{n\to\infty} \xi_{k,l}^{(n)}\right) = \mathbb{P}(\xi_{i,j} \leq \xi_{k,l})}_{\text{by Corollary 2.5.3}}$$

Then it follows

$$\mathbb{P}(\eta_{i,j} \geq \eta_{k,l}) = \mathbb{P}(1 - \xi_{i,j} \geq 1 - \xi_{k,l}) = \mathbb{P}(\xi_{i,j} \leq \xi_{k,l}) = 1.$$

■

Notice that in contrast with the previous statements Proposition 2.5.6 above states it for almost every realization. While the previous statements are true for every realization. It is needed because we only have almost sure convergence in the limit of the ratios.
We would like to demonstrate an already known property of the Beta-distribution as a short corollary of Proposition 2.5.6. For that, we have to recall the definition of stochastic dominance and a corresponding theorem. For more details check [7].

Theorem 2.5.7. *Let X and Y be two real valued random variables (not necessarily defined on the same space). Then*

$$\mathbb{P}(X \leq t) \leq \mathbb{P}(Y \leq t)$$

for any $t \in \mathbb{R}$ (i. e. Y stochastically dominates X) happens if and only if there exist a $\left(\tilde{X}, \tilde{Y}\right)$ coupling such that

$$X \sim \tilde{X}, \quad Y \sim \tilde{Y} \qquad \text{and} \qquad \mathbb{P}\left(\tilde{X} \leq \tilde{Y}\right) = 1.$$

Corollary 2.5.8. *Let $1 \leq i \leq k$ and $1 \leq l \leq j$. Then* $\mathrm{BETA}(k, l)$ *stochastically dominates* $\mathrm{BETA}(i, j)$.

Proof. We have the coupling $(\xi_{i,j}, \xi_{k,l})$, where

$$\xi_{i,j} \sim \mathrm{BETA}(i, j) \qquad \text{and} \qquad \xi_{k,l} \sim \mathrm{BETA}(k, l).$$

By Proposition 2.5.6 we have

$$\mathbb{P}(\xi_{i,j} \leq \xi_{k,l}) = 1.$$

Thus applying Theorem 2.5.7 finishes our proof. ■

In the last part we show a monotone property of the ratios which will come helpful in the upcoming Section 3.1 by letting us to apply the Harris-inequality. First let us recall the definition of increasing and decreasing events.

Definition 2.5.9 (Decreasing and increasing event)**.** *Let $(\Omega, \mathcal{F})$ be a measureable space and $\leq$ a partial order on Ω. We say that the event $A \in \mathcal{F}$ is decreasing if for any $\omega \in A$, $\omega' \leq \omega$ implies $\omega' \in A$. The event $B \in \mathcal{F}$ is increasing if for any $\omega \in B$, $\omega \leq \omega'$ implies $\omega' \in B$.*

.

Proposition 2.5.10. *For any (i,j) and $0 \le \alpha \le 1$ the events*

$$\left\{\liminf_{n\to\infty} \xi_{i,j}^{(n)} < \alpha\right\} \qquad \textit{and} \qquad \left\{\limsup_{n\to\infty} \xi_{i,j}^{(n)} < \alpha\right\}$$

are decreasing, while the events

$$\left\{\alpha < \liminf_{n\to\infty} \xi_{i,j}^{(n)}\right\} \qquad \textit{and} \qquad \left\{\alpha < \limsup_{n\to\infty} \xi_{i,j}^{(n)}\right\}$$

are increasing with respect to the order on Ω defined in Definition 2.2.3.

Proof. We only prove it for one of the event. The proof for the rest is analogous. Let

$$\omega \in \left\{\omega \in \Omega \ : \ \liminf_{n\to\infty} \xi_{i,j}^{(n)}\left(\omega\right) < \alpha\right\}$$

Now suppose $\omega' \multimapinv \omega$. Then Proposition 2.2.7 implies for any $n \in \mathbb{N}$

$$\mathrm{proj}_1 \circ S_{i,j}^{(n)}\left(\omega'\right) \le \mathrm{proj}_1 \circ S_{i,j}^{(n)}\left(\omega\right).$$

After dividing we get

$$\xi_{i,j}^{(n)}(\omega') = \frac{\mathrm{proj}_1 \circ S_{i,j}^{(n)}\left(\omega'\right)}{n} \le \frac{\mathrm{proj}_1 \circ S_{i,j}^{(n)}\left(\omega'\right)}{n} = \xi_{i,j}^{(n)}(\omega').$$

Then taking the $\liminf$ as $n \to \infty$

$$\liminf_{n\to\infty} \xi_{i,j}^{(n)}(\omega') \le \liminf_{n\to\infty} \xi_{i,j}^{(n)}(\omega) < \alpha.$$

This exaclty means

$$\omega' \in \left\{\omega \in \Omega \ : \ \liminf_{n\to\infty} \xi_{i,j}^{(n)}\left(\omega\right) < \alpha\right\}.$$

■

2.6 The limiting random variables and the trajectories

In this section we show some connection between the joint distribution of the limiting variables and the paths of the Pólya Walks. We also state a conjecture which we have not been able to prove yet.

Lemma 2.6.1. *For any $n \in \mathbb{N}$ and $(i,j),(k,l) \in \mathbb{N} \times \mathbb{N} \setminus \{(0,0)\}$*

$$\mathbb{P}\Big(\xi_{i,j} = \xi_{k,l} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\Big) = 1.$$

Proof. For any $m \geq \max\{i+j, k+l, n\}$ using Lemma 2.3.1 we have that

$$1 = \mathbb{P}\Big(S_{i,j}^{(m)} = S_{k,l}^{(m)} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\Big) = \mathbb{P}\left(\frac{S_{i,j}^{(m)}}{m} = \frac{S_{k,l}^{(m)}}{m} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\right)$$

$$= \mathbb{P}\Big(\xi_{i,j}^{(m)} = \xi_{k,l}^{(m)} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\Big).$$

Then applying Corollary 2.5.3

$$1 = \mathbb{P}\Big(\lim_{m\to\infty} \xi_{i,j}^{(m)} = \lim_{m\to\infty} \xi_{k,l}^{(m)} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\Big) = \mathbb{P}\Big(\xi_{i,j} = \xi_{k,l} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\Big).$$

■

Proposition 2.6.2.

$$\mathbb{P}\left(\xi_{i,j} = \xi_{k,l} \,\middle|\, \bigcup_{n=0}^{\infty} \left\{S_{i,j}^{(n)} = S_{k,l}^{(n)}\right\}\right) = 1.$$

Proof. Notice that by Lemma 2.3.1 $\left\{S_{i,j}^{(n)} = S_{k,l}^{(n)}\right\}_{n\in\mathbb{N}}$ is an increasing set of events. Thus

$$\mathbb{P}\left(\xi_{i,j} = \xi_{k,l} \,\middle|\, \bigcup_{n=0}^{\infty} \left\{S_{i,j}^{(n)} = S_{k,l}^{(n)}\right\}\right) = \lim_{n\to\infty} \underbrace{\mathbb{P}\Big(\xi_{i,j} = \xi_{k,l} \,\Big|\, S_{i,j}^{(n)} = S_{k,l}^{(n)}\Big)}_{=1,\text{ by Lemma 2.6.1}} = 1.$$

■

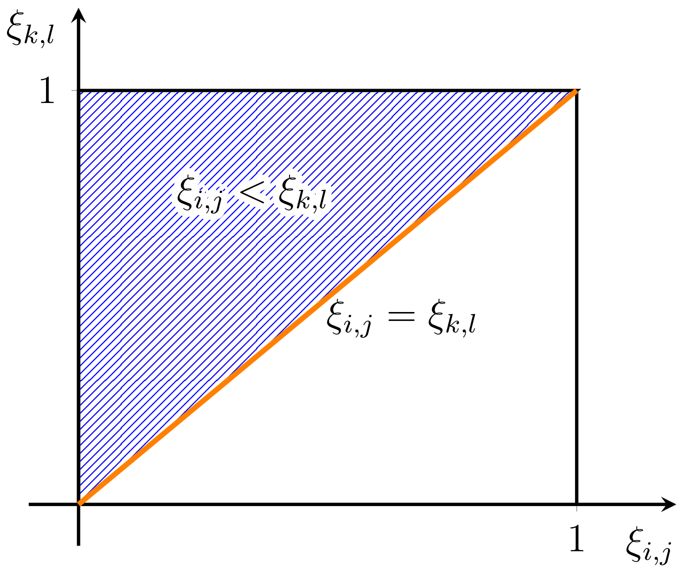

Figure 6: The joint distribution of $\xi_{i,j}$ and $\xi_{k,l}$ if $(i,j) \succ (k,l)$.

Conjecture 2.6.3.

$$\mathbb{P}\left(\bigcup_{n=0}^{\infty} \left\{S_{i,j}^{(n)} = S_{k,l}^{(n)}\right\} \mid \xi_{i,j} = \xi_{k,l}\right) = 1.$$

Proposition 2.6.2 states that if two Pólya Walks meet at one point they will have the same limiting ratio almost surely. While Conjecture 2.6.3 is that the equality of the limiting ratios happens almost surely if and only if the two walks meet at one point. Although we have not proved this up to this point.

To summarise the section so far if we take the joint distributions of $\xi_{i,j}$ and $\xi_{k,l}$ having $(i,j) \succ (k,l)$ on the unit square $[0,1] \times [0,1]$ then the total mass is located above the diagonal from $(0,0)$ to $(1,1)$. If our conjecture is true the mass on the diagonal equals with the probability that the Pólya Walks started from (i,j) and (k,l) meet (Figure 6).

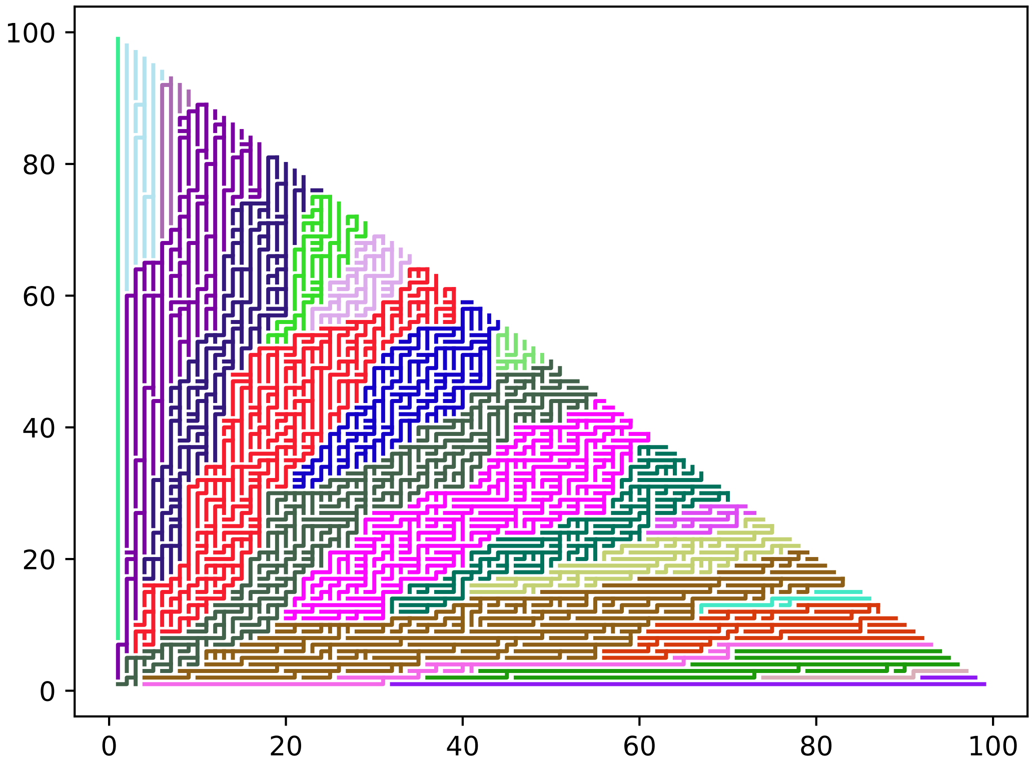

Figure 7: The random graph defined in Definition 2.6.4.

Definition 2.6.4 (The random oriented graph corresponding to the coupled Pólya Walks)**.** *Let us consider the following graph* $\mathcal{G}$*.*

$$V(\mathcal{G}) = \mathbb{N} \times \mathbb{N} \setminus \{(0,0)\},$$

and for $(i,j),(k,l) \in V(\mathcal{G})$

$$(i,j) \to (k,l) \in E(\mathcal{G})$$

if and only if

$$X_{i,j} = (k,l) - (i,j).$$

In the definition of $\mathcal{G}$ $X_{i,j}$ *is the random variable defined in Definition 2.1.1.*

We show a realization of the random graph (Figure 7). The colors denote different components of the graph, which is an oriented forest itself. The points with the same color have the same limiting ratio. If our conjecture is true different components have different limiting ratios. Notice that on the figure trajectories which do not have a common point might have the same color. This happens since they meet at a point which is not included in the picture (because of the obvious reason that we do not posses infinite time for simulation and cannot plot the whole $\mathbb{N} \times \mathbb{N}$ on a page). The Pólya Walks on the picture were simulated until $n = 1000$ and the picture shows the trajectories up to $n = 100$. Notice that we can also see the trajectories of the Dual Pólya Walks. Those are the white spaces between the trajectories of two Pólya Walks.

3 Local properties

In this chapter we will show properties of such Pólya Walks which start from neighboring points or points being relatively close to each other. We specify the meaning of the last statement later in a precise mathematical way.

3.1 Bounds on the pair (1,2) and (2,1)

First we show some bounds for the Pólya Walks started from the points (1,2) and (2,1). This can be easily generalized for any pairs of neighboring points. First we will prove a lower bound on the joint distributions.

Proposition 3.1.1. *For any*

$$(i,j) \succ (k,l)$$

we have

$$\mathbb{P}(\xi_{i,j} = \xi_{k,l}) \geq \int_0^1 \max\left\{\Big(1 - F_{\xi_{i,j}}(x)\Big)\, f_{\xi_{k,l}}(x),\ F_{\xi_{k,l}}(x) f_{\xi_{i,j}}(x)\right\} dx.$$

Proof. For any $0 \leq x \leq 1$

$$\mathbb{P}(\xi_{i,j} = \xi_{k,l} \in (x, x+dx)) \geq \mathbb{P}\Big(\tilde{\xi}_{i,j} > x,\ \xi_{k,l} \in (x, x+dx)\Big),$$

where $\tilde{\xi}_{i,j}$ is the ratio of blue balls started from the point (i,j) independently from $\xi_{k,l}$. Thus after using independence and the fact that $\tilde{\xi}_{i,j} \sim \xi_{i,j}$

$$\mathbb{P}(\xi_{i,j} = \xi_{k,l} \in (x, x+dx)) \geq \mathbb{P}\Big(\tilde{\xi}_{i,j} > x\Big) \cdot \mathbb{P}(\xi_{k,l} \in (x, x+dx)) =$$

$$\Big(1 - F_{\xi_{i,j}}(x)\Big) \cdot f_{\xi_{k,l}}(x)\, dx.$$

By a similar argument we get the following

$$\mathbb{P}(\xi_{i,j} = \xi_{k,l} \in (x, x+dx)) \geq F_{\xi_{k,l}}(x) \cdot f_{\xi_{i,j}}(x)\, dx.$$

After comparing the two inequalities

$$\mathbb{P}(\xi_{i,j} = \xi_{k,l} \in (x, x+dx)) \geq \max\left\{\Big(1 - F_{\xi_{i,j}}(x)\Big) \cdot f_{\xi_{k,l}}(x),\ F_{\xi_{k,l}}(x) \cdot f_{\xi_{i,j}}(x)\right\} dx$$

and integrating we get the desired result. ∎

Before stating our estimations we should recall the so called Harris-inequality [6].

Theorem 3.1.2 (Harris-inequality)**.** *Let* $\Omega = \{0,1\}^{\mathbb{N}}$ *and* $\mathbb{P} = \bigotimes_{i \in \mathbb{N}} \mu_i$ *where* μ_i *are Bernoulli measures on* $\{0,1\}$*. Then if* A *is an increasing and* B *is a decreasing event with the order defined on* Ω *as* $\omega \leq \omega'$ *if* $\omega_i \leq \omega_i'$ *for every* $i \in \mathbb{N}$*, we have*

$$\mathbb{P}(A \cap B) \leq \mathbb{P}(A)\mathbb{P}(B).$$

Notice that in our case we have as in Equation (3)

$$\Omega = \{(1,0),(0,1)\}^{\mathbb{N}\times\mathbb{N}\setminus\{(0,0)\}}$$

and the corresponding Bernoulli-measures are

$$\mu_{i,j}((1,0)) = \frac{i}{i+j} \qquad \text{and} \qquad \mu_{i,j}((0,1)) = \frac{j}{i+j}.$$

Our measure on Ω is exactly

$$\mathbb{P} = \bigotimes_{i,j} \mu_{i,j}.$$

Thus our setting coincides with the probability space (and also our order defined in Definition 2.2.3) stated in Theorem 3.1.2.

Corollary 3.1.3. *For any* $(i,j),(k,l)$ *and* $0 \leq \alpha, \beta \leq 1$ *we have*

$$\mathbb{P}(\alpha < \xi_{i,j},\ \xi_{k,l} < \beta) \leq \mathbb{P}(\alpha < \xi_{i,j})\mathbb{P}(\xi_{k,l} < \beta)$$

Proof. By Proposition 2.5.10 the event

$$\left\{\alpha < \liminf_{n\to\infty} \xi_{i,j}^{(n)}\right\}$$

is decreasing and the event

$$\left\{\limsup_{n\to\infty} \xi_{k,l}^{(n)} < \beta\right\}$$

is increasing. After applying Theorem 3.1.2

$$\mathbb{P}\Big(\alpha < \liminf_{n\to\infty} \xi_{i,j}^{(n)},\ \limsup_{n\to\infty} \xi_{k,l}^{(n)} < \beta\Big) \leq \mathbb{P}\Big(\alpha < \liminf_{n\to\infty} \xi_{i,j}^{(n)}\Big)\mathbb{P}\Big(\limsup_{n\to\infty} \xi_{k,l}^{(n)} < \beta\Big).$$

However by Corollary 2.5.3 we have

$$\mathbb{P}\Big(\alpha < \liminf_{n\to\infty} \xi_{i,j}^{(n)},\ \limsup_{n\to\infty} \xi_{k,l}^{(n)} < \beta\Big) = \mathbb{P}(\alpha < \xi_{i,j},\ \xi_{k,l} < \beta),$$

$$\mathbb{P}\Big(\alpha < \liminf_{n\to\infty} \xi_{i,j}^{(n)}\Big) = \mathbb{P}(\alpha < \xi_{i,j}) \qquad \text{and} \qquad \mathbb{P}\Big(\limsup_{n\to\infty} \xi_{k,l}^{(n)} < \beta\Big) = \mathbb{P}(\xi_{k,l} < \beta).$$

∎

Proposition 3.1.4.

$$\frac{11}{48} \leq \mathbb{P}(\xi_{1,2} = \xi_{2,1}) \leq \frac{2}{5}$$

Proof. We will prove the lower and upper bound separately.

1. To get the lower bound we just apply Proposition 3.1.1 to the pair $(1,2)$ and $(1,2)$. Then

$$\mathbb{P}(\xi_{1,2} = \xi_{2,1}) \geq \int_0^1 \max\left\{\Big(1 - F_{\xi_{1,2}}(x)\Big) f_{\xi_{2,1}}(x),\ F_{\xi_{2,1}}(x) f_{\xi_{1,2}}(x)\right\} dx =$$
$$\int_0^1 \max\left\{\Big(1 + x^2 - 2x\Big) \cdot 2x,\ x^2 \cdot (2 - 2x)\right\} dx = \frac{11}{48}.$$

2. For the upper bound let us suppose $0 \leq y \leq \frac{1}{2} \leq x \leq 1$ (Figure 8). First notice that

$$\underbrace{\mathbb{P}(\xi_{1,2} < x < \xi_{2,1}) = \mathbb{P}(\xi_{1,2} < x) - \mathbb{P}(\xi_{2,1} < x)}_{\text{by Proposition 2.5.6}} = 2x(1-x)$$

and also we have the equality

$$\mathbb{P}(\xi_{1,2} < x < \xi_{2,1}) = \mathbb{P}\Big(\xi_{2,1} < \frac{1}{2},\ x < \xi_{2,1}\Big) + \mathbb{P}\Big(\frac{1}{2} < \xi_{1,2} < x,\ x < \xi_{2,1}\Big).$$

Then after applying Proposition 2.5.10

$$\mathbb{P}\Big(\xi_{2,1} < \frac{1}{2},\ x < \xi_{2,1}\Big) \leq \mathbb{P}\Big(\xi_{2,1} < \frac{1}{2}\Big) \cdot \mathbb{P}(x < \xi_{2,1}) = \frac{3}{4}(1 - x^2),$$

we get the lower bound

$$\mathbb{P}\Big(\frac{1}{2} < \xi_{1,2} < x,\ x < \xi_{2,1}\Big) \geq 2x(1-x) - \frac{3}{4}(1-x^2) = \frac{1}{20} - \frac{5}{4}\Big(x - \frac{4}{5}\Big)^2 \geq \frac{1}{20}$$

with equality if and only if $x = \dfrac{4}{5}$.

Using the completely same argument we can archive the bound

$$\mathbb{P}\Big(y < \xi_{2,1} < \frac{1}{2},\ \xi_{1,2} < y\Big) \geq \frac{1}{20} - \frac{5}{4}\Big(y - \frac{1}{5}\Big)^2 \geq \frac{1}{20}$$

with equality if and only if $y = \dfrac{1}{5}$.

Thus we have the following bound

$$\mathbb{P}(\xi_{1,2} < \xi_{2,1}) \geq$$
$$\mathbb{P}\Big(\xi_{1,2} < \frac{1}{2} < \xi_{2,1}\Big) + \mathbb{P}\Big(\frac{1}{2} < \xi_{1,2} < \frac{4}{5},\ \frac{4}{5} < \xi_{2,1}\Big) + \mathbb{P}\Big(\frac{1}{5} < \xi_{2,1} < \frac{1}{2},\ \xi_{1,2} < \frac{1}{5}\Big) =$$
$$\frac{1}{2} + \frac{1}{20} + \frac{1}{20} = \frac{3}{5}.$$

From this it follows that

$$\mathbb{P}(\xi_{1,2} = \xi_{2,1}) = 1 - \mathbb{P}(\xi_{1,2} < \xi_{2,1}) \leq 1 - \frac{3}{5} = \frac{2}{5}.$$

∎

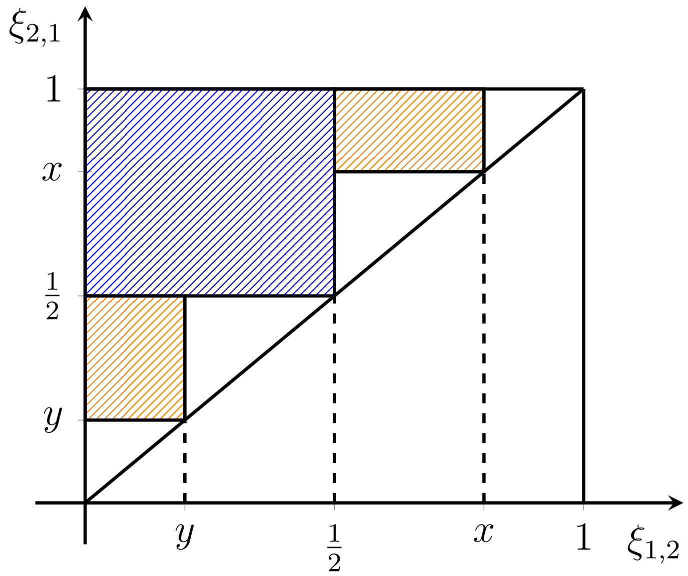

Figure 8: The upper bound on the probability $\mathbb{P}(\xi_{1,2} = \xi_{2,1})$.

At the end of this subsection we present some numerical empirical bounds on the probabilities above. We simulated the Pólya Walks for 10^7 steps with a sample size of 10^4. The empirical probability that the two ratios of equal was 0.328. This is a numerical upper bound to the probabilty $\mathbb{P}(\xi_{1,2} = \xi_{2,1})$.

3.2 The beta and the binomial distribution

In this section we will prove an already known formula describing the connection between the cumulative distribution function of a beta and a corresponding binomial distribution. However in our proof we will use some properties of the joint Pólya's urns.

Proposition 3.2.1. *The probability of the event, that the Dual Pólya Walk starting from* $(K, N-K)$ *hits the point* $(k, n-1-k)$ *is*

$$\mathbb{P}\Big(T^{(n-1)}_{N,N-K} = (k, n-1-k)\Big) = \binom{N-n+1}{K-k}\binom{N}{K}^{-1}\binom{n-1}{k}. \tag{4}$$

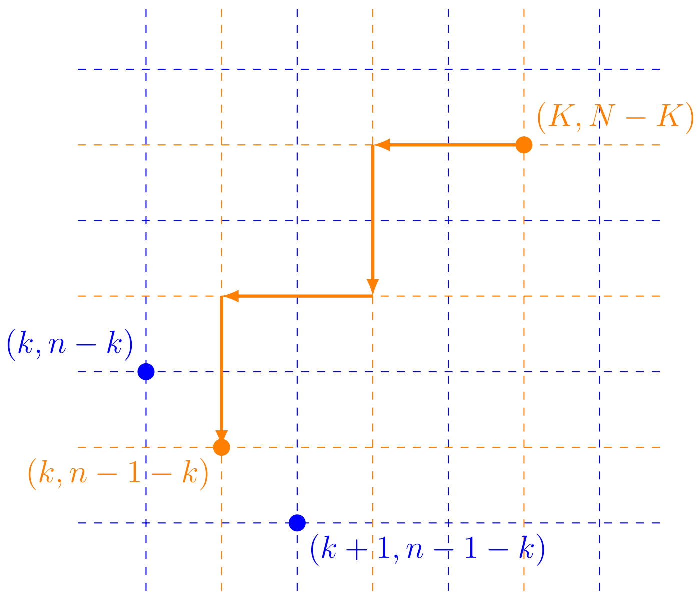

Figure 9: The Dual Pólya Walk started from $(N, N-K)$ hitting $(k, n-1-k)$.

Proof. Starting from $(K, N-K)$ for this it has to take $K-k$ steps in the vertical direction and all together $N-n+1$ steps (Figure 9). Thus the total number of possible paths is

$$\binom{N-n+1}{K-k}.$$

Now since all paths has the same probability, which equals to

$$\binom{N}{K}^{-1}\binom{n-1}{k},$$

multiplying the two we get Equation (4). ■

Lemma 3.2.2. *For any irrational* $0 < \alpha < 1$ *we have the following*

$$\mathbb{P}\Big(\xi^{(N)}_{k,n-k} < \alpha < \xi^{(N)}_{k+1,n-1-k}\Big) = \binom{\lfloor \alpha N\rfloor + \lfloor (1-\alpha)N\rfloor - n + 1}{\lfloor \alpha N\rfloor - k}\binom{\lfloor \alpha N\rfloor + \lfloor (1-\alpha)N\rfloor}{\lfloor \alpha N\rfloor}^{-1}\binom{n-1}{k} \quad (5)$$

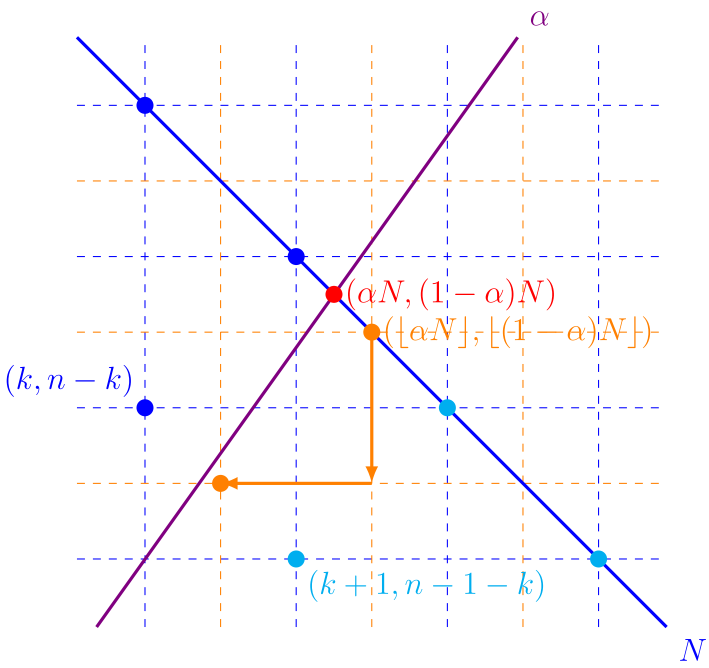

Figure 10: The event $\xi^{(N)}_{k,n-k} < \alpha < \xi^{(N)}_{k+1,n-1-k}$.

Proof. The event $\xi^{(N)}_{k,n-k} < \alpha < \xi^{(N)}_{k+1,n-1-k}$ means that

$$\mathrm{proj}_1 \circ S^{(N)}_{k,n-k} < \alpha N \qquad \text{and} \qquad \mathrm{proj}_2 \circ S^{(N)}_{k+1,n-1-k} \leq (1-\alpha)N$$

and since α is irrational αN and $(1-\alpha)N$ is never an integer. Thus the event happens if and only if

$$S^{(N)}_{k,n-k} \succ (\lfloor \alpha N\rfloor, \lfloor (1-\alpha)N\rfloor) \succ S^{(N)}_{k+1,n-1-k}.$$

Also notice that $\lfloor \alpha N\rfloor + \lfloor (1-\alpha)N\rfloor = N-1$. Thus after applying Lemma 2.4.2 and the formula Equation (4)

$$\mathbb{P}\Big(\xi^{(N)}_{k,n-k} < \alpha < \xi^{(N)}_{k+1,n-1-k}\Big) = \binom{\tilde N - n + 1}{\tilde K - k}\binom{\tilde N}{\tilde K}^{-1}\binom{n-1}{k}$$

with

$$\tilde N = \lfloor \alpha N\rfloor + \lfloor (1-\alpha)N\rfloor \qquad \text{and} \qquad \tilde K = \lfloor \alpha N\rfloor,$$

we get Equation (5). ■

Lemma 3.2.3.

$$\lim_{N\to\infty} \mathbb{P}\Big(\xi^{(N)}_{k,n-k} < \alpha < \xi^{(N)}_{k+1,n-1-k}\Big) = \mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}).$$

Proof. Notice that by Lemma 2.5.5

$$\mathbb{P}\Big(\xi^{(N)}_{k,n-k} > \alpha > \xi^{(N)}_{k+1,n-1-k}\Big) = 0,$$

thus it follows that for the complementary event

$$1 - \mathbb{P}\Big(\xi^{(N)}_{k,n-k} < \alpha < \xi^{(N)}_{k+1,n-1-k}\Big) = \mathbb{P}\Big(\alpha \leq \xi^{(N)}_{k,n-k}\Big) + \mathbb{P}\Big(\xi^{(N)}_{k+1,n-1-k} \leq \alpha\Big).$$

Using the convergence of the variables and the fact that the beta distribution has a continuous distribution function we have the following

$$\lim_{N\to\infty} \mathbb{P}\Big(\alpha \leq \xi^{(N)}_{k,n-k}\Big) + \mathbb{P}\Big(\xi^{(N)}_{k+1,n-1-k} \leq \alpha\Big) = \mathbb{P}(\alpha \leq \xi_{k,n-k}) + \mathbb{P}(\xi_{k+1,n-1-k} \leq \alpha).$$

Now also notice that by Proposition 2.5.6

$$\mathbb{P}(\xi_{k+1,n-1-k} < \xi_{k,n-k}) = 0,$$

so we have the equality

$$\mathbb{P}(\alpha \leq \xi_{k,n-k}) + \mathbb{P}(\xi_{k+1,n-1-k} \leq \alpha) = 1 - \mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}).$$

■

Proposition 3.2.4. *For any $0 \leq \alpha \leq 1$ we have*

$$\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}) = \binom{n-1}{k} \alpha^k (1-\alpha)^{n-1-k}. \tag{6}$$

Proof. Now let $0 < \alpha < 1$ be an arbitrary irrational number. Then considering Lemma 3.2.3 and Equation (5) we have that

$$\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}) = \lim_{N\to\infty} \binom{\tilde{N}-n+1}{\tilde{K}-k} \binom{\tilde{N}}{\tilde{K}}^{-1} \binom{n-1}{k},$$

where

$$\tilde{N} = \lfloor \alpha N \rfloor + \lfloor (1-\alpha)N \rfloor \qquad \text{and} \qquad \tilde{K} = \lfloor \alpha N \rfloor.$$

Using Stirling's formula we have the following

$$\binom{\tilde{N}-n+1}{\tilde{K}-k} \binom{\tilde{N}}{\tilde{K}}^{-1} \sim \frac{\tilde{K}^k (\tilde{N}-\tilde{K})^{n-1-k}}{\tilde{N}^{n-1}} = \left(\frac{\tilde{K}}{\tilde{N}}\right)^k \left(1 - \frac{\tilde{K}}{\tilde{N}}\right)^{n-1-k}$$

Now notice that

$$\lim_{N\to\infty}\frac{\tilde{K}}{\tilde{N}}=\lim_{N\to\infty}\frac{\lfloor \alpha N\rfloor}{\lfloor \alpha N\rfloor+\lfloor (1-\alpha)N\rfloor}=\alpha.$$

Thus we obtained the following result

$$\mathbb{P}(\xi_{k,n-k}<\alpha<\xi_{k+1,n-1-k})$$
$$=\binom{n-1}{k}\lim_{N\to\infty}\left(\frac{\tilde{K}}{\tilde{N}}\right)^k\left(1-\frac{\tilde{K}}{\tilde{N}}\right)^{n-1-k}=\binom{n-1}{k}\alpha^k(1-\alpha)^{n-1-k}.$$

Finally let $0\le\alpha\le 1$ be any real number. Then exits a $0<\alpha_j<1$ sequence of irrational numbers such that

$$\lim_{j\to\infty}\alpha_j=\alpha.$$

Using the same argument as previously in Lemma 3.2.3

$$\mathbb{P}(\xi_{k,n-k}<\alpha<\xi_{k+1,n-1-k})=1-\mathbb{P}(\alpha\le\xi_{k,n-k})-\mathbb{P}(\xi_{k+1,n-1-k}\le\alpha).$$

Since the cumulative distribution functions of $\xi_{k,n-k}$ and $\xi_{k+1,n-1-k}$ are both continuous the following identity holds

$$1-\mathbb{P}(\alpha\le\xi_{k,n-k})-\mathbb{P}(\xi_{k+1,n-1-k}\le\alpha)=\lim_{j\to\infty}1-\mathbb{P}(\alpha_j\le\xi_{k,n-k})-\mathbb{P}(\xi_{k+1,n-1-k}\le\alpha_j)=$$
$$\lim_{j\to\infty}\mathbb{P}(\xi_{k,n-k}<\alpha_j<\xi_{k+1,n-1-k}).$$

Now applying the formula we obtained above for irrational numbers

$$\mathbb{P}(\xi_{k,n-k}<\alpha<\xi_{k+1,n-1-k})=\lim_{j\to\infty}\binom{n-1}{k}\alpha_j^k(1-\alpha_j)^{n-1-k}=\binom{n-1}{k}\alpha^k(1-\alpha)^{n-1-k}.$$

■

The next result is already a well known one (for more detals check [2]). A different version was stated by Bayes [1]. However we will prove this using just the special behavior of the coupled Pólya Walks.

Corollary 3.2.5. *Let $X\sim BETA(k,n-k)$ and $Y\sim BIN(n-1,p)$, not necessarily jointly defined. Then the following holds*

$$\mathbb{P}(X<p)=\mathbb{P}(k\le Y). \tag{7}$$

Proof.

$$\mathbb{P}(X<p)=\sum_{j=k}^{n-1}\mathbb{P}(\xi_{j,n-j}<p<\xi_{j+1,n-1-j})=\sum_{j=k}^{n-1}\binom{n-1}{k}p^k(1-p)^{n-1-k}=\mathbb{P}(k\le Y),$$

where we used Equation (6). ■

Remark 3.2.6. *We proved our corollary in the case when the beta distribution has integer parameters. However this can be generalized for any pair of positive real numbers.*

3.3 Asymptotic bounds on the joint probabilities

In the first part of this section we investigate the allocation of mass in the joint distribution of the limiting ratios of Pólya Walks started next to each other. It turns out that the probability $\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k})$ is exponentially small for large enough n if α is different than the ratio $\frac{k}{n}$.

Proposition 3.3.1. *Let $0 \leq \alpha, \beta \leq 1$ and let*

$$k = \lfloor \beta n \rfloor,$$

then

$$\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}) \sim \frac{1}{\sqrt{2\pi\beta(1-\beta)}} e^{-(n-1)D(BER(\beta)\|BER(\alpha))} \tag{8}$$

where $D(\cdot\|\cdot)$ denotes the Kullbach-Leiber divergence.

The definition and the main properties of the Kullbach-Leiber divergence can be found in [3].

Proof. Using Equation (6) we have

$$\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}) = \binom{n-1}{\lfloor \beta n \rfloor} \alpha^{\lfloor \beta n \rfloor} (1-\alpha)^{n-1-\lfloor \beta n \rfloor}$$

After applying Stirling's formula we can approximate the probability with the following

$$\frac{1}{\sqrt{2\pi \frac{\lfloor \beta n \rfloor}{n-1}(1-\frac{\lfloor \beta n \rfloor}{n-1})}} \left(\frac{\lfloor \beta n \rfloor}{n-1}\right)^{-\frac{\lfloor \beta n \rfloor}{n-1}(n-1)} \left(1-\frac{\lfloor \beta n \rfloor}{n-1}\right)^{-\left(1-\frac{\lfloor \beta n \rfloor}{n-1}\right)(n-1)} \alpha^{\frac{\lfloor \beta n \rfloor}{n-1}(n-1)} (1-\alpha)^{\left(1-\frac{\lfloor \beta n \rfloor}{n-1}\right)(n-1)}.$$

Considering that

$$\lim_{n\to\infty} \frac{\lfloor \beta n \rfloor}{n-1} = \beta,$$

we have that

$$\begin{aligned}&\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}) \\ &\qquad \sim \frac{1}{\sqrt{2\pi\beta(1-\beta)}} \beta^{-\beta(n-1)} (1-\beta)^{-(1-\beta)(n-1)} \alpha^{\beta(n-1)} (1-\alpha)^{(1-\beta)(n-1)}.\end{aligned}$$

After rearranging

$$\begin{aligned}&\mathbb{P}(\xi_{k,n-k} < \alpha < \xi_{k+1,n-1-k}) \\ &\qquad \sim \frac{1}{\sqrt{2\pi\beta(1-\beta)}} \exp\left(-(n-1)\left(\beta \log\left(\frac{\beta}{\alpha}\right) + (1-\beta)\log\left(\frac{1-\beta}{1-\alpha}\right)\right)\right).\end{aligned}$$

In the exponent

$$\beta \log\left(\frac{\beta}{\alpha}\right) + (1-\beta)\log\left(\frac{1-\beta}{1-\alpha}\right) = \mathrm{D}(\mathrm{BER}(\beta)\,\|\mathrm{BER}(\alpha)\,).$$

■

Corollary 3.3.2. *If $\alpha \neq \beta$, then*

$$\lim_{n\to\infty} \mathbb{P}\Big(\xi_{\lfloor \beta n\rfloor, n-\lfloor \beta n\rfloor} < \alpha < \xi_{\lfloor \beta n\rfloor+1, n-1-\lfloor \beta n\rfloor}\Big) = 0$$

exponentially fast.

Proof. Since

$$\mathrm{D}(\mathrm{BER}(\beta)\,\|\mathrm{BER}(\alpha)\,) \geq 0$$

and equality holds if and only if $\alpha = \beta$, then for $\alpha \neq \beta$ it is strictly positive and by Equation (8) the probability tends to zero as we let $n \to \infty$. ■

In the second part of the section we focus on Pólya Walks started from points on the same line ($k + l = n$) with a distance $n^{\frac{1}{2}+\varepsilon}$.

Proposition 3.3.3. *Let $0 < \alpha < 1$, $a > 0$ and $\varepsilon \geq 0$ be arbitrary. Let us denote*

$$k^+ = \lceil \alpha n\rceil + \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor \qquad l^+ = \lceil (1-\alpha) n\rceil + \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor$$

$$k^- = \lfloor \alpha n\rfloor - \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor \qquad l^- = \lfloor (1-\alpha) n\rfloor - \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor$$

Then we have the inequality

$$\liminf_{n\to\infty} \mathbb{P}(\xi_{k^-,l^+} < \xi_{k^+,l^-}) \geq \begin{cases} 1 & \text{if } \varepsilon > 0 \\ 2\Phi\left(\frac{a}{\sqrt{\alpha(1-\alpha)}}\right) - 1 & \text{if } \varepsilon = 0 \end{cases}$$

Proof. Using Equation (7) we have that

$$\mathbb{P}(\xi_{k^-,l^+} < \xi_{k^+,l^-}) \geq \mathbb{P}\Big(\lfloor \alpha n\rfloor - \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor \leq X_n \leq \lceil \alpha n\rceil + \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor\Big),$$

where

$$X_n \sim \mathrm{BIN}(n-1, \alpha),$$

since

$$\lceil \alpha n\rceil + \lfloor (1-\alpha) n\rfloor = \lfloor \alpha n\rfloor + \lceil (1-\alpha) n\rceil = n.$$

After standardizing we get

$$\begin{aligned}&\mathbb{P}(\xi_{k^-,l^+} < \xi_{k^+,l^-})\\ &\geq \mathbb{P}\left(\frac{\lceil \alpha n\rceil - \alpha(n-1) - \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}} \leq \frac{X_n - \alpha(n-1)}{\sqrt{(n-1)\alpha(1-\alpha)}} \leq \frac{\lfloor \alpha n\rfloor - \alpha(n-1) + \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}}\right).\end{aligned}$$

Then using the Central Limit Theorem and the fact that the cumulative distribution function of the standard normal distribution is continuous we get

$$\begin{aligned}&\liminf_{n\to\infty} \mathbb{P}(\xi_{k^-,l^+} < \xi_{k^+,l^-})\\ &\quad \geq \lim_{n\to\infty} \Phi\left(\frac{\lceil \alpha n\rceil - \alpha(n-1) + \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}}\right) - \Phi\left(\frac{\lfloor \alpha n\rfloor - \alpha(n-1) - \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}}\right). \quad (9)\end{aligned}$$

Nocite that if $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \frac{\lceil \alpha n\rceil - \alpha(n-1) + \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}} = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{\lfloor \alpha n\rfloor - \alpha(n-1) - \left\lfloor a n^{\frac{1}{2}+\varepsilon}\right\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}} = -\infty.$$

In case of $\varepsilon = 0$ we have

$$\lim_{n\to\infty} \frac{\lceil \alpha n\rceil - \alpha(n-1) + \lfloor a\sqrt{n}\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}} = \frac{a}{\sqrt{\alpha(1-\alpha)}}$$

and

$$\lim_{n\to\infty} \frac{\lfloor \alpha n\rfloor - \alpha(n-1) - \lfloor a\sqrt{n}\rfloor}{\sqrt{(n-1)\alpha(1-\alpha)}} = -\frac{a}{\sqrt{\alpha(1-\alpha)}}.$$

After substituting into Equation (9) we get the desired formula. ■

4 Global properties

4.1 Bound on the expected number of components

In this section we consider the following problem. Let us fix an $n \in \mathbb{N}^+$ and start Pólya Walks from each (i, j) point with $i + j = n$. Consider the random graph we obtain this way (Figure 11). Let us denote the number of components of the graph by C_n. We show an asymptotic lower bound on the expected number of C_n.

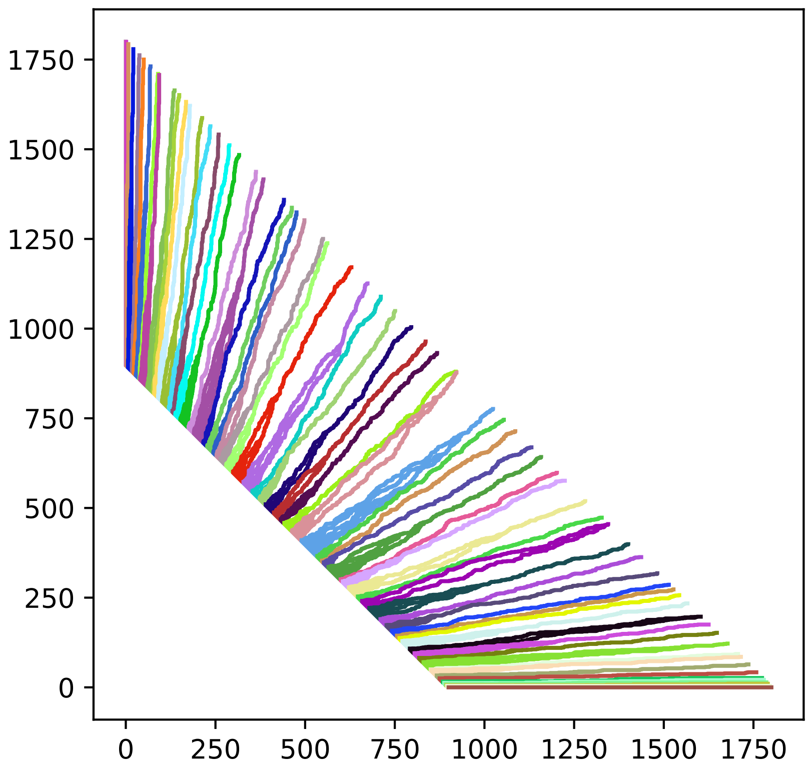

Figure 11: Random walks started simultaneously from the line $i + j = n = 900$. The different colors denote the different components of the graph.

Proposition 4.1.1. *The following inequality holds*

$$\sqrt{\frac{\pi}{2}} \leq \liminf_{n\to\infty} \frac{\mathbb{E}(C_n)}{\sqrt{n}}.$$

Proof. Notice that the following equality holds

$$C_n = \sum_{k=0}^{n-1} \mathbb{1}\left[\bigcup_{m=n}^{\infty} \left\{S^{(m)}_{k,n-k} \neq S^{(m)}_{k+1,n-1-k}\right\}\right].$$

Thus for the expected number of components we have the following equality

$$\mathbb{E}(C_n) = \sum_{k=0}^{n-1} \mathbb{P}\left(\bigcup_{m=n}^{\infty} \left\{ S_{k,n-k}^{(m)} \neq S_{k+1,n-1-k}^{(m)} \right\} \right).$$

Notice that by Proposition 2.6.2 we have the following inequality

$$\mathbb{P}\left(\bigcup_{m=n}^{\infty} \left\{ S_{k,n-k}^{(m)} \neq S_{k+1,n-1-k}^{(m)} \right\} \right) \geq \mathbb{P}(\xi_{k,n-k} < \xi_{k+1,n-1-k}).$$

Then for any $0 \leq \alpha_k \leq 1$

$$\mathbb{E}(C_n) \geq \sum_{k=0}^{n-1} \mathbb{P}(\xi_{k,n-k} < \alpha_k < \xi_{k+1,n-1-k}).$$

Now let $0 < \varepsilon_1 < \varepsilon_2 < 1$. Then let us denote

$$n_1 = \lceil \varepsilon_1(n-1) \rceil \qquad \text{and} \qquad n_2 = \lfloor \varepsilon_2(n-1) \rfloor.$$

In this case the following inequality also holds with choosing $\alpha_k = \frac{k}{n-1}$

$$\begin{aligned} \mathbb{E}(C_n) &\geq \sum_{k=n_1}^{n_2} \mathbb{P}\left(\xi_{k,n} < \frac{k}{n-1} < \xi_{k+1,n} \right) \\ &= \sum_{k=n_1}^{n_2} \binom{n-1}{k} \left(\frac{k}{n-1} \right)^k \left(1 - \frac{k}{n-1} \right)^{n-1-k} \end{aligned} \tag{10}$$

where for the last equality we used Equation (6). Using the following Stirling's approximation formula

$$\sqrt{2\pi n} n^n e^{-n} \cdot e^{\frac{1}{12n} - \frac{1}{360n^3}} < n! < \sqrt{2\pi n} n^n e^{-n} \cdot e^{\frac{1}{12n}},$$

we have the following for any $n_1 \leq k \leq n_2$

$$\binom{n-1}{k} \left(\frac{k}{n-1} \right)^k \left(1 - \frac{k}{n-1} \right)^{n-1-k} > a_{k,n} \cdot \frac{1}{\sqrt{2\pi(n-1)}} \cdot \frac{1}{\sqrt{\left(\frac{k}{n-1}\right)\left(1 - \frac{k}{n-1}\right)}},$$

where

$$a_{k,n} = \exp\left(\frac{1}{12(n-1)} - \frac{1}{12k} + \frac{1}{360k^3} - \frac{1}{12(n-1-k)} + \frac{1}{360(n-1-k)^3} \right).$$

Then for the sequence

$$a_n = \exp\left(\frac{1}{12(n-1)} - \frac{1}{12\varepsilon_1(n-1)} + \frac{1}{360(\varepsilon_2(n-1))^3} - \frac{1}{12(1-\varepsilon_2)(n-1)} + \frac{1}{360((1-\varepsilon_1)(n-1))^3} \right)$$

we have that

$$\lim_{n\to\infty} a_n = 1,$$

and for any $n_1 \leq k \leq n_2$

$$a_{k,n} \geq a_n.$$

Thus after substituting into Equation (10) we get

$$\mathbb{E}(C_n) \geq \sum_{k=n_1}^{n_2} a_n \cdot \frac{1}{\sqrt{2\pi(n-1)}} \cdot \frac{1}{\sqrt{\left(\frac{k}{n-1}\right)\left(1-\frac{k}{n-1}\right)}}.$$

After rearranging

$$\frac{\mathbb{E}(C_n)}{\sqrt{n}} \geq a_n \cdot \sqrt{\frac{n-1}{n}} \cdot \frac{1}{\sqrt{2\pi}} \sum_{k=n_1}^{n_2} \frac{1}{n-1} \cdot \frac{1}{\sqrt{\left(\frac{k}{n-1}\right)\left(1-\frac{k}{n-1}\right)}},$$

and taking the limit, the following holds for any $0 < \varepsilon_1 < \varepsilon_2 < 1$

$$\liminf_{n\to\infty} \frac{\mathbb{E}(C_n)}{\sqrt{n}} \geq \frac{1}{\sqrt{2\pi}} \int_{\varepsilon_1}^{\varepsilon_2} \frac{1}{\sqrt{x(1-x)}}\, dx.$$

In this case our proof is finished, since

$$\liminf_{n\to\infty} \frac{\mathbb{E}(C_n)}{\sqrt{n}} \geq \lim_{\varepsilon_1\to 0} \lim_{\varepsilon_2\to 1} \frac{1}{\sqrt{2\pi}} \int_{\varepsilon_1}^{\varepsilon_2} \frac{1}{\sqrt{x(1-x)}}\, dx = \frac{1}{\sqrt{2\pi}} \int_0^1 \frac{1}{\sqrt{x(1-x)}}\, dx = \sqrt{\frac{\pi}{2}}.$$

■

5 Outlook

In this short section we provide an outlook about the future researches about the Pólya Web we plan to do in the near future.

- Our first main objective is to prove Conjecture 2.6.3 in order to characterize the paths of the coupled Pólya Walks with only the coupling of the limiting distributions. Hopefully this will also bring us closer to give an upper bound in Proposition 4.1.1.
- A second goal is to find the "density" of the all the realizations of the limiting ratios. Our conjecture is that it follows a path on $[0, 1]$ similar to the curve

$$\frac{1}{\sqrt{x(1-x)}}$$

as seen in Figure 11. Notice that this is not a precise mathematical statement yet.

- Finally we wish to project $\mathbb{R}^+ \times \mathbb{R}^+$ on the triangle with edges $(0, 0), (1, 0)$ and $(1, 0)$ by the following map

$$(x, y) \mapsto \left(\frac{x}{x+y+1}, \frac{y}{x+y+1}\right).$$

By doing so we hope to be able to define and characterize the paths of Dual Pólya Walks started from "points at infinity". This approach also might help in proving our first goal stated above.

The Pólya Web has turned out to be a rich and interesting object. We are enthusiastic and determined to continue our research on the topic in the future.

References

[1] Thomas Bayes (1763) LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, Royal Society

[2] Joseph K. Blitzstein, Jessica Hwang (2019) Introduction to Probability, Second Edition, Chapman & Hall/CRC Texts in Statistical Science, Chapman and Hall/CRC

[3] Imre Csiszár, Paul C. Shields (2004) Information Theory and Statistics: A Tutorial, Foundations and Trends™ in Communications and Information, Vol 1, No 4 (2004), 417-528

[4] Rick Durrett (2010) Probability: Theory and Examples, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press

[5] Florian Eggenberger, György Pólya (1923) Über die Statistik verketteter Vorgänge, Z. Angew. Math. Mech.: (3), 279-289.

[6] Theodore Edward Harris (1960) A lower bound for the critical probability in a certain percolation process, Proc. Cambr. Philos. Soc. 56 13–20.

[7] Sébastien Roch (2024) Modern Discrete Probability. An Essential Toolkit, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press

[8] Bálint Tóth, Wendelin Werner (1998) The true self-repelling motion, Probability Theory and Related Fields 111: (3), 375-452.

[9] David Williams (1991) Probability with martingales, Cambridge Mathematical Textbooks, Cambridge University Press