1 Discrete Time Markov Chains (DTMC)

Definition 1 (Markov Chain)**.** *Let X_t for $t = 0, 1, 2, ...$ be a sequence of random variables. We say that X_t is a Markov Chain with state space Ω and transition matrix* $\mathbf{P}$ *if for any two states $x, y \in \Omega$:*

$$\begin{aligned} P\left(X_{t+1} = y | X_t = x, ..., X_1 = x_1, X_0 = x_0\right) &= P\left(X_{t+1} = y | X_t = x\right) \\ &\doteq P\left(x, y\right) \end{aligned}$$

This property is what we call the Markov Property, when the future depends on the past only through the present.

Remark 1. *We use the following notation for this chapter:*

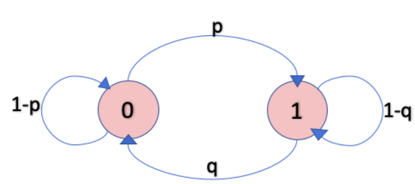
$$\begin{aligned} P\left(x, y\right) &= P\left(X_{t+1} | X_t\right) \\ &\neq P\left(X_{t+1}, X_t\right) \end{aligned}$$

Remark 2. *$P\left(x, y\right)$ is actually the element corresponding to the row of state x and the column of state y in what we call the transition matrix* $\mathbf{P}$.

$$\begin{array}{cc} \boxed{states} & y \downarrow \\ x \rightarrow & \begin{bmatrix} & & \\ & \boxed{\quad P\left(x, y\right) \quad} & \\ & & \end{bmatrix} \end{array} \quad \sum_y P\left(x, y\right) = 1 \, for \, all \, x \in \Omega$$

The sum of all elements in a raw of $\mathbf{P}$ *is 1 and all elements are positive, then* $\mathbf{P}$ *is a stochastic matrix.*
As we are dealing with time homogeneous chains then $\mathbf{P}$ *does not variate with t.*

Example 1. *Two states Markov chain: $\Omega = \{0, 1\}$*



The transition matrix for the chain represented is the above state diagram is

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & \quad\quad 1 \end{array} \\ \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \end{array}$$

Many applications could be modeled using a two states Markov Chain:

- *The Gilbert Elliot model for wireless channels:* $\Omega = \{good\, channel, bad\, channel\}$
- *Tomorrow's weather prediction based on today's weather:* $\Omega = \{snow, no\, snow\}$

Remark 3. *If the future depends on the present and one state before like* $P(X_{t+1}|X_t, X_{t-1})$*, then we can model the system where a state is not anymore the individual state but a pair* $(X_t,\, X_{t-1})$

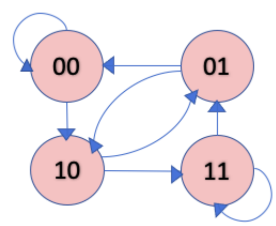


Figure 1: The state diagram of the Markov chain of the augmented space

Now the state space is of the size $|\Omega|^2 = 4$

Example 2. *Random Walk on* $\mathbb{Z}$

A random walk moves right or left by at most one step on each move.

A state X_t *is defined by*

$$X_t = W_0 + W_1 + W_2 + ... + W_t$$

where $W_i s$ *are iid random variables drawn from the following distribution:*

$$W_{i\neq 0} = \begin{cases} 1 & with\, probability\, p \\ -1 & with\, probability\, 1-p \end{cases}$$

The future state X_{t+1} *depends only on the current state* X_t

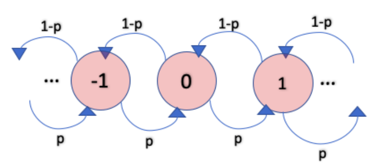


Figure 2: Random Walk on $\mathbb{Z}$

The state space $\Omega = \mathbb{Z}$*. It is infinite.*

Example 3. *Gambler's Ruin (Check Homework 1)*

A Gambler has \$$k$ initially. He flips a coin at each time and proceed as follow:

$$\begin{cases} \textit{if he gets a Head} & \rightarrow \textit{he wins\$1} \\ \textit{if he gets a Tail} & \rightarrow \textit{he loses \$1} \end{cases}$$

Once he reaches \$0 *or a maximum sum he wishes to win* \$$n$*, he stops palying.*

The Gambler's Ruin problem can be modeled as a random walk on a finite Markov chain bounded by the state 0 from below and the targeted sum n from above with an initial state X_0 equals to the initial sum k.

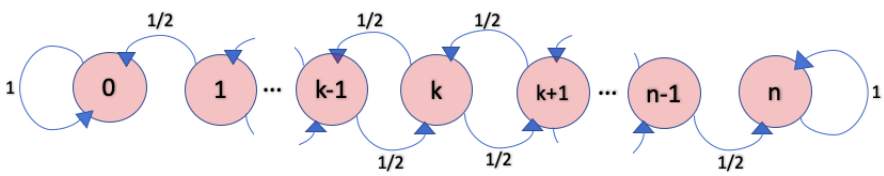


Figure 3: The state diagram of the Gambler's Ruin Markov chain

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \\ \\ k \\ \\ \\ n-1 \\ n \end{array} \begin{array}{c} \begin{array}{ccccccccccc} 0 & 1 & 2 & & & & & & & & \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & & & & & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & & & & & \\ & & \frac{1}{2} & 0 & \frac{1}{2} & & & & & \\ \vdots & & & & & & & & & \vdots \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & \frac{1}{2} & 0 & \frac{1}{2} & \\ & & & & & & & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Example 4. *Random Walk on Graph*

Starting at state 0, at each time instant, we jump to one of the neighbors with equal probability.

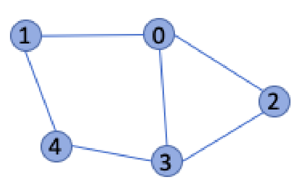


Figure 4: An undirected graph of 5 vertices

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

Special case: *Random walk on a cycle C_n*

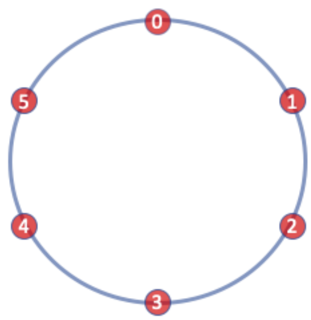


Figure 5: A cycle of 6 vertices

2 Two-state Markov Chains

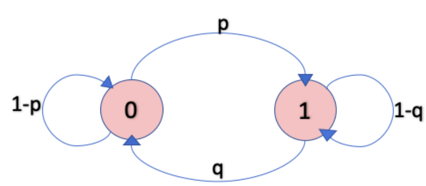


Figure 6: The state diagram of two states Markov chain

The question that we are seeking to answer is what is the $P(X_t|X_0)$?

- **For t = 1:**
  - $P(X_1 = 0|X_0 = 0) = 1 - p$
  - $P(X_1 = 1|X_0 = 0) = p$

- **For t = 2:** Conditionning on $X_0 = 0$
  - $P(X_2 = 0|X_0 = 0) = (1 - p)^2 + pq$
    It is the probability to stay at **0** at t = 1 and t = 2, or leave **0** at t = 0 and get back to it at t = 2
  - $P(X_2 = 1|X_0 = 0) = (1 - p)\,p + (1 - q)\,p$
    It is the probability to stay at **0** at t = 1 and reach **1** at t = 2, or leave **0** at t = 0 and stay at **1** at t = 2

What about any t?

First, for a deterministic initial state like $\mu_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}$, we get:

$$\begin{aligned} P(X_t) &= P(X_t|X_0=0)\,P(X_0=0) + P(X_t|X_0=1)\,P(X_0=1) \\ &= P(X_t|X_0=0) \end{aligned}$$

We note by $\mu_t = \begin{bmatrix} P(X_t=0) & P(X_t=1) \end{bmatrix}$

Suppose we know the distribution μ_t at time t, then

$$\begin{aligned} P(X_{t+1}=0) &= P(0,0)\,P(X_t=0) + P(1,0)\,P(X_t=1) \\ P(X_{t+1}=1) &= P(0,1)\,P(X_t=0) + P(1,1)\,P(X_t=1) \end{aligned}$$

We can condensate the equations above vectorially by

$$\mu_{t+1} = \mu_t \mathbf{P}$$

By iterating this equation starting at t = 0, we get

$$\begin{aligned} \mu_1 &= \mu_0 \mathbf{P} \\ \mu_2 &= \mu_0 \mathbf{P^2} \\ &\ldots \\ \mu_t &= \mu_0 \mathbf{P^t} \end{aligned}$$

Lemma 1. *The state distribution μ_t at time t of a Markov chain of transition matrix $\mathbf{P}$ is*

$$\mu_t = \mu_0 \mathbf{P^t}$$

3 Stationary distribution

We wish to know to what distribution π the chain is going to converge to after running it for enough time if such limiting distribution exists. So,

$$\pi = \lim_{t\to\infty} \mu_t$$

Since

$$\lim_{t\to\infty} \mu_t = \lim_{t\to\infty} \mu_{t+1} = \pi$$

then π should statisfy

$$\pi = \pi \mathbf{P}$$

Definition 2. *The stationary distribution π of a Markov chain with a transition matrix $\mathbf{P}$ is the solution of the equation*

$$\pi = \pi \mathbf{P}$$

Example 5. *The two-state Markov chain*

$$\begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix} = \begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix} \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$
$$\pi_0 = \pi_0 (1-p) + \pi_1 q$$
$$\pi_1 = \pi_0 p + \pi_1 (1-q)$$

$$\pi_0 = \frac{q}{p+q}$$
$$\pi_1 = \frac{p}{p+q}$$

Then,

$$\lim_{t\to\infty} P(X_t = 0) = \frac{q}{p+q}$$
$$\lim_{t\to\infty} P(X_t = 1) = \frac{p}{p+q}$$

4 Finite Time Analysis

The question to answer now is how fast we are converging to π?

Recall that $\mu_t = \mu_0 \mathbf{P}^t$ and suppose that $\mathbf{P}$ is diagonalizable so it can be written as:

$$\mathbf{P} = U\Lambda U^{-1} \quad for\ \Lambda = \begin{bmatrix} \lambda_0 & 0 & ... & 0 \\ 0 & \lambda_1 & & \\ \vdots & & & \\ 0 & & & \lambda_n \end{bmatrix}$$

Then,

$$\mathbf{P}^2 = U\Lambda U^{-1} U \Lambda U^{-1}$$
$$= U\Lambda^2 U^{-1}$$

for any t,

$$\mathbf{P}^t = U\Lambda^t U^{-1}$$

Now let us go back to the two-state Markov chain:

Finding the eigenvalues of $\mathbf{P}$ of the previous example for all p and β:

$$\det(\mathbf{P} - \lambda I) = \det\left(\begin{bmatrix} 1-p-\lambda & p \\ q & 1-q-\lambda \end{bmatrix}\right)$$
$$= (1-p-\lambda)(1-q-\lambda) - pq$$
$$= (\lambda - 1)(\lambda + p + q - 1)$$

Hence we get two eigenvalues, $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$. Their corresponding eigenvectors and the matrix U are the following:

$$\underset{\sim}{\Phi}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \underset{\sim}{\Phi}_2 = \begin{bmatrix} p \\ -q \end{bmatrix}$$

Therefore, if $p + q \neq 0$

$$U = \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix}$$
$$U^{-1} = \frac{1}{-p-q} \begin{bmatrix} -q & -p \\ -1 & 1 \end{bmatrix}$$

Otherwise, $\quad \underset{\sim}{\Phi}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. which is not linearly independent of $\underset{\sim}{\Phi}_1$

So, for $p + q \neq 0$ and with $\mu_0 = \begin{bmatrix} 1 & 0 \end{bmatrix}$:

$$\mu_t = \mu_0 \mathbf{P^t}$$
$$\mu_t = \mu_0 U \Lambda^t U^{-1}$$
$$\mu_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (1-(p+q))^t \end{bmatrix} \frac{1}{p+q} \begin{bmatrix} q & p \\ 1 & -1 \end{bmatrix}$$

Then,

$$\mu_t^T = \begin{bmatrix} \frac{q}{q+p} - \frac{q}{q+p}(1-p-q)^t \\ \\ \frac{p}{q+p} - \frac{q}{q+p}(1-p-q)^t \end{bmatrix}$$

For $|1 - (p+q)| < 1$

$$lim_{t \to \infty} \mu_t = \begin{bmatrix} \frac{q}{q+p} & \frac{p}{q+p} \end{bmatrix} = \pi$$

We can see that the chain converge exponentially to its stationary distribution π. This convergence is determined by λ_2: the second largest eigenvalue.

Now for $|1 - (p+q)| = 1$

- Case 1: $p = q = 1$

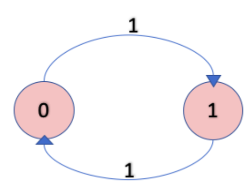


Figure 7: The state diagram of the case when $p = q = 1$

The chain never converge to a stationary distribution. Depending on the initial state, let us say $X_0 = 0$, the chain is going to be at 0 for any even t and at 1 for any odd t regardless since when the chain started.

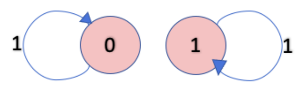


Figure 8: the state diagram for the case when $p = q = 0$

- Case 2: $p = q = 0$

  The chain will be stuck in 1 or 0 depending on the initial state and never get out of it. 1 and 0 are called absorbing states.

Remark 4. *When a matrix is diagonalizable?*

Theorem : *If* $\mathbf{P}$ *is a* $n \times n$ *matrix with* n *linearly independent eigenvectors* $\phi_0, \phi_1, ...\phi_n$*, then* $\mathbf{P}$ *is diagonalizable*

Fact*: If* $\mathbf{P}$ *has* n *distincts eigenvalues, then* $\mathbf{P}$ *has* n *linearly independent eigenvectors. (Refer to Linear Algerbra textbooks)*

Remark 5. *A stochastic matrix has always an eigenvalue* $\lambda_1 = 1$*. All other eigenvalues are in absolute value smaller than 1.*

Proof. *For the matrix* $\mathbf{P}$*, the sum of the row vectors is equal to 1. The matrix* $\mathbf{P}$ *has the eigenvalue 1 for the eigenvector* $\begin{bmatrix} 1 & 1 & ... & 1 \end{bmatrix}^T$

$$\mathbf{P}.\begin{bmatrix} 1 & 1 & ... & 1 \end{bmatrix}^{\mathbf{T}} = \begin{bmatrix} 1 & 1 & ... & 1 \end{bmatrix}^{\mathbf{T}}$$

Assume now that v *is an eigenvalue* $|\lambda| > 1$*. Then* $A^t = |\lambda|^t v$ *has exponentially growing length for* $t \to \infty$*. This implies that there is for large* t *one coefficient* $\left[A^t\right]_{ij}$ *which is larger than 1. but* A^t *is a stochastic matrix as well and has all entries* ≤ 1*(easy to prove). The assumption of an eigenvalue larger than 1 can not be valid.*

5 Convergence Analysis: Irreducibility and Aperiodicity

We have seen two cases where a 2 states Markov Chain does not converge to a stationary distribution. How about a chain with any number of states?

Irreducibility

Let us consider the following chain:

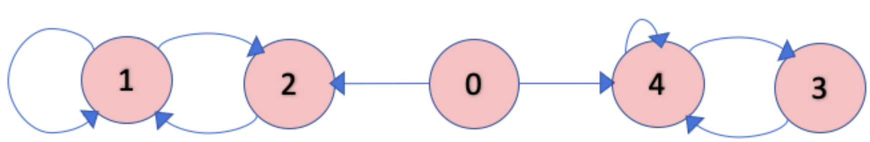


Figure 9: The state diagram of an irreducible Markov chain

Two possible scenarios are possible for the limiting distribution:

$$\pi = \left[\begin{array}{ccccc} * & * & 0 & 0 & 0 \end{array}\right]$$
$$or$$
$$\pi = \left[\begin{array}{ccccc} 0 & 0 & 0 & * & * \end{array}\right]$$

Suppose that $X_0 = 0$ the chain is reduced to one of two classes depending on whether we go from **0** to **2** or **4** at $t = 1$. Such a chain does not converge to a stationary distribution.

Definition 3. *Irreducibility*

A Markov chain is irreducible if for any two states $x\, and\, y \in \Omega$, it is possible to go from x to y in a finite time t:

$$P^t\left(x, y\right) > 0, for\, some\, t \geq 1\, for\, all\, x,\, y \in \Omega$$

Definition 4. *A class in a Markov chain is a set of states that are all reacheable from each other.*

Lemma 2. *Any transition matrix* $\mathbf{P}$ *of an irreducible Markov chain has a unique distribution stasfying* $\pi = \pi\mathbf{P}$.

Periodicity:

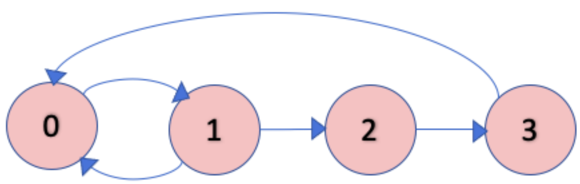


Figure 10: The state diagram of a periodic Markov chain

This chain is irreducible but that is not sufficient to prove the convergence

Starting at state 0:

- For all even time t: the chain can be only at 0 or 2
- For all odd time t: the chain can be at 1 or 3

Then, there is no convergence.

Definition 5. *The period of a state x is defined by*

$$d\left(x\right) = gcd\left\{t | P^t\left(x, x\right) > 0\right\}$$

Applying this definition to our example chain:

- $d\left(0\right) = gcd\left\{2,\, 4\right\} = 2$
- $d\left(1\right) = gcd\left\{2,\, 4\right\} = 2$

- $d(2) = gcd\{2, 4, 6, 8, ...\} = 2$
- $d(3) = gcd\{2, 4, 6, 8, ...\} = 2$

Definition 6. *A state is periodic if $d(x) > 1$. Otherwise, we call it aperiodic.*

Lemma 3. *All states of an irreducible chain has the same period.*

Example 6. *Consider the following chains:*

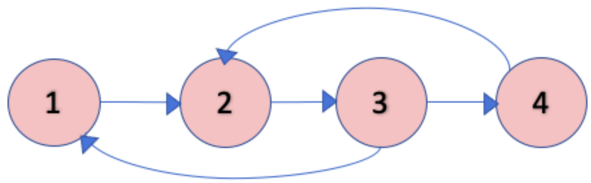


Figure 11: Chain 1

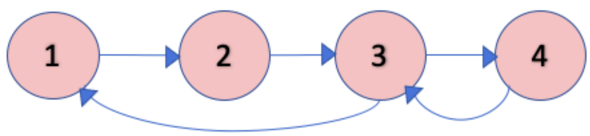


Figure 12: Chain 2

Questions

1. *Are these two chains irreducible?*
2. *Are they periodic?*

Anwers

1. *Both chains 1 and 2 are irreducible. We can reach any state starting by any other.*
2. *For the first chain: Starting at 1, it could take you $\{3, 6, ..\}$ to go back to 1. Then, the period of 1 = 3. The chain is irreducible which implies that all states are of period 3. So it is periodic.*

   For the second chain: Starting at 1, it could take you $\{3, 5, ..\}$ to go back to 1.Then, the period of 1 = 1. The chain is irreducible which implies that all states are of period 1. So it is aperiodic.

Theorem 1. *A finite state Markov chain that is irreducible and aperiodic converges to its stationary distribution π given by $\pi = \pi\mathbf{P}$, i.e., $\lim_{t\to\infty} \mu_t = \pi$*

6 Random Walk on Graph

Let $G(V, E)$ be a graph with $|E| = m$ edges.

The state diagram of a random walk on this graph is:

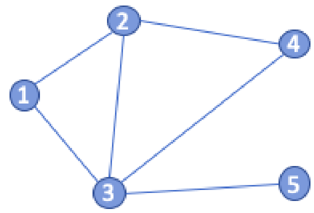


Figure 13: Undirected graph of 5 nodes

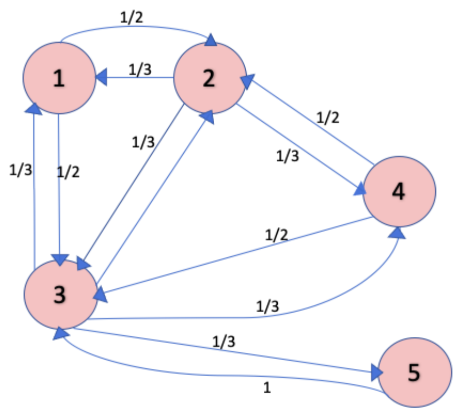


Figure 14: State diagram of the random walk on the previous graph

1. What is the stationary distribution of a random walk on a graph?

   Recall if $x,\, y\, \in V$, then,

$$P\left(x,\, y\right) = \begin{cases} \frac{1}{deg(x)} & if\, x\, is\, adjcent\, to\, y \\ 0 & otherwise \end{cases}$$

   where $deg\left(x\right)$ is the number of outgoing edges from node x. So, for the graph above:

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

   The stationary distribution π must satisfy $\pi = \pi\mathbf{P}$.

   The number of non-zero elements of the y^{th} column of $\mathbf{P}$ $= deg\left(y\right)$. For example, for y=2:

$$\frac{1}{2} * 2 + \frac{1}{3} * 3 + \frac{1}{2} * 2 = 3 = deg\left(2\right)$$

$$\frac{1}{2} deg\left(1\right) + \frac{1}{3} deg\left(3\right) + \frac{1}{2} deg\left(4\right) = deg\left(2\right)$$

In general for n vertices,

$$deg\,(y) = \sum_{x} deg\,(x)\,P\,(x,y)$$
$$Deg = Deg\,\mathbf{P}$$

where Deg is $[deg\,(1)\,,\,deg\,(2)\,,\,...,\,deg\,(n)]$ the degree vector.

But Deg is not a probability to consider it for a stationary distribution so we have to normalize it by $\sum_y deg\,(y) = 2|E| = 2m$

Then,

$$\pi = \frac{1}{2m}\,[deg\,(1)\,,\,...,\,deg\,(n)]$$

Lemma 4. *The stationary distribution of a random walk on an undirected graph $G\,(V,\,E)$ of n vertices and m edges is:*

$$\pi = \frac{1}{2m}\,[deg\,(1)\,,\,...,\,deg\,(n)]$$

2. Does this MC converge to its statinary distribution?

   - If the graph is connected then the MC is irreducible.
   - **Aperiodicity?** Let us focus on a particular family of graph: C_n cycles of n vertices:
     For n odd :

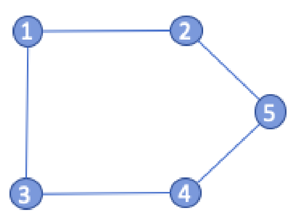


Figure 15: Odd cycle

To go from **1** to **1**:

- 2 steps: $\mathbf{1} - \mathbf{2} - \mathbf{1}$
- 5 steps: $\mathbf{1} - \mathbf{2} - \mathbf{5} - \mathbf{4} - \mathbf{3} - \mathbf{1}$
- ...

Then the period of $\mathbf{1}$ = gcd (2, 5, ..) = 1. The chain is aperiodic

For n even :

The period of **1** is 2. So this chain is periodic and does not converge. The stationary distribution is $\pi = \left[\frac{1}{4}\,\frac{1}{4}\,\frac{1}{4}\,\frac{1}{4}\right]$ and it is about the fraction of time that the chain spends on each state here. While the probability of being at a given state depends always on the first state and if that time is even or odd.

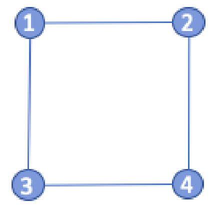


Figure 16: Even cycle

The point is a period of 2 exits always as we can leave a state to one of its neighbors and come back directly. So any graph that only has an even number of cycles is going to be periodic. We call these graphs: Bipartite graphs.

The get an aperiodic graph out of a periodic one, we can always add self loops to all states of the graph fairly and assigned a probability 1/2 to each. We call this graph the Lazy version.

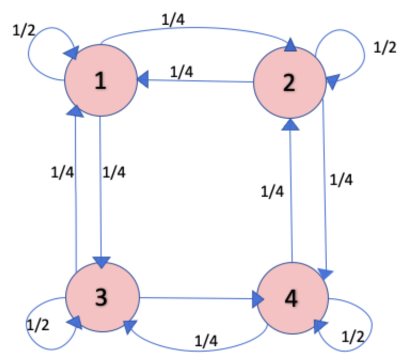


Figure 17: The lazy version of the 4-cycle

7 The Pòlya Urn Model

We consider an urn containing 2 balls: white and black at the beginning of the process at $k = 0$. At each time k, we choose uniformly a random ball from the urn, then return it and add a new ball of the same color. Let B_k the number of balck balls at time k.

1. Is B_k a Markov chain?
   - First B_k can take values: $1, 2, 3, \dots, k+1$ and the total number of balls is $k+1$
     - $P\left(B_{k+1} = j+1 | B_k = j\right) = \frac{j}{k+2}$
     - $P\left(B_{k+1} = j | B_k = j\right) = 1 - \frac{j}{k+2}$

     B_{k+1} just depends on the number of balls at time k regardless how we have reached B_k
2. What is the distribution of B_k on $\{1, 2, 3, \dots, k+1\}$?

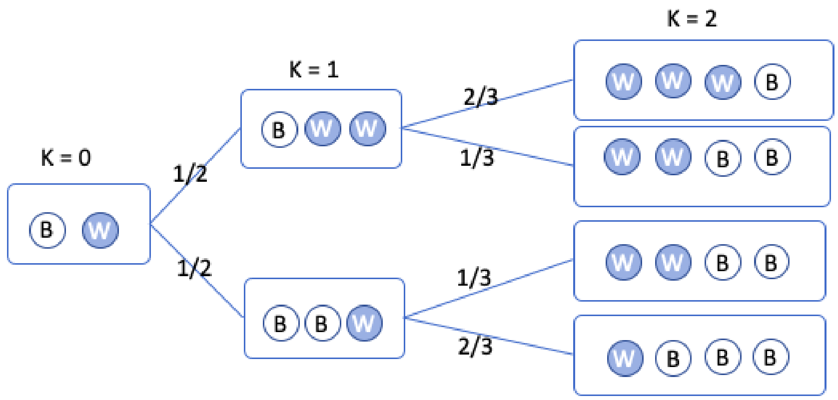


Figure 18: The possible outcomes for k=0,1 and 2 of the Pòlya urn model

- Checking First for B_2

$$
\begin{aligned}
P\left(B_2=1\right) &= \frac{2}{3}\frac{1}{2} = \frac{1}{3} \\
P\left(B_2=2\right) &= \frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{1}{3} = \frac{1}{3} \\
P\left(B_2=3\right) &= \frac{1}{2}\frac{2}{3} = \frac{1}{3}
\end{aligned}
$$

  We can see that B_2 is uniformly distributed over {1, 2, 3, ...}

- We assume that B_k is uniformly distributed on $\{1,2,3,...,k+1\}$, and we will proceed by induction to prove it.

  Let us prove that B_{k+1} is uniformly distributed over $\{1,2,3,...,k+2\}$

- $P\left(B_{k+1}=j\right)=?$
  - For $j=1$:

$$
\begin{aligned}
P\left(B_{k+1}=1\right) &= P\left(B_k=1\right)\ P\left(Picking\, a\, White\, ball\right) \\
&= \frac{1}{k+1}\frac{k+1}{k+2} \\
&= \frac{1}{k+2}
\end{aligned}
$$

  - For $1<j<k+2$:

$$
\begin{aligned}
P\left(B_{k+1}=j\right) &= P\left(B_k=j\right)\ P\left(Picking\, a\, White\, ball\right) + P\left(B_k=j-1\right)\ P\left(Picking\, a\, Black\, ball\right) \\
&= \frac{1}{k+1}\frac{k+2-j}{k+2} + \frac{1}{k+1}\frac{j-1}{k+2} \\
&= \frac{1}{k+1}\frac{k+1}{k+2} \\
&= \frac{1}{k+2}
\end{aligned}
$$

– For $j = k + 2$:

$$\begin{aligned} P\left(B_{k+1} = k+2\right) &= P\left(B_k = k+1\right)\, P\left(Picking\, a\, Black\, ball\right) \\ &= \frac{1}{k+1}\,\frac{k+1}{k+2} \\ &= \frac{1}{k+2} \end{aligned}$$

So the assumption is proved for $k + 1$.

8 The Ballot Markov Chain

We are intrested in counting the votes of two candidates A and B. We suppose that A got a votes and B got b votes with $b > a$. The ballots are mixed together, and the vote counting happens by picking one ballot randomly and updating the score according to it.

1. Show that the score after each ballot count a MC.

   Let $X_t = (a_t, b_t)$ where a_t is the number of revealed votes for candidate A up to t, and b_t is the votes of candidate B up to t.

   So, $t = a_t + b_t$, and the remaing votes to count equal $a + b - t$.

$$\begin{aligned} P\left(X_{t+1} = (x+1,\, t-x)\,|X_t = (x,\, t-x)\right) &= \frac{a-x}{a+b-t} \\ P\left(X_{t+1} = (x,\, t-x+1)\,|X_t = (x,\, t-x)\right) &= \frac{b-t+x}{a+b-t} \end{aligned}$$

2. Show that all paths from (0, 0) to $(a,\, b)$ have the same probability.

$$\begin{aligned} P\left(any\, path\right) &= \frac{a\,(a-1)\,(a-2)\,...\,1\,b\,(b-1)\,...\,1}{(a+b)\,(a+b-1)\,...\,1} \\ &= \frac{a!\,b!}{(a+b)!} \\ &= \frac{1}{\left(\begin{array}{c} a+b \\ b \end{array}\right)} \end{aligned}$$

3. What is the probability of B always leading in the vote count?

   This is the probability that the path from (0, 0) to $(a,\, b)$ never crosses the $x = y$ line. Since all paths have the same probability:

$$\begin{aligned} P\left(B\, always\, leading\right) &= P\left(the\, path\, never\, crosses\, x = y\right) \\ &= \frac{\#\, of\, paths\ (0,0)\ to\ (a,b)\, that\, do\, not\, cross\, x = y}{total\, number\, of\, paths} \end{aligned}$$

   where $total\, number\, of\, paths = \left(\begin{array}{c} a+b \\ b \end{array}\right)$

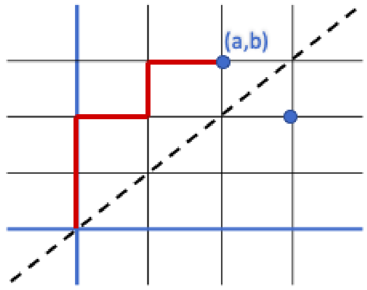


Figure 19: One good path to have for the votes counting

Let us count the number of paths that do not cross $x = y$. Any such path must start with a step up meaning a vote for B: # of such paths $= \begin{pmatrix} a+b-1 \\ a \end{pmatrix}$

Then I can take the part right after crossing $x = y$ and reflet it to get all paths from $(0, 1)$ to (b, a). There is a bijection between all paths from $(0, 0)$ to (a, b) that cross $x = y$ and all paths from $(0, 1)$ to (b, a).

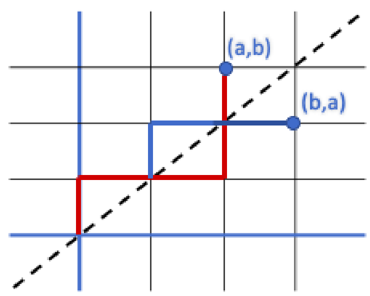


Figure 20: Reflexion Principle

#of all paths from $(0, 1)$ to $(b, a) = \begin{pmatrix} a+b-1 \\ b \end{pmatrix}$

Then,

$$\begin{aligned} P\left(B\, always\, leading\right) &= \frac{\begin{pmatrix} a+b-1 \\ a \end{pmatrix} - \begin{pmatrix} a+b-1 \\ b \end{pmatrix}}{\begin{pmatrix} a+b \\ b \end{pmatrix}} \\ &= \frac{b-a}{b+a}. \end{aligned}$$