1.5 **Arc Sine law** - last visit to origin in $2N$ time

$$L = \max\{0 \le n \le 2N \mid S_n = 0\}$$

(not a stopping time)

"**Order of L**" : time after which one player has gain in capital

$$L \approx 2N$$

Theorem 5: $n \in \mathbb{N}$ upto $n \le N$

$$\mathbb{P}(L = 2n) = \frac{1}{2^{2N}} \binom{2n}{n} \binom{2N-2n}{N-n}$$

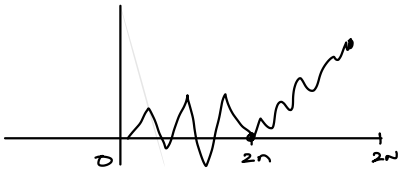
Proof: (next class)

(next class) $\mathbb{P}(L=2n)$ (stirling formula) $\simeq \frac{1}{N} f\left(\frac{n}{N}\right)$

$n! \sim n^n e^{-n}$

Proof :-

# of paths of length $2N$ with $L = 2n$



$$= \left(\begin{array}{c}\#\text{ of paths of length } 2n \\ \text{with } S_{2n} = 0\end{array}\right) \times \left(\begin{array}{c}\#\text{ of paths of length } 2N-2n \\ \text{with } \sigma_0 > 2N-2n\end{array}\right)$$

1st $2n$ tosses — next $2N-2n$ tosses

$$\therefore \quad \mathbb{P}(L = 2n) = \mathbb{P}(S_{2n} = 0)\ \mathbb{P}(\sigma_0 > 2N-2n)$$

$$\overset{\text{(lemma 4)}}{=} \mathbb{P}(S_{2n} = 0)\ \mathbb{P}(S_{2N-2n} = 0)$$

$$= \binom{2n}{n} \frac{1}{2^{2n}} \binom{2N-2n}{N-n} \frac{1}{2^{2N-2n}}$$

$$= \frac{1}{2^{2N}} \binom{2n}{n} \binom{2N-2n}{N-n}$$

Discrete arc sine law

□

Sketch (Ex: Make it precise)

$$\mathbb{P}(L = 2n) = \text{[Stirling's approximation } n! \sim n^n e^{-n}\text{]}$$

$$\overset{\text{Ex}}{\simeq} \frac{1}{\pi} \frac{1}{\sqrt{n(N-n)}}$$

$$= \frac{1}{\pi} \cdot \frac{1}{2} \frac{1}{\sqrt{\frac{n}{2N}\left(1 - \frac{n}{2N}\right)}}$$

$f : [0,1) \to \mathbb{R}$

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$

will involve constants

$$\mathbb{P}(L = 2n) \sim \frac{1}{2} f\left(\frac{n}{2N}\right)$$

Intuition : $\{S_n\}_{n \geq 1}$ interpretation vis a vis game

L – time at which particular takes the lead for ever.

$$E[S_{2N}] = 0 \Rightarrow \quad L \simeq 2N$$

Distribution function of scaled random variable

$$\mathbb{P}\left(\frac{L}{2N} \leq x\right) = \sum_{n \leq \lfloor Nx \rfloor} \mathbb{P}(L = 2Nn)$$

$$\sim \sum_{n \geq 1 \atop \frac{n}{N} \leq x} \frac{1}{N} f\left(\frac{n}{N}\right)$$

Riemann approximation $\left(\xrightarrow[N \to \infty]{}\right)$

$$\simeq \int_0^x f(y)\, dy$$

$$= \int_0^x \frac{1}{\pi \sqrt{y(1-y)}}\, dy$$

$$= \frac{2}{\pi} \arcsin\left(\sqrt{x}\right)$$

1.6 Infinite length random walks

2.1 Higher dimension finite length walks

$$\mathbb{Z}^d = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix} \;\middle|\; x_i \in \mathbb{Z} \quad 1 \le i \le d \right\}$$

Fix $N \ge 1$, $\qquad |x| = \sqrt{\sum_{i=1}^{d} x_i^2}$

$$\Omega_N = \{ \omega = (\omega_1 \dots \omega_N) : \omega_k \in \mathbb{Z}^d, \; |\omega_k| = 1, \; 1 \le k \le N \}$$

$$X_k : \Omega_N \to \mathbb{Z}^d \qquad X_k(\omega) = \omega_k$$

$$S_n = \sum_{i=1}^{n} X_i(\omega) \; ; \; 1 \le n \le N$$

$$S_n = \begin{bmatrix} S_n^{(1)} \\ \vdots \\ S_n^{(d)} \end{bmatrix} \qquad 0 \le n \le N, \; S_n^{(j)} \in \mathbb{Z}, \; 1 \le j \le d$$

$$\mathbb{P}_N(A) = \frac{|A|}{(2d)^N}, \quad A \subseteq \Omega_N$$

Observe that our construction of finite length walks of length N had the following property:

$0 < N < M$

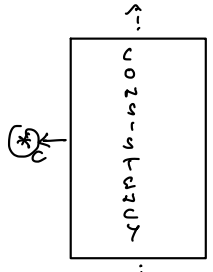
$\mathbb{P}_N(\cdot)$ $\qquad$ $\mathbb{P}_M(\cdot)$

$$\omega = (\underbrace{\omega_1, \dots, \omega_N}_{\tilde{\omega} \in \Omega_N}, \omega_{N+1}, \dots, \omega_M) \in \Omega_M$$

$$\Pi_N : \Omega_M \to \Omega_N, \quad \omega \to \tilde{\omega}$$

Fix $\tilde{\omega} \in \Omega_N$:

$$\mathbb{P}_M\left(\{\omega \in \Omega_M : \Pi_N(\omega) = \tilde{\omega}\}\right)$$



$$= \frac{\left|\{\omega \in \Omega_M : \Pi_N(\omega) = \tilde{\omega}\}\right|}{(2d)^M} = \frac{(2d)^{M-N}}{(2d)^M}$$

$$= (2d)^{-N}$$

$$\mathbb{P}_N(\{\tilde{\omega}\}) = \frac{1}{(2d)^N}$$

Construction for $N = \infty$ case

$$\Omega_\infty = \left\{ (\omega_m)_{m=1}^{\infty} \;\middle|\; \omega_m \in \mathbb{Z}^d,\ |\omega_m| = 1,\ 1 \le m \right\}$$

$$X_k : \Omega_\infty \to \mathbb{Z}^d \qquad X_k(\omega) = \omega_k \qquad 1 \le k$$

$$S_n = \sum_{k=1}^{n} X_k \qquad n \ge 1$$

define: $\mathbb{P}$ on $(\Omega_\infty, \mathcal{F}_\infty)$ $\longrightarrow$ { exist

$\forall N \ge 1$, $\tilde{\omega} \in \Omega_N$

$$\mathbb{P}\left(\{\omega \in \Omega_\infty \mid \Pi_N(\omega) = \tilde{\omega}\}\right) := \frac{1}{(2d)^N} = \mathbb{P}_N(\{\tilde{\omega}\})$$

Kolmogorov Consistency Theorem:

If $(\Omega_N, \mathcal{P}(\Omega_N), \mathbb{P}_N)$ satisfying (*)$_C$ Then

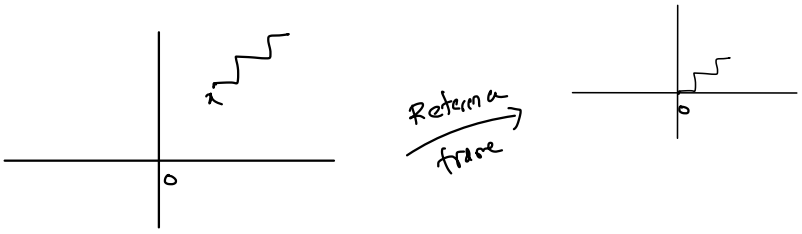
$\mathbb{P}$ on $(\Omega_\infty, \mathcal{F}_\infty)$ such $\mathbb{P}|_{\Omega_N} = \mathbb{P}_N$

Notation / Convention:

- $(\Omega_\infty, \mathbb{P})$ and $\{S_n\}_{n\geq 1}$ as simple random walk starting at 0.

- $\{S_n\}_{n\geq 1}$ to start at x, define another Probability $\mathbb{P}_x(\cdot)$

$$\mathbb{P}_x(S_1 = s_1, \dots, S_n = s_n) = \mathbb{P}(S_1 + x = s_1, \dots, S_n + x = s_n)$$



$$\mathbb{E}_x[f(S_1, \dots, S_n)] = \mathbb{E}[f(S_1 + x, \dots, S_n + x)]$$

Generalisations:

Independence

- (Both length & equal. probabilities can be tweaked)

- Each step is of size 1 / length 1
- Probability of each step = same

Strong law of large numbers: $\{S_n\}_{n\geq 1}$ is a simple random walk on $\mathbb{Z}^d$.

$$A = \{\omega \in \Omega_0 \mid \frac{S_n}{n} \longrightarrow 0 \text{ as } n \to \infty\}$$

then $\mathbb{P}(A) = 1$.

Typical position of the walk

$d=1$ $\{S_n\}_{n\geq 1}$ is simple random walk on $\mathbb{Z}$.

Central limit theorem:

$$\mathbb{P}\left(\frac{S_n}{\sqrt{n}} \leq x\right) \longrightarrow \int_{-\infty}^{x} \frac{e^{-y^2/2}}{\sqrt{2\pi}}\, dy$$

as $n \to \infty$

1

MARKOV CHAINS

Markov chains were invented by Russian mathematician Andrey Markov in 1906. He was trying to understand whether independence was essential to prove the law of large numbers for a sequence of random variables. The example Markov considered was the alternation of consonants and vowels in Pushkin's Eugene Omegin. The theory of Markov chains provided a challenging problem to many probabilists in the years to come. The simplcity of the model and the ability to resolve complex behviour in the real world makes it one of the most succesfully applied model from Probability theory with application in the social sciences, biology, physics, electrical engineering, computer science, and many other fields.

Earlier Francis Galton studied the survival of English peerage and posed the problem in Educational Times in 1887. These were independently solved by H.W. Watson and I.J. Bienaymé and in current literature known as the Galton-Watson branching chain. In the same time period Poincare also studided models for gambling tradition in France. The models can be viewed as Markov chains on finite groups. Paul and Tatiana Ehrenfest modeled diffusion of gasses using a finite state Markov chain for clarifying irreversibility in Thermodynamics.

In everyday life, we encounter systems which can be in various possible states. For example, the price of a stock at the Mumbai Stock exchange or the number of people waiting to purchase the first show ticket at a theatre screening Rajnikanth movie, or the number of tigers alive at Sariska national park. In each of the above cases the system moves from one possible state to another in a fashion affected by randomness. To analyze these systems mathematically we make the assumption of one crucial property – given the present, the past is independent of the future. This feature enables the systems to be modeled as Markov chains which allows us to make predictions about their long time behaviour.

In Section 1.1, we will define a Markov chain precisely. The present-day notations used to understand and work with Markov chains are due to A.N. Kolmogorov, widely regarded as one of the founders of modern Probability Theory.

1.1 DISCRETE TIME MC

We shall first fix some notation. Let S denote the set of possible states of the system. We shall assume that S is a countable set. Let $\mathcal{P}(S)$ be the power set of S. We shall say $\mu : \mathcal{P}(S) \to [0, 1]$ is a probability on S if

(1) $\mu(S) = 1$; and

(2) If $E_1, E_2, \ldots$ are a countable collection of disjoint events (that is, $E_i \cap E_j = \varnothing$ if $i \neq j$), then

$$\mu(\bigcup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} \mu(E_j). \tag{1.1.1}$$

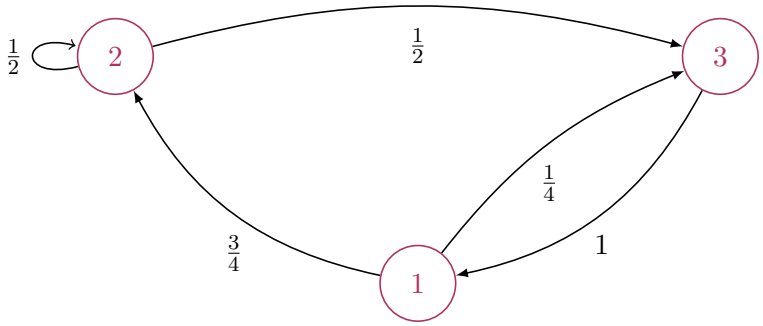
We shall also assume that "time", n, is given by the set of whole numbers Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ (one could think of units of time as minutes or hours or days etc.) and we let X_n denote the state of the system at time n. We shall refer to X_n as *Markov chain* or *chain*.

At each time step n, the chain X_n moves from a state $i \in S$ to another state $j \in S$ with probability p_{ij}. We allow for the possibility that $i = j$ which simply means that the chain did not change state at that time step. The value p_{ij} only depends on i and j, thus ensuring that once we know the present state of the chain, the future does not depend on the past. Note that for well definedness of the model, the probabilities need to satisfy two conditions. For every pair of states i and j we need $p_{ij} \in [0, 1]$ (so that these values indeed represent probabilities) and for every state i we have $\sum_{k \in S} p_{ik} = 1$ (so that when the chain transitions from state i it arrives at some other state in S with probability 1). The matrix $\mathbf{P} = [p_{ij}]_{i,j \in S}$ is referred to as the *one step transition matrix* or simply as *transition matrix*. Let us first consider a simple example to illustrate the notation.

EXAMPLE 1.1.1. Let $S = \{1, 2, 3\}$ and suppose the chain moves according to the following mechanism:

- from state 1 the chain will move to state 2 with probability $\frac{3}{4}$.
- from state 1 the chain will move to state 3 with probability $\frac{1}{4}$.
- from state 2 the chain will move to state 3 with probability $\frac{1}{2}$.
- from state 2 the chain will remain at state 2 with probability $\frac{1}{2}$.
- from state 3 the chain will move to state 1 with probability 1.

The above dynamics of the chain can be understood via graphs.



The graph is self-explanatory and at a glance describes how the chain will move at each time step. The arrows indicates the movement across states and the numbers on each arrow indicate the probability with which that move will happen. From the graph it is also readily seen that the transition matrix P is given by

$$P = \begin{bmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \end{bmatrix}$$

where the entries p_{ij} are the probability of moving from state $i \in S$ to state $j \in S$. ■

We are now ready to define precisely what is meant by a Markov chain.

DEFINITION 1.1.2. *Let μ be any probability on S. Let $\mathbf{P} = [p_{ij}]_{i,j\in S}$ be a matrix for which the entries satisfiy $p_{ij} \in [0,1]$ for all $i, j \in S$ and $\sum_{k=1}^{\infty} p_{ik} = 1$ for each $i \in S$.*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and for $n \geq 0$, let $X_n : \Omega \to S$ be a sequence of random variables whose joint distribution is given by

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = \mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{k-1} i_k}, \tag{1.1.2}$$

for all $k \geq 0$. Then X_n is said to be a Markov chain on state space S, with initial distribution μ, and transition matrix $\mathbf{P}$.

For the precise construction of the probability space and Markov chain in Definition 1.1.2, we refer the reader to [AS09, Theorem 4.3.5] when S is finite and [AS09, Theorem 4.4.4] when S is countably infinite. We will assume the result and shall now detail some of the properties of a Markov chain X_n.

(a) (Distribution of X_n) Using the definition of Probability, (1.1.2), and properties of the transition matrix $\mathbf{P}$ we have

$$\begin{aligned}\mathbb{P}(X_0 = i_0) &= \sum_{j\in S} \mathbb{P}(X_0 = i_0, X_1 = j) \\ &= \sum_{j\in S} \mu(\{i_0\}) p_{i_0,j} \\ &= \mu(\{i_0\}) \sum_{j\in S} p_{i_0,j} \\ &= \mu(\{i_0\}).\end{aligned}$$

Hence $X_0 \sim \mu$, i.e. the distribution of X_0 is given by μ which justifies describing μ as the initial distribution. To calculate the probability X_n is found in a state j we can decompose the event taking the disjoint union over all possible initial states and all possible paths from those initial states that finish at state j at time n. Specifically, for $n \geq 1$, using the definition of Probability and (1.1.2),

$$\begin{aligned}\mathbb{P}(X_n = j) &= \mathbb{P}(\cup_{i_0,i_1,\ldots i_{n-1}\in S}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1} X_n = j\}) \\ &= \sum_{i_0,i_1,\ldots i_{n-1}\in S} \mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_0,i_1,\ldots i_{n-1}\in S} \mu(\{i_0\}) \left(\prod_{k=1}^{n-1} p_{i_{k-1}i_k}\right) p_{i_{n-1},j}. \qquad (1.1.3)\end{aligned}$$

In particular, note that the distribution of X_n is determined solely by μ and the transition matrix $\mathbf{P}$.

(b) (Markov Property) Every Markov chain has an underlying memoryless property. For $n \in \mathbb{N}$ let $i_0, i_1, \ldots, i_{n-2}, i, j \in S$ is such that

$$\mathbb{P}(X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) > 0,$$

then

$$\begin{aligned}&\mathbb{P}(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) \\ &= \frac{\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0)}{\mathbb{P}(X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, X_0 = i_0)} \\ &= \frac{\mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{ij}}{\mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}}} \\ &= p_{ij} \qquad (1.1.4)\end{aligned}$$

and as $\mathbb{P}(X_{n-1} = i) \geq \mathbb{P}(X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) > 0$, we may calculate the conditional probability $\mathbb{P}(X_n = j | X_{n-1} = i) = \frac{\mathbb{P}(X_n=j, X_{n-1}=i)}{\mathbb{P}(X_{n-1}=i)}$ by using (1.1.3) and decomposing the probability into all possible paths for which $X_{n-1} = i$ and $X_n = j$. Specifically,

$$\begin{aligned}
&\mathbb{P}(X_n = j | X_{n-1} = i) \\
&= \frac{\mathbb{P}(X_n = j, X_{n-1} = i)}{\mathbb{P}(X_{n-1} = i)} \\
&= \frac{\mathbb{P}(\cup_{i_0,i_1,\ldots i_{n-2} \in S} \{X_n = j, X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = i_0\})}{\mathbb{P}(\cup_{i_0,i_1,\ldots i_{n-2} \in S} \{X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = i_0\})} \\
&= \frac{\sum_{i_0,i_1,\ldots,i_{n-2} \in S} \mu(\{j_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i} p_{ij}}{\sum_{i_0,i_1,\ldots,i_{n-2} \in S} \mu(\{j_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i}} \\
&= p_{ij} \cdot \frac{\sum_{i_0,i_1,\ldots,i_{n-2} \in S} \mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i}}{\sum_{i_0,i_1,\ldots,i_{n-2} \in S} \mu(\{i_0\}) p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-2} i}} \\
&= p_{ij}
\end{aligned} \tag{1.1.5}$$

Hence from (1.1.4) and (1.1.5) we have

$$\mathbb{P}(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0) = p_{ij} = \mathbb{P}(X_n = j | X_{n-1} = i) \tag{1.1.6}$$

So, it does not matter how the chain arrived at state i for deciding how likely it is that it will jump to state j in the next step. Thus for a Markov chain, the past and the future are independent given the present. This is referred to as the Markov property of X_n and can be rephrased below

(c) (n-th step Transition Probability Matrix) From (1.1.6), we have for $n \geq 1, i, j \in S$

$$\mathbb{P}(X_n = j | X_{n-1} = i) = p_{ij}$$

if $\mathbb{P}(X_{n-1} = i) > 0$. This is why the entries of $\mathbf{P}$ are called the (one-step) transition probabilities and the matrix $\mathbf{P}$ the (one-step) transition probability matrix of the Markov chain X_n.

Let $\mathbf{P}^n$ be the n-th power of the matrix $\mathbf{P}$ and $p_{ij}^{[n]}$ denote the (i,j)-th element of $\mathbf{P}^n$. Suppose $i \in S$ is such that $\mathbb{P}(X_0 = i) > 0$. For any $j \in S$, we have

$$\begin{aligned}
\mathbb{P}(X_n = j | X_0 = i) &= \frac{1}{\mathbb{P}(X_0 = i)} P(X_n = j, X_0 = i) \\
&= \frac{1}{\mu(\{i\})} \sum_{i_1,\ldots,i_{n-1} \in S} \mathbb{P}(X_n = j, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i) \\
&= \frac{1}{\mu(\{i\})} \sum_{i_1,\ldots,i_{n-1} \in S} \mu(\{i\}) p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j} \\
&= \sum_{i_1,\ldots,i_{n-1} \in S} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j}.
\end{aligned}$$

Now by definition of matrix multiplication we have that

$$p_{ij}^{[n]} = \sum_{i_1,\ldots,i_{n-1} \in S} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j}.$$

Thus,

$$\mathbb{P}(X_n = j \mid X_0 = i) = p_{ij}^{[n]} \tag{1.1.7}$$

Therefore the entries $p_{ij}^{[n]}$ are called the n-step transition probabilities for the chain since they are the probabilities that a chain in state i at time 0 will be in state j after exactly n steps. Likewise the matrix $\mathbf{P}^n$ is referred to as the n-step transition probability matrix of the chain.

This notation even makes sense when $n = 0$. The matrix $\mathbf{P}^0$ is the identity matrix. Therefore $p_{ij}^{[0]}$ will be 1 if $i = j$ and will be 0 if $i \lneq j$.

(d) (Chapman-Kolmogrov Equation) For $r, s \geq 0$ and $r + s = n$, we have

$$\mathbf{P}^n = \mathbf{P}^{r+s} = \mathbf{P}^r \mathbf{P}^s.$$

Standard rules governing multiplication of matrices implies that

$$p_{ij}^{[n]} = \sum_{k \in S} p_{ik}^{[r]} p_{kj}^{[s]}. \tag{1.1.8}$$

By (1.1.7), we can conclude

$$\mathbb{P}(X_n = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_r = k | X_0 = i) \mathbb{P}(X_s = j | X_0 = k) \tag{1.1.9}$$