# STA447
# Lecture Notes

Yuchen Wang

March 9, 2020

## Contents

# 1 Preliminary

**Theorem 1.1** (Dominated Convergence Theorem). If $\lim_{n \to \infty} X_n = X$, and there is some random variable $Y$ with $E|Y| < \infty$ and $|X_n| \leq Y$ for all $n$, then

$$\lim_{n \to \infty} E(X_n) = E(X)$$

**Definition 1.1** (weak convergence). $X_n$ converge to $X$ weakly if

$$\forall \epsilon > 0, \lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0$$

**Definition 1.2** (strong convergence). $X_n$ converge to $X$ strongly if

$$P(\lim_{n \to \infty} X_n = X) = 1$$

**Theorem 1.2** (Law of Large Numbers). If the sequence $\{X_n\}$ is i.i.d. with common mean $m$, then the sequence $\frac{1}{n} \sum_{i=1}^{n} X_i$ converges to $m$ (both weakly and strongly), i.e.,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = m \quad w.p.1$$

# 2 Markov Chain Probabilities

**Notation 2.1.**
$$P(X_{n+1} = j | X_n = i) = p_{ij}$$

**Definition 2.1** (Markov chain). A (discrete time, discrete space, time homogeneous) <u>Markov chain</u> is specified by three ingredients:

- A <u>state space</u> $S$, any non-empty finite or countable set.

- <u>Initial probabilities</u> $\{v_i\}_{i \in S}$, where $v_i$ is the probability of starting at $i$ (at time 0). (So $v_i \geq 0$ and $\sum_i v_i = 1$)

- <u>Transition probabilities</u> $\{p_{ij}\}_{i,j \in S}$, where $p_{ij}$ is the probability of jumping to $j$ if you start at $i$. (So $p_{ij} \geq 0$, and $\sum_j p_{ij} = 1$ for all $i$)

**Remark 2.1** (Markov property).

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = P(X_{n+1} = j | X_n = i_n) = p_{i_n j}$$

i.e. The probabilities at time $n+1$ depend only on the state at time $n$.

**Remark 2.2.**

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = v_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}$$

## 2.1 Markov Chain examples

**Example 2.1** (the Frog Walk). Let $X_n :=$ pad index the frog is at after $n$ steps.

$$S = \{1, 2, 3, \ldots, 20\}$$
$$v_{20} = 1, v_i = 0 \, \forall i \neq 20$$
$$p_{ij} = \begin{cases} \frac{1}{3}, & |j - i| \leq 1 \text{ or } |j - i| = 19 \\ 0, & \text{otherwise} \end{cases}$$

**Example 2.2** (Bernoulli process).

$$S = \{1, 2, 3, \ldots\}$$
$$v_0 = 1, v_i = 0 \, \forall i \neq 0$$
$$p_{ij} = \begin{cases} p, & j = i + 1 \\ 1 - p, & j = i \\ 0, & \text{otherwise} \end{cases}$$

where $0 < p < 1$.

**Example 2.3** (Simple random walk (s.r.w.)). Let $X_n :=$ net gain (in dollars) after $n$ bets

$$S = \{0, 1, 2, 3, \ldots\}$$
$$v_a = 1, v_i = 0 \, \forall i \neq a$$
$$p_{ij} = \begin{cases} p, & j = i + 1 \\ 1 - p, & j = i - 1 \\ 0, & \text{otherwise} \end{cases}$$

where $0 < p < 1, a \in \mathbb{Z}$.
**Special case:** When $p = 1/2$, call it <u>simple symmetric random walk</u>.

**Example 2.4** (Ehrenfest's Urn). Let $X_n :=$ # balls in Urn 1 at time $n$.
We have $d$ balls in total, divided into two urns. At each time, we choose one of the $d$ balls uniformly at random, and move it to the other urn.

$$S = \{1, 2, 3, \ldots, d\}$$
$$v_a = 1, v_i = 0 \, \forall i \neq a$$
$$p_{ij} = \begin{cases} (d - i)/d, & j = i + 1 \\ i/d, & j = i - 1 \\ 0, & \text{otherwise} \end{cases}$$

## 2.2 Elementary Computations

**Notation 2.2.**
$$\mu_i^{(n)} := P(X_n = i)$$

**Notation 2.3.**

$$m := |S| \qquad \text{(the number of elements in S, could be infinity)}$$
$$\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \mu_3^{(n)}, \dots) \qquad (m \times 1)$$
$$v = (v_1, v_2, v_3, \dots) \qquad (m \times 1)$$
$$P = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & \\ & & \ddots & \\ p_{m1} & \cdots & & p_{mm} \end{pmatrix} \qquad (m \times m \text{ matrix})$$

**Fact 2.1.**

$$\mu^{(1)} = vP = \mu^{(0)}P$$
$$\mu^{(n)} = vP^n = \mu^{(0)}P^n$$

**Notation 2.4.**

$$p_{ij}^{(n)} := P(X_n = j, X_0 = i) = P(X_{m+n} = j | X_m = i) \qquad \text{(for any } m \in \mathbb{N})$$

**Fact 2.2.**

$$\sum_{j \in S} p_{ij}^{(n)} = 1$$
$$p_{ij}^{(1)} = p_{ij}$$
$$P^{(n)} = P^n \qquad \text{(for all } n \in \mathbb{N})$$

**Notation 2.5.**

$$P^0 := I$$
$$P^{(0)} := I$$
$$p_{ij}^{(0)} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

**Theorem 2.1** (Chapman-Kolmogorov equations)**.**

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}$$
$$P_{ij}^{(m+s+n)} = \sum_{k \in S} \sum_{l \in S} p_{ik}^{(m)} p_{kl}^{(s)} p_{lj}^{(n)}$$

Matrix form:

$$P^{(m+n)} = P^{(m)} P^{(n)}$$
$$P^{(m+s+n)} = P^{(m)} P^{(s)} P^{(n)}$$

**Theorem 2.2** (Chapman-Kolmogorov Inequality)**.**

$$p_{ij}^{(m+n)} \geq p_{ik}^{(m)} p_{kj}^{(n)} \qquad \text{(for all } k \in S)$$

$$P_{ij}^{(m+s+n)} \geq p_{ik}^{(m)} p_{kl}^{(s)} p_{lj}^{(n)} \qquad \text{(for any } k, l \in S)$$

## 2.3  Recurrence and Transience

**Notation 2.6.**

$$P_i(\ldots) \equiv P(\ldots | X_0 = i)$$
$$E_i(\ldots) \equiv E(\ldots | X_0 = i)$$
$$N(i) = \#\{n \geq 1 : X_n = i\}$$

(total number of times that the chain hits $i$, not counting time 0)

**Definition 2.2** (return probability)**.** Let $f_{ij}$ be the <u>return probability</u> from $i$ to $j$.

$$f_{ij} := P_i(X_n = j \text{ for some } n \geq 1) \equiv P_i(N(j) \geq 1)$$

**Fact 2.3.**

$$1 - f_{ij} = P_i(X_n \neq j \text{ for all } n \geq 1) \tag{1}$$
$$P_i(N(i) \geq k) = (f_{ii})^k \tag{2}$$
$$P_i(N(j) \geq k) = f_{ij}(f_{jj})^{k-1} \tag{3}$$
$$f_{ik} \geq f_{ij} f_{jk} \tag{4}$$

**Fact 2.4.** $f_{ij} > 0$ iff $\exists m \geq 1$ with $p_{ij}^{(m)} > 0$, i.e., there is some time $m$ for which it is possible to get from $i$ to $j$ in $m$ steps.

**Definition 2.3** (recurrent and transient states)**.** A state $i$ of a Markov chain is <u>recurrent</u> if $f_{ii} = 1$. Otherwise, $i$ is <u>transient</u> if $f_{ii} < 1$.

**Proposition 2.1.** If $Z$ is a non-negative integer, then

$$E(Z) = \sum_{k=1}^{\infty} P(Z \geq k)$$

**Theorem 2.3** (Recurrent State Theorem)**.** As follows

- State $i$ is recurrent $\iff P_i(N(i) = \infty) = 1 \iff \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$

- State $i$ is transient $\iff P_i(N(i) = \infty) = 0 \iff \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$

*Proof.*

$$\begin{aligned} P_i(N(i) = \infty) &= \lim_{k \to \infty} P_i(N(i) \geq k) & \text{(by continuity of probabilities)} \\ &= \lim_{k \to \infty} (f_{ii})^k & (P_i(N(i) \geq k) = (f_{ii})^k) \\ &= \begin{cases} 1, & f_{ii} = 1 \\ 0, & f_{ii} < 1 \end{cases} \end{aligned}$$

∎

Therefore,

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} P_i(X_n = i)$$

$$= \sum_{n=1}^{\infty} E_i(\mathbb{1}\{X_n = i\})$$

$$= E_i(\sum_{n=1}^{\infty} \mathbb{1}\{X_n = i\})$$

$$= E_i(N(i))$$

$$= \sum_{k=1}^{\infty} P_i(N(i) \geq k) \qquad \text{(by proposition 1.1)}$$

$$= \sum_{k=1}^{\infty} (f_{ii})^k$$

$$= \begin{cases} \infty, & f_{ii} = 1 \\ \frac{f_{ii}}{1-f_{ii}} < \infty, & f_{ii} < 1 \end{cases}$$

**Example 2.5** (simple random walk). If $p = 1/2$ then $\forall i, f_{ii} = 1$. If $p \neq 1/2$, then $\forall i, f_{ii} < 1$

*Proof.* Consider state 0. We need to check if $\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$.
If $n$ is odd, then $p_{00}^{(n)} = 0$.
If $n$ is even, $p_{00}^{(n)} = P(\frac{n}{2}$ heads and $\frac{n}{2}$ tails on first $n$ tosses).
This is a Binomial$(n, p)$ distribution, so

$$p_{00}^{(n)} = \binom{n}{n/2} p^{n/2} (1-p)^{n/2}$$

$$= \frac{n!}{[(n/2)!]^2} p^{n/2} (1-p)^{n/2}$$

$$= \frac{(n/e)^n \sqrt{2\pi n}}{[(n/2e)^{n/2} \sqrt{2\pi n/2}]^2} p^{n/2} (1-p)^{n/2} \qquad \text{(Sirling's approximation)}$$

$$= [4p(1-p)]^{n/2} \sqrt{2/\pi n}$$

**Case 1:** If $p = 1/2$, then $4p(1-p) = 1$, so

$$\sum_{n=1} \infty p_{00}^{(n)} = \sum_{n=2,4,6,\dots} \sqrt{2/\pi n}$$

$$= \sqrt{2/\pi} \sum_{n=2,4,6,\dots} n^{-1/2}$$

$$= \sqrt{2/\pi} \sum_{n=1}^{\infty} 2k^{-1/2}$$

$$= \infty$$

Therefore, state 0 is recurrent.

**Case 2:** If $p \neq 1/2$, then $4p(1-p) < 1$, so

$$
\begin{aligned}
\sum_{n=1} \infty p_{00}^{(n)} &= \sum_{n=2,4,6,\dots} [4p(1-p)]^{n/2} \sqrt{2/\pi n} \\
&< \sum_{n=2,4,6,\dots} [4p(1-p)]^{n/2} \qquad \text{(Geometric Series)} \\
&= \frac{4p(1-p)}{1 - 4p(1-p)} \\
&< \infty
\end{aligned}
$$

Therefore, the state 0 is transient.

The same exact calculation applies to any other state $i$. ∎

**Theorem 2.4** (f-Expansion)**.**

$$
f_{ij} = p_{ij} + \sum_{k \in S, k \neq j} p_{ik} f_{kj}
$$

*Proof.*

$$
\begin{aligned}
f_{ij} &= P_i(\exists n \geq 1 : X_n = j) \\
&= \sum_{k \in S} P_i(X_1 = k, \exists n \geq 1 : X_n = j) \\
&= P_i(X_1 = j, \exists n \geq 1 : X_n = j) + \sum_{k \neq j} P_i(X_1 = k, \exists n \geq 1 : X_n = j) \\
&= P_i(X_1 = j)P_i(\exists n \geq 1 : X_n = j | X_1 = j) + \sum_{k \neq j} P_i(X_1 = k)P_i(\exists n \geq 1 : X_n = j | X_1 = k) \\
&= p_{ij}(1) + \sum_{k \neq j} p_{ik}(f_{kj})
\end{aligned}
$$

∎

**Remark 2.3.** The f-Expansion shows that $f_{ij} \geq p_{ij}$.

**Remark 2.4.** It essentially follows from logical reasoning: from $i$, to get to $j$ eventually, we have to either jump to $j$ immediately (with probability $p_{ij}$), or jump to some other state $k$ (with probability $p_{ik}$) and then get to $j$ eventually (with probability $p_{kj}$)

## 2.4 Communicating States and Irreducibility

**Definition 2.4** (communicating states)**.** State $i$ <u>communicates</u> with state $j$, written $i \to j$, if $f_{ij} > 0$.

**Remark 2.5.** i.e. if it is possible to get from $i$ to $j$.

**Notation 2.7.** Write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

**Definition 2.5** (irreducibility). A Markov chain is <u>irreducible</u> if $i \to j$ for all $i, j \in S$, i.e., if $f_{ij} > 0$ for all $i, j \in S$. Otherwise, the chain is <u>reducible</u>.

**Lemma 2.1** (Sum Lemma). If $i \to k$, and $l \to j$, and $\sum_{n=1}^{\infty} p_{kl}^{(n)} = \infty$, then $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$

*Proof.* Since $i \to k$, and $l \to j$, there exists $m, r \geq 1$ s.t. $p_{ik}^{(m)} > 0$ and $p_{lj}^{(r)} > 0$.
By the Chapman-Kolmogorov inequality,

$$p_{ij}^{(m+s+r)} \geq p_{ij}^{(m)} p_{kl}^{(s)} p_{lj}^{(r)}$$

Hence

$$
\begin{aligned}
\sum_{n=1}^{\infty} p_{ij}^{(n)} &\geq \sum_{n=m+r+1}^{\infty} p_{ij}^{(n)} \\
&= \sum_{s=1}^{\infty} p_{ij}^{(m+s+r)} \qquad\qquad (s = n - m - r) \\
&\geq \sum_{s=1}^{\infty} p_{ij}^{(m)} p_{kl}^{(s)} p_{lj}^{(r)} \\
&= \underbrace{p_{ij}^{(m)}}_{+} \underbrace{p_{lj}^{(r)}}_{+} \underbrace{\sum_{s=1}^{\infty} p_{kl}^{(s)}}_{=\infty} \\
&= \infty
\end{aligned}
$$

■

**Corollary 2.1** (Sum Corollary). If $i \leftrightarrow k$, then $i$ is recurrent iff $k$ is recurrent.

*Proof.* Setting $j = i$ and $l = k$ in the Sum Lemma: If $i \leftrightarrow k$, then $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty \iff \sum_{n=1}^{\infty} p_{kk}^{(n)} = \infty$. ■

**Theorem 2.5** (Cases Theorem). For an irreducible Markov chain, either

- (a) $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, and all states are recurrent (<u>recurrent Markov chain</u>); or

- (b) $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ for all $i, j \in S$, and all states are transient (<u>transient Markov chain</u>).

**Theorem 2.6** (Finite Space Theorem). An irreducible Markov chain on a finite state space always falls into case (a), i.e., $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ for all $i, j \in S$, and all states are recurrent.

*Proof.* Choose any state $i \in S$. We have

$$
\begin{aligned}
\sum_{j \in S} \sum_{n=1}^{\infty} p_{ij}^{(n)} &= \sum_{n=1}^{\infty} \sum_{j \in S} p_{ij}^{(n)} \qquad\qquad \text{(exchanging the sums)} \\
&= \sum_{n=1}^{\infty} 1 \\
&= \infty
\end{aligned}
$$

Then if $S$ is finite, it follows that there must exist at least one $j \in S$ with $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$. So we must be in case (a).

■

**Notation 2.8.** For $i \neq j$, let $H_{ij}$ be the event that the chain hits the state $i$ before returning to $j$, i.e.,

$$H_{ij} = \{\exists n \in \mathbb{N} : X_n = i, \text{ but } X_m \neq j \text{ for } 1 \leq m \leq n-1\}$$

**Lemma 2.2** (Hit Lemma)**.** If $j \to i$ with $j \neq i$, then $P_j(H_{ij}) > 0$.

*Proof.* Since $j \to i$, there is some possible path from $j$ to $i$. i.e., there is $m \in \mathbb{N}$ and $x_0, x_1, \ldots, x_m$ with $x_0 = j$ and $x_m = i$ and $p_{x_r x_{r+1}} > 0$ for all $0 \leq r \leq m-1$.
Let $S = \max\{r : x_r = j\}$ be the last time this path hits $j$.
Then $x_S, x_{S+1}, \ldots, x_m$ is a possible path which goes from $j$ to $i$ without first returning to $j$.
Hence $P_j(H_{ij}) \geq P(x_0, x_1, \ldots, x_m) = p_{x_S x_{S+1}} p_{x_{S+1} x_{S+2}} \cdots p_{x_{m-1} x_m} > 0$

■

**Remark 2.6.** If it is possible to get from $j$ to $i$ at all, then it is possible to get from $j$ to $i$ without first returning to $j$.
Intuitively obvious: If there is some path from $j$ to $i$, then the final part of the path (starting with the last time it visits $i$) is a possible path from $j$ to $i$ which does not return to $j$.

**Lemma 2.3** (f-Lemma)**.** If $j \to i$ and $f_{jj} = 1$, then $f_{ij} = 1$

*Proof.* If $i = j$ it is trivial, so assume $i \neq j$.
Since $j \to i$, we have $P_j(H_{ij}) > 0$ by the Hit Lemma.
But one way to never return to $j$ is to first hit $i$ and then from $i$ never return to $j$:

$$P_j(\text{never return to } j) \geq P_j(H_{ij}) P_i(\text{never return to } j)$$

Therefore

$$1 - f_{jj} \geq P_j(H_{ij})(1 - f_{ij})$$

Since $f_{jj} = 1$, then $\underbrace{P_j(H_{ij})}_{>0}(1 - f_{ij}) = 0$

Hence $f_{ij} = 1$.

■

**Lemma 2.4** (Infinite Returns Lemma)**.** For an irreducible Markov chain, if it is recurrent, then

$$P_i(N(j) = \infty) = 1$$

for all $i, j \in S$.
But if it transient, then $P_i(N(j) = \infty) = 0$ for all $i, j \in S$.

*Proof.* Let $i, j \in S$. If the chain is recurrent, then $f_{ij} = f_{jj} = 1$ by the f-Lemma. Then

$$
\begin{aligned}
P_i(N(j) = \infty) &= \lim_{k \to \infty} P_i(N(j) \geq k) \\
&= \lim_{k \to \infty} f_{ij}(f_{jj})^{k-1} \\
&= \lim_{k \to \infty} (1)(1)^{k-1} \\
&= 1
\end{aligned}
$$

If the chain is transient, then $f_{jj} < 1$, then

$$
\begin{aligned}
P_i(N(j) = \infty) &= \lim_{k \to \infty} P_i(N(j) \geq k) \\
&= \lim_{k \to \infty} f_{ij}(f_{jj})^{k-1} \\
&= \lim_{k \to \infty} (1)(f_{jj})^{k-1} \\
&= 0
\end{aligned}
$$

$\blacksquare$

**Theorem 2.7** (Recurrence Equivalence Theorem)**.** If a chain is irreducible, then the following are equivalent (and all correspond to case (a)):

1. There are $k, l \in S$ with $\sum_{n=1}^{\infty} p_{kl}^{(n)} = \infty$.

2. For all $i, j \in S$, we have $\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.

3. There is $k \in S$ with $f_{kk} = 1$, i.e. $k$ is recurrent.

4. For all $j \in S$, we have $f_{jj} = 1$, i.e. all states are recurrent.

5. For all $i, j \in S$, we have $f_{ij} = 1$.

6. There are $k, l \in S$ with $P_k(N(l) = \infty) = 1$.

7. For all $i, j \in S$, we have $P_i(N(j) = \infty) = 1$.

*Proof.* Follow from results that we have already proven

- $1 \implies 2$: Sum Lemma.

- $2 \implies 4$: Recurrent State Theorem (with $i = j$).

- $4 \implies 5$: f-Lemma.

- $5 \implies 3$: immediate.

- $3 \implies 1$: Recurrent State Theorem (with $l = k$).

- $4 \implies 7$: Infinite Returns Lemma.

- 7 $\implies$ 6: Immediate.

- 6 $\implies$ 3: Recurrent State Theorem (with $l = k$).

∎

**Theorem 2.8** (Transience Equivalence Theorem)**.** If a chain is irreducible, then the following are equivalent (and all correspond to case (b)):

1. There are $k, l \in S$ with $\sum_{n=1}^{\infty} p_{kl}^{(n)} < \infty$.

2. For all $i, j \in S$, we have $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$.

3. For all $k \in S$, we have $f_{kk} < 1$, i.e. $k$ is transient.

4. There is $j \in S$ with $f_{jj} < 1$, i.e. some state is recurrent.

5. There are $i, j \in S$ with $f_{ij} < 1$.

6. For all $k, l \in S$, $P_k(N(l) = \infty) = 0$.

7. There are $i, j \in S$ with $P_i(N(j) = \infty) = 0$.

**Remark 2.7** (closed subset note)**.** Suppose a chain is reducible, but it has a closed subset $C \subseteq S$ (i.e. $p_{ij} = 0$ for $i \in C$ and $j \notin C$) on which it is irreducible (i.e. $i \to j$ for all $i, j \in C$). Then, the Recurrence Equivalence Theorem and other results about irreducible chains still apply to the chain when restricted to $C$.

**Proposition 2.2.** For simple random walk with $p > 1/2$, $f_{ij} = 1$ whenever $j > i$. (Similarly, if $p < 1/2$ and $j < i$, then $f_{ij} = 1$.)

*Proof.* Let $X_0 = 0$, and $Z_n = X_n - X_{n-1}$ for $n = 1, 2, \ldots$, so that $X_n = \sum_{i=1}^{n} Z_i$.
Since $Z_n$s iid with $P(Z_n = 1) = p$ and $P(Z_n = -1) = 1 - p$, then by Law of Large Numbers,

$$\lim_{n \to \infty} \frac{1}{n}(Z_1 + Z_2 + \ldots + Z_n) \overset{p}{=} E(Z_1) = p(1) + (1 - p)(-1) = 2p - 1 > 0$$

$$\implies \infty = \lim_{n \to \infty}(Z_1 + Z_2 + \ldots + Z_n)$$
$$= \lim_{n \to \infty} X_n - X_0$$
$$= \lim_{n \to \infty} X_n$$

But if $i < j$, then to go from $i$ to $\infty$, the chain must pass through $j$, so $f_{ij} = 1$.          ∎

# 3   Markov Chain Convergence

## 3.1   Stationary Distributions

**Definition 3.1** (stationary distributions). If $\pi$ is a probability distribution on $S$ (i.e. $\pi_i \geq 0$ for all $i \in S$, and $\sum_{i \in S} \pi_i = 1$), then $\pi$ is <u>stationary</u> for a Markov chain with transition probabilities $(p_{ij})$ if $\sum_{i \in S} \pi_i p_{ij} = \pi_j$ for all $j \in S$ (or $\pi P = \pi$, in matrix notation).

**Remark 3.1.** Intuitively, $\pi$ being stationary means if the chain starts with probabilities $\{\pi_i\}$, then it will keep the same probabilities one time unit later.

**Definition 3.2** (doubly stochastic). A Markov Chain is <u>doubly stochastic</u> if in addition to the usual condition that $\sum_{j \in S} p_{ij} = 1$ for all $i \in S$, $\sum_{i \in S} p_{ij} = 1$ for all $j \in S$.

**Remark 3.2.** This holds for the Frog Example.

**Proposition 3.1.** If a Markov chain with states $S$ satisfies $|S| < \infty$ and is doubly stochastic, then the uniform distribution on $S$ is a stationary distribution.

*Proof.* Let $\{\pi_i\}$ be a distribution such that $\pi_i = \frac{1}{|S|}$.
Then

$$
\begin{aligned}
\sum_{i \in S} \pi_i p_{ij} &= \sum_{i \in S} \frac{1}{|S|} p_{ij} \\
&= \frac{1}{|S|} \sum_{i \in S} p_{ij} \\
&= \frac{1}{|S|}(1) \qquad \qquad \text{(doubly stochastic)} \\
&= \frac{1}{|S|} \\
&= \pi_j
\end{aligned}
$$

Then $\{\pi_i\}$ is stationary.                                           ∎

## 3.2   Searching for Stationary

**Definition 3.3** (reversibility). A Markov chain is <u>reversible</u> (or time reversible, or satisfies detailed balance) with respect to a probability distribution $\{\pi_i\}$ if $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i, j \in S$.

**Proposition 3.2.** If a chain is reversible with respect to $\pi$, then $\pi$ is a stationary distribution.

*Proof.* Reversibility means $\pi_i p_{ij} = \pi_j p_{ji}$, so then for $j \in S$,

$$
\sum_{i \in S} \pi_i p_{ij} = \sum_{i \in S} \pi_j p_{ji} = \pi_j \sum_{i \in S} p_{ji} = \pi_j(1) = \pi_j
$$

∎

**Lemma 3.1** (M-test). Let $\{x_{nk}\}_{n,k \in \mathbb{N}}$ be a collection of real numbers. Suppose that $\lim_{n \to \infty} x_{nk}$ exists for each fixed $k \in \mathbb{N}$. Suppose further that $\sum_{k=1}^{\infty} \sup_n |x_{nk}| < \infty$. Then $\lim_{n \to \infty} \sum_{k=1}^{\infty} x_{nk} = \sum_{k=1}^{\infty} \lim_{n \to \infty} x_{nk}$.

**Proposition 3.3** (Vanishing Probabilities Proposition)**.** If a Markov chain's transition probabilities satisfy that $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$, then the chain does not have a stationary distribution.

*Proof.* Suppose for contradiction that there is a stationary distribution $\pi$. Then we would have $\pi_j = \sum_{i \in S} \pi_i p_{ij}^{(n)}$ for any $n$, so

$$\pi_j = \lim_{n\to\infty} \pi_j = \lim_{n\to\infty} \sum_{i\in S} \pi_i p_{ij}^{(n)}$$

$$
\begin{aligned}
\pi_j &= \lim_{n\to\infty} \pi_j \\
&= \lim_{n\to\infty} \sum_{i\in S} \pi_i p_{ij}^{(n)} \\
&= \sum_{i\in S} \lim_{n\to\infty} \pi_i p_{ij}^{(n)} \qquad \text{(exchange the sum and the limit, which is valid by M-test)} \\
&= \sum_{i\in S} \pi_i \lim_{n\to\infty} p_{ij}^{(n)} \\
&= \sum_{i\in S} 0 \\
&= 0
\end{aligned}
$$

So we would have $\pi_j = 0$ for all $j$. But this means that $\sum_j \pi_j = 0$, which is a contradiction.

∎

**Lemma 3.2** (Vanishing Lemma)**.** If a Markov chain has some $k, l \in S$ with $\lim_{n\to\infty} p_{kl}^{(n)} = 0$, then for any $i, j \in S$ with $k \to i$ and $j \to l$, $\lim_{n\to\infty} p_{ij}^{(n)} = 0$.

*Proof.* Since $k \to i$ and $j \to l$, we can find $r, s \in \mathbb{N}$ with $p_{ki}^{(r)} > 0$ and $p_{jl}^{(s)} > 0$. Then by the Chapman-Kolmogorov Inequality,

$$p_{kl}^{(r+n+s)} \geq p_{ki}^{(r)} p_{ij}^{(n)} p_{jl}^{(s)}$$

Hence

$$p_{ij}^{(n)} \leq p_{kl}^{(r+n+s)} / p_{ki}^{(r)} p_{jl}^{(s)}$$

But the assumptions imply that

$$\lim_{n\to\infty} \left[ p_{kl}^{(r+n+s)} / p_{ki}^{(r)} p_{jl}^{(s)} \right] = 0$$

Hence

$$0 \leq \lim_{n\to\infty} p_{ij}^{(n)} \leq 0$$

$$\implies \lim_{n\to\infty} p_{ij}^{(n)} = 0$$

∎

**Corollary 3.1** (Vanishing Together Corollary)**.** For an irreducible Markov chain, either

1. $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$, or

2. $\lim_{n \to \infty} p_{ij}^{(n)} \neq 0$ for all $i, j \in S$

**Corollary 3.2** (Vanishing Probabilities Corollary)**.** If an irreducible Markov chain's transition probabilities satisfy that $\lim_{n \to \infty} p_{kl}^{(n)} = 0$ for some $k, l \in S$, then the chain does not have a stationary distribution.

**Lemma 3.3.** If the $x_n$s are non-negative, and $\sum_{n=1}^{\infty} x_n < \infty$, then $\lim_{n \to \infty} x_n = 0$.

**Corollary 3.3** (Transient Not Stationary Corollary)**.** A Markov chain which is irreducible and transient cannot have a stationary distribution.

*Proof.* If a chain is irreducible and transient, then by the Transience Equivalence Theorem, $\sum_{n=1}^{\infty} < \infty$ for all $i, j \in S$. Hence $\lim_{n \to \infty} p_{ij}^{(n)} = 0$ for all $i, j \in S$.
Thus by the Vanishing Probabilities Corollary, there is no stationary distribution.

∎

## 3.3  Obstacles to Convergence

**Definition 3.4** (period)**.** The period of a state $i$ is the greatest common divisor (gcd) of the set $\{n \geq 1 : p_{ii}^{(n)} > 0\}$, i.e. the largest number $m$ such that all the values of $n$ with $p_{ii}^{(n)} > 0$ are all integer multiples of $m$. If the period of each state is 1, we say the chain is aperiodic; otherwise we say the chain is periodic.

**Remark 3.3.** Intuitively, the period of a state $i$ is the pattern of returning to $i$ from $i$. e.g. If the period of $i$ is 2, then it is only possible to get from $i$ to $i$ in an even numbers of steps.

**Fact 3.1.** If state $i$ has period $t$, and $p_{ii}^{(m)} > 0$, then $m$ is an integer multiple of $t$, i.e., $t$ divides $m$.

**Fact 3.2.** If $p_{ii} > 0$, then the period of state $i$ is 1.

**Fact 3.3.** If $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$, then the period of state $i$ is 1.

**Lemma 3.4** (Equal Periods Lemma)**.** If $i \leftrightarrow j$, then the periods of $i$ and of $j$ are equal.

*Proof.* Let the periods of $i$ and $j$ be $t_i$ and $t_j$. Since $i \leftrightarrow j$, we can find $r, s \in \mathbb{N}$ with $p_{ij}^{(r)} > 0$ and $p_{ji}^{(s)} > 0$. Then

$$p_{ii}^{(r+s)} \geq p_{ij}^{(r)} p_{ji}^{(s)} > 0$$

Therefore by Fact 2.1, $t_i$ divides $r + s$.
Suppose now that $p_{jj}^{(n)} > 0$. Then

$$p_{ii}^{(r+n+s)} \geq p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0$$

So $t_i$ divides $r + n + s$.

Since $t_i$ divides both $r + n + s$ and $r + s$, then it must divide $n$ as well.

Since this is true for any $n$ with $p_{jj}^{(n)} > 0$, it follows that $t_i$ is a common divisor of $\{n \in \mathbb{N} : p_{jj}^{(n)} > 0\}$.

But $t_j$ is the greatest such common divisor, so $t_j \geq t_i$.

Similarly we can show that $t_i \geq t_j$, so we have $t_i = t_j$.

■

**Corollary 3.4** (Equal Periods Corollary)**.** If a chain is irreducible, then all states have the same period.

**Corollary 3.5.** If a chain is irreducible and $p_{ii} > 0$ for some state $i$, then the chain is aperiodic.

## 3.4   Convergence Theorem

**Theorem 3.1** (Markov Chain Convergence Theorem)**.** If a Markov chain is irreducible, aperiodic, and has a stationary distribution $\{\pi_i\}$, then $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$ for all $i, j \in S$, and $\lim_{n \to \infty} P(X_n = j) = \pi_j$ for any initial probabilities $\{v_i\}$.

**Theorem 3.2** (Stationary Recurrence Theorem)**.** If chain irreducible and has a stationary distribution, then it is recurrent.

*Proof.* The Transient Not Stationary Corollary says that a chain cannot be irreducible, transient and have a stationary distribution.

Therefore, if a chain is irreducible and has a stationary distribution, then it cannot be transient, i.e. it must be recurrent. ■

**Lemma 3.5** (Number Theory Lemma)**.** If a set $A$ of positive integers is non-empty, and satisfies additivity, and $gcd(A) = 1$, then there is some $n_0 \in \mathbb{N}$ s.t. for all $n \geq n_0$ we have $n \in A$ i.e. the set $A$ includes all of the integers $n_0, n_0 + 1, n_0 + 2, \ldots$

**Proposition 3.4.** If a state $i$ has $f_{ii} > 0$ and is aperiodic, then there is $n_0(i) \in \mathbb{N}$ such that $p_{ii}^{(n)} > 0$ for all $n \geq n_0(i)$

*Proof.* Let $A = \{n \geq 1 : p_{ii}^{(n)} > 0\}$. Since $f_{ii} > 0$, then $A$ is not empty.

If $m, n \in A$, then

$$p_{ii}^{(m+n)} \geq p_{ii}^{(m)} p_{ii}^{(n)} > 0$$

So $m + n \in A$, which shows that $A$ satisfies additivity. Also $gcd(A) = 1$ since the state $i$ is aperiodic. Hence from the Number Theory Lemma, there is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have $n \in A$ i.e. $p_{ii}^{(n)} > 0$.

■

**Corollary 3.6.** If a chain is irreducible and aperiodic, then for any states $i, j \in S$, there is $n_0(i, j) \in \mathbb{N}$ s.t. $p_{ij}^{(n)} > 0$ for all $n \geq n_0(i, j)$

*Proof.* Find $n_0(i)$ as in Proposition 2.3, and find $m \in \mathbb{N}$ with $p_{ij}^{(m)} > 0$.

Then let $n_0(i,j) = n_0(i) + m$

Then if $n \geq n_0(i,j)$, then $n - m \geq n_0(i)$, so $p_{ij}^{(n)} \geq p_{ii}^{(n-m)} p_{ij}^{(m)} > 0$. ∎

**Lemma 3.6** (Markov Forgetting Lemma)**.** If a Markov chain is irreducible and aperiodic, and has stationary distribution $\{\pi_i\}$, then for all $i, j, k \in S$,

$$\lim_{n \to \infty} \left| p_{ik}^{(n)} - p_{jk}^{(n)} \right| = 0$$

**Remark 3.4.** Intuitively, after a long time $n$, the chain "forgets" whether it started from state $i$ or from state $j$.

*Proof.* _____ ∎   `long`

**Proof of Markov Chain Convergence Theorem** _____   `long`

**Corollary 3.7.** If a chain is irreducible, then it has at most one stationary distribution.

*Proof.* By Markov Chain Convergence Theorem, any stationary distribution that ie has must be equal to $\lim_{n \to \infty} P(X_n = j)$, so it is unique. ∎

**Definition 3.5** (convergence in distribution)**.**

$$\forall a < b, \, P_{n \to \infty}(a < X_n < b) = P(a < X < b)$$

**Definition 3.6** (weak convergence)**.**

$$\forall \epsilon > 0, \, \lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0$$

**Remark 3.5.** This is "converge in probability".

**Definition 3.7** (strong convergence)**.**

$$P(\lim_{n \to \infty} X_n = X) = 1$$

**Remark 3.6.** This is "converge almost surely".

**Remark 3.7.** Strong convergence implies weak convergence, and weak convergence implies convergence in distribution.

**Proposition 3.5.** If $\{X_n\}$ is a simple symmetric random walk, then the absolute values $|X_n|$ converge weakly to positive infinity. _____   `prove this`

## 3.5   Periodic Convergence

**Theorem 3.3** (Periodic Convergence Theorem). Suppose a Markov chain is irreducible, with period $b \geq 2$, and stationary distribution $\{\pi_i\}$. Then for all $i, j \in S$,

$$\lim_{n \to \infty} \frac{1}{b}[p_{ij}^{(n)} + \ldots + p_{ij}^{(n+b-1)}] = \pi_j$$

and

$$\lim_{n \to \infty} \frac{1}{b}(P[X_n = j] + P[X_{n+1} = j] + \ldots + P[X_{n+b-1} = j]) = \pi_j$$

and also

$$\lim_{n \to \infty} \frac{1}{b}P(X_n = j \text{ or } X_{n+1} = j \text{ or } \ldots \text{ or } X_{n+b-1} = j) = \pi_j$$

**Theorem 3.4** (Average Probability Convergence). If a Markov chain is irreducible with stationary distribution $\{\pi_i\}$ (whether periodic or not), then

$$\forall i, j \in S, \lim_{n \to \infty} \frac{1}{n}[p_{ij}^{(1)} + p_{ij}^{(2)} + \ldots + p_{ij}^{(n)}] = \pi_j$$

i.e.,

$$\lim_{n \to \infty} \frac{1}{n}\sum_{l=1}^{n} p_{ij}^{(l)} = \pi_j$$

prove this

**Corollary 3.8** (Unique Stationary Corollary). If Markov chain $P$ is irreducible (whether periodic or not), then it has at most **one** stationary distribution.

## 3.6   Application - MCMC Algorithms

section missing...

## 3.7   Application - Random Walks on Graphs

Let $V$ be a non-empty finite or countable set. Let $w : V \times V \to [0, \infty)$ be a symmetric weight function so that $w(u, v) = w(v, u)$. (usual unweighted case: $w(u, v) = 1$ if there is an edge between $u$ and $v$, otherwise $w(u, v) = 0$).
Let $d(u) = \sum_{v \in V} w(u, v)$ be the degree of the vertex $u$. Assume that $d(u) > 0$ for all $u \in V$ (for example, by giving any isolated point a self-edge).

**Definition 3.8** ((simple) random walk on the (undirected) graph). Given a vertex set $V$ with symmetric weights $w$, the (simple) random walk on the (undirected) graph $(V, w)$ is the Markov chain with state space $S = V$ and transition probabilities $p_{uv} = \frac{w(u,v)}{d(u)}$ for all $u, v \in V$.

**Remark 3.8.** It follows that

$$\sum_{v \in V} p_{uv} = \frac{\sum_{v \in V} w(u, v)}{\sum_{v \in V} w(u, v)} = 1$$

**Remark 3.9.** The most common case is where each $w(u, v) = 0$ or 1, so from $u$, the chain moves to one of the $d(u)$ vertices connected to $u$ with equal probability.

**Theorem 3.5** (Graph Stationary Distribution)**.** Consider a random walk on a graph $V$ with degrees $d(u)$. Assume that $Z$ is finite. Then if $\pi_u = \frac{d(u)}{Z}$, then $\pi$ is a stationary distribution for this walk.

**Theorem 3.6** (Graph Convergence Theorem)**.** For a random walk on a connected non-bipartite graph, if $Z < \infty$, then $\lim\limits_{n \to \infty} p_{uv}^{(n)} = \frac{d(v)}{Z}$ for all $u, v \in V$, and $\lim\limits_{n \to \infty} P[X_n = v] = \frac{d(v)}{Z}$ (for any initial probabilities).

`prove this`

**Theorem 3.7** (Graph Average Convergence)**.** For a random walk on any connected graph with $Z < \infty$ (whether bipartite or not), for all $u, v \in V$,

$$\lim_{n \to \infty} \frac{1}{2} [p_{uv}^{(n)} + p_{uv}^{(n+1)}] = \frac{d(v)}{Z}$$

and

$$\lim_{n \to \infty} \frac{1}{n} \sum_{l=1}^{n} p_{uv}^{(l)} = \frac{d(v)}{Z}$$

`prove this`

## 3.8   Application - Gambler's Ruin

Consider the following gambling game:

Let $0 < a < c$ be integers, and let $0 < p < 1$. Suppose player $A$ starts with $a$ dollars, player $B$ starts with $c - a$ dollars, and they repeatedly bet. At each bet, $A$ wins \$1 from $B$ with probability $p$, or $B$ wins \$1 from $A$ with probability $1 - p$.

If $X_n$ is the amount of money that $A$ has at time $n$, then clearly $X_0 = a$, and $\{X_n\}$ follows a simple random walk.

Let $T_i = \inf\{n \geq 0 : X_n = i\}$ be the first time $A$ has $i$ dollars.

**The Gambler's Ruin question**   What is $P_a(T_c < T_0)$, i.e., what is the probability that $A$ reaches $c$ dollars before losing all their money?

*Answer:*   Define $s(a) := P_a(T_c < T_0)$, so that the probability we want to find is a function of the player's initial fortune $a$. Clearly $s(0) = 0$ and $s(c) = 1$.

For $1 \leq a \leq c - 1$, we have

$$
\begin{aligned}
s(a) &= P_a(T_c < T_0) \\
&= P_a(T_c < T_0, X_0 + 1) + P_a(T_c < T_0, X_1 = X_0 - 1) \\
&\qquad\qquad\qquad\qquad\qquad \text{(}A \text{ either wins or loses \$1 on the first bet)} \\
&= P(X_1 = X_0 + 1)P_a(T_c < T_0 | X_1 = X_0 + 1) + P(X_1 = X_0 - 1)P_a(T_c < T_0 | X_1 = X_0 - 1) \\
&= ps(a + 1) + (1 - p)s(a - 1)
\end{aligned}
$$

This gives $c - 1$ equations for the $c - 1$ unknowns, which can be solved by simple algebra:

$$
\begin{aligned}
ps(a) + (1 - p)s(a) &= ps(a + 1) + (1 - p)s(a - 1) \qquad\qquad \text{(re-arranging)} \\
\implies s(a + 1) - s(a) &= \frac{1 - p}{p}[s(a) - s(a - 1)]
\end{aligned}
$$

Suppose $s(1) = x$ for some $x \in \mathbb{R}$, then

$$s(1) - s(0) = x$$

$$s(2) - s(1) = \frac{1-p}{p}[s(1) - s(0)] = \frac{1-p}{p}x$$

$$s(3) - s(2) = \frac{1-p}{p}[s(2) - s(1)] = \left(\frac{1-p}{p}\right)^2 x$$

$$\implies s(a+1) - s(a) = \left(\frac{1-p}{p}\right)^a x \qquad \text{(for } 1 \le a \le c)$$

$$\implies s(a) = s(a) - s(0)$$

$$= [s(a) - s(a-1)] + [s(a-1) - s(a-2)] + \ldots + [s(1) - s(0)]$$

$$= \left[\left(\frac{1-p}{p}\right)^{a-1} + \left(\frac{1-p}{p}\right)^{a-2} + \ldots + \left(\frac{1-p}{p}\right) + 1\right]x$$

$$= \begin{cases} \left[\dfrac{(\frac{1-p}{p})^a - 1}{(\frac{1-p}{p}) - 1}\right]x, & p \ne \frac{1}{2} \\ ax, & p = \frac{1}{2} \end{cases}$$

Since $s(c) = 1$, we can solve for $x$:

$$x = \begin{cases} \dfrac{(\frac{1-p}{p}) - 1}{(\frac{1-p}{p})^c - 1}, & p \ne \frac{1}{2} \\ \dfrac{1}{c}, & p = \frac{1}{2} \end{cases}$$

We then obtain our final **Gambler's Ruin formula**:

$$s(a) = \begin{cases} \dfrac{(\frac{1-p}{p})^a - 1}{(\frac{1-p}{p})^c - 1}, & p \ne \frac{1}{2} \\ \dfrac{a}{c}, & p = \frac{1}{2} \end{cases}$$

**Remark 3.10.** We will sometimes write $s(a)$ as $s_{c,p}(a)$, to show the explicit dependence on $c$ and $p$.

**Example 3.1.** $c = 10,000, a = 9,700, p = 0.5$, then

$$s(a) = a/c = 0.97$$

**Example 3.2.** $c = 10,000, a = 9,700, p = 0.49$, then

$$s(a) \approx \frac{1}{163,000}$$

**Proposition 3.6 ().** Let $T = \min(T_0, T_c)$ be the time when the Gambler's Ruin game ends. Then $P(T > mc) \le (1 - p^c)^m$ where $m \in \mathbb{Z}^+$ and $P(T = \infty) = 0$, and $\mathbb{E}[T] < \infty$.

*Proof.* (1) If the player ever wins $c$ bets in a row, then the game must be over.
Then if $T > mc$, then the player has failed to win $c$ bets in a row, despite having $m$ independent attempts to do so.

But the probability of winning $c$ bets in a row is $p^c$. So the probability of failing to win $c$ bets in a row is $1 - p^c$. Therefore the probability of failing on $m$ independent attempts is $(1 - p^c)^m$, as claimed.

(2) Then by continuity of probabilities,

$$P(T = \infty) = \lim_{m \to \infty} P(T > mc) \leq \lim_{m \to \infty} (1 - p^c)^m = 0$$

(3) We have

$$
\begin{aligned}
E(T) &= \sum_{i=1}^{\infty} P(T \geq i) \\
&\leq \sum_{i=0}^{\infty} P(T \geq i) \\
&= P(T \geq 0) + P(T \geq 1) + P(T \geq 2) + P(T \geq 3) + P(T \geq 4) + \ldots \\
&\leq P(T \geq 0) + P(T \geq 0) + \ldots + P(T \geq 0) + P(T \geq c) + P(T \geq c) + \ldots \\
&= \sum_{j=0}^{\infty} cP(T \geq cj) \\
&\leq \sum_{j=0}^{\infty} c(1 - p^c)^j \\
&= \frac{c}{1 - (1 - p^c)} \\
&= \frac{c}{p^c} < \infty
\end{aligned}
$$

∎

**Remark 3.11.** This says that, with probability 1 the Gambler's Ruin game must eventually end, and the time it takes to end has finite expected value.

## 3.9   Mean Recurrence Times

**Definition 3.9** (mean recurrence time). The <u>mean recurrence time</u> of a state $i$ is

$$m_i = E_i(\inf\{n \geq 1 : X_n = i\}) = E_i(\tau_i)$$

where $\tau_i = \inf\{n \geq 1 : X_n = i\}$

**Remark 3.12.** That is, $m_i$ is the expected value of the time to return from $i$ back to $i$.

**Definition 3.10** (positive recurrence and null recurrence). A state is <u>positive recurrent</u> if $m_i < \infty$. It is <u>null recurrent</u> if it is recurrent but $m_i = \infty$.

**Theorem 3.8** (Recurrence Time Theorem). For an irreducible Markov chain, either

1. $m_i < \infty$ for all $i \in S$, and there is a unique stationary distribution given by $\pi_i = 1/m_i$; or

2. $m_i = \infty$ for all $i \in S$, and there is no stationary distribution.

**Proposition 3.7.** An irreducible Markov chain on a finite state space $S$ always falls into case $(i)$ above:
$m_i < \infty$ for all $i \in S$, and there is a unique stationary distribution given by $\pi_i = 1/m_i$.

**Remark 3.13.** The converse is false: There could be an example that has infinite state space $S = \mathbb{N}$, but still has a stationary distribution, so it falls into case $(i)$.

## 3.10   Application - Sequence Waiting Times

**Problem**   Suppose we repeatedly flip a fair coin and get Heads(H) or Tails(T) independently each time with probability $1/2$ each. Let $\tau$ be the first time the sequence $HTH$ is completed. What is $E[\tau]$?

To find $E[\tau]$, we can use Markov chains.

Let $X_n$ be the partial amount of the desired sequence $(HTH)$ that the chain has "achieved so far" after $n$ flips. Then we always have $X_\tau = 3$, since we "win" upon reaching state 3. Assume we "start over" right after we win ($X_{\tau+1} = 1$ if flip $(\tau + 1)$ is Heads, otherwise $X_{\tau+1} = 0$). Also, we take $X_0 = 0$, i.e., at the beginning we have not achieved any of the sequence.

Here, $\{X_n\}$ is a Markov chain with state space $S = \{0, 1, 2, 3\}$ and $P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \end{pmatrix}$.

The mean waiting time of $HTH$ is thus equal to the mean recurrence time of state 3.

Using the equation $\pi P = \pi$, it can be computed that the stationary distribution is $(0.3, 0.4, 0.2, 0.1)$. Therefore, by the Recurrence Time Theorem, the mean time to return from state 3 to state 3 (has the same probability as going from state 0 to state 3) is $1/\pi_3 = 10$.

# 4   Martingales

Roughly speaking, martingales are stochastic processes which "stays the same on average".

## 4.1   Martingale Definitions

For a formal definition, let $\{X_n\}_{n=0}^\infty$ be a sequence of random variables. We assume throughout that random variables $X_n$ have **finite expectation** (or are **integrable**): $E|X_n| < \infty \quad \forall n$.

**Definition 4.1** (Martingale)**.** A sequence $\{X_n\}_{n=0}^\infty$ is a <u>martingale</u> if for all $n$,

$$E(X_{n+1}|X_0, \dots, X_n) = X_n$$

**Remark 4.1.** No matter what has happened so far, the average of the next value will be equal to the most recent one.

**Special case: Markov chain**   If the sequence $\{X_n\}$ is a Markov chain, then we have

$$
\begin{aligned}
E[X_{n+1}|X_0 = i_0, \ldots, X_n = i_n] &= \sum_{j \in S} jP[X_{n+1}|X_0 = i_0, \ldots, X_n = i_n] \\
&= \sum_j jP[X_{n+1}|X_n = i_n] \\
&= \sum_j jp_{i_n, j}
\end{aligned}
$$

To be a martingale, this value must equal $i_n$. That is, a Markov chain (with $E|X_n| < \infty$) is a martingale if

$$
\sum_{j \in S} jp_{ij} = i
$$

for all $i \in S$.

**Example 4.1** (simple symmetric random walk)**.** Let $\{X_n\}$ be s.s.r.w. with $p = 1/2$. We always have $|X_n| \leq n$, so $E|X_n| \leq n < \infty$, so there is no problem with finite expectations. For all $i \in S$, we compute that $\sum_{j \in S} jp_{ij} = (i+1)(1/2) + (i-1)(1/2) = i$, so s.s.r.w. is indeed a martingale.

**Proposition 4.1.** If $\{X_n\}$ is a martingale, then by the Law of Total Expectation,

$$
E(X_{n+1}) = E[E(X_{n+1}|X_0, X_1, \ldots, X_n)] = E(X_n)
$$

$$
\implies E(X_n) = E(X_0) \quad \forall n
$$

This is not surprising, since martingales stay the same on average. However, this is not a sufficient condition for $\{X_n\}$ to be a martingale.

## 4.2   Stopping Times

We often want to consider $E(X_T)$ for a random time $T$. We need to prevent the random time $T$ from looking into the future of the process, before deciding whether to stop.

**Definition 4.2** (stopping time)**.** A non-negative-integer-valued random variable $T$ is a stopping time for $\{X_n\}$ if the event $\{T = n\}$ is determined by $X_0, X_1, \ldots, X_n$, i.e. if the indicator function $\mathbb{1}\{T = n\}$ is a function of $X_0, X_1, \ldots, X_n$.

**Remark 4.2.** Intuitively, this definition says that a stopping time $T$ must decide whether to stop at time $n$ based solely on what has happened up to time $n$, without first looking into the future.

**Example 4.2.** valid stopping times:
$T = 5, T = \inf\{n \geq 0 : X_n = 5\}, T = \inf\{n \geq 0 : X_n = 0 \vee X_n = c\}, T = \inf\{n \geq 2 : X_{n-2} = 5\}$ not valid stopping time: $T = \inf\{n \geq 0 : X_{n+1} = 5\}$ (since it looks into the future)

**Lemma 4.1** (Optional Stopping Lemma)**.** If $\{X_n\}$ is a martingale, and $T$ is a stopping time which is bounded (i.e., $\exists M < \infty$ with $P(T \leq M) = 1$), then

$$
E(X_T) = E(X_0)
$$

prove this!

**Example 4.3.** Consider s.s.r.w. with $X_0 = 0$, and let

$$T = \min\{10^{12}, \inf\{n \geq 0 : X_n = -5\}$$

Then $T$ is a bounded stopping time. Hence by the Optional Stopping Lemma,

$$E(X_T) = E(X_0) = E(0) = 0$$

But near always, we will have $X_T = -5$.
By the Law of Total Expectation,

$$
\begin{aligned}
0 &= E(X_T) \\
&= \underbrace{P(X_T = -5)}_{\approx 1} \underbrace{E(X_T | X_T = -5)}_{=-5} + \underbrace{P(X_T \neq -5)}_{\approx 0} \underbrace{E(X_T | X_T \neq -5)}_{huge}
\end{aligned}
$$

**Theorem 4.1** (Optional Stopping Theorem)**.** If $\{X_n\}$ is a martingale with stopping time $T$, and $P(T < \infty) = 1$, and $E|X_T| < \infty$, and $\lim_{n \to \infty} E(X_n \mathbb{1}\{T > n\}) = 0$, then

$$E(X_T) = E(X_0)$$

*Proof.* For each $m \in \mathbb{N}$, let $S_m = \min\{T, m\}$, so that $S_m$ is a bounded stopping time. Then by Optional Stopping Lemma, $E(X_{S_m}) = E(X_0)$ (for any $m$). Then for any $m$,

$$
\begin{aligned}
X_{S_m} &= X_{\min(T,m)} \\
&= X_T \mathbb{1}\{T \leq m\} + X_m \mathbb{1}\{T > m\} \\
&= X_T(1 - \mathbb{1}\{T > m\}) + X_m \mathbb{1}\{T > m\} \\
&= X_T - X_T \mathbb{1}\{T > m\} + X_m \mathbb{1}\{T > m\} \\
\implies X_T &= X_{S_m} + X_T \mathbb{1}\{T > m\} - X_m \mathbb{1}\{T > m\} \\
\implies E(X_T) &= E(X_{S_m}) + E(X_T \mathbb{1}\{T > m\}) - E(X_m \mathbb{1}\{T > m\}) \\
&= E(X_0) + E(X_T \mathbb{1}\{T > m\}) - E(X_m \mathbb{1}\{T > m\})
\end{aligned}
$$

Take $m \to \infty$. Since $P(T < \infty) = 1$, we have $\mathbb{1}\{T > m\} \to 0$.
Since $E|X_T| < \infty$ and $\mathbb{1}\{T > m\} \to 0$, we have

$$\lim_{m \to \infty} E(X_T \mathbb{1}\{T > m\}) = 0$$

by the Dominated Convergence Theorem 1.1
Also, $\lim_{m \to \infty} E(X_m \mathbb{1}\{T > m\}) = 0$ by assumption.
Hence $E(X_T) \to E(X_0)$, i.e. $E(X_T) = E(X_0)$.                                      ∎

**Corollary 4.1** (Optional Stopping Corollary)**.** If $\{X_n\}$ is a martingale with stopping time $T$, which is "bounded up to time $T$" (i.e., $\exists M < \infty$ with $P(|X_n|\mathbb{1}\{n \leq T\} \leq M) = 1$ for all $n$), and $P(T < \infty) = 1$, then

$$E(X_T) = E(X_0)$$

*Proof.* It follows that, $P(|X_T| \leq M) = 1$.

Hence, $E|X_T| \leq M < \infty$.

Also,

$$
\begin{aligned}
|E(X_n \mathbb{1}\{T > n\})| &\leq E(|X_n|\mathbb{1}\{T > n\}) \\
&= E(|X_n|\mathbb{1}\{n \leq T\}\mathbb{1}\{T > n\}) \\
&\leq E(M\mathbb{1}\{T > n\}) \\
&= MP(T > n) \to 0 \qquad\qquad \text{(Since } P(T < \infty) = 1\text{)}
\end{aligned}
$$

Hence the result follows from the Optional Stopping Theorem. ∎

**Example 4.4** (Gambler's Ruin problem - $p = 1/2$). Let $T = \inf\{n \geq 0 : X_n \vee X_n = c\}$ be the time when the game ends. Then $P(T < \infty) = 1$ by Proposition 3.6. Also, if the game has not yet ended, i.e. $n \leq T$, then $X_n$ must be between 0 and $c$. Hence $|X_n|\mathbb{1}\{n \leq T\} \leq c < \infty$ for all $n \leq T$.

So by the Optional Stopping Corollary 4.1, $E(X_T) = cs(a) + 0(1 - s(a) = E(X_0) = a \implies s(a) = a/c$.

**Example 4.5** (Gambler's Ruin problem - $p \neq 1/2$). Then $\{X_n\}$ is not a martingale since

$$
\sum_j jp_{ij} = p(i + 1) + (1 - p)(i - 1) = i + 2p - 1 \neq i
$$

Instead we use a trick: Let $Y_n := \left(\frac{1-p}{p}\right)^{X_n}$, then $\{Y_n\}$ is also a Markov chain, and

$$
\begin{aligned}
E(Y_{n+1}|Y_0, Y_1, \ldots, Y_n) &= p\left(\frac{1-p}{p}\right)^{X_n+1} + (1 - p)\left(\frac{1-p}{p}\right)^{X_n-1} \\
&= p\left[Y_n\left(\frac{1-p}{p}\right)\right] + (1 - p)\left[Y_n/\left(\frac{1-p}{p}\right)\right] \\
&= Y_n(1 - p) + Y_n(p) \\
&= Y_n
\end{aligned}
$$

So $\{Y_n\}$ is a martingale.

Again, $P(T < \infty) = 1$ by Proposition 3.6.

Also, $|Y_n|\mathbb{1}\{n \leq T\} \leq \max\left(\left(\frac{1-p}{p}\right)^0, \left(\frac{1-p}{p}\right)^c\right) := M < \infty$ for all $n$. Hence by the Optional Stopping Corollary 4.1,

$$
E(Y_T) = s(a)\left(\frac{1-p}{p}\right)^c + [1 - s(a)](1) = E(Y_0) = \left(\frac{1-p}{p}\right)^a
$$

$$
\implies s(a) = \frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1}
$$

## 4.3 Wald's Theorem

Suppose $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ are i.i.d. with finite mean $m$. Let $T$ be a stopping time for $\{X_n\}$ which has finite mean, i.e. $E(T) < \infty$. Then

$$E(X_T) = a + mE(T)$$

**Property 4.1** (Special case: $m = 0$). Then $\{X_n\}$ is a martingale, and Optional Stopping Theorem 4.1 says that $E(X_T) = a = E(X_0)$.

prove this!

**Corollary 4.2.** If $\{X_n\}$ is Gambler's Ruin with $p \neq 1/2$, and $T = \inf\{n \geq 0 : X_n = 0 \vee X_n = c\}$, then

$$E(T) = \frac{1}{2p-1}\left(c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} - a\right)$$

*Proof.* We again apply Wald's Theorem:
Here $Z_i = +1$ if you win the $i$th bet, otherwise $Z_i = -1$. So

$$m = E(Z_i) = p(1) + (1-p)(-1) = 2p - 1$$

Also, $E(T) < \infty$ by Proposition 3.6. Then by Wald's Theorem,

$$\begin{aligned}
E(X_T) &= a + mE(T) \\
&= cs(a) + 0(1 - s(a)) \\
&= c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} \\
\implies E(T) &= \frac{1}{m}(E(X_T) - a) \\
&= \frac{1}{2p-1}\left(c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} - a\right)
\end{aligned}$$

∎

**Lemma 4.2.** Let $X_n = a + Z_1 + \ldots + Z_n$, where $\{Z_i\}$ are i.i.d. with mean 0 and variance $v < \infty$. Let $Y_n = (X_n - a)^2 - nv = (Z_1 + \ldots + Z_n)^2 - nv$. Then $\{Y_n\}$ is a martingale.

prove this!

**Corollary 4.3.** If $\{X_n\}$ is Gambler's Ruin with $p = 1/2$, and $T = \inf\{n \geq 0 : X_n = 0 \vee X_n = c\}$, then

$$E(T) = Var(X_T) = a(c - a)$$

prove this!

## 4.4 Application - Sequence Waiting Times

Suppose at each time $n$, a new "player" appears, and bets \$1 on heads, then if they win they bet \$2 on tails, then if they win again they bet \$4 on heads. Each player stops betting as soon as they either lose once (and hence are down a total of \$1), or win three bets in a row (and hence are up a total of \$7.

Let $X_n$ be the total amount won by all the betters by time $n$. Then since the bets were fair, $\{X_n\}$ is a martingale with stopping time $\tau$.

## 4.5 Martingale Convergence Theorem

Suppose $\{X_n\}$ is a martingale. Then $\{X_n\}$ could have infinite fluctuations in both directions, as we have seen for s.s.r.w.; Or $\{X_n\}$ could converge with probability 1 to a fixed (perhaps random) value.

**Example 4.6.** Let $\{X_n\}$ be Gambler's Ruin with $p = 1/2$, where we stop as soon as we either win or lose. Then $X_n \to X$ with probability 1, where $P(X = c) = a/c$ and $P(X = 0) = 1 - a/c$.

**Example 4.7.** Let $\{X_n\}$ be a Markov chain on $S = \{2^m : m \in \mathbb{Z}\}$, with $X_0 = 1$, and $p_{i,2i} = 1/2$ and $p_{i,i/2} = 2/3$ for $i \in S$. This is a martingale, since $\sum_j j p_{ij} = (2i)(1/3) + (i/2)(2/3) = i$. Let $Y_n = \log_2 X_n$. Then $Y_0 = 0$, and $\{Y_n\}$ is s.r.w. with $p = 1/3$, $Y_n \to -\infty$ w.p. 1 by the Law of Large Numbers 1.2. Hence, $X_n = 2^{Y_n} \to 2^{-\infty} = 0$ w.p. 1.

**Theorem 4.2** (Martingale Convergence Theorem). Any non-negative martingale $\{X_n\}$ ($X_n \geq 0$) which is bounded below (i.e. $X_n \geq c$ for all $n$, for some finite number $c$), or is bounded above (i.e. $X_n \leq c$ for all $n$, for some finite number $c$), converges w.p. 1 to some random variable $X$.

**Remark 4.3.** The intuition behind this theorem is:

1. Since the martingale is bounded on one side, it cannot "spread out" forever.

2. Since it is a martingale, it cannot "drift" in a positive or negative direction.

3. So it has somewhere to go, and eventually has to stop somewhere.

**Remark 4.4.** If $\{X_n\}$ is not non-negative, then if $X_n \geq c$, then $\{X_n - c\}$ is a non-negative martingale, or if $X_n \leq c$, then $\{-X_n + c\}$ is a non-negative martingale, and in either case the non-negative martingale converges iff $\{X_n\}$ converges.

## 4.6 Application - Branching Processes

**Definition 4.3** (offspring distribution). Let $\mu$ be any prob dist on $\{0, 1, 2, \ldots\}$, the offspring distribution. Let $X_n$ be the number of individuals at time $n$. Start with $X_0 = a$ individuals. Assume $0 < a < \infty$, $Z_{n,i} \overset{i.i.d.}{\sim} \mu(i)$.

(i.e., Each of the $X_n$ individuals at time $n$ has a random number of offspring under the distribution $\mu$). Then

$$X_{n+1} = Z_{n,1} + Z_{n,2} + \ldots + Z_{n,X_n}$$

Here $\{X_n\}$ is a Markov chain, on the state space $\{0, 1, 2, \ldots\}$.

**Transition probabilties**   If $X_n$ ever reaches 0, then it stays there forever: $p_{0j} = 0 \quad \forall j \geq 0$.
This is called <u>extinction</u>.
Also, $p_{ij} = (\mu * \mu * \ldots * \mu)(j)$, a <u>convolution</u> of $i$ copies of $\mu$.

**Theorem 4.3.** Let $m = \sum_i \mu(i)$ be the mean of $\mu$, which is called the <u>reproductive number</u>.
If $m < 1$, then $E(X_n) \to 0$, and $P(X_n = 0) \to 1$.

*Proof.* Assume $0 < m < \infty$. Then

$$E(X_{n+1}|X_0,\ldots,X_n) = E(Z_{n,1} + Z_{n,2} + \ldots + Z_{n,X_n}|X_0,\ldots,X_n) = mX_n$$

$$\implies E(X_n) = m^n E(X_0) = m^n a < \infty$$

So if $m < 1$, then $E(X_n) = am^n \to 0$. Then we have

$$E(X_n) = \sum_{k=0}^{\infty} kP(X_n = k)$$
$$\geq \sum_{k=1}^{\infty} P(X_n = k)$$
$$= P(X_n \geq 1)$$
$$\implies P(X_n \geq 1) \leq E(X_n) = am^n \to 0$$
$$\implies P(X_n = 0) \to 1$$

∎

**Fact 4.1.** Let $m = \sum_i \mu(i)$ be the mean of $\mu$, which is called the <u>reproductive number</u>. If $m > 1$, then $E(X_n) \to \infty$, $P(X_n \to \infty) > 0$ and $P(X_n \to 0) > 0$, i.e., we have possible extinction but also possible flourishing.

**Theorem 4.4.** Let $m = \sum_i \mu(i)$ be the mean of $\mu$, which is called the <u>reproductive number</u>. If $m = 1$, and $\mu$ is non-degenerate (i.e. $\mu(1) < 1$, so that $\mu$ is not a constant), then $\{X_n\} \to 0$ w.p. 1.

*Proof.* If $m = 1$, then $E(X_n) = E(X_0) = a$ for all $n$. Then $E(X_{n+1}|X_0,\ldots,X_n) = mX_n - X_n$, so $\{X_n\}$ is a non-negative martingale.
Hence by the Martingale Convergence Theorem 4.2, we must have $X_n \to X$ for some random variable $X$. This could only happen if

1. $\mu(1) = 1$; or

2. $X = 0$

∎

## 4.7   Application - Stock Options (Discrete)

In mathematical finance, it is common to model the price of one share of some stock as a random process.

For now, we work in discrete time, and suppose that $X_n$ is the price of one share of the stock at each date $n$. If you buy the stock, then the situation is clear: if $X_n$ increases then you will make a profit, but if $X_n$ decreases then you will suffer a loss.

**Definition 4.4** (stock option). A stock option is the option to buy one share of the stock for some fixed strike price $K$ at some fixed future strike date (time) $S > 0$. If at the strike time $S$, the stock price $X_s$ is less than the strike price $K$, then the option would not be exercised, and would thus be worth exactly zero. If the stock price $X_S$ is more than $K$, then the option would be exercised to obtain a stock worth $X_S$ for a price of just $K$, for a net profit of $X_S - K$. Hence at time $S$, the stock option is worth $\max(0, X_S - K)$.

**Remark 4.5.** At time 0, $X_S$ is an unknown quantity. The fair price of a stock option is defined to be the no-arbitrage price, i.e., the price for the option which makes it impossible to make a guaranteed profit through any combination of buying or selling the option, and buying and selling the stock. At time 0, what is the fair price (no-arbitrage price of the stock option?

**Example 4.8** (naive example). Suppose that at time 0, you buy $x$ stock shares (for \$100 each), and $y$ option shares (for \$$c$ each) where $x, y \in \mathbb{R}$ (negative values indicates selling).

Then if the stock goes up to \$130, you make \$30 on each stock share and \$$(20 - c)$ on each option share for a total profit of $30x + (20 - c)y$.

But if the stock goes down to \$80, you lose \$20 on each stock share and \$$c$ on each option share for a total profit of $-20x - cy$.

To attempt to make a guaranteed profit, we could make these two different total profit amounts equal to each other $\implies y = (-5/2)x$, profits $= (5/2)(c - 8)x$.

If $c > 8$, then you buy $x > 0$ stock shares and $y = (-5/2)x < 0$ option shares and make a guaranteed profit of $(5/2)(c - 8)x > 0$.

If $c < 8$, then you buy $x < 0$ stock shares and $y = (-5/2)x < 0$ option shares and make a guaranteed profit of $(5/2)(c - 8)(-x) > 0$.

But if $c = 8$, then profits $= 0$.

In summary, there is no arbitrage iff $c = 8$.

**Example 4.9.** Suppose we assign the new probabilities $P(X_s = 80) = 3/5$ and $P(X_S = 130) = 2/5$. Then the stock price is a martingale since $E(X_S) = (3/5)80 + (2/5)(130) = 100 =$ initial price. The option price is a martingale since $option_value = (3/5)0 + (2/5)(130 - 110) = 8 = c =$ initial price.

Then the fair price is the martingale expected value, 8.

**Theorem 4.5** (Martingale Pricing Principle). The fair price of an option is equal to its expected value under the martingale probabilities.

**Proposition 4.2.** Suppose a stock price at time 0 equals $X_0 = a$, and at time $S > 0$ equals either $X_s = d$ (down) or $X_s = u$ (up), where $d < a < u$. Then if $d < K < u$ then at time 0, the fair (no-arbitrage) price of an option to buy the stock at time $S$ for $K$ is equal to $(a - d)(u - K)/(u - d)$.

prove this!