STA447
Lecture Notes

Yuchen Wang

April 2, 2020

Contents

All the copyrights to Professor Jeffery Rosenthal.

1 Preliminary

Proposition 1.1. If Z is a non-negative-integer-valued random variable, then

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

Proof.

$$
\begin{aligned}
\sum_{k=1}^{\infty} P(Z \geq k) &= \sum_{k=1}^{\infty}[P(Z = k) + P(Z = k+1) + \ldots] \\
&= [P(Z = 1) + P(Z = 2) + P(Z = 3) + \ldots] \\
&\quad + [P(Z = 2) + P(Z = 3) + P(Z = 4) + \ldots] \\
&\quad + [P(Z = 3) + P(Z = 4) + P(Z = 5) + \ldots] \\
&\quad + \ldots \\
&= P(Z = 1) + 2P(Z = 2) + 3P(Z = 3) + \ldots \\
&= \sum_{l=1}^{\infty} lP(Z = l) \\
&= E(Z)
\end{aligned}
$$

∎

Fact 1.1.

$$\sum_{n=1}^{\infty} \frac{1}{n^p} = \infty \iff p \leq 1$$

Fact 1.2. If the x_ns are non-negative, and $\sum_{n=1}^{\infty} x_n < \infty$, then

$$\lim_{n \to \infty} x_n = 0$$

Definition 1.1 (bounded random variable). X is a <u>bounded random variable</u> if there is $M < \infty$ with $P(|X| \leq M) = 1$, i.e. if it is always in some interval $[-M, M]$ for some finite number M.

Definition 1.2 (finite random variable). X is a <u>finite random variable</u> if $P(|X| \leq \infty) = 1$, i.e., if $P(|X| = \infty) = 0$, i.e. if it always takes on finite values.

Definition 1.3 (finite expectation). A random variable X has <u>finite expectation</u> if $E|X| < \infty$; this is also sometimes called integrable.

Fact 1.3. Bounded $\implies$ finite expectation.

Fact 1.4. Unbounded $\implies$ infinite expectation.

Fact 1.5. Finite expectation $\implies$ finite.

Theorem 1.1 (Law of Total Expectation)**.** If X and Y are discrete random variables, then

$$E(X) = \sum_y P(Y = y)E(X|Y = y)$$

i.e. we can compute $E(X)$ by averaging conditional expectations.

Theorem 1.2 (Double-expectation formula)**.**

$$E[E(X|Y)] = E(X)$$

i.e. the random variable $E(X|Y)$ equals X on average.

Proof. Since $E(X|Y)$ is equal to $E(X|Y = y)$ with probability $Y = y$, we compute that

$$E[E(X|Y)] = \sum_y P(Y = y)E(X|Y = y) = E(X)$$

which the results follows from Double-expectation formula 1.2. ∎

Definition 1.4 (weak convergence)**.** X_n converge to X weakly if

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

Definition 1.5 (strong convergence)**.** X_n converge to X strongly if

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

Theorem 1.3 (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.,

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

Theorem 1.4 (the Limit of Partial Averages Principle)**.** If

$$\lim_{n\to\infty} x_n = r$$

then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n x_i = r$$

Theorem 1.5 (Bounded Convergence Theorem)**.** If $\lim_{n\to\infty} X_n = X$, and the $\{X_n\}$ are uniformly bounded (i.e., there is $M < \infty$ with $|X_n| \leq M$ for all n), then

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

Theorem 1.6 (Monotone Convergence Theorem)**.** If $\lim_{n\to\infty} X_n = X$, and $0 \leq X_1 \leq X_2 \leq X_3 \leq \ldots$, then

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

Theorem 1.7 (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 1.6 (order notation). We say that a function $f = O(g)$ if

$$\lim_{h\searrow 0} \frac{f}{g} < \infty$$

We say that a function $f = o(g)$ if

$$\lim_{h\searrow 0} \frac{f}{g} = 0$$

Example 1.1. From a first-order Taylor series expansion,

$$e^x = 1 + x + o(x^2)$$

as $x \to 0$.

If time, please finish reading the preliminary, which I found useful. −Mar 9

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 simple symmetric random walk.

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.

$$1 - f_{ij} = P_i(X_n \neq j \text{ for all } n \geq 1) \tag{2.1}$$

$$P_i(N(i) \geq k) = (f_{ii})^k \tag{2.2}$$

$$P_i(N(j) \geq k) = f_{ij}(f_{jj})^{k-1} \tag{2.3}$$

$$f_{ik} \geq f_{ij} f_{jk} \tag{2.4}$$

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.

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

$\blacksquare$

Therefore,

$$
\begin{aligned}
\sum_{n=1}^{\infty} p_{ii}^{(n)} &= \sum_{n=1}^{\infty} P_i(X_n = i) \\
&= \sum_{n=1}^{\infty} E_i(\mathbb{1}\{X_n = i\}) \\
&= E_i\Big(\sum_{n=1}^{\infty} \mathbb{1}\{X_n = i\}\Big) \\
&= E_i(N(i)) \\
&= \sum_{k=1}^{\infty} P_i(N(i) \geq k) \qquad\qquad \text{(by proposition 1.1)} \\
&= \sum_{k=1}^{\infty} (f_{ii})^k \\
&= \begin{cases} \infty, & f_{ii} = 1 \\ \frac{f_{ii}}{1-f_{ii}} < \infty, & f_{ii} < 1 \end{cases}
\end{aligned}
$$

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

$$
\begin{aligned}
p_{00}^{(n)} &= \binom{n}{n/2} p^{n/2} (1-p)^{n/2} \\
&= \frac{n!}{[(n/2)!]^2} p^{n/2} (1-p)^{n/2} \\
&= \frac{(n/e)^n \sqrt{2\pi n}}{[(n/2e)^{n/2} \sqrt{2\pi n/2}]^2} p^{n/2} (1-p)^{n/2} \qquad \text{(Sirling's approximation)} \\
&= [4p(1-p)]^{n/2} \sqrt{2/\pi n}
\end{aligned}
$$

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

$$
\begin{aligned}
\sum_{n=1} \infty p_{00}^{(n)} &= \sum_{n=2,4,6,\dots} \sqrt{2/\pi n} \\
&= \sqrt{2/\pi} \sum_{n=2,4,6,\dots} n^{-1/2} \\
&= \sqrt{2/\pi} \sum_{n=1}^{\infty} 2k^{-1/2} \\
&= \infty
\end{aligned}
$$

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

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

∎

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

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

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 $l = k$).

$\blacksquare$

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_ns 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 ∞, the chain must pass through j, so $f_{ij} = 1$. $\blacksquare$

3 Markov Chain Convergence

3.1 Stationary Distributions

Definition 3.1 (stationary distributions)**.** If π 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 π is stationary 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, π 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 doubly stochastic 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

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

Then $\{\pi_i\}$ is stationary.   ∎

3.2 Searching for Stationary

Definition 3.3 (reversibility)**.** A Markov chain is reversible (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 π, then π 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 $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0$ 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 π. 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 $\lim\limits_{n\to\infty} p_{kl}^{(n)} = 0$, then for any $i, j \in S$ with $k \to i$ and $j \to l$, $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0$.

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_ns 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 <u>period</u> 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 <u>aperiodic</u>; otherwise we say the chain is <u>periodic</u>.

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:

$$\forall K > 0, \; \lim_{n \to \infty} P(|X_n| < L) = 0$$

Proof. We can write that $X_n = \sum_{i=1}^{n} Z_i$, where $\{Z_i\}$ are i.i.d. ± 1 with probability $1/2$ each, and hence common mean $m = 0$ and variance $v = 1$. Hence, the Central Limit Theorem says that

$$\lim_{n \to \infty} P(-b < \frac{X_n}{\sqrt{n}} < b) = \int_{-b}^{b} \phi(x)\, dx$$

where ϕ is the density function of the standard normal distribution.

Hence, $\lim_{b \searrow 0} P(-b < \frac{X_n}{\sqrt{n}} < b) = 0$.

But $P(|X_n| < K) = P(-K < X_n < K) = P(\frac{-K}{\sqrt{n}} < \frac{X_n}{\sqrt{n}} < \frac{K}{\sqrt{n}})$.

By monotonicity, for any finite K and any $b > 0$,

$$P(|X_n| < K) < P(-b < \frac{X_n}{\sqrt{n}} < b)$$

for all sufficiently large n.

It follows that $\lim_{n \to \infty} P(|X_n| < K) = 0$ for any finite K.

Therefore, $|X_n|$ converges weakly to positive infinity. ∎

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

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

Example 3.1. If $b = 2$, then $\lim_{n \to \infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j$ for all $i, j \in S$.

Example 3.2 (Ehrenfest's Urn). Ehrenfest's Urn has period $b = 2$, and oscillates between the two subsets $S_0 = \{\text{even } i \in S\}$ and $S_1 = \{\text{odd } i \in S\}$. It satisfies the periodic convergence property that

$$\lim_{n \to \infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j = 2^{-d}\binom{d}{j}$$

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.,

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

Proof. This follows from the usual Markov Chain Convergence Theorem by the Limit of Partial Averages Principle 1.4. ∎

Remark 3.8. This is a type of convergence that holds for all irreducible chains with stationary distributions, whether periodic or not.

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

Proof. By Average Probability Convergence, any π_j in a stationary distribution $\{\pi_i\}$ which exists must be equal to

$$\lim_{n \to \infty} \frac{1}{n}[p_{ij}^{(1)} + p_{ij}^{(2)} + \ldots + p_{ij}^{(n)}]$$

So if there exists two stationary distributions, then they are equal. ∎

3.6 Application - Markov Chain Monte Carlo Algorithms

' e.g. $S = \{1, 2, 3\}$, or $S = \{-5, -4, \ldots, 17\}$, or $S = \mathbb{N}$.

Let $\{\pi_i\}$ be any probability distribution on S. Assume for simplicity that $\pi_i > 0$ for all $i \in S$. Suppose we want to sample from π, i.e., create a random variable X with $P(X = i) \approx \pi_i$ for all $i \in S$.

Metropolis Algorithm   Let

$$p_{i,i+1} = \frac{1}{2} \min(1, \frac{\pi_{i+1}}{\pi_i})$$

$$p_{i,i-1} = \frac{1}{2} \min(1, \frac{\pi_{i-1}}{\pi_i})$$

and

$$p_{i,i} = 1 - p_{i,i+1} - p_{i,i-1}$$

Fact 3.4. This chain have $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$.

Proof.

$$\pi_i p_{i,i+1} = \pi_i \frac{1}{2} \min(1, \frac{\pi_{i+1}}{\pi_i}) = \frac{1}{2} \min(\pi_i, \pi_{i+1})$$

$$\pi_{i+1} p_{i+1,i} = \pi_{i+1} \frac{1}{2} \min(1, \frac{\pi_i}{\pi_{i+1}}) = \frac{1}{2} \min(\pi_{i+1}, \pi_i)$$

This implies that $\pi_i p_{ij} = \pi_j p_{ji}$ if $j = i + 1$, hence for all $i, j \in S$, since $p_{ij} = 0$ for all other cases.

Therefore, the chain is reversible w.r.t. $\{\pi_i\}$. So $\{\pi_i\}$ is stationary.

Also, the chain is easily checked to be irreducible and aperiodic.

Then by Markov Chain Convergence Theorem, $\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$, and $\lim_{n \to \infty} P[X_n = j] = \pi_j$, for all i, j and v. ∎

Hence, for "large enough" n, X_n is approximately a sample from π.

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 <u>degree</u> 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 <u>(simple) random walk on the (undirected) graph</u> (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 3.9. 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 3.10. 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.

Example 3.3 (Ring Graph)**.** Suppose $V = \{1, 2, 3, 4, 5\}$, with $w(i, i+1) = w(i+1, i) = 1$ for $i = 1, 2, 3, 4$, and $w(5, 1) = w(1, 5) = 1$, with $w(i, j) = 0$ otherwise.

Example 3.4 (Stick Graph)**.** Suppose $V = \{1, 2, \ldots, K\}$, with $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le K - 1$, with $w(i, j) = 0$ otherwise.

Example 3.5 (Star Graph)**.** Suppose $V = \{0, 1, 2, \ldots, K\}$, with $w(i, 0) = w(0, i) = 1$ for $i = 1, 2, 3$, with $w(i, j) = 0$ otherwise.

Example 3.6 (Infinite Graph)**.** Suppose $V = \mathbb{Z}$, with $w(i, i+1) = w(i+1, i) = 1$ for all $i \in V$, and $w(i, j) = 0$ otherwise. Random walk on this graph corresponds exactly to simple symmetric random walk.

Example 3.7. Frog Graph Suppose $V = \{1, 2, \ldots, K\}$, with $w(i, i) = 1$ for $1 \le i \le K$, and $w(i, i+1) = w(i+1, i) = 1$ for $1 \le i \le K - 1$, and $w(K, 1) = w(1, K) = 1$, and $w(i, j) = 0$ otherwise.

Remark 3.11. These walks have stationary distributions.
Let $Z = \sum_{u \in V} d(u) = \sum_{u,v \in V} w(u, v)$. In the unweighted case, Z equals two times the number of edges. (Except that self-edges are only counted once, not twice.) Z might be infinite, if V is infinite. But if $Z < \infty$, then we have a precise formula for a stationary distribution:

Theorem 3.5 (Graph Stationary Distribution Theorem)**.** 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 π is a stationary distribution for this walk.

Proof. It is easily checked that $\pi_u \ge 0$, and $\sum_u \pi_u = 1$, so π is a probability distribution on V.

Also,

$$\pi_u p_{uv} = \frac{d(u)}{Z} \frac{w(u,v)}{d(u)} = \frac{w(u,v)}{Z}$$

$$\pi_v p_{vu} = \frac{d(v)}{Z} \frac{w(v,u)}{d(v)} = \frac{w(v,u)}{Z} = \pi_u p_{uv}$$

So the chain is reversible w.r.t. π. So by Proposition 3.2, π is a stationary distribution. ∎

Fact 3.5 (periodicity)**.** The random walk can always return to any vertex u in 2 steps, by moving to any other vertex and then back along the same edge. Therefore, 1 and 2 are the only possible periods.
If the graph is a **bipartite** graph (i.e., it can be divided into two subsets such that all the links go from one subset to the other one), then the chain has period 2. But if the chain is not bipartite, then it is aperiodic.
Furthermore, any cycle of odd length is not bipartite (therefore aperiodic), any cycle of even length is bipartite.

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).

Proof. If the graph is connected, then the random walk is irreducible. If the graph is non-bipartite, then the random walk is aperiodic.
Also, if $Z < \infty$, then $\pi_v = \frac{d(v)}{Z}$ is a stationary distribution, by Graph Stationary Distribution Theorem 3.5. Hence, by the Markov Chain Convergence Theorem 3.1, $\lim\limits_{n \to \infty} p_{uv}^{(n)} = \frac{d(v)}{Z}$ for all $u, v \in V$, and $\lim\limits_{n \to \infty} P(X_n = v) = \frac{d(v)}{Z}$ (for any initial probabilities). ∎

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

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

Remark 3.12. For graphs which are bipartite, since the only possible period is 2, the Periodic Convergence Theorem and Average Probability Convergence leads to the theorem.

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

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

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

$$
s(a) = ps(a) + (1 - p)s(a) = ps(a + 1) + (1 - p)s(a - 1) \qquad \text{(re-arranging)}
$$

$$
\implies s(a + 1) - s(a) = \frac{1 - p}{p}[s(a) - s(a - 1)]
$$

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

$$
\begin{aligned}
s(1) - s(0) &= x \\
s(2) - s(1) &= \frac{1 - p}{p}[s(1) - s(0)] = \frac{1 - p}{p}x \\
s(3) - s(2) &= \frac{1 - p}{p}[s(2) - s(1)] = \left(\frac{1 - p}{p}\right)^2 x
\end{aligned}
$$

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

$$
\begin{aligned}
\implies s(a) &= s(a) - s(0) \\
&= [s(a) - s(a - 1)] + [s(a - 1) - s(a - 2)] + \ldots + [s(1) - s(0)] \\
&= \left[\left(\frac{1 - p}{p}\right)^{a-1} + \left(\frac{1 - p}{p}\right)^{a-2} + \ldots + \left(\frac{1 - p}{p}\right) + 1\right] x \\
&= \begin{cases} \left[\frac{(\frac{1-p}{p})^a - 1}{(\frac{1-p}{p}) - 1}\right] x, & p \neq \frac{1}{2} \\ ax, & p = \frac{1}{2} \end{cases}
\end{aligned}
$$

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

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

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

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

Remark 3.13. We will sometimes write $s(a)$ as $s_{c,p}(a)$, to show the explicit dependence on c and p.

Example 3.8. $c = 10,000, a = 9,700, p = 0.5$, then

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

Example 3.9. $c = 10,000, a = 9,700, p = 0.49$, 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 $P(T > mc) \leq (1 - p^c)^m$ 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

$$E[T] = \sum_{i=1}^{\infty} P(T \geq i)$$

$$\leq \sum_{i=0}^{\infty} P(T \geq i)$$

$$= P(T \geq 0) + P(T \geq 1) + P(T \geq 2) + P(T \geq 3) + P(T \geq 4) + \ldots$$

$$\leq \underbrace{P(T \geq 0) + P(T \geq 0) + \ldots + P(T \geq 0)}_{c \text{ terms}} + \underbrace{P(T \geq c) + P(T \geq c) + \ldots + P(T \geq c)}_{c \text{ terms}} + \ldots$$

$$= \sum_{j=0}^{\infty} cP(T \geq cj)$$

$$\leq \sum_{j=0}^{\infty} c(1 - p^c)^j$$

$$= \frac{c}{1 - (1 - p^c)}$$

$$= \frac{c}{p^c} < \infty$$

∎

Remark 3.14. 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 3.15. 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.

Proof. Suppose a chain starts at i, and $m_i < \infty$.

On average, the chain returns to i once every m_i steps. That is, for large r, by the Law of Large Numbers, it takes about rm_i steps for the chain to return to i a total of r times. So, in a large number $n \approx rm_i$ of steps, it will return to i about $r = \approx \frac{n}{m_i}$ times, i.e., about $\frac{1}{m_i}$ of the time. Hence, the limiting fraction of time it spends at i is equal to $\frac{1}{m_i}$:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}\{X_k = i\} = \frac{1}{m_i}, \ w.p. \ 1$$

It then follows from the Bounded Convergence Theorem 1.5 that

$$\lim_{n \to \infty} E_i\left(\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}\{X_k = i\}\right) = \frac{1}{m_i}$$

However, if the chain is irreducible, with stationary distribution π, then by finite linearity and Average Probability Convergence,

$$\lim_{n \to \infty} E_i\left(\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}\{X_k = i\}\right) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} E_i(\mathbb{1}\{X_k = i\})$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} P_i(X_k = i)$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} p_{ii}^{(k)}$$

$$= \pi_i$$

So we must have $\frac{1}{m_i} = \pi_i \implies \sum_{i \in S} \frac{1}{m_i} = 1$.

If $m_i = \infty$, then it follows as above that we must have $\pi_i = \frac{1}{m_i} = \frac{1}{\infty} = 0$. Hence if $m_j = \infty$ for all j, then we must have $\pi_j = 0$ for all j, which contradicts that $\sum_{j \in S} \pi_j = 1$. So in that case, there is no stationary distribution.

It turns out that, if a chain is irreducible, and $m_j < \infty$ for some $j \in S$, then $m_i < \infty$ for all $i \in S$, and $\{\frac{1}{m_i}\}$ is stationary. ∎

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 3.16. 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 τ 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,

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

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

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

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

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

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 $|X_n| \le n$, so $E|X_n| \le n < \infty$, so there is no problem with finite expectations. For all $i \in S$, we compute that $\sum_{j \in S} j p_{ij} = (i+1)(1/2) + (i-1)(1/2) = i$, 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 and 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:
$T = 5, T = \inf\{n \ge 0 : X_n = 5\}, T = \inf\{n \ge 0 : X_n = 0 \vee X_n = c\}, T = \inf\{n \ge 2 : X_{n-2} = 5\}$ not valid stopping time: $T = \inf\{n \ge 0 : X_{n+1} = 5\}$ (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 $P(T \le M) = 1$), then

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

Proof. We have

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

$$= E\left[\sum_{k=1}^{T}(X_k - X_{k-1})\right)$$

$$= E\left[\sum_{k=1}^{M}(X_k - X_{k-1})\mathbb{1}\{k \leq T\}\right]$$

$$= \sum_{k=1}^{M}E[(X_k - X_{k-1})\mathbb{1}\{k \leq T\}]$$

$$= \sum_{k=1}^{M}E[(X_k - X_{k-1})(1 - \mathbb{1}\{T \leq k - 1\})$$

$$= \sum_{k=1}^{M}E\left(E[(X_k - X_{k-1})(1 - \mathbb{1}\{T \leq k - 1\})|X_0, X_1, \ldots, X_{k-1}]\right)$$

$$\text{(Double-expectation formula)}$$

$$= \sum_{k=1}^{M}E\left((1 - \mathbb{1}\{T \leq k - 1\})E[(X_k - X_{k-1})|X_0, X_1, \ldots, X_{k-1}]\right)$$

Since $\{X_n\}$ is a martingale, then

$$E[X_k|X_0, X_1, \ldots, X_{k-1}] = X_{k-1}$$

Since X_{k-1} is a function of $X_0, X_1, \ldots, X_{k-1}$, then

$$E[X_{k-1}|X_0, X_1, \ldots, X_{k-1}] = X_{k-1}$$

Then we have

$$E(X_T) - E(X_0) = \sum_{k=1}^{M}E\left((1 - \mathbb{1}\{T \leq k - 1\})(X_{k-1} - X_{k-1})\right)$$

$$= \sum_{k=1}^{M}E\left((1 - \mathbb{1}\{T \leq k - 1\})(0)\right)$$

$$= 0$$

Therefore, $E(X_T) = E(X_0)$, as wanted. ∎

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,

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

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,

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

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 1.7

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. $E(X_T) = E(X_0)$. ∎

Remark 4.3. This is a more general result than Optional Stopping Lemma, in the sense that T can either be bounded or "unbounded but still finite" (e.g. the Gambler's Ruin problem).

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,

$$|E(X_n \mathbb{1}\{T > n\})| \leq E(|X_n| \mathbb{1}\{T > n\}) \qquad \text{(by the property of a Riemman Integral)}$$
$$= E(|X_n| \mathbb{1}\{n \leq T\} \mathbb{1}\{T > n\})$$
$$\leq E(M \mathbb{1}\{T > n\})$$
$$= MP(T > n) \to 0 \qquad\qquad \text{(Since } P(T < \infty) = 1)$$

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

Theorem 4.2 (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 $E(X_T) = a = E(X_0)$.

Proof. We compute that

$$
\begin{aligned}
E(X_T) - a &= E(X_T - a) \\
&= E(Z_1 + \ldots + Z_T) \\
&= E\left[\sum_{i=1}^{T} Z_i\right] \\
&= E\left[\sum_{i=1}^{\infty} Z_i \mathbb{1}\{T \geq i\}\right] \\
&= E\left[\lim_{N \to \infty} \sum_{i=1}^{N} Z_i \mathbb{1}\{T \geq i\}\right] \\
&= \lim_{N \to \infty} E\left[\sum_{i=1}^{N} Z_i \mathbb{1}\{T \geq i\}\right]
\end{aligned}
$$

(Assume we can interchange the expectation and limit)

$$
\begin{aligned}
&= \lim_{N \to \infty} \sum_{i=1}^{N} E[Z_i \mathbb{1}\{T \geq i\}] \\
&= \sum_{i=1}^{\infty} E[Z_i \mathbb{1}\{T \geq i\}] \\
&= \sum_{i=1}^{\infty} E[Z_i] E[\mathbb{1}\{T \geq i\}] \qquad (Z_i \text{ is independent of the event } \{T \geq i\}) \\
&= \sum_{i=1}^{\infty} m P[T \geq i] = m \sum_{i=1}^{\infty} P[T \geq i] = m E(T)
\end{aligned}
$$

It remains to justify interchanging the above expectation and limit. This follows from the Dominated Convergence Theorem 1.7 with sequence $X_N = \sum_{i=1}^{N} Z_i \mathbb{1}\{T \geq i\}$ and limit $X = \sum_{i=1}^{\infty} Z_i \mathbb{1}\{T \geq i\}$ and dominator $Y = \sum_{i=1}^{\infty} |Z_i| \mathbb{1}\{T \geq i\}$.
We can see that $|X_N| \leq Y$ for all N, and

$$
\begin{aligned}
E(Y) &= E[\sum_{i=1}^{\infty} |Z_i| \mathbb{1}\{T \geq i\}] \\
&= \sum_{i=1}^{\infty} E[|Z_i| \mathbb{1}\{T \geq i\}] \\
&= \sum_{i=1}^{\infty} E|Z_i| E[\mathbb{1}\{T \geq i\}] \\
&= |m| \sum_{i=1}^{\infty} P[T \geq i] \\
&= |m| E(T) < \infty
\end{aligned}
$$

■

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 ith 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.

Proof. We first check that $E|Y_n| \leq Var(X_n) + nv = 2nv < \infty$.
Then, since Z_{n+1} is independent of $Z_1, \ldots, Z_n, Y_0, \ldots, Y_n$, with $E(Z_{n+1}) = 0$ and $E(Z_{n+1}^2) = v$, we have

$$
\begin{aligned}
E[Y_{n+1}|Y_0, Y_1, \ldots, Y_n] &= E[(Z_1 + \ldots + Z_n + Z_{n+1})^2 - (n+1)v|Y_0, Y_1, \ldots, Y_n] \\
&= E[(Z_1 + \ldots + Z_n)^2 + (Z_{n+1})^2 + 2Z_{n+1}(Z_1 + \ldots + Z_n) - nv - v|Y_0, Y_1, \ldots, Y_n] \\
&= E[Y_n + (Z_{n+1})^2 - v + 2Z_{n+1}(Z_1 + \ldots + Z_n)|Y_0, Y_1, \ldots, Y_n] \\
&= Y_n + v - v + 2E(Z_{n+1})E[Z_1 + \ldots + Z_n|Y_0, Y_1, \ldots, Y_n] \\
&= Y_n + v - v + 0 \\
&= Y_n
\end{aligned}
$$

Therefore, $\{Y_n\}$ is a martingale as wanted.

∎

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

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

Proof. Let $Y_n = (X_n - a)^2 - n$ (since here $v = \frac{1}{2}(1-0)^2 + \frac{1}{2}(-1-0)^2 = 1$). Then $\{Y_n\}$ is a martingale by the above Lemma.

Choose $M > 0$, and let $S_M = \min(T, M)$. Then S_m is a stopping time, which is bounded by M. Hence by the Optional Stopping Lemma, $E[Y_{S_M}] = E[Y_0] = (a-a)^2 - 0 = 0$.

But $Y_{S_M} = (X_{S_M} - a)^2 - S_M$, so $E(S_M) = E[(X_{S_M} - a)^2]$.

As $M \to \infty$, S_M increases monotonically to T, so $E(S_M) \to E(T)$ by the Monotone Convergence Theorem 1.6.

Also, $E[(X_{S_M} - a)^2] \to E[(X_T - a)^2]$ by the Bounded Convergence Theorem 1.5, since for any n, $(X_{S_M} - a)^2 \le \max(a^2, (c-a)^2) < \infty$.

Hence, letting $M \to \infty$ gives

$$E(T) \leftarrow E(S_M) = E[(X_{S_M} - a)^2] \to E[(X_T - a)^2]$$

$$\begin{aligned}
\implies E(T) &= E[(X_T - a)^2] \\
&= Var(X_T) \\
&= (a/c)(c-a)^2 + (1-a/c)(0-a)^2 \qquad (\text{by } 4.4, E(X_T) = E(X_0) = a) \\
&= a(c-a)
\end{aligned}$$

∎

4.4   Application - Sequence Waiting Times

the Problem   Suppose we repeatedly flip a fair coin, and let τ be the number of flips until we first see the pattern HTH. Again we ask, what is the expected time $E(\tau)$?

the Martingale approach   Suppose at each time n, a new "player" appears, and bets \$1 on H, then if they win they bet \$2 on T, then if they win again they bet \$4 on H. 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 τ.

Each of the $\tau - 3$ players (all but the last three) has a net loss of \$1 (after either one or two or three bets).

Player $\tau - 2$ wins all three bets, for a gain of \$7.

Player $\tau - 1$ bets on "H" and loses, for a loss of \$1.

Player τ bets on "H" and wins, for a gain of \$1.

Hence $X_\tau = (\tau - 3)(-1) + (7) + (-1) + (1) = -\tau + 10$.

Let $T_m = \min(\tau, m)$, then $E(X_{T_m}) = 0$ by the Optional Stopping Lemma, and $\lim_{m \to \infty} E(X_{T_m}) = E(X_\tau)$ by the Dominated Convergence Theorem 1.7 with dominator $Y = 7\tau$ since $|X_n - X_{n-1}| \le 7$ and $E(\tau) < \infty$.

Hence

$$0 = E(X_\tau) = -E(\tau) + 10 \implies E(\tau) = 10$$

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.s.r.w. with $p = 1/3$, $Y_n \to -\infty$ w.p. 1 by the Law of Large Numbers 1.3. Hence, $X_n = 2^{Y_n} \to 2^{-\infty} = 0$ w.p. 1.

Theorem 4.3 (Martingale Convergence Theorem). Any 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 4.4. 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 4.5. 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.

Example 4.8. bounded below s.s.r.w. s.s.r.w. is a martingale, but it is not bounded above or below and does not converge.
If we modify s.s.r.w. to stop whenever it reaches 0 (after starting at a positive number), then it is still a martingale, and now it is non-negative and hence bounded below. But since $f_{i0=1}$ for all i, it will eventually reach 0, and hence it converges to $X = 0$.

4.6 Application - Branching Processes

Definition 4.3 (offspring distribution). Let μ be any prob dist on $\{0, 1, 2, \ldots\}$, the <u>offspring distribution</u>. 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 μ). 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 μ.

Theorem 4.4. Let $m = \sum_i i\mu(i)$ be the mean of μ, 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 $m = \sum_i i\mu(i)$ be the mean of μ, which is called the <u>reproductive number</u>. If $m > 1$ and $\mu(0) > 0$, 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 4.5. Let $m = \sum_i i\mu(i)$ be the mean of μ, which is called the <u>reproductive number</u>. If $m = 1$, and μ is non-degenerate (i.e. $\mu(1) < 1$, so that μ is not a constant), then $\{X_n\} \to 0$ w.p. 1.

Proof. If $m = 1$, 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 4.3, we must have $X_n \to X$ for some random variable X. This could only happen if

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

2. $X = 0$

Since μ is non-degenerate, then $X \equiv 0$. i.e., $\{X_n\} \to 0$ w.p. 1. ∎

In summary   Assuming $\mu(0) > 0$, then extinction is certain if $m \leq 1$, but both flourishing and extinction are possible if $m > 1$.

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 4.6. 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 4.9 (naive example)**.** Suppose a stock price X_0 at time 0 is equal to 100, and at time S the stock price X_S is random with $P(X_S = 80) = 9/10$ and $P(X_S = 130) = 1/10$. Suppose there is an option to buy the stock at time S for $K = 110$.

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 4.10. 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 4.6 (Martingale Pricing Principle)**.** The fair price of an option is equal to its expected value (worth) under the martingale probabilities.

Example 4.11. Suppose a stock price at strike data $S > 0$ equals either $X_S = d$ or $X_S = u$ under the martingale probabilities $P(X_S = d) = q_1$ and $P(X_S = u) = q_2$, and there is an option to buy the stock at time S for strike price K.
Then the fair price is

$$c = q_1(0) + q_2(u - K) = q_2(u - K)$$

Proposition 4.2. Suppose a stock price at time 0 equals $X_0 = a$, and at strike date $S > 0$ 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 strike price K is equal to $(a - d)(u - K)/(u - d)$.

Proof. **Profit Computation**
Suppose you buy x shares of the stock for \$$a$ per stock, plus y shares of the option for \$$c$ per share.
Then if the stock goes up to $X_S = u$, your profits is $x(u - a) + y(u - K - c)$. If the stock goes down to $X_S = d$, your profit is $x(d - a) + y(-c)$.
These are equal if $x(d - u) = y(u - K) \iff y = \frac{-x(u-d)}{u-K}$.
If there is no arbitrage, then your guaranteed profit is 0, which equals

$$x(d - a) - yc = x(d - a) + \frac{xc(u - d)}{u - K} = 0 \implies c = \frac{(a - d)(u - K)}{u - d}$$

■

Proof. **Martingale Pricing Principle**
We need to find martingale probabilities $q_1 = P(X_S = d)$ and $q_2 = P(X_S = u)$ to make the stock price a martingale. So we need that

$$dq_1 + uq_2 = a$$
$$\implies d_1 + u(1 - q_1) = a$$
$$\implies (d - u)q_1 + u = a$$
$$\implies q_1 = \frac{u - a}{u - d}, \quad q_2 = 1 - q_1 = \frac{a - d}{u - d}$$

Then by the Martingale Pricing Principle, the fair price of the option is the martingale expectation of the option's worth, which equals

$$q_1(0) + q_2(u - K) = \frac{(a - d)(u - K)}{u - d}$$

■

5 Continuous Processes

So far, we have mostly considered discrete processes, where the time is indexed by non-negative integers, and the process takes on a finite or countable number of different values.
We now consider various generalizations of this to continuous time and/or space. We begin with a continuous generalization of symmetric simple random walk, called Brownian motion.

5.1 Brownian Motion

Let $\{X_n\}_{n=0}^{\infty}$ be a symmetric simple random walk with $X_0 = 0$.
We have

$$X_n = \begin{cases} Z_1 + Z_2 + \ldots + Z_n = X_n + Z_n, & n \geq 1 \\ 0, & n = 0 \end{cases}$$

where $\{Z_i\}$ are i.i.d. with $P(Z_i = +1) = P(Z_i = -1) = \frac{1}{2}$.

Let M be a large integer, and let $\{Y_t^{(M)}\}$ be like $\{X_n\}$, except with time sped up by a factor of M, and space shrunk down by a factor of $\sqrt{M}$. We have $Y_0^{(M)} = 0$ and

$$Y_{\frac{i+1}{M}}^{(M)} = Y_{\frac{i}{M}}^{(M)} + \frac{1}{\sqrt{M}} Z_{i+1}$$

Fill in $\{Y_t^{(M)}\}_{t \geq 0}$ by linear interpolation.

<u>Brownian motion</u> $\{B_t\}_{t \geq 0}$ is (intuitively) the limit as $M \to \infty$ of $\{Y_t^{(M)}\}$.

Since $Y_0^{(M)} = 0$ for all M, also $B_0 = 0$. Also, note that $Y_t^{(M)} = \frac{1}{\sqrt{M}}(Z_1 + Z_2 + \ldots + Z_{tM})$ (at least if $tM \in \mathbb{Z}$, otherwise within $\mathcal{O}(1/\sqrt{M})$, which doesn't matter when $M \to \infty$)

Thus $E(Y_t^{(M)}) = E(Z_i) = 0$ and $Var(Y_t^{(M)}) = (\frac{1}{\sqrt{M}})^2(tM) = t$. As $M \to \infty$, by the Central Limit Theorem,

$$B_t \sim \mathcal{N}(0, t)$$

Also, if $0 < t < s$, then

$$Y_s^{(M)} - Y_t^{(M)} = \frac{1}{\sqrt{M}}(Z_{tM+1} + Z_{tM+2} + \ldots + Z_{sM}) \to \mathcal{N}(0, s - t)$$

which is independent of $Y_t^{(M)}$. Therefore,

$$B_s - B_t \sim \mathcal{N}(0, s - t)$$

which is independent of B_t.

More generally, if $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, then $B_{s_i} - B_{t_i} \sim \mathcal{N}(0, s_i - t_i)$, and $\{B_{s_i} - B_{t_i}\}_{i=1}^{k}$ are all independent. This is called "independent normal increments".

Finally, if $0 < t \leq s$, then

$$\begin{aligned} Cov(B_t, B_s) &= E(B_t B_s) \\ &= E(B_t[B_s - B_t + B_t]) \\ &= E(B_t[B_s - B_t]) + E((B_t)^2) \\ &= E(B_t)E(B_s - B_t) + E((B_t)^2) \quad \text{(Since } B_t \text{ and } B_s - B_t \text{ are independent)} \\ &= (0)(0) + t \\ &= t \end{aligned}$$

Similar for the case $0 < s \leq t$. Therefore

$$Cov(B_t, B_s) = \min(t, s)$$

Definition 5.1 (Brownian motion). <u>Brownian motion</u> is a process $\{B_t\}_{t\geq 0}$ with $B_0 = 0$ and continuous sample paths (the random function $t \mapsto B_t$ is always continuous) satisfying the following three properties

1. $B_t \sim \mathcal{N}(0, t)$

2. Independent normal increments: $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, then $B_{s_i} - B_{t_i} \sim \mathcal{N}(0, s_i - t_i)$, and $\{B_{s_i} - B_{t_i}\}_{i=1}^k$ are all independent.

3. $Cov(B_s, B_t) = \min(s, t)$

Remark 5.1. Such a process exists!

Fact 5.1. $\{B_t\}$ is a (continuous-time) martingale. Hence similar results apply just like for discrete-time martingales.

Proof. If $0 < t < s$, then

$$B_s | B_t = B_t + (B_s - B_t) | B_t$$
$$= B_t + \mathcal{N}(0, s - t)$$
$$\sim \mathcal{N}(B_t, s - t)$$

So in particular,

$$E[B_s | \{B_r\}_{0 \leq r \leq t}] = B_t$$

Hence $\{B_t\}$ is a (continuous-time) martingale. ∎

Example 5.1. Let $a, b > 0$, and let $\tau = \min\{t \geq 0 : B_t = -a \vee b\}$. What is $p \equiv P(B_\tau = b)$?
solution:
Here $\{B_t\}$ is a martingale, and τ is a stopping time. Furthermore, $\{B_t\}$ is bounded up to time τ:

$$|B_t| \mathbb{1}\{t \leq \tau\} \leq \max(|a|, |b|)$$

So just like for discrete martingales, must have

$$E(B_\tau) = E(B_0) = 0$$

Hence we have $p(b) + (1 - p)(-a) = 0 \implies p = \frac{a}{a+b}$.

Example 5.2. Let $a, b > 0$, and let $\tau = \min\{t \geq 0 : B_t = -a \vee b\}$. What is $e \equiv E(\tau)$?
solution:
Let $Y_t = B_t^2 - t$. Then for $0 < t < s$,

$$E[Y_s | \{B_r\}_{r \leq t}] = E[B_s^2 - s | \{B_r\}_{r \leq t}]$$
$$= Var[B_s | \{B_r\}_{r \leq t}] + (E[B_s | \{B_r\}_{r \leq t}])^2 - s$$
$$= (s - t) + (B_t)^2 - s$$
$$= (B_t)^2 - t$$
$$= Y_t$$

Then by the Double-expectation formula 1.2,

$$E[Y_s | \{Y_r\}_{r \leq t}] = E[E[Y_s | \{B_r\}_{r \leq t}] | \{Y_r\}_{r \leq t}] = E[Y_t | \{Y_r\}_{r \leq t}] = Y_t$$

So, $\{Y_t\}$ is also a martingale. Then

$$
\begin{aligned}
E(Y_\tau) &= E(B_\tau^2 - \tau) \\
&= E(B_\tau^2 - E(\tau)) \\
&= pb^2 + (1-p)(-a)^2 - e \\
&= \frac{a}{a+b}b^2 + \frac{b}{a+b}a^2 - e \\
&= ab - e
\end{aligned}
$$

Now we want to justify that $E(Y_\tau) = 0$.

Let $\tau_M = \min(\tau, M)$. Then τ_M is bounded, so $E(Y_{\tau_M}) = 0$. But $Y_{\tau_M} = B_{\tau_M}^2 - \tau_M$, so $E(\tau_M) = E(B_{\tau_M}^2)$. As $M \to \infty$, $E(\tau_M) \to E(\tau)$ by the Monotone Convergence Theorem, and $E(B_{\tau_M}^2) \to E(B_\tau^2)$ by the Bounded Convergence Theorem 1.5. Therefore,

$$E(\tau) = E(B_\tau^2) \implies E(Y_\tau) = 0$$

Then since $E(Y_\tau) = ab - e = 0$, we can solve this to get that $e = ab$.

Fact 5.2. The function $t \mapsto B_t$ is differentiable **nowhere**, despite the fact that it is continuous. reasons

5.2   Application - Stock Options (Continuous)

Assume now that the stock price X_t is equal to some (random) positive value, for each time $t \geq 0$. Common model for the stock price:

$$X_t = x_0 \exp(\mu t + \sigma B_t)$$

Also assume a risk-free interest rate r, so that \$1 today is worth \$$e^{rt}$ after t years. Equivalently, \$1 at a future time $t > 0$ is worth \$$e^{-rt}$ at time 0.

Definition 5.2 ("discounted" stock price). We define "discounted" stock price (in today's dollars) as

$$D_t \equiv e^{-rt} X_t = e^{-rt} x_0 \exp(\mu t + \sigma B_t) = x_0 \exp((\mu - r)t + \sigma B_t)$$

Recall that a stock option is the option to buy the stock for some amount \$$K$ at some fixed future time $S > 0$ years later. At time S, this is worth $\max(0, X_S - K)$. So at time 0, it's worth the discounted value $e^{-rS} \max(0, X_S - K)$. However, at time 0, X_S is unknown (random), so this discounted value is a random variable.

Remark 5.2. What is the fair price?

Fact 5.3. If $\mu = r - \frac{\sigma^2}{2}$, then it can be computed that $\{D_t\}$ becomes a martingale. Then similar to the discrete case, the fair price of the option at time 0 is the same as the expected value of the option at time S under the martingale probabilities, by the Martingale Pricing Principle.

i.e., if $X_S = x_0 \exp([r - \frac{\sigma^2}{2}]S + \sigma B_S)$, where $B_S \sim \mathcal{N}(0, S)$, then the fair price for the option equals

$$E[e^{-rs}\max(0, X_S - K)] = x_0 \Phi\left(\frac{(r + \frac{\sigma^2}{2}S - \log(K/x_0)}{\sigma\sqrt{S}}\right) - e^{-rS}K\Phi\left(\frac{(r - \frac{\sigma^2}{2}S - \log(K/x_0)}{\sigma\sqrt{S}}\right)$$

Remark 5.3. This price does not depend on the drift μ, since we first have to replace μ by $r - \frac{\sigma^2}{2}$. However, the price is an increasing function the volatility σ, since the option "protects" you against the large drops in the stock price.

5.3 Poisson Processes

Example 5.3. Suppose an average of $\lambda = 2.5$ fires in Toronto per day. Intuitively, this is caused by a very large number n of buildings, each of which has a very small probability p of having a fire. Then the mean $= np = \lambda$, so $p = \frac{\lambda}{n}$, and the number of fires today $X \sim Binom(n, p) = Binom(n, \lambda/n)$.

As $n \to \infty$, we have

$$P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$$

$$= \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1 - \left(\frac{\lambda}{n}\right)\right)^{n-k}$$

$$\approx \frac{n^k}{k!}\left(\frac{\lambda}{n}\right)^k\left(e^{-\lambda/n}\right)^n$$

$$= \frac{1}{k!}\lambda^k e^{-\lambda}$$

This is the *pdf* of $Poisson(\lambda)$.

Property 5.1. It is possible to show that if a distribution is continuous and memoryless, then it must have this pdf:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

which is the pdf of the exponential distribution.

In other words, memoryless $\iff$ exponential.

Remark 5.4. The rate parameter λ is some positive real number. You can conceive of it as the probability an event will occur in unit time. The larger λ is, the more rapidly the pdf descends with increasing x.

Construction of Poisson process   Let $\{Y_n\}_{n=1}^\infty \overset{i.i.d.}{\sim} Exp(\lambda)$ for some $\lambda > 0$.

Let $T_0 = 0$, and $T_n = Y_1 + Y_2 + \ldots + Y_n$ for $n \geq 1$ ("nth arrival time").

Let $N(t) = \max\{n \geq 0 : T_n \leq t\}$ be the number of arrivals up to time t.

$N(t)$ is called a counting process, since it counts number of arrivals.

Proposition 5.1. $N(t) \sim Poisson(\lambda t)$, so that $\forall m \in \mathbb{N}$,

$$P(N(t) = m) = \frac{(\lambda t)^m}{m!} e^{-\lambda t}$$

Proof. $N(t) = m$ iff both $T_m \leq t$ and $T_{m+1} > t$, iff there is $0 \leq s \leq t$ with $T_m = s$ and $T_{m+1} - T_m > t - s$.

Recall that $Y_n \sim Exp(\lambda) = Gamma(1, \lambda)$, so

$$T_m := Y_1 + Y_2 + \ldots + Y_m \sim Gamma(m, \lambda)$$

with density function

$$f_{T_m}(s) = \frac{\lambda^m}{\Gamma(m)} s^{m-1} e^{-\lambda s} = \frac{\lambda^m}{(m-1)!} s^{m-1} e^{-\lambda s}$$

Also, $P(T_{m+1} - T_m > t - s) = P(Y_{m+1} > t - s) = e^{-\lambda(t-s)}$. Then we have

$$
\begin{aligned}
P(N(t) = m) &= P(T_m \leq t, T_{m+1} > t) \\
&= P(\exists 0 \leq s \leq t : T_m = s, Y_{m+1} > t - s) \\
&= \int_0^t f_{T_m}(s) P(Y_{m+1} > t - s) \, ds \\
&= \int_0^t \frac{\lambda^m}{(m-1)!} s^{m-1} e^{-\lambda s} e^{-\lambda s} e^{-\lambda(t-s)} \, ds \\
&= \frac{\lambda^m}{(m-1)!} e^{-\lambda t} \int_0^t s^{m-1} \, ds \\
&= \frac{\lambda^m}{(m-1)!} e^{-\lambda t} \left[\frac{t^m}{m} \right] \\
&= \frac{(\lambda t)^m}{m!} e^{-\lambda t}
\end{aligned}
$$

∎

Property 5.2. If $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, then $N(s_i) - N(t_i) \sim Poisson(\lambda(s_i - t_i))$, and $\{N(s_i) - N(t_i)\}_{i=1}^k$ are all independent. This is called "independent Poisson increments".

Proof. Recall the memoryless property of $Exp(\lambda)$ distribution: for $a, b > 0$,

$$P(Y_n > b + a | Y_n > a) = P(Y_n > b) = e^{-\lambda b}$$

This means the process $\{N(t)\}$ "starts over" in each new time interval.

It follows that $N(t + s) - N(s) \sim N(t) \sim Poisson(\lambda t)$.

Also, it follows that if $0 \leq a < b \leq c < d$, then $N(d) - N(c)$ is independent of $N(b) - N(a)$, and similarly for multiple non-overlapping time intervals. ∎

Definition 5.3. A Poisson process with intensity λ is a collection $\{N(t)\}_{t \geq 0}$ of random non-decreasing integer counts $N(t)$, satisfying the following properties

1. $N(0) = 0$

2. $\forall t \geq 0, N(t) \sim Poisson(\lambda t)$

3. Independent Poisson increments: If $0 \leq t_1 \leq s_1 \leq t_2 \leq s_2 \leq \ldots \leq t_k \leq s_k$, then $N(s_i) - N(t_i) \sim Poisson(\lambda(s_i - t_i))$, and $\{N(s_i) - N(t_i)\}_{i=1}^{k}$ are all independent.

Example 5.4. Let $\{N(t)\}$ be a Poisson process with intensity λ. Then for $0 < t < s$,

$$P(N(t) = 1 | N(s) = 1) = \frac{P(N(t) = 1, N(s) = 1)}{P(N(s) = 1)}$$
$$= \frac{P(N(t) = 1, N(s) - N(t) = 0)}{P(N(s) = 1)}$$
$$= t/s$$

That is, condition on $N(s) = 1$, the first event is uniform over $[0, s]$. (Distribution does not depend on λ.

Example 5.5.

$$P(N(4) = 1 | N(5) = 3) = \frac{P(N(4) = 1, N(5) = 3)}{P(N(5) = 3)}$$
$$= \frac{P(N(4) = 1, N(5) - N(4) = 2)}{P(N(5) = 3)}$$
$$= 12/125$$

This also does not depend on λ, and equals $\binom{3}{1}(4/5)^1(1/5)^2$.
Intuition: 3 events occurred in 5 units of time, each event is uniform over $[0, 5]$, so for $i = 1, 2, 3, P(T_{e_i} \leq 4) = \frac{4}{5}$, where e represents an event, follows a Bernoulli distribution with $p = 4/5$. Then $P(N(4) = x | N(5) = 3) \sim Binom(3, 4/5)$.

Proposition 5.2. If $N(t)$ is a Poisson process with rate $\lambda > 0$, then as $h \searrow 0$:

1. $P(N(t + h) - N(t) = 1) = \lambda h + o(h)$

2. $P(N(t + h) - N(t) \geq 2) = o(h)$

Proof.

$$P(N(t + h) - N(t) = 1) = P(N(h) = 1) \tag{5.1}$$
$$= e^{-\lambda h}(\lambda h)^1/1! \tag{5.2}$$
$$= [1 - \lambda h + O(h^2)](\lambda h) \tag{5.3}$$
$$= \lambda h - \lambda^2 h^2 + O(h^3) \tag{5.4}$$
$$= \lambda h + O(h^2) \tag{5.5}$$
$$= \lambda h + o(h) \tag{5.6}$$

Also,

$$P(N(t+h) - N(t) \geq 2) = P(N(h) \geq 2) \tag{5.7}$$
$$= 1 - P(N(h) = 0) - P(N(h) = 1) \tag{5.8}$$
$$= 1 - e^{-\lambda h}(\lambda h)^0/0! - e^{-\lambda h}(\lambda h)^1/1! \tag{5.9}$$
$$= 1 - [1 - \lambda h + O(h^2)](1) - [1 - \lambda h + O(h^2)](\lambda h) \tag{5.10}$$
$$= 1 - 1 + \lambda h - O(h^2) - \lambda h + \lambda^2 h^2 - O(h^3) \tag{5.11}$$
$$= \lambda^2 h^2 + O(h^2) + O(h^3) \tag{5.12}$$
$$= O(h^2) \tag{5.13}$$
$$= o(h) \tag{5.14}$$

∎

Property 5.3 (Poisson clumping). Suppose we have a Poisson process on $[0, 100]$ with intensity $\lambda = 1$ event per day. Then the probability of seeing 4 events within some single day is

$$P(\exists r \in [1, 100] : N(r) - N(r-1) = 4) \geq P(\exists m \in \{1, 2, \ldots, 100\} : N(m) - N(m-1) = 4)$$
$$= 1 - P(\nexists m \in \{1, 2, \ldots, 100\} : N(m) - N(m-1) = 4)$$
$$= 1 - (P(N(1) \neq 4))^{100}$$
$$= 1 - (1 - P(N(1) = 4))^{100}$$
$$= 1 - (1 - e^{-1}(1^4/4!))^{100}$$
$$\approx 0.787$$

Therefore, the $\{T_i\}$ tend to "clump up" in various patterns, just by chance alone. This doesn't mean anything at all: they are still independent.

Proposition 5.3. Suppose $\{N_1(t)\}_{t\geq 0}$ and $\{N_2(t)\}_{t\geq 0}$ are independent Poisson processes, with rates λ_1 and λ_2 respectively. Let $N(t) = N_1(t) + N_2(t)$. Then $\{N(t)\}_{t\geq 0}$ is also a Poisson process, with rate $\lambda_1 + \lambda_2$.

Proposition 5.4 (thin). Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with rate λ. Suppose each arrival is independently of "type i" with probability p_i, for $i = 1, 2, 3, \ldots$ (could be countably infinite), where $\sum_i p_i = 1$. Let $N_i(t)$ be number of arrivals of type i up to time t. Then the different $\{N_i(t)\}$ are all independent, and furthermore each $\{N_i(t)\}$ is itself a Poisson process, with rate λp_i. `prove this!`

5.4   Continuous-Time, Discrete-Space Processes

Recall that a Markov chain $\{X_n\}_{n=0}^{\infty}$ is defined in discrete time. But Brownian motion $\{B_t\}_{t\geq 0}$ and Poisson processes $\{N(t)\}_{t\geq 0}$ are both defined in continuous time. So we want to define Markov processes in continuous time.

Definition 5.4 (continuous-time Markov process). A continuous-time (time-homogeneous, non-explosive) Markov process, on a countable (discrete) state space S, is a collection $\{X(t)\}_{t\geq 0}$ of random variables such that

$$P(X_0 = i_0, X_{t_1} = i_1, X_{t_2} = i_2, \ldots, X_{t_n} = i_n) = v_{i0} p_{i_0 i_1}^{(t_1)} p_{i_1 i_2}^{(t_2 - t_1)} \ldots p_{i_{n-1} i_n}^{(t_n - t_{n-1})}$$

for all $i_0, i_1, \ldots, i_n \in S$ and all $0 < t_1 < t_2 < \ldots < t_n$.

Property 5.4. We need some initial distribution $\{v_i\}_{i \in S}$ (with $v_i \geq 0$ and $\sum_{i \in S} v_i = 1$), and transition probabilities $\{p_{ij}^{(t)}\}_{i,j \in S, t \geq 0}$ (with $p_{ij}^{(t)} \geq 0$ and $\sum_{j \in S} p_{ij}^{(t)} = 1$)

Also, $\delta_{ij} := p_{ij}^{(0)} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

Remark 5.5. This is just like discrete-time chains, except need to keep track of elapsed time t too.

Property 5.5 (standard Markov process). In a standard Markov process,

$$\lim_{t \searrow 0} p_{ij}^{(t)} = p_{ij}^{(0)} = \delta_{ij}$$

That is, the mapping $t \mapsto p_{ij}^{(t)}$ is continuous.
Note that **this property does not always hold.** But from now on, assume our processes are standard.

Example 5.6. Let $S = \{1, 2\}$ with $p_{11}^{(0)} = p_{22}^{(0)} = 1$. Suppose that for all $t > 0, p_{11}^{(t)} = p_{21}^{(t)} = 1$. Then the process is not standard, since $p_{21}^{(0)} = 0$ so that p_{21} is not continuous at 0.

Property 5.6 (aperiodic). Continuous-time standard processes are always aperiodic.

Proof. For all small enough $h > 0, p_{ii}^{(h)} > 0$. Hence by Chapman-Kolmogorov Inequality 2.2,

$$p_{ii}^{(nh)} \geq \left(p_{ii}^{(h)} \right)^n > 0$$

for all small enough $h > 0$ and all $n \in \mathbb{N}$.
Hence $p_{ii}^{(t)} > 0$ for all $t \geq 0$.
Hence continuous-time standard processes are always aperiodic. ∎

Definition 5.5 (generator). Given a standard Markov process, its <u>generator</u> is the right-handed derivative

$$g_{ij} =: \lim_{t \searrow 0} \frac{p_{ij}^{(t)} - \delta_{ij}}{t} = p_{ij}'^{(0)}$$

In matrix form we have

$$G = (g_{ij})_{i,j \in S} = P'(0) = \lim_{t \searrow 0} \frac{P(t) - I}{t}$$

Property 5.7. $g_{ii} \leq 0$, while $g_{ij} \geq 0$ for $i \neq j$.

Proof. The result follows by the sign of the limits. ∎

Property 5.8. If the state space S is finite, then a generator G of a standard Markov process on S has row sums equal to 0. i.e., $\sum_{j \in S} g_{ij} = 0$.

Proof.

$$\sum_{j \in S} g_{ij} = \sum_{j \in S} \lim_{t \searrow 0} \frac{p_{ij}^{(t)} - \delta_{ij}}{t}$$

$$= \lim_{t \searrow 0} \frac{\sum_{j \in S} p_{ij}^{(t)} - \sum_{j \in S} \delta_{ij}}{t} \qquad \text{(exchange the sum and limit since } S \text{ finite)}$$

$$= \lim_{t \searrow 0} \frac{1 - 1}{t}$$

$$= 0$$

$\blacksquare$

Property 5.9. If $t > 0$ is small, then $\frac{P^{(t)} - I}{t} \approx G$, so

$$P^{(t)} \approx I + tG \iff p_{ij}^{(t)} \approx \delta_{ij} + tg_{ij}$$

Example 5.7.

$$S = \{1, 2\}, G = \begin{pmatrix} -3 & 3 \\ 6 & -6 \end{pmatrix}$$

G is a valid generator since

1. $g_{ii} \leq 0$, while $g_{ij} \geq 0$ for $i \neq j$

2. $\sum_{j \in S} g_{ij} = 0$

Then for small $t > 0, P^{(t)} \approx I + tG = \begin{pmatrix} 1 - 3t & 3t \\ 6t & 1 - 6t \end{pmatrix}$

Then for example, $p_{11}^{(0.02)} \approx 1 - 3(0.02) = 0.94, p_{12}^{(0.02)} \approx 3(0.02) = 0.06, P^{(0.02)} \approx \begin{pmatrix} 0.94 & 0.06 \\ 0.12 & 0.88 \end{pmatrix}$.

Theorem 5.1 (Continuous-Time Transitions Theorem). If a continuous-time Markov process has generator matrix G, then for any $t \geq 0$,

$$P^{(t)} = \exp(tG) := I + tG + \frac{t^2 G^2}{2!} + \frac{t^3 G^3}{3!} + \dots$$

Proof. Recall that $\lim_{n \to \infty} (1 + \frac{c}{n})^n = e^c = 1 + c + \frac{c^2}{2!} + \frac{c^3}{3!} + \dots$.

Similarly, for a matrix A, $\lim_{n \to \infty} (I + \frac{1}{n}A)^n = \exp(A)$ where $\exp(A) = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots$.

Hence, by Chapman-Kolmogorov equations 2.1,

$$P^{(t)} = [P^{(t/m)}]^m \qquad \text{for any } m \in \mathbb{N}$$

$$= \lim_{n \to \infty} [P^{(t/n)}]^n$$

$$= \lim_{n \to \infty} [I + (t/n)G]^n \qquad \text{by property 5.9 since } t/n \text{ is small}$$

$$= \exp(tG)$$

$$= I + tG + \frac{t^2 G^2}{2!} + \frac{t^3 G^3}{3!} + \dots$$

$\blacksquare$