Notes: MATH235 (F25) Markov Chains, Martingales, and Brownian Motion

Lecturer: Ewain Gwynne
Notes by: Aden Chen

Tuesday 18th November, 2025

Contents

1 Preliminaries

Definition 1.1 (Stochastic Process)**.** A **stochastic process** is a collection of random variables $\{X_t\}_{t \in T}$, where each X_t takes values in a **state space** S.

Remark 1.2. Alternatively, one may think of a random function $X : T \to S$.

Remark 1.3. We think of T as representing time. In this course, T will either be discrete ($T = \mathbb{N}_0$) or continuous ($T = [0, \infty)$).

Example 1.4.

- $\{X_n\}_{n \geq 0}$ is a sequence of independent random variables.
- Let $\{Y_n\}_{n \geq 0}$ are iid RVs in $\mathbb{R}$. We can consider $X_0 = 0$ and $X_n = \sum_{i=1}^n Y_i$ for $n \geq 1$.

Now recall that if Y is a RV in a countable set, the **distribution** of Y is the function $y \mapsto \mathbb{P}(Y = y)$. What is the analogue for a stochastic process? How to describe the distribution of $\{X_n\}_{n \geq 0}$. It suffices to describe

$$\mathbb{P}\left[X_0 = s_0, \ldots, X_n = s_n\right], \quad \forall n \in \mathbb{N}, \quad \forall s_0, \ldots, s_n \in S.$$

Definition 1.5 (Conditional Probability)**.** Let E and F be events such that $\mathbb{P}(F) > 0$. Then the **conditional probability** of E given F is

$$\mathbb{P}(E \mid F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}.$$

By definition, for each n and each $s_0, \ldots, s_n \in S$, we may write

$$\begin{aligned}\mathbb{P}\left[X_0 = s_0, \ldots, X_n = s_n\right] &= \mathbb{P}\left[X_n = s_n | X_0 = s_0, \ldots, X_{n-1} = s_{n-1}\right] \mathbb{P}\left[X_0 = s_0, \ldots, X_{n-1} = s_{n-1}\right] \\ &= \left(\prod_{i=1}^n \mathbb{P}\left[X_i = s_i | X_0 = s_0, \ldots, X_{i-1} = s_{i-1}\right]\right) \mathbb{P}\left[X_0 = s_0\right],\end{aligned}$$

assuming the conditional probabilities are well-defined. Thus it suffices to specify the initial distribution $\mathbb{P}(X_0 = s_0)$ and the conditional probabilities to describe the distribution of the stochastic process.

Without imposing any restrictions, there is little more we can say about the distribution of a stochastic process. The first restriction we will impose is the Markov property.

2 Markov Chains on Finite State Space

Definition 2.1 (Markov)**.** We say that a stochastic process $\{X_n\}_{n\geq 0}$ is a **Markov process (chain)** if for each n and each $s_0, \ldots, s_n, s_{n+1} \in S$, we have

$$\mathbb{P}\left[X_n = s_n | X_0 = s_0, \ldots, X_{n-1} = s_{n-1}\right] = \mathbb{P}\left[X_n = s_n | X_{n-1} = s_{n-1}\right].$$

We say $\{X_n\}_{n\geq 0}$ is **time-homogeneous** if for each n and each $s, s' \in S$, we have

$$\mathbb{P}\left[X_n = y | X_{n-1} = x\right] = \mathbb{P}\left[X_1 = y | X_0 = x\right], \quad \forall n \geq 1, \quad \forall x, y \in S.$$

In this class, we will assume all Markov processes are time-homogeneous.

To describe the distribution of a Markov process, we need only describe the distribution of X_0 together with the **transition probabilities**

$$P(x, y) := \mathbb{P}(X_1 = y | X_0 = x), \quad \forall x, y \in S.$$

Example 2.2.

- Let $\{Y_j\}_{j\geq 0}$ be iid RV in $\mathbb{Z}$. Let $X_0 = 0$ and $X_n = \sum_{j=1}^n Y_j$ for all $n \geq 1$. Then $\{X_n\}_{n\geq 0}$ is a Markov process: $X_n = X_{n-1} + Y_n$. The transition probabilities are given by $p(x, y) = \mathbb{P}(Y_1 = y - x)$.
- Let $S = \{0, 1\}$. We have the restrictions

  $$P(0,0) + P(0,1) = P(1,0) + P(1,1) = 1.$$

  The Markov chain is thus completely characterized by the values of $P(0,0)$ and $P(1,0)$. The transition probabilities can be represented by a graph with nodes S.

Example 2.3 (Random Walk on a Graph). A **graph** G is a collections of **vertices** $V(G)$ and **edges** $E(G)$ joining pairs of vertices. We assume each vertex is incident to finitely many edges, though we allow infinitely many vertices.

A random walk on G is the Markov chain with transition probabilities given by

$$P(x, y) = \begin{cases} 1/\deg(x), & \text{if } x \text{ and } y \text{ are joined by an edge,} \\ 0, & \text{otherwise,} \end{cases} \qquad \forall x, y \in V(G).$$

Here $\deg(x)$ is the **degree** of vertex x, i.e. the number of neighbors of x.

Example 2.4 (Simple Random Walk on $\mathbb{Z}$). Let $G = \mathbb{Z}$ with edges joining n and $n+1$ for each $n \in \mathbb{Z}$. We can equivalently describe the simple random walk on $\mathbb{Z}$ as follows: Let $\{Y_j\}_{j\geq 0}$ be iid RVs with $\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = 1/2$. Let $X_0 = 0$ and $X_n = \sum_{j=1}^n Y_j$ for $n \geq 1$. Then $\{X_n\}_{n\geq 1}$ is a simple random walk on $\mathbb{Z}$.

Example 2.5 (Non-example). Let $S = \mathbb{Z}$. Consider $X_0 = X_1 = X_2 = 0$, and for each $n \geq 3$,

$$X_n = \begin{cases} X_{n-1} + 1, & \text{wp } \frac{1}{3}, \\ X_{n-1} - 1, & \text{wp } \frac{1}{3}, \\ X_{n-3}, & \text{wp } \frac{1}{3}. \end{cases}$$

Then $\{X_n\}_{n\geq 0}$ is not a Markov process.

Definition 2.6. The ***n*-step transition probabilities** are

$$P^n(x,y) := \mathbb{P}(X_n = y | X_0 = x), \quad \forall x, y \in S.$$

Proposition 2.7. *For each $n, m \in \mathbb{N}$, $x, y \in S$, we have*

$$P^{n+m}(x,y) = \sum_{z \in S} P^n(x,z) P^m(z,y).$$

Proof.

$$\begin{aligned} P^n(x,z)P^m(z,y) &= \mathbb{P}(X_n = z | X_0 = x)\mathbb{P}(X_m = y | X_0 = z) \\ &= \mathbb{P}(X_n = z | X_0 = x)\mathbb{P}(X_{n+m} = y | X_n = z, X_0 = x) \\ &= \mathbb{P}(X_n = z, X_{n+m} = y | X_0 = x). \end{aligned}$$

The second equality follows from time-homogeneity, the third from the Markov property. Thus,

$$\begin{aligned} \sum_{z \in S} P^n(x,z)P^m(z,y) &= \sum_{z \in S} \mathbb{P}(X_n = z, X_{n+m} = y | X_0 = x) \\ &= \mathbb{P}(X_{n+m} = y | X_0 = x) = P^{n+m}(x,y). \end{aligned}$$

□

2.0.1 Transition Matrix

Assume now that S is finite. Without loss of generality, let $S = \{1, 2, \ldots, N\}$.

Definition 2.8. The **transition matrix** of a Markov chain $\{X_n\}_{n \geq 1}$ is the $N \times N$ matrix P such that $P_{i,j} = P(i,j)$:

$$P := \begin{pmatrix} P(1,1) & P(1,2) & \cdots & P(1,N) \\ P(2,1) & P(2,2) & \cdots & P(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ P(N,1) & P(N,2) & \cdots & P(N,N) \end{pmatrix}.$$

We write $\pi_j = \mathbb{P}(X_0 = j)$ and define the row vector $\pi = (\pi_1, \ldots, \pi_N)$.

Remark 2.9. Note that each row of P sums to 1. A matrix with this property is called a **stochastic matrix**.

Proposition 2.10. *For each $i = 1, \ldots, N$, the ith entry of the vector πP is $\mathbb{P}(X_1 = i)$.*

Proof.

$$\begin{aligned} (\pi P)_i &= \sum_{j=1}^{N} \pi_j P(j,i) = \sum_{j=1}^{N} \mathbb{P}(X_0 = j)\mathbb{P}(X_1 = i | X_0 = j) \\ &= \sum_j \mathbb{P}(X_0 = j, X_1 = i) = \mathbb{P}(X_1 = i). \end{aligned}$$

□

Proposition 2.11. *For each $n \geq 1$, we have $(P^n)_{i,j} = P^n(i,j)$.*

Proof. We induct on n. The base case $n = 1$ is true by definition. Now assume $n \geq 2$ and $(P^{n-1})_{i,j} = P^{n-1}(i,j)$. We have

$$(P^n)_{i,j} = (P^{n-1}P)_{i,j} = \sum_{k=1}^{N} (P^{n-1})_{i,k} P_{k,j}$$
$$= \sum_{k=1}^{N} P^{n-1}(i,k)P(k,j) = P^n(i,j).$$

In the second line, the first equality comes from the induction hypothesis, the second from a previous proposition. □

Example 2.12. Let $S = \{0,1\}$, $P(0,1) = 1/3$, and $P(1,0) = 1/2$. This information completely determines the transition matrix:

$$P = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Suppose we want to find the conditional distribution of X_3 given $X_0 = 0$.

$$P^3 = \begin{pmatrix} \frac{65}{108} & \frac{43}{108} \\ \frac{43}{72} & \frac{29}{72} \end{pmatrix}.$$

We have

$$\mathbb{P}(X_3 = 0 | X_0 = 0) = (P^3)_{1,1} = \frac{65}{108}, \quad \mathbb{P}(X_3 = 1 | X_0 = 0) = (P^3)_{1,2} = \frac{43}{108}.$$

2.0.2 Recurrent and Transient States

Definition 2.13. We say states $x, y \in S$ **communicate** (denoted $x \leftrightarrow y$) if there exist $m, n \geq 0$ such that $P^m(x,y) > 0$ and $P^n(y,x) > 0$. That is, if it is possible to reach y from x and x from y.

Proposition 2.14. *Communication is an equivalence relation on S.*

Proof.

(i) Reflexivity is clear since $P^0(x,x) = 1 \geq 0$.

(ii) Symmetry is clear from definition.

(iii) Transitivity: Choose n, m, l, k such that

$$P^n(x,y), P^m(y,x), P^l(y,z), P^k(z,y) > 0.$$

Then

$$P^{n+l}(x,z) = \sum_{w \in S} P^n(x,w)P^l(w,z) \geq P^n(x,y)P^l(y,z) > 0,$$

and similarly $P^{m+k}(z,x) > 0$.

$\square$

Definition 2.15 (Communication Classes, Recurrent and Transient for Finite State Space)**.** The equivalence classes induced by communication are called **communication classes**. That is, $x, y \in S$ belong to the same communication class if and only if $x \leftrightarrow y$.

A communication class $C \subset S$ is said to be **recurrent** if $P(x, y) = 0$ for each $x \in C$ and each $y \notin C$. Otherwise, C is said to be **transient**.

Remark 2.16. Intuitively, a recurrent class is one that the Markov chain cannot leave. Note that this definition only works for finite state spaces.

Definition 2.17. We say $\{X_n\}$ is **irreducible** if there is only one communication class.

Remark 2.18. An irreducible Markov chain has only one recurrent class.

Example 2.19. Fix graph G and consider the random walk $\{X_n\}$ on G. Then two vertices x, y are in the same communication class if and only if there exists a path from x to y. Thus the communication classes are exactly the connected components of G.

Example 2.20. Let $S = \{1, 2, 3, 4, 5\}$ and

$$P := \begin{pmatrix} \frac{1}{5} & \frac{1}{5} & 0 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \end{pmatrix}.$$

The matrix P can be described by a directed graph.

Example 2.21. Consider the random walk on $\{0, 1, \ldots, N\}$ with **absorbing boundary**. That is,

$$P(x, x+1) = P(x, x-1) = \frac{1}{2}, \quad \forall x = 1, \ldots, N-1,$$

and

$$P(0, 0) = P(N, N) = 1.$$

The random walk has the communication classes $\{0\}, \{N\}, \{1, 2, \ldots, N-1\}$. The classes $\{0\}$ and $\{N\}$ are recurrent, while $\{1, 2, \ldots, N-1\}$ is transient.

Proposition 2.22. *Assume S is finite. Let C be a recurrent communication class. Then for each $x, y \in C$, we have*

$$\mathbb{P}\left[\exists \text{ infinitely many } n \text{ such that } X_n = y | X_0 = x\right] = 1.$$

Proof. Fix $x, y \in C$ and assume $X_0 = x$. Since C is a communication class, for each $z \in C$, there exists $n_z \geq 0$ such that $P^{n_z}(z, y) > 0$. Define

$$n := \max_z n_z < \infty, \quad q := \min_z P^{n_z}(z, y) > 0.$$

For $k \in \mathbb{N}$, let

$$E_k = \left\{\exists j \in [n(k-1)+1, nk] \text{ s.t. } X_j = y\right\}$$

be the event that $X_j = y$ for some j in the kth block of length n. Note that for $s_0, \ldots, s_{nk} \in C$, we have

$$\begin{aligned}\mathbb{P}\left[E_{k+1}|X_0 = s_0, \ldots, X_{nk} = s_{nk}\right] &= \mathbb{P}\left[E_{k+1}|X_{nk} = s_{nk}\right] \\ &= \mathbb{P}\left[E_1|X_0 = s_{nk}\right] \geq P^{n_{s_{nk}}}(s_{nk}, y) \geq q.\end{aligned}$$

Let $M, N \in \mathbb{N}$ be such that $M > N$. Note that

$$\begin{aligned}\mathbb{P}\left[E_k \text{ does not occur } \forall k \in \{N, \ldots, M\}\right] &= \mathbb{P}\left[\bigcap_{k=N}^{M} E_k^{\mathsf{c}}\right] = \mathbb{P}\left[E_M^{\mathsf{c}} \middle| \bigcap_{k=N}^{M-1} E_k^{\mathsf{c}}\right] \mathbb{P}\left[\bigcap_{k=N}^{M-1} E_k^{\mathsf{c}}\right] \\ &\leq (1-q)\mathbb{P}\left[\bigcap_{k=N}^{M-1} E_k^{\mathsf{c}}\right] \\ &\vdots \\ &\leq (1-q)^{M-N},\end{aligned}$$

which converges to 0 as $M \to \infty$, with N fixed. Thus

$$\mathbb{P}\left[E_k \text{ does not occur for all } k \geq N\right] = 0.$$

from which

$$\mathbb{P}[\exists j \geq nN \text{ s.t. } X_j = y] = 1,$$

and so

$$\begin{aligned}&\mathbb{P}\left[\exists \text{ infinitely many } n \text{ such that } X_n = y|X_0 = x\right] \\ &= \mathbb{P}\left[\bigcap_n \left\{\exists j \geq nN : X_j = y\right\}\right] = 1,\end{aligned}$$

where the last line follows from continuity of probability measures and the fact that events $\left\{\exists j \geq nN : X_j = y\right\}$ are decreasing in n. □

Remark 2.23. We showed that there exists a $q \in (0,1)$ and $n \geq 1$ such that for each $k \in \mathbb{N}$, and each $x, y \in C$, we have

$$\mathbb{P}\left[X \text{ hits } y \text{ before time } nk|X_0 = x\right] \geq 1 - (1-q)^k.$$

This can be written in the following form: given $j \geq 1$, choose k such that $j \in [nk, n(k+1)]$ and set

$$c := -\frac{\log(1-q)}{n}.$$

Then,

$$\mathbb{P}\left[X \text{ hits } y \text{ before time } j\ |X_0 = x\right] \geq 1 - (1-q)^{j/n} = 1 - e^{-cj}$$

decays exponentially fast in time. ☕

Proposition 2.24. *Assume S is finite and let C be a transient communication class. Then, with probability* 1*, $\{X_n\}$ eventually leaves C and never returns.*

Proof. Similar to the previous proof; see lecture notes for details. □

Remark 2.25. There exists a positive $c > 0$ such that for each $x \in C$ and each $j \in \mathbb{N}$, we have

$$\mathbb{P}\left[\{X_n\} \text{ exits } C \text{ before time } j \mid X_0 = x\right] \geq 1 - e^{-cj}.$$

2.1 Stopping Times and the Strong Markov Property

Definition 2.26 (Stopping Time)**.** A random time $\tau \in \mathbb{N}_0 \cup \{+\infty\}$ is a **stopping time** if for each $n \in \mathbb{N}$, the event $\{\tau = n\}$ is determined by $X_0, \ldots, X_n$.

Remark 2.27. Equivalent definitions of stopping time:

- For each n, the event $\{\tau \leq n\}$ is determined by $X_0, \ldots, X_n$. Equivalent to above since $\{\tau \leq n\} = \cup_{j=0}^{n} \{\tau = j\}$ and $\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}$.
- For each n, the event $\{\tau > n\}$ is determined by $X_0, \ldots, X_n$. Equivalent to above since $\{\tau > n\} = \{\tau \leq n\}^{\mathsf{c}}$.

Example 2.28.

- $\tau = n$ is a non-random stopping time.
- $\tau = \min\{n \geq 0 : X_n = x\}$ for some $x \in S$ is a stopping time, since $\{\tau \leq n\} = \{\exists j \leq n \text{ s.t. } X_j = x\}$ is determined by $X_0, \ldots, X_n$.
- $\tau :=$ the k^{th} time such that $X_j \in A$ for fixed $k \in \mathbb{N}$ and $A \subset S$ is a stopping time.
- The minimum of two stopping times is a stopping time, since

$$\{\min\{\tau_1, \tau_2\} \leq n\} = \{\tau_1 \leq n\} \cup \{\tau_2 \leq n\}.$$

Example 2.29 (Non Example). Let $N \in \mathbb{N}$ and $x \in S$. Let

$$\tau := \text{last } n \leq N \text{ such that } X_n = x.$$

This is not a stopping time since if we see only $X_0, \ldots, X_n$ for some $n \leq N-1$, we cannot tell whether we visit x between time n and N.

Let $\{X_n\}$ be a Markov chain with a countable state space S. For each $x_0, \ldots, x_n \in S$ and $y_1, \ldots, y_m \in S$, we have

$$\begin{aligned}
&\mathbb{P}\left[X_{n+1} = y_1, \ldots, X_{n+m} = y_m | X_0 = x_0, \ldots, X_n = x_n\right] \\
&= \mathbb{P}\left[X_{n+1} = y_1, \ldots, X_{n+m} = y_m | X_n = x_n\right] \\
&= \mathbb{P}(x_n, y_1) \prod_i P(y_{i-1}, y_i).
\end{aligned}$$

It turns out that this property works also for random stopping times.

Theorem 2.30 (The Strong Markov Property)**.** *Let τ be a stopping time. Let $n \geq 0$, $m \geq 1$, $x_0, \dots, x_n \in S$ be such that $\mathbb{P}[X_0 = x_0, \dots, X_\tau = x_n] > 0$. Then,*

$$\begin{aligned}&\mathbb{P}[X_{\tau+1} = y_1, \dots, X_{\tau+m} = y_m | X_0 = x_0, \dots, X_\tau = x_n] \\ &= \mathbb{P}[X_{\tau+1} = y_1, \dots, X_{\tau+m} = y_m | X_\tau = x_n].\end{aligned}$$

Proof. Note that

$$\begin{aligned}\{X_0 = x_0, \dots, X_\tau = x_n\} &= \{\tau = n\} \cap \{X_0 = x_0, \dots, X_n = x_n\} \\ &= \{X_0 = x_0, \dots, X_n = x_n\},\end{aligned}$$

where the last equality follows since the event $\{\tau = n\}$ is determined by $X_0, \dots, X_n$. Thus,

$$\begin{aligned}&\mathbb{P}[X_{\tau+1} = y_1, \dots, X_{\tau+m} = y_m | X_0 = x_0, \dots, X_\tau = x_n] \\ &= \mathbb{P}[X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_0 = x_0, \dots, X_n = x_n] \\ &= \mathbb{P}[X_{n+1} = y_1, \dots, X_{n+m} = y_m | X_n = x_n],\end{aligned}$$

where the last equality following from the Markov property. ◻

Example 2.31. Let $x \in S$ and define $\tau := \min\{n \geq 0 : X_n = x\}$. Assume further that $\mathbb{P}(\tau < \infty) = 1$. For each $y_1, \dots, y_m \in S$ and $x_0, \dots, x_n \in S$ such that $\mathbb{P}(X_0 = x_0, \dots, X_\tau = x_n) > 0$. Note that $x_n = x$ by definition of τ. The strong Markov property gives

$$\mathbb{P}(X_{\tau+1} = y_1, \dots, X_{\tau+m} = y_m | X_0 = x_0, \dots, X_\tau = x_n) = \mathbb{P}(X_1 = y_1, \dots, X_m = y_m | X_0 = x).$$

Proposition 2.32. *Suppose $X_0 = x \in S$ and assume*

$$\mathbb{P}[\{X_n\} \text{ visits } x \text{ infinitely often}] = 1.$$

Let τ_k be the k^{th} time n such that $X_n = x$, and set $\tau_0 = 0$. Then the increments $\left\{(X_{\tau_k}, \dots, X_{\tau_{k+1}})\right\}_{k \in \mathbb{N}_0} \in \cup_{j=1}^{\infty} S^j$ are iid.

Proof. Observe first that each τ_k is a stopping time and $X_{\tau_k} = x$. This implies that the conditional distribution of $\{X_{\tau_k + j}\}_{j \geq 0}$ given everything before time τ_k is the same as the distribution of $\{X_j\}_{j \geq 0}$ given $X_0 = x$.

Observe also that $\tau_{k+1} - \tau_k$ is the first time $j \geq 1$ such that $X_{\tau_k + j} = x$. Thus the conditional distribution of $\left(X_{\tau_k}, \dots, X_{\tau_k + j}\right)$ given everything before time τ_k is the same as the distribution of $(X_0, \dots, X_{\tau_1})$, which implies that the increments are iid. ◻

Definition 2.33. We say a random variable M in $\mathbb{N}$ has the **geometric distribution** with **success probability** $p \in (0, 1)$ if

$$\mathbb{P}(M = m) = p(1-p)^{m-1}, \quad \forall m \geq 1.$$

Proposition 2.34. *Suppose $X_0 = x \in S$ and $\mathbb{P}(X_n$ visits x infinitely often$) = 1$. Let y be a state such that $x \leftrightarrow y$. Let M be the number of times we visit x before visiting y. Then M has a geometric distribution.*

Proof. Let τ_k be the k^{th} time n such that $X_n = x$, and set $\tau_0 = 0$. Then $\{(X_{\tau_k}, \dots, X_{\tau_{k+1}})\}_{k \geq 0}$ are iid. Now, M is the smallest k such that $(X_{\tau_k}, \dots, X_{\tau_{k+1}})$ visits y and is hence a geometric random variable. ◻

2.2 Periodicity

Let $\{X_n\}$ be a Markov chain and the state space S be countable.

Definition 2.35. For a state $x \in S$, the **period** of x is

$$d(x) \coloneqq \gcd(J_x), \quad \text{where } J_x \coloneqq \{n \geq 1 : P^n(x,x) > 0\}.$$

Note that if $P(x,x) > 0$, then $d(x) = 1$.

Example 2.36. A graph G is said to be **bipartite** if $V(G) = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$, V_1, V_2 nonempty, and every edge goes from V_1 to V_2.

Example 2.37. $\mathbb{Z}$ is bipartite with $V_1 = \{\text{odd}\}$ and $V_2 = \{\text{even}\}$.

Remark 2.38. For a random walk on a connected graph G, we always have $2 \in J_x$. It turns out that bipartite graphs are precisely those for which any path starting and ending at the same vertex must have even length.

- If G is connected and bipartite, then $d(x) = 2$ for the random walk on G for each $x \in V(G)$.
- If G is connected and not bipartite, then $d(x) = 1$ for each $x \in V(G)$, since J_x contains both 2 and an odd number.

Lemma 2.39 (J_m is closed under addition)**.** *If $n, m \in J_m$, then $n + m \in J_x$.*

Proof. We have $P^{n+m}(x,x) \geq P^n(x,x)P^m(x,x) > 0$. □

Proposition 2.40. *The set J_x contains $kd(x)$ for all sufficiently large k.*

Proof. Via previous lemma and Bezout's identity: if $d \coloneqq \gcd(a,b)$, then there exists integers x, y such that $xa + yb = d$. □

Proposition 2.41 (The Period is a Class Property)**.** *If $x \leftrightarrow y$, then $d(x) = d(y)$.*

Proof. Choose n, m such that $P^n(x,y) > 0$ and $P^m(y,x) > 0$. This implies that $P^{n+m}(x,x) \geq P^n(x,y)P^m(y,x) > 0$. Similarly $P^{n+m}(y,y) > 0$. This gives $n + m \in J_x \cap J_y$ and so $d(x)$ and $d(y)$ both divide $n + m$. Assume for contradiction that $d(x) < d(y)$. Then there exists $k \in J_x$ not divisible by $d(y)$. Observe that

$$P^{n+m+k}(y,y) \geq P^m(y,x)P^k(x,x)P^n(x,y) > 0,$$

implying $n + m + k \in J_y$. Since however that $n + m \in J_y$, we know $d(y)$ divides the difference k, a contradiction. □

Definition 2.42. We say a Markov chain $\{X_n\}$ is **aperiodic** if $d(x) = 1$ for all $x \in S$.

This is often a "nice" property.

Remark 2.43. The propositions above imply the following:

- If $\{X_n\}$ is irreducible (i.e., any two state communicate), then to check aperiodicity, it suffices to check $d(x) = 1$ for *any* $x \in S$.
- If $\{X_n\}$ is aperiodic, then $P^n(x,x) > 0$ for any large enough n (depending on x).
- For any Markov chain $\{X_n\}$, we can construct the following "lazy" Markov chain $\{\tilde{X}_n\}$ with transition matrix

$$\tilde{P}(x,y) := \begin{cases} \frac{1}{2}P(x,y), & x \neq y \\ \frac{1}{2} + \frac{1}{2}P(x,x), & x = y \end{cases},$$

which is always aperiodic since $\tilde{P}(x,x) > 0$ for each $x \in S$.

☕

2.3 Stationary Distribution, Finite State Space

Definition 2.44. Let $\pi : S \longrightarrow [0,1]$ be such that $\sum_{x\in S} \pi_x = 1$. We say that π is a **stationary (or invariant) distribution** for the Markov chain $\{X_n\}$ if $\pi_y = \sum_{x\in S} \pi_x P(x,y)$ for each $y \in S$.

Remark 2.45 (Equivalent definitions). Suppose $S = \{1,\dots,N\}$ and $\pi = (\pi_1,\dots,\pi_N)$ is a row vector.

(i) Then π is a stationary distribution if and only if $\sum_j \pi_j = 1$ and $\pi P = \pi$. That is, π is a stationary distribution if and only if it is a left eigenvector of P with eigenvalue 1.

(ii) If $\mathbb{P}(X_0 = x) = \pi_x$ for each $x \in S$, then $\mathbb{P}(X_1 = y) = \pi_y$.

To see the equivalence, note that if $X_1 \sim \pi$, then $\mathbb{P}(X_1 = y) = \sum_x \pi_x P(x,y)$ equals π_y if and only if π is a stationary distribution. Induction gives the following:

(iii) If $\mathbb{P}(X_0 = x) = \pi_x$ for all $x \in S$, then $X_n \sim \pi$ for each $n \in \mathbb{N}$.

☕

Recall the following:

- $\{X_n\}$ is irreducible if $\forall x,y \in S$ there exists $n \in \mathbb{N}_0$ such that $P^n(x,y) > 0$.
- $\{X_n\}$ is aperiodic if $\forall x \in S$, $\gcd(n \geq 1; P^n(x,x) > 0) = 1$.
- Period is a class property: if $x \leftrightarrow y$, then $d(x) = d(y)$.

Theorem 2.46. *Suppose S is finite and $\{X_n\}$ is irreducible and aperiodic. Then there exists a unique stationary distribution π. Moreover, for each $x, y \in S$, we have*

$$\lim_{n\to\infty} P^n(x,y) = \pi_y.$$

Remark 2.47.

- A probabilistic proof will be given below, though a linear algebra proof (using the Perron-Frobenius theorem) is also possible.
- π_x is the long-run proportion of time that the Markov chain spends in state x. This is useful for sampling algorithms.

  Example 2.48. Suppose we have a probability distribution π on a large state space S. We may find a Markov chain $\{X_n\}$ on S whose stationary distribution is π and run it for a long time to approximately sample from π.

Example 2.49. Let $S = \{1, 2, 3\}$ and set

$$P := \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \end{pmatrix}.$$

To find the stationary distribution, we need to solve $\pi P = \pi$ and $\sum_x \pi_x = 1$. This gives $\pi = (3/10, 1/5, 1/2)$.

Proposition 2.50 (The Stationary Distribution Exists)**.** *Suppose S is finite and $\{X_n\}$ is irreducible. Assume $X_0 = z \in S$ and set $T = \min\{n \geq 1 : X_n = z\}$. For $x \in S$, let $\tilde{\pi}(x) := \mathbb{E}\left[\#\{n \in \{0, \dots, T-1\} : X_n = x\}\right]$. Let $\pi_x := \tilde{\pi}_x/\mathbb{E}[T]$. Then π is a stationary distribution of $\{X_n\}$.*

Proof. Note first that

$$\sum_{x \in S} \tilde{\pi}_x = \mathbb{E}\left[\#n \in \{0, \dots, T-1\} : X_n \in S\right] = \mathbb{E}[T],$$

which gives $\sum \pi_x = 1$. It remains thus to show

$$\tilde{\pi}_y = \sum \tilde{\pi}_x P(x, y), \quad \forall y \in S.$$

Note that

$$\tilde{\pi}_x = \mathbb{E}\left[\sum_{n=0}^{T-1} \mathbb{1}_{\{X_n = x\}}\right] = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = x, T > n\}}\right] = \sum_{n=0}^{\infty} \mathbb{P}\left[X_n = x, T > n\right],$$

where the exchange of sum and expectation is justified since the summands are nonrandom. Now note that

$$\begin{aligned}
\sum_x \tilde{\pi}_x P(x, y) &= \sum_{x \in S} \sum_{n=0}^{\infty} \mathbb{P}[X_n = x, T > n] P(x, y) \\
&= \sum_{n=0}^{\infty} \sum_{x \in S} \mathbb{P}[X_n = x, T > n] \mathbb{P}\left[X_{n+1} = y | X_x = x, T > n\right] \\
&= \sum_{n=0}^{\infty} \sum_{x \in S} \mathbb{P}[X_{n+1} = y, X_n = x, T > n] = \sum_{n=0}^{\infty} \mathbb{P}[X_{n+1} = y, T > n],
\end{aligned}$$

where the second equality follows from the fact that T is a stopping time and depends only on $X_0, \ldots, X_n$. If $y \neq z$, the above can be simplified to

$$\begin{aligned}\sum_{x \in S} \tilde{\pi}_x P(x, y) &= \sum_{n=0}^{\infty} \mathbb{P}[X_{n+1} = y, T > n + 1] = \sum_{m=1}^{\infty} \mathbb{P}[X_m = y, T > m] \\ &= \tilde{\pi}_y - \mathbb{P}[T > 0, X_0 = y] = \tilde{\pi}_y,\end{aligned}$$

where the last equality follows since $X_0 = z \neq y$. If $y = z$, we have

$$\sum_{x \in S} \tilde{\pi}_x P(x, z) = \sum_{n=0}^{\infty} \mathbb{P}\left[T = n + 1\right] = 1 = \tilde{\pi}_z.$$

□

Proposition 2.51 (The Stationary Distribution is Unique)**.** *Let $\{X_n\}$ be irreducible and aperiodic with finite state space S. If π is a stationary distribution, then for each $x, y \in S$, we have $\lim_{n\to\infty} P^n(x, y) = \pi_y$. In particular, π is unique.*

Proof. We consider a coupling of the original Markov chain starting at x and the same Markov chain starting from $X_0 \sim \pi$. We will show that with probability 1, the two chains meet at some finite time and then move together forever after.

Specifically, consider the Markov chain (X_n, Y_n) on $S \times S$ with transition probability given by

$$\overline{P}\left((x, y), (x', y')\right) \coloneqq \begin{cases} P(x, y)P(x', y'), & x \neq y \\ P(x, x'), & x = y, x' = y' \\ 0, & \text{otherwise} \end{cases}.$$

We check that this is indeed a coupling:

$$\begin{aligned} &\mathbb{P}\left[X_1 = x' | X_0 = x, Y_0 = y\right] = \mathbb{P}(x, x') \sum_{y' \in S} P(y, y') = P(x, x'), && \text{if } x \neq y \\ &\mathbb{P}(X_1 = x' | X_0 = x, Y_0 = y) = P(x, x'), && \text{if } x = y \end{aligned}$$

A similar calculation holds for Y_1. Now let $\tau \coloneqq \min\{n \geq 0 : X_n = Y_n\}$. By definition, we know that $X_n = Y_n$ for each $n \geq \tau$.

Claim 2.52. $\mathbb{P}\left[\tau < \infty | X_0 = x, Y_0 = y\right] = 1$ for arbitrary $x, y \in S$.

Taking this claim as given for now, we consider the initial distribution where $X_0 = x \in S$ and $Y_0 \sim \pi$. Since π is a stationary distribution, we know that $Y_n \sim \pi$ for each n. Since also $X_n = Y_n$ for any large enough n, we have $\lim_{n\to\infty} \mathbb{P}[X_n = Y_n] = 1$. Now,

$$\lim_{n\to\infty} \left(\mathbb{P}\left[X_n = y | X_0 = x\right] - \pi_y\right) = \lim_{n\to\infty} \left(\mathbb{P}[X_n = y | X_0 = x] - \mathbb{P}[Y_n = y]\right) = 0.$$

We now return to the proof of Claim 2.52: We consider $\{\tilde{X}_n\}$ and $\{\tilde{Y}_n\}$, two independent copies of the original Markov chain with $\tilde{X}_0 = x$ and $\tilde{Y}_0 = y$. It suffices to show that

$$\mathbb{P}\left\{\exists n : \tilde{X}_n = \tilde{Y}_n\right\} = 1.$$

It suffices to show that $(\tilde{X}_n, \tilde{Y}_n)$ is an irreducible Markov chain on $S \times S$, since in that case it will visit the diagonal $\{(z,z) : z \in S\}$ infinitely often with probability 1 by irreducibility and finiteness of S.

Recall that aperiodicity implies that there exists a $k_0 \in \mathbb{N}$ such that $P^k(x,x) > 0$ for each $k \geq k_0$ and $x \in S$. Irreducibility implies that for each $x, x', y, y' \in S$, there exists n such that $P^n(x,x') > 0$. Furthermore, there exists a $m \geq n + k_0$ such that $P^m(y,y') > 0$. Since $\tilde{X}_m$ and $\tilde{Y}_m$ are independent by definition, we see that

$$\begin{aligned}
\mathbb{P}\left\{(\tilde{X}_m, \tilde{Y}_m) = (x', y') \middle| (\tilde{X}_0, \tilde{Y}_0) = (x, y)\right\} &= P^m(x,x')P^m(y,y') \\
&\geq P^{m-n}(x,x)P^n(x,x')P^m(y,y') > 0,
\end{aligned}$$

where the second line follows from $m - n \geq k_0$. This implies that $(\tilde{X}_n, \tilde{Y}_n)$ is irreducible. Thus with probability 1, there exists n such that $(\tilde{X}_n, \tilde{Y}_n) = (x, x)$. □

Example 2.53. The stationary distribution for the random walk on a finite connected nonbipartite graph G is given by

$$\pi_x = \frac{\deg(x)}{2\#E}, \quad x \in V(G).$$

We check that π_x is a distribution:

$$\sum_{x \in V} \pi_x = \frac{1}{2\#E} \sum_{x \in V} \deg(x) = 1.$$

And that it is stationary: for $y \in V$,

$$\sum_{x \in V} \pi_x P(x,y) = \sum_{x \sim y} \frac{\deg(x)}{2\#E} \cdot \frac{1}{\deg(x)} = \frac{1}{2\#E} \sum_{x \sim y} 1 = \frac{\deg(y)}{2\#E} = \pi_y.$$

3 Markov Chains with Countable State Space

Now consider the case where S is countably infinite.

3.1 Reducibility and Recurrence

Definition 3.1. We say $\{X_n\}$ is **irreducible** if for each $x, y \in S$, there exists $n \geq 0$ such that $P^n(x, y) > 0$.

Definition 3.2. We say a state $x \in S$ is **recurrent** if $\mathbb{P}[\exists$ infinitely many $n : X_n = x | X_0 = x] = 1$, and **transient** otherwise.

A particular difference of the case of countable state space is that irreducibility no longer imply recurrence, as we will see.

Proposition 3.3. *If $\{X_n\}$ is irreducible, then either all states are recurrent or all states are transient. In particular, it makes sense to say that $\{X_n\}$ is **recurrent** or **transient**.*

Proof. Assume first there exists recurrent state x. Let $\tau_0 = 0$ and set

$$\tau_k := \text{k}^{\text{th}} \text{ smallest } n \text{ such that } X_n = x.$$

By assumption, we have $\mathbb{P}(\tau_k < \infty | X_0 = x) = 1$ for each k. By the strong Markov property, we have

$$\left(X_{\tau_k}, \ldots, X_{\tau_{k+1}}\right) \in \bigcup_{j \in \mathbb{N}} S^j$$

are iid. Let $y \in S$. Since $\{X_n\}$ is irreducible, there exists n such that $P^n(x, y) > 0$. Thus there exists k such that $\mathbb{P}(y \in \{X_{\tau_k}, \ldots, X_{\tau_{k+1}}\}) > 0$. Since the intervals are iid, we have $q := \mathbb{P}(y \in \{X_{\tau_k}, \ldots, X_{\tau_{k+1}}\}) > 0$ for each k. The events

$$\left\{y \in \{X_{\tau_k}, \ldots, X_{\tau_{k+1}}\}\right\}$$

are iid, each with probability q. Thus with probability 1, infinitely many of these events occur, which implies that y is recurrent. That is,

$$\mathbb{P}[\exists \text{ infinitely many } n \text{ such that } X_n = y | X_0 = x] = 1.$$

Let $\sigma := \min\{n \geq 0 : X_n = y\}$. We know $\mathbb{P}[\sigma < \infty | X_n = k] = 1$. By the strong Markov property, $\{X_{\sigma+j}\}_{j \geq 0}$ has the same distribution as $\{X_j\}_{j \geq 0}$ given $X_0 = y$. This implies that the Markov chain starting at y visits x infinitely often with probability 1. □

Proposition 3.4. *A state x is recurrent if and only if $\sum_{n=0}^{\infty} P^n(x, x) = \infty$. Moreover, if x is transient, then with probability 1, the Markov chain $\{X_n\}$ visits x only finitely many times. Thus $\mathbb{P}(\exists$ infinitely many $n : X_n = x | X_0 = x)$ is either 0 or 1.*

Proof. Let $R_x := \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n = x\}}$ be the number of times $\{X_n\}$ visits x. Note that

$$\mathbb{E}[R_x | X_0 = x] = \sum_{n=0}^{\infty} \mathbb{P}[X_n = x | X_0 = x] = \sum_{n=0}^{\infty} P^n(x, x).$$

If $\sum_{n=0}^{\infty} P^n(x,x) < \infty$, then $\mathbb{E}[R_x] < \infty$, which implies that $R_x < \infty$ with probability 1.

For the converse. Assume x is a transient state. We will show that $\sum_{x=0}^{\infty} P^n(x,x) < \infty$. Let $\tau_0 = 0$ and set τ_k to be the kth time such that $X_n = x$. Since x is transient, there exists k such that $\mathbb{P}(\tau_k = \infty | X_0 = x) > 0$. By the strong Markov property, we have for each k that

$$\mathbb{P}\left[\tau_{k+1} - \tau_k = \infty | \tau_k < \infty\right] =: q > 0.$$

Note that the number of visits to x is $R_x := \min\{k : \tau_{k+1} = \infty\}$. Thus R_x has a geometric distribution with success probability q. In particular, $\mathbb{E}[R_x | X_0 = x] = 1/q < \infty$, which gives $\sum_{n=0}^{\infty} P^n(x,x) < \infty$. Note that this also implies that with probability 1, the Markov chain visits x only finitely many times. □

Note that in light of the previous two propositions, if $\{X_n\}$ is irreducible, to check recurrence, it suffices to check if $\sum_{n=0}^{\infty} P^n(x,x)$ is infinite for some $x \in S$.

Example 3.5. Let $S = \mathbb{N}_0$ and set

$$P(x,0) = \frac{1}{x+2}, \quad P(x,x+1) = 1 - \frac{1}{x+2}, \quad \forall x \geq 0.$$

This Markov chain is irreducible. To check recurrence, we compute $\sum P^n(0,0)$. Assume $X_0 = 0$ and note that $X_n \leq n$. Thus $P^n(0,0) = \mathbb{P}\left[X_n = 0 | X_0 = 0\right] \geq 1/(n+1)$, which gives $\sum P^n(0,0) = \infty$. Thus the Markov chain is recurrent.

Alternatively, let $\tau := \min\{n \geq 1 : X_n = 0\}$. It suffices to show that $\mathbb{P}[\tau < \infty] = 1$. To do this, we need only show $\lim_{N\to\infty} P(\tau > N) = 0$. Note that $\tau > N$ if and only if the first N steps are upward. Thus

$$\mathbb{P}\left[\tau > N\right] = \prod_{k=0}^{N-1} \left(1 - \frac{1}{k+2}\right).$$

Taking logs, we have

$$\log \mathbb{P}[\tau > N] = \sum_{k=0}^{N-1} \log\left(1 - \frac{1}{k+2}\right) = -\sum_{k=0}^{N-1} \left(\frac{1}{n+2} + O\left(\frac{1}{(n+2)^2}\right)\right) \longrightarrow -\infty.$$

This implies that $\mathbb{P}[\tau > N] \to e^{-\infty} = 0$ as $N \to \infty$.

3.2 Biased Random Walk

Proposition 3.6 (Biased Random Walk)**.** *Consider the biased random walk on* $\mathbb{Z}$ *with*

$$P(x,x+1) = p, \quad P(x,x-1) = 1-p, \quad \forall x \in \mathbb{Z},$$

Let $N \geq 1$. *For* $x \in \{0,\dots,N\}$, *we have*

$$\mathbb{P}\left[\{X_n\} \text{ hits } N \text{ before } 0 | X_0 = x\right] = \begin{cases} \frac{\left(\frac{1-p}{p}\right)^x - 1}{\left(\frac{1-p}{p}\right)^N - 1}, & p \neq \frac{1}{2}; \\ \frac{x}{N}, & p = \frac{1}{2}. \end{cases}$$

Proof. Let $\alpha(x) := \mathbb{P}\left[\{X_n\} \text{ hits } N \text{ before } 0 \,|X_0 = x\right]$. Note that $\alpha(0) = 0$, $\alpha(N) = 1$, and for each $x \in \{1, \ldots, N-1\}$, we have

$$\alpha(x) = p\alpha(x+1) + (1-p)\alpha(x-1).$$

This gives a system of $N+1$ equations in $N+1$ unknowns.

When $p = 1/2$, we have

$$\alpha(x) = \frac{\alpha(x+1) + \alpha(x-1)}{2}, \quad x \in \{1, \ldots, N-1\}.$$

Thus $\alpha(x)$ is affine in x. Suppose $\alpha(x) = \alpha + \beta x$ and plugging in the boundary conditions gives $\alpha = 0$ and $\beta = 1/N$, as desired.

When $p \neq 1/2$, we use the ansatz $\alpha(x) = b^x$. Plugging this in gives

$$b^x = pb^{x+1} + (1-p)b^{x-1} \iff pb^2 - b + (1-p) = 0.$$

Solving this quadratic gives roots $b = 1$ and $b = (1-p)/p$. The general solution is thus

$$\alpha(x) = c_1 + c_2 \left(\frac{1-p}{p}\right)^x.$$

Plugging in the boundary conditions gives

$$c_1 = -c_2 = \left(1 - \left(\frac{1-p}{p}\right)^N\right)^{-1}.$$

□

Corollary 3.7. *For each $x \geq 1$*

$$\mathbb{P}\left[\{X_n\} \textit{ hits } 0 | X_0 = x\right] = \begin{cases} 1, & p \leq \frac{1}{2}; \\ \left(\frac{1-p}{p}\right)^x, & p > \frac{1}{2}. \end{cases}.$$

Proof. Send $N \to \infty$ in the previous proposition. □

Corollary 3.8. *A biased random walk is recurrent if $p = 1/2$ and transient otherwise.*

Proof. Note first that the biased random walk is irreducible. If $p > 1/2$, then $\{X_n\}$ has a positive chance to never hit 0 and so is transient. If $p < 1/2$, then $\{-X_n\}$ is a biased random walk with parameter $1 - p > 1/2$ and so is transient. If $p = 1/2$, then $\{X_n\}$ has by the previous corollary a probability 1 of hitting 0 no matter where it starts, and so is recurrent. □

3.3 A Queuing Model

At each time $n \geq 1$, the following events occur independently from each other and from the past, in the following order:

- With probability q, if there is at least one person, then one person is served and leaves the queue.
- With probability p, a new person arrives and joins the queue.

Let X_n denote the number of people in queue at time n. Note that $\{X_n\}$ is a Markov chain on $S := \mathbb{N}_0$ with transition probabilities

$$P(0,1) = p, \quad P(0,0) = 1-p,$$
$$P(x,x-1) = q(1-p), \quad P(x,x+1) = p(1-q), \quad \forall x \geq 1.$$

Observe that this Markov chain is irreducible.

Proposition 3.9. *$\{X_n\}$ is recurrent if $q \geq p$.*

Proof. We will reduce to the biased random walk. Let τ_k be the k^{th} time such that $X_{n-1} \neq X_n$. Note that τ_k is a stopping time, and thus we have

$$\begin{aligned}\mathbb{P}\left[X_{\tau_k} = x+1 \middle| X_{\tau_{k-1}} = x\right]&\mathbb{P}\left[X_{\tau_k} = x+1 \middle| X_{\tau_k - 1} = x\right] \\ &= \mathbb{P}\left[X_1 = x+1 | X_0 = x, X_1 \neq X_0\right] \\ &= \frac{P(x,x+1)}{1-P(x,x)} = \frac{p(1-q)}{p(1-q)+q(1-p)}.\end{aligned}$$

Since

$$\mathbb{P}\left[X_{\tau_k} = x+1 \middle| X_{\tau_k} = x\right] = \mathbb{P}\left[X_{\tau_k} = x+1 \middle| X_{\tau_k - 1} = x\right],$$

and $\tau_k - 1$ is not a stopping time, we cannot omit the extra conditioning above. The calculation also implies

$$\mathbb{P}\left[X_{\tau_k} = x+1 \middle| X_{\tau_k} = x\right] = 1 - \frac{p(1-q)}{p(1-q)+q(1-p)}.$$

Thus $\{X_{\tau_k}\}_{k\geq 0}$ is a biased random walk with parameter

$$\frac{p(1-q)}{p(1-q)+q(1-p)}$$

until it hits 0. Thus $\{X_{\tau_k}\}$ eventually hits 0 with probability 1 if and only if

$$\frac{p(1-q)}{p(1-q)+q(1-p)} \leq \frac{1}{2} \iff q \geq p,$$

and so $\{X_n\}$ has probability 1 of eventually hitting 0 if and only if $q \geq p$. $\square$

3.4 Stationary Distribution

Definition 3.10. Let S be countable. We say $\pi : S \to [0,1]$ is a **stationary distribution** for the Markov chain $\{X_n\}$ if $\sum_{x\in S} \pi_x = 1$ and for each $y \in S$, we have $\sum_{x\in S} \pi_x P(x,y) = \pi_y$.

Recall that for a *finite* S, there exists a unique stationary distribution if $\{X_n\}$ is irreducible and aperiodic. When S is countably infinite, we need stronger assumptions.

Proposition 3.11. *If $\{X_n\}$ is irreducible and transient, then there does not exist a stationary distribution.*

Proof. Then for each $x, y \in S$,

$$\mathbb{P}\left[\{X_n\} \text{ visits } y \text{ finitely many times } | X_0 = x\right] = 1.$$

This implies that $P^n(x, y) \to 0$ as $n \to \infty$. Now if π is a stationary distribution, then for each $n \in \mathbb{N}$,

$$\pi_y = \sum_{x \in S} \pi_x P^n(x, y) \longrightarrow 0 \text{ as } n \to \infty,$$

which implies that $\pi_y = 0$ for each $y \in S$, a contradiction. □

Definition 3.12. Assume $\{X_n\}$ is irreducible and recurrent. We say $\{X_n\}$ is **null recurrent** if $\lim_{n\to\infty} P^n(x, y) = 0$ for each $x, y \in S$. We say $\{X_n\}$ is **positive recurrent** otherwise.

Note that the same argument above shows that null recurrent Markov chains do not have a stationary distribution.

Proposition 3.13. *Assume $\{X_n\}$ is irreducible. Then the following are equivalent:*

(i) $\{X_n\}$ *is positive recurrent.*

(ii) $\{X_n\}$ *has a stationary distribution.*

(iii) $\limsup_{n\to\infty} P^n(x, y) > 0$ *for all (or equivalently, some)* $x, y \in S$.

(iv) If $T_x \coloneqq \min\{n \geq 1 : X_n = x\}$. *Then* $\mathbb{E}[T_x | X_0 = x] < \infty$ *for all* $x \in S$.

Furthermore, if $\{X_n\}$ is aperiodic and positively recurrent, then the stationary distribution is unique and

$$\pi_y = \lim_{n\to\infty} P^n(x, y) = \frac{1}{\mathbb{E}\left[T_y \middle| X_0 = y\right]}, \quad \forall x, y \in S.$$

Proof. Basically the same as the finite case. □

We summarize the characterization of recurrence in the following proposition: If $\{X_n\}$ is irreducible, then for any $x \in S$, we have

(i) Transient $\iff \sum P^n(x, x) < \infty$.

(ii) Null recurrent $\iff \sum P^n(x, x) = \infty$ and $P^n(x, x) \to 0$.

(iii) Positive recurrent $\iff \limsup P^n(x, x) > 0 \iff \mathbb{E}[T_x | X_0 = x] < \infty \iff$ there exists a stationary distribution.

Example 3.14 (Biased Random Walk with Partially Reflecting Boundary). Consider the based random walk on $\mathbb{N}_0$ with partially reflected boundary:

$$\begin{aligned} P(0,0) &= 1-p, & P(0,1) &= p, & \\ P(x,x-1) &= 1-p, & P(x,x+1) &= p, & \forall x \geq 1. \end{aligned}$$

It is clear that this Markov chain is irreducible. Since $P(0,0) = 1-p$, it is also aperiodic.

It is positive recurrent, null recurrent, or transient?

(i) If $p > 1/2$, it is transient since it has a positive probability of never hitting 0 if $X_0 \geq 1$.

(ii) If $p \leq 1/2$, we know that it is recurrent. To see if it is null or positive recurrent, we try to find a stationary distribution π. Recall that π must satisfy

$$\begin{aligned} \pi_0 &= (1-p)\pi_0 + p_1\pi_1, \\ \pi_y &= p\pi_{y-1} + (1-p)\pi_{y+1}, \quad \forall x \geq 1. \end{aligned}$$

If $p < 1/2$, it turns out that the general solution to the above is given by

$$\pi_y = c_1 + c_2\left(\frac{p}{1-p}\right)^y, \quad y \geq 1.$$

Since also $\sum_y \pi_y = 1$, we have $c_1 = 0$ and

$$1 = c_2 \sum \left(\frac{p}{1-p}\right)^y = \frac{c_2}{1-\frac{p}{1-p}}.$$

Thus $c_2 = 1 - p/(1-p)$ and

$$\pi_y = \left(1 - \frac{p}{1-p}\right)\left(\frac{p}{1-p}\right)^y$$

is the stationary distribution, and so the Markov chain is positive recurrent.

When $p = 1/2$, the general solution is given by $\pi_y = c_1 + c_2 y$. Since $\sum_y \pi_y = 1$, we must have $c_1 = c_2 = 0$, which is not a distribution. Thus the Markov chain is null recurrent.

3.5 Random Walk on the Integer Lattice

Consider the random walk on $\mathbb{Z}^d$. We view $\mathbb{Z}^d$ as a graph where $x, y \in \mathbb{Z}^d$ are joined by an edge if and only if $|x-y| = 1$.

Theorem 3.15. *The random walk on $\mathbb{Z}^d$ is null recurrent if $d = 1, 2$ and transient if $d \geq 3$.*

Proof. Consider first $d = 1$. Since $\{X_n\}$ is irreducible, we need only consider $X_0 = 0$. If n is odd, $X_n \neq 0$ since $\mathbb{Z}$ is bipartite. For each n, $X_{2n} = 0$ if and only if there are n positive steps and n negative steps in the first $2n$ steps. Thus

$$\mathbb{P}\left[X_{2n} = 0\right] = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}.$$

To determine whether the expression above is summable, we recall **Stirling's formula**:

$$n! \sim \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n},$$

where $a_n \sim b_n$ means $a_n/b_n \to 1$ as $n \to \infty$. This approximation gives

$$\begin{aligned}
\mathbb{P}\left(X_{2n} = 0\right) &\sim \frac{\sqrt{2\pi}(2n)^{2n+\frac{1}{2}} e^{-2n}}{\left(\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}\right)^2} \cdot 2^{-2n} \\
&= \frac{1}{\sqrt{2\pi}} 2^{2n+\frac{1}{2}} n^{-\frac{1}{2}} 2^{-2n} = \frac{1}{\sqrt{\pi n}}.
\end{aligned}$$

From this we know that $P^{2n}(0,0) \to 0$ and so $\{X_n\}$ is not positive recurrent. But since $\sum_{n\geq 1} \frac{1}{\sqrt{\pi n}} = \infty$, we know that $\{X_n\}$ is recurrent. Thus we can conclude that $\{X_n\}$ is null recurrent.

Now consider the case $d \geq 2$. Since there are d components, by the law of large numbers, in $2n$ steps, there will be around $2n/d$ steps in each component. For $k = 1, \ldots, d$, each step in the k^{th} component is $+1$ with probability $1/2$ and -1 with probability $1/2$. By the $d = 1$ case,

$$\mathbb{P}\left[k\text{th component of } X_{2n} \text{ is } 0\right] = \frac{1}{\sqrt{\pi(2n/d)}}.$$

Thus

$$\mathbb{P}\left[X_{2n} = 0\right] = \left(\frac{d}{2\pi n}\right)^{d/2} = \text{constant} \cdot n^{-d/2},$$

which goes to 0 as $n \to \infty$ and is summable if $d \geq 3$ (thus transient) and not summable if $d = 2$ (thus null recurrent). $\square$

4 Branching (or Galton-Watson) Processes

Given the offspring distribution $\{p_k\}_{k\geq 0}$ with $\sum_{k=0}^{\infty} p_k = 1$ (where p_k models the probability of having k children), we define the branching process $\{X_n\}_{n\geq 0}$ (modelling the total number of offspring in each generation) as:

$$X_{n+1} = \sum_{i=1}^{X_n} \xi_j,$$

where ξ_j are conditional independent given X_n and for each $x \geq 0$, $\mathbb{P}\left[\xi_j = k \middle| X_n = x\right] = p_k$.

Remark 4.1. In addition to modelling population growth, branching processes can also be used to model the spread of epidemics, nuclear chain reactions, and the propagation of information in networks.

What is the extinction probability $a := \mathbb{P}[\exists n \geq 1 : X_n = 0 | X_0 = 1]$? We define $\mu := \sum_{k=0}^{\infty} k p_k$ to be the expected number of offspring per individual. Observe that

$$\mathbb{E}[X_{n+1}|X_n = m] = \mathbb{E}\left[\sum_{j=1}^{m} \xi_j \middle| X_n = m\right] = \mu m.$$

Thus $\mathbb{E}[X_{n+1}] = \mu \mathbb{E}[X_n]$, and we have $\mathbb{E}[X_n] = \mu^n \mathbb{E}[X_0]$. Thus we have the following

Proposition 4.2. *If $\mu < 1$, then the extinction probability is $a = 1$.*

Proof. Assume $X_0 = 1$. We have as $n \to \infty$ that $\mathbb{P}[X_n \geq 1] \leq \mathbb{E}[X_n] = \mu^n \to 0$. □

What if $\mu \geq 1$? Observe that if $X_1 = k$, then the descendants of the k individuals are k independent branching processes with the same offspring distribution. From this we see that

$$\mathbb{E}[\text{extinct}|X_1 = k] = a^k.$$

Thus, assuming $X_0 = 1$, we have

$$a := \mathbb{P}[\text{extinction}] = \sum_{k=0}^{\infty} \mathbb{P}[X_1 = k]\, \mathbb{P}[\text{extinction}|X_1 = k] = \sum_{k=0}^{\infty} p_k a^k.$$

Thus $a = \varphi(a)$, where φ is the generating function for $\{p_k\}$, defined as:

Definition 4.3. Let Y be a random variable in $\mathbb{N}_0$. The **generating function** of Y is defined as

$$\begin{aligned} \varphi = \varphi_Y : [0, \infty] &\longrightarrow [0, \infty] \\ s &\longmapsto \mathbb{E}\left[s^Y\right] = \sum_{k=0}^{\infty} \mathbb{P}[Y = k] s^k. \end{aligned}$$

Proposition 4.4 (Properties of Generating Functions)**.**

(i) *We allow $\varphi(s) = \infty$, but note that we have $\varphi(s) < \infty$ for all $s \in [0, 1]$.*

(ii) $\varphi(1) = \sum \mathbb{P}[Y = k] = 1$.

(iii) $\varphi(0) = \mathbb{P}[Y = 0] \cdot 0^0 = \mathbb{P}[Y = 0]$.

(iv) $\varphi'(s) = \sum_{k=0}^{\infty} k\mathbb{P}[Y = k]s^{k-1}$. *Thus* $\varphi'(1) = \mathbb{E}[Y]$.

(v) *If $Y_1, \ldots, Y_m$ are independent, then*

$$\varphi_{Y_1+\cdots+Y_m}(s) = \mathbb{E}\left[\prod s^{Y_j}\right] = \prod \mathbb{E}\left[s^{Y_j}\right] = \prod_{i=1}^{m} \varphi_{Y_i}(s).$$

Proposition 4.5. *Consider the branching process $\{X_n\}$ with offspring distribution $\{p_k\}$. Let φ be the generating function for $\{p_k\}$ and let $\varphi^{(n)} := \underbrace{\varphi \circ \cdots \circ \varphi}_{n \text{ times}}$. Then $\varphi_{X_n}(s) = \varphi^{(n)}(s)$.*

Proof. It suffices to show that $\varphi_{X_{n+1}}(s) = \varphi_{X_n}(\varphi(s))$. Note that

$$\begin{aligned}\varphi_{X_{n+1}}(s) &= \sum_{k\geq 0} \mathbb{P}\left[X_{n+1} = k\right] s^k \\ &= \sum_{k\geq 0}\sum_{j\geq 0} \mathbb{P}\left[X_{n+1} = k | X_n = j\right] \mathbb{P}\left[X_n = j\right] s^k \\ &= \sum_{j\geq 0} \mathbb{P}\left[X_n = j\right] \sum_{k\geq 0} \mathbb{P}\left[X_{n+1} = k | X_n = j\right] s^k.\end{aligned}$$

If $X_n = j$, we have $X_{n+1} = \sum_{i=1}^{j} \xi_i$, where $\{\xi_i\}$ are iid with distribution $\{p_k\}$. Thus $\sum_{k\geq 0} \mathbb{P}[X_{n+1} = k|X_n = j]s^k$ is the generating function of $\sum \xi_i$, and thus can be written as $\varphi(s)^j$. Thus we have

$$\varphi_{X_{n+1}}(s) = \sum_{j\geq 0} \mathbb{P}\left[X_n = j\right] \varphi(s)^j = \varphi_{X_n}(\varphi(s)),$$

□

Proposition 4.6. *Assume $0 < p_0 < 1$. Then the extinction probability a is the smallest positive solution to the equation $s = \varphi(s)$.*

Proof. We know that $\varphi(a) = a$. We first check that there exists a smallest positive solution. Note that $\varphi(s) - s$ as a power series is continuous within its radius of convergence, which in particular includes $[0, 1]$. Thus $\{s \in [0, 1] : \varphi(s) = s\}$ is closed. Since $\varphi(0) = p_0$, we have that $\varphi(0) \neq 0$. Thus there exists a smallest $s_0 := \min\{s \in [0, 1] : \varphi(s) = s\}$.

We claim that $\varphi_{X_n}(0) < s_0$ for any $n \geq 0$. We prove this by induction. Recall that $\varphi_{X_0}(s) = s$, so $\varphi(0) = 0 < S_0$. Assume now that $n \geq 0$ and $\varphi_{X_n}(0) < s_0$. We have

$$\varphi_{X_{n+1}}(0) = \varphi\left(\varphi_{X_n}(0)\right) < \varphi(s_0) = s_0,$$

where the inequality comes from the fact that φ is strictly increasing on $[0,1]$ (since it has nonnegative derivative).

Now since $\varphi_{X_n}(0) = \mathbb{P}[X_n = 0]$, we have

$$a = \lim_{n\to\infty} \mathbb{P}[X_n = 0] \le s_0.$$

Since s_0 is the smallest positive solution for $\varphi(a) = a$, we have $a = s_0$. $\square$

Proposition 4.7. *Assume* $0 < p_0 < 1$. *If* $\mu := \sum k p_k > 1$, *then the extinction probability* $a < 1$. *If* $\mu \le 1$, *then* $a = 1$.

Proof. We saw that $a = 1$ if $\mu < 1$, so we need only consider the case $\mu \ge 1$. Note that

$$\varphi''(s) = \sum_{k=2}^{\infty} k(k-1)p_k s^{k-2} \ge 0,$$

If $p_0 > 0$, and $\mu \ge 1$, $p_k > 0$ for some $k \ge 2$. This implies that $\varphi''(s) > 0$ for all $s \in (0,1)$, so φ' is strictly increasing on $(0,1)$. Recall that $\varphi(1) = 1$ and $\varphi'(1) = \mu$.

If $\mu = 1$, we have

$$1 - \varphi(s) = \int_s^1 \varphi'(t)\,\mathrm{d}t < \int_s^1 1\,\mathrm{d}t = 1 - s,$$

which gives $\varphi(s) > s$ for all $s \in [0,1)$. Since a is the smallest positive solution to $\varphi(s) = s$, we have $a = 1$.

If $\mu > 1$, we have $\varphi'(1) > 1$ and $\varphi(1) = 1$. By a Taylor expansion of φ around $s = 1$, we have

$$\varphi(1-\varepsilon) = 1 - \varphi'(1)\varepsilon + O(\varepsilon).$$

Since $\varphi'(1) > 1$, we see that there exists $\varepsilon \in (0,1)$ such that $\varphi(1-\varepsilon) < 1-\varepsilon$. But note also that $\varphi(0) = p_0 > 0$. Since $\varphi(s) - s$ is continuous on $[0,1]$, by the intermediate value theorem, there exists $s_0 \in (0, 1-\varepsilon)$ such that $\varphi(t) = t$. This gives $a \le t < 1-\varepsilon$. $\square$

Remark 4.8. The graph of $\varphi(s)$ can be described as follows:

$\mu < 1$ It intersects the y-axis at p_0, is convex, and intersects the line $y = s$ at $s = 1$ (with slope less than 1) and nowhere else in $(0,1)$.

$\mu > 1$ It intersects the y-axis at p_0, is convex, and intersects the line $y = s$ at $s = 1$ (at which point it has slope grater than 1) and at some $s = a < 1$.

Definition 4.9. We say a branching process $\{X_n\}$ is

1. **supercritical** if $\mu > 1$;
2. **critical** if $\mu = 1$;
3. **subcritical** if $\mu < 1$.

Note that in case 1, $a < 1$; in cases (2) and (3), $a = 1$.

Example 4.10. $p_0 = 1/10, p_1 = 3/5, p_2 = 3/10$. Then $\mu = 0 \cdot \frac{1}{10} + 1 \cdot \frac{3}{5} + 2 \cdot \frac{3}{10} = 1.2 > 1$. This is a supercritical branching process.

To find the extinction probability a, we need to solve $\varphi(s) = s$.

$$\varphi(s) = \frac{1}{10} + \frac{3}{5}s + \frac{3}{10}s^2 = s \iff 3s^2 - 4s + 1 = 0 \iff s = 1, \frac{1}{3}.$$

Thus $a = 1/3$.

Remark 4.11. Consider a supercritical branching process $\{X_n\}$. Conditioning on $E :=$ {extinction}, we have the conditional distribution of $\{X_n\}$ is a branching process with offspring distribution

$$\tilde{p}_k = a^{k-1} m p_k.$$

This is sometimes called the **conjugate branching process**.

Now consider a subcritical branching process $\{\tilde{X}_n\}$. Is it the conjugate of a supercritical branching process? The answer is "usually." The smallest solution to $\varphi(s) = s$ for a subcritical branching process is $s = 1$, but it usually has an extra solution $A > 1$, using which we can define the conjugate supercritical branching process with offspring distribution

$$\hat{p}_k = A^{k-1} \tilde{p}_k.$$

5 Poisson Processes

Consider a process $\{X_t\}_{t\geq 0}$ in $\mathbb{N}_0$ that models the number of phone line calls arrive at same rate at all hours. Let's suppose there are $\lambda > 0$ calls per hour on average, and that the number of calls during disjoint time intervals are independent. Let X_t be the number of calls before time t (hours).

Definition 5.1. A random variable Y in $\mathbb{N}_0$ has the **Poisson distribution** with mean $\lambda > 0$ if

$$\mathbb{P}[Y = k] = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k \in \mathbb{N}_0.$$

One can show that $\mathbb{E}[Y] = \lambda$.

Lemma 5.2. *Let Y_1, Y_2 be independent Poisson random variables with means λ_1, λ_2 respectively. Then $Y_1 + Y_2 \sim$ Poisson$(\lambda_1 + \lambda_2)$.*

Proof.

$$\begin{aligned}\mathbb{P}[Y_1 + Y_2 = k] &= \sum_{j=0}^{k} \mathbb{P}[Y_1 = j]\mathbb{P}[Y_2 = k - j] = \sum_{j=0}^{k} \frac{\lambda_1^j e^{-\lambda_1}}{j!} \cdot \frac{\lambda_2^{k-j} e^{-\lambda_2}}{(k-j)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{j=0}^{k} \frac{k!}{j!(k-j)!} \lambda_1^j \lambda_2^{k-j} = \frac{e^{-(\lambda_1+\lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k,\end{aligned}$$

where the last equality follows from the binomial theorem. □

Definition 5.3. The **Poisson process** with rate $\lambda > 0$ is the continuous time stochastic process $\{X_t\}_{t\geq 0}$ such that $X_0 = 0$ and for each $0 \leq s_1 \leq t_1 \leq \cdots \leq s_k \leq t_k$, the random variables $X_{t_1} - X_{s_1}, \ldots, X_{t_k} - X_{s_k}$ are independent and each has distribution Poisson$(\lambda(t_j - s_j))$.

A Poisson process models the number of "arrivals" until time t, when the number of arrivals in disjoint time intervals are independent.

Proposition 5.4. *The Poisson process of rate λ exists and is unique.*

We omit the proof.

An equivalent way to model Process is to consider the inter-arrival times: Let $\tau_0 = 0$ and for each $j \geq 1$, let $\tau_j := \min\{t \geq 0 : X_t = j\}$ be the jth arrival time. Note that $X_t = \max\{j : \tau_j \leq t\}$. It turns out that the inter-arrival times $T_j := \tau_j - \tau_{j-1}$ are iid with exponential distribution with parameter λ:

Definition 5.5. A random variable T in $[0, \infty)$ has the **exponential distribution** with parameter $\lambda > 0$ if for each $t > 0$,

$$\mathbb{P}[T \geq t] = e^{-\lambda t}.$$

We compute the mean of the exponential distribution:

$$\mathbb{E}[T] = \mathbb{E}\left[\int_0^\infty \mathbb{1}(t \leq T)\, \mathrm{d}t\right] = \int_0^\infty \mathbb{E}[\mathbb{1}(t \leq T)]\, \mathrm{d}t = \int_0^\infty e^{-\lambda t}\, \mathrm{d}t = \frac{1}{\lambda}.$$

Proposition 5.6. *The inter-arrival times* $\tau_j - \tau_{j-1}$ *are iid with distribution* $\text{Exponential}(\lambda)$.

We will provide the following "hand-waving" proof that captures the main idea, but does not deal with the measure-theoretic details.

Proof. Note that

$$\mathbb{P}[\tau_1 > t] = \mathbb{P}[X_t = 0] = e^{-\lambda t}.$$

Thus $\tau_1 \sim \text{Exponential}(\lambda)$. By the independent increments property, for each $t \geq 0$, we have that $\{X_{s+t} - X_t\}_{s \geq 0}$ is a Poisson process with rate λ independent from $\{X_s\}_{s \leq t}$. Time τ_1 is a stopping time for $\{X_s\}_{s \geq 0}$ since $\{\tau_1 \leq t\}$ is determined by $\{X_t\}_{s \leq t}$ for each $t \geq 0$.

By a version of the strong Markov property for $\{X_s\}$,[1] we have $\{X_{s+\tau_1} - X_{\tau_1}\}_{s \geq 0}$ is a Poisson process independent from $\{X_s\}_{s \leq \tau_1}$. Note that

$$\tau_2 - \tau_1 = \min\{s \geq 0 : X_{s+\tau_1} - X_{\tau_1} = 1\}.$$

By the τ_1 case,

$$\tau_2 - \tau_1 \sim \text{Exponential}(\lambda)$$

is independent from τ_1. We can apply the same argument iteratively to show that $\tau_j - \tau_{j-1} \sim \text{Exponential}(\lambda)$ and is independent from $\{\tau_i\}_{i<j}$ for each $j \geq 1$. □

Remark 5.7 (Some intuition of why the Poisson Process exists). Let $\{\tau_j\}_{j \geq 1}$ be iid $\text{Exponential}(\lambda)$, and for each t, let

$$X_t := \max\left\{k \geq 0 : \sum_{j=1}^{k} \tau_j \leq t\right\}.$$

It is possible to show that $\{X_t\}$ is a Poisson process. ☕

Example 5.8. Consider a bus stop at which buses arrive according to a Poisson process. On average, two buses arrive per hour. We arrive at the bus stop at time $t = 0$.

(i) If we wait for 2 hours, what is the probability that we do not see a bus?

Note that $X_2 \sim \text{Poisson}(4)$, and so

$$\mathbb{P}[X_2 = 0] = e^{-4} \approx 0.0183.$$

(ii) Given that no bus arrives in the first two hours, what is the conditional probability that a bus arrives in the next hour?

$$\mathbb{P}[X_3 - X_2 \geq 0 | X_2 = 0] = \mathbb{P}[X_3 - X_2 \geq 0] = 1 - \mathbb{P}[X_3 - X_2 = 0] = 1 - e^{-2} \approx 0.864.$$

(iii) What is the expected arrival time of the second bus?

$$\mathbb{E}[\tau_2] = \mathbb{E}[T_1 + T_2] = \mathbb{E}[T_1] + \mathbb{E}[T_2] = \frac{1}{2} + \frac{1}{2} = 1.$$

[1] This is where the hand-waving happens.

Proposition 5.9 (Two Properties of the Expoenetial Distribution)**.**

(i) The memoryless property: for each $t, s \geq 0$*,*

$$\mathbb{P}\left[T - t \geq s | T \geq t\right] = e^{-\lambda s}.$$

(ii) The minimum property: Let $T_1, \ldots, T_n$ *be independent random variables with* $T_j \sim \text{Exponential}(\lambda_j)$*. Then* $T \coloneqq \min_{1 \leq j \leq n} T_j \sim \text{Exponential}\left(\sum_{j=1}^{n} \lambda_j\right)$*. Moreover,*

$$\mathbb{P}\left[T_j = \min\{T_1, \ldots, T_n\}\right] = \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_n}.$$

Proof.

(i)

$$\mathbb{P}[T \geq t + s | T \geq t] = \frac{\mathbb{P}[T \geq t + s]}{\mathbb{P}[T \geq t]} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s}.$$

(ii) We have

$$\begin{aligned}\mathbb{P}[\min\{T_1, \ldots, T_n\} \geq t] &= \mathbb{P}[T_1 \geq t, \ldots, T_n \geq t] \\ &= \prod_{j=1}^{n} \mathbb{P}[T_j \geq t] = e^{-(\lambda_1 + \cdots + \lambda_n)t}.\end{aligned}$$

For the second statement, assume without loss of generality that $j = 1$. Then

$$\begin{aligned}\mathbb{P}[T_1 = \min\{T_1, \ldots, T_n\}] &= \mathbb{P}[T_1 \leq T_j, \forall j \geq 2] \\ &= \mathbb{E}\left[\mathbb{P}\left[T_j \geq T_1, \forall j \geq 2 \middle| T_1\right]\right] = \mathbb{E}\left[e^{-(\lambda_1 + \cdots + \lambda_n)T_1}\right] \\ &= \int_0^\infty \lambda_1 e^{-\lambda_1 t} e^{-(\lambda_2 + \cdots + \lambda_n)t} \, \mathrm{d}t = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_n}.\end{aligned}$$

□

6 Continuous Time Markov Chains

Let S be finite. We will define a **continuous time Markov chain** $\{X_t\}_{t \geq 0}$ in S. The **rates** $\alpha(x, y) \geq 0$ for distinct $x, y \in S$ will describe the "rate at which we jump from x to y."

Assume $X_0 = x \in S$. For $y \in S$ with $\alpha(x, y) \neq 0$, let $T_j \sim \text{Exponential}(\alpha(x, y))$ be independent for different y. Think of this intuitively as "alarm clocks" which rings at time T_y. At time

$$T := \min \left\{T_y : \alpha(x, y) \neq 0\right\},$$

we move to state y, where y is such that $T_y = T$. By the minimum property of the exponential distribution, we have $T \sim \text{Exponential}\,(\alpha(x))$, where $\alpha(x) := \sum_{y \neq x} \alpha(x, y)$ and $\mathbb{P}[X_T = y] = \alpha(x, y)/\alpha(x)$. After time T, we forget the T_j's and choose the next step using exponential random variables of parameters $\alpha(X_T, z)$ for $z \neq X_T$ and $\alpha(X_T, z)$. We continue iteratively to define X_t for $t \geq 0$. If $\alpha(x, y) = 0$ for all $y \neq x$, then once we reach x, we stay there forever.

Proposition 6.1 (The Strong Markov Property for Continuous Time Markov Chains)**.** *Let $t > 0$ and $x \in S$. The conditional distribution of $\{X_{t+s}\}_{s \geq 0}$ given $\{X_t = x\}$ and everything before time t is the same as the distribution of $\{X_t\}$ started at x.*

Proof. Omitted. Key ingredient is the memoryless property of the exponential distribution. □

Let $\tau_0 = 0$ and $\tau_1, \tau_2, \ldots$ be the jump times of $\{X_t\}$. By construction, if we condition on $\{X_{\tau_0}, X_{\tau_1}, \ldots, X_{\tau_{n-1}}\}$, then the conditional probability that $X_{\tau_n} = y$ (if $X_{\tau_{n-1}} = x$)) is $\alpha(x, y)/\alpha(x)$. Equivalently, if $\tilde{X}_n := X_{\tau_n}$ for $n \geq 0$, then $\{\tilde{X}_n\}$ is a discrete time Markov chain with transition probabilities

$$P(x, y) = \begin{cases} \frac{\alpha(x,y)}{\alpha(x)}, & x \neq y, \\ 0, & x = y. \end{cases}$$

This is called the **associated discrete time Markov chain** of $\{X_t\}$. Note that from the distribution of the discrete time Markov chain, we cannot recover the rates $\alpha(x, y)$ since the inter-jump times $\tau_n - \tau_{n-1}$ are lost.

Definition 6.2. The **time t transition probabilities** of the continuous time Markov chain $\{X_t\}$ are defined as

$$P_t(x, y) := \mathbb{P}\left[X_t = y | X_0 = x\right], \quad x, y \in S, \quad t \geq 0.$$

Since without loss of generality we can assume $S = \{1, \ldots, N\}$, we can define the **time t transition matrix**

$$P_t := \left(P_t(x, y)\right)_{1 \leq x, y \leq N} \in \mathbb{R}^{N \times N}.$$

The transition matrix at different times satisfy a differential equation, which we now derive.

For $x \neq y$, we have

$$\begin{aligned}\mathbb{P}\left[X_{t+\varepsilon} = y | X_t = x\right] &= \mathbb{P}\left[X_\varepsilon = y | X_0 = x\right] \\ &= \mathbb{P}\left[T_y < \varepsilon \middle| X_0 = x\right] + o(\varepsilon) = 1 - e^{-\alpha(x,y)\varepsilon} + o(\varepsilon),\end{aligned}$$

where $o(\varepsilon)$ denotes a term such that $o(\varepsilon)/\varepsilon \to 0$ as $\varepsilon \to 0$. Using a first order Taylor expansion of the exponential function, we have

$$\mathbb{P}\left[X_{t+\varepsilon} = y | X_t = x\right] = \alpha(x, y)\varepsilon + o(\varepsilon),$$

where we absorb the higher order terms into $o(\varepsilon)$. Summing over $y \neq x$, we also have

$$\begin{aligned}\mathbb{P}\left[X_{t+\varepsilon} \neq x | X_t = x\right] &= \sum_{y \neq x} \mathbb{P}\left[X_{t+\varepsilon} = y | X_t = x\right] \\ &= \sum_{y \neq x} \alpha(x, y)\varepsilon + o(\varepsilon) = \alpha(x)\varepsilon + o(\varepsilon).\end{aligned}$$

Now observe that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} P_t(x, y) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left\{\mathbb{P}\left[X_{t+\varepsilon} = y | X_0 = x\right] - \mathbb{P}\left[X_t = y | X_0 = x\right]\right\} \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \Big\{\mathbb{P}\left[X_{t+\varepsilon} = y, X_t = y | X_0 = x\right] + \mathbb{P}\left[X_{t+\varepsilon} = y, X_t \neq y | X_0 = x\right] \\ &\qquad - \mathbb{P}\left[X_{t+\varepsilon} = y, X_t = y | X_0 = x\right] - \mathbb{P}\left[X_{t+\varepsilon} \neq y, X_t = y | X_0 = x\right]\Big\} \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left\{\sum_{z \neq y} \mathbb{P}\left[X_{t+\varepsilon} = y | X_t = z\right] P_t(x, z) - \mathbb{P}\left[X_{t+\varepsilon} \neq y, X_t = y | X_0 = x\right]\right\} \\ &= \sum_{z \neq y} \alpha(z, y) P_t(x, z) - \alpha(y) P_t(x, y).\end{aligned}$$

This is a system of ordinary differential equations called the **Kolmogorov forward equations**. We have the following initial conditions:

$$P_0(x, x) = 1, \quad P_0(x, y) = 0 \text{ for } y \neq x.$$

Definition 6.3. The **infinitesimal generator** is the $N \times N$ matrix A defined as

$$A_{x,y} := \begin{cases} \alpha(x, y), & x \neq y, \\ -\alpha(x), & x = y. \end{cases}$$

$$A = \begin{pmatrix} -\alpha(1) & \alpha(1,2) & \alpha(1,3) & \cdots & \alpha(1,N) \\ \alpha(2,1) & -\alpha(2) & \alpha(2,3) & \cdots & \alpha(2,N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha(N,1) & \alpha(N,2) & \alpha(N,3) & \cdots & -\alpha(N) \end{pmatrix}.$$

Using the infinitesimal generator, we can write the Kolmogorov forward equations as

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = P_t A, \quad P_0 = I.$$

Remark 6.4. Recall that for $a \in \mathbb{R}$, the ODE $f'(t) = af(t), f(0) = 1$ has solution $f(t) = e^{at}$.

It is thus natural to guess that the solution to the matrix ODE is given by

$$P_t = e^{tA} := \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}.$$

By differentiating term-by-term, we can check that

$$\frac{\mathrm{d}}{\mathrm{d}t} e^{tA} = e^{tA} A = A e^{tA},$$

where we know e^{tA} and A commute since they are both polynomials in A. Since also $e^{0 \cdot A} = I$, by the existence and uniqueness theorem for ODEs, we have the unique solution $P_t = e^{tA}$. We summarize this in the following proposition:

Proposition 6.5. *The time t transition matrix of the continuous time Markov chain with infinitesimal generator A is given by*

$$P_t = e^{tA} = \sum_{n=0}^{\infty} \frac{(tA)^n}{n!}.$$

How do we compute e^{tA}? For $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, it is clear that

$$e^{tD} = \begin{pmatrix} e^{t\lambda_1} & & & \\ & e^{t\lambda_2} & & \\ & & \ddots & \\ & & & e^{t\lambda_n} \end{pmatrix}.$$

Thus if $A = QDQ^{-1}$ is diagonalizable, then for any n, $A^n = QD^nQ^{-1}$ and consequently,

$$e^{tA} = Qe^{tD}Q^{-1}.$$

For a general matrix A, we can use the Jordan normal form to compute e^{tA}.

Example 6.6. $S = \{0, 1\}$, $\alpha(0, 1) = 2$, $\alpha(1, 0) = 3$. The infinitesimal generator is

$$A = \begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}.$$

To diagonalize A, we first find the eigenvalues of A by solving

$$0 = \det(A - \lambda I) = (-2 - \lambda)(-3 - \lambda) - 6 \implies \lambda_1 = 0, \lambda_2 = -5.$$

The corresponding eigenvectors can be found by solving $Av = -5v$ and $Av = 0$, which yields

$$v_{-5} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}, \quad v_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Thus we have

$$A = \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} -5 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix}^{-1}$$

and so

$$e^{tA} = \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} e^{-5t} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -3 & 1 \end{pmatrix}^{-1} = \frac{1}{5} \begin{pmatrix} 3 + 2e^{-5t} & 2 - 2e^{-5t} \\ 3 - 3e^{-5t} & 2 + 3e^{-5t} \end{pmatrix}.$$

In particular, we have say $P_t(0, 1) = \left(2 - 2e^{-5t}\right)/5$.

6.1 Stationary Distributions

Definition 6.7. A continuous time Markov chain $\{X_t\}$ is **irreducible** if can get from any state to any other state.

Remark 6.8. $\{X_t\}$ is irreducible if $P_t(x, y) > 0$ for some or equivalently all $t > 0$ and all $x, y \in S$.

Definition 6.9. We say $\pi : S \to [0, 1]$ with $\sum \pi_x = 1$ is a **stationary distribution** of the continuous time Markov chain $\{X_t\}$ if whenever $X_0 \sim \pi$, we have $X_t \sim \pi$ for all $t \geq 0$.

Note that stationarity is equivalent to $\pi P_t = \pi$ for each $t \geq 0$. Since $P_0 = I$, this condition is equivalent to πP_t does not depend on 0, or

$$0 = \frac{\mathrm{d}}{\mathrm{d}t} \pi P_t = \pi \frac{\mathrm{d}}{\mathrm{d}t} e^{tA} = \pi A e^{tA} \iff \pi A = 0.$$

Note that throughout, we have assumed that $\sum \pi_x = 1$. Thus we have the following proposition:

Proposition 6.10. *A distribution π is a stationary distribution of the continuous time Markov chain $\{X_t\}$ with infinitesimal generator A if and only if*

$$\pi A = 0, \quad \sum_{x \in S} \pi_x = 1.$$

Moreover,

Proposition 6.11. *If $\{X_t\}$ is irreducible, then there exists a unique stationary distribution π for $\{X_t\}$. Also,*

$$\lim_{t \to \infty} P_t(x, y) = \pi_y, \quad \forall x, y \in S.$$

Proof. Omitted. Similar to the discrete time case. □

Example 6.12.

$$A = \begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}.$$

The stationary distribution $\pi = (\pi_0, \pi_1)$ satisfies

$$\begin{cases} -2\pi_0 + 3\pi_1 = 0 \\ 2\pi_0 - 3\pi_1 = 0 \end{cases}, \quad \pi_0 + \pi_1 = 1 \implies \pi_0 = \frac{3}{5}, \quad \pi_1 = \frac{2}{5}.$$

A comparison of the continuous time and discrete time cases:

Discrete Time	Continuous Time
Transition probabilities $P(x, y)$	Rates $\alpha(x, y)$
Matrix P	Infinitesimal generator A
$P^n(x, y) = (P^n)_{x,y}$	$P_t(x, y) = (P_t)_{x,y} = (e^{tA})_{x,y}$
Stat. dist. solves $\pi P = \pi$	Stat. dist. solves $\pi A = 0$, $\sum \pi_x = 1$

7 Conditional Expectation

The conditional distribution of a random variable X given Y, $\mathbb{E}[X|Y]$, is the best guess for X if we know Y.

Example 7.1. Let X, Y be random variables in countable sets $S \subset \mathbb{R}$ and T. For $s \in S$, $y \in T$, write $f(s,t) := \mathbb{P}[X = s, Y = y]$. Then, for any $y \in T$,

$$\mathbb{E}[X|Y = y] = \sum_{s \in S} s\mathbb{P}[X = s|Y = y] = \frac{\sum_{s \in S} sf(s,y)}{\sum_{s \in S} f(s,y)},$$

provided the denominator is non-zero. In general, we have the random variable

$$\mathbb{E}[X|Y] = \frac{\sum_{s \in S} sf(s,Y)}{\sum_{s \in S} f(s,Y)}.$$

Example 7.2. Let X, Y be random variables in $\mathbb{R}$. Assume there exists a joint density $f(x,y)$ such that for any measurable $A, B \subset \mathbb{R}$,

$$\mathbb{P}[X \in A, Y \in B] = \int_B \int_A f(x,y)\,\mathrm{d}x\,\mathrm{d}y.$$

The conditional density of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{\int_{\mathbb{R}} f(u,y)\,\mathrm{d}u},$$

The conditional expectation is

$$\mathbb{E}[X|Y] := \frac{\int_{\mathbb{R}} xf(x,Y)\,\mathrm{d}x}{\int_{\mathbb{R}} f(u,Y)\,\mathrm{d}u}.$$

Definition 7.3. Let X, Y be random variables. Assume X take values in $\mathbb{R}$ and $\mathbb{E}[|X|] < \infty$. The **conditional expectation** $\mathbb{E}[X|Y]$ is the unique random variable in $\mathbb{R}$ such that:

- $\mathbb{E}[X|Y]$ is a function of Y,
- if $F(y)$ is a function of Y in $\mathbb{R}$ with $\mathbb{E}[|F(Y)|] < \infty$, then

$$\mathbb{E}[XF(Y)] = \mathbb{E}\big[\mathbb{E}[X|Y]F(Y)\big].$$

Theorem 7.4. *Conditional expectation exists and is unique.*

Proof. Full proof in MATH314. Uses the Radon-Nikodym theorem, or an orthogonal projection argument in L^2. $\square$

Definition 7.5. The **conditional probability** of an event E given Y is

$$\mathbb{P}[E|Y] := \mathbb{E}[\mathbb{1}_E|Y].$$

Theses definitions generalize the usual definitions of conditional probability and expectation:

Example 7.6. If X takes values in a countable set S, then if $F(Y)$ is a function of Y in $\mathbb{R}$ with $\mathbb{E}[|F(Y)|] < \infty$,

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}[X|Y]F(Y)\right] &= \sum_{u\in S}\sum_{v\in T} f(u,v)F(v)\frac{\sum_{s\in S} sf(s,v)}{\sum_{s\in S} f(s,v)} \\
&= \sum_{v\in T}\left(\sum_{u\in S} f(u,v)\right)F(v)\frac{\sum_{s\in S} sf(s,v)}{\sum_{s\in S} f(s,v)} \\
&= \sum_{v\in T}\sum_{s\in S} F(v)sf(s,v) = \mathbb{E}[XF(Y)].
\end{aligned}$$

Assume all random variables X in $\mathbb{R}$ satisfy $\mathbb{E}[|X|] < \infty$. Consider a sequence of random variables $\{Y_n\}_{n\geq 0}$. We can define $\mathbb{E}X|Y_0,\dots,Y_n]$ by taking $Y := (Y_0,\dots,Y_n$.

Definition 7.7. Let $\mathcal{F}_n$ denote the information in $Y_0,\dots,Y_n$, called the **filtration** generated by $Y_0,\dots,Y_n$. That is,

$$\mathbb{E}[X|\mathcal{F}_n] := \mathbb{E}[X|Y_0,\dots,Y_n].$$

We say a random variable X is $\boldsymbol{\mathcal{F}_n}$**-measurable** if $X = f(Y_0,\dots,Y_n)$ for some function f.

Remark 7.8. $\mathcal{F}$ is rigorously defined as the σ-algebra generated by $Y_0,\dots,Y_n$.

We can reformulate the definition of conditional expectation using $\mathcal{F}_n$:

Proposition 7.9. $\mathbb{E}[X|\mathcal{F}_n]$ *is the unique random variable in* $\mathbb{R}$ *such that*

- $\mathbb{E}[X|\mathcal{F}_n]$ *is* $\mathcal{F}_n$*-measurable,*
- *if* Z *is* $\mathcal{F}_n$*-measurable and takes values in* $\mathbb{R}$*, then*

$$\mathbb{E}[XZ] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{F}_n]Z\right].$$

7.1 Properties of Conditional Expectation

We now state some properties of conditional expectation (note how they reflect conditional expectation as the "best guess" of X given Y):

Proposition 7.10. *If* X_1, X_2 *are random variables in* $\mathbb{R}$ *and* $a, b \in \mathbb{R}$ *are constants, then*

$$\mathbb{E}[aX_1 + bX_2|\mathcal{F}_n] = a\mathbb{E}[X_1|\mathcal{F}_n] + b\mathbb{E}[X_2|\mathcal{F}_n].$$

Proof. Let $W := a\mathbb{E}[X_1|\mathcal{F}_n] + b\mathbb{E}[X_2|\mathcal{F}_n]$. We check that Wsatisfied the definition of $\mathbb{E}[aX_1 + bX_2|\mathcal{F}_n]$.

- W is $\mathcal{F}_n$-measurable since it is a linear combination of $\mathcal{F}_n$-measurable random variables.

- Let Z be $\mathcal{F}_n$-measurable and take values in $\mathbb{R}$. Then

$$\begin{aligned}\mathbb{E}[(aX_1 + bX_2)Z] &= a\mathbb{E}[X_1 Z] + b\mathbb{E}[X_2 Z] \\ &= a\mathbb{E}\left[\mathbb{E}[X_1|\mathcal{F}_n]Z\right] + b\mathbb{E}\left[\mathbb{E}[X_2|\mathcal{F}_n]Z\right] \\ &= \mathbb{E}\left[WZ\right].\end{aligned}$$

□

Proposition 7.11. *If X is $\mathcal{F}_n$-measurable, then* $\mathbb{E}[X|\mathcal{F}_n] = X$.

Proof. X is $\mathcal{F}_n$-measurable by assumption. If Z is $\mathcal{F}_n$-measurable and takes values in $\mathbb{R}$, then $\mathbb{E}[XZ] = \mathbb{E}[XZ]$. □

Proposition 7.12. *If X is independent from $\mathcal{F}_n$, then* $\mathbb{E}[X|\mathcal{F}_n] = \mathbb{E}[X]$.

Proof. $\mathbb{E}[X]$ is $\mathcal{F}_n$-measurable since it is a constant. If Z is $\mathcal{F}_n$-measurable and takes values in $\mathbb{R}$, then

$$\mathbb{E}[XZ] = \mathbb{E}[X]\mathbb{E}[Z] = \mathbb{E}[\mathbb{E}[X]Z].$$

□

Proposition 7.13. *If Z is $\mathcal{F}_n$-measurable then* $\mathbb{E}[ZX|\mathcal{F}_n] = Z\mathbb{E}[X|\mathcal{F}_n]$.

Proof. Write $W := Z\mathbb{E}[X|\mathcal{F}_n]$. W is $\mathcal{F}_n$-measurable since it is a product of $\mathcal{F}_n$-measurable random variables. If Z' is $\mathcal{F}_n$-measurable and takes values in $\mathbb{R}$, then

$$\mathbb{E}[WZ'] = \mathbb{E}\big[\underbrace{Z'Z}_{\mathcal{F}_n\text{-measurable}} \mathbb{E}[X|\mathcal{F}_n]\big] = \mathbb{E}[ZXZ'].$$

□

Proposition 7.14.

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}_n]\big] = \mathbb{E}[X].$$

Proof. Take $Z = 1$ in the definition of conditional expectation. □

Proposition 7.15 (Tower Property)**.** *If $m \leq n$, then*

$$\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}_n]\big|\mathcal{F}_m\big] = \mathbb{E}[X|\mathcal{F}_m].$$

Proof. $\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}_n]\big|\mathcal{F}_m\big]$ is $\mathcal{F}_m$-measurable since it is a conditional expectation given $\mathcal{F}_m$. If Z is $\mathcal{F}_m$-measurable and takes values in $\mathbb{R}$, then

$$\mathbb{E}\left[\mathbb{E}\big[\mathbb{E}[X|\mathcal{F}_n]\big|\mathcal{F}_m\big]Z\right] = \mathbb{E}\left[Z\mathbb{E}[X|\mathcal{F}_n]\right] = \mathbb{E}[ZX].$$

□

Proposition 7.16.

$$|\mathbb{E}[X|\mathcal{F}_n]| \leq \mathbb{E}[|X||\mathcal{F}_n].$$

Proof. Let Z be $\mathcal{F}_n$-measurable and take values in $[0, \infty)$ We have

$$\mathbb{E}[Z\mathbb{E}[X|\mathcal{F}_n]] = \mathbb{E}[ZX] \leq \mathbb{E}[Z|X|] = \mathbb{E}[Z\mathbb{E}[|X||\mathcal{F}_n]].$$

Since this holds for all such Z, we have $\mathbb{E}[X|\mathcal{F}_n] \leq \mathbb{E}[|X||\mathcal{F}_n]$. Similarly, we can show that $-\mathbb{E}[X|\mathcal{F}_n] \leq \mathbb{E}[|X||\mathcal{F}_n]$. $\square$

Example 7.17. Let $\{X_j\}_{j\geq 1}$ be independent random variables with mean μ. Denote the information in $X_1, \ldots, X_m$ as $\mathcal{F}_m$. Then, for $m < n$, we have

$$\begin{aligned}
\mathbb{E}[X_1 + \cdots + X_n|\mathcal{F}_m] &= \mathbb{E}[X_1 + \cdots + X_m|\mathcal{F}_m] + \mathbb{E}[X_{m+1} + \cdots + X_n|\mathcal{F}_m] \\
&= X_1 + \cdots + X_m + \mathbb{E}[X_{m+1}] + \cdots + \mathbb{E}[X_n] \\
&= X_1 + \cdots + X_m + (n-m)\mu.
\end{aligned}$$

Example 7.18. $\{X_j\}_{j\geq 1}$ are iid with mean 0 and variance σ^2. Let $\mathcal{F}_n$ denote the information in $X_1, \ldots, X_n$ and write $S_n := X_1 + \cdots + X_n$. Note that for $m < n$,

$$S_n^2 = (S_n - S_m + S_m)^2 = S_m^2 + 2S_m(S_n - S_m) + (S_n - S_m)^2.$$

Thus

$$\begin{aligned}
\mathbb{E}[(X_1 + \cdots + X_n)^2|\mathcal{F}_m] &= \mathbb{E}[S_m^2|\mathcal{F}_m] + 2S_m\mathbb{E}[S_n - S_m|\mathcal{F}_m] + \mathbb{E}[(S_n - S_m)^2|\mathcal{F}_m] \\
&= S_m^2 + 2S_m\mathbb{E}[S_n - S_m] + \mathbb{E}[(S_n - S_m)^2] \\
&= (X_1 + \cdots + X_m)^2 + (n-m)\sigma^2.
\end{aligned}$$

8 Martingales

Definition 8.1. A sequence of random variables in $\mathbb{R}$, $\{M_n\}$, is a **martingale** with respect to the filtration $\{\mathcal{F}_n\}$ if

(i) $\{M_n\}$ is $\mathcal{F}_n$-measurable for each n,

(ii) $\mathbb{E}[|M_n|] < \infty$ for each n,

(iii) For each $m \leq n$, $\mathbb{E}[M_n|\mathcal{F}_m] = M_m$, or equivalently, $\mathbb{E}[M_n - M_m|\mathcal{F}_m] = 0$m.

Remark 8.2.

- Intuitively, a martingale is a sequence of random variables which models a fair game: the best guess for the future value given the past is the current value.
- If we don't specify $\{\mathcal{F}_n\}$, assume $\mathcal{F}_n$ contains the information in $M_0, \ldots, M_n$.
- Note that $\mathbb{E}[M_n] = \mathbb{E}[M_0]$ for each n.

Proposition 8.3. *Property (iii) in the definition of martingale is equivalent to the condition that for each $n \geq 1$,*

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = M_{n-1}.$$

Proof. The forward direction $\implies$ is immediate.

For the reverse direction, we apply induction on $n - m$. Base case $n - m = 1$ holds by assumption. For $n - m = 2$,

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_{n-2}] &= \mathbb{E}\left[\mathbb{E}[M_n|\mathcal{F}_{n-1}]|\mathcal{F}_{n-2}\right] \\ &= \mathbb{E}[M_{n-1}|\mathcal{F}_{n-2}] = M_{n-2},\end{aligned}$$

where the first equality uses the tower property. □

Example 8.4. Let $\{X_j\}$ be independence random variables in $\mathbb{R}$ with a common mean μ. Let $\mathcal{F}_n$ denote the information in $X_1, \ldots, X_n$ and write $S_n := X_1 + \cdots + X_n$. Then $M_n := S_n - n\mu$ is a martingale with respect to $\{\mathcal{F}_n\}$.

The first property is clear. For the second,

$$\mathbb{E}[|M_n|] \leq \sum_{j=1}^{n} \mathbb{E}[|X_j|] + n|\mu| = 2n\mathbb{E}[|X_1|] < \infty.$$

For the third, note that

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_{n-1}] &= \mathbb{E}[S_{n-1} + X_n - n\mu|\mathcal{F}_{n-1}] \\ &= S_{n-1} + \mathbb{E}[X_n] - n\mu = S_{n-1} - (n-1)\mu = M_{n-1}.\end{aligned}$$

Example 8.5. Let $\{X_j\}$ be independent random variables in $\mathbb{R}$ with mean 0 and variance σ^2. Consider $M_n := S_n^2 - \sigma^2 n$. This is a martingale with respect to the filtration $\{\mathcal{F}_n\}$.

To see this, note that

$$\mathbb{E}[|M_n|] \leq \mathbb{E}[S_n^2] + \sigma^2 n = n\sigma^2 + n\sigma^2 < \infty$$

and

$$\begin{aligned}\mathbb{E}[M_n|\mathcal{F}_{n-1}] &= \mathbb{E}[S_{n-1}^2 + 2S_{n-1}X_n + X_n^2 - \sigma^2 n|\mathcal{F}_{n-1}] \\ &= S_{n-1}^2 + \mathbb{E}[X_n^2] - \sigma^2 n \\ &= S_{n-1}^2 - \sigma^2(n-1) = M_{n-1}.\end{aligned}$$

Example 8.6. Let $\{F_n\}$ be any filtration. Let X be any random variable in $\mathbb{R}$ with $\mathbb{E}[|X_n|] < \infty$. Then $M_n := \mathbb{E}[X|\mathcal{F}_n]$ is a martingale with respect to $\{\mathcal{F}_n\}$.

To see this, note that M_n is $\mathcal{F}_n$-measurable by definition of conditional expectation

$$\mathbb{E}[|M_n|] \leq \mathbb{E}\left[\mathbb{E}[|X||\mathcal{F}_n]\right] = \mathbb{E}[|X|] < \infty$$

by the property of conditional expectation. Finally, for any n, by the tower property we have

$$\mathbb{E}[M_n|\mathcal{F}_{n-1}] = \mathbb{E}[X|\mathcal{F}_{n-1}] = M_{n-1}.$$

8.1 The Optional Stopping Theorem

Let $\{M_n\}$ be a martingale with respect to $\{\mathcal{F}_n\}$. Recall that $\mathbb{E}[M_n] = \mathbb{E}[M_0]$ for each n.

But what about a random time? Recall the following

Definition 8.7. A random variable τ in $\mathbb{N}_0 \cup \{\infty\}$ is a **stopping time** for $\{\mathcal{F}_n\}$ if for each $n \in \mathbb{N}_0$, $\{\tau = n\}$ is $\mathcal{F}_n$-measurable.

Recall also that $\{\tau = n\}$ can be replaced by $\{\tau \leq n\}$ or $\tau > n$ in the definition.

Example 8.8. Let $\{X_n\}$ be the random walk on $\mathbb{Z}$ with $X_0 = 0$. It is the martingale with respect to the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ is the information in $X_0, \ldots, X_n$.

Now consider $\tau := \min\{n \geq 0 : X_n = 1\}$. Note that by recurrence of $\{X_n\}$, we have $\mathbb{P}[\tau < \infty] = 1$. However,

$$\mathbb{E}[X_\tau] = 1 \neq 0 = \mathbb{E}[X_0].$$

We will see that this property is related to the fact that $\{X_n\}$ is null recurrent, or that τ has infinite expectation.

Theorem 8.9 (Optional Stopping Theorem, Version 1)**.** *Let $\{M_n\}$ be a martingale with respect to $\{\mathcal{F}_n\}$. Let τ be a stopping time that is bounded, i.e., such that there exists a non-random $K \in \mathbb{N}_0$ such that $\mathbb{P}[\tau \leq K] = 1$. Then, $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.*

Proof. Note first that $M_\tau = \sum_{n=0}^{K} M_n \mathbb{1}_{\{\tau=n\}}$. We have

$$\mathbb{E}[M_\tau|\mathcal{F}_{K-1}] = \sum_{n=0}^{K-1} \mathbb{E}\left[M_n \mathbb{1}_{\{\tau=n\}}\middle|\mathcal{F}_{K-1}\right] + \mathbb{E}[M_K \mathbb{1}_{\{\tau=K\}}|\mathcal{F}_{K-1}]$$

where the expectations in the first term can be dropped. For the second term, note that $\{\tau = K\} \equiv \{\tau > K-1\}$ is $\mathcal{F}_{K-1}$-measurable. Thus,

$$\begin{aligned}
\mathbb{E}[M_\tau|\mathcal{F}_{K-1}] &= \sum_{n=0}^{K-1} M_n \mathbb{1}_{\{\tau=n\}} + M_{k-1}\mathbb{1}_{\{\tau>K-1\}} \\
&= \sum_{n=0}^{K-2} M_n \mathbb{1}_{\{\tau=n\}} + M_{K-1}\left(\mathbb{1}_{\{\tau=K-1\}} + \mathbb{1}_{\{\tau\geq K-1\}}\right) \\
&= \sum_{n=0}^{K-2} M_n \mathbb{1}_{\{\tau=n\}} + M_{K-1}\mathbb{1}_{\{\tau>K-2\}}.
\end{aligned}$$

Now, applying the same argument again, we get

$$\begin{aligned}
\mathbb{E}[M_\tau|\mathcal{F}_{K-2}] &= \mathbb{E}\left[\mathbb{E}[M_\tau|\mathcal{F}_{K-1}]|\mathcal{F}_{K-2}\right] \\
&= \sum_{n=0}^{K-2} \mathbb{E}\left[M_n \mathbb{1}_{\{\tau=n\}}\middle|\mathcal{F}_{K-2}\right] + \mathbb{E}\left[M_{K-2}\mathbb{1}_{\{\tau>K-2\}}\middle|\mathcal{F}_{K-2}\right] \\
&= \sum_{n=0}^{K-2} M_n \mathbb{1}_{\{\tau=n\}} + M_{K-2}\mathbb{1}_{\{\tau>K-2\}} \\
&= \sum_{n=0}^{K-3} M_n \mathbb{1}_{\{\tau=n\}} + M_{K-2}\mathbb{1}_{\{\tau>K-3\}}.
\end{aligned}$$

Repeating this process, we get $\mathbb{E}[M_\tau|\mathcal{F}_0] = M_0 \mathbb{1}_{\{\tau>-1\}} = M_0$. □

What if τ is not bounded? We take limits! Write

$$\tau \wedge n := \min\{\tau, n\}.$$

As the minimum of τ and a nonrandom stopping time n, $\tau \wedge n$ is a stopping time. Since $\tau \wedge n \leq n$, we have

$$\mathbb{E}[M_n] = \mathbb{E}[M_{\tau\wedge n}] = \mathbb{E}[M_\tau \mathbb{1}_{\{\tau\leq n\}}] + \mathbb{E}[M_n \mathbb{1}_{\{\tau>n\}}].$$

If τ is almost surely finite, then as $n \to \infty$, the indicator $\mathbb{1}_{\{\tau\leq n\}}$ should converge to 1, and $\mathbb{1}_{\{\tau>n\}}$ should converge to 0, though this is not enough for the expectations to converge. For the first term, we will use the following

Theorem 8.10 (Dominated Convergence Theorem)**.** *Let $\{X_n\}$ be random variables in $\mathbb{R}$ such that $\lim_{n\to\infty} X_n = X$ with probability 1. Assume that there exists a random variable Y in $\mathbb{R}$ with $\mathbb{E}[|Y|] < \infty$ such that $|X_n| \leq |Y|$ for all n with probability 1. Then,*

$$\mathbb{E}[X_n] \longrightarrow \mathbb{E}[X].$$

Example 8.11 (Nonexample). Let $\{X_n\}$ be random variables such that $\mathbb{P}[X_n = n^2] = n^{-2}$ and $\mathbb{P}[X_n = 0] = 1 - n^{-2}$. Note that $\mathbb{P}[\exists j \geq n : X_j \neq 0] \leq \sum_{j \geq n} j^{-2} \to 0$. In particular, $\lim_{n\to\infty} X_n = 0$ with probability 1. However, $\mathbb{E}[X_n] = 1$ for each n, so $\mathbb{E}[X_n] \not\to 0 = \mathbb{E}[0]$.

Assume that $\mathbb{E}[|M_\tau|] < \infty$ and $\mathbb{P}[\tau < \infty] = 1$. Then, $M_\tau \mathbb{1}_{\{\tau \leq n\}} \to M_\tau$ with probability 1. Further, we have $|M_\tau \mathbb{1}_{\{\tau \leq n\}}| \leq |M_\tau|$. Thus by the dominated convergence theorem (using $Y := |M_\tau|$), $\mathbb{E}[M_\tau \mathbb{1}_{\{\tau \leq n\}}] \longrightarrow \mathbb{E}[M_\tau]$. From this discussion we have:

Theorem 8.12 (Optional Stopping Theorem, Version 2)**.** *Let $\{M_n\}$ be a martingale with respect to $\{\mathcal{F}_n\}$. Let τ be a stopping time such that $\mathbb{P}[\tau < \infty] = 1$, $\mathbb{E}[|M_\tau|] < \infty$, and $\lim_{n\to\infty} \mathbb{E}\left[M_n \mathbb{1}_{\{\tau > n\}}\right] = 0$. Then, $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$.*

The condition $\lim_{n\to\infty} \mathbb{E}[M_n \mathbb{1}_{\{\tau > n\}}] = 0$ is usually the hardest to verify. We discuss some sufficient conditions:

Proposition 8.13. *If the first or second moments of M_n are uniformly bounded and $\mathbb{P}(\tau < \infty) = 1$, then*

$$\lim_{n\to\infty} \mathbb{E}[M_n \mathbb{1}_{\{\tau > n\}}] = 0.$$

Proof. If $\mathbb{P}[|M_n| \leq C] = 1$ for each n, then

$$|\mathbb{E}[M_n \mathbb{1}_{\{\tau > n\}}]| \leq \mathbb{E}[|M_n| \mathbb{1}_{\{\tau > n\}}] \leq C\mathbb{P}[\tau > n] \to 0.$$

If $\mathbb{E}[M_n^2] \leq C$ for each n, then by Cauchy-Schwarz,

$$|\mathbb{E}[M_n \mathbb{1}_{\{\tau > n\}}]| \leq \mathbb{E}[M_n^2]^{1/2} \mathbb{P}[\tau > n]^{1/2} \leq C^{1/2} \mathbb{P}[\tau > n]^{1/2} \to 0.$$

□

Example 8.14. Let $\{X_n\}$ be the unbiased random walk on $\mathbb{Z}$ starting at $X_0 = 0$. Let $a, b \in \mathbb{N}$, and set

$$\tau := \min\{n \geq 0 : X_n = -a \text{ or } X_n = b\}.$$

Recall that we obtained $\mathbb{P}[X_\tau = b] = a/(a+b)$ previously by solving a system of linear equations. We give an alternative proof using the optional stopping theorem:

Recall that the random walk $\{X_n\}$ is a martingale, and τ is a stopping time. We have $\mathbb{P}[\tau < \infty] = 1$ by recurrence of the random walk; $|X_\tau| < \max\{a, b\}$; and $\mathbb{E}[|X_\tau|] \leq \max\{a, b\}$ for each n. The optional stopping theorem applies, giving

$$\begin{aligned} 0 = \mathbb{E}[X_\tau] &= b\mathbb{P}[X_\tau = b] - a\mathbb{P}[X_\tau = -a] \\ &= b\mathbb{P}[X_\tau = b] - a(1 - \mathbb{P}[X_\tau = b]). \end{aligned}$$

Thus,

$$\mathbb{P}[X_\tau = b] = \frac{a}{a+b}.$$

Example 8.15. Let $\{X_n\}$ be a symmetric random walk on $\mathbb{Z}$ starting at $X_0 = 0$. Let $a, b \in \mathbb{N}$, and set $\tau := \min\{n \geq 0 : X_n = -a \text{ or } X_n = b\}$. Then, $\mathbb{E}[\tau] = ab$.

Note first that $M_n := X_n^2 - n$ is a martingale. As before, $\mathbb{P}[\tau < \infty] = 1$ by recurrence of the random walk. We claim that $\mathbb{P}[\tau > n] \leq e^{-cn}$ for some constant $c > 0$.

To see this, note that $Y_n := X_{\tau\wedge n}$, the random walk stopped at time τ, is a Markov chain on $\{-a, \ldots, b\}$ with three communicating classes $\{-a\}, \{b\}, \{-a+1, \ldots, b-1\}$. The first two classes are recurrent, and the last class is transient. By remark on recurrence and transience of Markov chains, the time before exit from the transient class can be bounded by a geometric random variable.

Next, note that

$$\mathbb{E}[|M_\tau|] \leq \mathbb{E}[X_\tau^2] + \mathbb{E}[\tau] \leq \max\{a^2, b^2\} + \mathbb{E}[\tau] < \infty,$$

where we used $\mathbb{E}[\tau] < \infty$ because of the exponential tail bound on τ. We thus have

$$\mathbb{E}[|M_n|\mathbb{1}_{\{\tau>n\}}] \leq \mathbb{E}\left[((\max\{a,b\})^2 + n)\mathbb{1}_{\{\tau>n\}}\right] = ((\max\{a,b\})^2 + n)\underbrace{\mathbb{P}[\tau > n]}_{\leq e^{-cn}} \to 0.$$

The optional stopping theorem applies, giving

$$\begin{aligned} 0 = \mathbb{E}[M_\tau] = \mathbb{E}[X_\tau^2] - \mathbb{E}[\tau] &= a^2\mathbb{P}[X_\tau = -a] + b^2\mathbb{P}[X_\tau = b] - \mathbb{E}[\tau] \\ &= a^2\frac{b}{a+b} + b^2\frac{a}{a+b} - \mathbb{E}[\tau] = ab - \mathbb{E}[\tau]. \end{aligned}$$

Proposition 8.16. *Let $\{M_n\}$ be a martingale with respect to $\{\mathcal{F}_n\}$. Let τ be a stopping time. Then, $\{M_{n\wedge\tau}\}$ is also a martingale with respect to $\{\mathcal{F}_n\}$.*

Proof. Note that $n \wedge \tau$ is a bounded stopping time. Thus the optional stopping theorem (version 1) applies. We check the three properties of martingales:

(i)
$$M_{n\wedge\tau} = M_n\mathbb{1}_{\{\tau>n\}} + \sum_{k=0}^{n} M_k\mathbb{1}_{\{\tau=k\}}$$
is $\mathcal{F}_n$-measurable since it is a combination of $\mathcal{F}_n$-measurable random variables.

(ii)
$$\mathbb{E}[|M_{n\wedge\tau}|] \leq \mathbb{E}[|M_n|] + \sum_{k=0}^{n}\mathbb{E}[|M_k|] < \infty.$$

(iii) Let $m < n$.
$$\mathbb{E}[M_{n\wedge\tau}|\mathcal{F}_m] = \mathbb{E}[M_\tau\mathbb{1}_{\{\tau\leq m\}}|\mathcal{F}_m] + \mathbb{E}[M_{n\wedge\tau}|\mathcal{F}_m]\mathbb{1}_{\{\tau>m\}}.$$
By $M_\tau\mathbb{1}_{\{\tau\leq m\}}$ being $\mathcal{F}_m$-measurable, we have $\mathbb{E}[M_\tau\mathbb{1}_{\{\tau\leq m\}}|\mathcal{F}_m] = M_\tau\mathbb{1}_{\{\tau\leq m\}}$. By the optional stopping theorem, $\mathbb{E}[M_{n\wedge\tau}|\mathcal{F}_m]\mathbb{1}_{\{\tau>m\}} = M_m\mathbb{1}_{\{\tau>m\}}$. Combining,
$$\mathbb{E}[M_{n\wedge\tau}|\mathcal{F}_m] = M_\tau\mathbb{1}_{\{\tau\leq m\}} + M_m\mathbb{1}_{\{\tau>m\}} = M_{m\wedge\tau}.$$

□

8.2 The Martingale Convergence Theorem

Theorem 8.17. *Let $\{M_n\}$ be a Martingale with respect to $\{\mathcal{F}_n\}$, and assume that there exists constant $C > 0$ such that $\mathbb{E}[|M_n|] \le C$ for each n. Then,*

$$\mathbb{P}\left[M_\infty := \lim_{n\to\infty} M_n \text{ exists}\right] = 1.$$

Proposition 8.18. *Other sufficient conditions for the martingale convergence theorem:*

(i) M_n is unbounded below or above uniformly in n.

(ii) The second moments of M_n are uniformly bounded in n.

Proof. Assume that there exists $K > 0$ such that $\mathbb{P}[M_n \ge -K] = 1$ for each n. Then,

$$\begin{aligned}\mathbb{E}[|M_n|] &= \mathbb{E}[M_n \mathbb{1}_{\{M_n \ge 0\}}] - \mathbb{E}[M_n \mathbb{1}_{\{M_n < 0\}}] \\ &\le \mathbb{E}[M_n \mathbb{1}_{\{M_n \ge 0\}}] + K \\ &\le \mathbb{E}[M_n + K] + K = \mathbb{E}[M_0] + 2K],\end{aligned}$$

the last expression being a constant independent of n. The bounded above case is similar.

Assume that there exists some $C > 0$ such that $\mathbb{E}[M_n^2] \le C$ for each n. Then, by Cauchy-Schwarz,

$$\mathbb{E}[|M_n|] \le \mathbb{E}[M_n^2]^{1/2} \cdot 1^{1/2} \le C^{1/2}.$$

□

Example 8.19 (M_∞ can be random). Let $\{X_n\}$ be a random walk on $\mathbb{Z}$ starting at $X_0 = 0$, and let τ be the first time it hits $-a$ or b, where $a, b \in \mathbb{N}$. Recall that $\{X_{n\wedge\tau}\}$ is a martingale such that $|X_{n\wedge\tau}| \le \max\{a, b\}$. By the martingale convergence theorem, $\lim_{n\to\infty} X_{n\wedge\tau} = X_\tau$ exists, is random, and takes values in $\{-a, b\}$.

Example 8.20 (Expectation of a martingale need not be preserved in the limit). Let $\{X_n\}$ be the random walk on $\mathbb{Z}$ starting at $X_0 = 0$. Let $\tau := \min\{n : X_n = 1\}$. Note that $\{X_{n\wedge\tau}\}$ is a martingale and $X_{n\wedge\tau} \le 1$. We have

$$\lim_{n\to\infty} X_{n\wedge\tau} = X_\tau = 1 \ne 0 = \mathbb{E}[X_0].$$

Example 8.21 (Not every martingale converges).

(i) Let $\{Y_n\}$ be iid with $\mathbb{P}[Y_n = 1] = \mathbb{P}[Y_n = -1] = 1/2$. The martingale $\sum_{k=1}^n kY_n$.

(ii) Let $\{X_n\}$ be the random walk on $\mathbb{Z}$. The limit $\lim_{n\to\infty} X_n$ does not exist by recurrence.

Example 8.22. Recall that $\sum n^{-1} = \infty$ but $\sum (-1)^n n^{-1} < \infty$. What if we randomize the signs? More precisely, let $\{Y_n\}$ be iid with $\mathbb{P}[Y_n = 1] = \mathbb{P}[Y_n = -1] = 1/2$ and consider the series $\sum Y_n n^{-1}$.

Consider the partial sums $M_n := \sum_{j=1}^n Y_j/j$. Note that $\{M_n\}$ is a martingale with respect to the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n$ is the information in $Y_1, \ldots, Y_n$. Now,

$$\mathbb{E}[M_n^2] = \sum_i \sum_j \mathbb{E}\left[\frac{Y_i Y_j}{ij}\right] = \sum_{j=1}^n \frac{\mathbb{E}[Y_j^2]}{j^2} = \sum_{j=1}^n \frac{1}{j^2} \le \sum_{j=1}^\infty \frac{1}{j^2} = \frac{\pi^2}{6} < \infty.$$

The martingale convergence theorem applies, giving that $\lim_{n\to\infty} M_n = M_\infty$ exists with probability 1. Thus, the series $\sum Y_n n^{-1}$ converges with probability 1.

The limit is a random variable, since

$$M_\infty = \sum_{j=2}^\infty \frac{Y_j}{j} + Y_1,$$

where the first term is independent of the second term, which is random.

Example 8.23 (Polya's urn). Consider an urn with red and blue balls. At time $n = 0$, we have one of each color. At time n, we draw one ball from the urn and look at the color. We return the ball to the urn along with another ball of the same color.

Let X_n be the number of red balls at time n. Note that X_n can be written as a tie inhomogeneous Markov process with transition probabilities

$$\mathbb{P}[X_n = x+1 | X_{n-1} = x] = \frac{x}{n+1}, \quad \mathbb{P}[X_n = x | X_{n-1} = x] = \frac{n+1-x}{n+1}.$$

Let $M_n := X_n/(n+2)$ be the fraction of red balls at time n. We claim that $\{M_n\}$ is a martingale with respect to the natural filtration $\{\mathcal{F}_n\}$. It is clear that M_n is $\mathcal{F}_n$-measurable and $\mathbb{E}[|M_n|] \le 1$. We check the third property:

$$\begin{aligned}
\mathbb{E}[M_n | \mathcal{F}_{n-1}] &= \frac{X_{n-1}}{n+1} \cdot \frac{X_{n-1}+1}{n+2} + \frac{n+1-X_{n-1}}{n+1} \cdot \frac{X_{n-1}}{n+2} \\
&= \frac{X_{n-1}\,(X_{n-1}+1+n+1-X_{n-1})}{(n+1)(n+2)} \\
&= \frac{X_{n-1}}{(n+1)} = M_{n-1}.
\end{aligned}$$

Since $M_n \in [0,1]$, the Martingale convergence theorem implies that $M_\infty := \lim_{n\to\infty} M_n$ exists. In fact, M_∞ has a Uniform$(0,1)$ distribution.

Example 8.24. Let X be a random variable with $\mathbb{E}[|X|] < \infty$. Let $M_n := \mathbb{E}[X|\mathcal{F}_n]$, where $\{\mathcal{F}_n\}$ is any filtration. Recall that M_n is a martingale with respect to $\{\mathcal{F}_n\}$. Further,

$$\mathbb{E}[|M_n|] \le \mathbb{E}\left[\mathbb{E}[|X||\mathcal{F}_n]\right] = \mathbb{E}[|X|] < \infty.$$

Thus by the martingale convergence theorem, $\lim_{n\to\infty} \mathbb{E}[X|\mathcal{F}_n]$ exists with probability 1.

This has an important interpretation in the context of statistics: Let X be some quantity we want to estimate, and $\mathcal{F}_n$ the information in $Y_1, \dots, Y_n$, n independent samples from the population. Then, $\mathbb{E}[X|\mathcal{F}_n]$ is the best grues for X using n samples, and $X = \mathbb{E}[X|\mathcal{F}_\infty]$.

We finally prove the Martingale convergence theorem.

Proof. If $\{M_n\}$ does not converge, then $\liminf_{n\to\infty} M_n < \limsup_{n\to\infty} M_n$. Consequently, there exists numbers a, b which we can choose to be rational, and a sequence of times

$$m_1 < n_1 < m_2 < n_2 < \dots$$

such that

$$M_{m_j} \le a, \quad M_{n_j} \ge b, \quad \forall j.$$

Note that M_n cannot blowup to infinity since $\mathbb{E}[|M_n|] \le C$ for each n. We say an **upcrossing** of $[a, b]$ is an interval $\{m, \dots, n\}$ such that $M_m \le a$ and $M_n \ge b$, with $M_k \in (a, b)$ for each $k \in \{m+1, \dots, n-1\}$.

Thus if M_n does not converge, there exists some rational $a < b$ such that there are infinitely many upcrossings of $[a, b]$. Since we chose a, b to be rational, it is thus sufficient to show the following lemma. ◻

Lemma 8.25 (Doob's Upcrossing Lemma)**.** *Fix $a < b$. Then with probability* 1*, there are only finitely many upcrossings of* $[a, b]$.

Proof. We consider the random variable defined by

$$B_0 = 1, \quad B_n = \begin{cases} B_{n-1} + 1, & M_{n-1} \le a, B_{n-1} = 0 \\ B_{n-1} - 1, & M_{n-1} \ge b, B_{n-1} = 1 \\ B_{n-1}, & \text{otherwise.} \end{cases}$$

Intuitively, we think of M_n as the price of a stock, and B_n as the number of stock we hold using the following strategy: we start with one stock, and we buy one stock when the price is at most a and we have no stock, and we sell one stock when the price is at least b and we have one stock.

Note that B_n is $\mathcal{F}_{n-1}$-measurable. Let

$$W_n := \sum_{j=1}^{n} B_j (M_j - M_{j-1})$$

denote the profit from this strategy up to time n. We claim that W_n is a martingale with respect to $\{\mathcal{F}_n\}$ (in particular, $\mathbb{E}[W_n] = 0$). Note that

$$\mathbb{E}[|W_n|] \le \sum_{j=1}^{n} \mathbb{E}[|M_j| + |M_{j-1}|]$$

and

$$\mathbb{E}[W_n|\mathcal{F}_{n-1}] = W_{n-1} + B_n \mathbb{E}[M_n - M_{n-1}|\mathcal{F}_{n-1}] = W_{n-1}.$$

Now let U_n be the number of upcrossings of $[a, b]$ by time n. Note that if $B_n = 0$, then $W_n \geq (b - a)U_n$. If $B_n = 1$ and k is the last time before n such that $B_k = 0$, we bought at time $k + 1$ and so $M_k \leq a$. Note also that $U_n = U_k$ since no new upcrossing can be completed between times k and n. We thus have the bound

$$\begin{aligned} W_n &= W_k + (M_n - M_k) \\ &\geq (b - a)U_k + \sum_{j=k+1}^{n} B_j(M_j - M_{j-1}) \\ &\geq (b - a)U_n + M_n - M_k \geq (b - a)U_n - |M_n| - |a|. \end{aligned}$$

Taking expectations, we get

$$0 = \mathbb{E}[W_0] = \mathbb{E}[W_n] \geq (b - a)\mathbb{E}[U_n] - C - |a|.$$

Rearranging gives $\mathbb{E}[U_n] \leq \frac{C+|a|}{b-a}$. By the monotone convergence theorem,

$$\mathbb{E}\left[\lim_{n\to\infty} U_n\right] \leq \frac{C + |a|}{b - a} < \infty.$$

Thus with probability 1, $\lim_{n\to\infty} U_n < \infty$, i.e., there are only finitely many upcrossings of $[a, b]$. □

Theorem 8.26 (Monotone Convergence Theorem)**.** *Let $\{X_n\}$ be nonnegative random variables in $\mathbb{R}$ such that $X_n \leq X_{n+1}$ with probability 1 for each n. Then,*

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}\left[\lim_{n\to\infty} X_n\right].$$

9 Brownian Motion

A Brownian motion is a continuous-time stochastic process $\{B_t\}_{t\geq 0}$ taking values in $\mathbb{R}$. Alternatively, we may think of $B : [0,\infty) \to \mathbb{R}$ as a random function.

It has been used to model motion of particles in water, fluctuation of stock prices, among others.

We expect the Brownian motion to have the following properties:

(i) $B_0 = 0$.

(ii) $\{B_t\}$ has stationary increments. That is, for any $0 \leq s < t$, $B_t - B_s$ has the same distribution as B_{t-s}.

(iii) $\{B_t\}$ has independent increments; That is, for any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$, the random variables $B_{t_1} - B_{s_1}, B_{t_2} - B_{s_2}, \ldots, B_{t_n} - B_{s_n}$ are independent.

(iv) $\{B_t\}$ is continuous in t.

Definition 9.1. A random variable X in $\mathbb{R}$ has the **Gaussian (normal)** distribution if its density is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

for some $\mu \in \mathbb{R}$ and $\sigma > 0$. In this case we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

Proposition 9.2 (Properties of the Normal Distribution)**.**

- *If $X \sim \mathcal{N}(0,1)$ (the **standard normal**), then $\sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$.*
- *If $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, and X, Y are independent, then $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Theorem 9.3. *Let $B : [0,\infty) \to \mathbb{R}$ be a random continuous function with independent, stationary increments, and such that $B_0 = 0$. Then there exists a $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ such that for any $t \geq 0$, $B_t \sim \mathcal{N}(\mu t, \sigma^2 t)$. Further, μ and σ^2 uniquely characterize the distribution of B_t.*

This gives the uniqueness of the Brownian motion up to two parameters μ and σ^2.

Remark 9.4. The proof of this theorem is beyond the scope of this course (it will be prove in MATH385). To get some intuition, write $B_t = \sum_{j=1}^{n} (B_{tj/n} - B_{t(j-1)/n})$, for $n \in \mathbb{N}$. By independence and stationarity, the summands are iid. Further, they have to be small when n is large, since B is continuous. We may then use a version of the central limit theorem to conclude that B_t is approximately normal for large n. ☕

Definition 9.5. The **(standard) Brownian motion** is the process $\{B_t\}_{t\geq 0}$ such that $B_0 = 0$ and

(i) B_t is continuous.

(ii) For each $s < t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$.

(iii) For each $s_1 \leq t_1 \leq \cdots \leq s_k \leq t_k$, the increments $B_{t_j} - B_{s_j}$ are independent.

Theorem 9.6. *Such a process exists.*

Proof in MATH314.

It turns out that the Brownian motion has the following properties, some of which we will prove later:

(i) The Brownian notion is nowhere differentiable with probability 1.

(ii) For each $t > 0$ and $\varepsilon > 0$, there exists $s_1, s_2 \in t, t + \varepsilon]$ such that $B_{s_1} < B_t$ and $B_{s_2} > B_t$.

(iii) $\limsup_{t\to\infty} B_t = \infty$, $\liminf_{t\to\infty} B_t = -\infty$.

Example 9.7. Let B be a standard Brownian motion. We will compute $\mathbb{P}(B_1 \geq 1, B_3 \geq B_1 + 1)$. Note that

$$\begin{aligned}\mathbb{P}(B_1 \geq 1, B_3 \geq B_1 + 1) &= \mathbb{P}(B_1 \geq 1, B_3 - B_1 \geq 1) \\ &= \mathbb{P}(B_1 \geq 1)\mathbb{P}(B_3 - B_1 \geq 1) \\ &= \mathbb{P}\big(\mathcal{N}(0,1) \geq 1\big) \cdot \mathbb{P}\big(\mathcal{N}(0,2) \geq 1\big) \\ &= (1 - \Phi(1)) \cdot \left(1 - \Phi\left(\frac{1}{\sqrt{2}}\right)\right).\end{aligned}$$

Example 9.8. $\mathbb{P}[B_t \geq 0] = 1/2$ by symmetry.

Proposition 9.9 (Brownian Scaling)**.** *Let B be a standard Brownian motion. Let $C > 0$ be a constant. Then $\left\{C^{-1/2}B_{ct}\right\}_{t\geq 0}$ is again a standard Brownian motion.*

Proof. Note first that $C^{-1/2}B_{c\dot{0}} = 0$. Next, $\left\{C^{-1/2}B_{ct}\right\}_{t\geq 0}$ inherits independent increments and continuity from $\{B_t\}$. The increments has distribution

$$C^{-1/2}(B_{ct} - B_{cs}) \sim C^{-1/2}\mathcal{N}(0, c(t-s)) = \mathcal{N}(0, t-s).$$

□

Proposition 9.10. *For each $t \geq 0$,*

$$\mathbb{P}\left[B \textit{ is not differentiable at } t\right] = 1.$$

Note that this is weaker than B being nowhere differentiable with probability 1.

Proof. Let $\varepsilon > 0$. Note that $\varepsilon^{-1/2}\left(B_{t+\varepsilon} - B_t\right) \sim \mathcal{N}(0,1)$. For $C > 0$, note that

$$\mathbb{P}\left[\left|\frac{B_{t+\varepsilon} - B_\varepsilon}{\varepsilon}\right| > C\right] = \mathbb{P}\left[\left|\varepsilon^{-1/2}\mathcal{N}(0,1)\right| > C\right] = \mathbb{P}\left[|\mathcal{N}(0,1)| > C\varepsilon^{1/2}\right],$$

which converges to 1 as $\varepsilon \to 0$. From this, we conclude that

$$\mathbb{P}\left[\limsup_{\varepsilon\to 0}\left|\frac{B_{t+\varepsilon} - B_t}{\varepsilon}\right| = \infty\right] = 1.$$

Thus B is not differentiable at t. □

Remark 9.11. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables in S. If $A \subset S$ and $\mathbb{P}[X_n \in A] = 0$ for each n, then $\mathbb{P}[\exists n : X_n \in A] \le \sum_{n=1}^{\infty} \mathbb{P}[X_n \in A] = 0$.

This does not work for uncountable collections of random variables. We have for instance $\mathbb{P}[B_t = 0] = 0$ for each t, but as we will later see, $\mathbb{P}[\exists t > 0 : B_t = 0] = 1$. ☕

Brownian motion can be thought of as the scaling limit of the random walk on $\mathbb{Z}$. The relevant limit here is that of weak convergence of probability measures:

Definition 9.12. Fix $k \in \mathbb{N}$. Let $\{(X_1^n, \dots, X_k^n)\}_{n\in\mathbb{N}}$ be a sequence of random variables in $\mathbb{R}^k$, and $(X_1, \dots, X_k)$ be another random variable in $\mathbb{R}^k$. We say $(X_1^n, \dots, X_k^n) \to (X_1, \dots, X_k)$ **in distribution** if for any $a_1 < b_1, a_2 < b_2, \dots, a_k < b_k$ such that

$$\mathbb{P}\left[X_j = a_j \text{ or } X_j = b_j\right] = 0, \quad \forall j \in \{1, \dots, k\}$$

we have

$$\mathbb{P}\left[X_1^n \in (a_1, b_1), \dots, X_k^n \in (a_k, b_k)\right] \longrightarrow \mathbb{P}\left[X_1 \in (a_1, b_1), \dots, X_k \in (a_k, b_k)\right].$$

Note that $\mathbb{P}[X_j \in \{a, b\}] = 0$ is automatic is $(X_1, \dots, X_n)$ has a continuous distribution.

Note that this notion of convergence depends only on the distributions of the random variables. It turns out that it is equivalent to requiring for each bounded and continuous $f : \mathbb{R}^k \to \mathbb{R}$ that

$$\mathbb{E}[f(X_1^n, \dots, X_k^n)] \longrightarrow \mathbb{E}[f(X_1, \dots, X_k)].$$

We may now state the convergence of the random walk to Brownian motion.

Theorem 9.13. *Let $\{X_j\}_{j\ge 0}$ be the (unbiased) random walk on $\mathbb{Z}$ with $X_0 = 0$. We extend to $X : [0, \infty) \to \mathbb{R}$ by linear interpolation. That is, define for each $t \in [j, j+1)$,*

$$X_t := (t - j)X_{j+1} + (j + 1 - t)X_j.$$

Then, $\{n^{-1/2} X_{nt}\}_{t>0}$ converges in distribution to the Brownian motion in the following sense: for each $t_1 \le t_2 \le \dots \le t_k$, we have

$$\left(n^{-1/2} X_{nt_1}, \dots, n^{-1/2} X_{nt_k}\right) \longrightarrow \left(B_{t_1}, \dots, B_{t_k}\right)$$

in distribution.

Remark 9.14.

- Note that this can be thought of as a generalization of the central limit theorem to functions.
- There exists stronger versions of this theorem. In particular, one can show that the convergence holds in the space of continuous functions equipped with the uniform norm on compact sets.
- From this it should not be surprising that Brownian motion acts like a random walk, as we will see.

☕

Proof. Recall that $\{X_j - X_{j-1}\}_{j\ge 1}$ are iid with mean 0 and variance 1. □