Stochastic Processes

Derek Li

Contents

1 Review

1.1 Basic Probability Theory

Property 1.1. If Z is non negative integer valued, then

$$\mathbb{E}[Z] = \sum_{k=1}^{\infty} P(Z \geqslant k).$$

Property 1.2. Probabilities are monotone, i.e.,

$$A \subseteq B \Rightarrow P(A) \leqslant P(B).$$

1.2 Standard Probability Distribution

Definition 1.1. The ***continuous uniform distribution*** on an interval $[L, R]$ where $L < R$ gives probability $\frac{b-a}{R-L}$ to $[a, b]$ whenever $L \leqslant a \leqslant b \leqslant R$ with mean $\frac{a+b}{2}$.

Example 1.1. Let $a = 0, b = y$, then

$$P(X \leqslant y) = \min(1, y), \forall y \geqslant 0.$$

Definition 1.2. The ***normal distribution*** $\mathcal{N}(m, v)$ with mean m and variance v has density

$$f(x) = \frac{1}{\sqrt{2\pi v}} e^{-(x-m)^2/2v}.$$

Definition 1.3. The ***standard normal distribution***$\mathcal{N}(0, 1)$ has density

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

1.3 Infinite Series and Limit

Property 1.3. If $x_n \geqslant 0$ and $\sum\limits_{n=1}^{\infty} x_n < \infty$, then $\lim\limits_{n\to\infty} x_n = 0$.

Property 1.4. It is possible that $\sum\limits_{n=1}^{\infty} x_n = \infty$ even if $\lim\limits_{n\to\infty} x_n = 0$.

Property 1.5 (Cesaro Sum Principle)**.** If $\lim\limits_{n\to\infty} x_n = r$, then $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=1}^{n} x_i = r$.

Property 1.6 (Squeeze's Theorem)**.** If $\{a_n\}, \{b_n\}$ and $\{c_n\}$ are three sequences with $a_n \leqslant b_n \leqslant c_n$ and $\lim\limits_{n\to\infty} a_n = \lim\limits_{n\to\infty} c_n = L$, then $\lim\limits_{n\to\infty} b_n = L$.

Definition 1.4. A quantity is $O(\cdot)$ if in a certain limit, it is bounded above by a finite constant times "$\cdot$".

1.4 Bounded, Finite, and Infinite

Property 1.7. If $P(|X| = \infty) > 0$, then $\mathbb{E}|X| = \infty$.

1.5 Conditioning

Definition 1.5. If $P(B) > 0$, then the ***conditional probability*** of A given B is

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

from which it follows that

$$P(A \cap B) = P(B)P(A|B).$$

If Y is a discrete r.v., and $P(Y = y) > 0$, then

$$P(A|Y = y) = \frac{P(A, Y = y)}{P(Y = y)}.$$

Definition 1.6. If X is discrete, with finite expectation, then we define its ***conditional expectation*** as

$$\mathbb{E}[X|A] = \sum_x xP(X = x|A).$$

If X and Y are discrete, and $P(Y = y) > 0$, then

$$\mathbb{E}[X|Y = y] = \sum_x xP(X = x|Y = y).$$

Property 1.8 (Law of Total Expectation)**.** If X and Y are discrete random variables, then

$$\mathbb{E}[X] = \sum_y P(Y = y)\mathbb{E}[X|Y = y].$$

Property 1.9. If $X = \mathbf{1}_A$, then

$$P(A) = \sum_y P(Y = y)P(A|Y = y).$$

Property 1.10 (Double-Expectation Formula)**.** $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$, i.e., the r.v. $\mathbb{E}[X|Y]$ equals X on average.

Property 1.11 (Conditional Factoring)**.** If $h(Y)$ is a function of Y, then

$$\mathbb{E}[h(Y)X|Y] = h(Y)\mathbb{E}[X|Y],$$

i.e., when conditioning on Y, we can treat any function of Y as a constant and factor it out.

Property 1.12 (Memoryless Property of the Exponential Distribution)**.** If $X \sim \text{Exponential}(\lambda)$ and $a, b > 0$, then

$$P(X > b + a|X > a) = P(X > b).$$

1.6 Convergence of Random Variable

Definition 1.7. ***Convergence in distribution*** iff $\forall a < b$,

$$\lim_{n\to\infty} P(a < X_n < b) = P(a < X < b).$$

Definition 1.8. ***Weak convergence*** means $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \geqslant \varepsilon) = 0.$$

Definition 1.9. ***Strong convergence*** or ***convergence w.p. 1*** means

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

i.e., the random sequence of values always converge.

Property 1.13 (Law of Large Numbers)**.** If the sequence $\{X_n\}$ is i.i.d. with mean m, then the sequence $\frac{1}{n}\sum\limits_{i=1}^{n} X_i$ converges to m (both weakly and strongly), i.e.,

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

Property 1.14 (Central Limit Theorem)**.** If the sequence $\{X_n\}$ is i.i.d., with mean m and finite variance v, then the sequence $\frac{1}{\sqrt{nv}}\sum\limits_{i=1}^{n}(X_i - m)$ converges in distribution to $\mathcal{N}(0,1)$, i.e., for any $a < b$,

$$\lim_{n\to\infty} P\left(a < \frac{1}{\sqrt{nv}}\sum_{i=1}^{n}(X_i - m) < b\right) = \int_a^b \phi(x)\mathrm{d}x,$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is the $\mathcal{N}(0,1)$ density.

1.7 Continuity of Probability

Property 1.15.

$$\lim_{n\to\infty} P(X \geqslant n) = P(X = \infty).$$

1.8 Exchanging Sum and Expectation

Property 1.16 (Countable Linearity)**.** If $\{Y_n\}$ is a sequence of non negative random variables, then

$$\sum_{n=1}^{\infty} \mathbb{E}[Y_n] = \mathbb{E}\left[\sum_{n=1}^{\infty} Y_n\right].$$

Property 1.17. If x_{nk} are non negative real numbers, then

$$\sum_n \sum_k x_{nk} = \sum_k \sum_n x_{nk}.$$

1.9 Exchanging Expectation and Limit

Property 1.18 (Bounded Convergence Theorem)**.** If $X_n \to X$ w.p. 1, and the $\{X_n\}$ are uniformly bounded, i.e., there is $M < \infty$ with $|X_n| \leqslant M, \forall n$, then $\lim\limits_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Property 1.19 (Monotone Convergence Theorem)**.** If $X_n \to X$ w.p. 1 and $0 \leqslant X_1 \leqslant X_2 \leqslant \cdots$, then $\lim\limits_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

Property 1.20 (Dominated Convergence Theorem)**.** If $X_n \to X$ w.p. 1 and there is some r.v. Y with $\mathbb{E}[|Y|] < \infty$ and $|X_n| \leqslant Y, \forall n$, then $\lim\limits_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X]$.

1.10 Exchanging Limit and Sum

Property 1.21 (M-Test)**.** Let $\{x_{nk}\}_{n,k\in\mathbb{N}}$ be real numbers. Suppose that $\lim\limits_{n\to\infty} x_{nk}$ exists for each fixed $k \in \mathbb{N}$, and that $\sum\limits_{k=1}^{\infty} \sup\limits_{n} |x_{nk}| < \infty$. Then

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} x_{nk} = \sum_{k=1}^{\infty} \lim_{n\to\infty} x_{nk}.$$

1.11 Basic Linear Algebra

1.12 Mathematical Fact

Property 1.22 (Stirling's Approximation)**.** If n is large, then

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n},$$

or

$$\lim_{n\to\infty} \frac{n!}{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}} = 1.$$

Property 1.23 (Number Theory Lemma)**.** If a set A of positive integers is non-empty, and satisfies additivity and $\gcd(A)$, then $\exists n_0 \in \mathbb{N}$ s.t. $n \in A, \forall n \geqslant n_0$.

2 Markov Chain Probability

2.1 Markov Chain

Definition 2.1. A discrete-time, discrete-space, time-homogeneous ***Markov chain*** is specified by three ingredients:

(1) A ***state space*** S, any non-empty finite or countable set.

(2) ***Initial probabilities*** $\{v_i\}_{i\in S}$, where v_i is the probability of starting at i (at time 0). (So $v_i \geqslant 0$ and $\sum\limits_i v_i = 1$.)

(3) ***Transition probabilities*** $\{p_{ij}\}_{i,j\in S}$, where p_{ij} is the probability of jumping to j if you start at i. (So, $p_{ij} \geqslant 0$ and $\sum\limits_j p_{ij} = 1, \forall i$.)

Note. (1) Given any Markov chain, let X_n be the Markov chain's state at time n and thus $X_0, X_1, \cdots$ are random variables.
(2) At time 0, we have $P(X_0 = i) = v_i, \forall i \in S$.
(3) p_{ij} can be interpreted as conditional probabilities, i.e., if $P(X_n = i) > 0$, then

$$P(X_{n+1} = j | X_n = i) = p_{ij}, \forall i, j \in S, n = 0, 1, \cdots,$$

which does not depend on n because of time-homogeneous property.
(4) The probabilities at time $n+1$ depend only on the state at time n, i.e.,

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

which is called the Markov property.
(5) The joint probabilities can be computed by relating them to conditional probabilities:

$$\begin{aligned} P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) &= P(X_0 = i_0)P(X_1 = i_1 | X_0 = i_0) \cdots P(X_n = i_n | X_{n-1} = i_{n-1}) \\ &= v_{i0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}, \end{aligned}$$

which completely defines the probabilities of the sequence $\{X_n\}_{n=0}^{\infty}$. The random sequence $\{X_n\}_{n=0}^{\infty}$ is the Markov chain.

Example 2.1 (Bernoulli Process)**.** Let $0 < p < 1$. Suppose repeatedly flip a p-coin at times $1, 2, \cdots$. Let X_n be the number of heads on the first n flips, then $\{X_n\}$ is a Markov chain, with $S = \{0, 1, \cdots\}, X_0 = 0$ (i.e., $v_0 = 1$ and $v_i = 0, \forall i \neq 0$), and

$$p_{ij} = \begin{cases} p, & j = i+1 \\ 1-p, & j = i \\ 0, & \text{otherwise} \end{cases}.$$

Example 2.2 (Simple Random Walk)**.** Let $0 < p < 1$. Suppose repeatedly bet \$1. Each time, you have probability p of winning \$1 and probability $1-p$ of losing \$1. Let X_n be the net gain after n bets, then $\{X_n\}$ is a Markov chain, with $S = \mathbb{Z}, X_0 = a$ for some $a \in \mathbb{Z}$ (i.e., $v_a = 1$), and

$$p_{ij} = \begin{cases} p, & j = i+1 \\ 1-p, & j = i-1 \\ 0, & \text{otherwise} \end{cases}.$$

If $p = \frac{1}{2}$, we call it simple symmetric random walk since $p = 1 - p$.

Example 2.3 (Ehrenfest's Urn)**.** Suppose we have d balls, divided into two urns. At each time, we choose one of the d balls uniformly at random, and move it to the other urn. Let X_n be the number of balls in Urn 1 at time n, then $\{X_n\}$ is a Markov chain, with $S = \{0, 1, \cdots, d\}$, and

$$p_{ij} = \begin{cases} \frac{i}{d}, & j = i - 1 \\ \frac{d-i}{d}, & j = i + 1 \\ 0, & \text{otherwise} \end{cases}.$$

2.2 Multi-Step Transition

Let $\mu_i^{(n)} = P(X_n = i)$ be the probabilities at time n : at time 0, $\mu_i^{(0)} = P(X_0 = i) = v_i$; at time 1, $\mu_j^{(1)} = P(X_1 = j) = \sum\limits_{i \in S} P(X_0 = i, X_1 = j) = \sum\limits_{i \in S} v_i p_{ij} = \sum\limits_{i \in S} \mu_i^{(0)} p_{ij}$ by the Law of Total Probability; at time 2, $\mu_k^{(2)} = \sum\limits_{i \in S} \sum\limits_{j \in S} v_i p_{ij} p_{jk}$, etc.

Let $m = |S|$ be the number of elements in S (could be infinity), $v = (v_1, v_2, \cdots, v_m)$ be a $1 \times m$ row vector, $\mu^{(n)} = (\mu_1^{(n)}, \mu_2^{(n)}, \cdots, \mu_m^{(n)})$ be a $1 \times m$ row vector, and

$$P = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}$$

be an $m \times m$ matrix. Therefore, in matrix form: $\mu^{(1)} = vP = \mu^{(0)} P, \mu^{(2)} = vPP = vP^2 = \mu^{(0)} P^2$. By induction, we have

$$\mu^{(n)} = vP^n = \mu^{(0)} P^n, n \in \mathbb{N}.$$

By convention, let $P^0 = I$, then $\mu^{(n)} = vP^n$ holds for $n = 0$.

Another way to track the probabilities of a Markov chain is with n-step transitions

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

Since the chain is time-homogeneous, $p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i), \forall m \in \mathbb{N}$. Note that $p_{ij}^{(n)} \geqslant 0$ and $\sum\limits_{j \in S} p_{ij}^{(n)} = \sum\limits_{j \in S} P_i(X_n = j) = P_i(X_n \in S) = 1$. We have $p_{ij}^{(1)} = P(X_1 = j | X_0 = i) = p_{ij}$, and $p_{ij}^{(2)} = P(X_2 = j | X_0 = i) = \sum\limits_{k \in S} P(X_2 = j, X_1 = k | X_0 = i) = \sum\limits_{k \in S} p_{ik} p_{kj}, p_{ij}^{(3)} = \sum\limits_{k \in S} \sum\limits_{l \in S} p_{ik} p_{kl} p_{lj}$, etc. Therefore, in matrix form: $P^{(2)} = (p_{ij}^(2)) = PP = P^2, P^{(3)} = P^3$. By induction we have

$$P^{(n)} = P^n, n \in \mathbb{N}.$$

By convention, let $P^{(0)} = I$, then $P^{(n)} = P^n$ holds for $n = 0$.

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

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

Proof. By the Law of Total Probability,

$$p_{ij}^{(m+n)} = P(X_{m+n} = j | X_0 = i) = \sum_{k \in S} P(X_{m+n} = j, X_m = k | X_0 = i) = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}.$$

□

In matrix form: $P^{(m+n)} = P^{(m)}P^{(n)}, P^{(m+s+n)} = P^{(m)}P^{(s)}P^{(n)}$, etc.

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

$$p_{ij}^{(m+n)} \geqslant p_{ik}^{(m)} p_{kj}^{(n)},$$

for any fixed state $k \in S$, etc.

2.3 Recurrence and Transience

Let $N(i) = |\{n \geqslant 1 : X_n = i\}|$ be the total number of times that the chain hits i (not counting time 0) and so $N(i)$ is a random variable, possibly infinite. Let f_{ij} be the ***return probability*** from i to j, i.e., f_{ij} is the probability, starting from i, that the chain will eventually visit j at least once:

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

Thus, we have

$$1 - f_{ij} = P_i(X_n \neq j \text{ for all } n \geqslant 1).$$

Also, we have

$$P_i(\text{Chain will eventually visit } j, \text{ and then eventually visit } k) = f_{ij} f_{jk}, \text{ etc.}$$

Hence, $P_i(N(i) \geqslant k) = (f_{ii})^k, P_i(N(j) \geqslant k) = f_{ij}(f_{jj})^{k-1}$.

Property 2.1. $f_{ik} \geqslant f_{ij} f_{jk}$, etc.

Definition 2.2. A state i of a Markov chain is ***recurrent*** or ***persistent*** if

$$P_i(X_n = i \text{ for some } n \geqslant 1) = 1, \text{ i.e., } f_{ii} = 1.$$

Otherwise, if $f_{ii} < 1$, then i is ***transient***.

Theorem 2.3 (Recurrent State Theorem)**.** State i is recurrent iff $P_i(N(i) = \infty) = 1$ iff $\sum\limits_{n=1}^{\infty} p_{ii}^{(n)} = \infty$. State i is transient iff $P_i(N(i) = \infty) = 0$ iff $\sum\limits_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Proof. We have

$$P_i(N(i) = \infty) = \lim_{k \to \infty} P_i(N(i) \geqslant k) = \lim_{k \to \infty} (f_{ii})^k = \begin{cases} 1, & f_{ii} = 1 \\ 0, & f_{ii} < 1 \end{cases}.$$

Also, we have

$$\begin{aligned} \sum_{n=1}^{\infty} p_{ii}^{(n)} &= \sum_{n=1}^{\infty} P_i(X_n = i) = \sum_{n=1}^{\infty} \mathbb{E}_i[\mathbf{1}_{X_n = i}] = \mathbb{E}_i\left[\sum_{n=1}^{\infty} \mathbf{1}_{X_n = i}\right] \\ &= \mathbb{E}_i[N(i)] = \sum_{k=1}^{\infty} P_i(N(i) \geqslant k) = \sum_{k=1}^{\infty} (f_{ii})^k = \begin{cases} \infty, & f_{ii} = 1 \\ \frac{f_{ii}}{1 - f_{ii}} < \infty, & f_{ii} < 1 \end{cases}. \end{aligned}$$

□

Corollary 2.1.

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

Example 2.4 (Simple Random Walk)**.** State 0 is recurrent only if $p = \frac{1}{2}$.

Proof. If n is odd, then $p_{00}^{(n)} = 0$.

If n is even,

$$p_{00}^{(n)} = \binom{n}{\frac{n}{2}} p^{\frac{n}{2}} (1-p)^{\frac{n}{2}} = \frac{n!}{\left[\left(\frac{n}{2}\right)!\right]^2} p^{\frac{n}{2}} (1-p)^{\frac{n}{2}}.$$

By Stirling's approximation, we have

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}, \left(\frac{n}{2}\right)! \approx \left(\frac{n}{2e}\right)^{\frac{n}{2}} \sqrt{\pi n}$$

and thus

$$p_{00}^{(n)} \approx [4p(1-p)]^{\frac{n}{2}} \sqrt{\frac{2}{\pi n}}.$$

If $p = \frac{1}{2}$, then

$$\sum_{n=1}^{\infty} p_{00}^{(n)} \approx \sqrt{\frac{2}{\pi}} \sum_{n=2,4,6,\cdots} n^{-\frac{1}{2}} = \infty.$$

If $p \neq \frac{1}{2}$, then $4p(1-p) < 1$ and thus

$$\sum_{n=1}^{\infty} p_{00}^{(n)} \approx \sum_{n=2,4,6,\cdots} [4p(1-p)]^{\frac{n}{2}} \sqrt{\frac{2}{\pi n}} < \sum_{n=2,4,6,\cdots} [4p(1-p)]^{\frac{n}{2}} < \infty.$$

Therefore, if $p = \frac{1}{2}$. then state 0 is recurrent and the chain will return to state 0 infinitely often with probability 1; if $p \neq \frac{1}{2}$, then state 0 is transient and the chain will note return to state 0 infinitely often. □

Property 2.2 (f-Expansion)**.**

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

Proof. We have

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

□

Corollary 2.2. $f_{ij} \geqslant p_{ij}$.

2.4 Communicating States and Irreducibility

Definition 2.3. State i ***communicates*** with state j, written $i \to j$, if $f_{ij} > 0$, i.e., if it is possible to get from i to j. $f_{ij} > 0$ iff $\exists m \geqslant 1$ s.t. $p_{ij}^{(m)} > 0$, i.e., there is some time m for which it is possible to get from i to j in m steps.

We will write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Definition 2.4. A Markov chain is ***irreducible*** if $i \to j$ for all $i, j \in S$, i.e., if $f_{ij} > 0, \forall i, j \in S$. Otherwise, it is reducible.

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

Proof. Since $i \to k, l \to j$, then $\exists m, r \geqslant 1$ s.t. $p_{ik}^{(m)}, p_{lj}^{(r)} > 0$. By Chapman-Kolmogorov inequality, we have $p_{ij}^{(m+s+r)} \geqslant p_{ik}^{(m)} p_{kl}^{(s)} p_{lj}^{(r)}$. Thus,

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} \geqslant \sum_{n=m+1+r}^{\infty} p_{ij}^{(n)} = \sum_{s=1}^{\infty} p_{ij}^{(m+s+r)} \geqslant \sum_{s=1}^{\infty} p_{ik}^{(m)} p_{kl}^{(s)} p_{lj}^{(r)} = \infty.$$

□

Corollary 2.3 (Sum Corollary)**.** If $i \leftrightarrow k$, then i is recurrent iff k is recurrent.

Proof. By Sum Lemma, we have if $i \to k, k \to i$ and then

$$\sum_{n=1}^{\infty} p_{kk}^{(n)} = \infty \Leftrightarrow \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty.$$

□

Theorem 2.4 (Cases Theorem)**.** For an irreducible Markov chain, either

(a) $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty, \forall i, j \in S$, and all states are recurrent;

or (b) $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} < \infty, \forall i, j \in S$, and all states are transient.

Example 2.5 (Simple Random Walk)**.** Simple random walk is irreducible. If $p = \frac{1}{2}$, state 0 is recurrent, then all states are recurrent and $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty, \forall i, j \in S$. If $p \neq \frac{1}{2}$, state 0 is transient, then all states are transient and $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} < \infty, \forall i, j \in S$.

Theorem 2.5 (Finite Space Theorem)**.** An irreducible Markov chain on a finite state space always falls into case (a), i.e., $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty, \forall i, j \in S$ and all states are recurrent.

Proof. Choose any state $i \in S$. Since $\sum\limits_{j \in S} p_{ij}^{(n)} = 1$, we have

$$\sum_{j \in S} \sum_{n=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{j \in S} p_{ij}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty.$$

Since S is finite, then we have at least one $j \in S$ s.t. $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ and thus we must be in case (a). □

Lemma 2.2 (Hit Lemma)**.** Let H_{ij} be the event that the chain hits the state i before returning to j, i.e.,

$$H_{ij} = \{\exists n \in \mathbb{N} : X_n = i, X_m \neq j, 1 \leqslant m \leqslant n-1\}.$$

If $j \to i$ with $j \neq i$, then $P_j(H_{ij}) > 0$. In other words, if it is possible to get from j to i at all, then it is possible to get from j to i without first returning to j.

Proof (optional). Since $j \to i$, there is some possible path R from j to i, i.e., $\exists m \in \mathbb{N}$ and $x_0, \cdots, x_m$ s.t. $x_0 = j, x_m = i, p_{x_r x_{r+1}} > 0, \forall 0 \leqslant r \leqslant m-1$. Let $S = \max\{r : x_r = j\}$ be the last time path R hits j, then $x_S, x_{S+1}, \cdots, x_m$ is a possible path which goes from j to i without first returning to j. So $P_j(H_{ij}) \geqslant P_j(R) = p_{x_S x_{S+1}} p_{x_{S+1} x_{S+2}} \cdots p_{x_{m-1} x_m} > 0$. □

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

Proof. If $i = j$, it is obvious. Now assume $i \neq j$. Since $j \to i$, we have $P_j(H_{ij}) > 0$. But one way to never return to j is to first hit i and then from i never return to j, i.e.,

$$P_j(\text{Never return to } j) = 1 - f_{jj} \geqslant P_j(H_{ij}) P_i(\text{Never return to } j) = P_j(H_{ij})(1 - f_{ij}).$$

Since $f_{jj} = 1$, then

$$P_j(H_{ij})(1 - f_{ij}) = 0.$$

Since $P_j(H_{ij}) > 0$, then $1 - f_{ij} = 0 \Rightarrow f_{ij} = 1$. □

Lemma 2.4 (Infinite Returns Lemma)**.** For an irreducible Markov chain, if it is recurrent then $P_i(N(j) = \infty) = 1, \forall i, j \in S$; if it is transient then $P_i(N(j) = \infty) = 0, \forall i, j \in S$.

Proof. If the chain is recurrent, then $f_{ij} = f_{jj} = 1$. We have

$$P_i(N(j) = \infty) = \lim_{k \to \infty} P_i(N(j) \geqslant k) = \lim_{k \to \infty} f_{ij}(f_{jj})^{k-1} = 1.$$

If the chain is transient, then $f_{jj} < 1$, then similarly, $P_i(N(j) = \infty) = 0$. □

Theorem 2.6 (Recurrence Equivalences Theorem)**.** If a chain is irreducible, the following are equivalent:

(1) $\exists k, l \in S$ s.t. $\sum_{n=1}^{\infty} p_{kl}^{(n)} = \infty$.
(2) $\forall i, j \in S, \sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty$.
(3) $\exists k \in S$ s.t. $f_{kk} = 1$, i.e., k is recurrent.
(4) $\forall j \in S, f_{jj} = 1$, i.e., all states are recurrent.
(5) $\forall i, j \in S, f_{ij} = 1$.
(6) $\exists k, l \in S$ s.t. $P_k(N(l) = \infty) = 1$.
(7) $\forall i, j \in S, P_i(N(j) = \infty) = 1$.

Proof. We have (1) $\Rightarrow$ (2) : sum lemma; (2) $\Rightarrow$ (4) : recurrent state theorem; (4) $\Rightarrow$ (5) : f-lemma; (5) $\Rightarrow$ (3) : immediate; (3) $\Rightarrow$ (1) : recurrent state theorem with $l = k$; (4) $\Rightarrow$ (7) : infinite returns lemma; (7) $\Rightarrow$ (6) : immediate; (6) $\Rightarrow$ (3) : infinite returns lemma. □

Theorem 2.7 (Transience Equivalences Theorem)**.** If a chain is irreducible, the following are equivalent:

(1) $\forall k, l \in S, \sum_{n=1}^{\infty} p_{kl}^{(n)} < \infty$.
(2) $\exists i, j \in S$ s.t. $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$.
(3) $\forall k \in S, f_{kk} < 1$, i.e., k is transient.
(4) $\exists j \in S$ s.t. $f_{jj} = 1$, i.e., some state is transient.
(5) $\exists i, j \in S$ s.t. $f_{ij} < 1$.
(6) $\forall k, l \in S, P_k(N(l) = \infty) = 0$.
(7) $\exists i, j \in S$ s.t. $P_i(N(j) = \infty) = 0$.

Example 2.6 (Simple Symmetric Random Walk)**.** We know the simple symmetric ($p = \frac{1}{2}$) random walk is recurrent and thus $f_{ij} = 1, \forall i, j \in S$, i.e., for any conceivable pattern of values, with probability 1, the chain will eventually hit each of them in sequence. We say the chain has ***infinite fluctuations***.

Property 2.3 (Closed Subset)**.** Suppose a chain is reducible, but has a closed subset $C \subseteq S$ (i.e., $p_{ij} = 0, i \in C, j \notin C$), on which it is irreducible (i.e., $i \to j, \forall i, j \in C$). Then the recurrence equivalences theorem and all results about irreducible chains stall apply to the chain restricted to C.

Example 2.7. For simple random walk with $p > \frac{1}{2}, f_{ij} = 1$ whenever $j > i$. Or if $p < \frac{1}{2}$ and $j < i$, then $f_{ij} = 1$.

Proof. Let $X_0 = 0$ and $Z_n = X_n - X_{n-1}$ for $n \in \mathbb{N}$. By construction, $X_n = \sum_{i=1}^{n} Z_i$. $\{Z_n\}$ are i.i.d. with $P(Z_n = +1) = p$ and $P(Z_n = -1) = 1 - p$. Thus by the law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n}(Z_1 + \cdots + Z_n) = \mathbb{E}[Z_1] = p \cdot 1 + (1-p) \cdot (-1) = 2p - 1 > 0.$$

Therefore,

$$\lim_{n\to\infty} Z_1 + \cdots + Z_n = \infty \Rightarrow X_n - X_0 \to \infty \Rightarrow X_n \to \infty.$$

So starting from i, the chain will converge to ∞. If $i < j$, then to go from i to ∞, the chain must pass through j, i.e., $f_{ij} = 1$. $\square$

2.5 Application: Gambler's Ruin

Let $0 < a < c$ be integers, and $0 < p < 1$. Suppose player A starts with a dollars, B starts with $c - a$ dollars and they repeatedly bet. At each bet, A wins \$1 from B with probability p or B wins \$1 from A with probability $1 - p$. If X_n is the amount of money that A has at time n, then $X_0 = a$ and $\{X_n\}$ follows a simple random walk. Let $T_i = \inf\{n \geqslant 0 : X_n = i\}$ be the first time A has i dollars.

First consider what is the probability that A reaches c dollars before losing all their money, i.e., $P_a(T_c < T_0)$.

Write $P_a(T_c < T_0)$ as $s(a)$ and consider it to be a function of the player's initial fortune a. It is obvious that $s(0) = 0, s(c) = 1$. On the first bet, for $1 \leqslant a \leqslant c - 1$,

$$\begin{aligned} s(a) &= P_a(T_c < T_0) = P_a(T_c < T_0, X_1 = X_0 + 1) + P_a(T_c < T_0, X_1 = X_0 - 1) \\ &= P(X_1 = X_0 + 1)P_a(T_c < T_0 | X_1 = X_0 + 1) + P(X_1 = X_0 - 1)P_a(T_c < T_0 | X_1 = X_0 - 1) \\ &= ps(a+1) + (1-p)p(a-1). \end{aligned}$$

Therefore,

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

Let $x = s(1)$, then $s(1) - s(0) = x, s(2) - s(1) = \frac{1-p}{p}x, \cdots$. In general, for $0 \leqslant a \leqslant c-1$, we have $s(a+1) - s(a) = \left(\frac{1-p}{p}\right)^a x$ and thus for $0 \leqslant a \leqslant c$,

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

Since $s(c) = 1$, then

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

We then obtain Gambler's Ruin formula:

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

We can also consider $r_{c,p}(a) = P_a(T_0 < T_c) = P_a(\text{Ruin})$:

$$r_{c,p}(a) = s_{c,1-p}(c-a) = \begin{cases} \frac{c-a}{c}, & p = \frac{1}{2} \\ \frac{\left(\frac{1-p}{p}\right)^{c-a} - 1}{\left(\frac{1-p}{p}\right)^c - 1}, & p \neq \frac{1}{2} \end{cases}.$$

Now we consider $P_a(T_0 < \infty)$, the probability of eventual ruin:

$$\begin{aligned} P_a(T_0 < \infty) &= \lim_{K \to \infty} P_a(T_0 < K) = \lim_{c \to \infty} P_a(T_0 < T_c) = \lim_{c \to \infty} r_{c,p}(a) \\ &= \begin{cases} 1, & p \leqslant \frac{1}{2} \\ \left(\frac{1-p}{p}\right)^{-a}, & p > \frac{1}{2} \end{cases}. \end{aligned}$$

Thus, eventual ruin is certain if $p \leqslant \frac{1}{2}$.

Finally, we consider the time $T = \min(T_0, T_c)$ when the Gambler's Ruin game ends.

Property 2.4. $P(T > mc) \leqslant (1-p^c)^m, P(T = \infty) = 0$, and $\mathbb{E}[T] < \infty$.

Proof. If the player ever wins c bets in a row, then the game must be over. But if $T > mc$, then the player has failed to win c bets in a row, despite having m independent attempts to do so. The probability of winning c bets in a row is p^c and the probability of failing to win c bets in a row is $1 - p^c$. Thus, the probability of failing on m independent attempts is $(1-p^c)^m$ and

$$P(T > mc) \leqslant (1-p^c)^m.$$

Then

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

And

$$\begin{aligned} \mathbb{E}[T] &= \sum_{i=1}^{\infty} P(T \geqslant i) \leqslant \sum_{i=0}^{\infty} P(T \geqslant i) \\ &\leqslant P(T \geqslant 0) + \cdots + P(T \geqslant 0) + P(T \geqslant c) + \cdots = \sum_{j=0}^{\infty} cP(T \geqslant cj) \\ &\leqslant \sum_{j=0}^{\infty} c(1-p^c)^j = \frac{c}{1-(1-p^c)} = \frac{c}{p^c} < \infty. \end{aligned}$$

Hence, with probability 1, the Gambler's Ruin game must eventually end, and the time it takes to end has finite expected value. ∎

3 Markov Chain Convergence

3.1 Stationary Distributions

Definition 3.1. If π is a probability distribution on S (i.e., $\pi_i \geqslant 0, \forall i \in S, \sum_{i\in S} \pi_i = 1$), then π is ***stationary*** for a Markov chain with transition probabilities (p_{ij}) is $\sum_{i\in S} \pi_i p_{ij} = \pi_j, \forall j \in S$. In matrix notation: $\pi P = \pi$ or π is a left eigenvector for the matrix P with eigenvalue 1.

Intuitively, if the chain starts with probabilities $\{\pi_i\}$, then it will keep the same probabilities one time unit later.

Example 3.1. Suppose $|S| < \infty$, we sat a chain is doubly stochastic if $\sum_{i\in S} p_{ij} = 1, \forall j \in S$ (in addition to the usual condition that $\sum_{j\in S} p_{ij} = 1, \forall i \in S$). Let π be the uniform distribution on S, i.e., $\pi_i = \frac{1}{|S|}, \forall i \in S$. Then

$$\sum_{i\in S} \pi_i p_{ij} = \frac{1}{|S|} \sum_{i\in S} p_{ij} = \frac{1}{|S|} = \pi_j, \forall j \in S.$$

Thus, $\{\pi_i\}$ is stationary.

3.2 Searching for Stationarity

Definition 3.2. A Markov chain is ***reversible*** (or ***time reversible***, or satisfies ***detailed balance***) with respect to a probability distribution $\{\pi_i\}$ if $\pi_i p_{ij} = \pi_j p_{ji}, \forall i, j \in S$.

Property 3.1. If a chain is reversible with respect to π, then π is a stationary distribution.

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

$$\sum_{i\in S} \pi_i p_{ij} = \sum_{i\in S} \pi_j p_{ji} = \pi_j \sum_{i\in S} p_{ji} = \pi_j \cdot 1 = \pi_j.$$

□

Note that the converse is false: it is possible for a chain to have a stationary distribution if it is not reversible.

Example 3.2 (Ehrenfest's Urn)**.** Let

$$\pi_i = \binom{d}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{d-i} = 2^{-d} \frac{d!}{i!(d-i)!},$$

since $\pi_i \geqslant 0$ and $\sum_i \pi_i = 1$, then π is a distribution. To check if π stationary, we can check the reversibility, i.e., to check if $\pi_i p_{ij} = \pi_j p_{ji}, \forall i, j \in S$.

Clearly, both sides are 0 unless $j = i + 1$ or $j = i - 1$.

If $j = i + 1$, then

$$\pi_i p_{ij} = 2^{-d} \frac{d!}{i!(d-i)!} \frac{d-i}{d} = 2^{-d} \frac{(d-1)!}{i!(d-i-1)!}$$

and
$$\pi_j p_{ji} = 2^{-d}\frac{d!}{j!(d-j)!}\frac{j}{d} = \frac{(d-1)!}{(j-1)!(d-j)!} = \frac{(d-1)!}{i!(d-i-1)!} = \pi_i p_{ij}.$$
If $j = i-1$, similarly, we have $\pi_i p_{ij} = \pi_j p_{ji}$. Hence, it is reversible w.r.t. π and thus π is a stationary distribution.

Property 3.2 (Vanishing Probabilities Proposition)**.** IF a Markov chain's transition probabilities have $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0, \forall i,j \in S$, then the chain does not have a stationary distribution.

Proof. If there were a stationary distribution π, then we would have $\pi_j = \sum\limits_{i\in S} \pi_i p_{ij}^{(n)}$ for any n and thus
$$\pi_j = \lim_{n\to\infty} \pi_j = \lim_{n\to\infty} \pi_j p_{ij}^{(n)}.$$
By M-test, we have
$$\pi_j = \lim_{n\to\infty}\sum_{i\in S}\pi_i p_{ij}^{(n)} = \sum_{i\in S}\lim_{n\to\infty}\pi_i p_{ij}^{(n)} = \sum_{i\in S} 0 = 0, \forall j.$$
But then $\sum\limits_j \pi_j = 0$, which is a contraction. Therefore, there is no stationary distribution. □

Lemma 3.1 (Vanishing Lemma)**.** If a Markov chain has some $k,l \in S$ with $\lim\limits_{n\to\infty} p_{kl}^{(n)} = 0$, then for any $i,j \in S$ with $k \to i$ and $j \to l$, $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0$.

Proof. Since $k \to i, j \to l$, then we can find $r,s \in \mathbb{N}$ s.t. $p_{ki}^{(r)} > 0$ and $p_{jl}^{(s)} > 0$. By Chapman-Kolmogorov, $p_{kl}^{(r+n+s)} \geqslant p_{ki}^{(r)} p_{ij}^{(n)} p_{jl}^{(s)}$ and hence
$$p_{ij}^{(n)} \leqslant \frac{p_{kl}^{r+n+s}}{p_{ki}^{(r)} p_{jl}^{(s)}} \to 0$$
Since $p_{ij}^{(n)} \geqslant 0$, then we have
$$\lim_{n\to\infty} p_{ij}^{(n)} = 0.$$
□

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

(i) $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0, \forall i,j \in S$;

or (ii) $\lim\limits_{n\to\infty} p_{ij}^{(n)} \neq 0, \forall i,j \in S$.

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

Example 3.3 (Simple Random Walk)**.** Simple random walk is irreducible. It may be recurrent or transient, depending on p. But $\lim\limits_{n\to\infty} p_{00}^{(n)} = 0$ and thus by the vanishing together corollary, $p_{ij}^{(n)} \to 0, \forall i,j \in S$ and by vanishing probabilities corollary, simple random walk does not have a stationary distribution.

Note that for all $i,j \in S$, $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty$ even though $\lim\limits_{n\to\infty} p_{ij}^{(n)} = 0$.

Corollary 3.3 (Transient not Stationary Corollary)**.** An irreducible and transient Markov chain cannot have a stationary distribution.

Proof. If a chain is irreducible and transient, then by the transience equivalence theorem, $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty, \forall i, j \in S$. Hence, $\lim_{n \to \infty} p_{ij}^{(n)} = 0, \forall i, j \in S$. Then by the vanishing probabilities proposition, there is no stationary distribution. □

3.3 Obstacles to Convergence

Definition 3.3. The ***period*** of a state i is the greatest common divisor (gcd) of the set $\{n \geqslant 1 : p_{ii}^{(n)} > 0\}$, i.e., the largest number m s.t. all values of n with $p_{ii}^{(n)} > 0$ are integer multiples of m. If the period of every state is 1, the chain is ***aperiodic***; otherwise, the chain is ***periodic***.

Property 3.3. If state i has period t and $p_{ii}^{(m)} > 0$, then m is an integer multiple of t, i.e., t divides m.

Property 3.4. If $p_{ii} > 0$, then the period of state i is 1.

Property 3.5. If $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$ for some n, then the period of state i is 1.

Lemma 3.2 (Equal Periods Lemma)**.** If $i \leftrightarrow j$, then the periods of i and of j are equal.

Proof. Let the periods of i and j be t_i and t_j respectively. Since $i \leftrightarrow j, \exists r, s \in \mathbb{N}$ s.t. $p_{ij}^{(r)} > 0$ and $p_{ji}^{(s)} > 0$. Also, $p_{ii}^{(r+s)} \geqslant p_{ij}^{(r)} p_{ji}^{(s)} > 0$ and t_i divides $r + s$.

Suppose now that $p_{jj}^{(n)} > 0$, then $p_{ii}^{r+n+s} \geqslant p_{ij}^{(r)} p_{jj}^{(n)} p_{ji}^{(s)} > 0$ so t_i divides $r + n + s$. Thus, t_i divides n. Since it holds for any n with $p_{jj}^{(n)} > 0$, it follows that t_i is a common divisor of $\{n \in \mathbb{N} : p_{jj}^{(n)} > 0\}$. We know t_j is the gcd so that $t_j \geqslant t_i$.

Similarly, $t_i \geqslant t_j$ so $t_i = t_j$. □

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

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

3.4 Convergence Theorem

Theorem 3.1 (Stationary Recurrence Theorem)**.** If a Markov chain is irreducible and has a stationary distribution, then it is recurrent.

Proof. The transient not stationary corollary says that a chain cannot be irreducible and transient had have a stationary distribution. So if a chain is irreducible and has a stationary distribution, then it cannot be transient, i.e., it must be recurrent. □

Property 3.6. If a state i has $f_{ii} > 0$ and is aperiodic, then $\exists n_0(i) \in \mathbb{N}$ s.t. $p_{ii}^{(n)} > 0, \forall n \geqslant n_0(i)$.

Proof. Let $A = \{n \geqslant 1 : p_{ii}^{(n)} > 0\}$. If $m, n \in A$, then $p_{ii}^{(m)} > 0$ and $p_{ii}^{(n)} > 0$. Thus $p_{ii}^{(m+n)} \geqslant p_{ii}^{(m)} p_{ii}^{(n)} > 0$, so $m + n \in A$, which shows that A satisfies additivity. Since the state i is aperiodic, then $\gcd(a) = 1$. Hence, from the number theory lemma, $\exists n_0 \in \mathbb{N}$ s.t. $\forall n \geqslant n_0, n \in A$, i.e., $p_{ii}^{(n)} > 0$. □

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

Proof. Let $n_0(i)$ and $m \in \mathbb{N}$ s.t. $p_{ij}^{(m)} > 0$. Let $n_0(i,j) = n_0(i) + m$. If $n \geqslant n_0(i,j)$, then $n - m \geqslant n_0(i)$ and thus $p_{ij}^{(n)} \geqslant p_{ii}^{(n-m)} p_{ij}^{(m)} > 0$. □

Lemma 3.3 (Markov Forgetting Lemma). If a Markov chain is irreducible and aperiodic, and has stationary distribution $\{\pi_i\}$, then

$$\forall i, j, k \in S, \lim_{n\to\infty} |p_{ik}^{(n)} - p_{jk}^{(n)}| = 0.$$

Intuitively, after a long time n, the chain forgets whether it started from state i or from state j.

Proof. Define a new Markov chain $\{(X_n, Y_n)\}_{n=0}^{\infty}$ with state space $\overline{S} = S \times S$, and transition probabilities $\overline{p}_{(ij),(kl)} = p_{ik}p_{jl}$. Because of independence, new chain has a joint stationary distribution given by $\overline{\pi}_{(ij)} = \pi_i \pi_j$ for $i, j \in S$. Whenever $n \geqslant \max[n_0(i,k), n_0(j,l)], \overline{p}_{(ij),(kl)}^{(n)} = p_{ik}^{(n)} p_{jl}^{(n)} > 0$. Hence the new chain is irreducible and is aperiodic. By the stationary recurrence theorem, the new chain is recurrent.

Choose any $i_0 \in S$ and let $\tau = \inf\{n \geqslant 1 : X_n = Y_n = i_0\}$ be the first time that the new chain hits (i_0, i_0), i.e., that both chains equal i_0 at the same time. Since the new chain is irreducible and recurrent, the recurrence equivalences theorem says that $\overline{f}_{(ij),(i_0 i_0)} = 1$, i.e., starting from (i,j), the new chain must eventually hit (i_0, i_0), so $P_{(ij)}(\tau < \infty) = 1$. By the law of total probability,

$$p_{ij} = P_{(ij)}(X_n = k) = \sum_{m=1}^{\infty} P_{(ij)}(X_n = k, \tau = m) = \sum_{m=1}^{n} P_{(ij)}(X_n = k, \tau = m) + P_{(ij)}(X_n = k, \tau > n).$$

Similarly,

$$p_{jk}^{(n)} = \sum_{m=1}^{n} P_{ij}(Y_n = k, \tau = m) + P_{(ij)}(Y_n = k, \tau > n).$$

If $n \geqslant m$, then

$$\begin{aligned} P_{(ij)}(X_n = k, \tau = m) &= P_{(ij)}(\tau = m)P_{(ij)}(X_n = k | \tau = m) = P_{(ij)}(\tau = m)P_{(ij)}(X_n = k | X_m = Y_m = i_0) \\ &= P_{(ij)}(\tau = m)P(X_n = k | X_m = i_0) = P_{(ij)}(\tau = m)p_{i_0,k}^{(n-m)}. \end{aligned}$$

Similarly,

$$P_{(ij)}(Y_n = k, \tau = m) = P_{(ij)}(\tau = m)p_{i_0,k}^{(n-m)}.$$

So, $P_{(ij)}(X_n = k, \tau = m) = P_{(ij)}(Y_n = k, \tau = m)$. Hence, for all $i, j, k \in S$,

$$\begin{aligned} |p_{ik}^{(n)} - p_{jk}^{(n)}| &= \Bigg| \sum_{m=1}^{n} P_{(ij)}(X_n = k, \tau = m) + P_{(ij)}(X_n = k, \tau > n) \\ &\quad - \sum_{m=1}^{n} P_{(ij)}(Y_n = k, \tau = m) - P_{(ij)}(Y_n = k, \tau > n) \Bigg| \\ &= |P_{(ij)}(X_n = k, \tau > n) - P_{(ij)}(Y_n = k, \tau > n)| \\ &\leqslant |P_{(ij)}(X_n = k, \tau > n)| + |P_{(ij)}(Y_n = k, \tau > n)| \\ &\leqslant P_{(ij)}(\tau > n) + P_{(ij)}(\tau > n) = 2P_{(ij)}(\tau > n). \end{aligned}$$

Thus, $\lim_{n\to\infty} 2P_{(ij)}(\tau > n) = 2P_{(ij)}(\tau = \infty) = 0$ since $P_{(ij)}(\tau < \infty) = 1$. □

Theorem 3.2 (Markov Chain Convergence Theorem)**.** If a Markov chain is irreducible, aperiodic and has a stationary distribution $\{\pi_i\}$, then $\lim\limits_{n\to\infty} p_{ij}^{(n)} = \pi_j, \forall i, j \in S$ and $\lim\limits_{n\to\infty} P(X_n = j) = \pi_j$ for any initial distribution $\{v_i\}$.

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

Proof. By Markov chain convergence theorem, any stationary distribution must be equal to $\lim\limits_{n\to\infty} P(X_n = j)$, so they are all equal. □

Example 3.4. If $\{X_n\}$ is simple symmetric random walk, then the absolute values $|X_n|$ converge weakly to positive infinity.

Proof. We write $X_n = \sum\limits_{i=1}^{n} Z_i$ where $\{Z_i\}$ are i.i.d. ± 1 with probability $\frac{1}{2}$ each, and hence mean $m = 0$ and variance $v = 1$. Hence, by Central Limit Theorem, for any $b > 0$,

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

where ϕ is the standard normal distribution's density function. Hence, $\lim\limits_{b\downarrow 0}\lim\limits_{n\to\infty} P\left(-b < \frac{X_n}{\sqrt{n}} < b\right) = 0$.

Since $P(|X_n| < K) = P(-K < X_n < K) = P\left(-\frac{K}{\sqrt{n}} < \frac{X_n}{\sqrt{n}} < \frac{K}{\sqrt{n}}\right)$, then by monotonicity, for any fixed finite K and any $b > 0$,

$$P(|X_n| < K) \leqslant P\left(-b < \frac{X_n}{\sqrt{n}} < b\right), \forall n \geqslant \frac{K^2}{b^2},$$

i.e., all sufficiently large n. It follows that $\lim\limits_{n\to\infty} P(|X_n| < K) = 0$ for any fixed finite K and thus $|X_n|$ converges weakly to positive infinity. □

3.5 Periodic Convergence

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

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

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

and

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

Lemma 3.4 (Cyclic Decomposition Lemma)**.** There is a disjoint partition $S = S_0 \dot\cup S_1 \dot\cup \cdots \dot\cup S_{b-1}$ s.t. w.p. 1, the chain always moves from S_0 to S_1, and then to S_2, and so on to S_{b-1}, and then back to S_0.

Example 3.5 (Ehrenfest's Urn)**.** Ehrenfest's Urn has period $b = 2$ and oscillates between the two subsets $S_0 = \{\text{Even } i \in S\}$ and $S_1 = \{\text{Odd } i \in S\}$. It satisfies the periodic convergence theorem that

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

Corollary 3.8 (Average Probability Convergence)**.** If a Markov chain is irreducible with stationary distribution $\{\pi_i\}$ (whether periodic or not), then for all $i, j \in S$,

$$\lim_{n\to\infty} \frac{1}{n}[p_{ij}^{(1)} + p_{ij}^{(2)} + \cdots + p_{ij}^{(n)}] = \lim_{n\to\infty} \frac{1}{n}\sum_{l=1}^{n} p_{ij}^{(l)} = \pi_j.$$

Proof. This follows from either Markov chain convergence theorem (for aperiodic chains) or the periodic convergence theorem (for chains with period $b \geqslant 2$), by the Cesaro sum principle: if a sequence converges, then its partial averages also converge to the same value. □

Corollary 3.9 (Unique Stationary Corollary)**.** An irreducible chain, whether periodic or not, has at most one stationary distribution.

Proof. By average probability convergence, any stationary distribution which exists must be equal to $\lim\limits_{n\to\infty} \frac{1}{n}[p_{ij}^{(1)} + p_{ij}^{(2)} + \cdots + p_{ij}^{(n)}]$, so they are all equal. □

3.6 Application: Markov Chain Monte Carlo Algorithm (Discrete)

Let $S = \mathbb{Z}$ or more generally let S be any contiguous subset of $\mathbb{Z}$ (e.g., $S = \{1, 2, 3\}, \{-5, -4, \cdots, 18\}, \mathbb{N}$). Let $\{\pi_i\}$ be any probability distribution on S. Assume for simplicity that $\pi_i > 0, \forall i \in S$. Suppose we want to sample from π, i.e., create random variables X with $P(X = i) \approx \pi_i, \forall i \in S$.

One popular method is ***Markov chain Monte Carlo*** (***MCMC***): Create a Markov chain $X_0, X_1, \cdots$ s.t. $\lim\limits_{n\to\infty} P(X_n = i) = \pi_i, \forall i \in S$, so for large $n, P(X_n = i) \approx \pi_i, \forall i \in S$.

We can create Markov chain transitions $\{p_{ij}\}$ s.t. $\lim\limits_{n\to\infty} p_{ij}^{(n)} = \pi_j$ by ***Metropolis algorithm***: Let $p_{i,i+1} = \frac{1}{2}\min\left[1, \frac{\pi_{i+1}}{\pi_i}\right], p_{i,i-1} = \frac{1}{2}\min\left[1, \frac{\pi_{i-1}}{\pi_i}\right]$, and $p_{i,i} = 1 - p_{i,i+1} - p_{i,i-1}$ with $p_{ij} = 0$ otherwise (where $\pi_j = 0$ if $j \notin S$). An equivalent algorithmic version is: Given X_{n-1}, let Y_n equal $X_{n-1} \pm 1$ (w.p. $\frac{1}{2}$ each), and let $U_n \sim \text{Uniform}[0, 1]$ with the $\{U_n\}$ chosen i.i.d., and then let

$$X_n = \begin{cases} Y_n, & U_n \leqslant \frac{\pi Y_n}{\pi X_{n-1}} \quad \text{("accept")} \\ X_{n-1}, & \text{otherwise} \quad \text{("reject")} \end{cases}.$$

Theorem 3.4 (MCMC Convergence Theorem)**.** Metropolis algorithm Markov chain has the property that $\lim\limits_{n\to\infty} p_{ij}^{(n)} = \pi_j$, and $\lim\limits_{n\to\infty} P(X_n = j) = \pi_j, \forall i, j \in S$ and all initial distribution v.

Proof. We have

$$\pi_i p_{i,i+1} = \pi_i \frac{1}{2}\min\left[1, \frac{\pi_{i+1}}{\pi_i}\right] = \frac{1}{2}\min\left[\pi_i, \pi_{i+1}\right]$$

and

$$\pi_{i+1} p_{i+1,i} = \pi_{i+1} \frac{1}{2}\min\left[1, \frac{\pi_i}{\pi_{i+1}}\right] = \frac{1}{2}\min[\pi_{i+1}, \pi_i].$$

Thus, $\pi_i p_{ij} = \pi_j p_{ji}$ if $j = i + 1$ and similarly if $j = i - 1$. Also, $\pi_i p_{ij} = \pi_j p_{ji}$ if $|j - i| \geqslant 2$ since both sides are 0. Hence, $\pi_i p_{ij} = \pi_j p_{ji}, \forall i, j \in S$ and thus the chain is reversible w.r.t. $\{\pi_i\}$ and $\{\pi_i\}$ is stationary. Since the chain is irreducible and aperiodic then the results follows from Markov chain convergence theorem. □

By MCMC convergence theorem, we know for large enough n, X_n is approximately a sample from π. Such Markov chains are very widely used to sample from complicated distributions $\{\pi_i\}$ to estimate probabilities and expectations.

3.7 Application: Random Walk on Graph

Let V be a non-empty finite or countable set. Let $w : V \times V \to [0, \infty)$ be a ***weight function*** which is symmetric, i.e., $w(u, v) = w(v, u), \forall u, v \in V$. The usual ***unweighted graph*** case is $w(u, v) = 1$ if there is an ***edge*** between u and v, otherwise $w(u, v) = 0$. We can also have other weights, multiple edges, self-edges, etc. Let $d(u) = \sum\limits_{v \in V} w(u, v)$ be the ***degree*** of the vertex u. Assume $d(u) > 0, \forall u \in V$ (for example, by giving any isolated point a self-edge).

Definition 3.4. Given a vertex set V and a symmetric weight function w, the ***simple random walk on the undirected graph*** (V, w) is the Markov chain with state space $S = V$ and transition probabilities $p_{uv} = \frac{w(u,v)}{d(u)}, \forall u, v \in V$.

Note if follows that $\sum\limits_{v \in V} p_{uv} = \frac{\sum_{v \in V} w(u,v)}{\sum_{v \in V} w(u,v)} = 1$. In the usual case where each $w(u, v) = 0$ or 1, from u the chain moves to one of the $d(u)$ vertices connected to u, each w.p. $\frac{1}{d(u)}$.

Example 3.6 (Ring Graph)**.** Suppose $V = \{1, 2, 3, 4, 5\}$ with $w(i, i + 1) = w(i + 1, i) = 1$ for $i = 1, 2, 3, 4$, $w(5, 1) = w(1, 5) = 1$, and $w(i, j) = 0$ otherwise.

Example 3.7 (Stick Graph)**.** Suppose $V = \{1, 2, \cdots, K\}$ with $w(i, i + 1) = w(i + 1, i) = 1$ for $1 \leqslant i \leqslant K - 1$, and $w(i, j) = 0$ otherwise.

Example 3.8 (Star Graph)**.** Suppose $V = \{0, 1, 2, \cdots, K\}$ with $w(i, 0) = w(0, i) = 1$ for $i = 1, 2, \cdots, K$, and $w(i, j) = 0$ otherwise.

Example 3.9 (Infinite Graph)**.** Suppose $V = \mathbb{Z}$ with $w(i, i + 1) = w(i + 1, i) = 1, \forall i \in V$, and $w(i, j) = 0$ otherwise. Random walk on this graph corresponds exactly to simple symmetric random walk.

Example 3.10 (Frog Graph)**.** Suppose $V = \{1, 2, \cdots, K\}$ with $w(i, i) = 1$ for $1 \leqslant i \leqslant K$, $w(i, i + 1) = w(i + 1, i) = 1$ for $1 \leqslant i \leqslant K - 1$, $w(K, 1) = w(1, K) = 1$, and $w(i, j) = 0$ otherwise.

Property 3.7 (Graph Stationary Distribution)**.** Consider a random walk on an undirected graph V with degrees $d(u)$. Assume Z is finite, then if $\pi_u = \frac{d(u)}{Z}, \forall u \in V$, then π is a stationary distribution.

Proof. It is easily to check that $\pi_u \geqslant 0$ and $\sum\limits_u \pi_u = 1$, so π is a probability distribution on V. We have

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

Hence, the chain is reversible w.r.t. π and thus π is a stationary distribution. $\square$

From the property above, a precise formula for the stationary distribution of almost any simple random walk on any undirected graph is given.

Theorem 3.5 (Graph Convergence Theorem)**.** For a random walk on a connected non-bipartite graph, if $Z < \infty$, then $\lim\limits_{n \to \infty} p_{uv}^{(n)} = \frac{d(v)}{Z}, \forall u, v \in V$, and $\lim\limits_{n \to \infty} P(X_n = v) = \frac{d(v)}{Z}$ for any initial probabilities.

Proof. If the graph is ***connected***, i.e., it is possible to move from any vertex to any other vertex through edges of positive weight, then the chain is irreducible.

The walk can always return to any vertex u in 2 steps by moving to any other vertex and then back along the same edge, and thus 1 and 2 are the only possible periods. If the graph is a ***bipartite graph***, i.e., it can be divided into two subsets s.t. all the links go from one subset to the other one, then the chain has period 2. If the chain is not bipartite, then it is aperiodic. Furthermore, if there is any cycle of odd length, then the graph cannot be bipartite, so it must be aperiodic.

Since the graph is connected and non-bipartite, then it is irreducible and aperiodic. If $Z < \infty$ then $\pi_v = \frac{d(v)}{Z}$ is a stationary distribution and thus by the Markov chain convergence theorem, for all $u, v \in V$, $\lim_{n\to\infty} p_{uv}^{(n)} = \pi_v = \frac{d(v)}{Z}$. □

Property 3.8 (Graph Average Convergence)**.** A random walk on a connected graph with $Z < \infty$ (whether bipartite or not) has $\lim_{n\to\infty} \frac{1}{2}[p_{uv}^{(n)} + p_{uv}^{(n+1)}] = \frac{d(v)}{Z}$, and $\lim_{n\to\infty} \frac{1}{n}\sum_{l=1}^{n} p_{uv}^{(l)} = \frac{d(v)}{Z}, \forall u, v \in V$.

Example 3.11 (Stick Graph)**.** Suppose $V = \{1, \cdots, K\}$ and $w(i, i+1) = w(i+1, i) = 1$ for $1 \leqslant i \leqslant K-1$, and $w(i, j) = 0$ otherwise. The graph is connected but is bipartite since every edge goes between an even vertex and an odd vertex. Also $d(i) = 1$ for $i = 1$ or $K, d(i) = 2$ otherwise. Hence $Z = 1 + 2 + \cdots + 2 + 1 = 1 + 2(K-2) + 1 = 2K - 2$. Then if $\pi_i = \frac{d(i)}{Z} = \frac{1}{2K-2}$ for $i = 1$ or K, and $\pi_i = \frac{2}{2K-2}$ for $2 \leqslant i \leqslant K-1$, then π is a stationary distribution and we must have $\lim_{n\to\infty} \frac{1}{2}[p_{ij}^{(n)} + p_{ij}^{(n+1)}] = \pi_j, \forall j \in V$.

3.8 Mean Recurrence Time

Definition 3.5. The ***mean recurrence time*** of a state i is $m_i = \mathbb{E}_i[T_i]$ where $T_i = \inf\{n \geqslant 1 : X_n = i\}$, i.e., m_i is the expected time to return from i back to i.

Note that if the chain never returns to i, then $T_i = \infty$. If $P(T_i = \infty) > 0$ (which holds iff i is transient), then $m_i = \infty$, i.e., if i is transient, then $m_i = \infty$. Hence, if $m_i < \infty$, then i must be recurrent. However, even if i is recurrent, i.e., $P_i(T_i < \infty) = 1$, it is still possible that $\mathbb{E}_i[T_i] = \infty$.

Definition 3.6. A state is ***positive recurrent*** if $m_i < \infty$. A state is ***null recurrent*** if it is recurrent but $m_i = \infty$.

Theorem 3.6 (Recurrence Time Theorem)**.** For an irreducible Markov chain, either
(α) $m_i < \infty, \forall i \in S$ and there is a unique stationary distribution given by $\pi_i = \frac{1}{m_i}$;
or (β) $m_i = \infty, \forall i \in S$ and there is no stationary distribution.

Proof. Suppose a chain starts at i and $m_i < \infty$. On average, the chain returns to i once every m_i steps, i.e., for large r, by the law of large numbers, it takes about rm_i steps for the chain to return a total of r times. So in a large number $n \approx rm_i$ of steps, it will return to i about r times, i.e., about $\frac{n}{m_i}$ times or about $\frac{1}{m_i}$ of the time. Hence, the limiting fraction of time it spends at i is $\frac{1}{m_i}$:

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

Then by the bounded convergence theorem,

$$\lim_{n\to\infty} \mathbb{E}_i\left[\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i}\right] = \frac{1}{m_i}.$$

Since the chain is irreducible, with stationary distribution π, then by finite linearity and average probability convergence,

$$\lim_{n\to\infty} \mathbb{E}_i\left[\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_k=i}\right] = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}_i[\mathbf{1}_{X_k=i}] = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} P_i(X_k = i) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_{ii}^{(k)} = \pi_i.$$

Hence, we must have $\frac{1}{m_i} = \pi_i$, i.e., $m_i = \frac{1}{\pi_i}$. It also follows that $\sum\limits_{i \in S} \frac{1}{m_i} = 1$.

If $m_i = \infty$, then $\pi_i = \frac{1}{m_i} = 0$. Hence, if $m_j = \infty, \forall j$, then we must have $\pi_j = 0, \forall j$, which contradicts that $\sum\limits_{j \in S} \pi_j = 1$. In this case, there is no stationary distribution.

Furthermore, if a chain is irreducible and $m_j < \infty$ for some $j \in S$, then $m_i < \infty, \forall i \in S$, and $\left\{ \frac{1}{m_i} \right\}$ is stationary. □

Property 3.9. An irreducible Markov chain on a finite state space S always falls into case (α), i.e., has $m_i < \infty, \forall i \in S$, and a unique stationary distribution given by $\pi_i = \frac{1}{m_i}$.

Note that the converse is false.

Example 3.12. Simple random walk, for any $0 < p < 1$, has infinite mean recurrence times, i.e., has $m_i = \infty, \forall i \in S$.

Proof. We know that sample random walk is irreducible. However, s.r.w. has no stationary distribution and thus in cannot be in case (α). Hence, it must be in case (β), i.e., $m_i = \infty, \forall i$. □

On average, simple random walk (including simple symmetric random walk) takes an infinite amount of time to return to where it started. Although simple symmetric random walk is recurrent and usually returns quickly, the small chance of taking a very long time to return is sufficient to make the expected return time infinite even though the actual return time is always finite.

In summary, simple symmetric random walk is: recurrent and hence in case (a) $f_{ij} = 1$ and $\sum\limits_{n=1}^{\infty} p_{ij}^{(n)} = \infty, \forall i, j \in S$; but in case (i) $\lim\limits_{n \to \infty} p_{ij}^{(n)} = 0, \forall i, j \in S$ and no stationary distribution; and in case $(\beta) m_i = \infty, \forall i \in S$.

Example 3.13. For simple random walk, we consider $\mathbb{E}_i[T_{i+1}]$, i.e., the expected time starting from i to first hit the state $i+1$. At time 0, simple random walk moves from i to either $i+1$ w.p. p, or $i-1$ w.p. $1-p$, which takes one step. Hence,

$$m_i := \mathbb{E}_i[T_i] = 1 + p\mathbb{E}_{i+1}[T_i] + (1-p)\mathbb{E}_{i-1}[T_i].$$

By shift-invariance,

$$m_i = 1 + p\mathbb{E}_i[T_{i-1}] + (1-p)\mathbb{E}_i[T_{i+1}].$$

We know $m_i = \infty$ and thus

$$\infty = 1 + p\mathbb{E}_i[T_{i-1}] + (1-p)\mathbb{E}_i[T_{i+1}].$$

Hence, at least one of $\mathbb{E}_i[T_{i-1}]$ and $\mathbb{E}_i[T_{i+1}]$ is infinite. If $p = \frac{1}{2}$, then by symmetry $\mathbb{E}_i[T_{i+1}] = \mathbb{E}_i[T_{i-1}]$ and thus $\mathbb{E}_i[T_{i+1}] = \mathbb{E}_i[T_{i-1}] = \infty$, i.e., on average it takes an infinite amount of time for simple symmetric random walk to reach the state just to the right or left. For simple random walk with $p > \frac{1}{2}, \mathbb{E}_i[T_{i+1}] < \infty$, and in that case $f_{i,i-1} < 1$, so $P_i(T_{i-1} = \infty) > 0 \Rightarrow \mathbb{E}_i[T_{i-1}] = \infty$.

3.9 Application: Sequence Waiting Time

Example 3.14. Suppose we repeatedly flip a fair coin, and get H or T independently each time w.p. $\frac{1}{2}$ each. Let τ be the first time the sequence HTH is completed. Find $\mathbb{E}[\tau]$.

Solution. Let X_n be the amount of the desired sequence that the chain has achieved at the nth flip. For example, if the flips begin with HHTTHTH, then $X_1 = 1, X_2 = 1, X_3 = 2, X_4 = 0, X_5 = 1, X_6 = 2$ and $X_7 = 3$. Hence $\tau = 7$. We always have $X_\tau = 3$ since we win upon reaching state 3. Assume we start over right after we win, i.e., $X_{\tau+1} = 1$ if flip $\tau + 1$ is H, otherwise $X_{\tau+1} = 0$. We take $X_0 = 0$, i.e., at the beginning we have not achieved anything, which is equivalent to $X_0 = 3$ since after 3 we start over anyway. Hence, the mean waiting time of HTH is equal to the mean recurrence time of the state 3.

Let $\{X_n\}$ be a Markov chain with state space $S = \{0, 1, 2, 3\}$. The transitions are

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

With $\pi P = \pi$, we compute the stationary distribution is $(0.3, 0.4, 0.2, 0.1)$. By the recurrence time theorem, the mean time to return from state 3 to state 3 is $\frac{1}{\pi_3} = \frac{1}{0.1} = 10$. Returning from state 3 to state 3 has the same probability as going from state 0 to state 3, i.e., as achieving $X_n = 3$ from $X_0 = 0$. Hence the mean time to go from state 0 to state 3 is also 10, i.e., the mean waiting time for HTH is 10.

Example 3.15. Suppose we repeatedly flip a fair coin, and get H or T independently each time w.p. $\frac{1}{2}$ each. Let τ be the first time the sequence THH is completed. Find $\mathbb{E}[\tau]$.

Solution. The transitions are

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

The stationary distribution is $(0.125, 0.5, 0.25, 0.125)$ and the expected waiting time is $\frac{1}{\pi_3} = 8$.

4 Martingale

4.1 Martingale Definition

Let $\{X_n\}_{n=0}^{\infty}$ be a sequence of r.v.s., and assume that r.v.s. X_n have finite expectation (or are integrable), i.e., $\mathbb{E}[|X_n|] < \infty, \forall n$.

Definition 4.1. A sequence $\{X_n\}_{n=0}^{\infty}$ is a ***martingale*** if

$$\mathbb{E}[X_{n+1}|X_0, \cdots, X_n] = X_n, n = 0, 1, \cdots,$$

i.e., no matter what has happened so far, the conditional average of the next value will be equal to the most recent value.

For discrete r.v.s., the definition means that

$$\mathbb{E}[X_{n+1}|X_0 = i_0, \cdots, X_n = i_n] = i_n$$

for all $i_0, i_1, \cdots, i_n$, in terms of discrete conditional expectation.

If the sequence $\{X_n\}$ is a Markov chain and is a martingale, then

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

i.e.,

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

Example 4.1 (Simple Symmetric Random Walk)**.** Let $\{X_n\}$ be simple symmetric random walk with state space $S = \mathbb{Z}$. We have $|X_n| \leqslant n$ and so $\mathbb{E}[|X_n|] \leqslant n < \infty$. For all $i \in S$,

$$\sum_{j \in S} jp_{ij} = (i+1)\frac{1}{2} + (i-1)\frac{1}{2} = i$$

and thus simple symmetric random walk is a martingale.

Property 4.1. If $\{X_n\}$ is a martingale, then $\mathbb{E}[X_n] = \mathbb{E}[X_0], \forall n$.

Proof. If $\{X_n\}$ is a martingale, then by double-expectation formula,

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[\mathbb{E}[X_{n+1}|X_0, \cdots, X_n]] = \mathbb{E}[X_n].$$

□

Note that it is not a sufficient condition for $\{X_n\}$ to be a martingale.

4.2 Stopping Time and Optional Stopping

Definition 4.2. A non-negative-integer-valued r.v. T is a ***stopping time*** for $\{X_n\}$ if the event $\{T = n\}$ is determined by $X_0, \cdots, X_n$, i.e., if the indicator function $\mathbf{1}_{T=n}$ is a function os $X_0, \cdots, X_n$.

Intuitively, a stopping time T must decide whether to stop at time n based solely on what has happened up to time n, i.e., without first looking into the future.

Example 4.2. (1) $T = 5$ is a valid stopping time.
(2) $T = \inf\{n \geqslant 0 : X_n = 5\}$ is a valid stopping time. Note that $T = \infty$ if $\{X_n\}$ never hits 5.
(3) $T = \inf\{n \geqslant 0 : X_n = 0 \text{ or } X_n = c\}$ is a valid stopping time.
(4) $T = \inf\{n \geqslant 2 : X_{n-2} = 5\}$ is a valid stopping time.
(5) $T = \inf\{n \geqslant 2 : X_{n-1} = 5, X_n = 6\}$ is a valid stopping time.
(6) $T = \inf\{n \geqslant 0 : X_{n+1} = 5\}$ is not a valid stopping time, since it looks into the future to stop one step before reaching 5.

Lemma 4.1 (Optional Stopping Lemma)**.** If $\{X_n\}$ is a martingale and T is a stopping time which is bounded, i,e., $\exists M < \infty$ s.t. $P(T \leqslant M) = 1$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Proof. We have

$$\begin{aligned}\mathbb{E}[X_T] - \mathbb{E}[X_0] = \mathbb{E}[X_T - X_0] &= \mathbb{E}\left[\sum_{k=1}^{T}(X_k - X_{k-1})\right] = \mathbb{E}\left[\sum_{k=1}^{M}(X_k - X_{k-1})\mathbf{1}_{k\leqslant T}\right] \\ &= \sum_{k=1}^{M}\mathbb{E}[(X_k - X_{k-1})\mathbf{1}_{k\leqslant T}] = \sum_{k=1}^{M}\mathbb{E}[(X_k - X_{k-1})(1 - \mathbf{1}_{T\leqslant k-1})].\end{aligned}$$

Using double-expectation formula and the fact that $1-\mathbf{1}_{T\leqslant k-1}$ is completely determined by $X_0, \cdots, X_{k-1}$, we have

$$\begin{aligned}\mathbb{E}[X_T] - \mathbb{E}[X_0] &= \sum_{k=1}^{M}\mathbb{E}[\mathbb{E}[(X_k - X_{k-1})(1 - \mathbf{1}_{T\leqslant k-1})|X_0, \cdots, X_{k-1}]] \\ &= \sum_{k=1}^{M}\mathbb{E}[(1 - \mathbf{1}_{T\leqslant k-1})\mathbb{E}[(X_k - X_{k-1}|X_0, \cdots, X_{k-1})]] \\ &= \sum_{k=1}^{M}\mathbb{E}[(1 - \mathbf{1}_{T\leqslant k-1})\mathbb{E}[X_k|X_0, \cdots, X_{k-1}]] - \sum_{k=1}^{M}\mathbb{E}[(1 - \mathbf{1}_{T\leqslant k-1})\mathbb{E}[X_{k-1}|X_0, \cdots, X_{k-1}]].\end{aligned}$$

Since $\{X_n\}$ is a martingale, then $\mathbb{E}[X_k|X_0, \cdots, X_{k-1}] = X_{k-1}$ and $\mathbb{E}[X_{k-1}|X_0, \cdots, X_{k-1}] = X_{k-1}$ and thus $\mathbb{E}[X_T] - \mathbb{E}[X_0] = 0$, i.e., $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. $\square$

Note that if $M = \infty$, the proof fails since we cannot change sum and expectation operator.

Example 4.3. Consider simple symmetric random walk with $X_0 = 0$ and let $T = \inf\{n \geqslant 0 : X_n = -5\}$, which is a stopping time. $P(T < \infty) = 1$ since the Markov chain is recurrent. However, in this case, we always have $X_T = -5$ and thus $\mathbb{E}[X_T] = -5 \neq 0 = \mathbb{E}[X_0]$.

Let $T = \min(10^12, \inf\{n \geqslant 0 : X_n = -5\})$, then T is a stopping time, and $T \leqslant 10^{12}$ so T is bounded. By optional stopping lemma, $\mathbb{E}[X_T] = \mathbb{E}[X_0] = 0$. BUT nearly always (i.e., whenever the process hits -5 some time within the first 10^{12} steps), we will have $X_T = -5$.

By the law of total expectation

$$0 = \mathbb{E}[X_T] = P(X_T = -5)\mathbb{E}[X_T|X_T = -5] + P(X_T \neq -5)\mathbb{E}[X_T|X_T \neq -5].$$

We have $\mathbb{E}[X_T|X_T = -5] = -5$ and $q := P(X_T = -5) \approx 1$ so $P(X_T \neq 5) = 1 - q \approx 0$. Hence,

$$0 = q(-5) + (1-q)\mathbb{E}[X_T|X_T \neq -5]$$

where $q \approx 1$. We must have $\mathbb{E}[X_T|X_T \neq -5] = \frac{5q}{1-q}$ is huge, which is plausible since if the process did not hit -5 within the first 10^{12} steps, then probably it instead got very large.

Theorem 4.1 (Optional Stopping Theorem)**.** If $\{X_n\}$ is a martingale with stopping time T, $P(T < \infty) = 1$, $\mathbb{E}[|X_T|] < \infty$, and $\lim\limits_{n\to\infty} \mathbb{E}[X_n \mathbf{1}_{T>n}] = 0$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Proof. For each $m \in \mathbb{N}$, let $S_m = \min(T, m)$, then S_m is a stopping time, and it is bounded. By the optional stopping lemma, $\mathbb{E}[X_{S_m}] = \mathbb{E}[X_0]$. We also have

$$X_{S_m} = X_{\min(T,m)} = X_T \mathbf{1}_{T \leqslant m} + X_m \mathbf{1}_{T>m} = X_T(1 - \mathbf{1}_{T>m}) + X_m \mathbf{1}_{T>m} = X_T - X_T \mathbf{1}_{T>m} + X_m \mathbf{1}_{T>m}.$$

Hence, $X_T = X_{S_m} + X_T \mathbf{1}_{T>m} - X_m \mathbf{1}_{T>m}$ and

$$\mathbb{E}[X_T] = \mathbb{E}[X_{S_m}] + \mathbb{E}[X_T \mathbf{1}_{T>m}] - \mathbb{E}[X_m \mathbf{1}_{T>m}] = \mathbb{E}[X_0] + \mathbb{E}[X_T \mathbf{1}_{T>m}] - \mathbb{E}[X_m \mathbf{1}_{T>m}]$$

for any m. As $m \to \infty$, $\lim\limits_{m\to\infty} \mathbb{E}[X_m \mathbf{1}_{T>m}] = 0$. Also, $\lim\limits_{m\to\infty} \mathbb{E}[X_T \mathbf{1}_{T>m}]$ by dominated convergence theorem, since $\mathbb{E}[|X_T|] < \infty$ and $\mathbf{1}_{T>m} \to 0$. Therefore, $\mathbb{E}[X_T] \to \mathbb{E}[X_0]$, i.e., $\mathbb{E}[X_T] = \mathbb{E}[X_0]$. □

Corollary 4.1 (Optional Stopping Corollary)**.** If $\{X_n\}$ is a martingale with stopping time T, which is bounded up to time T, i.e., $\exists M < \infty$ s.t. $P(|X_n|\mathbf{1}_{n \leqslant T} \leqslant M) = 1, \forall n$, and $P(T < \infty) = 1$, then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

Proof. First, we have $P(|X_T| \leqslant M) = 1$ since

$$\begin{aligned} P(|X_T| > M) &= \sum_n P(T = n, |X_T| > M) = \sum_n P(T = n, |X_n|\mathbf{1}_{n \leqslant T} > M) \\ &\leqslant \sum_n P(|X_n|\mathbf{1}_{n \leqslant T} > M) = \sum_n 0 = 0. \end{aligned}$$

Hence, $\mathbb{E}[|X_T|] \leqslant M < \infty$. Also,

$$\begin{aligned} |\mathbb{E}[X_n \mathbf{1}_{T>n}]| &\leqslant \mathbb{E}[|X_n|\mathbf{1}_{T>n}] = \mathbb{E}[|X_n|\mathbf{1}_{n \leqslant T}\mathbf{1}_{T>n}] \\ &\leqslant \mathbb{E}[M\mathbf{1}_{T>n}] = MP(T > n), \end{aligned}$$

which converges to 0 as $n \to \infty$ since $P(T < \infty) = 1$. The result then follows from optional stopping theorem. □

Example 4.4 (Gambler's Ruin with $p = \frac{1}{2}$)**.** Let $T = \inf\{n \geqslant 0 : X_n = 0 \text{ or } X_n = c\}$ be the time the game ends. we have $P(T < \infty) = 1$. Also, if the game has not yet ended, i.e., $n \leqslant T$, then X_n must be between 0 and c. Hence, $|X_n|\mathbf{1}_{n \leqslant T} \leqslant c < \infty, \forall n$. By the optional stopping corollary, $\mathbb{E}[X_T] = \mathbb{E}[X_0] = a$. Thus, $a = cs(a) + 0(1 - s(a))$ and we must have $s(a) = \frac{a}{c}$ ($s(a) = P_a(T_c < T_0)$).

Example 4.5 (Gambler's Ruin with $p \neq \frac{1}{2}$)**.** If $p \neq \frac{1}{2}$, then $\{X_n\}$ is not a martingale since $\sum\limits_j jp_{ij} = p(i+1) + (1-p)(i-1) = i + 2p - 1 \neq i$. Let

$$Y_n = \left(\frac{1-p}{p}\right)^{X_n},$$

then $\{Y_n\}$is a Markov chain with

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

Hence, $\{Y_n\}$ is a martingale. We have $P(T < \infty) = 1$ and $|Y_n|\mathbf{1}_{n\leqslant T} \leqslant \max\left[\left(\frac{1-p}{p}\right)^0, \left(\frac{1-p}{p}\right)^c\right] < \infty, \forall n$. By the optional stopping corollary, $\mathbb{E}[Y_T] = \mathbb{E}[Y_0] = \left(\frac{1-p}{p}\right)^a$. We also have $Y_T = \left(\frac{1-p}{p}\right)^c$ if we win or $\left(\frac{1-p}{p}\right)^0 = 1$ if we lose. Therefore,

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

and thus

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

4.3 Wald's Theorem

Theorem 4.2 (Wald's Theorem)**.** Suppose $X_n = a + Z_1 + \cdots + Z_n$ where $\{Z_i\}$ are i.i.d. with finite mean m. Let T be a stopping time for $\{X_n\}$ which has finite mean, i.e., $\mathbb{E}[T] < \infty$, then $\mathbb{E}[X_T] = a + m\mathbb{E}[T]$.

Proof. We compute that

$$\begin{aligned}\mathbb{E}[X_T] - a = \mathbb{E}[X_T - a] = \mathbb{E}[Z_1 + \cdots + Z_T] &= \mathbb{E}\left[\sum_{i=1}^{T} Z_i\right] \\ &= \mathbb{E}\left[\sum_{i=1}^{\infty} Z_i \mathbf{1}_{T\geqslant i}\right] = \mathbb{E}\left[\lim_{N\to\infty}\sum_{i=1}^{N} Z_i \mathbf{1}_{T\geqslant i}\right].\end{aligned}$$

We can interchange expectation and limit by dominated convergence theorem, then

$$\mathbb{E}[X_T] - a = \lim_{N\to\infty}\mathbb{E}\left[\sum_{i=1}^{N} Z_i\mathbf{1}_{T\geqslant i}\right] = \lim_{N\to\infty}\sum_{i=1}^{N}\mathbb{E}[Z_i\mathbf{1}_{T\geqslant i}] = \sum_{i=1}^{\infty}\mathbb{E}[Z_i\mathbf{1}_{T\geqslant i}].$$

We have $\{T \geqslant i\} = \{T \leqslant i-1\}^C$, since T is a stopping time, determined by $Z_1, \cdots, Z_{i-1}$ and Z_i is independent of $\{T \geqslant i\}$. Hence,

$$\mathbb{E}[X_T] - a = \sum_{i=1}^{\infty}\mathbb{E}[Z_i]\mathbb{E}[\mathbf{1}_{T\geqslant i}] = \sum_{i=1}^{\infty} mP(T \geqslant i) = m\sum_{i=1}^{\infty} P(T \geqslant i) = m\mathbb{E}[T].$$

□

A special case is when $m = 0$, them $\{X_n\}$ is a martingale and by Wald's Theorem, $\mathbb{E}[X_T] = a = \mathbb{E}[X_0]$ but under different assumptions (an i.i.d. sum with $\mathbb{E}[T] < \infty$).

Example 4.6 (Simple Symmetric Random Walk)**.** Consider $\{X_n\}$ with $X_0 = 0$ and $T = \inf\{n \geqslant 0 : X_n = -5\}$ is the first time we hit -5. In Wald's Theorem, $a = 0$ and $m = 0$. Also, $P(T < \infty) = 1$ since the chain is recurrent. But $X_T = -5$ so $\mathbb{E}[X_T] = -5 \neq 0 = \mathbb{E}[X_0]$, which does not contradict Wald's Theorem since $\mathbb{E}[T] = \infty$.

Example 4.7. Suppose we repeatedly roll a fair six-sided die. Let Z_n be the result of the nth roll, $R = \inf\{n \geqslant 1 : Z_n = 5\}$ be the first time we roll 5, and A be the sum of all the numbers rolled up to time R, i.e., $A = \sum_{n=1}^{R} Z_n$. Let $S = \inf\{n \geqslant 1 : Z_n = 3\}$ and $B = \sum_{n=1}^{S} Z_n$. Let $A' = \sum_{n=1}^{R-1} Z_n$ and $B' = \sum_{n=1}^{S-1} Z_n$ be the sums not counting the final roll.

In this case, $\{Z_i\}$ are i.i.d. with mean $m = 3.5$. Clearly, R and S are stopping times. Also R ans S have Geometric $\left(\frac{1}{6}\right)$ distributions and so $\mathbb{E}[R] = \mathbb{E}[S] = 6$. In the notation of Wald's Theorem, $a = 0, A = X_R, B = X_S$. So

$$\mathbb{E}[A] = \mathbb{E}[X_R] = a + m\mathbb{E}[R] = 3.5 \times 6 = 21, \mathbb{E}[B] = \mathbb{E}[X_S] = a + m\mathbb{E}[S] = 21.$$

Besides, $A' = X_{R-1}$ and $B' = X_{S-1}$ but $R-1$ and $S-1$ are not stopping times, so we cannot apply Wald's Theorem. On the other hand, $A' = A - 5$ and $B' = B - 3$ and thus $\mathbb{E}[A'] = 21 - 5 = 16$ and $\mathbb{E}[B'] = 21 - 3 = 18$.

Example 4.8 (Gambler's Ruin with $p \neq \frac{1}{2}$)**.** Let $T = \inf\{n \geqslant 0 : X_n = 0 \text{ or } c\}$, the expected number of bets in the game

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

Proof. Let $Z_i = +1$ if wining the ith bet, otherwise $Z_i = -1$. So $m = \mathbb{E}[Z_i] = p(1) + (1-p)(-1) = 2p - 1 \neq 0$. We know $\mathbb{E}[T] < \infty$. Hence by Wald's Theorem, $\mathbb{E}[X_T] = a + m\mathbb{E}[T]$ and thus $\mathbb{E}[T] = \frac{1}{m}(\mathbb{E}[X_T] - a)$. We have

$$\mathbb{E}[X_T] = cs(a) + 0(1 - s(a)) = c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1}$$

and so

$$\mathbb{E}[T] = \frac{1}{m}(\mathbb{E}[X_T] - a) = \frac{1}{2p-1}\left(c\frac{\left(\frac{1-p}{p}\right)^a - 1}{\left(\frac{1-p}{p}\right)^c - 1} - a\right).$$

□

For example, if $p = 0.49, a = 9700$ and $c = 10000$, then $\mathbb{E}[T] = 484997$.

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

Proof. First, $\mathbb{E}[|Y_n|] \leqslant \mathrm{Var}[X_n] + nv = 2nv < \infty$. Then, since Z_{n+1} is independent of $Z_1, \cdots, Z_n, Y_0, \cdots, Y_n$ with $\mathbb{E}[Z_{n+1}] = 0$ and $\mathbb{E}[(Z_{n+1})^2] = v$, we have

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

□

Example 4.9 (Gambler's Ruin with $p = \frac{1}{2}$). $\mathbb{E}[T] = \text{Var}[X_t] = a(c-a)$.

Proof. Let $Y_n = (X_n - a)^2 - n$ ($v = 1$). Then $\{Y_n\}$ is a martingale. Take $M > 0$, let $S_M = \min(T, M)$, and thus S_M is a stopping time which is bounded by M. Hence, by the optional stopping lemma, $\mathbb{E}[Y_{S_M}] = \mathbb{E}[Y_0] = (a-a)^2 - 0 = 0$. Also, $Y_{S_M} = (X_{S_M} - a)^2 - S_M$, so $\mathbb{E}[S_M] = \mathbb{E}[(X_{S_M} - a)^2]$. As $M \to \infty$, S_M increases monotonically to T, so $\mathbb{E}[S_M] \to \mathbb{E}[T]$ by the monotone convergence theorem. Also $\mathbb{E}[(X_{S_M} - a)^2] \to \mathbb{E}[(X_T - a)^2]$ by the bounded convergence theorem since $(X_{S_M} - a)^2 \leqslant \max(a^2, (c-a)^2) < \infty$. Hence as $M \to \infty, \mathbb{E}[T] = \mathbb{E}[(X_T - a)^2]$. Since $\mathbb{E}[X_T] = a, \mathbb{E}[(X_T - a)^2] = \text{Var}[X_T]$. Finally,

$$\text{Var}[X_T] = \frac{a}{c}(c-a)^2 + \left(1 - \frac{a}{c}\right) a^2 = a(c-a).$$

□

For example, if $c = 10000, a = 9700$ and $p = \frac{1}{2}$, then $\mathbb{E}[T] = a(c-a) = 2910000$.

Example 4.10 (Martingale Strategy). A gambling strategy involves wagering different amounts on different bets. One common example is the double until you win strategy (also called a martingale strategy though it is different from the martingales discussed above): First you wager \$1, then if you lose you wager \$2, then if you lose again you wager \$4, etc. As soon as you win any one bet you stop, with total net gain always equal to \$1. For any $p > 0$ w.p. 1 you will eventually win a bet, so this appears to guarantee a \$1 profit. But you might have a run of very bad luck, reach your credit limit, and have to stop with a very negative net gain.

Now let $p \leqslant \frac{1}{2}$ be the probability of winning each bet. Let $\{Z_i\}$ be i.i.d. with $P(Z_i = +1) = p$ and $P(Z_i = -1) = 1 - p$. Then $m := \mathbb{E}[Z_i] = p - (1-p) = 2p - 1 \leqslant 0$. Suppose we bet \$$2^{i-1}$ on the ith bet, then the winnings after n bets is $X_n = \sum_{i=1}^{n} 2^{i-1} Z_i$. Let $T = \inf\{n \geqslant 1 : Z_n = +1\}$ be time of first win and suppose we stop betting at time T. T has a geometric distribution with $\mathbb{E}[T] = \frac{1}{p} < \infty$. Also w.p. 1, $T < \infty$ and $X_T = -1 - 2 - \cdots - 2^{T-1} + 2^T = +1$, i.e., we are guaranteed to be up \$1 at time T, even though we start at $a = 0$ and have average gain $m \leqslant 0$ on each bet. So, if X_n is the total amount won by time n, then $\mathbb{E}[X_T] = +1$. However, $a + m\mathbb{E}[T] = \frac{2p-1}{p} \leqslant 0$.

Note that it does not contradict Wald's Theorem since $\{2^{i-1} Z_i\}$ is not i.i.d..

4.4 Application: Sequence Waiting Time

Example 4.11 (HTH). Suppose that at each time n, a new player appears and bets \$1 on H, then if they win, they bet \$2 on T; then if they win again, they bets \$4 on H. Each player stops betting as soon as they either lose once (and hence are down a total of \$1) or win three bets in a row (and hence are up a total of \$7) as table shown (one possible).

Let X_n be the total amount won by all payers up to time n, then since the bets were fair, $\{X_n\}$ is a martingale, and τ is a stopping time for $\{X_n\}$. Each of the first $\tau - 3$ players (i.e., all but the last three) each has a net loss of \$1 (after either one or two or three bets). Player $\tau - 2$ wins all three bets for a gain of \$7. Player $\tau - 1$ bets on T and loses, for a loss of \$1. Player τ bets on H and wins, for a gain of \$1. Hence $X_\tau = (\tau - 3)(-1) + 7 + (-1) + 1 = -\tau + 10$.

Let $T_m = \min(\tau, m)$, then $\mathbb{E}[X_{T_m}] = 0$ by optional stopping lemma and $\lim_{m\to\infty} \mathbb{E}[X_{T_m}] = \mathbb{E}[X_\tau]$ by dominated convergence theorem with dominator $Y = 7\tau$ since $|X_n - X_{n-1}| \leqslant 7$ and $\mathbb{E}[\tau] < \infty$. Thus, $\mathbb{E}[X_\tau] = \mathbb{E}[X_0] = 0$. Therefore, $\mathbb{E}[\tau] = 10$.

n	1	2	3	4	5	6	7	8	Total
Flip	H	H	T	T	H	H	T	H	
P1	+1	−2	0	0	0	0	0	0	−1
P2	0	+1	+2	−4	0	0	0	0	−1
P3	0	0	−1	0	0	0	0	0	−1
P4	0	0	0	−1	0	0	0	0	−1
P5	0	0	0	0	+1	−2	0	0	−1
P6	0	0	0	0	0	+1	+2	+4	+7
P7	0	0	0	0	0	0	−1	0	−1
P8	0	0	0	0	0	0	0	+1	+1

Example 4.12 (THH)**.** Similarly, we have $X_\tau = (\tau - 3)(-1) + 7 + (-1) + (-1) = -\tau + 8$ and thus $\mathbb{E}[\tau] = 8$.

4.5 Martingale Convergence Theorem

Theorem 4.3 (Martingale Convergence Theorem)**.** Any martingale $\{X_n\}$ which is bounded below (i.e., $X_n \geqslant c$ for all n and for some finite c), or is bounded above (i.e., $X_n \leqslant c$ for all n and for some finite c), converges w.p. 1 to some r.v. X.

Example 4.13. (1) Simple symmetric random walk is a martingale which does not converge and it is not bounded.

(2) If we modify simple symmetric random walk to stop whenever it reaches 0 after starting at a positive number, then it is still a martingale, and now it is non-negative and hence bounded below. Since $f_{i0} = 1, \forall i$, it will eventually reach 0, and hence it converges to $X = 0$.

(3) If we modify it so that from 0 it always moves to 1, then it is again non-negative, but it does not converge since the modification is not a martingale.

4.6 Application: Branching Process

A ***branching process*** is defined as follows: Let X_n equal the number of individuals (e.g., people with colds) which are present at time n. Start with $X_0 = a$ individuals for some $0 < a < \infty$, then at each time n, each of the X_n individuals creates a random number of offspring to appear at time $n+1$ (and then vanishes). The number of offspring of each individual follows i.i.d. μ where μ is any probability distribution on $\{0, 1, \cdots\}$, called the ***offspring distribution***, i.e., each individual has i children w.p. $\mu\{i\}$. (Note we assume there is just one parent per offspring, i.e., ***asexual reproduction***. Hence, $X_{n+1} = Z_{n,1} + Z_{n,2} + \cdots + Z_{n,X_n}$ where $\{Z_{n,i}\}_{i=1}^{X_n}$ follows i.i.d. μ.

For example ,if we start with $X_0 = 2$ individuals, and the first individual has 1 offspring while the second individual has 4 offspring, then at time 1 we would have $X_1 = 1 + 4 = 5$.

$\{X_n\}$ is Markov chain on the state space $S = \{0, 1, 2, \cdots\}$. Also, $p_{00} = 1$ and $p_{0j} = 0, \forall j \geqslant 1$, which is called ***extinction***. But p_{ij} for $i \neq 0$ is complicated: p_{ij} is the probability that a sum of i different r.v.s. which follows i.i.d. μ is equal to j, i.e., $p_{ij} = (\mu * \mu \cdots * \mu)(j)$, a ***convolution*** of i copies of μ.

Let $m = \sum_i i\mu\{i\}$ be the mean of μ, called the ***reproductive number***. Assume $0 < m < \infty$, then

$$\mathbb{E}[X_{n+1}|X_0, \cdots, X_n] = \mathbb{E}[Z_{n,1} + \cdots + Z_{n,X_n}|X_0, \cdots, X_n] = mX_n.$$

Hence, by induction, $\mathbb{E}[X_n] = m^n \mathbb{E}[X_0] = m^n a$.

Example 4.14. Consider a branching process, with reproductive number m. Assume that $\mu\{0\} > 0$ (and so $\mu\{1\} < 1$), then extinction is certain if $m \leqslant 1$, but both flourishing and extinction are possible if $m > 1$.

Proof. If $m < 1$, then $\mathbb{E}[X_n] \to 0$. Since

$$\mathbb{E}[X_n] = \sum_{k=0}^{\infty} kP(X_n = k) \geqslant \sum_{k=1}^{\infty} P(X_n = k) = P(X_n \geqslant 1).$$

Hence $P(X_n \geqslant 1) \leqslant \mathbb{E}[X_n] \to 0$, i.e., $P(X_n = 0) \to 1$. Therefore, if $m < 1$, extinction is certain.

If $m > 1$, then $\mathbb{E}[X_n] \to \infty$ and thus $P(X_n \to \infty) > 0$, i.e., there is a positive probability that X_n converges to infinity, which is called ***flourishing***. If assuming $\mu\{0\} > 0$, we still have $P(X_n \to 0) > 0$ (for example, this will happen if no one has any offspring at all on the first iteration, which has probability $(\mu\{0\})^a > 0$). So, $P(X_n \to \infty) \leqslant 1 - P(X_n \to 0) < 1$. Hence, if $m > 1$ and $\mu\{0\} > 0$, then we have possible extinction and flourishing.

If $m = 1$, then $\mathbb{E}[X_n] = \mathbb{E}[X_0] = a, \forall n$. Also, $\mathbb{E}[X_{n+1}|X_0, \cdots, X_n] = mX_n = X_n$, i.e., $\{X_n\}$ is a martingale. Since it is non-negative, then by the martingale convergence theorem, we have $X_n \to X$ w.p. 1 for some r.v. X. After reaching X, the process $\{X_n\}$ would still continue to fluctuate, i.e., would not converge w.p. 1, unless either $\mu\{1\} = 1$, i.e., the branching process is ***degenerate***, or $X = 0$. In conclusion, if μ is non-degenerate, i.e., $\mu\{1\} < 1$, then $X = 0$, i.e., $\{X_n\} \to 0$ w.p. 1. In the non-degenerate case, there is certain extinction, even in the borderline case when $m = 1$. □

4.7 Application: Stock Options (Discrete)

Definition 4.3. Suppose in discrete time, and X_n is the price of one share of the stock at each time n. A ***stock option*** is the option to buy one share of a stock for a fixed ***strike price*** (or ***exercise price***) \$$K$ at a fixed future ***strike time*** (or ***maturity time***) $S > 0$.

If at the strike time S, the stock price X_S is less than the strike price K, the option would not be exercised, and would thus be worth zero. But if the stock price X_S is more than K, then the option would be exercised to obtain a stock worth X_S for a price of just K, for a net profit of $X_S - K$. Hence at time S, the stock option is worth $\max(0, X_S - K)$. But at time 0, when the option is to be purchased, X_S is an random quantity.

Definition 4.4. The ***fair price*** is defined in terms of the concept of ***arbitrage***, i.e., a portfolio of investments which guarantees a positive profit no matter how the stock performs. Specifically, the fair price of a stock option is defined to be the ***no-arbitrage price***, i.e., the price for the option which makes it impossible to make a guaranteed profit through any combination of buying or selling the option, and/or buying and selling the stock.

Example 4.15. Suppose at time 0 the stock price $X_0 = 100$, and at time S the stock price X_S is random with $P(X_S = 80) = 0.9$ and $P(X_S = 130) = 0.1$. Consider an option to buy the stock at time S for $K = 110$.

At time S, the option is worth

$$\max(0, X_S - K) = \begin{cases} 0, & X_S = 80 \\ 20, & X_S = 120 \end{cases}.$$

Hence, the option's expected value at time S is given by $\mathbb{E}[\text{Option}] = 0.9 \times 0 + 0.1 \times 20 = 2$, but it is not a fair price.

Suppose that at time 0, you buy x stock shares for \$100 each, and y option shares for some price \$$c$ each. Note that we allow for the possibility that x or y is negative, corresponding to selling (i.e., ***shorting***) the stock or option instead of buying positive amounts. Then if the stock goes up to \$130, then you make \$130 − \$100 = \$30 on each stock share, and make \$$(20 - c)$ on each option share, for a total profit of $30x + (20 - c)y$. But if the stock instead goes down to \$80, you lose \$20 on each stock share, and lose \$$c$ on each option share, for a total profit of $-20x - cy$.

To attempt to make a guaranteed profit, regardless of how the stock performs, we could attempt to make these two different total profit amounts equal to each other: If $y = -\frac{5}{2}x$, then these profits both equal $\frac{5(c-8)}{2}x$, i.e., net profit is no longer random.

Thus if $c > 8$, then if you buy $x > 0$ stock shares and $y = -\frac{5}{2}x < 0$ option shares, you will make a guaranteed profit $\frac{5}{2}(c - 8)x > 0$. If $c < 8$, then if you buy $x < 0$ stock shares and $y = -\frac{5}{2}x > 0$ option shares, you will make a guaranteed profit $\frac{5}{2}(8 - c)(-x) > 0$. Either way, there is arbitrage, i.e., a guaranteed profit. But if $c = 8$, then these both equal 0, so there is no arbitrage, i.e., there is no choice of x and y which makes both possible profits positive. Hence, $c = \$8$ is the unique fair (no-arbitrage) price.

Property 4.3 (Martingale Pricing Principle)**.** The fair price of an option equals its expected value under the martingale probabilities, i.e., under the probabilities which make the stock price a martingale.

Property 4.4. Suppose a stock price at time 0 equals $X_0 = a$, and at time $S > 0$ equals either $X_S = d$ (down) or $X_S = u$ (up). where $d < a < u$. If $d < K < u$, then at time 0, the fair (no-arbitrage) price of an option to buy the stock at time S for K is

$$\frac{(a-d)(u-K)}{u-d}.$$

Proof. We can use profit computation or martingale pricing principle to prove.

We need to find martingale probabilities $q_1 = P(X_S = d)$ and $q_2 = P(X_S = u)$ to make the stock price a martingale. So we need $dq_1 + uq_2 = a$, i.e, $d(q_1) + u(1 - q_1) = a \Rightarrow (d - u)q_1 + u = a$. Hence,

$$q_1 = \frac{u-a}{u-d}, q_2 = 1 - q_1 = \frac{a-d}{u-d}.$$

Then by the martingale pricing principle, the fair price is the martingale expectation of the option's worth, which equals $q_1(0) + q_2(u - K) = \frac{a-d}{u-d}(u - K)$. □

5 Continuous Process

5.1 Brownian Motion

Definition 5.1. ***Brownian motion*** is a continuous-time process $\{B_t\}_{t\geqslant 0}$ satisfying the properties that:
(i) $B_0 = 0$.
(ii) $B_t \sim \mathcal{N}(0, t)$, i.e., normally distributed.
(iii) $B_s - B_T \sim \mathcal{N}(0, s-t)$ and is independent of B_t. More generally, if $0 \leqslant t_1 \leqslant s_1 \leqslant t_2 \leqslant s_2 \leqslant \cdots \leqslant t_k \leqslant s_k$, then $B_{s_i} - B_{t_i} \sim \mathcal{N}(0, s_i - t_i)$ and $\{B_{s_i} - B_{t_i}\}_{i=1}^k$ are all independent, i.e., independent normal increments.
(iv) Covariance structure: $\mathrm{Cov}(B_s, B_t) = \min(s, t)$.
(v) Continuous sample paths (the function $t \to B_t$ is continuous).

Note that although the function $t \to B_t$ is continuous everywhere w.p. 1, it is differentiable nowhere, i.e., it does not have any derivatives.

Example 5.1. Let $\alpha > 0, W_t = \alpha B_{t/\alpha^2}$, then $B_{t/\alpha^2} \sim \mathcal{N}(0, t/\alpha^2)$. Hence, $W_t \sim \mathcal{N}(0, \alpha^2(t/\alpha^2)) = \mathcal{N}(0, t)$. For $0 < t < s, \mathbb{E}[W_t, W_s] = \alpha^2\mathbb{E}[B_{t/\alpha^2}B_{s/\alpha^2}] = \alpha^2(t/\alpha^2) = t$. In fact, $\{W_t\}$ has all the same properties as $\{B_t\}$ so it is Brownian motion. Indeed, $\{W_t\}$ is a ***transformation*** of original Brownian motion $\{B_t\}$.

Property 5.1. Let $Y_t = B_t^2 - t$, then $\{Y_t\}$ is a martingale.

Proof. For $0 < t < s$,

$$\begin{aligned}\mathbb{E}[Y_s|\{B_r\}_{r\leqslant t}] &= \mathbb{E}[B_s^2 - s|\{B_r\}_{r\leqslant t}] = \mathrm{Var}[B_s|\{B_r\}_{r\leqslant t}] + (\mathbb{E}[B_s|\{B_r\}_{r\leqslant t}])^2 - s\\ &= s - t + (B_t)^2 - s = (B_t)^2 - t = Y_t.\end{aligned}$$

By the double-expectation formula,

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

□

Example 5.2. Let $a, b > 0, \tau = \min\{t \geqslant 0 : B_t = -a \text{ or } b\}$. Find $s = P(B_\tau = b)$ and the expected time $\mathbb{E}[\tau]$ that it takes for Brownian motion to hit $-a$ or b.

Solution. $\{B_t\}$ is martingale and τ is stopping time. $\{B_t\}$ is bounded up to time τ, i.e.,

$$|B_t|\mathbf{1}_{t\leqslant\tau} \leqslant \max(|a|, |b|).$$

By optional stopping corollary, we have $\mathbb{E}[B_\tau] = \mathbb{E}[B_0] = 0$. Hence, $sb + (1-s)(-a) = 0$. It follows that $s = \frac{a}{a+b}$.

Let $Y_t = B_t^2 - t$, then $\{Y_t\}$ is a martingale. We have

$$\mathbb{E}[Y_\tau] = \mathbb{E}[B_\tau^2 - \tau] = \mathbb{E}[B_\tau^2] - \mathbb{E}[\tau] = pb^2 + (1-p)(-a)^2 - \mathbb{E}[\tau] = \frac{a}{a+b}b^2 + \frac{b}{a+b}a^2 - \mathbb{E}[\tau] = ab - \mathbb{E}[\tau].$$

Let $\tau_M = \min(\tau, M)$, then τ_M is bounded, so $\mathbb{E}[Y_{\tau_M}] = 0$. Since $Y_{\tau_M} = B_{\tau_M}^2 - \tau_M$, then $\mathbb{E}[\tau_M] = \mathbb{E}[B_{\tau_M}^2]$. As $M \to \infty, \mathbb{E}[\tau_M] \to \mathbb{E}[\tau]$ by monotone convergence theorem, and $\mathbb{E}[B_{\tau_M}^2] \to \mathbb{E}[B_\tau^2]$ by bounded convergence theorem. Therefore, $\mathbb{E}[\tau] = \mathbb{E}[B_\tau^2]$, i.e., $\mathbb{E}[Y_\tau] = 0$. Therefore, $\mathbb{E}[\tau] = ab$.

Definition 5.2. Suppose $X_t = x_0 + \mu t + \sigma B_t, t \geqslant 0$ for some constants $x_0, \mu, \sigma \geqslant 0$. We say $\{X_t\}$ is a ***diffusion*** and call x_0 the ***initial value***, μ the ***drift***, and σ the ***volatility***.

Note that μ and σ affect the behavior of the resulting diffusion. We can also write

$$X_0 = x_0, \mathrm{d}X_t = \mu \mathrm{d}t + \sigma \mathrm{d}B_t.$$

We then can compute that $\mathbb{E}[X_t] = x_0 + \mu t, \mathrm{Var}[X_t] = \sigma^2 t, \mathrm{Cov}(X_t, X_s) = \sigma^2 \min(s,t)$, and $X_t \sim \mathcal{N}(x_0 + \mu t, \sigma^2 t)$.

Example 5.3. Suppose $X_t = 2 + 5t + 3B_t$ for $t \geqslant 0$. We have $\mathbb{E}[X_t] = 2 + 5t$ and $\mathrm{Var}[X_t] = 3^2 \mathrm{Var}[B_t] = 9t$. For $0 < t < s, \mathrm{Cov}(X_t, X_s) = \mathbb{E}[(X_t - 5t - 2)(X_s - 5s - 2)] = 9\mathbb{E}[B_t B_s] = 9t$. It follows that $X_t \sim \mathcal{N}(2 + 5t, 9t)$. We can also write $X_0 = 2, \mathrm{d}X_t = 5\mathrm{d}t + 3\mathrm{d}B_t$.

5.2 Application: Stock Options (Continuous)

A common model is to assume that

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

i.e., changes occur proportional to the total price. If $Y_t = \ln(X_t)$, then $Y_t = y_0 + \mu t + \sigma B_t$, i.e., $\mathrm{d}Y_t = \mu \mathrm{d}t + \sigma \mathrm{d}B_t$. So $\{Y_t\} = \{\ln(X_t)\}$ is a diffusion with drift (***appreciation rate***) μ and volatility σ.

We also assume a ***risk-free interest rate*** r so that \$1 at time 0 (i.e., today) is worth $\$e^{rt}$ at a time t years later. Equivalently, \$1 at a future time $t > 0$ is wroth $\$e^{-rt}$ at time 0. Thus the discounted stock price in today's dollars is

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

As a special case, if $r = 0$ then there is no discounting.

Definition 5.3. A ***stock option*** is the option to buy one share of a stock for a fixed ***strike price*** (or ***exercise price***) $\$K$ at a fixed future ***strike time*** (or ***maturity time***) $S > 0$.

Property 5.2. The fair price is equal to $\mathbb{E}[e^{-rS} \max(0, X_S - K)]$, but only after replacing μ by $r - \frac{\sigma^2}{2}$, i.e., s.t. $X_S = x_0 \exp\left[\left(r - \frac{\sigma^2}{2}\right) S + \sigma B_S\right]$.

Proof. If $\mu = r - \frac{\sigma^2}{2}$, then it can be computed that $\{D_t\}$ becomes a martingale and then the option values become a martingale. Hence, the martingale pricing principle holds: the fair price of the option at time zero is the same as the expected value of the option at time S under the martingale probabilities where $\mu = r - \frac{\sigma^2}{2}$. □

Theorem 5.1 (Black-Scholes Formula)**.** The fair (no-arbitrage) price at time 0 of an option to buy one share of a stock governed by $X_t = x_0 \exp(\mu t + \sigma B_t)$ at maturity time S for exercise price $\$K$ is equal to

$$x_0 \Phi \left(\frac{\left(r + \frac{\sigma^2}{2}\right) S - \ln\left(\frac{K}{x_0}\right)}{\sigma \sqrt{S}} \right) - e^{-rS} K \Phi \left(\frac{\left(r - \frac{\sigma^2}{2}\right) S - \ln\left(\frac{K}{x_0}\right)}{\sigma \sqrt{S}} \right)$$

where

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x.$$

Note that the assumptions (e.g., no transaction fees) is unrealistic, but it is widely used to price actual stock options.

The price does not depend on the appreciation rate μ since we first replace μ by $r - \frac{\sigma^2}{2}$: Intuitively, if μ is large, then we can already make a large profit by buying the stock itself, so the option does not add much in the way of additional value (despite the large μ).

The price is an increasing function of the volatility σ since the option protects you against potential large drops in the stock price.

Example 5.4. (1) Suppose $x_0 = \$100, K = \$110, S = 1$ (years), $r = 0.05$, i.e., 5% per year, and $\sigma = 0.3$, i.e., 30% per year. The fair option price is \$10.02.
(2) $x_0 = 200, K = 250, S = 2, r = 0.1, \sigma = 0.5$, the fair option price is \$53.60.
(3) $x_0 = 200, K = 250, S = 2, r = 0.1, \sigma = 0.8$, the fair option price is \$84.36.

5.3 Poisson Process

Property 5.3. $N(t) \sim \text{Poisson}(\lambda t)$, i.e.,

$$P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, 2, \cdots$$

and thus $\mathbb{E}[N(t)] = \text{Var}[N(t)] = \lambda t$.

Proof. $N(t) = k$ iff both $T_k \leqslant t$ and $T_{k+1} > t$, i.e., there is $0 \leqslant s \leqslant t$ with $T_k = s$ and $T_{k+1} - T_k > t - s$. Thus,

$$\begin{aligned} P(N(t) = k) &= P(T_k \leqslant t, T_{k+1} > t) = P(\exists 0 \leqslant s \leqslant t : T_k = s, Y_{k+1} > t - s) \\ &= \int_0^t f_{T_k}(s) P(Y_{k+1} > t - s) \mathrm{d}s = \int_0^t f_{T_k}(s) e^{-\lambda(t-s)} \mathrm{d}s \end{aligned}$$

where f_{T_k} is the density function for T_k.

Since $Y_n \sim \text{Exponential}(\lambda) = \text{Gamma}(1, \lambda)$ then $T_k := Y_1 + \cdots + Y_k \sim \text{Gamma}(k, \lambda)$ with density function

$$f_{T_k}(s) = \frac{\lambda^k}{\Gamma(k)} s^{k-1} e^{-\lambda s} = \frac{\lambda^k}{(k-1)!} s^{k-1} e^{-\lambda s}.$$

Hence

$$\begin{aligned} P(N(t) = k) &= \int_0^t \frac{\lambda^k}{(k-1)!} s^{k-1} e^{-\lambda s} e^{-\lambda(t-s)} \mathrm{d}s \\ &= \frac{\lambda^k}{(k-1)!} e^{-\lambda t} \int_0^t s^{k-1} \mathrm{d}s = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \end{aligned}$$

□

Definition 5.4. A ***Poisson process*** with intensity $\lambda > 0$ is a collection $\{N(t)\}_{t \geqslant 0}$ of non-decreasing integer-valued r.v.s. s.t.
(1) $N(0) = 0$;
(2) $N(t) \sim \text{Poisson}(\lambda t), \forall t \geqslant 0$;
(3) Independent Poisson increments: If $0 \leqslant t_1 \leqslant s_1 \leqslant t_2 \leqslant s_2 \cdots \leqslant t_k \leqslant s_k$ then $N(s_i) - N(t_i) \sim \text{Poisson}(\lambda(s_i - t_i))$.

Example 5.5.

$$\begin{aligned} P(N(4)=1|N(5)=3) &= \frac{P(N(4)=1,N(5)=3)}{P(N(5)=3} \\ &= \frac{P(N(4)=1,N(5)-N(4)=2)}{P(N(5)=3)} = \frac{12}{125}. \end{aligned}$$

Property 5.4. If $N(t)$ is a Poisson process with rate $\lambda > 0$, then as $h \downarrow 0$:
(1) $P(N(t+h)-N(t)=1) = \lambda h + o(h)$.
(2) $P(N(t+h)-N(t) \geqslant 2) = o(h)$.

Proof. We have

$$\begin{aligned} P(N(t+h)-N(t)=1) = P(N(h)=1) &= \frac{e^{\lambda h}(\lambda h)^1}{1!} = [1+\lambda h + O(h^2)](\lambda h) \\ &= \lambda h + \lambda^2 h^2 + O(h^3) = \lambda h + O(h^2) = \lambda h + o(h). \end{aligned}$$

Also,

$$\begin{aligned} P(N(t+h)-N(t) \geqslant 2) &= P(N(h) \geqslant 2) = 1 - P(N(h)=0) - P(N(h)=1) \\ &= 1 - \frac{e^{\lambda h}(\lambda h)^0}{0!} - \frac{e^{\lambda h}(\lambda h)^1}{1!} \\ &= 1 - [1+\lambda h + O(h^2)](1) - [1+\lambda h + O(h^2)](\lambda h) \\ &= 1 - 1 + \lambda h - \lambda h + \lambda^2 h^2 + O(h^2) + O(h^3) = O(h^2) = o(h). \end{aligned}$$

□

Property 5.5. Suppose $\{N_1(t)\}_{t\geqslant 0}$ and $\{N_2(t)\}_{t\geqslant 0}$ are independent Poisson processes with rates λ_1 and λ_2 respectively. Let $N(t) = N_1(t) + N_2(t)$, then $\{N(t)\}_{t\geqslant 0}$ is also a Poisson process with rate $\lambda_1 + \lambda_2$.

Property 5.6. Let $\{N(t)\}_{t\geqslant 0}$ be a Poisson process with rate λ. Suppose each arrival is independently of type i w.p. p_i for $i = 1, 2, \cdots$ where $\sum_i p_i = 1$. Let $N_i(t)$ be number of arrivals of type i up to time t, then the different $\{N_i(t)\}$ are all independent and each $\{N_i(t)\}$ is a Poisson process with rate λp_i.

5.4 Continuous-Time, Discrete-Space Process

Definition 5.5. A continuous-time (time-homogeneous, non-explosive) ***Markov process*** on a countable (discrete) state space S, is a collection $\{X(t)\}_{t\geqslant 0}$ of r.v.s. s.t.

$$P(X_{t_0} = i_0, \cdots, X_{t_n} = i_n) = v_{i0} p_{i_0 i_1}^{(t_1)} p_{i_1 i_2}^{(t_2 - t_1)} \cdots p_{i_{n-1} i_n}^{(t_n - t_{n-1})}$$

for all $i_0, \cdots, i_n \in S$ and all $0 < t_1 < \cdots < t_n$.

Definition 5.6. Given a standard Markov process, its ***generator*** is

$$g_{ij} = \lim_{t\downarrow 0} \frac{p_{ij}^{(t)} - \delta_{ij}}{t} = p'_{ij}(0)$$

where $'$ means the right-handed derivative with matrix

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

Theorem 5.2 (Continuous-Time Transitions Theorem)**.** If a continuous-time Markov process has generator matrix G, then for any $t \geqslant 0$,

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

Proof. Recall that for $c \in \mathbb{R}$,

$$\lim_{n\to\infty}\left(1 + \frac{c}{n}\right)^n = e^c = 1 + c + \frac{c^2}{2!} + \frac{c^3}{3!} + \cdots .$$

Similarly, for any matrix A,

$$\lim_{n\to\infty}\left(I + \frac{1}{n}A\right)^n = \exp(A)$$

where we define $\exp(A) := I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$. Hence, using the Chapman-Kolmogorov equation, for any $m \in \mathbb{N}$,

$$\begin{aligned} P^{(t)} &= [P^{(t/m)}]^m = \lim_{n\to\infty}[P^{(t/n)}]^n \\ &= \lim_{n\to\infty}\left(I + \frac{t}{n}G\right)^n = \exp(tG). \end{aligned}$$

□

Definition 5.7. A probability $\{\pi_i\}$ is a ***stationary distribution*** for a continuous-time Markov process with generator $\{g_{ij}\}$ if

$$\sum_{i\in S}\pi_i g_{ij} = 0, \forall j \in S,$$

i.e., $\pi G = 0$.

Definition 5.8. A continuous-time Markov process with generator $\{g_{ij}\}$ is ***reversible*** w.r.t. $\{\pi_i\}$ if $\pi_i g_{ij} = \pi_j g_{ji}, \forall i, j \in S$.

Property 5.7. If a continuous-time Markov process with generator $\{g_{ij}\}$ is reversible w.r.t. $\{\pi_i\}$, then $\{\pi_i\}$ is a stationary distribution for the process.

Theorem 5.3 (Continuous-Time Markov Convergence)**.** If a continuous-time standard Markov process is irreducible and has a stationary distribution π, then

$$\lim_{t\to\infty} p_{ij}^{(t)} = \pi_j, \forall i, j \in S.$$

Property 5.8. The rate λ exponential holding times version of a discrete-time Markov chain with transitions $\{\widehat{p}_{ij}\}_{i,j\in S}$ has generator $G = \lambda(\widehat{P} - I)$, i.e., $g_{ii} = \lambda(\widehat{p}_{ii} - 1)$, and for $i \neq j, g_{ij} = \lambda\widehat{p}_{ij}$.

Corollary 5.1. A Poisson process $\{N(t)\}_{t\geqslant 0}$ with rate $\lambda > 0$ has generator given by

$$G = \lambda(\widehat{P} - I) = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ 0 & -\lambda & \lambda & 0 & \cdots \\ 0 & 0 & -\lambda & \lambda & \cdots \end{pmatrix}.$$

Definition 5.9. Suppose $\{X_t\}$ is a continuous-time process. Let $\{\widehat{X}_n\}$ be the corresponding ***jump chain***, i.e., the discrete-time Markov chain consisting of each new state that $\{X_t\}$ visits.

Property 5.9. If a continuous-time Markov process X_t has generator G, then its corresponding jump chain $\{\widehat{X}_n\}$ has transition probabilities $\widehat{p}_{ii} = 0$, and for $j \neq i$,

$$\widehat{p}_{ij} = \frac{g_{ij}}{\sum\limits_{k\neq i} g_{ik}}.$$