

MATH 171 Stochastic Processes Lecture Note

Hanbaek Lyu

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, LOS ANGELES, CA 90095

Email address: `hlyu@math.ucla.edu`

`WWW.HANAEKLYU.COM`

Contents

Chapter 1. Markov chains	3
1. Definition and examples	3
2. Stationary distribution and examples	7
3. Existence of stationary distribution	9
4. Uniqueness of stationary distribution	11
5. Convergence to the stationary distribution	13
6. Markov chain Monte Carlo	17
Chapter 2. Poisson Processes	22
1. Recap of exponential and Poisson random variables	22
2. Poisson processes as an arrival process	24
3. Memoryless property and stopping times	25
4. Merging and splitting of Poisson process	28
5. M/M/1 queue	31
6. Poisson process as a counting process	33
7. Nonhomogeneous Poisson process	35
Chapter 3. Renewal Processes	37
1. Definition of renewal processes	37
2. Renewal reward processes	40
3. Little's Law	43
Chapter 4. Martingales	49
1. Conditional expectation	49
2. Definition and examples of martingales	50
3. Basic properties of martingales	54
4. Gambling strategies and stopping times	56
5. Applications of martingales at stopping times	58
Chapter 5. Introduction to mathematical finance	62
1. Hedging and replication in the two-state world	62
2. The fundamental theorem of asset pricing	63
3. The binomial model	65
Bibliography	71

CHAPTER 1

Markov chains

Say we would like to model the USD price of bitcoin. We could observe the actual price at every hour and record it by a sequence of real numbers $x_1, x_2, \dots$. However, it is more interesting to build a ‘model’ that could predict the price of bitcoin at time t , or at least give some meaningful insight how the actual bitcoin price behaves over time. Since there are so many factors affecting the price at every time, it might be reasonable that the price at time t should be given by a certain RV, say X_t . Then our sequence of predictions would be a sequence of RVs, $(X_t)_{t \geq 0}$. This is an example of *stochastic processes*. Here ‘process’ means that we are not interest in just a single RV, that their sequence as a whole: ‘stochastic’ means that the way the RVs evolve in time might be random.

In this note, we will be studying a very important class of stochastic processes called *Markov chains*. The importance of Markov chains lies two places: 1) They are applicable for a wide range of physical, biological, social, and economical phenomena, and 2) the theory is well-established and we can actually compute and make predictions using the models.

1. Definition and examples

Roughly speaking, *Markov processes* are used to model temporally changing systems where future state only depends on the current state. For instance, if the price of bitcoin tomorrow depends only on its price today, then bitcoin price can be modeled as a Markov process. (Of course, the entire history of price often affects decisions of buyers/sellers so it may not be a realistic assumption.)

Even though Markov processes can be defined in vast generality, we concentrate on the simplest setting where the state and time are both discrete. Let $\Omega = \{1, 2, \dots, m\}$ be a finite set, which we call the *state space*. Consider a sequence $(X_t)_{t \geq 0}$ of Ω -valued RVs, which we call a *chain*. We call the value of X_t the *state* of the chain at time t . In order to narrow down the way the chain $(X_t)_{t \geq 0}$ behaves, we introduce the following properties:

- (i) (Markov property) The distribution of X_{t+1} given the history $X_0, X_1, \dots, X_t$ depends only on X_t . That is, for any values of $j_0, j_1, \dots, j_t, k \in \Omega$,

$$\mathbb{P}(X_{t+1} = k | X_t = j_t, X_{t-1} = j_{t-1}, \dots, X_1 = j_1, X_0 = j_0) = \mathbb{P}(X_{t+1} = k | X_t = j_t). \quad (1)$$

- (ii) (Time-homogeneity) The transition probabilities

$$p_{ij} = \mathbb{P}(X_{t+1} = j | X_t = i) \quad i, j \in \Omega \quad (2)$$

do not depend on t .

When the chain $(X_t)_{t \geq 0}$ satisfies the above two properties, we say it is a (discrete-time and time-homogeneous) *Markov chain*. We define the *transition matrix* P to be the $m \times m$ matrix of transition probabilities:

$$P = (p_{ij})_{1 \leq i, j \leq m} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}. \quad (3)$$

Finally, since the state X_t of the chain is a RV, we represent its probability mass function (PMF) via a row vector

$$\mathbf{r}_t = [\mathbb{P}(X_t = 1), \mathbb{P}(X_t = 2), \dots, \mathbb{P}(X_t = m)]. \quad (4)$$

Example 1.1. Let $\Omega = \{1, 2\}$ and let $(X_t)_{t \geq 0}$ be a Markov chain on Ω with the following transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \quad (5)$$

We can also represent this Markov chain pictorially as in Figure 4, which is called the ‘state space diagram’ of the chain $(X_t)_{t \geq 0}$.

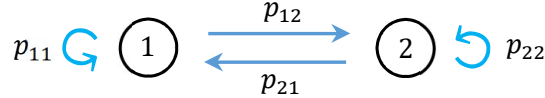


FIGURE 1. State space diagram of a 2-state Markov chain

For some concrete example, suppose

$$p_{11} = 0.2, \quad p_{12} = 0.8, \quad p_{21} = 0.6, \quad p_{22} = 0.4. \quad (6)$$

If the initial state of the chain X_0 is 1, then

$$\mathbb{P}(X_1 = 1) = \mathbb{P}(X_1 = 1 | X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 1 | X_0 = 2)\mathbb{P}(X_0 = 2) \quad (7)$$

$$= \mathbb{P}(X_1 = 1 | X_0 = 1) = p_{11} = 0.2 \quad (8)$$

and similarly,

$$\mathbb{P}(X_1 = 2) = \mathbb{P}(X_1 = 2 | X_0 = 1)\mathbb{P}(X_0 = 1) + \mathbb{P}(X_1 = 2 | X_0 = 2)\mathbb{P}(X_0 = 2) \quad (9)$$

$$= \mathbb{P}(X_1 = 2 | X_0 = 1) = p_{12} = 0.8. \quad (10)$$

Also we can compute the distribution of X_2 . For example,

$$\mathbb{P}(X_2 = 1) = \mathbb{P}(X_2 = 1 | X_1 = 1)\mathbb{P}(X_1 = 1) + \mathbb{P}(X_2 = 1 | X_1 = 2)\mathbb{P}(X_1 = 2) \quad (11)$$

$$= p_{11}\mathbb{P}(X_1 = 1) + p_{21}\mathbb{P}(X_1 = 2) \quad (12)$$

$$= 0.2 \cdot 0.2 + 0.6 \cdot 0.8 = 0.04 + 0.48 = 0.52. \quad (13)$$

In general, the distribution of X_{t+1} can be computed from that of X_t via a simple linear algebra. Note that for $i = 1, 2$,

$$\mathbb{P}(X_{t+1} = i) = \mathbb{P}(X_{t+1} = i | X_t = 1)\mathbb{P}(X_t = 1) + \mathbb{P}(X_{t+1} = i | X_t = 2)\mathbb{P}(X_t = 2) \quad (14)$$

$$= p_{1i}\mathbb{P}(X_t = 1) + p_{2i}\mathbb{P}(X_t = 2). \quad (15)$$

This can be written as

$$[\mathbb{P}(X_{t+1} = 1), \mathbb{P}(X_{t+1} = 2)] = [\mathbb{P}(X_t = 1), \mathbb{P}(X_t = 2)] \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}. \quad (16)$$

That is, if we represent the distribution of X_t as a row vector, then the distribution of X_{t+1} is given by multiplying the transition matrix P to the left. $\blacktriangle$

Example 1.2 (Gambler's ruin). Suppose a gambler has fortune of k dolars initially and starts gambling. At each time he wins or loses 1 dolar independently with probability p and $1 - p$, respectively. The game ends when his fortune reaches either 0 or N dolars. What is the probability that he wins N dolars and goes home happy?

We use Markov chains to model his fortune after betting t times. Namely, let $\Omega = \{0, 1, 2, \dots, N\}$ be the state space. Let $(X_t)_{t \geq 0}$ be a sequence of RVs where X_t is the gambler's fortune after betting t times. We first draw the state space diagram for $N = 4$ below: Next, we can write down its transition probabilities as

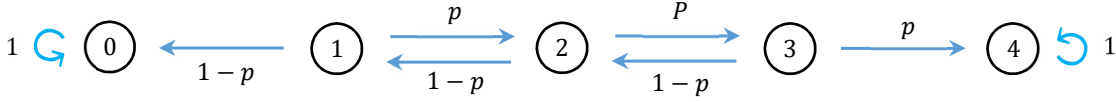


FIGURE 2. State space diagram of a 5-state gambler's chain

$$\begin{cases} \mathbb{P}(X_{t+1} = k+1 | X_t = k) = p & \forall 1 \leq k < N \\ \mathbb{P}(X_{t+1} = k | X_t = k+1) = 1-p & \forall 1 \leq k < N \\ \mathbb{P}(X_{t+1} = 0 | X_t = 0) = 1 \\ \mathbb{P}(X_{t+1} = N | X_t = N) = 1. \end{cases} \quad (17)$$

For example, the transition matrix P for $N = 5$ is given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1-p & 0 & p & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & 0 & 0 \\ 0 & 0 & 1-p & 0 & p & 0 \\ 0 & 0 & 0 & 1-p & 0 & p \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (18)$$

We call the resulting Markov chain $(X_t)_{t \geq 0}$ the *gambler's chain*. ▲

Example 1.3 (Ehrenfest Chain). This chain is originated from the physics literature as a model for two cubical volumes of air connected by a thin tunnel. Suppose there are total N indistinguishable balls split into two “urns” A and B . At each step, we pick up one of the N balls uniformly at random, and move it to the other urn. Let X_t denote the number of balls in urn A after t steps. This is a Markov chain called the *Ehrenfest chain*. (See the state space diagram in Figure 3.)

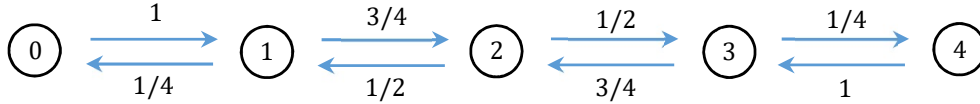


FIGURE 3. State space diagram of the Ehrenfest chain with 4 balls

It is easy to figure out the transition probabilities by considering different cases. If $X_t = k$, then urn B has $N - k$ balls at time t . If $0 < k < N$, then with probability k/N we move one ball from A to B and with probability $(N - k)/N$ we move one from B to A . If $k = 0$, then we must pick up a ball from urn B so $X_{t+1} = 1$ with probability 1. If $k = N$, then we must move one from A to B and $X_{t+1} = N - 1$ with probability 1. Hence, the transition kernel is given by

$$\begin{cases} \mathbb{P}(X_{t+1} = k+1 | X_t = k) = (N - k)/N & \forall 0 \leq k < N \\ \mathbb{P}(X_{t+1} = k-1 | X_t = k) = k/N & \forall 0 < k \leq N \\ \mathbb{P}(X_{t+1} = 1 | X_t = 0) = 1 \\ \mathbb{P}(X_{t+1} = N-1 | X_t = N) = 1. \end{cases} \quad (19)$$

For example, the transition matrix P for $N = 5$ is given by

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/5 & 0 & 4/5 & 0 & 0 & 0 \\ 0 & 2/5 & 0 & 3/5 & 0 & 0 \\ 0 & 0 & 3/5 & 0 & 2/5 & 0 \\ 0 & 0 & 0 & 4/5 & 0 & 1/5 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \quad (20)$$



Exercise 1.4. Repeat rolling two four sided dices with numbers 1,2,3, and 4 on them. Let Y_k be the some of the two dice at the k th roll. Let $S_n = Y_1 + Y_2 + \dots + Y_n$ be the total of the first n rolls, and define $X_t = S_t \pmod{6}$. Show that $(X_t)_{t \geq 0}$ is a Markov chain on the state space $\Omega = \{0, 1, 2, 3, 4, 5\}$. Furthermore, identify its transition matrix.

We generalize our observation in Example 1.1 in the following exercise.

Exercise 1.5. Let $(X_t)_{t \geq 0}$ be a Markov chain on state space $\Omega = \{1, 2, \dots, m\}$ with transition matrix $P = (p_{ij})_{1 \leq i, j \leq m}$. Let $\mathbf{r}_t = [\mathbb{P}(X_t = 1), \dots, \mathbb{P}(X_t = m)]$ denote the row vector of the distribution of X_t .

(i) Show that for each $i \in \Omega$,

$$\mathbb{P}(X_{t+1} = i) = \sum_{j=1}^m p_{ji} \mathbb{P}(X_t = j). \quad (21)$$

(ii) Show that for each $t \geq 0$,

$$\mathbf{r}_{t+1} = \mathbf{r}_t P. \quad (22)$$

(iii) Show by induction that for each $t \geq 0$,

$$\mathbf{r}_t = \mathbf{r}_0 P^t. \quad (23)$$

While right-multiplication of P advances a given row vector of distribution one step forward in time, left-multiplication of P on a column vector computes the expectation of a given function with respect to the future distribution. This point is clarified in the following exercise.

Exercise 1.6. Let $(X_t)_{t \geq 0}$ be a Markov chain on a state space $\Omega = \{1, 2, \dots, m\}$ with transition matrix P . Let $f : \Omega \rightarrow \mathbb{R}$ be a function. Suppose that if the chain X_t has state x at time t , then we get a ‘reward’ of $f(x)$. Let $\mathbf{r}_t = [\mathbb{P}(X_t = 1), \dots, \mathbb{P}(X_t = m)]$ be the distribution of X_t . Let $\mathbf{v} = [f(1), f(2), \dots, f(m)]^T$ be the column vector representing the reward function f .

(i) Show that the expected reward at time t is given by

$$\mathbb{E}[f(X_t)] = \sum_{i=1}^m f(i) \mathbb{P}(X_t = i) = \mathbf{r}_t \mathbf{v}. \quad (24)$$

(ii) Use part (i) and Exercise 1.5 to show that

$$\mathbb{E}[f(X_t)] = \mathbf{r}_0 P^t \mathbf{v}. \quad (25)$$

(iii) The total reward up to time t is a RV given by $R_t = \sum_{k=0}^t f(X_k)$. Show that

$$\mathbb{E}[R_t] = \mathbf{r}_0 (I + P + P^2 + \dots + P^t) \mathbf{v}. \quad (26)$$

Exercise 1.7. Suppose that the probability it rains today is 0.4 if neither of the last two days was rainy, but 0.5 if at least one of the last two days was rainy. Let $\Omega = \{S, R\}$, where S =sunny and R =rainy. Let W_t be the weather of day t .

(i) Show that $(W_t)_{t \geq 0}$ is *not* a Markov chain.

(ii) Expand the state space into the set of pairs $\Sigma := \Omega^2$. For each $t \geq 0$, define $X_t = (W_{t-1}, W_t) \in \Sigma$. Show that $(X_t)_{t \geq 0}$ is a Markov chain on Σ . Identify its transition matrix.

(iii) What is the two-step transition matrix?

(iv) What is the probability that it will rain on Wednesday if it didn’t rain on Sunday and Monday?

2. Stationary distribution and examples

Let $(X_t)_{t \geq 0}$ be a Markov chain on state space $\Omega = \{1, 2, \dots, m\}$ with transition matrix P . If π is a distribution on Ω such that

$$\pi = \pi P, \quad (27)$$

then we say π is a *stationary distribution* of the Markov chain $(X_t)_{t \geq 0}$.

In Exercise 1.5, we observed that we can simply multiply the transition matrix P to a given row vector $\mathbf{r}_t$ of distribution on the state space Ω in order to get the next distribution $\mathbf{r}_{t+1}$. Hence if the initial distribution of the chain is π , then its distribution is invariant in time.

Example 2.1. Consider the 2-state Markov chain $(X_t)_{t \geq 0}$ with transition matrix (as in Exercise 1.1)

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}. \quad (28)$$

Then $\pi = [3/7, 4/7]$ is a stationary distribution of X_t . Indeed,

$$[3/7, 4/7] = [3/7, 4/7] \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}. \quad (29)$$

Furthermore, this is the unique stationary distribution. To see this, let $\pi = [\pi_1, \pi_2]$ be a stationary distribution of X_t . Then $\pi = \pi P$ gives

$$0.2\pi_1 + 0.6\pi_2 = \pi_1 \quad (30)$$

$$0.8\pi_1 + 0.4\pi_2 = \pi_2. \quad (31)$$

These equations lead to

$$4\pi_1 = 3\pi_2. \quad (32)$$

Since π is a probability distribution, $\pi_1 + \pi_2 = 1$ so $\pi = [3/7, 4/7]$ is the only solution. This shows the uniqueness of the stationary distribution for X_t . $\blacktriangle$

A Markov chain may have multiple stationary distributions, as the following example illustrates.

Example 2.2. Let $(X_t)_{t \geq 0}$ be a 2-state Markov chain with transition matrix

$$P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (33)$$

Then any distribution $\pi = [p, 1 - p]$ is a stationary distribution for the chain $(X_t)_{t \geq 0}$.

Stationary distributions are closely related with eigenvectors and eigenvalues of the transition matrix P . Namely, by taking transpose,

$$\pi^T = P^T \pi^T. \quad (34)$$

Hence, the transpose of any stationary distribution is an eigenvector of P^T associated with eigenvalue 1. More properties of stationary distribution in this line of thought are given by the following exercise.

Exercise 2.3. Given a matrix A , a row vector v , and a real number λ , we say v is a *left eigenvector* of A associated with *left eigenvalue* λ if

$$\mathbf{v}A = \lambda \mathbf{v}. \quad (35)$$

If $\mathbf{v}$ is a column vector and if $A\mathbf{v} = \lambda \mathbf{v}$, then we say v is a (*right*) *eigenvector* associated with (*right*) *eigenvalue* λ . Let $(X_t)_{t \geq 0}$ be a Markov chain on state space $\Omega = \{1, 2, \dots, m\}$ with transition matrix $P = (p_{ij})_{1 \leq i, j \leq m}$.

(i) Show that a distribution π on Ω is a stationary distribution for the chain $(X_t)_{t \geq 0}$ if and only if it is a left eigenvector of P associated with left eigenvalue 1.

(ii) Show that 1 is a right eigenvalue of P with right eigenvector $[1, 1, \dots, 1]^T$.

- (iii) Recall that a square matrix and its transpose have the same (right) eigenvalues and corresponding (right) eigenspaces have the same dimension. Show that the Markov chain $(X_t)_{t \geq 0}$ has a unique stationary distribution if and only if $[1, 1, \dots, 1]^T$ spans the (right) eigenspace of P associated with (right) eigenvalue 1.

In the following exercise, we compute the stationary distribution of the so-called birth-death chain.

Exercise 2.4 (Birth-Death chain). Let $\Omega = \{0, 1, 2, \dots, N\}$ be the state space. Let $(X_t)_{t \geq 0}$ be a Markov chain on Ω with transition probabilities

$$\begin{cases} \mathbb{P}(X_{t+1} = k+1 | X_t = k) = p & \forall 0 \leq k < N \\ \mathbb{P}(X_{t+1} = k-1 | X_t = k) = 1-p & \forall 1 \leq k \leq N \\ \mathbb{P}(X_{t+1} = 0 | X_t = 0) = 1-p \\ \mathbb{P}(X_{t+1} = N | X_t = N) = p. \end{cases} \quad (36)$$

This is called a Birth-Death chain. Its state space diagram is as below.

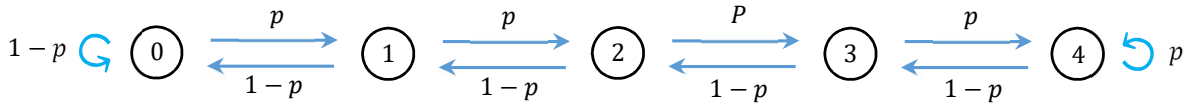


FIGURE 4. State space diagram of a 5-state Birth-Death chain

- (i) Let $\pi = [\pi_0, \pi_1, \dots, \pi_N]$ be a distribution on Ω . Show that π is a stationary distribution of the Birth-Death chain if and only if it satisfy the following ‘balance equation’

$$p\pi_k = (1-p)\pi_{k+1} \quad 0 \leq k < N. \quad (37)$$

- (ii) Let $\rho = p/(1-p)$. From (ii), deduce that $\pi_k = \rho^k \pi_0$ for all $0 \leq k < N$.

- (iii) Using the normalization condition $\pi_0 + \pi_1 + \dots + \pi_N = 1$, show that $\pi_0 = 1/(1 + \rho + \rho^2 + \dots + \rho^N)$. Conclude that

$$\pi_k = \frac{\rho^k}{1 + \rho + \rho^2 + \dots + \rho^N} \quad 0 \leq k \leq N. \quad (38)$$

Conclude that the Birth-Death chain has a unique stationary distribution given by (38), which becomes the uniform distribution on Ω when $p = 1/2$.

Next, we take a look at an example of an important class of Markov chains, which is called the random walk on graphs. This is the basis of many algorithms involving machine learning on networks (e.g., Google’s PageRank).

Example 2.5 (Random walk on graphs). We first introduce some notions in graph theory. A *graph* G consists of a pair (V, E) of sets of nodes V and edges $E \subseteq V^2$. A graph G can be concisely represented as a $|V| \times |V|$ matrix A_G , which is called the *adjacency matrix* of G . Namely, the (i, j) entry of A_G is defined by

$$A_G(i, j) = \mathbf{1}(\text{nodes } i \text{ and } j \text{ are adjacent in } G). \quad (39)$$

We say G is *simple* if $(i, j) \in E$ implies $(j, i) \in E$ and $(i, i) \notin E$ for all $i \in V$. For a simple graph $G = (V, E)$, we say a node j is *adjacent* to i if $(i, j) \in E$. We denote $\deg_G(i)$ the number of neighbors of i in G , which we call the *degree* of i .

Consider we hop around the nodes of a given simple graph $G = (V, E)$: at each time, we jump from one node to one of the neighbors with equal probability. For instance, if we are currently at node 2 and if 2 is adjacent to 3, 5, and 6, then we jump to one of the three neighbors with probability $1/3$. The location

of this jump process at time t can be described as a Markov chain. Namely, a Markov chain $(X_t)_{t \geq 0}$ on the node set V is called a *random walk on G* if

$$\mathbb{P}(X_{t+1} = j | X_t = i) = \frac{A_G(i, j)}{\deg_G(i)}. \quad (40)$$

Note that its transition matrix P is obtained by normalizing each row of the adjacency matrix A_G by the corresponding degree.

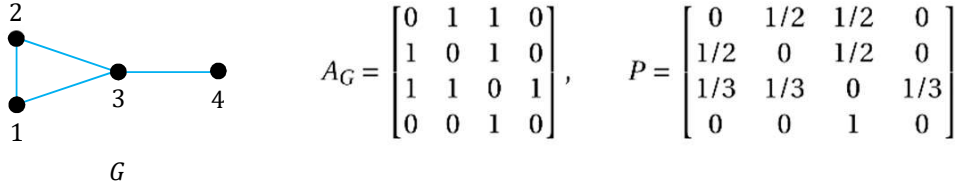


FIGURE 5. A 4-node simple graph G , its adjacency matrix A_G , and associated random walk transition matrix P

What is the stationary distribution of random walk on G ? There could be many (see Exercise 4.1), but here is a typical one that always works. Define a probability distribution π on V by

$$\pi(i) = \frac{\deg_G(i)}{\sum_{j \in V} \deg_G(j)} = \frac{\deg_G(i)}{2|E|}. \quad (41)$$

Namely, the probability that $X_t = i$ is proportional to the degree of node i . Then π is a stationary distribution for P . Indeed,

$$\sum_{j \in V} \pi(j) P(j, i) = \sum_{j \in V} \frac{\deg_G(j)}{2|E|} \frac{A_G(j, i)}{\deg_G(j)} \quad (42)$$

$$= \frac{1}{2|E|} \sum_{j \in V} A_G(j, i) = \frac{\deg_G(i)}{2|E|} = \pi(i). \quad (43)$$

Hence if we perform a random walk on facebook network, then we are about ten times more likely to visit a person of 100 friends than to visit a person of 10 friends. $\blacktriangle$

3. Existence of stationary distribution

Does a finite-state Markov chain always have a stationary distribution? If yes, when there is a unique stationary distribution? In Exercise 2.3, we have exploited the relation between stationary distribution and eigenvectors associated with eigenvalue 1 to study existence and uniqueness of stationary distribution. Namely, the all-one column vector is a right-eigenvector associated with eigenvalue 1 of a given transition matrix P . Hence its transpose P^T has an eigenvector v associated with eigenvalue 1. That is, $P^T v = v$. Then taking transpose, we get $v^T P = v^T$, so $\pi = v^T / D$, where D is the sum of all entries in v , is a stationary distribution of P . For the uniqueness, see part (iii) of Exercise 2.3.

However, the linear algebra approach do not provide us a useful intuition for the behavior of the Markov chain itself. In Theorem 3.1, we give an alternative and constructive argument to show that every Markov chain has a stationary distribution. This is valuable since we can write down what the stationary distribution is, whereas the linear algebra argument doesn't give this information.

Theorem 3.1 (Existence of stationary distribution). *Let $(X_t)_{t \geq 0}$ be a Markov chain on a finite state space Ω with transition matrix P .*

(i) For each $x, y \in \Omega$, let $V_t(x, y)$ be the number of visits to y in the first t steps given that $X_0 = x$,

$$V_t(x, y) = \sum_{k=1}^t \mathbf{1}(X_k = y | X_0 = x). \quad (44)$$

Define

$$\pi_x(y) = \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}[V_t(x, y)], \quad (45)$$

which is the expected proportion of times spending at y starting at x . Then π_x is a probability distribution on Ω .

(ii) For each $x \in \Omega$, π_x is a stationary distribution for P .

PROOF. For (i), note that $0 \leq V_t(x, y) \leq t$, so $\pi_x(y) \in [0, 1]$ for all $y \in \Omega$. To show $\sum_y \pi_x(y) = 1$, since the chain has to be at some unique state at each time, we have

$$\sum_{y \in \Omega} V_t(x, y) = \sum_{y \in \Omega} \sum_{k=1}^t \mathbf{1}(X_k = y | X_0 = x) \quad (46)$$

$$= \sum_{k=1}^t \sum_{y \in \Omega} \mathbf{1}(X_k = y | X_0 = x) = \sum_{k=1}^t 1 = t. \quad (47)$$

Hence $t^{-1} \sum_y V_t(x, y) = 1$, so by taking expectation and letting $t \rightarrow \infty$, we get

$$1 = \lim_{t \rightarrow \infty} \sum_{y \in \Omega} \frac{1}{t} \mathbb{E}[V_t(x, y)] = \sum_{y \in \Omega} \lim_{t \rightarrow \infty} \frac{1}{t} \mathbb{E}[V_t(x, y)] = \sum_{y \in \Omega} \pi_x(y). \quad (48)$$

This shows (i).

Next, we show (ii). It amounts to show the matrix equation $\pi_x P = \pi_x$, viewing π_x as a row vector. Fix $y \in \Omega$. By considering the y -th entry of $\pi_x P$, our goal is to show

$$\sum_{z \in \Omega} \pi_x(z) P(z, y) = \pi_x(y). \quad (49)$$

Indeed, first observe that

$$\mathbb{P}(X_k = z | X_0 = x) P(z, y) = \mathbb{P}(X_k = z | X_0 = x) \mathbb{P}(X_{k+1} = y | X_k = z) \quad (50)$$

$$= \mathbb{P}(X_k = z | X_0 = x) \mathbb{P}(X_{k+1} = y | X_k = z, X_0 = x) \quad (51)$$

$$= \mathbb{P}(X_{k+1} = y, X_k = z | X_0 = x). \quad (52)$$

Hence we have

$$\sum_{z \in \Omega} \pi_x(z) P(z, y) = \sum_{z \in \Omega} \lim_{t \rightarrow \infty} \frac{\mathbb{E}[V_t(x, z)]}{t} P(z, y) \quad (53)$$

$$= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{z \in \Omega} \mathbb{E}[V_t(x, z)] P(z, y) \quad (54)$$

$$= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{z \in \Omega} \sum_{k=1}^t \mathbb{P}(X_k = z | X_0 = x) P(z, y) \quad (55)$$

$$= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t \sum_{z \in \Omega} \mathbb{P}(X_{k+1} = y, X_k = z | X_0 = x) \quad (56)$$

$$= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=2}^{t+1} \mathbb{P}(X_k = y | X_0 = x) = \pi_x(y). \quad (57)$$

This shows (ii). □

4. Uniqueness of stationary distribution

Next, we consider the uniqueness of stationary distribution. Before we state our main result, we first take a look at random walk on disconnected graphs and that it must have multiple stationary distributions.

Exercise 4.1 (Random walk on disconnected graphs). Let $G = (V, E)$ be a graph with two disjoint components G_1 and G_2 . Let $(X_t)_{t \geq 0}$ be a random walk on G and denote by P its transition matrix.

(i) Let P_1 and P_2 be the transition matrices for random walks on G_1 and G_2 , respectively. Let $V_1 = \{1, 2, \dots, n\}$ and $V_2 = \{n+1, n+2, \dots, n+m\}$ be the set of nodes in G_1 and G_2 . Then show that P is of the following block diagonal form

$$P = \begin{bmatrix} P_1 & O \\ O & P_2 \end{bmatrix}. \quad (58)$$

(ii) Let $\pi_1 = [\pi_1^{(1)}, \dots, \pi_1^{(n)}]$ and $\pi_2 = [\pi_2^{(1)}, \dots, \pi_2^{(m)}]$ be any stationary distributions for P_1 and P_2 . Let

$$\bar{\pi}_1 = [\pi_1^{(1)}, \dots, \pi_1^{(n)}, 0, \dots, 0] \quad (59)$$

$$\bar{\pi}_2 = [0, \dots, 0, \pi_2^{(1)}, \dots, \pi_2^{(m)}] \quad (60)$$

be distributions on $V = V_1 \cup V_2 = \{1, 2, \dots, n, n+1, \dots, n+m\}$. Show that for any $\lambda \in [0, 1]$, $\pi := \lambda \bar{\pi}_1 + (1 - \lambda) \bar{\pi}_2$ is a stationary distribution for P .

(iii) Recall that random walk on arbitrary graph has a stationary distribution, where probability of each node is proportional to its degree. Conclude that the random walk $(X_t)_{t \geq 0}$ on G has infinitely many stationary distributions.

A graph G is said to be *connected* if there exists a ‘path’ between any given two nodes. A *path* in G is a sequence of distinct nodes $v_0, v_1, \dots, v_k$ such that consecutive nodes are adjacent in G . In Exercise 4.1, we have seen that the random walk on a disconnected graph has infinitely many stationary distributions. How about random walk on connected graphs? We will see that they have a unique stationary distribution, which must be the typical one that we know.

Proposition 4.2 (RW on connected graphs). *Let $G = (V, E)$ be a connected graph. Then the random walk on G has a unique stationary distribution π , which is given by*

$$\pi(i) = \frac{\deg_G(i)}{2|E|}. \quad (61)$$

PROOF. Let P denote the transition matrix of random walk on G . Let $V = \{1, 2, \dots, N\}$ and let E_1 be the column eigenspace of P associated with eigenvalue 1. That is,

$$E_1 = \{\mathbf{v} \in \mathbb{R}^N \mid P\mathbf{v} = \mathbf{v}\}. \quad (62)$$

Note that $[1, 1, \dots, 1]^T \in E_1$, so E_1 has dimension ≥ 1 . According to Exercise 2.3, it suffices to show that E_1 has dimension 1. This is equivalent to say that E_1 is spanned by $[1, \dots, 1]^T$.

The argument uses the ‘maximum principle’ for ‘harmonic functions’. Let $\mathbf{x} \in E_1$ be arbitrary. We will view $\mathbf{x}$ as a function on the node set V . So we want to show that $\mathbf{x}$ is a constant function on V . Since V is finite, there is some node k where $\mathbf{x}$ attains its global maximum. Suppose node k has some neighbor j such that $\mathbf{x}(j) < \mathbf{x}(k)$. Then from the eigenvector condition $P\mathbf{x} = \mathbf{x}$,

$$\mathbf{x}(k) = \sum_{i=1}^N P(k, i) \mathbf{x}(i) \quad (63)$$

$$= P(k, j) \mathbf{x}(j) + \sum_{i \neq j} P(k, i) \mathbf{x}(i) \quad (64)$$

$$< P(k, j) \mathbf{x}(k) + \sum_{i \neq j} P(k, i) \mathbf{x}(i) \quad (65)$$

$$\leq P(k, j)\mathbf{x}(k) + \sum_{i \neq j} P(k, i)\mathbf{x}(k) = \mathbf{x}(k), \quad (66)$$

which is a contradiction. This implies that $\mathbf{x}$ also attains its global maximum. Apply the similar argument to the neighbors of k . Repeating the same argument, we see that $\mathbf{x}$ must be constant on a ‘connected component’ of G containing k . But since G is connected, it follows that $\mathbf{x}$ is constant on V . This shows the assertion. $\square$

If we carefully examine the proof of above result, we find that we haven’t really used the value of entries of the transition matrix P . Rather, we only used an abstract property following from the connectivity of G : We can reach any node from any other node. We extract this as a general property of Markov chains.

Definition 4.3 (Irreducibility). Let P be the transition matrix of a Markov chain $(X_t)_{t \geq 0}$ on a finite state space Ω . We say the chain (or P) is *irreducible* if for any $i, j \in \Omega$,

$$\mathbb{P}(X_k = j \mid X_0 = i) > 0 \quad (67)$$

for some integer $k \geq 0$.

In words, a Markov chain is irreducible if every state is ‘accessible’ from any other state.

Exercise 4.4 (RW on connected graphs is irreducible). Let $G = (V, E)$ be a connected graph. Let $(X_t)_{t \geq 0}$ be a random walk on G and denote by P its transition matrix. Let $v_0, v_1, \dots, v_k$ such that consecutive nodes are adjacent in G . Show that

$$P^k(v_0, v_k) = \mathbb{P}(X_k = v_k \mid X_0 = v_0) \geq P(v_0, v_1)P(v_1, v_2) \cdots P(v_{k-1}, v_k) \quad (68)$$

$$= \frac{1}{\deg_G(v_0)} \frac{1}{\deg_G(v_1)} \cdots \frac{1}{\deg_G(v_k)} > 0. \quad (69)$$

Conclude that random walks on connected graphs are irreducible.

Now we state the general uniqueness theorem of stationary distribution.

Theorem 4.5 (Uniqueness of stationary distribution). Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain on a finite state space Ω . Then it has a unique stationary distribution, which is given by (45).

PROOF. Exercise. Mimic the proof of Proposition 4.2. $\square$

Exercise 4.6 (Hitting times). Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain on a finite state space Ω with transition matrix P . Let $X_0 = x \in \Omega$. For any $y \in \Omega$ and $t \geq 0$, define the *first hitting time* of y by

$$\tau(t, y) = \min\{k \geq 1 \mid X_{t+k} = y\}. \quad (70)$$

In words, $\tau(t, y)$ is the minimum number of steps needed to hit y starting from time t .

(i) Recall that by the irreducibility, for any $z, y \in \Omega$, $P^r(z, y) > 0$ for some $r \geq 1$ (r may depend on z and y). Let

$$d = \max_{z, y \in \Omega} \min\{r \geq 1 \mid P^r(z, y) > 0\}. \quad (71)$$

Show that there exists some constant $\delta > 0$ such that

$$\mathbb{P}(X_s \text{ visits } y \text{ some time during } [t, t+d] \mid X_t = z) \geq \delta > 0 \quad (72)$$

for all $t \geq 0$ and $y, z \in \Omega$.

(ii) Show that for any $k \geq 1$,

$$\mathbb{P}(\tau(0, y) > kd) \leq (1 - \delta)\mathbb{P}(\tau(0, y) > (k-1)d). \quad (73)$$

Use induction to conclude that

$$\mathbb{P}(\tau(0, y) > kd) \leq (1 - \delta)^k. \quad (74)$$

(iii) Denote $\tau_y = \tau(0, y)$. Noting that $\mathbb{P}(\tau_y \geq t)$ is a non-increasing function in t , show that

$$\mathbb{E}[\tau_y] = \sum_{t=1}^{\infty} \mathbb{P}(\tau_y \geq t) \quad (75)$$

$$\leq d \sum_{k=0}^{\infty} \mathbb{P}(\tau_y \geq kd) \quad (76)$$

$$\leq d \sum_{k=0}^{\infty} (1 - \delta)^k = d/\delta < \infty. \quad (77)$$

Hence the expected hitting time of $(X_t)_{t \geq 0}$ to any state is finite.

Exercise 4.7 (Gambler's ruin). Let $(X_t)_{t \geq 0}$ be the Gambler's chain on state space $\Omega = \{0, 1, 2, \dots, N\}$ introduced in Example 1.1. (see Lecture note 1)

- (i) Show that any distribution $\pi = [a, 0, 0, \dots, 0, b]$ on Ω is stationary with respect to the gambler's chain. Also show that any stationary distribution of this chain should be of this form.
- (ii) Clearly the gambler's chain eventually visits state 0 or N , and stays at that boundary state thereafter. This is called *absorption*. Let τ_i denote the time until absorption starting from state i :

$$\tau_i = \min\{t \geq 0 : X_t \in \{0, N\} \mid X_0 = i\}. \quad (78)$$

We are going to compute the 'winning probabilities': $q_i := \mathbb{P}(X_{\tau_i} = N)$.

By considering what happens in one step, show that they satisfy the following recursion

$$\begin{cases} q_i = p q_{i+1} + (1-p) q_{i-1} & \forall 1 \leq i < N \\ q_0 = 0, \quad q_N = 1 \end{cases}. \quad (79)$$

(iii) Denote $\rho = (1-p)/p$. Show that

$$q_{i+1} - q_i = \rho(q_i - q_{i-1}) \quad \forall 1 \leq i < N. \quad (80)$$

Deduce that

$$q_{i+1} - q_i = \rho^i (q_1 - q_0) = \rho^i q_1 \quad \forall 1 \leq i < N, \quad (81)$$

and that

$$q_i = q_1 (1 + \rho + \dots + \rho^{i-1}) \quad \forall 1 \leq i \leq N. \quad (82)$$

(iv)* Use $q_N = 1$ to deduce

$$q_1 = \frac{1}{1 + \rho + \dots + \rho^{N-1}}. \quad (83)$$

Conclude that

$$q_i = \frac{1 + \rho + \dots + \rho^{i-1}}{1 + \rho + \dots + \rho^{N-1}} = \begin{cases} \frac{1 - \rho^i}{1 - \rho^N} & \text{if } p \neq 1/2 \\ \frac{i}{N} & \text{if } p = 1/2. \end{cases} \quad (84)$$

(Remark: Unlike the Birth-Death chain problem, q_i 's do not have to add up to 1)

5. Convergence to the stationary distribution

Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain on a finite state space Ω . By Theorems 3.1 and 4.5, we know that the chain has unique stationary distribution π . Hence if we denote by π_t the distribution of X_t , then we should expect that π_t 'converges' to π as $t \rightarrow \infty$ in some sense. We will prove a precise version of this claim in this section.

We start with an example, where we show the convergence of a 2-state chain using a diagonalization of its transition matrix.

Exercise 5.1. Let $(X_t)_{t \geq 0}$ be a Markov chain on $\Omega = \{1, 2\}$ with the following transition matrix

$$P = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}. \quad (85)$$

(i) Show that P admits the following diagonalization

$$P = \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -2/5 \end{bmatrix} \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix}^{-1}. \quad (86)$$

(ii) Show that P^t admits the following diagonalization

$$P^t = \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (-2/5)^t \end{bmatrix} \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix}^{-1}. \quad (87)$$

(iii) Let $\mathbf{r}_t$ denote the row vector of distribution of X_t . Show that

$$\mathbf{r}_t = \mathbf{r}_0 \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (-2/5)^t \end{bmatrix} \begin{bmatrix} 1 & -4/3 \\ 1 & 1 \end{bmatrix}^{-1}. \quad (88)$$

Also show that

$$\lim_{t \rightarrow \infty} \mathbf{r}_t = \mathbf{r}_0 \begin{bmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{bmatrix} = [3/7, 4/7]. \quad (89)$$

(iv) Show that $\pi = [3/7, 4/7]$ is the unique stationary distribution for P . Conclude that regardless of the initial distribution $\mathbf{r}_0$, the distribution of the Markov chain $(X_t)_{t \geq 0}$ converges to $[3/7, 4/7]$.

Next, we observe that an irreducible MC may not always converge to the stationary distribution. The key issue there is the notion of ‘periodicity’.

Example 5.2. Let $(X_t)_{t \geq 0}$ be a 2-state MC on state space $\Omega = \{0, 1\}$ with transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (90)$$

Namely, the chain deterministically alternates between the two states. Note that it is irreducible and has a unique stationary distribution

$$\pi = [1/2, 1/2]. \quad (91)$$

Let π_t be the distribution of X_t , where the initial distribution is given by $\pi_0 = [1, 0]$. Then we have

$$\pi_1 = [0, 1] \quad (92)$$

$$\pi_2 = [1, 0] \quad (93)$$

$$\pi_3 = [0, 1] \quad (94)$$

$$\pi_4 = [1, 0], \quad (95)$$

and so on. Hence the distributions π_t do not converge to the stationary distribution π . ▲

Example 5.3 (RW on torus). Let $\mathbb{Z}_n$ be the set of integers modulo n . Let $G = (V, E)$ be a graph where $V = \mathbb{Z}_n \times \mathbb{Z}_n$ and two nodes $u = (u_1, u_2)$ and $v = (v_1, v_2)$ are adjacent if and only if

$$|u_1 - v_1| + |u_2 - v_2| = 1. \quad (96)$$

Such a graph G is called the $n \times n$ torus and we write $G = \mathbb{Z}_n \times \mathbb{Z}_n$. Intuitively, it is obtained from the $n \times n$ square grid by adding boundary edges to wrap around (see Figure 6 left).

Now let $(X_t)_{t \geq 0}$ be a random walk on G . Since G is connected, X_t is irreducible. Since all nodes in G have degree 4, the uniform distribution on $\mathbb{Z}_n \times \mathbb{Z}_n$, which we denote by π , is the unique stationary distribution of X_t . Let π_t denote the distribution of X_t .

For instance, consider $G = \mathbb{Z}_6 \times \mathbb{Z}_6$. As illustrated in Figure 6 below, observe that if X_0 is one of the red nodes (where sum of coordinates is even), then X_t is at a red node for any $t = \text{even}$ and at a black node (where sum of coordinates is odd) at $t = \text{odd}$. Hence, π_t is supported only on the ‘even’ nodes for even

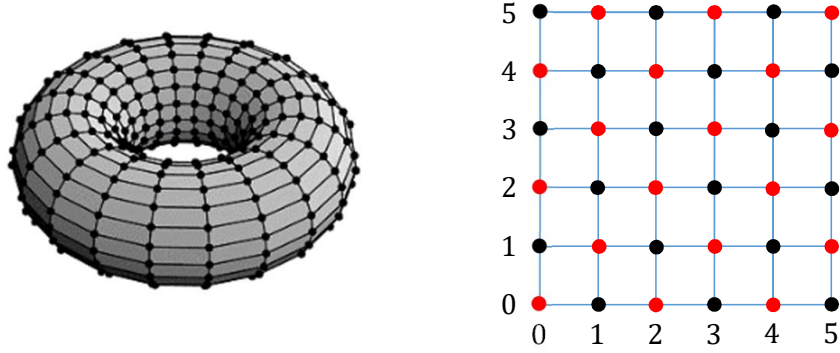


FIGURE 6. (Left) Torus graph (Figure excerpted from [LP17]). (Right) RW on torus $G = \mathbb{Z}_6 \times \mathbb{Z}_6$ has period 2.

times and on the ‘odd’ nodes for the odd times. Hence π_t does not converge in any sense to the uniform distribution π . ▲

The key issue in the 2-periodicity of RW on $G = \mathbb{Z}_6 \times \mathbb{Z}_6$ is that it takes even number of steps to return to any given node. Generalizing this observation, we introduce the following notion of periodicity.

Definition 5.4. Let P be the transition matrix of a Markov chain $(X_t)_{t \geq 0}$ on a finite state space Ω . For each state $x \in \Omega$, let $\mathcal{T}(x) = \{t \geq 1 \mid P^t(x, x) > 0\}$ be the set of times when it is possible for the chain to return to starting state x . We define the *period* of x by the greatest common divisor of $\mathcal{T}(x)$. We say the chain X_t is *aperiodic* if all states have period 1.

Example 5.5. Let $\mathcal{T}(x) = \{4, 6, 8, 10, \dots\}$. Then the period of x is 2, even though it is not possible to go from x to x in 2 steps. If $\mathcal{T}(x) = \{4, 6, 8, 10, \dots\} \cup \{3, 6, 9, 12, \dots\}$, then the period of x is 1. This means the return times to x is irregular. For the RW on $G = \mathbb{Z}_6 \times \mathbb{Z}_6$ in Example 5.3, all nodes have period 2. ▲

Exercise 5.6 (Aperiodicity of RW on graphs). Let $(X_t)_{t \geq 0}$ be a random walk on a connected graph $G = (V, E)$.

- (i) Show that all nodes have the same period.
- (ii) If G contains an odd cycle C (e.g., triangle), show that all nodes in C have period 1.
- (iii) Show that X_t is aperiodic if and only if G contains an odd cycle.
- (iv)* Show that X_t is periodic if and only if G is *bipartite*. (A graph G is bipartite if there exists a partition $V = A \cup B$ of nodes such that all edges are between A and B .)

Remark 5.7. If $(X_t)_{t \geq 0}$ is an irreducible Markov chain on a finite state space Ω , then all states $x \in \Omega$ must have the same period. The argument is similar to that for Exercise 5.6 (i).

We now state the convergence theorem for random walks on graphs.

Theorem 5.8 (Convergence of RW on graphs). Let $(X_t)_{t \geq 0}$ be a random walk on a connected graph $G = (V, E)$ with an odd cycle. Let π denote the unique stationary distribution π . Then for each $x, y \in V$,

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t = y \mid X_0 = x) = \pi(y). \quad (97)$$

PROOF. Since G contains an odd cycle, random walk on G is aperiodic according to Exercise 5.6. Let $(Y_t)_{t \geq 0}$ be another RW on G . We evolve X_t and Y_t simultaneously independently until they meet, after which we evolve them in union. Namely, let τ be the ‘meeting time’ of these two walkers:

$$\tau = \min\{t \geq 0 \mid X_t = Y_t\}. \quad (98)$$

Then we let $X_t = Y_t$ for all $t \geq \tau$. Still, if we disguise one random walk, then the other behaves exactly as it should do. (This is a ‘coupling’ between the two random walks)

Now suppose $X_0 = x$ for some $x \in V$ and Y_0 is drawn from the stationary distribution π . In particular, the distribution of Y_t is π for all t . Let π_t denote the distribution of X_t . Then for any $y \in V$,

$$|\pi_t(y) - \pi(y)| = |\mathbb{P}(X_t = y) - \mathbb{P}(Y_t = y)| \quad (99)$$

$$= |\mathbb{P}(X_t = y, Y_t = y) + \mathbb{P}(X_t = y, Y_t \neq y) - \mathbb{P}(X_t = y, Y_t = y) - \mathbb{P}(X_t \neq y, Y_t = y)| \quad (100)$$

$$\leq \mathbb{P}(X_t = y, Y_t \neq y) + \mathbb{P}(X_t \neq y, Y_t = y) \quad (101)$$

$$\leq \mathbb{P}(X_t \neq Y_t) \quad (102)$$

$$= \mathbb{P}(\tau > t). \quad (103)$$

Hence it suffices to show that $\mathbb{P}(\tau > t) \rightarrow 0$ as $t \rightarrow \infty$, as this will yield $|\pi_t(y) - \pi(y)| \rightarrow 0$ as $t \rightarrow \infty$. This follows from the fact that $\mathbb{P}(\tau < \infty) = 1$, that is, the two independent random walks on G eventually meet with probability 1 (see Exercise 5.11). $\square$

The following exercise shows that if a RW on G is irreducible and aperiodic, then it is possible to reach any node from any other node in a fixed number of steps.

Exercise 5.9. Let $(X_t)_{t \geq 0}$ be a RW on a connected graph $G = (V, E)$. Let P denote its transition matrix. Suppose G contains an odd cycle, so that the walk is irreducible and aperiodic. For each $x \in V$, let $\mathcal{T}(x)$ denote the set of all possible return times to the state x .

- (i) For any $x \in V$, show that $a, b \in \mathcal{T}(x)$ implies $a + b \in \mathcal{T}(x)$.
- (ii) For any $x \in V$, show that $\mathcal{T}(x)$ contains 2 and some odd integer b .
- (iii) For each $x \in V$, let b_x be the smallest odd integer contained in $\mathcal{T}(x)$. Show that $m \in \mathcal{T}(x)$ whenever $m \geq b_x$.
- (iv) Let $b^* = \max_{x \in V} b_x$. Show that $m \in \mathcal{T}(x)$ for all $x \in V$ whenever $m \geq b^*$.
- (v) Let $r = |V| + b^*$. Show that $P^r(x, y) > 0$ for all $x, y \in V$.

With a little more work, one can also show a similar statement for general irreducible and aperiodic Markov chains.

Exercise 5.10. Let P be the transition matrix of a Markov chain $(X_t)_{t \geq 0}$ on a finite state space Ω . Show that (i) implies (ii):

- (i) P is irreducible and aperiodic.
- (ii) There exists an integer $r \geq 0$ such that every entry of P^r is positive.

Furthermore, show that (ii) implies (i) if X_t is a RW on some graph $G = (V, E)$.

(Hint for (i) $\Rightarrow$ (ii): You may use the fact that if a subset $\mathcal{T}$ of positive integers $\mathbb{N}$ is closed under addition with $\gcd(\mathcal{T}) = 1$, then it must contain all but finitely many integers. Proof is similar to Exercise 5.9, but a bit more number-theoretic. See [LP17, Lem. 1.27].)

Exercise 5.11. Let (X_t) and (Y_t) be independent random walks on a connected graph $G = (V, E)$. Suppose that G contains an odd cycle. Let P be the transition matrix of the random walk on G .

- (i) Let $t \geq 0$ be arbitrary. Use Exercise 5.9 to deduce that for some integer $r \geq 1$, we have

$$\mathbb{P}(X_{t+r} = Y_{t+r} | X_t = x, Y_t = y) = \sum_{z \in V} \mathbb{P}(X_{t+r} = z = Y_{t+r} | X_t = x, Y_t = y) \quad (104)$$

$$= \sum_{z \in V} \mathbb{P}(X_{t+r} = z | X_t = x) \mathbb{P}(Y_{t+r} = z | Y_t = y) \quad (105)$$

$$= \sum_{z \in V} P^r(x, z) P^r(y, z) > 0. \quad (106)$$

- (ii) Let $r \geq 1$ be as in (i) and let

$$\delta = \min_{x, y \in V} \sum_{z \in V} P^r(x, z) P^r(y, z) > 0. \quad (107)$$

Use (i) and Markov property that in every r steps, X_t and Y_t meet with probability $\geq \delta > 0$.

(iii) Let τ be the first time that X_t and Y_t meet. From (ii) and Markov property, deduce that

$$\mathbb{P}(\tau \geq kr) = \mathbb{P}(\tau \geq r) \mathbb{P}(X_t \text{ and } Y_t \text{ never meet during } [r, kr) \mid X_{r-1} \neq Y_{r-1}) \quad (108)$$

$$\leq (1 - \delta) \mathbb{P}(X_t \text{ and } Y_t \text{ never meet during } [r, kr) \mid X_{r-1} \neq Y_{r-1}). \quad (109)$$

By an induction on k , conclude that

$$\mathbb{P}(\tau \geq kr) \leq (1 - \delta)^k \rightarrow 0 \quad \text{as } k \rightarrow \infty. \quad (110)$$

The general convergence theorem is stated below.

Theorem 5.12 (Convergence Thm). *Let $(X_t)_{t \geq 0}$ be an irreducible aperiodic Markov chain on a finite state space Ω . Let π denote the unique stationary distribution of X_t . Then for each $x, y \in V$,*

$$\lim_{t \rightarrow \infty} \mathbb{P}(X_t = y \mid X_0 = x) = \pi(y). \quad (111)$$

PROOF. As always, there is a linear algebra proof to this result (see, e.g., [LP17, Thm. 4.9]) Mimic the argument for Theorem 5.8 for a coupling based proof. $\square$

Exercise 5.13 (Convergence of empirical distribution). Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain on a finite state space Ω . Fix a state $x \in \Omega$, and let T_k be the k th return time to x .

(i) Let $\tau_k = T_k - T_{k-1}$ for $k \geq 2$. Show that τ_k 's are i.i.d. and $0 < \mathbb{E}[\tau_k] < \infty$. (See Exercise 4.6.)

(ii) Use Markov property and strong law of large numbers to show

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{\tau_2 + \cdots + \tau_n}{n} = \mathbb{E}[\tau_2]\right) = 1. \quad (112)$$

(iii) Noting that $T_n = T_1 + \tau_2 + \cdots + \tau_n$, show that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{T_n}{n} = \mathbb{E}[\tau_2]\right) = 1. \quad (113)$$

(vi) Let $V_x(t) = V(t)$ be the number of visits to x up to time t . Using the fact that $\mathbb{E}[\tau_k] < \infty$, show that $V(t) \rightarrow \infty$ as $t \rightarrow \infty$ a.s. Also, noting that $T_{V(t)} \leq t < T_{V(t)+1}$, use (iii) and (iv) to deduce

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{V_x(t)}{t} = \frac{1}{\mathbb{E}[\tau_2]}\right) = 1. \quad (114)$$

(v) Suppose $(X_t)_{t \geq 0}$ is also aperiodic and let π denote the unique stationary distribution. Use Theorem 3.1 to deduce

$$\pi(x) = \lim_{t \rightarrow \infty} \frac{\mathbb{E}[V_x(t)]}{t} = \mathbb{E}\left[\lim_{t \rightarrow \infty} \frac{V_x(t)}{t}\right] = \frac{1}{\mathbb{E}[\tau_2]}. \quad (115)$$

Conclude that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{V_x(t)}{t} = \pi(x)\right) = 1. \quad (116)$$

6. Markov chain Monte Carlo

So far, we were given a Markov chain $(X_t)_{t \geq 0}$ on a finite state space Ω and studied existence and uniqueness of its stationary distribution and convergence to it. In this section, we will consider the reverse problem. Namely, given a distribution π on a sample space Ω , can we construct a Markov chain $(X_t)_{t \geq 0}$ such that π is a stationary distribution? If in addition the chain is irreducible and aperiodic, then by the convergence theorem (Theorem 5.12), we know that the distribution π_t of X_t converges to π . Hence if we run the chain for long enough, the state of the chain is asymptotically distributed as π . In other words, we can sample a random element of Ω according to the prescribed distribution π by emulating it through a suitable Markov chain. This method of sampling is called *Markov chain Monte Carlo* (MCMC).

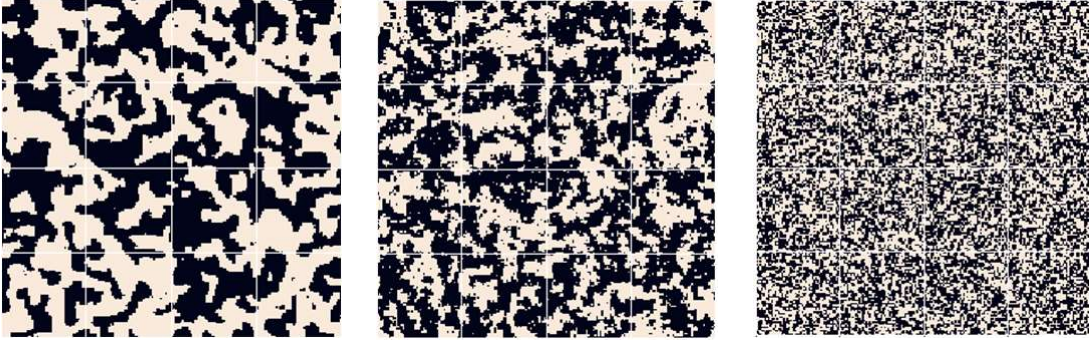


FIGURE 7. MCMC simulation of Ising model on 200 by 200 torus at temperature $T = 1$ (left), 2 (middle), and 5 (right).

Example 6.1 (Uniform distribution on regular graphs). Let $G = (V, E)$ be a connected *regular* graph, meaning that all nodes have the same degree. Let μ be the uniform distribution on the node set V . How can we sample a random node according to μ ? If we have a list of all nodes, then we can label them from 1 to $N = |V|$, choose a random number between 1 and N , and find corresponding node. But often times, we do not have the full list of nodes, especially when we want to sample a random node from a social network. Given only local information (set of neighbors for each given node), can we still sample a uniform random node from G ?

One answer is given by random walk. Indeed, random walks on graphs are defined by only using local information of the underlying graph: Choose a random neighbor and move there. Moreover, since G is connected, there is a unique stationary distribution π for the walk, which is given by

$$\pi(x) = \frac{\deg_G(x)}{2|E|}. \quad (117)$$

Since G is regular, any two nodes have the same degree, so $\pi(x) = \pi(y)$ for all $x, y \in V$. This means π equals the uniform distribution μ on V . Hence the sampling algorithm we propose is as follows:

(*) *Run a random walk on G for $t \gg 1$ steps, and take the random node that the walk sits on.*

However, there is a possible issue of convergence. Namely, if the graph G does not contain any odd cycle, then random walk on G is periodic (see Exercise 5.6), so we are not guaranteed to have convergence. We can overcome this by using a lazy random walk instead, which is introduced in Exercise 6.2. We know that the lazy RW on G is irreducible, aperiodic, and has the same set of stationary distribution as the ordinary RW on G . Hence we can apply the sampling algorithm (*) above for lazy random walk on G to sample a uniform random node in G . ▲

Exercise 6.2 (Lazy RW on graphs). Let $G = (V, E)$ be a graph. Let $(X_t)_{t \geq 0}$ be a Markov chain on the node set V with transition probabilities

$$\mathbb{P}(X_{t+1} = j | X_t = i) = \begin{cases} 1/2 & \text{if } j = i \\ 1/(2 \deg_G(i)) & \text{if } j \text{ is adjacent to } i \\ 0 & \text{otherwise.} \end{cases} \quad (118)$$

This chain is called the *lazy random walk* on G . In words, the usual random walker on G now flips a fair coin to decide whether it stays at the same node or make a move to one of its neighbors.

- (i) Show that for any connected graph G , the lazy random walk on G is irreducible and aperiodic.
- (ii) Let P be the transition matrix for the usual random walk on G . Show that the following matrix

$$Q = \frac{1}{2}(P + I) \quad (119)$$

is the transition matrix for the lazy random walk on G .

- (iii) For any distribution π on V , show that $\pi Q = \pi$ if and only if $\pi P = \pi$. Conclude that the usual and lazy random walks on G have the same set of stationary distribution.

Example 6.3 (Finding local minima). Let $G = (V, E)$ be a connected graph and let $f : V \rightarrow [0, \infty)$ be a ‘cost’ function. The objective is to find a node $x^* \in V$ such that f takes global minimum at x^* . This problem has a lot of application in machine learning, for example. Note that if the domain V is very large, then an exhaustive search is too expensive to use.

Here is simple form of the popular algorithm of *stochastic gradient descent*, which lies at the heart of most of the important machine learning algorithms.

- (i) Initialize the first guess $X_0 = x_0 \in V$.
- (ii) Suppose $X_t = x \in V$ is chosen. Let

$$D_t = \{y \text{ a neighbor of } x \text{ or } x \text{ itself} \mid f(y) \leq f(x)\}. \quad (120)$$

Define X_{t+1} to be a uniform random node from D_t .

- (iii) The algorithm terminates if it finds a local minima.

In words, at each step we move to a random neighbor which could possibly decrease the current value of f . It is easy to see that one always converges to a local minima, which may not be a global minimum. In a very complex machine learning task (e.g., training a deep neural network), this is often good enough. Is this a Markov chain? Irreducible? Aperiodic? Stationary distribution? $\blacktriangle$

There is a neat solution to finding global minimum. The idea is to allow that we go uphill with a small probability.

Example 6.4 (Finding global minimum). Let $G = (V, E)$ be a connected regular graph and let $f : V \rightarrow [0, \infty)$ be a cost function. Let

$$V^* = \left\{x \in V \mid f(x) = \min_{y \in V} f(y)\right\} \quad (121)$$

be the set of all nodes where f attains global minimum.

Fix a parameter $\lambda \in (0, 1]$, and define a probability distribution π_λ on V by

$$\pi_\lambda(x) = \frac{\lambda^{f(x)}}{Z_\lambda}, \quad (122)$$

where $Z_\lambda = \sum_{x \in V} \lambda^{f(x)}$ is the normalizing constant. Since $\pi_\lambda(x)$ is decreasing in $f(x)$, it favors nodes x for which $f(x)$ is small.

Let $(X_t)_{t \geq 0}$ be Markov chain on V , whose transition is defined as follows. If $X_t = x$, then let y be a uniform random neighbor of x . If $f(y) \leq f(x)$, then move to y ; If $f(y) > f(x)$, then move to y with probability $\lambda^{f(y)-f(x)}$ and stay at x with probability $1 - \lambda^{f(y)-f(x)}$. We analyze this MC below:

- (i) (Irreducibility) Since G is connected and we are allowing any move (either downhill or uphill) we can go from one node to any other in some number of steps. Hence the chain $(X_t)_{t \geq 0}$ is irreducible.
- (ii) (Aperiodicity) By (i) and Remark 5.7, all nodes have the same period. Moreover, let $x \in V^*$ be an arbitrary node where f takes global minimum. Then all return times are possible, so x has period 1. Hence all nodes have period 1, so the chain is aperiodic.
- (iii) (Stationarity) Here we show that π_λ is a stationary distribution of the chain. To do this, we first need to write down the transition matrix P . Namely, if we let $A_G(y, z)$ denote the indicator that y and z are adjacent, then

$$P(x, y) = \begin{cases} \frac{A_G(x, y)}{\deg_G(x)} \min(1, \lambda^{f(y)-f(x)}) & \text{if } x \neq y \\ 1 - \sum_{y \neq x} P(x, y) & \text{if } y = x. \end{cases} \quad (123)$$

To show $\pi_\lambda P = \pi_\lambda$, it suffices to show for any $y \in V$ that

$$\sum_{z \in V} \pi_\lambda(z) P(z, y) = \pi_\lambda(y). \quad (124)$$

Note that

$$\sum_{z \in V} \pi_\lambda(z) P(z, y) = \pi_\lambda(y) P(y, y) + \sum_{z \neq y} \pi_\lambda(z) P(z, y) \quad (125)$$

$$= \pi_\lambda(y) - \sum_{z \neq y} \pi_\lambda(y) P(y, z) + \sum_{z \neq y} \pi_\lambda(z) P(z, y). \quad (126)$$

Hence it suffices to show that

$$\pi_\lambda(y) P(y, z) = \pi_\lambda(z) P(z, y) \quad (127)$$

for any $z \neq y$. Indeed, considering the two cases $f(z) \leq f(y)$ and $f(z) > f(y)$, we have

$$\pi_\lambda(y) P(y, z) = \frac{\lambda^{f(y)}}{Z_\lambda} \frac{A_G(y, z)}{\deg_G(y)} \min(1, \lambda^{f(z)-f(y)}) = \frac{1}{Z_\lambda} \frac{A_G(z, y)}{\deg_G(z)} \lambda^{\max(f(y), f(z))}, \quad (128)$$

$$\pi_\lambda(z) P(z, y) = \frac{\lambda^{f(z)}}{Z_\lambda} \frac{A_G(z, y)}{\deg_G(z)} \min(1, \lambda^{f(y)-f(z)}) = \frac{1}{Z_\lambda} \frac{A_G(y, z)}{\deg_G(y)} \lambda^{\max(f(y), f(z))}. \quad (129)$$

Now since $A_G(z, y) = A_G(y, z)$ and we are assuming G is a regular graph, this yields (127), as desired. Hence π_λ is a stationary distribution for the chain $(X_t)_{t \geq 0}$.

- (iv) (Convergence) By (i), (iii), Theorems 4.5, we see that π_λ is the unique stationary distribution for the chain X_t . Furthermore, by (i)-(iii) and Theorem 5.12, we conclude that the distribution of X_t converges to π_λ .
- (v) (Global minimum) Let $f_* = \min_{x \in V} f(x)$ be the global minimum of f . Note that by definition of V^* , we have $f(x) = f_*$ for any $x \in V^*$. Then observe that

$$\lim_{\lambda \rightarrow 0} \pi_\lambda(x) = \lim_{\lambda \rightarrow 0} \frac{\lambda^{f(x)}}{\sum_{y \in V} \lambda^{f(y)}} = \lim_{\lambda \rightarrow 0} \frac{\lambda^{f(x)} / \lambda^{f_*}}{\sum_{y \in V} \lambda^{f(y)} / \lambda^{f_*}} \quad (130)$$

$$= \lim_{\lambda \rightarrow 0} \frac{\lambda^{f(x)-f_*}}{|V^*| + \sum_{y \in V \setminus V^*} \lambda^{f(y)-f_*}} = \frac{\mathbf{1}(x \in V^*)}{|V^*|}. \quad (131)$$

Thus for λ very small, π_λ is approximately the uniform distribution on the set of all nodes V^* where f attains global minimum. $\blacktriangle$

Exercise 6.5 (Detailed Balance equation). Let P be a transition matrix of a Markov chain on state space Ω . Suppose π is a probability distribution on Ω that satisfies the following *detailed balance equation*

$$\pi(x) P(x, y) = \pi(y) P(y, x) \quad \text{for all } x, y \in \Omega. \quad (132)$$

- (i) Show that for all $x \in \Omega$,

$$\sum_{y \in \Omega} \pi(y) P(y, x) = \sum_{y \in \Omega} \pi(x) P(x, y) = \pi(x) \sum_{y \in \Omega} P(x, y) = \pi(x). \quad (133)$$

- (ii) Conclude that π is a stationary distribution for P , that is, $\pi P = \pi$.

Exercise 6.6 (Metropolis-Hastings algorithm). Let P be a transition matrix of a Markov chain on state space $\Omega = \{1, 2, \dots, m\}$. Let π be a probability distribution on Ω , which is not necessarily a stationary distribution for P . Our goal is to design a Markov chain on Ω that has π as a stationary distribution. Below we will derive the famous *Metropolis-Hastings algorithm* for MCMC sampling.

Fix a $m \times m$ matrix A of entries from $[0, 1]$. Consider a Markov chain $(X_t)_{t \geq 0}$ on Ω that uses additional rejection step on top of the transition matrix P as follows:

(Generation) Suppose the current location $X_t = a$. Generate a candidate $b \in \Omega$ from the conditional distribution $P(a, \cdot)$.

(Rejection) Flip an independent coin with success probability $A(a, b)$. Upon success, accept the proposed move and set $X_{t+1} = b$; Otherwise, reject the move and set $X_{t+1} = a$.

Here, the (a, b) entry $A(a, b)$ is called the *acceptance probability* of the move $a \rightarrow b$.

(i) Let Q denote the transition matrix of the chain $(X_t)_{t \geq 0}$ defined above. Show that

$$Q(a, b) = \begin{cases} P(a, b)A(a, b) & \text{if } b \neq a \\ 1 - \sum_{c: c \neq a} P(a, c)A(a, c) & \text{if } b = a. \end{cases} \quad (134)$$

(ii) Show that $\pi(x)Q(x, y) = \pi(y)Q(y, x) \forall x, y \in \Omega, x \neq y$ implies $\pi Q = \pi$. Deduce that if

$$\pi(x)P(x, y)A(x, y) = \pi(y)P(y, x)A(y, x) \quad \forall x, y \in \Omega, x \neq y, \quad (135)$$

then π is a stationary distribution for $(X_t)_{t \geq 0}$.

(iii) Since we also want the Markov chain to converge fast, we want to choose the acceptance probability $A(a, b) \in [0, 1]$ as large as possible for each $a, b \in \Omega$. Show that the following choice (denoting $\alpha \wedge \beta := \min(\alpha, \beta)$)

$$A(x, y) = \frac{\pi(y)P(y, x)}{\pi(x)P(x, y)} \wedge 1 \quad \forall x, y \in \Omega, x \neq y \quad (136)$$

satisfies the condition in (ii) and each $A(x, y)$ is maximized for all $x \neq y$.

(iv) Let $(Y_t)_{t \geq 0}$ be a random walk on the 5-wheel graph $G = (V, E)$ as shown in Figure 8. Show that $\pi =$

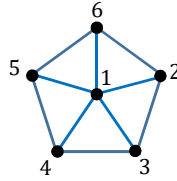


FIGURE 8. 5-wheel graph G

$\left[\frac{5}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}, \frac{3}{20}\right]$ is the unique stationary distribution of Y_t . Apply the Metropolis-Hastings algorithm derived in (i)-(iii) above to modify Y_t to obtain a new Markov chain X_t on V that converges to $\text{Uniform}(V)$ in distribution.

Poisson Processes

1. Recap of exponential and Poisson random variables

In this section, we study two important properties of exponential and Poisson random variables, which will be crucial when we study Poisson processes from the following section.

Example 1.1 (Exponential RV). X is an *exponential* RV with rate λ (denoted by $X \sim \text{Exp}(\lambda)$) if it has PDF

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}(x \geq 0). \quad (137)$$

Integrating the PDF gives its CDF

$$\mathbb{P}(X \leq x) = (1 - e^{-\lambda x}) \mathbf{1}(x \geq 0). \quad (138)$$

The following complimentary CDF for exponential RVs will be useful:

$$\mathbb{P}(X \geq x) = e^{-\lambda x} \mathbf{1}(x \geq 0). \quad (139)$$

▲

Exercise 1.2. Let $X \sim \text{Exp}(\lambda)$. Show that $\mathbb{E}(X) = 1/\lambda$ and $\text{Var}(X) = 1/\lambda^2$.

Minimum of independent exponential RVs is again an exponential RV.

Exercise 1.3. Let $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$ and suppose they are independent. Define $Y = \min(X_1, X_2)$. Show that $Y \sim \text{Exp}(\lambda_1 + \lambda_2)$. (Hint: Compute $\mathbb{P}(Y \geq y)$.)

The following example is sometimes called the ‘competing exponentials’.

Example 1.4. Let $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$ be independent exponential RVs. We will show that

$$\mathbb{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2} \quad (140)$$

using the iterated expectation. Using iterated expectation for probability,

$$\mathbb{P}(X_1 < X_2) = \int_0^\infty \mathbb{P}(X_1 < X_2 | X_1 = x_1) \lambda_1 e^{-\lambda_1 x_1} dx_1 \quad (141)$$

$$= \int_0^\infty \mathbb{P}(X_2 > x_1) \lambda_1 e^{-\lambda_1 x_1} dx_1 \quad (142)$$

$$= \lambda_1 \int_0^\infty e^{-\lambda_2 x_1} e^{-\lambda_1 x_1} dx_1 \quad (143)$$

$$= \lambda_1 \int_0^\infty e^{-(\lambda_1 + \lambda_2)x_1} dx_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \quad (144)$$

▲

When we add up independent continuous RVs, we can compute the PDF of the resulting RV by using the convolution formula or moment generating functions. In the following exercise, we compute the PDF of the sum of independent exponentials.

Exercise 1.5 (Sum of i.i.d. Exp is Erlang). Let $X_1, X_2, \dots, X_n \sim \text{Exp}(\lambda)$ be independent exponential RVs.

(i) Show that $f_{X_1 + X_2}(z) = \lambda^2 z e^{-\lambda z} \mathbf{1}(z \geq 0)$.

(ii) Show that $f_{X_1+X_2+X_3}(z) = 2^{-1} \lambda^3 z^2 e^{-\lambda z} \mathbf{1}(z \geq 0)$.

(iii) Let $S_n = X_1 + X_2 + \cdots + X_n$. Use induction to show that $S_n \sim \text{Erlang}(n, \lambda)$, that is,

$$f_{S_n}(z) = \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!}. \quad (145)$$

Exponential RVs will be the building block of the Poisson processes, because of their ‘memoryless property’.

Exercise 1.6 (Memoryless property of exponential RV). A continuous positive RV X is said to have *memoryless property* if

$$\mathbb{P}(X \geq t_1 + t_2) = \mathbb{P}(X \geq t_1) \mathbb{P}(X \geq t_2) \quad \forall x_1, x_2 \geq 0. \quad (146)$$

(i) Show that (146) is equivalent to

$$\mathbb{P}(X \geq t_1 + t_2 | X \geq t_2) = \mathbb{P}(X \geq t_1) \quad \forall x_1, x_2 \geq 0. \quad (147)$$

(ii) Show that exponential RVs have memoryless property.

(iii) Suppose X is continuous, positive, and memoryless. Let $g(t) = \log \mathbb{P}(X \geq t)$. Show that g is continuous at 0 and

$$g(x+y) = g(x) + g(y) \quad \text{for all } x, y \geq 0. \quad (148)$$

Using Exercise 1.7, conclude that X must be an exponential RV.

Exercise 1.7. Let $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$ be a function with the property that $g(x+y) = g(x) + g(y)$ for all $x, y \geq 0$. Further assume that g is continuous at 0. In this exercise, we will show that $g(x) = cx$ for some constant c .

(i) Show that $g(0) = g(0+0) = g(0) + g(0)$. Deduce that $g(0) = 0$.

(ii) Show that for all integers $n \geq 1$, $g(n) = ng(1)$.

(iii) Show that for all integers $n, m \geq 1$,

$$ng(1) = g(n \cdot 1) = g(m(n/m)) = mg(n/m). \quad (149)$$

Deduce that for all nonnegative rational numbers r , we have $g(r) = rg(1)$.

(iv) Show that g is continuous.

(v) Let x be nonnegative real number. Let r_k be a sequence of rational numbers such that $r_k \rightarrow x$ as $k \rightarrow \infty$. By using (iii) and (iv), show that

$$g(x) = g\left(\lim_{k \rightarrow \infty} r_k\right) = \lim_{k \rightarrow \infty} g(r_k) = g(1) \lim_{k \rightarrow \infty} r_k = x \cdot g(1). \quad (150)$$

Lastly, we introduce the Poisson RVs and record some of their nice properties.

Example 1.8 (Poisson RV). A RV X is a *Poisson RV* with rate $\lambda > 0$ if

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (151)$$

for all nonnegative integers $k \geq 0$. We write $X \sim \text{Poisson}(\lambda)$.

Exercise 1.9. Let $X \sim \text{Poisson}(\lambda)$. Show that $\mathbb{E}(X) = \text{Var}(X) = \lambda$.

Exercise 1.10 (Sum of ind. Poisson RVs is Poisson). Let $X \sim \text{Poisson}(\lambda_1)$ and $Y \sim \text{Poisson}(\lambda_2)$ be independent Poisson RVs. Show that $X + Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

2. Poisson processes as an arrival process

An *arrival* process is a sequence of strictly increasing RVs $0 < T_1 < T_2 < \dots$. For each integer $k \geq 1$, its k th *inter-arrival time* is defined by $\tau_k = T_k - T_{k-1} \mathbf{1}(k \geq 2)$. For a given arrival process $(T_k)_{k \geq 1}$, the associated *counting process* $(N(t))_{t \geq 0}$ is defined by

$$N(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t) = \#(\text{arrivals up to time } t). \quad (152)$$

Note that these three processes (arrival times, inter-arrival times, and counting) determine each other:

$$(T_k)_{k \geq 1} \iff (\tau_k)_{k \geq 1} \iff (N(t))_{t \geq 0}. \quad (153)$$

Exercise 2.1. 6.12 Let $(T_k)_{k \geq 1}$ be any arrival process and let $(N(t))_{t \geq 0}$ be its associated counting process. Show that these two processes determine each other by the following relation

$$\{T_n \leq t\} = \{N(t) \geq n\}. \quad (154)$$

In words, n th customer arrives by time t if and only if at least n customers arrive up to time t .

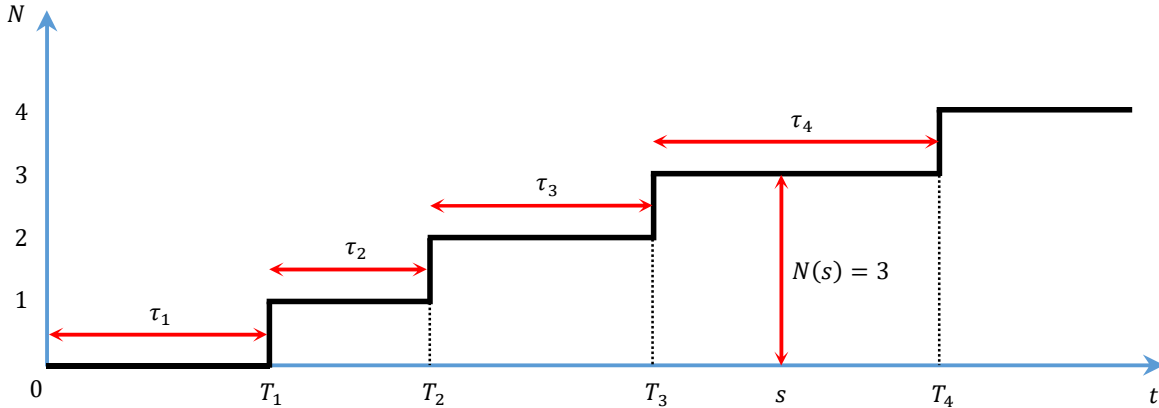


FIGURE 1. Illustration of a continuous-time arrival process $(T_k)_{k \geq 1}$ and its associated counting process $(N(t))_{t \geq 0}$. τ_k 's denote inter-arrival times. $N(t) \equiv 3$ for $T_3 < t \leq T_4$.

Now we define Poisson process.

Definition 2.2 (Poisson process). An arrival process $(T_k)_{k \geq 1}$ is a *Poisson process of rate λ* if its inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ RVs. In this case, we denote $(T_k)_{k \geq 1} \sim \text{PP}(\lambda)$ and equivalently $(N(t))_{t \geq 0} \sim \text{PP}(\lambda)$ for its associated counting process $(N(t))_{t \geq 0}$.

The choice of exponential inter-arrival times is special due to the memoryless property of exponential RVs (Exercise 1.6).

Exercise 2.3. Let $(T_k)_{k \geq 1}$ be a Poisson process with rate λ . Show that $\mathbb{E}[T_k] = k/\lambda$ and $\text{Var}(T_k) = k/\lambda^2$. Furthermore, show that $T_k \sim \text{Erlang}(k, \lambda)$, that is,

$$f_{T_k}(z) = \frac{\lambda^k z^{k-1} e^{-\lambda z}}{(k-1)!}. \quad (155)$$

The following exercise explains what is ‘Poisson’ about the Poisson process.

Exercise 2.4. Let $(T_k)_{k \geq 1}$ be a Poisson process with rate λ and let $(N(t))_{t \geq 0}$ be the associated counting process. We will show that $N(t) \sim \text{Poisson}(\lambda t)$.

(i) Using the relation $\{T_n \leq t\} = \{N(t) \geq n\}$ and Exercise 2.3, show that

$$\mathbb{P}(N(t) \geq n) = \mathbb{P}(T_n \leq t) = \int_0^t \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!} dz. \quad (156)$$

(ii) Let $G(t) = \sum_{m=n}^{\infty} (\lambda t)^m e^{-\lambda t} / m! = \mathbb{P}(\text{Poisson}(\lambda t) \geq n)$. Show that

$$\frac{d}{dt} G(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} = \frac{d}{dt} \mathbb{P}(T_n \leq t). \quad (157)$$

Conclude that $G(t) = \mathbb{P}(T_n \leq t)$.

(iii) From (i) and (ii), conclude that $N(t) \sim \text{Poisson}(\lambda t)$.

3. Memoryless property and stopping times

Let $(N(t))_{t \geq 0} \sim \text{PP}(\lambda)$. For any fixed $u \geq 0$, $(N(t+u) - N(t))_{t \geq 0}$ is a counting process itself, which counts the number of arrivals during the interval $[u, u+t]$. We call this the counting process $(N(t))_{t \geq 0}$ *restarted at time u* . One of the remarkable properties of a Poisson process is that their counting process restarted at any given time u is again the counting process of a Poisson process, which is independent of what have happened up to time u . This is called the memoryless property of Poisson processes.

Memoryless property of Poisson processes not only applies for a deterministic restart times, but also some class of random times called ‘stopping times’. We introduce this notion in a greater generality.

Definition 3.1 (Stopping time). Let $\mathcal{F}_t$ be the collection of all events that one could observe by time t ¹. The sequence $(\mathcal{F}_t)_{t \geq 0}$ is called a *filtration*². A random variable $T \geq 0$ is a *stopping time* w.r.t. a given filtration $(\mathcal{F}_t)_{t \geq 0}$ if for each $u \geq 0$,

$$\{T \leq u\} \in \mathcal{F}_u \quad \text{or} \quad \{T > u\} \in \mathcal{F}_u. \quad (158)$$

The condition above is interpreted as follows: By using all information we gathered up to time u , we can determine whether $T \leq u$ or $T > u$, not using any future information that is yet to be observed.

Theorem 3.2 (Memoryless property of PP). *Let $(N(t))_{t \geq 0} \sim \text{PP}(\lambda)$ and let $T \geq 0$ be a stopping time w.r.t. a filtration $(\mathcal{F}_t)_{t \geq 0}$ such that $\mathcal{F}_t$ contains all information of $(N(u))_{0 \leq u \leq t}$. Then the following holds.*

- (i) $(N(T+t) - N(T))_{t \geq 0} \sim \text{PP}(\lambda)$.
- (ii) $(N(t))_{0 \leq t \leq T}$ and $(N(T+t) - N(T))_{t \geq 0}$ are independent.

The following is a very useful observation, which allows us to restart a Poisson process at any stopping time and still get an independent Poisson process.

Lemma 3.3. *Let $(X(t))_{t \geq 0}$ be any stochastic processes and let T be a stopping time for $(X(t))_{t \geq 0}$. Suppose the followings hold:*

- (i) (Independent increments) *For any $u \geq 0$, the processes $(X(t))_{0 \leq t \leq u}$ and $(X(t) - X(u))_{t > u}$ are independent.*
- (ii) (Stationary increments) *The distribution of the process $(X(t) - X(u))_{t > u}$ does not depend on u (stationary increment).*

Then $(X(t) - X(T))_{t > T}$ has the same distribution as $(X(t) - X(0))_{t > 0}$ and is independent of $(X(t))_{t \leq T}$.

PROOF. Let E, E' be any events for real stochastic processes. For any stochastic process $W = (W(t))_{t \geq 0}$, we write $W \in E$ if the event E occurs for the process W . We want to show

$$\mathbb{P}((X(t) - X(T))_{t > T} \in E \mid (X(t))_{0 \leq t \leq T}) = \mathbb{P}((X(t) - X(T))_{t > T} \in E). \quad (159)$$

¹Proper name: σ -field generated by information up to time t

²Since one can only gain information, we have $\mathcal{F}_a \subseteq \mathcal{F}_b$ for each $0 \leq a \leq b$.

We show the above equation under the conditioning on $\{T = u\}$, and then undo the conditioning by iterated expectation.

Since T is a stopping time for $(X(t))_{t \geq 0}$, for any $u \geq 0$, the event

$$\{(X(t))_{0 \leq t \leq T} \in E', T = u\} = \{(X(t))_{0 \leq t \leq u} \in E', T = u\} \quad (160)$$

is independent of the process $(X(t) - X(u))_{t > u}$ on the time interval (u, ∞) . Hence by the independent increments assumption (i), the above event is independent from the event $\{(X(t) - X(u))_{t > u} \in E\}$. This gives that, for any $u \geq 0$,

$$\mathbb{P}\left((X(t) - X(T))_{t > T} \in E \mid (X(t))_{0 \leq t \leq T} \in E', T = u\right) \quad (161)$$

$$= \mathbb{P}\left((X(t) - X(u))_{t > u} \in E \mid (X(t))_{0 \leq t \leq u} \in E', T = u\right) \quad (162)$$

$$= \mathbb{P}\left((X(t) - X(u))_{t > u} \in E\right) \quad (163)$$

$$= \mathbb{P}\left((X(t) - X(0))_{t > 0} \in E\right), \quad (164)$$

where for the last equality, we have used the stationary increments assumption (ii).

Hence by iterated expectation,

$$\mathbb{P}\left((X(t) - X(T))_{t > T} \in E \mid (X(t))_{0 \leq t \leq T} \in E'\right) \quad (165)$$

$$= \mathbb{E}\left[\mathbb{P}\left((X(t) - X(T))_{t > T} \in E \mid (X(t))_{0 \leq t \leq T} \in E', T\right) \mid T\right] \quad (166)$$

$$= \mathbb{E}\left[\mathbb{P}\left((X(t) - X(0))_{t > 0} \in E\right) \mid T\right] \quad (167)$$

$$= \mathbb{P}\left((X(t) - X(0))_{t > 0} \in E\right). \quad (168)$$

To finish the proof, let E' be the entire sample space. Then the above yields

$$\mathbb{P}\left((X(t) - X(T))_{t > T} \in E\right) = \mathbb{P}\left((X(t) - X(0))_{t > 0} \in E\right). \quad (169)$$

This shows that $(X(t) - X(T))_{t > T}$ has the same distribution as $(X(t) - X(0))_{t > 0}$. Furthermore, we conclude

$$\mathbb{P}\left((X(t) - X(T))_{t > T} \in E \mid (X(t))_{0 \leq t \leq T} \in E'\right) = \mathbb{P}\left((X(t) - X(0))_{t > 0} \in E\right) \quad (170)$$

$$= \mathbb{P}\left((X(t) - X(T))_{t > T} \in E\right). \quad (171)$$

Since E, E' were arbitrary, this also shows that $(X(t) - X(T))_{t > T}$ and $(X(t))_{0 \leq t \leq T}$ are independent. $\square$

According to Lemma 3.3, Theorem 3.2 follows once we show that the counting process $(N(t))_{t \geq 0}$ of $\text{PP}(\lambda)$ satisfies the independent and stationary increment properties stated in Lemma 3.3. This is verified by the following proposition.

Proposition 3.4 (PP has independent and stationary increments). *Let $(T_k)_{k \geq 1}$ be a Poisson process of rate λ and let $(N(t))_{t \geq 0}$ be the associated counting process.*

- (i) *For any $t \geq 0$, let $Z(t) = \inf\{s > 0 : N(t+s) > N(t)\}$ be the waiting time for the first arrival after time t . Then $Z(t) \sim \text{Exp}(\lambda)$ and it is independent of the process up to time t .*
- (ii) *For any $s \geq 0$, $(N(t+s) - N(t))_{s \geq 0}$ is the counting process of a Poisson process of rate λ , which is independent of the process $(N(u))_{t \leq u}$.*

PROOF. We first show (ii). Note that

$$T_{N(t)} \leq t < T_{N(t)+1}. \quad (172)$$

Hence we may write

$$Z(t) = T_{N(t)+1} - t = \tau_{N(t)+1} - (t - T_{N(t)}). \quad (173)$$

Namely, $Z(t)$ is the remaining portion of the $N(t) + 1$ st inter-arrival time $\tau_{N(t)+1}$ after we waste the first $t - T_{N(t)}$ of it. (See Figure 4).

Now consider restarting the arrival process $(T_k)_{k \geq 0}$ at time t . The first inter-arrival time is $T_{N(t)+1} - t = Z(t)$, which is $\text{Exp}(\lambda)$ and independent from the past by (i). The second inter-arrival time is $T_{N(t)+2} - T_{N(t)+1}$, which is $\text{Exp}(\lambda)$ and is independent of everything else by assumption. Likewise, the following inter-arrival times for this restarted arrival process are i.i.d. $\text{Exp}(\lambda)$ variables. This shows (ii).

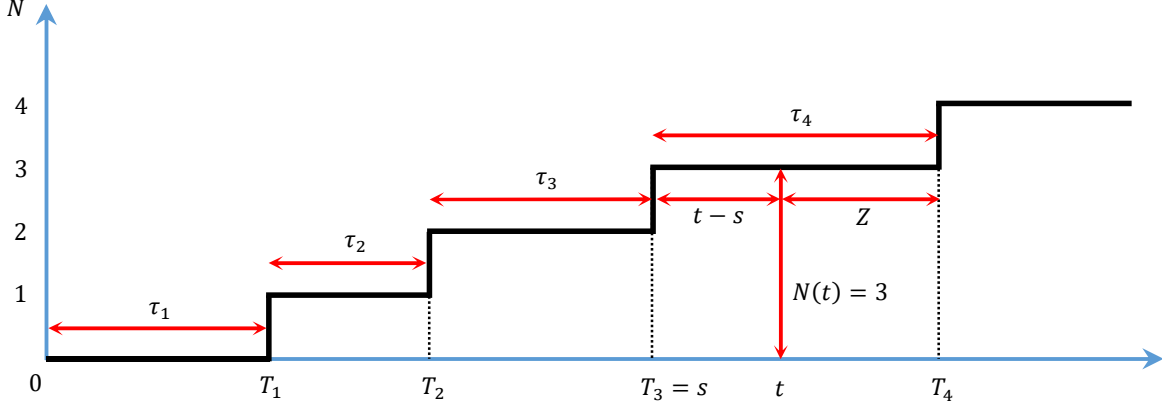


FIGURE 2. Assuming $N(t) = 3$ and $T_3 = s \leq t$, we have $Z = \tau_4 - (t - s)$. By memoryless property of exponential RV, Z follows $\text{Exp}(\lambda)$ on this conditioning.

Next, we show (i). Let E be any event for the counting process $(N(s))_{0 \leq s \leq t}$ up to time t . In order to show the remaining waiting time $Z(t)$ and the past process up to time t are independent and $Z(t) \sim \text{Exp}(\lambda)$, we want to show that

$$\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E) = \mathbb{P}(Z(t) \geq x) = e^{-\lambda x}. \quad (174)$$

for any $x \geq 0$. To this end,

As can be seen from (3), $Z(t)$ depends on three random variables: $\tau_{N(t)+1}$, $N(t)$, and $T_{N(t)}$. To show, we argue by conditioning the last two RVs and use iterated expectation. Using 3, note that

$$\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, N(t) = n, T_{N(t)} = s) \quad (175)$$

$$= \mathbb{P}(\tau_{n+1} - (t - s) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, N(t) = n, T_n = s) \quad (176)$$

$$= \mathbb{P}(\tau_{n+1} - (t - s) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, T_{n+1} > t, T_n = s) \quad (177)$$

$$= \mathbb{P}(\tau_{n+1} - (t - s) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, \tau_{n+1} > t - s, T_n = s). \quad (178)$$

Conditioned on $N(t) = n$, the event that $(N(s))_{0 \leq s \leq t} \in E$ is determined by the arrival times $T_1, \dots, T_n$ and the fact that $T_{n+1} \geq t$. Hence we can rewrite

$$\{(N(s))_{0 \leq s \leq t} \in E, \tau_{n+1} > t - s, T_n = s\} = \{(\tau_1, \dots, \tau_n) \in E', \tau_{n+1} > t - s\} \quad (179)$$

for some event E' to be satisfied by the first n inter-arrival times. Since inter-arrival times are independent, this gives

$$\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, N(t) = n, T_{N(t)} = s) \quad (180)$$

$$= \mathbb{P}(\tau_{n+1} - (t - s) \geq x \mid \tau_{n+1} \geq t - s) \quad (181)$$

$$= \mathbb{P}(\tau_{n+1} \geq x) = e^{-\lambda x}, \quad (182)$$

where we have used the memoryless property of exponential variables. Hence by iterated expectation,

$$\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, N(t) = n) = \mathbb{E}_{T_{N(t)}} \left[\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, N(t) = n, T_{N(t)} = s) \right] \quad (183)$$

$$= \mathbb{E}_{N(t)}[e^{-\lambda x}] = e^{-\lambda x}. \quad (184)$$

By using iterated expectation once more,

$$\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E) = \mathbb{E}_{N(t)} \left[\mathbb{P}(Z(t) \geq x \mid (N(s))_{0 \leq s \leq t} \in E, N(t)) \right] \quad (185)$$

$$= \mathbb{E}_{N(t)}[e^{-\lambda x}] = e^{-\lambda x}. \quad (186)$$

By taking E to be the entire sample space, this also gives

$$\mathbb{P}(Z(t) \geq x) = e^{-\lambda x}. \quad (187)$$

This shows (174). $\square$

Exercise 3.5 (Sum of independent Poisson RV's is Poisson). Let $(T_k)_{k \geq 1}$ be a Poisson process with rate λ and let $(N(t))_{t \geq 0}$ be the associated counting process. Fix $t, s \geq 0$.

- (i) Use memoryless property to show that $N(t)$ and $N(t+s) - N(t)$ are independent Poisson RVs of rates λt and λs .
- (ii) Note that the total number of arrivals during $[0, t+s]$ can be divided into the number of arrivals during $[0, t]$ and $[t, t+s]$. Conclude that if $X \sim \text{Poisson}(\lambda t)$ and $Y \sim \text{Poisson}(\lambda s)$ and if they are independent, then $X + Y \in \text{Poisson}(\lambda(t+s))$.

Exercise 3.6 (Poisson process restarted at stopping times). Let $(T_k)_{k \geq 1} \sim \text{PP}(\lambda)$ and let $(N(t))_{t \geq 0}$ be its counting process.

- (i) Let T be any stopping time for $(N(t))_{t \geq 0}$. Use the memoryless property of Poisson processes and Lemma 3.3 to show that $(N(t) - N(T))_{t \geq T}$ is the counting process of a $\text{PP}(\lambda)$, which is independent of $(N(t))_{0 \leq t < T}$.
- (ii) Let T be the first time that we see three arrivals during a unit time. That is,

$$T = \inf\{t \geq 0 \mid N(t) - N(t-1) = 3\}. \quad (188)$$

Show that T is a stopping time. According to (i), $(N(t) - N(T))_{t \geq T}$ is the counting process of a $\text{PP}(\lambda)$, which is independent of what have happened during $[0, T]$.

- (iii) Let T be the first time t that there is no arrival during the interval $[t, t+1]$. Is T a stopping time? Is $(N(t) - N(T))_{t \geq T}$ the counting process of a $\text{PP}(\lambda)$?

4. Merging and splitting of Poisson process

If customers arrive at a bank according to $\text{Poisson}(\lambda)$ and if each one is male or female independently with probability q and $1 - q$, then the 'thinned out' process of only male customers is a $\text{Poisson}(q\lambda)$; the process of female customers is a $\text{Poisson}((1 - q)\lambda)$.

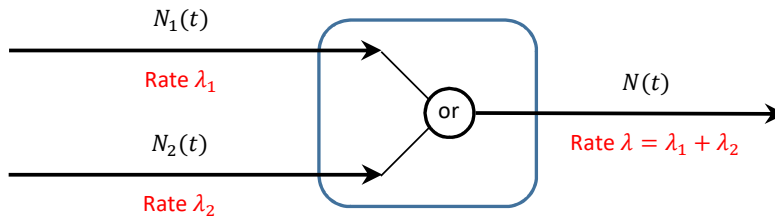


FIGURE 3. Merging two independent Poisson processes of rates λ_1 and λ_2 gives a new Poisson process of rate $\lambda_1 + \lambda_2$.

The reverse operation of splitting a given PP into two complementary PPs is call the 'merging'. Namely, imagine customers arrive at a register through two doors A and B independently according to PPs of rates λ_A and λ_B , respectively. Then the combined arrival process of entire customers is again a PP of the added rate.

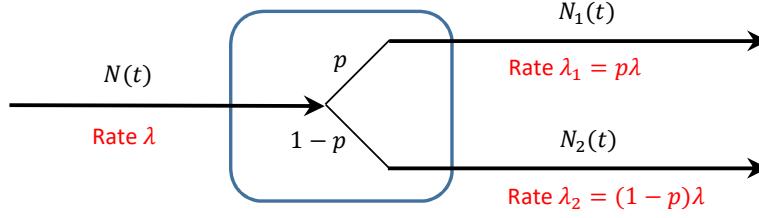


FIGURE 4. Splitting of Poisson process $N(t)$ of rate λ according to an independent Bernoulli process of parameter p .

Exercise 4.1 (Excerpted from [BT02]). Transmitters A and B independently send messages to a single receiver according to Poisson processes with rates $\lambda_A = 3$ and $\lambda_B = 4$ (messages per min). Each message (regardless of the source) contains a random number of words with PMF

$$\mathbb{P}(1 \text{ word}) = 2/6, \quad \mathbb{P}(2 \text{ words}) = 3/6, \quad \mathbb{P}(3 \text{ words}) = 1/6, \quad (189)$$

which is independent of everything else.

- (i) Find $\mathbb{P}(\text{total nine messages are received during } [0, t])$.
- (ii) Let $M(t)$ be the total number of words received during $[0, t]$. Find $\mathbb{E}[M(t)]$.
- (iii) Let T be the first time that the receiver receives exactly three messages consisting of three words from transmitter A . Find distribution of T .
- (iv) Compute $\mathbb{P}(\text{exactly seven messages out of the first ten messages are from } A)$.

Exercise 4.2 (Order statistics of i.i.d. Exp RVs). One hundred light bulbs are simultaneously put on a life test. Suppose the lifetimes of the individual light bulbs are independent $\text{Exp}(\lambda)$ RVs. Let T_k be the k th time that some light bulb fails. We will find the distribution of T_k using Poisson processes.

- (i) Think of T_1 as the first arrival time among 100 independent PPs of rate λ . Show that $T_1 \sim \text{Exp}(100\lambda)$.
- (ii) After time T_1 , there are 99 remaining light bulbs. Using memoryless property, argue that $T_2 - T_1$ is the first arrival time of 99 independent PPs of rate λ . Show that $T_2 - T_1 \sim \text{Exp}(99\lambda)$ and that $T_2 - T_1$ is independent of T_1 .
- (iii) As in the coupon collector problem, we break up

$$T_k = \tau_1 + \tau_2 + \cdots + \tau_k, \quad (190)$$

where $\tau_i = T_i - T_{i-1}$ with $\tau_1 = T_1$. Note that τ_i is the waiting time between $i-1$ st and i th failures. Using the ideas in (i) and (ii), show that τ_i 's are independent and $\tau_i \sim \text{Exp}((101-i)\lambda)$. Deduce that

$$\mathbb{E}[T_k] = \frac{1}{\lambda} \left(\frac{1}{100} + \frac{1}{99} + \cdots + \frac{1}{(100-k+1)} \right), \quad (191)$$

$$\text{Var}[T_k] = \frac{1}{\lambda^2} \left(\frac{1}{100^2} + \frac{1}{99^2} + \cdots + \frac{1}{(100-k+1)^2} \right). \quad (192)$$

- (iv) Let $X_1, X_2, \dots, X_{100}$ be i.i.d. $\text{Exp}(\lambda)$ variables. Let $X_{(1)} < X_{(2)} < \cdots < X_{(100)}$ be their order statistics, that is, $X_{(k)}$ is the k th smallest among the X_i 's. Show that $X_{(k)}$ has the same distribution as T_k , the k th time some light bulb fails. (So we know what it is from the previous parts.)

In the next two exercises, we rigorously justify merging and splitting of Poisson processes.

Exercise 4.3 (Merging of independent PPs). Let $(N_1(t))_{t \geq 0}$ and $(N_2(t))_{t \geq 0}$ be the counting processes of two independent PPs of rates λ_1 and λ_2 , respectively. Define a new counting process $(N(t))_{t \geq 0}$ by

$$N(t) = N_1(t) + N_2(t). \quad (193)$$

In this exercise, we show that $(N(t))_{t \geq 0} \sim \text{PP}(p\lambda)$.

- (i) Let $\tau_k^{(1)}, \tau_k^{(2)}$, and τ_k be the k th inter-arrival times of the counting processes $(N_1(t))_{t \geq 0}$, $(N_2(t))_{t \geq 0}$, and $(N(t))_{t \geq 0}$. Show that $\tau_1 = \min(\tau_1^{(1)}, \tau_1^{(2)})$. Conclude that $\tau_1 \sim \text{Exp}(\lambda_1 + \lambda_2)$.
- (ii) Let T_k be the k th arrival time for the joint process $(N(t))_{t \geq 0}$. Use memoryless property of PP and Exercise 3.6 to deduce that N_1 and N_2 restarted at time T_k are independent PPs of rates λ_1 and λ_2 , which are also independent from the past (before time T_k).
- (iii) From (ii), show that

$$\tau_{k+1} = \min(\tilde{\tau}_1, \tilde{\tau}_2), \quad (194)$$

where $\tilde{\tau}_1$ is the waiting time for the first arrival after time T_k for N_1 , and similarly for $\tilde{\tau}_2$. Deduce that $\tau_{k+1} \sim \text{Exp}(\lambda_1 + \lambda_2)$ and it is independent of $\tau_1, \dots, \tau_k$. Conclude that $(N(t))_{t \geq 0} \sim \text{PP}(\lambda_1 + \lambda_2)$.

Exercise 4.4 (Splitting of PP). Let $(N(t))_{t \geq 0}$ be the counting process of a $\text{Poisson}(\lambda)$, and let $(X_k)_{k \geq 0}$ be a sequence of i.i.d. Bernoulli(p) RVs. We define two counting processes $(N_1(t))_{t \geq 0}$ and $(N_2(t))_{t \geq 0}$ by

$$N_1(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t) \mathbf{1}(X_k = 1) = \#(\text{arrivals with coin landing on heads up to time } t), \quad (195)$$

$$N_2(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t) \mathbf{1}(X_k = 0) = \#(\text{arrivals with coin landing on tails up to time } t). \quad (196)$$

In this exercise, we show that $(N_1(t))_{t \geq 0} \sim \text{Poisson}(p\lambda)$ and $(N_2(t))_{t \geq 0} \sim \text{Poisson}((1-p)\lambda)$.

- (i) Let τ_k and $\tau_k^{(1)}$ be the k th inter-arrival times of the counting processes $(N(t))_{t \geq 0}$ and $(N_1(t))_{t \geq 0}$. Let Y_k be the location of k th 1 in $(X_i)_{i \geq 0}$. Show that

$$\tau_1^{(1)} = \sum_{i=1}^{Y_1} \tau_i. \quad (197)$$

- (ii) Show that

$$\tau_2^{(1)} = \sum_{k=Y_1+1}^{Y_2} \tau_i. \quad (198)$$

- (iii) Show that in general,

$$\tau_k^{(1)} = \sum_{i=Y_{k-1}+1}^{Y_k} \tau_i. \quad (199)$$

- (iv) Recall that $Y_k - Y_{k-1}$'s are i.i.d. $\text{Geom}(p)$ RVs. Use Example 4.5 and (iii) to deduce that $\tau_k^{(1)}$'s are i.i.d. $\text{Exp}(p\lambda)$ RVs. Conclude that $(N_1(t)) \sim \text{Poisson}(p\lambda)$. (The same argument shows $(N_2(t))_{t \geq 0} \sim \text{Poisson}((1-p)\lambda)$.)

Example 4.5 (Sum of geometric number of Exp. is Exp.). Let $X_i \sim \text{Exp}(\lambda)$ for $i \geq 0$ and let $N \sim \text{Geom}(p)$. Let $Y = \sum_{k=1}^N X_k$. Suppose all RVs are independent. Then $Y \sim \text{Exp}(p\lambda)$.

To see this, recall that their moment generating functions are

$$M_{X_1}(t) = \frac{\lambda}{\lambda - t}, \quad M_N(t) = \frac{pe^t}{1 - (1-p)e^t}. \quad (200)$$

Hence (see Remark 4.6)

$$M_Y(t) = M_N(\log M_{X_1}(t)) = \frac{p \frac{\lambda}{\lambda - t}}{1 - (1-p) \frac{\lambda}{\lambda - t}} = \frac{p\lambda}{(\lambda - t) - \lambda(1-p)} = \frac{p\lambda}{p\lambda - t}. \quad (201)$$

Notice that this is the MGF of an $\text{Exp}(p\lambda)$ variable. Thus by uniqueness, we conclude that $Y \sim \text{Exp}(p\lambda)$.

▲

Remark 4.6. Let $Y = X_1 + X_2 + \dots + X_N$, where X_i 's are i.i.d. RVs and N is an independent RV taking values from positive integers. By iterated expectation, we have

$$M_Y(t) = \mathbb{E}[e^{tY}] = \mathbb{E}[e^{tX_1} e^{tX_2} \dots e^{tX_N}] \quad (202)$$

$$= \mathbb{E}_N[\mathbb{E}[e^{tX_1} e^{tX_2} \dots e^{tX_N} | N]] \quad (203)$$

$$= \mathbb{E}_N[\mathbb{E}[e^{tX_1}]^N | N] \quad (204)$$

$$= \mathbb{E}_N[M_{X_1}(t)^N] \quad (205)$$

$$= \mathbb{E}_N[e^{N \log M_{X_1}(t)}] \quad (206)$$

$$= M_N(\log M_{X_1}(t)). \quad (207)$$

5. M/M/1 queue

Suppose customers arrive at a single server according to a Poisson process with rate $\lambda > 0$. Customers get serviced one by one according to the first-in-first-out ordering, and suppose there is no cap in the queue. Finally, assume that each customer in the top of the queue takes an independent $\text{Exp}(\mu)$ time to get serviced and exit the queue.

This is called the *M/M/1 queue* in queuing theory. Here the name of this model follows Kendall's notation, where the first 'M' stands for memorylessness of the arrival process, the second 'M' stands for memorylessness of the service times, and '1' means there is a single server. One can think of *M/M/c* queue for c servers, in general.

The main quantity of interest is the number of customers waiting in the queue at time t , which we denote by $Y(t)$. Then $(Y(t))_{t \geq 0}$ is a continuous-time stochastic process, which changes by 1 whenever a new customer arrives or the top customer in the queue leaves the system. In fact, this system can be modeled as a Markov chain, if we only think of the times when the queue state changes. Namely, let $T_1 < T_2 < \dots$ denote the times when the queue length changes. Let $X_k := Y(T_k)$. Then $(X_k)_{k \geq 0}$ forms a Markov chain. In fact, it is the Birth-Death chain we have seen before.

- (i) Let $(T_k^a)_{k \geq 0} \sim \text{PP}(\lambda)$ and $(T_k^d)_{k \geq 0} \sim \text{PP}(\mu)$. These are sequences of arrival and departure times, respectively. Let $\tilde{T}_i$ be the i th smallest time among all such arrival and departure times. In the next section, we will learn $(\tilde{T}_i)_{i \geq 0} \sim \text{PP}(\lambda + \mu)$. This is called 'merging' of two independent Poisson processes (see Exercise 4.3). Note that $\tilde{T}_i$ is the i th time that 'something happens' to the queue.
- (ii) Define a Markov chain $(X_k)_{k \geq 0}$ on state space $\Omega = \{0, 1, 2, \dots\}$ by

$$X_k = Y(\tilde{T}_k). \quad (208)$$

Namely, X_k is the number of customers in the queue at k th time that something happens to the queue.

- (iii) What is the probability that $X_2 = 2$ given $X_1 = 1$? As soon as a new customer arrives at time T_1 , she gets serviced and it takes an independent $\sigma_1 \sim \text{Exp}(\mu)$ time. Let τ_1 be the inter-arrival time between the first and second customer. Then by Example 1.4 (competing exponentials),

$$\mathbb{P}(X_2 = 2 | X_1 = 1) = \mathbb{P}(\text{New customer arrives before the previous customer exits}) \quad (209)$$

$$= \mathbb{P}(\tau_1 < \sigma_1) = \frac{\lambda}{\mu + \lambda}. \quad (210)$$

Similarly,

$$\mathbb{P}(X_2 = 0 | X_1 = 1) = \mathbb{P}(\text{New customer arrives after the previous customer exits}) \quad (211)$$

$$= \mathbb{P}(\tau_1 > \sigma_1) = \frac{\mu}{\mu + \lambda}. \quad (212)$$

(iii) In general, consider what has to happen for $X_{k+1} = n + 1$ given $X_k = n \geq 1$:

$$\mathbb{P}(X_{k+1} = n + 1 | X_k = n) = \mathbb{P}\left(\begin{array}{c} \text{remaining service time} \\ \text{after time } T_k \end{array} > \begin{array}{c} \text{remaining time until first} \\ \text{arrival after time } T_k \end{array}\right) \quad (213)$$

Note that the remaining service time after time T_k is still an $\text{Exp}(\mu)$ variable due to the memoryless property of exponential RVs. Moreover, by the memoryless property of Poisson processes, the arrival process restarted at time T_k is a $\text{Poisson}(\lambda)$ process that is independent of the past. Hence (213) is the probability that an $\text{Exp}(\mu)$ RV is less than an independent $\text{Exp}(\lambda)$ RV. Thus we are back to the same computation as in (ii). Similar argument holds for the other possibility $\mathbb{P}(X_{k+1} = n - 1 | X_k = n)$.

(iv) From (i)-(iii), we conclude $(X_k)_{k \geq 0}$ is a Birth-Death chain on state space $\Omega = \{0, 1, 2, \dots\}$. By com-

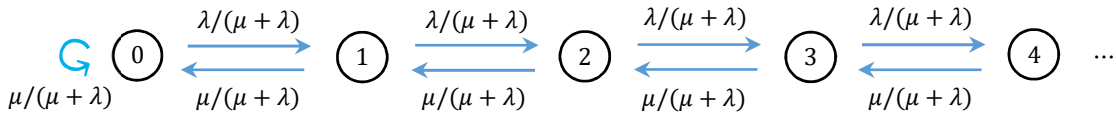


FIGURE 5. State space diagram of the M/M/1 queue

puting the transition matrix P (which is of infinite size!) and solving $\pi P = \pi$, one obtains that the stationary distribution for the M/M/1 queue is unique and it is a geometric distribution. Namely, if we write $\rho = \lambda/\mu$, then

$$\pi(n) = \rho^n (1 - \rho) \quad \forall n \geq 0. \quad (214)$$

Namely, π is the (shifted) geometric distribution with parameter ρ , which is well-defined if and only if $\rho < 1$, that is, $\mu > \lambda$. In words, the rate of service times should be larger than that of the arrival process in order for the queue to not to blow up.

(v) Where does the loop at state 0 in Figure 5 come from? Don't we have no service whatsoever when there is no customer in the queue? The loop is introduced in order to have a consistent time scale to emulate the continuous time Markov process using a discrete time Markov chain.

To illustrate this point, let $\mu = \lambda = 1$. Then the merged Poisson process $(\tilde{T}_k)$ has rate 2. In other words, something happens after the minimum of two independent exponential 1 RVs, which is an exponential RV with rate 2 (Exercise 1.3). Hence all the transitions except from 0 to 1 takes 1/2 unit time on average. On the other hand, we $X_k = 0$ and if we are waiting for the next arrival, this will happen according to $\text{Exp}(1)$ time, which has mean 1. So if we want to emulate the continuous time process by chopping it up at random times with mean 1/2, we need to imagine that there the server takes $\text{Exp}(1)$ times for service regardless of whether there is a customer. This explains the loop at state 0 in Figure 5.

Exercise 5.1 (RW perspective of M/M/1 queue). Consider a M/M/1 queue with service rate μ and arrival rate λ . Denote $p = \lambda/(\mu + \lambda)$. Let $(X_k)_{k \geq 0}$ be the Markov chain where X_k is the number of customers in the queue at k th time that either departure or arrival occurs. (c.f. [Dur10, Example 6.2.4])

(i) Let $(\xi_k)_{k \geq 1}$ be a sequence of i.i.d. RVs such that

$$\mathbb{P}(\xi_k = 1) = p, \quad \mathbb{P}(\xi_k = -1) = 1 - p \quad (215)$$

Define a sequence of RVs $(Z_k)_{k \geq 0}$ by

$$Z_{k+1} = \max(0, Z_k + \xi_k). \quad (216)$$

Show that X_k and Z_k have the same distribution for all $k \geq 1$.

(ii) Define a simple random walk $(S_n)_{n \geq 0}$ by $S_n = \xi_1 + \dots + \xi_n$ for $n \geq 1$ and $S_0 = 0$. Show that Z_k can also be written as

$$Z_k = S_k - \min_{0 \leq i \leq k} S_i. \quad (217)$$

- (iii) The simple random walk $(S_n)_{n \geq 0}$ is called *subcritical* if $p < 1/2$, *critical* if $p = 1/2$, and *supercritical* if $p > 1/2$. Below is a plot of $(Z_k)_{k \geq 0}$ when $p < 1/2$, in which case Z_k is more likely to decrease than to increase when $Z_k \geq 1$. Does it make sense that the $M/M/1$ queue has a unique stationary distribution for $p < 1/2$? Draw plots of Z_k for the critical and supercritical case. Convince yourself that the $M/M/1$ queue should not have a stationary distribution for $p > 1/2$. How about the critical case?



FIGURE 6. Plot of $Z_k = S_k - \min_{0 \leq i \leq k} S_i$ for the subcritical case $p < 1/2$.

6. Poisson process as a counting process

In Section 2, we have defined an arrival process $(T_k)_{k \geq 1}$ to be a Poisson process of rate λ if its inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ variables (Definition 2.2). In this section, we provide equivalent definitions of Poisson processes in terms of the associated counting process (see (249)). This new perspective has many complementary advantages. Most importantly, this allows us to define the time-inhomogeneous Poisson processes, where the rate of arrival changes in time.

Definition 6.1 (Def of PP:counting1). An arrival process $(T_k)_{k \geq 1}$ is said to be a *Poisson process* of rate $\lambda > 0$ if its associated counting process $(N(t))_{t \geq 0}$ satisfies the following properties:

- (i) $N(0) = 0$;
- (ii) (Independent increment) For any $t, s \geq 0$, $N(t+s) - N(t)$ is independent of $(N(u))_{u \leq t}$;
- (iii) For any $t, s \geq 0$, $N(t+s) - N(t) \sim \text{Poisson}(\lambda s)$.

Proposition 6.2. The two definitions of Poisson process in Definitions 2.2 and 6.1 are equivalent.

PROOF. Let $(N(t))_{t \geq 0}$ be a counting process with the properties (i)-(iii) in Def 6.1. We want to show that the inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ RVs. This is the content of Exercise 6.3.

Conversely, let $(T_k)_{k \geq 1}$ be an arrival process. Suppose its inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ RVs. Let $(N(t))_{t \geq 0}$ be its associated counting process. Clearly $N(0) = 0$ by definition so (i) holds. By the memoryless property (Proposition 3.4), $(N_u)_{u \geq t}$ is the counting process of a Poisson process of rate λ (in the sense of Def 2.2) that is independent of the past $(N(u))_{u \leq t}$. In particular, the increment $N(t+s) - N(t)$ during time interval $[t, t+s]$ is independent of the past process $(N(u))_{u \leq t}$, so (ii) holds. Lastly, the increment $N(t+s) - N(t)$ has the same distribution as $N(s) = N(s) - N(0)$ by the memoryless property. Since Exercise 2.4 shows that $N(t) \sim \text{Poisson}(\lambda t)$, we have (iii). $\square$

Exercise 6.3. Let $(N(t))_{t \geq 0}$ be a counting process with the properties (i)-(iii) in Def 6.1. Let $T_k = \inf\{u \geq 0 \mid N(u) = k\}$ be the k th arrival time and let $\tau_k = T_k - T_{k-1}$ be the k th inter-arrival time.

- (i) Use the fact that T_k is a stopping time for $(N(t))_{t \geq 0}$ and Lemma 3.3 to deduce that $(N(T_k + t) - N(T_k))_{t \geq 0}$ is the counting process of a $\text{PP}(\lambda)$ that is independent of $(N(t))_{t \leq T_k}$.
- (ii) Let $Z(t) = \inf\{u \geq 0 \mid N(t+u) > N(t)\}$ be the waiting time for the first arrival after time t . Show that $Z(t) \sim \text{Exp}(\lambda)$ for all $t \geq 0$.
- (iii) Use (ii) and conditioning on T_{k-1} to show that $\tau_k \sim \text{Exp}(\lambda)$ for all $k \geq 1$.

Next, we give yet another definition of Poisson process in terms of the asymptotic properties of its counting process. For this, we need something called the ‘small-o’ notation. We say a function $f(t)$ is of order $o(t)$ or write $f(t) = o(t)$ if

$$\lim_{t \rightarrow 0} \frac{f(t)}{t} = 0. \quad (218)$$

Definition 6.4 (Def of PP:counting2). A counting process $(N(t))_{t \geq 0}$ is said to be a *Poisson process* with rate $\lambda > 0$ if it satisfies the following conditions:

- (i) $N(0) = 0$;
- (ii) $\mathbb{P}(N(t) = 0) = 1 - \lambda t + o(t)$;
- (iii) $\mathbb{P}(N(t) = 1) = \lambda t + o(t)$;
- (iv) $\mathbb{P}(N(t) \geq 2) = o(t)$;
- (v) (Independent increment) For any $t, s \geq 0$, $N(t+s) - N(t)$ is independent of $(N(u))_{u \leq t}$;
- (vi) (Stationary increment) For any $t, s \geq 0$, the distribution of $N(t+s) - N(t)$ is invariant under t .

It is easy to see that our usual definition of Poisson process in Definition 2.2 satisfies the properties (i)-(vi) above.

Proposition 6.5. Let $(T_k)_{k \geq 1}$ be a Poisson process of rate λ in the sense of Definition 6.1 and let $(N(t))_{t \geq 0}$ be its associated counting process. Then $(N(t))_{t \geq 0}$ is a Poisson process in the sense of Definition 6.4.

PROOF. Using Taylor expansion of exponential function, note that

$$e^{-\lambda t} = 1 - \lambda t + o(t) \quad (219)$$

for all $t > 0$. Hence

$$\mathbb{P}(N(t) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad (220)$$

$$= (1 - \lambda t + o(t)) \frac{(\lambda t)^n}{n!}. \quad (221)$$

So plugging in $n = 0$ and 1 gives (ii) and (iii). For (iv), we use (ii) and (iii) to get

$$\mathbb{P}(N(t) \geq 2) = 1 - \mathbb{P}(N(t) \leq 1) = 1 - (1 - \lambda t + o(t)) - (\lambda t + o(t)) = o(t). \quad (222)$$

Lastly, (v) and (iv) follows from the memoryless property of Poisson process (Proposition 3.4). $\square$

Next, we consider the converse implication. We will break this into several exercises.

Exercise 6.6. Let $(N(t))_{t \geq 0}$ is the Poisson process with rate $\lambda > 0$ in the sense of Definition 6.4. In this exercise, we will show that $\mathbb{P}(N(t) = 0) = e^{-\lambda t}$.

- (i) Use independent/stationary increment properties to show that

$$\mathbb{P}(N(t+h) = 0) = \mathbb{P}(N(t) = 0, N(t+h) - N(t) = 0) \quad (223)$$

$$= \mathbb{P}(N(t) = 0) \mathbb{P}(N(t+h) - N(t) = 0) \quad (224)$$

$$= \mathbb{P}(N(t) = 0) (1 - \lambda h + o(h)). \quad (225)$$

- (ii) Denote $f_0(t) = \mathbb{P}(N(t) = 0)$. Use (i) to show that

$$\frac{f_0(t+h) - f_0(t)}{h} = \left(-\lambda + \frac{o(h)}{h} \right) f_0(t). \quad (226)$$

By taking limit as $h \rightarrow 0$, show that $f(t)$ satisfies the following differential equation

$$\frac{df_0(t)}{dt} = -\lambda f_0(t). \quad (227)$$

- (iii) Conclude that $\mathbb{P}(N(t) = 0) = e^{-\lambda t}$.

Next, we generalize the ideas used in the previous exercise to compute the distribution of $N(t)$.

Exercise 6.7. Let $(N(t))_{t \geq 0}$ is the Poisson process with rate $\lambda > 0$ in the sense of Definition 6.4. Denote $f_n(t) = \mathbb{P}(N(t) = n)$ for each $n \geq 0$.

(i) Show that

$$\mathbb{P}(N(t) \leq n-2, N(t+h) = n) \leq \mathbb{P}(N(t+h) - N(t) \geq 2). \quad (228)$$

Conclude that

$$\mathbb{P}(N(t) \leq n-2, N(t+h) = n) = o(h). \quad (229)$$

(ii) Use (i) and independent/stationary increment properties to show that

$$f_n(t+h) = \mathbb{P}(N(t+h) = n) = \mathbb{P}(N(t) = n, N(t+h) - N(t) = 0) \quad (230)$$

$$+ \mathbb{P}(N(t) = n-1, N(t+h) - N(t) = 1) \quad (231)$$

$$+ \mathbb{P}(N(t) \leq n-2, N(t+h) = n) \quad (232)$$

$$= f_n(t)(1 - \lambda h + o(h)) + f_{n-1}(t)(\lambda h + o(h)) + o(h). \quad (233)$$

(iii) Use (ii) to show that the following differential equation holds:

$$\frac{df_n(t)}{dt} = -\lambda f_n(t) + \lambda f_{n-1}(t). \quad (234)$$

(iv) By multiplying the integrating factor $\mu(t) = e^{\lambda t}$ to (234), show that

$$(e^{\lambda t} f_n(t))' = \lambda e^{\lambda t} f_{n-1}(t). \quad (235)$$

Use the initial condition $f_n(0) = \mathbb{P}(N(0) = n) = 0$ to derive the recursive equation

$$f_n(t) = \lambda e^{-\lambda t} \int_0^t e^{\lambda s} f_{n-1}(s) ds. \quad (236)$$

(v) Use induction to conclude that $f_n(t) = (\lambda t)^n e^{-\lambda t} / n!$.

(vi) Conclude that for all $t, s \geq 0$ and $n \geq 0$,

$$N(t+s) - N(s) \sim \text{Poisson}(\lambda t). \quad (237)$$

7. Nonhomogeneous Poisson process

In this section, we introduce Poisson process with time-varying rate $\lambda(t)$.

Example 7.1. Consider an counting process $(N(t))_{t \geq 0}$, which follows $\text{PP}(\lambda_1)$ on interval $[1, 2)$, $\text{PP}(\lambda_2)$ on interval $[2, 3)$, and $\text{PP}(\lambda_3)$ on interval $[3, 4]$. Further assume that the increments $N(2) - N(1)$, $N(3) - N(2)$, and $N(4) - N(3)$ are independent. Then what is the distribution of the total number of arrivals during $[1, 4]$? Since we can add independent Poisson RVs and get a Poisson RV with added rates, we get

$$N(4) - N(1) = [N(4) - N(3)] + [N(3) - N(2)] + [N(2) - N(1)] \sim \text{Poisson}(\lambda_1 + \lambda_2 + \lambda_3). \quad (238)$$

Note that the combined rate $\lambda_1 + \lambda_2 + \lambda_3$ can be seen as the integral of the step function $f(t) = \lambda_1 \mathbf{1}(t \in [1, 2)) + \lambda_2 \mathbf{1}(t \in [2, 3)) + \lambda_3 \mathbf{1}(t \in [3, 4])$. ▲

In general, suppose we have a concatenation of Poisson processes on disjoint intervals of very small lengths. Then $N(t) - N(s)$ can be seen as the sum of independent increments over the interval $[t, s]$, and by additivity of independent Poisson increments, it follows a Poisson distribution with rate given by the 'Riemann sum'. As the lengths of the intervals go to zero, this Riemann sum of rates tend to the integral of the rate function $\lambda(r)$ over the interval $[s, t]$. This suggests the following definition of nonhomogeneous Poisson processes.

Definition 7.2. An arrival process $(T_k)_{k \geq 1}$ is said to be a Poisson process with rate $\lambda(t)$ if its counting process $(N_t)_{t \geq 0}$ satisfies the following properties:

(i) $N(0) = 0$.

- (ii) $(N(t))_{t \geq 0}$ has independent increments.
 (iii) For any $0 \leq s < t$, $N(t) - N(s) \sim \text{Poisson}(\mu)$ where

$$\mu = \int_s^t \lambda(r) dr. \quad (239)$$

Example 7.3. A store opens at 8 AM. From 8 until 10 AM, customers arrive at a Poisson rate of four per hour. Between 10 AM and 12 PM, they arrive at a Poisson rate of eight per hour. From 12 PM to 2 PM the arrival rate increases steadily from eight per hour at 12 PM to ten per hour at 2 PM; and from 2 PM to 5 PM, the arrival rate drops steadily from ten per hour at 2 PM to four per hour at 5 PM. Let us determine the probability distribution of the number of customers that enter the store on a given day (between 8 AM and 5 PM).

Let $\lambda(t)$ be the rate function in the statement and let $N(t)$ be a nonhomogeneous Poisson process with this rate function. From the description above, we can compute $m = \int_8^{17} \lambda(s) ds = 63$. Hence

$$N(17) - N(8) \sim \text{Poisson}(63). \quad (240)$$

▲

Exercise 7.4. Let $(T_k)_{k \geq 1} \sim \text{PP}(\lambda(t))$. Let $(\tau_k)_{k \geq 1}$ be the inter-arrival times.

- (i) Let $Z(t)$ be the waiting time for the first arrival after time t . Show that

$$\mathbb{P}(Z(t) \geq x) = \exp\left(-\int_t^{t+x} \lambda(t) dt\right). \quad (241)$$

- (ii) From (i), deduce that τ_1 has PDF

$$f_{\tau_1}(t) = \lambda(t) e^{-\int_0^t \lambda(r) dr}. \quad (242)$$

- (iii) Denote $\mu(t) = \int_0^t \lambda(s) ds$. Use (i) and conditioning to show

$$\mathbb{P}(\tau_2 > x) = \mathbb{E}_{\tau_1}[\mathbb{P}(\tau_2 > x | \tau_1) | \tau_1] \quad (243)$$

$$= \int_0^\infty \mathbb{P}(\tau_2 > x | \tau_1 = t) f_{\tau_1}(t) dt \quad (244)$$

$$= \int_0^\infty e^{-(\mu(t+x) - \mu(t))} \lambda(t) e^{-\mu(t)} dt \quad (245)$$

$$= \int_0^\infty \lambda(t) e^{-\mu(t+x)} dt. \quad (246)$$

Conclude that τ_1 and τ_2 do not necessarily have the same distribution.

The following exercise shows that the nonhomogeneous Poisson process with rate $\lambda(t)$ can be obtained by a time change of the usual Poisson process of rate 1.

Exercise 7.5. Let $(N_0(t))_{t \geq 0}$ be the counting process of a Poisson process of rate 1. Let $\lambda(t)$ denote a non-negative function of t , and let

$$m(t) = \int_0^t \lambda(s) ds. \quad (247)$$

Define $N(t)$ by

$$N(t) = N_0(m(t)) = \# \text{ arrivals during } [0, m(t)]. \quad (248)$$

Show that $(N(t))_{t \geq 0}$ is the counting process of a Poisson process of rate $\lambda(t)$.

Renewal Processes

1. Definition of renewal processes

Recall that an *arrival process* is a sequence of strictly increasing RVs $0 < T_1 < T_2 < \dots$. For each integer $k \geq 1$, its k th *inter-arrival time* is defined by $\tau_k = T_k - T_{k-1} \mathbf{1}(k \geq 2)$. For a given arrival process $(T_k)_{k \geq 1}$, the associated *counting process* $(N(t))_{t \geq 0}$ is defined by

$$N(t) = \sum_{k=1}^{\infty} \mathbf{1}(T_k \leq t) = \#(\text{arrivals up to time } t). \quad (249)$$

Note that these three processes (arrival times, inter-arrival times, and counting) determine each other:

$$(T_k)_{k \geq 1} \iff (\tau_k)_{k \geq 1} \iff (N(t))_{t \geq 0}. \quad (250)$$

Also note that we have the following relation

$$\{T_n \leq t\} = \{N(t) \geq n\}. \quad (251)$$

That is, the n th customer arrives by time t if and only if at least n customers arrive up to time t .

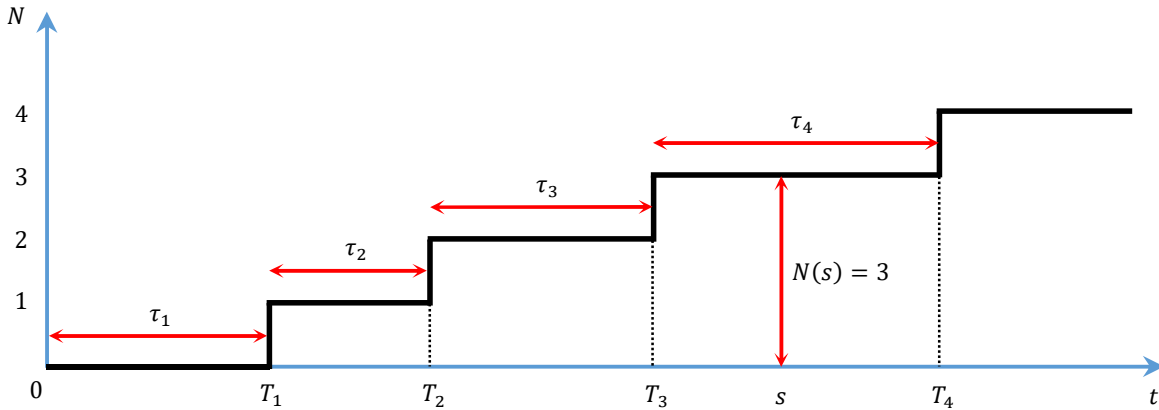


FIGURE 1. Illustration of an arrival process $(T_k)_{k \geq 1}$ and its associated counting process $(N(t))_{t \geq 0}$. τ_k 's denote inter-arrival times. $N(t) \equiv 3$ for $T_3 < t \leq T_4$.

Definition 1.1 (Renewal process). A counting process $(N(t))_{t \geq 0}$ is called a *renewal process* if its inter-arrival times $\tau_1, \tau_2, \dots$ are i.i.d. with $\mathbb{E}[\tau_1] < \infty$.

Example 1.2. Let $(X_t)_{t \geq 0}$ be an irreducible Markov chain on a finite state space Ω . Let $X_0 = x \in \Omega$ and let T_k be the k th time that the chain returns to x . By the strong Markov property, the inter-arrival times $\tau_k = T_k - T_{k-1}$ are i.i.d. We also know that $\mathbb{E}[\tau_k] < \infty$ since hitting times of finite-state irreducible chains have finite mean (see Exercise 4.6 in Lecture note 1). Hence $(T_k)_{k \geq 0}$ is a renewal process. $\blacktriangle$

Example 1.3. Let $(\xi_k)_{k \geq 0}$ be a sequence of i.i.d. RVs with

$$\mathbb{P}(\xi_k = 1) = \mathbb{P}(\xi_k = -1) = 1/2. \quad (252)$$

Define a random walk $(S_n)_{n \geq 0}$ by $S_0 = 0$ and $S_n = \xi_1 + \dots + \xi_n$. Let T_k denote the k th time that the walk visits the origin. Note that one can view S_n as a Markov chain on state space $\Omega = \mathbb{Z}$. Hence again by the strong Markov property, the inter-arrival times $\tau_k = T_k - T_{k-1}$ are i.i.d. However, later we will see that $\mathbb{E}[\tau_k] = \infty$. Hence $(T_k)_{k \geq 0}$ is not a renewal process as we defined above. However, we can still call it a 'renewal process with infinite expected inter-arrival times'. This will make more sense since, albeit having infinite mean, the inter-arrival times are finite almost surely. In other words, S_n always returns to the origin with probability 1.

How do we know that the walk S_n will eventually return back to the origin? That is, do we know that the inter-arrival times are finite almost surely? It is easy to see that S_n is an irreducible Markov chain. However, since the state space Ω for S_n is infinite, this does not guarantee that S_n returns to zero with probability 1.

In order to see this, recall the Gambler's ruin problem (Exercise 7.1 in Lecture note 1). Namely, let $(X_t)_{t \geq 0}$ be a Markov chain on state space $\Omega_N = \{0, 1, \dots, N\}$ with transition probabilities

$$\mathbb{P}(X_{t+1} = k+1 | X_t = k) = p, \quad \mathbb{P}(X_{t+1} = k-1 | X_t = k) = 1-p \quad \forall 0 \leq k < N. \quad (253)$$

Let $\rho = (1-p)/p$. We have seen that, for any $0 < i < N$,

$$\mathbb{P}(X_t \text{ hits } N \text{ before } 0 | X_0 = i) = \frac{1 + \rho + \dots + \rho^{i-1}}{1 + \rho + \dots + \rho^{N-1}}. \quad (254)$$

For our case, let $p = 1/2$ so that $\rho = 1$. We can view X_t as the random walk S_n during it lies in the interval $[0, N]$. Hence

$$\mathbb{P}(S_n \text{ hits } N \text{ before } 0 | S_0 = i) = i/N. \quad (255)$$

Now by taking the limit $N \rightarrow \infty$, which makes the right barrier at N to fade away, we deduce

$$\mathbb{P}(S_n \text{ never hits } 0 | S_0 = i) = \lim_{N \rightarrow \infty} \mathbb{P}(S_n \text{ hits } N \text{ before } 0 | S_0 = i) = 0. \quad (256)$$

This shows S_0 eventually returns to 0 with probability 1 starting from any initial state $S_0 = i > 0$. Since S_n is symmetric, the same conclusion holds for negative starting location. Hence the inter-arrival times are finite with probability 1. $\blacktriangle$

A cornerstone in the theory of renewal processes is the following strong law of large numbers for renewal processes. We first recall the strong law of large numbers.

Theorem 1.4 (SLLN). *Let $(X_k)_{k \geq 1}$ be i.i.d. RVs and let $S_n = \sum_{k=1}^n X_i$, $n \geq 1$ be a random walk. Suppose $\mathbb{E}[|X_1|] < \infty$ and let $\mathbb{E}[X_1] = \mu$. Then*

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = \mu\right) = 1. \quad (257)$$

PROOF. See Durrett [Dur10, Thm. 2.4.1]. $\square$

Exercise 1.5. Let $(X_k)_{k \geq 1}$ be i.i.d. RVs and let $S_n = \sum_{k=1}^n X_i$, $n \geq 1$ be a random walk. Suppose $\mathbb{E}[|X_1|] < \infty$ and let $\mathbb{E}[X_1] = \mu$. Let $(N(t))_{t \geq 0}$ be any counting process such that $N(t) \rightarrow \infty$ as $t \rightarrow \infty$. Show that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{S_{N(t)}}{N(t)} = \mu\right) = 1. \quad (258)$$

Theorem 1.6 (Renewal SLLN). *Let $(T_k)_{k \geq 0}$ be a renewal process and let $(\tau_k)_{k \geq 0}$ and $(N(t))_{t \geq 0}$ be the associated inter-arrival times and counting process, respectively. Let $\mathbb{E}[\tau_k] = \mu$ be the mean inter-arrival time. If $0 < \mu < \infty$, then*

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu}\right) = 1. \quad (259)$$

PROOF. First, write $T_k = \tau_1 + \tau_2 + \cdots + \tau_k$. Since the inter-arrival times are i.i.d. with mean $\mu < \infty$, the strong law of large numbers imply

$$\mathbb{P}\left(\lim_{k \rightarrow \infty} \frac{T_k}{k} = \mu\right) = 1. \quad (260)$$

Next, fix $t \geq 0$ and let $N(t) = n$, so that there are total n arrivals up to time t . Then the n th arrival time T_n must occur by time t , whereas the $n+1$ st arrival time T_{n+1} must occur after time t . Hence $T_n \leq t < T_{n+1}$. In general, we have

$$T_{N(t)} \leq t < T_{N(t)+1}. \quad (261)$$

Dividing by $N(t)$, we get

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{T_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)}. \quad (262)$$

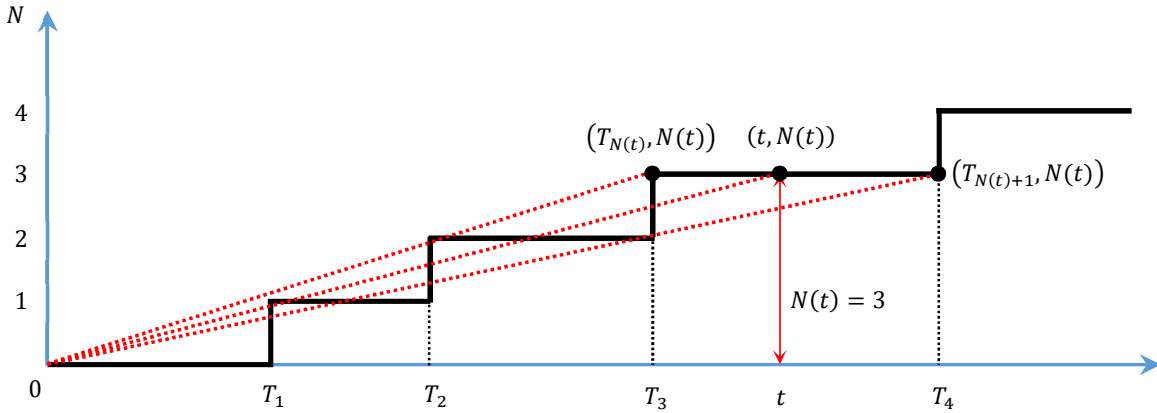


FIGURE 2. Illustration of the inequalities (262).

To take the limit as $t \rightarrow \infty$, we note that $\mathbb{P}(T_k < \infty) = 1$ for all k since $\mathbb{E}[\tau_k] = \mu < \infty$. This yields $N(t) \geq k$ for some large enough t . Since k was arbitrary, this yields $N(t) \nearrow \infty$ as $t \rightarrow \infty$ with probability 1. Therefore, according to (260) and Exercise 1.5, we get

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{T_{N(t)}}{N(t)} = \mu\right) = \mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{T_{N(t)+1}}{N(t)+1} = \mu\right) = \mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{N(t)+1}{N(t)} = 1\right) = 1. \quad (263)$$

Hence (262) gives

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{t}{N(t)} = \mu\right) = 1. \quad (264)$$

Since $\mu > 0$, we can take the reciprocal inside the above probability. This shows the assertion. $\square$

Example 1.7 (Poisson process). Let $(T_k)_{k \geq 1} \sim \text{PP}(\lambda)$. Since the inter-arrival times are i.i.d. $\text{Exp}(\lambda)$ RVs, $(T_k)_{k \geq 0}$ is a renewal process. Moreover, since the mean inter-arrival time is $1/\lambda$, the renewal SLLN yields

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \lambda\right) = 1. \quad (265)$$

Namely, with probability 1, we tend to see about λt arrivals during $[0, t]$ as $t \rightarrow \infty$. In other words, we tend to see λ arrivals during an interval of unit length. Hence it makes sense to call the parameter λ of $\text{PP}(\lambda)$ as its ‘rate’. $\blacktriangle$

2. Renewal reward processes

In this section, we consider a renewal process together with rewards, which are given for each inter-arrival times. This simple extension of the renewal processes will greatly improve the applicability of our theory.

Let $(T_k)_{k \geq 0}$ be a renewal process and let $(\tau_k)_{k \geq 0}$ and $(N(t))_{t \geq 0}$ be the associated inter-arrival times and counting process, respectively. We define the *reward process* $(R(t))_{t \geq 0}$ associated with the renewal process $(T_k)_{k \geq 0}$ and fixed reward function $g : [0, \infty) \rightarrow \mathbb{R}$ by

$$R(t) = \sum_{k=1}^{N(t)} g(\tau_k). \quad (266)$$

Namely, upon the k th arrival at time T_k , we receive a reward $g(\tau_k)$. Then $R(t)$ is the total reward up to time t .

As we looked at the average number of arrivals $N(t)/t$ as $t \rightarrow \infty$, a natural quantity to look at for the reward process is the ‘average reward’ $R(t)/t$ as $t \rightarrow \infty$. Intuitively, since everything refreshes upon new arrivals, we should expect

$$\frac{R(t)}{t} \rightarrow \frac{\text{expected reward during one ‘cycle’}}{\text{expected duration of one ‘cycle’}} \quad (267)$$

as $t \rightarrow \infty$ almost surely. This is made precise by the following result.

Theorem 2.1 (Renewal reward SLLN). *Let $(T_k)_{k \geq 0}$ be a renewal process and let $(R(t))_{t \geq 0}$ be the associated reward process with reward function g . Suppose $0 < \mathbb{E}[|g(\tau_1)|] < \infty$, where τ_1 is the first inter-arrival time. Then*

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{R(t)}{t} = \frac{\mathbb{E}[g(\tau_1)]}{\mathbb{E}[\tau_1]} \right) = 1. \quad (268)$$

PROOF. Let $(\tau_k)_{k \geq 0}$ denote the inter-arrival times for the renewal process $(T_k)_{k \geq 0}$. Note that

$$\frac{R(t)}{t} = \left(\frac{1}{N(t)} \sum_{k=1}^{N(t)} g(\tau_k) \right) \frac{N(t)}{t}. \quad (269)$$

Hence by Exercise 1.5, the ‘average reward’ up to time t in the bracket converges to $\mathbb{E}[g(\tau_1)]$ almost surely. Moreover, the average number of arrivals $N(t)/t$ converges to $1/\mathbb{E}[\tau_1]$ by Theorem 1.6. Hence the assertion follows. $\square$

Remark 2.2. Theorem 1.6 can be obtained as a special case of the above reward version of SLLN, simply by choosing $g \equiv 1$ so that $R(t) = N(t)$.

Example 2.3 (Long run car costs). This example is excerpted from [Dur99]. Mr. White do not drive the same car more than t^* years, where $t^* > 0$ is some fixed number. He changes to a new car when the old one breaks down or reaches t^* years. Let X_k be the life time of the k th car that Mr. White drives, which are i.i.d. with finite expectation. Let τ_k be the duration of his k th car. According to his policy, we have

$$\tau_k = \min(X_k, t^*). \quad (270)$$

Let $T_k = \tau_1 + \dots + \tau_k$ be the time that Mr. White is done with the k th car. Then $(T_k)_{k \geq 0}$ is a renewal process. Note that the expected running time for the k th car is

$$\mathbb{E}[\tau_k] = \mathbb{E}[\tau_k | X_k < t^*] \mathbb{P}(X_k < t^*) + \mathbb{E}[\tau_k | X_k \geq t^*] \mathbb{P}(X_k \geq t^*) \quad (271)$$

$$= \mathbb{E}[X_k | X_k < t^*] \mathbb{P}(X_k < t^*) + t^* \mathbb{P}(X_k \geq t^*). \quad (272)$$

Suppose that the car cost g during each cycle is given by

$$g(t) = \begin{cases} A + B & \text{if } t < t^* \\ A & \text{if } t \geq t^*. \end{cases} \quad (273)$$

Namely, if the car breaks down by t^* years, then Mr. White has to pay $A + B$ dolars; otherwise, the cost is only A dolars. Then the expected cost for one cycle is

$$\mathbb{E}[g(\tau_k)] = A + B\mathbb{P}(\tau_k < t^*) = A + B\mathbb{P}(X_k < t^*). \quad (274)$$

Thus by Theorem 2.1, the long-run car cost of Mr. White is

$$\lim_{t \rightarrow \infty} \frac{R(t)}{t} = \frac{\mathbb{E}[g(\tau_k)]}{\mathbb{E}[\tau_k]} = \frac{A + B\mathbb{P}(X_k < t^*)}{\mathbb{E}[X_k | X_k < t^*]\mathbb{P}(X_k < t^*) + t^*\mathbb{P}(X_k \geq t^*)}. \quad (275)$$

For more concrete example, let $X_k \sim \text{Uniform}([0, 10])$ and let $A = 10$ and $B = 3$. Then

$$\mathbb{E}[g(\tau_k)] = 10 + 3t^*/10. \quad (276)$$

On other other hand,

$$\mathbb{E}[\tau_k] = \mathbb{E}[X_k | X_k < t^*]\mathbb{P}(X_k < t^*) + t^*\mathbb{P}(X_k \geq t^*) \quad (277)$$

$$= \frac{t^*}{2} \frac{t^*}{10} + t^* \frac{10 - t^*}{10} = t^* - (t^*)^2/20. \quad (278)$$

Note that for $\mathbb{E}[X_k | X_k < t^*] = t^*/2$, observe that a uniform RV over $[0, 10]$ conditioned on being $[0, t^*]$ is uniformly distributed over $[0, t^*]$. This yields

$$\frac{\mathbb{E}[g(\tau_k)]}{\mathbb{E}[\tau_k]} = \frac{10 + 0.3t^*}{t^*(1 - t^*/20)}. \quad (279)$$

Lastly, in order to minimize the above long-run cost, we differentiate it in t^* and find global minimum. A straightforward computation shows that the long-run cost is minimized at

$$t^* = \frac{-1 + \sqrt{1.6}}{0.03} \approx 8.83. \quad (280)$$

Thus the optimal strategy for Mr. White in this situation is to drive each car up to 8.83 years. $\blacktriangle$

Exercise 2.4 (Reward from Markov process). Let $(X_k)_{k \geq 0}$ be an irreducible and aperiodic Markov chain on state space $\Omega = \{1, 2, \dots, m\}$ with transition matrix $P = (p_{ij})$. Let π be the unique stationary distribution of the chain.

Suppose the chain spends an independent amount of time at each state $x \in \Omega$, whose distribution F_x may depend only on x . For each real $t \geq 0$, let $Y(t) \in \Omega$ denote the state of the chain at time t . (This is a continuous-time Markov process.)

- (i) Fix $x \in \Omega$, and let $T_k^{(x)}$ denote the k th time that the Markov process $(Y(t))_{t \geq 0}$ returns to x . Let $(\tau_k^{(x)})_{k \geq 1}$ and $(N^{(x)}(t))_{t \geq 0}$ be the associated inter-arrival times and the counting process, respectively. Then

$$N^{(x)}(t) = \text{number of visits to } x \text{ that } (Y(t))_{t \geq 0} \text{ makes up to time } t. \quad (281)$$

Show that $(T_k^{(x)})_{k \geq 2}$ is a renewal process. Moreover, show that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{N^{(x)}(t)}{t} = \frac{1}{\mathbb{E}[\tau_1^{(x)}]}\right) = 1. \quad (282)$$

- (ii) Let T_k denote the k th time that the Markov process $(Y(t))_{t \geq 0}$ jumps. Let $(\tau_k)_{k \geq 1}$ and $(N(t))_{t \geq 0}$ be the associated inter-arrival times and the counting process, respectively. Show that

$$N(t) = N^{(1)}(t) + N^{(2)}(t) + \dots + N^{(m)}(t). \quad (283)$$

Use (i) to derive that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[\tau_1^{(1)}]} + \dots + \frac{1}{\mathbb{E}[\tau_1^{(m)}]}\right) = 1. \quad (284)$$

(iii) Using the fact that (see Exercise 5.13 in Lecture note 2)

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{1}(X_k = x) = \pi(x)\right) = 1, \quad (285)$$

show that

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{N^{(x)}(t)}{N(t)} = \pi(x)\right) = 1. \quad (286)$$

(iv) Let $g : [0, \infty) \rightarrow \mathbb{R}$ be a reward function and fix $x \in \Omega$. Use strong law of large numbers to show

$$\lim_{t \rightarrow \infty} \frac{1}{N^{(x)}(t)} \sum_{k=1}^{N(t)} g(\tau_k) \mathbf{1}(X_k = x) = \mathbb{E}[g(\tau_k) | X_k = x] \quad \text{a.s.} \quad (287)$$

(v) Define

$$R^{(x)}(t) = \sum_{k=1}^{N(t)} g(\tau_k) \mathbf{1}(X_k = x) \quad (288)$$

Namely, every time the Markov process $(Y(t))_{t \geq 0}$ visits x and spends τ_k amount of time, we get a reward of $g(\tau_k)$. Writing

$$\frac{R^{(x)}(t)}{t} = \frac{N(t)}{t} \frac{N^{(x)}(t)}{N(t)} \frac{1}{N^{(x)}(t)} \sum_{k=1}^{N(t)} g(\tau_k) \mathbf{1}(X_k = x), \quad (289)$$

show that as $t \rightarrow \infty$,

$$\lim_{t \rightarrow \infty} \frac{R^{(x)}(t)}{t} = \left(\frac{1}{\mathbb{E}[\tau_1^{(1)}]} + \cdots + \frac{1}{\mathbb{E}[\tau_1^{(m)}]} \right) \pi(x) \mathbb{E}[g(\tau_k) | X_k = x] \quad \text{a.s.} \quad (290)$$

Exercise 2.5 (Alternating renewal process). Let $(\tau_k)_{k \geq 1}$ be a sequence of independent RVs where

$$\mathbb{E}[\tau_{2k-1}] = \mu_1, \quad \mathbb{E}[\tau_{2k}] = \mu_2 \quad \forall k \geq 1. \quad (291)$$

Define an arrival process $(T_k)_{k \geq 1}$ by $T_k = \tau_1 + \cdots + \tau_k$ for all $k \geq 1$.

(i) Is $(T_k)_{k \geq 1}$ a renewal process?

(ii) Let $(X_k)_{k \geq 0}$ be a Markov chain on state space $\Omega = \{1, 2\}$ with transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (292)$$

Show that the chain has $\pi = [1/2, 1/2]$ as the unique stationary distribution.

(iii) Suppose the chain spends time τ_k at state $X_k \in \Omega$ in between the $k-1$ st and k th jump. For each real $t \geq 0$, let $Y(t) \in \Omega$ denote the state of the chain at time t . Let $N(t)$ be the number of jumps that $Y(t)$ makes up to time t . Define

$$R^{(1)}(t) = \sum_{k=1}^{N^{(1)}(t)} \tau_k \mathbf{1}(X_k = 1), \quad (293)$$

which is the total amount of time that $(Y(t))_{t \geq 0}$ spends at state 1. Use Exercise 2.4 to deduce

$$\mathbb{P}\left(\lim_{t \rightarrow \infty} \frac{R^{(1)}(t)}{t} = \frac{\mu_1}{\mu_1 + \mu_2}\right) = 1. \quad (294)$$

Exercise 2.6 (Poisson janitor). (Excerpted from [Dur99]) A light bulb has a random lifespan with distribution F and mean μ_F . A janitor comes at times according to $\text{PP}(\lambda)$ and checks and replace the bulb if it is burnt out. Suppose all bulbs have independent lifespans with the same distribution F .

(i) Let T_k be the k th time that the janitor arrives and replaces the bulb. Show that $(T_k)_{k \geq 0}$ with $T_0 = 0$ is a renewal process.

(ii) Let $(\tau_k)_{k \geq 1}$ be the inter-arrival times of the renewal process defined in (i). Using the memoryless property of Poisson processes to show that

$$\mathbb{E}[\tau_k] = \mu_F + 1/\lambda \quad \forall k \geq 1. \quad (295)$$

(iii) Let $N(t)$ be the number of bulbs replaced up to time t . Show that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{N(t)}{t} = \frac{1}{\mu_F + 1/\lambda}\right) = 1. \quad (296)$$

(iv) Let $B(t)$ be the total duration that bulb is working up to time t , that is,

$$B(t) = \int_0^t \mathbf{1}(\text{Bulb is on at time } s) ds. \quad (297)$$

Use renewal reward process to show that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{B(t)}{t} = \frac{\mu_F}{\mu_F + 1/\lambda}\right) = 1. \quad (298)$$

(v) Let $V(t)$ denote the total number of visits that the janitor has made by time t . Show that

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{N(t)}{V(t)} = \frac{1/\lambda}{\mu_F + 1/\lambda}\right) = 1. \quad (299)$$

That is, the fraction of times that the janitor replaces the bulb converges to $\frac{1/\lambda}{(\mu_F + 1/\lambda)}$ almost surely, which is also the fraction of times that the bulb is off by (iv).

3. Little's Law

In this section, we will learn one of the cornerstones of queuing theory, which is called Little's Law. Roughly speaking, this reads

$$(\text{average size of the system}) = (\text{arrival rate}) \times (\text{average time spent in the system}), \quad (300)$$

or $\ell = \lambda w$ for short. The power of Little's law lies in its applicability in a very general situation; one can even choose a portion of a gigantic queuing network and apply Little's law to analyze the local behavior of the system. We also remark that Little's law applies for deterministic queuing system: For stochastic ones satisfying certain conditions, it will hold with probability 1.

Consider a queuing system, where the k th customer arrives at time t_k , spends w_k units of time, and then exits the system. We assume no two customers arrive at the same time, that is, $t_1 < t_2 < \dots$. Let $N(t)$ and $N^d(t)$ denote the number of arrivals and departures up to time t , respectively. Finally, let $L(t)$ denote the number of customers in the system at time t . To summarize:

$$t_k = \text{Time that the } k\text{th customer enters the system.} \quad (301)$$

$$w_k = \text{Time that the } k\text{th customer spends in the system until he exits.} \quad (302)$$

$$N(t) = \sum_{k=1}^{\infty} \mathbf{1}(t_k \leq t) = \text{Number of arrivals up to time } t \quad (303)$$

$$N^d(t) = \sum_{k=1}^{\infty} \mathbf{1}(t_k + w_k \leq t) = \text{Number of departures up to time } t \quad (304)$$

$$L(t) = N(t) - N^d(t) = \text{Size of the system at time } t. \quad (305)$$

Now we introduce three key quantities which describe the average behavior of the system. Define the following quantities, whenever their limit exist:

$$\ell = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t L(s) ds = \text{Average size of the queue} \quad (306)$$

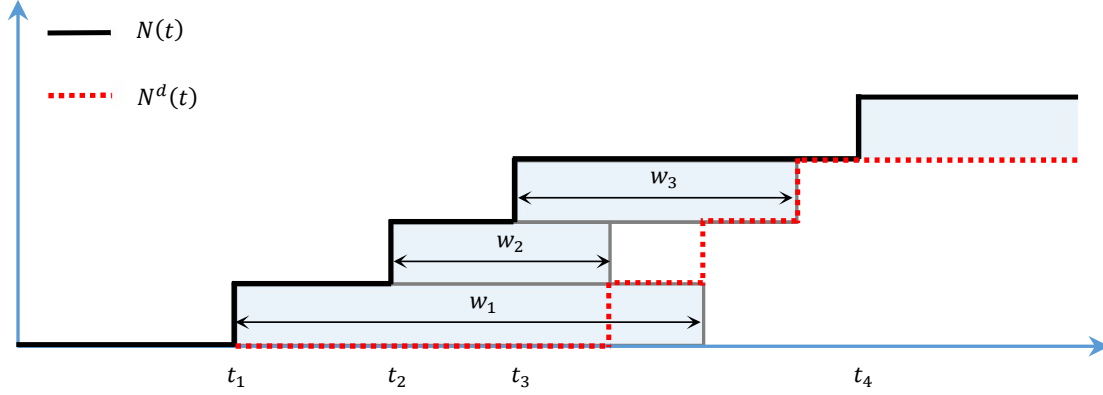
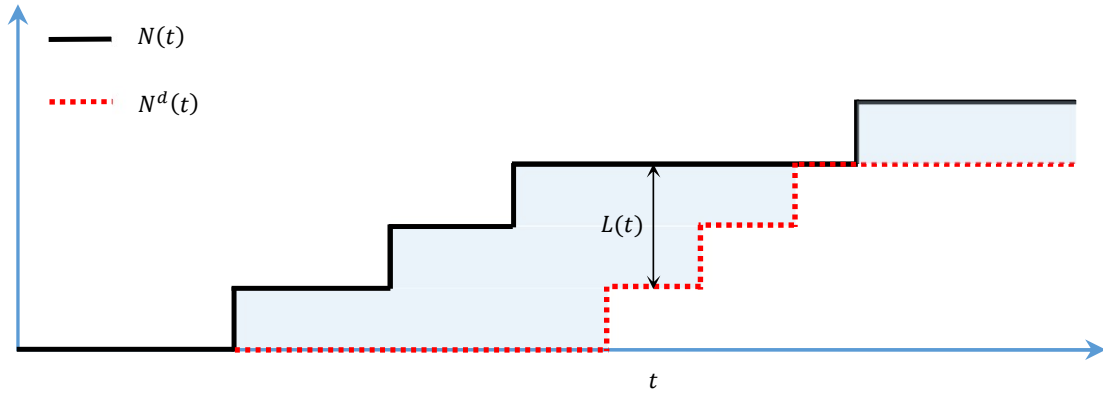


FIGURE 3. Arrival times, wait times, number of arrivals, and number of departures.

FIGURE 4. System size at time t

$$\lambda = \lim_{t \rightarrow \infty} \frac{N(t)}{t} = \text{Arrival rate} \quad (307)$$

$$w = \lim_{t \rightarrow \infty} \frac{1}{N(t)} \sum_{k=1}^{N(t)} w_k = \text{Average wait time.} \quad (308)$$

Little's law gives a very simple relation between the above three average quantities.

Theorem 3.1 (Little's law). *If both λ and w exist and are finite, then so does ℓ and*

$$\ell = \lambda w. \quad (309)$$

PROOF. Note that

$$\text{The } k\text{th customer is in the queue at time } t \iff t_k \leq t < t_k + w_k. \quad (310)$$

Hence we may write

$$L(t) = \sum_{k=1}^{\infty} \mathbf{1}(t_k \leq t < t_k + w_k). \quad (311)$$

Thus by Fubini's theorem,

$$\int_0^T L(t) dt = \int_0^T \sum_{k=1}^{\infty} \mathbf{1}(t_k \leq t < t_k + w_k) dt \quad (312)$$

$$= \sum_{k=1}^{\infty} \int_0^T \mathbf{1}(t_k \leq t < t_k + w_k) dt \quad (313)$$

$$= \sum_{k=1}^{N(T)} \min(T, t_k + w_k) - t_k. \quad (314)$$

This yields

$$\sum_{k=1}^{N^d(T)} w_k \leq \int_0^T L(t) dt \leq \sum_{k=1}^{N(T)} w_k. \quad (315)$$

Now observe that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^{N(T)} w_k = \lim_{T \rightarrow \infty} \frac{N(T)}{T} \left(\frac{1}{N(T)} \sum_{k=1}^{N(T)} w_k \right) = \lambda w. \quad (316)$$

On the other hand, using Exercise 3.2 we also have

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{k=1}^{N^d(T)} w_k = \lim_{T \rightarrow \infty} \frac{N^d(T)}{T} \left(\frac{1}{N^d(T)} \sum_{k=1}^{N^d(T)} w_k \right) = \lambda w. \quad (317)$$

Hence the assertion follows from (315). $\square$

Exercise 3.2. Let $(t_k)_{k \geq 0}$ be a sequence of arrival times and let w_k be the time that the k th customer spends in the system. Also, let $N(t)$ and $N^d(t)$ denote the number of arrivals and departures up to time t . Suppose the following limit exist and finite:

$$\lambda := \lim_{t \rightarrow \infty} \frac{N(t)}{t}, \quad w := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n w_k. \quad (318)$$

We will show that

$$\lim_{T \rightarrow \infty} \frac{N^d(t)}{t} = \lambda. \quad (319)$$

(i) Write

$$\frac{w_n}{n} = \left(\frac{1}{n} \sum_{k=1}^n w_k \right) - \frac{n-1}{n} \left(\frac{1}{n-1} \sum_{k=1}^{n-1} w_k \right). \quad (320)$$

Deduce that $\lim_{n \rightarrow \infty} w_n/n = 0$. Hence for each $\delta > 0$, there exists $M(\delta) > 0$ such that

$$w_n \leq \delta n \quad \forall n \geq M(\delta). \quad (321)$$

(ii) Show that for each $\varepsilon > 0$, there exists $T_1(\varepsilon) > 0$ such that for all $t > T_1(\varepsilon)$,

$$(\lambda - \varepsilon)\varepsilon t \leq N(\varepsilon t) \leq (\lambda + \varepsilon)\varepsilon t, \quad (322)$$

$$(\lambda - \varepsilon)(1 - \varepsilon)t \leq N((1 - \varepsilon)t) \leq (\lambda + \varepsilon)(1 - \varepsilon)t. \quad (323)$$

Also deduce that for all $t > T_1(\varepsilon)$,

$$\varepsilon t \leq t_n < (1 - \varepsilon)t \implies (\lambda - \varepsilon)\varepsilon t \leq n \leq (\lambda + \varepsilon)(1 - \varepsilon)t. \quad (324)$$

(iii) Using (i) and (ii), show that

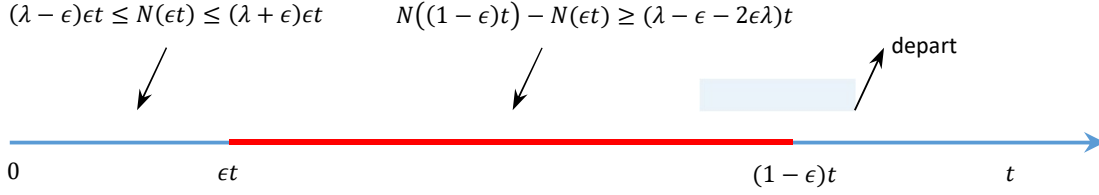
$$t > \max\left(T_1(\varepsilon), \frac{1}{(\lambda - \varepsilon)\varepsilon} M\left(\frac{\varepsilon}{2(\lambda + \varepsilon)(1 - \varepsilon)}\right)\right) \quad \text{and} \quad \varepsilon t \leq t_n < (1 - \varepsilon)t \implies w_n \leq (\varepsilon/2)t. \quad (325)$$

Deduce that for each t large enough, there are at least $(\lambda - \varepsilon - 2\varepsilon\lambda)t$ arrivals during $[\varepsilon t, (1 - \varepsilon)t]$ and they all depart by time t .

(iv) From (iii), show that for each $\varepsilon > 0$,

$$\liminf_{t \rightarrow \infty} \frac{N^d(t)}{t} \geq \lambda - \varepsilon - 2\varepsilon\lambda. \quad (326)$$

Finally, conclude (319).

FIGURE 5. All customers arriving during $[\epsilon t, (1 - \epsilon)t]$ departs by t

Example 3.3 (Housing Market). Suppose the local real estate in Westwood estimates that it takes 120 days on average to sell a house; This number does fluctuate with the economy and season, but it has been fairly stable over the past decade. We found out that at any given day last year, the number of houses for sale has ranged from 20 to 30, with average of 25. What can we say about the average number of transaction last year?

In order to apply Little's law, we view the housing market as a queuing system. Namely, we regard houses being put up for sale as an arrival to the system. The queue consists of unsold houses, and when houses are being sold, we regard them as exiting the queue. Now from the description above, we set the average wait time $w = 120$, and average queue length $\ell = 25$. Then by Little's law, we infer that the arrival rate is $\lambda = \ell / w = 25/120$ houses per day, or 75 houses per year. ▲

Exercise 3.4 (SLLN for weighted sum). Let $(X_k)_{k \geq 1}$ be a sequence of i.i.d. RVs with finite variance. Let $(w_k)_{k \geq 1}$ be a sequence of real numbers such that

$$\bar{w} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n w_k < \infty, \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n w_k^2 < \infty. \quad (327)$$

Define $S_n = \sum_{k=1}^n X_k w_k$. In this exercise, we will show that, almost surely,

$$\lim_{n \rightarrow \infty} S_n / n = \mathbb{E}[X_1] \bar{w}. \quad (328)$$

(i) Write

$$\frac{S_n}{n} = \frac{1}{n} \sum_{k=1}^n \mathbb{E}[X_k] w_k + \frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X_k]) w_k. \quad (329)$$

Show that it suffices to show the assertion assuming $\mathbb{E}[X_k] = 0$ for all $k \geq 1$. We may assume this for the following steps.

(i) Use Chebyshev's inequality to show that

$$\mathbb{P}(S_n \geq t) \leq t^{-2} \mathbb{E}[X_1^2] \left(\sum_{k=1}^n w_k^2 \right). \quad (330)$$

(ii) Use (i) to conclude that

$$\mathbb{P}\left(\frac{S_n}{n} \geq \frac{1}{\log n}\right) \leq \frac{(\log n)^2}{n} \mathbb{E}[X_1^2] \left(\frac{1}{n} \sum_{k=1}^n w_k^2 \right). \quad (331)$$

Deduce that

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\frac{S_n^2}{n^2} \geq \frac{1}{2 \log n}\right) < \infty. \quad (332)$$

By Borel-Cantelli Lemma, this yields

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \frac{S_n^2}{n^2} = 0\right) = 1. \quad (333)$$

(iii) Use (ii) to show that, for any sequence (n_k) of integers such that $n_k \rightarrow \infty$ as $k \rightarrow \infty$, we can choose a further subsequence $(n_{k(r)})$ such that

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{S_{n_{k(r)}}^2}{n_{k(r)}^2} = 0 \right) = 1. \quad (334)$$

This implies that, almost surely as $n \rightarrow \infty$,

$$0 = \liminf_{n \rightarrow \infty} \frac{S_n}{n} \leq \limsup_{n \rightarrow \infty} \frac{S_n}{n} = 0. \quad (335)$$

This implies $\lim_{n \rightarrow \infty} S_n/n = 0$ almost surely. This shows (328).

Exercise 3.5 (Expected load in the server). Consider a single-server queuing system, which is determined by the arrival times $(t_k)_{k \geq 1}$ and total times spent in the system $(w_k)_{k \geq 0}$ (a.k.a. sojourn times). Let $N(t)$ and $N^d(t)$ denote the number of arrivals and departures up to time t , respectively.

Each customer may wait for $\tilde{w}_k$ time in the queue, and then spends $\tilde{s}_k$ time the the server to get serviced, so that

$$w_k = \tilde{w}_k + \tilde{s}_k. \quad (336)$$

We may assume the following limits exist:

$$\lambda := \lim_{t \rightarrow \infty} \frac{N(t)}{t}, \quad w := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n w_k, \quad \tilde{s}^2 := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \tilde{s}_k^2 < \infty. \quad (337)$$

(i) Show that the k th customer is in the server if and only if

$$t_k + \tilde{w}_k \leq t < t_k + \tilde{w}_k + \tilde{s}_k. \quad (338)$$

(ii) Let $R(t)$ denote the remaining service time of the current customer in the server. Use (i) to show that

$$R(t) = \sum_{k=1}^{\infty} (t_k + \tilde{w}_k + \tilde{s}_k - t) \mathbf{1}(t_k + \tilde{w}_k \leq t < t_k + \tilde{w}_k + \tilde{s}_k). \quad (339)$$

(iii) Use Fubini's theorem to justify the following steps:

$$\int_0^T R(t) dt = \sum_{k=1}^{N(T)} \int_0^T (t_k + \tilde{w}_k + \tilde{s}_k - t) \mathbf{1}(t_k + \tilde{w}_k \leq t < t_k + \tilde{w}_k + \tilde{s}_k) dt \quad (340)$$

$$= \sum_{k=1}^{N(T)} \int_0^{\min(\tilde{s}_k, T - t_k - \tilde{w}_k)} (\tilde{s}_k - t) dt. \quad (341)$$

(iv) From (iii), deduce

$$\sum_{k=1}^{N^d(T)} \frac{\tilde{s}_k^2}{2} \leq \int_0^T R(t) dt \leq \sum_{k=1}^{N(T)} \frac{\tilde{s}_k^2}{2}. \quad (342)$$

Finally, derive the following formula for the average load in the server:

$$r := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T R(t) dt = \lambda \tilde{s}^2 / 2. \quad (343)$$

Exercise 3.6 (Pollaczek–Khinchine formula). Consider a $G/G/1$ queue, where arrivals are given by a renewal process of rate λ and service times are i.i.d. copies of a RV S with finite mean and variance. We use the following notations:

$$T_k = k\text{th arrival time} \quad (344)$$

$$\tilde{W}_k = \text{Time that the } k\text{th customer spends in the queue} \quad (345)$$

$$\tilde{S}_k = \text{Time that the } k\text{th customer spends in the server} \quad (346)$$

$$\tilde{W}(t) = \text{Remaining time until exit of the last customer in the queue at time } t. \quad (347)$$

Note that $\lim_{h \searrow 0} \tilde{W}(T_k - h)$ equals the waiting time $\tilde{W}_k$ of the k th customer in the queue. We will assume that the following limit exists almost surely:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \tilde{W}_k < \infty. \quad (348)$$

The goal of this exercise is to show the following Pollaczek–Khinchine formula for the mean waiting time: Almost surely,

$$\tilde{w} := \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \tilde{W}(s) ds = \frac{\lambda \mathbb{E}[S^2]}{2(1 - \lambda \mathbb{E}[S])}. \quad (349)$$

(i) Let $S(t)$ denote the sum of service times of all customers in the queue at time t . Show that

$$S(t) = \sum_{k=1}^{\infty} \tilde{S}_k \mathbf{1}(t_k \leq t < t_k + \tilde{W}_k). \quad (350)$$

(ii) Let $N(t)$ and $N^d(t)$ denote the number of arrivals and departures up to time t . Use Fubini's theorem to show that

$$\int_0^T S(t) dt = \sum_{k=1}^{\infty} \int_0^T \tilde{S}_k \mathbf{1}(t_k \leq t < t_k + \tilde{W}_k) dt \quad (351)$$

$$= \sum_{k=1}^{N(T)} \tilde{S}_k [\min(T, t_k + \tilde{W}_k) - t_k]. \quad (352)$$

Also, deduce that

$$\sum_{k=1}^{N^d(T)} \tilde{S}_k \tilde{W}_k \leq \int_0^T S(t) dt \leq \sum_{k=1}^{N(T)} \tilde{S}_k \tilde{W}_k. \quad (353)$$

(iii) From (ii) and Exercise 3.4, show that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T S(t) dt = \lambda \mathbb{E}[S] \tilde{w}. \quad (354)$$

(iv) Let $R(t)$ denote the remaining service time of the current customer in the server. Then $\tilde{W}(t) = R(t) + S(t)$. Using (iii) and Exercise (3.5), conclude the PK formula (349).

Martingales

1. Conditional expectation

Let X, Y be discrete RVs. Recall that the expectation $\mathbb{E}(X)$ is the ‘best guess’ on the value of X when we do not have any prior knowledge on X . But suppose we have observed that some possibly related RV Y takes value y . What should be our best guess on X , leveraging this added information? This is called the *conditional expectation of X given $Y = y$* , which is defined by

$$\mathbb{E}[X|Y = y] = \sum_x x \mathbb{P}(X = x|Y = y). \quad (355)$$

This best guess on X given $Y = y$, of course, depends on y . So it is a function in y . Now if we do not know what value Y might take, then we omit y and $\mathbb{E}[X|Y]$ becomes a RV, which is called the *conditional expectation of X given Y* .

Exercise 1.1. Let X, Y be discrete RVs. Show that for any function $g : \mathbb{R} \rightarrow \mathbb{R}$,

$$\mathbb{E}_X[Xg(Y)|Y] = g(Y)\mathbb{E}_X[X|Y]. \quad (356)$$

Exercise 1.2 (Iterated expectation). Let X, Y be discrete RVs. Use Fubini’s theorem to show that

$$\mathbb{E}[X] = \mathbb{E}_Y[\mathbb{E}_X[X|Y]]. \quad (357)$$

In case when X or Y are continuous RVs, we simply replace the sum by integral and PMF by PDF. For instance, if both X and Y are continuous with PDFs f_X, f_Y and joint PDF $f_{X,Y}$, then

$$\mathbb{E}[X|Y = y] = \int_{-\infty}^{\infty} x f_{X|Y=y}(x) dx, \quad (358)$$

where $f_{X|Y=y}$ is the conditional PDF of X given $Y = y$ defined by

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}. \quad (359)$$

To summarize how we compute the iterated expectations when we condition on discrete and continuous RV:

$$\mathbb{E}[\mathbb{E}[X|Y]] = \begin{cases} \sum_y \mathbb{E}[X|Y = y] \mathbb{P}(Y = y) & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathbb{E}[X|Y = y] f_Y(y) dy & \text{if } Y \text{ is continuous.} \end{cases} \quad (360)$$

An important use of iterated expectation is that we can compute probabilities using conditioning, since probability of an event is simply the expectation of the corresponding indicator variable.

Exercise 1.3 (Iterated expectation for probability). Let X, Y be RVs.

(i) For any $x \in \mathbb{R}$, show that $\mathbb{P}(X \leq x) = \mathbb{E}[\mathbf{1}(X \leq x)]$.

(ii) By using iterated expectation, show that

$$\mathbb{P}(X \leq x) = \mathbb{E}_Y[\mathbb{P}(X \leq x|Y)]. \quad (361)$$

In order to properly develop our discussion on martingales in the following sections, we need to generalize the notion of conditional expectation $\mathbb{E}[X|Y]$ of a RV X given another RV Y . Recall that this was the a collection of ‘best guesses’ of X given $Y = y$ for all y . But what if we only know, say, $Y \geq 1$? Can we condition on this event as well?

More concretely, suppose Y takes values from $\{1, 2, 3\}$. Regarding Y , the following outcomes are possible:

$$\mathcal{E}_Y := \{\{Y = 1\}, \{Y = 2\}, \{Y = 3\}, \{Y = 1, 2\}, \{Y = 2, 3\}, \{Y = 1, 3\}, \{Y = 1, 2, 3\}\}. \quad (362)$$

For instance, the information $\{Y = 1, 2\}$ could yield some nontrivial implication on the value of X , so our best guess in this scenario should be

$$\mathbb{E}[X | \{Y = 1, 2\}] = \sum_x x \mathbb{P}(X = x | \{Y = 1, 2\}). \quad (363)$$

More generally, for each $A \in \mathcal{E}_Y$, the best guess of X given $A \in \mathcal{E}_Y$ is the following conditional expectation

$$\mathbb{E}[X | A] = \sum_x x \mathbb{P}(X = x | A). \quad (364)$$

Now, what if we don't know which event in the collection $\mathcal{E}_Y$ to occur? As we did before to define $\mathbb{E}[X | Y]$ from $\mathbb{E}[X | Y = y]$ by simply not specifying what value y that Y takes, we simply do not specify which event $A \in \mathcal{E}_Y$ to occur. Namely,

$$\mathbb{E}[X | \mathcal{E}_Y] = \text{best guess on } X \text{ given the information in } \mathcal{E}_Y. \quad (365)$$

In general, this could be defined for any collection of events $\mathcal{E}$ in place of $\mathcal{E}_Y$. Mathematically, we understand $\mathbb{E}[X | \mathcal{E}]$ as¹

$$\mathbb{E}[X | \mathcal{E}] = \text{the collection of } \mathbb{E}[X | A] \text{ for all } A \in \mathcal{E}. \quad (366)$$

2. Definition and examples of martingales

Let $(X_t)_{t \geq 0}$ be the sequence of observations of the price of a particular stock over time. Suppose that an investor has a strategy to adjust his portfolio $(M_t)_{t \geq 0}$ according to the observation $(X_t)_{t \geq 0}$. Namely,

$$M_t = \text{Net value of portfolio after observing } (X_k)_{0 \leq k \leq t}. \quad (367)$$

We are interested in the long-term behavior of the 'portfolio process' $(M_t)_{t \geq 0}$. Martingales provide a very nice framework for this purpose.

Martingale is a class of stochastic processes, whose expected increment conditioned on the past is always zero. Recall that the simple symmetric random walk has this property, since each increment is i.i.d. and has mean zero. Martingales do not assume any kind of independence between increments, but it turns out that we can proceed quite far with just the unbiased conditional increment property.

In order to define martingales properly, we need to introduce the notion of 'information up to time t '. Imagine we are observing the stock market starting from time t . We define

$$\mathcal{E}_t := \text{collection of all possible events we can observe at time } t \quad (368)$$

$$\mathcal{F}_t := \bigcup_{k=1}^t \mathcal{E}_k = \text{collection of all possible events we can observe up to time } t. \quad (369)$$

In words, $\mathcal{E}_t$ is the information available at time t and $\mathcal{F}_t$ contains all possible information that we can obtain by observing the market up to time t . We call $\mathcal{F}_t$ the *information* up to time t . As a collection of events, $\mathcal{F}_t$ needs to satisfy the following properties²:

- (i) (*closed under complementation*) $A \in \mathcal{F}_t \implies A^c \in \mathcal{F}_t$;
- (ii) (*closed under countable union*) $A_1, A_2, A_3, \dots \in \mathcal{F}_t \implies \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}_t$.

¹For more details, see [Dur10].

²We are requiring $\mathcal{F}_t$ to be a σ -algebra, but we avoid using this terminology.

Note that as we gain more and more information, we have

$$\mathcal{F}_s \subseteq \mathcal{F}_t \quad \forall t \geq s \geq 0. \quad (370)$$

In other words, $(\mathcal{F}_t)_{t \geq 0}$ is an increasing set of information, which we call a *filtration*. The roll of a filtration is to specify what kind of information is observable or not, as time passes by.

Example 2.1. Suppose $(\mathcal{F}_t)_{t \geq 0}$ is a filtration generated by observing the stock price $(X_t)_{t \geq 0}$ of company A in New York. Namely, $\mathcal{E}_t$ consists of the information on the values of the stock price X_t at day t . Given $\mathcal{F}_{10}$, we know the actual values of $X_0, X_1, \dots, X_{10}$. For instance, X_8 is not random given $\mathcal{F}_{10}$, but X_{11} could still be random. On the other hand, if $(Y_t)_{t \geq 0}$ is the stock price of company B in Hong Kong, then we may have only partial information for $Y_0, \dots, Y_{10}$ given $\mathcal{F}_t$. ▲

Now we define martingales.

Definition 2.2. Let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration and $(M_t)_{t \geq 0}$ be discrete-time stochastic processes. We call $(M_t)_{t \geq 0}$ a *martingale* with respect to $(\mathcal{F}_t)_{t \geq 0}$ if the following conditions are satisfied: For all $t \geq 0$,

- (i) (*finite expectation*) $\mathbb{E}[|M_t|] < \infty$.
- (ii) (*measurability*³) $\{M_t = m\} \in \mathcal{F}_t$ for all $m \in \mathbb{R}$.
- (iii) (*conditional increments*) $\mathbb{E}[M_{t+1} - M_t | \mathcal{A}] = 0$ for all $\mathcal{A} \in \mathcal{F}_t$.

If (iii) holds with “=” replaced by “ $\leq$ ” (resp., “ $\geq$ ”), then $(M_t)_{t \geq 0}$ is called a *supermartingale* (resp., *submartingale*) with respect to $(\mathcal{F}_t)_{t \geq 0}$, respectively.

When appropriate, we will abbreviate the condition (iii) as

$$\mathbb{E}[M_{t+1} - M_t | \mathcal{F}_t] = 0. \quad (371)$$

If martingale is a fair gambling strategy, then one can think of supermartingale and submartingale as unfavorable and favorable gambling strategies, respectively. For instance, expected winning in gambling on an unfavorable game should be non-increasing in time. This is an immediate consequence of the definition and iterated expectation.

Proposition 2.3. Let $(M_t)_{t \geq 0}$ be a stochastic process and let $(\mathcal{F}_t)_{t \geq 0}$ be a filtration.

- (i) If $(M_t)_{t \geq 0}$ is a supermartingale w.r.t. filtration $(\mathcal{F}_t)_{t \geq 0}$, then $\mathbb{E}[M_n] \leq \mathbb{E}[M_m]$ for all $n \geq m \geq 0$.
- (ii) If $(M_t)_{t \geq 0}$ is a submartingale w.r.t. filtration $(\mathcal{F}_t)_{t \geq 0}$, then $\mathbb{E}[M_n] \geq \mathbb{E}[M_m]$ for all $n \geq m \geq 0$.
- (iii) If $(M_t)_{t \geq 0}$ is a martingale w.r.t. filtration $(\mathcal{F}_t)_{t \geq 0}$, then $\mathbb{E}[M_n] = \mathbb{E}[M_m]$ for all $n \geq m \geq 0$.

PROOF. (ii) and (iii) follows directly from (i), so we only show (i). Let $(M_t)_{t \geq 0}$ be a supermartingale w.r.t. $(\mathcal{F}_t)_{t \geq 0}$. Recall that for each $m \in \mathbb{R}$, $\{M_t = m\} \in \mathcal{F}_t$. Hence

$$\mathbb{E}[M_{t+1} - M_t | M_t = m] \leq 0 \quad (372)$$

since M_t is a supermartingale. Hence by conditioning on the values of M_t ,

$$\mathbb{E}[M_{t+1} - M_t] = \mathbb{E}[\mathbb{E}[M_{t+1} - M_t | M_t]] \leq 0. \quad (373)$$

Then the assertion follows by an induction. □

In order to get familiar with martingales, it is helpful to envision them as a kind of simple symmetric random walk. In general, one can subtract off the mean of a given random walk to make it a martingale.

Example 2.4 (Random walks). Let $(X_t)_{t \geq 1}$ be a sequence of i.i.d. increments with $\mathbb{E}[X_i] = \mu < \infty$. Let $S_t = S_0 + X_1 + \dots + X_t$. Then $(S_t)_{t \geq 0}$ is called a *random walk*. Define a stochastic process $(M_t)_{t \geq 0}$ by

$$M_t = S_t - \mu t. \quad (374)$$

³In this case, we say “ M_t is measurable w.r.t. $\mathcal{F}_t$ ”, but we avoid using this terminology.

For each $t \geq 0$, let $\mathcal{F}_t$ be the information obtained by observing $S_0, S_1, \dots, S_t$. Then $(M_t)_{t \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Indeed, we have

$$\mathbb{E}[|M_t|] = \mathbb{E}[|S_t - \mu t|] \leq \mathbb{E}[|S_t| + |\mu t|] = \mathbb{E}[|S_t|] + \mu t < \infty, \quad (375)$$

and for any $m \in \mathbb{R}$,

$$\{M_t = m\} = \{S_t - \mu t = m\} = \{S_t = m + \mu t\} \in \mathcal{F}_t. \quad (376)$$

Furthermore, Since X_{t+1} is independent from $S_0, \dots, S_t$, it is also independent from any $A \in \mathcal{F}_t$. Hence

$$\mathbb{E}[M_{t+1} - M_t | A] = \mathbb{E}[X_{t+1} - \mu | A] \quad (377)$$

$$= \mathbb{E}[X_{t+1} - \mu] = \mathbb{E}[X_{t+1}] - \mu = 0. \quad (378)$$

▲

Example 2.5 (Products of indep. RVs). Let $(X_t)_{t \geq 0}$ be a sequence of independent RVs such that $X_t \geq 0$ and $\mathbb{E}[X_t] = 1$ for all $t \geq 0$. For each $t \geq 0$, let $\mathcal{F}_t$ be the information obtained by observing $M_0, X_0, \dots, X_t$. Define

$$M_t = M_0 X_1 X_2 \cdots X_t, \quad (379)$$

where M_0 is a constant. Then $(M_t)_{t \geq 0}$ is a martingale with respect to $(\mathcal{F}_t)_{t \geq 0}$. Indeed, the assumption implies $\mathbb{E}[|M_t|] < \infty$ and that $\{M_t = m\} \in \mathcal{F}_t$ for all $m \in \mathbb{R}$ since M_t is determined by $M_0, X_1, \dots, X_t$. Furthermore, since X_{t+1} is independent from $X_1, \dots, X_t$, for each $A \in \mathcal{F}_t$,

$$\mathbb{E}[M_{t+1} - M_t | A] = \mathbb{E}[M_t X_{t+1} - M_t | A] \quad (380)$$

$$= \mathbb{E}[(X_{t+1} - 1)(M_0 X_1 \cdots X_t) | A] \quad (381)$$

$$= \mathbb{E}[X_{t+1} - 1 | A] \mathbb{E}[(M_0 X_1 \cdots X_t) | A] \quad (382)$$

$$= \mathbb{E}[X_{t+1} - 1] \mathbb{E}[(M_0 X_1 \cdots X_t) | A] = 0. \quad (383)$$

This multiplicative model is reasonable for the stock market since the changes in stock prices are believed to be proportional to the current stock price. Moreover, it also guarantees that the price will stay positive, in comparison to additive models. ▲

Exercise 2.6 (Exponential martingale). Let $(X_t)_{t \geq 0}$ be a sequence of i.i.d. RVs such that their moment generating function exists, namely, there exists $\theta > 0$ for which

$$\varphi(\theta) := \mathbb{E}[\exp(\theta X_k)] < \infty \quad \forall k \geq 0. \quad (384)$$

Let $S_t = S_0 + X_1 + \cdots + X_t$. Define

$$M_t = \exp(\theta S_t) / \varphi(\theta)^t. \quad (385)$$

Show that $(M_t)_{t \geq 0}$ is a martingale with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t$ is the information obtained by observing $S_0, \dots, S_t$.

The following lemma allows us to provide many examples of martingales from Markov chains.

Lemma 2.7. Let $(X_t)_{t \geq 0}$ be a Markov chain on state space Ω with transition matrix P . For each $t \geq 0$, let f_t be a function $\Omega \rightarrow \mathbb{R}$ such that

$$f_t(x) = \sum_{y \in \Omega} P(x, y) f_{t+1}(y) \quad \forall x \in \Omega. \quad (386)$$

Then $M_t = f_t(X_t)$ defines a martingale with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t$ is the information obtained by observing $X_0, \dots, X_t$.

PROOF. First note that for any $x \in \Omega$,

$$\mathbb{E}[M_{t+1} - M_t | X_t = x] = \mathbb{E} \left[\left(\sum_{y \in \Omega} P(x, y) f_{t+1}(y) \right) - f_t(x) \right] = 0. \quad (387)$$

Now fix $A \in \mathcal{F}_t$. By conditioning on the value of X_t and using Markov property,

$$\mathbb{E}[M_{t+1} - M_t | A] = \mathbb{E}[f_{t+1}(X_{t+1}) - f_t(X_t) | A] \quad (388)$$

$$= \mathbb{E}[\mathbb{E}[f_{t+1}(X_{t+1}) - f_t(X_t) | A, X_t] | A] \quad (389)$$

$$= \mathbb{E}[\mathbb{E}[f_{t+1}(X_{t+1}) - f_t(X_t) | X_t] | A] = 0. \quad (390)$$

This shows the assertion. $\square$

Example 2.8 (Simple random walk). Let $(X_t)_{t \geq 1}$ be a sequence of i.i.d. RVs with

$$\mathbb{P}(X_k = 1) = p, \quad \mathbb{P}(X_k = -1) = 1 - p. \quad (391)$$

Let $S_t = S_0 + X_1 + \cdots + X_t$. Note that $(S_t)_{t \geq 0}$ is a Markov chain on $\mathbb{Z}$. Define

$$M_t = \left(\frac{1-p}{p} \right)^{S_t}. \quad (392)$$

Then $(M_t)_{t \geq 0}$ is a martingale with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t$ is the information obtained by observing $S_0, \dots, S_t$.

In order to see this, define a function $h_t(x) = ((1-p)/p)^x$. According to Lemma 2.7, it suffice to show that h is a harmonic function with respect to the Gambler's chain. Namely,

$$\sum_{y \in \mathbb{Z}} P(x, y) h_t(y) = p h(x+1) + (1-p) h(x-1) \quad (393)$$

$$= p \left(\frac{1-p}{p} \right)^{x+1} + (1-p) \left(\frac{1-p}{p} \right)^{x-1} \quad (394)$$

$$= (1-p) \left(\frac{1-p}{p} \right)^x + p \left(\frac{1-p}{p} \right)^x = \left(\frac{1-p}{p} \right)^x = h_t(x). \quad (395)$$

Hence by Lemma 2.7, $(M_t)_{t \geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. $\blacktriangle$

Example 2.9 (Simple symmetric random walk). Let $(X_t)_{t \geq 1}$ be a sequence of i.i.d. RVs with

$$\mathbb{P}(X_k = 1) = \mathbb{P}(X_k = -1) = 1/2. \quad (396)$$

Let $S_t = S_0 + X_1 + \cdots + X_t$. Note that $(S_t)_{t \geq 0}$ is a Markov chain on $\mathbb{Z}$. Define

$$M_t = S_t^2 - t. \quad (397)$$

Then $(M_t)_{t \geq 0}$ is a martingale with respect to $(S_t)_{t \geq 0}$.

For each $t \geq 0$, define a function $f_t: \mathbb{Z} \rightarrow \mathbb{R}$ by $f_t(x) = x^2 - t$. By Lemma 2.7, it suffices to check if $f_t(x)$ is the average of $f_{t+1}(y)$ with respect to the transition matrix of S_t . Namely,

$$\sum_{y \in \mathbb{Z}} P(x, y) f_{t+1}(y) = \frac{1}{2} f_{t+1}(x+1) + \frac{1}{2} f_{t+1}(x-1) \quad (398)$$

$$= \frac{(x+1)^2 - (t+1)}{2} + \frac{(x-1)^2 - (t+1)}{2} = x^2 - t = f_t(x). \quad (399)$$

Hence by Lemma 2.7, $(M_t)_{t \geq 0}$ is a martingale with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_t$ is the information obtained by observing $S_0, \dots, S_t$. $\blacktriangle$

3. Basic properties of martingales

In this section, we study some of the basic properties of martingales. We begin with the relation between martingale increments and convex functions. Recall that a function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is *convex* if

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)(\varphi(y)), \quad \forall \lambda \in [0, 1] \text{ and } x, y \in \mathbb{R}. \quad (400)$$

For instance, $\varphi(x) = x^2$ and $\exp(x)$ are convex functions.

Exercise 3.1 (Jensen's inequality). Let X be any RV with $\mathbb{E}[X] < \infty$. Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be any convex function, that is,

$$\varphi(\lambda x + (1 - \lambda)y) \leq \lambda\varphi(x) + (1 - \lambda)(\varphi(y)), \quad \forall \lambda \in [0, 1] \text{ and } x, y \in \mathbb{R}. \quad (401)$$

Jensen's inequality states that

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]. \quad (402)$$

(i) Let $c := \mathbb{E}[X] < \infty$. Show that there exists a line $f(x) = ax + b$ such that $f(c) = \varphi(c)$ and $\varphi(x) \geq f(x)$ for all $x \in \mathbb{R}$.

(ii) Verify the following and prove Jensen's inequality:

$$\mathbb{E}[\varphi(X)] \geq \mathbb{E}[f(X)] = a\mathbb{E}[X] + b = f(c) = \varphi(c) = \varphi(\mathbb{E}[X]). \quad (403)$$

(iii) Let X be RV, A an event, φ be the convex function as before. Show the Jensen's inequality for the conditional expectation:

$$\varphi(\mathbb{E}[X|A]) \leq \mathbb{E}[\varphi(X)|A]. \quad (404)$$

Proposition 3.2. Let $(M_t)_{t \geq 0}$ be a submartingale with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$. Let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function. Suppose that $\mathbb{E}[|\varphi(M_t)|] < \infty$ for all $t \geq 0$.

(i) If $(M_t)_{t \geq 0}$ is a martingale, then $(\varphi(M_t))_{t \geq 0}$ is a submartingale w.r.t. $(\mathcal{F}_t)_{t \geq 0}$.

(ii) If $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is a non-decreasing function, then $(\varphi(M_t))_{t \geq 0}$ is a submartingale w.r.t. $(\mathcal{F}_t)_{t \geq 0}$.

PROOF. We first show (i). Fix $A \in \mathcal{F}_t$. Since $(M_t)_{t \geq 0}$ is a martingale, for each $A' \in \mathcal{F}_t$,

$$\mathbb{E}[M_{t+1} | A'] - \mathbb{E}[M_t | A'] = \mathbb{E}[M_{t+1} - M_t | A'] = 0. \quad (405)$$

Note that for each $m \in \mathbb{R}$, since $\{M_t = m\} \in \mathcal{F}_t$, by using Jensen's inequality,

$$\mathbb{E}[\varphi(M_{t+1}) - \varphi(M_t) | A \cap \{M_t = m\}] = \mathbb{E}[\varphi(M_{t+1}) | A \cap \{M_t = m\}] - \varphi(m) \quad (406)$$

$$\geq \varphi(\mathbb{E}[M_{t+1} | A \cap \{M_t = m\}]) - \varphi(m) \quad (407)$$

$$= \varphi(\mathbb{E}[M_t | A \cap \{M_t = m\}]) - \varphi(m) = \varphi(m) - \varphi(m) = 0. \quad (408)$$

Then by iterated expectation and using the fact that $A \cap \{M_t = m\} \in \mathcal{F}_t$,

$$\mathbb{E}[\varphi(M_{t+1}) - \varphi(M_t) | A] = \mathbb{E}[\mathbb{E}[\varphi(M_{t+1}) - \varphi(M_t) | A, M_t]] = 0. \quad (409)$$

So $(\varphi(M_t))_{t \geq 0}$ is a submartingale. This shows (i).

To show (ii), note that since $(M_t)_{t \geq 0}$ is a submartingale, for any $A' \in \mathcal{F}_t$,

$$\mathbb{E}[M_{t+1} | A'] - \mathbb{E}[M_t | A'] = \mathbb{E}[M_{t+1} - M_t | A'] \geq 0. \quad (410)$$

By Jensen's inequality and since φ is non-decreasing,

$$\mathbb{E}[\varphi(M_{t+1}) | A'] \geq \varphi(\mathbb{E}[M_{t+1} | A']) \geq \varphi(\mathbb{E}[M_t | A']). \quad (411)$$

Then following the similar argument as before shows that $(\varphi(M_t))_{t \geq 0}$ is a submartingale. $\square$

Example 3.3. Let $(M_t)_{t \geq 0}$ be a martingale. Since $\varphi(x) = x^2$ is a convex function, $(M_t^2)_{t \geq 0}$ is a submartingale. $\blacktriangle$

The following observation is a martingale analogue of the formula $\mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{Var}(X)$.

Proposition 3.4. *If $(M_t)_{t \geq 0}$ is a martingale with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$, then we have*

$$\mathbb{E}[M_{t+1}^2 - M_t^2 | \mathcal{F}_t] = \mathbb{E}[(M_{t+1} - M_t)^2 | \mathcal{F}_t] \geq 0. \quad (412)$$

PROOF. By the martingale condition, we have

$$\mathbb{E}[M_{t+1} - M_t | \mathcal{F}_t] = 0 \quad (413)$$

Hence by expanding the square, we get

$$\mathbb{E}[(M_{t+1} - M_t)^2 | \mathcal{F}_t] = \mathbb{E}[M_{t+1}^2 | \mathcal{F}_t] - 2M_t \mathbb{E}[M_{t+1} | \mathcal{F}_t] + \mathbb{E}[M_t^2 | \mathcal{F}_t] \quad (414)$$

$$= \mathbb{E}[M_{t+1}^2 | \mathcal{F}_t] - \mathbb{E}[M_t^2 | \mathcal{F}_t]. \quad (415)$$

□

Exercise 3.5 (Long range martingale condition). Let $(M_t)_{t \geq 0}$ be a martingale with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$. For any $0 \leq k < n$, we will show that

$$\mathbb{E}[(M_n - M_k) | \mathcal{F}_k] = 0. \quad (416)$$

(i) Suppose (416) holds for fixed $0 \leq k < n$. For each $A \in \mathcal{F}_k$, to show that

$$\mathbb{E}[M_{n+1} - M_k | A] = \mathbb{E}[M_{n+1} - M_n | A] + \mathbb{E}[M_n - M_k | A] \quad (417)$$

$$= \mathbb{E}[M_{n+1} - M_n | A] = 0. \quad (418)$$

(ii) Conclude (416) for all $0 \leq k < n$ by induction.

Proposition 3.6 (Orthogonality of martingale increments). *Let $(M_t)_{t \geq 0}$ be a martingale with respect to filtration $(\mathcal{F}_t)_{t \geq 0}$.*

(i) *For each $0 \leq j \leq k < n$, we have*

$$\mathbb{E}[(M_n - M_k)M_j] = 0. \quad (419)$$

(ii) *For each $0 \leq j \leq k < n$, we have*

$$\mathbb{E}[(M_n - M_k)(M_j - M_i)] = 0. \quad (420)$$

PROOF. To show (i), we use conditioning on the values of M_j . By noting that $\{M_j = m\} \in \mathcal{F}_j \subseteq \mathcal{F}_k$ for all $m \in \mathbb{R}$, we have

$$\mathbb{E}[(M_n - M_k)M_j | M_j = m] = m\mathbb{E}[M_n - M_k | M_j = m] = 0, \quad (421)$$

where we have used Exercise 3.5 for the last equality. Hence

$$\mathbb{E}[(M_n - M_k)M_j] = \mathbb{E}[\mathbb{E}[(M_n - M_k)M_j | M_j]] = 0. \quad (422)$$

Moreover, (ii) follows from (i) immediately since

$$\mathbb{E}[(M_n - M_k)(M_j - M_i)] = \mathbb{E}[(M_n - M_k)M_j] - \mathbb{E}[(M_n - M_k)M_i]. \quad (423)$$

This shows the assertion. □

Exercise 3.7 (Pythagorean theorem for martingales). Let $(M_t)_{t \geq 0}$ be a martingale with respect to a filtration $(\mathcal{F}_t)_{t \geq 0}$. Use Proposition 3.6 to show that

$$\mathbb{E}[(M_t - M_0)^2] = \sum_{k=1}^t \mathbb{E}[(M_k - M_{k-1})^2]. \quad (424)$$

4. Gambling strategies and stopping times

In this section, we derive some properties of martingales in relation to stopping times viewed as gambling strategies. The following observation is natural if we think of supermartingale as a random walk with negative drift; or as betting on an unfavorable game.

Next, we prove one of the most famous result in martingale theory, which says

$$\text{“You can’t beat an unfavorable game.”} \quad (425)$$

To formulate this, let $(M_t)_{t \geq 0}$ be a supermartingale w.r.t. $(\mathcal{F}_t)_{t \geq 0}$. Think of $(X_t)_{t \geq 0}$ as the progression of stock market, and $(M_t)_{t \geq 0}$ is the price of the stock of company A. Say we have an investment strategy so that we can determine

$$H_{t+1} = \text{Amount of stock of company A that we hold between time } t \text{ and } t+1 \quad (426)$$

based on the observation $M_0, X_0, \dots, X_t$. Then the stock price will change from M_t to M_{t+1} ,

$$H_{t+1}(M_{t+1} - M_t) = \text{Net gain occurred between time } t \text{ and } t+1. \quad (427)$$

Hence if we let W_0 denote the initial fortune, then

$$W_t := W_0 + \sum_{k=1}^t H_k(M_k - M_{k-1}) = \text{Total fortune at time } t. \quad (428)$$

The following theorem tells us that, if $(M_t)_{t \geq 0}$ is declining stock on average, then no matter what strategy $(H_t)_{t \geq 0}$ we use, we will always lose our fortune.

Theorem 4.1. *Let $(M_t)_{t \geq 0}$ be a supermartingale w.r.t. a filtration $(\mathcal{F}_t)_{t \geq 0}$. Suppose $(H_t)_{t \geq 0}$ is such that*

- (i) *(Predictability) $\{H_{t+1} = h\} \in \mathcal{F}_t$ for all $h \in \mathbb{R}$.*
- (ii) *(Boundedness) $0 \leq H_t \leq c_t$ for some constant $c_t \geq 0$ for all $t \geq 0$.*

Let $(W_t)_{t \geq 0}$ be as defined at (428). Then $(W_t)_{t \geq 0}$ is a supermartingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$.

PROOF. By (ii) and a triangle inequality,

$$|W_t| \leq |W_0| + \sum_{k=1}^t c_k(|M_k| + |M_{k-1}|). \quad (429)$$

Taking expectation shows $\mathbb{E}[|W_t|] < \infty$. Moreover, for each $w \in \mathbb{R}$,

$$\{W_t = w\} = \left\{ W_0 + \sum_{k=1}^t H_k(M_k - M_{k-1}) = w \right\} \in \mathcal{F}_t \quad (430)$$

since the event on the right hand side can be written as the union of suitable events involving the values of $H_0, \dots, H_t$ and $M_0, \dots, M_t$, which are all events belonging to $\mathcal{F}_t$.

Lastly, fix $A \in \mathcal{F}_t$. For each $h \in [0, c_t]$, since $\{H_{t+1} = h\} \in \mathcal{F}_t$, we also have $A \cap \{H_{t+1} = h\} \in \mathcal{F}_t$. Since M_t is a supermartingale, we have

$$\mathbb{E}[W_{t+1} - W_t | A \cap \{H_{t+1} = h\}] = \mathbb{E}[H_{t+1}(M_{t+1} - M_t) | A \cap \{H_{t+1} = h\}] \quad (431)$$

$$= h \mathbb{E}[M_{t+1} - M_t | A] \leq 0. \quad (432)$$

Hence, by conditioning on the values of H_{t+1} , we get

$$\mathbb{E}[W_{t+1} - W_t | A] = \mathbb{E}[\mathbb{E}[W_{t+1} - W_t | A, H_{t+1}] | A] \leq 0. \quad (433)$$

This shows the assertion. $\square$

Next, we define stopping times with respect to a given filtration. This generalizes the version of stopping times we introduced for Poisson processes (See Def. 3.3 in Lecture note 3). The definition is as you would expect:

Definition 4.2. A RV $T \geq 0$ is a *stopping time* w.r.t. a filtration $(\mathcal{F}_t)_{t \geq 0}$ if

$$\{T = s\}, \{T \neq s\} \in \mathcal{F}_s \quad \forall s \geq 0. \quad (434)$$

Example 4.3. Let $(N_1(t))_{t \geq 0}$ and $(N_2(t))_{t \geq 0}$ be the counting processes of $\text{PP}(\lambda_1)$ and $\text{PP}(\lambda_2)$ (not necessarily indep.). For each $t \geq 0$, let $\mathcal{F}_t$ be the information obtained by observing $(N_i(t))_{0 \leq s \leq t}$ for $i = 1, 2$. Let $T_k^{(i)}$ denote the k th arrival time for the i th process.

Clearly, for each $i \in \{1, 2\}$ and $k \geq 0$, $T_k^{(i)}$ is a stopping time w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$. Namely, for each $t \geq 0$,

$$\{T_k^{(i)} = t\} = \{N^{(i)}(t) = k, N^{(i)}(t^-) = k - 1\} \in \mathcal{F}_t. \quad (435)$$

Also, let

$$\tilde{T}_k := k\text{th arrival time for the merged process } N_1 + N_2. \quad (436)$$

This is also a stopping time w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$. Namely, for each $t \geq 0$,

$$\{\tilde{T}_k = t\} = \{N^{(1)}(t) + N^{(2)}(t) = k, N^{(1)}(t^-) + N^{(2)}(t^-) = k - 1\} \in \mathcal{F}_t. \quad (437)$$

On the other hand, let $(N^{(3)}(t))_{t \geq 0}$ be the counting process of another Poisson process of rate λ_3 , which is independent from the other processes. Let $T_k^{(3)}$ denote the k th arrival time of this process. Is this a stopping time w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$? No, since for each $t \geq 0$, the event

$$\{T_k^{(3)} = t\} = \{N^{(3)}(t) = k, N^{(3)}(t^-) = k - 1\} \quad (438)$$

cannot be determined from observing the other two processes $N^{(1)}$ and $N^{(2)}$ up to time t . $\blacktriangle$

Example 4.4 (Constant betting up to a stopping time). Let $(M_t)_{t \geq 0}$ be a supermartingale w.r.t. $(X_t)_{t \geq 0}$. Let T be a stopping time for $(X_t)_{t \geq 0}$. Consider the gambling strategy of betting \$1 up to time T . That is,

$$H_k = \begin{cases} 1 & \text{if } k \leq T \\ 0 & \text{if } k > T \end{cases}. \quad (439)$$

Since T is a stopping time, the above H_k defines a predictable sequence as in Theorem 4.1. Then the wealth at time t is

$$W_t = W_0 + \sum_{k=1}^t H_k(M_k - M_{k-1}) = W_0 + M_{T \wedge t} - M_0, \quad (440)$$

where $T \wedge t = \min(T, t)$. The above relation can be easily verified by considering whether $t > T$ or $t \leq T$. According to Theorem 4.1, we know that W_t is a supermartingale. It follows that $M_{T \wedge t}$ is also a supermartingale. $\blacktriangle$

Theorem 4.5. Let $(M_t)_{t \geq 0}$ be a supermartingale w.r.t. a filtration $(\mathcal{F}_t)_{t \geq 0}$. Let T be a stopping time for $(\mathcal{F}_t)_{t \geq 0}$. Then $M_{T \wedge t}$ is a supermartingale w.r.t. $(\mathcal{F}_t)_{t \geq 0}$. Furthermore,

$$\mathbb{E}[M_{T \wedge t}] \leq \mathbb{E}[M_0] \quad \forall t \geq 0. \quad (441)$$

In particular, if $(M_t)_{t \geq 0}$ is a martingale, then

$$\mathbb{E}[M_{T \wedge t}] = \mathbb{E}[M_0] \quad \forall t \geq 0. \quad (442)$$

PROOF. Let W_t be as defined as in Example (4.4). Since $W_t = W_0 + M_{T \wedge t} - M_0$ and since W_t is a supermartingale by Theorem 4.1, and since W_0 and M_0 are constants, it follows that $M_{T \wedge t}$ is also a supermartingale. Then by Proposition (2.3), we have

$$\mathbb{E}[M_{T \wedge t}] \leq \mathbb{E}[W_t] \leq \mathbb{E}[W_0] = \mathbb{E}[M_0]. \quad (443)$$

The same argument holds for the submartingale case with the reversed inequalities. Then the martingale case follows. $\square$

Exercise 4.6. Let $(M_t)_{t \geq 0}$ be a supermartingale w.r.t. a filtration $(\mathcal{F}_t)_{t \geq 0}$. We will directly show that $(M_{T \wedge t})_{t \geq 0}$ is a supermartingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$.

(i) For each fixed $A \in \mathcal{F}_t$, show that

$$A \cap \{T \leq t\} = A \cap \left(\bigcup_{k=1}^t \{T = k\} \right) \in \mathcal{F}_t, \quad (444)$$

$$A \cap \{T \geq t+1\} = A \cap \left(\bigcap_{k=1}^t \{T \neq k\} \right) \in \mathcal{F}_t. \quad (445)$$

(ii) Use conditioning on whether $T \leq t$ or not to show

$$\mathbb{E}[M_{T \wedge (t+1)} - M_{T \wedge t} | A] = \mathbb{E}[M_T - M_t | A \cap \{T \leq t\}] \mathbb{P}(T \leq t) \quad (446)$$

$$+ \mathbb{E}[M_{t+1} - M_t | A \cap \{T \geq t+1\}] \mathbb{P}(T \geq t+1) \quad (447)$$

$$\leq 0. \quad (448)$$

Conclude that $(M_{T \wedge t})_{t \geq 0}$ is a supermartingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$.

5. Applications of martingales at stopping times

If a martingale is bounded up to a stopping time, then the expected value of the martingale at stopping time equals the initial expectation. This observation will be useful in some applications.

Proposition 5.1. *Let $(Y_t)_{t \geq 0}$ be a stochastic process such that $\mathbb{E}[|Y_t|] < \infty$ for all $t \geq 0$. Let $T \geq 0$ be a RV. Suppose*

$$\mathbb{P}(T < \infty) = 1, \quad |\mathbb{E}[Y_T | T > t]| < \infty, \quad |\mathbb{E}[Y_{T \wedge t} | T > t]| < \infty. \quad (449)$$

Then

$$\mathbb{E}[Y_T] = \lim_{t \rightarrow \infty} \mathbb{E}[Y_{T \wedge t}]. \quad (450)$$

PROOF. We first condition on whether $T \leq t$ or $T > t$ to write

$$\mathbb{E}[Y_T] = \mathbb{E}[Y_T | T \leq t] \mathbb{P}(T \leq t) + \mathbb{E}[Y_T | T > t] \mathbb{P}(T > t) \quad (451)$$

$$= \mathbb{E}[Y_{T \wedge t} | T \leq t] \mathbb{P}(T \leq t) + \mathbb{E}[Y_T | T > t] \mathbb{P}(T > t). \quad (452)$$

Similarly, we also write

$$\mathbb{E}[Y_{T \wedge t}] = \mathbb{E}[Y_{T \wedge t} | T \leq t] \mathbb{P}(T \leq t) + \mathbb{E}[Y_{T \wedge t} | T > t] \mathbb{P}(T > t). \quad (453)$$

Subtracting these two equations, we get

$$|\mathbb{E}[Y_T] - \mathbb{E}[Y_{T \wedge t}]| \leq |\mathbb{E}[Y_T | T > t]| \mathbb{P}(T > t) + |\mathbb{E}[Y_{T \wedge t} | T > t]| \mathbb{P}(T > t). \quad (454)$$

Note that since $\mathbb{P}(T \leq t) \rightarrow \mathbb{P}(T < \infty) = 1$ as $t \rightarrow \infty$, $\mathbb{P}(T > t) \rightarrow 0$ as $t \rightarrow \infty$. Then by the hypothesis, the right hand side converges to zero as $t \rightarrow \infty$. This shows the assertion. $\square$

An immediate consequence for martingales is the following:

Proposition 5.2. *Suppose $(M_t)_{t \geq 0}$ is a martingale and T is a stopping time.*

$$\mathbb{P}(T < \infty) = 1, \quad |\mathbb{E}[M_T | T > t]| < \infty, \quad |\mathbb{E}[M_{T \wedge t} | T > t]| < \infty. \quad (455)$$

Then

$$\mathbb{E}[M_T] = \mathbb{E}[M_0]. \quad (456)$$

PROOF. By Theorem 4.5 and Proposition 5.1,

$$\mathbb{E}[M_0] = \lim_{t \rightarrow \infty} \mathbb{E}[M_0] = \lim_{t \rightarrow \infty} \mathbb{E}[M_{T \wedge t}] = \mathbb{E}[M_T]. \quad (457)$$

$\square$

Example 5.3 (Gambler's ruin). Let $(X_t)_{t \geq 1}$ be i.i.d. RVs with

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = -1) = 1 - p. \quad (458)$$

Let $S_t = S_0 + X_1 + \cdots + X_t$ with $S_0 = i$. Let $\mathcal{F}_t$ be the information obtained by observing $S_0, \dots, S_t$. Let $h(x) = ((1-p)/p)^x$ as in Exercise 2.8. We have seen $M_t := h(S_t)$ is a martingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$.

Fix integers $a < S_0 < b$, and define

$$T = \min\{t \geq 0 \mid S_t = a \text{ or } S_t = b\}. \quad (459)$$

So T is the first time that the random walk hits a or b . Note that T is a stopping time, and since S_t is an irreducible Markov chain on $\mathbb{Z}$, its hitting time to any state is almost surely finite, so $\mathbb{P}(T < \infty) = 1$. Moreover, $S_T, S_{T \wedge t} \in [a, b]$. Hence by Proposition 5.2, we have

$$\left(\frac{1-p}{p}\right)^i = \mathbb{E}[h(S_T)] = \left(\frac{1-p}{p}\right)^a \mathbb{P}(S_T = a) + \left(\frac{1-p}{p}\right)^b (1 - \mathbb{P}(S_T = a)). \quad (460)$$

Assuming $p \neq 1/2$, solving for $\mathbb{P}(S_T = a)$, this gives

$$\mathbb{P}(S_T = a) = \frac{\left(\frac{1-p}{p}\right)^b - \left(\frac{1-p}{p}\right)^i}{\left(\frac{1-p}{p}\right)^b - \left(\frac{1-p}{p}\right)^a}. \quad (461)$$

▲

Example 5.4 (Duration of fair games). Let $(X_t)_{t \geq 1}$ be i.i.d. RVs with

$$\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2. \quad (462)$$

Let $S_t = S_0 + X_1 + \cdots + X_t$ with $S_0 = 0$. Let $\mathcal{F}_t$ be the information obtained by observing $S_0, \dots, S_t$. Note that $(S_t)_{t \geq 0}$ itself is a martingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$. Also, we have seen that $M_t := S_t^2 - t$ is a martingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$ in Exercise 2.9.

Fix integers $a < 0 < b$, and define

$$T = \min\{t \geq 0 \mid S_t = a \text{ or } S_t = b\}. \quad (463)$$

Again T is a stopping time w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$. As before, we also have $\mathbb{P}(T < \infty) = 1$ and $S_T, S_{T \wedge t} \in [a, b]$. Hence by Proposition 5.2, we get

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_T] = a\mathbb{P}(S_T = a) + b\mathbb{P}(S_T = b). \quad (464)$$

Since $\mathbb{P}(S_T = b) = 1 - \mathbb{P}(S_T = a)$, the first equation yield

$$\mathbb{P}(S_T = a) = \frac{b}{b-a}, \quad \mathbb{P}(S_T = b) = \frac{-a}{b-a}. \quad (465)$$

For the other martingale, we need to stop it at time $T \wedge t$ since $M_{T \wedge t} = S_{T \wedge t}^2 - (T \wedge t)$ may depend on t . Proposition 5.2 yields

$$0 = \mathbb{E}[M_0] = \mathbb{E}[M_{T \wedge t}] = \mathbb{E}[S_{T \wedge t}^2 - (T \wedge t)] = \mathbb{E}[S_{T \wedge t}^2] - \mathbb{E}[T \wedge t]. \quad (466)$$

Notice that, as $t \rightarrow \infty$,

$$\mathbb{E}[T \wedge t] = \sum_{k=0}^t \mathbb{P}(T \geq k) \nearrow \sum_{k=0}^{\infty} \mathbb{P}(T \geq k) = \mathbb{E}[T]. \quad (467)$$

Also, since $S_{T \wedge t}^2 \leq \max(a^2, b^2)$, we can use Proposition 5.1 to get

$$\mathbb{E}[S_T^2] = \lim_{t \rightarrow \infty} \mathbb{E}[S_{T \wedge t}^2]. \quad (468)$$

Hence (466) yields $\mathbb{E}[T] = \mathbb{E}[S_T^2]$. Thus

$$\mathbb{E}[T] = a^2 \mathbb{P}(S_T = a) + b^2 \mathbb{P}(S_T = b) \quad (469)$$

$$= \frac{a^2 b}{b-a} - \frac{b^2 a}{b-a} = \frac{ab(a-b)}{b-a} = -ab. \quad (470)$$

This shows the expected duration of a fair game is $\mathbb{E}[T] = |ab|$.

Observe that for each fixed $a < 0$, $\mathbb{E}[T] = |ab| \rightarrow \infty$ as $b \rightarrow \infty$. In other words, if one starts gambling on fair coin flips, each time winning or losing \$1, then the expected time $\mathbb{E}[T]$ that he reaches $-\$1$, or ruins, is infinity. In terms of random walk, this shows that the first return time of a simple symmetric random walk has infinite expected value. In terms of $M/M/1$ queue, this shows that if the arrival and service rate is the same, then there is no stationary distribution for the queue size. ▲

Next, we recall the following computation of normal MGF

Exercise 5.5 (MGF of normal RVs). Let $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0, 1)$. Using the fact that $\mathbb{E}[e^{tZ}] = e^{t^2/2}$, show that

$$\mathbb{E}[e^{tY}] = \exp(\sigma^2 t^2/2 + t\mu). \quad (471)$$

The following example illustrates how martingales can be used in a risk management situation.

Example 5.6 (Cramér's estimate of ruin). Let S_n denote the total assets of an insurance company at the end of year n . During year n , premiums totaling $c > 0$ dollars are received, while claims totaling Y_n dollars are paid. Hence

$$S_n = S_{n-1} + c - Y_n. \quad (472)$$

Let $X_n = c - Y_n$ be the net profit during year n . We may assume that X_n 's are i.i.d. with distribution $N(\mu, \sigma^2)$. We are interested in estimating the probability of bankruptcy. Namely, let

$$B = \{S_n < 0 \text{ for some } n \geq 1\} = \{\text{Bankruptcy}\}. \quad (473)$$

We will show that

$$\mathbb{P}(B) \leq \exp\left(-\frac{2\mu S_0}{\sigma^2}\right). \quad (474)$$

Hence, the company will avoid bankruptcy by maximizing the mean profit per year μ and the initial asset S_0 , while minimizing the uncertainty σ of profit per year.

To begin, notice that $S_n = S_0 + X_1 + \dots + X_n$ is a random walk. Let $\mathcal{F}_t$ be the information obtained by observing $S_0, S_1, \dots, S_t$. According to Exercise 5.5, we have

$$\varphi(\theta) := \mathbb{E}[e^{\theta X_k}] = \exp(\sigma^2 \theta^2/2 + \theta \mu) < \infty. \quad (475)$$

Hence we can use the following exponential martingale (see Exercise 2.6)

$$M_t(\theta) := \frac{\exp(\theta S_n)}{\varphi(\theta)} \quad (476)$$

with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$. Moreover, note that

$$\varphi(-2\mu/\sigma^2) = \exp(2\mu^2/\sigma^2 - 2\mu^2/\sigma^2) = \exp(0) = 1. \quad (477)$$

Hence

$$M_t := M_t(-2\mu/\sigma^2) = \exp\left(-\frac{2\mu S_n}{\sigma^2}\right) \quad (478)$$

is a martingale w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$.

Let $T = \min\{k \geq 1 \mid S_k < 0\}$ be the first time of hitting negative asset, which is a stopping time w.r.t. the filtration $(\mathcal{F}_t)_{t \geq 0}$. Also note that

$$M_T = \exp\left(-\frac{2\mu S_T}{\sigma^2}\right) > 1 \quad \text{a.s.}, \quad (479)$$

since $S_T < 0$. Since S_t is a supercritical random walk ($\mu > 0$), $\mathbb{P}(T = \infty) > 0$. Hence we need to stop the martingale at truncated stopping time $T \wedge t$, using Proposition 5.2. Noting that $M_t \geq 0$ for all $t \geq 0$, this yields

$$\exp\left(-\frac{2\mu S_0}{\sigma^2}\right) = \mathbb{E}[M_0] = \mathbb{E}[M_{T \wedge t}] = \mathbb{E}[M_T | T \leq t] \mathbb{P}(T \leq t) + \mathbb{E}[M_t | T > t] \mathbb{P}(T > t) \quad (480)$$

$$\geq \mathbb{E}[M_T | T \leq t] \mathbb{P}(T \leq t) \geq \mathbb{P}(T \leq t). \quad (481)$$

Note that the last equality follows since S_T given $T \leq t$ is zero almost surely and $M_t \geq 0$ for all $t \geq 0$. To finish, note that $\mathbb{P}(T \leq t)$ is the probability of having bankruptcy by time t . Hence

$$\mathbb{P}(B) = \mathbb{P}(T < \infty) = \lim_{n \rightarrow \infty} \mathbb{P}(T \leq n) \leq \exp\left(-\frac{2\mu S_0}{\sigma^2}\right). \quad (482)$$

▲

Introduction to mathematical finance

We introduce basic notions of no arbitrage principle, binomial model, and connect to martingales.

1. Hedging and replication in the two-state world

We model the market as a probability space $(\Omega, \mathbb{P})$, where Ω consists of sample paths ω of the market, which describes a particular time evolution scenario. For each event $E \subseteq \Omega$, $\mathbb{P}(E)$ gives the probability that the event E occurs. A *portfolio* is a collection of assets that one has at a particular time. The value of a portfolio A at time t is denoted by V_t^A . If t denotes the current time, then V_t^A is a known quantity. However, at a future time $T \geq t$, V_T^A depends on how the market evolves during $[t, T]$, so it is a random variable. Also recall the definition of an arbitrage portfolio:

Definition 1.1. A portfolio A at current time t is said to be an *arbitrage portfolio* if its value V^A satisfies the followings:

- (i) $V^A(t) \leq 0$.
- (ii) There exists a future time $T \geq t$ such that $\mathbb{P}(V^A(T) \geq 0) = 1$ and $\mathbb{P}(V^A(T) > 0) > 0$.

Example 1.2 (A 1-step binomial model). Suppose we have an asset with price $(S_t)_{t \geq 0}$. Consider a European call option at time $t = 0$ with strike $K = 110$ and maturity $t = 1$ (year). Suppose that $S_0 = 100$ and at time 1, S_1 takes one of the two values 120 and 90 according to a certain distribution. One can imagine flipping a coin with unknown probability, and according to whether it lands heads (H) or tail (T), the stock value S_1 takes values $S_1(H) = 120$ and $S_1(T) = 90$. Assume annually compounded interested rate $r = 4\%$. Can we determine its current value $c = C_{110}(0, 1)$? We will show $c = 14/3 \approx 4.48$ by using two arguments – hedging and replication.

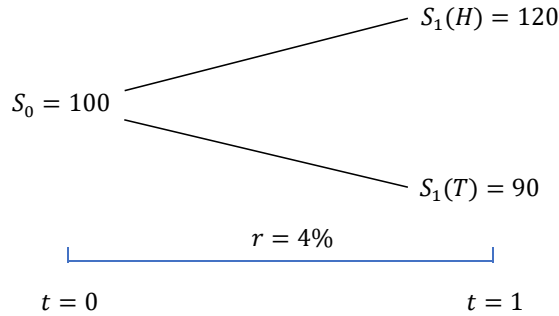


FIGURE 1. 1-step binomial model

Here we give a ‘hedging argument’ for option pricing. Consider the following portfolio at time $t = 0$:

Portfolio A: $[x \text{ shares of the stock}] + [y \text{ European call options with strike 110 and maturity 1}]$.

The cost of entering into this portfolio (at time $t = 0$) is $100x + cy$. Hence the profit of this portfolio takes the following two values

$$\begin{cases} V_1^A(H) - (100x + cy)(1.04) = [120x + y(120 - 110)^+] - [104x + (1.04)cy] = 16x + (10 - (1.04)c)y \\ V_1^A(T) - (100x + cy)(1.04) = [90x + y(90 - 110)^+] - [104x + (1.04)cy] = -14x - (1.04)cy. \end{cases} \quad (483)$$

In order for a perfect hedging, consider choosing the values of x and y such that the profit of this portfolio at maturity is the same for the two outcomes of the stock. Hence we must have

$$16x + (10 - (1.04)c)y = -14x - (1.04)cy. \quad (484)$$

Solving this, we find

$$3x + y = 0. \quad (485)$$

Hence if the above equation is satisfied, the profit of portfolio A is

$$V_1^A - (104x + (1.04)c)y = -14x - (1.04)c(-3x) = ((3.12)c - 14)x. \quad (486)$$

If $(3.12)c > 14$, then portfolio A is an arbitrage portfolio; If $(3.12)c < 14$, then the ‘dual’ of portfolio A , which consists of $-x$ shares of the stock and $-y$ European call options, is an arbitrage portfolio. Hence assuming no-arbitrage, the only possible value of c is $c = 14/(3.12)$. ▲

2. The fundamental theorem of asset pricing

The observation we made in Example 1.2 can be generalized into the so-called ‘fundamental theorem of asset pricing’. For its setup, consider a market where there are n different time evolution $\omega_1, \dots, \omega_n$ between time $t = 0$ and $t = 1$, each occurs with a positive probability. Suppose there are assets $A^{(1)}, \dots, A^{(m)}$, whose price at time t is given by $S_t^{(i)}$ for $i = 1, 2, \dots, m$ (see Figure 2). For each $1 \leq i \leq m$ and $1 \leq j \leq n$, define

$$\alpha_{i,j} = \begin{pmatrix} \text{profit at time } t = 1 \text{ of buying one share of asset } A^{(i)} \\ \text{at time } t = 0 \text{ when the market evolves via } \omega_j. \end{pmatrix} \quad (487)$$

Let $\mathbf{A} = (\alpha_{i,j})$ denote the $(m \times n)$ matrix of profits:

$$\mathbf{A} := \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,n} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \cdots & \alpha_{m,n} \end{bmatrix}. \quad (488)$$

Consider the following portfolio at time $t = 0$:

Portfolio A: $[x_1 \text{ shares of asset } A^{(1)}] + \cdots + [x_m \text{ shares of asset } A^{(m)}]$.

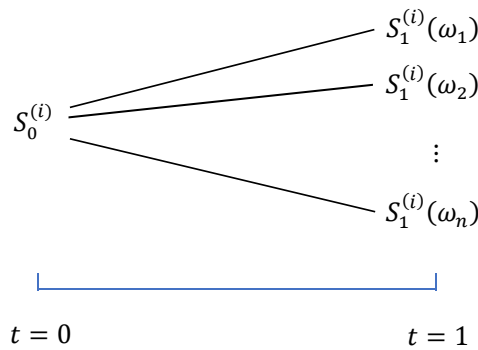


FIGURE 2. 1-step n -state model for asset $A^{(i)}$.

Theorem 2.1 (The fundamental theorem of asset pricing). *Consider portfolio A and the profit matrix $\mathbf{A} = (\alpha_{i,j})$ as above. Then exactly one of the followings hold:*

- (i) There exists an investment allocation $(x_1, \dots, x_m)$ such that portfolio A is an arbitrage portfolio, that is, the n -dimensional row vector

$$[x_1, x_2, \dots, x_m] \begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,n} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \cdots & \alpha_{m,n} \end{bmatrix} \quad (489)$$

has nonzero coordinates and at least one strictly positive coordinate.

- (ii) There exists a strictly positive probability distribution $\mathbf{p}^* = (p_1^*, \dots, p_n^*)$ under which the expected profit of each asset is zero:

$$\begin{bmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,n} \\ \alpha_{2,1} & \alpha_{2,2} & \cdots & \alpha_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ \alpha_{m,1} & \alpha_{m,2} & \cdots & \alpha_{m,n} \end{bmatrix} \begin{bmatrix} p_1^* \\ p_2^* \\ \vdots \\ p_n^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (490)$$

Remark 2.2. Theorem 2.1 (i) states that the portfolio A is an arbitrage portfolio for some $(x_1, \dots, x_m)$. The probability distribution $\mathbf{p}^*$ in the above theorem is called the *risk-neutral probability distribution*. Hence Theorem 2.1 states that there is no way to make A into an arbitrage portfolio if and only if there exists a risk-neutral probability distribution under which the expected profit of each asset $A^{(i)}$ is zero.

Example 2.3 (Example 1.2 revisited). Consider the situation described in Example 1.2. Let $A^{(1)}$ be the asset of price $(S_t)_{t \geq 0}$ and $A^{(2)}$ denote the European call with strike $K = 110$ on this asset with maturity T . Then the matrix $\mathbf{A}$ of profits is given by

$$\mathbf{A} = \begin{bmatrix} 16 & -14 \\ 10 - (1.04)c & -(1.04)c \end{bmatrix}, \quad (491)$$

where $c = C_{110}(0, 1)$ denotes the price of this European option. Assuming no-arbitrage, the fundamental theorem implies that there exists risk-neutral probability distribution $\mathbf{p}^* = (p_1^*, p_2^*)$ such that $\mathbf{A}(\mathbf{p}^*)^T = \mathbf{0}^T$. Namely,

$$\begin{cases} 16p_1^* - 14p_2^* = 0 \\ (10 - (1.04)c)p_1^* - (1.04)c p_2^* = 0 \end{cases} \quad (492)$$

Since $p_1^* + p_2^* = 1$, the first equation implies $p_1^* = 7/15$ and $p_2^* = 8/15$. Then from the second equation, we get

$$(1.04)c = 10p_1^* = \frac{14}{3}. \quad (493)$$

This gives $c = 14/(3.12)$. ▲

Exercise 2.4. Rework Examples 1.2 and 2.3 with following parameters:

$$S_0 = 100, \quad S_1(H) = 130, \quad S_1(T) = 80, \quad r = 5\%, \quad K = 110. \quad (494)$$

PROOF OF THEOREM 2.1. Suppose (i) holds with $\mathbf{x} = (x_1, \dots, x_n)$. We want to show that (ii) cannot hold. Fix a strictly positive probability distribution $\mathbf{p} = (p_1, \dots, p_n)'$, where $'$ denotes the transpose so that $\mathbf{p}$ is an n -dimensional column vector. By (i), we have

$$\mathbf{x}(\mathbf{A}\mathbf{p}) = (\mathbf{x}\mathbf{A})\mathbf{p} > 0. \quad (495)$$

It follows that $\mathbf{A}\mathbf{p}$ cannot be the zero vector in $\mathbb{R}^n$. Hence (490) cannot hold for $\mathbf{p}^* = \mathbf{p}$, as desired.

Next, suppose that (ii) holds for some strictly positive probability distribution $\mathbf{p}^* = (p_1^*, \dots, p_n^*)'$. We use a linear algebra argument to show that (i) does not hold. For each m -dimensional row vector $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^m$, one can correspond a n -dimensional column vector $\mathbf{x}\mathbf{A} \in \mathbb{R}^n$. The condition (490) says

that $\mathbf{A}\mathbf{p}^* = \mathbf{0}$, where T denotes the transpose. Hence for each $\mathbf{x} \in \mathbb{R}^m$, by using associativity of matrix multiplication,

$$(\mathbf{x}\mathbf{A})\mathbf{p}^* = \mathbf{x}(\mathbf{A}\mathbf{p}^*) = \mathbf{x}\mathbf{0} = \mathbf{0}. \quad (496)$$

This shows that the image of the linear map $\mathbf{x} \mapsto \mathbf{x}\mathbf{A}$, which is a linear subspace of $\mathbb{R}^n$, is orthogonal to the strictly positive vector $\mathbf{p}^*$. Hence this linear subspace intersects with the positive orthant $\{(y_1, \dots, y_n) \mid y_1, \dots, y_n \geq 0\}$ only at the origin. This shows that (i) does not hold, as desired. $\square$

3. The binomial model

3.1. 1-step binomial model. Suppose we have an asset with price $(S_t)_{t \geq 0}$. Assume that a future time $t = 1$, the stock price S_1 can take two values $S_1(H) = S_0u$ and $S_1(T) = S_0d$ according to an invisible coin flip, where $u, d > 0$ are multiplicative factors for upward and downward moves for the stock price during the period $[0, 1]$. Assume the aggregated interest rate during $[0, 1]$ is $r > 0$, so that the value of ZCB (Zero Cupon Bond) maturing at 1 is given by $Z(0, 1) = 1/(1 + r)$.

Consider a general European option on this stock, whose value at time $t = 0, 1$ is denoted V_t . We would like to determine its initial value (price at $t = 0$) V_0 in terms of its payoff V_1 . In the previous section, we have seen that there are three ways to proceed: 1) hedging argument, 2) replication argument, and 3) risk-neutral probability distribution.

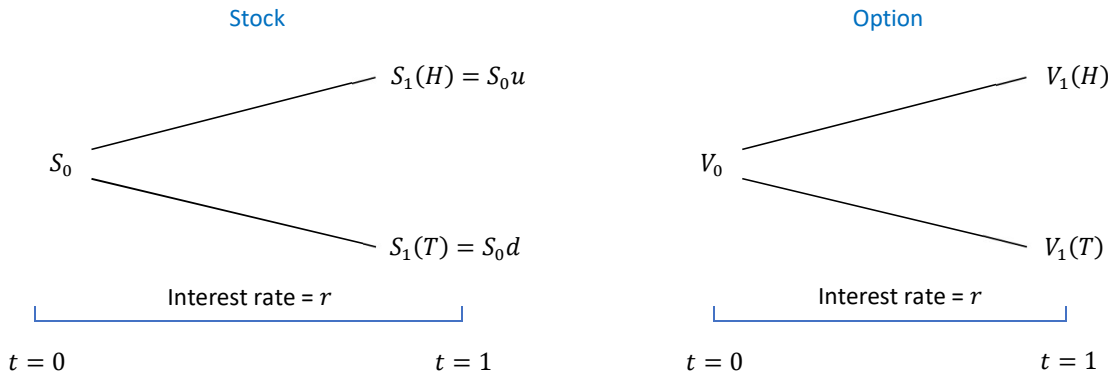


FIGURE 3. 1-step binomial model with general European option

Proposition 3.1. *In the above binomial model, the followings hold.*

(i) *There exists a risk-neutral probability distribution $\mathbf{p}^* = (p_1^*, p_2^*)$ if and only if*

$$0 < d < 1 + r < u. \quad (497)$$

Furthermore, if the above condition holds, $\mathbf{p}^*$ is uniquely given by

$$p_1^* = \frac{(1+r)-d}{u-d}, \quad p_2^* = \frac{u-(1+r)}{u-d}. \quad (498)$$

(ii) *Suppose (497) holds. Then the initial value V_0 of the European option is given by*

$$V_0 = \frac{1}{1+r} \mathbb{E}_{\mathbf{p}^*}[V_1] = \frac{1}{1+r} \left(\frac{(1+r)-d}{u-d} V_1(H) + \frac{u-(1+r)}{u-d} V_1(T) \right). \quad (499)$$

PROOF. To begin, we first need to compute the (2×2) profit matrix $\mathbf{A} = (\alpha_{i,j})$, whose rows and columns correspond to the two kinds of assets (stock and European option) and outcomes (coin flips), respectively. We find

$$\mathbf{A} = \begin{bmatrix} S_1(H) - S_0 \cdot (1+r) & S_1(T) - S_0 \cdot (1+r) \\ V_1(H) - V_0 \cdot (1+r) & V_1(T) - V_0 \cdot (1+r) \end{bmatrix}. \quad (500)$$

The risk-neutral probability $\mathbf{p}^* = (p_1^*, p_2^*)'$ satisfies $\mathbf{A}\mathbf{p}^* = \mathbf{0}$, so

$$\begin{cases} [S_1(H) - S_0 \cdot (1+r)]p_1^* + [S_1(T) - S_0 \cdot (1+r)]p_2^* = 0 \\ [V_1(H) - V_0 \cdot (1+r)]p_1^* + [V_1(T) - V_0 \cdot (1+r)]p_2^* = 0. \end{cases} \quad (501)$$

Using the fact that $p_1^* + p_2^* = 1$, the first equation gives

$$p_1^* = \frac{S_0(1+r) - S_1(T)}{S_1(H) - S_1(T)} = \frac{(1+r) - d}{u - d}, \quad p_2^* = \frac{S_1(H) - S_0(1+r)}{S_1(H) - S_1(T)} = \frac{u - (1+r)}{u - d}. \quad (502)$$

Hence the desired expression for the risk-neutral probabilities p_1^* and p_2^* holds. Note that this gives a strictly positive probability distribution if and only if (497) holds. This shows (i). (Why does this condition make sense?)

Assuming (497), the second equation in (501) then gives

$$V_0 \cdot (1+r) = V_1(H)p_1^* + V_1(T)p_2^*. \quad (503)$$

The right hand side can be regarded as the expectation $\mathbb{E}_{\mathbf{p}^*}[V_1]$ of the value V_1 of the European option at time $t = 1$ under the risk-neutral probability distribution $\mathbf{p}^*$. Then (ii) follows from (i). $\square$

Proposition 3.2. *In the 1-step binomial model as before, consider the following portfolios:*

Portfolio A: $[\Delta_0 \text{ shares of the stock}] + [\text{Short one European option}]$.

Portfolio B: $[\Delta_0 \text{ shares of the stock}] + [x \text{ cash}]$,

Then the followings hold:

(i) *Portfolio A is perfectly hedged (i.e., constant payoff at time $t = 1$) if and only if*

$$\Delta_0 = \frac{V_1(H) - V_1(T)}{S_1(H) - S_1(T)}. \quad (504)$$

(ii) *Portfolio B replicates long one European option if and only if we have (504) and*

$$x = \frac{1}{1+r} \frac{S_1(H)V_1(T) - S_1(T)V_1(H)}{S_1(H) - S_1(T)}. \quad (505)$$

Furthermore,

$$V_0 = x + \Delta_0 S_0 = \frac{1}{1+r} \left(V_1(H) \frac{(1+r) - u}{u - d} + V_1(T) \frac{u - (1+r)}{u - d} \right). \quad (506)$$

PROOF. To show (i), we equate the two payoffs of portfolio A at time $t = 1$ and obtain

$$\Delta_0 S_1(H) - V_1(H) = \Delta_0 S_1(T) - V_1(T). \quad (507)$$

Solving this for Δ_0 shows the assertion.

To show (ii), note that portfolio A replicates long one European option if and only if

$$\begin{cases} \Delta_0 S_1(H) + x \cdot (1+r) = V_1(H) \\ \Delta_0 S_1(T) + x \cdot (1+r) = V_1(T), \end{cases} \quad (508)$$

or in matrix form,

$$\begin{bmatrix} S_1(H) & 1+r \\ S_1(T) & 1+r \end{bmatrix} \begin{bmatrix} \Delta_0 \\ x \end{bmatrix} = \begin{bmatrix} V_1(H) \\ V_1(T) \end{bmatrix}. \quad (509)$$

This is equivalent to

$$\begin{bmatrix} \Delta_0 \\ x \end{bmatrix} = \frac{1}{(1+r)(S_1(H) - S_1(T))} \begin{bmatrix} 1+r & -1-r \\ -S_1(T) & S_1(H) \end{bmatrix} \begin{bmatrix} V_1(H) \\ V_1(T) \end{bmatrix}, \quad (510)$$

which is also equivalent to (504) and (505), as desired.

Lastly, suppose portfolio B replicates long one European option. Then by the monotonicity theorem, initial value V_0 of the European option should equal to the value of portfolio A at time $t = 0$. This shows

$$V_0 = x + \Delta_0 S_0. \quad (511)$$

Using (505), we also have

$$x + \Delta_0 S_0 = \frac{1}{1+r} \frac{S_1(H)V_1(T) - S_1(T)V_1(H)}{S_1(H) - S_1(T)} + \frac{V_1(H)S_0 - V_1(T)S_0}{S_1(H) - S_1(T)} \quad (512)$$

$$= \frac{1}{1+r} \left(V_1(H) \frac{S_0 \cdot (1+r) - S_1(H)}{S_1(H) - S_1(T)} + V_1(T) \frac{S_1(H) - S_0 \cdot (1+r)}{S_1(H) - S_1(T)} \right) \quad (513)$$

$$= \frac{1}{1+r} \left(V_1(H) \frac{(1+r) - u}{u - d} + V_1(T) \frac{u - (1+r)}{u - d} \right). \quad (514)$$

This shows (ii). (Remark: By using Proposition 3.1, one can avoid using the monotonicity theorem here.) $\square$

Example 3.3 (Excerpted from [Dur99]). Suppose a stock is selling for \$60 today. A month from now it will be either \$80 or \$50, i.e., $u = 4/3$ and $d = 5/6$. Assume the interest rate $r = 1/18$ for this period. Then according to Proposition 3.2, the risk-neutral probability distribution $\mathbf{p}^* = (p_1^*, p_2^*)$ is given by

$$p_1^* = \frac{(1 + \frac{1}{18}) - \frac{5}{6}}{\frac{4}{3} - \frac{5}{6}} = \frac{4}{9}, \quad p_2^* = \frac{5}{9}. \quad (515)$$

Now consider a European call option on this stock with strike $K = 65$ maturing in a month. Then $V_1(H) = (80 - 65)^+ = 15$ and $V_1(T) = (50 - 65)^+ = 0$. By Proposition 3.2, the initial value V_0 of this European option is

$$V_0 = \frac{1}{1 + (1/18)} \mathbb{E}_{\mathbf{p}^*}[V_1] = \frac{18}{19} \cdot 15 \cdot \frac{4}{9} = \frac{120}{19} = 6.3158. \quad (516)$$

Working in an investment bank, you were able to sell 10,000 calls to a customer for a slightly higher price each at \$6.5, receiving up-front payment of \$65,000. At maturity, the overall profit is given by

$$\begin{cases} (19/18) \cdot \$65,000 - 10,000 \cdot (80 - 65)^+ = -\$81,389 & \text{if stock goes up} \\ (19/18) \cdot \$65,000 - 10,000 \cdot (50 - 65)^+ = \$68,611 & \text{if stock goes down.} \end{cases} \quad (517)$$

Being worried about losing a huge amount if the stock goes up, you decided to hedge and lock the profit. According to Proposition 3.2, the hedge ratio Δ_0 is given by

$$\Delta_0 = \frac{(80 - 65)^+ - (50 - 65)^+}{80 - 50} = \frac{15}{30} = \frac{1}{2}. \quad (518)$$

Since you have shorted 10,000 calls, this means you need to buy 5,000 shares of the stock owing $5,000 \cdot 60 - \$65,000 = \$235,000$ to the bank. This forms a portfolio of

$$[5,000 \text{ shares of stock}] + [10,000 \text{ short calls}]. \quad (519)$$

The overall profit at maturity is then

$$\begin{cases} 5,000 \cdot \$80 - 10,000 \cdot (80 - 65)^+ - (19/18) \cdot \$235,000 = \$1,944 & \text{if stock goes up} \\ 5,000 \cdot \$50 - 10,000 \cdot (50 - 65)^+ - (19/18) \cdot \$235,000 = \$1,944 & \text{if stock goes down.} \end{cases} \quad (520)$$

▲

Exercise 3.4. Rework Example 3.3 for the following parameters:

$$S_0 = 50, \quad S_1(T) = 70, \quad S_1(H) = 40, \quad r = 4\%, \quad K = 60. \quad (521)$$

3.2. The N -step binomial model. In this subsection, we consider the general N -step binomial model. Namely, starting at the current time $t = 0$, we flip N coins at times $t = 1, 2, \dots, N$ to determine the market evolution. More precisely, now the sample space of the outcomes is $\Omega = \{H, T\}^N$, which consists of sequences of length N strings of H 's or T 's. We assume constant interest rate for each periods $[k, k + 1]$ for $k = 0, 1, \dots, N - 1$.

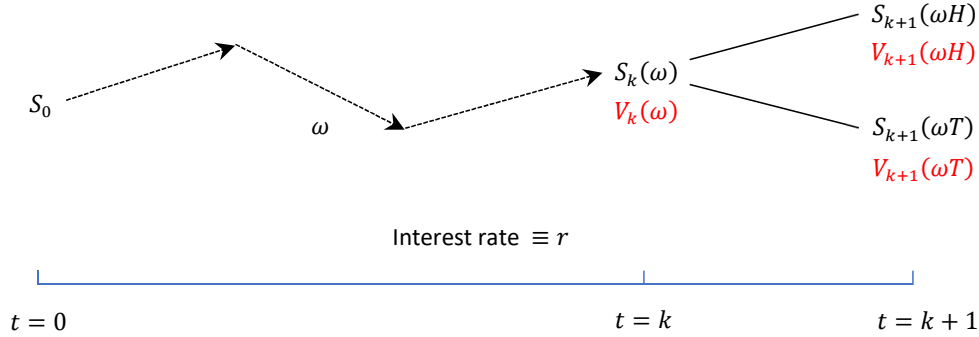


FIGURE 4. Illustration of the N -step binomial model. ω is a sample path of the market evolution in the first k steps. In the next period $[k, k + 1]$, either up or down evolution occurs and the same path is extended to ωH or ωT accordingly. S_t and V_t denote the stock price and option payoff.

The following result for the general N -step binomial model is a direct analog of the 2-step case we have discussed in the previous subsection. To better understand its second part, recall the remark on a probabilistic interpretation on the 2-step value formula as in Proposition ??.

Proposition 3.5. Consider the N -step binomial model as above. Consider a European option on this stock with value $(V_t)_{0 \leq t \leq N}$.

(i) For each integer $0 \leq k < N$ and a sample path $\omega \in \{H, T\}^k$ for the first k steps, define the risk-neutral probability distribution $\mathbf{p}^*(\omega) = (p_1^*(\omega), p_2^*(\omega))$ by

$$p_1^*(\omega) = \frac{(1+r)S_k(\omega) - S_{k+1}(\omega T)}{S_{k+1}(\omega H) - S_{k+1}(\omega T)}, \quad p_2^*(\omega) = \frac{S_{k+1}(\omega H) - (1+r)S_k(\omega)}{S_{k+1}(\omega H) - S_{k+1}(\omega T)}. \quad (522)$$

If $0 < p_1^*(\omega) < 1$, then

$$V_k(\omega) = \frac{1}{(1+r)} \mathbb{E}_{\mathbf{p}^*(\omega)}[V_{k+1} | \text{first } k \text{ coin flips} = \omega] \quad (523)$$

$$= \frac{1}{(1+r)} (V_{k+1}(\omega H)p_1^*(\omega) + V_{k+1}(\omega T)p_2^*(\omega)). \quad (524)$$

(ii) Consider N consecutive coin flips such that given any sequence $x_1 x_2 \dots x_k$ of the first k flips, the $(k + 1)$ st coin lands on heads with probability $p_1^*(x_1 x_2 \dots x_k)$. Let $\mathbb{P}^*$ denote the induced probability measure (risk-neutral probability measure) on the sample space $\Omega = \{H, T\}^N$. Then

$$V_0 = \frac{1}{(1+r)^N} \mathbb{E}_{\mathbb{P}^*}[V_N]. \quad (525)$$

PROOF. The argument for (i) is exactly the same as in the proof of Proposition ??. For (ii) we use an induction on the number of steps. The base case is verified by (i). For the induction step, we first use (i) to write

$$V_0 = \frac{1}{1+r} (V_1(H)p_1^*(\emptyset) + V_1(T)p_2^*(\emptyset)). \quad (526)$$

Let $X_1, X_2, \dots, X_N$ be a sequence of N (random) coin flips given by the risk-neutral probability measure $\mathbb{P}^*$. Denote the expectation under $\mathbb{P}^*$ by $\mathbb{E}^*$. By the induction hypothesis, we have

$$V_1(H) = \frac{1}{(1+r)^{N-1}} \mathbb{E}^*[V_N | X_1 = H], \quad V_1(T) = \frac{1}{(1+r)^{N-1}} \mathbb{E}^*[V_N | X_1 = T]. \quad (527)$$

Hence we have

$$V_0 = \frac{1}{(1+r)^N} [\mathbb{E}^*[V_N | X_1 = H] p_1^*(\phi) + \mathbb{E}^*[V_N | X_1 = T] p_2^*(\phi)] \quad (528)$$

$$= \frac{1}{(1+r)^N} \mathbb{E}^*[\mathbb{E}^*[V_N | X_1]] = \frac{1}{(1+r)^N} \mathbb{E}^*[V_N], \quad (529)$$

where we have used iterated expectation for the last equality. This shows the assertion. $\square$

3.3. The N -step binomial model revisited. In this subsection, we revisit the N -step binomial model in the framework of martingales. First recall the model. Starting at the current time $t = 0$, we flip N coins at times $t = 1, 2, \dots, N$ to determine the market evolution. The sample space of the outcomes is $\Omega = \{H, T\}^N$, which consists of sequences of length N strings of H 's or T 's. We assume constant interest rate for each periods $[k, k+1]$ for $k = 0, 1, \dots, N-1$.

Let $\mathcal{F}_t$ denote the information that we can obtain by observing the market up to time t . For instance, $\mathcal{F}_t$ contains the information of the first t coin flips, stock prices $S_0, S_1, \dots, S_t$, European option values $V_0, V_1, \dots, V_t$, and so on. Then $(\mathcal{F}_t)_{t \geq 0}$ defines a natural filtration for the N -step binomial model. Below we reformulate Proposition 3.5.

Proposition 3.6. *Consider the N -step binomial model as above. Let $\mathbb{P}^*$ denote the risk-neutral probability measure defined in Proposition 3.5. Consider a European option on this stock with value $(V_t)_{0 \leq t \leq N}$.*

(i) *The process $((1+r)^{-t} V_t)_{0 \leq t \leq N}$ forms a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ under the risk-neutral probability measure $\mathbb{P}^*$. That is,*

$$\mathbb{E}_{\mathbb{P}^*} \left[(1+r)^{-(t+1)} V_{t+1} - (1+r)^{-t} V_t \mid \mathcal{F}_t \right] = 0, \quad (530)$$

which is also equivalent to

$$V_t = \frac{1}{1+r} \mathbb{E}_{\mathbb{P}^*} \left[V_{t+1} \mid \mathcal{F}_t \right]. \quad (531)$$

(ii) *We have*

$$V_0 = \mathbb{E}_{\mathbb{P}^*} \left[(1+r)^{-N} V_N \right]. \quad (532)$$

PROOF. To show (i), we first note that, conditioning on the information $\mathcal{F}_t$ up to time t , we know all the coin flips, stock prices, and European option values up to time t . Hence we have

$$\mathbb{E}_{\mathbb{P}^*} \left[(1+r)^{-(t+1)} V_{t+1} \mid \mathcal{F}_t \right] = (1+r)^{-(t+1)} \mathbb{E}_{\mathbb{P}^*} \left[V_{t+1} \mid \mathcal{F}_t \right] \quad (533)$$

$$= (1+r)^{-(t+1)} \cdot (1+r) V_t = (1+r)^{-t} V_t. \quad (534)$$

This shows that $(1+r)^{-t} V_t$ is a martingale with respect to the natural filtration $(\mathcal{F}_t)_{t \geq 0}$ under the risk-neutral probability measure $\mathbb{P}^*$.

Now (ii) follows from (i) and Exercise 3.5. Namely, for each $A \in \mathcal{F}_0$, by Exercise 3.5,

$$\mathbb{E}_{\mathbb{P}^*} \left[(1+r)^{-N} V_N - (1+r)^0 V_0 \mid \mathcal{F}_0 \right] = 0. \quad (535)$$

This yields

$$V_0 = \mathbb{E}_{\mathbb{P}^*} [V_0 \mid \mathcal{F}_0] = \mathbb{E}_{\mathbb{P}^*} [(1+r)^{-N} V_N \mid \mathcal{F}_0], \quad (536)$$

as desired. $\square$

Exercise 3.7 (Risk-neutral probabilities make the stock price a martingale). Consider the N -step binomial model with stock price $(S_t)_{0 \leq t \leq N}$. Let $\mathbb{P}^*$ denote the risk-neutral probability measure defined in Proposition 3.10 in Lecture note 2. 3.5.

- (i) Show that the discounted stock price $(1+r)^{-t}S_t$ forms a martingale with respect to the filtration $(\mathcal{F}_t)_{t \geq 0}$ under the risk-neutral probability measure $\mathbb{P}^*$. That is,

$$\mathbb{E}_{\mathbb{P}^*} \left[(1+r)^{-(t+1)} S_{t+1} - (1+r)^{-t} S_t \mid \mathcal{F}_t \right] = 0, \quad (537)$$

which is also equivalent to

$$S_t = \frac{1}{1+r} \mathbb{E}_{\mathbb{P}^*} \left[S_{t+1} \mid \mathcal{F}_t \right]. \quad (538)$$

(Hint: Use the fundamental thm of asset pricing.)

- (ii) Show that

$$S_0 = \mathbb{E}_{\mathbb{P}^*} \left[(1+r)^{-N} S_N \right]. \quad (539)$$

Bibliography

- [BT02] Dimitri P Bertsekas and John N Tsitsiklis, *Introduction to probability*, vol. 1, Athena Scientific Belmont, MA, 2002.
- [Dur99] Rick Durrett, *Essentials of stochastic processes*, vol. 1, Springer, 1999.
- [Dur10] ———, *Probability: theory and examples*, Cambridge university press, 2010.
- [LP17] David A Levin and Yuval Peres, *Markov chains and mixing times*, vol. 107, American Mathematical Soc., 2017.