Lecturenotes on

Stochastic Processes

Chung-Chin Lu

Department of Electrical Engineering
National Tsing Hua University
Hsinchu 30013, Taiwan

September 13, 2001

Contents

Chapter 1

A Review of Probability Theory

1.1 Measurable Spaces

Let Ω be a set. In a stochastic context, Ω is commonly used to denote the set of all possible outcomes in a random experiment. And Ω is usually called the sample space of that experiment.

A description of the outcome of doing the experiment, i.e. an event, can usually be represented by a subset of the sample space Ω. For example, the description that the outcome of throwing a dice is red-colored can be represented as the subset $\{1, 4\}$ of the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. An event, represented by a subset A, is said to occur if and only if the observed outcome ω of the experiment is indeed an element of A. An interesting question is that is any subset of the sample space Ω meaningful to represent an event of the experiment? The answer is quite general that it is up to your interests to select a collection $\mathcal{F}$ of subsets of Ω such that each member of $\mathcal{F}$ is meaningful (to you) to represent an event, in case that the logical consistency is maintained in $\mathcal{F}$ as discussed in below.

If a subset A of the sample space Ω represents an event as you select to be in $\mathcal{F}$, then the subset $A^c = \Omega - A$ represents the description that event A does not occur, which is logically an event too. Thus it is logical to require that A^c is a member of $\mathcal{F}$. Furthermore, if two subsets A and B represents events as you wish, then the union $A \cup B$ of A and B represents the description that either event occurs, which is again logically an event. And it is again logical to require that $A \cup B$ is a member of $\mathcal{F}$.

Definition 1.1.1 *A non-empty collection $\mathcal{F}$ of subsets of Ω is a (Boolean) algebra if*

1. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
2. $A_1, A_2 \in \mathcal{F} \Rightarrow A_1 \cup A_2 \in \mathcal{F}$. □

In fact, an algebra is closed under finite number of logical operations as stated in the following theorem, which can be proved by induction.

Theorem 1.1.2 *Let $A_1, A_2, \ldots, A_n$ be in an algebra $\mathcal{F}$. Then, $\cup_{j=1}^{n} A_j$ and $\cap_{j=1}^{n} A_j$ are also in $\mathcal{F}$.* □

For convenience, we shall identify a member of an algebra $\mathcal{F}$ as the event it represents. An algebra may not be closed under countably infinite number of logical operations as illustrated in the following example.

Example 1.1.3 *Let Ω be the set Z of all integers. Let $\mathcal{F}$ be the collection of all finite subsets of Z and their complements. It can be shown that $\mathcal{F}$ is an algebra. And we can perform a finite number of logical operations on events in $\mathcal{F}$ and obtain an event in $\mathcal{F}$ consistently. Now, let $A_i = \{2i\}$, $i = 1, 2, \ldots$, in $\mathcal{F}$. Then $\cup_{i=1}^{\infty} A_i$ is the set of all positive even integers and is clearly not in $\mathcal{F}$. Thus the subset $\cup_{i=1}^{\infty} A_i$ is not meaningful to us in this setting.* □

As we shall see, it is required to perform countably infinite number of logical operations on events in dealing with various problems.

Definition 1.1.4 *A non-empty collection $\mathcal{F}$ of subsets of Ω is a (Boolean) σ-algebra if*

1. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$.
2. $A_j \in \mathcal{F}, j = 1, 2, \ldots \Rightarrow \cup_{j=1}^{\infty} A_j \in \mathcal{F}$. □

The next theorem states some properties of a σ-algebra.

Theorem 1.1.5 *Let $\mathcal{F}$ be a σ-algebra. Then*

1. *Ω and $\emptyset$ are in $\mathcal{F}$,*
2. *For $A_j \in \mathcal{F}, j = 1, 2, \ldots$, $\cap_{j=1}^{\infty} A_j$ is in $\mathcal{F}$.* □

By taking $A_{n+1} = A_{n+2} = \cdots = \emptyset$ in Definition 1.1.4, it can be seen that a σ-algebra is an algebra. Thus we can perform countable, finite or infinite, number of logical operations on events in a σ-algebra.

Example 1.1.6 *The power set 2^{Ω} (i.e. the collection of all subsets of Ω) is a σ-algebra.* □

Theorem 1.1.7 *Let $\mathcal{F}_\alpha$ be a σ-algebra on Ω for each α in an index set I. Then the intersection $\cap_{\alpha \in I} \mathcal{F}_\alpha$ of all σ-algebra $\mathcal{F}_\alpha$ is also a σ-algebra on Ω.* □

For any collection $\mathcal{G}$ of subsets of Ω, there exists a smallest σ-algebra $\mathcal{F}$ containing $\mathcal{G}$. This smallest σ-algebra $\mathcal{F}$ is just the intersection of all σ-algebra containing $\mathcal{G}$. $\mathcal{F}$ is called the σ-algebra generated by $\mathcal{G}$.

Example 1.1.8 *The σ-algebra generated by the algebra in Example 1.1.3 is just the power set of Ω.* □

Example 1.1.9 *Let Ω be the set R of all real numbers. Let $\mathcal{G}$ be the set of all open intervals in R. And let $\mathcal{B}$ be the σ-algebra generated by $\mathcal{G}$. A member in $\mathcal{B}$ is called a* Borel set. *Since, for $-\infty \le a < b \le +\infty$,*

$$[a,b] = \cap_{n=1}^{\infty}(a-\frac{1}{n}, b+\frac{1}{n}), \quad (a,b] = \cap_{n=1}^{\infty}(a, b+\frac{1}{n}), \quad [a,b) = \cap_{n=1}^{\infty}(a-\frac{1}{n}, b),$$

any interval is a Borel set in R. □

Definition 1.1.10 *A measurable space $(\Omega, \mathcal{F})$ is a sample space Ω together with a σ-algebra $\mathcal{F}$ on Ω.* □

The measurable space $(R, \mathcal{B})$ in the above example is usually called the Borel measurable space. Now consider a countable [1] sample space Ω. It is usually desirable to have every singleton [2] as an event in each interesting σ-algebra $\mathcal{F}$. It is clear that any such an $\mathcal{F}$ is just the power set 2^{Ω} of Ω. $(\Omega, 2^{\Omega})$ is called a discrete measurable space.

1.2 Probability Spaces

It is now ready to assign a probability to each event in a measurable space $(\Omega, \mathcal{F})$.

Definition 1.2.1 *A probability measure $\mathcal{P}$ on a measurable space $(\Omega, \mathcal{F})$ is a set function from $\mathcal{F}$ to R which satisfies:*

1. *$\forall A \in \mathcal{F}$, $\mathcal{P}(A) \ge 0$.*
2. *(Countable additivity) If $A_j, j = 1, 2, \ldots$ are mutually disjoint events in $\mathcal{F}$, then*

$$\mathcal{P}(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathcal{P}(A_j).$$

3. *$\mathcal{P}(\Omega) = 1$.* □

A probability space $(\Omega, \mathcal{F}, \mathcal{P})$ is a measurable space $(\Omega, \mathcal{F})$ together with a probability measure $\mathcal{P}$. We next list some useful properties of a probability measure $\mathcal{P}$ in the following, where all sets are events:

1. $\mathcal{P}(A) \le 1$.

[1]A set A is *countable* if there exists a one-to-one mapping from A into the set N of positive integers. Otherwise, A is *uncountable*. For example, N, Z and Q are countable sets but R and C are uncountable. A countable set may be finite or infinite. For example, N, Z and Q are countable infinite sets but $\{a, b, c, d\}$ is countably finite.

[2]A singleton is a set containing exactly one element.

2. $\mathcal{P}(\emptyset) = 0$.

3. $\mathcal{P}(A^c) = 1 - \mathcal{P}(A)$.

4. $\mathcal{P}(A \cup B) + \mathcal{P}(A \cap B) = \mathcal{P}(A) + \mathcal{P}(B)$.

5. $A \subseteq B \Rightarrow \mathcal{P}(A) = \mathcal{P}(B) - \mathcal{P}(B - A) \leq \mathcal{P}(B)$.

6. (Monotone property) $A_j \uparrow A$ or $A_j \downarrow A \Rightarrow \mathcal{P}(A_j) \to \mathcal{P}(A)$ as $j \to \infty$.

7. (Boole's inequality) $\mathcal{P}(\cup_{j=1}^{\infty} A_j) \leq \sum_{j=1}^{\infty} \mathcal{P}(A_j)$.

Remark 1.2.2 *$A_j \uparrow A$ means that $\{A_j, j = 1, 2, \ldots\}$ is a monotone increasing sequence of subsets, i.e. $A_1 \subseteq A_2 \subseteq \ldots$, and $A = \cup_{j=1}^{\infty} A_j$. Also $A_j \downarrow A$ means that $\{A_j, j = 1, 2, \ldots\}$ is a monotone decreasing sequence of subsets, i.e. $A_1 \supseteq A_2 \supseteq \ldots$, and $A = \cap_{j=1}^{\infty} A_j$.*

Proof. Properties 1–5 are trivial. Firstly, we consider $A_j \downarrow A$. It is obvious that

$$A_j = \cup_{k=j}^{\infty} (A_k - A_{k+1}) \cup A,$$

a countable union of mutually disjoint events. From the countable additivity axiom of probability measure,

$$\mathcal{P}(A_j) = \sum_{k=j}^{\infty} \mathcal{P}(A_k - A_{k+1}) + \mathcal{P}(A).$$

Especially, the series $\sum_{k=1}^{\infty} \mathcal{P}(A_k - A_{k+1})$ converges and the tail sum $\sum_{k=j}^{\infty} \mathcal{P}(A_k - A_{k+1})$ converges to 0 as j goes to infinity. This proves that $\lim_{j\to\infty} \mathcal{P}(A_j) = \mathcal{P}(A)$. Since $A_j \uparrow A$ is equivalent to $A_j^c \downarrow A^c$, we have $\lim_{j\to\infty} \mathcal{P}(A_j^c) = \mathcal{P}(A^c)$. By Property 3, we then have $\lim_{j\to\infty} \mathcal{P}(A_j) = \mathcal{P}(A)$. This proves the monotone property. Now, by repeatedly applying Property 4, we have $\mathcal{P}(\cup_{j=1}^{n} A_j) \leq \sum_{j=1}^{n} \mathcal{P}(A_j) \leq \sum_{j=1}^{\infty} \mathcal{P}(A_j)$. Since $(\cup_{j=1}^{n} A_j) \uparrow (\cup_{j=1}^{\infty} A_j)$, we have

$$\mathcal{P}\left(\cup_{j=1}^{\infty} A_j\right) = \lim_{n\to\infty} \mathcal{P}\left(\cup_{j=1}^{n} A_j\right) \leq \sum_{j=1}^{\infty} \mathcal{P}(A_j)$$

by the monotone property. This proves the Boole's inequality. □

A discrete measurable space $(\Omega, 2^{\Omega})$, where Ω is countable, together with a probability measure $\mathcal{P}$ on it is called a *discrete* probability space. Since the collection of all singletons (elementary events) of Ω generates 2^{Ω}, the probability measure $\mathcal{P}$ can be completely specified by its assignment on all singletons, i.e. $\mathcal{P}(\{\omega\}), \forall \omega \in \Omega$. If Ω is countably infinite and samples are indexed by the positive integers, then $\{\mathcal{P}(\{\omega_i\}), i = 1, 2, \ldots\}$ is a sequence of non-negative numbers and the series $\sum_{i=1}^{\infty} \mathcal{P}(\{\omega_i\})$ converges to 1. Conversely, given a sequence $\{a_i, i = 1, 2, \ldots\}$ of non-negative number such that the series $\sum_{i=1}^{\infty} a_i$ converges to 1. Then the assignment $\mathcal{P}(\{\omega_i\}) = a_i$, $i = 1, 2, \ldots$, completely specifies a probability measure on the discrete measurable space $(\Omega, 2^{\Omega})$.

There are many experiments in each of which the sample space is uncountable such as the lifetime of a bulb, the length of a telephone call, the end-to-end voltage of a resistor, where the sample space is an interval of the real line R. It is often that the probability $\mathcal{P}(\{\omega\})$ of an elementary event is zero.

Consider the Borel measurable space $(R, \mathcal{B})$. A probability measure on $(R, \mathcal{B})$ is directly related to a distribution function $F(x)$ on R. Let μ be a probability measure on $(R, \mathcal{B})$. Define a function $F(x)$ of x in R as

$$F(x) \equiv \mu((-\infty, x]),\ \forall x \in R. \tag{1.1}$$

It is easy to check that $F(x)$ is a monotone increasing right-continuous function over R and $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$. Such a function is called a *distribution function* on R. Conversely, given a distribution function $F(x)$ on R, we can define a set function μ on any interval of R as [3]

$$\begin{aligned} \mu((a,b]) &= F(b) - F(a), && (1.2a)\\ \mu((a,b)) &= F(b-) - F(a), && (1.2b)\\ \mu([a,b)) &= F(b-) - F(a-), && (1.2c)\\ \mu([a,b]) &= F(b) - F(a-) && (1.2d) \end{aligned}$$

for $-\infty \le a < b \le +\infty$. Such an assignment can be extended to any Borel set in R and μ becomes a probability measure on the Borel measurable space $(R, \mathcal{B})$ uniquely as stated in the following theorem [4]:

Theorem 1.2.3 *Given a probability measure on $(R, \mathcal{B})$. There is a unique distribution function $F(x)$ on R satisfying (1.1). Conversely, given a distribution function $F(x)$. there is a unique probability measure μ satisfying (1.1).* □

We shall call $F(x)$ the distribution function of μ and μ the probability measure of $F(x)$. Note that μ is a set function, while $F(x)$ is a point function.

There is a convenience way to construct a distribution function on R.

Definition 1.2.4 *A probability density function $f(x)$ on R is a non-negative integrable function on R such that*

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

□

Given a probability density function $f(x)$, the function F defined by

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

[3] Since $F(x)$ is monotone increasing, the left limit $\lim_{x\to b-} F(x)$ at every point $x = b$ exists.

[4] See K. L. Chung, *A Course in Probability Theory,* 2nd edn. New York: Academic Press, 1974, pp. 24–28.

is easily seen to be a distribution function. Such a distribution function and its corresponding probability measure are called *absolutely continuous* (in R and with respect to the Lebesque measure [5]).

1.3 Random Variables

Definition 1.3.1 *Let $(\Omega, \mathcal{F})$ and $(\Lambda, \mathcal{G})$ be two measurable spaces. A mapping X from Ω into Λ is measurable if $\forall A \in \mathcal{G}, X^{-1}(A) \in \mathcal{F}$.* □

Note that the inverse mapping X^{-1} is regarded as a set function [6].

Definition 1.3.2 *A (real-valued) random variable on a measurable space $(\Omega, \mathcal{F})$ is a measurable function from $(\Omega, \mathcal{F})$ into $(R, \mathcal{B})$.* □

Example 1.3.3 *Any mapping from a discrete measurable space $(\Omega, 2^\Omega)$ into a measurable space is measurable. In particular, any real-valued function on $(\Omega, 2^\Omega)$ is a random variable.* □

Theorem 1.3.4 *X is an r.v. X on a measurable space $(\Omega, \mathcal{F})$ if and only if*

$$\{\omega : X(\omega) \leq x\} \in \mathcal{F}, \ \forall x \in R.$$

Proof. The "only if" part of the theorem follows from the definition of a random variable. We now consider the "if" part. Let $\mathcal{G}$ be the collection of all subsets A of R such that $X^{-1}(A)$ is in $\mathcal{F}$. By properties of the inverse mapping X^{-1} listed in Footnote 6, we have

1. if A in $\mathcal{G}$, then $X^{-1}(A^c) = (X^{-1}(A))^c$ in $\mathcal{F}$ and A^c in $\mathcal{G}$;
2. if A_j in $\mathcal{G}$ for all j, then $X^{-1}(\cup_j A_j) = \cup_j X^{-1}(A_j)$ in $\mathcal{F}$ and $\cup_j A_j$ in $\mathcal{G}$.

[5] The Lebesque measure m on R is the set function from $\mathcal{B}$ to R which assigns to each Borel set its length. In particular, we have $m([a, b]) = b - a$. The Lebesque measure m satisfies axioms 1 and 2 in the definition of a probability measure. Also note that $m(R) = \infty$.

[6] Let X be a mapping from a *set* Ω into another *set* Λ. The inverse mapping X^{-1} is a set function from the power set 2^Λ of Λ to the power set 2^Ω of Ω defined as

$$X^{-1}(A) = \{\omega \in \Omega | X(\omega) \in A\}$$

for any subset A of Λ. Let A_α, α in an index set I, be subsets of Λ. Then

$$\begin{aligned} X^{-1}(A^c) &= (X^{-1}(A))^c, \\ X^{-1}(\cap_\alpha A_\alpha) &= \cap_\alpha X^{-1}(A_\alpha), \\ X^{-1}(\cup_\alpha A_\alpha) &= \cup_\alpha X^{-1}(A_\alpha). \end{aligned}$$

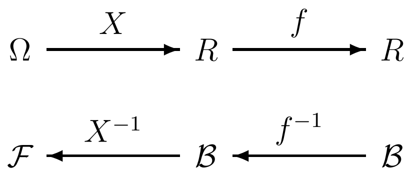

Figure 1.1: Composition of measurable functions.

Thus $\mathcal{G}$ is a σ-algebra. By hypothesis, $\mathcal{G}$ contains all intervals of the form $(-\infty, x]$, $\forall x \in R$, which generate the Borel σ-algebra $\mathcal{B}$. Thus $\mathcal{B} \subseteq \mathcal{G}$ and X is an r.v. by definition. □

In particular, a random variable on the measurable space $(R, \mathcal{B})$ (i.e. a measurable function from $(R, \mathcal{B})$ into $(R, \mathcal{B})$) is called a Borel measurable function. Borel measurable functions are plentiful as stated in the next theorem.

Theorem 1.3.5 *Any real-valued function on R with countably many discontinuities is Borel measurable.* □

Given a random variable X, the event

$$(X \in B) \equiv X^{-1}(B) = \{\omega \in \Omega | X(\omega) \in B\}$$

for a Borel set B is called an event induced by the r.v. X. By properties of the inverse mapping X^{-1}, the collection of all events $(X \in B)$ induced by X, $\forall\ B \in \mathcal{B}$, is a σ-algebra on Ω and is denoted as $\mathcal{F}(X)$. $\mathcal{F}(X)$ contains all information about the r.v. X in the measurable space $(\Omega, \mathcal{F})$. It is clear that $\mathcal{F}(X) \subseteq \mathcal{F}$.

Theorem 1.3.6 *Let X be a random variable on a measurable space $(\Omega, \mathcal{F})$ and $f(x)$ be a Borel measurable function. Then $Y = f(X)$ is a random variable on $(\Omega, \mathcal{F})$.*

Proof. Let B be a Borel set in $\mathcal{B}$. Then

$$Y^{-1}(B) = (f \circ X)^{-1}(B) = X^{-1}(f^{-1}(B))$$

is in $\mathcal{F}$ since $f^{-1}(B)$ is in $\mathcal{B}$. The behavior of inverse mappings X^{-1} and f^{-1} as set functions can be seen in Figure 1.1. □

It can be seen that the σ-algebra $\mathcal{F}(Y)$ generated by the r.v. $Y = f(X)$ is contained in the σ-algebra $\mathcal{F}(X)$ generated by the r.v. X. This reflects the fact that data processing usually loses information about the raw data.

An operation on a finite number of random variables usually results in a random variable. To discuss such operations, we first generalize the concepts of intervals and Borel sets in n-dimensional Euclidean space R^n. An n-cell W in R^n is a subset of R^n of the form

$$W = I_1 \times I_2 \times \cdots \times I_n$$

where I_i's are intervals in R. W is called an open n-cell if all I_i's are open intervals. Let $\mathcal{B}^n$ be the σ-algebra generated by all open n-cells in R^n. Any member in $\mathcal{B}^n$ is called a Borel set in R^n. A function $f(x_1, x_2, \dots, x_n)$ from R^n into R is called Borel measurable if $f^{-1}(B)$ is in $\mathcal{B}^n$ for all $B \in \mathcal{B}$. In particular, any continuous function from R^n into R is Borel measurable. The following is an extension of Theorem 1.3.6.

Theorem 1.3.7 . *Let $X_1, X_2, \dots, X_n$ be r.v.'s on a measurable space $(\Omega, \mathcal{F})$ and $f(x_1, x_2, \dots, x_n)$ be a Borel measurable function. Then $Y = f(X_1, X_2, \dots, X_n)$ is an r.v. on $(\Omega, \mathcal{F})$.*

Proof. As an exercise. □

Define
$$X \vee Y = \max(X, Y) \quad \text{and} \quad X \wedge Y = \min(X, Y).$$

Theorem 1.3.8 *Let X and Y be random variables. Then*
$$X \vee Y, \quad X \wedge Y, \quad X + Y, \quad X - Y, \quad X \cdot Y, \quad X/Y$$
are r.v.'s, provided Y does not vanish in the last one,

Proof. This follows from Theorem 1.3.7 and the continuity of the functions
$f(x, y) = \max(x, y)$, $\min(x, y)$, $x + y$, $x - y$, xy and x/y. □

To discuss operations on a countably infinite number of r.v.'s, we need the following concepts from analysis.

Definition 1.3.9 *Given a sequence $\{a_j, j \geq 1\}$ of real numbers, the supremum $\sup_j a_j$ of a_j's is the least upper bound of a_j's and the infimum $\inf_j a_j$ of a_j's is the greatest lower bound of a_j's.* □

Both supremum $\sup_j a_j$ and infimum $\inf_j a_j$ of an arbitrary sequence $\{a_j, j \geq 1\}$ always exist and may take values at $-\infty$ or $+\infty$.

Definition 1.3.10 *Given a sequence $\{a_j, j \geq 1\}$ of real numbers, the limit superior (upper limit) $\limsup_j a_j$ and the limit inferior (lower limit) $\liminf_j a_j$ of a_j's are defined as*
$$\limsup_j a_j = \lim_{n\to\infty} \left(\sup_{j \geq n} a_j \right), \quad \liminf_j a_j = \lim_{n\to\infty} \left(\inf_{j \geq n} a_j \right).$$
□

Since $\{\sup_{j \geq n} a_j, n \geq 1\}$ is a monotone decreasing sequence and $\{\inf_{j \geq n} a_j, n \geq 1\}$ is a monotone increasing sequence, both limit superior $\limsup_j a_j$ and limit inferior $\liminf_j a_j$ exist and may take values at $-\infty$ or $+\infty$. We list main properties of limit superior and limit inferior as follows [7]:

[7] For detailed discussion and proofs, please see
1. T. M. Apostol, *Mathematical Analysis,* 2nd edn. Reading, Mass.: Addison-Wesley, 1974, pp. 184–185.
2. W. Rudin, *Principles of Mathematical Analysis,* 3rd edn. New York: McGraw-Hill, 1976, pp. 55–57.

1. $\limsup_j a_j = \inf_n \left(\sup_{j\geq n} a_j\right)$ and $\liminf_j a_j = \sup_n \left(\inf_{j\geq n} a_j\right)$.
2. $\limsup_j a_j = -\liminf_j(-a_j)$ and $\liminf_j a_j = -\limsup_j(-a_j)$.
3. $\liminf_j a_j \leq \limsup_j a_j$.
4. Let E be the set of all subsequential limits [8] on the extended real line R^* [9]. Then both $\limsup_j a_j$ and $\liminf_j a_j$ are in E and
$$\limsup_j a_j = \sup E \quad \text{and} \quad \liminf_j a_j = \inf E.$$
5. $\limsup_j a_j$ is a finite number u if and only if
   (a) for every $\epsilon > 0$, there exists an integer $J = J(\epsilon)$ such that $a_j < u + \epsilon \ \forall\ j \geq J$;
   (b) given $\epsilon > 0$ and given an integer n, there exists an integer $j > n$ such that $a_j > u - \epsilon$.
6. Both limit superior and limit inferior are independent of the ordering of a_j's.
7. The sequence $\{a_j, j \geq 1\}$ converges on the extended real line R^* if and only if $\limsup_j a_j = \liminf_j a_j$, in which case $\lim_{j\to\infty} a_j = \limsup_j a_j = \liminf_j a_j$.
8. If $a_j \leq b_j,\ \forall\ j \geq 1$, then
$$\liminf_j a_j \leq \liminf_j b_j \quad \text{and} \quad \limsup_j a_j \leq \limsup_j b_j.$$

It is now clear that it is needed to consider the extended real line $R^* = [-\infty, +\infty]$ and the extended Borel σ-algebra $\mathcal{B}^*$, where an extended Borel set is a Borel set possibly enlarged by one or both infinite points $\pm\infty$. An extended-valued r.v. is a measurable function from $(\Omega, \mathcal{F})$ into $(R^*, \mathcal{B}^*)$.

Theorem 1.3.11 *If $\{X_j, j \geq 1\}$ is a sequence of extended-valued r.v.'s, then*
$$\sup_j X_j, \quad \inf_j X_j, \quad \limsup_j X_j, \quad \liminf_j X_j$$
are also real extended-valued r.v.'s

Proof. Let $Y(\omega) = \sup_j X_j(\omega),\ \forall\ \omega \in \Omega$. Since
$$\{\sup_j X_j \leq x\} = \cap_j \{X_j \leq x\},\ \forall\ x \in R,$$
$Y^{-1}((-\infty, x])$ is in $\mathcal{F}$. By Theorem 1.3.4, Y is an extended-valued random variable. Since $\inf_j X_j = -\sup_j(-X_j)$ and $-X_j$'s are extended-valued random variables, it is cleat that

[8] A subsequential limit is the limit of a convergent subsequence of $\{a_j, j \geq 1\}$.

[9] A sequence $\{b_j, j \geq 1\}$ is said to converges to $+\infty$ (or $-\infty$) if for every real M, there is an integer J such that for all $j \geq J$, we have $b_j \geq M$ (or $b_j \leq M$).

$\inf_j X_j$ is an extended-valued random variable. Since $\limsup_j X_j$ and $\liminf_j X_j$ can be obtained by combinations of sup and inf operations on X_j's as

$$\limsup_j X_j = \inf_n \left(\sup_{j\geq n} X_j \right) \quad \text{and} \quad \liminf_j X_j = \sup_n \left(\inf_{j\geq n} X_j \right),$$

they are extended-valued random variables. □

Since $\limsup_j X_j$ and $\liminf_j X_j$ are extended-valued random variables, the set

$$\begin{aligned}
\Lambda &= \left(\limsup_j X_j = \liminf_j X_j \right) \\
&= \left(\limsup_j X_j = \liminf_j X_j = \infty \right) \cup \left(\limsup_j X_j = \liminf_j X_j = -\infty \right) \\
&\cup \left(\limsup_j X_j \in R, \liminf_j X_j \in R, (\limsup_j X_j - \liminf_j X_j) = 0 \right)
\end{aligned}$$

is an event. It is clear that for $\omega \in \Lambda$, the limit $\lim_{j\to\infty} X_j(\omega)$ exists and equals to $\limsup_j X_j(\omega)$. If $\mathcal{P}(\Lambda) = 1$, we say that the sequence $\{X_j, j \geq 1\}$ converges w.p.1 to an extended-valued r.v. X defined by

$$X(\omega) = \begin{cases} \limsup_j X_j(\omega), & \text{if } \omega \in \Lambda, \\ 0, & \text{otherwise.} \end{cases}$$

This limiting r.v. X will be denoted as $\lim_{j\to\infty} X_j$.

Definition 1.3.12 *An r.v. X is called* discrete *if there is a countable set B in R such that $X^{-1}(B) = \Omega$. If B is finite, then X is also called simple.* □

We next discuss a more direct way to look at a discrete random variable. Let Λ be an event in $\mathcal{F}$. The *indicator function* 1_Λ of Λ is the random variable defined by

$$1_\Lambda(\omega) = \begin{cases} 1, & \text{if } \omega \in \Lambda, \\ 0, & \text{otherwise.} \end{cases}$$

Let X be a discrete r.v. with values x_j's. Let Λ_j be the event $(X = x_j)$ for all j. Λ_j's form a countable measurable partition [10] of Ω by definition. Thus X can be represented as

$$X(\omega) = \sum_j x_j 1_{\Lambda_j}(\omega).$$

We shall say that X belongs to the weighted partition $\{\Lambda_j, x_j\}$. It can be seen that the σ-algebra $\mathcal{F}(X)$ generated by X is just the collection of all unions of Λ_j's. Thus, any event induced by X can be specified by elementary events Λ_j in the partition.

[10] A partition of Ω is a collection $\{A_j, j \in I\}$ of subsets of Ω indexed by a set I such that

1. A_j's are mutually disjoint,
2. $\cup_{j\in I} A_j = \Omega$.

A partition $\{A_j, j \in I\}$ is countable if the index set I is a countable set. A partition $\{A_j, j \in I\}$ is measurable if each A_j is an event.

1.4 Expectation

It is much easier to deal with simple r.v.'s at first.

Definition 1.4.1 *For a simple r.v. Y belonging to a weighted finite partition $\{\Lambda_j; y_j\}_{j=1}^n$, the expectation $\mathcal{E}(Y)$ of Y is defined by*

$$\mathcal{E}(Y) = \sum_{j=1}^{n} y_j \mathcal{P}(\Lambda_j). \tag{1.3}$$

□

Since

$$\mathcal{P}(A) = \mathcal{E}(1_A)$$

where 1_A is the indicator function of event A, probabilities can be treated as expectations. It is usually convenient to denote $\mathcal{E}(Y)$ as an abstract integral [11]

$$\mathcal{E}(Y) = \int_\Omega Y(\omega)\mathcal{P}(d\omega).$$

Lemma 1.4.2 *Let X and Y be two simple r.v.'s belonging to $\{A_i; x_i\}_{i=1}^n$ and $\{B_j; y_j\}_{j=1}^m$ respectively. Let a, b be two real numbers. Then*

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y)$$

or in abstract integral form

$$\int_\Omega (aX + bY)(\omega)\mathcal{P}(d\omega) = a\int_\Omega X(\omega)\mathcal{P}(d\omega) + b\int_\Omega Y(\omega)\mathcal{P}(d\omega).$$

Proof. Let $E_{ij} = A_i \cap B_j$, $\forall\, i, j$. Then $aX + bY$ is a simple r.v. belonging to the weighted finite partition $\{E_{ij}; ax_i + by_j\}$. Thus

$$\begin{aligned}
\int_\Omega (aX + bY)(\omega)\mathcal{P}(d\omega) &= \sum_{i,j}(ax_i + by_j)\mathcal{P}(E_{ij}) \\
&= a\sum_i x_i \sum_j \mathcal{P}(E_{ij}) + b\sum_j y_j \sum_i \mathcal{P}(E_{ij}) \\
&= a\sum_i x_i \mathcal{P}(A_i) + b\sum_j y_j \mathcal{P}(B_j) \\
&= a\int_\Omega X(\omega)\mathcal{P}(d\omega) + b\int_\Omega Y(\omega)\mathcal{P}(d\omega),
\end{aligned}$$

where the third equality follows from the countable additivity of $\mathcal{P}$. This completes the proof. □

11

$$\int_\Omega Y(\omega)\mathcal{P}(d\omega)$$

is called the Lebesgue integral of Y over Ω with respect to the p.m. $\mathcal{P}$.

We next consider non-negative extended-valued random variables.

Definition 1.4.3 *Let Y be a non-negative extended-valued random variable and $\mathcal{S}(Y)$ be the set of all non-negative simple r.v.'s Z such that $0 \le Z \le Y$. The expectation $\mathcal{E}(Y)$ of Y is defines as*

$$\mathcal{E}(Y) = \sup_{Z \in \mathcal{S}(Y)} \mathcal{E}(Z). \tag{1.4}$$

□

It can be seen that $\mathcal{E}(Y)$ exists and may be finite or infinite for any $Y \ge 0$. The following properties of expectation can be deduced directly from Definition 1.4.3.

Lemma 1.4.4 *Let X, Y be extended-valued non-negative r.v.'s and c a non-negative number. Then*

1. *if $X \le Y$, then $\mathcal{E}(X) \le \mathcal{E}(Y)$;*
2. *$\mathcal{E}(cX) = c\mathcal{E}(X)$.*

Proof. For the first property, we have

$$\mathcal{E}(X) = \sup_{Z \in \mathcal{S}(X)} \mathcal{E}(Z) \le \sup_{Z \in \mathcal{S}(Y)} \mathcal{E}(Z) = \mathcal{E}(Y)$$

since $\mathcal{S}(X) \subseteq \mathcal{S}(Y)$. From Lemma 1.4.2, the second property holds for a simple non-negative r.v. Z. Then for a general non-negative r.v. X, we have

$$\mathcal{E}(cX) = \sup_{Z' \in \mathcal{S}(cX)} \mathcal{E}(Z') = \sup_{Z \in \mathcal{S}(X)} \mathcal{E}(cZ) = c \sup_{Z \in \mathcal{S}(X)} \mathcal{E}(Z) = c\mathcal{E}(X)$$

since $\mathcal{S}(cX) = c\mathcal{S}(X)$. □

In general, the definition in (1.4) is conceptually simple but technically hard. To develop a workable way to compute $\mathcal{E}(Y)$, we first prove an important theorem.

Theorem 1.4.5 [Lebesgue's monotone convergence theorem] *Let $\{Y_n\}$ be an increasing sequence of non-negative extended-valued r.v.'s*

$$0 \le Y_1(\omega) \le Y_2(\omega) \le \ldots \le \infty, \; \forall\, \omega \in \Omega.$$

Let

$$Y(\omega) = \lim_{n \to \infty} Y_n(\omega).$$

Then Y is an extended-valued r.v. and

$$\mathcal{E}(Y) = \lim_{n \to \infty} \mathcal{E}(Y_n).$$

or in abstract integral form

$$\int_\Omega Y(\omega)\mathcal{P}(d\omega) = \lim_{n \to \infty} \int_\Omega Y_n(\omega)\mathcal{P}(d\omega).$$

Proof. The measurability of Y follows from Theorem 1.3.11 and the discussion followed. We assume that $\mathcal{P}(\Lambda) > 0$. By Property 1 of Lemma 1.4.4, $\{\mathcal{E}(Y_n)\}$ is an increasing sequence of non-negative numbers and then converges to a number $\alpha \in [0, \infty]$. Also since $Y_n \leq Y$ and then $\mathcal{E}(Y_n) \leq \mathcal{E}(Y)$ for all n, we have

$$\alpha = \lim_{n\to\infty} \mathcal{E}(Y_n) \leq \mathcal{E}(Y). \tag{1.5}$$

Now let Z be any non-negative simple r.v. in $\mathcal{S}(Y)$. Given a constant c, $0 < c < 1$, define

$$A_n = \{\omega \in \Omega | Y_n(\omega) \geq cZ(\omega)\}.$$

It is clear that $\{A_n\}$ is a monotone increasing sequence of events. Also if $Y(\omega) = 0$ then $\omega \in A_1$ and if $Y(\omega) > 0$ then $cZ(\omega) < Y(\omega)$ since $0 < c < 1$, hence $\omega \in A_n$ for some n. Thus we have $A_n \uparrow \Omega$. Since $Y_n \geq 1_{A_n} Y_n \geq 1_{A_n} cZ$, we have

$$\mathcal{E}(Y_n) \geq \mathcal{E}(1_{A_n} Y_n) \geq \mathcal{E}(1_{A_n} cZ) \tag{1.6}$$

by Property 1 of Lemma 1.4.4. But,

$$\mathcal{E}(1_{A_n} cZ) = \mathcal{E}(cZ) - \mathcal{E}(1_{A_n^c} cZ) \tag{1.7}$$

by $1 = 1_\Lambda + 1_{\Lambda^c}$ and Lemma 1.4.2, and

$$\mathcal{E}(1_{A_n^c} cZ) \leq d \cdot \mathcal{E}(1_{A_n^c}) = d \cdot \mathcal{P}(A_n^c), \tag{1.8}$$

where $d = c \cdot \max_j z_j$ with z_j's being possible values of Z. By (1.6)–(1.8), we have

$$\mathcal{E}(Y_n) \geq \mathcal{E}(cZ) - d \cdot \mathcal{P}(A_n^c).$$

By the monotone property of $\mathcal{P}$, we have

$$\alpha = \lim_{n\to\infty} \mathcal{E}(Y_n) \geq c\mathcal{E}(Z) - d \lim_{n\to\infty} \mathcal{P}(A_n^c) = c\mathcal{E}(Z).$$

Since the above inequality holds for any $c < 1$, we have

$$\alpha \geq \mathcal{E}(Z)$$

and then

$$\alpha \geq \sup_{Z \in \mathcal{S}(Y)} \mathcal{E}(Z) = \mathcal{E}(Y). \tag{1.9}$$

By (1.5) and (1.9), the proof is now completed. □

Here is a workable way to compute expectation as an application of the monotone convergence theorem.

Lemma 1.4.6 *A non-negative extended-valued random variable Y can be approximated by an increasing sequence of non-negative simple r.v.'s Y_n defined as*

$$Y_n(\omega) = \begin{cases} \frac{i}{2^n}, & \text{if } \frac{i}{2^n} \leq Y(\omega) < \frac{i+1}{2^n} \text{ and } 0 \leq i < n2^n, \\ n, & \text{otherwise,} \end{cases} \tag{1.10}$$

and

$$\mathcal{E}(Y) = \lim_{n\to\infty} \mathcal{E}(Y_n).$$

□

Theorem 1.4.7 [Linearity] *Let X and Y be two non-negative r.v.'s and a, b be two non-negative numbers. Then*

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y)$$

or in abstract integral form

$$\int_\Omega (aX + bY)(\omega)\mathcal{P}(d\omega) = a\int_\Omega X(\omega)\mathcal{P}(d\omega) + b\int_\Omega Y(\omega)\mathcal{P}(d\omega).$$

Proof. The case that both X and Y are non-negative simple r.v.'s has been proved in Lemma 1.4.2. Now for general non-negative r.v.'s X and Y, let $\{X_i\}$ and $\{Y_i\}$ be two approximation sequences of non-negative simple r.v.'s as in Lemma 1.4.6 for X and Y respectively. Let $Z_i = aX_i + bY_i$. Then $\{Z_i\}$ is an approximation sequence of $aX + bY$ and

$$\int_\Omega (Z_i)(\omega)\mathcal{P}(d\omega) = a\int_\Omega X_i(\omega)\mathcal{P}(d\omega) + b\int_\Omega Y_i(\omega)\mathcal{P}(d\omega)$$

by Lemma 1.4.2. Finally by monotone convergence theorem, we have

$$\int_\Omega (cX + dY)(\omega)\mathcal{P}(d\omega) = a\int_\Omega X(\omega)\mathcal{P}(d\omega) + b\int_\Omega Y(\omega)\mathcal{P}(d\omega).$$

□

The following theorem is an extension.

Theorem 1.4.8 *If X_n's be a sequence of non-negative r.v.'s and*

$$X(\omega) = \sum_{n=1}^\infty X_n(\omega),$$

then

$$\mathcal{E}(X) = \sum_{n=1}^\infty \mathcal{E}(X_n)$$

or in abstract integral form

$$\int_\Omega X(\omega)\mathcal{P}(d\omega) = \sum_{n=1}^\infty \int_\Omega X_n(\omega)\mathcal{P}(d\omega).$$

Proof. Let

$$Y_m(\omega) = \sum_{n=1}^{m} X_n(\omega).$$

By Theorem 1.4.7, we have

$$\int_\Omega Y_m(\omega)\mathcal{P}(d\omega) = \sum_{n=1}^{m} \int_\Omega X_n(\omega)\mathcal{P}(d\omega).$$

The proof is now completed by the monotone convergence theorem. $\square$

Theorem 1.4.9 [Countable additivity of abstract integral] *Let X be a non-negative r.v. and Λ be the union of countably many disjoint events Λ_i in Ω. Then*

$$\mathcal{E}(1_\Lambda X) = \sum_i \mathcal{E}(1_{\Lambda_i} X)$$

or in abstract integral form

$$\int_\Lambda X(\omega)\mathcal{P}(d\omega) = \sum_i \int_{\Lambda_i} X(\omega)\mathcal{P}(d\omega).$$

Proof. We first note that

$$(1_\Lambda X)(\omega) = \sum_{i=1}^{\infty} (1_{\Lambda_i} X)(\omega).$$

By Theorem 1.4.8, we have

$$\mathcal{E}(1_\Lambda X) = \sum_{i=1}^{\infty} \mathcal{E}(1_{\Lambda_i} X).$$

This completes the proof. $\square$

Corollary 1.4.10 *For a non-negative r.v. X, $\mathcal{E}(X) < +\infty$ if and only if $\mathcal{E}(1_\Lambda X) < +\infty$ for any event Λ in $\mathcal{F}$.*

Proof. Since Ω is a disjoint union of Λ and Λ^c, we have $1 = 1_\Lambda + 1_{\Lambda^c}$ and

$$\mathcal{E}(X) = \mathcal{E}(1_\Lambda X) + \mathcal{E}(1_{\Lambda^c} X)$$

by Theorem 1.4.9. Thus $\mathcal{E}(X) < +\infty$ if and only if $\mathcal{E}(1_\Lambda X) < +\infty$ and $\mathcal{E}(1_{\Lambda^c} X) < +\infty$ for any Λ in $\mathcal{F}$. This completes the proof. $\square$

We now consider the expectation of a general r.v. Y. By letting

$$Y^+ = Y \vee 0 \quad \text{and} \quad Y^- = (-Y) \vee 0,$$

we have

$$Y = Y^+ - Y^-.$$

The expectation $\mathcal{E}(Y)$ of Y is said to be well-defined if either $\mathcal{E}(Y^+)$ or $\mathcal{E}(Y^-)$ is finite and is defined as

$$\mathcal{E}(Y) = \mathcal{E}(Y^+) - \mathcal{E}(Y^-). \tag{1.11}$$

Since $|Y| = Y^+ + Y^-$, we have $\mathcal{E}(Y) < +\infty$ if and only if $\mathcal{E}(|Y|) < +\infty$.

Theorem 1.4.11 [Linearity] *Let X and Y be two r.v.'s with finite $\mathcal{E}(X)$ and $\mathcal{E}(Y)$ and a, b be two real numbers. Then $aX + bY$ has finite $\mathcal{E}(aX + bY)$ and*

$$\mathcal{E}(aX + bY) = a\mathcal{E}(X) + b\mathcal{E}(Y) \tag{1.12}$$

or in abstract integral form

$$\int_\Omega (aX + bY)(\omega)\mathcal{P}(d\omega) = a\int_\Omega X(\omega)\mathcal{P}(d\omega) + b\int_\Omega Y(\omega)\mathcal{P}(d\omega).$$

Proof. Without loss of generality, we may assume that a, b are non-negative. Let $Z = aX + bY$. Then $|Z| \le a|X| + b|Y|$ and by Lemma 1.4.4, we have

$$\mathcal{E}(|Z|) \le a\mathcal{E}(|X|) + b\mathcal{E}(|Y|) < +\infty$$

which implies that Z has finite expectation. Since

$$Z^+ - Z^- = aX^+ - aX^- + bY^+ - bY^-$$

or

$$Z^+ + aX^- + bY^- = aX^+ + bY^+ + Z^-,$$

we have

$$\mathcal{E}(Z^+) + a\mathcal{E}(X^-) + b\mathcal{E}(Y^-) = a\mathcal{E}(X^+) + b\mathcal{E}(Y^+) + \mathcal{E}(Z^-)$$

by Theorem 1.4.7 and the second property of Lemma 1.4.4. By re-arranging the equality, we obtain (1.12). □

1.5 Conditional Expectation

In a random experiment, observations (i.e. r.v.'s) commonly relate to each other. An useful way to characterize such dependence is *conditional expectation.* We shall introduce the concept of conditional expectation in two stages, from local to global.

As a passage, we firstly introduce *conditional probability measure.* Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space and Λ an event with $\mathcal{P}(\Lambda) \neq 0$.

Definition 1.5.1 *For any event $A \in \mathcal{F}$, the* conditional probability $\mathcal{P}(A \mid \Lambda)$ *of* A *on (assuming)* Λ *is the ratio*

$$\mathcal{P}(A \mid \Lambda) = \frac{\mathcal{P}(A \cap \Lambda)}{\mathcal{P}(\Lambda)}.$$

□

This conditional probability assignment $\mathcal{P}(A \mid \Lambda)$ for every event $A \in \mathcal{F}$ turns out to be a probability measure on the measurable space $(\Omega, \mathcal{F})$ as can be seen from the following three fundamental properties:

1. $\mathcal{P}(A \mid \Lambda) \geq 0, \forall A \in \mathcal{F}$.
2. $\mathcal{P}(\Omega \mid \Lambda) = 1$.
3. For mutually disjoint events $A_j, j = 1, 2, \cdots \in \mathcal{F}$,

$$\mathcal{P}(A_1 \cup A_2 \cup \cdots \mid \Lambda) = \mathcal{P}(A_1 \mid \Lambda) + \mathcal{P}(A_2 \mid \Lambda) + \cdots .$$

This probability measure will be called the conditional p.m. relative to the event Λ and denoted as $\mathcal{P}_\Lambda$. $\mathcal{P}_\Lambda$ is a p.m. concentrated on Λ in the measurable space $(\Omega.\mathcal{F})$.

Conditional expectation of an r.v. Y relative to an event Λ will be treated as the expectation of Y relative to the conditional p.m. $\mathcal{P}_\Lambda$. Thus for a simple r.v. Y belonging to a weighted finite partition $\{\Lambda_j; y_j\}_{j=1}^n$, the conditional expectation $\mathcal{E}_\Lambda(Y)$ of Y relative to an event Λ is defined, as in Definition 1.4.1,

$$\mathcal{E}_\Lambda(Y) \equiv \sum_{j=1}^n y_j \mathcal{P}_\Lambda(\Lambda_j) = \frac{1}{\mathcal{P}(\Lambda)} \sum_{j=1}^n y_j \mathcal{P}(\Lambda \cap \Lambda_j). \tag{1.13}$$

For convenience, we shall define $\mathcal{E}_\Lambda(Y) = 0$ when $\mathcal{P}(\Lambda) = 0$. In this case, $\mathcal{P}_\Lambda \equiv 0$ is clearly not a probability measure. Since

$$\mathcal{P}_\Lambda(A) = \mathcal{E}_\Lambda(1_A)$$

where 1_A is the indicator function of event A, conditional probabilities can be treated as conditional expectations. Next for a non-negative extended-valued random variable Y, the conditional expectation $\mathcal{E}_\Lambda(Y)$ of Y relative to Λ is

$$\mathcal{E}_\Lambda(Y) = \sup_{Z \in \mathcal{S}(Y)} \mathcal{E}_\Lambda(Z) \tag{1.14}$$

by (1.4).

Theorem 1.5.2 *For a non-negative extended-valued r.v.* Y*, we have*

$$\mathcal{P}(\Lambda)\mathcal{E}_\Lambda(Y) = \mathcal{E}(1_\Lambda Y),$$

which, in abstract integral form, is

$$\mathcal{E}_\Lambda(Y) = \frac{1}{\mathcal{P}(\Lambda)} \int_\Lambda Y(\omega)\mathcal{P}(d\omega).$$

if $\mathcal{P}(\Lambda) > 0$.

Proof. We first consider a simple non-negative r.v. Z belonging to a weighted finite partition $\{\Lambda_i; z_i\}_{i=1}^n$. Then $1_\Lambda Z$ is a simple r.v. belonging to weighted finite partition $\{\Lambda_1 \cap \Lambda, \ldots, \Lambda_n \cap \Lambda, \Lambda^c; z_1, \ldots, z_n, 0\}$. Thus we have

$$\mathcal{E}(1_\Lambda Z) = \sum_{i=1}^n z_i \mathcal{P}(\Lambda_i \cap \Lambda) = \mathcal{P}(\Lambda)\mathcal{E}_\Lambda(Z),$$

where the second equality is from (1.13). Then for a general non-negative r.v. Y, we have

$$\mathcal{E}(1_\Lambda Y) = \sup_{Z' \in \mathcal{S}(1_\Lambda Y)} \mathcal{E}(Z') = \sup_{Z \in \mathcal{S}(Y)} \mathcal{E}(1_\Lambda Z) = \mathcal{P}(\Lambda) \sup_{Z \in \mathcal{S}(Y)} \mathcal{E}_\Lambda(Z) = \mathcal{P}(\Lambda)\mathcal{E}_\Lambda(Y).$$

since $\mathcal{S}(1_\Lambda Y) = 1_\Lambda \mathcal{S}(Y)$. □

Finally for a general random variable Y, the conditional expectation $\mathcal{E}_\Lambda(Y)$ of Y relative to Λ is

$$\mathcal{E}_\Lambda(Y) = \mathcal{E}_\Lambda(Y^+) - \mathcal{E}_\Lambda(Y^-) \tag{1.15}$$

by (1.11) if either $\mathcal{E}_\Lambda(Y^+)$ or $\mathcal{E}_\Lambda(Y^-)$ is finite. It is easy to see that Theorem 1.5.2 is valid for a general random variable Y with well-defined conditional expectation $\mathcal{E}_\Lambda(Y)$ of Y relative to Λ. This theorem says that conditional expectation of an r.v. relative to an event can be treated as the expectation of a related random variable.

The usefulness of conditional expectations can be easily seen from the notion of partition. Let $\{\Lambda_i\}_{i=1}^\infty$ be a countable measurable partition of the sample space Ω. In many cases, the *a priori* probabilities $\mathcal{P}(\Lambda_i)$ about the partition $\{\Lambda_i\}_{i=1}^\infty$ is known. Furthermore, it may be easy to obtain the conditional expectations $\mathcal{E}_{\Lambda_i}(Y)$ of an r.v. Y relative to each event Λ_i, instead of the expectation $\mathcal{E}(Y)$ of the r.v. Y itself. An application of Theorem 1.4.9 to this partition gives

$$\mathcal{P}(A) = \mathcal{E}(1_A) = \sum_i \mathcal{P}(\Lambda_i)\mathcal{E}_{\Lambda_i}(1_A) = \sum_i \mathcal{P}(\Lambda_i)\mathcal{P}_{\Lambda_i}(A) \tag{1.16}$$

which is usually called the *total probability theorem* for an event A. In fact, there is much more information which can be deduced from $\mathcal{E}_{\Lambda_i}(Y)$'s.

Let $\mathcal{G}$ be the σ-algebra generated by the partition $\{\Lambda_i, i \geq 1\}$, i.e. the collection of all unions of Λ_i's. $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$ and contains all events related to members (sometimes called atoms) of the partition. The conditional expectation $\mathcal{E}_{\Lambda_i}(Y)$ of an r.v. Y relative to Λ_i can be thought as a smoothed (averaged) version of Y observed by all samples ω in Λ. To describe the totality of these smoothed versions (i.e. conditional expectations) of Y over various parts Λ_i of the sample space Ω, we shall introduce a new r.v. as follows.

Definition 1.5.3 *Let Y be a non-negative r.v. or an r.v. with finite expectation. The conditional expectation $\mathcal{E}(Y|\mathcal{G})$ of Y relative to the σ-algebra $\mathcal{G}$, which is generated by a countable measurable partition $\{\Lambda_i\}$ of Ω, is defined as the following r.v.*

$$\mathcal{E}(Y|\mathcal{G})(\omega) = \sum_i \mathcal{E}_{\Lambda_i}(Y) 1_{\Lambda_i}(\omega).$$

□

If Y is non-negative, then $\mathcal{E}(Y|\mathcal{G})$ is non-negative and may be extended-valued. If Y has finite expectation, then $\mathcal{E}_{\Lambda_i}(Y)$ is finite for each Λ_i by Corollary 1.4.10 and then $\mathcal{E}(Y|\mathcal{G})$ is a finite-valued random variable. Thus, we have

$$\mathcal{E}(Y|\mathcal{G})(\omega) = \mathcal{E}(Y^+|\mathcal{G})(\omega) - \mathcal{E}(Y^-|\mathcal{G})(\omega) \tag{1.17}$$

where both $\mathcal{E}(Y^+|\mathcal{G})$ and $\mathcal{E}(Y^-|\mathcal{G})$ are finite-valued. Furthermore, $\mathcal{E}(Y|\mathcal{G})$ has finite expectation, equal to $\mathcal{E}(Y)$ as will be shown in Theorem 1.5.8.

Definition 1.5.4 *Let $\mathcal{G}$ be a σ-subalgebra of $\mathcal{F}$. An (extended-valued) r.v. Y is said to be $\mathcal{G}$-measurable, denoted as $Y \in \mathcal{G}$ (reads as Y belongs to $\mathcal{G}$), if $\mathcal{F}(Y) \subseteq \mathcal{G}$.* □

Theorem 1.5.5 *Let Y be a non-negative r.v. or an r.v. with finite expectation and $\mathcal{G}$ be the σ-algebra generated by a countable measurable partition $\{\Lambda_i\}$ of Ω. If Y belongs to $\mathcal{G}$, then Y is a discrete r.v. belonging to a weighted partition $\{\Lambda_i; y_i\}$ and $Y = \mathcal{E}(Y|\mathcal{G})$ w.p.1, in particular, $y_i = \mathcal{E}_{\Lambda_i}(Y)$ if $\mathcal{P}(\Lambda_i) > 0$.*

Proof. Since Y belongs to $\mathcal{G}$, each event $(Y = y)$, $y \in [-\infty, +\infty]$, is a member of $\mathcal{G}$ and is the union of a (possibly empty) subcollection of Λ_i's. This implies that Y is constant over each Λ_i. Thus, we have

$$Y = \sum_i y_i 1_{\Lambda_i}$$

belonging to a weighted partition $\{\Lambda_i; y_i\}$. Since for $\mathcal{P}(\Lambda_i) > 0$, $\mathcal{E}_{\Lambda_j}(Y) = \mathcal{E}(1_{\Lambda_j} Y)/\mathcal{P}(\Lambda_j)$ $= \mathcal{E}(y_j 1_{\Lambda_j})/\mathcal{P}(\Lambda_j) = y_j$, we have $Y = \mathcal{E}(Y|\mathcal{G})$ w.p.1. This completes the proof. □

We now prove the fundamental properties of conditional expectation $\mathcal{E}(Y|\mathcal{G})$.

Theorem 1.5.6 *Let Y be a non-negative r.v. or an r.v. with finite expectation and $\mathcal{G}$ be the σ-algebra generated by a countable measurable partition $\{\Lambda_i\}$ of Ω. Then*

1. *$\mathcal{E}(Y|\mathcal{G})$ belongs to $\mathcal{G}$.*
2. *$\mathcal{E}_\Lambda(Y) = \mathcal{E}_\Lambda(\mathcal{E}(Y|\mathcal{G}))$ for any $\Lambda \in \mathcal{G}$.*

Furthermore, any non-negative r.v. or r.v. with finite expectation, which satisfies the above two properties, must be equal to $\mathcal{E}(Y|\mathcal{G})$ w.p.1.

Proof. Since $\mathcal{E}(Y|\mathcal{G})$ is a discrete r.v. belonging to the weighted partition $\{\Lambda_i; \mathcal{E}_{\Lambda_i}(Y)\}$ by definition, it belongs to $\mathcal{G}$. Let $\Lambda = \cup_j \Lambda_{i_j}$ be in $\mathcal{G}$ and $\mathcal{P}(\Lambda) > 0$. Then

$$\mathcal{E}_\Lambda(1_{\Lambda_i}) = \mathcal{P}_\Lambda(\Lambda_i) = \frac{\mathcal{P}(\Lambda \cap \Lambda_i)}{\mathcal{P}(\Lambda)} = \begin{cases} \frac{\mathcal{P}(\Lambda_i)}{\mathcal{P}(\Lambda)}, & \text{if } \Lambda_i \subseteq \Lambda, \\ 0, & \text{otherwise.} \end{cases}$$

and for non-negative r.v. Y,

$$\mathcal{E}_\Lambda(\mathcal{E}(Y|\mathcal{G})) = \sum_i \mathcal{E}_{\Lambda_i}(Y)\mathcal{E}_\Lambda(1_{\Lambda_i}) = \frac{1}{\mathcal{P}(\Lambda)}\sum_j \mathcal{P}(\Lambda_{i_j})\mathcal{E}_{\Lambda_{i_j}}(Y) = \mathcal{E}_\Lambda(Y)$$

where the first equality follows from Theorem 1.4.8 and the last equality from Theorem 1.4.9. Next for an r.v. Y with finite expectation, we have

$$\begin{aligned}
\mathcal{E}_\Lambda(Y) &= \mathcal{E}_\Lambda(Y^+) - \mathcal{E}_\Lambda(Y^-) \ \text{ by Corollary 1.4.10 and (1.15)} \\
&= \mathcal{E}_\Lambda(\mathcal{E}(Y^+|\mathcal{G})) - \mathcal{E}_\Lambda(\mathcal{E}(Y^-|\mathcal{G})) \ \text{ as just proved} \\
&= \mathcal{E}_\Lambda(\mathcal{E}(Y^+|\mathcal{G}) - \mathcal{E}(Y^-|\mathcal{G})) \ \text{ by Theorem 1.4.11} \\
&= \mathcal{E}_\Lambda(\mathcal{E}(Y|\mathcal{G})) \ \text{ by (1.17).}
\end{aligned}$$

Suppose that Z is a non-negative r.v. or an r.v. with finite expectation which belongs to $\mathcal{G}$ and

$$\mathcal{E}_\Lambda(Y) = \mathcal{E}_\Lambda(Z) \tag{1.18}$$

for any $\Lambda \in \mathcal{G}$. By Theorem 1.5.5, we have $Z = \sum_i z_i 1_{\Lambda_i}$ and then for $\mathcal{P}(\Lambda_i) > 0$,

$$z_i = \mathcal{E}_{\Lambda_i}(Z) = \mathcal{E}_{\Lambda_i}(Y)$$

by (1.18) with $\Lambda = \Lambda_i$. Thus $Z = \mathcal{E}(Y|\mathcal{G})$ w.p.1, which completes the proof. □

The following theorem is a further characterization of conditional expectation.

Theorem 1.5.7 *Let Y be a non-negative r.v. or an r.v. with finite expectation and $\mathcal{F}(X)$ be the σ-algebra generated by a discrete r.v. X. Then the conditional expectation $\mathcal{E}(Y|\mathcal{F}(X))$ is a function of X*

$$\mathcal{E}(Y|\mathcal{F}(X)) = f(X)$$

for some Borel measurable function f.

Proof. Let $\{x_i\}$ be the set of all possible values taken by X. Thus $\mathcal{F}(X)$ is generated by the countable measurable partition $\{(X = x_i)\}$ of Ω. By definition, the conditional expectation $\mathcal{E}(Y|\mathcal{G})$ is a function $f(X)$ of X where

$$f(x) = \begin{cases} \mathcal{E}_{(X=x_i)}(Y), & \text{if } x = x_i, \\ 0, & \text{otherwise,} \end{cases}$$

is Borel measurable by Theorem 1.3.5. □

The conditional expectation $\mathcal{E}(Y|\mathcal{F}(X))$ is also denoted as $\mathcal{E}(Y|X)$ and called the conditional expectation of Y relative to X.

We now list some basic properties of conditional expectation in the following theorem.

Theorem 1.5.8 *Let Y, Y_i's be non-negative r.v.'s or r.v.'s with finite expectations, a, b be non-negative or arbitrary real numbers and $\mathcal{G}$ be the σ-algebra generated by a countable measurable partition $\{\Lambda_i\}$ of Ω. Then*

1. (Linearity) $\mathcal{E}(aY_1 + bY_2|\mathcal{G}) = a\mathcal{E}(Y_1|\mathcal{G}) + b\mathcal{E}(Y_2|\mathcal{G})$.
2. (Comparability) *If* $Y_1 \leq Y_2$, *then* $\mathcal{E}(Y_1|\mathcal{G}) \leq \mathcal{E}(Y_2|\mathcal{G})$.
3. (Absolute-valued dominancy) $|\mathcal{E}(Y|\mathcal{G})| \leq \mathcal{E}(|Y| \mid \mathcal{G})$.
4. (Generalized total probability theorem) $\mathcal{E}(Y) = \mathcal{E}\left(\mathcal{E}(Y|\mathcal{G})\right)$.
5. (Monotone convergence property) *If* $Y_n \geq Z$ *where* $\mathcal{E}(Z) > -\infty$ *and* $Y_n \uparrow Y$, *then* $\mathcal{E}(Y_n|\mathcal{G}) \geq \mathcal{E}(Z|\mathcal{G})$ *and* $\mathcal{E}(Y_n|\mathcal{G}) \uparrow \mathcal{E}(Y|\mathcal{G})$.
6. (Series convergence property) *If* $Y_n \geq 0$ *and* $Y = \sum_n Y_n$, *then* $\mathcal{E}(Y|\mathcal{G}) = \sum_n \mathcal{E}(Y_n|\mathcal{G})$.

Proof.

1. The linearity property is a direct extension of Theorems 1.4.7 and 1.4.11.
2. The comparability property is a direct extension of the first property of Lemma 1.4.4 for non-negative Y_i's. For Y_i's with finite expectations, $0 \leq (Y_2 - Y_1)$ implies $0 \leq \mathcal{E}(Y_2 - Y_1|\mathcal{G})$. By linearity property, we have $\mathcal{E}(Y_1|\mathcal{G}) \leq \mathcal{E}(Y_2|\mathcal{G})$.
3. The absolute-valued dominancy property is trivial for non-negative Y. For Y with finite expectation, we have from (1.17)
$$\begin{aligned} |\mathcal{E}(Y|\mathcal{G})| &= |\mathcal{E}(Y^+|\mathcal{G}) - \mathcal{E}(Y^-|\mathcal{G})| \\ &\leq \max(\mathcal{E}(Y^+|\mathcal{G}), \mathcal{E}(Y^-|\mathcal{G})) \\ &\leq \mathcal{E}(|Y| \mid \mathcal{G}), \quad \text{since } |Y| \geq \max(Y^+, Y^-) \text{ and by the comparability property.} \end{aligned}$$
4. The generalized total probability theorem is a direct application of Theorem 1.4.9 with $\Lambda = \Omega$ for non-negative Y. For Y with finite expectation,
$$\mathcal{E}(Y) = \mathcal{E}(Y^+) - \mathcal{E}(Y^-) = \mathcal{E}(\mathcal{E}(Y^+|\mathcal{G})) - \mathcal{E}(\mathcal{E}(Y^-|\mathcal{G})) = \mathcal{E}(\mathcal{E}(Y^+|\mathcal{G}) - \mathcal{E}(Y^-|\mathcal{G})) = \mathcal{E}(\mathcal{E}(Y|\mathcal{G})).$$
5. The monotone convergence property is a direct extension of Theorem 1.4.5.
6. The series convergence property is a direct extension of Theorem 1.4.8. □

The following definition generalizes the concept of statistical independence among events.

Definition 1.5.9 *A collection $\{\mathcal{G}_i, i \in I\}$ of σ-subalgebras of $\mathcal{F}$ is called statistically independent if for any $A_j \in \mathcal{G}_j$, j in a finite subset J of I, we have*
$$\mathcal{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathcal{P}(A_j).$$
In particular, a collection of r.v.'s X_i , $i \in I$, is statistically independent if $\{\mathcal{F}(X_i), i \in I\}$ is statistically independent. □

Theorem 1.5.10 (Independence property) *Let Y be a non-negative r.v. or an r.v. with finite expectation and $\mathcal{G}$ be the σ-algebra generated by a countable measurable partition $\{\Lambda_i\}$ of Ω. If $\mathcal{F}(Y)$ is statistically independent of $\mathcal{G}$, then $\mathcal{E}(Y|\mathcal{G}) = \mathcal{E}(Y)$ w.p.1.*

Proof. We first consider an indicator function 1_A with $A \in \mathcal{F}(Y)$. For $\mathcal{P}(\Lambda_i) > 0$, we have

$$\mathcal{E}_{\Lambda_i}(1_A) = \mathcal{P}_{\Lambda_i}(A) = \frac{\mathcal{P}(\Lambda_i \cap A)}{\mathcal{P}(\Lambda_i)} = \frac{\mathcal{P}(\Lambda_i)\mathcal{P}(A)}{\mathcal{P}(\Lambda_i)} = \mathcal{P}(A) = \mathcal{E}(1_A).$$

Now for any non-negative simple r.v. $Z = \sum_{j=1}^{n} z_j 1_{A_j}$ with $A_j \in \mathcal{F}(Y)$, we have

$$\mathcal{E}_{\Lambda_i}(Z) = \sum_{j=1}^{n} z_j \mathcal{E}_{\Lambda_i}(1_{A_j}) = \sum_{j=1}^{n} z_j \mathcal{P}(A_j) = \mathcal{E}(Z)$$

for $\mathcal{P}(\Lambda_i) > 0$. Suppose that $Y \geq 0$ and $Y_n \uparrow Y$, where Y_n's are non-negative simple r.v.'s as in Lemma 1.4.6. By monotone convergence theorem, we have

$$\mathcal{E}_{\Lambda_i}(Y) = \lim_{n\to\infty} \mathcal{E}_{\Lambda_i}(Y_n) = \lim_{n\to\infty} \mathcal{E}(Y_n) = \mathcal{E}(Y)$$

for $\mathcal{P}(\Lambda_i) > 0$. By definition, we have

$$\mathcal{E}(Y|\mathcal{G}) = \mathcal{E}(Y) \text{ w.p.1.}$$

For a general Y with finite expectation, we have

$$\begin{aligned} \mathcal{E}(Y|\mathcal{G}) &= \mathcal{E}(Y^+|\mathcal{G}) - \mathcal{E}(Y^-|\mathcal{G}) \\ &= \mathcal{E}(Y^+) - \mathcal{E}(Y^-) \text{ w.p.1} \\ &= \mathcal{E}(Y), \end{aligned}$$

where the second equality follows from the fact that $\mathcal{F}(Y^+)$, $\mathcal{F}(Y^-) \subseteq \mathcal{F}(Y)$. This completes the proof. □

Theorem 1.5.11 *Let Y be a non-negative r.v. or an r.v. with finite expectation and $\mathcal{G}$ be the σ-algebra generated by a countable measurable partition $\{\Lambda_i\}$ of Ω. If Z is $\mathcal{G}$-measurable, non-negative (if Y non-negative) or with $\mathcal{E}(|YZ|) < \infty$ (if $\mathcal{E}(|Y|) < \infty$), then $\mathcal{E}(YZ|\mathcal{G}) = Z\mathcal{E}(Y|\mathcal{G})$.*

Proof. We first suppose $Y \geq 0$ and $Z \geq 0$ and then $\mathcal{E}(Y|\mathcal{G}) \geq 0$. Since Z belongs to $\mathcal{G}$, we have

$$Z = \sum_j z_j 1_{\Lambda_j}$$

by Theorem 1.5.5, where $z_j \geq 0$. Observe that for $\mathcal{P}(\Lambda_i) > 0$,

$$\mathcal{E}_{\Lambda_i}(Y 1_{\Lambda_j}) = \frac{\mathcal{E}(Y 1_{\Lambda_i} 1_{\Lambda_j})}{\mathcal{P}(\Lambda_i)} = \frac{\mathcal{E}(Y 1_{\Lambda_i \cap \Lambda_j})}{\mathcal{P}(\Lambda_i)} = \begin{cases} \mathcal{E}_{\Lambda_i}(Y), & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases}$$

By Theorem 1.4.8 and Lemma 1.4.4, we have

$$\mathcal{E}_{\Lambda_i}(YZ) = \sum_j z_j \mathcal{E}_{\Lambda_i}(Y 1_{\Lambda_j}) = z_i \mathcal{E}_{\Lambda_i}(Y)$$

and then

$$\mathcal{E}(YZ|\mathcal{G}) = \sum_i \left(z_i \mathcal{E}_{\Lambda_i}(Y)\right) 1_{\Lambda_i} = \left(\sum_i z_i 1_{\Lambda_i}\right) \cdot \left(\sum_i \mathcal{E}_{\Lambda_i}(Y) 1_{\Lambda_i}\right) = Z\mathcal{E}(Y|\mathcal{G}).$$

Now suppose that Y, YZ have finite expectations. Since Y^+Z^+, Y^-Z^+, Y^+Z^- and Y^-Z^- are all less than or equal to $|YZ|$, they have finite expectations. Then

$$\begin{aligned} \mathcal{E}(YZ|\mathcal{G}) &= \mathcal{E}(Y^+Z^+ - Y^-Z^+ - Y^+Z^- + Y^-Z^-|\mathcal{G}) \\ &= Z^+\mathcal{E}(Y^+|\mathcal{G}) - Z^+\mathcal{E}(Y^-|\mathcal{G}) - Z^-\mathcal{E}(Y^+|\mathcal{G}) + Z^-\mathcal{E}(Y^-|\mathcal{G}) \\ &= (Z^+ - Z^-) \cdot (\mathcal{E}(Y^+ - Y^-|\mathcal{G})) \\ &= Z\mathcal{E}(Y|\mathcal{G}), \end{aligned}$$

which completes the proof. □

Definition 1.5.12 *A σ-algebra $\mathcal{H}$ is called finer than another σ-algebra $\mathcal{G}$ if $\mathcal{H} \supseteq \mathcal{G}$.*

A countable measurable partition $\{A_i, i \in I\}$ of Ω is called finer than another countable measurable partition $\{B_j, j \in J\}$ of Ω if each B_j is a union of A_i's. Thus, the σ-algebra generated by A_i's is finer than the σ-algebra generated by B_j's.

Theorem 1.5.13 *Let Y be a non-negative r.v. or an r.v. with finite expectation and $\mathcal{G}$, $\mathcal{H}$ be σ-algebras generated by countable measurable partitions $\{\Lambda_i\}$, $\{A_j\}$ of Ω respectively. If $\mathcal{H}$ is finer than $\mathcal{G}$, then*

$$\mathcal{E}(Y|\mathcal{G}) = \mathcal{E}(\mathcal{E}(Y|\mathcal{H})|\mathcal{G}) \text{ w.p.1 } = \mathcal{E}(\mathcal{E}(Y|\mathcal{G})|\mathcal{H}) \text{ w.p.1.}$$

Proof. Since $\mathcal{E}(Y|\mathcal{G})$ is $\mathcal{G}$-measurable, it is also $\mathcal{H}$-measurable and then by Theorem 1.5.5,

$$\mathcal{E}(Y|\mathcal{G}) = \mathcal{E}(\mathcal{E}(Y|\mathcal{G})|\mathcal{H}) \text{ w.p.1.}$$

To show the first equality, we let $\Lambda \in \mathcal{G}$, then $\Lambda \in \mathcal{H}$. And by applying Theorem 1.5.6 twice, we have

$$\mathcal{E}_\Lambda(\mathcal{E}(\mathcal{E}(Y|\mathcal{H})|\mathcal{G})) = \mathcal{E}_\Lambda(\mathcal{E}(Y|\mathcal{H})) = \mathcal{E}_\Lambda(Y).$$

Thus

$$\mathcal{E}(Y|\mathcal{G}) = \mathcal{E}(\mathcal{E}(Y|\mathcal{H})|\mathcal{G}) \text{ w.p.1.}$$

This completes the proof. □

Up to now, we have considered conditional expectations of r.v.'s relative only to σ-subalgebras of $\mathcal{F}$ generated by countable measurable partitions of Ω. The following theorem [12] establishs the foundation to consider conditional expectations of r.v.'s relative to general σ-subalgebras of $\mathcal{F}$.

Theorem 1.5.14 *If Y is a non-negative r.v. or an r.v. with finite expectation and $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$, then there exists a non-negative r.v. or an r.v. with finite expectation, unique w.p.1 and denoted as $\mathcal{E}(Y|\mathcal{G})$, such that*

1. *$\mathcal{E}(Y|\mathcal{G})$ belongs to $\mathcal{G}$;*
2. *$\mathcal{E}_\Lambda(Y) = \mathcal{E}_\Lambda(\mathcal{E}(Y|\mathcal{G}))$, for any $\Lambda \in \mathcal{G}$.* □

In fact, $\mathcal{E}(Y|\mathcal{G})$ represents any member in an equivalent class of r.v.'s which satisfy the two properties in the above theorem and is also called the conditional expectation of Y relative to σ-algebra $\mathcal{G}$. We may regard $\mathcal{E}(Y|\mathcal{G})$ as a smoothed version of Y on events in $\mathcal{G}$.

Compared with the above theorem, the r.v. given in Definition 1.5.3 is just an explicit version of the conditional expectation of an r.v. relative to a σ-algebra generated by a countable measurable partition of Ω, as proved by Theorem 1.5.6.

We have seen that if $Y = f(X)$, then Y belongs to $\mathcal{F}(X)$ since $\mathcal{F}(Y) \subseteq \mathcal{F}(X)$. The following theorem [13] gives the converse.

Theorem 1.5.15 *Let X, Y be two random variables. If Y belongs to $\mathcal{F}(X)$, then $Y = f(X)$ for some (possibly extended-valued) Borel measurable function f.* □

Thus the conditional expectation $\mathcal{E}(Y|\mathcal{F}(X))$ is a function of X and will also be denoted as $\mathcal{E}(Y|X)$.

Theorems 1.5.8, 1.5.10, 1.5.11 and 1.5.13 still hold for the general case [14]. Finally we note that all of the statements in these theorems hold w.p.1 since the conditional expectation can only be uniquely specified w.p.1 as stated in Theorem 1.5.14.

---

[12] This theorem follows from the Radon-Nikodym theorem, See W. Rudin, *Real and Complex Analysis,* 2nd edn. New York: McGraw-Hill, 1974, pp. 129–132.

[13] See K. L. Chung, *A Course in Probability Theory,* 2nd edn. New York: Academic Press, 1974, page 299.

[14] Detailed proofs can be found in:

1. K. L. Chung, *A Course in Probability Theory,* 2nd edn. New York: Academic Press, 1974, pp. 300–304.
2. A. N. Shiryayev, *Probability.* New York: Springer-Verlag, 1984, pp. 213–217.

Chapter 2

Discrete-Time Markov Chains

2.1 Introduction

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a given probability space. And let $(S, 2^S)$ be a discrete measurable space. A measurable function from $(\Omega, \mathcal{F})$ into $(S, 2^S)$ will be called an S-valued random variable.

Definition 2.1.1 *A discrete-time* Markov chain *is a sequence* $\{X_n, n = 0, 1, 2, \ldots\}$ *of S-valued random variables such that*

$$\mathcal{P}\{X_{n+1} = s_{n+1} | X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n\} = \mathcal{P}\{X_{n+1} = s_{n+1} | X_n = s_n\}$$

for all time index $n \geq 0$ *and all* $s_0, \ldots, s_n, s_{n+1}$ *in* S*, whenever* $\mathcal{P}\{X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n\} > 0$. □

Remark 2.1.2

1. *The discrete measurable space* $(S, 2^S)$ *is called the state space of the Markov chain.*
2. *The conditional probabilities* $\mathcal{P}\{X_{n+1} = s_{n+1} | X_n = s_n\}$ *are called one-step transition probabilities and usually denoted as* $P_n(s_n, s_{n+1})$.
3. *If the conditional probabilities* $P_n(s, s')$ *are independent of the time index* n*, then the Markov chain is called homogeneous and* $P_n(s, s')$ *will be rewritten as* $P(s, s')$. □

Example 2.1.3 [Weather forcasting] *We use* X_n *to denote the weather of the* n*th day, which is rainy or fair. For simplicity, we use "1" to denote that the weather is fair and "0" that the weather is rainy. We then model* X_n*'s as a homogeneous Markov chain with state space* S *being the set* $\{0, 1\}$. *It is usually useful to specify one-step transition probabilities by a*

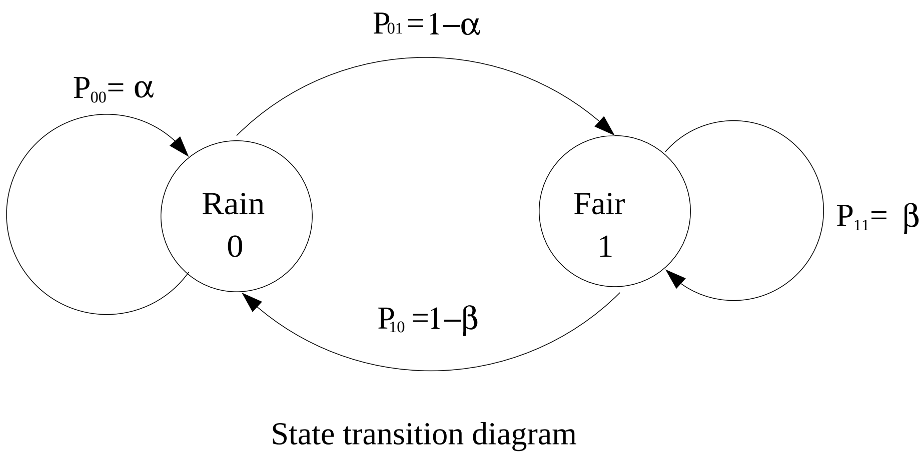

Figure 2.1: The state-transition diagram of weather forcasting.

state-transition diagram as shown in Figure 2.1. In addition, a matrix can be formed by arranging the one-step transition probabilities as follows:

$$P = \begin{bmatrix} P(0,0) & P(0,1) \\ P(1,0) & P(1,1) \end{bmatrix} = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{bmatrix}$$

which is called the one-step transition probablity matrix of the Markov chain. Note that in the second equality, we have used the fact that $P(0,0)+P(0,1) = 1$ *and* $P(1,0)+P(1,1) = 1$, *which will be shown in the following lemma.* □

Lemma 2.1.4 *For any* $s \in S$ *with* $\mathcal{P}(X_n = s) > 0$, *we have*

$$\sum_{s' \in S} P_n(s, s') = 1.$$

proof. Since $\{(X_{n+1} = s'), s' \in S\}$ is a countable partition of the sample space Ω, we have

$$\mathcal{P}_{(X_n=s)}(\Omega) = \sum_{s' \in S} \mathcal{P}_{(X_n=s)}(X_{n+1} = s')$$

by the countable additivity of the conditional p.m. $\mathcal{P}_{(X_n=s)}$ relative to the event $X_n = s$. Since $\mathcal{P}_{(X_n=s)}(\Omega) = 1$ and $\mathcal{P}_{(X_n=s)}(X_{n+1} = s') = P_n(s, s')$, the proof is completed. □

The above lemma implies that each row sum of a one-step transition probability matrix is equal to one.

Example 2.1.5 [An error model of a communication channel] *Let* X_n *denote the error status of the* n*th transmission in a digital communication system.* $X_n = 0$ *means that the* n*th transmission has been received error-freely and* $X_n = 1$ *erroneously. The state space* S *is* $\{0, 1\}$. *An error model, called Gilbert model, specifies* X_n*'s as a Markov chain with state-transition diagram similar to that in Figure 2.1.* □

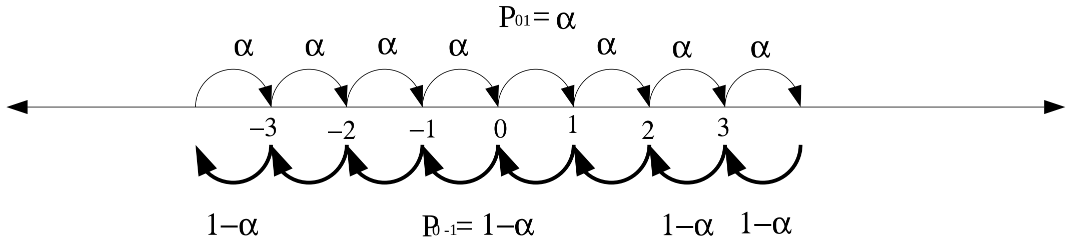

Figure 2.2: The state-transition diagram of a random walk.

Example 2.1.6 [Random walk] *Let $\{V_i, i = 1, 2, \ldots\}$ be a sequence of independent identically distributed (i.i.d.) r.v.'s with $\mathcal{P}(V_i = +1) = \alpha$ and $\mathcal{P}(V_i = -1) = 1 - \alpha$ for all i. Let X_0 be an arbitrary integer-valued random variable, independent of the random sequence $\{V_i, i = 1, 2, \ldots\}$, and*

$$X_n = \sum_{i=1}^{n} V_i, \ \forall \ n \geq 1.$$

The sequence $\{X_n, n = 0, 1, \ldots\}$ of r.v.'s is called a random walk with increment ± 1 and initial position X_0. Note that the state space S is the set of all integers. We now show that a random walk is a Markov chain. Consider $s_0, s_1, \ldots, s_{n+1}$ in S such that $\mathcal{P}\{X_0 = s_0, X_1 = s_1, \cdots, X_n = s_n\} > 0$. Let $v_i = s_i - s_{i-1}$ for all $1 \leq i \leq n+1$. Then we have

$$\begin{aligned}
& \mathcal{P}\{X_{n+1} = s_{n+1} | X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n\} \\
= & \frac{\mathcal{P}\{X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n, X_{n+1} = s_{n+1}\}}{\mathcal{P}\{X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n\}} \\
= & \frac{\mathcal{P}\{X_0 = s_0, V_1 = v_1, \ldots, V_n = v_n, V_{n+1} = v_{n+1}\}}{\mathcal{P}\{X_0 = s_0, V_1 = v_1, \ldots, V_{n-1} = v_{n-1}, V_n = v_n\}} \\
= & \frac{\mathcal{P}(X_0 = s_0)\mathcal{P}(V_1 = v_1) \ldots \mathcal{P}(V_n = v_n)\mathcal{P}(V_{n+1} = v_{n+1})}{\mathcal{P}(X_0 = s_0)\mathcal{P}(V_1 = v_1) \ldots \mathcal{P}(V_{n-1} = v_{n-1})\mathcal{P}(V_n = v_n)} \\
= & \mathcal{P}(V_{n+1} = v_{n+1}).
\end{aligned}$$

Also, we have

$$\begin{aligned}
& \mathcal{P}\{X_{n+1} = s_{n+1} | X_n = s_n\} = \frac{\mathcal{P}\{X_{n+1} = s_{n+1}, X_n = s_n\}}{\mathcal{P}\{X_n = s_n\}} \\
= & \frac{\mathcal{P}(V_{n+1} = v_{n+1})\mathcal{P}(X_n = s_n)}{\mathcal{P}(X_n = s_n)} = \mathcal{P}(V_{n+1} = v_{n+1})
\end{aligned}$$

where the second equality follows from the independence of V_{n+1} and X_n. Thus

$$\mathcal{P}\{X_{n+1} = s_{n+1} | X_0 = s_0, \ldots, X_{n-1} = s_{n-1}, X_n = s_n\} = \mathcal{P}\{X_{n+1} = s_{n+1} | X_n = s_n\}.$$

and then a random walk is a homogeneous Markov chain with one-step transition probability

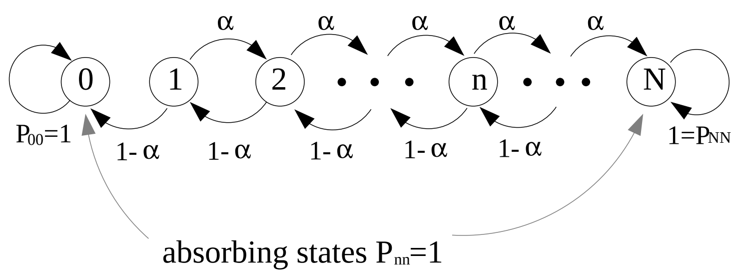

Figure 2.3: The state-transition diagram of the gambler's model.

matrix

$$P = \begin{bmatrix} \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \\ \cdots & 0 & \alpha & 0 & 0 & \cdots \\ \cdots & 1-\alpha & 0 & \alpha & 0 & \cdots \\ \cdots & 0 & 1-\alpha & 0 & \alpha & \cdots \\ \cdots & 0 & 0 & 1-\alpha & 0 & \cdots \\ \cdots & \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}$$

which is a doubly infinite matrix. □

Example 2.1.7 [A gambler's model] *Let V_n be the gain of a gambler at the nth paly, $n \geq 1$. He wins one unit, i.e. $V_n = 1$, with probability p and loses one unit, i.e. $V_n = -1$, with probability $1 - p$. Assuming that successive plays of the game are statistically independent. The gambler will stop playing the game if either his fortune reaches a preset number N or he has no money left. To model his game, we let X_n be the fortune of the gambler after the nth paly. We also assume that the initial fortune X_0 of the gambler is a fixed number between 0 and N. It can be shown that X_n's form a Markov chain with state space $S = \{0, 1, \ldots, N\}$ and state-transition diagram as shown in Figure 2.3.* □

2.2 Fundamental Properties

The aim of this seciton is to develop more delicate properties of a discrete-time Markov chain $\{X_n, n = 0, 1, \ldots\}$ with state space S. The following theorem is a stronger version of Definition 2.1.1.

Theorem 2.2.1 *For any non-negative r.v. $Y \in \mathcal{F}(X_{n+1})$ or any real-valued r.v. $Y \in \mathcal{F}(X_{n+1})$ with finite expectation, we have*

$$\mathcal{E}(Y|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) = \mathcal{E}(Y|X_n = s_n),$$

for all $n \geq 0$ and all $s_0, s_1, \ldots, s_n$ in S, whenever $\mathcal{P}\{X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n\} > 0$.

Proof. We first consider $Y = 1_{(X_{n+1}=s_{n+1})}$, the indicator function of the event $(X_{n+1} = s_{n+1})$. Since $\mathcal{P}(X_0 = s_0, \ldots, X_n = s_n) > 0$, we have

$$\begin{aligned}
\mathcal{E}(1_{(X_{n+1}=s_{n+1})}|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) &= \mathcal{E}_{(X_0=s_0,X_1=s_1,\ldots,X_n=s_n)}(1_{(X_{n+1}=s_{n+1})}) \\
&= \mathcal{P}_{(X_0=s_0,X_1=s_1,\ldots,X_n=s_n)}(X_{n+1} = s_{n+1}) \\
&= \mathcal{P}_{(X_n=s_n)}(X_{n+1} = s_{n+1}) \ \text{ by Definition 2.1.1} \\
&= \mathcal{E}_{(X_n=s_n)}(1_{(X_{n+1}=s_{n+1})}) \\
&= \mathcal{E}(1_{(X_{n+1}=s_{n+1})}|X_n = s_n).
\end{aligned}$$

This concludes that the theorem is true for indicator functions $Y = 1_{(X_{n+1}=s_{n+1})}$. We next consider a non-negative $Y \in \mathcal{F}(X_{n+1})$. By Theorem 1.5.5, we have

$$Y = \sum_{s_{n+1}} y_{s_{n+1}} 1_{(X_{n+1}=s_{n+1})}, \tag{2.1}$$

where $y_{s_{n+1}} \geq 0$ for all s_{n+1}. By the series convergence property of expectation in Theorem 1.4.8 and (2.1), we have

$$\begin{aligned}
\mathcal{E}(Y|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) &= \sum_{s_{n+1}} y_{s_{n+1}} \mathcal{E}(1_{(X_{n+1}=s_{n+1})}|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) \\
&= \sum_{s_{n+1}} y_{s_{n+1}} \mathcal{E}(1_{(X_{n+1}=s_{n+1})}|X_n = s_n) \\
&= \mathcal{E}(Y|X_n = s_n).
\end{aligned}$$

Finally, we consider $Y \in \mathcal{F}(X_{n+1})$ with finite expectation. Then, $Y^+, Y^- \in \mathcal{F}(X_{n+1})$ (as an exercise) with finite expectation and by the linearity of expectation in Theorem 1.4.11,

$$\begin{aligned}
& \mathcal{E}(Y|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) \\
= \ & \mathcal{E}(Y^+|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) - \mathcal{E}(Y^-|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) \\
= \ & \mathcal{E}(Y^+|X_n = s_n) - \mathcal{E}(Y^-|X_n = s_n) \\
= \ & \mathcal{E}(Y|X_n = s_n).
\end{aligned}$$

This completes the proof. □

The above theorem can be described in a compact form. At first, we need some definitions of σ-algebras. The σ-algebra generated by the union $\cup_{i=m}^{n} \mathcal{F}(X_i)$ of $\mathcal{F}(X_i)$'s is called the σ-algebra generated by r.v.'s $X_m, X_{m+1}, \ldots, X_n$, denoted as $\mathcal{F}(X_m, X_{m+1}, \ldots, X_n)$ or $\mathcal{F}_{[m,n]}$ for brevity. It can be seen that $\mathcal{F}_{[m,n]}$ is the σ-algebra generated by the countable measurable partition $\{(X_m = s_m, \ldots, X_n = s_n)\}_{s_m,\ldots,s_n \in S}$ of Ω. Now the conditional expectation $\mathcal{E}(Y|\mathcal{F}_{[0,n]})$ of Y relative to the σ-algebra $\mathcal{F}_{[0,n]}$ is, as defined in 1.5.3,

$$\mathcal{E}(Y|\mathcal{F}_{[0,n]}) = \sum_{s_0,s_1,\ldots,s_n \in S} \mathcal{E}(Y|X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) 1_{(X_0=s_0,X_1=s_1,\ldots,X_n=s_n)}$$

and the conditional expectation $\mathcal{E}(Y|\mathcal{F}_{\{n\}})$ of Y relative to the σ-algebra $\mathcal{F}_{\{n\}}$ is

$$\mathcal{E}(Y|\mathcal{F}_{\{n\}}) = \sum_{s_n \in S} \mathcal{E}(Y|X_n = s_n) 1_{(X_n = s_n)}.$$

By Theorem 2.2.1, we have

Corollary 2.2.2 *If $\{X_n, n = 0, 1, \ldots\}$ is a Markov chain with state space S, then*

$$\mathcal{E}(Y|\mathcal{F}_{[0,n]}) = \mathcal{E}(Y|\mathcal{F}_{\{n\}}) \quad \text{w.p.1, i.e.,}$$

$$\mathcal{E}(Y|X_0, X_1, \ldots, X_n) = \mathcal{E}(Y|X_n) \quad \text{w.p.1,}$$

for any non-negative r.v. $Y \in \mathcal{F}(X_{n+1})$ or any real-valued r.v. $Y \in \mathcal{F}(X_{n+1})$ with finite expectation.

We now need the concept of a monotone class.

Definition 2.2.3 *A collection $\mathcal{C}$ of subsets of Ω is called a monotone class in Ω if*

1. $A_n \in \mathcal{C}$ *and* $A_n \subseteq A_{n+1}$, $n = 1, 2, \ldots \Rightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{C}$;
2. $A_n \in \mathcal{C}$ *and* $A_n \supseteq A_{n+1}$, $n = 1, 2, \ldots \Rightarrow \cap_{n=1}^{\infty} A_n \in \mathcal{C}$. □

It is clear that any one of the above two properties, together with the complement property: $A \in \mathcal{C} \Rightarrow A^c \in \mathcal{C}$, implies the other property. The following theorem is useful [1].

Theorem 2.2.4 *[Monotone class theorem] If a monotone class $\mathcal{C}$ in Ω contains an algebra $\mathcal{F}_0$ in Ω, then $\mathcal{C}$ contains the σ-algebra $\mathcal{F}$ generated by the algebra $\mathcal{F}_0$.*

Let $\mathcal{F}_{[n,\infty)}$ be the σ-algebra generated by r.v.'s $X_i, i = n, n+1, \ldots$. The following theorem is an extension of Corollary 2.2.2.

Theorem 2.2.5 *If $\{X_n, n = 0, 1, \ldots\}$ is a Markov chain with state space S, then*

$$\mathcal{E}(Y|X_0, \ldots, X_{n-1}, X_n) = \mathcal{E}(Y|X_n) \quad \text{w.p.1} \tag{2.2}$$

for any non-negative r.v. $Y \in \mathcal{F}_{[n,\infty)}$ or any real-valued r.v. $Y \in \mathcal{F}_{[n,\infty)}$ with finite expectation.

Proof. We shall use induction to show that (2.2) is true for any non-negative $Y \in \mathcal{F}_{[n,n+k]}$ or any $Y \in \mathcal{F}_{[n,n+k]}$ with finite expectation for all $k = 0, 1, \ldots$. For $k = 0$, we have Y belongs to $\mathcal{F}(X_n)$ and then to $\mathcal{F}(X_0, X_1, \ldots, X_n)$. By Theorem 1.5.5, we have

[1]For a proof, please see K. L. Chung, *A Course in Probability Theory,* 2nd edn. New York: Academic Press, 1974, pp. 16–18.

$\mathcal{E}(Y|X_0,\ldots,X_{n-1},X_n) = Y = \mathcal{E}(Y|X_n)$ w.p.1. Now suppose that it is true up to some k. We first consider an r.v. $Y = Y_1 \cdot Y_2$ with $Y_1, Y_2 \geq 0$ and $Y_1 \in \mathcal{F}_{[n,n+k]}$, $Y_2 \in \mathcal{F}_{\{n+k+1\}}$. Then

$$\begin{aligned}
&\mathcal{E}(Y|\mathcal{F}_{[0,n]}) = \mathcal{E}\{\mathcal{E}(Y|\mathcal{F}_{[0,n+k]})|\mathcal{F}_{[0,n]}\} \text{ by Theorem 1.5.13} \\
= \ &\mathcal{E}\{Y_1\mathcal{E}(Y_2|\mathcal{F}_{[0,n+k]})|\mathcal{F}_{[0,n]}\} \text{ since } Y_1 \in \mathcal{F}_{[0,n+k]} \text{ and by Theorem 1.5.11} \\
= \ &\mathcal{E}\{Y_1\mathcal{E}(Y_2|\mathcal{F}_{\{n+k\}})|\mathcal{F}_{[0,n]}\} \text{ by Corollary 2.2.2} \\
= \ &\mathcal{E}\{Y_1\mathcal{E}(Y_2|\mathcal{F}_{\{n+k\}})|\mathcal{F}_{\{n\}}\} \text{ since } Y_1 \cdot \mathcal{E}(Y_2|\mathcal{F}_{\{n+k\}}) \in \mathcal{F}_{[n,n+k]} \text{ and by the induction step} \\
= \ &\mathcal{E}\{Y_1\mathcal{E}(Y_2|\mathcal{F}_{[0,n+k]})|\mathcal{F}_{\{n\}}\} \text{ by Corollary 2.2.2} \\
= \ &\mathcal{E}\{\mathcal{E}(Y_1Y_2|\mathcal{F}_{[0,n+k]})|\mathcal{F}_{\{n\}}\} \text{ since } Y_1 \in \mathcal{F}_{[0,n+k]} \text{ and by Theorem 1.5.11} \\
= \ &\mathcal{E}(Y_1Y_2|\mathcal{F}_{\{n\}}) \text{ by Theorem 1.5.13} \\
= \ &\mathcal{E}(Y|\mathcal{F}_{\{n\}}).
\end{aligned}$$

We next consider a general non-negative $Y \in \mathcal{F}_{[n,n+k+1]}$. By Theorem 1.5.5, we have

$$\begin{aligned}
Y &= \sum_{s_n,\ldots,s_{n+k},s_{n+k+1}} y_{s_n,\ldots,s_{n+k},s_{n+k+1}} 1_{(X_n=s_n,\ldots,X_{n+k}=s_{n+k},X_{n+k+1}=s_{n+k+1})} \\
&= \sum_{s_n,\ldots,s_{n+k},s_{n+k+1}} y_{s_n,\ldots,s_{n+k},s_{n+k+1}} 1_{(X_n=s_n,\ldots,X_{n+k}=s_{n+k})} \cdot 1_{(X_{n+k+1}=s_{n+k+1})}
\end{aligned}$$

with $y_{s_n,\ldots,s_{n+k},s_{n+k+1}} \geq 0$. Thus

$$\begin{aligned}
&\mathcal{E}(Y|\mathcal{F}_{[0,n]}) \\
= \ &\sum_{s_n,\ldots,s_{n+k},s_{n+k+1}} y_{s_n,\ldots,s_{n+k},s_{n+k+1}} \mathcal{E}(1_{(X_n=s_n,\ldots,X_{n+k}=s_{n+k})} \cdot 1_{(X_{n+k+1}=s_{n+k+1})}|\mathcal{F}_{[0,n]}) \\
&\quad \text{by the series convergence property of conditional expectation in Theorem 1.5.8} \\
= \ &\sum_{s_n,\ldots,s_{n+k},s_{n+k+1}} y_{s_n,\ldots,s_{n+k},s_{n+k+1}} \mathcal{E}(1_{(X_n=s_n,\ldots,X_{n+k}=s_{n+k})} \cdot 1_{(X_{n+k+1}=s_{n+k+1})}|\mathcal{F}_{\{n\}}) \text{ as proved in above} \\
= \ &\mathcal{E}(Y|\mathcal{F}_{\{n\}}) \text{ w.p.1.}
\end{aligned}$$

Finally, the proof for a r.v. $Y \in \mathcal{F}_{[n,n+k+1]}$ with finite expectation can be done as that in Theorem 2.2.1. This completes the induction process. Now, let $\mathcal{C}$ be the collection of all events Λ in $\mathcal{F}$ such that

$$\mathcal{E}(1_\Lambda|\mathcal{F}_{[0,n]}) = \mathcal{E}(1_\Lambda|\mathcal{F}_{\{n\}}) \text{ w.p.1.}$$

Since 1_Λ belongs to $\mathcal{F}_{[n,n+k]}$ for any Λ in $\mathcal{F}_{[n,n+k]}$, we have $\cup_{k=0}^{\infty}\mathcal{F}_{[n,n+k]} \subseteq \mathcal{C}$. Furthermore, we have

1. $\Lambda \in \mathcal{C} \Rightarrow \Lambda^c \in \mathcal{C}$, since

   $$\mathcal{E}(1_{\Lambda^c}|\mathcal{F}_{[0,n]}) = \mathcal{E}(1_\Omega|\mathcal{F}_{[0,n]}) - \mathcal{E}(1_\Lambda|\mathcal{F}_{[0,n]}) = \mathcal{E}(1_\Omega|\mathcal{F}_{\{n\}}) - \mathcal{E}(1_\Lambda|\mathcal{F}_{\{n\}}) = \mathcal{E}(1_{\Lambda^c}|\mathcal{F}_{\{n\}}) \text{ w.p.1;}$$

2. $\Lambda_n \in \mathcal{C}$ and $\Lambda_n \subseteq \Lambda_{n+1}$, $n = 1, 2, \ldots \Rightarrow \cup_{n=1}^{\infty}\Lambda_n \in \mathcal{C}$, since $1_{\Lambda_n} \uparrow 1_{\cup_{n=1}^{\infty}\Lambda_n}$ and

   $$\mathcal{E}(1_{\cup_{i=1}^{\infty}\Lambda_i}|\mathcal{F}_{[0,n]}) = \lim_{n\to\infty} \mathcal{E}(1_{\Lambda_n}|\mathcal{F}_{[0,n]}) = \lim_{n\to\infty} \mathcal{E}(1_{\Lambda_n}|\mathcal{F}_{\{n\}}) = \mathcal{E}(1_{\cup_{i=1}^{\infty}\Lambda_i}|\mathcal{F}_{\{n\}}) \text{ w.p.1}$$

   by the monotone convergence property of conditional expectation in Theorem 1.5.8.

Thus $\mathcal{C}$ is a monotone class containing $\cup_{k=0}^{\infty}\mathcal{F}_{[n,n+k]}$ which is an algebra and by the monotone class theorem, $\mathcal{C}$ contains the σ-algebra $\mathcal{F}_{[n,\infty)}$ generated by the algebra $\cup_{k=0}^{\infty}\mathcal{F}_{[n,n+k]}$, i.e.,

$$\mathcal{E}(1_\Lambda|\mathcal{F}_{[0,n]}) = \mathcal{E}(1_\Lambda|\mathcal{F}_{\{n\}}) \text{ w.p.1}, \quad \forall\, \Lambda \in \mathcal{F}_{[n,\infty)}.$$

By linearity, (2.2) holds for any non-negative simple r.v. Y belonging to $\mathcal{F}_{[n,\infty)}$. Since any non-negative r.v. $Y \in \mathcal{F}_{[n,\infty)}$ is the limit of a monotonely increasing sequence of non-negative simple r.v.'s belonging to $\mathcal{F}_{[n,\infty)}$, (2.2) holds for any non-negative $Y \in \mathcal{F}_{[n,\infty)}$, by the monotone convergence property of conditional expectation. The proof for an r.v. $Y \in \mathcal{F}_{[n,\infty)}$ with finite expectation is similar to that in Theorem 2.2.1. This completes the proof. □

Theorem 2.2.6 *If $\{X_n, n = 0, 1, \ldots\}$ is a Markov chain with state space S, then*

$$\mathcal{E}(Y_1Y_2|X_n) = \mathcal{E}(Y_1|X_n)\cdot\mathcal{E}(Y_2|X_n) \quad \textit{w.p.1} \tag{2.3}$$

for any non-negative r.v.'s $Y_1 \in \mathcal{F}_{[0,n]}$, $Y_2 \in \mathcal{F}_{[n,\infty)}$ or any real-valued r.v.'s $Y_1 \in \mathcal{F}_{[0,n]}$, $Y_2 \in \mathcal{F}_{[n,\infty)}$ such that Y_1, Y_2 and Y_1Y_2 all have finite expectations.

Remark. If Y_1Y_2 has finite expectation, then we have

$$\begin{aligned}
\mathcal{E}(|Y_1\mathcal{E}(Y_2|X_n)|) &\le \mathcal{E}(|Y_1|\mathcal{E}(|Y_2||X_n)) \text{ since } |\mathcal{E}(Y_2|X_n)| \le \mathcal{E}(|Y_2||X_n)\\
&= \mathcal{E}(|Y_1|\mathcal{E}(|Y_2||\mathcal{F}_n)) \text{ by Theorem 2.2.5}\\
&= \mathcal{E}(\mathcal{E}(|Y_1Y_2||\mathcal{F}_n)) = \mathcal{E}(|Y_1Y_2|) < \infty.
\end{aligned}$$

Proof. Since

$$\begin{aligned}
&\mathcal{E}(Y_1|X_n)\cdot\mathcal{E}(Y_2|X_n)\\
=\;& \mathcal{E}\left\{Y_1\cdot\mathcal{E}(Y_2|X_n)|X_n\right\} \text{ since } \mathcal{E}(Y_2|X_n) \in \mathcal{F}_{\{n\}} \text{ and by Theorem 1.5.11}\\
=\;& \mathcal{E}\left\{Y_1\cdot\mathcal{E}(Y_2|\mathcal{F}_{[0,n]})|X_n\right\} \text{ by Theorem 2.2.5}\\
=\;& \mathcal{E}\left\{\mathcal{E}(Y_1Y_2|\mathcal{F}_{[0,n]})|X_n\right\} \text{ since } Y_1 \in \mathcal{F}_{[0,n]} \text{ and by Theorem 1.5.11}\\
=\;& \mathcal{E}(Y_1Y_2|X_n) \text{ w.p.1 by Theorem 1.5.13,}
\end{aligned}$$

the proof is completed. □

A particular application of the above theorem is that for any event $\Lambda_1 \in \mathcal{F}_{[0,n]}$ and any event $\Lambda_2 \in \mathcal{F}_{[n,\infty)}$, we have

$$\mathcal{P}(\Lambda_1\cap\Lambda_2|X_n) = \mathcal{E}(1_{\Lambda_1}\cdot 1_{\Lambda_2}|X_n) = \mathcal{E}(1_{\Lambda_1}|X_n)\cdot\mathcal{E}(1_{\Lambda_2}|X_n) = \mathcal{P}(\Lambda_1|X_n)\cdot\mathcal{P}(\Lambda_2|X_n) \text{ w.p.1.} \tag{2.4}$$

The above equation can be informally interpreted as follows: "The past and the future of the Markov chain are conditionally independent given the present". From (2.4), we have

$$\mathcal{P}(\Lambda_1\cap\Lambda_2|X_n = s) = \mathcal{P}(\Lambda_1|X_n = s)\cdot\mathcal{P}(\Lambda_2|X_n = s), \tag{2.5}$$

for all $s \in S$ whenever $\mathcal{P}(X_n = s) > 0$. Since

$$\mathcal{P}(\Lambda_1 \cap \Lambda_2 | X_n = s) = \mathcal{P}(\Lambda_1 | X_n = s) \cdot \mathcal{P}(\Lambda_2 | \Lambda_1, X_n = s) = \mathcal{P}(\Lambda_2 | X_n = s) \cdot \mathcal{P}(\Lambda_1 | \Lambda_2, X_n = s),$$

we have

$$\mathcal{P}(\Lambda_2 | \Lambda_1, X_n = s) = \mathcal{P}(\Lambda_2 | X_n = s) \quad \text{and} \quad \mathcal{P}(\Lambda_1 | \Lambda_2, X_n = s) = \mathcal{P}(\Lambda_1 | X_n = s) \tag{2.6}$$

for any $\Lambda_1 \in \mathcal{F}_{[0,n]}$, $\Lambda_2 \in \mathcal{F}_{[n,\infty)}$ and $s \in S$ whenever $\mathcal{P}(\Lambda_1, X_n = s) > 0$ and $\mathcal{P}(\Lambda_2, X_n = s)$ respectively.

2.3 Chapman-Kolmogorov Equations

Consider a discrete-time Markov chain $\{X_n, n = 0, 1, \ldots\}$ with state space S. The joint distribution $\mathcal{P}(X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n)$ of r.v.'s X_0, X_1, $\ldots$, X_n on the state space S can be calculated as

$$\begin{aligned}
& \mathcal{P}(X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) \\
= \; & \mathcal{P}(X_0 = s_0)\mathcal{P}(X_1 = s_1 | X_0 = s_0)\mathcal{P}(X_2 = s_2 | X_0 = s_0, X_1 = s_1) \cdots \\
& \mathcal{P}(X_n = s_n | X_0 = s_0, X_1 = s_1, \ldots, X_{n-1} = s_{n-1}) \\
= \; & \mathcal{P}(X_0 = s_0)\mathcal{P}(X_1 = s_1 | X_0 = s_0)\mathcal{P}(X_2 = s_2 | X_1 = s_1) \cdots \mathcal{P}(X_n = s_n | X_{n-1} = s_{n-1}) \\
= \; & \pi_0(s_0) P_0(s_0, s_1) P_1(s_1, s_2) \cdots P_{n-1}(s_{n-1}, s_n),
\end{aligned} \tag{2.7}$$

where $\pi_0(s_0) \equiv \mathcal{P}(X_0 = s_0)$. Thus the finite-dimensional joint distributions of the Markov chain $\{X_n, n = 0, 1, \ldots\}$ is completely specified by the initial distribution $\pi_0 = (\pi_0(s)), s \in S$, (as a row vector) and the one-step transition probability matrices $P_m = [P_m(s, s')], s, s' \in S$, at time m, $m = 0, 1, \ldots$. In particular, a homogeneous Markov chain is completely specified by the initial distribution π_0 and the one-step transition probability matrix P.

We next investigate the n-step transition probability from state s at time m to state s' at time $m + n$, defined as

$$P_m^{(n)}(s, s') \equiv \mathcal{P}\{X_{m+n} = s' \mid X_m = s\}.$$

Note that $P_m^{(1)}(s, s') = P_m(s, s')$. Then for any l, $1 \le l \le n - 1$, we have

$$\begin{aligned}
& P_m^{(n)}(s, s') \\
= \; & \sum_{\tilde{s} \in S} \mathcal{P}\{X_{m+l} = \tilde{s}, X_{m+n} = s' \mid X_m = s\} \quad \text{by the countable additivity of conditional p.m.} \\
= \; & \sum_{\tilde{s} \in S} \mathcal{P}\{X_{m+l} = \tilde{s} \mid X_m = s\}\mathcal{P}\{X_{m+n} = s' \mid X_m = s, X_{m+l} = \tilde{s}\} \\
= \; & \sum_{\tilde{s} \in S} \mathcal{P}\{X_{m+l} = \tilde{s} \mid X_m = s\}\mathcal{P}\{X_{m+n} = s' \mid X_{m+l} = \tilde{s}\} \quad \text{by (2.6)} \\
= \; & \sum_{\tilde{s} \in S} P_m^{(l)}(s, \tilde{s}) P_{m+l}^{(n-l)}(\tilde{s}, s')
\end{aligned} \tag{2.8}$$

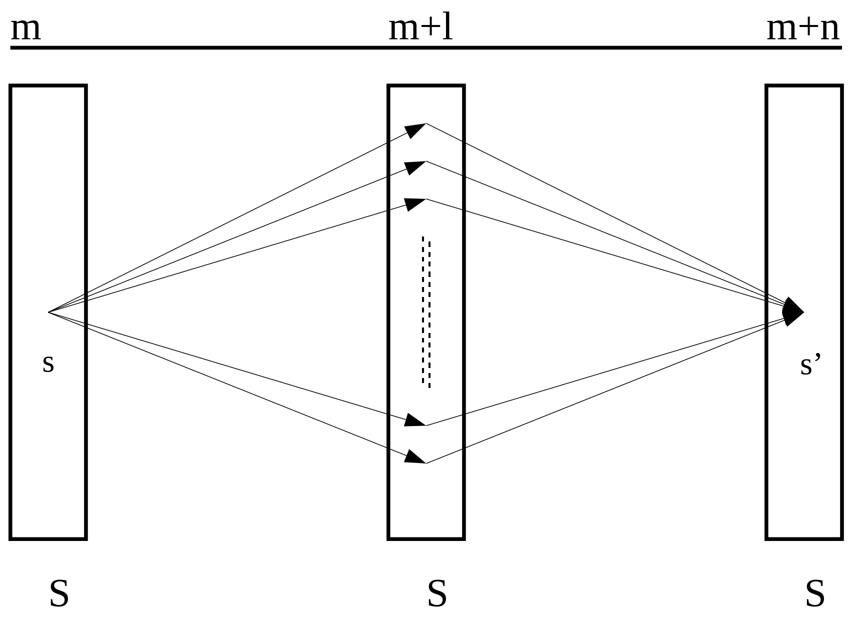

Figure 2.4: The Chapman-Kolmogorov equation.

The above equation is called the Chapman-Kolmogorov equation and illustrated in Figure 2.4. In particular, there are two important cases:

1. The backward equation $P_m^{(n+1)}(s,s') = \sum_{\tilde{s}\in S} P_m(s,\tilde{s})P_{m+1}^{(n)}(\tilde{s},s')$;
2. The forward equation $P_m^{(n+1)}(s,s') = \sum_{\tilde{s}\in S} P_m^{(n)}(s,\tilde{s})P_{m+n}(\tilde{s},s')$.

To rewrite the Chapman-Kolmogorov equations in a compact form, we define the n-step transition probability matrix at time m as

$$P_m^{(n)} = [P_m^{(n)}(s,s')], s, s' \in S.$$

Note that $P_m^{(1)} = P_m$. Then we have

$$P_m^{(n)} = P_m^{(l)} P_{m+l}^{(n-l)}, \ \forall\ 0 \le l \le n, \tag{2.9}$$

where $P_m^{(0)}$ is defined to be the identity matrix I for any m. And the backward and forward equations can be rewritten in the following matrix forms

$$P_m^{(n+1)} = P_m P_{m+1}^{(n)} \quad \text{and} \quad P_m^{(n+1)} = P_m^{(n)} P_{m+n}.$$

By iteratively using either the backward equation or the forward equation, we have the following theorem.

Theorem 2.3.1 *The n-step transition probability matrix $P_m^{(n)}$ at time m can be calculated as*

$$P_m^{(n)} = P_m P_{m+1} \dots P_{m+n-1}.$$

□

In particular, for a homogeneous Markov chain with $P = P_m$ for all m, we have

$$P_m^{(n)} = P^n$$

and the n-step transition probabilities are independent of the initial time m. Finally, let $\pi_m = (\pi_m(s)), s \in S$ be the distribution of the Markov chain at time m where $\pi_m(s) = \mathcal{P}(X_m = s)$, we have

Corollary 2.3.2

$$\pi_{m+n} = \pi_m P_m^{(n)}$$

for all $m, n \geq 0$. □

Example 2.3.3 *Consider a homogeneous Markov chain with state space* $S = \{1, 2, 3\}$ *and one-step transition probability matrix*

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

Since P *has eigenvectors*

$$[1/3, 1/3, 1/3], \quad [-1/2, (1+i\sqrt{3})/4, (1-i\sqrt{3})/4], \quad [-1/2, (1-i\sqrt{3})/4, (1+i\sqrt{3})/4]$$

corresponding to eigenvalues $1, (1+i\sqrt{3})/4, (1-i\sqrt{3})/4$*, it can be diagonalized as*

$$P = M^{-1}\Lambda M$$

where

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1+i\sqrt{3}}{4} & 0 \\ 0 & 0 & \frac{1-i\sqrt{3}}{4} \end{bmatrix}, \quad M = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ -\frac{1}{2} & \frac{1+i\sqrt{3}}{4} & \frac{1-i\sqrt{3}}{4} \\ -\frac{1}{2} & \frac{1-i\sqrt{3}}{4} & \frac{1+i\sqrt{3}}{4} \end{bmatrix} \text{ and } M^{-1} = \begin{bmatrix} 1 & -\frac{2}{3} & -\frac{2}{3} \\ 1 & \frac{1-i\sqrt{3}}{3} & \frac{1+i\sqrt{3}}{3} \\ 1 & \frac{1+i\sqrt{3}}{3} & \frac{1-i\sqrt{3}}{3} \end{bmatrix}.$$

Thus, we have the n*-step transition probability matrix* $P^{(n)}$ *as*

$$P^{(n)} = P^n = M^{-1}\Lambda^n M = M^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1+i\sqrt{3}}{4}\right)^n & 0 \\ 0 & 0 & \left(\frac{1-i\sqrt{3}}{4}\right)^n \end{bmatrix} M.$$

Suppose that the initial distribution is

$$\pi_0 = [p_1 \; p_2 \; p_3].$$

Then the distribution π_n of the Markov chain at time n is

$$\pi_n = \pi_0 P^{(n)} = [p_1\ p_2\ p_3] M^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \left(\frac{1+i\sqrt{3}}{4}\right)^n & 0 \\ 0 & 0 & \left(\frac{1-i\sqrt{3}}{4}\right)^n \end{bmatrix} M$$

$$= [p_1\ p_2\ p_3] \begin{bmatrix} \frac{1}{3} + \frac{\cos n\theta}{3\cdot 2^{n-1}} & \frac{1}{3} - \frac{\cos(n+1)\theta}{3\cdot 2^{n-1}} & \frac{1}{3} - \frac{\cos(n-1)\theta}{3\cdot 2^{n-1}} \\ \frac{1}{3} - \frac{\cos(n-1)\theta}{3\cdot 2^{n-1}} & \frac{1}{3} + \frac{\cos n\theta}{3\cdot 2^{n-1}} & \frac{1}{3} - \frac{\cos(n-2)\theta}{3\cdot 2^{n-1}} \\ \frac{1}{3} - \frac{\cos(n+1)\theta}{3\cdot 2^{n-1}} & \frac{1}{3} - \frac{\cos(n+2)\theta}{3\cdot 2^{n-1}} & \frac{1}{3} + \frac{\cos n\theta}{3\cdot 2^{n-1}} \end{bmatrix}$$

$$= \left[\frac{1}{3} + \frac{p_1 \cos n\theta - p_2 \cos(n-1)\theta - p_3 \cos(n+1)\theta}{3\cdot 2^{n-1}}, \frac{1}{3} + \frac{p_1 \cos(n+1)\theta - p_2 \cos n\theta - p_3 \cos(n+2)\theta}{3\cdot 2^{n-1}}, \right.$$
$$\left. \frac{1}{3} + \frac{p_1 \cos(n-1)\theta - p_2 \cos(n-2)\theta - p_3 \cos n\theta}{3\cdot 2^{n-1}} \right],$$

where $e^{i\theta} = (1 + i\sqrt{3})/2$. It can be seen that

$$\lim_{n\to\infty} \pi_n = [1/3\ 1/3\ 1/3]$$

no matter what the initial distribution π_0 is. □

2.4 Classification of States

In this section, we consider a homogeneous Markov chain $\{X_n, n = 0, 1, \ldots\}$ with state space S and one-step transition probability matrix P.

Definition 2.4.1 *A state s' is called accessible (reachable) from state s, denoted as $s \to s'$, if $P^{(n)}(s, s') > 0$ for some $n \geq 0$.* □

Definition 2.4.2 *States s and s' are said to communicate with each other, denoted as $s \leftrightarrow s'$, if $s \to s'$ and $s' \to s$.* □

Theorem 2.4.3 *Communication is an equivalence relation among states in S.*

Proof. We need to show that

1. reflectivity: for any $s \in S$, $s \leftrightarrow s$;

2. symmetry: if $s \leftrightarrow s'$, then $s' \leftrightarrow s$;

3. transitivity: if $s \leftrightarrow s'$ and $s' \leftrightarrow s''$, then $s \leftrightarrow s''$.

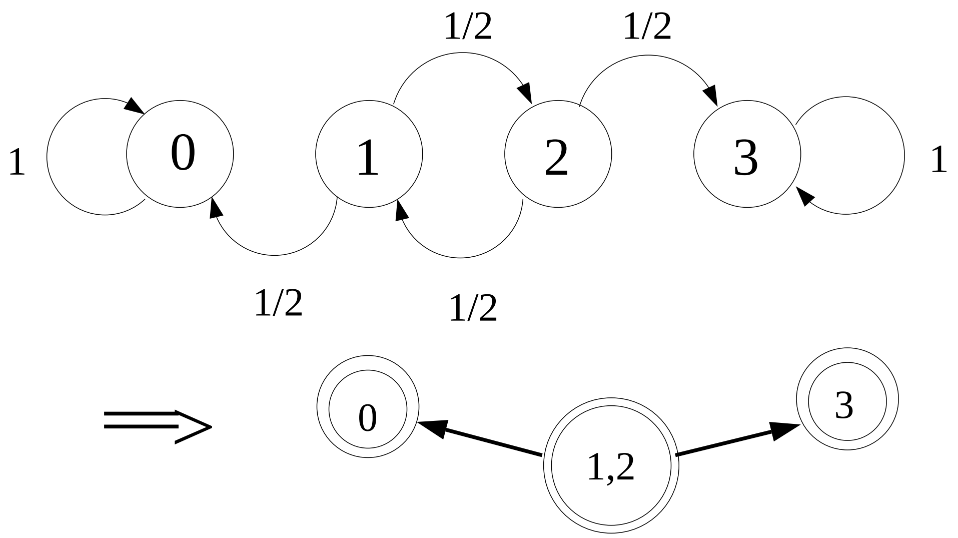

Figure 2.5: State-transition diagram of the Markov chain in Example 2.4.4.

.

The reflectivity and symmetry property of the communication relation is obvious. To show the transitivity property, we note that if $s \to s'$ and $s' \to s''$, then there exist $m, n \geq 0$ such that $P^{(m)}(s, s') > 0$ and $P^{(n)}(s', s'') > 0$. Thus by Chapman-Kolmogorov equation, we have

$$P^{(n+m)}(s, s'') = \sum_{t \in S} P^{(m)}(s, t) P^{(n)}(t, s'') \geq P^{(m)}(s, s') P^{(n)}(s', s'') > 0$$

which, by definition, means that $s \to s''$. Similarly $s'' \to s'$ and $s' \to s$ imply $s'' \to s$. This completes the proof. □

Now we can partition the state space S into equivalence classes under the communication relation. Each of such classes is called a communication class.

Example 2.4.4 . *Consider a Markov chain with state transition diagram as shown in Figure 2.5. Then the state space $S = \{0, 1, 2, 3\}$ can be partitioned into three communication classes $\{0\}$, $\{1, 2\}$ and $\{3\}$. Note that the arrows in the class transition diagram is always one-directional.* □

Definition 2.4.5 *A Markov chain is called irreducible if the whole state space is a communication class.* □

We shall investigate class properties of a Markov chain. A property of states in the state space S is called a class property if a state s satisfies this property, then any other state in the same class as s also satisfies this property. Before this, we need to consider the so called first passage probabilities.

2.4.1 First passage probabilities

Let s be a given state in the state space S. Let $T_m^{(s)}$, $m = 1, 2, \ldots$, be the successive times at which the Markov chain visits state s after the initial time $n = 0$. For a sample $\omega \in \Omega$, if there do not exist such times after the k-th visit, i.e. only $T_1^{(s)}(\omega), T_2^{(s)}(\omega), \ldots, T_k^{(s)}(\omega)$ are finite, we shall let $T_{m+1}^{(s)}(\omega) - T_m^{(s)}(\omega) = \infty$ for all $m \geq k$. This implies that $T_m^{(s)}(\omega) = \infty$ for all $m \geq k+1$.

The conditional probability $\mathcal{P}\left(T_1^{(s)} = k | X_0 = s'\right)$ is the conditional probability that the Markov chain will visit the state s firstly at the time $n = k$, $k \geq 1$, relative to the event that the initial state X_0 is s'. These first passage probabilities in k steps will be denoted as

$$f_k(s', s) \equiv \mathcal{P}\left(T_1^{(s)} = k | X_0 = s'\right) = \mathcal{P}\left(X_1 \neq s, \ldots, X_{k-1} \neq s, X_k = s | X_0 = s'\right)$$

for $k \geq 1$ and $s, s' \in S$. To compute $f_k(s', s)$'s, we first note that for $k = 1$,

$$f_1(s', s) = \mathcal{P}\left(X_1 = s | X_0 = s'\right) = P(s', s).$$

And for $k \geq 2$, we have

$$\begin{aligned} f_k(s', s) &= \mathcal{P}\left(X_1 \neq s, \ldots, X_{k-1} \neq s, X_k = s | X_0 = s'\right) \\ &= \sum_{s'' \in S - \{s\}} \mathcal{P}\left(X_1 = s'' | X_0 = s'\right) \mathcal{P}\left(X_2 \neq s, \ldots, X_{k-1} \neq s, X_k = s | X_1 = s''\right) \\ &= \sum_{s'' \in S - \{s\}} P(s', s'') \mathcal{P}\left(X_1 \neq s, \ldots, X_{k-2} \neq s, X_{k-1} = s | X_0 = s''\right) \\ &\qquad \text{by the homogeneity of the Markov chain} \\ &= \sum_{s'' \in S - \{s\}} P(s', s'') f_{k-1}(s'', s). \end{aligned}$$

We summarize the above discussions in the following theorem.

Theorem 2.4.6 *The first passage probabilities $f_k(s, s')$ in k steps satisfy the recursive formula*

$$f_k(s', s) = \sum_{s'' \in S - \{s\}} P(s', s'') f_{k-1}(s'', s)$$

for all $k \geq 2$ and $s, s' \in S$ with initial conditions

$$f_1(s', s) = P(s', s)$$

for all $s, s' \in S$. □

Example 2.4.7 *Consider a Markov chain with state space $S = \{1, 2, 3\}$ and one-step transition probability matrix*

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let $v_k = [f_k(1,1), f_k(2,1), f_k(3,1)]^t$ be a column vector for each $k = 1, 2, \ldots$. By Theorem 2.4.6, v_1 is the first column of P, i.e. $v_1 = [1/2, 1/4, 0]^t$ and

$$v_k = Qv_{k-1}$$

for all $k \geq 2$ with the matrix Q obtained from P by replacing its first column by zeros as follows:

$$Q = \begin{bmatrix} 0 & 1/4 & 1/4 \\ 0 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{bmatrix}.$$

By diagonalizing Q, we have

$$Q = M \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix} M^{-1}$$

with

$$M = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 8 \end{bmatrix} \quad \text{and} \quad M^{-1} = \begin{bmatrix} 1 & -1/2 & -1/8 \\ 0 & 1/2 & -1/4 \\ 0 & 0 & 1/8 \end{bmatrix}$$

Thus for $k \geq 2$,

$$v_k = Q^{k-1} \begin{bmatrix} 1/2 \\ 1/4 \\ 0 \end{bmatrix} = M \begin{bmatrix} 0 & 0 & 0 \\ 0 & (1/2)^{k-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} M^{-1} \begin{bmatrix} 1/2 \\ 1/4 \\ 0 \end{bmatrix} = \begin{bmatrix} (1/2)^{k+2} \\ (1/2)^{k+1} \\ 0 \end{bmatrix}.$$

Thus if $T_1^{(1)}$ is the first time at which the Markov chain visits the state 1 after the initial time $n = 0$, then

$$\mathcal{P}(T_1^{(1)} = \infty | X_0 = 1) = 3/8 \ \ , \mathcal{P}(T_1^{(1)} = \infty | X_0 = 2) = 1/2 \ \text{ and } \ \mathcal{P}(T_1^{(1)} = \infty | X_0 = 3) = 1.$$

□

Next we define the first passage probability

$$f(s', s) \equiv \mathcal{P}\left(T_1^{(s)} < \infty | X_0 = s'\right)$$

to be the conditional probability that the Markov chain ever visits the state s from time $n = 1$ relative to the event that the initial state X_0 is s'. (It is a probability under the conditional probability measure $\mathcal{P}_{\{X_0 = s'\}}$ relative to the event $\{X_0 = s'\}$.) If $s = s'$, then $f(s, s)$ is also called the first return probability of the state s. It is clear that

$$f(s', s) = \sum_{k=1}^{\infty} f_k(s', s)$$

and by Theorem 2.4.6, we have

Theorem 2.4.8 *The first passage probabilities $f(s', s)$ satisfy*

$$f(s', s) = P(s', s) + \sum_{s'' \in S - \{s\}} P(s', s'')f(s'', s)$$

for all $s, s' \in S$. □

The above theorem provides systems of linear equations to solve first passage probabilities as given in the following example.

Example 2.4.9 *Consider the same Markov chain as in Example 2.4.7. Let v be the column vector $[f(1,1), f(2,1), f(3,1)]^t$ and p_1 be the first column of the one-step transition probability matrix P. By Theorem 2.4.8, we have*

$$v = p_1 + Qv$$

which is a system of linear equations, where the matrix Q is as before. By solving the matrix equation, we have

$$v = \begin{bmatrix} f(1,1) \\ f(2,1) \\ f(3,1) \end{bmatrix} = \begin{bmatrix} \frac{3\alpha+5}{8} \\ \frac{\alpha+1}{2} \\ \alpha \end{bmatrix}$$

with α in $[0, 1]$. Similarly we have

$$\begin{bmatrix} f(1,2) \\ f(2,2) \\ f(3,2) \end{bmatrix} = \begin{bmatrix} \frac{\beta+1}{2} \\ \frac{3\beta+5}{8} \\ \beta \end{bmatrix} \text{ and } \begin{bmatrix} f(1,3) \\ f(2,3) \\ f(3,3) \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

with β in $[0, 1]$. The uncertainty of α and β should be resolved by other means. In general, a thorough investigation of the recurrency of each state in the state space will provide enough information. In this case, it is easy to see, from the state transition diagram, that state 3 never reach either state 1 or state 2. Thus, $\alpha = f(3,1) = 0 = f(3,2) = \beta$ and then $f(1,1) = 5/8 = f(2,2)$, $f(2,1) = 1/2 = (1,2)$. These results will be justified later by other methods. □

Definition 2.4.10 *A state s in S is called recurrent if its first return probability is one, i.e $f(s, s) = 1$. Otherwise, it is called transient.* □

This says that a recurrent state will be visited again for sure if it has been visited before. In fact, a recurrent state will be visited again and again forever. To give a more precise meaning, we define r.v. $N^{(s)}$ to be the total number of times at which the Markov chain visits the state s. Then for each non-negative integer m, we have

$$\begin{aligned}
& \{N^{(s)} = m\} \\
= \; & \{X_0 \neq s, T_1^{(s)} < \infty, \ldots, T_m^{(s)} < \infty, T_{m+1}^{(s)} = \infty\} \\
& \cup \{X_0 = s, T_1^{(s)} < \infty, \ldots, T_{m-1}^{(s)} < \infty, T_m^{(s)} = \infty\} \\
= \; & \{X_0 \neq s, T_1^{(s)} < \infty, T_2^{(s)} - T_1^{(s)} < \infty, \ldots, T_{m+1}^{(s)} - T_m^{(s)} = \infty\} \\
& \cup \{X_0 = s, T_1^{(s)} < \infty, T_2^{(s)} - T_1^{(s)} < \infty, \ldots, T_m^{(s)} - T_{m-1}^{(s)} = \infty\} \qquad (2.10)
\end{aligned}$$

and

$$\{N^{(s)} = \infty\} = \cap_{m=1}^{\infty}\{T_m^{(s)} < \infty\} = \cap_{m=1}^{\infty}\{T_m^{(s)} - T_{m-1}^{(s)} < \infty\}$$

where $T_0^{(s)}$ is defined to be 0. Since the first passage probability $f(s', s)$ is equal to $\mathcal{P}(T_1^{(s)} < \infty | X_0 = s')$ and the event $(T_1^{(s)} < \infty)$ is equivalent to the event that there is a $k \geq 1$ such that $X_k = s$ which in turn is equivalent to the event $(N^{(s)} > 0)$ if $s' \neq s$ and $(N^{(s)} > 1)$ if $s' = s$, we have

$$f(s', s) = \begin{cases} \mathcal{P}\left(N^{(s)} > 0 | X_0 = s'\right), & \text{if } s' \neq s, \\ \mathcal{P}\left(N^{(s)} > 1 | X_0 = s\right), & \text{if } s' = s. \end{cases} \tag{2.11}$$

For further investigation, we need the following property about successive visiting times $T_m^{(s)}$ to state s.

Lemma 2.4.11 *For any s', s in S, $1 \leq n_1 < n_2 < \ldots < n_m$, and $k, m \geq 1$,*

$$\mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} = k | X_0 = s', T_1^{(s)} = n_1, T_2^{(s)} = n_2, \ldots, T_m^{(s)} = n_m\right) = f_k(s, s). \tag{2.12}$$

Proof. We fix a state s in S and let $T_m = T_m^{(s)}$ for simplicity. Note that the event $(X_0 = s', T_1 = n_1, T_2 = n_2, \ldots, T_m = n_m)$ is the intersection $(X_{n_m} = s) \cap \Lambda$ of the event $(X_{n_m} = s)$ and an event Λ in $\mathcal{F}_{[0, n_m]}$. Thus for $k \geq 1$, we have

$$\begin{aligned} & \mathcal{P}\left(T_{m+1} - T_m = k | X_0 = s', T_1 = n_1, T_2 = n_2, \ldots, T_m = n_m\right) \\ = \; & \mathcal{P}\left(X_{n_m+1} \neq s, \ldots, X_{n_m+k-1} \neq s, X_{n_m+k} = s | \Lambda, X_{n_m} = s\right) \\ = \; & \mathcal{P}\left(X_{n_m+1} \neq s, \ldots, X_{n_m+k-1} \neq s, X_{n_m+k} = s | X_{n_m} = s\right) \quad \text{by (2.6)} \\ = \; & \mathcal{P}\left(X_1 \neq s, \ldots, X_{k-1} \neq s, X_k = s | X_0 = s\right) \quad \text{by the homogeneity of the Markov chain} \\ = \; & f_k(s, s). \end{aligned}$$

This completes the proof. □

From (2.12), we have

$$\mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} < \infty | X_0 = s', T_1^{(s)} = n_1, T_2^{(s)} = n_2, \ldots, T_m^{(s)} = n_m\right) = f(s, s) \tag{2.13}$$

and

$$\mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} = \infty | X_0 = s', T_1^{(s)} = n_1, T_2^{(s)} = n_2, \ldots, T_m^{(s)} = n_m\right) = (1 - f(s, s)) \tag{2.14}$$

for all $m \geq 1$ and $1 \leq n_1 < n_2 < \ldots < n_m$. If Λ is in the σ-algebra $\mathcal{F}(X_0, T_1^{(s)}, \ldots, T_m^{(s)})$ and is contained in the event $\{T_m^{(s)} < \infty\}$, (2.13) implies that

$$\mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} < \infty | \Lambda\right) = f(s, s) \tag{2.15}$$

and (2.14) implies that

$$\mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} = \infty | \Lambda\right) = 1 - f(s, s). \tag{2.16}$$

for all $m \geq 1$. If $s' \neq s$, we have

$$\begin{aligned}
& \mathcal{P}\left(N^{(s)} = m | X_0 = s'\right) \\
= \ & \mathcal{P}\left(T_1^{(s)} < \infty, T_2^{(s)} - T_1^{(s)} < \infty, \ldots, T_m^{(s)} - T_{m-1}^{(s)} < \infty, T_{m+1}^{(s)} - T_m^{(s)} = \infty | X_0 = s'\right) \\
= \ & \mathcal{P}\left(T_1^{(s)} < \infty | X_0 = s'\right) \mathcal{P}\left(T_2^{(s)} - T_1^{(s)} < \infty | X_0 = s', T_1^{(s)} < \infty\right) \cdots \\
& \mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} = \infty | X_0 = s', T_1^{(s)} < \infty, T_2^{(s)} - T_1^{(s)} < \infty, \ldots, T_m^{(s)} - T_{m-1}^{(s)} < \infty\right) \\
= \ & f(s', s) f(s, s)^{m-1}(1 - f(s, s)) \qquad (2.17)
\end{aligned}$$

for all $m \geq 1$ by (2.15) and (2.16) and

$$\mathcal{P}\left(N^{(s)} = 0 | X_0 = s'\right) = \mathcal{P}\left(T_1^{(s)} = \infty | X_0 = s'\right) = 1 - f(s', s). \qquad (2.18)$$

Similarly if $s' = s$, we have

$$\mathcal{P}\left(N^{(s)} = m | X_0 = s\right) = \begin{cases} f(s, s)^{m-1}(1 - f(s, s)), & \text{if } m \geq 1, \\ 0, & \text{if } m = 0. \end{cases} \qquad (2.19)$$

Finally we have

$$\mathcal{P}\left(N^{(s)} = \infty | X_0 = s'\right) = \begin{cases} f(s', s), & \text{if } f(s, s) = 1, \\ 0, & \text{if } f(s, s) < 1. \end{cases} \qquad (2.20)$$

Note that (2.11) can be verified by (2.17) - (2.20). By letting

$$g(s', s) \equiv \mathcal{P}\left(N^{(s)} = \infty | X_0 = s'\right)$$

denote the conditional probability that the Markov chain will visit the state s infinitely many times relative to the event that the initial state is s', we have

Theorem 2.4.12 *For any $s, s' \in S$,*

1. *$g(s, s) = 1$ if and only if $f(s, s) = 1$,*
2. *$g(s, s) = 0$ if and only if $f(s, s) < 0$,*
3. *$g(s', s) = f(s', s) g(s, s)$.*

Proof. This is a direct consequence of (2.20). □

Define

$$R(s', s) \equiv \mathcal{E}_{\{X_0 = s'\}}\left(N^{(s)}\right)$$

to be the conditional expectation of the total number $N^{(s)}$ of times at which the Markov chain will visit the state s relative to the event that the initial state is s'. By (2.17) – (2.20), we have

Theorem 2.4.13 *For $s \in S$,*

$$R(s,s) = 1/(1 - f(s,s))$$

and for $s' \neq s$,

$$R(s',s) = f(s',s)R(s,s).$$

□

In the above theorem, we have used the conventions $1/0 = \infty$ and $0 \cdot \infty = 0$. The matrix $R = [R(s',s)]$ is called the *potential matrix* of the Markov chain. If the first passage probabilities $f(s',s)$ are known, Theorem 2.4.13 can be used to find the potential matrix. In practice, it is much easier to find the potential matrix first and to use Theorem 2.4.13 to find the first passage probabilities. The following theorem is useful.

Theorem 2.4.14 *The potential matrix R and the one-step transition probability matrix P are related as*

$$R = \sum_{n=0}^{\infty} P^n. \tag{2.21}$$

Proof. We first note that for each sample ω in Ω,

$$N^{(s)}(\omega) = \sum_{n=0}^{\infty} 1_{\{X_n = s\}}(\omega).$$

By the monotone convergence theorem of conditional expectation, we have

$$R(s',s) = \mathcal{E}_{\{X_0=s'\}}\left(N^{(s)}\right) = \sum_{n=0}^{\infty} \mathcal{E}_{\{X_0=s'\}}\left(1_{\{X_n=s\}}\right) = \sum_{n=0}^{\infty} P^{(n)}(s',s).$$

By Theorem 2.3.1, we have

$$R = I + P + P^2 + R^3 + \cdots.$$

□

Example 2.4.15 *Consider the same Markov chain as in Example 2.4.7 with one-step transition probability matrix*

$$P = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{bmatrix}.$$

The matrix P can be diagonalized into

$$P = M^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3/4 & 0 \\ 0 & 0 & 1/4 \end{bmatrix} M$$

with

$$M = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \text{ and } M^{-1} = \begin{bmatrix} 1 & 1/2 & 1/2 \\ 1 & 1/2 & -1/2 \\ 1 & 0 & 0 \end{bmatrix}.$$

Thus,

$$\begin{aligned} R &= \sum_{n=0}^{\infty} P^n \\ &= M^{-1} \left(\sum_{n=0}^{\infty} \Lambda^n \right) M \\ &= M^{-1} \begin{bmatrix} \sum_{n=0}^{\infty} 1 & 0 & 0 \\ 0 & \sum_{n=0}^{\infty} (3/4)^n & 0 \\ 0 & 0 & \sum_{n=0}^{\infty} (1/4)^n \end{bmatrix} M \\ &= M^{-1} \begin{bmatrix} +\infty & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4/3 \end{bmatrix} M \\ &= \begin{bmatrix} 8/3 & 4/3 & +\infty \\ 4/3 & 8/3 & +\infty \\ 0 & 0 & +\infty \end{bmatrix}. \end{aligned}$$

From the diagonal entries of R and by Theorem 2.4.13, we have first passage probabilities as follows:

$$\begin{array}{lll} f(1,1) = 5/8, & f(1,2) = 1/2, & f(1,3) > 0, \\ f(2,1) = 1/2, & f(2,2) = 5/8, & f(2,3) > 0, \\ f(3,1) = 0, & f(3,2) = 0, & f(3,3) = 1. \end{array}$$

It can be seen that the uncertainty with $\alpha = f(3,1)$ and $\beta = f(3,2)$ from solving the first passage probabilities by systems of linear equations in Example 2.4.9 can be resolved here. However, the uncertainty with $f(1,3)$ and $f(2,3)$ here in calculating the first passage probabilities from the potential matrix R cab be resolved by solving systems of linear equations on first passage probabilities as in Example 2.4.9. □

Corollary 2.4.16 *A state s is recurrent if and only if $\sum_{n=0}^{\infty} P^{(n)}(s,s) = \infty$.*

Proof. This is followed from Theorems 2.4.13 and 2.4.14. □

Note that although, by (2.21), we have

$$PR = RP = R - I, \tag{2.22}$$

we cannot subtract the matrix PR from both sides of (2.22) and then add the identity matrix I to both sides to form the matrix equation

$$I = (I - P)R$$

and to claim that $R = (I - P)^{-1}$, even the size of the state space is finite. The reason is that the potential matrix may contain extended value ∞ and we do not know what is ∞ minus ∞.

The first passage probabilities $f_k(s', s)$ from state s' to state s in k steps $k = 1, 2, \ldots, +\infty$ form a probability distribution. (Indeed, it is a probability distribution under the conditional probability measure $\mathcal{P}_{\{X_0=s'\}}$.) We define

$$m(s', s) \equiv \sum_{1 \le k \le \infty} k \cdot f_k(s', s)$$

to be the *mean first passage time* from state s' to state s. If $s' = s$, $m(s, s)$ is also called the mean return time of the state s. It is clear that if s is a transient state, i.e. $f_\infty(s, s) > 0$, then its mean return time is ∞. If s is a recurrent state, its mean return time $m(s, s)$ is also called the *mean recurrent time.*

Definition 2.4.17 *A recurrent state is called* positive *recurrent if its mean recurrent time is finite. Otherwise, it is called* null *recurrent.* □

Theorem 2.4.18 *For* $s, s' \in S$,

$$m(s', s) = 1 + \sum_{s'' \in S - \{s\}} P(s', s'') m(s'', s).$$

Proof. From Theorem 2.4.6, we have

$$\begin{aligned} m(s', s) &= P(s', s) + \sum_{2 \le k \le \infty} k \cdot \sum_{s'' \in S - \{s\}} P(s', s'') f_{k-1}(s'', s) \\ &= P(s', s) + \sum_{s'' \in S - \{s\}} P(s', s'') \sum_{1 \le k \le \infty} (k+1) f_k(s'', s) \\ &= P(s', s) + \sum_{s'' \in S - \{s\}} P(s', s'') \left(m(s'', s) + 1\right) \\ &= 1 + \sum_{s'' \in S - \{s\}} P(s', s'') m(s'', s). \end{aligned}$$

This completes the proof. □

2.4.2 Class properties

Theorem 2.4.19 *If a state* s *communicates with a recurrent state* s', *then it is also recurrent.*

Proof. Since state s' is recurrent, we have

$$\sum_{n=0}^{\infty} P^{(n)}(s', s') = +\infty$$

by Corollary 2.4.16. Since $s \leftrightarrow s'$, there exist n and m such that $P^{(n)}(s,s') > 0$ and $P^{(m)}(s',s) > 0$. By Chapman-Kolmogorov equation, we have

$$\begin{aligned} P^{(n+k+m)}(s,s) &= \sum_{s''\in S} P^{(n)}(s,s'')P^{(k+m)}(s'',s) \\ &\geq P^{(n)}(s,s')P^{(k+m)}(s',s) \\ &= P^{(n)}(s,s')\sum_{s''\in S} P^{(k)}(s',s'')P^{(m)}(s'',s) \\ &\geq P^{(n)}(s,s')P^{(k)}(s',s')P^{(m)}(s',s) \end{aligned}$$

and then

$$\begin{aligned} &\sum_{i=0}^{\infty} P^{(i)}(s,s) \geq \sum_{k=0}^{\infty} P^{(n+k+m)}(s,s) \geq \sum_{k=0}^{\infty} P^{(n)}(s,s')P^{(k)}(s',s')P^{(m)}(s',s) \\ = \ & P^{(n)}(s,s')\left(\sum_{k=0}^{\infty} P^{(k)}(s',s')\right) P^{(m)}(s',s) = +\infty. \end{aligned}$$

The proof is now completed by Theorem 2.4.16. □

The above theorem says that recurrency is a communication class property. Note that a state is either recurrent or transient, but not both, If a state in a communication class is transient, then all other states in the same class must be transient. Otherwise, all states in the class must be recurrent, a contradiction. Thus, transiency is also a class property. Now it is proper to say that a communication class is recurrent or transient.

Lemma 2.4.20 *If s is reachable from a recurrent state s' ($s' \to s$), then s' is also reachable from s ($s \to s'$) and $f(s,s') = 1$.*

Proof. If $s = s'$, it is trivial. We assume $s' \neq s$. Let N be the set of all positive integers n with $P^{(n)}(s',s) > 0$. Since $s' \to s$, N is a non-empty set. Let n_0 be the smallest integer in N. Then s is reached from s' in n_0 steps without returning to s', i.e.

$$\mathcal{P}(X_{n_0} = s, T_1^{(s')} \leq n_0 | X_0 = s') = 0.$$

Otherwise there is an n, $1 \leq n < n_0$, such that

$$\begin{aligned} 0 &< \mathcal{P}(X_{n_0} = s, T_1^{(s')} = n | X_0 = s') \leq \mathcal{P}(X_{n_0} = s, X_n = s' | X_0 = s') \\ &= \mathcal{P}(X_n = s' | X_0 = s') \cdot \mathcal{P}(X_{n_0} = s | X_n = s') \end{aligned}$$

by the Markov property. This implies that

$$0 < \mathcal{P}(X_{n_0} = s | X_n = s') = \mathcal{P}(X_{n_0 - n} = s | X_0 = s') = P^{(n_0 - n)}(s', s)$$

and then $(n_0 - n)$ in N, a contradiction to the minimal property of n_0 in N. Thus we have

$$\begin{aligned} \mathcal{P}(X_{n_0} = s | X_0 = s') &= \mathcal{P}(X_{n_0} = s, T_1^{(s')} \geq n_0 + 1 | X_0 = s') \\ &= \mathcal{P}(X_1 \neq s', \ldots, X_{n_0 - 1} \neq s', X_{n_0} = s | X_0 = s') \end{aligned} \tag{2.23}$$

Now, we have

$$\begin{aligned} & \mathcal{P}(T_1^{(s')} = \infty | X_0 = s') \\ \geq \; & \mathcal{P}(T_1^{(s')} = \infty, X_{n_0} = s | X_0 = s') \\ = \; & \mathcal{P}(X_1 \neq s', \ldots, X_{n_0-1} \neq s', X_{n_0} = s, X_{n_0+1} \neq s', \ldots | X_0 = s') \\ = \; & \mathcal{P}(X_1 \neq s', \ldots, X_{n_0-1} \neq s', X_{n_0} = s | X_0 = s') \cdot \\ & \mathcal{P}(X_{n_0+1} \neq s', \ldots | X_0 = s', X_1 \neq s', \ldots, X_{n_0-1} \neq s', X_{n_0} = s) \\ = \; & \mathcal{P}(X_{n_0} = s | X_0 = s') \mathcal{P}(X_{n_0+1} \neq s', \ldots | X_{n_0} = s) \text{ by (2.23) and the Markov property} \\ = \; & \mathcal{P}(X_{n_0} = s | X_0 = s') \mathcal{P}(X_1 \neq s', \ldots | X_0 = s) \\ = \; & \mathcal{P}(X_{n_0} = s | X_0 = s') \mathcal{P}(T_1^{(s')} = \infty | X_0 = s) \end{aligned}$$

which says that

$$1 - f(s', s') \geq \mathcal{P}(X_{n_0} = s | X_0 = s') \cdot (1 - f(s, s')) \geq 0.$$

Since s' is recurrent, i.e. $f(s', s') = 1$, we have

$$\mathcal{P}(X_{n_0} = s | X_0 = s') \cdot (1 - f(s, s')) = 0$$

and then $f(s, s') = 1$ for $\mathcal{P}(X_{n_0} = s | X_0 = s') > 0$. This implies that there is an n such that

$$0 < \mathcal{P}(T^{(s')} = n | X_0 = s) \leq \mathcal{P}(X_n = s' | X_0 = s) = P^{(n)}(s, s')$$

which says that $s \to s'$. The proof is now completed. □

Definition 2.4.21 *A squared matrix (finite-dimensional or infinite-dimensional) is called a stochastic (or Markov) matrix if its entries are all non-negative and its row sums are all equal to one.* □

The one-step transition probability matrix (at time n) of an (inhomogeneous) Markov chain is a stochastic matrix. Conversely, for a stochastic matrix, we can always construct a homogeneous Markov chain with one-step transition probability matrix to be that matrix.

Definition 2.4.22 *A set C of states is called* closed *if any state s in the state space S which is reachable from a state s' in C must be in C. Otherwise, it is called* open. □

The state space itself is closed. From Lemma 2.4.20, the set of all recurrent states is closed. A communication class may be closed or open.

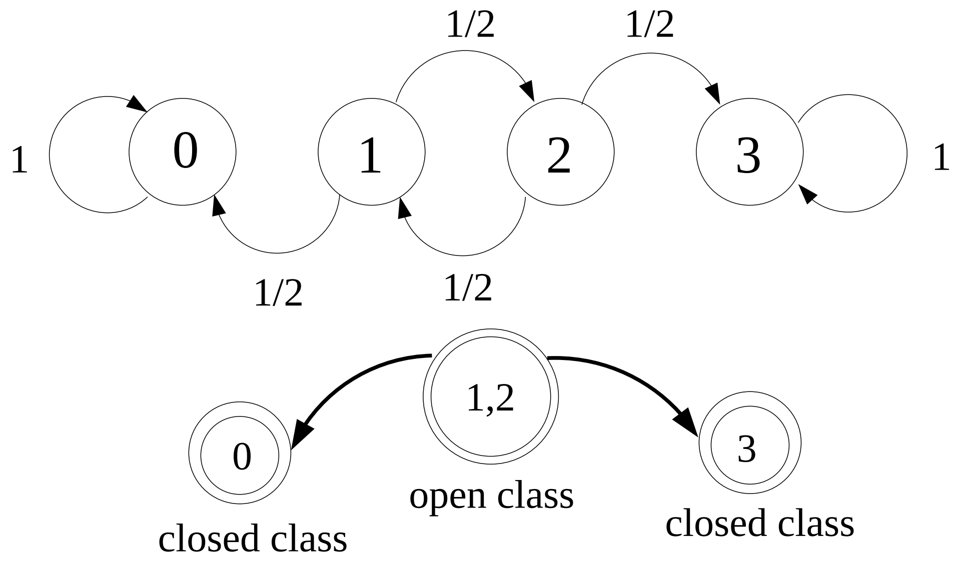

Figure 2.6: Open and closed communication classes in a Markov chain.

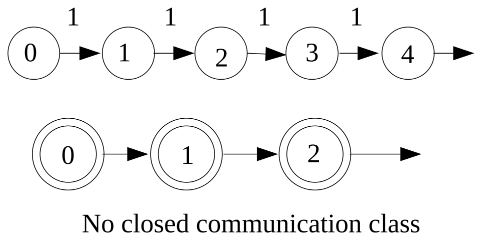

Figure 2.7: A Markov chain without any closed communication class.

Example 2.4.23 *The state-transition diagram in Figure 2.6 has shown that there are two closed and one open communication classes. And in Figure 2.7, a Markov chain without any closed communication class is demonstrated.*

Also by Lemma 2.4.20, every recurrent class is closed. A transient class may be closed or open. But if a transient class has a finite number of states, it must be open, as implied by the following theorem.

Theorem 2.4.24 *If a closed communication class is transient, then it must have an infinite number of states.*

Proof. Let C be a closed transient communication class. Let s' be a state in C. Since $P^{(n)}(s', s) = 0$ for any $n \geq 0$ and any $s \notin C$, we have

$$\mathcal{P}\left(\cap_{n=0}^{\infty}\{X_n \in C\} | X_0 = s'\right) = 1$$

which implies that

$$\mathcal{P}\left(\sum_{s\in C} N^{(s)} = \infty | X_0 = s'\right) = 1. \tag{2.24}$$

Suppose that C is finite. The event $\{\sum_{s\in C} N^{(s)} = \infty\}$ is equal to the union $\cup_{s\in C}\{N^{(s)} = \infty\}$ of events $\{N^{(s)} = \infty\}$. Thus

$$\begin{aligned}
\mathcal{P}\left(\sum_{s\in C} N^{(s)} = \infty | X_0 = s'\right) &= \mathcal{P}\left(\cup_{s\in C}\{N^{(s)} = \infty\} | X_0 = s'\right) \\
&\leq \sum_{s\in C} \mathcal{P}\left(N^{(s)} = \infty | X_0 = s'\right) \\
&= \sum_{s\in C} g(s', s) \\
&= \sum_{s\in C} f(s', s) g(s, s) \\
&= 0
\end{aligned} \tag{2.25}$$

by Theorem 2.4.12 and the transiency of states in C. It can be seen that (2.25) is a contradiction to (2.24). Thus C must be infinite. □

Let P be the one-step transition probability matrix of a Markov chain and C be a closed subset of the state space S. By rearranging the order of states in S such that states in C are labeled at first, the one-step transition probability matrix must be of the form

$$P = \begin{bmatrix} P_C & 0 \\ \star & \star \end{bmatrix}$$

where the entries of the submatrix P_C are those one-step transition probabilities from states to states in the closed subset C. It can be seen that P_C is a stochastic matrix and is the one-step transition probability matrix of the Markov chain obtained from the original Markov chain by reducing the state space to C.

Definition 2.4.25 *A closed set of states is called irreducible if it has no proper closed subset.* □

Lemma 2.4.26 *A closed communication class is irreducible.*

Proof. Let P_C be the stochastic matrix associated with the considered closed communication class C. Suppose that C has a proper closed subset D. By rearranging the states in C, P_C can be of the form

$$P_C = \begin{bmatrix} P_D & 0 \\ \star & \star \end{bmatrix}.$$

Then we have

$$P_C^n = \begin{bmatrix} P_D^n & 0 \\ \star & \star \end{bmatrix}$$

which implies that $P^{(n)}(s,s') = 0$ for all $n \geq 1$ and for any $s \in D$ and $s' \in C - D$. This is a contradiction to $s \leftrightarrow s'$. Thus C has no proper closed subset and is irreducible. □

Let C_1, C_2, …, be all recurrent communication classes in the state space S and D be the set of all transient states. Then by rearranging the order of states in S, the one-step transition probability matrix P can be of the form

$$P = \begin{bmatrix} P_{C_1} & 0 & 0 & \cdots & 0 \\ 0 & P_{C_2} & 0 & \cdots & 0 \\ 0 & 0 & P_{C_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_1 & L_2 & L_3 & \cdots & Q \end{bmatrix}.$$

It is clear that

$$P^n = \begin{bmatrix} P_{C_1}^n & 0 & 0 & \cdots & 0 \\ 0 & P_{C_2}^n & 0 & \cdots & 0 \\ 0 & 0 & P_{C_3}^n & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ L_{1,n} & L_{2,n} & L_{3,n} & \cdots & Q^n \end{bmatrix}.$$

By Theorem 2.4.14, the potential matrix R of the Markov chain is

$$R = \begin{bmatrix} R_{C_1} & 0 & 0 & \cdots & 0 \\ 0 & R_{C_2} & 0 & \cdots & 0 \\ 0 & 0 & R_{C_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{n=0}^{\infty} L_{1,n} & \sum_{n=0}^{\infty} L_{2,n} & \sum_{n=0}^{\infty} L_{3,n} & \cdots & \sum_{n=0}^{\infty} Q^n \end{bmatrix}$$

where R_{C_i} is the potential matrix associated with the recurrent communication class C_i. It can be seen that

$$R(s,s') = 0, \quad \forall\ s \in C_i \text{ and } s' \notin C_i$$

for any recurrent communication class C_i. From Theorem 2.4.13, we have

$$f(s,s') = 0, \quad \forall\ s \in C_i \text{ and } s' \notin C_i$$

for any recurrent communication class C_i. Since each class C_i is recurrent, it is clear from Lemma 2.4.20 and Theorem 2.4.13 that

$$f(s,s') = 1 \quad \text{and} \quad R(s,s') = +\infty$$

for all $s, s' \in C_i$. There are two remaining cases to be investigated. We first consider $R(s,s')$ with s, s' both transient. Let

$$U = \sum_{n=0}^{\infty} Q^n.$$

Since $U(s,s') = R(s,s')$ for all $s,s' \in D$, each entry in U is finite. Thus we have

$$(I-Q)U = U(I-Q) = I. \tag{2.26}$$

Theorem 2.4.27 *If there are only finitely many transient states in the state space S, then*

$$U = (I-Q)^{-1}.$$

□

In general cases where there may exist infinitely many transient states, the following theorem shows that U has a minimal property.

Theorem 2.4.28 *The matrix U is the minimal solution of the matrix equation in* (2.26) *among all non-negative solutions Y*

$$(I-Q)Y = I, \quad Y \geq 0.$$

Proof. As discussed in above, U is a solution of the matrix equation. Suppose that Y is a non-negative solution, i.e. $Y = I + QY$. By iteration, we have

$$Y = I + QY = I + Q + Q^2 Y = I + Q + \ldots + Q^m + Q^{m+1} Y \geq \sum_{n=0}^{m} Q^n$$

for all $m \geq 1$. By taking $m \to \infty$, we have $Y \geq U$. □

Suppose that Y is a non-negative solution of the matrix equation in (2.26), then

$$Y = I + QY \quad \text{and} \quad U = I + QU$$

and then

$$H = QH$$

with $H = Y - U \geq 0$. Thus every column of H satisfies

$$h = Qh, \quad h \geq 0. \tag{2.27}$$

A matrix (possibly having infinite dimension) is called column-bounded if every column of this matrix is a bounded vector. Since $U(s,s') = R(s,s') = f(s,s')R(s',s') = f(s,s')U(s',s') \leq U(s',s')$, every column of the matrix U is bounded and then U is a column-bounded matrix.

Theorem 2.4.29 *The matrix U is the unique column-bounded non-negative solution of the matrix equation in* (2.26) *if and only if the only solution of the system of linear equations in* (2.27)

$$h = Qh, \quad 0 \leq h \leq 1$$

is $h = 0$, where 1 is the column vector whose components are all 1's. □

We next develop a probabilistic non-negative solution for the system of linear equations in (2.27) in a more general setting.

Let A be a subset of the state space S. And let Q be the matrix obtained from the one-step transition probability matrix P by deleting all rows and columns corresponding to states which are not in A. Then the (s,s')-entry of the nth power Q^n of Q is

$$\begin{aligned} Q^n(s,s') &= \sum_{s_1\in A}\sum_{s_2\in A}\cdots\sum_{s_{n-1}\in A} Q(s,s_1)Q(s_1,s_2)\cdots Q(s_{n-1},s') \\ &= \mathcal{P}\left(X_1\in A,\ldots,X_{n-1}\in A, X_n=s'|X_0=s\right) \end{aligned}$$

by the Markov property. Thus we have

$$\sum_{s'\in A} Q^n(s,s') = \mathcal{P}\left(X_1\in A,\ldots,X_{n-1}\in A, X_n\in A|X_0=s\right).$$

Since $\{X_1\in A\}$, $\{X_1\in A, X_2\in A\}$, ... is a decreasing monotone sequence and converges to the event $\cap_{n=1}^{\infty}\{X_n\in A\}$, we have

$$y(s)\equiv \mathcal{P}\left(\cap_{n=1}^{\infty}\{X_n\in A\}|X_0=s\right) = \lim_{n\to\infty}\sum_{s'\in A} Q^n(s,s')$$

for all $s\in A$ by the monotone property of probability measure, where $y(s)$ is the conditional probability that the Markov chain stays in the set A forever given the initial state X_0 to be s. The column vector $y=(y(s), s\in A)^t$ has a maximal property as stated in the following theorem.

Lemma 2.4.30 *Let A be a subset of the state space S. And let Q be the matrix obtained from the one-step transition probability matrix P by deleting all rows and columns corresponding to states which are not in A. The vector $y=(y(s), s\in A)^t$ with $y(s)=\lim_{n\to\infty}\sum_{s'\in A} Q^n(s,s')$ is the maximal solution of the system of linear equations*

$$h = Qh, \quad 0\le h\le 1.$$

And, either $y=0$ or $\sup_{s\in A} y(s)=1$.

Proof. Note that

$$\begin{aligned} y(s) &= \mathcal{P}\left(X_1\in A, X_2\in A,\ldots|X_0=s\right) \\ &= \sum_{s'\in A}\mathcal{P}\left(X_1=s', X_2\in A,\ldots|X_0=s\right) \\ &= \sum_{s'\in A}\mathcal{P}\left(X_1=s'|X_0=s\right)\mathcal{P}\left(X_2\in A, X_3\in A,\ldots|X_0=s, X_1=s'\right) \\ &= \sum_{s'\in A} P(s,s')\mathcal{P}\left(X_2\in A, X_3\in A,\ldots|X_1=s'\right) \quad \text{by Markov property} \\ &= \sum_{s'\in A} P(s,s')\mathcal{P}\left(X_1\in A, X_2\in A,\ldots|X_0=s'\right) \quad \text{by the homogeneity} \\ &= \sum_{s'\in A} P(s,s')y(s'), \end{aligned}$$

which says that $y = Qy$. Since $0 \le y \le 1$, y is a non-negative bounded solution of the system of linear equations. Suppose that h is also a non-negative bounded solution, $0 \le h \le 1$. Then by iteration, we have

$$h = Qh = Q^2 h = \ldots = Q^n h \le Q^n 1$$

for all $n \ge 1$. Since $y = \lim_{n\to\infty} Q^n 1$, we have $h \le y$. Now suppose that $y \ne 0$. Let $c = \sup_{s\in A} y(s) > 0$. Since $y \le c1$, again by iteration, we have

$$y = Qy = Q^2 y = \ldots = Q^n y \le cQ^n 1$$

and by taking $n \to \infty$,

$$y \le cy$$

which implies that $c = 1$. This completes the proof. □

By Theorem 2.4.29 and Lemma 2.4.30, we have

Theorem 2.4.31 *The matrix U is the unique column-bounded non-negative solution of the matrix equation in* (2.26) *if and only if*

$$\sup_{s\in D} \lim_{n\to\infty} \sum_{s'\in D} Q^n(s, s') < 1,$$

where D is the set of all transient states and Q is the matrix obtained from the one-step transition probability matrix P by deleting all rows and columns corresponding to states which are not in D. □

We now return to the previous discussion on potential matrix R and first passage probabilities. With the computed submatrix U of R, we have

$$f(s, s') = \frac{R(s, s')}{R(s', s')} = \frac{U(s, s')}{U(s', s')}$$

for all transient states s, s'. The final case to be discussed is $R(s, s')$ with s a transient state and s' a recurrent state. It is clear that

$$R(s, s') = \begin{cases} 0, & \text{if } f(s, s') = 0, \\ +\infty, & \text{if } f(s, s') > 0, \end{cases}$$

which shows that there still exists uncertainty with $f(s, s')$ for s a transient state and s' a recurrent state, even when $R(s, s')$ is known as in Example 2.4.15. We now need to compute first passage probabilities $f(s, s')$ with s a transient state and s' a recurrent state directly by linear equations provided in Theorem 2.4.8.

Theorem 2.4.32 *Let C be a recurrent communication class and D be the set of all transient states. For each $s' \in C$, the set of first passage probabilities $\{f(s, s'), s \in D\}$ satisfies the system of linear equations*

$$f(s, s') = \sum_{s''\in D} P(s, s'') f(s'', s') + \sum_{s''\in C} P(s, s'').$$

Proof. By Theorem 2.4.8, we have

$$f(s,s') = P(s,s') + \sum_{s'' \in S-\{s'\}} P(s,s'')f(s'',s').$$

If $s'' \notin C$ and $s'' \notin D$, we have $f(s'',s') = 0$. Also, note that $f(s'',s') = 1$ for all $s'' \in C$. The system of linear equations is now followed. □

Let $f(D,s') = (f(s,s'), s \in D)^t$ and $b(D,C) = (\sum_{s' \in C} P(s,s'), s \in D)^t$ be column vectors. Then from Theorem 2.4.32, we have

$$f(D,s') = Qf(D,s') + b(D,C). \tag{2.28}$$

A more informative derivation of (2.28) can be done as follows. For a transient state s in D and a recurrent state s' in the recurrent communication class C, the first passage probability

$$\begin{aligned}
f(s,s') &\equiv \mathcal{P}\left(T_1^{(s')} < +\infty | X_0 = s\right) \\
&= \sum_{k=1}^{\infty} \mathcal{P}\left(X_1 \neq s', \ldots, X_{k-1} \neq s', X_k = s' | X_0 = s\right) \\
&= \sum_{k=1}^{\infty}\sum_{n=1}^{k-1} \mathcal{P}\left(X_1 \in D, \ldots, X_{n-1} \in D, X_n \in C\backslash\{s'\}, \ldots, X_{k-1} \in C\backslash\{s'\}, X_k = s' | X_0 = s\right) \\
&= \sum_{n=1}^{\infty} \mathcal{P}\left(X_1 \in D, \ldots, X_{n-1} \in D, X_n \in C | X_0 = s\right)
\end{aligned}$$

is independent of the state s' in C and will be denoted as $f(s,C)$, which is also called the absorbing probability to the recurrent class C given an initial transient state $X_0 = s$. The column vector $f(D,s')$, $s' \in C$, will also be denoted as $f(D,C)$. Now

$$\begin{aligned}
&\mathcal{P}\left(X_1 \in D, \ldots, X_{n-1} \in D, X_n \in C | X_0 = s\right) \\
=& \sum_{s'' \in D} \mathcal{P}\left(X_1 \in D, \ldots, X_{n-2} \in D, X_{n-1} = s'', X_n \in C | X_0 = s\right) \\
=& \sum_{s'' \in D} \mathcal{P}\left(X_1 \in D, \ldots, X_{n-2} \in D, X_{n-1} = s'' | X_0 = s\right) \mathcal{P}\left(X_n \in C | X_{n-1} = s''\right) \\
=& \sum_{s'' \in D} Q^{n-1}(s,s'')\mathcal{P}\left(X_1 \in C | X_0 = s''\right),
\end{aligned}$$

by the Markov property and the homogeneity of the Markov chain, where Q is the matrix obtained from the one-step transition proabability matrix P by deleting rows and columns corresponding to states not in D, and then we have

$$f(D,C) = \sum_{n=1}^{\infty} Q^{n-1} b(D,C),$$

which implies that the absorbing probability vector $f(D,C)$ satisfies (2.28). If the Markov chain has only finitely many transient states, i.e., the square matrix Q is finite-dimensional, then the matrix $I-Q$ has the inverse matrix U by (2.26) and we have

$$f(D,C) = (I-Q)^{-1}b(D,C) = Ub(D,C).$$

In general, we will show that the absorbing probability vector $f(D,C)$ is the minimal non-negative solution of (2.28). Let y be a non-negative solution of the inhomogeneous system of linear equations in (2.28)

$$y = Qy + b(D,C).$$

Then

$$y = b(D,C) + Qy = b(D,C) + Qb(D,C) + \cdots + Q^{n-1}b(D,C) + Q^n y \geq \sum_{k=0}^{n-1} Q^k b(D,C)$$

for all $n \geq 1$ and by taking $n \to \infty$, we have

$$y \geq f(D,C).$$

Similar to Theorem 2.4.29, we now have

Theorem 2.4.33 *The absorbing probabilities $\{f(s,C), s \in D\}$ from transient states to a recurrent communication class C is the unique bounded non-negative solution of the inhomogeneous system of linear equations in Theorem 2.4.32 if and only if the homogeneous system of linear equations $h = Qh$, $0 \leq h \leq 1$, has only trivial solution $h = 0$.* □

Combined with Lemma 2.4.30, we have

Theorem 2.4.34 *The absorbing probabilities $\{f(s,C), s \in D\}$ from transient states to a recurrent communication class C is the unique bounded non-negative solution of the inhomogeneous system of linear equations in Theorem 2.4.32 if and only if*

$$\sup_{s\in D} \lim_{n\to\infty} \sum_{s'\in D} Q^n(s,s') < 1,$$

□

The following theorem is useful to verify a closed communication class to be recurrent.

Theorem 2.4.35 *Let C be a closed communication class in S and let s be a fixed state in C. Then C is recurrent if and only if for every $s' \in C$ and $s' \neq s$, we have $f(s',s) = 1$.*

Proof. Let $C' = C - \{s\}$. If C is recurrent, then $f(s',s) = 1$ for all $s' \in C'$ by Lemma 2.4.20. Conversely, from Theorem 2.4.8, we have

$$f(s,s) = P(s,s) + \sum_{s'\in C'} P(s,s')f(s',s) = \sum_{s'\in C} P(s,s') = 1$$

since $P(s,s') = 0$ for all $s' \notin C$. This completes the proof. □

The next theorem is an application of Lemma 2.4.30 and can also be used to verify a closed communication class to be recurrent.

Theorem 2.4.36 *Let P_C be the one-step transition probability matrix corresponding to a closed communication class C. Let s be a fixed state in C. And let Q be the matrix obtained from P_C by deleting the row and column corresponding to the state s. Then C is recurrent if and only if the only solution of the system of linear equations*

$$h = Qh, \quad 0 \le h \le 1$$

is $h = 0$.

Proof. Let $C' = C - \{s\}$ for convenience. If C is recurrent, then $f(s', s) = 1$ for all $s' \in C'$. Thus, the conditional probability that the Markov chain will stay in C' forever given initial state $X_0 = s'$ in C' is zero. By Lemma 2.4.30, the only non-negative bounded solution of the system of linear equations is the trivial solution. Conversely, if the only non-negative bounded solution of the system of linear equations is the trivial solution, then the conditional probability that the Markov chain will stay in C' forever given initial state $X_0 = s'$ in C' is zero by Lemma 2.4.30. Thus $f(s', s) = 1$ for all $s' \in C'$. Then by Theorem 2.4.35, C is recurrent. □

Before leaving this section, we shall discuss three more class properties. We first define the period of a state s.

Definition 2.4.37 *The period $d(s)$ of a state s is the greatest common divisor of the set of positive integers n for which $P^{(n)}(s, s) > 0$.* □

A state s is called a non-return state if $P^{(n)}(s, s) = 0$ for all $n \ge 1$. For a non-return state s, we define its period $d(s)$ to be $+\infty$. A return state s is called aperiodic if its period $d(s)$ is one and periodic if $d(s) > 1$. We need the following technical lemma.

Lemma 2.4.38 *Let M be a non-empty set of positive integers, closed under addition. And let d be the greatest common divisor of all integers in M. Then there exists a positive integer k_0 such that $kd \in M$ for all $k \ge k_0$.*

Proof. Let E be the set of all positive integers which are finite linear combinations

$$e_1 n_1 + e_2 n_2 + \cdots + e_k n_k,$$

where $n_1, n_2, \ldots, n_k$ belong to M and $e_1, e_2 \ldots, e_k$ are positive or negative integers. It is clear that M is a subset of E. Let d' be the smallest member in E with

$$d' = a_1 n_1 + a_2 n_2 + \cdots + a_l n_l \tag{2.29}$$

for some integers $n_1, n_2, \ldots, n_l$ in M and positive or negative integers $a_1, a_2, \ldots, a_l$. It is clear that $d \mid d'$. Furthermore, d' is a common divisor of all integers in M. Otherwise, if n

is in M and $d' \not| n$, then there exists positive integers q and r such that $n = qd' + r$ with $1 \leq r < d'$. Thus

$$r = n - qd' = n + (-qa_1)n_1 + (-qa_2)n_2 + \cdots + (-qa_l)n_l$$

is in E and is less than the smallest member d' of E, which is a contradiction. Thus we have $d = d'$. By rearranging the terms in (2.29) such that terms with positive coefficients are written first, we have

$$d = x - y$$

for some positive integers x and y in M. Since d is the greatest common divisor of all integers in M, there exists positive integers x' and y' such that $x = x'd$, $y = y'd$ and then $x' - y' = 1$. Now for any $k \geq y'^2$, there exist unique non-negative integers a and b such that $k = ay' + b$ with $a \geq y'$ and $0 \leq b < y'$. Thus we have

$$k = ay' + b(x' - y') = bx' + (a - b)y'$$

and then

$$kd = bx + (a - b)y \in M.$$

The proof is now completed by letting $k_0 = y'^2$. □

Theorem 2.4.39 *For a return state s with period d, there exists a positive integer k_0 such that $P^{(k \cdot d)}(s, s) > 0$ for all $k \geq k_0$.*

Proof. Let M be the set of all positive integers n such that $P^{(n)}(s, s) > 0$. If n_1 and n_2 are in M, i.e. $P^{(n_1)}(s, s) > 0$ and $P^{(n_2)}(s, s) > 0$, then

$$P^{(n_1+n_2)}(s, s) \geq P^{(n_1)}(s, s)P^{(n_2)}(s, s) > 0$$

where the first inequality is from the Chapman-Kolmogorov equation. Thus $n_1 + n_2$ is in M. This shows that M is closed under addition. Since s is a return state, M is non-empty and the period d of s is the greatest common divisor of all integers in M. By Lemma 2.4.38, there exists a positive integer k_0 such that $k \cdot d \in M$ for all $k \geq k_0$ which says that $P^{(k \cdot d)}(s, s) > 0$ for all $k \geq k_0$. □

Theorem 2.4.40 *Two return states which communicate have the same period.*

Proof. Suppose that $s \leftrightarrow s'$. Then there exist $n, m \geq 1$ such that $P^{(n)}(s, s') > 0$ and $P^{(m)}(s', s) > 0$. As in the proof of Theorem 2.4.19, we have

$$P^{(n+k+m)}(s, s) \geq P^{(n)}(s, s')P^{(k)}(s', s')P^{(m)}(s', s)$$

and

$$P^{(n+2k+m)}(s, s) \geq P^{(n)}(s, s')P^{(k)}(s', s')P^{(k)}(s', s')P^{(m)}(s', s)$$

for all $k \geq 1$. Now for any k such that $P^{(k)}(s', s') > 0$, we have

$$P^{(n+k+m)}(s, s) > 0 \quad \text{and} \quad P^{(n+2k+m)}(s, s) > 0,$$

which implies that $d(s)|(n + k + m)$ and $d(s)|(n + 2k + m)$ and then $d(s)|k$. Since $d(s')$ is the g.c.d. of all such k's, we have $d(s)|d(s')$. Similarly, by changing the roles of s and s', we have $d(s')|d(s)$. We then conclude that $d(s) = d(s')$. □

From the above theorem, period is a class property and we can define the period of a communication class to be the common period of states in the class.

Let s be a recurrent state in the state space S. Let the initial state of the Markov chain be set to s with probability one, i.e., $\mathcal{P}(X_0 = s) = 1$. Consider the successive visiting times $T_1^{(s)}, T_2^{(s)}, \ldots$ of state s. Since s is recurrent, we have $f(s, s) = 1$ and then $\mathcal{P}(N^{(s)} = \infty | X_0 = s) = g(s, s) = 1$ by Theorem 2.4.12. Thus we have

$$\mathcal{P}(N^{(s)} = \infty) = \mathcal{P}(X_0 = s)\mathcal{P}(N^{(s)} = \infty | X_0 = s) = 1$$

and then $\mathcal{P}(T_m^{(s)} < +\infty) = 1$ for all $m \geq 1$. By Lemma 2.4.11, we have

$$\mathcal{P}\left(T_{m+1}^{(s)} - T_m^{(s)} = k | T_1^{(s)} = n_1, T_2^{(s)} = n_2, \ldots, T_m^{(s)} = n_m\right) = f_k(s, s)$$

for all $m \geq 0$. This implies that the intervisit times $T_n^{(s)} - T_{n-1}^{(s)}$, $n \geq 1$, of state s are independent and identically distributed. Furthermore, the first return probabilities in k steps $\{f_k(s, s), k = 1, 2, \ldots\}$ form the distribution of the intervisit times $T_n^{(s)} - T_{n-1}^{(s)}$ and the mean return time $m(s, s)$ is the mean of the distribution.

A sequence $\{T_m, m = 0, 1, \ldots\}$ of finite random times, i.e., $\mathcal{P}(T_m < +\infty) = 1 \; \forall \, m$, taking values in the set of all finite non-negative integers with

$$T_0 \equiv 0 \text{ and } T_0 < T_1 < T_2 < \cdots$$

is called an ordinary recurrent renewal process if the differences $T_m - T_{m-1}$ of successive times are independent and identically distributed. T_m's are called renewal times and $T_m - T_{m-1}$ are called interarrival times. Let $\{f_k, k = 1, 2, \ldots\}$ be the distribution of interarrival times. Let d be the greatest common divisor of all k such that $f_k > 0$. Then the interarrival times only take the multipliers of d as values. And d is called the period of the ordinary recurrent renewal process. The mean m of the distribution of interarrival times is called the mean return time of the renewal process. We next state a general theorem without a proof [2].

Theorem 2.4.41 *[Blackwell's Theorem] For an ordinary recurrent renewal process with period d and mean return time m, we have*

$$\lim_{k \to \infty} \mathcal{P}\left(\textit{there is a renewal at time } kd\right) = \frac{d}{m}.$$

□

As shown in above, the successive visiting times $T_0^{(s)}, T_1^{(s)}, T_2^{(s)}, \ldots$ of a reccurent state s in the Markov chain with initial state set to s, i.e., $\mathcal{P}(X_0 = s) = 1$, form an ordinary recurrent renewal process with the first return probabilities in k steps $\{f_k(s, s), k = 1, 2, \ldots\}$ as the distribution of the intervisit times $T_n^{(s)} - T_{n-1}^{(s)}$ and the mean return time $m(s, s)$ as the mean of this distribution. To apply Blackwell's Theorem, we will show that the period of this ordinary recurrent renewal process is just the period $d(s)$ of the state s. At first, we need the following lemma.

[2] For a proof, please see W. Feller, *An Introduction to Probability Theory and Its Applications,* vol. 2, 2nd edn. New York: Wiley, 1971, pp. 364–366.

Lemma 2.4.42 *For any s, s' in the state space S,*

$$P^{(n)}(s,s') = \sum_{k=1}^{n} f_k(s,s')P^{(n-k)}(s',s')$$

for all $n \geq 1$.

Proof. As an exercise. $\square$

Let $\tilde{d}$ be the period of the ordinary recurrent renewal process $T_0^{(s)}, T_1^{(s)}, T_2^{(s)}, \ldots$. Let n be a positive integer such that $f_n(s,s) > 0$. By Lemma 2.4.42, we have

$$P^{(n)}(s,s) = \sum_{k=1}^{n} f_k(s,s)P^{(n-k)}(s,s) > f_n(s,s) > 0$$

and then n is divisible by the period $d(s)$ of the state s, which implies $d(s)|\tilde{d}$ and

$$P^{(nd(s))}(s,s) = \sum_{k=1}^{n} f_{kd(s)}(s,s)P^{((n-k)d(s))}(s,s) \tag{2.30}$$

again by Lemma 2.4.42. Suppose that $\tilde{d} = \ell d(s)$ for some $\ell > 1$. We will show that $P^{((k\ell+1)d(s))}(s,s) = 0$ for all $k \geq 0$ by induction. When $k = 0$, we know that

$$P^{(d(s))}(s,s) = f_{d(s)}(s,s) = 0$$

by (2.30). Assume that $P^{((i\ell+1)d(s))}(s,s) = 0$ for all $i = 0, 1, \ldots, k-1$. Then again by (2.30), we have

$$P^{((k\ell+1)d(s))}(s,s) = \sum_{i=1}^{k} f_{i\ell d(s)}(s,s)P^{(((k-i)\ell+1)d(s))}(s,s) = 0$$

by the induction step. Thus we have $P^{((k\ell+1)d(s))}(s,s) = 0$ for all $k \geq 0$, which is a contradiction to Theorem 2.4.39 and then we must have $\tilde{d} = d(s)$.

Now for the ordinary recurrent renewal process $T_0^{(s)}, T_1^{(s)}, T_2^{(s)}, \ldots$, we have

$$\begin{aligned}
\mathcal{P}(\text{there is a renewal at time } kd(s)) &= \mathcal{P}\left(X_{kd(s)} = s\right) \\
&= \mathcal{P}(X_{kd(s)} = s | X_0 = s) \\
&= P^{(kd(s))}(s,s)
\end{aligned}$$

and by Blackwell's Theorem, we have the following theorem.

Theorem 2.4.43 *Let s be a recurrent state with period $d(s)$. Then*

$$\lim_{k\to\infty} P^{(kd(s))}(s,s) = \frac{d(s)}{m(s,s)}.$$

$\square$

If s is periodic positive recurrent with period $d > 1$, then

$$\liminf_n P^{(n)}(s,s) = 0 \quad \text{and} \quad \limsup_n P^{(n)}(s,s) = \frac{d}{m(s,s)} > 0.$$

Thus the limit of the sequence $\{P^{(n)}(s,s), n = 0, 1, \ldots\}$ does not exist. If s is periodic null recurrent with period $d > 1$, then

$$\liminf_n P^{(n)}(s,s) = 0 \quad \text{and} \quad \limsup_n P^{(n)}(s,s) = \frac{d}{m(s,s)} = 0.$$

Thus the limit of the sequence $\{P^{(n)}(s,s), n = 0, 1, \ldots\}$ exists and equals to 0. Finally, if s is aperiodic recurrent, then

$$\lim_{n\to\infty} P^{(n)}(s,s) = \frac{1}{m(s,s)}.$$

We now summarize the above discussion in the following corollary.

Corollary 2.4.44 *Let s be a recurrent state. Then s is null recurrent if and only if* $\lim_{n\to\infty} P^{(n)}(s,s) = 0$. □

Theorem 2.4.45 *If s is a null recurrent state and s communicates with s', then s' is also a null recurrent state.*

Proof. As in the proof of Theorem 2.4.19, we have

$$P^{(n+k+m)}(s,s) \geq P^{(n)}(s,s')P^{(k)}(s',s')P^{(m)}(s',s)$$

where $P^{(n)}(s,s') > 0$ and $P^{(m)}(s',s) > 0$ for some $n, m \geq 1$ and for all $k \geq 0$. By taking limsup as $k \to \infty$ on both sides, we have

$$0 = \limsup_{k\to\infty} P^{(n+k+m)}(s,s) \geq P^{(n)}(s,s')P^{(m)}(s',s) \limsup_{k\to\infty} P^{(k)}(s',s') \geq 0$$

which implies $\lim_{k\to\infty} P^{(k)}(s',s') = 0$ and then s' is a null recurrent state by Theorem 2.4.19 and Corollary 2.4.44. □

The above theorem says that null recurrency and positive recurrency are class properties.

Corollary 2.4.46 *If s' is a null recurrent state, then for any state s in the state space S,*

$$\lim_{n\to\infty} P^{(n)}(s,s') = 0.$$

Proof. From Lemma 2.4.42, we have

$$\begin{aligned} P^{(n)}(s,s') &= \sum_{k=1}^{n} f_k(s,s')P^{(n-k)}(s',s') \\ &= \sum_{k=1}^{N} f_k(s,s')P^{(n-k)}(s',s') + \sum_{k=N+1}^{n} f_k(s,s')P^{(n-k)}(s',s') \\ &\le \sum_{k=1}^{N} f_k(s,s')P^{(n-k)}(s',s') + \sum_{k=N+1}^{n} f_k(s,s') \quad \text{since } P^{(n-k)}(s',s') \le 1 \end{aligned}$$

for any given N and for all $n \ge N$. We take limit superior over n first and obtain

$$\begin{aligned} &\limsup_n P^{(n)}(s,s') \\ \le\ & \sum_{k=1}^{N} f_k(s,s') \limsup_n P^{(n-k)}(s',s') + \limsup_n \sum_{k=N+1}^{n} f_k(s,s') \\ =\ & \sum_{k=N+1}^{\infty} f_k(s,s') \quad \text{by Corollary 2.4.44} \end{aligned}$$

for all N. By taking $N \to \infty$, we have $\limsup_n P^{(n)}(s,s') = 0$ and then $\lim_{n\to\infty} P^{(n)}(s,s') = 0$. □

Theorem 2.4.47 *If C is a null recurrent communication class, then C has infinitely many states.*

Proof. For any $s \in C$, we have

$$1 = \sum_{s' \in S} P^{(n)}(s,s') = \sum_{s' \in C} P^{(n)}(s,s')$$

for all $n \ge 1$, since C is closed. Thus we have

$$\lim_{n\to\infty} \sum_{s' \in C} P^{(n)}(s,s') = 1.$$

Suppose that C is finite. Then

$$\lim_{n\to\infty} \sum_{s' \in C} P^{(n)}(s,s') = \sum_{s' \in C} \lim_{n\to\infty} P^{(n)}(s,s') = 0$$

by Corollary 2.4.46, which is a contradiction. Thus C must be infinite. □

In Corollary 2.4.46, we have shown that for a null recurrent state s',

$$\lim_{n\to\infty} P^{(n)}(s,s') = 0 \tag{2.31}$$

for any state $s \in S$. Furthermore, for a transient state s', we have

$$\lim_{n\to\infty} P^{(n)}(s, s') = 0 \tag{2.32}$$

for any state $s \in S$, since $R(s, s') = \sum_{n=0}^{\infty} P^{(n)}(s, s')$ is finite by Theorem 2.4.13. The next theorem gives similar results for positive recurrent states.

Theorem 2.4.48 *For a periodic positive recurrent state s' with period $d(s')$, we have*

$$\lim_{n\to\infty} P^{(nd(s')+i)}(s, s') = \frac{f^{(i)}(s, s')d(s')}{m(s', s')}$$

for any state $s \in S$ and $0 \le i \le d(s') - 1$, where $f^{(i)}(s, s') = \sum_{k=0}^{\infty} f_{kd(s')+i}(s, s')$.

Proof. From Theorem 2.4.43, we have

$$\lim_{n\to\infty} P^{(nd(s'))}(s', s') = \frac{d(s')}{m(s', s')}.$$

And from Lemma 2.4.42, we have

$$P^{(nd(s')+i)}(s, s') = \sum_{k=1}^{n} f_{kd(s')+i}(s, s')P^{((n-k)d(s'))}(s', s') \ge \sum_{k=1}^{N} f_{kd(s')+i}(s, s')P^{((n-k)d(s'))}(s', s')$$

for all $n \ge N \ge 1$. By fixing N and taking $\liminf$ over all $n \ge N$, we have

$$\liminf_{n\ge N} P^{(nd(s')+i)}(s, s') \ge \sum_{k=1}^{N} f_{kd(s')+i}(s, s') \liminf_{n\ge N} P^{((n-k)d(s'))}(s', s') = \frac{d(s')}{m(s', s')} \sum_{k=1}^{N} f_{kd(s')+i}(s, s')$$

for all $N \ge 1$. Since

$$\liminf_{n} P^{(nd(s')+i)}(s, s') = \liminf_{n\ge N} P^{(nd(s')+i)}(s, s')$$

for any $N \ge 1$, we have

$$\liminf_{n} P^{(nd(s')+i)}(s, s') \ge \frac{f^{(i)}(s, s')d(s')}{m(s', s')}$$

by taking $N \to \infty$. On the other hand, as in the proof of Corollary 2.4.46, we have

$$\limsup_{n} P^{(nd(s')+i)}(s, s') \le \frac{f^{(i)}(s, s')d(s')}{m(s', s')}.$$

We now conclude that

$$\lim_{n\to\infty} P^{(nd(s')+i)}(s, s') = \frac{f^{(i)}(s, s')d(s')}{m(s', s')}.$$

□

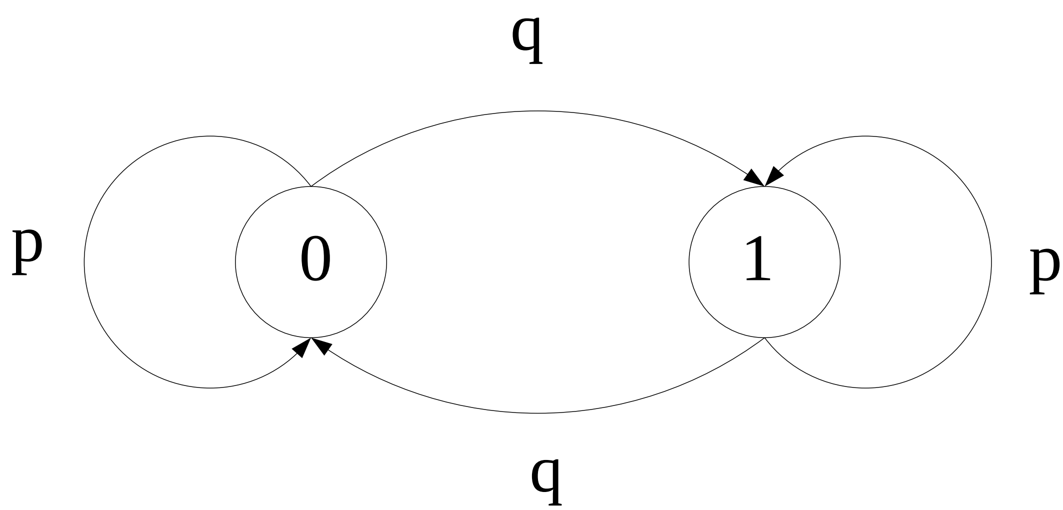

Figure 2.8: The state transition diagram of Example 2.5.2.

2.5 Limiting Behavior

In this section, we will develop the limiting distribution of a homogeneous Markov chain.

Definition 2.5.1 *A (homogeneous) Markov chain is said to have an invariant distribution (or a stationary distribution) if there exists a probability vector* π *(a non-negative vector with the sum of all its components to be 1) such that*

$$\pi = \pi P.$$

Any such a probability vector π *is called an invariant distribution of the Markov chain.* □

Example 2.5.2 *Consider a Markov chain with state transition diagram as shown in Figure 2.8, whose one-step transition probability matrix is*

$$\mathbf{P} = \begin{bmatrix} p & q \\ q & p \end{bmatrix},$$

With $\pi = [1/2, 1/2]$*, we have*

$$\pi P = [1/2\ 1/2] \begin{bmatrix} p & q \\ q & p \end{bmatrix} = [1/2\ 1/2] = \pi$$

and then $\pi = [1/2\ 1/2]$ *is an invariant distribution of the Markov chain.* □

Suppose that a Markov chain has an invariant distribution π. Let π be the initial distribution of the Markov chain, i.e. $\pi_0 = \pi$. Then the distribution π_n of states at time n is

$$\pi_n = \pi_0 P^n = (\pi P) P^{n-1} = \pi P^{n-1} = \cdots = \pi \tag{2.33}$$

for all $n \geq 0$. Furthermore,

$$\begin{aligned} & \mathcal{P}(X_m = s_0, X_{m+1} = s_1, \ldots, X_{m+n} = s_n) \\ = \; & \mathcal{P}(X_m = s_0) \mathcal{P}(X_{m+1} = s_1, \ldots, X_{m+n} = s_n | X_m = s_0) \\ = \; & \mathcal{P}(X_0 = s_0) \mathcal{P}(X_1 = s_1, \ldots, X_n = s_n | X_0 = s_0) \quad \text{by (2.33) and the homogeneity} \\ = \; & \mathcal{P}(X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n) \end{aligned}$$

which says that the joint distribution of $X_0, X_1, \ldots, X_n$ is the same as the joint distribution of $X_m, X_{m+1}, \ldots, X_{m+n}$ for any $n, m \geq 0$.

Definition 2.5.3 *A stochastic process $\{X_n, n = 0, 1, \ldots\}$ is said to be strictly stationary if the joint distribution of $X_{n_1}, X_{n_2}, \ldots, X_{n_k}$ is the same as the joint distribution of $X_{m+n_1}, X_{m+n_2}, \ldots, X_{m+n_k}$ for all $0 \leq n_1 < n_2 < \ldots < n_k$, for all k and for all m.* □

It is now clear that a Markov chain with an invariant initial distribution is strictly stationary. We next prove several convergence theorems which are needed for future purposes.

Theorem 2.5.4 [Fatou's lemma] *Let $Y_1, Y_2, \ldots$ be a sequence of non-negative r.v.'s. Then*

$$\mathcal{E}\left(\liminf_n Y_n\right) \leq \liminf_n \mathcal{E}(Y_n).$$

Proof. Let

$$Z_k = \inf_{n \geq k} Y_n, \ \forall\, k \geq 1.$$

Then $Z_k \leq Y_k$ and $\mathcal{E}(Z_k) \leq \mathcal{E}(Y_k)$ for all $k \geq 1$ by Lemma 1.4.4. Thus we have

$$\liminf_k \mathcal{E}(Z_k) \leq \liminf_k \mathcal{E}(Y_k) \tag{2.34}$$

by the eighth property listed after Definition 1.3.10. Since $Z_1, Z_2, \ldots$ is an increasing sequence of r.v.'s and converges to $\liminf_n Y_n$ by Definition 1.3.10, we have

$$\liminf_k \mathcal{E}(Z_k) = \lim_{k \to \infty} \mathcal{E}(Z_k) = \mathcal{E}\left(\lim_{k \to \infty} Z_k\right) = \mathcal{E}\left(\liminf_k Y_k\right) \tag{2.35}$$

by the Lebesgue's monotone convergence theorem. Combining (2.34) and (2.35), the proof is completed. □

Theorem 2.5.5 [Lebesgue's dominated convergence theorem] *Let $Y_1, Y_2, \ldots$ be a sequence of r.v.'s such that there exists a r.v. Y with*

$$Y = \lim_{n \to \infty} Y_n \ \textit{w.p.1.}$$

If there exists a non-negative r.v. Z with finite expectation such that

$$|Y_n| \leq Z \text{ w.p.1} \quad \forall\, n \geq 1,$$

then the r.v. Y has finite expectation, $\lim_{n\to\infty} \mathcal{E}\left(|Y_n - Y|\right) = 0$ *and*

$$\lim_{n\to\infty} \mathcal{E}\left(Y_n\right) = \mathcal{E}\left(Y\right) = \mathcal{E}\left(\lim_{n\to\infty} Y_n\right).$$

Proof. Since $|Y_n| \leq Z$ w.p.1, we have $\mathcal{E}(|Y_n|) \leq \mathcal{E}(Z) < \infty$ by Lemma 1.4.4. Also since $|Y_n| \leq Z$ w.p.1, we have $|Y| \leq Z$ w.p.1 and then $\mathcal{E}(|Y|) < \infty$, $|Y_n - Y| \leq 2Z$ w.p.1 and $\mathcal{E}(|Y_n - Y|) < \infty$. Now we have

$$\begin{aligned}
& \mathcal{E}\left(2Z\right) \\
= \; & \mathcal{E}\left(\liminf_n \left\{2Z - |Y_n - Y|\right\}\right) \\
\leq \; & \liminf_n \mathcal{E}\left(2Z - |Y_n - Y|\right) \quad \text{by applying Fatou's lemma to} \\
& \text{the sequence } \{2Z - |Y_n - Y|, n = 1, 2 \ldots\} \text{ of r.v.'s} \\
= \; & \mathcal{E}\left(2Z\right) + \liminf_n \left\{-\mathcal{E}\left(|Y_n - Y|\right)\right\} \quad \text{by Theorem 1.4.11} \\
= \; & \mathcal{E}\left(2Z\right) - \limsup_n \mathcal{E}\left(|Y_n - Y|\right),
\end{aligned}$$

which implies that

$$\limsup_n \mathcal{E}\left(|Y_n - Y|\right) \leq 0.$$

Since $|Y_n - Y| \geq 0$ for all n, we have

$$\liminf_n \mathcal{E}\left(|Y_n - Y|\right) \geq 0$$

and then

$$\lim_{n\to\infty} \mathcal{E}\left(|Y_n - Y|\right) = 0.$$

Finally since

$$|\mathcal{E}(Y_n) - \mathcal{E}(Y)| \leq \mathcal{E}(|Y - Y_n|)$$

by the absolute-valued dominancy property of expectation in Theorem 1.5.8, we have

$$\lim_{n\to\infty} \mathcal{E}(Y_n) = \mathcal{E}(Y).$$

This completes the proof. □

Theorem 2.5.6 [Bounded convergence theorem] *Let $Y_1, Y_2, \ldots$ be a sequence of r.v.'s such that there exists a r.v. Y with*

$$Y = \lim_{n\to\infty} Y_n \text{ w.p.1.}$$

If there exists a constant M such that

$$|Y_n| \leq M \text{ w.p.1} \quad \forall\, n \geq 1,$$

then the r.v. Y has finite expectation, $\lim_{n\to\infty} \mathcal{E}\left(|Y_n - Y|\right) = 0$ *and*

$$\lim_{n\to\infty} \mathcal{E}\left(Y_n\right) = \mathcal{E}\left(Y\right) = \mathcal{E}\left(\lim_{n\to\infty} Y_n\right).$$

Proof. Apply the dominated convergence theorem to the sequence $Y_1, Y_2, \ldots$ of r.v.'s with $Z = M$ a constant r.v. which has finite expectation clearly. □

Corollary 2.5.7 *Let S be a countable set. Let $f_1(s), f_2(s), \ldots$ be a sequence of real-valued functions on S such that the limit $f(s) = \lim_{n\to\infty} f_n(s)$ exists for each $s \in S$. Let $w(s)$ be a non-negative function on S with $\sum_{s\in S} w(s) < \infty$. If there exists a constant M such that $|f_n(s)| \le M$ for all $n = 1, 2, \ldots$ and all $s \in S$, then the sums $\sum_{s\in S} w(s)f(s)$, $\sum_{s\in S} w(s)f_n(s)$, $n = 1, 2, \ldots$, exist and are finite such that*

$$\lim_{n\to\infty} \sum_{s\in S} w(s)f_n(s) = \sum_{s\in S} w(s)f(s) = \sum_{s\in S} w(s) \lim_{n\to\infty} f_n(s).$$

Proof. We first construct a discrete probability space $(\Omega, \mathcal{F}, \mathcal{P})$ as follows. Let $\Omega = S$ and $\mathcal{F} = 2^S$. Define a p.m. $\mathcal{P}$ by assigning the singleton $\{s\}$ with probability

$$\mathcal{P}(\{s\}) = \frac{w(s)}{\sum_{s'\in S} w(s')}$$

for any $s \in S$. Then f and f_n's become r.v.'s on this probability space and

$$\mathcal{E}(f_n) = \frac{1}{\sum_{s'\in S} w(s')} \sum_{s\in S} w(s)f_n(s), \quad \mathcal{E}(f) = \frac{1}{\sum_{s'\in S} w(s')} \sum_{s\in S} w(s)f(s).$$

By applying the bounded convergence theorem to the sequence $f_1, f_2, \ldots$ of r.v.'s with $Z = M$, the proof is completed. □

Theorem 2.5.8 *For an irreducible aperiodic Markov chain, all states are positive recurrent if and only if the Markov chain has an invariant distribution. Furthermore, if such an invariant distribution π exists, then it is unique and*

$$\pi(s) = \lim_{n\to\infty} P^{(n)}(s', s) = \frac{1}{m_(s, s)},$$

for any $s', s \in S$.

Proof. We first assume that all states are positive recurrent. Then by Theorem 2.4.48, we have

$$\lim_{n\to\infty} P^{(n)}(s', s) = \frac{1}{m(s, s)} > 0 \tag{2.36}$$

for any $s', s \in S$. For convenience, let $\gamma(s) = 1/m(s, s)$, $\forall\ s \in S$. Now by the Chapman-Kolmogorov equation, we have

$$P^{(n+1)}(s', s) = \sum_{s''\in S} P^{(n)}(s', s'')P(s'', s)$$

for $s', s \in S$. By letting $n \to \infty$, we have

$$\gamma(s) = \lim_{n\to\infty} P^{(n+1)}(s', s) = \lim_{n\to\infty} \sum_{s''\in S} P^{(n)}(s', s'')P(s'', s). \tag{2.37}$$

With $f_n(s'') = P^{(n)}(s', s'')$ and $w(s'') = P(s'', s)$, we have $\lim_{n\to\infty} f_n(s'') = \gamma(s'')$ by (2.36), $\sum_{s''\in S} w(s'') = 1 < \infty$, and $|f_n(s'')| < 2$. By Corollary 2.5.7, (2.37) becomes

$$\gamma(s) = \lim_{n\to\infty} \sum_{s''\in S} P^{(n)}(s', s'')P(s'', s) = \sum_{s''\in S} \left(\lim_{n\to\infty} P^{(n)}(s', s'')\right) P(s'', s) = \sum_{s''\in S} \gamma(s'')P(s'', s)$$

and in matrix form, we have

$$\gamma = \gamma P$$

where $\gamma = (\gamma(s), s \in S)$. We next show that $\alpha = \sum_{s\in S} \gamma(s)$ is finite. Let A be a finite subset of S. Then

$$1 = \sum_{s\in S} P^{(n)}(s', s) \geq \sum_{s\in A} P^{(n)}(s', s)$$

and by letting $n \to \infty$ on both side, we have

$$1 \geq \sum_{s\in A} \gamma(s).$$

Now by a given increasing sequence $A_1 \subseteq A_2 \subseteq \ldots$ of finite subsets of S with $\lim_{i\to\infty} A_i = \cup_{i=1}^{\infty} A_i = S$, we have

$$\sum_{s\in A_1} \gamma(s) \leq \sum_{s\in A_2} \gamma(s) \leq \ldots$$

and then

$$1 \geq \lim_{i\to\infty} \sum_{s\in A_i} \gamma(s) = \sum_{s\in S} \gamma(s).$$

Now α is a positive number and then $\alpha^{-1}\gamma$ becomes to a probability vector and satisfies the matrix equation $(\alpha^{-1}\gamma) = (\alpha^{-1}\gamma)P$, which implies that $\alpha^{-1}\gamma = (\alpha^{-1}\gamma(s)), s \in S$, is an invariant distribution of the Markov chain. Suppose that μ be another invariant distribution of the Markov chain. Since

$$\mu(s) = \sum_{s'\in S} \mu(s')P^{(n)}(s', s)$$

for all $n \geq 1$, we have

$$\begin{aligned}
& \mu(s) \\
= \; & \lim_{n\to\infty} \sum_{s'\in S} \mu(s')P^{(n)}(s', s) \\
= \; & \sum_{s'\in S} \mu(s') \lim_{n\to\infty} P^{(n)}(s', s) \quad \text{by Corollary 2.5.7} \\
= \; & \sum_{s'\in S} \mu(s')\gamma(s) \\
= \; & \gamma(s),
\end{aligned}$$

which implies that α is equal to 1, γ is a probability vector, and there is one and only one invariant distribution of the Markov chain, which is equal to γ. Conversely, we assume that the Markov chain has an invariant distribution π. If all states are transient or null recurrent, then we have

$$\begin{aligned} & \pi(s) \\ = & \sum_{s' \in S} \pi(s') \lim_{n \to \infty} P^{(n)}(s', s) \quad \text{again by Corollary 2.5.7} \\ = & \sum_{s' \in S} \pi(s') 0 \quad \text{by (2.32) and (2.31)} \\ = & \; 0, \end{aligned}$$

for any $s \in S$, which is a contradiction to the fact that π is a probability vector. Thus all states are positive recurrent. This completes the proof. □

To extend the above results to irreducible Markov chain with period d greater than 1, we note that from Theorem 2.4.48, we have

$$\lim_{n \to \infty} \frac{1}{d(s)} \sum_{i=0}^{d(s)-1} P^{(nd(s)+i)}(s', s) = \frac{f(s', s)}{m(s, s)} \tag{2.38}$$

for any $s' \in S$ and for any positive recurrent state s with period $d(s)$. With (2.38), we have the following extension of Theorem 2.5.8.

Theorem 2.5.9 *For an irreducible Markov chain with period d, all states are positive recurrent if and only if the Markov chain has an invariant distribution. Furthermore, if such an invariant distribution π exists, then it is unique and*

$$\pi(s) = \lim_{n \to \infty} \frac{1}{d} \sum_{i=0}^{d-1} P^{(nd+i)}(s', s) = \frac{1}{m_(s, s)},$$

for any $s', s \in S$.

Proof. Left as an exercise. □

Corollary 2.5.10 *For an (aperiodic or periodic) irreducible Markov chain with finite state space S, the system of homogeneous linear equations*

$$\pi = \pi P, \quad \pi \geq 0, \quad \sum_{s \in S} \pi(s) = 1,$$

has a unique solution, which is the invariant distribution of the Markov chain.

Proof. Left as an exercise. □

We shall call an irreducible aperiodic positive recurrent Markov chain as an *ergodic* Markov chain. It is important to note that for an ergodic Markov chain with one-step transition probability matrix P and an arbitrarily given initial distribution π_0, the limit of the distribution $\pi_n = (\pi_n(s), s \in S)$ at time n as $n \to \infty$ is

$$\lim_{n\to\infty} \pi_n(s) = \lim_{n\to\infty} \sum_{s'\in S} \pi_0(s')P^{(n)}(s', s) = \sum_{s'\in S} \pi_0(s') \lim_{n\to\infty} P^{(n)}(s', s) = \sum_{s'\in S} \pi_0(s') \frac{1}{m(s, s)} = \frac{1}{m(s, s)}$$

for all $s \in S$, which is independent of π_0.

Definition 2.5.11 *A Markov chain is said to have a long-run (or steady-state) distribution,* $\gamma = (\gamma(s), s \in S)$, $\gamma(s) \geq 0$, $\sum_{s\in S} \gamma(s) = 1$, *if*

$$\gamma(s) = \lim_{n\to\infty} \pi_n(s), \ \forall \ s \in S,$$

independent of the initial distribution π_0. □

Note that long-run distribution, if exists, must be unique.

Lemma 2.5.12 *A Markov chain has a long-run distribution* $\gamma = (\gamma(s), s \in S)$ *if and only if* $\gamma(s) = \lim_{n\to\infty} P^{(n)}(s', s)$ *for all* $s', s \in S$, *i.e.*

$$\lim_{n\to\infty} P^{(n)} = \begin{bmatrix} \gamma \\ \gamma \\ \vdots \\ \gamma \end{bmatrix}.$$

Proof. We first assume that a long-run distribution γ exists. Fix a state s_0 and let π_0 be an initial distribution with

$$\pi_0(s) = \begin{cases} 1, & \text{if } s = s_0, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\pi_n(s) = P^{(n)}(s_0, s)$$

and as $n \to \infty$,

$$\gamma(s) = \lim_{n\to\infty} \pi_n(s) = \lim_{n\to\infty} P^{(n)}(s_0, s),$$

independent of s_0. Conversely, we assume that $\gamma(s) = \lim_{n\to\infty} P^{(n)}(s', s)$ for all $s', s \in S$. Then we have

$$\lim_{n\to\infty} \pi_n(s) = \lim_{n\to\infty} \sum_{s'\in S} \pi_0(s')P^{(n)}(s', s) = \sum_{s'\in S} \pi_0(s') \lim_{n\to\infty} P^{(n)}(s', s) = \sum_{s'\in S} \pi_0(s')\gamma(s) = \gamma(s)$$

for all $s \in S$, which is independent of π_0. This completes the proof. □

It is now clear that an ergodic Markov chain has a long-run distribution.

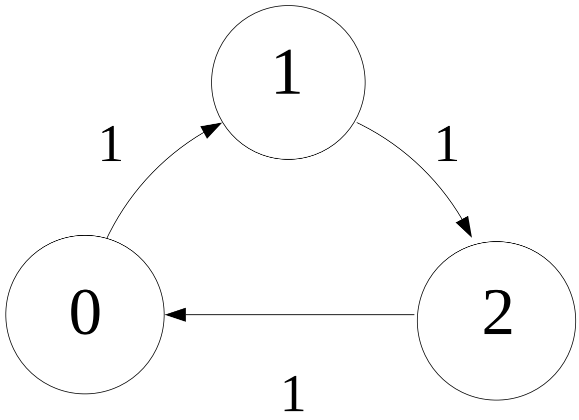

Figure 2.9: An irreducible Markov chain with periodic states.

Theorem 2.5.13 *If a Markov chain has a long-run distribution* $\gamma = (\gamma(s), s \in S)$, *then* γ *is the only invariant distribution of the Markov chain.*

Proof. Left as an exercise. □

The following example shows that aperiodicity is required for a positive recurrent irreducible Markov chain to have a long-run distribution.

Example 2.5.14 *Consider a Markov chain with state space* $\{0, 1, 2\}$ *and state-transition diagram as shown in Figure 2.9. It can be seen that this Markov chain is irreducible, positive recurrent and periodic with period 3. This Markov chain has an invariant distribution* $(1/3, 1/3, 1/3)$. *But for* $m \geq 0$, *we have*

$$\mathbf{P}^{(3m+1)} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{P}^{(3m+2)} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{P}^{(3m)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which implies that $\lim_{n\to\infty} P^{(n)}$ *does not exist. And then this Markov chain has no long-run distribution.* □

Corollary 2.5.15 *For an aperiodic irreducible Markov chain with finite state space* S, *the system of homogeneous linear equations*

$$\pi = \pi P, \quad \pi \geq 0, \quad \sum_{s \in S} \pi(s) = 1,$$

has a unique solution, which is the long-run distribution of the Markov chain.

Proof. This is a direct consequence of Corollary 2.5.10. □

The following example demonstrates a typical limiting behavior of a Markov chain with more than one positive recurrent classes.

Example 2.5.16 *Consider a Markov chain with state space $S = \{1, 2, 3, 4\}$ and one-step transition probability matrix*

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}.$$

It can be easily found that

$$P^{(n)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3(1-1/4^n) & 1/4^n & 1/3(1-1/4^n) & 1/3(1-1/4^n) \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}, \quad \forall\, n \geq 1.$$

Thus we have

$$\lim_{n\to\infty} P^{(n)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix},$$

which says that this Markov chain has no long-run distribution. On the other hand, this Markov chain has two aperiodic positive recurrent classes $C_1 = \{1\}$ and $C_2 = \{3, 4\}$ which have long-run distributions $\gamma_1 = (1)$ and $\gamma_2 = (1/2, 1/2)$ associated with the two restricted irreducible positive recurrent aperiodic Markov chains with state spaces C_1 and C_2 respectively. Now for any given α, $0 \leq \alpha \leq 1$, the probability vector

$$\mu = \alpha(1, 0, 0, 0) + (1 - \alpha)(0, 0, 1/2, 1/2)$$

is an invariant distribution of the original Markov chain since

$$\mu = \mu P.$$

Thus this Markov chain has uncountably many invariant distributions. □

Chapter 3

Second-order Processes

3.1 Hilbert Spaces

A complex inner product space H is a vector space over the complex numbers C together with an additional binary operation in H, called an inner product, which associates each ordered pair of vectors x and y in H with a complex number (x, y) such that the following properties hold:

1. $(x, x) \geq 0$ for all $x \in H$ and $(x, x) = 0$ if and only if $x = 0$. (Non-negativity)
2. $(y, x) = \overline{(x, y)}$. (Hermitian symmetry)
3. $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$ for all x, y, z in H and α, β in C. (Linearity)

Example 3.1.1 *The vector space C^n of all n-tuples over C becomes a complex inner product space if we define an inner product in C^n as follows:*

$$(x, y) = \sum_{i=1}^{n} \eta_i \overline{\xi_i}$$

where $x = (\eta_1, \eta_2, \ldots, \eta_n)$ and $y = (\xi_1, \xi_2, \ldots, \xi_n)$. □

Example 3.1.2 *The vector space $C[0, 1]$ of all continuous complex functions on $[0, 1]$ is a complex inner product space if*

$$(f, g) = \int_0^1 f(t)\overline{g(t)}dt$$

is defined as an inner product of f and g in $C[0, 1]$. (Please verify it.) Contrast to C^n which is a finite-dimensional complex vector space, the vector space $C[0, 1]$ is infinite-dimensional. Suppose that $C[0, 1]$ has a basis $\{f_1, f_2, \ldots, f_n\}$, i.e., for each $g \in C[0, 1]$, g is a linear *combination of f_i, i.e., $g(t) = \sum_{i=1}^{n} \alpha_i f_i(t)$ for some complex numbers α_i's. Picking up $N > n$ distinct t such as $t_1, t_2, ..., t_N$, we must have*

$$\begin{bmatrix} g(t_1) \\ g(t_2) \\ \vdots \\ g(t_N) \end{bmatrix} = \begin{bmatrix} f_1(t_1) & \cdots & f_n(t_1) \\ f_1(t_2) & \cdots & f_n(t_2) \\ \vdots & \ddots & \vdots \\ f_1(t_1) & \cdots & f_n(t_1) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix},$$

where the $N \times n$ matrix will be denoted as A. Regarding A as a linear transformation from C^n to C^N, the vector $[g(t_i)]$ must be in the range Im(A) of A for each $g \in C[0,1]$. Since $n < N$, the dimension of Im(A) is less than N and there is a vector $u \in C^N$ such that $u \notin$ Im(A). By Lagrange's interpolation formula, we can construct a polynomial function $p(t)$ such that the vector $[p(t_i)]$ is equal to $u = [\nu_i]$ as

$$p(t) = \sum_{i=1}^{N} \nu_i \prod_{j=1, j\neq i}^{N} \frac{t - t_j}{t_i - t_j}.$$

Since $P(t) \in C[0,1]$ and $[p(t_i)] \notin$ Im(A), we have a contradiction. □

Two vectors x, y in a complex inner product space H is called orthogonal to each other, denoted as $x \perp y$, if their inner product is zero. Let M be a vector subspace of H. The orthogonal complement $M^\perp$ of M is the set of all vectors x in H such that x is orthogonal to every vector y in M. It is clear that $M^\perp$ is also a vector subspace of H.

Since (x, x) is non-negative, we define the norm $||x||$ of a vector x to be the square root of (x, x), i.e. $||x||^2 = (x, x)$. The application of the following lemma is far-reaching.

Lemma 3.1.3 [Schwarz Inequality] *For x, y in a complex inner product space H,*

$$|(x, y)| \leq ||x||\ ||y||.$$

Proof. If $y = 0$, the inequality is trivially true. Noe suppose $y \neq 0$. Since (x, y) is in general a comlex number, we find a complex number α with $|\alpha| = 1$ such that $\bar{\alpha}(x, y) = |(x, y)|$. Since $\alpha(y, x) = \overline{\bar{\alpha}(x, y)}$, we have $\alpha(y, x) = |(x, y)|$. Now for each real r, we have

$$\begin{aligned} 0 \leq (x - r\alpha y, x - r\alpha y) &= ||x||^2 - r(\bar{\alpha}(x, y) + \alpha(y, x)) + r^2||y||^2 = ||x||^2 - 2|(x, y)|r + ||y||^2 r^2 \\ &= \left(||x||^2 - \frac{|(x, y)|^2}{||y||^2}\right) + \left(||y||r - \frac{|(x, y)|}{||y||}\right)^2 \end{aligned}$$

which implies that

$$0 \leq ||x||^2 - \frac{|(x, y)|^2}{||y||^2}, \text{i.e., } |(x, y)|^2 \leq ||x||^2\ ||y||^2.$$

This completes the proof. □

Example 3.1.4 *Let A be any set, countable or uncountable. Let $\ell^2(A)$ be the complex vector space of all complex-valued functions on A such that $\phi(\iota) = 0$ for all but countably many ι in A and $\sum_{\iota \in A} |\phi(\iota)|^2 < \infty$ [1] . The space $\ell^2(A)$ becomes to a complex inner product space when defining an inner product of ϕ and ρ as*

$$(\phi, \rho) = \sum_{\iota \in A} \phi(\iota)\overline{\rho(\iota)}.$$

It can be shown that the inner product is well-defined, i.e. $|\sum_{\iota \in A} \phi(\iota)\overline{\rho(\iota)}| < \infty$. □

Lemma 3.1.5 [Triangle Inequality] *For x, y in a complex inner product space H,*

$$||x + y|| \leq ||x|| + ||y||.$$

Proof. As an exercise. □

The norm $|| \cdot ||$ in a complex inner product space H has the following metric properties:

1. $||x - y|| \geq 0$ and $||x - y|| = 0$ if and only if $x = y$. (Non-negativity)
2. $||x - y|| = ||y - x||$. (Symmetry)
3. $||x - y|| \leq ||x - z|| + ||z - y||$. (Triangle Inequality)

Thus we may regard the quantity $||x - y||$ (sometimes denoted as $d(x, y)$) as a distance measure between two vectors x and y in H.

A sequence of vectors x_n, $n = 1, 2 \ldots$, in a complex inner product space H is said to converge in H if there exists a vector x in H such that $||x_n - x|| \to 0$ as $n \to \infty$, i.e., for all $\epsilon > 0$, there exists a positive integer $N(\epsilon)$, which depends on ϵ, such that $||x_n - x|| < \epsilon$ for all $n \geq N(\epsilon)$. It can be shown that x is unique. We call x as the limit of the convergent sequence $\{x_n\}$.

A sequence $\{x_n\}$ in H is called a Cauchy sequence if $||x_m - x_n|| \to 0$ as $m, n \to \infty$, i.e, for every $\epsilon > 0$, there exists a positive integer $N(\epsilon)$ such that $||x_m - x_n|| < \epsilon$ for all $n, m \geq N(\epsilon)$. It is clear that every convergent sequence in H is a Cauchy sequence. A complex inner product space H is called complete if every Cauchy sequence in H converges in H. A complete complex inner product space is called a Hilbert space. The space C^n in Example 3.1.1 is a Hilbert space, while the space $C[0, 1]$ in Example 3.1.2 is not. Consider the sequence of continuous functions on $[0, 1]$

$$f_n(t) = \begin{cases} 1 - 2nt, & t \in [0, 1/2n] \\ 0, & t \in [1/2n, 1 - 1/2n] \\ 2nt - (2n - 1), & t \in [1 - 1/2n, 1] \end{cases}$$

[1] Since $\phi(\iota) \neq 0$ only for ι in a countable subset A_0 of A, the sum $\sum_{\iota \in A} |\phi(\iota)|^2$ is (defined to be) the countable sum $\sum_{\iota \in A_0} |\phi(\iota)|^2$. The evaluation of such a countable sum of non-negative numbers is independent of the way of enumerating the countable set A_0, i.e. independent of the arrangement of elements in A_0 as a list $\iota_1, \iota_2, \ldots$. Please see W. Rudin, *Principles of Mathematical Analysis*, 3rd edn. New York: McGraw-Hill, 1976, page 78.

for $n = 1, 2, \ldots$. It can be seen that f_n is a Cauchy sequence in $C[0,1]$ but cannot converges in $C[0,1]$ since f_n would converges to the function f with $f(0) = f(1) = 1$ and $f(t) = 0$ for $0 < t < 1$ and f is obviously not a continuous function on $[0,1]$.

Example 3.1.6 *In this example, we will show that C is a Hilbert space. Suppose that $\{c_n\}$ is a Cauchy sequence in C. It is clear that $\{c_n\}$ is a bounded sequence, i.e., there exists a positive number s such that $|c_n| < s$ for all n. Let $c_n = a_n + ib_n$ where a_n and b_n are real and imaginary parts of c_n respectively. Thus both real sequences $\{a_n\}$ and $\{b_n\}$ are bounded and then $\limsup_n a_n = \inf_n\{\sup_{m\geq n} a_m\}$, $\liminf_n a_n = \sup_n\{\inf_{m\geq n} a_m\}$, $\limsup_n b_n = \inf_n\{\sup_{m\geq n} b_m\}$, and $\liminf_n b_n = \sup_n\{\inf_{m\geq n} b_m\}$ are all finite. Given an $\epsilon > 0$, there are a_i and a_j such that (a) $|a_i - a_j| < \epsilon/3$, (b) $\limsup_n a_n - \epsilon/3 < a_i < \limsup_n a_n + \epsilon/3$, (c) $\liminf_n a_n - \epsilon/3 < a_j < \liminf_n a_n + \epsilon/3$, where (a) is due to the facts that $|a_i - a_j| \leq |c_i - c_j|$ and that $\{c_n\}$ is a Cauchy sequence, (b) and (c) are from the properties of limsup and liminf. Now by triangle inequality, we have*

$$|\limsup_n a_n - \liminf_n a_n| \leq |\limsup_n a_n - a_i| + |a_i - a_j| + |a_j - \liminf_n a_n| < \epsilon.$$

By letting $\epsilon \to 0$, we have $\limsup_n a_n = \liminf_n a_n$ and then $\lim_{n\to\infty} a_n$ exists. Similarly $\lim_{n\to\infty} b_n$ exists. It is then clear that the sequence $\{c_n\}$ converges to $\lim_{n\to\infty} a_n + i\lim_{n\to\infty} b_n$.
□

Hereafter, H will be refered as a Hilbert space. Let S be a subset of H. We define the closure $\bar{S}$ of S as the subset of H which consists of the limits of all Cauchy sequences in S. It is clear that $\bar{S}$ contains S. And S is called closed in H if $S = \bar{S}$.

Lemma 3.1.7 *For a vector subspace M of H, its closure $\bar{M}$ is also a vector subspace of H.*

Proof. Let x, y be in $\bar{M}$. Then there are sequences $\{x_n\}$ and $\{y_n\}$ in M which converge to x and y respectively. Since M is a vector subspace of H, $\alpha x_n + \beta y_n$ is in M for $\alpha, \beta \in C$. The sequence $\{\alpha x_n + \beta y_n\}$ converges to $\alpha x + \beta y$ since It is clear that $\{\alpha x_n + \beta y_n\}$ is in M □

Theorem 3.1.8 *For a vector subspace M of H, we have $\bar{\bar{M}} = \bar{M}$, i.e., $\bar{M}$ is a complete complex inner product space.*

Proof. As an exercise. □

The above theorem shows that the closure of a vector subspace of H is closed and a closed subspace of H is a Hilbert space itself. In particular, every finite-dimensional vector subspace of H is closed.

Let H and K be two Hilbert spaces. A mapping f from H to K is said to be continuous at $x \in H$ if for every $\epsilon > 0$, there is a $\delta > 0$ such that $||f(y) - f(x)||_K < \epsilon$ whenever $||y - x||_H < \delta$.

Proposition 3.1.9 *A mapping f from H to K is continuous at $x \in H$ if and only if for every sequence $\{x_n\}$ which converges to x in H, the sequence $\{f(x_n)\}$ converges to $f(x)$ in K.*

Proof. As an exercise. □

The mapping f is called continuous over H if f is continuous at every $x \in H$. And f is called uniformly continuous over H if the selection of δ is independent of $x \in H$ in the ϵ-δ definition of continuity.

Lemma 3.1.10 *For any fix $z \in H$, the mappings*

$$x \mapsto (x, z) \text{ and } x \mapsto (z, x)$$

from H to C are uniformly continuous over H.

Proof. By the linearity of inner product and the Schwarz Inequality, we have

$$|(y, z) - (x, z)| = |(y - x, z)| \leq ||y - x|| \, ||z|| \text{ and} |(z, y) - (z, x)| = |(z, y - x)| \leq ||y - x|| \, ||z||$$

for all $x, y \in H$. It is clear that we can select $\delta = \epsilon / ||z||$ which is indepedent of $x \in H$. □

Let S be a subset of H. The orthogonal complement $S^{\perp}$ of S in H is the set of all vectors x in H such that $(x, y) = 0$ for all $y \in S$. It is easy to show that $S^{\perp}$ is a vector subspace of H. The following corollary is an application of the continuity of inner product.

Corollary 3.1.11 *The orthogonal complement $M^{\perp}$ of a vector subspace M in H is closed.*

Proof. Let $\{x_n\}$ be a Cauchy sequence in $M^{\perp}$. Since $\{x_n\}$ is also a Cauchy sequence in H, the sequence has a limit x in H. We will show that $x \in M^{\perp}$. Fix a y in M. Since $\{x_n\}$ converges to x, we have $\{(x_n, y)\}$ converges to (x, y) by the continuity of inner product. But $(x_n, y) = 0$ since $x_n \in M^{\perp}$ and $y \in M$, we have $(x, y) = \lim_n 0 = 0$. This completes the proof. □

Let S be a subset of H. We next investigate the existence and uniqueness of a unique x in S such that $||x|| = \min\{||y|| \mid y \in S\}$. We will show that closedness and convexity of S are sufficient conditions for such an x in S. S is called convex if for $x, y \in S$, $tx + (1 - t)y$ is in S for $0 \leq t \leq 1$. We need the following simple lemma.

Lemma 3.1.12 [Parallelogram Law] *For x, y in a complex inner product space, we have*

$$||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2.$$

Proof. As an exercise. □

Theorem 3.1.13 *For a non-empty, closed and convex subset S of H, there is a unique vector of the smallest norm in S.*

Proof. Let

$$\eta = \inf\{||y|| \mid y \in S\}$$

which must be finite since S is non-empty. Then there exists a sequence $\{y_n\}$ in S such that $||y_n||$ converges to η as $n \to \infty$. From the parallelogram law, we have

$$||y_n - y_m||^2 = 2||y_n||^2 + 2||y_m||^2 - 4||\frac{y_n + y_m}{2}||^2 \leq 2||y_n||^2 + 2||y_m||^2 - 4\eta^2$$

since $(y_n + y_m)/2$ is in S by the convexity of S. When n, m go to infinity, we have $||y_n||, ||y_m||$ go to η and then $||y_n - y_m||$ goes to zero. Thus $\{y_n\}$ is a Cauchy sequence in S and converges to an x in S by the closedness of S. Thus $||x|| \geq \eta$. But $||x|| \leq ||x - y_n|| + ||y_n||$ and by letting $n \to \infty$, we have $||x|| \leq \eta$. This proves the existence of an x in S with $||x|| = \eta$. To prove the uniqueness of x, we suppose that there exists x' in S such that $||x'|| = \eta$. Then again by the parallelogram law, we have

$$||x - x'||^2 = 2||x||^2 + 2||x'||^2 - 4||\frac{x + x'}{2}||^2 \leq 2\eta^2 + 2\eta^2 - 4\eta^2 = 0$$

which says that $x = x'$. □

We are now able to state a key theorem of orthogonal projection.

Theorem 3.1.14 *Let M be a closed vector subspace of H. Then for each x in H, there exists a unique decomposition*

$$x = y + z$$

where $y \in M$ and $z \in M^\perp$. And the above y and z have the following properties:

1. $||x - y|| = \min\{||x - u|| \mid u \in M\}$, *i.e., y is the nearest vectors in M to x.*
2. $||x - z|| = \min\{||x - v|| \mid v \in M^\perp\}$, *i.e., z is the nearest vectors in $M^\perp$ to x.*
3. [Pythogorean Identity] $||x||^2 = ||y||^2 + ||z||^2$.

Proof. Consider the subset $S = \{x - u | u \in M\}$ of H. Since the zero vector 0 is in M, S is a non-empty set. For $u, u' \in M$ and $t \in [0, 1]$, $(1 - t)u + tu'$ is in M since M is a vector space. Then

$$(1 - t)(x - u) + t(x - u') = x - ((1 - t)u + tu')$$

is in S which implies that S is convex. If $\{x - u_n\}$ is a Cauchy sequence in S, then $\{u_n\}$ is a Cauchy sequence in M since $||(x - u_m) - (x - u_n)|| = ||u_n - u_m||$. Since M is closed, the sequence $\{u_n\}$ converges to a limit u in M and then the sequence $\{x - u_n\}$ converges to a limit $x - u$ in S. Thus S is closed. By Theorem 3.1.13, S has a unique element $x - y$, $y \in M$, with

$$||x - y|| = \min\{||x - u|| \mid u \in M\}.$$

Let $z = x - y$. We will show that $z \in M^\perp$, i.e., $(z, u) = 0$ for all $u \in M$. If $u = 0$, then it is clear that $(z, u) = 0$. For each $u \neq 0$ in M and α in C, $z - \alpha\frac{u}{||u||}$ is in S and then $||z||^2 \leq ||z - \alpha\frac{u}{||u||}||^2$ which implies that

$$0 \leq -\overline{\alpha}(z, \frac{u}{||u||}) - \alpha\overline{(z, \frac{u}{||u||})} + |\alpha|^2.$$

By taking $\alpha = (z, \frac{u}{||u||})$, we have $0 \leq -|(z, \frac{u}{||u||})|^2$, i.e., $(z, u) = 0$. Now for each $v \in M^\perp$, we have

$$||x - v||^2 = ||(x - z) + (z - v)||^2 = ||x - z||^2 + ||z - v||^2$$

since $x - z = y \in M$ and $z - v \in M^\perp$. Thus we have $||x - z|| \leq ||x - v||$ for all $v \in M^\perp$, where the equality holds if and only if $v = z$. Thus we conclude that

$$||x - z|| = \min\{||x - v|| \mid v \in M^\perp\}.$$

If there exist $y' \in M$ and $z' \in M^\perp$ such that $x = y' + z'$, then we have $y - y' = z' - z$ which is in $M \cap M^\perp = \{0\}$, i.e., $y' = y$ and $z' = z$. This proves the uniqueness of the decomposition $x = y + z$. The Pythogorean identity follows since $y \perp z$. □

As in the above theorem, we let $y = P_M(x)$ and $z = P_{M^\perp}(x)$. The mappings P_M and $P_{M^\perp}$ are called the projections of H onto M and $M^\perp$ respectively. And y and z are called the projections of x in M and in $M^\perp$ respectively.

Corollary 3.1.15 *The mappings P_M and $P_{M^\perp}$ are linear transformations from H onto M and $M^\perp$ respectively.*

Proof. For x, y in H and α, β in C, we have

$$P_M(\alpha x + \beta y) + P_{M^\perp}(\alpha x + \beta y) = \alpha x + \beta y = \alpha\left(P_M(x) + P_{M^\perp}(x)\right) + \beta\left(P_M(y) + P_{M^\perp}(y)\right).$$

Thus we have

$$P_M(\alpha x + \beta y) - \alpha P_M(x) - \beta P_M(y) = \alpha P_{M^\perp}(x) + \beta P_{M^\perp}(y) - P_{M^\perp}(\alpha x + \beta y),$$

where the left-hand side is in M and the right-hand side in $M^\perp$. Since $M \cap M^\perp = \{0\}$, we have

$$P_M(\alpha x + \beta y) = \alpha P_M(x) + \beta P_M(y) \text{ and } P_{M^\perp}(\alpha x + \beta y) = \alpha P_{M^\perp}(x) + \beta P_{M^\perp}(y),$$

which completes the proof. □

A subset $\{u_\kappa | \kappa \in A\}$ of vectors in a Hilbert space H with index set A is called orthonormal if $(u_\kappa, u_\iota)) = 0$ for all $\kappa \neq \iota$ and $||u_\kappa|| = 1$ for all κ. Given an orthonormal set $\{u_\kappa | \kappa \in A\}$ in H, each $x \in H$ is associated with a function $\hat{x}$ on the index set A, defined as

$$\hat{x}(\kappa) = (x, u_\kappa), \ \forall\ \kappa \in A.$$

The complex numbers $\hat{x}(\kappa)$ are called Fourier coefficients of x relative to the orthonormal set $\{u_\kappa | \kappa \in A\}$. It is clear that the orthonormal set $\{u_\kappa | \kappa \in A\}$ is a linearly independent set in H. Let M_A be the subspace of H spanned by the set $\{u_\kappa | \kappa \in A\}$. We first consider the case that the index set A is finite.

Lemma 3.1.16 *Let $\{u_\kappa | \kappa \in A\}$ be a finite orthonormal set in a Hilbert space H.*

1. *For each $x \in H$, the projection $P_{M_A}(x)$ of x in the subspace M_A is $P_{M_A}(x) = \sum_{\kappa \in A} \hat{x}(\kappa) u_\kappa$.*
2. [Bessel Inequality] *For each $x \in H$, we have*
$$\sum_{\kappa \in A} |\hat{x}(\kappa)|^2 \leq ||x||^2.$$
3. *The mapping $f : x \mapsto \hat{x}$ from H to $\ell^2(A)$ is linear and onto. And the restriction of f to M_A is one-to-one. Furthermore, for $x, y \in M_A$, we have $(x, y) = (\hat{x}, \hat{y})$.*

Proof. As an exercise. □

A linear, one-to-one and onto mapping f from a Hilbert space H to a Hilbert space K which also preserves inner product, i.e., $(x, y) = (f(x), f(y))$ for all $x, y \in H$, is called a Hilbert space isomorphism. And the two Hilbert space H and K are called equivalent. The above lemma asserts that the mapping $f : x \mapsto \hat{x}$ from M_A to $\ell^2(A)$ is a Hilbert space isomorphism if A is a finite set. We will extend these results to a general A.

A subset S of a Hilbert space H is called dense in H if $H = \bar{S}$, i.e., H is the closure of S.

Example 3.1.17 *Let F be a subset of $\ell^2(A)$ which consists of all functions ϕ on A such that $\phi(\kappa) \neq 0$ only for a finitely many $\kappa \in A$. We now show that F is a dense set of $\ell^2(A)$. This is equivalent to show that each ϕ in $\ell^2(A)$ but not in F is the limit of a Cauchy sequence in F. Such a ϕ takes non-zero values only in a countably infinite subset $\{\kappa_1, \kappa_2, \ldots\}$ of A. For each positive integer n, we define ϕ_n to be the function on A such that $\phi_n(\kappa_i) = \phi(\kappa_i)$ for $1 \leq i \leq n$ and $\phi_n(\kappa) = 0$ otherwise. It is clear that ϕ_n is in F for each n. And we have*
$$||\phi_n - \phi||^2 = \sum_{i=n+1}^{\infty} |\phi(\kappa_i)|^2$$
which converges to zero as $n \to \infty$ since $\sum_{i=1}^{\infty} |\phi(\kappa_i)|^2 = ||\phi||^2$ is finite. This concludes that ϕ is the limit of the Cauchy sequence $\{\phi_n\}$ in F. □

A mapping f from a subset S of a Hilbert space H to a Hilbert space K is called an isometry if $||f(x) - f(y)||_K = ||x - y||_H$ for all $x, y \in S$, i.e., an isometry preserves the distance between vectors. It is clear that an isometry is one-to-one and uniformly continuous. If f is a linear transformation from a vector subspace S of a Hilbert space H to a Hilbert space K and preserves the norm, i.e., $||x||_H = ||f(x)||_K$ for all $x \in S$, then f is an isometry from S to K since $||x - y||_H = ||f(x - y)||_K = ||f(x) - f(y)||_K$. Conversely, if f is a linear isometry from S to K, then f preserves the norm since $f(0) = 0$ and
$$||x||_H = ||x - 0||_H = ||f(x) - f(0)||_K = ||f(x) - 0||_K = ||f(x)||_K.$$

The following extension theorem is useful.

Theorem 3.1.18 *Let S be a dense subset of a Hilbert spaces H. Let f be an isometry from S to a Hilbert space K with the range $f(S)$ of S a dense set of K. Then there is a unique isometry g from H onto K such that $f = g|_S$, the restriction of g on S. If S is a vector subspace of H and f is linear, then g is also linear.*

Proof. For an x in H, there exists a Cauchy sequence $\{x_n\}$ in S which converges to x, since S is dense in H. Since f is an isometry and preserves the distance between vectors, $\{f(x_n)\}$ is a Cauchy sequence in K and then converges to a z in K. This z is independent of the choice of $\{x_n\}$. For if $\{x'_n\}$ is another Cauchy sequence in S which converges to x, then the combined sequence

$$x_1, x'_1, x_2, x'_2, \ldots$$

converges to x and then is a Cauchy sequence in S. This implies that the sequence

$$f(x_1), f(x'_1), f(x_2), f(x'_2), \ldots$$

is a Cauchy sequence in K and this sequence and all its subsequences converge to the same limit in K. In particular, subsequences $\{f(x_n)\}$ and $\{f(x'_n)\}$ converge to the same limit which must be z. We now define $g(x) = z$ and g is a well-defined function from H to K. And for an $x \in S$, the sequence $\{x_n\}$ in S with $x_n = x$ converges to x and then $g(x) = \lim_n f(x_n) = \lim_n f(x) = f(x)$. Thus we have $f = g|_S$, the restriction of g on S. We next show that g is an isometry. Consider x, y in H and two Cauchy sequences $\{x_n\}$ and $\{y_n\}$ in S which converge to x and y respectively. By the definition of g, $\{f(x_n)\}$ and $\{f(y_n)\}$ converge to $g(x)$ and $g(y)$ in K respectively. Then the sequence $\{x_n - y_n\}$ converges to $x - y$ and the sequence $\{f(x_n) - f(y_n)\}$ converges to $g(x) - g(y)$. By the continuity of the norm $||\cdot||$ (as an exercise), we have

$$||x_n - y_n||_H \to ||x - y||_H \text{ and } ||f(x_n) - f(y_n)||_K \to ||g(x) - g(y)||_K$$

as $n \to \infty$. But $||f(x_n) - f(y_n)||_K = ||x_n - y_n||_H$, we have $||g(x) - g(y)||_K = ||x - y||_H$. Thus g is an isometry from H to K. Suppose that there is another isometry g' from H to K such that $f = g'|_S$. For an $x \in H$ and a Cauchy sequence $\{x_n\}$ in S which converges to x, we have

$$g(x_n) \to g(x), \quad g'(x_n) \to g'(x) \quad \text{as} \quad n \to \infty$$

by the continuity of g and g' on H. Since $g(x_n) = f(x_n) = g'(x_n)$, we have $g(x) = g'(x)$. This shows the uniquness of g. To show that g is onto, we consider an arbitrary $z \in K$. Since $f(S)$ is dense in K, there is a Cauchy sequence $\{z_n\}$ in $f(S)$ which converges to z. For each n, there exists an x_n in S such that $z_n = f(x_n)$. Since f is an isometry on S, the sequence $\{x_n\}$ is also a Cauchy sequence and converges to a limit, says x, in H. By the definition of g, we have $g(x) = \lim_n f(x_n) = \lim_n z_n = z$ and g is onto. Assume that S is a subspace of H and f is linear on S. Consider $x, y \in H$, $\alpha, \beta \in C$ and two Cauchy sequences $\{x_n\}$ and $\{y_n\}$ in S which converge to x and y respectively. Then $\{\alpha x_n + \beta y_n\}$ in S converges to $\alpha x + \beta y$ and then

$$g(\alpha x + \beta y) = \lim_n f(\alpha x_n + \beta y_n) = \alpha \lim_n f(x_n) + \beta \lim_n f(y_n) = \alpha g(x) + \beta g(y),$$

which shows that g is linear. □

Corollary 3.1.19 *Let f be a continuous mapping from a Hilbert space H to a Hilbert space K. Let S be a dense subset of H. If*

1. *f is an isometry when restricted to S, i.e., $||f(x) - f(y)||_K = ||x - y||_H$ for all $x, y \in S$,*
2. *$f(S)$ is a dense subset of K,*

then f is an isometry from H onto K.

Proof. Since the restriction $f|_S$ of f to S is an isometry from S to K and $f(S)$ is a dense set of K, there is a unique isometry g from H onto K such that $g|_S = f|_S$ by Theorem 3.1.18. The proof will be completed if $f(x) = g(x)$ for all $x \in H$. Let x be in H. Then there exists a Cauchy sequence $\{x_n\}$ in S which converges to x. By continuity of f and g, we have $f(x_n) \to f(x)$ and $g(x_n) \to g(x)$ as $n \to \infty$. Since $f(x_n) = g(x_n)$ for all n, we must have $f(x) = g(x)$. □

The following lemma will be needed which expresses the inner product in terms of norms.

Lemma 3.1.20 [Polarization Identity] *For x, y in a complex inner product space, we have*

$$4(x, y) = ||x + y||^2 - ||x - y||^2 + i||x + iy||^2 - i||x - iy||^2.$$

Proof. As an exercise. □

Here is an application of the polarization identity.

Corollary 3.1.21 *A mapping g is a linear isometry from a Hilbert space H onto a Hilbert space K if and only if g is a Hilbert space isomorphism from H to K.*

Proof. As an exercise. □

Theorem 3.1.22 *Let $\{u_\kappa | \kappa \in A\}$ be an orthonormal set in a Hilbert space H. Then*

1. *$\hat{x}$ is in $\ell^2(A)$ for all x in H,*
2. [Bessel Inequality] $\left(||\hat{x}||^2_{\ell^2(A)} =\right) \sum_{\kappa \in A} |\hat{x}(\kappa)|^2 \le ||x||^2_H$ *for all x in H,*
3. *the mapping $f : x \mapsto \hat{x}$ is a Hilbert space isomorphism from the closure $\overline{M_A}$ of M_A to $\ell^2(A)$.*

Proof. We first show that $\hat{x}$ is in $\ell^2(A)$ for each x in H. For an $\epsilon > 0$, let A_ϵ be the subset of κ in the index set A such that $|\hat{x}(\kappa)| > \epsilon$. We claim that A_ϵ is a finite set. If not, there exists a countably infinite subset $\{\kappa_1, \kappa_2, \ldots\}$ of A_ϵ. By the Bessel inequality in Lemma 3.1.16, we have

$$\sum_{i=1}^{n} |\hat{x}(\kappa_i)|^2 \leq ||x||^2$$

for all n. Since $\sum_{i=1}^{n} |\hat{x}(\kappa_i)|^2 > n\epsilon^2$, the sum $\sum_{i=1}^{n} |\hat{x}(\kappa_i)|^2$ is greater than $||x||^2$ for sufficient large n, a contradiction. Now, let A_0 be the subset of κ in A such that $|\hat{x}(\kappa)| > 0$. It is clear that A_0 is the union $\cup_{m=1}^{\infty} A_{1/m}$ of the sets $A_{1/m}$, $m = 1, 2, \ldots$. Since each $A_{1/m}$ is finite, the set A_0 is a countable set, says $\{\iota_1, \iota_2, \ldots\}$. Again by the Bessel inequality in Lemma 3.1.16, we have

$$\sum_{i=1}^{k} |\hat{x}(\iota_i)|^2 \leq ||x||^2$$

for all k. By letting $k \to \infty$, we have

$$\sum_{\kappa \in A} |\hat{x}(\kappa)|^2 = \sum_{i=1}^{\infty} |\hat{x}(\iota_i)|^2 \leq ||x||^2.$$

We conclude that $\hat{x}$ is in $\ell^2(A)$ and $||\hat{x}||^2_{\ell^2(A)} \leq ||x||^2_H$. Let f be the mapping $x \mapsto \hat{x}$ from H to $\ell^2(A)$. Since

$$(\widehat{\alpha x + \beta y})(\kappa) = (\alpha x + \beta y, u_\kappa) = \alpha(x, u_\kappa) + \beta(y, u_\kappa) = \alpha\hat{x}(\kappa) + \beta\hat{y}(\kappa),$$

we have $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$ and f is a linear transformation from H to $\ell^2(A)$. Also since

$$||\hat{x} - \hat{y}||^2_{\ell^2(A)} = ||\widehat{x - y}||^2_{\ell^2(A)} \leq ||x - y||_H,$$

the mapping $f : x \mapsto \hat{x}$ is continuous from H to $\ell^2(A)$. In particular, f is a continuous linear transformation from the closure $\overline{M_A}$ of M_A to $\ell^2(A)$. Note that M_A is dense in its closure $\overline{M_A}$. An x in M_A is a linear combination $\sum_{i=1}^{n} \alpha_i u_{\kappa_i}$ of finitely many u_{κ_i} in the orthonormal set, where α_i are complex numbers. Thus $\hat{x}(\kappa) = \alpha_i$ if $\kappa = \kappa_i$, $i = 1, 2, \ldots, n$ and $\hat{x}(\kappa) = 0$, otherwise. This says that $\hat{x}$ is in the dense subset F of $\ell^2(A)$ in Example 3.1.17 for all $x \in M_A$. Conversely, for a ϕ in F with $\phi(\kappa) \neq 0$ only for $\kappa = \kappa_i$, $i = 1, 2, \ldots, n$ for some n, the element $x = \sum_{i=1}^{n} \phi(\kappa_i) u_{\kappa_i}$ is in M_A and has $\hat{x} = \phi$. Thus $f(M_A)$ is equal to F and is dense in $\ell^2(A)$. Furthermore for each $x \in M_A$, we have

$$||x||^2_H = \sum_{i=1}^{n} |\alpha_i|^2 = \sum_{i=1}^{n} |\hat{x}(\kappa_i)|^2 = ||\hat{x}||_{\ell^2(A)},$$

which implies that

$$||\hat{x} - \hat{y}||_{\ell^2(A)} = ||\widehat{x - y}||_{\ell^2(A)} = ||x - y||_H$$

for $x, y \in M_A$, i.e., f is an isometry when restricted to M_A. By Corollary 3.1.19, f is an isometry from $\overline{M_A}$ onto $\ell^2(A)$. By Corollary 3.1.21, f is a Hilbert space isomorphism from $\overline{M_A}$ to $\ell^2(A)$. □

An orthonormal set in a Hilbert space H is called maximal if there is no other orthonormal set in H which contains this set properly. We will state the following fundamental theorem without giving any proof [2].

Theorem 3.1.23 *Every orthonormal set in a Hilbert space H is contained in a maximal orthonormal set in H.*

The importance of a maximal orthonormal set in a Hilbert space can be seen from the following theorem.

Theorem 3.1.24 *Let $\{u_\kappa | \kappa \in A\}$ be an orthonormal set in a Hilbert space H. The following statements are equivalent:*

1. *$\{u_\kappa | \kappa \in A\}$ is a maximal orthonormal set in H.*
2. *M_A is a dense set in H.*
3. *The mapping $f : x \mapsto \hat{x}$ is a Hilbert space isomorphism from H to $\ell^2(A)$.*

Proof. ("1⇒2") If M_A is not dense in H, then there exists an x in H but not in $\overline{M_A}$. Consider the decomposition $x = y + z$ of x with $y \in \overline{M_A}$ and $z \in \overline{M_A}^\perp$ as promised in Theorem 3.1.14. Then we have $z \neq 0$ and $(z, u_\kappa) = 0$ for all $\kappa \in A$. Thus we can extend the orthonormal set $\{u_\kappa | \kappa \in A\}$ by adding $z/||z||$ to it, which is a contradiction to the maximality of the orthonormal set $\{u_\kappa | \kappa \in A\}$. ("2⇒3") This is just a consequence of Theorem 3.1.22. ("3⇒1") Suppose that $\{u_\kappa | \kappa \in A\}$ is not a maximal orthonormal set in H. Then there is a non-zero x in H such that $(x, u_\kappa) = 0$ for all $\kappa \in A$, i.e. $\hat{x}$ is the zero vector in $\ell^2(A)$. But

$$0 < ||x||_H = ||\hat{x}||_{\ell^2(A)} = ||0||_{\ell^2(A)} = 0,$$

which is a contradiction, where the first equality is due to the fact that $f : x \mapsto \hat{x}$ is a Hilbert space isomorphism. □

Let $\{u_\kappa | \kappa \in A\}$ be a maximal orthonormal set in a Hilbert space H. For an x in H, the function $\hat{x}$ in $\ell^2(A)$ takes non-zero values only over a countably infinite subset A_0 of A. With a specific enumeration $\kappa_1, \kappa_2, \ldots$ of A_0, we define ϕ_n in $\ell^2(A)$ to be

$$\phi_n(\kappa) = \begin{cases} \hat{x}(\kappa_i), & \text{if } \kappa = \kappa_i, 1 \leq i \leq n, \\ 0, & \text{otherwise,} \end{cases}$$

for each $n = 1, 2, \ldots$. It is clear that the sequence $\{\phi_n\}$ converges to $\hat{x}$ in $\ell^2(A)$. Since $\sum_{i=1}^n \hat{x}(\kappa_i) u_{\kappa_i}$ corresponds to ϕ_n under the Hilbert space isomorphism $f : x \mapsto \hat{x}$, we have $\sum_{i=1}^n \hat{x}(\kappa_i) u_{\kappa_i} \to x$ as $n \to \infty$. If we rearrange the enumeration list $\kappa_1, \kappa_2, \ldots$ of A_0 to obtain another enumeration list $\iota_1, \iota_2, \ldots$ of A_0, we also have $\sum_{i=1}^n \hat{x}(\iota_i) u_{\iota_i} \to x$ as $n \to \infty$. Thus we

[2] For a proof, please see W. Rudin, *Real and Complex Analysis*, 3rd edn. New York: McGraw-Hill, 1987, page 87.

can denote x as the countable sum $\sum_{\kappa\in A_0} \hat{x}(\kappa)u_\kappa$ to illustrate that the evaluation of this sum is independent of the enumeration of the countable set A_0. Since $\hat{x}(\kappa) = 0$ for all $\kappa \notin A_0$, we finally represent x as

$$x = \sum_{\kappa\in A} \hat{x}(\kappa)u_\kappa. \tag{3.1}$$

3.2 The Space $L^2(\Omega, \mathcal{F}, \mathcal{P})$

Consider a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with sample space Ω, σ-algebra $\mathcal{F}$, and probability measure $\mathcal{P}$. Let $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$ be the set of all complex-valued random variables X on the probability space $(\Omega, \mathcal{F}, \mathcal{P})$ with second moments, i.e., $\mathcal{E}(|X|^2) < \infty$.

Lemma 3.2.1 [Schartz Inequality for Random Variables] *If X, Y are in $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$, then*

$$\mathcal{E}(|XY|) \leq \mathcal{E}(|X|^2)^{1/2}\mathcal{E}(|Y|^2)^{1/2}.$$

Proof. Let $A = \mathcal{E}(|X|^2)^{1/2}$ and $B = \mathcal{E}(|Y|^2)^{1/2}$. Since the geometric average of two non-negative numbers is no more than their arithmetic average, we have

$$\sqrt{\frac{|X(\omega)|^2}{A^2}\frac{|Y(\omega)|^2}{B^2}} \leq \frac{\frac{|X(\omega)|^2}{A^2} + \frac{|Y(\omega)|^2}{B^2}}{2}$$

for all $\omega \in \Omega$. By taking expectation, we have

$$\frac{\mathcal{E}(|XY|)}{AB} \leq \frac{\frac{\mathcal{E}(|X|^2)}{A^2} + \frac{\mathcal{E}(|Y|^2}{B^2}}{2} = 1,$$

which completes the proof. □

Let X, Y be in $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$. By triangle inequality for complex numbers, we have $|X(\omega) + Y(\omega)|^2 \leq (|X(\omega)| + |Y(\omega)|)^2$ for all $\omega \in \Omega$, and then

$$\mathcal{E}(|X + Y|^2) \leq \mathcal{E}(|X|^2) + 2\mathcal{E}(|XY|) + \mathcal{E}(|Y|^2) < \infty$$

by Lemma 3.2.1. And for a complex number α, we have

$$\mathcal{E}(|\alpha X|^2) = |\alpha|^2\mathcal{E}(|X|^2) < \infty.$$

Thus $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$ is a vector space over C. We define

$$(X, Y) = \mathcal{E}(X\overline{Y})$$

which satisfies the three properties of an inner product except that $(X,X) = \mathcal{E}(|X|^2) = 0$ may not imply that $X = 0$. Consider such an X with $\mathcal{E}(|X|^2) = 0$. Since for each $\epsilon > 0$,

$$0 = \int_\Omega |X(\omega)|^2 \mathcal{P}(d\omega) \geq \int_{(|X|\geq\epsilon)} |X(\omega)|^2 \mathcal{P}(d\omega) \geq \epsilon^2 \int_{(|X|\geq\epsilon)} \mathcal{P}(d\omega) = \epsilon^2 \mathcal{P}((|X| \geq \epsilon)),$$

we have $\mathcal{P}(|X| \geq \epsilon) = 0$. Since the event $(X \neq 0)$ is the union $\cup_{m=1}^\infty (|X| \geq 1/m)$ of the events $(|X| \geq 1/m)$, we have

$$\mathcal{P}(X \neq 0) \leq \sum_{m=1}^\infty \mathcal{P}(|X| \geq 1/m) = \sum_{m=1}^\infty 0 = 0$$

which says that X is equal to 0 with probability one (abbreviated as w.p.1). Conversely, if a r.v. X is equal to 0 w.p.1, then we have $\mathcal{E}(|X|^2) = 0$. Thus, the ambiguity between a r.v. X with $(X,X) = 0$ and the zero random variable 0 is that X is equal to 0 w.p.1. Two r.v.s X and Y are said to be equivalent, denoted as $X \equiv Y$, if $X = Y$ w.p.1, i.e., $X - Y = 0$ w.p.1. It is clear that (a) $X \equiv X$ (b) $X \equiv Y \Rightarrow Y \equiv X$ (c) $X \equiv Y$ and $Y \equiv Z \Rightarrow X \equiv Z$. Thus $\equiv$ is an equivalent relation among random variables and partitions the vector space $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$ into equivalent classes. We will denote $[X]$ as the equivalent class which the r.v. X belongs to. And we call X as a representative of the class $[X]$. Let $L^2(\Omega, \mathcal{F}, \mathcal{P})$ be the set of all equivalent classes in $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$. We next show that $L^2(\Omega, \mathcal{F}, \mathcal{P})$ is a complex vector space by defining

$$[X] + [Y] \equiv [X + Y] \text{ and } \alpha[X] \equiv [\alpha X] \tag{3.2}$$

for all classes $[X]$ and $[Y]$ and for all complex number α. The operations in (3.2) is well defined as follows. Assume that $[X] = [X']$ and $[Y] = [Y']$, i.e., there exist $\Omega_1, \Omega_2 \in \mathcal{F}$ with $\mathcal{P}(\Omega_1) = \mathcal{P}(\Omega_2) = 1$ such that

$$X(\omega) = X'(\omega) \ \forall\ \omega \in \Omega_1,\ Y(\omega) = Y'(\omega) \ \forall\ \omega \in \Omega_2.$$

Now for $\omega \in \Omega_1 \cap \Omega_2$, we have

$$X(\omega) + Y(\omega) = X'(\omega) + Y'(\omega).$$

Since $\mathcal{P}(\Omega_1 \cap \Omega_2) = 1$, we have $X + Y = X' + Y'$ w.p.1, i.e., $[X + Y] = [X' + Y']$. Similarly, for $\omega \in \Omega_1$, we have

$$\alpha X(\omega) = \alpha X'(\omega)$$

for all complex α and then $[\alpha X] = [\alpha X']$. Furthermore, an inner product can be defined in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ as

$$([X], [Y]) \equiv \mathcal{E}(X\overline{Y})$$

which is clearly well-defined as in above and satisfies the three properties of an inner product in the previous section. Thus $L^2(\Omega, \mathcal{F}, \mathcal{P})$ is a complex inner product space. Hereafter, we will identify an equivalent class in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ with any of its representative random variables in $\tilde{L}^2(\Omega, \mathcal{F}, \mathcal{P})$. In particular, we will say that $X = 0$ in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ to mean that $[X] = [0]$. We now show that the space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ is a Hilbert space.

Let $\{X_n\}$ be a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathcal{P})$. Then there exist $0 < n_1 < n_2 < \cdots < n_i < \cdots$ such that

$$||X_n - X_{n_i}|| < 2^{-i}, \ \forall\, n \geq n_i,$$

for each $i \geq 1$. In particular, we have

$$||X_{n_{i+1}} - X_{n_i}|| < 2^{-i}, \ \forall\, i \geq 1.$$

Now we define $Y_k = \sum_{i=1}^{k} |X_{n_{i+1}} - X_{n_i}|$ for all $k \geq 1$. It is clear that $\{Y_k\}$ is a monotone increasing sequence of non-negative r.v.'s and converges pointwise to the r.v.

$$Y = \sum_{i=1}^{\infty} |X_{n_{i+1}} - X_{n_i}|.$$

By the Lebesgue's monotone convergence theorem (Theorem 1.4.5), we have

$$\mathcal{E}(Y^2) = \mathcal{E}(\lim_{n\to\infty} Y_k^2) = \lim_{n\to\infty} \mathcal{E}(Y_k^2). \tag{3.3}$$

And by the triangle inequality of the norm $||\cdot||$, we have

$$||Y_k|| \leq \sum_{i=1}^{k} ||X_{n_{i+1}} - X_{n_i}|| < \sum_{i=1}^{k} 2^{-i} < 1$$

for all $k \geq 1$ and then $\mathcal{E}(Y_k^2) = ||Y_k||^2 < 1$ which implies that $\mathcal{E}(Y^2) < 1$ by (3.3)and then

$$Y = \sum_{i=1}^{\infty} |X_{n_{i+1}} - X_{n_i}| < +\infty \ \text{ w.p.1.}$$

Thus the series $X_{n_1} + \sum_{i=1}^{\infty}(X_{n_{i+1}} - X_{n_i})$ converges absolutely w.p.1. Let $X(\omega)$ be the limit of the series $X_{n_1}(\omega) + \sum_{i=1}^{\infty}(X_{n_{i+1}}(\omega) - X_{n_i}(\omega))$ when the series converges absolutely. Otherwise, we let $X(\omega) = 0$. Note that the partial sum of the series $X_{n_1} + \sum_{i=1}^{\infty}(X_{n_{i+1}} - X_{n_i})$ is

$$X_{n_{k+1}} = X_{n_1} + \sum_{i=1}^{k}(X_{n_{i+1}} - X_{n_i})$$

which says that the sequence $\{X_{n_i}\}$ of r.v.'s converges pointwise to the r.v. X w.p.1. We next show that X is the limit of the Cauchy sequence $\{X_n\}$ in L^2-norm, i.e., $X \in L^2(\Omega, \mathcal{F}, \mathcal{P})$ and $||X_n - X|| \to 0$ as $n \to \infty$. Since $\{X_n\}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, for $\epsilon > 0$, there exists $N(\epsilon)$ such that $||X_n - X_m|| < \epsilon$ for all $n, m \geq N(\epsilon)$. In particular, we have $||X_{n_i} - X_m|| < \epsilon$ for all $n_i, m \geq N(\epsilon)$. Since

$$X - X_m = \lim_{i\to\infty}(X_{n_i} - X_m) \text{ w.p.1,}$$

we have by Fatou's Lemma (Theorem 2.5.4),

$$\mathcal{E}(|X - X_m|^2) = \mathcal{E}(\liminf_i |X_{n_i} - X_m|^2) \leq \liminf_i \mathcal{E}(|X_{n_i} - X_m|^2) = \liminf_i ||X_{n_i} - X_m||^2 < \epsilon^2, \tag{3.4}$$

for all $m \in N(\epsilon)$. Since $|X| = |(X - X_m) + X_m| \leq |X - X_m| + |X_m|$, we have

$$||X|| = \mathcal{E}(|X|^2)^{1/2} \leq \mathcal{E}((|X-X_m|+|X_m|)^2)^{1/2} = |||X-X_m|+|X_m||| \leq ||X-X_m||+||X_m|| < +\infty$$

by the triangle inequality of the L^2-norm $||\cdot||$. Thus X is in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ and from (3.4), X is the limit of the Cauchy sequence $\{X_n\}$ in the complex inner product space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ with L^2-norm. This completes the proof of the following theorem.

Theorem 3.2.2 *The complex inner product space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ is a Hilbert space.*

A useful side result in the above proof will be stated in the following theorem.

Theorem 3.2.3 *Every Cauchy sequence in the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$ has a subsequence which converges pointwise to a r.v. in $L^2(\Omega, \mathcal{F}, \mathcal{P})$.*

The following theorem provides a dense subset of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$.

Theorem 3.2.4 *The set S of all complex simple r.v.'s is a dense subset of $L^2(\Omega, \mathcal{F}, \mathcal{P})$.*

proof. It is clear that every complex simple r.v. has 2nd-moment and thus S is a subset of $L^2(\Omega, \mathcal{F}, \mathcal{P})$. Consider a non-negative r.v. X in $L^2(\Omega, \mathcal{F}, \mathcal{P})$. As stated in Lemma 1.4.6, there is a monotone increasing sequence $\{X_n\}$ of non-negative simple r.v.'s

$$0 \leq X_1 \leq X_2 \leq \cdots \leq X$$

which converges to X pointwise. Since $X - X_n \leq X$ for all $n \geq 1$ and $\mathcal{E}(|X|^2) < \infty$, we have

$$\lim_{n\to\infty} ||X - X_n||^2 = \lim_{n\to\infty} \mathcal{E}(|X - X_n|^2) = \mathcal{E}(\lim_{n\to\infty} |X - X_n|^2) = \mathcal{E}(0) = 0$$

by Lebesgue's dominated convergence theorem (Theorem 2.5.5) which shows that $\{X_n\}$ is a Cauchy sequence in S and converges to X in L^2-norm. Since a complex r.v. X in $L^2(\Omega, \mathcal{F}, \mathcal{P})$ is a linear combination of non-negative r.v.'s in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, X is the L^2-limit of a Cauchy sequence in S by a similar argument as in above. □

3.3 Differential Calculus on $L^2(\Omega, \mathcal{F}, \mathcal{P})$

A 2nd-order process is a mapping $X(t)$ from a subset T of the real line $\mathcal{R}$ into the hilbert spapce $L^2(\Omega, \mathcal{F}, \mathcal{P})$. Thus a 2nd-order process is a "curve" in the Hilbert space $L^2(\Omega, \mathcal{F}, \mathcal{P})$, just like a curve in the complex Euclidean space $\mathcal{C}^n$. In this section , we will assume that T is an interval of $\mathcal{R}$ such as $T = \mathcal{R}$, $T = [0, \infty)$ or $T = [a, b]$.

A 2nd-order process $X(t)$ over T is said to be differentiable at an interior point t_0 of T (i.e., $(t_0 - \epsilon, t_0 + \epsilon) \subseteq T$ for some $\epsilon > 0$) if

$$X'(t_0) = \lim_{h\to 0} \frac{X(t_0 + h) - X(t_0)}{h} \text{ in } L^2\text{-norm}$$

exists in $L^2(\Omega, \mathcal{F}, \mathcal{P})$m, which means that for any sequence $\{h_n\}$ in R with $h_n \neq 0$ and $\lim_{n\to\infty} h_n = 0$, the sequence $\{\frac{X(t_0+h_n)-X(t_0)}{h_n}\}$ converges in L^2-norm to an r.v. $X'(t_0)$ in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, independent of the sequence $\{h_n\}$ chosen.

Recall that $X(t)$ is continuous at a point t_0 in T if $\lim_{n\to\infty} X(t_n) = X(t_0)$ in L^2-norm for any sequence $\{t_n\}$ in T with $\lim_{n\to\infty} t_n = t_0$.

Theorem 3.3.1 *If $X(t)$ is differentiable at an interior point t_0 of T, then $X(t)$ is continuous at t_0.*

Proof. As an exercise. □

The covariance function of a 2nd-order process $X(t)$ over T is defined as

$$K(s,t) \triangleq \mathcal{E}(X(s)\overline{X(t)}) = (X(s), X(t))$$

which is a complex-valued function on $T \times T$. We will show that the properties of a 2nd-order process is closely related to the properties of its covariance function. We first give a result on the continuity of inner product.

Lemma 3.3.2 *If $\lim_{m\to\infty} x_m = x$ and $\lim_{n\to\infty} y_n = y$ in a complex inner product space H, then we have*

$$\lim_{m,n\to\infty} (x_m, y_n)_H \ = (x,y)_H = (\lim_{m\to\infty} x_m, \lim_{n\to\infty} y_n)_H.$$

Proof. Since $\lim_{m\to\infty} x_m = x$ and $\lim_{n\to\infty} y_n = y$, for $\epsilon > 0$, there exists a positive integer $N(\epsilon)$ such that

$$\|x_m - x\|_H < \epsilon, \quad \|y_n - y\|_H < \epsilon$$

for all $m, n \geq N(\epsilon)$. Now for $\delta > 0$, we select an $\epsilon > 0$ such that $\epsilon^2 + (\|x\|_H + \|y\|_H)\epsilon < \delta$ and then we have

$$\begin{aligned}
|(x_m, y_n)_H - (x,y)_H| &= |(x_m, y_n)_H - (x, y_n)_H + (x, y_n)_H - (x,y)_H| \\
&\leq |(x_m - x, y_n)_H| + |(x, y_n - y)_H| \\
&\leq \|x_m - x\|_H \|y_n\|_H + \|x\|_H \|y_n - y\|_H \\
&\leq \|x_m - x\|_H (\|y_n - y\|_H + \|y\|_H) + \|x\|_H \|y_n - y\|_H \\
&< \epsilon^2 + (\|x\|_H + \|y\|_H)\epsilon < \delta
\end{aligned}$$

for all $m, n \geq N(\epsilon) \equiv N(\delta)$. This proves $\lim_{m,n\to\infty} (x_m, y_n)_H \ = (x,y)_H$. □

Theorem 3.3.3 *Let $X(t)$ be a 2nd-order process over T and $K(s,t)$ be the covariance function of $X(t)$. Let t_0 be an interior point of T.*

1. *$\lim_{t\to t_0} X(t)$ exists in L^2-norm if and only if $\lim_{s,t\to t_0} K(s,t)$ exists.*
2. *$X(t)$ is continuous at t_0 if and only if $K(s,t)$ is continuous at (t_0, t_0).*

Proof. Assume that $\lim_{t\to t_0} X(t) = Z$ in L^2-norm for some Z in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, i.e., for any sequence $\{t_n\}$ in T with $t_n \neq t_0$ and $\lim_{n\to\infty} t_n = t_0$, we have

$$\lim_{n\to\infty} X(t_n) = Z \text{ in } L^2\text{-norm.} \tag{3.5}$$

Now consider two sequences $\{s_m\}$, $\{t_n\}$ in T with $s_m \neq t_0$, $t_n \neq t_0$ and $\lim_{m\to\infty} s_m = t_0$, $\lim_{n\to\infty} t_n = t_0$. By (3.5), we have

$$\lim_{m\to\infty} X(s_m) = Z, \ \lim_{n\to\infty} X(t_n) = Z \text{ in } L^2\text{-norm}$$

and then by the continuity of inner product, we have

$$\lim_{m,n\to\infty} K(s_m, t_n) = \lim_{m,n\to\infty} (X(s_m), X(t_n)) = (Z, Z),$$

which proves that $\lim_{s,t\to t_0} K(s,t) = (Z, Z)$. Furthermore, if $X(t)$ is continuous at t_0, then for any sequence $\{(s_n, t_n)\}$ in $T \times T$ with $\lim_{n\to\infty}(s_n, t_n) = (t_0, t_0)$, we have $\lim_{n\to\infty} s_n = t_0$ and $\lim_{n\to\infty} t_n = t_0$. Then $\lim_{n\to\infty} X(s_n) = X(t_0)$ and $\lim_{n\to\infty} X(t_n) = X(t_0)$. Again by the continuity of inner product, we have

$$\lim_{n\to\infty} K(s_n, t_n) = \lim_{n\to\infty} (X(s_n), X(t_n)) = (\lim_{n\to\infty} X(s_n), \lim_{n\to\infty} X(t_n)) = (X(t_0), X(t_0)) = K(t_0, t_0)$$

which proves that $K(s,t)$ is continuous at (t_0, t_0). Conversely, assume that $\lim_{s,t\to t_0} K(s,t) = a$ for some complex number a. Consider a sequence $\{t_n\}$ in T with $t_n \neq t_0$ and $\lim_{n\to\infty} t_n = t_0$. Since

$$\begin{aligned}
&\lim_{n,m\to\infty} \|X(t_n) - X(t_m)\|^2 \\
= \ &\lim_{n,m\to\infty} ((X(t_n), X(t_n)) - (X(t_n), X(t_m)) - (X(t_m), X(t_n)) + (X(t_m), X(t_m))) \\
= \ &\lim_{n,m\to\infty} (K(t_n, t_n) - K(t_n, t_m) - K(t_m, t_n) + K(t_m, t_m)) \\
= \ &a - a + a - a = 0,
\end{aligned}$$

$\{X(t_n)\}$ is a Cauchy sequence and then converges in L^2-norm to a r.v. Z in $L^2(\Omega, \mathcal{F}, \mathcal{P})$. Now consider another sequence $\{s_m\}$ in T with $s_m \neq t_0$ and $\lim_{m\to\infty} s_m = t_0$. With similar argument, $\{X(s_m)\}$ is a Cauchy sequence and converges in L^2-norm to a r.v. Y in $L^2(\Omega, \mathcal{F}, \mathcal{P})$. Define a new sequence $\{u_k\}$ in T with $u_{2n-1} = s_n$ and $u_{2n} = t_n$ for $n \geq 1$. Then $u_k \neq t_0$, $\lim_{k\to\infty} u_k = t_0$ and $\{X(u_k)\}$ is a Cauchy sequence and converges in L^2-norm to a r.v. W in $L^2(\Omega, \mathcal{F}, \mathcal{P})$. Since $\{X(s_m\}$ and $\{X(t_n)\}$ are subsequences of $\{X(u_k)\}$, we must have $W = Y = Z$. This proves that $\lim_{t\to t_0} X(t) = Z$. Furthermore, if $K(s,t)$ is continuous at (t_0, t_0), then for a sequence $\{t_n\}$ in T with $\lim_{n\to\infty} t_n = t_0$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \|X(t_n) - X(t_0)\|^2 \\
= \ &\lim_{n\to\infty} ((X(t_n), X(t_n)) - (X(t_n), X(t_0)) - (X(t_0), X(t_n)) + (X(t_0), X(t_0))) \\
= \ &\lim_{n\to\infty} (K(t_n, t_n) - K(t_n, t_0) - K(t_0, t_n) + K(t_0, t_0)) \\
= \ &K(t_0, t_0) - K(t_0, t_0) + K(t_0, t_0) - K(t_0, t_0) = 0,
\end{aligned}$$

which proves that $\lim_{n\to\infty} X(t_n) = X(t_0)$, i.e., $X(t)$ is continuous at t_0. $\square$

Lemma 3.3.4 *Let $f(s,t)$ be a complex-valued function over $(a,b)\times(a,b)$. If $\frac{\partial f(s,t)}{\partial s}$, $\frac{\partial f(s,t)}{\partial t}$ and $\frac{\partial^2 f(s,t)}{\partial s\partial t}$ exist and are continuous over $(a,b)\times(a,b)$, then*

$$\lim_{h,k\to 0}\frac{1}{hk}(f(s+h,t+k)-f(s+h,t)-f(s,t+k)+f(s,t))=\frac{\partial^2 f(s,t)}{\partial s\partial t}$$

for all $s,t\in(a,b)$.

Theorem 3.3.5 *Let $X(t)$ be a 2nd-order process over T and $K(s,t)$ be the covariance function of $X(t)$. Let t_0 be an interior point of T. Then $X(t)$ is differentiable at t_0 if and only if the limit*

$$\lim_{h,k\to 0}\frac{1}{hk}\{K(t_0+h,t_0+k)-K(t_0+h,t_0)-K(t_0,t_0+k)+K(t_0,t_0)\}$$

exists. Moreover, $X(t)$ is continuously differentiable over an interval (a,b) of T if and only if $\frac{\partial K(s,t)}{\partial s}$, $\frac{\partial K(s,t)}{\partial t}$ and $\frac{\partial^2 K(s,t)}{\partial s\partial t}$ exist and are continuous over $(a,b)\times(a,b)$.

Proof. Since t_0 is an interior point of T, there exists an $\epsilon>0$ such that $(t_0-\epsilon,t_0+\epsilon)\in T$. Define a 2nd-order process $Y(h)$ over $(-\epsilon,\epsilon)$ as

$$Y(h)\triangleq\frac{X(t_0+h)-X(t_0)}{h}$$

for all $h\in(-\epsilon,0)\cup(0,\epsilon)$ and $Y(0)\triangleq 0$. Thus $X(t)$ is differentiable at t_0, i.e., $X'(t_0)$ exists if and only if

$$\lim_{h\to 0}Y(h)=\lim_{h\to 0}\frac{X(t_0+h)-x(t_0)}{h}$$

exists and by Theorem 3.3.3, if and only if $\lim_{h,k\to 0}K_Y(h,k)$ exists, where $K_Y(h,k)$ is the covariance function of the process $Y(h)$

$$\begin{aligned}
K_Y(h,k) &\triangleq (Y(h),Y(k))\\
&= \left(\frac{X(t_0+h)-X(t_0)}{h},\frac{X(t_0+k)-X(t_0)}{k}\right)\\
&= \frac{1}{hk}(X(t_0+h)-X(t_0),X(t_0+k)-X(t_0))\\
&= \frac{1}{hk}((X(t_0+h),X(t_0+k))-(X(t_0+h),X(t_0))-(X(t_0),X(t_0+k))\\
&\quad +(X(t_0),X(t_0)))\\
&= \frac{1}{hk}\left(K(t_0+h,t_0+k)-K(t_0+h,t_0)-K(t_0,t_0+k)-K(t_0,t_0)\right).
\end{aligned}$$

This completes the proof of the first part. For the second part, we assume that $X(t)$ is continuously differentiable over an interval (a,b) of T. Since

$$\begin{aligned}
\lim_{h\to 0}\frac{K(s+h,t)-K(s,t)}{h} &= \lim_{h\to 0}\left(\frac{X(s+h)-X(s)}{h},X(t)\right)\\
&= \left(\lim_{h\to 0}\frac{X(s+h)-X(s)}{h},X(t)\right)\\
&= (X'(s),X(t))
\end{aligned}$$

for all $s, t \in (a,b)\times(a,b)$ by the continuity of inner product and $X'(t)$ is continuous over (a,b), the partial derivative $\frac{\partial K(s,t)}{\partial s}$ exists and is continuous over $(a,b)\times(a,b)$. Similarly, the partial derivative $\frac{\partial K(s,t)}{\partial t}$

$$\frac{\partial K(s,t)}{\partial t} = (X(s), X'(t))$$

also exists and is continuous over $(a,b)\times(a,b)$. Furthermore,

$$\begin{aligned}
\lim_{h\to 0}\frac{\partial K(s+h,t)/\partial t - \partial K(s,t)/\partial t}{h} &= \lim_{h\to 0}\left(\frac{X(s+h)-X(s)}{h}, X'(t)\right) \\
&= \left(\lim_{h\to 0}\frac{X(s+h)-X(s)}{h}, X'(t)\right) \\
&= (X'(s), X'(t)) \qquad (3.6)
\end{aligned}$$

by the continuity of inner product. Thus the partial derivative $\frac{\partial^2 K(s,t)}{\partial s\partial t}$ exists and is continuous over $(a,b)\times(a,b)$. Conversely, assume that $\frac{\partial K(s,t)}{\partial s}$, $\frac{\partial K(s,t)}{\partial t}$ and $\frac{\partial^2 K(s,t)}{\partial s\partial t}$ exist and are continuous over $(a,b)\times(a,b)$. By Lemma 3.3.4, we have

$$\lim_{h,k\to 0}\frac{1}{hk}(f(t_0+h,t_0+k) - f(t_0+h,t_0) - f(t_0,t_0+k) + f(t_0,t_0)) = \frac{\partial^2 f(s,t)}{\partial s\partial t}|_{s=t=t_0}$$

for all $t_0 \in (a,b)$, which implies that the derivative $X'(t)$ exists for all $t \in (a,b)$ by the first part of the proof. Now let $Y(t) = X'(t)$ over (a,b). Since the covariance function $K_Y(s,t)$ of $Y(t)$

$$K_Y(s,t) = (X'(s), X'(t)) = \frac{\partial^2 K(s,t)}{\partial s\partial t},$$

by (3.6), is continuous over $(a,b)\times(a,b)$, $Y(t) = X'(t)$ is continuous over (a,b) by Theorem 3.3.3. □

The following corollary is a side result of the proof of the above theorem.

Corollary 3.3.6 *If $X(t)$ is a continuously differentiable 2nd-order process over $t \in (a,b)$, then we have*

$$\begin{bmatrix} \mathcal{E}(X(s)\overline{X(t)}) & \mathcal{E}(X(s)\overline{X'(t)}) \\ \mathcal{E}(X'(s)\overline{X(t)}) & \mathcal{E}(X'(s)\overline{X'(t)}) \end{bmatrix} = \begin{bmatrix} K(s,t) & \frac{\partial K(s,t)}{\partial t} \\ \frac{\partial K(s,t)}{\partial s} & \frac{\partial^2 K(s,t)}{\partial s\partial t} \end{bmatrix}.$$

□

A stochastic process $X(t)$, $t \in T$, (not necessarily a 2nd-order process) is said to have independent increments if for $t_0 < t_1 < t_2 < ... < t_{n-1} < t_n$ in T, the increments $X(t_1) - X(t_0)$, $X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ are statistically independent r.v.s. The process $X(t)$ is said to have stationary increments if the joint distribution of the increments $X(t_1+h) - X(t_0+h), X(t_2+h) - X(t_1+h), \ldots, X(t_n+h) - X(t_{n-1}+h)$ is independent of h.

A Brownian motion $X(t), t \in [0,\infty)$, is a real-valued 2nd-order process which has

1. independent and stationary increments,
2. $X(0) = 0$,
3. $X(t)$ normally distributed with zero mean and variance $\sigma^2 t$,
4. $X(t)$ continuous over $[0, \infty)$.

If $\sigma = 1$, then $X(t)$ is called a standard Brownian motion.

Theorem 3.3.7 *The covariance function of a Brownian motion $X(t)$ is*

$$K(s,t) = \sigma^2 \min(s,t).$$

Proof. For $0 \leq s \leq t$, we have

$$\begin{aligned} K(s,t) &= \mathcal{E}(X(s)X(t)) \\ &= \mathcal{E}(X(s)(X(s) + X(t) - X(s))) \\ &= \mathcal{E}(X^2(s)) + \mathcal{E}(X(s)(X(t) - X(s))) \end{aligned}$$

and with $X(0) = 0$,

$$\begin{aligned} \mathcal{E}(X(s)(X(t) - X(s))) &= \mathcal{E}((X(s) - X(0))(X(t) - X(s))) \\ &= \mathcal{E}(X(s) - X(0))\mathcal{E}(X(t) - X(s)) \quad \because \text{independent increments} \\ &= \mathcal{E}(X(s))\mathcal{E}(X(t) - X(s)) \\ &= 0 \cdot \mathcal{E}(X(t) - X(s)) = 0. \end{aligned}$$

Thus we have $K(s,t) = \mathcal{E}(X^2(s)) = \sigma^2 s \ \ \forall\, 0 \leq s \leq t$. Similarly, $K(s,t) = \sigma^2 t \ \ \forall\, 0 \leq t \leq s$. Then we have $K(s,t) = \sigma^2 \min(s,t) \ \ \forall\, s,t \geq 0$. □

Remark 3.3.8 *Assume that the condition 3 is missing in the definition of a Brownian motion. Then define*

$$\begin{aligned} m(t) &= \mathcal{E}(X(t)), \\ v(t) &= \mathcal{E}((X(t) - m(t))^2) = \|X(t) - m(t)\|^2, \end{aligned}$$

for all $t \in [0, \infty)$. We first show that $m(t)$ and $v(t)$ are continuous functions over $[0, \infty)$. Note that for a r.v. Z in $L^2(\Omega, \mathcal{F}, \mathcal{P})$, we have

$$|\mathcal{E}(Z)| = |\mathcal{E}(Z \cdot 1)| \leq \sqrt{\mathcal{E}(|Z|^2)\mathcal{E}(1)} = \sqrt{\mathcal{E}(|Z|^2)} = \|Z\|$$

by Schwartz inequality. Now for any $t \in [0, \infty)$ and any sequence $\{t_n\}$ in $[0, \infty)$ such that

$$t_n \to t \text{ as } n \to \infty,$$

we have

$$\|X(t_n) - X(t)\| \to 0 \text{ as } n \to \infty,$$

by the (L^2-)continuity of $X(t)$ over $[0,\infty)$, which implies that

$$|m(t_n) - m(t)| \to 0 \text{ as } n \to \infty,$$

since

$$|m(t_n) - m(t)| = |\mathcal{E}(X(t_n) - X(t))| \leq \|X(t_n) - X(t)\|.$$

Thus $m(t)$ is continuous over $T = [0,\infty)$. With similar arguments, we can show that the continuity of $X(t)$ implies the continuity of $v(t)$. Now for $s, t \geq 0$, we have

$$\begin{aligned} m(s+t) &= \mathcal{E}(X(s+t)) \\ &= \mathcal{E}(X(s) + X(s+t) - X(s)) \\ &= \mathcal{E}(X(s)) + \mathcal{E}(X(s+t) - X(s)) \\ &= m(s) + \mathcal{E}(X(t) - X(0)) \quad \text{by stationary increments} \\ &= m(s) + m(t) \end{aligned}$$

and by the continuity of $m(t)$ and $m(0) = 0$, we have

$$m(t) = at,$$

where a is the expectation of $X(1)$. Similarly, for $t, s \geq 0$, we have

$$\begin{aligned} v(s+t) &= \mathcal{E}((X(s+t) - m(s+t))^2) \\ &= \mathcal{E}(\{(X(s) - m(s)) + (X(s+t) - X(s) - m(t))\}^2) \\ &= \mathcal{E}((X(s) - m(s))^2) + 2\mathcal{E}((X(s) - m(s))((X(s+t) - X(s) - m(t)))) \\ &\quad + \mathcal{E}((X(s+t) - X(s) - m(t))^2) \\ &= v(t) + v(s), \end{aligned}$$

from the property of independent and stationary increments, and by the continuity of $v(t)$ and $v(0) = 0$, we have

$$v(t) = \sigma^2 t,$$

where σ^2 is the variance of $X(1)$. Thus, if condition 3 is replaced by

3'. $X(t)$ normally distributed,

then $X(t)$ will be called a Brownian motion with linear drift for the mean function $m(t)$ is a linear function $m(t) = at$. And a Brownian motion in our previous defintion is just a Brownian motion with zero drift. □

It can be seen from the covariance function of a Brownian motion and Theorem 3.3.5 that a Brownian motion is nowhere differentiable.