1

Introduction

Two main topics are covered in this book. One is machine-type communication (MTC) and the other is non-orthogonal multiple access (NOMA). Each topic has its own foundations and applications. In this chapter, we briefly explain each of them, and then explain why both topics should be covered in this book.

1.1 Machine-Type Communication

It may not be easy to imagine our daily life and business without the Internet although it began to appear as a backbone network in the 1970s to interconnect a small number of academic and military networks. The Internet is a network of networks and allows to exchange information between servers, computers, mobile phones, and so on restlessly. The Internet-of-Things (IoTs) is a natural extension of the Internet as machines, devices, and sensors are connected to the Internet to exchange information without human intervention in a number of applications such as smart factory.

As the number of devices connected to the Internet grows, their connectivity becomes important. Private and public networks can be used for their connectivity. For example, for smart home applications, a private network can be used at home to allow a small number of devices to be connected. For smart city applications, a large-scale public network would be preferable. Thus, the deployment of IoT networks depend on applications.

MTC is to support communications between machines or devices without human intervention. Unlike human-type communication (HTC), MTC mainly focuses on uplink transmission rather than downlink transmission (this is one of the main differences between MTC and HTC, where it can be seen that MTC's design principles must be different from those of HTC) and will support sporadic traffic in the form of short data packets. As a result, in order to keep signaling

Massive Connectivity: Non-Orthogonal Multiple Access to High Performance Random Access,
First Edition. Jinho Choi.

overhead low, the random access channel (RACH) procedure is used for MTC in Long-Term Evolution (LTE) systems. In the fifth generation (5G) system, a new random access scheme consisting of two steps, which is more efficient than the RACH procedure in LTE consisting of four steps, has been standardized.

MTC can provide connectivity for a large number of devices in a cell, paving the way for IoT applications to interact with devices deployed over a large area via cellular systems. This means that MTC becomes essential in various IoT applications such as smart cities and intelligent transport systems.

Furthermore, the global number of connected devices is expected to exceed 500 billion by 2030, while the human population is predicted to be 8.5 billion by the United Nations (UN). This means that IoT devices will outnumber human population by approximately 60-folds in 2030, and these devices will be used in a variety of IoT applications requiring heterogeneous connectivity demand. As a result, MTC will play a more prominent role in 5G and beyond (i.e. the sixth generation (6G)) and thus new MTC schemes need to be developed to meet the diverse requirements for future IoT applications.

1.2 Non-orthogonal Multiple Access

Various multiple access schemes have been used to support multiple users in a cellular system. For example, time division multiple access (TDMA) is adopted in the global system for mobile communications (GSM), which is regarded as a second generation (2G) system. In the third generation (3G) to 5G systems, orthogonal frequency division multiple access (OFDMA) is used. In general, most multiple access schemes used in cellular systems are orthogonal multiple access (OMA) schemes that allocate orthogonal channel resources to different users. To increase the spectral efficiency, NOMA schemes have been considered, where multiple users share the same channel resource.

NOMA has been extensively studied for cellular systems since the 2010s. In particular, for downlink transmissions, various NOMA schemes are studied using the difference in propagation loss between users near the center of the cell (where a base station (BS) is located) and users far from the center. The resulting NOMA is often referred to as power-domain NOMA.

It is necessary to distinguish between power-domain NOMA and NOMA in a broad sense. For example, code-division multiple access (CDMA) and interleave-division multiple access (IDMA) can be seen as NOMA schemes, because multiple users' signals can co-exist in a shared radio resource block, where one user signal can see the other users' signals as interfering signals. In CDMA, each user's signal is spread by a dedicated sequence, which is called the spreading sequence. Due to spreading sequences, CDMA has a bandwidth

expansion. In particular, the bandwidth of CDMA increases by a factor of the processing gain or the length of the spreading sequence, while IDMA is a generalization of CDMA with forward error correcting codes. On the other hand, power-domain NOMA does not use spreading sequences. As a result, there is no bandwidth expansion and a high spectral efficiency can be achieved.

In power-domain NOMA, however, the transmit power levels and transmission rates should be carefully decided so that successive interference cancellation (SIC) can be used to remove other users' signals once they are decoded.

1.3 NOMA for MTC

In general, power-domain NOMA for downlink requires coordinated transmissions by a BS in terms of transmit powers and rates. Thus, it seems difficult to use power-domain NOMA for uplink as coordinated transmission by distributed users is not easy to implement. In other words, the gain of NOMA can be offset by excessive signaling overhead to perform coordinated transmissions by distributed users. As a result, the use of NOMA for random access seems quite challenging. On the contrary, NOMA is well-suited to random access as we will illustrate with two users.

Suppose that two users want to access a given channel without coordination. If two users always transmit their signals, they experience collisions and no user succeeds to transmit. Thus, they need to transmit with a certain probability. To this end, each user is to decide the access probability, denoted by p_k, $k = 1, 2$, for user k. The probability that at least one user succeeds to transmit a packet is given by

$$p_{\text{succ}} = p_1(1 - p_2) + p_2(1 - p_1).$$

If $p = p_k$, $p_{\text{succ}} = 2p(1 - p)$, which is maximized when $p = \frac{1}{2}$ and the maximum of p_{succ} is $\frac{1}{2}$, which is also the maximum average number of successfully transmitted packets. To consider random access with NOMA, we can assume two different power levels, P_{H} and P_{L} with $P_{\text{H}} > P_{\text{L}}$, and the receiver is able to decode both the signals if one user transmits a signal with transmit power P_{H} and the other with P_{L}. Let q be the probability that a user chooses the high transmit power when transmitting (with probability p). Then, the average number of successfully transmitted packets is given as follows:

$$\begin{aligned}\eta_2 &= \underbrace{\binom{2}{1} p(1-p)}_{\text{one user transmits}} + 2 \underbrace{\binom{2}{2} p^2}_{\text{two users transmit}} \binom{2}{1} q(1-q) \\ &= 2p(1-p) + 4p^2q(1-q),\end{aligned}$$

where the first term is the average number of successfully transmitted packets when only one user transmits and the second term is the average number of successfully transmitted packets when two users transmit simultaneously. It is easy to show that $q = \frac{1}{2}$ maximizes η_2. Then, we have $\eta_2 = 2p(1-p) + p^2 = 2p - p^2$. Thus, $p = 1$ maximizes η_2, which is 1. In other words, the maximum average number of successfully transmitted packets can be doubled if NOMA is used for uncoordinated transmissions of two users in uplink transmissions.

As more power levels are considered, the average number of successfully transmitted packets can increase. However, this comes at the cost of high transmit power by devices.

In this book, we mainly focus on the principles of NOMA and the application of NOMA to MTC. In particular, we discuss how NOMA can help improve the performance of random access in MTC once we present key ideas of random access including its stability. Game theory will also be used to understand the nature of random access where users compete for common radio resources in MTC.

1.4 An Overview of Probability and Random Processes

Prior to the main parts of this book, we present an overview of probability and random processes in this section, which can be used to see the required background in terms of probability and random processes. The reader is referred to well-known textbooks such as Papoulis and Pillai (2002); Ross (1995); Mitzenmacher and Upfal (2005) if not well equipped with theory of probability and random processes.

1.4.1 Review of Probability

A sample space Ω is the set of all possible outcomes (or events) of an experiment. Let A_m be an event m, which is a subset of Ω. A probability measure $\Pr(\cdot)$ is a mapping from Ω to the real line with the following properties:

1. $\Pr(A_m) \geq 0, A_m \in \Omega$
2. $\Pr(\Omega) = 1$
3. For a countable set of events, $\{A_m\}$, if $A_l \cup A_m = \emptyset$, for $l \neq m$, then

$$\Pr\left(\bigcup_{l=1}^{\infty} A_l\right) = \sum_{l=1}^{\infty} \Pr(A_l).$$

The joint probability of two events A and B is expressed as $\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$, where the conditional probability of A given B is

expressed as

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}, \quad \Pr(B) > 0.$$

The two events A and B are independent if and only if

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

and this implies $\Pr(A|B) = \Pr(A)$.

In addition, for any two events A and B, we have

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \leq \Pr(A) + \Pr(B),$$

where the equality holds if $A \cap B = \emptyset$. Thus, for a set of events, it can be shown that

$$\Pr\left(\bigcup_{l=1}^{\infty} A_l\right) \leq \sum_{l=1}^{\infty} \Pr(A_l), \tag{1.1}$$

which is called the union bound.

1.4.2 Random Variables

A random variable is a mapping from an event ω in Ω to a real number, denoted by $X(\omega)$. We first consider continuous random variables. The *cumulative distribution function* (cdf) of X is defined as

$$\begin{aligned} F_X(x) &= \Pr(\{\omega | X(\omega) \leq x\}) \\ &= \Pr(X \leq x) \end{aligned}$$

and the *probability density function* (pdf) is defined as

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x),$$

where the subscript X on F and f identifies the random variable. If the random variable is obvious, the subscript is often omitted.

Note that we use capital letters to denote random variables in this book if necessary. For example, X is a random variable, while x is a variable. Random vectors will be written in boldface letters. For example, $\mathbf{x}$ is a random vector. This could lead to a confusion, because a vector (not a random vector) will also be written in boldface letters. To avoid this confusion, we will make clear indication if necessary.

There are some well-known pdfs of continuous random variables as follows:

1. Gaussian pdf with mean μ and variance σ^2:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right).$$

As the Gaussian pdf is frequently used in this book, it is denoted by $\mathcal{N}(\mu, \sigma^2)$. If X is a Gaussian random variable with mean μ and variance σ^2, we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

2. Exponential pdf ($a > 0$):

$$f(x) = \begin{cases} a\, e^{-ax}, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

3. Rayleigh pdf ($b > 0$):

$$f(x) = \begin{cases} \frac{x}{b}\, e^{-x^2/b}, & \text{if } x \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

4. Chi-square pdf of n degrees of freedom:

$$f(x) = \begin{cases} \frac{x^{(n-2)/2} e^{-x/2}}{2^{n/2}\Gamma(n/2)}, & \text{if } x \geq 0; \\ 0, & \text{otherwise,} \end{cases}$$

where $\Gamma(x)$ is the gamma function.

A joint cdf of random variables, $X_1, X_2, \ldots, X_n$, is expressed as

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \Pr(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n).$$

If the random variables are independent, the cdf is given by

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i).$$

The joint pdf of random variables, $X_1, X_2, \ldots, X_n$, is then expressed as

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n).$$

For independent random variables, the joint pdf becomes a product of individual pdfs, i.e. $f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$. The conditional pdf of X_1 given $X_2 = x_2$ is

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}, \quad f_{X_2}(x_2) > 0.$$

Thus, if X_1 and X_2 are independent, it can be shown that

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)} = f_{X_1}(x_1).$$

Let us consider the expectation and variance of X. First, the expectation of X is defined as

$$\mathbb{E}[X] = \int x f_X(x) \mathrm{d}x.$$

In addition, the expectation of $g(X)$, a function of X, is

$$\mathbb{E}[g(X)] = \int g(x) f_X(x) \mathrm{d}x.$$

The variance of X is defined as

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \int (x - \mathbb{E}[X])^2 f_X(x) \mathrm{d}x.\end{aligned}$$

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

$$\mathbb{E}[X|Y = y] = \int x f_{X|Y}(x|Y = y) \mathrm{d}x.$$

The pdf of a real-valued joint Gaussian random vector, $\mathbf{x}$, is given as

$$f(\mathbf{x}) = \frac{1}{\sqrt{\det(2\pi \mathbf{C}_\mathbf{x})}} \exp\left(-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}} \mathbf{C}_\mathbf{x}^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right),$$

where $\bar{\mathbf{x}} = \mathbb{E}[\mathbf{x}]$ and $\mathbf{C}_\mathbf{x} = \mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}}]$. Here, the superscript T stands for the transpose. For convenience, if $\mathbf{x}$ is a Gaussian random vector with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{C}_\mathbf{x}$, we write $\mathbf{x} \sim \mathcal{N}(\bar{\mathbf{x}}, \mathbf{C}_\mathbf{x})$.

If $\mathbf{x}$ is a circularly symmetric complex Gaussian (CSCG) random vector,

$$f(\mathbf{x}) = \frac{1}{\det(\pi \mathbf{C}_\mathbf{x})} \exp\left(-(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{H}} \mathbf{C}_\mathbf{x}^{-1} (\mathbf{x} - \bar{\mathbf{x}})\right),$$

where $\bar{\mathbf{x}} = \mathbb{E}[\mathbf{x}]$ and $\mathbf{C}_\mathbf{x} = \mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{H}}]$. For a CSCG random vector, we can show that

$$\mathbb{E}[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^{\mathrm{T}}] = 0.$$

If $\mathbf{x}$ is a CSCG random vector with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{C}_\mathbf{x}$, we write $\mathbf{x} \sim \mathcal{CN}(\bar{\mathbf{x}}, \mathbf{C}_\mathbf{x})$.

While a continuous random variable X is real-valued (i.e. $X \in \mathbb{R}$), a discrete random variable X has a countable image. For example, if X can be either 0 or 1, it is a discrete random variable. The cdf of a discrete random variable X be defined as that of a continuous random variable. However, the pdf cannot be defined. Instead, the probability mass function (pmf), can be defined as

$$P_m = \Pr(X = x_m), \quad x_m \in \mathcal{X}, \tag{1.2}$$

where $\mathcal{X}$ represents of the image of X. The mean and variance of a discrete random variable are given by

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x \in \mathcal{X}} x \Pr(X = x) = \sum_m x_m P_m, \\ \mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_m (x_m - \mathbb{E}[X])^2 P_m.\end{aligned} \tag{1.3}$$

There are a number of different discrete probability distributions. For example, the Bernoulli distribution is the discrete probability distribution of a random variable X that takes 1 with probability p and 0 with probability $1-p$. The binomial distribution is the discrete probability distribution of the random variable that is the sum of the outcomes of n independent Bernoulli trials, which is given by

$$\Pr(X = i) = \binom{n}{i} p^i (1-p)^{n-1}, \quad X \in \mathcal{X} = \{0, \dots, n\}. \tag{1.4}$$

In this case, we write $X \sim X \sim \mathrm{Bin}(n, p)$. Another important distribution is the Poisson distribution that is given by

$$\Pr(X = i) = \frac{\lambda^i}{i!} e^{-\lambda}, \quad X \in \mathcal{X} = \{0, \dots\}, \tag{1.5}$$

where $\lambda > 0$, and we write $X \sim \mathrm{Pois}(\lambda)$. It can be readily shown that $\mathbb{E}[X] = \lambda$ and $\mathrm{Var}(X) = \lambda$. If X follows the Poisson distribution with parameter λ, we write $X \sim \mathrm{Pois}(\lambda)$.

A sequence of random variables, X_n, $n = 1, 2, \dots$, is called an independent and identically distributed (iid) sequence if the X_n's are independent and their distribution are identical.

Let X_n be a sequence of independent random variables with $\bar{x}_n = \mathbb{E}[X_n]$ and $\sigma_n^2 = \mathrm{Var}(X_n)$. Consider the normalized sum as follows:

$$\begin{aligned} Y_k &= \frac{S_k - \mathbb{E}[S_k]}{\mathrm{Var}(S_k)} \\ &= \frac{\sum_{n=1}^k X_n - \bar{x}_n}{\sum_{n=1}^k \sigma_n^2}, \end{aligned}$$

where $S_k = \sum_{n=1}^k X_n$. If

$$\lim_{k\to\infty} \Pr(a \le Y_k \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} \, \mathrm{d}y \quad (a < b),$$

the sequence X_n, $n = 1, 2, \dots$, is said to satisfy the central limit theorem (CLT). If a sequence satisfies the following condition, called the Lyapunov condition:

$$\lim_{k\to\infty} \frac{\sum_{n=1}^k \mathbb{E}[|X_n - \bar{x}_n|^3]}{\left(\sum_{n=1}^k \sigma_n^2\right)^{\frac{3}{2}}} = 0,$$

this sequence satisfies the CLT. An example is a binary iid sequence, where $X_n \in \{0, 1\}$ and $\Pr(X_n = 1) = \frac{1}{2}$. In this case, $\mathbb{E}[|X_n - \bar{x}_n|^3] = 0$ and it satisfies the Lyapunov condition.

For a nonnegative random variable X and $a > 0$, it can be shown that

$$\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \tag{1.6}$$

which is called Markov's inequality. If X is not a nonnegative random variable, we can consider $Y = e^{\lambda X}$, which is a nonnegative random variable. Then, it can be shown that

$$\Pr(e^{\lambda X} \geq a) \leq \frac{\mathbb{E}[e^{\lambda X}]}{a} = \frac{M_X(\lambda)}{a}, \tag{1.7}$$

where $M_X(\lambda) = \mathbb{E}[e^{\lambda X}]$ is the moment generating function (MGF). Furthermore, let $a = e^{\lambda\tau}$. Then, we have

$$\Pr(X \geq \tau) \leq e^{-\lambda\tau}\ \mathbb{E}[e^{\lambda X}]. \tag{1.8}$$

A tight upper-bound on the tail probability $\Pr(X \geq \tau)$ can be obtained by minimizing the right-hand side (RHS) term with respect to $\lambda \geq 0$, i.e.

$$\Pr(X \geq \tau) \leq \min_{\lambda \geq 0} e^{-\lambda\tau}\ \mathbb{E}[e^{\lambda X}], \tag{1.9}$$

which is called the Chernoff bound.

1.4.3 Random Processes

A (discrete-time) random process is a sequence of random variables, $\{X_m\}$. The mean and autocorrelation function of $\{X_m\}$ are denoted by $\mathbb{E}[X_m]$ and $R_X(l, m) = \mathbb{E}[X_l X_m^*]$, respectively. A random process is called wide-sense stationary (WSS) if

$$\mathbb{E}[X_l] = \mu, \quad \forall l;$$
$$\mathbb{E}[X_l X_m^*] = \mathbb{E}[X_{l+p} X_{m+p}^*], \quad \forall l, m, p.$$

For a WSS random process, we have

$$R_X(l, m) = R_X(l - m).$$

The power spectrum of a WSS random process, $\{X_m\}$, is defined as

$$S_X(z) = \sum_{k=-\infty}^{\infty} z^{-k} R_X(k).$$

In addition, we can show that

$$R_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(e^{\mathrm{j}\omega})\ e^{\mathrm{j}k\omega}\ \mathrm{d}\omega,$$

where $\mathrm{j} = \sqrt{-1}$. A zero-mean WSS random process is called white if

$$R_X(l) = \begin{cases} \sigma_x^2, & \text{if } l = 0; \\ 0, & \text{otherwise,} \end{cases}$$

where $\sigma_x^2 = \mathbb{E}[|X_l|^2]$.

If X_l is an output of the linear system whose impulse response is $\{h_p\}$ with a zero-mean white random process input, n_l, the autocorrelation function of $\{X_m\}$ is given by

$$R_X(l) = \sigma_n^2 \sum_l h_l h_{l-m},$$

where $\sigma_n^2 = \mathbb{E}[|n_l|^2]$. In addition, its power spectrum is given as

$$S_X(z) = \sigma_n^2 H(z) H(z^{-1})$$

or

$$S_X(e^{\mathrm{j}\omega}) = \sigma_n^2 |H(e^{\mathrm{j}\omega})|^2, \quad z = e^{\mathrm{j}\omega}.$$

1.4.4 Markov Chains

In this subsection, we briefly present some key results of Markov chains. The reader is referred to Norris (1998) for more details on Markov chains.

Suppose a discrete-time random process X_n for $n = 0, 1, \ldots$ takes on a finite or countable number of possible values from a set of nonnegative integers, i.e. $X_n \in \{0, 1, \ldots\}$. If $X_n = i$, it is said that the process is in state i at time n. A process has the Markov property if the state at time $n + 1$ depends only on the state at time n, i.e.

$$\begin{aligned} &\Pr(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ &\quad = \Pr(X_{n+1} = j \mid X_n = i), \end{aligned}$$

and such a process is known as a Markov chain. Let us denote the transition probability from state i at time n to state j at time $n + 1$ by

$$\Pr(X_{n+1} = j \mid X_n = i) = p_{ij}.$$

It satisfies the following properties:

$$p_{ij} \geq 0 \text{ and } \sum_j p_{ij} = 1.$$

To describe the transition probabilities for an arbitrary number of steps, let us consider the n-step transition probability as

$$p_{ij}^n = \Pr(X_{n+m} = j \mid X_m = i) \quad \text{for } n \geq 1. \tag{1.10}$$

As an example, let us consider a two-step transition probability as

$$\begin{aligned} p_{ij}^2 &= \Pr(X_{2+m} = j \mid X_m = i) \\ &= \Pr(X_2 = j \mid X_0 = i) \\ &= \sum_k \Pr(X_2 = j \mid X_1 = k, X_0 = i) \Pr(X_1 = k \mid X_0 = i) \\ &= \sum_k \Pr(X_2 = j \mid X_1 = k) \Pr(X_1 = k \mid X_0 = i) \\ &= \sum_k p_{ik} p_{kj}, \end{aligned} \tag{1.11}$$

where the second line shows that the transition probabilities do not depend on time and Markov property has been used from the third line to the fourth line.

Let us consider the probability that the process goes from state i at time $t = 0$ and passes through state k at time $t = m$, and ends up at state j at time $t = n + m$. Similar to (1.11), this probability is obtained as

$$p_{ij}^{n+m} = \sum_k p_{ik}^n p_{kj}^m, \tag{1.12}$$

which is known as the Chapman–Kolmogorov equation.

Let us consider classes of states: State j is said to be accessible from state i if $p_{ij}^n > 0$ for some $n \geq 1$. It is said that two states i and j communicate if they are accessible to each other. This communication relation is denoted by $i \leftrightarrow j$. In addition, if two states communicate with each other, we say that two states belong to the same class. Accordingly, two different classes of states should be disjoint because they can communicate otherwise.

A Markov chain is said to be irreducible if all states communicate with each other.

If $P_{ii}^n = 0$ for all n that are not divisible by d, it is said that state i has period d. State i with $d = 1$ is said to be aperiodic. If two states i and j communicate with each other, the periods of the two states are the same.

Let ρ_{ij} denote the probability that the process makes a transition into state j given that the process starts in state i. State j is said to be recurrent if $\rho_{ij} = 1$ (for any i), and transient otherwise (i.e. $\rho_{ij} < 1$). Let ρ_{ij}^n represent the probability that the process makes the first transition into state j at time n since it starts in state i. Thus, $\rho_{ij} = \sum_{n=1}^{\infty} \rho_{ij}^n$. In addition, denote by μ_i the expected number of transitions to return to state i given that the process starts in state i. Then, for a recurrent state i, it can be shown that

$$\mu_i = \sum_{n=1}^{\infty} n \rho_{ii}^n. \tag{1.13}$$

State i is said to be positive recurrent if $\mu_i < \infty$, and null recurrent otherwise (i.e. $\mu_i = \infty$).

A positive recurrent and aperiodic state is called ergodic. For an irreducible Markov chain, if one state is ergodic, the other states are also ergodic. For an ergodic Markov chain, the stationary distribution, $\{\pi_i\}$, exists, which is defined as

$$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}. \tag{1.14}$$

In addition, for an ergodic Markov chain,

$$\pi_j = \lim_{n\to\infty} P_{ij}^n > 0. \tag{1.15}$$