SCMA469 Actuarial Statistics

Pairote Satiracoo

2021-08-24

Contents

Chapter 1

Introduction to Stochastic Processes

The course will cover the probabilistic framework for stochastic models of real-world applications with emphasis on actuarial work. We will illustrate some practical actuarial problems for which we will develop mathematical models, tools and techniques for analysing and quantifying the uncertainty of the problems.

Here are some of the examples which will be covered later in the course.

1.1 Examples of real world processes

Example 1.1. *(**No claims discount systems (NCD)**) A well-known model widely used by auto insurance companies is the **no claims discount system**, in which an insured receives a discount for a claim free year, while the insured is penalised by an additional premium when one or more accidents occur.*

An example of the NCD system in UK may be structured as follows:

Level	7	6	5	4	3	2	1
Premium	100%	75%	65%	55%	45%	40%	33%

The rules for moving between these levels are as follows:

- *For a claim-free year, a policyholder moves down 1 level.*
- *Levels 4-7:*
  - *For every one claim, the policyholder moves up 1 level or remains at level 7.*
  - *For every two or more claims, move to, or remains at, level 7.*
- *Levels 2-3:*
  - *For every one claim, move up 2 levels.*
  - *For every two claims, move up 4 levels.*
  - *For every three or more claims, move to level 7.*
- *Level 1:*
  - *For every one claim, move to level 4.*
  - *For every two claims, move to level 6.*
  - *For every three or more claims, move to level 7.*

The no claims discount system is a form of experience rating consisting of a finite number of levels (or classes), each with its own premium. The 7 levels are experience-rated as described above.

For the NCD model, questions of interest may include:

1. *For 10,000 policyholders at level 7, estimate the* expected numbers* at each discount level at a given time, or once stability has been achieved.

2. *What is the* probability* that a policyholder who is at a specific discount level (i.e. one of the levels 1-6) has no discount after 2 years?

3. *What is the* distribution* of being in one of the levels at time 5 years?

4. *Suppose a large number of people having the same claim probabilities take out policies at the same time. What is the proportion would you expect to be in each discount level in the long run?*

What would be a suitable model to study the NCD system? As opposed to a **deterministic model** *for which its outcomes are fixed, the outcomes of the NCD model are uncertain. It turns out that the NCD system can be studied within the framework of Markov chains, which are examples of stochastic processes. The use of matrix algebra provides a powerful tool to understand and analyse the processes.*

The evolution of the states or levels can be described the random variables $X_0, X_1, X_2, ...$ and probability distributions, where X_n is the level of the policyholder at time n. In this example, the set of all states called the state space is discrete, which consists of seven levels, and the time variable is also discrete. This is an example of a **discrete time, discrete state space stochastic process***.*

Example 1.2. *(**Poisson processes**) Consider the number of claims that occur up to time t (denoted by N_t) from a portfolio of health insurance policies (or other types of insurance products). Suppose that the average rate of occurrence of claims per time unit (e.g. day or week) is given by λ.*

Here are some questions of interest:

1. *On average, 20 claims arrive every day, what is the probability that more than 100 claims arrive within a week?*

2. *What is the expected time until the next claim?*

In this example, the state space consists of all whole numbers $\{0, 1, 2, ...\}$, while the time variable is continuous. The process is a **continuous-time stochastic process with discrete state space**. The model used to model the insurance claims is an example of **Poisson processes**. The Poisson process is one of the most widely-used counting processes. Even thought we know that claims occur at a certain rate, but completely at random. Moreover, the timing between claims seem to be completely random.

Later, we will see that there are several ways to describe this process. One can focus on the number of claims that occur up to time t or the times between those claims when they occur. Many important properties of Poisson processes will be discussed.

Example 1.3. *(**Markov processes**) Suppose that we observe a total of n independent lives all aged between x and x + 1. For life i, we define the following terms:*

- $x + a_i$ *is the age at which observation begins,* $0 \le a_i < 1$.

- $x + b_i$ *is the age at which observation ends, if life does not die,* $0 \le b_i < 1$.

- $x + t_i$ *is the age at which observation stops, by death or censoring.*

- $d_i = 1$*, if life i dies, otherwise $d_i = 0$, if life i censored.*

For example, consider the following mortality data on eight lives all aged between 70 *and* 71.

Life	a_i	b_i	d_i	t_i
1	0	1	1	0.25
2	0	1	1	0.75

Life	a_i	b_i	d_i	t_i
3	0	1	0	1
4	0.1	0.6	1	0.5
5	0.2	0.7	1	0.6
6	0.2	0.4	0	0.4
7	0.5	1	1	0.75
8	0.5	0.75	0	0.75

How would one use this dataset to estimate the probability that a life aged 70 dies before age $70+t$ or survives to at least age $70+t$, for $t \in [0,1)$?

In this example, we can represent the process by $\{X_t\}_{t\geq 0}$ with two possible states (alive or dead). This model is also an example of a **continuous-time stochastic process with discrete state space**.

Here, we illustrate three actuarial applications which can be modelled by some **stochastic processes**. *We should also emphasis that the outcome of one of the above processes is not fixed or uncertain. The course will provide important tools and techniques to analyse the problems with the goal of quantifying the uncertainty in the system.*

Chapter 2

Review of probability theory

2.1 Random variables

The dynamics of a stochastic process are describes by random variables and probability distributions. This section provides a brief discussion of the properties of random variables.

The probability theory is about random variables. Roughly speaking, a random variable can be regarded as an uncertain, numerical quantity (i.e. the value in $\mathbb{R}$) whose possible values depend on the outcomes of a certain random phenomenon. The random variable is usually denoted by a capital letter $X, Y, \ldots$, etc..

More precisely, let S be a sample space. A **random variable** X is a real-valued function defined on the sample space S,

$$X : S \to \mathbb{R}.$$

Hence, the random variable X is a function that maps outcomes to real values.

Example 2.1. *Two coins are tossed simultaneously and the outcomes are HH, HT, TH and TT. We can associate the outcomes of this experiment with the set $A = \{1, 2, 3, 4\}$, where $X(HH) = 1, X(HT) = 2, X(TH) = 3$ and $X(TT) = 4$. Assume each of the outcomes has an equal probability of $1/4$. Here, we can associate a function P (known as a probability measure) defined on $S = \{HH, HT, TH, TT\}$ by

$$P(HH) = 1/4, P(HT) = 1/4, P(TH) = 1/4, P(TT) = 1/4.$$

A probability measure $P : \mathcal{A} \to [0, 1]$, where $\mathcal{A}$ is a collection of subsets of S, has the following properties

1. $0 \leq P(A), \quad A \subset S$.

2. $P(S) = 1$.

3. If $A_i \cap A_j = \emptyset$ for $i, j = 1, 2, \ldots$, and $i \neq j$ where $A_j \subset S$, then

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$$

Random variables can be discrete or continuous. If the range of a random variable is finite or countably infinite, then the random variable is a **discrete random variable**. Otherwise, if its range is an uncountable set, then it is a **continuous random variable**.

2.2   Probability distribution

The probability distribution of a random variable X is a function describing all possible values of X and their corresponding probabilities or the likelihood of obtaining those values of X. Functions that define the probability measure for a discrete or a continuous random variable are the **probability mass function (pmf)** and the **probability density function (pdf)**, respectively.

Suppose X is a discrete random variable. Then the function

$$f(x) = P(X = x)$$

that is defined for each x in the range of X is called the **probability mass function** (p.m.f) of a random variable X.

Suppose X is a continuous random variable with c.d.f F and there exists a nonnegative, integrable function f, $f : \mathbb{R} \to [0, \infty)$ such that

$$F(x) = \int_{-\infty}^{x} f(y)\, dy$$

Then the function f is called the **probability density function** (p.d.f) of a random variable X.

Examples of discrete and continuous random variables

The main quantities of interest in a portfolio of motor insurance are the number of claims arriving in a fixed time period and the sizes of those claims. Clearly, the number of claims can be describe by a discrete random variable, whose range is finite or countably infinite. On the other hand, the claim sizes can be describe by a continuous random variable defined over continuous sample spaces.

Example 2.2. *Let N denote the number of claims which arise up to a given time. The range of all possible values N is $\mathbf{N} \cup \{0\}$. Here N is an example of discrete random variable. We could model the number of claims by the Poisson family of distributions. Recall that a random variable N has a Poisson distribution with the parameter λ if its probability distribution is given by*

$$f(n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad for\ n = 0, 1, \dots.$$

Now suppose further that the number of claims N which arise on a portfolio in a week has a Poisson(λ) where $\lambda = 5$. Calculate the following quantities:

1. $\Pr(N \geq 6)$.

2. $\mathrm{E}[N]$.

1. $\Pr(N \geq 6) = 1 - \Pr(N \leq 5) = 1 - \sum_{n=0}^{5} f(n) = 0.3840393$.

2. Clearly, $\mathrm{E}[N] = \lambda = 5$.

In R, density, distribution function, for the Poisson distribution with parameter λ is shown as follows:

Distribution	Density function: $P(X = x)$	Distribution function: $P(X \leq x)$
Poisson	`dpois(x, lambda, log = FALSE)`	`ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)`

When lower.tail is set to be TRUE (or default), probabilities are $P(X \leq x)$, otherwise, $P(X > x)$.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJwcG9pcyg1LCBsYW1iZGEgPSA1LCBsb3dlci50YWlsID0gRkFMU0UpIn0=

Example 2.3. *Let X denote the claim sizes in a given time period. The range of all possible values X is the set of all non-negative numbers. Here X is an example of a continuous random variable. Suitable families of distributions which could be used to modelled claim sizes are "fat tails" distribution. They allow for possibilities of large claim sizes.*

Examples of fat-tailed distributions include

- *the Pareto distribution,*

- *the Log-normal distribution,*

- *the Weibull distribution with shape parameter greater than 0 but less than 1, and*

- *the Burr distribution.*

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJsaWJyYXJ5KGdncGxvdDIpXG5kZiA8LSBkYXRhLmZyYW1lKHg9c2VxKD

See https://dk81.github.io/dkmathstats_site/rvisual-cont-prob-dists.html for more details.

The course "SCMA 470 Risk Analysis and Credibility" provides more details about the loss distribution.

2.3 Conditional probability

A stochastic process can be defined as a collection or sequence of random variables. The concept of **conditional probability** plays an important role to analyse dependency between random variables in the process. Roughly speaking, conditional probability is the probability of seeing some event knowing that some other event has actually occurred.

Let A and B be two events (elements of $\mathcal{A}$). The conditional probability of event A given B denoted by $P(A|B)$ is defined as

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

Note that $P(A \cap B)$ is often called the joint probability of A and B, and $P(A)$ and $P(B)$ are often called the marginal probabilities of A and B, respectively.

The events A and B are independent if the occurrence of either one of the events does not affect the probability of occurrence of the other. More precisely, the events A and B are independent if

$$P(A \cap B) = P(A)P(B),$$

or equivalently,

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

2.4 Law of total probability

Suppose there are three events: A, B, and C. Events B and C are distinct from each other while event A intersects with both events. We do not know the probability of event A. However, partial information and dependencies between events can be used to calculate the probability of event A, i.e. we know the probability of event A under condition B and the probability of event A under condition C.

The total probability rule states that by using the two conditional probabilities, we can find the probability of event A, which is

$$P(A) = P(A \cap B) + P(A \cap C).$$

In general, suppose $B_1, B_2, \ldots B_n$ be a collection of events that partition the sample space. Then for any event A,

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Example 2.4. *Suppose in a particular study area, the vaccination rate for the yearly flu virus is 70%. Suppose of those vaccinated, 10% of the residents still get the flu that year. Calculate the conditional probability of someone getting the flu in this area given that the person was vaccinated.*

Example 2.5. *You are an invester buying shares of a company. You have discovered that the company is planning to introduce a new project that is likely to affect the company's stock price. You have determined the following probabilities:*

- *There is a 80% probability that the new project will be launched.*

- *If a company launches the project, there is a 85% probability that the company's stock price will increase.*

- *If a company does not launch the project, there is a 30% probability that the company's stock price will increase.*

Calculate the probability that the company's stock price will increase.

2.5   Conditional distribution and conditional expectation

Let X and Y be two discrete random variables with joint probability mass function

$$f(x, y) = P(X = x, Y = y).$$

If X and Y are continuous random variables, the joint probability density function $f(x, y)$ satisfies

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \, du \, dv.$$

When no information is given about the value of Y, the marginal probability density function of X, $f_X(x)$ is used to calculate the probabilities of events concerning X. However, when the value of Y is known, to find such probabilities, $f_{X|Y}(x|y)$, the conditional probability density function of X given that $Y = y$ is used and is defined as follows:

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

provided that $f_Y(y) > 0$. The conditional mass function of X is defined in a similar manner.

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

Note also that the conditional probability density function of X given that $Y = y$ is itself a probability density function, i.e.

$$\int_{-\infty}^{\infty} f_{X|Y}(x|y) \, dx = 1.$$

Note that the conditional probability distribution function of X given that $Y = y$, the conditional expectation of X given that $Y = y$ can be as follows:

$$F_{Y|X}(x|y) = P(X \leq x|Y = y) = \int_{-\infty}^{x} f_{X|Y}(t|y) \, dt$$

and

$$\mathrm{E}(X|Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx,$$

where $f_Y(y) > 0$.

Note that if X and Y are independent, then $f_{X|Y}$ coincides with f_X because

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{f_X(x)f_Y(y)}{f_Y(y)} = f_X(x).$$

2.6 Central Limit Theorem

This section introduces the Central Limit Theorem, which is an important theorem in probability theory. It states that the mean of n independent and identically distributed random variables has an approximate normal distribution given a sufficiently large n. This applies to a collection of random variables from any distribution with a finite mean and variance. In summary we can use the Central Limit Theorem to extract probabilistic information about the sums of independent and identical random variables.

Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables with a finite mean $\mathrm{E}[X_i] = \mu$ and finite variance $\mathrm{Var}[X_i] = \sigma^2$. Let Z_n be the normalised average of the first n random variables

$$\begin{aligned} Z_n &= \frac{\sum_{i=1}^n X_i/n - \mu}{\sigma/\sqrt{n}} \\ &= \frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma\sqrt{n}}. \end{aligned}$$

Then Z_n converges in distribution to a standard normal distribution.

Chapter 3

Stochastic processes

Evolution of a random process is at least partially random, and each run the process leads to potentially a different outcome. It is of great interest to understand or model the behaviour of a random process by describing how different states, represented by random variables X's, evolve in the system over time. Just as probability theory is considered as the study of mathematical models of random phenomena, the theory of stochastic processes plays an important role in the study of time-dependent random phenomena. Stochastic processes can be used to represent many different random phenomena from different fields such as science, engineering, finance, and economics.

A **stochastic process** is a collection of random variables $\{X_t : t \in T\}$ defined on a common sample space, where

- t is a parameter running over some index set T, called the **time domain**.

- The common sample space of the random variables (the range of possible values for X_t) denoted by S is called the **state space** of the process.

1. The set of random variables may be dependent or need not be identically distributed.

2. Techniques used to study stochastic processes depend on whether the state space or the index set (the time domain) are discrete or continuous.

3.1   Classification of stochastic processes

Stochastic processes can be classified on the basis of the nature of their state space and index set.

1. **Discrete state space with discrete time changes** : No claims discount (NCD) policy: A car owner purchases a motor insurance policy for which the premium charged depends on the claim record. Let X_t denote the discount status of a policyholder with three levels of discount, i.e. $S = \{0, 1, 2\}$ corresponding to three discount levels of 0%, 20% and 40% and the time set is $T = \{0, 1, 2, ...\}$. Both time and state space are discrete.

2. **Discrete state space with continuous time changes** : In a health insurance system, an insurance company may classify policyholders as Healthy, Sick or Dead, i.e. $S = \{H, S, D\}$. The time domain can be taken as continuous, $T = [0, \infty)$.

3. **Continuous state space with continuous time changes** : Let S_t be the total amount claimed by time $t \in T$ where $T = [0, \infty)$ and the state space is $\mathbb{R}$. Both time and state space are continuous. Some continuous time stochastic process taking value in a continuous state space will be studied in Risk Analysis and Credibility course.

4. **Continuous state space with discrete time changes** : The outcomes of the above claim process S_t could be recorded continuously, however, we may choose to model the values only at discrete time, for e.g. the total claim amounts at the end of each day. This may due to the limitation of the measurement process (for e.g. expensive to measure). Hence, the time domain is discrete but the state space is continuous.

In case that claim amounts are recorded to the nearest baht or in satang, i.e. discrete state space, we could also approximate or model the process by using a discrete state space, rather than continuous.

3.2 Random walk: an introductory example

One of the simplest examples of a stochastic process is a simple random walk. Consider a simple model of the price of a stock measured in baht. For each trading day $n = 0, 1, 2, ...$, the stock price increases by 1 baht with probability p or decreases by 1 baht with probability $q = 1 - p$. Let X_n denote the stock price at day n and $X_0 = 100$. This simple model is called a simple random walk.

In the simple random walk process, time is discrete (as observed at the end of each day) and the state space is discrete. The stochastic model has an infinite number of outcomes known as **stochastic realisations or sample paths**. A **sample path** is then just the sequence of a particular set of experiments. Graphs of some stochastic realisations of the simple random walk with $p = 0.5$ and $a = 100$ are shown in Figure 1.
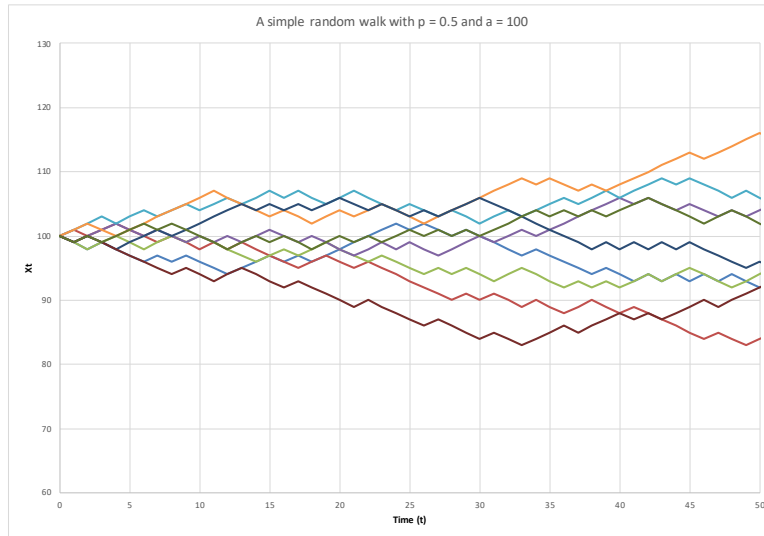

Figure 3.1: Some stochastic realisations of the simple random walk

We can use R to generate sample paths of this random walk.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJsaWJyYXJ5KHRpZHlKVxuc2V0LnNlZWQoMSlcbiAgPC0gMC41XG5pPC

A complete description of the simple random walk, observed as a collection of n random variables at time points $t_1, t_2, ..., t_n$ can be specified by the joint distribution of $X_{t_1}, X_{t_2}, ..., X_{t_n}$, i.e.

$$F(x_1, x_2, ..., x_n) = \Pr(X_{t_1} \leq x_1, X_{t_2} \leq x_2, ..., X_{t_n} \leq x_n).$$

However, the **multidimensional distribution function cannot be easily written in a closed form** unless the random variables have a multivariate normal distribution. In practice, it is more convenient to deal with some stochastic processes via some simple **intermediary processes** or under some addition assumptions.

In general, a simple random walk X_n is a discrete-time stochastic process defined by

- $X_0 = a$ and

- for $n \geq 1$,

$$X_n = a + \sum_{i=1}^{n} Z_i, \text{ where } Z_i = \begin{cases} 1, & \text{with probability } p \\ -1, & \text{with probability } q = 1 - p. \end{cases}$$

1. When $p = 1/2$, the value of the process increases or decreases randomly by 1 unit with equal probability. In this case, the process is known as a **symmetric** random walk.

2. The (intermediary) process $\{Z_i : i \in \mathbb{N}\}$ is a sequence of independent identically distributed (i.i.d.) random variables. The process X_t themselves are neither independent nor identically distributed. This process Z_i is also known as **white noise process**.

Example 3.1. *Explain why the simple random process X_n is neither independent nor identically distributed.*

Suppose that $X_0 = 100$. Firstly, we will show that X_n is not independent. From definition, the process X_n can be written as

$$X_n = X_{n-1} + Z_n.$$

That is, the value of X_n is the previous value plus a random change of 1 or -1. Therefore, the value of the process depends on the previous value and they are not independent. For e.g.,

$$\Pr(X_2 = 102) > 0 \quad \text{but} \quad \Pr(X_2 = 102 | X_1 = 99) = 0.$$

The process X_n cannot be identically distributed. For e.g. X_1 can take the values of 99 and 101, while X_2 can take three different values of 98, 100 and 102.

Example 3.2. *Let*

$$\mu = \mathrm{E}[Z_i]$$
$$\sigma^2 = \mathrm{Var}[Z_i]$$

Calculate the expectation (μ) and variance (σ^2) of the random variable Z_i.

$$\mu = \mathrm{E}[Z_i] = 1 \cdot p + (-1) \cdot q = p - q.$$

$$\begin{aligned} \sigma^2 &= \mathrm{Var}[Z_i] \\ &= \mathrm{E}[Z_i^2] - (\mathrm{E}[Z_i])^2 \\ &= 1 - (p - q)^2 = (p + q)^2 - (p - q)^2 \\ &= 4pq. \end{aligned}$$

Example 3.3. *Calculate the expectation and variance of the process X_n at time n.*

$$\mathrm{E}[X_n] = \mathrm{E}[a + \sum_{i=1}^{n} Z_i] = a + n\mu.$$

$$\mathrm{Var}[X_n] = \mathrm{Var}[a + \sum_{i=1}^{n} Z_i] = n\sigma^2.$$

It should be noted that the variance of X_n increases with time.

Example 3.4. *For the random process, calculate*

$$\Pr(X_2 = 98, X_5 = 99 | X_0 = 100).$$

The process X_n must decrease on the first two days, which happens with probability $(1-p)^2$. Independently, it must then increases on another two days and decrease on one day (not necessarily in that order), giving three different possibilities. Each of these has probability $p^2(1-p)$. So

$$\Pr(X_2 = 98, X_5 = 99 | X_0 = 100) = (1-p)^2 \cdot 3p^2(1-p) = 3p^2(1-p)^3.$$

In what follows, we will see that exact calculations are possible for the simple random walk process. Note also that it is sufficient to understand the behaviour of the random walk when it starts at $X_0 = 0$.

Example 3.5. *For the random process with $X_0 = 0$, X_{2n} is always even and X_{2n+1} is always odd. Based on the binomial distribution, show that*

$$\Pr(X_{2n} = 2m) = \binom{2n}{n+m} p^{n+m} q^{n-m}, \quad -n \le m \le n$$

$$\Pr(X_{2n+1} = 2m+1) = \binom{2n+1}{n+m+1} p^{n+m+1} q^{n-m}, \quad -n-1 \le m \le n.$$

Let A denote the number of $+1$ and B denote the number of -1. Then $A + B = 2n$ and $X_{2n} = A - B$ (i.e. the position at time $2n$). Hence,

$$\begin{aligned}
\Pr(X_{2n} = 2m) &= \Pr(A - B = 2m) \\
&= \Pr(A - (2n - A) = 2m) = \Pr(2A - 2n = 2m) = \Pr(A = m + n) \\
&= \binom{2n}{n+m} p^{n+m} q^{2n-(n+m)}, \quad -n \le m \le n \\
&= \binom{2n}{n+m} p^{n+m} q^{n-m}, \quad -n \le m \le n.
\end{aligned}$$

->

->

Chapter 4

Discrete-time Markov chains

Recall the simple random walk model of the price of the stock. Suppose the stock price for the first four days are

$$(X_0, X_1, X_2, X_3) = (100, 99, 98, 99).$$

Based on this past information, what can we say about the price at day 4, X_4? Although, we completely know the whole past price history, the only information relevant for predicting their future price is the price on the previous day, i.e. X_3. This means that

$$\Pr(X_4 = j | X_0 = 100, X_1 = 99, X_2 = 98, X_3 = 99) = \Pr(X_4 = j | X_3 = 99).$$

Given the current price X_3, the price X_4 at day 4 is independent of the history prices X_0, X_1, X_2. The sequence of stock prices $X_0, X_1, \ldots, X_n$ is an example of a **Markov chain**.

A Markov process is a special type of stochastic processes with the property that the future evolution of the process depends only on its current state and not on its past history. That is given the value of X_t, the values of X_s for $s > t$ do not depend on the values of X_u for $u < t$. This property is called the **Markov property**.

A **discrete-time Markov chain** is a discrete-time stochastic process that satisfies the Markov property:

$$\Pr(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \Pr(X_{n+1} = j | X_n = i),$$

for all time points n and all states $i_0, i_1, \ldots, i_{n-1}, i, j$. It is convenient to assume that the state space of the Markov chain is a subset of non-negative integers, i.e. $S \subseteq \{0, 1, \ldots\}$.

Example 4.1. *A process with independent increments has the Markov property.*

Recall the following definitions. An increment of a process is the amount by which its value changes over a period of time, for e.g. $X_{t+u} - X_t$ where $u > 0$.

A process X_t is said to have independent increments if for all t and every $u > 0$, the increment $X_{t+u} - X_t$ is independent of all the past of the process $\{X_s : 0 \le s \le t\}$).

In order to show that a process with independent increments has the Markov property, we proceed as follows:

$$\begin{aligned}
\Pr(X_t \in A | X_{s_1} = x_1, X_{s_2} = x_2, \ldots, X_s = x) &= \Pr(X_t - X_s + x \in A | X_{s_1} = x_1, X_{s_2} = x_2, \ldots, X_s = x) \\
&= \Pr(X_t - X_s + x \in A | X_s = x)(\text{by independence of the past}) \\
&= \Pr(X_t \in A | X_s = x).
\end{aligned}$$

The random walk process has the Markov property.

4.1   One-step transition probabilities

The conditional probability that X_{n+1} is in state j given that X_n is in state i is called **one-step transition probability** and is denoted by

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

Note that the transition probabilities depend not only on the current and future states, **but also on the time of transition** n.

If the transition probabilities $p_{ij}^{n,n+1}$ in a Markov chain do not depend on time n, the Markov chain is said to be **time-homogeneous or stationary or simply homogeneous**. Then

$$p_{ij}^{n,n+1} = \Pr(X_{n+1} = j | X_n = i) = \Pr(X_1 = j | X_0 = i) = p_{ij}.$$

Otherwise, it is said to be **nonstationary** or **nonhomogeneous**.

Unless stated otherwise, it shall be assumed that the Markov chain is stationary. The matrix P whose elements are p_{ij} is called the **transition probability matrix** of the process.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Note that the elements of the matrix P satisfies the following properties:

$$0 \le p_{ij} \le 1, \quad \text{and} \quad \sum_{j \in S} p_{ij} = 1,$$

for all $i, j \in S$. A matrix that satisfies these properties is called a **stochastic matrix**.

Example 4.2. *No claims discount (NCD) policy: Let X_n be the discount status of a policyholder at time n. There are three levels of discount, i.e. $S = \{0, 1, 2\}$ corresponding to three discount levels of 0, 20% and 40%. The following rules are assumed:*

- *For a claim-free year, the policyholder moves up a level or remains in state 2 (the maximum discount state).*

- *If there is at least one claim, the policyholder moves down one level or remains in state 0.*

Suppose also that the probability of a claim-free year is p and is independent of n. The transition probability matrix is given by

$$P = \begin{bmatrix} 1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 1-p & p \end{bmatrix}.$$

The transition diagram is illustrated in the following figure.

The following questions are of interest.

1. *What is the probability*

$$\Pr(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n)?$$

2. *What is the probability*

$$\Pr(X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)?$$

3. *What is the probability of transferring from state i to state j in n steps*

$$\Pr(X_{m+n} = j | X_m = i)?$$

4. *What is the long-term behavior of the Makov chain, i.e.* $\lim_{n \to \infty} \Pr(X_n = j), j = 0, 1, 2$ *given that* $\Pr(X_0 = 0)$.

Later we will apply matrix algebra to compute these types of probabilities and long-term probabilities.

Example 4.3. *In a health insurance system, at the end of each day an insurance company classifies policy-holders as Healthy, Sick or Dead, i.e. $S = \{H, S, D\}$. The following transition matrix P for a healthy-sick-dead model is given by*

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix}.$$

The transition diagram is shown below.

It turns out that the probabilistic description of the Markov chain is completely determined by its transition probability matrix and its initial probability distribution X_0 at time 0.

Example 4.4. *By using the definition of conditional probabilities, show that*

$$\Pr(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \mu_{i_0} \, p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \, p_{i_{n-1} i_n},$$

where $\mu = \mu^{(0)}$ is the distribution of initial random variable X_0, i.e. $\mu_i = \Pr(X_0 = i)$ (the probability mass function of X_0.)

$$\begin{aligned}
&\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) \\
&= \Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \cdot \Pr(X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \\
&= \Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \cdot \Pr(X_n = i_n | X_{n-1} = i_{n-1}) \\
&= \Pr(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) \cdot p_{i_{n-1} i_n} \\
&\vdots \\
&= \mu_{i_0} \, p_{i_0 i_1} \cdots p_{i_{n-2} i_{n-1}} \, p_{i_{n-1} i_n}.
\end{aligned}$$

Example 4.5. *Show that*

$$\begin{aligned}
\Pr(X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\
= \Pr(X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_n = i_n) \\
= p_{i_n i_{n+1}} \cdots p_{i_{n+m-2} i_{n+m-1}} \, p_{i_{n+m-1} i_{n+m}}.
\end{aligned}$$

$$\begin{aligned}
&\Pr(X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\
&= \Pr(X_{n+1} = i_{n+1}, X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_n = i_n) \\
&= \Pr(X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_{n+1} = i_{n+1}, X_n = i_n) \cdot \Pr(X_{n+1} = i_{n+1} | X_n = i_n) \\
&= \Pr(X_{n+2} = i_{n+2}, \dots, X_{n+m} = i_{n+m} | X_{n+1} = i_{n+1}) \cdot p_{i_n i_{n+1}} \\
&\vdots \\
&= \Pr(X_{n+m} = i_{n+m} | X_{n+m-1} = i_{n+m-1}) \cdots \Pr(X_{n+2} = i_{n+2} | X_{n+1} = i_{n+1}) \cdot p_{i_n i_{n+1}} \\
&= p_{i_{n+m-1} i_{n+m}} \cdot p_{i_{n+m-2} i_{n+m-1}} \cdots p_{i_n i_{n+1}}.
\end{aligned}$$

More general probabilities of the possible realisations of the process can be calculated by summing the probabilities of elementary elements of these forms.

Example 4.6. *For the NCD system defined on the state space $S = \{0, 1, 2\}$ as given in ExampleÂ Example Â 2, suppose that the probability of a claim-free year $p = 3/4$, and the distribution of the initial discount rate $\mu = (0.5, 0.3, 0.2)$. Find the following:*

1. $\Pr(X_0 = 2, X_1 = 1, X_2 = 0)$.

2. $\Pr(X_1 = 1, X_2 = 0 | X_0 = 2)$.

3. $\Pr(X_{10} = 2, X_{11} = 1, X_{12} = 0)$.

4. $\Pr(X_{11} = 1, X_{12} = 0 | X_{10} = 2)$.

The corresponding transition matrix is

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 1/4 & 0 & 3/4 \\ 0 & 1/4 & 3/4 \end{bmatrix}.$$

1. Denote $\mu = (\mu_1, \mu_2, \mu_3) = (0.5, 0.3, 0.2)$

$$
\begin{aligned}
\Pr(X_0 = 2, X_1 = 1, X_2 = 0) &= \Pr(X_0 = 2, X_1 = 1) \cdot \Pr(X_2 = 0 | X_0 = 2, X_1 = 1) \\
&= \Pr(X_0 = 2, X_1 = 1) \cdot \Pr(X_2 = 0 | X_1 = 1) \quad \text{(by Markov property)} \\
&= \Pr(X_0 = 2) \cdot \Pr(X_1 = 1 | X_0 = 2) \cdot \Pr(X_2 = 0 | X_1 = 1) \quad \text{(again by Markov property)} \\
&= \mu_3 p_{32} p_{21} = 0.2 \cdot (1/4) \cdot (1/4) = 1/80.
\end{aligned}
$$

   Alternatively, it follows from ExampleÂ ExampleÂ 5 that,

$$\Pr(X_0 = 2, X_1 = 1, X_2 = 0) = \mu_3 p_{32} p_{21} = 0.2 \cdot (1/4) \cdot (1/4) = 1/80.$$

2.

$$
\begin{aligned}
\Pr(X_1 = 1, X_2 = 0 | X_0 = 2) &= \Pr(X_1 = 1 | X_0 = 2) \cdot \Pr(X_2 = 0 | X_1 = 1, X_0 = 2) \\
&= \Pr(X_1 = 1 | X_0 = 2) \cdot \Pr(X_2 = 0 | X_1 = 1) \quad \text{(by Markov property)} \\
&= p_{32} p_{21} = (1/4) \cdot (1/4) = 1/16.
\end{aligned}
$$

3. Following conditional probability, the Markov property, and time-homogeneity (to be discussed later) results in

$$
\begin{aligned}
\Pr(X_{10} = 2, X_{11} = 1, X_{12} = 0) &= \Pr(X_{10} = 2) \cdot \Pr(X_{11} = 1, X_{12} = 0 | X_{10} = 2) \\
&= \Pr(X_{10} = 2) \cdot \Pr(X_{11} = 1 | X_{10} = 2) \cdot \Pr(X_{12} = 0 | X_{10} = 2, X_{11} = 1) \\
&= \Pr(X_{10} = 2) \cdot \Pr(X_{11} = 1 | X_{10} = 2) \cdot \Pr(X_{12} = 0 | X_{11} = 1) \\
&= \Pr(X_{10} = 2) \cdot \Pr(X_1 = 1 | X_0 = 2) \cdot \Pr(X_1 = 0 | X_0 = 1) \\
&= \Pr(X_{10} = 2) \cdot p_{32} p_{21} = 0.6922 \cdot (1/4) \cdot (1/4) = 0.0433.
\end{aligned}
$$

   Later we will show that $\Pr(X_{10} = 2) = (\mu P^{10})_3 = 0.6922$ (here $(\mu P^{10})_i$ denotes the i-th entry of the vector μP^{10}.

4. From conditional probability, the Markov property, and time-homogeneity, it follows that

$$
\begin{aligned}
\Pr(X_{11} = 1, X_{12} = 0 | X_{10} = 2) &= \Pr(X_{11} = 1 | X_{10} = 2) \cdot \Pr(X_{12} = 0 | X_{11} = 1, X_{10} = 2) \\
&= \Pr(X_{11} = 1 | X_{10} = 2) \cdot \Pr(X_{12} = 0 | X_{11} = 1) \quad \text{(by Markov property)} \\
&= \Pr(X_1 = 1 | X_0 = 2) \cdot \Pr(X_1 = 0 | X_0 = 1) \quad \text{(by time-homogeneity)} \\
&= p_{32} p_{21} = (1/4) \cdot (1/4) = 1/16.
\end{aligned}
$$

4.2 The Chapman-Kolmogorov equations and n-step transition probabilities

The n-step transition probability denoted by $p_{ij}^{(n)}$ is the probability that the process goes from state i to state j in n transitions, i.e.

$$p_{ij}^{(n)} = \Pr(X_{m+n} = j | X_m = i).$$

Note that for homogeneous process, the left hand side of the above equation does not depend on m. Suppose that the transition from state i at time m to state j at time $m + n$ (i.e. in n steps), going via state k in l steps. One needs to examine all possible paths (from i to k and then k to j) and hence the n-step transition probability $p_{ij}^{(n)}$ can be expressed as the sum of the product of the transition probabilities $p_{ik}^{(l)} p_{kj}^{(n-l)}$.

$$p_{ij}^{(n)} = \sum_{k \in S} p_{ik}^{(l)} p_{kj}^{(n-l)}, \quad 0 < l < n$$

To derive

$$Chapman$$

, we proceed as follows:

$$
\begin{aligned}
p_{ij}^{(n)} &= \Pr(X_n = j | X_0 = i) \\
&= \sum_{k \in S} \Pr(X_n = j, X_l = k | X_0 = i) \\
&= \sum_{k \in S} \Pr(X_n = j | X_l = k, X_0 = i) \cdot \Pr(X_l = k | X_0 = i) \\
&= \sum_{k \in S} \Pr(X_n = j | X_l = k) \cdot \Pr(X_l = k | X_0 = i) \\
&= \sum_{k \in S} p_{kj}^{(n-l)} p_{ik}^{(l)}, = \sum_{k \in S} p_{ik}^{(l)} p_{kj}^{(n-l)}, \quad 0 < l < n.
\end{aligned}
$$

This result is known as the Chapman-Kolmogorov equation. This relation can be expressed in terms of matrix multiplication as

$$P^n = P^l P^{n-l}.$$

The n-**step transition probabilities** $p_{ij}^{(n)}$ are the ij elements of P^n.

Example 4.7. *For the NCD system given in ExampleÂ ExampleÂ 2, suppose that $p = 3/4$, the probability of a claim-free year and the initial discount level of a policyholder is 1 (with 20% discount).*

1. *Calculate the probability of starting with a discount level of 20% and ending up 3 years later at the same level.*

2. *Calculate the policyholder's expected level of discount after 3 years.*

The transition matrix is

$$P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 1/4 & 0 & 3/4 \\ 0 & 1/4 & 3/4 \end{bmatrix},$$

and

$$P^3 = \begin{bmatrix} 7/64 & 21/64 & 9/16 \\ 7/64 & 3/16 & 45/64 \\ 1/16 & 15/64 & 45/64 \end{bmatrix}.$$

1. The probability of starting with a discount level of 20% and ending up 3 years later at the same level is equal to $p_{11}^{(3)} = 3/16$, which is the element in the second row and second column of the matrix P^3 (not to be confused with the indices used) .

2. The policy's expected level of discount after 3 years is

$$E[X_3|X_0 = 1] = \sum_{j=0}^{2} j \cdot \Pr(X_3 = j|X_0 = 1)$$
$$= 0 \cdot (7/64) + 1 \cdot (3/16) + 2 \cdot (45/64)$$
$$= 51/32 = 1.59375.$$

4.3   Distribution of X_n

Let $\mu^{(n)}$ be the vector of probability mass function of X_n, i.e.

$$\mu^{(n)} = (\mu_1, \mu_2, ...),$$

where $\mu_i = \Pr(X_n = i)$. It follows that

$$\mu^{(n+1)} = \mu^{(n)} P$$

and, in general,

$$\mu^{(n+m)} = \mu^{(n)} P^m.$$

Example 4.8.   1. *Show that*

$$Pr(X_1 = i) = \sum_{k \in S} \mu_k p_{ki} = (\mu P)_i,$$

which is the ith element of the vector μP. Here $\mu = \mu^0$ is the distribution of initial random variable X_0 with $\mu_i = \Pr(X_0 = i)$.

2. *In general, show that the distribution of X_n is given by*

$$Pr(X_n = i) = (\mu P^n)_i.$$

1.

$$\Pr(X_1 = i) = \sum_{k \in S} \Pr(X_1 = i|X_0 = k) \cdot \Pr(X_0 = k)$$
$$= \sum_{k \in S} \mu_k \cdot p_{ki}$$
$$= (\mu P)_i.$$

Example 4.9. *The simple weather pattern can be classified into three types including rainy (R), cloudy (C) and sunny (S). The weather is observed daily. The following information is provided.*

- *On any given rainy day, the probability that it will rain the next day is 0.7; the probability that it will be cloudy the next day 0.2.*

- *On any given cloudy day, the probability that it will rain the next day is 0.75; the probability that it will be sunny the next day 0.1.*

- *On any given sunny day, the probability that it will rain the next day is 0.2; the probability that it will be sunny the next day 0.4.*

The weather forecast for tomorrow shows that there is a 40% chance of rain and a 60% chance of cloudy. Find the probability that it will sunny 2 days later.

As the ordered state of the chain is R, C, S, the initial distribution is $\mu = (0.4, 0.6, 0)$. The transition matrix P is given by

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.75 & 0.15 & 0.1 \\ 0.2 & 0.4 & 0.4 \end{bmatrix},$$

and

$$P^2 = \begin{bmatrix} 0.66 & 0.21 & 0.13 \\ 0.6575 & 0.2125 & 0.13 \\ 0.52 & 0.26 & 0.22 \end{bmatrix}.$$

This gives

$$\mu \cdot P^2 = (0.6585, 0.2115, 0.13).$$

Hence, the desired probability of sunny is

$$\Pr(X_2 = S) = (\mu \cdot P^2)_S = (\mu \cdot P^2)_3 = 0.13.$$

4.4   Joint Distribution

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix P and initial distribution μ. For all $0 \le n_1 \le n_2 < \cdots < n_{k-1} < n_k$ and states $i_1, i_2, \ldots, i_{k-1}, i_k$,

$$P(X_{n_1} = i_1, X_{n_2} = i_2, \ldots, X_{n_k} = i_k 1, X_{n_k} = i_k) = (\mu P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1 i_2} \cdots (P^{n_k - n_{k-1}})_{i_{k-1} i_k}.$$

From the above result, the joint probability is obtained from just the initial distribution μ and the transition matrix P.

Example 4.10. *In ExampleÂ ExampleÂ 9, on Sunday, the chances of rain, cloudy and sunny have the same probabilities. Find the probability that it will be sunny on the following Wednesday and Friday, and cloudy on Saturday.*

We are given that $\mu = (1/3, 1/3, 1/3)$. From

$$P^3 = \begin{bmatrix} 0.645500 & 0.215500 & 0.139 \\ 0.645625 & 0.215375 & 0.139 \\ 0.603000 & 0.231000 & 0.166 \end{bmatrix},$$

the required probability is

$$\begin{aligned}
\Pr(X_3 = S, X_5 = S, X_6 = C) &= (\mu \cdot P^3)_S \cdot P_{SS}^2 \cdot P_{SC} \\
&= (\mu \cdot P^3)_3 \cdot P_{33}^2 \cdot P_{32} \\
&= 0.148 \cdot 0.22 \cdot 0.4 = 0.013024.
\end{aligned}$$

4.5   Random walk with absorbing and reflecting barrier(s)

Example 4.11. *A one-dimensional random walk $\{X_n\}$ is defined on a finite or infinite subset of integers in which the process in state i can either stay in i or move to its neighbouring states $i-1$ and $i+1$. Suppose that given that $X_n = i$ at time n,*

- *the probability of moving to state $i+1$ is p_i,*

- *the probability of remaining in state i is r_i, and*

- *the probability of moving to state $i-1$ is q_i,*

where $p_i + q_i + r_i = 1$ for all i.

1. *Write down the transition matrix.*

2. *Show that the random walk has Markov property.*

1. The transition diagram and the transition matrix are infinite:

$$P = \begin{bmatrix} \ddots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & q_{-1} & r_{-1} & p_{-1} & \cdots & \cdots & \cdots \\ \cdots & \cdots & q_0 & r_0 & p_0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & q_1 & r_1 & p_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \ddots \end{bmatrix}.$$

2. Clearly, the Markov property holds because

$$\Pr(X_{n+1} = k | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0)$$

$$= \Pr(X_{n+1} = k | X_n = i) = \begin{cases} p_i, & k = i+1 \\ r_i, & k = i \\ q_i, & k = i-1 \\ 0, & \text{otherwise.} \end{cases}$$

Example 4.12. *The random walk can be used to model the fortune of a gambler. The gambler bets per game and the probability of winning is p and the probability of losing is q where $p + q = 1$. In addition, the gambler is ruined (or goes broke) if he reaches state 0, and also stops the game if he reaches state N. Therefore, the state space is $S = \{0, 1, \dots, N\}$. Note that*

$$p_{00} = 1 \ and \ p_{NN} = 1.$$

*The states 0 and N are referred to as **absorbing boundaries (absorbing states)** and the remaining states $1, 2, \dots, N-1$ are **transient**. Roughly speaking, if a state is known as transient if there is a possibility of leaving the state and never returning.*

The transition diagram and the transition matrix of the simple random walk with absorbing boundaries (states) are given as follows:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ q & 0 & p & 0 & \cdots & 0 & 0 & 0 \\ 0 & q & 0 & p & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

1. In general, a state i is called **absorbing** if $p_{ii} = 1$

2. The simple random walk as given in Example ExampleÂ 12 can be modified so that whenever the process is in state 0 (or state N),

   - the probability of remaining in state 0 is α, and

   - the probability of moving to state 1 is $1 - \alpha$.

   In this case, the state 0 is referred to as a **reflecting barrier** for the chain. The process might be used to model the fortune of an individual when negative fortune is reset to zero.

4.6   An example of nonhomogeneous Markov chain

In this section, we give an example of a discrete-time nonhomogeneous Markov chain. Again, without stated otherwise, we shall assume that the discrete-time Markov chains are homogeneous.

Example 4.13. *(Adapted from W.J.Stewart)*

$$exampleStationary2$$

A Markov chain $X_0, X_1, \ldots$ consists of two states $\{1, 2\}$. At time step n, the probability that the Markov chain remains in its current state is given by

$$p_{11}(n) = p_{22}(n) = 1/n,$$

while the probability that it changes state is given by

$$p_{12}(n) = p_{21}(n) = 1 - 1/n.$$

1. *Draw a transition diagram of the Markov chain.*

2. *Write down the transition matrix.*

3. *Calculate $\Pr(X_5 = 2, X_4 = 2, X_3 = 1, X_2 = 1 | X_1 = 1)$.*

1. The transition diagram and the transition matrix are dependent of the time step n, and are given as follows:

$$P(n) = \begin{bmatrix} \frac{1}{n} & \frac{n-1}{n} \\ \frac{(n-1)}{n} & \frac{1}{n} \end{bmatrix}.$$

2. The probability of taking a particular part can be calculated by

$$\Pr(X_5 = 2, X_4 = 2, X_3 = 1, X_2 = 1 | X_1 = 1) = p_{11}(1) \cdot p_{11}(2) \cdot p_{12}(3) \cdot p_{22}(4)$$
$$= 1 \cdot 1/2 \cdot 2/3 \cdot 1/4 = 1/12.$$

Other paths lead to state 2 after four transitions, and have different probabilities according to the route they follow. What is important is that, no matter which route is chosen, once the Markov chain arrives in state 2 after four steps, the future evolution is specified by $P(5)$, and not any other $P(i), i \leq 4$.

4.7 Simulation

Simulation is a powerful tool for studying Markov chains. For many Markov chains in real-world applications, state spaces are large and matrix methods may not be practical.

A Markov chain can be simulated from an initial distribution and transition matrix. To simulate a Markov sequence $X_0, X_1, \ldots$, simulate each random variable sequentially conditional on the outcome of the previous variable. That is, first simulate X_0 according to the initial distribution. If $X_0 = i$, then simulate X_1 from the $i - th$ row of the transition matrix. If $X_1 = j$, then simulate X_2 from the $j - th$ row of the transition matrix, and so on.

```
[frame=single, escapeinside={(*}{*)}, caption={Algorithm for Simulating a Markov Chain}]

Input: (i) initial distribution  (*$\boldsymbol{\mu}$*), (ii) transition matrix (*$P$*), (iii) number o:
Output: (*$X_0, X_1, \ldots , X_n$*)
Algorithm:
    Generate (*$X_0$*) according to (*$\boldsymbol{\mu}$*)
    FOR (*$i = 1, \ldots ,n$*)
        Assume that (*$X_{i-1} = j$*)
        Set (*$\boldsymbol p = j-$*)th row of (*$P$*)
        Generate (*$X_i$*) according to (*$\boldsymbol p$*)
    END FOR
```

4.8   Monte Carlo Methods

Monte Carlo methods are simulation-based algorithms that rely on generating a large set of samples from a statistical model to obtain the behaviour of the model and estimate the quantities of interest. For a large sample set of a random variable representing a quantity of interest, the law of large numbers allows to approximate the expectation by the average value from the samples.

Consider repeated independent trials of a random experiment. We will need to generate a large number of samples $X_1, X_2, ...$ from the model. A Monte Carlo method for estimating the expectation $\mathrm{E}(X)$ is a numerical method based on the approximation

$$\mathrm{E}(X) \approx \frac{1}{N} \sum_{i=1}^{N} X_i,$$

where $X_1, X_2, ...$ are i.i.d. with the same distribution as X.

While computing expectations and computing probabilities at first look like different problems, the latter can be reduced to the former: if X is a random variable, we have

$$\Pr(X \in A) = \mathrm{E}(1_A(X)).$$

Using this equality, we can estimate $\Pr(X \in A)$ by

$$\Pr(X \in A) = \mathrm{E}(1_A(X)) = \frac{1}{N} \sum_{i=1}^{N} 1_A(X_i).$$

Recall that the indicator function of the set A is the defined as

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

The following user-defined function in Excel can be used to simulate random numbers from a discrete distribution.

```
[frame=single, escapeinside={(*}{*)}, caption={A user-defined function in Excel to simulate random numb
}]
Public Function Discrete(value As Variant, prob As Variant)

Dim i As Integer
Dim cumProb As Single
Dim uniform As Single
Randomize
'Randomize Statement
'Initializes the random-number generator.

Application.Volatile

' This example marks the user-defined function Discrete as volatile.
' The function will be recalculated when any cell in any workbook
' in the application window changes value worksheet.

uniform = Rnd
cumProb = prob(1)
i = 1
```

```
Do Until cumProb > uniform
    i = i + 1
    cumProb = cumProb + prob(i)
Loop
Discrete = value(i)

End Function
```

Example 4.14. *(R or Excel) A gambler starts with and plays a game where the chance of winning each round is 60%. The gambler either wins or loses on each round. The game stops when the gambler either gains or goes bust.*

1. *Develop an Excel worksheet or create an R code to simulate 50 steps of the finite Markov chain of the random walk X_n given in ExampleÂ ExampleÂ 12. Repeat the simulation 10 times. How many of your simulations end at 0.*

2. *Use the results from the simulations to estimate the mean and variance of X_5.*

3. *Use the results from the simulations to estimate the probability that the gambler is eventually ruined.*

Example 4.15. *(R or Excel) A Markov chain $X_0, X_1, \ldots$ on states $\{1, 2\}$ has the following transition matrix*

$$P = \begin{bmatrix} 1 - a & a \\ b & 1 - b \end{bmatrix},$$

where $0 < a, b < 1$.

1. *Use either Excel or R to estimate the long-term distribution of the Markov chain. (Hint: consider the n-step transition matrix for several increasing values of n). Comments on the results obtained.*

   Later we will see that in many cases, a Markov chain exhibits a long-term limiting behaviour. The chain settles down to an equilibrium distribution, which is independent of its initial state.

2. *Use simulations to estimate the long-term probability that a Markov chain hits each of the states. (Hint: simulate the Markov chain 1000 steps and calculate the proportion of visits to each state)*

4.9 Classification of states

Throughout this section $\{X_n\}_{n \geq 0}$ is a time homogeneous Markov chain with state space S and transition matrix $P = (p_{ij})_{i,j \in S}$.

For any $i, j \in S$,

- The state j can be **reached** from the state i, denoted by $i \to j$ if there is a nonzero probability $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

- The states i and j are said to **communicate**, or to be **in the same class**, and denoted by $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

Note that $i \to j$ if and only if there exist states $k_1, k_2, \ldots, k_r$ such that

$$p_{ik_1} p_{k_1 k_2} \cdots p_{k_r j} > 0,$$

i.e. it is not necessary that j can be reached from the state i in one single step.

- The relation $\leftrightarrow$ is an equivalence relation and partition the state space S into equivalence classes, which are known as **classes (or communication classes)** of the Markov chain. Thus in any class all the states communicate, but none of them communicates with any state outside the class.

Additional properties for a communication class are defined as follows:

- The class C is said to be **closed** if it is impossible to reach any state outside C from any state in C, i.e. escape from C is impossible. Otherwise, the class C is said to be **non-closed**, i.e. escape from C is possible.

- If the entire state space S is only one communication class (all states communicate), then it is necessarily closed and the Markov chain is said to be **irreducible**. Otherwise, the Markov chain is said to be **reducible**.

- A closed class consisting of a single state is an **absorbing state**.

Example 4.16. *Consider each of the following Markov chains:*

(a) *NCD system (Example ExampleÂ 2),*

(b) *the health insurance system (Example ExampleÂ 3), and*

(c) *a simple random walk (Example ExampleÂ 12),*

Identify the communication classes. Is the Markov chain irreducible?

It is a good practice to use transition diagram and also verify the answers.

(a) Every two states are intercommunicating, so $\{0, 1, 2\}$ is a single closed class, and hence the Markov chain is irreducible. This is because $0 \to 1$ and $1 \to 0$ ($p_{01} > 0$ and $p_{10} > 0$), and $1 \to 2$ and $2 \to 1$ ($p_{12} > 0$ and $p_{21} > 0$).

(b) There are two classes of intercommunicating states, $O = \{H, S\}$ is non-closed, and $C = \{D\}$ is closed (and also an absorbing state). Clearly, the Markov chain is not irreducible. This is because

  - $p_{DD} = 1$, i.e. C is a class.

  - O is open class because $p_{HS} > 0$, $p_{SH} > 0$, so this is a class, and for example $p_{HD} > 0$ but $p_{DH}^{(n)} = 0$ for all n (i.e. one cannot leave C starting from the state D), so O is an open class.

(c) The simple random walk with absorbing boundaries has three classes, $\{1, 2, ..., N - 1\}$ is non-closed class, $\{0\}$ and $\{N\}$ are two closed classes.

Example 4.17. *A Markov chain with state space $S = \{1, 2, 3, 4, 5\}$ has the following transition matrix:*

$$
P = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\
1/3 & 1/3 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1/2 & 1/2
\end{bmatrix}.
$$

1. *Draw a transition diagram.*

2. *Identify the communication classes. Is the Markov chain irreducible?*

There are two closed classes $C_1 = \{1\}$ and $C_2 = \{4, 5\}$ and one non-closed class $O = \{2, 3\}$. This is because

- C_1 is closed because $p_{11} = 1$.

- C_2 is a class because $4 \to 5$ ($p_{45} > 0$) and $5 \to 4$ ($p_{54} > 0$), and is closed because $p_{ij} = 0$ for all $i \in C_2$ and $j \notin C_2$.

- O is a class because $2 \to 3$ ($p_{23} > 0$) and $3 \to 2$ ($p_{32} > 0$), and is non-closed because $p_{21} > 0$, but $p_{11} = 1$.

4.10 Absorption probabilities and expected time to absorption

For the random walk with absorbing boundaries (i.e. 0 and N), two questions arises, in which state, 0 or N is the process eventually absorbed (or trapped) and on the average how long does it take to reach one of these absorbing states? We first define the following terms which applies to the random walk process with absorbing boundaries.

The time of absorption T is defined as

$$T = \min\{n \geq 0 | X_n = 0 \text{ or } X_n = N\}$$

and the **probability of eventually absorption** in state 0 is given by

$$u_i = \Pr\{X_T = 0 | X_0 = i\}, \text{ for } i = 1, 2, \dots, N-1.$$

The **mean time to absorption** of the process is given by

$$\mathrm{E}[T | X_0 = i] \text{ for } i = 1, 2, \dots, N-1.$$

4.11 First step analysis

First step analysis allows us to evaluate quantities of interest from the Markov chain, for e.g. the absorption probabilities and the mean duration until absorption. The method is based on considering all possibilities at the end of the first transition and then apply the law of total probability to formulate equations involved all unknown quantities. We illustrate how to use the first step analysis in the following Markov chain.

Example 4.18. *Consider the Markov chain with state space $S = \{0, 1, 2\}$ and transition probability matrix given by*

$$P = \begin{bmatrix} 1 & 0 & 0 \\ p_{10} & p_{11} & p_{12} \\ 0 & 0 & 1 \end{bmatrix}.$$

The classes and types are as follows:

- Two closed classes are $C_1 = \{0\}$ and $C_2 = \{2\}$.

- $\{1\}$ is a non-closed class.

Let us consider the problem of evaluating the absorption probabilities. For any closed class C, define

$$u_i^C = \Pr(\text{Markov chain eventually absorbed in } C | X_0 = i).$$

Clearly, the absorption probabilities also depend on the initial states. A vector of absorption probabilities is then given by $\mathbf{u}^C = (u_i^C)_{i \in S}$ We suppress the superscript C and simply write $u_i^C = u_i$ and $\mathbf{u}^C = \mathbf{u}$.

Consider the closed class $C_1 = \{0\}$. We have

$$u_0 = \Pr(\text{Markov chain eventually absorbed in } C_1 | X_0 = 0) = 1,$$
$$u_2 = \Pr(\text{Markov chain eventually absorbed in } C_1 | X_0 = 2) = 0,$$
$$u_1 = \Pr(\text{Markov chain eventually absorbed in } C_1 | X_0 = 1) = u_1.$$

By considering the first transition from state 1 to either state 0, 1 and 2, and using the Markov property, the law of total probability gives

$$u_1 = \Pr(\text{Markov chain eventually absorbed in } C_1 | X_0 = 1)$$

$$= \sum_{k=0}^{2} \Pr(\text{Markov chain eventually absorbed in } C_1 | X_0 = 1, X_1 = k) \Pr(X_1 = k | X_0 = 1)$$

$$= \sum_{k=0}^{2} \Pr(\text{Markov chain eventually absorbed in } C_1 | X_1 = k) \Pr(X_1 = k | X_0 = 1)$$

$$= (p_{10}) \cdot u_0 + (p_{11}) \cdot u_1 + (p_{12}) \cdot u_2$$

$$= (p_{10}) \cdot 1 + (p_{11}) \cdot u_1 + (p_{12}) \cdot 0.$$

Solving for u_1 gives

$$u_1 = u_1^{C_1} = \frac{p_{10}}{1 - p_{11}} = \frac{p_{10}}{p_{10} + p_{12}}.$$

Similarly, we have

$$u_0 = p_{00} \cdot u_0 + p_{01} \cdot u_1 + p_{02} \cdot u_2$$

$$u_1 = p_{10} \cdot u_0 + p_{11} \cdot u_1 + p_{12} \cdot u_2$$

$$u_2 = p_{20} \cdot u_0 + p_{21} \cdot u_1 + p_{22} \cdot u_2,$$

where the first and the last equations reduce to $u_0 = u_0$ and $u_2 = u_2$, respectively. In general, for a closed class C, the vector of absorption probabilities $\mathbf{u}$ satisfies the following system of linear equations:

1. $\mathbf{u} = P\mathbf{u}$ (here $\mathbf{u}$ is treated as a column vector),

2. $u_i = \Pr(\text{Markov chain eventually absorbed in } C) | X_0 = i) = 1$ for all $i \in C$, and

3. $u_i = \Pr(\text{Markov chain eventually absorbed in } C) | X_0 = i) = 0$ for all i in any other close classes.

Example 4.19. *In this example, consider the closed class $C_2 = \{2\}$. Find the absorption probabilities $u_0^{C_2}, u_1^{C_2}$ and $u_2^{C_2}$. Comment on these results.*

For the closed class C_2, we proceed in the same way as in the closed class C_1. Let $\mathbf{u} = \mathbf{u}^{C_2} = (u_0, u_1, u_2)^T$ be the vector of absorption probabilities in the closed class $C_2 = \{2\}$ with $u_0 = 0$ and $u_2 = 1$. It follows that

$$u_1 = p_{10} \cdot u_0 + p_{11} \cdot u_1 + p_{12} \cdot u_2$$

$$= p_{11} \cdot u_1 + p_{12}.$$

Hence, $u_1 = \frac{p_{12}}{1 - p_{11}} = \frac{p_{12}}{p_{10} + p_{12}}$. It should be emphasised that

$$\mathbf{u}^{C_1} + \mathbf{u}^{C_2} = \mathbf{1}.$$

1. For any initial state i, the sum of the absorption probabilities over all closed classes is 1 (as verified in Example Example 19). In particular, when a Markov chain has two closed classes C_1 and C_2, $\mathbf{u}^{C_2} = \mathbf{1} - \mathbf{u}^{C_1}$.

2. In the case when S is finite or when the set of states in non-closed classes is finite, the vector $\mathbf{u}$ is the unique solution of the above system of linear equations.

Example 4.20. *Consider the Markov chain defined in Example Example 17. Find the absorption probabilities in the closed class $C_1 = \{1\}$ and $C_2 = \{4, 5\}$.*

Here $C_1 = \{1\}$ and $C_2 = \{4, 5\}$ are closed classes and $0 = \{2, 3\}$ is an open class.

Let $\mathbf{u} = \mathbf{u}^{C_1}$ be the vector of absorption probabilities in the closed class $C_1 = \{1\}$. Write $\mathbf{u} = (u_1, u_2, \dots, u_5)^T$ and $u_1 = 1$ and $u_4 = u_5 = 0$, From $\mathbf{u} = P \cdot \mathbf{u}$,

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_5 \end{pmatrix} = P \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_5 \end{pmatrix} \text{ gives}$$

$$u_2 = \frac{1}{5} + \frac{1}{5}u_2 + \frac{1}{5}u_3$$

$$u_3 = \frac{1}{3} + \frac{1}{3}u_2.$$

Solving the linear system for u_2 and u_3 yields $u_2 = 4/11$ and $u_3 = 5/11$. Hence, the absorption probabilities in the closed class C_1 is

$$\mathbf{u} = (1, 4/11, 5/11, 0, 0)^T.$$

In addition, since there are two closed classes, $\mathbf{u}^{C_2} = \mathbf{1} - \mathbf{u}^{C_1} = (0, 7/11, 6/11, 1, 1)^T$.

4.12 The expected time to absorption

The expected time to absorption can be determined by analysing all possibilities occurring in the first step. We again consider the process defined in Example ExampleÂ 18 on the set $\{0, 1, 2\}$ with the transition matrix

$$P = \begin{bmatrix} 1 & 0 & 0 \\ p_{10} & p_{11} & p_{12} \\ 0 & 0 & 1 \end{bmatrix}.$$

The time of absorption T is defined as

$$T = \min\{n \geq 0 | X_n = 0 \text{ or } X_n = 2\}$$

and the mean time to absorption of the process is given by $v = \mathrm{E}[T|X_0 = 1]$.

The following observations can be made:

1. The absorption time T is always at least 1.

2. If either $X_1 = 0$ or $X_1 = 2$, then no further steps are required.

3. If $X_1 = 1$, then the process is back at its starting point and on the average v additional steps are required for absorption.

Weighting all these possibilities by their respective probabilities, we obtain the following equation

$$v = 1 + p_{10} \cdot 0 + p_{11} \cdot v + p_{12} \cdot 0$$
$$= 1 + p_{11} \cdot v,$$

which results in

$$v = \frac{1}{1 - p_{11}}.$$

Example 4.21. *Consider the Markov chain with state space $S = \{0, 1, 2, 3\}$ and transition probability matrix given by*

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Let T be the time of absorption defined by

$$T = \min\{n \geq 0 | X_n = 0 \text{ or } X_n = 3\}$$

*and the **absorption probabilities** given by*

$$u_i = \Pr\{X_T = 0 | X_0 = i\}, \ for \ i = 1, 2$$

and the mean time to absorption of the process is given by

$$v_i = \mathrm{E}[T | X_0 = 1], \ for \ i = 1, 2.$$

Calculate the absorption probabilities and the mean time to absorption.

There are 2 closed classes including $C_1 = \{0\}$, and $C_2 = \{3\}$, and one non-closed class $O = \{1, 2\}$. Here,

$$u_i = u_i^{C_1} = \Pr\{X_T = 0 | X_0 = i\} = \Pr(\text{Markov chain eventually absorbed in } C_1 | X_0 = i), \ for \ i = 1, 2$$

By conditioning on the first step from state i and using the Markov property, we have

$$u_i^{C_1} = \sum_{j \in S} p_{ij} u_j^{C_1}.$$

Clearly, $u_0^{C_1} = 1$ and $u_3^{C_1} = 0$. In particular, we have

$$u_1 = p_{10} \cdot 1 + p_{11} \cdot u_1 + p_{12} \cdot u_2$$
$$u_2 = p_{20} \cdot 1 + p_{21} \cdot u_1 + p_{22} \cdot u_2,$$

which can also be obtained from the matrix equation $\mathbf{u} = P\mathbf{u}$, where $\mathbf{u} = (1, u_1, u_2, 0)^T$. The solution to the system of linear equations is

$$u_1 = \frac{p_{10}(p_{22} - 1) - p_{12}p_{20}}{p_{11}(-p_{22}) + p_{11} + p_{12}p_{21} + p_{22} - 1},$$
$$u_2 = \frac{(p_{11} - 1)p_{20} - p_{10}p_{21}}{p_{11}(-p_{22}) + p_{11} + p_{12}p_{21} + p_{22} - 1}.$$

Similarly, the mean time to absorption also depends on the starting state. By the first step analysis, we have for $v_i = \mathrm{E}[T | X_0 = 1]$,

$$v_1 = 1 + p_{11} \cdot v_1 + p_{12} \cdot v_2$$
$$v_2 = 1 + p_{21} \cdot v_1 + p_{22} \cdot v_2.$$

Here the absorption time T is always at least 1. If either $X_1 = 0$ or $X_1 = 3$, then no further steps are required. On the other hand, if $X_1 = 1$ or $X_1 = 2$, then the process will require additional steps, and on the average, these are v_1 and v_2. Weighting these two possibilities, i.e. whether $X_1 = 1$ or $X_1 = 2$, by their respective probabilities and summing according to the law of total probability result in the above system of equations.

Solving the equations for v_1 and v_2 give the mean time to absorption

$$v_1 = \frac{-p_{12} + p_{22} - 1}{p_{11}(-p_{22}) + p_{11} + p_{12}p_{21} + p_{22} - 1},$$
$$v_2 = \frac{p_{11} - p_{21} - 1}{p_{11}(-p_{22}) + p_{11} + p_{12}p_{21} + p_{22} - 1}.$$

4.13   The long-term distribution of a Markov chain

In this section, we present another important property concerning limiting behaviour of P^n as $n \to \infty$ and hence the long-term distribution of a Markov chain satisfying some certain conditions. In particular, some Markov chains will converge to an equilibrium (limiting) distribution, which is independent of its initial state.

We also assume that the Markov chain with **a single closed class** S.

4.14 Stationary and limiting distributions for a single closed class

Stationary distributions

Throughout this section, we consider a Markov chain whose transition probability matrix is P and state space S is a single close class. Then S is necessarily closed and hence irreducible.

A probability distribution $\pi = (\pi)_{i \in S}$ on S is **stationary** if the following conditions hold (here π is a row vector):

1. $\pi_j = \sum_{i \in S} \pi_i p_{ij}$ or equivalently $\pi P = \pi$,

2. $\pi_j \geq 0$,

3. $\sum_{j \in S} \pi_j = 1$.

1. For any stationary distribution π, for all $n \geq 1$,

$$\pi P^n = \pi.$$

   Therefore, if we take π as the initial probability distribution, i.e. $\Pr(X_0 = i) = \pi_i$, then then the distribution of X_n is also π, i.e. $\Pr(X_n = i) = \pi_i$

2. The probability distribution π is said to be an invariant probability distribution.

3. The most important property concerning the stationary distribution(which will be made formal) is that it gives the **long-term (limiting) distribution** of a Markov chain. In addition, π_j also gives **the long run mean fraction of time** that the process $\{X_n\}$ is in state j.

Example 4.22. *A Markov chain $X_0, X_1, \dots$ on states $\{1, 2\}$ has the following transition matrix*

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix},$$

where $0 < a, b < 1$.

1. *Show that*

$$P^n = \frac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix} + \frac{(1-a-b)^n}{a+b} \begin{bmatrix} a & -a \\ -b & b \end{bmatrix}.$$

2. *Show that the stationary probability distribution is*

$$\pi = \left(\frac{b}{a+b}, \frac{a}{a+b} \right).$$

3. *Show that*

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j > 0, \ \text{for } j \in \{1, 2\}.$$

Proportion of Time in Each State

The limiting distribution provides the long-term behaviour of the Markov chain, i.e. it is the long-term probability that a Markov chain hits each state. In this section, it can be shown that it also gives the long-term proportion of time that the chain visits each state. Let us consider a Markov chain $X_0, X_1, \dots$ whose transition probability matrix is P and its limiting distribution is π. Note that the limiting distribution for the Markov chain satisfies

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j.$$

For each state j, define indicator random variable

$$I_k = \begin{cases} 1, & \text{if } X_k = j \\ 0, & \text{otherwise,} \end{cases}$$

for $k = 0, 1, \dots$. Hence, the number of times that the Markov chain visits j in the first n steps is given by $\sum_{k=0}^{n-1} I_k$ and the expected long-term proportion of time that the chain visits state j given that its initial state is i is

$$\lim_{n \to \infty} E\left(\frac{1}{n} \sum_{k=0}^{n-1} I_k \mid X_0 = i \right) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} E(I_k \mid X_0 = i)$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \Pr(X_k = j | X_0 = i)$$

$$= \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$$

$$= \lim_{n \to \infty} p_{ij}^{(n)} = \pi_j.$$

Here we use the fact that if the sequence of numbers converges to a limit, i.e. $x_n \to x$ as $n \to \infty$, then the sequence of partial averages also converges to that limit, i.e. $(x_1 + x_2 + \cdots x_n)/n \to x$ as $n \to \infty$. This result is known as Cesaro's lemma.

Example 4.23. *Recall from ExampleÂ ExampleÂ 9, the simple weather pattern can be classified into three types including rainy (R), cloudy (C) and sunny (S). The weather is observed daily and can be modelled by the Markov transition matrix*

$$P = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.75 & 0.15 & 0.1 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}.$$

It can be shown that the stationary distribution and also the limiting distribution of the Markov chain is

$$\pi = (94/147, 32/147, 1/7)$$

which gives the proportions of visits to rainy, cloudy and sunny states are 94/147, 32/147, 1/7, respectively.

The method of finding the stationary distribution

To find the stationary distribution, we simply solve the linear equations $\pi P = \pi$ (note that one of the equations can be discarded), together with the condition $\sum_{j \in S} \pi_j = 1$.

Example 4.24. *For the NCD process in Example ExampleÂ 2, the Markov chain has the following transition probability matrix*

$$P = \begin{bmatrix} 1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 1-p & p \end{bmatrix}.$$

Find the stationary probability distribution of this chain.

Denote the stationary probability distribution by $\pi = (\pi_1, \pi_2, \pi_3)$. From $\pi P = \pi$ and $\pi_1 + \pi_2 + \pi_3 = 1$,

$$(\pi_1, \pi_2, \pi_3) \begin{bmatrix} 1-p & p & 0 \\ 1-p & 0 & p \\ 0 & 1-p & p \end{bmatrix} = (\pi_1, \pi_2, \pi_3),$$

which is equivalent to

$$(1-p)\pi_1 + (1-p)\pi_2 = \pi_1$$
$$p\pi_1 + (1-p)\pi_3 = \pi_2$$
$$p\pi_2 + p\pi_3 = \pi_3$$

By discarding one of the equations and adding the condition that $\pi_1 + \pi_2 + \pi_3 = 1$, one can solve for π_1, π_2, π_3:

$$\pi_1 = \frac{(p-1)^2}{p^2 - p + 1}, \quad \pi_2 = \frac{-p^2 + p}{p^2 - p + 1}, \quad \pi_3 = \frac{p^2}{p^2 - p + 1}.$$

1. In the above two examples, it can be shown that

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j > 0, \text{ for } j \in S,$$

or, in terms of the Markov chain $\{X_n\}$,

$$\lim_{n\to\infty} \Pr(X_n = j | X_0 = i) = \pi_j > 0, \text{ for } j \in S.$$

This means that in the long run (as $n \to \infty$), the probability of finding Markov chain in state j is approximately π_j **no matter in which state the chain began at time 0**. This property holds for some Markov chains which satisfy **"certain conditions"**.

4.15 Sufficient conditions for the long-run behaviour of a Markov chain

In what follows, we will establish a set of sufficient conditions for the long-run behaviour of a Markov chain. Two important results are stated without proof.

Theorem 4.1. *Theorem 1.* A Markov chain with a finite state space has at least one stationary probability distribution.

Theorem 4.2. *Theorem 2.* An irreducible Markov chain with a finite state space has a unique stationary probability distribution.

Example 4.25. *The simple random walks X_n on $S = \{..., -2, -1, 0, 1, 2, ...\}$ is defined as*

$$X_n = X_0 + \xi_1 + \xi_2 + ... + \xi_n,$$

where the random variables ξ_j are independent identically distributed with

$$\Pr(\xi = 1) = p, \quad \Pr(\xi = -1) = 1 - p.$$

*We can check that this Markov chain is irreducible. However, the state space S is infinite. It can be checked directly from the equations $\pi P = \pi$ that there is **no** stationary distribution (given as an exercise).*

First we know that the entries of a stationary distribution sum to one. Suppose the contrary that there is a stationary distribution π for the simple random walk. Then by spatial invariance of the simple random walk, π_i is constant for all $i \in S$ and also $\sum_{i \in S} \pi_i = 1$, which is impossible because S is infinite. Hence, there is **no** stationary distribution for the simple random walk.

4.16   Limiting distributions

One of the important properties of stationary distributions is that the distribution of the Markov chain satisfying certain conditions converges to the stationary distribution. This result provides the long-term behaviour of the Markov chain. In order to state the main result of this section, we need to introduce another concept, namely the period of a state.

A state i is said to be **periodic** with period $d > 1$ if a return to i is possible only in a number of steps that is a multiple of d. Equivalently, the period d is the greatest common divisor of all integers n for which $p_{ii}^{(n)} > 0$. If the greatest common divisor is 1, the state has period 1 and is said to be **aperiodic**.

A Markov chain in which each state has period 1 is called **aperiodic**. Most Markov chains in applications are aperiodic.

Example 4.26. *Is the NCD system in ExampleÂ ExampleÂ 2 aperiodic?*

The entire state space of the NCD system is a single class, and is necessarily closed. Note also that $p_{00} > 0$ (and also $p_{22} > 0$), i.e. the state 0 (and state 2) has an arrow back to itself. Consequently, it is aperiodic because a return to this state is possible in any number of steps (or the system can remain in this state in any length of time).

Similarly, a return to state 1 is possible in $2, 3, \ldots$ steps. Therefore, the NCD system is aperiodic.

1. Periodicity is a class property, i.e. all states in one class have the same period (or if $i \leftrightarrow j$, then i and j have the same period).

2. The Markov chain $\{X_n\}_{n \geq 0}$ is aperiodic if and only if there exists some $n > 0$ such that

$$p_{ij}^{(n)} > 0 \text{ for all } i, j \in S.$$

   Such Markov chain is also called **regular**.

Example 4.27. *In the random walk model on a finite state space $S = \{0, 1, \ldots, N\}$ with absorbing boundaries in ExampleÂ ExampleÂ 12, determine the period of each state.*

The transition diagram and the transition matrix of the simple random walk with absorbing boundaries (states) are given as follows:

The simple random walk with absorbing boundaries has three classes, $O = \{1, 2, \ldots, N-1\}$ is non-closed class, $C_1 = 0$ and $C_2 = N$ are two closed classes. For each state i in O, $p_{ii}^{(2n)} > 0$ and $p_{ii}^{(2n+1)} = 0$ for $n = 1, 2, \ldots$. Therefore, the state i in this open communication class has period 2.

On the other hand, $p_{00} = 1$ (and also $p_{NN} = 1$) and hence, the states 0 and N are aperiodic because a return to each of these states is possible in any number of steps (or the system can remain in this state in any length of time).

Example 4.28. *In the random walk model on an infinite state space $S = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ in ExampleÂ ExampleÂ 25, determine the period of each state.*

The entire state space is a single class. Note also that $p_{ii}^{(2n)} > 0$ and $p_{ii}^{(2n+1)} = 0$ for $n = 1, 2, \ldots$. Therefore each state in the random walk on an infinite set S is periodic with period 2.

Example 4.29. *Suppose the states of the system are $\{1, 2, 3, 4\}$ and the transition matrix is*

$$P = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}.$$

Determine the period of each state.

The entire state space is a single class. Note also that $p_{ii}^{(2n)} > 0$ and $p_{ii}^{(2n+1)} = 0$ for $n = 1, 2, \ldots$. Therefore each state in the random walk on an infinite set S is periodic with period 2.

Example 4.30. *A Markov chain $X_0, X_1, \ldots$ on states $\{1, 2\}$ has the following transition matrix*

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

1.

<div align="center">

example1

</div>

Find the stationary distribution(s) of this Markov chain.

2. *Describe the long-term behaviour of the Markov chain. Does the distribution of the chain tend to the stationary distribution(s) found in*

<div align="center">

example1

</div>

.

1. To find the stationary distribution $\pi = (\pi_1, \pi_2)$, we need to solve

$$(\pi_1, \pi_2) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = (\pi_1, \pi_2),$$

and

$$\pi_1 + \pi_2 = 1.$$

This gives $\pi = (1/2, 1/2)$.

2. There is an equal chance of being in either state. Note that for any initial probability distribution $\mu = (\mu_1, \mu_2)$ with $\mu_1 + \mu_2 = 1$, we have

$$\mu \cdot P = \mu \cdot P^3 = \mu \cdot P^5 = \ldots = (\mu_2, \mu_1)$$

and

$$\mu P^2 = \mu P^4 = \mu P^6 = \ldots = (\mu_1, \mu_2).$$

The process does not settle down to an equilibrium position. Note also that the chain is not aperiodic, i.e. each state is periodic of period 2. The process does not conform to stationary in the long run.

Example 4.31. *A Markov chain $X_0, X_1, \ldots$ on states $\{1, 2\}$ has the following transition matrix*

$$P = \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix},$$

Answer the same questions as given in the Example ExampleÂ 30

1. The process is finite and irreducible, so a unique stationary distribution exists. To find the stationary distribution $\pi = (\pi_1, \pi_2)$, we need to solve

$$(\pi_1, \pi_2) \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix} = (\pi_1, \pi_2),$$

and

$$\pi_1 + \pi_2 = 1.$$

Solving the system of linear equations gives $\pi = (1/2, 1/2)$.

2. The chain is aperiodic. Therefore, according to the results from ExampleÂ ExampleÂ 22, if follows that

$$\lim_{n \to \infty} p_{i1}^{(n)} = 1/2 > 0, \text{ and } \lim_{n \to \infty} p_{i2}^{(n)} = 1/2,$$

which is independent of i. This is contrast to the process given in ExampleÂ ExampleÂ 30, i.e. the process in this example reaches the stationary probability distribution in the long run.

4.17   Main result

The main result in this section can be stated as follows:

Theorem 4.3. ***Theorem 3***. Let P be the transition probability matrix of a homogeneous discrete-time Markov chain $\{X_n\}_{n\geq 0}$. If the Markov chain is

- finite,

- irreducible and

- aperiodic,

then there is the unique probability distribution $\pi = (\pi_j)_{j\in S}$ such that

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j > 0, \text{ for any } j \in S$$

and

$$\sum_{j\in S} \pi_j = 1,$$

and this distribution is independent of the initial state. Such probability distribution π is called the **limiting probability distribution**. In addition, the limiting distribution $\pi = (\pi_j)_{j\in S}$ is the stationary probability distribution of the Markov chain, i.e. it also satisfies $\pi P = \pi$.

Example 4.32. *Recall the Markov chain as given in ExampleÂ ExampleÂ 22. The Markov chain is finite, irreducible and aperiodic. We have also shown that*

$$\pi = \left(\frac{b}{a+b}, \frac{a}{a+b} \right)$$

is the limiting distribution, i.e.

$$\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j > 0, \text{ for } j \in \{1,2\},$$

independent of i. This limiting distribution is also the unique stationary distribution of the Markov chain, which can be verified by

$$\left(\frac{b}{a+b}, \frac{a}{a+b} \right) \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \left(\frac{b}{a+b}, \frac{a}{a+b} \right),$$

where $0 < a, b < 1$.

Example 4.33. *The above result can be applied to the NCD system because it is finite, irreducible and aperiodic. Indeed, there is the unique limiting probability distribution*

$$\pi = \left(\frac{(p-1)^2}{p^2-p+1}, \frac{p-p^2}{p^2-p+1}, \frac{p^2}{p^2-p+1} \right),$$

which is the stationary distribution of the chain. This gives the long-term behaviour of the Markov chain, i.e. the probability of finding the Markov chain in state j is approximately π_j independent of the initial distribution.

For example, let $p = 0.8$ in the transition probability matrix P. We compute several powers of P as follows:

$$P = \begin{bmatrix} 0.2 & 0.8 & 0 \\ 0.2 & 0 & 0.8 \\ 0 & 0.2 & 0.8 \end{bmatrix}, \qquad P^2 = \begin{bmatrix} 0.2 & 0.16 & 0.64 \\ 0.04 & 0.32 & 0.64 \\ 0.04 & 0.16 & 0.8 \end{bmatrix},$$

$$P^4 = \begin{bmatrix} 0.072 & 0.1856 & 0.7424 \\ 0.0464 & 0.2112 & 0.7424 \\ 0.0464 & 0.1856 & 0.768 \end{bmatrix}, \qquad P^8 = \begin{bmatrix} 0.0482432 & 0.1903514 & 0.7614054 \\ 0.04758784 & 0.1910067 & 0.7614054 \\ 0.04758784 & 0.1903514 & 0.7620608 \end{bmatrix}$$

$$P^{16} = \begin{bmatrix} 0.04761946 & 0.1904761 & 0.7619044 \\ 0.04761903 & 0.1904765 & 0.7619044 \\ 0.04761903 & 0.1904761 & 0.7619049 \end{bmatrix}, \quad P^{32} = \begin{bmatrix} 0.04761905 & 0.1904762 & 0.7619048 \\ 0.04761905 & 0.1904762 & 0.7619048 \\ 0.04761905 & 0.1904762 & 0.7619048 \end{bmatrix}.$$

The limiting probability distribution is

$$\lim_{n \to \infty} P^n = \begin{bmatrix} 0.04761905 & 0.1904762 & 0.7619048 \\ 0.04761905 & 0.1904762 & 0.7619048 \\ 0.04761905 & 0.1904762 & 0.7619048 \end{bmatrix}.$$

Applications of Markov chains to NCD systems

Example 4.34. *A no-claims discount system for motor insurance has three levels of discount:*

Level	1	2	3
Discount	0%	30%	50%

The rules for moving between these levels are given as follows:

- *Following a claim-free year, move to the next higher level, or remain at level 3.*

- *Following a year with one claim, move to the next lower level, or remain at level 1.*

- *Following a year with two or more claims, move to level 1, or remain at level 1.*

A portfolio consists of 10000 policyholders, of which

- *5000 policyholders are classified as good drivers. The number of claims per year in this group is Poisson(0.1).*

- *5000 policyholders are classified as bad drivers. The number of claims per year in this group is Poisson(0.2).*

1. *Calculate* $\Pr[N = 0]$, $\Pr[N = 1]$, *and* $\Pr[N \geq 2]$ *for each group.*

2. *Write down the transition probability matrix of this no-claims discount system for each group.*

3. *Calculate the expected number of policyholders at each level for each group once stability has been achieved.*

4. *Calculate the expected premium income per driver from each group once stability has been achieved.*

5. *Calculate the ratio of the expected premium income per driver from the group of good drivers to that from the group of bad drivers once stability has been achieved.*

6. *Comments on the results obtained. Does this NCD system encourage good driving?*

Chapter 5

Applications

Some *significant* applications are demonstrated in this chapter.

5.1   DataCamp Light

By default, `tutorial` will convert all R chunks.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJhIDwtIDJcbmIgPC0gM1xuXG5hICsgYiJ9

Chapter 6

Final words

Some *significant* applications are demonstrated in this chapter.

6.1   DataCamp Light

By default, `tutorial` will convert all R chunks.

eyJsYW5ndWFnZSI6InIiLCJzYW1wbGUiOiJhIDwtIDJcbmIgPC0gM1xuYSArIGIifQ==