SCMA469 Actuarial Statistics

Pairote Satiracoo

2021-08-23

Contents

Chapter 1

Prerequisites

This is a *sample* book written in **Markdown**. You can use anything that Pandoc's Markdown supports, e.g., a math equation $a^2 + b^2 = c^2$.

The **bookdown** package can be installed from CRAN or Github:

```
install.packages("bookdown")
# or the development version
# devtools::install_github("rstudio/bookdown")
```

Remember each Rmd file contains one and only one chapter, and a chapter is defined by the first-level heading `#`.

To compile this example to PDF, you need XeLaTeX. You are recommended to install TinyTeX (which includes XeLaTeX): https://yihui.name/tinytex/.

Chapter 2

Introduction to Stochastic Processes

You can label chapter and section titles using `{#label}` after them, e.g., we can reference Chapter **??**. If you do not manually label them, there will be automatic labels anyway, e.g., Chapter **??**.

Figures and tables with captions will be placed in `figure` and `table` environments, respectively.

```
par(mar = c(4, 4, .1, .1))
plot(pressure, type = 'b', pch = 19)
```

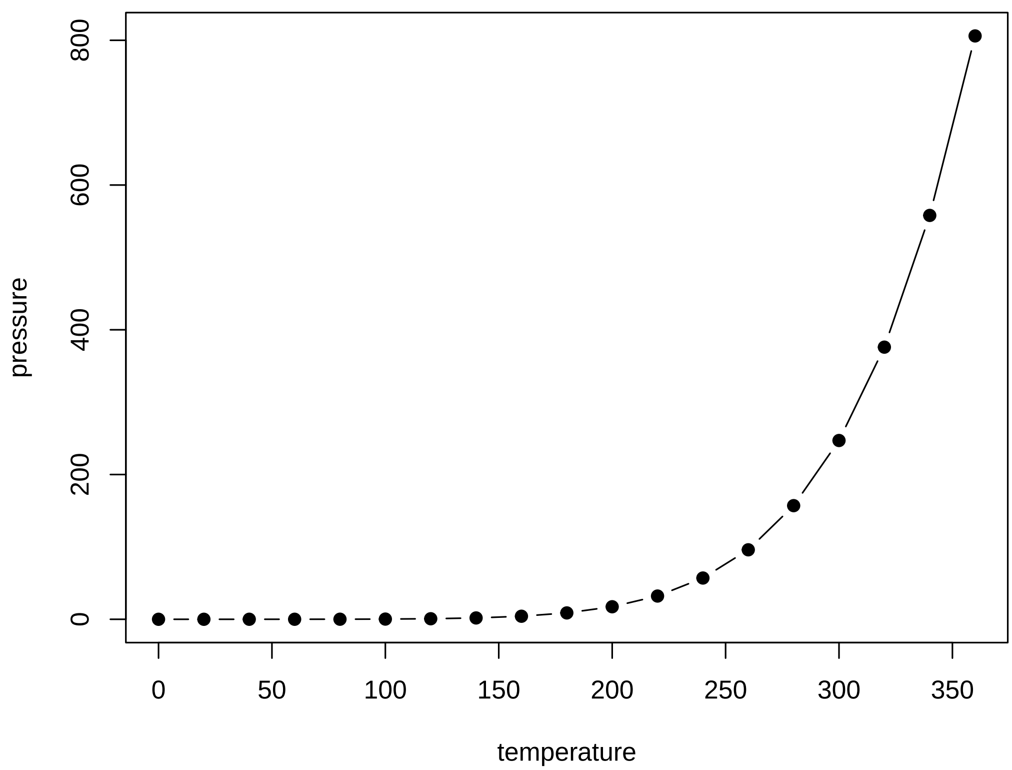

Figure 2.1: Here is a nice figure!

Reference a figure by its code chunk label with the `fig:` prefix, e.g., see Figure **??**. Similarly, you can reference tables generated from `knitr::kable()`, e.g., see Table **??**.

Table 2.1: Here is a nice table!

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa

```
knitr::kable(
  head(iris, 20), caption = 'Here is a nice table!',
  booktabs = TRUE
)
```

You can write citations, too. For example, we are using the **bookdown** package (**?**) in this sample book, which was built on top of R Markdown and **knitr** (**?**).

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.

Chapter 3

Examples of real world processes

Example 3.1. Example 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 $X_0, X_1, X_2, \ldots$ *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 3.2. Example 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 $\{0, 1, 2, \ldots\}$*, 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 3.3. Example 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,* $0 \leq a_i < 1$.
- $x + b_i$ *is the age at which observation ends, if life does not die,* $0 \leq b_i < 1$.
- $x + t_i$ *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
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 4

Review of probability theory

4.1 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 4.1. Example 4. *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 $i, j = 1, 2, ...,$ 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**.

4.2 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 4.2. Example 5. *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*

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

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

Example 4.3. Example 6. *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.*

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

4.3 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,

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

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 $B_1, B_2, \dots B_n$ 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 4.4. Example 7. *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 4.5. Example 8. *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.

4.4 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

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

4.5 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.

Central Limit Theorem

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 5

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.

5.1 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.

5.2 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 satang. For each trading day $n = 0, 1, 2, \ldots$, the stock price increases by 1 satang with probability p or decreases by 1 satang with probability $q = 1 - p$. Let X_n denote the stock price at day n and $X_0 = 100$, i.e. $X_0 =$. 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.

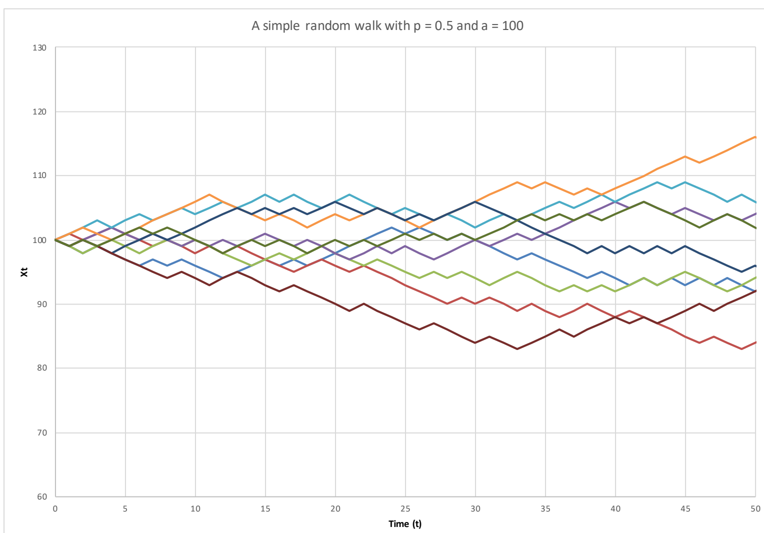

Figure 5.1: Some stochastic realisations of the simple random walk

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

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

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 5.1. Example 9. *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 5.2. Example 10. *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 5.3. Example 11. *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 5.4. Example 12. *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 5.5. Example 13. *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*

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

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 6

Discrete-time Markov chains

Chapter 7

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 $X_0, X_1, \dots, X_n$ 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:

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

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

Example 7.1. Example 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:

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

The random walk process has the Markov property.

7.1 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 7.2. Example 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 7.3. Example 3. *In a health insurance system, at the end of each day an insurance company classifies policyholders 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 7.4. Example 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*.)*

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

Example 7.5. Example 5. *Show that*

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

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

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

Example 7.6. Example 6. *For the NCD system defined on the state space $S = \{0, 1, 2\}$ as given in Example Example 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 Example Example 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

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

Chapter 8

The Chapman-Kolmogorov equations

8.1 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 8.1. Example 7. *For the NCD system given in Example Example 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

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

8.2 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 8.2. Example 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.

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

Example 8.3. Example 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

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

8.3 Joint Distribution

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

$$P(X_{n_1} = i_1, X_{n_2} = i_2, \dots, X_{n_k} = i_{k1}1, X_{n_k} = i_k) = (\mu P^{n_1})_{i_1} (P^{n_2 - n_1})_{i_1 i_2} \cdots (P^{n_k - n_{k-1}})_{i_{k-1} i_k}.$$

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

Example 8.4. Example 10. *In Example Example 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

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

8.4 Random walk with absorbing and reflecting barrier(s)

Example 8.5. Example 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

$$\begin{aligned} &\Pr(X_{n+1} = k | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= \Pr(X_{n+1} = k | X_n = i) = \begin{cases} p_i, & k = i+1 \\ r_i, & k = i \\ q_i, & k = i-1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Example 8.6. Example 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*

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

*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 Example 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.

8.5 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 8.7. Example 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

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

   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 $P(i), i \leq 4$.