Applied Probability III

APP MTH 3001/7056

#APPMTH3001

Lewis Mitchell

lewis.mitchell@adelaide.edu.au

Room 6.46

Ingkarni Wardli

THE UNIVERSITY of ADELAIDE

Course outline

- **Section 00:** Probability Review *(Assumed Knowledge)*
- **Section 01:** Introduction
- **Section 02:** Probability and measure
  - Sample space
  - Algebras and σ-algebras of events
  - Probability measure
- **Section 03:** Discrete Time Markov Chains
  - Basic definitions
  - Hitting probabilities and hitting times
  - Classification of states
  - Recurrence and transience
  - Limiting behaviour
- **Section 04:** Martingales
  - Definition of a Martingale
  - Stopping times and the Optional stopping theorem
- **Section 05:** Brownian motion

Sample space

A formal treatment of probability requires a study of the axiomatic structure of probability theory, which is beyond the scope of this course.

The interested reader is referred to one of the references listed in the introduction.

In this section, we review some of the concepts pertaining to events and probabilities that you are likely to be already familiar with.

Let Ω be a **sample space**, whose elements ω correspond to the possible outcomes of an experiment.

Example 0.1:

Consider an experiment which consists of tossing a fair coin twice.
Then the sample space is

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\} = \{HH, HT, TH, TT\}.$$

Events

Here are some examples of events

- the event that a "head" comes up when a coin is tossed,
- the event that two "heads" come up when a coin is tossed twice,
- the event that a particular candidate wins an election,
- the event that the sun rises tomorrow.

In general, if A and B are events, they can be thought of as subsets of the sample space Ω. That is, we have $A, B \subseteq \Omega$.

The event that at least one of A or B occurs is called the **union** of A and B, and is written $A \cup B$.

The event that both occur is called the **intersection** of A and B, and is written $A \cap B$.

This notation extends to all countable collections of events $A_i, i = 1, 2, \ldots$, for which we write

$$A_1 \cup A_2 \cup \cdots = \cup_{i=1}^{\infty} A_i,$$

$$A_1 \cap A_2 \cap \cdots = \cap_{i=1}^{\infty} A_i.$$

Events and probability

- The probability of an event A is written $\mathrm{P}(A)$.
- The **certain** event, denoted by Ω, always occurs, so $\mathrm{P}(\Omega) = 1$
- The **impossible** event, denoted by $\emptyset$, never occurs, so $\mathrm{P}(\emptyset) = 0$.

It is always the case that $0 \leq \mathrm{P}(A) \leq 1$ for any event A, so that if you ever calculate a probability and it is negative or greater than one, then you know you have made a mistake !

Definition 0.1: Disjoint events

> Events A and B are said to be disjoint if $A \cap B = \emptyset$.
> That is, if A and B cannot both occur.

For disjoint events A and B, we have $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$.
A more general statement of the **addition law** is as follows.
For events $A_i, i = 1, 2, \ldots,$ with A_j and A_k disjoint for all $j \neq k$, we have

$$\mathrm{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)\,. \tag{0.1}$$

Conditional probability

Many of the probabilities we will encounter in this course, and indeed many statements about chance in general, are conditional probabilities, taking the form

"The probability of A, given that B occurs ",

where A and B are events.

Definition 0.2: Conditional probability

If $\mathrm{P}(B) > 0$ then the conditional probability that A occurs given that B occurs is defined to be

$$P\left(A \,\middle|\, B\right) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}. \tag{0.2}$$

Note: For event A to occur given event B has occurred, then both must occur. That is, event $A \cap B$, and since we know that B has occurred, the event $B \subseteq \Omega$ must be our new sample space. Hence the the conditional probability of A given B must be the probability of A and B relative to the probability of B.

Conditional probability

Note: the definition of conditional probability can also be motivated as follows.

Given that an event B occurs, then event A occurs if and only if event $\{A \cap B\}$ occurs.

Therefore the conditional probability of A given B must be proportional to $\mathrm{P}(A \cap B)$, and we write $\mathrm{P}(A \mid B) = \alpha\, \mathrm{P}(A \cap B)$ for some constant α.

In order to determine α, we exploit the fact that we must have $\mathrm{P}(\Omega \mid B) = 1$.

That is,

$$\begin{aligned} 1 = \mathrm{P}(\Omega \mid B) &= \alpha\, \mathrm{P}(\Omega \cap B) \\ &= \alpha\, \mathrm{P}(B)\,, \end{aligned}$$

$$\text{yielding } \alpha = \frac{1}{\mathrm{P}(B)}.$$

The Law of total probability

The formula for conditional probability can be used to derive a very important result in probability, known as the *law of total probability*, which we shall make regular use of in this course.

Definition 0.3: Partition

A family $B_1, B_2, \ldots, B_n$ of events is a partition of Ω if

$$B_i \cap B_j \;=\; \emptyset \text{ when } i \neq j, \quad \text{and} \quad \cup_{i=1}^{n} B_i \;=\; \Omega.$$

Theorem 0.1: The Law of total probability

Let $B_1, B_2, \ldots, B_n$ be a partition of Ω. Then

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) \tag{0.3}$$

$$= \sum_{i=1}^{n} P(A \mid B_i)\, P(B_i)\,. \tag{0.4}$$

The Law of total probability

A proof of the Law of total probability is left as a question in class exercise 1.

Hint: use the fact that if the set $\{B_1, B_2, \ldots B_n\}$ forms a partition of Ω then

$$\mathrm{P}(A) \quad = \quad \mathrm{P}(\cup_{i=1}^{n}(A \cap B_i))$$

as a starting point.

We will also show later that for events A, B and C, such that $\mathrm{P}(B \cap C) > 0$

$$\mathrm{P}(\, A \cap B \,|\, C) \quad = \quad \mathrm{P}(\, A \,|\, B \cap C)\, \mathrm{P}(\, B \,|\, C)\,. \tag{0.5}$$

Note that this expression will be used often in the course, and will play a fundamental role in several derivations and proofs.

Independent events

Definition 0.4: Independent events

Events A and B are called independent if

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)\,. \tag{0.6}$$

This follows from first principles by noting that if A and B are independent, then

$$\mathrm{P}(\,A \mid B) = \mathrm{P}(A)\,, \tag{0.7}$$

which substituted into the formula for conditional probability gives (0.6).
More generally, a family $A_i, i \in I$ is called independent if

$$\mathrm{P}(\cap_{i\in J} A_i) = \prod_{i\in J} \mathrm{P}(A_i)$$

for all finite subsets J of I.

Common notation

Notation	Set jargon	Probability jargon
Ω	Collection of objects (whole space)	Sample space (certain event)
ω	Member of Ω	Elementary event, outcome
A	Subset of Ω	Event that (some outcome in) A occurs
A^c	Complement of A	Event that A does not occur
$A \cap B$	Intersection	Event that both A and B occur
$A \cup B$	Union	Event that either A or B, or both occur
$\emptyset$	Empty set	Impossible event

Table 1: Summary of common notation

Random variables

A random variable (r.v.) can be thought of informally as a variable that takes its values "by chance".
Recall that Ω is used to denote a sample space, whose elements ω correspond to the possible outcomes of an experiment, so that a little more formally.

Definition 0.5: Random Variable

A real-valued **random variable** is a (measurable) function $X(\omega)$ which maps Ω onto the real number line $\mathbb{R}$.

That is, $X : \Omega \to \mathbb{R}$.

The idea of "measurability" as it pertains to the function $X(\omega)$ will be introduced later in this course.

For a complete discussion of axiomatic probability theory the interested reader can refer to, for example,

Probability: theory and examples by Rick Durrett.

State space

It is a common convention to denote

- random variables using capital letters such as $X(\omega), Y(\omega), Z(\omega)$,
- specific values of the r.v.'s by the corresponding lower case letters x, y, z.

In Example 0.1, if the random variable $Y(\omega)$ is the total number of "heads", then

$$Y(HH) = 2,\ Y(HT) = Y(TH) = 1,\ Y(TT) = 0.$$

Note that the set $Y(\omega)$ takes values in the set $\{0, 1, 2\} \subseteq \mathbb{R}$.
This leads to the notion of the state space of a random variable.

Definition 0.6: State Space of a random variable

The **state space** of a random variable $X(\omega)$ is the set $\mathcal{S}_X \subseteq \mathbb{R}$ of all possible values that can be taken by $X(\omega)$.

Typically, a r.v. $X(\omega)$ maps Ω onto a restricted subset of $\mathbb{R}$. The above is an example of a discrete random variable, where the state space of $Y(\omega)$ is clearly the set of discrete values $\mathcal{S}_Y = \{0, 1, 2\}$, so that $Y : \Omega \to \mathbb{R}$.

Discrete random variables

Definition 0.7: Discrete random variables

A **discrete random variable** $X(\omega)$ takes values in some countable set $\{x_1, x_2, \ldots\}$. That is, it has a countable state space $\mathcal{S}$.

Recall that a set $\mathcal{S}$ is countable if every element of the set can be placed in a one-to-one correspondence with the set of natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$.

Note that:

- all finite sets are countable, eg: $\{H, T\}, \{1, 2, 3, 4, 5, 6\}$.
- infinite sets may be
  - countable (countably infinite), such as $\mathbb{N}, \mathbb{Z}$, or
  - uncountable (uncountably infinite), such as $[0, 1], \mathbb{R}$.

In this course, we deal most often with discrete random variables which, by definition, have a countable state space.

In that case, we can always represent the state space of a random variable $X(\omega)$ by $\mathcal{S} = \{x_1, x_2, \ldots\}$.

Discrete random variables

Example 0.2:

Consider an experiment which consists of tossing a fair coin repeatedly until the first "tail" turns up, so that

$$\Omega = \{T, HT, HHT, HHHT, \ldots\}.$$

Let the random variable $Z(\omega)$ count the number of tosses that occur before the experiment stops (equivalently, the duration of the experiment). Then

$$Z(T) = 1, \quad Z(HT) = 2, \quad Z(HHT) = 3, \ldots$$

and the state space of $Z(\omega)$ is $\mathcal{S}_Z = \{1, 2, 3, \ldots\}$.

In the above example, both the sample space Ω and the state space $\mathcal{S}$ are (countably) infinite.

Note: uncountably infinite states spaces arise exclusively in the context of continuous random variables, only briefly covered in this course (covered extensively in the second semester subject Random Processes).

Probability measure

From here on, we abbreviate $X = X(\omega)$, $Y = Y(\omega)$, etc. when convenient.

When the outcome of an experiment is known, a random variable X takes some numerical value, which is often more likely to lie in certain subsets of the state space than others.

We wish to describe the distribution of likelihoods of possible values of X, so this is where the notion of a probability measure comes in.

Recall Example 0.1.
Since the coin is fair, the possible outcomes in $\Omega = \{HH, HT, TH, TT\}$ are equally likely, each having probability

$$\frac{1}{2} \times \frac{1}{2} = \frac{1}{4} \, .$$

We can write

$$\mathrm{P}(HH) = \mathrm{P}(HT) = \mathrm{P}(TH) = \mathrm{P}(TT) = \frac{1}{4}.$$

Probability mass

We now turn our attention to the random variable Y, which counts the total number of "heads", and find that

$$\begin{aligned} \mathrm{P}(Y = 0) &= \mathrm{P}(TT) = \frac{1}{4} \\ \mathrm{P}(Y = 1) &= \mathrm{P}(HT \cup TH) \\ &= \mathrm{P}(HT) + \mathrm{P}(TH) = \frac{1}{2} \\ \mathrm{P}(Y = 2) &= \mathrm{P}(HH) = \frac{1}{4} \end{aligned}$$

The above probabilities constitute the "probability masses" associated with each possible value of Y.

This motivates the following definition on the next slide.

Probability mass and distribution

Definition 0.8: Probability mass function

A discrete random variable X has a **probability mass function** $f_X : \mathcal{S}_X \to [0, 1]$, given by

$$f_X(x_i) = \mathrm{P}(X = x_i) = \mathrm{P}((\omega \in \Omega : X(\omega) = x_i)), \qquad i = 1, 2, \ldots.$$

such that $\sum_i f_X(x_i) = 1$.

(Cf. the probability density function for a continuous random variable).

An alternative way to characterise a probability distribution is by its distribution function.

Definition 0.9: Probability distribution function

The **distribution function** of a random variable X is the function $F_X : \mathbb{R} \to [0, 1]$, given by

$$F_X(x) = \mathrm{P}(X \leq x) = \mathrm{P}((\omega \in \Omega : X(\omega) \leq x)).$$

Probability mass and distribution

Distribution functions become particularly important in the study of continuous random variables.

However, for most applications involving discrete random variables, as is our situation, it is more convenient to utilise the probability mass function.

For completeness, we note that the probability mass function and the distribution function of X are related in the following way:

$$\begin{aligned} F_X(x) &= \sum_{i:x_i \leq x} f_X(x_i), \\ f_X(x_i) &= F_X(x_i) - \lim_{\epsilon \to 0} F_X(x_i - \epsilon), \quad \text{for } \epsilon > 0. \end{aligned}$$

Probability mass and distribution

Example 0.3:

For the coin tossing experiment in Example 0.1, the distribution function of Y is given by

$$F_Y(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{4} & \text{if } 0 \leq x < 1 \\ \frac{3}{4} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2, \end{cases}$$

For a random variable X, the abbreviations $F \equiv F_X$ and $f \equiv f_X$ can be used whenever no confusion is likely to arise.

Of course, a study of random variables is not complete without a discussion of expected values.

Expected values

Definition 0.10: Expectation

The **expectation** of a random variable (or expected value) X with probability mass function f is defined to be

$$E(X) = \sum_i x_i f(x_i), \tag{0.8}$$

provided that this sum is absolutely convergent.

Note: the sum (0.8) has the potential to diverge if it contains infinitely many terms, as we shall see later in the course.

If X is a random variable and g is a function, then $Y = g(X)$ is also a random variable, whose expectation is given

$$E(g(X)) = \sum_i g(x_i) f(x_i), \tag{0.9}$$

again provided that this sum is absolutely convergent.

Expected values

Example 0.4:

Let's calculate two expectations using the expectation operator on the random variable Y defined on Example 0.1.

First the expected value of Y

$$\begin{aligned} E(Y) &= 0 \times P(Y=1) + 1 \times P(Y=1) + 2 \times P(Y=2) \\ &= 0\left(\frac{1}{4}\right) + 1\left(\frac{1}{2}\right) + 2\left(\frac{1}{4}\right) = 1\,. \end{aligned}$$

which is one of the values that Y can take.
(this is not always true, e.g. expected value when you roll a die is $\frac{7}{2}$)

Now the expectation of the function Y^2

$$\begin{aligned} E(Y^2) &= 0^2 \times P(Y=1) + 1^2 \times P(Y=1) + 2^2 \times P(Y=2) \\ &= 0\left(\frac{1}{4}\right) + 1\left(\frac{1}{2}\right) + 4\left(\frac{1}{4}\right) = \frac{3}{2}\,. \end{aligned}$$

Higher moments

Definition 0.11: Moments and central moments

For a positive integer k, the k^{th} **moment** m_k of X is given by

$$m_k = E(X^k).$$

The k^{th} **central moment** σ_k is

$$\sigma_k = E((X - m_1)^k). \tag{0.10}$$

The two moments most commonly used in probability and statistics are

- $m_1 = E(X)$ (expectation) and
- $\sigma_2 = Var(X)$ (variance).

σ_3 skewness

σ_4 kurtosis

When obtaining the variance of a r.v. X, it is often useful to use the fact that

$$\sigma_2 = m_2 - (m_1)^2, \quad \left(\text{that is, } Var(X) = E(X^2) - (E(X))^2 \right).$$

Variance

We can prove the last expression for variance directly using (0.10):

$$\sigma_2 \;=\; \sum_i (x_i - m_1)^2 f(x_i) \;=\; \sum_i x_i^2 f(x_i) - 2m_1 \sum_i x_i f(x_i) + m_1^2 \sum_i f(x_i)$$

$$\;=\; m_2 - 2m_1 m_1 + m_1^2 \times 1 \;=\; m_2 - m_1^2.$$

An alternative way to prove the result is by making use of the fact that taking the expectation of a r.v. is a linear operation.

In particular, E is a linear operator because it satisfies

1. $E(aX) = aE(X)$, for an arbitrary real constant a and random variable X.
2. $E(X + Y) = E(X) + E(Y)$, for random variables X and Y.

We can then use this property to show that

$$Var(X) \;=\; E((X - E(X))^2) \;=\; E(X^2 - 2XE(X) + E(X)E(X))$$

$$\;=\; E(X^2) - 2\big(E(X)\big)^2 + \big(E(X)\big)^2 \;=\; E(X^2) - \big(E(X)\big)^2.$$

Common discrete distributions

Example 0.5: The Bernoulli Distribution.

A random variable X following the Bernoulli distribution with parameter p has only two possible values, 0 and 1, so that $\Omega = \{0, 1\}$.

The probability mass function is $p(1) = p$ and $p(0) = 1 - p$, where $0 < p < 1$.

The mean of the Bernoulli Distribution is given by

$$\begin{aligned} E(X) &= 0(1-p) + 1(p) \\ &= p, \end{aligned}$$

and the variance is given by

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 \\ &= 0^2(1-p) + 1^2(p) - p^2 \\ &= p(1-p). \end{aligned}$$

Common discrete distributions

Bernoulii random variables arise frequently as indicators of events.

The **indicator** of an event A is the random variable

$$1_A = \begin{cases} 1 & \text{if A occurs} \\ 0 & \text{if A does not occur} \end{cases}$$

Then 1_A is a Bernoulli random variable with parameter p, which has the convenient property that

$$\begin{aligned} E(1_A) &= p \\ &= \mathrm{P}(A)\,. \end{aligned}$$

This property is sometimes useful for simplifying certain calculations which would otherwise be quite complicated, as we shall see shortly.

Common discrete distributions

Example 0.6: Binomial Distribution.

Consider independent events $A_1, \ldots A_n$, all having the same probability $p = P(A_i)$ of occurrence (or "success").
In this case, $\Omega = \{0,1\}^n = \{0,1\} \times \{0,1\} \times \cdots \times \{0,1\}$ (n times).
If we let Y count the total number of events among $A_1, \ldots A_n$, that occur, then Y has a binomial distribution with parameters n and p.
The probability mass function is

$$f_Y(k) = P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, 2, \ldots, n,$$

because there are exactly $\binom{n}{k}$ points in Ω which give a total of k occurrences (successes), and each of these occur with probability $p^k(1-p)^{n-k}$.

We also have that $E(Y) = np$ *and* $Var(Y) = np(1-p)$.

We can think of a binomial random variable as counting the number of "successes" in n independent trials where there is a constant probability p of success on any single trial.

Common discrete distributions

Exercise:

Find $E(Y)$

Hint: by writing Y as a sum of the appropriate indicator random variables.

$$Y = 1_{A_1} + 1_{A_2} + \cdots + 1_{A_n}.$$

Common discrete distributions

Example 0.7: Geometric Distribution.

Let $A_1, A_2, \ldots$ be independent events, all having a common probability $p = P(A_i)$ of occurrence.
Say trial k is a success or failure according as whether A_i occurs or not, and let Z count the number of failures prior to the first success.
Then Z has a geometric distribution with parameter p, and has probability mass function

$$f_Z(k) = p(1-p)^k, \qquad k = 0, 1, 2, \ldots.$$

We also have that $E(Z) = \dfrac{1-p}{p}$ and $Var(Z) = \dfrac{1-p}{p^2}$.

Exercise: show the above using the following facts

$$\frac{d}{dx}\left(\frac{1}{1-x}\right) = \frac{1}{(1-x)^2}$$

and

$$\frac{d^2}{dx^2}\left(\frac{1}{1-x}\right) = \frac{2}{(1-x)^3}\,.$$

Common discrete distributions

Example 0.8: Poisson Distribution.

If a random variable X takes values in the set $\{0, 1, 2, \ldots\}$ with probability mass function

$$f_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \qquad k = 0, 1, 2 \ldots$$

where $\lambda > 0$, then X is said to have the Poisson distribution with parameter λ. We also have that $E(X) = Var(X) = \lambda$.

The Poisson distribution arises in many models of physical systems, and also has some simple and elegant mathematical properties.

Exercise: show the above expected value and variance using the Taylor series expansion of the exponential

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} .$$

Common discrete distributions

The Poisson distribution arises as a limiting case of the binomial distribution.

Specifically, the binomial distribution with parameters n and p converges to the Poisson distribution with parameter λ if $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains constant.

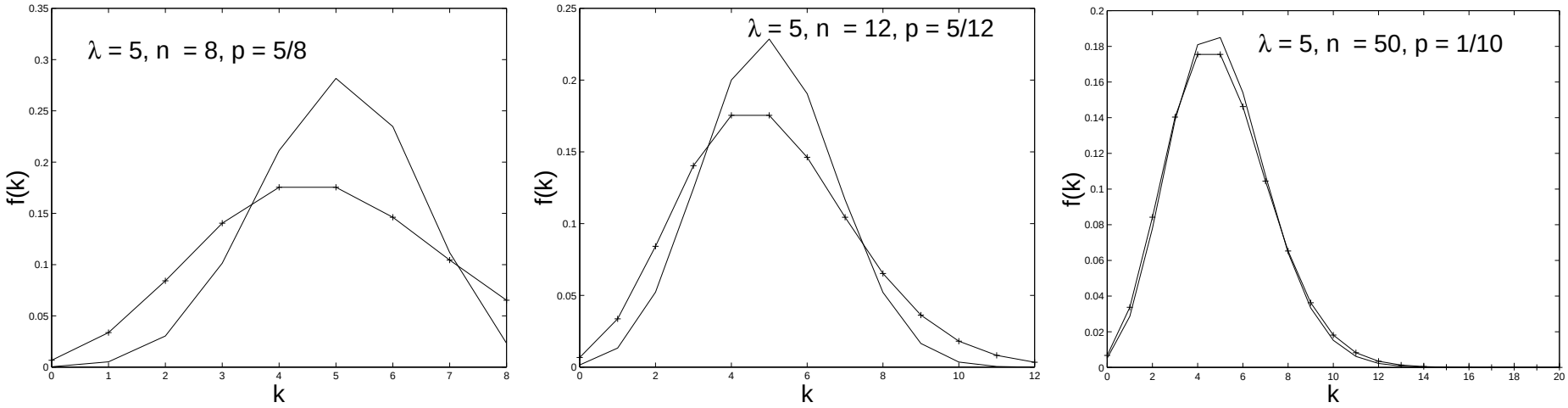
Exercise: show this by making using of the identity

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Common discrete distributions

In words, given an indefinitely large number of independent trials, where success on each trial occurs with the same arbitrarily small probability, then the total number of successes will follow, approximately, a Poisson distribution.
This can be seen in the sequence of figures.
The Poisson distribution with parameter $\lambda = 5 = np$ is plotted (crosses) with the binomial distribution with parameters $n, p = \dfrac{\lambda}{n}$.

"Approximations of a Poisson distribution (crosses) with parameter $\lambda = 5$ by a sequence of binomial distributions".



Common discrete distributions

Example 0.9:

If a Poisson distributed random variable X has a mean of $\lambda = 2$.

What is the probability that X is less than or equal to 2 ?

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} \\ &= 5e^{-2} = 0.6767. \end{aligned}$$

What is the probability that X is greater than 2 ?

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - 5e^{-2} = 0.3233. \end{aligned}$$

Course outline:

- **Section 00:** Probability Review *(Assumed Knowledge)*
- **Section 01:** Introduction
- **Section 02:** Probability and measure
  - Sample space
  - Algebras and σ-algebras of events
  - Probability measure
- **Section 03:** Discrete Time Markov Chains
  - Basic definitions
  - Hitting probabilities and hitting times
  - Classification of states
  - Recurrence and transience
  - Limiting behaviour
- **Section 04:** Martingales
  - Definition of a Martingale
  - Stopping times and the Optional stopping theorem
- **Section 05:** Brownian motion

General Information

APP MTH 3001 Applied Probability III

Semester 1, 2017.

Lecture times:

- 3pm on Monday, Engineering Sth S111
- 3pm on Tuesday, Barr Smith Sth 534 (Forum)
- 3pm on Friday, Lower Napier LG28

Consulting time: Friday 10am-12pm.

Tutorials will be held in weeks $3, 5, 7, 9, 11$.
They will take the place of the Monday lecture. You will be randomly allocated to either Engineering Sth S111 or Engineering & Maths EM218 for each tutorial. You will also be randomly assigned to groups and expected to present your group solutions to the tutorial

Class Exercises will be due at the end of weeks $2, 4, 6, 10, 12$.
They are to be handed in at the **beginning** of the Friday lecture at 3pm or there is a handin box on level 6 of Ingkarni Wardli which will be cleared immediately *before* the lecture.

Group project

There will be a group project that will be due in Week 10. No extensions will be considered.

You are expected to form your own groups, which must be of size 2 or 3. Please inform me of your groups via email (with an email from every group participant, so that I know that you have agreed) within the next week. At the end of that period, I will assign everyone who has not informed me of their group choice into random groups. **Warning:** There is clear evidence that the randomly assigned groups under-perform those that have nominated themselves!

I'll hand out the project sometime in Week 2 and discuss my expectations at that time.

Assessment

Final 2 hour Examination 70%, Project and assignments 30%.

Grading Scheme

High distinction	85 – 100%
Distinction	75 – 84%
Credit	65 – 74%
Pass	50 – 64%
Fail	0 – 49%

Supplementary Examinations:

Students are advised that the University of Adelaide has an official policy on supplementary, replacement or additional assessment, which they are advised to read at

http://www.adelaide.edu.au/student/exams/supps.html

The Course website for this course is on MyUni

Plagiarism and Collusion

Policy on Academic Honesty:

Students are advised that the University of Adelaide has an official policy on academic honesty, which they are advised to read at

http://www.adelaide.edu.au/policies/230/

Policy on assignments for this course only:

Students are encouraged to work together in order to enhance their understanding of the subject matter. To this end they may work together on assignments. However, students are required to have, and may be required to demonstrate, a complete understanding of their submitted assignments. Failure to demonstrate a complete understanding of a submitted assignment may be interpreted as evidence of plagiarism.

Hint: One easy way to work together and yet avoid any issues with plagiarism is to plan how to do the assignment together, but then work on the assignment itself separately. This way, you are able to learn together, but are not tempted to **copy** each other's actual assignment.

"It is remarkable that a science which began with the consideration of games of chance should have become the most important object of human knowledge."

- Pierre Simon, Marquis de Laplace, 1749–1827 (Mathematician)

"Fate laughs at probabilities."

- Edward Bulwer-Lytton, (1803–1873) (Writer)

"The conception of chance enters in the very first steps of scientific activity in virtue of the fact that no observation is absolutely correct. I think chance is a more fundamental conception than causality; for whether in a concrete case, a cause-effect relation holds or not can only be judged by applying the laws of chance to the observation."

- Max Born, 1882–1970 (Physicist)

---

http://http://probweb.berkeley.edu/quotes.html

Introduction

- Probability theory provides us with a way of investigating phenomena in which “chance” plays some part.

  For example, the experiment of flipping a coin is best studied by treating the outcome as a random one, rather than trying to solve the (deterministic) Newtonian equations of motion for the coin and predicting whether it comes up heads or tails.

- The study of “experiments” or phenomena where the outcome is not known for sure until after the experiment has been carried out, provided the initial motivation for the development of probabilistic concepts.

  Indeed, mathematical probability has its origins in games of chance with dice and cards, originating in the fifteenth and sixteenth centuries.

  Important advances were made in the 1600’s (Pascal, Fermat).

Introduction

- In modern probability theory, probability is defined in terms of a set of axioms which do not need to be tied to any specific physical interpretation or experiment (Kolmogorov, 1933).

  As a result, modern probability theory is an extremely rich and interesting area of mathematics, which we could not hope to cover in one course.

  You are encouraged to consult the references if you are interested in the foundations of modern probability theory.

- In this course, we shall initially look at some measure theoretic concepts that form the basis of probability theory, but will then move on to a particular and important branch of applied probability known as Markov processes, rather than continue studying the theory of probability itself.

- We will conclude the course by touching on another important branch of probability known as Martingales and then follow by introducing Brownian motion, which is observed ubiquitously in nature and is therefore used in modelling frequently by mathematicians, economists and physicists.

Introduction

Methods in applied probability can be used to create models and analyse a wide variety of systems which exhibit some degree of random behaviour.

For example:

- Mathematical biology - modelling the population size of a given species in an ecosystem.

- Queueing Theory - modelling the number of data packets waiting to be transmitted on a communications link, or the number of callers waiting at a call centre queue.

- Finance - modelling the time-evolution of a stock price.

Introduction

Using such models, we can then ask and hopefully obtain answers to questions such as:

- what is the probability that a species becomes extinct (under given environmental conditions) ?
- how long will a data packet have to wait in a queue before it is processed by a server ?
- what is the probability that a given stock price exceeds a specific threshold value over a given time period ?

The majority of this course is concerned with a particular type of random process, called a Markov process.

Markov processes form a cornerstone of applied probability.

Introduction

A **random process** is a family, or sequence, of random variables X_n, where n is a parameter running over a suitable index set T.

In this course, we predominantly focus our attention to the case where

- the process has the Markov "memoryless" property, which we define in Section 03.
- the index n corresponds to discrete units of time, and $T = \{0, 1, 2, \ldots\}$ (the non-negative integers $\{0 \cup \mathbb{Z}^+\}$)

  - Cf. continuous time random processes, the subject of the second semester course Random Processes III.
- the random variables X_n take discrete values $\{x_1, x_2, \ldots\}$

  - Cf. continuous random variables, as seen in other 1^{st} and 2^{nd} year courses.

Useful References

1. *Introduction to Probability Models*
   Sheldon Ross (Academic Press, 2010).

2. *An introduction to Stochastic Modelling.*
   Taylor and Karlin (Academic Press, 1998).

3. *A First Course in Stochastic Processes.*
   Karlin and Taylor (Academic Press, 1975).

4. *An Introduction to Probability Theory and its Applications.*
   Feller (Wiley, 1968).

5. *Introduction to Stochastic Models.* *(good background in probability)*
   Roe Goodman (2nd edition, Dover, 2006).

There are many good books on probability theory in the Barr Smith Library.

Try browsing books with call numbers beginning with 519.2.

Course outline

- **Section 00:** Probability Review *(Assumed Knowledge)*
- **Section 01:** Introduction
- **Section 02:** Probability and measure
  - Sample space
  - Algebras and σ-algebras of events
  - Probability measure
- **Section 03:** Discrete Time Markov Chains
  - Basic definitions
  - Hitting probabilities and hitting times
  - Classification of states
  - Recurrence and transience
  - Limiting behaviour
- **Section 04:** Martingales
  - Definition of a Martingale
  - Stopping times and the Optional stopping theorem
- **Section 05:** Brownian motion

Measure & Probability Theory

> Probability theory has a right and a left hand. On the left is the rigourous foundational work using the tools of measure theory. The right hand 'thinks probabilistically,' reduces problems to gambling situations, coin tossing, and motions of a physical particle.
>
> A quote from Rick Durrett's book on Probability Theory.

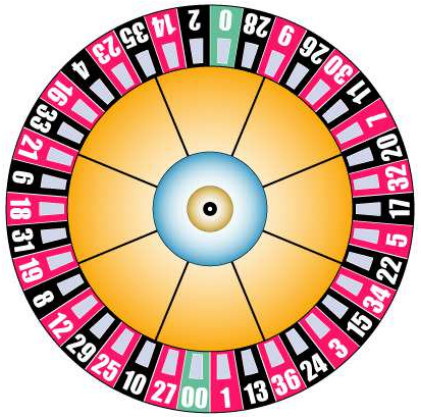
The above gives a reasonable account of the existence of two distinct treatments of probability where on the one hand we use a measure theoretic basis and on the other a more physical interpretative basis.

Left in Latin is "sinister", whereas right is "dexter". This by no means implies that the measure theory basis is sinister, but it does however require an extensive set of definitions before it may be effectively used to describe 'probability'. This is a major reason why it is not often encountered in an undergraduate course.

We will begin by considering the Weak and Strong Laws of Large Numbers and the Central Limit Theorem by way of example using the interpretative right hand to see what information they convey to us and because we need them in what follows.

Laws of Large Numbers

Example 2.1: Roulette



An American Roulette wheel has 38 slots comprised of 18 red, 18 black and 2 green slots numbered 0 and 00.

A gambler who bets \$1 on red being realised on a spin of the wheel gets \$1 and his original bet if it does, or loses his \$1 bet if it does not.

Let's assume that the wheel is unbiased so that the gambler wins with measure $\frac{18}{38}$ and loses with measure $\frac{20}{38}$. Let $X_1, X_2, \ldots$ be a sequence of random variables that are the outcomes of the first, second and subsequent bets. The gambler is interested in $S_n = X_1 + X_2 + \cdots + X_n$, which represents his winnings at time n. Without any justification, the facts are:

- the average winnings on one play $= \$\left(\frac{18}{38} - \frac{20}{38}\right) = -\$\frac{1}{19}$
- the average winnings after n-plays $= -\$\frac{n}{19}$.

◁

Laws of Large Numbers

For most values of n the probability of the gambler having lost exactly $\$\frac{n}{19}$ is essentially zero. For example, let's look at the possible realisations after

1. a single play:

2. ten plays:

We could consider that a gambler could have won $\$10$ if we are optimistic or $-\$10$ if we are pessimistic. These outcomes have respective probabilities

how close to the average is the gambler's experience?

The answer to this is given by the Weak Law of Large Numbers (see next slide).

Laws of Large Numbers

Theorem 2.2 Weak Law of Large Numbers

If $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean $E[X_i] = \mu$, then for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \epsilon\right) \quad = \quad 0.$$

This gives some information, but what if the gambler wrote down the sequence of numbers $\frac{S_1}{1}, \frac{S_2}{2}, \ldots, \frac{S_n}{n}$; would it converge to the value $-\frac{1}{19}$?

The answer is given by the Strong Law of Large Numbers (see next slide).

Laws of Large Numbers

Theorem 2.3 Strong Law of Large Numbers

If $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean $E[X_i] = \mu$, then

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

The immediate consequence of this is that the gambler eventually becomes bankrupt irrespective of how much money he starts with, as $S_n \to -\infty$ as $n \to \infty$.

These two laws of large numbers tell us what will happen in the long run, but do not give much information about what happens in the short run.

This gap is filled by the Central Limit Theorem (see next slide).

Central Limit Theorem

Theorem 2.4 Central Limit Theorem

If $X_1, X_2, \ldots$ are independent and identically distributed random variables with mean $E[X_i] = \mu$ and variance $\sigma^2 = E[(X_i - \mu)^2]$, then

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq y\right) \quad \rightarrow \quad \Phi(y),$$

where $\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} dx$ is the standard Normal distribution, or $N(0, 1)$.

Using this, we can informally write

where $Z \sim N(0, 1)$.

In our roulette wheel example we have that

Central Limit Theorem

If for simplicity we assume that $\sigma^2 = 1$, then we may write

$$S_n \approx -\frac{n}{19} + \sqrt{n}Z,$$

so that for example $S_{100} \approx -5.26 + 10Z$ and

$$\mathrm{P}(-5.26 + 10Z \geq 0) = \mathrm{P}(Z \geq 0.526) \approx 0.30.$$

Hence after 100 plays the gambler has lost \$5.26 on average and has probability 0.30 of being in front.
Continuing on in this manner, 100 gamblers who play 100 times by our calculations yields

In fact, $\mathrm{P}(Z \leq 2.33) = 0.9901$, which implies that $\mathrm{P}(S_{10,000} \leq -293) = 0.9901$ and so the Casino is slowly but surely making money.

Measure & Probability Theory

Now let's begin to formalise things a little more by considering some concepts which we should be familiar with...

- Ω is a set of outcomes (sample space) of a stochastic process
- $\mathcal{F}$ is a set of events defined on the set of outcomes
- $P : \mathcal{F} \to [0, 1]$ is a function that assigns probabilities to events.

Collectively, the two sets Ω and $\mathcal{F}$ along with the probability measure P form what is known in the measure theoretic realm as a triple $(\Omega, \mathcal{F}, P)$, that describes the probability space.

1. If $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$.
2. If $\{A_i : A_i \in \mathcal{F}\}$ is a countable sequence of sets then

$$\cup_i A_i \in \mathcal{F} \ .$$

Note that $\cap_i A_i = \left(\cup_i A_i^C\right)^C$, which implies that a σ-field is also closed under countable intersections.

A most important σ-field is the Borel σ-field which contains sets which are known as Borel sets.

Definition 2.12: Borel σ-field

The smallest σ-field $\mathcal{B}$ that contains all open intervals (a, b), such that $(-\infty \leq a < b \leq \infty)$, is called a Borel σ-field.

A set $B \in \mathcal{B}$ is called a Borel set.

It turns out that a number of different intervals such as

- $[a, b) : -\infty < a < b \leq \infty$,
- $(a, b] : -\infty \leq a < b < \infty$,
- $[a, b] : -\infty < a < b < \infty$,
- $(a, b) : -\infty \leq a < b \leq \infty$, or
- $(-\infty, b] : -\infty < b < \infty$

can be used to generate the same Borel σ-field.
If we are able to define a probability measure w.r.t. any of these intervals, a probability measure is then automatically defined for $\mathcal{B}$.

We could continue on in this way and build up a concise statement of measure theoretic probability, but that is not our objective.

Why algebras, σ-algebras, σ-fields or Borel σ-fields?

Isn't a probability measure simple to establish on the outcomes of an experiment?

The answer is of course not always, as some measures can be undefined!

For most everyone in this class, that the probability of a particular occurrence of a continuous random variable is zero (or has measure zero) even if it is not an impossible event should not be surprising, but what about an infinite collection of such events?

The need for such things as σ-fields or Borel σ-fields can be motivated by some nice examples, and so we will start with a particular example (again a gambling example as they are readily accessible and well defined) taken from SA Lotteries.

Example 2.2: X Lotto

Let's consider a simplified version of the game X Lotto offered by SA Lotteries, where you essentially pick 6 numbers between 1 and 45 and when the game is played, SA Lotteries draws 6 numbers from the 45 and if yours match, you win!

There are 45 ways of drawing the first number, 44 ways of drawing the second number and so since the order does not matter, we conclude that there are

ways of choosing a set of 6 numbers.

The sample space, Ω, is therefore the union of all these $8,145,060$ distinct 6-tuples $\omega_j = \{\omega_{j_1}, \omega_{j_2}, \ldots, \omega_{j_6}\}$, since each is a possible outcome of the game.

Each of these 6-tuples is different in so far as $\omega_j \cap \omega_k \notin \Omega$ for $j \neq k$.

Example 2.2: X Lotto (cont)

But what are the possible *events* in this game? If I had the money and patience, I could choose to bet on any particular subset of Ω (i.e. collection of the above 6-tuples). So any subset of Ω is certainly a potential event. Therefore, let $\mathcal{F}$ - the set of all possible events - be the family of all subsets of the sample space Ω, also known as the *power set* of Ω. Note that $\mathcal{F}$ also includes Ω itself and the empty set $\emptyset$. This is because, you could either decide not to play ($\emptyset$) or you could choose every possible 6-tuple (Ω). Now because the collection $\mathcal{F}$ contains all subsets of Ω, it satisfies the following conditions

(2.11)

(2.12)

By Induction, (2.12) implies that if

Now for a few definitions based on what we have seen.

Definition 2.13: An algebra

A collection of subsets $\mathcal{F}$ of a non-empty set Ω satisfying equations (2.11)

$$\text{If} \quad A \in \mathcal{F} \quad \text{then} \quad A^C \in \mathcal{F}$$

and (2.12),

$$\text{If} \quad A, B \in \mathcal{F} \quad \text{then} \quad A \cup B \in \mathcal{F}$$

is called an algebra (or sometimes a field).

In the X Lotto example the sample space Ω is finite and so the collection of subsets $\mathcal{F}$ is also finite and so we may write

(2.13)

where not all of the A_i can be distinct, since $\mathcal{F}$ is finite.

Definition 2.14: A σ-algebra

A collection of subsets $\mathcal{F}$ of a non-empty set Ω satisfying equations (2.11)

$$\text{If} \quad A \in \mathcal{F} \quad \text{then} \quad A^C \in \mathcal{F}$$

and (2.13)

$$\text{If} \;\; A_i \in \mathcal{F} \;\; \text{for} \;\; i = 1, 2, \ldots, \;\; \text{then} \;\; \cup_{i=1}^{\infty} A_i \in \mathcal{F}$$

is called a σ-algebra (or sometimes a σ-field or Borel σ-field).

Note the key difference: we now need $\mathcal{F}$ to be closed under a *countable*-union of sets, not just under a *finite*-union of sets.

Measure & Probability Theory

Now let's consider the probability of winning our X Lotto game. Each number (that has not already been selected) has an equal chance of being chosen, and so each 6-tuple, ω_j, has the same chance of occurrence

There are $N = 8,145,060$ possible ω_j, and so a single 6-tuple has a $\frac{1}{N}$ chance of occurrence. Therefore if you choose n 6-tuples, you have probability $\frac{n}{N}$ of winning.

To put this another way, let A be the set of 6-tuples that you have chosen. Then the probability $P(A)$ for $A \in \mathcal{F}$ is given by the number n of elements in the set A divided by the total number of elements in Ω. In particular $P(\Omega) = 1$ (if you choose every 6-tuple) and $P(\emptyset) = 0$ (if you do not choose any 6-tuples).

$P(A)$ for $A \in \mathcal{F}$ is a function called a probability measure, which assigns a number $P(A) \in [0, 1]$ to each set $A \in \mathcal{F}$. It is not true that every function that assigns a number in $[0, 1]$ to the sets in $\mathcal{F}$ is a probability measure, so we need a definition to describe appropriate functions of this kind.

Definition 2.15: Probability measure

A mapping $P : \mathcal{F} \to [0,1]$ from a σ-algebra $\mathcal{F}$ of subsets of a set Ω into the unit interval is a probability measure on the measurable space $\{\Omega, \mathcal{F}\}$ if it satisfies the following three conditions

$$P(A) \geq 0 \quad \text{for all } A \in \mathcal{F}, \tag{2.14}$$

$$P(\Omega) = 1 \quad \text{and} \tag{2.15}$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{for all disjoint sets } A_i \in \mathcal{F}. \tag{2.16}$$

In the X Lotto example we only have a finite number of disjoint sets but equation (2.16) is still satisfied since we can always include an infinite number of empty sets in this union since $\emptyset \cap \emptyset = \emptyset$.

Then, in general, if $\mathcal{F}$ is finite, it is sufficient to check that

In our X Lotto example, if disjoint sets A_1 and A_2 have n_1 and n_2 elements respectively, then $P(A_1 \cup A_2) = \frac{n_1+n_2}{N} = \frac{n_1}{N} + \frac{n_2}{N} = P(A_1) + P(A_2)$.

Our X Lotto game is therefore completely described by the triple $\{\Omega, \mathcal{F}, P\}$, which is called the probability space.

In general we may state for a probability space that it consists of

- The sample space Ω (the set of all possible outcomes of the experiment)
- A σ-algebra (a collection of subsets of Ω where equations (2.11) and (2.13) are satisfied)
- A probability measure $P : \mathcal{F} \to [0, 1]$, satisfying (2.14), (2.15) and (2.16).

In the X Lotto game, the collection $\mathcal{F}$ of events is, in fact, an algebra, but because $\mathcal{F}$ is finite, it is automatically a σ-algebra as well.

Theorem 2.5

If an algebra contains only a finite number of sets, then it is a σ-algebra.

But why a σ-algebra?

Consider the following example, about which we will assume a great deal.

Example 2.3: Tossing a fair coin

Toss a fair coin infinitely many times and if the result is heads at each toss you receive a reward of 1 unit and nothing otherwise. In this case, we can establish a state space Ω, which consists of an infinite number of infinite strings of 0s and 1s. For example, $\omega = (0, 1, 1, 1, 0, \dots)$.

Now consider the event $A_{k,n}$ which consists of all those outcomes $\omega \in \Omega$ where the winnings after n tosses is k units; an example of which is $A_{1,2}$, which consists of all $\omega \in \Omega$ of the form $(0, 1, \dots)$ and $(1, 0, \dots)$.

The probability $P(A_{k,n})$ is therefore given by

Example 2.3: Tossing a fair coin (cont)

For $q = 1, 2, \ldots$, consider now the events that after n tosses, the average win per coin toss, $\frac{k}{n}$, is contained in the interval $\left[\frac{1}{2} - \frac{1}{q}, \frac{1}{2} + \frac{1}{q}\right]$. These events correspond to the sets $B_{q,n}$, where

Then the set $\bigcap_{m=n}^{\infty} B_{q,m}$ corresponds to the event that from the n^{th} tossing onwards, the average win per coin toss will stay in the interval $\left[0.5 - \frac{1}{q}, 0.5 + \frac{1}{q}\right]$.

The set $\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} B_{q,m}$ corresponds to the event that there exists an n such that from the n^{th} tossing onwards , the average win per coin toss will stay in the interval $\left[0.5 - \frac{1}{q}, 0.5 + \frac{1}{q}\right]$

Example 2.3: Tossing a fair coin (cont)

Then finally, the set $\bigcap_{q=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} B_{q,m}$ corresponds to the event that the average win per coin toss converges to $\frac{1}{2}$ as $n \to \infty$. Now the Strong Law of Large Numbers (Theorem 2.3) therefore states that the latter event has probability 1. That is,

which is only defined if

To guarantee this, $\mathcal{F}$ must be a σ-algebra. ◁

Definition 2.16: Smallest σ-algebra

The smallest σ-algebra containing a given collection $\mathcal{C}$ of sets is called the σ-algebra generated by $\mathcal{C}$ and is usually denoted by $\sigma(\mathcal{C})$.

The idea of a smallest σ-algebra of subsets of Ω is always relative to a collection of subsets of Ω, because without reference to such a collection of sets the smallest σ-algebra of subsets of Ω is just $\{\phi, \Omega\}$.

An important special case of a smallest σ-algebra is where $\Omega = \mathbb{R}$ and $\mathcal{C}$ is the collection of all open intervals:

$$\mathcal{C} = \{(a, b) : a < b, \quad \text{where } a, b \in \mathbb{R}\} \ .$$

Definition 2.17: Euclidean Borel field (see Definition (2.12))

The σ-algebra generated by the collection of all open intervals $\{(a, b) : a < b, \quad \text{where } a, b \in \mathbb{R}\}$ in $\mathbb{R}$ is called the Euclidean Borel field, denoted by $\mathcal{B}$ and its members are called the Borel sets.

Note that although $\mathcal{B}$ is generally defined in terms of the open intervals as we have just seen, the σ-algebras generated by the

- open intervals $\{(a, b) : a < b, \text{ where } a, b \in \mathbb{R}\}$,
- closed intervals $\{[a, b] : a < b, \text{ where } a, b \in \mathbb{R}\}$ and
- half-open intervals
  - $\{[a, b) : a < b, \text{ where } a, b \in \mathbb{R}\}$ and
  - $\{(a, b] : a < b, \text{ where } a, b \in \mathbb{R}\}$

are all the same. The notion of Borel sets extends to the higher dimensions as well.

Definition 2.18: k-dimensional Euclidean Borel field

$$\mathcal{B}^k = \sigma\left(\left\{(a_1, b_1) \times \cdots \times (a_k, b_k) : a_j < b_j, \text{ where } a_j, b_j \in \mathbb{R}\right\}\right)$$

is the k-dimensional Euclidean Borel field
and its members are the Borel sets in $\mathbb{R}^k$.

Measure & Probability Theory

Let's now revisit our probability measure given in Definition 2.15 and look at the implied properties of the three axiomatic equations (2.14), (2.15) and (2.16).

Theorem 2.6

Let $\{\Omega, \mathcal{F}, P\}$ be a probability space, then the following hold for sets in $\mathcal{F}$

(a) $P(\emptyset) = 0$,

(b) $P(A^C) = 1 - P(A)$,

(c) $A \subset B$ *implies that* $P(A) \leq P(B)$,

(d) $P(A \cup B) + P(A \cap B) = P(A) + P(B)$,

(e) *If* $A_n \subset A_{n+1}$ *for* $n = 1, 2, \ldots,$ *then* $P(A_n) \uparrow P\left(\bigcup_{n=1}^{\infty} A_n\right)$,

(f) *If* $A_n \supset A_{n+1}$ *for* $n = 1, 2, \ldots,$ *then* $P(A_n) \downarrow P\left(\bigcap_{n=1}^{\infty} A_n\right)$,

(g) $P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$.

Example 2.4: The Uniform Probability Measure

Place ten balls, numbered $0, 1, \ldots, 9$, in a bowl.

Draw a ball randomly from this bowl and write down the corresponding number in the tenths place of a decimal number between 0 and 1. Put the ball back in the bowl and repeat this experiment, except this time write the corresponding number of the chosen ball after the previous number so that it is in the hundredths place.

For example, if I had chosen 5 and 7 respectively I would have scribed the number 0.57.

Repeating this experiment infinitely many times yields a **random number** between 0 and 1 such that the sample space is the unit interval $[0, 1]$.

In this example, for a given number $x \in [0, 1]$, the probability that a **random number** chosen as above is less than or equal to x is in fact x.

To see this is true, we will consider a progressive experiment. ◁

Example 2.4: The Uniform Probability Measure (cont)

Suppose we draw only two balls as above and end up with 0.57, then

1.

2.

3.

Hence we see that there are 58 ways of choosing a number less than or equal to 0.57 and so in this experiment where we only choose two balls, the probability we obtain a number less than or equal to 0.57 is $58/100 = 0.58$. Similarly if we choose ten balls with replacement, the probability we obtain a number less than or equal to 0.5713492765 is 0.5713492766 and in the limit we see that for $x \in [0, 1]$ we have that $P\big([0, x]\big) = x$.

It follows that $P\big([x]\big) = P\big([x, x]\big) = 0$ and $P\big([0, 1]\big) = 1$. ◁

Example 2.4: The Uniform Probability Measure (cont)

More generally

$$P\big((0,x]\big) \;=\; P\big([0,x)\big) \;=\; P\big((0,x)\big) \;=\; x \quad \text{and}$$

$$P\big((a,b]\big) \;=\; P\big([a,b)\big) \;=\; P\big((a,b)\big) \;=\; b-a, \;\; \text{for } \; 0 \le a < b \le 1\,.$$

Any finite union of intervals can be written as a finite union of disjoint intervals by cutting out the overlap. Therefore this probability measure extends to finite unions of intervals by adding up the lengths of the associated disjoint intervals.

Moreover the collection of all finite unions of subintervals in $[0,1]$ including $[0,1]$ and the empty set, is closed under complement and finite unions. Hence, we have derived the probability measure P for an algebra $\mathcal{F}_0$ of subsets of $[0,1]$ given by

$$\mathcal{F}_0 \;=\; \left\{ \begin{array}{c} (a,b),[a,b],(a,b],[a,b), \text{ where } a,b \in [0,1] \text{ and } a \le b, \\ \text{along with their finite unions} \end{array} \right\} \; \triangleleft$$

Measure & Probability Theory

The probability measure we have just derived is a special case of Lebesgue measure, which assigns the length of an interval.

The algebra $\mathcal{F}_0$ does not however contain all of the Borel sets in $[0,1]$ and therefore not all probability measures are defined. In particular, for a countable sequence of sets $A_j \in \mathcal{F}_0$, the probability $P\left(\bigcup_{j=1}^{\infty} A_j\right)$ is not always defined because there is no guarantee that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}_0$.

We therefore need to extend the probability measure P on $\mathcal{F}_0$ to a probability measure defined on the Borel sets in $[0,1]$, which is normally achieved using the concept of "outer measure".

Any subset A of $[0,1]$ can always be completely covered by a finite or countably infinite union of sets in the algebra $\mathcal{F}_0 \;:\; A \subset \bigcup_{j=1}^{\infty} A_j, \;\text{for } A_j \in \mathcal{F}_0$.

The "probability" of A is therefore bounded above by $\sum_{j=1}^{\infty} P(A_j)$ and the infimum of this bound taken across all sequences of sets $A_j \in \mathcal{F}_0$ where $A \subset \bigcup_{j=1}^{\infty} A_j,$ yields the "outer measure".

Definition 2.19: Outer measure

If $\mathcal{F}_0$ is an algebra of subsets of Ω, then the outer measure P^* of an arbitrary subset A of Ω is

$$P^*(A) \quad = \quad \inf_{A \subset \left(\bigcup_{j=1}^{\infty} A_j\right),\, A_j \in \mathcal{F}_0} \sum_{j=1}^{\infty} P(A_j) \; .$$

Note that $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}_0$ is not a requirement in Definition 2.19.

Without loss of generality we may also assume that the infimum of Definition 2.19 is taken over all disjoint sets $A_j \in \mathcal{F}_0$, in which case if it happens that $A \in \mathcal{F}_0$ then $P^*(A) = P(A)$.

It seems reasonable to now also consider for what other sets in Ω does the outer measure correspond to a probability measure.

Note that the outer measure P^* satisfies the first two equations (2.14) and (2.15) of Definition 2.15 of probability measure, but in general the third equation (2.16) does not hold for arbitrary sets. It is however possible to extend the outer measure to a probability measure on a σ-algebra $\mathcal{F}$ containing $\mathcal{F}_0$.

Theorem 2.7

If P is a probability measure on $\{\Omega, \mathcal{F}_0\}$, where $\mathcal{F}_0$ is an algebra, let $\mathcal{F} = \sigma(\mathcal{F}_0)$ (the smallest σ-algebra containing $\mathcal{F}_0$). Then the outer measure P^* is the unique probability measure on $\{\Omega, \mathcal{F}\}$ which coincides with P on $\mathcal{F}_0$.

Hence for the experiment in Example 2.4, there exists a σ-algebra $\mathcal{F}$ of subsets of $\Omega = [0, 1]$ containing the algebra $\mathcal{F}_0$, for which the outer measure $P^* \ : \ \mathcal{F} \to [0, 1]$ is the unique probability measure that coincides with P. This probability measure assigns to each interval in $[0, 1]$ its length as its probability and is called the *Uniform* probability measure.

A probability space becomes decidedly more interesting when we define a random variable on it, as we shall now see.

Measure & Probability Theory

In broad terms a random variable is just a numerical translation of the outcomes of the sample space.

Example 2.5: Toss a coin

Here $\Omega = \{H, T\}$ and the σ-algebra is , with corresponding probability measure .

$$\text{Define} \quad X(\omega) \quad = \quad \begin{cases} 1 & \text{if} \quad \omega = H \\ 0 & \text{if} \quad \omega = T \end{cases}$$

For an arbitrary Borel set B we have $P\big(\{\omega \in \Omega : X(\omega) \in B\}\big)$

Example 2.5: Toss a coin (cont)

In this example, the set $\{\omega \in \Omega : X(\omega) \in B\}$ is automatically equal to one of the elements of $\mathcal{F}$ and so $P\big(X(\omega) \in B\big) = P\big(\{\omega \in \Omega : X(\omega) \in B\}\big)$ is well defined. In general however, we need to confine our random variables $X : \Omega \to \mathbb{R}$ to those for which we can make probability statements about events of the type $\{\omega \in \Omega : X(\omega) \in B\}$, where B is an arbitrary Borel set, which is only possible if these sets are members of $\mathcal{F}$. $\triangleleft$

Definition 2.20: Random variable

If $\{\Omega, \mathcal{F}, P\}$ is a probability space, then a mapping $X : \Omega \to \mathbb{R}$ is called a random variable defined on $\{\Omega, \mathcal{F}, P\}$, if X is $\mathcal{F}$-measurable. This means that for every Borel set B, $\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$.

Theorem 2.8

A mapping $X : \Omega \to \mathbb{R}$ is $\mathcal{F}$-measurable if and only if for all $x \in \mathbb{R}$, the sets $\{\omega \in \Omega : X(\omega) \leq x\}$ are members of $\mathcal{F}$.

In the previous example 2.4, consider that I set $X(\omega) = 1$ if $\omega \in \Omega = [0, 1]$ is rational and $X(\omega) = 0$ otherwise.

Would such a random variable be $\mathcal{F}$-measurable? The answer in short is no!

Definition 2.21: Generated σ-algebra

If X is a random variable, then define $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. Then the σ-algebra $\mathcal{F}_X = \{X^{-1}(B), \text{ for all } B \in \mathcal{B}\}$, is called the σ-algebra generated by X.

In our coin tossing example 2.5, the mapping X is one-to-one, and therefore $\mathcal{F}_X = \mathcal{F}$, but in general $\mathcal{F}_X$ is much smaller than $\mathcal{F}$.

Example 2.6: Roll a die

Consider the roll of a die where we set X to be 1 if the die shows an even face or 0 if it shows an odd face. Then

Example 2.6: Roll a die (cont)

You now decide to conduct a further investigation where you need to use the random variable Y which has value 1 if the die shows a number less than 4 and value 0 otherwise. Thus

Of course, Y is $\mathcal{F}_Y$ - measurable, but Y is *not* $\mathcal{F}_X$ - measurable. Therefore, if you only record X, you cannot say anything about Y, and vice-versa. To be able to answer questions about both X and Y, you need to work with the σ-algebra generated by

Moral: You need to know what investigation you need to do *before* conducting the experiment, so that you know what information you need to record. Otherwise you will not have the information necessary to make meaningful probability statements. ◁

Definition 2.22: Induced probability measure

Given a random variable X, define for arbitrary $B \in \mathcal{B}$

$$\mu_X(B) \;=\; P\left(X^{-1}(B)\right) \;=\; P\left(\left\{\omega \in \Omega : X(\omega) \in B\right\}\right) \;.$$

Then $\mu_X(B)$ is a probability measure on $\{\mathbb{R}, \mathcal{B}\}$ induced by the random variable X, where

(a) for all $B \in \mathcal{B}$, $\mu_X(B) \geq 0$,

(b) $\mu_X(\mathbb{R}) = 1$ and

(c) for all disjoint $B_j \in \mathcal{B}$, $\mu_X\left(\bigcup_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} \mu_X(B_j)$.

Therefore, the random variable X maps the probability space $\{\Omega, \mathcal{F}, P\}$ into a new probability space $\{\mathbb{R}, \mathcal{B}, \mu_X\}$, which in turn is mapped back by X^{-1} into (the possibly smaller) probability space $\{\Omega, \mathcal{F}_X, P\}$.

Definition 2.23: Distribution function

If X is a random variable, which induces probability measure μ_X,
the function $F(x) = \mu_X\big((-\infty, x]\big)$, for $x \in \mathbb{R}$
is called the distribution function of X.

Theorem 2.9

A distribution function of a random variable is always right continuous and monotonic non-decreasing. That is, for all $x \in \mathbb{R}$,

$$\lim_{\delta \downarrow 0} F(x+\delta) \quad = \quad F(x) \quad \textit{and}$$

$$F(x_1) \leq F(x_2) \quad \textit{if} \quad x_1 < x_2, \quad \textit{with}$$

$$\lim_{x \downarrow -\infty} F(x) = 0 \quad \textit{and} \quad \lim_{x \uparrow \infty} F(x) = 1 \ .$$

References

For those of you who are captivated by measure theoretic probability and would like to take it further, the following books are useful.

1. *A First L$\tilde{\odot}\tilde{\odot}$k at Rigorous Probability Theory*
   Jeffrey S. Rosenthal, World Scientific Press, (2006).

2. *Probability : Theory and Examples*
   Rick Durrett, Duxbury Advanced Series, (2005).

3. *Introduction to the Mathematical and Statistical Foundations of Econometrics*
   Herman J. Bierens, Cambridge University Press (2004)

We will now leave the "sinister" and consider the "dexter" in Section 3, although at times we will make backward reference to the former.

Course outline

- **Section 00:** Probability Review *(Assumed Knowledge)*
- **Section 01:** Introduction
- **Section 02:** Probability and measure
  - Sample space
  - Algebras and σ-algebras of events
  - Probability measure
- **Section 03:** Discrete Time Markov Chains
  - Basic definitions
  - Hitting probabilities and hitting times
  - Classification of states, recurrence and transience
  - Branching processes
  - Limiting behaviour
- **Section 04:** Martingales
  - Definition of a Martingale
  - Stopping times and the Optional stopping theorem
- **Section 05:** Brownian motion

DTMCs

In this part of the course, we consider a particular class of random process, known as a Markov chain. As we have already mentioned, Markov chains have the property of being "memoryless", which means that the next state that is visited by the Markov chain depends only on the present state, and not on any previous states.

We begin by providing some general definitions, then go on to define a discrete time Markov chain.

Definition 3.24: Random processes (stochastic processes)

> A **random process** is a family, or sequence, of random variables X_n, where n is a parameter running over a suitable index set T.

Each $X_n, n \in T$, is a (measurable) function which maps the sample space Ω into a state space $\mathcal{S}$.
For any fixed ω, there is a corresponding family $\{X_n(\omega), n \in T\}$ of members of $\mathcal{S}$. This family is called a **realisation** or **sample path** of X at ω.

DTMCs

Let's consider the case where the index n corresponds to discrete units of time, so that $T = \{0, 1, 2, \ldots\}$. This type of random process is therefore said to occur in **discrete time**. Furthermore, we restrict our attention to discrete random variables, which always have a countable state space.

Let $X_n, n \in \mathbb{N}$, be a sequence of random variables with a countable state space $\mathcal{S}$.

Definition 3.25: Discrete time Markov chain (DTMC)

The discrete time random process $X_n, n \in \mathbb{N}$, is a DTMC if it satisfies

$$\begin{aligned} \mathrm{P}(X_n = s \,|\, X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) \\ = \mathrm{P}(X_n = s \,|\, X_{n-1} = x_{n-1}) \end{aligned} \tag{3.17}$$

for all $n \geq 1$ and all $s, x_0, x_1, \ldots, x_{n-1} \in \mathcal{S}$.

That is, the probability of moving from state x_{n-1} to state s depends only on the state x_{n-1}, and not on the history of the process $X_0, X_1, \ldots, X_{n-2}$ (i.e.: how the process got to state x_{n-1}). This is the "memoryless" nature of Markov processes.

DTMCs

Since $\mathcal{S} = \{x_0, x_1, x_2, \ldots, x_i, \ldots\}$ is countable, it can be put into a one-to-one correspondence with some subset of the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$.

Without loss of generality, we can assume that the state space, $\mathcal{S}$, is a subset of $\mathbb{N}$.

This simplifies our notation considerably as we essentially can give up the physical interpretation of the states for a numerical one, and we can simply write more general expressions like

$$\mathrm{P}(X_{n+1} = j \mid X_n = i), \quad \text{for } i, j \in \mathcal{S}\ .$$

The **transition probabilities**

$$\mathrm{P}(X_{n+1} = j \mid X_n = i), \quad \text{for } i, j \in \mathcal{S}$$

govern the evolution of the Markov chain, that is, the evolution of the sequence $X_1, X_2, \ldots$ from some starting state X_0.

Note that the "starting state" X_0 can be specified deterministically, or it can be chosen randomly from some distribution across $\mathcal{S}$.

Time Homogeneous DTMCs

Definition 3.26: Time homogeneous Markov chain

A Markov chain X_n is called **time-homogeneous** if we have

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) \;=\; \mathrm{P}(X_1 = j \mid X_0 = i) \text{ for all } n \text{ and } i, j \in S.$$

in which case we write

In contrast, if the transition probabilities depend on the current time n as well as on i and j, then the Markov chain is called time-inhomogeneous.

Definition 3.27: Transition matrix

The **transition matrix** $\mathbb{P}$ of a time-homogeneous Markov chain is the $|\mathcal{S}| \times |\mathcal{S}|$ matrix of transition probabilities

$$p_{i,j} = \mathrm{P}(X_{n+1} = j \mid X_n = i) \qquad \rightarrow \qquad \mathbb{P} = [p_{i,j}].$$

DTMCs

Transition probabilities satisfy the following properties

1. $0 \leq p_{i,j} \leq 1, \qquad \forall i, j \in \mathcal{S}$
2. $\sum_{j \in \mathcal{S}} p_{i,j} = 1, \qquad \forall i \in \mathcal{S}$ (the row sum is 1)

Definition 3.28: m-step transition matrix

The **m-step transition matrix** $\mathbb{P}^{(m)} = [p_{i,j}^{(m)}]$ of a time-homogeneous Markov chain is the $|\mathcal{S}| \times |\mathcal{S}|$ matrix of transition probabilities

$$p_{i,j}^{(m)} = \mathrm{P}(\, X_{n+m} = j \mid X_n = i) \,.$$

The m-step transition probability $p_{i,j}^{(m)}$ is the probability that the process starting in state i at time n, finds itself in state j at time $n + m$.

In general, there are multiple possible "paths" that result in this outcome.

The probability $p_{i,j}^{(m)}$ is essentially the sum of probabilities of all such paths.

DTMCs

In order to have a complete picture of the evolution of our Markov process, a matrix of such probabilities is somewhat fundamental to our needs, but we do not wish to measure the $p_{i,j}^{(m)}$ for all values of m. The following theorem gives us the answer without having to separately establish the m-step probabilities.

Theorem 3.10: Calculating the m-step transition matrix

$$\mathbb{P}^{(m)} = \mathbb{P}^m \ .$$

This theorem is useful because it tells us that the entry (i, j) of the matrix $\mathbb{P}^m$ is equal to the m-step transition probability $p_{i,j}^{(m)}$.

A proof of this theorem follows, which gathers together everything that we have discussed in this course to date and in some sense covers many things we need to know about Markov chains. The method of proof is logically constructed from the actual statement of the theorem and should be well understood. For example, we use induction on m because the thing we wish to prove has a general pattern in terms of powers that sits naturally with such a method of proof.

Proof of Theorem 3.10: We use induction on m.

Since it is true for $m = 1$, we assume it is true for $m - 1$, so $\mathbb{P}^{(m-1)} = \mathbb{P}^{m-1}$.
We then condition $p_{i,j}^{(m)}$ on the state after $m - 1$ steps, so that

DTMCs

Example 3.10: Numerical example using Theorem 3.10.

If $\mathbb{P} = \begin{pmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$, *what is* $p_{1,2}^{(3)}$ *?* *(row sums and col sums)*

By Theorem 3.10 we have that $\mathbb{P}^{(3)} = \mathbb{P}^3$, *so that*

Example 3.10: (cont)

What about $p_{1,1}^{(n)}$ for general values of n ? We can use a program like Matlab to plot $p_{1,1}^{(n)}$ versus n for a series of values of n.

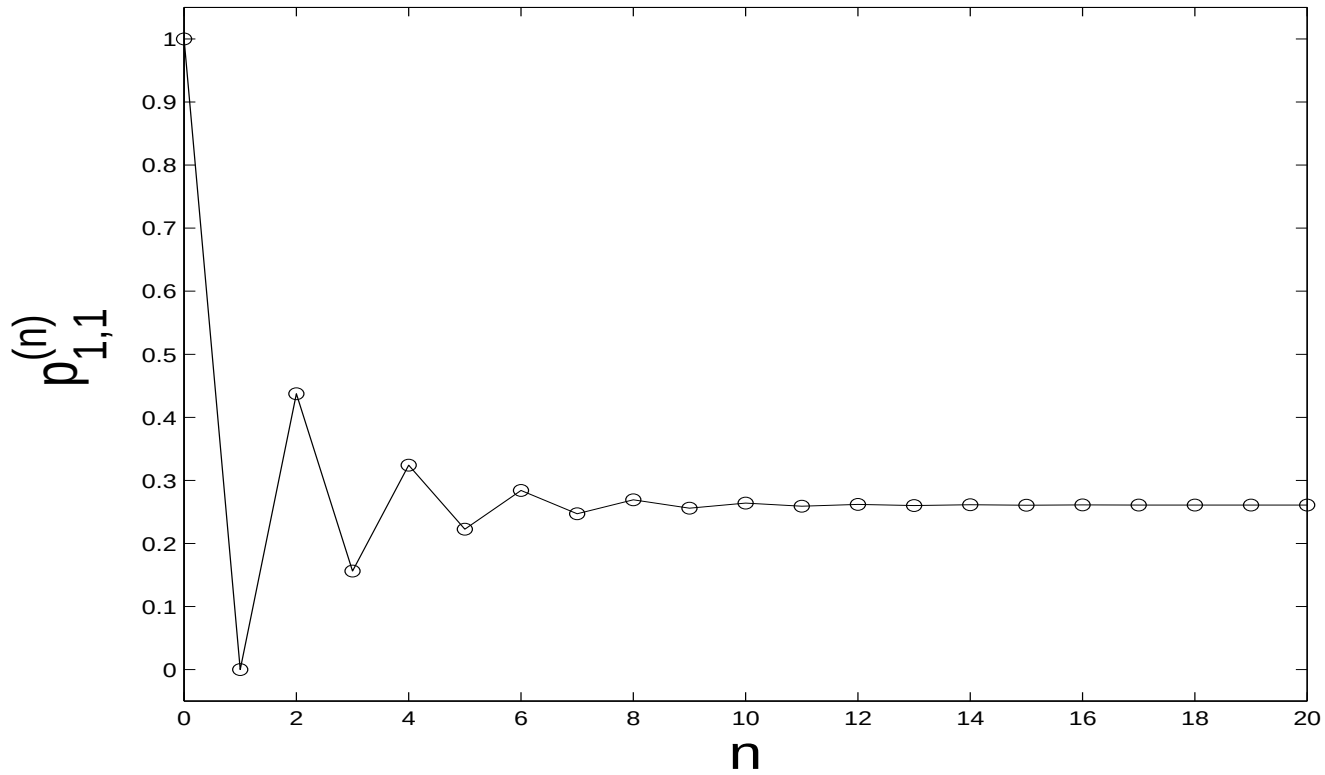


Figure 3.1: Plot of $p_{1,1}^{(n)}$ versus n for $n = 0, 1, \ldots, 20$.

The two gamblers

Example 3.11: Two gamblers

Player A and Player B are engaged in a contest comprised of a series of games, in which Player A has probability p of winning any given game, and Player B has probability $q = 1 - p$ of winning any given game.

Each time a player wins, the player's fortune is increased by one dollar, and conversely, the player's fortune decreases by one dollar if the player loses.

The player's combined fortune is N dollars. Player A starts with a fortune of k, and Player B starts with a fortune of $N - k$.

The players stop playing when one of them has lost all of their money.

There is a definite advantage to Player A in each individual trial if $p > q$, and a definite disadvantage if $p < q$.

If $p = q$ then the contest is regarded as "fair".

The two gamblers

Example 3.11: (cont)

Consider the Markov chain X_n representing Player A's fortune after n games, for $n \in \mathbb{N}$.

The state space of the process X_n is clearly

Player A is ruined when the state of the process is 0, and Player B is ruined when the state of the process is N.

The transition probabilities for Player A's fortune for $0 < i < N$, are given by

The two gamblers

Example 3.11: (cont)

States 0 *and* N *are called "absorbing states", because once the process enters one of these states, it cannot leave.*

The transition probability matrix for the process X_n *is the* $(N+1) \times (N+1)$ *matrix*

The two gamblers

Example 3.11: (cont)

The future value of a player's fortune generally does depend on the player's history. For example, if we know that $X_5 = 1$, then the probability that Player A is ruined before Player B will be higher than if we know that $X_5 = N - 1$. However, the fortune of Player A is still a Markov chain because the

fortune

does not depend on the history

given the

That is, they do not depend on the history of wins and losses, but only depend on the player's current fortune.

Note that the fortune of Player A (or Player B) is a

Markov chain, because the transition probabilities do not depend on the time index $n \in \mathbb{N}$.

The two gamblers

Example 3.11: (cont)

What is the probability that Player A eventually loses all of her money (i.e. is ruined before Player B)?

We can answer this by letting W be the event that Player A eventually loses all of her money, and then let

We can immediately identify the "boundary conditions" that are relevant for this problem:

In order to obtain u_i for general values of i, we employ a first step analysis.

The two gamblers

Example 3.11: (cont)

Here, we make use of the Law of Total Probability to simplify the calculation. To do so, we need to chose a suitable set of partitioning events. In this case, we use the events defined by the state at the end of the first transition.

In our problem, for $0 < i < N$ *we have*

Analysis of the two gamblers

So now we have the system of difference equations

$$u_i = pu_{i+1} + qu_{i-1}, \quad 0 < i < N, \tag{3.18}$$

with boundary conditions $u_0 = 1, u_N = 0$.

There are several possible approaches we can use to solve this system, but we shall apply a very systematic, standard method for solving them.

Note that (3.18) is a system of difference equations which are

1.

2.

3.

Difference equations can be thought of as the discrete analogue of differential equations.

Analysis of the two gamblers

Recall that for DEs, we "guess" a form of solution, then check that it satisfies the DE and it turns out that we can apply a similar technique with difference equations. Let's try a solution of the form

Substitute into (3.18) to get

divide through by w^{i-1} and rearrange to get the characteristic equation

The two roots of the characteristic equation are

Note that if $p = q$, we have a repeated root, so that we must study the cases $p \neq q$ and $p = q$ separately.

Analysis of the two gamblers

1. Case $p \neq q$: Recall the principle of superposition of solutions of linear DEs. This suggests a general solution of the form

(3.19)

Exercise: check that this satisfies (3.18).

Now use boundary conditions with (3.19) to determine the values of A_1 and A_2.

Substituting this back into (3.19), after a bit of algebra, we obtain

(3.20)

Analysis of the two gamblers

2. Case $p = q$: We have repeated roots $w_1 = w_2 = 1$ and so in a manner that is analogous to the solution of DEs, we guess

Exercise: check that this general form satisfies (3.18).
Now use boundary conditions to find A_1 and A_2.

The solution can then be written compactly for $0 \leq i \leq N$ as follows.

Analysis of the two gamblers

Consider the plots in Figure 3.2.

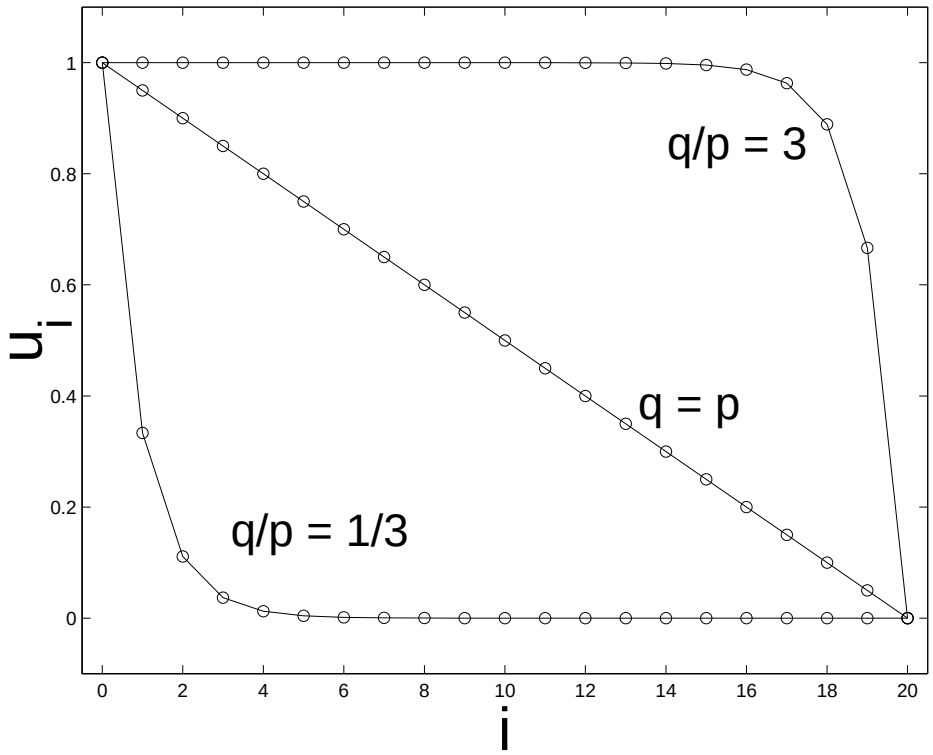


Figure 3.2: Plot of u_i versus i for $N = 20$

When Player A has a definite advantage, eg: $p = \frac{3}{4}, q = \frac{1}{4}, q/p = \frac{1}{3}$, then the probability u_i that A loses all of her money drops off dramatically for small values of i and is very low thereafter (and is zero at $i = N = 20$).

Analysis of the two gamblers

Conversely, when Player B has a definite advantage, eg: $p = \frac{1}{4}, q = \frac{3}{4}, q/p = 3$, then the probability u_i that A loses all of her money remains high for most values of $i < N$, and only drops off dramatically when i is very close to N.

Note that the event "Player A loses the contest" is equivalent to the event "Player B wins the contest", hence the symmetry observed in Figure 3.2.

Now, let us consider the probability that Player A eventually *wins* all of the money (that is, Player A wins the contest and B loses all of his money).

That is, we wish to find $v_i = \mathrm{P}(\, X_n \;$ reaches state N before state $0 \mid X_0 = i)$.

We could solve using the method of solution used above, with the appropriate modified boundary conditions or we could use a simple argument that exploits the symmetry of the problem. This is left as a class exercise question.
The answer is of course, that for $0 \leq i \leq N$,

Analysis of the two gamblers

We are now in a position to easily answer the following important question:

What is the probability that the contest eventually comes to an end?

This is equivalent to the question:

What is the probability that Player A eventually wins or loses all of the money?

Answer:

Note that while it is a *logical possibility* that the players can play forever:
(for example, it is possible to construct a sample path where the players take turns in winning, i.e.: win, lose, win, lose, ... , such that i never hits the boundaries),
the probability of the event that "the players play forever" is equal to zero.

Analysis of the two gamblers

Example 3.12: Gamblers again

We found in Example 3.11 that the contest eventually comes to an end with probability 1.

What is the expected duration of the contest, given that Player A has initial fortune i?

Let H denote the (random) time until Player A wins or loses the contest (that is, the process reaches state 0 or state N), and let

Note that H is a random variable taking values on the countably infinite state space $\{0, 1, 2, \ldots\} \cup \{\infty\}$. Instead, k_i is its (conditional) expectation, and is thus a constant value, not a random variable.

We solve this problem by employing a first step analysis along the lines of Example 3.11.

Analysis of the two gamblers

Example 3.12: (cont)

In particular, the expectations k_i satisfy the system of difference equations

(3.21)

The "1" on the RHS of (3.21) is the unit of time spent in playing one game, (which is also the minimum amount of time which must be spent playing if the contest has not yet ended).
This is a system of second order non-homogeneous linear difference equations.

Writing

(3.22)

we see that if the RHS were zero, this would be identical to the homogenous equation (3.18).

The "-1" on the RHS makes this an non-homogenous difference equation.

Analysis of the two gamblers

Example 3.12: (cont)

We construct a solution as follows (similar to differential equations):

[solution] = [general solution to the associated homogeneous equation] + [particular solution]

1. Case $p \neq q$: From example 3.11 the solution to the homogeneous equation is

(3.23)

We know that when substituted into the LHS of (3.22), the solution (3.23) yields zero and so the particular solution must provide the -1 when substituted into (3.22).
Strategy: try polynomial solutions $f_i = C_1 + C_2 i + C_3 i^2 + \ldots$, starting with the simplest possible choice: $f_i = C$ (constant). Substituting into (3.22) yields

Analysis of the two gamblers

Example 3.12: (cont)

Try the next simplest form of particular solution, $f_i = C_1 + C_2 i$.
Substituting into (3.22) yields

Note that a constant A_1 *occurs in the homogeneous form of solution and so* C_1 *not surprisingly cancels again. So the general form of the solution to the inhomogeneous equation is*

(3.24)

Analysis of the two gamblers

Example 3.12: (cont)

Finally, use the boundary conditions to determine A_1 *and* A_2.

Putting it all together, we have for $0 \leq i \leq N$,

$$k_i = \frac{1}{p-q}\left[N\frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} - i\right], \quad p \neq q. \tag{3.25}$$

Analysis of the two gamblers

Example 3.12: (cont)

When player A has the advantage ($\frac{q}{p} = \frac{1}{3}$), the expected duration is maximised for low values of i (see LHS of figure 3.3 below).
What do we expect when the game is fair (i.e. when $p = q$)?

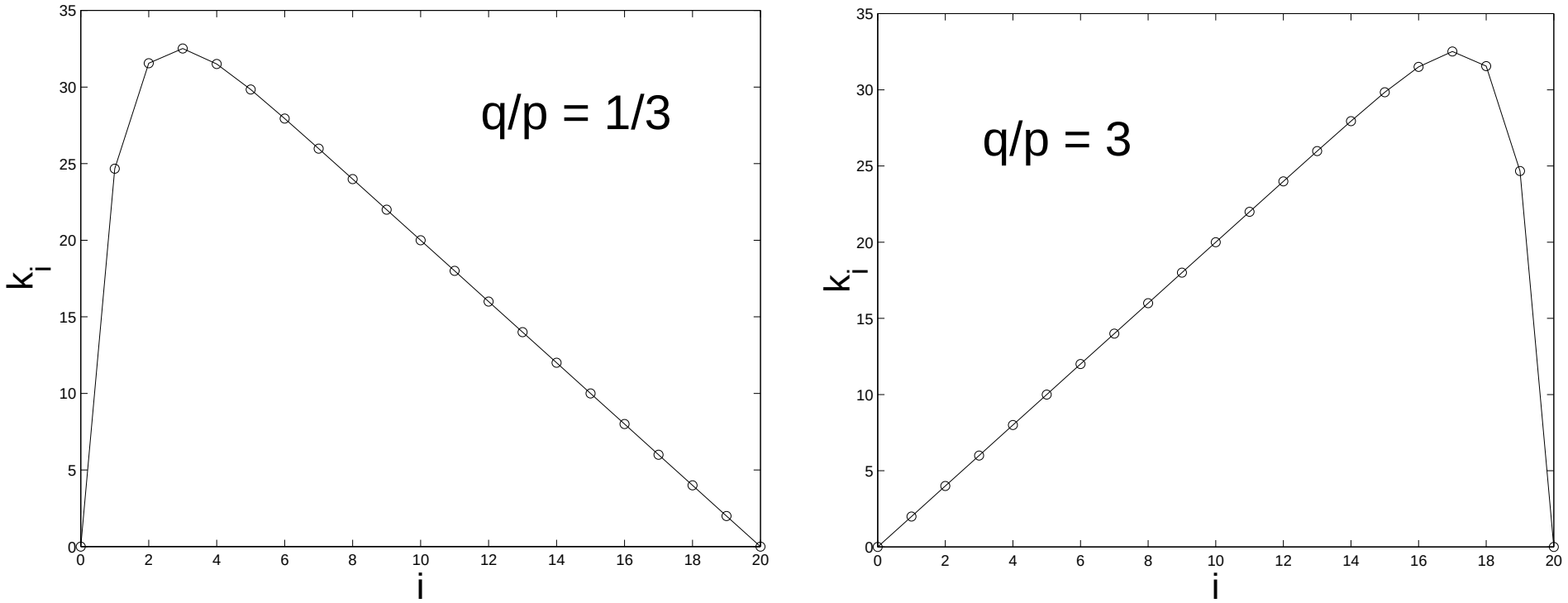


Figure 3.3: Expected duration of contest, k_i starting in state $i = 1, 2, \ldots, 20$.

Random walks

The gambling example is a special case of a class of Markov chain known as **random walks**.

Instead of representing the fortune or capital of a player, the state i can represent the position of a particle or anyother quantity you like.

Other examples include random walks on

- $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$
- $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2 \ldots\}$
- $\mathbb{Z}^2$, or higher dimensions $\mathbb{Z}^n$.

Hitting probabilities and hitting times

The probabilities u_i that we obtained in Example 3.11 were examples of hitting probabilities. In particular, u_i corresponds to the probability of hitting state 0 (an absorbing state) before hitting state N (also an absorbing state), given that the starting state is i.
Also, the random time H in Example 3.12 is an example of a hitting time, that is, the time at which the process first "hits" state 0 or N.

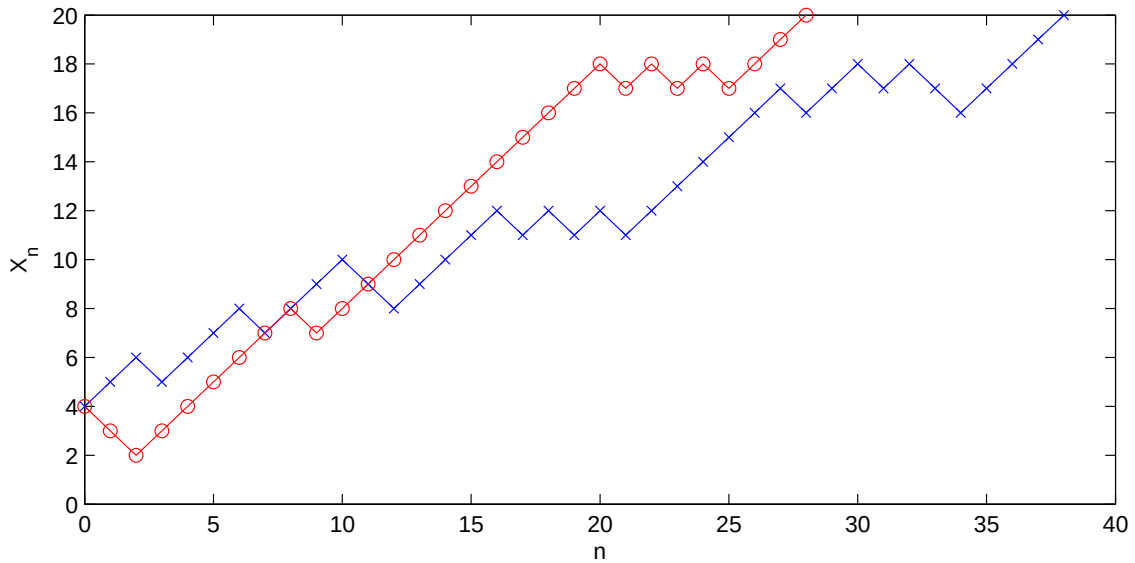


Figure 3.4: Two sample paths from $X_n = 4$ with $p = 3/4$

In Example 3.12, we solved the system of difference equations for the expectations of these hitting times, labelled $k_i = E(H|X_0 = i)$.

Hitting probabilities and hitting times

We now define the notion of hitting probabilitiy and hitting time in more general terms.

Let $X_n, n \in \mathbb{N}$, be a Markov chain with transition matrix $\mathbb{P}$, and state space $\mathcal{S}$. Let $\mathcal{A}$ be a subset of $\mathcal{S}$, which contains the state or states for which we want to know the hitting probability.

Definition 3.29: Hitting time

The **hitting time** of a subset $\mathcal{A}$ of $\mathcal{S}$ is the random variable

$$H^{\mathcal{A}} : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}, \text{ given by}$$

$$H^{\mathcal{A}} = \inf\{n \geq 0 : X_n \in \mathcal{A}\}.$$

Note, in this case, Ω corresponds to the set of all possible sample paths. We use the "infimum" $(\inf)$ in the above definition instead of the "minimum" $(\min)$, because in general we cannot be sure, for all sample paths, that such a value of n exists. We are therefore forced to use the convention that the infimum of the empty set $\emptyset$ is ∞.

Hitting probabilities and hitting times

Note that in the "Two Gamblers" problem (Example 3.12), we had $\mathcal{A} = \{0, N\}$, and we knew for sure that there exists a finite value of n such that $X_n \in \mathcal{A}$, because we showed that with probability one, the contest ends after a finite number of games (time-steps).

Definition 3.30: Hitting probability

The **hitting probability** *of a subset $\mathcal{A}$ of $\mathcal{S}$ is defined as*

$$u_i^{\mathcal{A}} = P\big(H^{\mathcal{A}} < \infty \,\big|\, X_0 = i\big)\,.$$

In words, the hitting probability $u_i^{\mathcal{A}}$ is the probability that the process ever "hits" (visits) some state $j \in \mathcal{A}$, given that it starts in state $i \in \mathcal{S}$.

When all of the states in the set $\mathcal{A}$ are absorbing states (or when $\mathcal{A}$ is an absorbing class - see later), then $u_i^{\mathcal{A}}$ is a special case of a hitting probability which is often referred to as an **absorption probability**.

Hitting probabilities and hitting times

Note that in the first "Two Gamblers" Example (3.11), we first set $\mathcal{A} = \{0\}$, and $u_i = u_i^{\{0\}}$ was the absorption probability for state 0 starting from state i.
We then chose $\mathcal{A} = \{N\}$, and $v_i = u_i^{\{N\}}$ was the absorption probability for state N starting from state i.
Note that $\mathcal{A}$ need not consist of absorbing states - we can investigate the hitting probability for any state or set of states.

Theorem 3.11: Minimal non-negative solution for hitting probabilities

The hitting probabilities $u_i^{\mathcal{A}}, i \in \mathcal{S}$, are given by the minimal non-negative solution to the system of equations

$$u_i^{\mathcal{A}} = \begin{cases} 1 & \text{for} \quad i \in \mathcal{A} \\ \sum_{j \in \mathcal{S}} p_{i,j} u_j^{A} & \text{for} \quad i \notin \mathcal{A}. \end{cases} \tag{3.26}$$

In general terms, this means that if $x_i, i \in \mathcal{S}$, is another solution to (3.26) with $x_i \geq 0$ for all i, then $x_i \geq u_i$ for all i.

Hitting probability

Proof of Theorem 3.11:

First we show that $u_i^{\mathcal{A}}$, $i \in \mathcal{S}$, satisfy (3.26).

If $X_0 = i \in A$, then $H^{\mathcal{A}} = 0$, so $u_i^{\mathcal{A}} = 1$.

If $X_0 = i \notin \mathcal{A}$, then $H^{\mathcal{A}} \geq 1$, and we have

$$u_i^{\mathcal{A}} \quad = \quad \mathrm{P}\big(H^{\mathcal{A}} < \infty \,\big|\, X_0 = i\big) \qquad \text{which by the Law of Total Probability yields}$$

Hitting probability

Proof of Theorem 3.11 (cont)

Suppose now that $x_i, i \in \mathcal{S}$ is any solution to (3.26).

Then $u_i^{\mathcal{A}} = x_i = 1$ for $i \in \mathcal{A}$. For $i \notin A$, we have

since $\mathcal{A}$ and $\mathcal{A}^c$ form a partition of $\mathcal{S}$ (see Definition 0.3).

Substitute for x_j to obtain

Hitting probability

Proof of Theorem 3.11 (cont)

Substituting again for x in the final term, we obtain

$$
\begin{aligned}
x_i &= \sum_{j\in\mathcal{A}} p_{i,j} + \sum_{j\notin\mathcal{A}} p_{i,j} \left[\sum_{k\in\mathcal{A}} p_{j,k} + \sum_{k\notin\mathcal{A}} p_{j,k} \left(\sum_{\ell\in\mathcal{A}} p_{k,\ell} + \sum_{\ell\notin\mathcal{A}} p_{k,\ell} x_\ell \right) \right] \\
&= \mathrm{P}(\, X_1 \in \mathcal{A} \mid X_0 = i) + \mathrm{P}(\, X_1 \notin \mathcal{A}, X_2 \in \mathcal{A} \mid X_0 = i) \\
&\qquad + \mathrm{P}(\, X_1 \notin \mathcal{A}, X_2 \notin \mathcal{A}, X_3 \in \mathcal{A} \mid X_0 = i) + \sum_{j\notin\mathcal{A}} \sum_{k\notin\mathcal{A}} \sum_{\ell\notin\mathcal{A}} p_{i,j} p_{j,k} p_{k,\ell} x_\ell . \\
\vdots &\qquad \vdots \qquad \text{(continuing in this manner)} \\
&= \mathrm{P}(\, X_1 \in A \mid X_0 = i) + \ldots + \mathrm{P}(\, X_1 \notin \mathcal{A}, \ldots, X_{n-1} \notin \mathcal{A}, X_n \in \mathcal{A} \mid X_0 = i) \\
&\qquad + \sum_{j_1\notin\mathcal{A}} \cdots \sum_{j_n\notin\mathcal{A}} p_{i,j_1} p_{j_1,j_2} \cdots p_{j_{n-1},j_n} x_{j_n} . \qquad (3.27)
\end{aligned}
$$

Hitting probability

Proof of Theorem 3.11 (cont)

Now if x is non-negative, so is the last term on the right of (3.27)

(since $p_{i,j} \geq 0$ for all i and j).

Furthermore, all of the other terms in (3.27) sum to $\mathrm{P}\big(H^{\mathcal{A}} \leq n \,\big|\, X_0 = i \big)$.

Hitting time

A similar result holds for the expected hitting times $k_i^{\mathcal{A}}$.

As before, let $\mathcal{A}$ be a subset of $\mathcal{S}$.

Theorem 3.12: Minimal non-negative solution for hitting times

The mean hitting times $k_i^{\mathcal{A}}, i \in \mathcal{S}$, are given by the minimal non-negative solution to the system of equations

$$k_i^{\mathcal{A}} = \begin{cases} 0 & \text{for} \quad i \in \mathcal{A} \\ 1 + \sum_{j \notin \mathcal{A}} p_{i,j} k_j^{\mathcal{A}} & \text{for} \quad i \notin \mathcal{A}. \end{cases}$$

Proof: This is very similar to the proof of Theorem 3.11, except that because we are dealing with units of "expected time" rather than with hitting probabilities, we must account for the unit of time that passes with each state transition. ■

Hitting probabilities and hitting times

In problems with a specific number of states, say $|\mathcal{S}| = 4$, the equations in Theorem 3.11 for hitting probabilities yield a system of simultaneous equations which we can solve explicitly using the usual techniques from linear algebra. The same applies for the expected hitting time equations given in Theorem 3.12.

The minimal non-negative solution specified in both of Theorems 3.11 and 3.12 does not come into play, since there is a unique solution to the finite system of linear equations.

That is, we get a system of linear equations in the unknowns u_i^A or k_i^A, $i \in \mathcal{S} \setminus \mathcal{A}$, and some boundary conditions corresponding to states $i \in \mathcal{A}$.
See class exercise/tutorial for an example.

How can we make use of the minimality property to find the hitting probabilities and expected hitting times?

The minimality property can be useful when the state space is not finite, and where there are multiple possible solutions from which we must choose, as demonstrated in the following example.

Hitting probabilities and hitting times

Example 3.13: Infinitely rich Player B

Consider a modified version of the contest in Example 3.11, in which Player B is infinitely rich.

The state space then becomes $\mathcal{S} = \{0, 1, 2, \ldots\}$.

As before, Player A has lost all of her money when the state of the process reaches 0.

On the other hand, Player B cannot ever be bankrupt, and there is no bound on how much money can potentially be won by Player A.

This could be an approximation to the situation where Player A with finite capital plays against a casino (Player B), which for all practical purposes, is infinitely rich compared to Player A ! (But casinos evict you, if ever you win too much money!)

As before, we associate the state of the process with the finite current capital, or fortune, of Player A.

Hitting probability

Example 3.13: (cont)

Let $u_i = u_i^{\mathcal{A}} = u_i^{\{0\}}$ *be the probability of absorption in state* 0*, given that the process starts in state* i*.*
Given our new state space $\mathcal{S} = \{0, 1, 2, \ldots\}$*, we have the system of difference equations*

Recall that in Example 3.11, the general solution of the system of difference equations (3.18) involved two coefficients A_1 *and* A_2*, and that we required two boundary conditions in order to completely determine the values of these coefficients.*

In the current problem, we are missing the second boundary condition ($u_N = 0$) *because there is no second boundary point !*

The result of Theorem 3.11 can help us in this situation.

Hitting probability

Example 3.13: (cont)

Consider the case $p \neq q$.

We found that the general form of the solution was

except that now, this is valid for all

The boundary condition $u_0 = 1$ *gives*

How can we determine A_2*, given that we do not have another boundary condition ?*

Hitting probability

Example 3.13: (cont)

Proceed as follows:

The constraints $0 \leq u_i \leq 1$ *for all* $i \geq 0$ *imply that*

Using Theorem 3.11, we find the minimal non-negative solution by selecting a value of A_2 *from the infinite number of possible values on the interval* $[0, 1]$.

In particular, since $-1 \leq \left(\left(\frac{q}{p} \right)^i - 1 \right) \leq 0$ *when* $q < p$, *then* u_i *is minimised (in the non-negative sense) over all possible choices of*

Hitting probability

Example 3.13: (cont)

On the other hand, if $q > p$*, then for all* $i \geq 1$*,*

The minimal non-negative solution is obtained by setting

Finally, consider the case $q = p$*.*
We found in Example 3.11 that the general form of the solution for the case of repeated roots $\omega_1 = \omega_2 = 1$ *is*

The boundary condition $u_0 = 1$ *gives* $A_1 = 1$*, so*

Hitting probability

Example 3.13: (cont)

Once again, the minimal non-negative solution is obtained by setting $A_2 = 0$*, so*

Note that we could also have deduced that $A_2 = 0$ *in this case just from the constraints* $0 \leq u_i \leq 1$.

Putting it all together, we have for $i \geq 0$*,*

This means that if Player B *(the casino) has a definite advantage (*$q > p$ *– the games are specified to ensure this), or if the contest is fair (*$p = q$*), then Player* A *will eventually lose all of her money with probability 1.*

Hitting probability

Example 3.13: (cont)

On the other hand, if Player A has a definite advantage ($q < p$), then there is a positive probability $1 - \left(\frac{q}{p}\right)^i$, that she plays forever and does not lose all of her money.

In particular, if Player A starts with 1 dollar, the probability that she escapes ruin forever is $1 - \frac{q}{p}$.

It may be surprising to find that Player A loses the contest with probability 1, in the case that the contest is fair.

This phenomenon is related to the fact that Player B is able to lose an arbitrarily large amount of money without going bankrupt (and hence ending the contest), whereas the same is not true for Player A.

We shall now work towards introducing the notions of recurrence and transience, which provide a framework within which to better understand this type of situation.

Classification of states

In this part, we will look at ways of classifying different states of a DTMC.

Consider a DTMC with transition matrix $\mathbb{P}$.

Definition 3.31: Accessible states

State j is said to be **accessible** *from state* i *if* $p_{i,j}^{(m)} > 0$ *for some integer* $m > 0$.

In words: state j is accessible from i if there is positive probability that state j can be reached from state i in a finite number of steps.

The notation $i \to j$ is used to signify that j is accessible from i.

Definition 3.32: Ephemeral states

If state i does not have access to itself, we say that i is **ephemeral.**

Such a state is never revisited beyond the time of any initial occupation. In general, we shall disregard such states.

Classification of states

Definition 3.33: Communicating states

States i and j are said to **communicate** *if each state is accessible from the other.*

The notation $i \leftrightarrow j$ is used to signify that i and j communicate.
Note: if i and j do not communicate, then either

$$p_{i,j}^{(m)} = 0 \text{ for all } m > 0, \; p_{j,i}^{(m)} = 0 \text{ for all } m > 0, \text{ or both.}$$

The concept of communication is an equivalence relation, which means that it satisfies the following three properties:

1. Excluding ephemeral states, $i \leftrightarrow i$ (reflexivity), is a consequence of the definition of an ephemeral state!
2. If $i \leftrightarrow j$, then $j \leftrightarrow i$ (symmetry), by the definition of communication.
3. If $i \leftrightarrow j$ and if $j \leftrightarrow k$, then $i \leftrightarrow k$ (transitivity).

Classification of states

Proof of transitivity :

$i \leftrightarrow j$ and $j \leftrightarrow k$ imply that there exist integers n and m such that

Conditioning on the n^{th} state, we can then write

Similarly, we can show that $p_{k,i}^{(r)} > 0$ for some integer r. ■

Since the relation $\leftrightarrow$ is an equivalence relation, it partitions the set of (non-ephemeral) states of the process into mutually disjoint classes and therefore

1.

2.

Classification of states

Note: suppose that two classes overlap, so that state $j \in C_1$ and $j \in C_2$, for some state j, and $C_1 \neq C_2$.
If k is any state in C_1 and i any state in C_2, we have $j \leftrightarrow k$ and $j \leftrightarrow i$.
By transitivity, we would then have $k \leftrightarrow i$ and so k and i must be in the same class, which implies $C_1 = C_2$, which contradicts our assumption $C_1 \neq C_2$.
Therefore classes do not overlap.

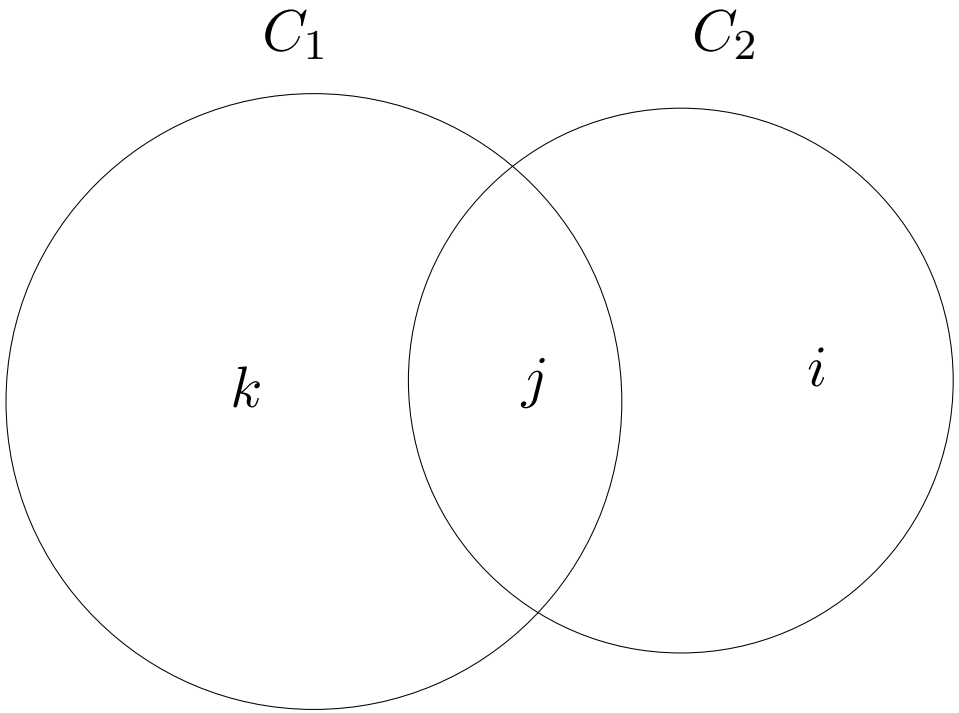


Figure 3.5: Communicating classes do not overlap.

Classification of states

Example 3.14: Communicating classes

Consider the six-state Markov chain with transition matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

To see what classes exist in this DTMC, we draw directional arrows on a state diagram to indicate "accessibility" and build up a picture of what is going on

(see next slide).

For example, the first column consists of all zeros, and so state 1 *is ephemeral. Certainly, therefore, state* 1 *can not communicate with itself.*

Classification of states

Example 3.14: (cont)

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Classification of states

Example 3.14: **(cont)**

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Classification of states

Example 3.14: (cont)

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

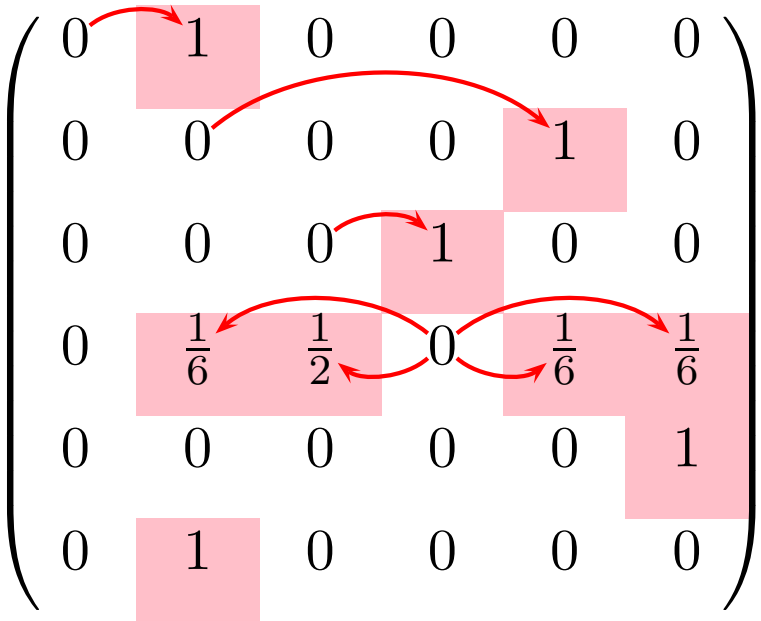
Classification of states

Example 3.14: **(cont)**

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

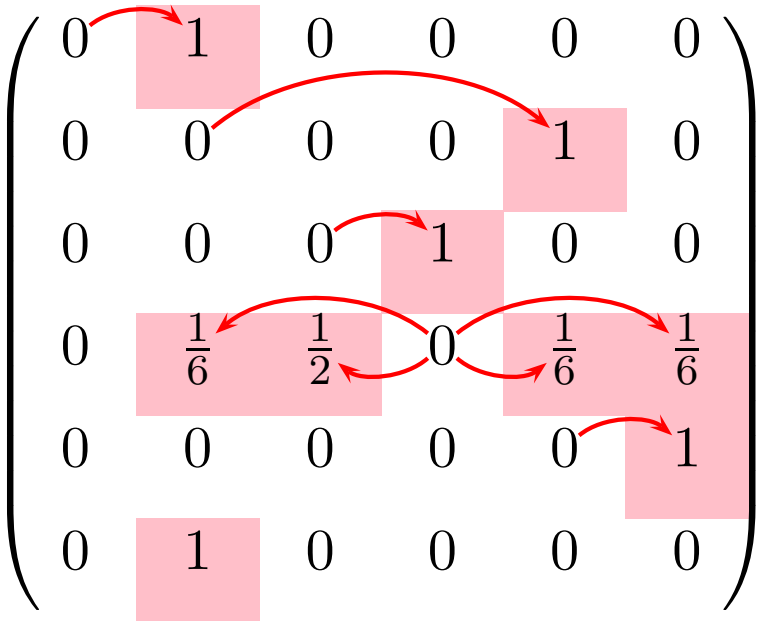
Classification of states

Example 3.14: (cont)



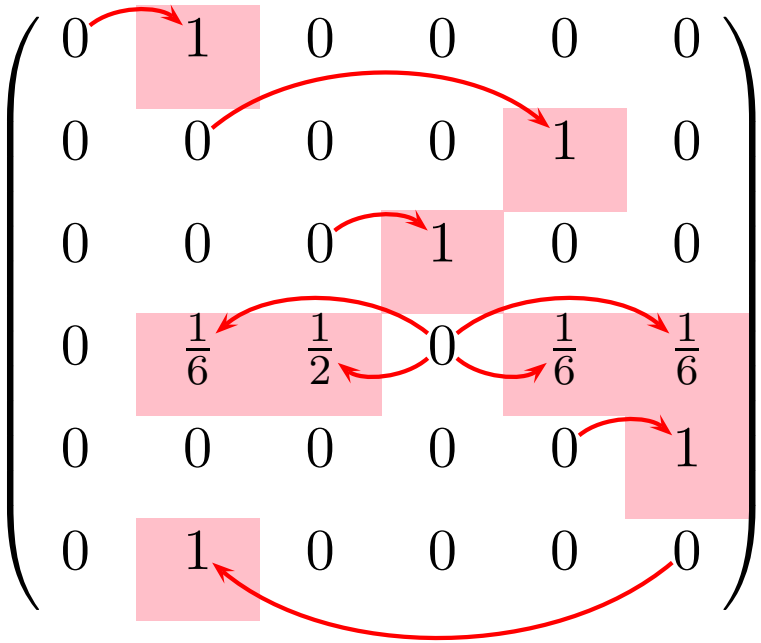
Classification of states

Example 3.14: (cont)



Classification of states

Example 3.14: (cont)



Classification of states

Example 3.14: (cont)

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

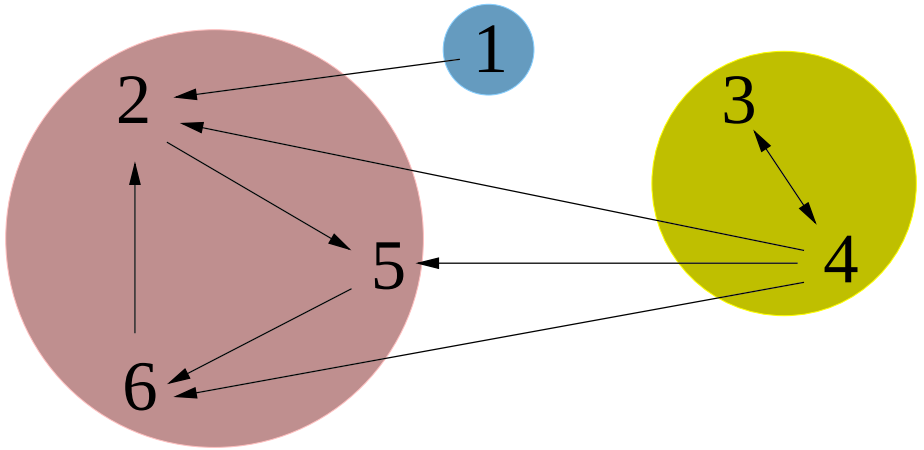


Figure 3.6: Communicating classes.

We see that states $\{3, 4\}$ *form a communicating class, and states* $\{2, 5, 6\}$ *form another communicating class.*

Note that once the Markov chain leaves the class $\{3, 4\}$*, it cannot return there. On the other hand, once the Markov chain enters the class* $\{2, 5, 6\}$*, it stays there forever.*
This motivates the concepts of recurrence and transience which we now explore.

Recurrence and transience

Informally, a recurrent state of a Markov chain is one which is revisited infinitely often, whereas a transient state is one which is only revisited a finite number of times, and is never revisited thereafter.

This classification is important, because every state of a Markov chain is either recurrent or transient.

From a practical point of view, this type of information can be very useful in the modelling of physical systems using Markov chains.

For example, suppose that in the earlier Gambling problem (Exercise 3.13) with Player B being infinitely rich, Player A has a rich uncle who gives Player A one dollar whenever she loses all of her money.

In this case, the "boundary point" 0 becomes a "reflecting barrier", rather than an absorbing state.

It is in the interest of the rich uncle to know whether Player A will "hit" the point 0 (i.e. lose all her money) an infinite number of times.

That is, is state 0 a recurrent or transient state?

Recurrence and transience

Let $f_{i,i}^{(n)}$ be the probability that starting from state i, the first return to state i occurs at the n^{th} transition, so that for a fixed state i, and for each $n \geq 1$

(3.30)

Physically, it makes sense that we also need $f_{i,i}^{(0)} = 0$ for all states i, because if no transitions have occurred ($n = 0$), then the probability of a revisit having occurred is trivially zero as well.
Define for an arbitrary, fixed state i,

(3.31)

We see that $f_{i,i}$ is the probability that when the process starts in state i, it returns to state i after a finite number of transitions.
That is, $f_{i,i}$ is the probability of ever revisiting state i.

Equivalently, $f_{i,i}$ is the probability that the process revisits state i *at least once*.

Recurrence and transience

We can now formally define recurrence and transience.

Definition 3.34: Recurrence and Transience

A state i is

$$\underline{\textbf{recurrent}} \quad \textit{if} \quad f_{i,i} = 1, \textit{ or}$$

$$\underline{\textbf{transient}} \quad \textit{if} \quad f_{i,i} < 1.$$

In addition, if for a transient state i, we have $f_{i,i} = 0$, then state i is **ephemeral**.

In the following theorem, we show that if a state is recurrent, then it will be occupied infinitely often with probability 1.

Theorem 3.13:

For a Markov process $X_n, n \geq 0$,

$$P(\, X_n = i \ \text{ for infinitely many } \ n \,|\, X_0 = i) \quad = \quad \begin{cases} 1 & \text{if state } i \text{ is recurrent,} \\ \\ 0 & \text{if state } i \text{ is transient.} \end{cases}$$

Recurrence and transience

To understand why this is true, consider a recurrent state i (hence $f_{i,i} = 1$ by definition), which means that the process revisits state i at least once with probability one.
Consider the first revisit, following which a second revisit also occurs with probability one, and so forth.
Since the Markov process is "memoryless", the dynamics of the process do not depend on the number of previous visits to state i (otherwise, the process would not be Markovian), and thus the process visits i infinitely often with probability one.
On the other hand, if i is transient, then $f_{i,i} < 1$, by definition.
Therefore, $1 - f_{i,i} > 0$, which means that there is a positive probability that the process *never* returns to state i.
If it does make a first return to state i (with probability $f_{i,i}$), then it once again has probability $1 - f_{i,i} > 0$ of never returning.

Because this is true following any visit to state i (due to the "memoryless" property of the Markov process), then eventually the process leaves state i and never returns.

Recurrence and transience

Proof of Theorem 3.13

Let $A_i^{m,j}$ be the *event* that the process visits state i *at least* m times after time $j \geq 0$ and define

Let E^k be the *event* that the first return to state i occurs at time $k > 0$, that is,

$$E^k \quad = \quad \{X_k = i,\ X_l \neq i,\ l = 1, \ldots, k-1\}, \qquad k > 0,$$

and so, by (3.30),

Clearly, the events E^k are disjoint.

Recurrence and transience

Proof of Theorem 3.13 (cont)

We begin by considering all possible realisations of the process for which $X_0 = i$, and the first return to state i occurs at time $k > 0$.

Partitioning on the time k of first return to i we have (by the Law of Total Probability) that

Note, this is a bit different to the first-step analysis seen in other proofs, where we partitioned on the state of the process after one transition.

Here, we have partitioned on the first return time k, but we can still employ the Law of Total Probability since the events E^k are disjoint.

Recurrence and transience

Proof of Theorem 3.13 (cont) Consider the first conditional probability.

Therefore, using (3.31) and proceeding recursively, we have

Recurrence and transience

Proof of Theorem 3.13 (cont)

But by definition

$$Q_i^1 = \mathrm{P}\Big(A_i^{1,0} \,\Big|\, X_0 = i \Big) = f_{i,i}$$

and recall that $f_{i,i}$ is the probability that the process revisits state i at least once after time 0.
Hence

Taking limits then gives

proving the result. ■

Recurrence and transience

The definition of recurrence or transience given above was in terms of $f_{i,i}$. Often, we can evaluate this easily by looking at the structure of the DTMC, as we did in Example 3.14.

Sometimes, however, the $f_{i,i}$ are difficult to calculate and we need a characterisation in terms of the one-step transition matrix $\mathbb{P}$.

The following theorem gives us a way of doing this.

Theorem 3.14:

If state i is transient, then

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} < \infty.$$

If state i is recurrent, then

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty.$$

Recurrence and transience

Proof of Theorem 3.14

The proof involves generating functions and a few preliminaries.

(3.32)

It is left as an exercise to show this. Hint: *use the Law of Total Probability and consider all possible realisations of the process for which* $X_0 = i$ *and* $X_n = i$*, and the first return to state* i *occurs at the* k^{th} *transition.*
Expression (3.32) is a *convolution*.
These are best handled using generating functions and so we define

(3.33)

(3.34)

Recurrence and transience

Proof of Theorem 3.14

We need to know that these are well-defined.
The following arguments show that they are.

Therefore, $P_{i,i}(z) < \infty$ for $|z| < 1$.

Recurrence and transience

Proof of Theorem 3.14 (cont)

Similarly, we have

Therefore, $F_{i,i}(z) < \infty$ for $|z| \leq 1$.

Recurrence and transience

Proof of Theorem 3.14 (cont)

Now, multiply (3.32) by z^n and sum from $n = 1$ to $n = \infty$.
This gives

Recurrence and transience

Proof of Theorem 3.14 (cont)

We are working towards writing everything in terms of the generating functions (3.33) and (3.34).
For the LHS, note that $p_{i,i}^{(0)} = 1$, so that

giving

We can now prove the theorem itself.
We exploit the fact that (we will justify this later)

and $f_{i,i} < 1$ if i is transient.

Recurrence and transience

Proof of Theorem 3.14 (cont)

Similarly, it follows that if i is recurrent (in which case $f_{i,i} = 1$), then

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty.$$

Recurrence and transience

Theorem 3.15: Transience theorem

In a DTMC, state i is transient if and only if

Proof: *(exercise, similar techniques as in the proof of Theorem 3.14.)*

Example 3.15: Random walk on $\mathbb{Z}$

Consider a DTMC with state space $\mathbb{Z}$ and transition probabilities

$$p_{i,j} = \begin{cases} p & \text{for } j = i+1 \\ q & \text{for } j = i-1 \\ 0 & \text{otherwise.} \end{cases}$$

with $p + q = 1, p > 0, q > 0$.

Recurrence and transience

Example 3.15: Random walk on $\mathbb{Z}$ (cont)

Think of a particle moving on the line according to the previous probabilities.

How many communicating classes are there?

The entire state space is a single communicating class.

Is state 0 *recurrent or transient ?*

By Theorem 3.14, state 0 *is recurrent if*

Recurrence and transience

Example 3.15: Random walk on $\mathbb{Z}$ (cont)

Let's think about $p_{0,0}^{(n)}$.

Clearly, $p_{0,0}^{(n)} = 0$ *if* n *is odd.*

The only number of transitions starting from 0 *for which it is possible to be back in state* 0 *are even numbers of transitions, so* $p_{0,0}^{(2n)} > 0, n = 1, 2, \ldots$

Specifically, of the $2n$ *transitions, we require a total of* n *transitions to the "right" and* n *transitions to the "left" in order to get back to where we started (at* 0*).*

The probability of each instance of this is $p^n q^n$.

We must then multiply this probability by the number of ways that the left moves and right moves can occur.

Therefore,

$$p_{0,0}^{(2n)} = \binom{2n}{n} p^n q^n.$$

Recurrence and transience

Example 3.15: Random walk on $\mathbb{Z}$ (cont)

Hence, state 0 *is transient if*

and is recurrent otherwise.

Recall the ratio test from first year, let $U_n = \binom{2n}{n} p^n q^n$.

The ratio test says that if

$$\begin{aligned}
&\lim_{n\to\infty} \left|\frac{U_{n+1}}{U_n}\right| < 1, && \text{then } U_n \text{ converges,} \\
&\lim_{n\to\infty} \left|\frac{U_{n+1}}{U_n}\right| = 1, && \text{then the test is inconclusive,} \\
&\lim_{n\to\infty} \left|\frac{U_{n+1}}{U_n}\right| > 1, && \text{then } U_n \text{ diverges .}
\end{aligned} \tag{3.35}$$

Recurrence and transience

Example 3.15: Random walk on $\mathbb{Z}$ (cont)

Now, $4pq = 4p(1-p)$ is a quadratic function which is equal to zero when $p = 0, 1$, and has a single maximum point when $p = q = \frac{1}{2}$, in which case $4pq = 1$. Hence, for $p \neq \frac{1}{2}$, the ratio test given in (3.35) says that the sequence $\{U_n\}$ is convergent and state 0 is transient.

Recurrence and transience

Example 3.15: Random walk on $\mathbb{Z}$ (cont)

When $p = \frac{1}{2}$, we have $4pq = 1$ and the ratio test does not tell us anything about the convergence of the series and we must try and use another test or comparison with a series for which we do know the convergence properties.

Recurrence and transience

Example 3.15: Random walk on $\mathbb{Z}$ (cont)

Note that in the last step we have used the fact that $n - \frac{1}{2} > n - 1$, $n - \frac{3}{2} > n - 2$, and so forth.

Therefore, we conclude that the sequence $\{U_n\}$ is divergent if $p = q = \frac{1}{2}$ and hence that state 0 is recurrent.

We may ask, in the random walk on $\mathbb{Z}$ example, what is special about state 0 ? After all, all states form part of the same communicating class.

With a little thought, it is not difficult to see that we could have chosen any other state $i \in \mathbb{Z}$ and derived the same conditions for transience and recurrence.

So in this Markov chain, all states are either transient or recurrent together.

It turns out that this is true for all communicating classes of a Markov chain.

Recurrence and transience

Theorem 3.16: Solidarity theorem

In a communicating class, either all states are recurrent together, or all states are transient together.

(Note: A property that is automatically shared by all states of a communicating class is called a *solidarity property*.)

Proof of theorem 3.16:

Let i, j be in the same communicating class C.
It suffices to prove that if i is recurrent then j is also recurrent.
Now,

which implies that there exist $m, n > 0$ such that

Recurrence and transience

Proof of Theorem 3.16

This then implies that

by picking out the specific term in the double sum corresponding to $h = k = i$, because any term $p_{rs}^{(t)} \geq 0$.

Therefore

because the middle term on the RHS diverges since i is assumed to be recurrent.

Recurrence and transience

Proof of Theorem 3.16

Therefore

Comment: $\sum_{n=1}^{\infty} p_{i,i}^{(n)}$ being finite/infinite is a useful theoretical way of expressing the fact that i is transient/recurrent.
However, it can have limitations for testing this in practical situations.

Theorem 3.17:

A Markov chain with a finite number of states has at least one recurrent state.

Recurrence and transience

Proof of Theorem 3.17

Suppose there are N states, $j = 1, \ldots N$, then we have

$$\sum_{j=1}^{N} p_{i,j}^{(n)} = 1 \ , \tag{3.36}$$

since the DTMC must be in *some* state j after n steps.

Recurrence and transience

Proof of Theorem 3.17 (cont)

Assume now that all states are transient, so that $\sum_{n=1}^{\infty} p_{i,j}^{(n)}$ converges for all i, j and

which contradicts (3.36).

Therefore, the DTMC has at least one recurrent state. ■

Irreducible Markov chains

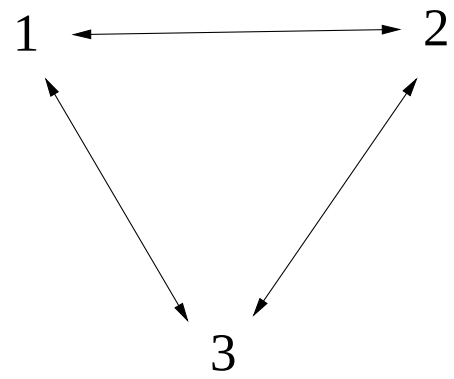
Definition 3.35: Irreducible DTMC

A DTMC is **irreducible** *if all states communicate with each other.*

In other words, a Markov chain consisting of a single communicating class is called irreducible.

Example 3.16:

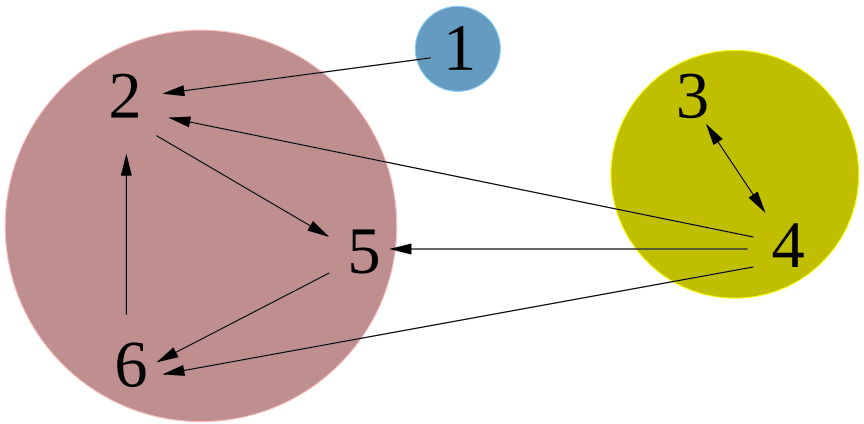
A three-state markov chain in which all possible transitions have a positive probability.
It is possible to reach any state from any other.
Since the Markov chain is irreducible, we cannot simplify the problem by focusing on a smaller subset of recurrent states.



Irreducible Markov chains

In contrast, if we add an absorbing state 4, then the Markov chain is not irreducible, since the absorbing state does not communicate with states $1, 2, 3$. However, the "trivial" Markov chain consisting of state 4 alone, with $p_{44} = 1$, is irreducible.

Recall Example 3.14 with transition diagram:



This is not irreducible, but the Markov chain consisting of only the states $\{2, 5, 6\}$ is an irreducible Markov chain.
We cannot isolate the class $\{3, 4\}$ because of the non-zero probabilities p_{42}, p_{45}, p_{46}.

However, if we re-define $p_{43} = 1$, say, then we cut the access between the classes and the new Markov chain on the states $\{3, 4\}$ is irreducible.
If we were to do this, we would have two distinct and independent Markov chains which we would treat as separate problems.

Irreducible Markov chains

Theorem 3.18:

An irreducible finite-state DTMC is recurrent

Proof of Theorem 3.18

Theorem 3.17 says that a finite-state DTMC has at least one recurrent state.
Theorem 3.16 says that recurrence is a solidarity property, ie: if one state in a communicating class is recurrent, then all states in the class are recurrent.
Theorem 3.18 is a direct consequence of these two previous theorems. ■

Definition 3.36: Closed communicating class

A communicating class C is **closed** if

$$\sum_{j \in C} p_{i,j} = 1, \qquad \text{for all } \; i \in C.$$

That is, no state in C has access to any state outside of C.
Note: An absorbing state is a special case, consisting of just one state.

Recurrence

In the "Two Gamblers" Example 3.11, the (singleton) class $\{0\}$ is a closed communicating class, as is $\{N\}$. Since the states 0 and N do not communicate, then they do not form a closed class together.

In Example 3.14, the states $\{2, 5, 6\}$ form a closed class, because once the process enters this class it cannot leave. Note however, that the process can continue to move around on the states within the class.

Theorem 3.19:

If C is a recurrent class, then it is closed.

Proof of Theorem 3.19

If C is not closed, then $p_{i,j} > 0$ for some $i \in C$ and $j \notin C$, so $i \to j$ (i has access to j).
We cannot have $j \to k$ for any $k \in C$, otherwise j would be in C,

Recurrence

Proof of Theorem 3.19 (cont)

So if C is not closed, this implies that there is a positive probability of never returning to i if the process leaves C via state $j \notin C$, so that $f_{i,i} < 1$, which means that i is transient.

Since transience is a solidarity property (by Theorem 3.16), then C must be transient, which contradicts the original assumption of this theorem.

Therefore, if C is recurrent, then it is closed. ■

It is important to note that the converse of Theorem 3.19 is not always true.

For instance, in the random walk on $\mathbb{Z}$ (Example 3.15), the entire state space $\mathbb{Z}$ forms a closed communicating class (since the process stays in $\mathbb{Z}$ forever), but in the case $p \neq q$, all states are transient, so in this example, it is not true that closed implies recurrent.

Branching Processes

We are now going to consider a particular class of DTMC known as Branching Processes. They have wide application in biology, sociology and engineering.

Consider a population consisting of individuals able to reproduce and suppose that each individual will, by the end of its lifetime, have produced j new offspring with probability $p_{1,j}$ for $j \geq 0$, independently of all other individuals.

This means that the lines of descent from different individuals of the same generation are independent.

Let X_0 be the number of individuals initially present, known as the zeroth generation. All their offspring form the first generation, and their number is denoted X_1.

In general, let X_n denote the size of the n^{th} generation.

Suppose that $p_{1,0} > 0$, or that the probability of an individual producing zero offspring is greater than zero and hence that a line of descent of an individual and hence a population of individuals may become extinct.

Branching Processes

To find the probability that the line of descent from an individual eventually becomes extinct, we may model the population process on a state space

This process has a countable collection of states, changes only occur at a countable collection of time points corresponding to the $1^{st}, 2^{nd}, \ldots$ generation, the process is memoryless and time homogeneous.

Hence;

- this process may be modelled as a Discrete Time Markov Chain.
- the probability of extinction is then a hitting probability on state 0.
- the time to extinction is also a hitting time on state 0.

Branching Processes

Assuming that $X_0 > 0$, we have the following cases arising

1. $(\mathbf{0 < p_{1,0} < 1})$: State 0 is a class on its own, it is a recurrent and absorbing state because if there are no individuals, there can be no offspring. In the population sense, state zero represents extinction.
   The remainder of the states $\{1, 2, 3, \dots\}$ form a single transient class such that there is a positive probability that the population will leave this set at the end of the n^{th} generation. That is, using the independence of each individual in the n^{th} generation, $(p_{1,0})^{X_n} > 0$.

2. $(\boldsymbol{p_{1,0} = 1})$: The trivial case where state 0 will be immediately occupied at the first generation, so that state 0 is again a recurrent and absorbing state, except now the states $\{1, 2, 3, \dots\}$ are all ephemeral.

3. $(\boldsymbol{p_{1,0} = 0})$: State 0 is unreachable, the states $\{1, 2, 3, \dots\}$ are transient if $p_{1,1} < 1$, else they are all individual absorbing classes, but either way there is no possibility of extinction.

Branching Processes

We will therefore always consider the case where $0 < p_{1,0} < 1$, so that if we let $U_i^{(0)}$ be the probability of eventually hitting state 0 given that we start in state $i \in S$, it is also an absorption probability that represents the extinction probability of a population of size i. Independence of distinct lines of descent implies that since $U_1^{(0)}$ is the probability of any individual's line of descent becoming extinct, that

(3.37)

Clearly we also have that $U_0^{(0)} = 1$, which means

We will now use a first step analysis to consider $U_{X_0}^{(0)}$, where $X_0 = 1$. That is,

Branching Processes

As before, we will look for the minimal non-negative solution to this system of equations. Now if $X_0 = 1$,

but what values does $U_1^{(0)}$ take? We know, for $0 < p_{1,0} < 1$, that the states $\{1, 2, 3, \dots\}$ are transient and so these states must eventually be vacated. In fact we have that

$$\lim_{n\to\infty} \mathrm{P}(X_n > N) = 1 - U_1^{(0)}, \qquad \text{for any finite value of } N.$$

That is, there is a probability of $1 - U_1^{(0)}$ that the population grows without bound. Now, the state of the process at generation n, is just the sum of the number of offspring Z_i from each individual of the $(n-1)^{st}$ generation, so that

(3.39)

Branching Processes

Let's now evaluate the expected number in the n^{th} generation. To do this, we condition on the number of individuals in the $(n-1)^{st}$ generation,

To make progress with this, we also need an expression for the expected number of offspring, μ, that any individual may produce. That is,

(3.40)

Then we may write

3.39

3.40

Branching Processes

From this result we see that for $X_0 = 1$

$$\begin{aligned} E[X_1] &= \mu E[X_0] = \mu \\ E[X_2] &= \mu E[X_1] = \mu^2 \\ &\vdots \\ E[X_n] &= \mu E[X_{n-1}] = \mu^n. \end{aligned}$$

So what does the value of the expected number of offspring for an individual μ imply for the survival or extinction of a population?

1. $(\boldsymbol{\mu < 1})$: means that on average an individual is replaced by less than one individual and we expect the population to decline and so $U_1^{(0)}$ should be 1.
2. $(\boldsymbol{\mu > 1})$: means that on average an individual is replaced by more than one individual and we would expect that $U_1^{(0)}$ should be less than 1.
3. $(\boldsymbol{\mu = 1})$: means that on average an individual is replaced by one individual and the situation is not entirely clear. (although we will show that $U_1^{(0)}$ is still 1 in this case).

Branching Processes

Consider

Then if $(\mu < 1)$, we have that

For the other two cases where $(\mu \geq 1)$, we will initially consider an example of a binary branching process and then extend the ideas to the more general case after we have gained some intuitive facts.

Branching Processes

Example 3.17: Binary branching process

Let X_n be the number of individuals in a population at time n, where the population evolves according to the following transition probabilities

$$\begin{aligned} p_{1,0} &= p \\ p_{1,1} &= r \\ p_{1,2} &= q, \end{aligned}$$

such that $p + r + q = 1$, where $p, q > 0$ and $r \geq 0$.

Then using a first step analysis, we see that

which is quadratic in $U_1^{(0)}$, for which we seek the minimal non-negative solution.

Branching Processes

Example 3.17: (cont)

Let's consider quadratics of the form $y(x) = p + rx + qx^2$ *and find the points for which* $y(x) = x$. *Clearly*

- $y(0) = p > 0$ *and*
- $y(1) = 1$.

Consider the derivative,

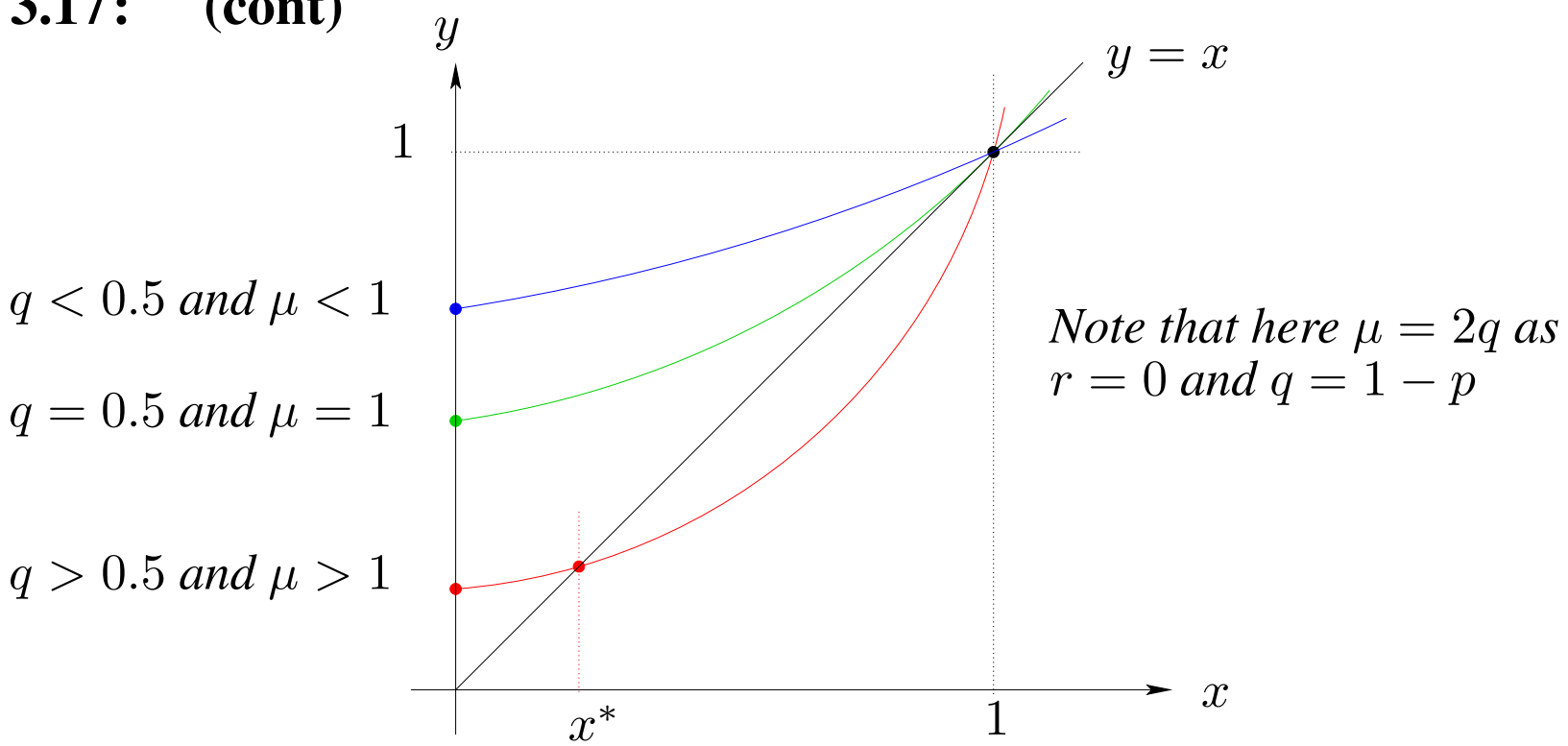
$$\left(= \sum_{j=0}^{\infty} j \; p_{1,j} \; x^{j-1} \quad \textit{in general} \right)$$

which is the mean number of offspring μ *from an individual at* $x = 1$.

Hence we have that the quadratic $y(x) = p + rx + qx^2$ *equals* $p > 0$ *at* $x = 0$, *intersects the line* $y = x$ *at* $x = 1$ *and the derivative at* $x = 1$ *is* μ. *The curve is also convex since it is a quadratic and so there are three possibilities as given in the figure on the next slide.*

Branching Processes

Example 3.17: (cont)



- $(\mu > 1)$: $y(x)$ *crosses the line* $y = x$ *at* $x = 1$ *from below to above and so* $\exists\, x^* \in (0,1)$ *such that* $y(x^*) = x^*$, *which means* $U_1^{(0)} < 1$.
- $(\mu = 1)$: $y(x)$ *is a tangent to the line* $y = x$ *at* $x = 1$ *and so* $U_1^{(0)} = 1$.
- $(\mu < 1)$: $y(x)$ *crosses the line* $y = x$ *at* $x = 1$ *from above to below and so* $U_1^{(0)} = 1$.

Branching Processes

For $p_{1,0} < 1$, re-write equation (3.38) in the form

$$y(x) \;=\; \sum_{j=0}^{\infty} p_{1,j}\; x^j \qquad \text{for all } \; x \in [0,1].$$

As all the $p_{1,j} \geq 0$, the polynomial $y(x)$ is monotonic non-decreasing on $[0,1]$ and the derivative

Hence all of the arguments employed in the binary branching process example hold for all branching processes where $p_{1,0} < 1$ and so we have that

- $(\boldsymbol{\mu > 1})$: means that $U_1^{(0)} < 1$.
- $(\boldsymbol{\mu = 1})$: means that $U_1^{(0)} = 1$.
- $(\boldsymbol{\mu < 1})$: means that $U_1^{(0)} = 1$.

Branching Processes

Example 3.18

Suppose an individual has $0, 1, 2$ *or* 3 *offspring with probabilities* $\frac{1}{3}, \frac{1}{3}, \frac{1}{6}$, *and* $\frac{1}{6}$ *respectively and we wish to find the probability of a line of descent becoming extinct. That is, we want to find the minimum non-negative solution to the system of equations*

$x = 1$ *is clearly a solution giving us that*

The minimal non-negative solution to this is

Note that $\mu = 1.\frac{1}{3} + 2.\frac{1}{6} + 3.\frac{1}{6} = \frac{7}{6} > 1.$

Example 3.19

Modify the probabilities so that an individual has $0, 1, 2$ *or* 3 *offspring with probabilities* $\frac{7}{18}, \frac{1}{3}, \frac{1}{6}$, *and* $\frac{1}{9}$ *respectively and find the probability of a line of descent becoming extinct. Again we want to find the minimum non-negative solution to the system of equations*

$x = 1$ *is clearly a solution giving us that*

The minimal non-negative solution to this is $U_1^{(0)} = 1.$

Note that $\mu = 1.\frac{1}{3} + 2.\frac{1}{6} + 3.\frac{1}{9} = 1.$

Branching Processes

Example 3.20

Modify the probabilities yet again so that an individual has $0, 1, 2$ *or* 3 *offspring with probabilities* $\frac{4}{9}, \frac{1}{3}, \frac{1}{9}$, *and* $\frac{1}{9}$ *respectively and find the probability of a line of descent becoming extinct. Again we want to find the minimum non-negative solution to the system of equations*

$x = 1$ *is clearly a solution giving us that*

The minimal non-negative solution to this is $U_1^{(0)} = 1.$

Note that $\mu = 1.\frac{1}{3} + 2.\frac{1}{9} + 3.\frac{1}{9} = \frac{8}{9} < 1.$

Periodicity

We have seen that recurrence/transience are solidarity properties of a communicating class. Another solidarity property is the *period* of a state. To discuss this, we need the following result:

Theorem 3.20: Kolmogorov

Suppose W is a set of integers with the property that

$$m, n \in W \Rightarrow m + n \in W.$$

Define d to be the greatest common divisor (gcd) of all elements of W.
Then

1. $nd \in W$ *for all sufficiently large* n

2. $d \nmid m \Rightarrow m \notin W$.
(i.e.: if m is not exactly divisible by d, then m is not in W)

Periodicity

Example 3.21:

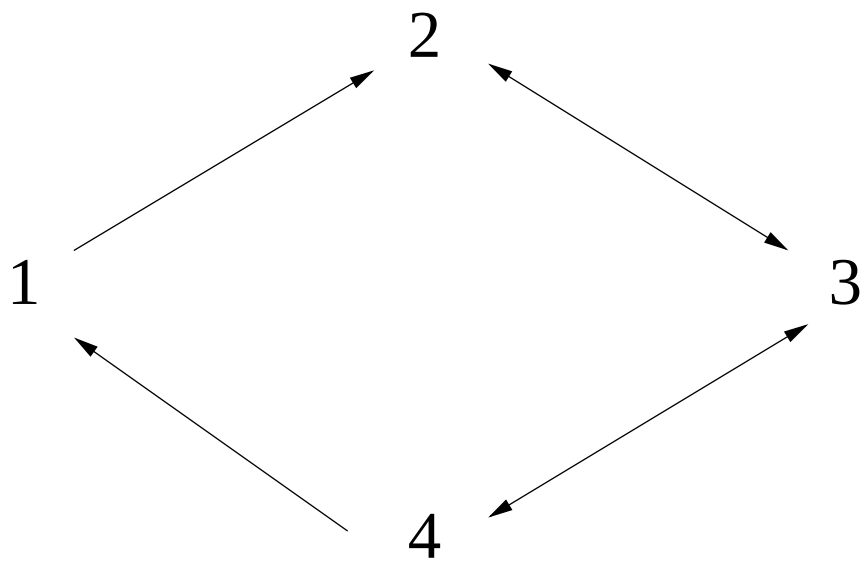
Let $W = \{n : p_{1,1}^{(n)} > 0\}$ *in the Markov chain representation shown below.*

Then $W = \{4, 6, 8, \ldots\}$.

To see this, start at 1 *and walk on the chain and count the number of transitions of paths that bring us back to* 1.

The shortest path involves 4 *transitions - the path is given by the sequence of states* $(1, 2, 3, 4, 1)$.

The next-longest paths involve 6 *transitions, these are (1, 2, 3, 2, 3, 4, 1) and (1, 2, 3, 4, 3, 4, 1), and so forth.*



The greatest common divisor of the elements of W *is* $d = 2$, $nd \in W$ *for* $n \geq 2$ *and if* $2 \nmid m \quad \Rightarrow m \notin W$.
That is, if 2 *does not divide an arbitrary integer* m *exactly, then* m *is not in* W.

Periodicity

Example 3.21: (cont)

Now let $W = \{n : p_{1,1}^{(n)} > 0\}$ *for the slightly modified Markov chain shown below.*

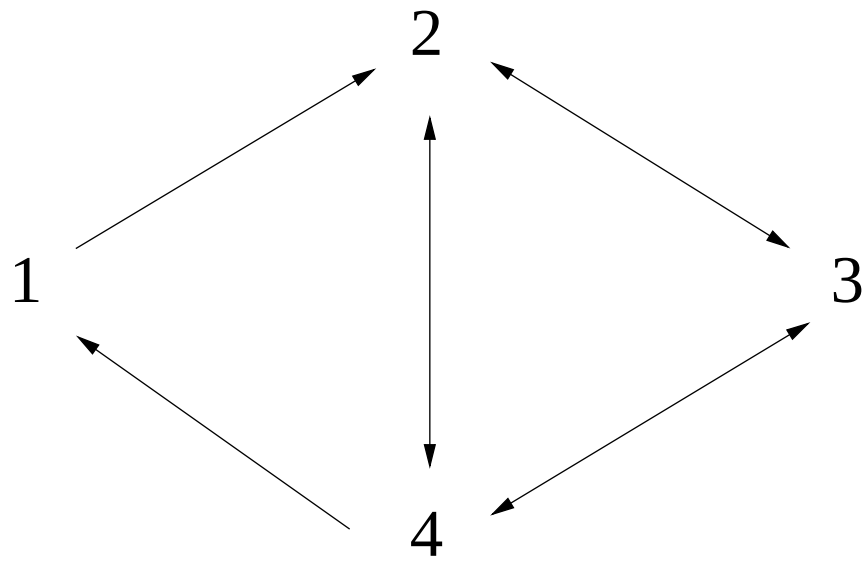
In this case, $W = \{3, 4, 5, 6, 7, \ldots\}$.

To see this, start at 1 *and walk on the chain as before.*

The shortest path that brings us back to 1 *has* 3 *transitions and is* $(1, 2, 4, 1)$.

Similarly, we can construct paths with $n \geq 3$ *transitions.*

The greatest common divisor of the elements of W *is* 1, $nd \in W$ *for* $n \geq 3$ *and* $1 \nmid m \Rightarrow m \notin W$.



That is, if 1 *does not divide an arbitrary integer* m *exactly, then* m *is not in* W, *which is trivial in this case.*

Periodicity

Definition 3.37: Period of a state

The **period** $d(i)$ of a state i in an arbitrary DTMC is the greatest common divisor of the set $W = \{n : p_{i,i}^{(n)} > 0\}$.

From Theorem 3.20, we have as an immediate consequence

1. $p_{i,i}^{(nd(i))} > 0$ for sufficiently large n
2. $d(i) \nmid m \quad \Rightarrow p_{i,i}^{(m)} = 0.$
   That is, pick an arbitrary integer m; if it is not exactly divisible by $d(i)$, then the probability of being in state i after m steps, having started in state i, is zero.

The following theorem states that *period* is a solidarity property.

Theorem 3.21:

If i and j belong to the same communicating class, then $d(i) = d(j)$.

Periodicity

Proof of Theorem 3.21:

Suppose i and j are both members of a communicating class C.

Then there exist $m, n > 0$ such that $p_{i,j}^{(m)} > 0$ and $p_{j,i}^{(n)} > 0$, so that

The middle term on the RHS is greater than zero for all l sufficiently large, say $l \geq l_0$ and the other two terms are strictly positive, by assumption.

Therefore

Periodicity

Proof of Theorem 3.21:

Hence

Similarly, we can show that $d(j)|d(i)$.
If $d(i)$ divides $d(j)$ exactly and vice versa, then the only logical possibility is that $d(i) = d(j)$.

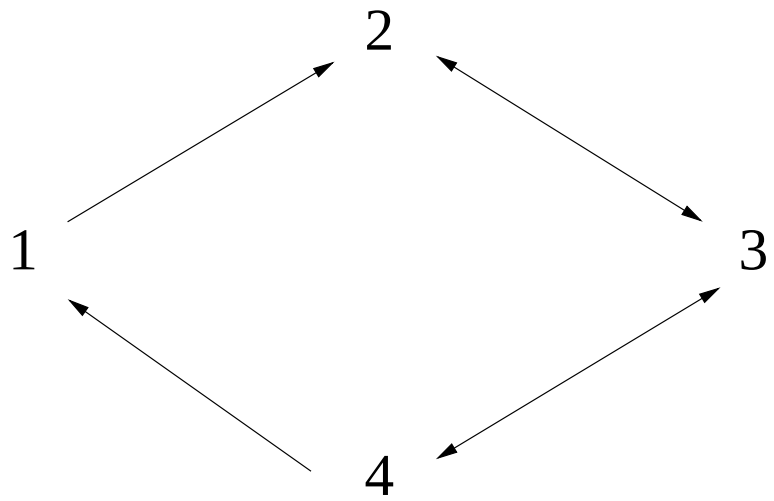
■

Since period is a solidarity property, it makes sense to talk about the period of a communicating class, and hence of an irreducible Markov chain.

Periodicity

Example 3.22:

We return to the Markov chains in Example 3.21.
For the first chain



we saw that $W = W_1 = \{4, 6, 8, \ldots\}$.
Similarly, for the other states, we have $W_2 = W_3 = W_4 = \{2, 4, 6, 8, \ldots\}$.

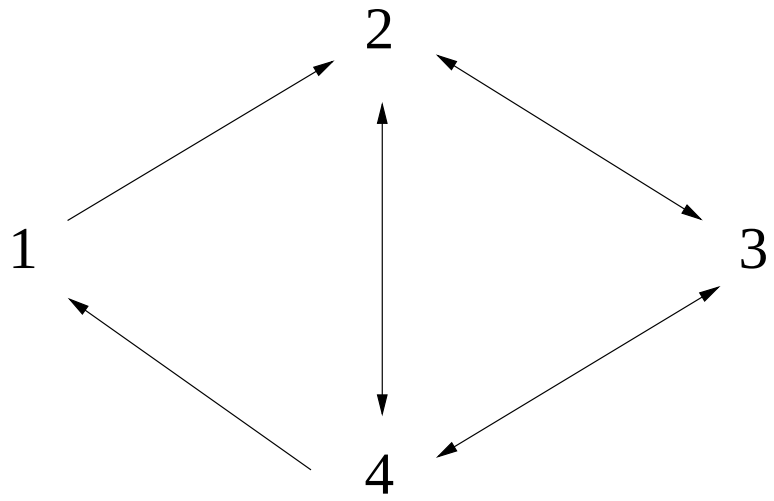
We see that the period of all states is 2*, since for every set* W_1, W_2, W_3 *and* W_4*, we clearly have that* 2 *is the greatest common divisor of the elements of each set.*

Then since this Markov chain is irreducible, we say that it has period 2*.*

Periodicity

Example 3.22: (cont)

For the slightly modified Markov chain



$$\begin{aligned} W_1 &= \{3,4,5,6,7,8,\ldots\} \\ W_2, W_3, W_4 &= \{2,3,4,5,6,7,8,\ldots\}. \end{aligned}$$

Hence, the period of all states is 1*, in accordance with the result of Theorem 3.21.*

Then since this Markov chain is also irreducible, we say that it has period 1.

Periodicity

Definition 3.38: Aperiodic state

State i is called **aperiodic** if it has period $d(i) = 1$.
The communicating class to which i belongs is also called aperiodic.

Recall that in Example 3.10, we found that $p_{1,1}^{(n)}$ settled down to a constant value by "time" $n \approx 15$.

This suggests that in certain types of Markov chains, there exists some kind of limiting behaviour, meaning that $\lim_{n\to\infty} p_{i,i}^{(n)}$ exists for all states i.

In other types of Markov chains, this limit does not exist.
For example (see again later)

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{has period 2 and} \quad p_{1,1}^{(n)} = \begin{cases} 1 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$$

It turns out that the types of Markov chains for which the limit $\lim_{n\to\infty} p_{i,i}^{(n)}$ exists are *aperiodic* ($d = 1$) Markov chains, and the types of Markov chains for which this limit does not exist are *periodic* ($d > 1$) Markov chains.

Limiting behaviour

In general, we are often interested in the limiting behaviour of $p_{i,j}^{(n)}$ as $n \to \infty$.
If state i is transient, we have already seen in (Theorem 3.14) that

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} < \infty.$$

Therefore, $p_{i,i}^{(n)} \to 0$ as $n \to \infty$, otherwise the sum could not be finite.
On the other hand, if state i is recurrent, then

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty,$$

but this is still compatible with either

Limiting behaviour

The following theorem addresses the previous possibilities.

Theorem 3.22:

For an irreducible recurrent DTMC of period d,

$$p_{i,i}^{(n)} = 0 \qquad \text{if } d \nmid n$$

and

$$\lim_{n\to\infty} p_{i,i}^{(nd)} = \lambda_i = \frac{d}{\sum_{n=1}^{\infty} n f_{i,i}^{(n)}}.$$

Proof omitted:

The first expression is a trivial consequence of periodicity.
A proof of the latter expression is quite involved, and can be found for example in Karlin and Taylor. The proof is simpler in the aperiodic case. ■

Limiting behaviour

Example 3.23:

This example illustrates why the period d plays such a prominent role in the above theorem.
Consider the irreducible Markov chain with transition probability matrix

This Markov chain has no limiting distribution, since

(3.41)

That is,
This is due to the fact that $d = 2$, which can be verified by noting that

Limiting behaviour

Example 3.23: (cont)

Similarly, due to the deterministic nature of this Markov chain, it is easy to see without performing any calculation that

and therefore it is also true when we take the limit

Limiting behaviour

Example 3.23: (cont)

What about the limiting quantities $\lambda_i = \dfrac{d}{\sum_{n=1}^{\infty} n f_{i,i}^{(n)}}$*?*

For this chain, we know that $d = 2$*. So what is* $\sum_{n=1}^{\infty} n f_{i,i}^{(n)}$*?*

where μ_i *is the mean number of steps required for first return to* i*.*

For this chain, it is clear that

Hence, $\lambda_i = 1$ *for* $i = 1, 2$ *and so Theorem 3.22 is confirmed.*

Limiting behaviour

In the above example, we were able to determine the limiting quantities λ_1, λ_2 by inspection. In general, this is not possible.

Most of this section will deal with

- Interpreting the physical meaning of the quantities λ_i
- Finding a general method for calculating the λ_i

Interpretation of Theorem 3.22:

Imagine that the process starts in state i, and a very long time later is observed. Intuitively, suppose we can define

$$\pi_i = \mathrm{P}(\text{we find the process in state } i)\,.$$

Limiting behaviour

We don't know whether or not a multiple of d steps has occurred and so we need to determine the *probability* that the number of transitions since time 0 is a multiple of d.

To address this, we apply the principle of "maximum entropy", where we make the assumption which maximises the uncertainty – thus minimising the assumed prior knowledge – of the number of transitions since time 0, which we shall denote by n.

That is, in the absence of any specific knowledge about the time n that has elapsed, we assume all values of

$$n \pmod{d} \in [0, d-1]$$

to be equally likely.

Limiting behaviour

Hence, we can state that the

$$\text{P(number of transitions since time } 0 \text{ is not a multiple of } d) \quad =$$

$$\text{P(number of transitions since time } 0 \text{ is a multiple of } d) \quad = \qquad .$$

Using Theorem 3.22, we have

so that

where μ_i = mean number of steps required for first return to i.

Positive and null recurrence

So $\pi_i = \frac{1}{\mu_i}$ can also be interpreted as the mean proportion of time that the process spends in i given that it starts in i. Note, it is also the long-term probability of finding the process in state i given that it starts in i.

Also, for an irreducible, aperiodic ($d = 1$) Markov chain, the informal argument given above suggests that $\lambda_i = \pi_i$, so λ_i is the long term probability of being in state i given that the process started in state i.
There are two cases:

- $\lambda_i > 0 \quad \Rightarrow \quad \pi_i > 0.$
  -
  -
  -

- $\lambda_i = 0 \quad \Rightarrow \quad \pi_i = 0.$
  -
  -
  -

Positive and null recurrence

So given a recurrent state i, we have a couple of ways of determining whether the state is positive recurrent or null recurrent.
Specifically, we determine this by calculating μ_i, or $\lim_{n\to\infty} p_{i,i}^{(nd)}$.

Theorem 3.23:

Null recurrence and positive recurrence are solidarity properties.

Proof: tutorial or CE.

Because of solidarity, in order to determine whether a communicating class is transient, positive recurrent or null recurrent, it is possible to determine the nature of just one state in the communicating class.

An (irreducible) DTMC can therefore be called transient, positive recurrent or null recurrent, in which case all states of the DTMC have the same property.
Computationally, sometimes this provides an easy way of determining the nature of a communicating class, if some states are easier to analyse than others.
That is, it may be easier to derive an expression for $p_{i,i}^{(n)}$ for some state(s) i, and more difficult for others.

Solidarity summary

Solidarity properties so far:

- period
- transience
- positive recurrence
- null recurrence

We have investigated some properties of $\lim_{n\to\infty} p_{i,i}^{(n)}$ for a state i.

Now we shall investigate some properties of $\lim_{n\to\infty} p_{i,j}^{(n)}$ for states $i \neq j$.

We saw earlier that if j is transient, then $\sum_{n=1}^{\infty} p_{i,j}^{(n)} < \infty$ for any state i.

In order for this series to be finite, it must be true that $\lim_{n\to\infty} p_{i,j}^{(n)} = 0$.

Now let's look at this limit where j is recurrent, by first defining

$$f_{i,j}^{(n)} = \mathrm{P}(\, X_n = j, X_l \neq j, l = 1, \ldots, n-1 \,|\, X_0 = i) \ . \tag{3.42}$$

Long term behaviour

In words, $f_{i,j}^{(n)}$ is the probability that the process first reaches state j in exactly n transitions, having started in state i
(this is a generalisation of $f_{i,i}^{(n)}$ introduced earlier).

Let's define

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)} \ , \tag{3.43}$$

which is the probability of eventually reaching state j starting from i.

It is also equal to $u_i^{\{j\}}$, the probability of ever hitting (visiting) state j, given that the process starts in i, so $f_{i,j}$ is a hitting probability.

The next few results are derived only for the aperiodic case ($d = 1$).

Extensions of these results to the periodic case are more difficult (see for example Karlin and Taylor), and will not be presented in this course.

Long term behaviour

Theorem 3.24:

Let j be an aperiodic recurrent state in a DTMC and let i be any other state. Then

$$p_{i,j}^{(n)} \to f_{i,j}\lambda_j \qquad as\ n \to \infty$$

Proof of Theorem 3.24:

A simple variant of expression (3.32) gives

for any i, j.
We know that for any $\epsilon > 0$, there exists N such that

Long term behaviour

Proof of Theorem 3.24 (cont):

We know this because $\sum_{k=1}^{\infty} f_{i,j}^{(k)} \leq 1$, by definition, which implies that $\lim_{k\to\infty} f_{i,j}^{(k)} = 0$, and therefore such an N must exist.

Then for $n > N$, we have

Long term behaviour

Aside:

It is possible that $\lim_{n\to\infty} p_{i,j}^{(n)}$ does not exist, eg: $p_{i,j}^{(n)}$ could oscillate.

However, $\inf_{k\geq n} p_{i,j}^{(k)}$ is an increasing sequence (an increasing function of n), and the limit of this sequence

$$\lim_{n\to\infty} \left\{ \inf_{k\geq n} p_{i,j}^{(k)} \right\}$$

does exist, which is known as the $\lim\inf$ (limit inferior) of the sequence $p_{i,j}^{n}$.

Similarly, $\sup_{k\geq n} p_{i,j}^{(k)}$ is a decreasing sequence (a decreasing function of n), and

$$\lim_{n\to\infty} \left\{ \sup_{k\geq n} p_{i,j}^{(k)} \right\}$$

exists, and is known as the $\lim\sup$ (limit superior) of the sequence $p_{i,j}^{n}$.

Long term behaviour

Proof of Theorem 3.24 (cont):

Now take the limit $n \to \infty$ in the above expression.
Theorem 3.22 tells us that

so

Long term behaviour

Proof of Theorem 3.24 (cont):

So

for all $\epsilon > 0$, so

Hence, $\lim_{n\to\infty} p_{i,j}^{(n)}$ exists and let's call it $\lambda_{i,j}$.

Then

Long term behaviour

Now let $N \to \infty$,

where $f_{i,j} = \sum_{k=1}^{\infty} f_{i,j}^{(k)}$.

Therefore, since ϵ can be made arbitrarily small, we have that

■

A more complicated analysis proves the similar result when j is periodic.

Long term behaviour

Corollary 3.1:

For an irreducible, aperiodic, recurrent DTMC,

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \lambda_j.$$

Proof of Corollary 3.1 :

■

Application:

Consider a K state, aperiodic, irreducible DTMC, which we have seen must be recurrent. Further, we know that

$$\sum_{j=1}^{K} p_{i,j}^{(n)} = 1, \qquad i \in \mathcal{S}, \forall n \geq 1.$$

Long term behaviour

Let $n \to \infty$, to show that

$$\sum_{j=1}^{K} \lambda_j = 1,$$

which implies that there exists at least one j such that $\lambda_j > 0$, which in turn implies that j is positive recurrent.
Since the chain is irreducible, then all states are positive recurrent and we write $\lambda_j = \pi_j$, the long term probability of being in state j.

Example 3.24: **Recall Example 3.10**

$$\mathbb{P} = \begin{pmatrix} 0 & \frac{1}{4} & \frac{3}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Long term behaviour

Example 3.24: (cont)

Plotting $p_{1,1}^{(n)}, p_{2,1}^{(n)}, p_{3,1}^{(n)}$ *for some values of* n*, we see that there is a limiting behaviour for the probability of finding the process in state* 1 *that is independent of the starting state* $1, 2$ *or* 3*.*

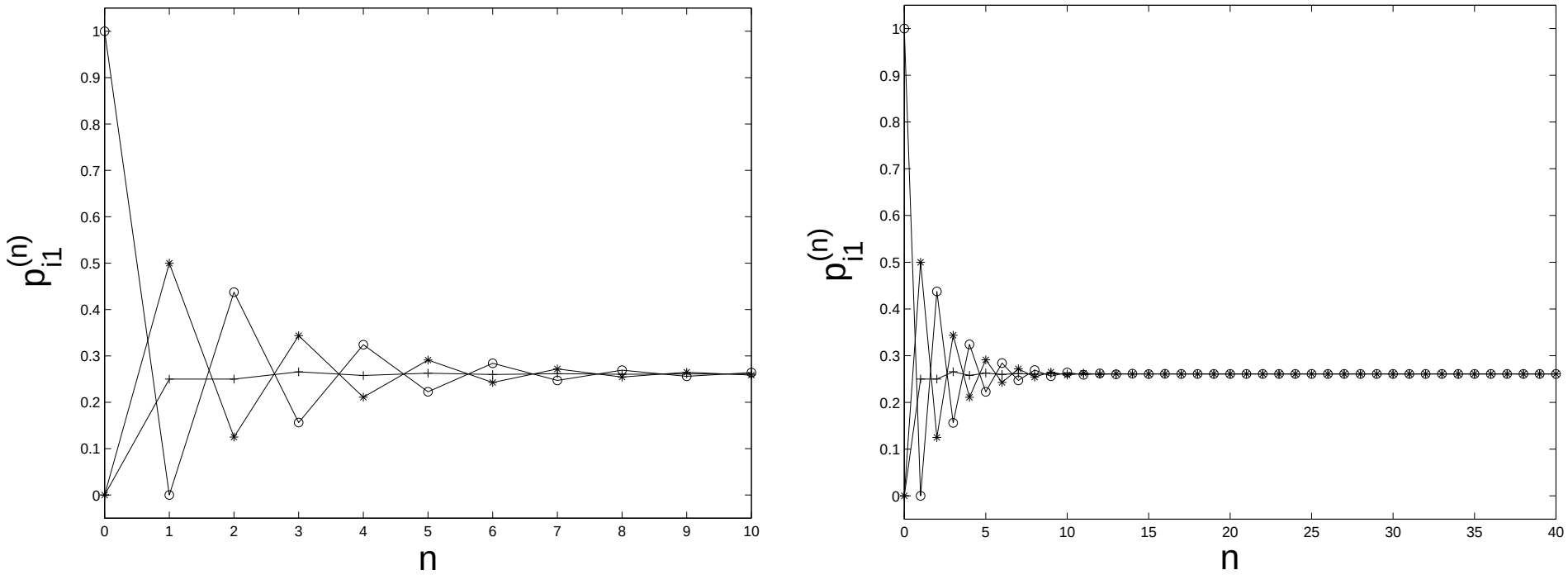


Figure 3.7: Plot of $p_{1,1}^{(n)}, p_{2,1}^{(n)}, p_{3,1}^{(n)}$ *versus* n*.*

Long term behaviour

In the previous example, we estimated λ_1 by computational means, that is, by calculating some values of $p_{i,1}^{(n)}$.
However, we still do not know how to calculate the λ_j analytically.

We now derive a set of equations which can be solved analytically to find λ_j.

Theorem 3.25:

For an aperiodic, irreducible, recurrent DTMC, the λ_j satisfy

$$\lambda_j = \sum_{i \in \mathcal{S}} \lambda_i p_{i,j}, \quad j \in \mathcal{S}.$$

Proof of Theorem 3.25:

Condition on the state of the process i after n transitions,

Long term behaviour

Proof of Theorem 3.25 (cont):

To prove the result, we would like to simply take the limit $n \to \infty$, thus giving

In order to do this, we require that the interchange of limit and summation be justified. How might we justify it ?

Aside: We have a set of numbers $f_i^{(n)}$ with the property that

$$\lim_{n\to\infty} f_i^{(n)} = f_i$$

for all i and we want to know something about

$$\lim_{n\to\infty} \sum_i f_i^{(n)}.$$

We would essentially like it to be $\sum_i f_i$, but when is this the case?

Long term behaviour

1. If the sum is over a finite number of terms
   $\big($Maths I: $\lim(a+b) = \lim a + \lim b\big)$
2. If $f_i^{(n)} \to f_i$ monotonically
   $\big($Monotone Convergence Theorem$\big)$
3. If $|f_i^{(n)}| < g_i$ for all i, where $\sum_i g_i < \infty$
   $\big($Lebesgue Dominated Convergence Theorem$\big)$

In our case,

- the sum might be infinite (infinite number of states),
- the limit is not monotone
  (eg: see figures for $p_{i,1}^{(n)}$ vs n, does not tend to λ_1 monotonically), and
- we can't find a family of summable dominating functions g_i.

However, we can use Fatou's Lemma, which states that if the $f_i^{(n)}$ are positive,

$$\liminf_{n\to\infty} \sum_i f_i^{(n)} \geq \sum_i f_i \ .$$

Long term behaviour

In what follows, we will always know that

$$\lim_{n\to\infty} \sum_{i\in\mathcal{S}} f_i^{(n)}$$

exists (see Corollary 3.1 where we see that $\lim_{n\to\infty} \sum_j p_{i,j}^{(n)}$ exists), and so we can use Fatou's Lemma with $\liminf$ replaced by $\lim$.

Proof of Theorem 3.25 (cont):

Long term behaviour

Proof of Theorem 3.25 (cont):

Recall,

Let $n \to \infty$, and use Fatou's Lemma to give

Now multiply by $p_{j,k}$ and sum over j, to get

Long term behaviour

Proof of Theorem 3.25 (cont):

The LHS $\leq \lambda_k$ by the previous relation and hence the RHS is an absolutely convergent double series and we can change the order of summation (Maths I result). Thus,

By induction, we have

Suppose this inequality is strict for some $n > 1$ and $k \in \mathcal{S}$, so that summing over k we have

$$\sum_{k \in \mathcal{S}} \lambda_k \quad > \quad \sum_{k \in \mathcal{S}} \sum_{i \in \mathcal{S}} \lambda_i \; p_{i,k}^{(n)} .$$

Long term behaviour

Proof of Theorem 3.25 (cont):

The LHS is ≤ 1, as we saw earlier, and so the series on the RHS is absolutely convergent, and we can change the order of summation, to yield

which is a contradiction.
Therefore, the inequality (3.44) cannot be strict, so that

(3.45)

and it is also true for $n = 1$:

$$\lambda_k = \sum_{i \in \mathcal{S}} \lambda_i \; p_{i,k}.$$

■

Long term behaviour

Note: We did not need (3.45) - the same would work for $n = 1$ directly. However, we will use it in the next proof.

Theorem 3.26:

For an irreducible, aperiodic, positive recurrent DTMC $\sum_{k \in \mathcal{S}} \lambda_k = 1.$

Proof of Theorem 3.26 :

In the proof of Theorem 3.25, we found that

(a)

(b)

We can get the result if we take the limit $n \to \infty$ in the above expression, and justify the swap of limit and summation as follows.

Long term behaviour

Proof of Theorem 3.26 :

Therefore, rearranging we see that

Now $\lambda_k = 0$ is impossible for a positive recurrent state k (and all states are positive recurrent), so it must be true that $\sum_{i \in \mathcal{S}} \lambda_i = 1$. ■

Global balance equations

For the aperiodic, positive recurrent case, this result ties down the interpretation of the λ_i as *steady state probabilities*, also known as *stationary probabilities, or equilibrium probabilities.* The usual notation for these is $\lambda_i = \pi_i$.
We now have a systematic way of calculating these probabilities, given by the solution to the system of linear equations known as the **global balance equations**.

$$\pi_k = \sum_{i \in \mathcal{S}} \pi_i \, p_{i,k}, \; k \in \mathcal{S}, \qquad \text{subject to } \sum_i \pi_i = 1 \tag{3.46}$$

Note that equilibrium in a Markov chain does not imply that things stop moving (compare with equilibrium in a mechanical system, eg: pendulum has points of static equilibrium).
By equilibrium, we mean that certain statistical quantities settle down to constant values.
For example, the average amount of time that the process spends in each state i of a Markov chain settles to an equilibrium value, even though the process keeps moving around on the state space.

Long term behaviour

Example 3.25: Finite number of states

The Markov chain whose transition probability matrix is

$$\mathbb{P} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

is aperiodic when $0 < a, b < 1$ *(why?).*
The global balance equations are

One of the first two equations is redundant (this is always true when the number of states is finite - see Class Exercise).
We can arbitrarily strike out the first equation, leaving us with two equations in two unknowns.

The solution is

Long term behaviour

Note: for a general finite state DTMC with N states, the global balance equations yield $N + 1$ equations in N unknowns.
Removing one of the redundant equations (that is, any equation other than the normalisation constraint !!!), leaves us with N equations in N unknowns, which we can solve using the usual techniques from linear algebra.

In the last few results and the last example, we have restricted our attention to the case where the Markov chain is aperiodic ($d = 1$).
However, with a little more work, the same results can be derived for the periodic case.

Most importantly, for an irreducible, positive recurrent DTMC with period $d > 1$, the system (3.46) still has a solution.
However, the interpretation should be that

$$\begin{aligned} \lim_{n\to\infty} p_{j,j}^{(nd)} &= \lambda_j = d\pi_j \\ p_{j,j}^{(m)} &= 0 \qquad \text{if } d \nmid m \ . \end{aligned}$$

Long term behaviour

Example 3.26: Limiting versus stationary distribution

Recall Example 3.23, where

$$\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

We saw that this Markov chain has no limiting *distribution because* $d = 2$. *However, it does have a* stationary *distribution, which is given by the solution to the system of equations*

Recall that for this example, we have $\lambda_1 = \lambda_2 = 1$.

Long term behaviour

Example 3.26: Limiting versus stationary distribution (cont)

Our solution to the global balance equations coincides with the expression

The above example demonstrates that an irreducible, positive recurrent DTMC with period $d > 1$ does not have a limiting distribution (because $\lim_{n\to\infty} \mathbb{P}^{(n)}$ DNE).

However, the DTMC *does* have a stationary distribution π.

This makes sense, because even if a given $p_{i,i}^{(n)}$ oscillates indefinitely, we can still ask the following question:

"after a long period of time,
what is the probability π_i of finding the process in state i?"

Similarly, we can ask, what is the mean proportion of time that the process spends in state i, and we have seen that this is given by $\mu_i = \frac{1}{\pi_i}$.

Long term behaviour

Theorem 3.27:

For an irreducible DTMC, if the equations

$$x_k \quad = \quad \sum_{i \in \mathcal{S}} x_i \, p_{i,k} \quad \textit{such that} \quad \sum_{i \in \mathcal{S}} x_i = 1 \textit{ and } x_i > 0 \; \forall i \in \mathcal{S},$$

have a solution, then the DTMC is positive recurrent, and furthermore, the stationary distribution is given by $\pi_i = x_i, \; i \in \mathcal{S}$.

Proof of Theorem 3.27:

We prove the result for the aperiodic case only, since the results that we draw upon from earlier in this section all pertain to the aperiodic case.
We showed earlier that

Long term behaviour

Proof of Theorem 3.27 (cont):

Now suppose the DTMC is either transient or null recurrent and we have already seen that $\lim_{n\to\infty} p_{i,k}^{(n)} = 0$ for all $k \in \mathcal{S}$.
Therefore

But the LHS > 0 since $x_k > 0$ for at least some k (normalisation), so we have a contradiction.
Therefore, the DTMC is positive recurrent.

Now since all states are positive recurrent, we know that

Long term behaviour

Proof of Theorem 3.27 (cont):

Therefore, $\forall k \in \mathcal{S}$,

■

This theorem is very important, because it says that to test if an irreducible DTMC is positive recurrent, all we have to do is see whether there exists a positive solution to the global balance equations (3.46).
At the same time, we get the stationary distribution.

Long term behaviour

Example 3.27: Infinite number of states

Consider the DTMC with state space $\mathbb{Z}^+$ and transition probabilities

$$\begin{aligned} p_{i,i+1} &= p \\ p_{i,i-1} &= q \\ p_{0,1} &= 1 \\ p_{i,j} &= 0 \qquad \textit{otherwise}. \end{aligned}$$

Does this look familiar?
It is a slightly modified version of the "Two Gamblers" problem with an infinitely rich Player B.
The slight modification is that state 0 is no longer an absorbing state, so that instead of $p_{0,0} = 1$, we have $p_{0,0} = 0$ and $p_{0,1} = 1$.
Therefore, all states form a single communicating class.

A physical interpretation for this is that whenever Player A loses all of her money (hits state 0), then she is given one dollar to keep playing by a rich benefactor.

Long term behaviour

Example 3.27: Infinite number of states (cont)

The cases where the communicating class $\mathbb{Z}^+$ is positive recurrent, null recurrent or transient, each have a different implication for the benefactor (what are they ?).
A solution to the global balance equations for this problem exists only for the case $p < q$,

$$\pi_i = \frac{1}{2p}\left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^i,$$

and there is no solution otherwise

Hint for CE or tutorial: the global balance equations yield a system of difference equations, with special equations for $i = 0, 1$.
Try a solution of the form $\pi_i = Aw^i$.

Thus, the DTMC is positive recurrent if and only if $p < q$.
Which property does the DTMC possess when $p = q$?
What about $p > q$?

Long term behaviour

Solving the global balance equations:

There is no all-purpose method for solving the global balance equations.
However, some guidelines are as follows:

- **Finite number of states:**
  We have a system of $N + 1$ linear equations in N unknowns, discard one of the redundant ones (not the normalisation constraint !!), use techniques from linear algebra to solve.
  We have seen a couple of examples of this already.

- **Countably infinite number of states:**

  - If the transition probabilities $p_{i,j}$ **do not** depend on i for $i \geq I$, for some finite integer I, then we can often employ difference equation methods. Another method is to employ *generating functions.*

  - If the transition probabilities $p_{i,j}$ **do** depend on i (the most general case), then it is often not possible to solve the equations analytically.
    There are, however, some exceptions, which we characterise in the next section.

Partial balance equations

Consider an irreducible, positive recurrent Markov chain.
Partition the state space $\mathcal{S}$ into two sets $\mathcal{B}$ and $\mathcal{B}^c$.
Then the global balance equations for state j read Flux out $=$ Flux in)

Partial balance equations

The equations

$$\sum_{j\in\mathcal{B}} \pi_j \sum_{i\in\mathcal{B}^c} p_{j,i} \;=\; \sum_{j\in\mathcal{B}} \sum_{i\in\mathcal{B}^c} \pi_i p_{i,j} \tag{3.47}$$

are known as the **partial balance equations** and have the following physical interpretation

$$\text{“total flux from } \mathcal{B} \text{ to } \mathcal{B}^c\text{”} \;=\; \text{“total flux from } \mathcal{B}^c \text{ to } \mathcal{B}.\text{”}$$

Global balance and partial balance equations contain the same “information” about the Markov chain.
That is, all of the above operations, are derived by manipulating the global balance equations, but no special assumptions have been made.
In fact, setting $\mathcal{B} = \{j\}$ in the above partial balance equations, gives you the global balance equations $\pi_j \left(1 - p_{j,j}\right) = \sum_{i\in\mathcal{B}^c} \pi_i p_{i,j}$.

Sometimes, the Markov chain is such that the partial balance equations are easier to solve than the global balance equations.

Partial balance equations

Example 3.28: Skipfree irreducible chain on $\mathbb{Z}$

The process can only move up or down by one unit per transition.

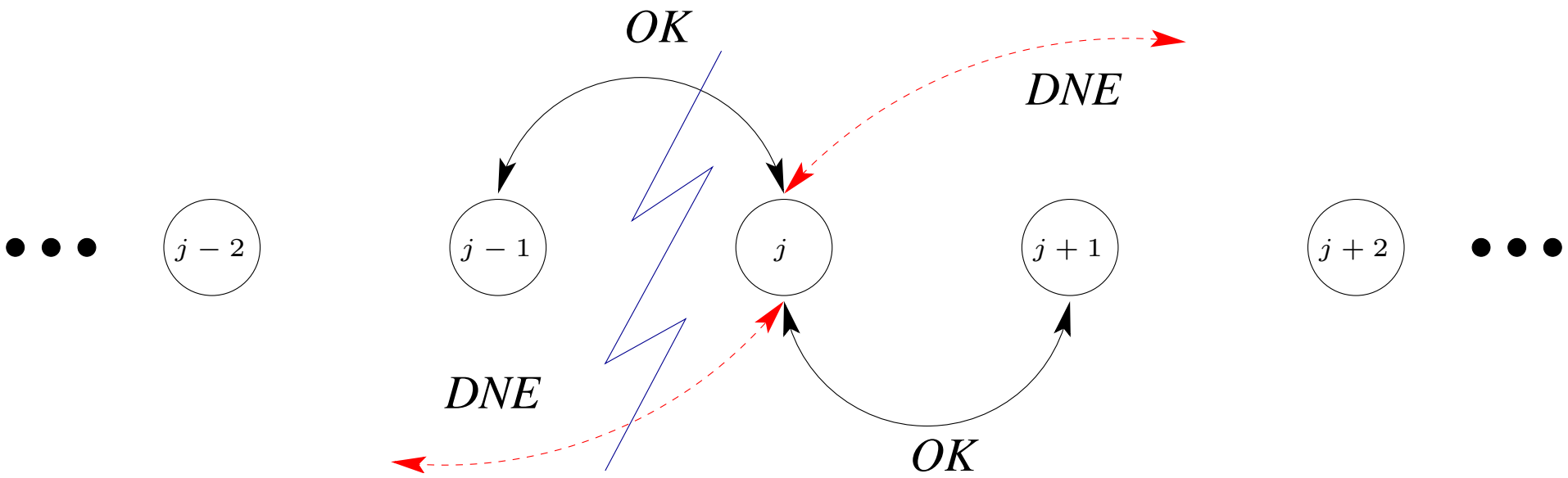


Figure 3.8: State diagram of a skip free Markov chain

Partition into sets $\mathcal{B}$ and $\mathcal{B}^c$, by making a "cut" between the states $j-1$ and j. Since the chain is skipfree, the only way that the process can move between $\mathcal{B}$ to $\mathcal{B}^c$ is by going from $j-1$ to j, and vice versa.

Partial balance equations

Example 3.28: Skipfree irreducible chain on $\mathbb{Z}$ (cont)

The partial balance equations for this given choice of partition $\mathcal{B}$ and $\mathcal{B}^c$ reduce to

since j can be chosen arbitrarily for this skipfree chain.
These equations are also known as the "cut equations" in some texts.

In particular, because they relate to just two states in the state space $\mathcal{S}$, they are often known as the "local balance" or "detailed balance" equations, which occur for particular types of DTMC (see later).

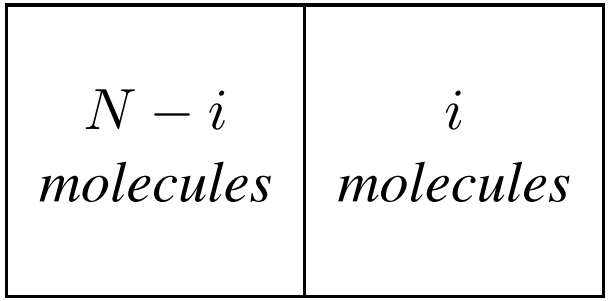
Compare with the more complicated global balance equations:

Partial balance equations

Example 3.29: Ehrenfest Diffusion Model

A famous instance of a skipfree Markov chain in physics is the Ehrenfest Diffusion Model for diffusion of gas through a membrane separating two chambers.

In this discrete Markov chain model we essentially represent each chamber as an urn, with the gas molecules represented as balls (numbered $1, 2, ..., N$*) distributed between the two urns.*



Let X_n *be the number of molecules in the right chamber after the* n^{th} *passage through the membrane, so*

$$0 \leq X_n \leq N.$$

The physical assumption is that for some constant C*,*

$$\begin{aligned} P(\textit{transition right to left}) &= Ci \\ P(\textit{transition left to right}) &= C(N-i)\ . \end{aligned}$$

That is, the rate of transitions is proportional to the "concentration gradient".

Partial balance equations

Example 3.29: Ehrenfest Diffusion Model (cont)

Normalisation gives

$$
\begin{aligned}
Ci + C(N-i) &= 1 \\
\textit{so that} \quad C &= \frac{1}{N} \quad \textit{and hence} \\
p_{i,i+1} &= \frac{N-i}{N} \\
p_{i,i-1} &= \frac{i}{N} \\
p_{i,k} &= 0. \quad \textit{otherwise.}
\end{aligned}
$$

We could write down and solve the global balance equations, but the partial balance equations (shown on the next slide) provide a slightly easier way.

Partial balance equations

Example 3.29: Ehrenfest Diffusion Model (cont)

Partial balance equations

Example 3.29: Ehrenfest Diffusion Model (cont)

To find π_0, we use the normalisation condition:

The solution shows that π_j has the binomial distribution

$$\pi_j = \frac{1}{2^N}\binom{N}{j} = \binom{N}{j}\left(\frac{1}{2}\right)^j\left(\frac{1}{2}\right)^{N-j}.$$

Partial balance equations

Example 3.29: Ehrenfest Diffusion Model (cont)

An interesting thing is that if N is reasonably large, even of the order of 30 or so, then $\binom{N}{\frac{N}{2}}$ is much larger than $\binom{N}{0}, \binom{N}{1}, \binom{N}{N-1}, \binom{N}{N}$. This means that we will almost never see the system in a state with all molecules in one chamber. For example consider the Figure 3.9 for $N = 30$, where it is clear that the probability at either extreme is essentially 0.

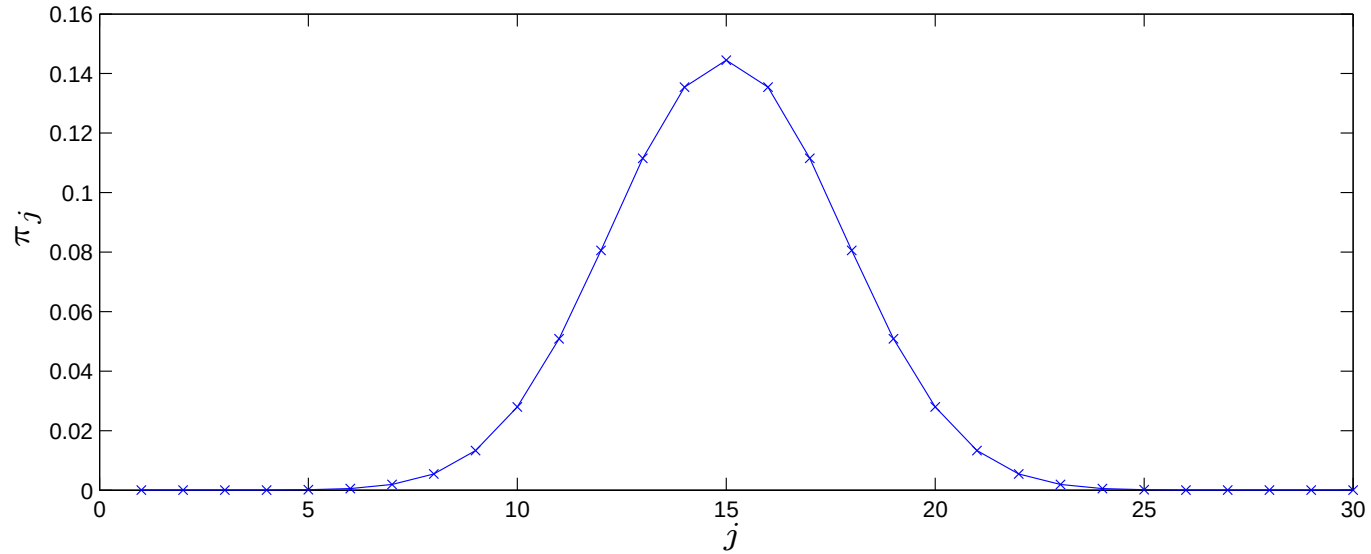


Figure 3.9: Probability of j molecules in the right chamber with $N = 30$

Partial balance equations

That this happens is purely a combinatorial phenomenon, and provides an insight into the combinatorial aspect of statistical mechanics (thermodynamics).

The above technique of "cutting" works whenever the state diagram forms a tree, that is, a connected graph without any closed loops.

On the other hand, if the Markov chain is not skipfree, then of course the partial balance equations still hold, but they are no simpler than the global balance equations, and this technique is not useful!

We now discuss the concept of running a Markov chain in reverse. From this idea we will then consider a special form where we consider when a DTMC is reversible in the sense that it is appears to be the same in forward time as in reverse time. Reversibility is a useful characteristic of DTMCs, where detailed balance holds for all i, j.

Time Reversed Markov chains

Consider an irreducible positive recurrent DTMC $\{X_n : n \in \mathbb{N}\}$, and its associated reverse time DTMC $\{X_n^* : n \in \mathbb{N}\}$, where the DTMC is considered in reverse time.

That is, "the film is run backwards" so to speak $\Rightarrow X_n^* = X_{\tau - n}$.

Time reversal is essentially mirroring with respect to a time τ, which in itself is unimportant as it only defines where the origin is in reversed time.

Why do we study the reversed time DTMC?

1. With the aid of the reversed time DTMC, we can often get more insight into the properties of the DTMC.

2. By considering the reversed time DTMC, we can often very simply and elegantly derive results, where direct derivation might be quite complicated.

3. For example, the balance equations of a complex DTMC can be derived by "guessing" the reversed time DTMC.

Time Reversed Markov chains

Is the Time-reversed chain Markov? Yes!

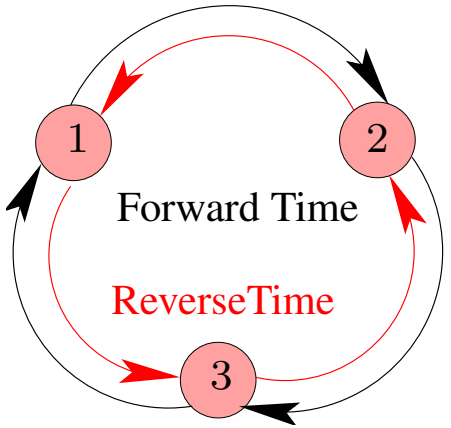
Time Reversed Markov chains

In general, a DTMC run in reverse time will be very different from that which is run in forward time.

Example 3.7: A cyclic process

Consider the following DTMC of period 3, where

$$\mathbb{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$



The processes here are very different as in

- forward time the process goes from $1 \to 2 \to 3 \to 1 \ldots$
- reverse time the process goes from $1 \to 3 \to 2 \to 1 \ldots$

If we assume that the process has equilibrium distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$ in forward time and $\boldsymbol{\pi}^* = (\pi_1^*, \pi_2^*, \pi_3^*)$ in reverse time, we note that the proportion of time spent in state $i = 1, 2, 3$ will be the same in forward or reverse time, so that $\boldsymbol{\pi} = \boldsymbol{\pi}^* = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. ◁

Time Reversed Markov chains

Consider now the reversed time probabilities $p^*_{i,j}$ of a more general stationary, irreducible, positive recurrent DTMC.

Reversed Markov chains

We know that an irreducible DTMC $\{X_n : n \in \mathbb{N}\}$, may be considered in reverse time $\{X_n^* : n \in \mathbb{N}\}$ and we also have expressions for the transition probabilities $p_{i,j}^*$ in reverse time.

This brings us to consider a special case, where the forward time DTMC and it's associated reverse time DTMC are such that they are statistically indistinguishable. This means that an outside observer cannot tell whether "the film is run in the forward or backward direction".

If this is the case, then we say that the DTMC $\{X_n : n \in \mathbb{N}\}$ is time reversible.

Reversibility

Definition 3.39: Reversibility

An irreducible discrete time Markov chain $\{X_n : n \in \mathbb{N}\}$ is reversible if

$$\mathrm{P}(X_n = i, X_{n+1} = j) \;=\; \mathrm{P}(X_n = j, X_{n+1} = i) \quad \forall i, j \in \mathcal{S} \text{ and } \forall n$$

$$\mathrm{P}(X_n = j) \;>\; 0 \qquad \forall j \in \mathcal{S} \text{ and } n\ .$$

This tells us that a reversible DTMC running in forward time appears identical to when it is running in reverse time.

Theorem 3.28:

An irreducible Markov chain X_n is reversible if and only if there exists a collection of positive numbers x_j such that

$$\begin{aligned} x_i\, p_{i,j} &= x_j\, p_{j,i} \quad \forall i, j \in \mathcal{S} \text{ and} \\ \sum_j x_j &= 1. \end{aligned}$$

Reversibility

Proof of Theorem 3.28 $\Longrightarrow$

If the DTMC is reversible, by the first condition of Definition 3.39

Now, for some fixed n let $x_j = \mathrm{P}(X_n = j)\,,\ \forall j \in \mathcal{S}$, which is a set of positive numbers by the second condition of Definition 3.39. Then

Reversibility

Proof of Theorem 3.28 $\Longrightarrow$ (cont.)

The last step is to show, under the first condition of Definition 3.39, that $\mathrm{P}(X_n = j)\,,\ j \in \mathcal{S}$ are independent of the value of n.
We therefore have

$$\mathrm{P}(X_n = i, X_{n+1} = j) \;=\; \mathrm{P}(X_n = j, X_{n+1} = i)\,.$$

Sum this over all $j \in \mathcal{S}$ to get

for all $i \in \mathcal{S}$.

Reversibility

Proof of Theorem 3.28 $\Longleftarrow$

If the x_j are positive and both

hold, then from Theorem 3.27 we see that the DTMC is positive recurrent with

$$x_j \quad = \quad \pi_j, \quad \forall j \in \mathcal{S} \; .$$

In the steady state, $\mathrm{P}(X_n = j) \; = \; \pi_j \; > \; 0$, so that reversing the algebra on the previous part of the proof, we have that

$$\mathrm{P}(X_n = i, X_{n+1} = j) \quad = \quad \mathrm{P}(X_n = j, X_{n+1} = i) \quad \forall i, j \in \mathcal{S} \text{ and } \forall n$$

Reversibility

Example 3.30: Reversibility

Suppose an irreducible positive recurrent DTMC has a state diagram that is a tree, for example like that seen in Figure 3.10

Then the detailed balance (or local balance) equations

$$\pi_i \, p_{i,j} \quad = \quad \pi_j \, p_{j,i}$$

hold for all $i, j \in \mathcal{S}$*, with*

$$\sum_j \pi_j \quad = \quad 1 \, ,$$

and hence is reversible.

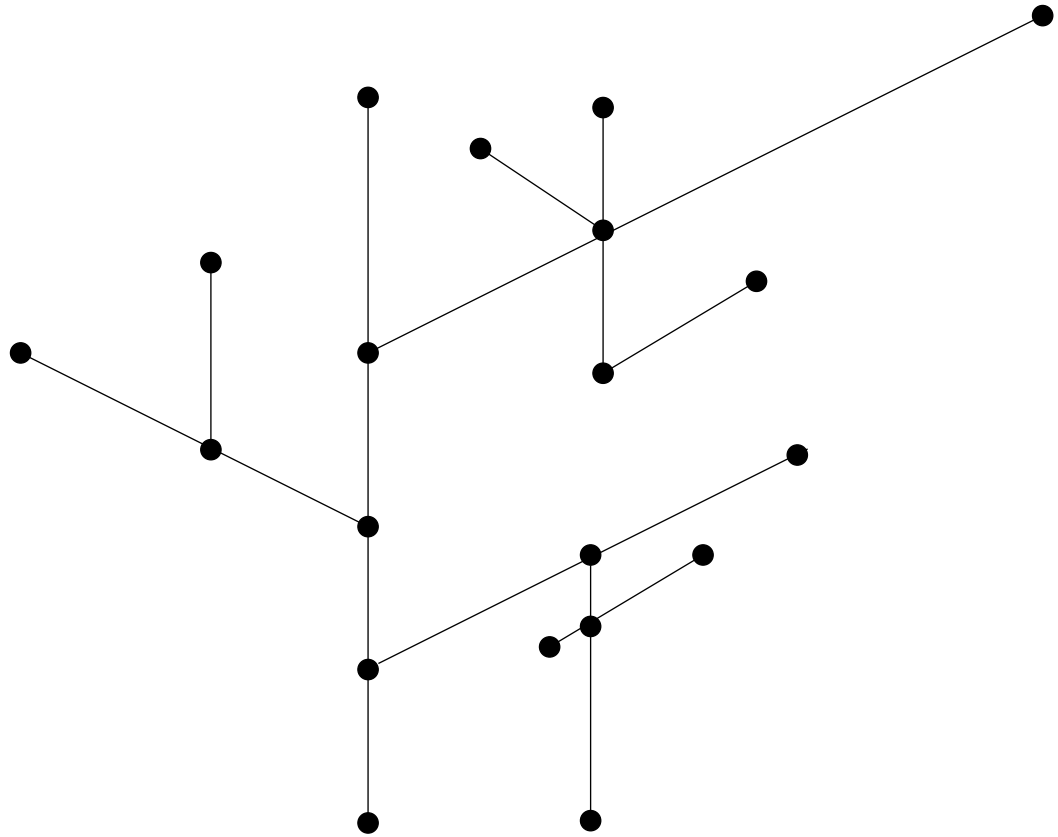


Figure 3.10: State diagram

Kolmogorov's criterion

Theorem 3.29: Kolmogorov's criterion

An irreducible positive recurrent DTMC is reversible if and only if

$$p_{j_1,j_2}\, p_{j_2,j_3} \cdots p_{j_{n-1},j_n}\, p_{j_n,j_1} \quad = \quad p_{j_1,j_n}\, p_{j_n,j_{n-1}} \cdots p_{j_3,j_2}\, p_{j_2,j_1}$$

for all sequences of disjoint states $j_1, \ldots, j_n \in \mathcal{S}$.

Proof of Theorem 3.29: $\Longrightarrow$

If reversibility holds then we have $\pi_i\, p_{i,j} = \pi_j\, p_{j,i} \quad \forall i, j \in \mathcal{S}$ and hence

Dividing both sides by the positive constant $\pi_{j_1} \pi_{j_2} \ldots \pi_{j_{n-1}} \pi_{j_n}$ yields

$$p_{j_1,j_2}\, p_{j_2,j_3} \cdots p_{j_{n-1},j_n}\, p_{j_n,j_1} \quad = \quad p_{j_1,j_n}\, p_{j_n,j_{n-1}} \cdots p_{j_3,j_2}\, p_{j_2,j_1} \ .$$

Kolmogorov's criterion

Proof of Theorem 3.29: $\Longleftarrow$

Consider states $j, k \in \mathcal{S}$.

If $p_{j,k} = p_{k,j} = 0$, then regardless of the choose of x_j and x_k, we have that $x_j p_{j,k} = x_k p_{k,j}$, as required.

Suppose the cycle condition

$$p_{j_1,j_2}\, p_{j_2,j_3} \cdots p_{j_{n-1},j_n}\, p_{j_n,j_1} \quad = \quad p_{j_1,j_n}\, p_{j_n,j_{n-1}} \cdots p_{j_3,j_2}\, p_{j_2,j_1}$$

holds for all sequences of disjoint states $j_1, \ldots, j_n \in \mathcal{S}$.

Assume now that $p_{j,k} > 0$. Then irreducibility implies that there must be a disjoint cycle, with all edges marked by positive probabilities, containing the edge $\{j, k\}$, as k must be able to access j. The cycle condition then implies that $p_{k,j} > 0$.

So we are left with only considering the states $j, k \in \mathcal{S}$ such that $p_{j,k} > 0$ and $p_{k,j} > 0$.

Kolmogorov's criterion

Proof of Theorem 3.29: $\Longleftarrow$ (cont)

Choose an arbitrary reference state i_0 and consider $j \in \mathcal{S}$. Then irreducibility implies that there exist disjoint states $\{j_1, \ldots, j_n, j_1^*, \ldots j_m^*\} \in \mathcal{S} \setminus \{i_0, j\}$ such that

The cycle condition then says that

$$\begin{aligned} &\left(p_{i_0,j_1}\, p_{j_1,j_2} \cdots p_{j_{n-1},j_n}\, p_{j_n,j}\right)\left(p_{j,j_m^*}\, p_{j_m^*,j_{m-1}^*} \cdots p_{j_2^*,j_1^*}\, p_{j_1^*,i_0}\right) \\ &\quad = \quad \left(p_{i_0,j_1^*}\, p_{j_1^*,j_2^*} \cdots p_{j_{m-1}^*,j_m^*}\, p_{j_m^*,j}\right)\left(p_{j,j_n}\, p_{j_n,j_{n-1}} \cdots p_{j_2,j_1}\, p_{j_1,i_0}\right) \end{aligned}$$

or

$$\frac{p_{i_0,j_1}\, p_{j_1,j_2} \cdots p_{j_{n-1},j_n}\, p_{j_n,j}}{p_{j,j_n}\, p_{j_n,j_{n-1}} \cdots p_{j_2,j_1}\, p_{j_1,i_0}} \quad = \quad \frac{p_{i_0,j_1^*}\, p_{j_1^*,j_2^*} \cdots p_{j_{m-1}^*,j_m^*}\, p_{j_m^*,j}}{p_{j,j_m^*}\, p_{j_m^*,j_{m-1}^*} \cdots p_{j_2^*,j_1^*}\, p_{j_1^*,i_0}}.$$

Kolmogorov's criterion

Proof of Theorem 3.29: $\Longleftarrow$ (cont)

Then, for some positive constant B, write

by the above equation. Hence we see that the value of x_j, $j \in \mathcal{S}$ is independent of the path chosen to define it.

Kolmogorov's criterion

Proof of Theorem 3.29: $\Longleftarrow$ (cont)

Therefore, since $p_{j,k} > 0$ and $p_{k,j} > 0$, we can also write

Hence, for all $j, k \in \mathcal{S}$,

Kolmogorov's criterion

Proof of Theorem 3.29: $\Longleftarrow$ (cont)

Summing over all $k \in \mathcal{S}$ implies

so that the DTMC is positive recurrent and these equations have a unique solution up to a scalar factor. Then by a suitable choice of B we can ensure that

which says the DTMC is reversible by Theorem 3.28. ■

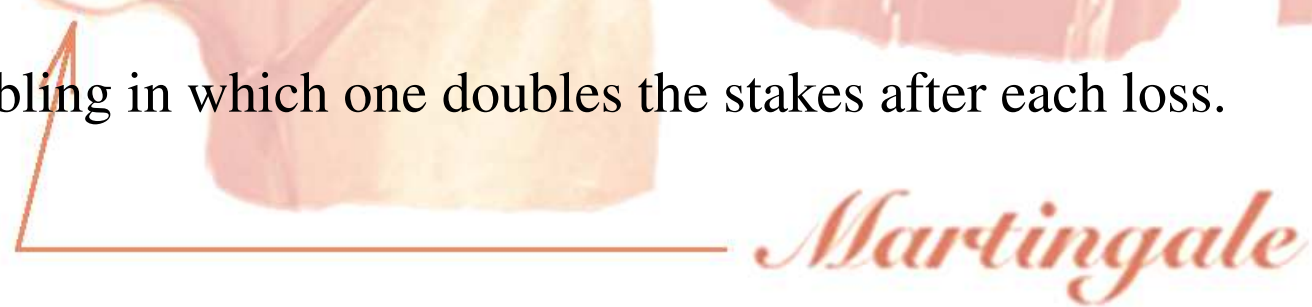
Course outline

- **Section 00:** Probability Review *(Assumed Knowledge)*
- **Section 01:** Introduction
- **Section 02:** Probability and measure
  - Sample space
  - Algebras and σ-algebras of events
  - Probability measure
- **Section 03:** Discrete Time Markov Chains
  - Basic definitions
  - Hitting probabilities and hitting times
  - Classification of states
  - Recurrence and transience
  - Limiting behaviour
- **Section 04:** Martingales
  - Definition of a Martingale
  - Stopping times and the Optional stopping theorem
- **Section 05:** Brownian motion

Martingales

Etymology of Martingale

- A part of a harness designed to prevent a horse from throwing it's head back.
- *(Nautical)* Any of several parts of standing rigging strengthening the bowsprit and jib boom against the force of the head stays.
- *(Fencing)* A leather strap attached to the handle of a foil or epeé and looped around the the hand as a safety precaution in the event of a disarm.
- A loose half belt or strap placed on the back of a garment, such as a coat or jacket.
- A method of gambling in which one doubles the stakes after each loss.



Martingales

Stochastic processes are characterised by the dependence relationships that exist among their variables. In the previous section, we considered a discrete time random process X_n, where each $X_n, n \in \mathbb{N} = \{0, 1, 2, \ldots\}$, is a measurable function mapping a sample space Ω into a state space $\mathcal{S}$ such that

$$\mathrm{P}(X_n = s \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) \quad = \quad \mathrm{P}(X_n = s \mid X_{n-1} = x_{n-1}) \ ,$$

for all $n \geq 1$ and all $s, x_0, \ldots, x_{n-1} \in \mathcal{S}$.

The increments of a random process are the differences $X_s - X_t$ between its values at different times $t < s$.

The Markovian description for $\{X_n : n \in \mathbb{N}\}$ above, implies that the random process has independent increments. This means that the increments $X_s - X_t$ and $X_u - X_v$ for $s, t, u, v \in \mathbb{N}$ are independent random variables whenever the two time intervals do not overlap and, more generally, any finite number of increments assigned to pairwise non-overlapping time intervals are mutually (not just pairwise) independent.

Martingales

Almost sure events and events of probability 0.

In elementary examples, a random variable X can take on only a finite number of distinct values from a finite state space Ω and

More commonly however, for most applications, uncountability is the usual situation. Then, when X can take on uncountably many different values, the implications are very different. For example, if we choose a point ω from $[0, 1]$ with a uniform distribution, we have that

Similarly if we set $\mathbb{Q}_1$ to be the set of all rational numbers in $[0, 1]$, we have that $P(X \in \mathbb{Q}_1) = 0$, but $X \in \mathbb{Q}_1$ is not an impossible event...

Martingales

Almost sure events and events of probability 0.

In general we therefore say that if $P(A) = 1$, then

and if $P(A) = 0$, then

As we will see later, $X = Y$ a.s., or $P(X = Y) = 1$, implies that

Martingales

Lemma 3.1

$$
\begin{aligned}
E[T] &= \sum_{k=1}^{\infty} P(T \geq k) \qquad \textit{discrete rv} \\
E[T] &= \int_{t=0}^{\infty} P(T \geq t)\, dt \qquad \textit{continuous rv.}
\end{aligned}
$$

Proof:

$$
\begin{aligned}
E[T] &= \sum_{j=1}^{\infty} jP(T=j), \\
&= \sum_{j=1}^{\infty} P(T=j) + \sum_{j=2}^{\infty} P(T=j) + \cdots + \sum_{j=k}^{\infty} P(T=j) + \cdots \\
&= P(T \geq 1) + P(T \geq 2) + \cdots + P(T \geq k) + \cdots \\
&= \sum_{k=1}^{\infty} \mathrm{P}(T \geq k)\,.
\end{aligned}
$$

Martingales

Corollary 3.2

1.

$$E[T] < \infty \quad \Rightarrow \quad P[T \geq k] \to 0 \quad \textit{as} \quad k \to \infty,$$

$$\Rightarrow \quad P(T < \infty) = 1.$$

2.

$$E[T] < \infty \quad \Rightarrow \quad \sum_{k=n}^{\infty} k\, P(T = k) \to 0 \quad \textit{as} \quad n \to \infty,$$

$$\Rightarrow \quad n \sum_{k=n}^{\infty} P(T = k) \to 0 \quad \textit{as} \quad n \to \infty,$$

$$\Rightarrow \quad n\, P(T \geq n) \to 0 \quad \textit{as} \quad n \to \infty.$$

Martingales

Definition 3.40: Martingale (a)

A random process $\{X_n : n \in \mathbb{N}\}$ is a martingale if for all n

$$E[|X_n|] < \infty \qquad \text{Absolutely convergent} \tag{3.48}$$

$$E\left[X_{n+1}\middle|X_0, ..., X_n\right] = X_n \quad \text{a.s.} \tag{3.49}$$

Recall that $E[X_n]$ is a number, but notice that $E\left[X_n\middle|X_0, ..., X_{n-1}\right]$ is a random variable.
Because $E[X_n|X_n] = X_n$, we can re-write condition (3.49) as

$$E\left[X_{n+1} - X_n\middle|X_0, ..., X_n\right] = 0 \text{ a.s.}$$

The definition above is the earliest definition which bears historical interest, but a more general definition follows on the next slide.

Martingales

Definition 3.41: Martingale (b)

A random process $\{X_n : n \in \mathbb{N}\}$ is called a martingale with respect to a random process $\{Y_n : n \in \mathbb{N}\}$, if

$$E[|X_n|] < \infty \qquad \text{Absolutely convergent} \tag{3.50}$$

$$E\left[X_{n+1}\middle|Y_0, ..., Y_n\right] = X_n \quad \text{a.s.} \tag{3.51}$$

Of course in classical theory (3.51) really means that

where $\mathcal{F}_n$ is the σ-algebra generated by $(Y_0, \ldots, Y_n)$ or $\sigma(Y_0, \ldots, Y_n)$.

That is, X_n is *adapted* to $\mathcal{F}_n$, so X_n is measurable with respect to $\mathcal{F}_n$, for all n.

The random process $\{X_n : n \in \mathbb{N}\}$ is said to be a martingale (with respect to $\mathcal{F}_n$).

Martingales

Note that (3.51) implies that X_n is, in fact, a *measurable* function of $Y_0, \ldots, Y_n$.

That is,

and $g(\cdot)$ must be measurable with respect to $Y_0, \ldots, Y_n$, by the above comments.

Hence

Martingales

Definition 3.42: Expectation of a conditional expectation

Let X and Y be random variables, then

$$E\Big[E\left[X\Big|Y\right]\Big] \quad = \quad E[X]\,. \tag{3.52}$$

whenever the expectations exist.

To see why this is so, let X and Y be random variables such that

$$F_Y(y) \equiv P(Y \leq y), \quad \text{and} \quad f_Y(y) = \frac{dF_Y(y)}{dy}\,.$$

Note that $P(X \leq x\Big|Y = y)$ is well-defined.

Martingales

Then

Martingales

Hence use of (3.51) implies that

Therefore, by applying this recursively, we get that

The martingale property, in terms of increments, is that the future increments of X_n have conditional mean zero, given the past and present values of the process.

Example 3.8: A fair game

Suppose a gambler has initial wealth X_0 and makes bets with various odds, such that the bets made are all for fair games, in the sense that the expected net gains are zero. Then the wealth of the gambler at time m, given by X_m, is a Martingale.

The random walk $\{X_n : n \in \mathbb{N}\}$ arising in the gamblers ruin problem is an independent increment process, and if $p = \frac{1}{2}$ (which corresponds to a fair game), it is also a Martingale. ◁

Martingales

The Martingale is thus sometimes related to the concept of fairness in gambling. Some early Martingale Theory was motivated by considering how an astute choice of betting could turn a fair game into a favourable one.

The name ‘Martingale’ comes from the French term for a strategy which involved doubling one’s bet until a win was secured (see Example 3.14).

Although we will consider some gambling examples, as they are easily understood, Martingales have a much wider application.

The definition of a Martingale is a restriction on the conditional means of some of the random variables given values of others and does not otherwise depend on the actual distribution of the random variables in the stochastic process.

The simplicity of the Martingale definition yields processes that are both inherently interesting and that provide a very powerful tool for analysing a variety of stochastic processes.

Identifying Martingales in processes then becomes a very useful pursuit.

Martingales

Example 3.9: The sum of independent random variables

If X_j for $j \in \mathbb{N}$ are independent random variables with $E[X_j] = 0$ and $E[|X_j|] < \infty$, then

$$Z_n = \sum_{j=0}^{n} X_j \quad \text{for} \quad n \geq 0$$

is a Martingale. This follows because

Martingales

Example 3.10: The product of independent random variables

If X_j for $j \in \mathbb{N}$ are independent random variables with $E[X_j] = 1$ and $E[|X_j|] < \infty$, then

$$Z_n \quad = \quad \prod_{j=0}^{n} X_j \quad \text{for} \quad n \geq 0$$

is a Martingale. This follows because

Example 3.11: A marble game (Pólya's Urn)

A bowl initially contains n red and m green marbles. A marble is drawn from the bowl at random and then replaced along with another marble of the same colour. Hence, after the i^{th} draw, there will be $n + m + i$ marbles in the bowl. If we let Y_i be the number of red marbles in the bowl at time i, then the conditional distribution of Y_{i+1} is

$$Y_{i+1} = \begin{cases} Y_i + 1 & \text{with probability} \quad \frac{Y_i}{n+m+i} \\ Y_i & \text{with probability} \quad 1 - \frac{Y_i}{n+m+i} \end{cases}$$

The conditional expectation

Martingales

Example 3.11: A marble game (Pólya's Urn) (cont)

If we let X_i be the proportion of red marbles in the bowl at time i, then $\{X_i \big| i \in \mathbb{N}\}$ is a Martingale wrt $\{Y_i \big| i \in \mathbb{N}\}$. This follows because

and so $E[|X_i|]$ is finite for all i (in fact $0 \leq X_i \leq 1$ for all $i \geq 0$) and

Martingales

Example 3.12: Stock prices

Let X_n be the closing price at the end of day n of a certain publicly traded security such as a share of stock. Now, the daily prices fluctuate, but many believe that in a perfect market, these price sequences $\{X_n : n \in \mathbb{N}\}$ should be Martingales. That is, in a perfect market, it should not be possible to predict with any degree of accuracy whether a future price X_{n+1} will be higher or lower than the current price X_n. Otherwise there will be a no risk win. ◁

Example 3.13: Branching processes

Consider a population branching process as seen in the previous section. Let X_n be the size of the n^{th} generation and let m be the mean number of offspring per individual. If we define the random variable $Z_n = \dfrac{X_n}{m^n}$ for $n \in \mathbb{N}$, then

$$\{Z_n : n \in \mathbb{N}\} \text{ is a martingale wrt } \{X_n : n \in \mathbb{N}\}.$$

(see class exercise or tutorial) ◁

Martingales

Example 3.14: Eponymous example

At each toss of an unbiased coin, a bet is placed such that if a head is realised, the gambler gains as much as has been bet and loses that bet if a tail is seen. Let the random variable

$$\beta_k = \begin{cases} -1 & \text{if a tail is observed at the } k^{th} \text{ toss} \\ +1 & \text{if a head is observed at the } k^{th} \text{ toss.} \end{cases}$$

Let the amount bet on the first game be $f_1 = 1$ unit and then define

$$f_k = f_k(\beta_1, \ldots, \beta_{k-1}) \quad \text{for } k \geq 2$$

as the k^{th} bet placed according to some prescribed function f_k. The net gain of the gambler after the n^{th} bet is then a random variable

Example 3.14: Eponymous example (cont)

If we assume that $\left|f_k\right| < \infty$ for each k, $E[|X_n|]$ is finite for all $n \geq 1$ and

Example 3.14: Eponymous example (cont)

Since the outcomes of each successive toss of the coin are independent,

and if the coin is unbiased,

so that

and hence $\{X_n \big| n \geq 0\}$ is a Martingale with respect to $\{\beta_n, n \geq 0\}$. ◁

Example 3.14: Eponymous example (cont)

Consider the following gambling strategy on hand k:

- If you won on hand $k-1$, then you bet 1 unit.
- If your last win was on hand $k-1-j$, then you bet 2^j units.

This is the eponymous martingale and defines a messy $f_k(\beta_1, \ldots, \beta_{k-1})$.

Then, if the first head occurs on the T^{th} bet, the net gain X_T from the first T tosses is given by

◁

Example 3.14: Eponymous example (cont)

Note that if p is the probability that a head is realised at each toss (and the eponymous strategy is still played), then

Now, for every finite value of T,

and so

That is, T is almost surely finite and X_T is almost surely defined, with $E[X_T] = 1$.

Example 3.14: Eponymous example (cont)

Note that in the doubling strategy, we assume that the gambler can lose an arbitrarily large amount before the first head occurs.

Now in contrast, for a fixed time n, in the case of an unbiased coin, we have

Martingales

Definition 3.43: Stopping time

If $\{X_n, n \in \mathbb{N}\}$ is a stochastic process, then a random variable T is a Markov (stopping) time w.r.t. $\{X_n\}$ if

1. the range of $T \subset \mathbb{N}$.
2. for each $n \geq 0$, whether or not the event $\{T = n\}$ occurs is determined by $(X_0, X_1, \ldots, X_n)$.

If in addition to the above two conditions we have

3. $T < \infty$ a.s. $(\equiv P(T < \infty) = 1)$, then T is an almost sure finite Markov time or stopping time.

The description is commonly given by the indicator function

$$I_{\{T=n\}} = I_{\{T=n\}}(X_0, \ldots, X_n) \quad \big(\text{a function of } (X_0, \ldots, X_n)\big)$$

$$= \begin{cases} 1 & \text{if event } T = n \\ 0 & \text{otherwise.} \end{cases}$$

Martingales

Example 3.15:

i. The k^{th} time a process visits a collection of states K is a stopping time (it only looks at past history).

ii. The last time a process visits a collection of states K is not in general a stopping time (as it needs to look into the future).

◁

Elementary properties

If S and T are stopping times for a given process, then so are

i. $S + T$

ii. $S \wedge T \equiv \min(S, T)$ 'S wedge T '

iii. $S \vee T \equiv \max(S, T)$ 'S vee T '.

In particular, for a given n, $T \wedge n$ is a stopping time.

Martingales

Proofs : (Obvious, but based on being a function of past history)

i.

ii.

iii.

Martingales

Elementary properties of martingales

We know that

Let's assume, for some given value of $k \geq 0$, that

$$E\left[X_{n+k} \middle| Y_0, \ldots, Y_n\right] = X_n \text{ a.s.}$$

Then

Martingales

Lemma 3.2

Suppose that $\{X_n : n \in \mathbb{N}\}$ *is a martingale and that* T *is a stopping time wrt* $\{Y_n : n \in \mathbb{N}\}$*, then for each* $n \geq 0$*,*

$$E[X_n] \;=\; E[X_{T\wedge n}] \;=\; E[X_0] \;. \tag{3.53}$$

Proof:

for $n \geq k$, we have

Martingales

Proof: (cont)

3.53

■

Martingales

Theorem 3.30 (Lebesgue Dominated Convergence Theorem (in this context)**)**

$$\textit{If} \quad \begin{cases} a. & Z_n \textit{ is a real valued random variable for all } n \geq 0 \\ b. & |Z_n| \leq V \textit{ a.s. and } E[V] < \infty \\ c. & Z_n \to Z \textit{ as } n \to \infty \end{cases}$$

then $E[Z]$ *exists and* $E[Z_n] \to E[Z]$ *a.s. as* $n \to \infty$.

This tells us when $Z_n \to Z$ implies $E[Z_n] \to E[Z]$, which is non-trivial.

Example 3.16: Condition (b) is crucial!

$$\text{Define} \quad Z_n = \begin{cases} 0 & \text{with probability } \left(1 - \frac{1}{n}\right) \\ n & \text{with probability } \frac{1}{n} \end{cases}$$

Then $E[Z_n] = 0\left(1 - \frac{1}{n}\right) + n\left(\frac{1}{n}\right) = 1 \ \ \forall n$,

but $Z_n \to Z = 0$ a.s. as $n \to \infty$ and so $E[Z] = 0$

Now $Z_n \to Z$, but $E[Z_n] = 1 \not\to E[Z] = 0$. ◁

Martingales

Example 3.16: (cont)

Can we find $V \geq |Z_n|,$ a.s. ?

$$
\begin{aligned}
P(Z_1 = 1) = 1 \quad &\Rightarrow \quad P(V \geq 1) = 1. \\
P(Z_2 = 2) = \frac{1}{2} \quad &\Rightarrow \quad P(V \geq 2) = \frac{1}{2}, \\
P(Z_3 = 3) = \frac{1}{3} \quad &\Rightarrow \quad P(V \geq 3) = \frac{1}{3}, \\
P(Z_4 = 4) = \frac{1}{4} \quad &\Rightarrow \quad P(V \geq 4) = \frac{1}{4}, \\
&\vdots \\
P(Z_k = k) = \frac{1}{k} \quad &\Rightarrow \quad P(V \geq k) = \frac{1}{k},
\end{aligned}
$$

$$
V = k, \text{with probability} \frac{1}{k(k+1)} \; k \geq 1.
$$

Martingales

Example 3.16: (cont)

Is $E[V] < \infty$?

Therefore Condition (b) above does not hold. ◁

Martingales

Theorem 3.31 (Optional Stopping Theorem)

Suppose $\{X_n : n \in \mathbb{N}\}$ *is a martingale and* T *is an almost sure finite stopping time wrt* $\{Y_n : n \in \mathbb{N}\}$*, then*

$$\text{if} \quad \begin{cases} \textit{(i)} & E\left[|X_T|\right] < \infty \\ \textit{(ii)} & E\left[X_n I_{\{T>n\}}\right] \to 0 \quad \textit{as} \quad n \to \infty \end{cases}$$

then $E\left[X_T\right] = E\left[X_0\right]$.

Proof:

Set $Z_n = X_T I_{\{T \le n\}}$, $Z = X_T$ and $V = \left|X_T\right|$, then

- Z_n is a real valued random variable for all $n \ge 0$
- $|Z_n| \le V$ a.s. and $E\left[V\right] < \infty$
- $I_{\{T \le n\}} \to 1$ a.s. as $n \to \infty$,

so that $Z_n = X_T I_{\{T \le n\}} \to X_T = Z$ a.s. as $n \to \infty$. (cont)

Martingales

Proof: (cont)

All conditions of the DCT are satisfied, so that $E[Z_n] \to E[Z]$.

Now we have that

3.2

Therefore, $E[X_T] = E[X_0]$. ■

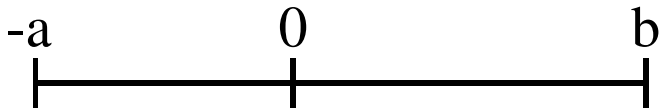
Martingales

Example 3.17: Random walk on the integers

Let $\{X_n : n \in \mathbb{N}\}$ be an IID sequence of random variables that take the values $\{-1, 1\}$ for each $n \in \mathbb{N}$, as follows

$$P(X_n = 1) = \frac{1}{2} = P(X_n = -1) \, .$$

Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$ describe the position of the walk at time $n \in \mathbb{N}$ and for $a, b \in \mathbb{Z}^+$, we bound the walk as follows



Then let T be the first time that n is such that $S_n = -a$ or b

(the time the random walk first hits the boundary $-a$ or b)

It is then clear that S_n is a martingale wrt. X_n. ◁

Example 3.17: Random walk on the integers (cont)

Setting $-a$ and b as absorbing states, the random walk is also a Markov chain with all the interior states being transient. Therefore with probability 1, the process spends only a finite time on these interior states, so that $T < \infty$ is an almost sure finite stopping time.

Further, since a, b are finite, it is clear that

and

Hence the condition of the OST holds and so the OST implies that

◁

Martingales

Example 3.17: Random walk on the integers (cont)

Let $V_a = P\left(S_n \text{ reaches } -a \text{ before } b\right)$,

Then $S_T = \begin{cases} -a & \text{with probability } V_a \\ b & \text{with probability } 1 - V_a. \end{cases}$

Hence,

Therefore,

Example 3.17: Random walk on the integers (cont)

Let's now consider the random variable $Z_n = S_n^2 - n$.

Note that $S_{n+1} = S_n + X_{n+1} = \begin{cases} S_n + 1, & \text{with probability } 1/2, \\ S_n - 1, & \text{with probability } 1/2. \end{cases}$

Martingales

Example 3.17: Random walk on the integers (cont)

Further,

Hence, Z_n is a martingale wrt. S_n. ◁

Example 3.17: Random walk on the integers (cont)

As above, since a, b are finite, the conditions of the OST are satisfied.

Therefore,

where

$$Z_T = \begin{cases} & \text{with probability} \quad , \\ & \text{with probability} \quad , \end{cases}$$

and $V_a = \dfrac{b}{a+b}$. ◁

Example 3.17: Random walk on the integers (cont)

Therefore,

Hence

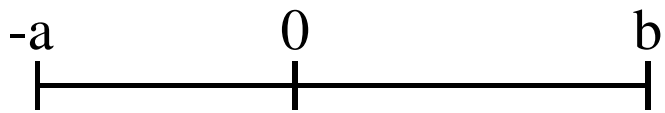
Martingales

Example 3.18: Biased random walk on the integers

Let $\{X_n : n \in \mathbb{N}\}$ be an IID sequence of random variables that take the values $\{-1, 1\}$ for each $n \in \mathbb{N}$, as follows

$$\begin{aligned} P(X_n = 1) &= p > 0 \\ P(X_n = -1) &= q = 1 - p > 0 \,. \end{aligned}$$

Let $S_0 = 0$ and $S_n = \sum_{i=1}^{n} X_i$ describe the position of the walk at time $n \in \mathbb{N}$ and for $a, b \in \mathbb{Z}^+$, we bound the walk as follows



Let $\theta_n = \left(\frac{q}{p}\right)^{S_n}$ for all $n \geq 0$, then we showed, in Tutorial 5, that θ_n is a martingale wrt X_n.

◁

Example 3.18: Biased random walk on the integers (cont)

The conditions of the Optional Stopping Theorem (OST) are satisfied for T, since $a, b < \infty$. Therefore,

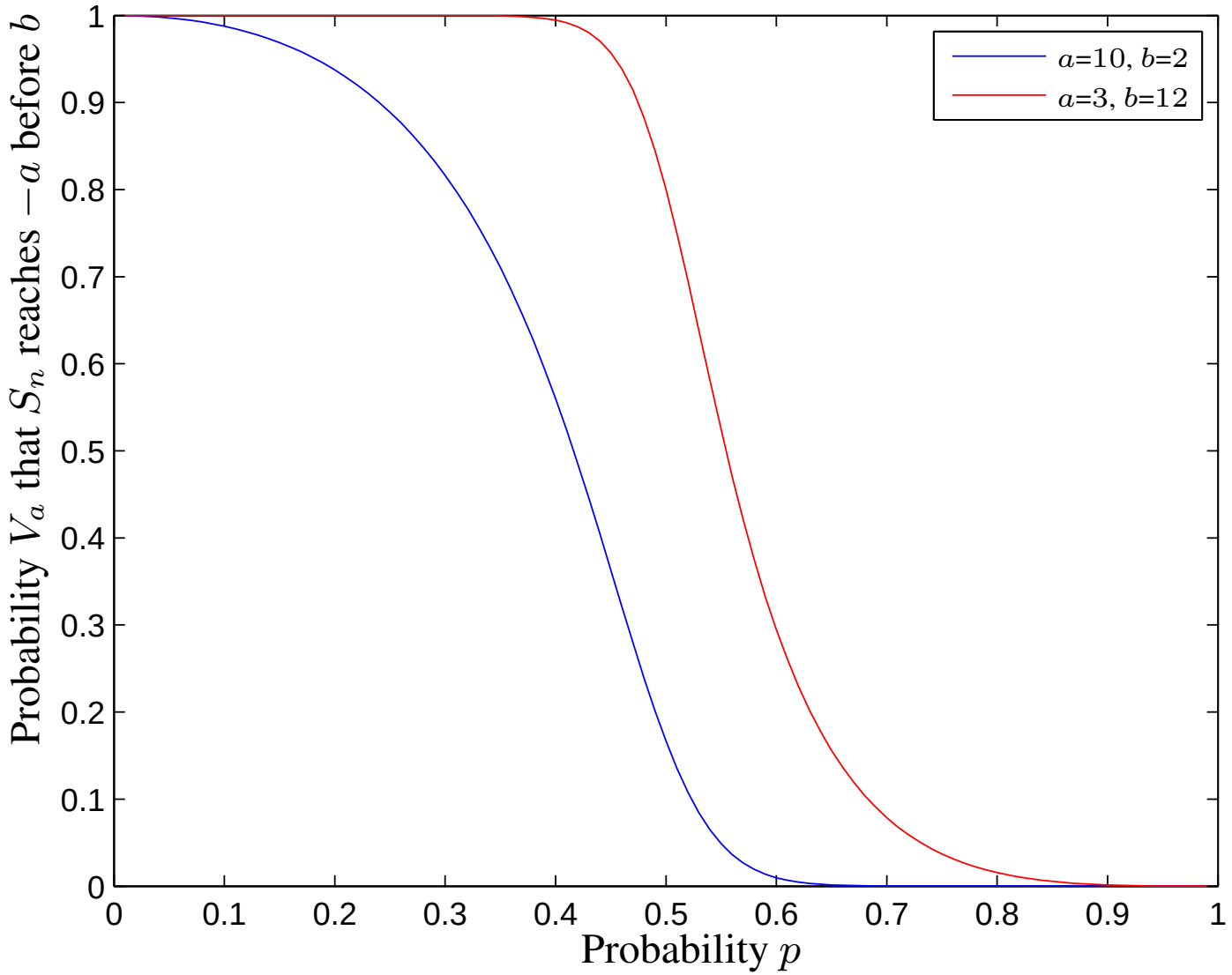


Figure 3.11: Bounded random walk for different values of a and b.

Example 3.19: Biased random walk on the integers (yet again!)

Consider again the biased and bounded random walk on the integers of Example 3.18, but now let $Y_n = S_n - n\mu$, where $\mu = p - q \neq 0$.

Then Y_n is a martingale wrt X_n and again the OST holds, so that

and substitution of V_a from Example 3.18, completes the argument.

◁

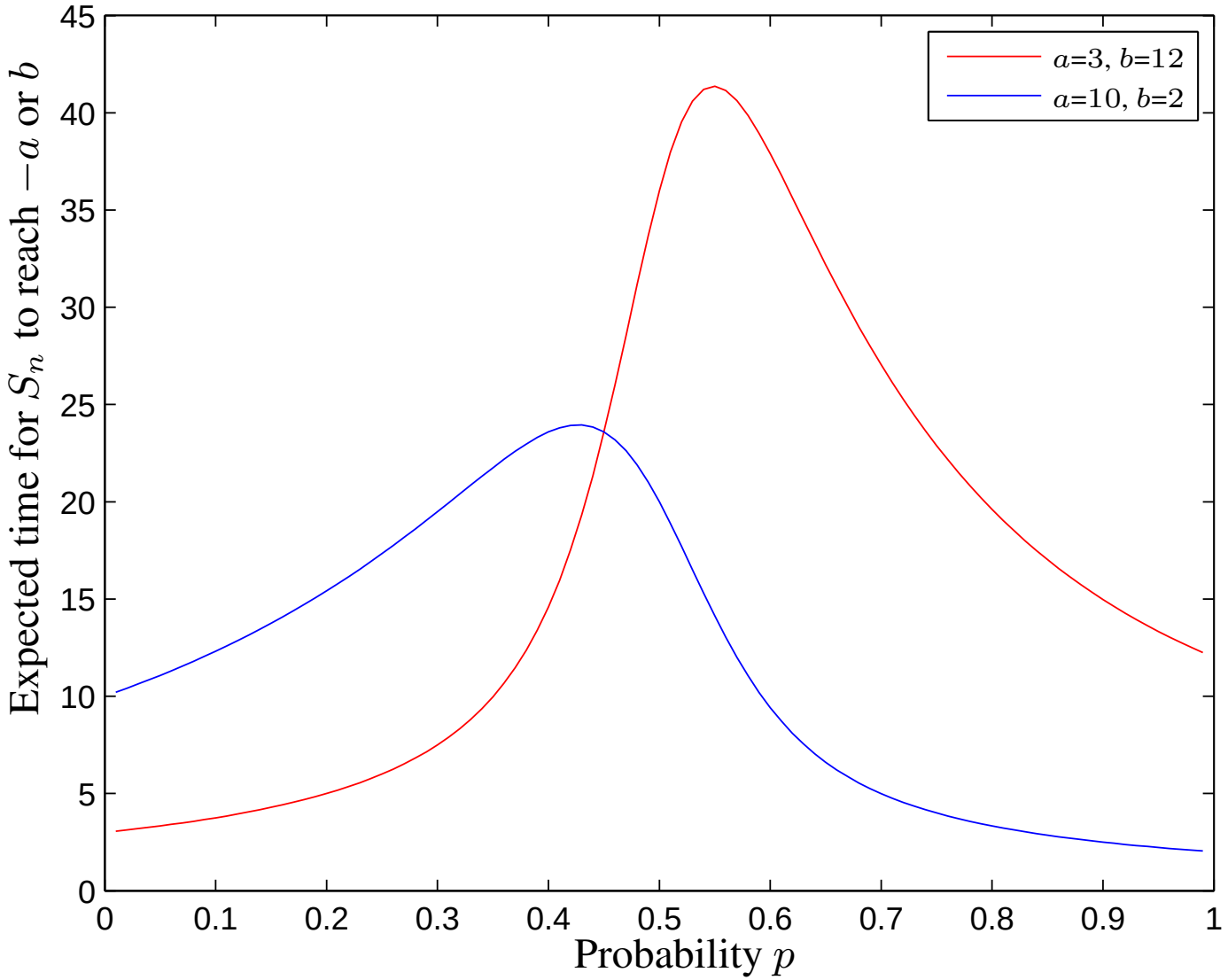


Figure 3.12: Bounded random walk for different values of a and b.

Martingales

Thus far we have "waved our hands" and claimed that the conditions of the OST were satisfied. Now we will consider some examples of the actual verification of the OST conditions. That is, if T is an almost sure finite stopping time then we want to verify that

(i) $E[|Y_T|] < \infty$.

(ii) $E[Y_n I_{\{T \geq n\}}] \to 0, \quad \text{as} \quad n \to \infty$

We first note that $E[|Y_T|] < \infty$ is not an immediate consequence of $E[|Y_n|] < \infty$ for all n.

Example 3.20:

In the problem of the random walk on $[-a, b]$, for the almost sure finite stopping time T it is obvious that

(condition (i) of the OST)

and so we need to show that the second condition is true. ◁

Example 3.20: (cont)

We know that

Since $\left|Y_n\right| < (a \vee b)$ for all n,

since $\mathrm{P}(T < n) \to 1$ a.s. as $n \to \infty$.

Hence all of the OST conditions are satisfied. ◁

Martingales

As a martingale embodies the concept of a fair game, there are two modifications to the standard martingale definition that encapsulates the ideas of super-fair (winning) and sub-fair (losing) games.

These ideas are represented in the sub-martingale and super-martingale as given in the following definitions.

Definition 3.44: Submartingale

A random process $\{X_n : n \in \mathbb{N}\}$ is called a submartingale with respect to a random process $\{Y_n : n \in \mathbb{N}\}$, if

$$E\left[\left|X_n\right|\right] < \infty \quad \text{for all } n,$$

$$E\left[X_{n+1}\middle|Y_0, ..., Y_n\right] \geq X_n \quad \text{a.s. and}$$

$$X_n \quad \text{is a (measurable) function of } Y_0, Y_1, \ldots, Y_n \text{ a.s.}$$

Martingales

Definition 3.45: Supermartingale

A random process $\{X_n : n \in \mathbb{N}\}$ is called a supermartingale with respect to a random process $\{Y_n : n \in \mathbb{N}\}$, if

$$E\left[\left|X_n\right|\right] < \infty \quad \text{for all } n,$$

$$E\left[X_{n+1}\middle|Y_0, ..., Y_n\right] \leq X_n \quad \text{a.s. and}$$

X_n is a (measurable) function of $Y_0, Y_1, \ldots, Y_n$ a.s.

Notes:

- Every martingale is also a submartingale and a supermartingale.
- Conversely, any stochastic process that is both a submartingale and a supermartingale is a martingale.

Martingales

Some more observations:

Consider again the gambler who wins \$1 when a coin comes up heads and loses \$1 when the coin comes up tails. Suppose now that the coin may be biased, so that it comes up heads with probability p.

If $p = \frac{1}{2}$: the gambler on average neither wins nor loses money, and the gambler's fortune over time is a martingale.

If $p < \frac{1}{2}$: the gambler loses money on average, and the gambler's fortune over time is a *super*martingale.

If $p > \frac{1}{2}$: the gambler wins money on average, and the gambler's fortune over time is a *sub*martingale.

Martingales

Elementary properties of super-martingales

If $\{X_n : n \in \mathbb{N}\}$ is a super-martingale and T is a stopping time with respect to a random process $\{Y_n : n \in \mathbb{N}\}$, then

1. $E[X_{n+1}] \;=\; E\left[E\left[X_{n+1}\middle|Y_0, \ldots, Y_n\right]\right] \;\leq\; E\left[X_n\right]$

   so that $E[X_0] \geq E[X_1] \geq E[X_2] \geq \ldots$

2. $E\left[X_{n+k}\middle|Y_0 \ldots, Y_n\right] \leq X_n$ a.s. for all $k \geq 0$.

3. $E\left[X_n \; I_{\{T=k\}}\right] \;\leq\; E\left[X_k \; I_{\{T=k\}}\right]$ for all $n \geq k$.

Note that in the case of a sub-martingale we can just use the opposite inequality signs.

Martingales

Example 3.21: Construction of martingales

Given that $\{X_n : n \in \mathbb{N}\}$ is a super-martingale (or sub-martingale) with respect to a random process $\{Y_n : n \in \mathbb{N}\}$, let

$$Z_0 = 0 \text{ and } Z_n = \sum_{k=1}^{n} \left(X_k - E\left[X_k \middle| Y_0, \ldots, Y_{k-1}\right]\right)$$

for all $k \geq 1$. Then $\{Z_n : n \in \mathbb{N}\}$ is a martingale wrt $\{Y_n : n \in \mathbb{N}\}$.

To see this, consider

Satisfying equation (3.50) in the definition of a martingale. ◁

Example 3.21: Construction of martingales (cont)

because X_k is a function of $Y_0, \ldots, Y_k$.

Thus, we also have satisfied equation (3.51) in the definition of a martingale.

Hence $\{Z_n : n \in \mathbb{N}\}$ is a martingale wrt $\{Y_n : n \in \mathbb{N}\}$. ◁

Course outline

- **Section 00:** Probability Review *(Assumed Knowledge)*
- **Section 01:** Introduction
- **Section 02:** Probability and measure
  - Sample space
  - Algebras and σ-algebras of events
  - Probability measure
- **Section 03:** Discrete Time Markov Chains
  - Basic definitions
  - Hitting probabilities and hitting times
  - Classification of states
  - Recurrence and transience
  - Limiting behaviour
- **Section 04:** Martingales
  - Definition of a Martingale
  - Stopping times and the Optional stopping theorem
- **Section 05:** Brownian motion

Brownian motion

Brownian motion, sometimes called the Wiener process, is one of the most useful stochastic processes in applied probability theory, named after its discoverer, the English Botanist Robert Brown. He described the motion exhibited by a small particle which is totally immersed in a liquid or gas and subsequently has been used in such fields as quantum mechanics and even to analyse price levels in the stock market. From a purely mathematical perspective, Brownian motion gave rise to the study of continuous-time martingales.

The first explanation of the phenomenon of Brownian motion was given by Albert Einstein in 1905, when he claimed that the small immersed particle was continually being bombarded by molecules of the surrounding medium (liquid or gas). The mathematical definition however, was due to Norbert Wiener, given in a series of papers originating in 1918.

Predating all of this is the fact that in 1900, in Paris, Louis Bachelier proposed what we would now call a "Brownian motion model" for the movement of stock prices in the French bond market. His work was largely ignored at the time, but now stands as the innovative first step in a mathematical theory of stock markets that has greatly altered the current financial world.

Brownian motion

Consider the symmetric random walk on $\mathbb{Z}$ with $p = \frac{1}{2}$ (a fair game). This is a Markov chain $\{X_n : n \in \mathbb{Z}\}$, with $P_{i,i+1} = \frac{1}{2} = P_{i,i-1}$ for $i = 0, \pm 1, \pm 2, \ldots$

Now, speed up the process by taking smaller and smaller steps in smaller and smaller time intervals. If we take this to a limit in the correct manner, what we obtain is known as Brownian motion.

More precisely, suppose that at each time unit Δt, we take a step of size Δx either to the left or right with equal probability. If we let $S(t)$ denote the position of the random walk at time t, then

where $\left\lfloor \frac{t}{\Delta t} \right\rfloor$ is the largest integer less than or equal to $\frac{t}{\Delta t}$.

Brownian motion

By the Central Limit Theorem, $S(t)$ is then *normally* distributed with variance $\sigma^2 t$ and the process $\{S(t) : t \geq 0\}$, is such that for all $t_1 < t_2 < \cdots < t_n$, the increments

$$S(t_n) - S(t_{n-1}), \ldots, S(t_2) - S(t_1), S(t_1)$$

are *independent* (by construction). Then, since the distribution of the change in position depends only on the length of the interval over time, $\{S(t)|t \geq 0\}$ also must have *stationary* increments, such that $S(t+s) - S(t)$ does not depend on t.

SHOW MOVIE.

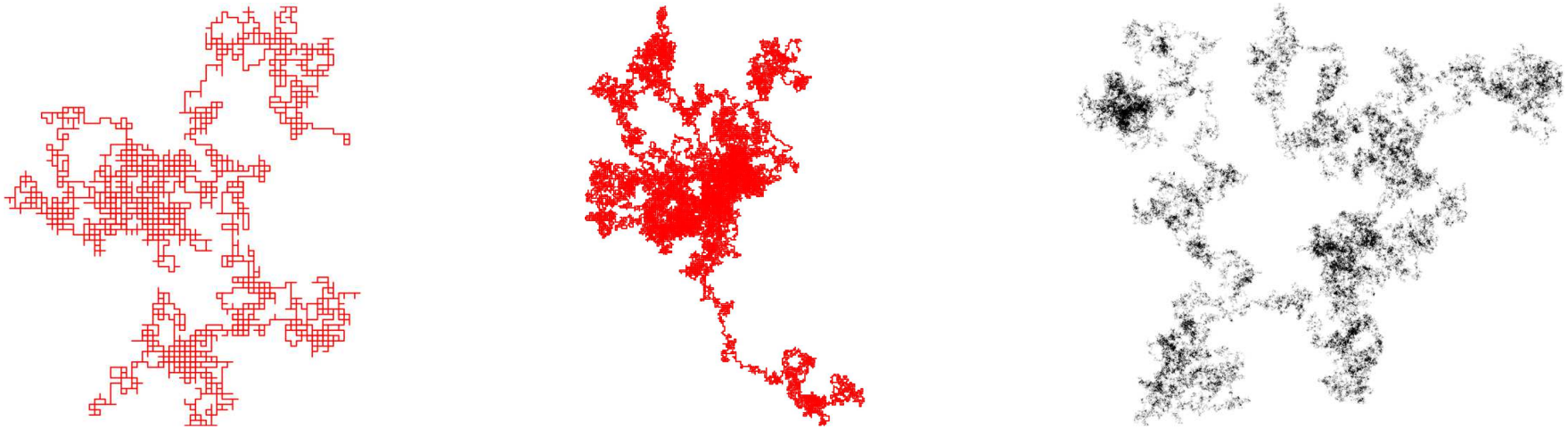


Figure 3.13: 2D random walks with progressively smaller steps

Brownian motion

Observe that the increments in X_n are independent and such that

Therefore,

If we let $\Delta x = \sigma \Delta t$ for some $\sigma > 0$, taking limits as $\Delta t \to 0$ yields

However, if we let $\Delta x = \sigma \sqrt{\Delta t}$ for some $\sigma > 0$, taking limits as $\Delta t \to 0$ yields

Brownian motion

Definition 3.46: Brownian motion

A stochastic process $\{X(t)|t \geq 0\}$ is said to be Brownian motion if

a. $X(0) = 0$

b. $\{X(t)|t \geq 0\}$ has stationary and independent increments, and

c. for every $t > 0$, $X(t) \sim N(0, \sigma^2 t)$.

When $\sigma = 1$, the process is called a standard Brownian motion $B(t)$. Any Brownian process $X(t)$ with $\sigma \neq 1$ can be rescaled to a standard Brownian motion by setting $B(t) = \frac{X(t)}{\sigma}$

As $B(t)$ is a continuous function of t (a fact that is non-trivial to prove), with mean 0 and variance t, its density function is given by

$$f_t(x) = \frac{1}{\sqrt{2\pi t}} e^{\frac{-x^2}{2t}}$$

Brownian motion

Interesting to note is that Brownian motion is an example of a process which is everywhere continuous but nowhere differentiable.

Example 3.22: Hitting times

Let T_b denote the first time that a Brownian motion hits $b > 0$, conditional on starting at 0.

The last term is equal to 0 because if the process was at a point greater than or equal to b at time t, by continuity, it must have reached b at a time less than or equal to time t and therefore $T_b \leq t$. ◁

Brownian motion

Example 3.22: Hitting times (cont)

If $T_b \leq t$, then the process hits position b at some time point in $[0, t]$ and could equally likely either be above or below position b at time t.
That is,

so we see that

◁

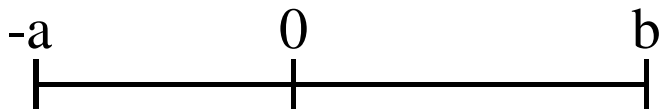
Brownian motion

Example 3.22: Hitting times (cont)

The distribution of the maximum value that the process achieves in $[0, t]$ is also given by

◁

Recall in the random walk the conditional probability that the process hits $-a$ before b for $a, b > 0$, given that it starts in state 0.



In the symmetric random walk, P (process reaches $-a$ before $b) = \frac{b}{a+b}$, which when we take limit as $\Delta x = \sigma\sqrt{\Delta t} \to 0$ yields the same result in the case of Brownian motion.

Brownian motion

There are many variations on Brownian motion, two of which are given below.

Example 3.23: Brownian motion with drift

A random process $\{X(t) : t \geq 0\}$ is called a Brownian motion with drift coefficient μ and variance parameter σ^2 if

i. $X(0) = 0$

ii. $\{X(t) : t \geq 0\}$ has stationary and independent increments, and

iii. for every $t > 0$, $X(t) \sim N(\mu\, t, \sigma^2\, t)$.

Equivalently, we set

$$X(t) = \sigma B(t) + \mu\, t\ ,$$

where $\{B(t) : t \geq 0\}$ is a standard Brownian motion.

◁

Brownian motion

Example 3.24: Geometric Brownian motion

If $\{Y(t) : t \geq 0\}$ is a Brownian motion with drift coefficient μ and variance parameter σ^2, then the process $\{X(t) : t \geq 0\}$ defined by

$$X(t) = e^{Y(t)}$$

is a geometric Brownian motion.

◁

Geometric Brownian motion is useful for example in the modelling of stock prices over time when the percentage changes are independent and identically distributed.

Brownian motion in all its forms is an example of a continuous random process which we have considered as the limiting case of a discrete time random process (the random walk). Many other continuous random processes may also be similarly derived. To continue an excursion into continuous random processes, see the second semester course Random Processes III, which considers in particular the Continuous Time Markov Chain (CTMC).