MATH60045 and MATH70045
Applied Probability

Almut E. D. Veraart
Room 551, Huxley Building
Department of Mathematics
Imperial College London
180 Queen's Gate, London, SW7 2AZ
E-Mail: a.veraart@imperial.ac.uk

Autumn 2022

(Last Update: Tuesday 13th September, 2022)

Contents

Chapter 1

Course information

1.1 Course content

Course details: This course is focussed upon applied probability, which is a rather complex academic discipline. We will study discrete and continuous-time stochastic processes and their applications, that is random variables that will vary according to a *time* parameter. This parameter can change either discretely or continuously. In particular, we will focus upon three broad areas:

- Markov chains in discrete time
- Markov chains in continuous time (incl. Poisson, Birth-death processes)
- Brownian Motion: A Markov process in continuous time on a continuous state space

Applications: During the course we will look at genuine applications of these ideas, within the context of finance, population genetics. Examples include

- The evolution of insurance claims over time.
- The time evolution of DNA sequences back to a most recent common ancestor.
- The evolution of a stock over a day of financial trading.

1.1.1 Syllabus

This module introduces stochastic processes and their applications. The theory of different kinds of processes will be described and illustrated with applications in several areas.

The course covers the following topics:

- Discrete-time Markov chains: Chapman-Kolmogorov equations. Recurrent, transient, periodic, aperiodic chains. Returning probabilities and times. Communicating classes. The basic limit theorem. Stationarity. Ergodic Theorem. Time-reversibility. Random walks. Gambler's ruin.
- Poisson processes: Poisson processes and their properties; Superposition, thinning of Poisson processes; non-homogeneous, compound, and doubly stochastic Poisson processes. Autocorrelation functions. Generating functions and how to use them.
- General continuous-time Markov chains: generator, forward and backward equations, holding times, stationarity, long-term behaviour, jump chain, explosion; birth, death, immigration, emigration processes. Differential and difference equations and pgfs. Finding pgfs. Embedded processes. Time to extinction.
- Brownian motion and its properties.

All parts of the lecture notes and problem sheets are potentially examinable.

1.1.2 Recent updates

The lecture notes were originally based on a set of notes by Prof. Ajay Jasra, who taught the course in 2010 and 2011, but they have been regularly updated subsequently.

The lecture notes have been revised substantially for the academic year 2020-2021, when more examples, case studies and implementations of examples in R have been included[1]. In particular, most of the material which had typically been covered in the live lectures (but was not necessarily included in the typed notes) is now included in the lecture notes to make the notes more self-contained and more suitable for mixed-mode delivery.

In the academic year 2020-2021, further corrections have been made and a chapter on Brownian motion has been added to reflect the new syllabus.

For the academic year 2022-2023, the material on discrete-time Markov chains has been shortened to reflect the fact that both Y3 and Y4 students had a first introduction to this topic in the Y2 course *Probability for Statistics* and the material on Brownian motion has been further extended.

1.2 Complementary reading

The lecture notes are self-contained. They are mainly based on the following textbooks:

- Grimmett & Stirzaker (2001*b*): At times, the course relies quite heavily on this text, and reading this book, as well as doing the exercises, will help quite a lot.
- Grimmett & Stirzaker (2001*a*): This is the solutions manual for the book Grimmett & Stirzaker (2001*b*).
- Norris (1998): This books is very suitable for independent complementary reading of the theoretical aspects of Markov chains in both discrete and continuous time.
- Dobrow (2016): This book explains the underlying theory at a more intuitive level and contains plenty of examples and R code. Students who are more interested in applications will most likely find this book very enjoyable to read.

Other very good textbooks include Ross (2010), Pinsky & Karlin (2011), Durrett (2016).

For very motivated students, the references Billingsley (2012), Kallenberg (2002), Shiryaev (1996), Williams (1991) might be of interest, but note that they are very advanced.

[1] However, note that programming skills in R are not required for any assessments.

Chapter 2

Preliminaries [Mainly reading material]

There are number of concepts that need to be known. You should be familiar with the material in this section from your first year probability and statistics course. If you have forgotten some of the material, however, then it is a good idea to start revising the material now before we start with the new material!

2.1 Cardinality of sets

Definition 2.1.1 (Finite, countably infinite and uncountably infinite sets)**.**

- *A set A is said to be **finite**, if it has a finite number of elements.*
- *A set A is called **countably infinite** if there is a bijection between the elements of A and the natural numbers $\mathbb{N} = \{1, 2, \dots\}$.*
- *A set A is called **countable** if it is either finite or countably infinite.*
- *If the set A is neither finite nor countably infinite, we call it **uncountable** or **uncountably infinite**.*

2.2 The probability space

Recall that probability theory starts with a **probability space** which consists of a triple $(\Omega, \mathcal{F}, \mathrm{P})$, where

- the set Ω is the sample-space (or state-space), which is the set of all possible outcomes. E. g. when we role a die once, the corresponding sample space is given by $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- $\mathcal{F}$ is a collection of events/sets we can make probability statements about (we will make this precise below).
- P is a probability measure which 'measures' the probability of each set $A \in \mathcal{F}$.

Let Ω be a set.

Definition 2.2.1. *A collection $\mathcal{F}^*$ of subsets of Ω is called **an algebra** on Ω if it satisfies the following three conditions:*

1. $\emptyset \in \mathcal{F}^*$;
2. *if $A \in \mathcal{F}^*$, then $A^c \in \mathcal{F}^*$;*
3. *if $A_1, A_2 \in \mathcal{F}^*$, then $A_1 \cup A_2 \in \mathcal{F}^*$.*

Note that the above definition implies that $\Omega = \emptyset^c \in \mathcal{F}^*$, and that if $A_1, A_2 \in \mathcal{F}^*$, then $A_1 \cap A_2 = (A_1^c \cup A_2^c)^c \in \mathcal{F}^*$. So we have seen that an algebra is *stable under finitely many set operations*.

Definition 2.2.2. *A collection $\mathcal{F}$ of subsets of Ω is called a $\sigma-$**algebra on** Ω if $\mathcal{F}$ is an algebra such that if $A_1, A_2, \ldots, \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

Note that if $\mathcal{F}$ is a σ-algebra and $A_1, A_2, \ldots, \in \mathcal{F}$, then $\cap_{i=1}^{\infty} A_i = (\cup_{i=1}^{\infty} A_i^c)^c \in \mathcal{F}$. Hence a σ-algebra is *stable under countably infinite set operations*.

Definition 2.2.3. *Let Ω denote a set and let $\mathcal{F}$ be a σ-algebra on Ω. We call the pair $(\Omega, \mathcal{F})$ a **measurable space**. An element of $\mathcal{F}$ is called a $\mathcal{F}$-measurable subset of Ω.*

Definition 2.2.4. *Let $\mathcal{A}$ denote an arbitrary family of subsets of Ω. Then the **σ-algebra generated by** $\mathcal{A}$ is defined as the smallest σ-algebra which contains $\mathcal{A}$, i.e.*

$$\sigma(\mathcal{A}) := \cap\{\mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra}, \mathcal{A} \subset \mathcal{G}\}.$$

Example 2.2.5. *The smallest $\sigma-$algebra associated with Ω is the collection $\mathcal{F} = \sigma(\Omega) = \{\emptyset, \Omega\}$.*

Example 2.2.6. *Let A be any subset of Ω. Then $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$.*

Example 2.2.7. *A very important example is the σ–algebra generated by the open subsets of $\mathbb{R}$, which we call the Borel σ–algebra and which is denoted by $\mathcal{B} = \mathcal{B}(\mathbb{R})$. One can show that $\mathcal{B} = \sigma(\{(-\infty, y] : y \in \mathbb{R}\})$.*

Note that $(-\infty, y] = \cap_{n\in\mathbb{N}}(-\infty, y + \frac{1}{n})$ is a countable intersection of open sets and hence is in $\mathcal{B}$.

Definition 2.2.8 (Probability measure)**.** *A mapping $\mathrm{P} : \mathcal{F} \to \mathbb{R}$ is called a **probability measure** on $(\Omega, \mathcal{F})$ if it satisfies three conditions:*

(i) $\mathrm{P}(A) \geq 0$ *for all events* $A \in \mathcal{F}$,

(ii) $\mathrm{P}(\Omega) = 1$,

(iii) For any sequence of disjoint events $A_1, A_2, A_3, \cdots \in \mathcal{F}$ we have

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i).$$

[Note that by "disjoint events" we mean that $A_i \cap A_j = \emptyset$ for all $i \neq j$.]

Summarising, we re-iterate that throughout the course, we always assume that we are on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, where

- the set Ω is the sample space and $\omega \in \Omega$ is called a sample point;
- $\mathcal{F}$ is a σ-algebra on Ω which describes the family of events. An event is defined as an element of $\mathcal{F}$, i.e. an event is a $\mathcal{F}$-measurable subset of Ω;
- P is a probability measure on the measurable space $(\Omega, \mathcal{F})$.

2.3 Some rules of probability

As a reminder, and in a non-rigorous way, we have the following identities. All notation is as above and we refer to any sets as above. Note that we use these ideas, without further reference, later in the notes.

- Complement: $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$ for all $A \in \mathcal{F}$,
- Addition law: $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$, for all $A, B \in \mathcal{F}$,
- Conditional probability: For $B \in \mathcal{F}$ with $\mathrm{P}(B) > 0$, define

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

In this course, we will deal with conditional probabilities a lot! When we write down a conditional probability, we will always (implicitly) assume that, the event on which we condition, does not have zero probability!

Theorem 2.3.1 (Multiplication rule). *Let $n \in \mathbb{N}$, then for any events $A_1, \ldots, A_n$ with $\mathrm{P}(A_2 \cap \cdots \cap A_n) > 0$, we have*

$$\mathrm{P}(A_1 \cap \cdots \cap A_n) = \mathrm{P}(A_1|A_2 \cap \cdots \cap A_n)\mathrm{P}(A_2|A_3 \cap \cdots \cap A_n) \cdots \\ \mathrm{P}(A_{n-2}|A_{n-1} \cap A_n)\mathrm{P}(A_{n-1}|A_n)\mathrm{P}(A_n),$$

where the right hand side is a product of n terms.

Note that the ordering of the events in the theorem above can be changed.

2.3.1 The law of total probability

Definition 2.3.2 (Partition). *A **partition** of the sample space Ω is a collection $\{B_i : i \in \mathcal{I}\}$ (for a countable index set $\mathcal{I}$) of disjoint events (meaning that $B_i \in \mathcal{F}$ and $B_i \cap B_j = \emptyset$ for $i \neq j$) such that $\Omega = \bigcup_{i \in \mathcal{I}} B_i$.*

Remark 2.3.3. *We note that a partition of the sample space is often not unique and the choice of the particular partition typically very much depends on the problem we want to solve!*

Theorem 2.3.4 (Law of total probability). *Let $\{B_i : i \in \mathcal{I}\}$ denote a partition of Ω, with $\mathrm{P}(B_i) > 0$ for all $i \in \mathcal{I}$. Then, for all $A \in \mathcal{F}$,*

$$\mathrm{P}(A) = \sum_{i \in \mathcal{I}} \mathrm{P}(A \cap B_i) = \sum_{i \in \mathcal{I}} \mathrm{P}(A|B_i)\mathrm{P}(B_i).$$

Example 2.3.5. *Given two events $A, B \in \mathcal{F}$ such that $0 < \mathrm{P}(B) < 1$, we have*

$$\mathrm{P}(A) = \mathrm{P}(A|B)\mathrm{P}(B) + \mathrm{P}(A|B^c)\mathrm{P}(B^c).$$

Theorem 2.3.6 (Law of total probability with additional conditioning). *Consider events A, E with $\mathrm{P}(E) > 0$ and let $\{B_i : i \in \mathcal{I}\}$ denote a partition of Ω, with $\mathrm{P}(B_i \cap E) > 0$ for all $i \in \mathcal{I}$. Then,*

$$\mathrm{P}(A|E) = \sum_{i \in \mathcal{I}} \frac{\mathrm{P}(A \cap B_i \cap E)}{\mathrm{P}(E)} = \sum_{i \in \mathcal{I}} \mathrm{P}(A|B_i \cap E)\mathrm{P}(B_i|E).$$

2.3.2 Independence

Definition 2.3.7. *Events A and B are **independent** if*

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

*More generally, a family $\{A_i : i \in \mathcal{I}\}$ is called **independent** if*

$$\mathrm{P}\left(\cap_{i \in \mathcal{J}} A_i\right) = \prod_{i \in \mathcal{J}} \mathrm{P}(A_i),$$

for all finite subsets $\mathcal{J}$ of $\mathcal{I}$.

2.4 Random variables

Let us recall further important definitions.

2.4.1 Pre-images

Definition 2.4.1. *Consider a function with domain $\mathcal{X}$ and co-domain $\mathcal{Y}$, i.e. $f : \mathcal{X} \to \mathcal{Y}$.*

- *For any subset $A \subseteq \mathcal{X}$, we define the **image of** A* under *f as*
$$f(A) = \{y \in \mathcal{Y} : \exists x \in A : f(x) = y\} = \{f(x) : x \in A\}.$$
*If $A = \mathcal{X}$, then we call $f(\mathcal{X}) = \mathrm{Im} f$ the **image of** f.*
- *For any subset $B \subseteq \mathcal{Y}$, we define the **pre-image of** B **under** f as*
$$f^{-1}(B) = \{x \in \mathcal{X} : f(x) \in B\}.$$

Please not that the pre-image should not be confused with the inverse function (despite the fact that we are using the same notation). The pre-image is well-defined for any function, whereas the inverse function obviously only exists when the function f is invertible.

The definition of the pre-image implies that
$$x \in f^{-1}(B) \Leftrightarrow f(x) \in B.$$

Note that in the case when B is a singleton, i.e. $B = \{b\}$ for an element $b \in \mathcal{Y}$, then we often simplify the notation to $f^{-1}(\{b\}) = f^{-1}(b)$.

2.4.2 Random variables

Definition 2.4.2 (Random variable)**.** *A **random variable** on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is defined as the mapping $X : \Omega \to \mathbb{R}$ which satisfies*
$$X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \qquad \text{for all } x \in \mathbb{R}. \tag{2.4.1}$$

Definition 2.4.3 (Cumulative distribution function (c.d.f.))**.** *Suppose that X is a random variable on $(\Omega, \mathcal{F}, \mathrm{P})$, then the **cumulative distribution function** (c.d.f.) of X is defined as the mapping $F_X : \mathbb{R} \to [0, 1]$ given by*
$$F_X(x) = \mathrm{P}(\{\omega \in \Omega : X(\omega) \leq x\}) = \mathrm{P}(X^{-1}((-\infty, x])),$$
which is typically abbreviated to $F_X(x) = \mathrm{P}(X \leq x)$.

Definition 2.4.4 (Discrete random variable)**.** *A **discrete random variable** on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is defined as a mapping $X : \Omega \to \mathbb{R}$ such that*

*(i) the **image/range of** Ω **under** X denoted by $\mathrm{Im}X = \{X(\omega) : \omega \in \Omega\}$ is a countable subset of $\mathbb{R}$,*

(ii) $X^{-1}(x) = \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ for all $x \in \mathbb{R}$.

Definition 2.4.5 (Probability mass function)**.** *The **probability mass function** (pmf) of the discrete random variable X is defined as the function $p_X : \mathbb{R} \to [0, 1]$ given by*
$$p_X(x) = \mathrm{P}(\{\omega \in \Omega : X(\omega) = x\}) = \mathrm{P}(X^{-1}(x)). \tag{2.4.2}$$

We typically shorten the notation significantly and write $p_X(x) = \mathrm{P}(X = x)$.

Definition 2.4.6 (Continuous random variable and probability density function)**.** *A random variable X is called **continuous** if its c.d.f. can be written as*
$$F_X(x) = \mathrm{P}(X \leq x) = \int_{-\infty}^{x} f_X(u)du, \qquad \text{for all } x \in \mathbb{R}, \tag{2.4.3}$$
where the function $f_X : \mathbb{R} \to \mathbb{R}$ satisfies

(i) $f_X(u) \geq 0$ for all $u \in \mathbb{R}$,

(ii) $\int_{-\infty}^{\infty} f_X(u)du = 1$.

*We call f_X the **probability density function** (p.d.f.) of X (or just the **density**).*[1]

[1] In a later analysis/measure course we will say that equation 2.4.3 means that the "c.d.f. of a continuous random variable is **absolutely continuous** with respect to the Lebesgue measure".

2.4.3 Independence of random variables

Recall that in Definition 2.3.7 we defined independent events. Next, we want to define the concept of independence of random variables.

Discrete random variables X and Y are **independent** if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all x and y. More generally:

Definition 2.4.7. *Let $\mathcal{I} \subset \mathbb{R}$ denote an index set. A family $\{X_i : i \in \mathcal{I}\}$ of discrete random variables is **independent** if for all finite subsets $\mathcal{J} \subseteq \mathcal{I}$ and all $x_j \in \mathbb{R}, j \in \mathcal{J}$, the following product rule holds:*

$$\mathrm{P}\left(X_i = x_i, \text{ for all } i \in \mathcal{J}\right) = \prod_{i \in \mathcal{J}} \mathrm{P}(X_i = x_i).$$

Recall that we cannot define the independence of *continuous* random variables X and Y in terms of events such as $\{X = x\}$ and $\{Y = y\}$, since these events have zero probabilities and are hence trivially independent.

We now state a definition of independence which is valid for any pair of random variables, regardless of their types (discrete, continuous, etc.).

Random variables X and Y are called **independent** if the events $\{X \leq x\}$ and $\{Y \leq y\}$ are independent for all $x, y \in \mathbb{R}$. More generally, we have:

Definition 2.4.8 (Independence of a family of random variables)**.** *Let $\mathcal{I} \subset \mathbb{R}$ denote an index set. A family of random variables $\{X_i : i \in \mathcal{I}\}$ is said to be **independent** if for all finite subsets $\mathcal{J} \subseteq \mathcal{I}$ and all $x_j \in \mathbb{R}, j \in \mathcal{J}$, the following product rule holds:*

$$\mathrm{P}\left(\cap_{j \in \mathcal{J}} \{X_j \leq x_j\}\right) = \prod_{j \in \mathcal{J}} \mathrm{P}(X_j \leq x_j).$$

2.5 Continuity of the probability measure

Definition 2.5.1 (Increasing and decreasing sets)**.** *A sequence of sets $(A_i)_{i=1}^{\infty}$ is said to* increase *to A, i.e. $A_i \uparrow A$, if $A_1 \subseteq A_2 \subseteq \cdots$ and $\cup_{i=1}^{\infty} A_i = A$. Similarly, a sequence of sets $(A_i)_{i=1}^{\infty}$ is said to* decrease *to A, i.e. $A_i \downarrow A$, if $A_1 \supseteq A_2 \supseteq \cdots$ and $\cap_{i=1}^{\infty} A_i = A$.*

Note that $A_i \uparrow A$ if and only if $A_i^c \downarrow A^c$.

Next we will state and prove the continuity property of the probability measure[2].

Theorem 2.5.2. *If $A_1, A_2, \cdots \in \mathcal{F}$ and $A_i \uparrow A$ or $A_i \downarrow A$, then*

$$\lim_{i \to \infty} \mathrm{P}(A_i) = \mathrm{P}(A).$$

The above theorem states that, for increasing or decreasing sets, we can interchange the limit operation and the probability measure, i.e. we have

$$\lim_{i \to \infty} \mathrm{P}(A_i) = \mathrm{P}(\lim_{i \to \infty} A_i),$$

where the set limit on the right hand side needs to be understood as taking an infinite union or intersection for increasing and decreasing sequences, respectively.

2.6 Expectation of random variables

2.6.1 Definition of the expectation

Next we define the expectation of a discrete random variable.

[2] Recall that a sequence of real numbers (x_n) is said to converge to a real number x if for all $\epsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $|x_n - x| < \epsilon$.

Definition 2.6.1 (Expectation of discrete random variable)**.** *Let X denote a discrete random variable, then the* expectation *of X is defined as*

$$\mathrm{E}(X) = \sum_x x\mathrm{P}(X = x)$$

whenever the sum on the right hand side converges absolutely, i.e. when we have $\sum_x |x|\mathrm{P}(X = x) < \infty$.[3]

Definition 2.6.2 (Expectation of a continuous random variable)**.** *For a continuous random variable X with density f_X, we define the* expectation *of X as*

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f_X(x)dx,$$

provided that $\int_{-\infty}^{\infty} |x| f_X(x)dx < \infty$.

2.6.2 Law of the unconscious statistician (LOTUS)

Theorem 2.6.3 (LOTUS: Discrete case)**.** *Let X be a discrete random variable and $g : \mathbb{R} \to \mathbb{R}$, then*

$$\mathrm{E}(g(X)) = \sum_{x \in \mathrm{Im}X} g(x)\mathrm{P}(X = x),$$

whenever the sum on the right hand side converges absolutely.

Theorem 2.6.4 (LOTUS: Continuous case)**.** *Let X be a continuous random variable with density f_X, consider a function $g : \mathbb{R} \to \mathbb{R}$, then*

$$\mathrm{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)dx,$$

provided that $\int_{-\infty}^{\infty} |g(x)| f_X(x)dx < \infty$.

2.7 Conditional expectation and the law of total expectation

2.7.1 The discrete case

Definition 2.7.1 (Conditional distribution and conditional expectation)**.** *Let X denote a discrete random variable on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Consider an event $B \in \mathcal{F}$ such that $\mathrm{P}(B) > 0$. The* conditional distribution of X given B *is defined as*

$$\mathrm{P}(X = x|B) = \frac{\mathrm{P}(\{X = x\} \cap B)}{\mathrm{P}(B)}, \text{ for } x \in \mathbb{R}.$$

Further, the conditional expectation of X given B *is defined as*

$$\mathrm{E}(X|B) = \sum_{x \in \mathrm{Im}X} x\mathrm{P}(X = x|B),$$

provided the sum is absolutely convergent.

Similarly to the ideas presented in the law of total probability, it can often be useful to consider a partition of the probability space to compute an (unconditional) expectation via conditional expectations as we describe in the following theorem.

Theorem 2.7.2. *[Law of total expectation] Consider a partition $\{B_i : i \in \mathcal{I}\}$ of Ω with $\mathrm{P}(B_i) > 0$ for all $i \in \mathcal{I}$. Let X denote a discrete random variable with finite expectation. Then*

$$\mathrm{E}(X) = \sum_{i \in \mathcal{I}} \mathrm{E}(X|B_i)\mathrm{P}(B_i),$$

whenever the sum converges absolutely.

[3] This assumption matters in the case when $\mathrm{Im}X$ is infinite. If the sum converges absolutely, then the sum takes the same value irrespectively of the order of summation.

Conditioning on a random variable

Suppose (X, Y) are jointly discrete random variables. In the above definition, consider the event $B = \{X = x\}$ for some $x \in \mathbb{R}$ such that $p_X(x) = \mathrm{P}(X = x) > 0$. Then the *conditional distribution/probability mass function of* Y *given* $X = x$ is given by

$$p_{Y|X}(y|x) = \mathrm{P}(Y = y|X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}, \quad \text{for } y \in \mathbb{R}.$$

Also, the *conditional expectation of* Y *given* $X = x$ is given by

$$\psi(x) = \mathrm{E}(Y|X = x) = \sum_y y p_{Y|X}(y|x),$$

provided the sum is absolutely convergent.

We call the random variable $\psi(X) = \mathrm{E}(Y|X)$ the *conditional expectation of* Y *given* X.

Also, the law of the unconscious statistician (LOTUS) for conditional expectations says that

$$\mathrm{E}(g(Y)|X = x) = \sum_y g(y) p_{Y|X}(y|x).$$

Note that we can also formulate an independence condition in terms of conditional p.m.f.s: Discrete X and Y are independent if and only if

$$\mathrm{P}(Y = y|X = x) = \mathrm{P}(Y = y)$$

for all x, y such that $\mathrm{P}(X = x) > 0$.

2.7.2 Continuous case

Let X, Y denote jointly continuous random variables with joint density function denoted by $f_{(X,Y)}$ and marginal densities denoted by f_X and f_Y, respectively. Then the *conditional density function of* Y *given* X is defined as

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

for any x, such that $f_X(x) > 0$. Also, the *conditional distribution function of* Y *given* $X = x$ is defined as

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

for any x, such that $f_X(x) > 0$.

Note that we can also formulate an independence condition in terms of conditional p.d.f.s: Jointly continuous random variables X and Y are independent if and only if

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

for all x, y such that $f_X(x) > 0$.

Note that, by Tonelli's theorem, see Theorem 2.8.1, we obtain the continuous version of the law of total probability:

$$\begin{aligned}
\mathrm{P}(Y > y) &= \int_y^\infty f_Y(v)dv = \int_y^\infty \int_{-\infty}^\infty f_{(X,Y)}(x, v)dxdv \\
&= \int_{-\infty}^\infty \int_y^\infty f_{(X,Y)}(x, v)dvdx \\
&= \int_{-\infty}^\infty \int_y^\infty f_{Y|X}(v|x) f_X(x)dvdx
\end{aligned}$$

$$= \int_{-\infty}^{\infty} \int_{y}^{\infty} f_{Y|X}(v|x)dv f_X(x)dx$$
$$= \int_{-\infty}^{\infty} \mathrm{P}(Y > y|X = x) f_X(x)dx.$$

Definition 2.7.3 (Conditional expectation)**.** *For two jointly continuous random variables X, Y, we define the* conditional expectation of Y given $X = x$ *as*

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

provided that $f_X(x) > 0$.

The continuous analogue of the law of total expectation reads as follows:

Theorem 2.7.4 (Law of total expectation)**.** *For jointly continuous random variable X, Y with $\mathrm{E}|Y| < \infty$, we have*

$$\mathrm{E}(Y) = \int_{\{x: f_X(x) > 0\}} \mathrm{E}(Y|X = x) f_X(x)dx.$$

Another variant of the law of total expectation/continuous law of total probability, which we will be using frequently, is given as follows.

Suppose that A is an event and $\mathbb{I}_A$ is the indicator variables associated with the event A. Let X be a continuous random variable with probability density function f_X. Then we have

$$\mathrm{E}(\mathbb{I}_A) = \mathrm{E}(\mathrm{E}(\mathbb{I}_A|X)) = \int \mathrm{E}(\mathbb{I}_A|X = x) f_X(x)dx. \tag{2.7.1}$$

Recalling that $\mathrm{E}(\mathbb{I}_A) = \mathrm{P}(A)$, we get

$$\mathrm{P}(A) = \int \mathrm{P}(A|X = x) f_X(x)dx, \tag{2.7.2}$$

which we often also refer to as a continuous version of the law of total probability.

2.8 Interchanging sums, integrals, limits

Throughout the course, we will often need to interchange limits/sums/integrals/expectations. We will now state important theorems which will justify such interchanges under suitable conditions.

I will not assume that you have seen these theorems before. We will now state these theorems (without proofs), and we will apply them throughout the course.

Theorem 2.8.1 (Tonelli's theorem)**.** *The order of integration, countable summation and (conditional) expectation can be interchanged whenever the integrand/summands/random variables are non-negative.*

Proof. See (Kallenberg 2002, Theorem 1.27). □

There is a related result, called Fubini's theorem, which applies in the setting when integrands/summands/random variables are not necessarily non-negative. However, an additional integrability condition is needed for this more general theorem to hold.

We will now state the famous dominated and monotone convergence theorems. We note that they can be expressed more concisely using measure theory. However, since knowledge of measure theory is not assumed for this course, we will just state the versions of these theorems which will be needed during the course, even if this appears rather repetitive!

We first state the dominated convergence theorem for real (deterministic) sequences, there will be no randomness! See (Kallenberg 2002, Theorem 1.21).

Theorem 2.8.2 (Dominated Convergence Theorem). *Let $\mathcal{I}$ denote a countable index set. If $\sum_{i\in\mathcal{I}} a_i(n)$ is an absolutely convergent series for all $n \in \mathbb{N}$ such that*

1. *for all $i \in \mathcal{I}$ the limit $\lim_{n\to\infty} a_i(n) = a_i$ exists,*
2. *there exists a sequence $(b_i)_{i\in\mathcal{I}}$, such that $b_i \geq 0$ for all $i \in \mathcal{I}$ and $\sum_{i\in\mathcal{I}} b_i < \infty$ such that for all $n, i : |a_i(n)| \leq b_i$.*

Then $\sum_{i\in\mathcal{I}} |a_i| < \infty$ and

$$\sum_{i\in\mathcal{I}} a_i = \sum_{i\in\mathcal{I}} \lim_{n\to\infty} a_i(n) = \lim_{n\to\infty} \sum_{i\in\mathcal{I}} a_i(n).$$

Next, we study theorems which apply to sequences of random variables and will enable us to interchange limits and expectations.

Theorem 2.8.3. *Suppose $\{Z_n\}$ is a sequence of random variables with $\mathrm{P}(\lim_{n\to\infty} Z_n = Z) = 1$, then*

1. *Monotone convergence (MON): If $\mathrm{P}(Z_n \geq 0) = 1$ and $\mathrm{P}(Z_n \leq Z_{n+1}) = 1$ for all n, then*
$$\lim_{n\to\infty} \mathrm{E}(Z_n) = \mathrm{E}(\lim_{n\to\infty} Z_n) = \mathrm{E}(Z).$$
2. *Monotone convergence (MON2): If $\mathrm{P}(Z_n \geq 0) = 1$ and $\mathrm{P}(Z_n \geq Z_{n+1}) = 1$ for all n, $\mathrm{E}(Z_1) < \infty$, then*
$$\lim_{n\to\infty} \mathrm{E}(Z_n) = \mathrm{E}(\lim_{n\to\infty} Z_n) = \mathrm{E}(Z).$$
3. *Dominated convergence (DOM): If $\mathrm{P}(|Z_n| \leq Y) = 1$ for all n, and $\mathrm{E}|Y| < \infty$, then*
$$\lim_{n\to\infty} \mathrm{E}(Z_n) = \mathrm{E}(\lim_{n\to\infty} Z_n) = \mathrm{E}(Z).$$
4. *Bounded convergence (BC) (special case of DOM): If $\mathrm{P}(|Z_n| \leq c) = 1$ for some constant c for all n, then*
$$\lim_{n\to\infty} \mathrm{E}(Z_n) = \mathrm{E}(\lim_{n\to\infty} Z_n) = \mathrm{E}(Z).$$

See Grimmett & Stirzaker (2001*b*, p.179–180) and Kallenberg (2002, Theorem 1.19, Theorem 1.21, Corollary 17.13, Theorem 6.1) for details and note that these convergence results can be extended to conditional expectations as well!

An important consequence of the monotone convergence theorem is that, for non-negative random variables $\{Z_n\}$ with finite expectation,

$$\mathrm{E}\left(\sum_{i=1}^{\infty} Z_i\right) = \mathrm{E}\left(\lim_{N\to\infty} \sum_{i=1}^{N} Z_i\right) \overset{\text{MON}}{=} \lim_{N\to\infty} \mathrm{E}\left(\sum_{i=1}^{N} Z_i\right) = \lim_{N\to\infty} \sum_{i=1}^{N} \mathrm{E}\left(Z_i\right) = \sum_{i=1}^{\infty} \mathrm{E}\left(Z_i\right),$$

whether or not the summation if finite.

2.9 Stochastic processes and their realisations/sample paths

In the first year probability courses we have only focused on random variables, now we extend this notion to stochastic/random processes.

Definition 2.9.1. *A stochastic process $X = (X_t)_{t\in\mathcal{T}}$ on $(\Omega, \mathcal{F}, \mathrm{P})$ is a collection of random variables. I.e.*

$$X : \Omega \times \mathcal{T} \to E, \qquad (\omega, t) \mapsto X_t(\omega),$$

where $\mathcal{T}$ is some time domain (e.g. $\mathcal{T} = \mathbb{N} \cup \{0\}, \mathcal{T} = [0, T]$ or $\mathcal{T} = [0, \infty)$) and E denotes the common state space of the random variables.

Note that a stochastic process is a function in two variables: ω and t, but we typically just write $X = (X_t)_{t\in\mathcal{T}}$.

Note that we will be using the notation (X_t), $(X_t)_{t\in\mathcal{T}}$, $\{X_t\}$ and $\{X_t\}_{t\in\mathcal{T}}$ interchangeably for stochastic processes.

Recall that the randomness in the random variable comes from ω, more precisely, from the fact that we do not know which ω will appear in an experiment. For a random variable Y, say, as soon as ω is known, then $Y(\omega)$ is known and typically called the **realisation** of the random variable Y.

In the context of stochastic processes, for any $\omega \in \Omega$, we call the function

$$X(\omega) : \mathcal{T} \to E, \qquad t \mapsto X_t(\omega)$$

a **realisation/sample path/trajectory** of $X = (X_t)_{t\in\mathcal{T}}$. I.e. when we talk about a sample path/realisation, then the ω is fixed and we consider only a function in t and no longer a function in both t and ω.

End of lecture 1.

Chapter 3

Discrete-time Markov chains

Markov processes are one of the most important (if not the most important) of all stochastic processes. Examples include Poisson processes, Brownian motion, diffusion processes etc. The underlying idea is to encapsulate the dependence structure of the process using a simple property.

Andrei Markov (1856–1922) was a Russian mathematician, who is well–known for his influential work on stochastic processes.

Figure 3.1: Andrei Markov

Informally, a Markov process has the property that **conditional on the present value, the future is independent of the past**.

We will, of course, formalise this concept as the chapter progresses. Markov processes, are often the term for a continuous time process, on a general (continuous) state-space. We will consider one such a process towards the end of this course, but this chapter is focussed upon Markov chains, which, for our purposes are Markov processes in discrete (or continuous time) on a discrete state–space. We begin with discrete-time, discrete space Markov chains, and we will investigate a variety of properties, including stationarity. In the next chapter, we then turn to continuous-time, discrete state, Markov chains.

3.1 Definition of discrete-time Markov chains

We begin by giving the basic set-up:

- Consider a probability space $(\Omega, \mathcal{F}, \mathrm{P})$.
- Let $X_0, X_1, \ldots$ be a sequence of discrete variables, each variable taking some value in a state space E.

- We always assume that E is **countable**. Often we assume that $E \subseteq \mathbb{Z}$, but E can be any finite or countably infinite set.
- Let $K = \text{card}(E)$ (i.e. the number of elements in the set E). If the state space E is **finite**, then $K < \infty$, if it is **countably infinite**, then $K = \infty$.
- We write $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.
- A collection of random variables denoted by $X = \{X_n\}_{n \in \mathbb{N}_0}$ is called a discrete-time stochastic process.

We already mentioned that, informally, a Markov process has the property that **conditional on the present value, the future is independent of the past**. We will now describe this property mathematically.

Definition 3.1.1. *A discrete-time stochastic process $X = \{X_n\}_{n \in \mathbb{N}_0}$ taking values in a countable state space E is a **Markov chain** if it satisfies the **Markov condition***

$$\mathrm{P}(X_n = j | X_{n-1} = i, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \mathrm{P}(X_n = j | X_{n-1} = i),$$

for all $n \in \mathbb{N}$ and for all $x_0, \dots, x_{n-2}, i, j \in E$.

Here we see that the dependence, of X_n, *conditional* upon the sequence $\mathbf{X}_{0:n-1} = (X_0, \dots, X_{n-1})$, is only on X_{n-1}.

An important point to note is that X_n is *not* in general independent of (say) X_{n-2}.

In the following, we will always work with time-homogeneous transition probabilities (unless stated otherwise):

Definition 3.1.2. 1. *The Markov chain $\{X_n\}_{n \in \mathbb{N}_0}$ is **time-homogeneous** if*

$$\mathrm{P}(X_{n+1} = j | X_n = i) = \mathrm{P}(X_1 = j | X_0 = i)$$

for every $n \in \mathbb{N}_0$, $i, j \in E$.

2. *The **transition matrix** $\mathbf{P} = (p_{ij})_{i,j \in E}$ is the $K \times K$ matrix of transition probabilities*

$$p_{ij} = \mathrm{P}(X_{n+1} = j | X_n = i).$$

Note that, if $X_n = i$, we typically say that the Markov chain *visits* state i, or *hits* i, at time n.

Definition 3.1.3 (Stochastic matrix)**.** *A square matrix $\mathbf{P}$ is called a **stochastic matrix** if*

1. *$\mathbf{P}$ has non-negative entries, $p_{ij} \geq 0$ for all i, j.*
2. *$\mathbf{P}$ has row sums equal to 1; $\sum_j p_{ij} = 1$ for all i.*

Theorem 3.1.4. *The transition matrix $\mathbf{P}$ is a stochastic matrix.*

Proof. 1. p_{ij} is a conditional probability and hence it is trivially non-negative.

2. Note that for each $i \in E$, we have

$$\sum_{j \in E} p_{ij} = \sum_{j \in E} \mathrm{P}(X_1 = j | X_0 = i) = \sum_{j \in E} \frac{\mathrm{P}(X_1 = j, X_0 = i)}{\mathrm{P}(X_0 = i)} = \sum_{j \in E} \frac{\mathrm{P}(X_0 = i | X_1 = j)\mathrm{P}(X_1 = j)}{\mathrm{P}(X_0 = i)}$$
$$= \frac{1}{\mathrm{P}(X_0 = i)} \sum_{j \in E} \mathrm{P}(X_0 = i | X_1 = j)\mathrm{P}(X_1 = j) = \frac{1}{\mathrm{P}(X_0 = i)} \mathrm{P}(X_0 = i) = 1,$$

where we used the law of total probability.

□

A useful tool in the context of Markov chains are so-called **transition diagrams**, where we draw a node for each state in E and a directed edges between the nodes i and j ($i, j \in E$) if $p_{ij} > 0$.

Example 3.1.5. *Let us consider an example of a Markov chain with three possible states* $\{1,2,3\}$ *and strictly positive transition probabilities* p_{ij} *for* $i,j \in \{1,2,3\}$. *The corresponding transition diagram is given by*

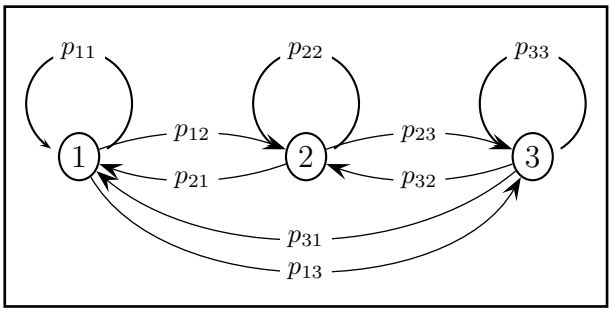


Figure 3.2: Example of a transition diagram of a Markov chain with state space $E = \{1,2,3\}$.

Example 3.1.6. *Let us assume that the weather in London can be modelled by a Markov chain* $(X_n)_{n\in\{0,1,2,\dots\}}$, *and for simplicity we assume that there are only three possible states:* $E = \{1,2,3\}$, *where the three states can be interpreted as follows: State 1: Rain; state 2: sun; state 3: fog. We assume the following:*

- *There are never two sunny days in a row. If it is a sunny day, then it is equally likely that the next day will be rainy or foggy.*
- *If it is rainy or foggy, then there is a 50% chance of having the same weather condition the next day.*
- *If the weather changes from rain or fog, only half of the time is this a change to a sunny day.*

Write down the transition matrix $\mathbf{P}$.

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}.$$

Example 3.1.7. *Suppose that* $X = \{X_n\}_{n\in\mathbb{N}_0}$ *is a sequence of independent and identically distributed discrete random variables which take values in* $E = \{1,\dots,K\}$. *We write*

$$\mathrm{P}(X_n = j) = p_j, \qquad \text{for all } j \in E, n \in \mathbb{N}_0.$$

Then $\sum_{j=1}^K p_j = 1$, *and the transition probabilities are given by*

$$p_{ij} = \mathrm{P}(X_1 = j | X_0 = i) = \mathrm{P}(X_1 = j) = p_j, \qquad \text{for all } i,j \in E,$$

by independence. Hence

$$\mathbf{P} = \begin{pmatrix} p_1 & p_2 & \cdots & p_K \\ p_1 & p_2 & \cdots & p_K \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_K \end{pmatrix}.$$

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

We have looked at the 1-step transition dynamics of a Markov chain. However, it is often of interest to consider the n-step transition dynamics, defined as follows.

Definition 3.2.1. *Let $n \in \mathbb{N}$. The n-step transition matrix $\mathbf{P}_n = (p_{ij}(n))$ is the matrix of n-step transition probabilities*

$$p_{ij}(n) = \mathrm{P}(X_{m+n} = j | X_m = i),$$

with $m \in \mathbb{N}_0$.

Note that $p_{ij}(n)$ is the probability that a process which currently is in state i will be in state j after n steps.

For a Markov chain, we know that, given the present, the future is independent of the past. We can show that *given the most recent past*, the future is independent of the past. We will now describe this property mathematically.

Lemma 3.2.2. *For a discrete Markov chain $(X_n)_{n\geq 0}$ on the state space E, we have*

$$\mathrm{P}(X_{n+m} = x_{n+m} | X_n = x_n, \ldots, X_0 = x_0) = \mathrm{P}(X_{n+m} = x_{n+m} | X_n = x_n),$$

for $m \in \mathbb{N}$ and for all $x_{n+m}, x_n, \ldots, x_0 \in E$.

Proof. See Exercise 1- 1. □

We can now formulate the Chapman-Kolmogorov equations, which can be used for computing n–step transition probabilities.

Theorem 3.2.3. *Let $m \in \mathbb{N}_0, n \in \mathbb{N}$. Then we have for any $i, j \in E$ that*

$$p_{ij}(m+n) = \sum_{l \in E} p_{il}(m) p_{lj}(n)$$

that is

$$\mathbf{P}_{m+n} = \mathbf{P}_m \mathbf{P}_n,$$

and

$$\mathbf{P}_n = \mathbf{P}^n.$$

Remark 3.2.4. *Note that in the case that $K < \infty$ matrix multiplication is well–defined. For the general case we extend the definition in the natural way: Let $\mathbf{x}$ be a K–dimensional row vector and let $\mathbf{P}$ be a $K \times K$–matrix where $K = \infty$. Then*

$$(\mathbf{xP})_j := \sum_{i \in E} x_i p_{ij}, \qquad (\mathbf{P}^2)_{ik} := \sum_{j \in E} p_{ij} p_{jk},$$

for $i, j, k \in \mathbb{N}$. Similarly, we define $\mathbf{P}^n$ for any $n \geq 0$. Also $\mathbf{P}^0$ is the identity matrix, where $(\mathbf{P}^0)_{ij} = \delta_{ij}$.

Proof of Theorem 3.2.3. First we show that $\mathbf{P}_{m+n} = \mathbf{P}_m \mathbf{P}_n$ for all $m \in \mathbb{N}_0, n \in \mathbb{N}$:

For any $i, j \in E$ and integers $m \in \mathbb{N}_0, n \in \mathbb{N}$ we have

$$\begin{aligned} p_{ij}(m+n) &= \mathrm{P}(X_{m+n} = j | X_0 = i) \\ &\overset{(*)}{=} \sum_{l \in E} \mathrm{P}(X_{m+n} = j | X_m = l, X_0 = i) \mathrm{P}(X_m = l | X_0 = i) \end{aligned}$$

$$\overset{(**)}{=} \sum_{l\in E} \mathrm{P}(X_{m+n} = j|X_m = l)\mathrm{P}(X_m = l|X_0 = i)$$
$$= \sum_{l\in E} p_{lj}(n)p_{il}(m) = \sum_{l\in E} p_{il}(m)p_{lj}(n).$$

where we applied the law of total probability with additional conditioning (Theorem 2.3.6) in (*) and the Markov property in (**).

Next we show that $\mathbf{P}_n = \mathbf{P}^n$ for all $n \in \mathbb{N}$. We prove this by induction in n. For $n = 1$ the claim is trivially true. For any $i, j \in E$ and any $n \in \mathbb{N}$ we have

$$\begin{aligned} p_{ij}(n+1) &= \mathrm{P}(X_{n+1} = j|X_0 = i) \\ &\overset{(*)}{=} \sum_{l\in E} \mathrm{P}(X_{n+1} = j|X_n = l, X_0 = i)\mathrm{P}(X_n = l|X_0 = i) \\ &\overset{(**)}{=} \sum_{l\in E} \mathrm{P}(X_{n+1} = j|X_n = l)\mathrm{P}(X_n = l|X_0 = i) \\ &= \sum_{l\in E} p_{lj}p_{il}(n) = \sum_{l\in E} p_{il}(n)p_{lj}. \end{aligned}$$

where we applied the law of total probability with additional conditioning (Theorem 2.3.6) in (*) and the Markov property in (**). Now we can apply the induction hypothesis to conclude that $\mathbf{P}_{n+1} = \mathbf{P}_n\mathbf{P} = \mathbf{P}^n\mathbf{P} = \mathbf{P}^{n+1}$, which concludes the proof.

□

Example 3.2.5. *Consider the following transition matrix of a two–state Markov chain on the state space* $E = \{0, 1\}$,

$$\mathbf{P} = \begin{pmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{pmatrix},$$

for $\alpha, \beta \in (0, 1)$. *(Check whether this is indeed a transition matrix!). Let* $\alpha = 0.7, \beta = 0.4$. *What is* $p_{00}(4)$*? Note that* $p_{00}(4)$ *is the probability, that we will be in state 0 in four steps given that we are in state 0 now. We need to compute* $\mathbf{P}^4$. *Then*

$$\mathbf{P}^2 = \begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix}\begin{pmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix},$$
$$\mathbf{P}^4 = \begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix}\begin{pmatrix} 0.61 & 0.39 \\ 0.52 & 0.48 \end{pmatrix} = \begin{pmatrix} 0.5749 & 0.4251 \\ 0.5668 & 0.4332 \end{pmatrix}.$$

Hence we have $p_{00}(4) = 0.5749$.

End of lecture 2. .....

3.3 Dynamics of a Markov chain

The equations are called the Chapman-Kolmogorov (CK) equations and are the most basic ingredient of Markov chains. They help to relate the long-term behaviour of the Markov chain, to the local transition dynamics (or transition matrix) of the chain. We can also use the CK equations to describe the marginal distribution of the chain $X = \{X_n\}_{n\in\mathbb{N}_0}$, for any time $n \in \mathbb{N}_0$.

Definition 3.3.1. *We denote the probability mass function of* X_n *for* $n \in \mathbb{N}_0$ *by*

$$\nu_i^{(n)} = \mathrm{P}(X_n = i), \qquad i \in E.$$

Let $K = \mathrm{card}(E)$. *We denote by* $\boldsymbol{\nu}^{(n)}$ *the* K*-dimensional row vector with elements* $\nu_i^{(n)}$ *for* $i \in E$*; it is called the* ***marginal distribution*** *of the chain at time* $n \in \mathbb{N}_0$.

Clearly, $\nu_i^{(n)} \geq 0$ for all $i \in E$ and $\sum_{i\in E} \nu_i^{(n)} = 1$.

Example 3.3.2. *Let $E = \{1, \ldots, K\}$. Then*

$$\boldsymbol{\nu}^{(n)} = \left(\nu_1^{(n)}, \ldots, \nu_K^{(n)}\right) = (\mathrm{P}(X_n = 1), \mathrm{P}(X_n = 2), \ldots, \mathrm{P}(X_n = K)).$$

Theorem 3.3.3. *With the notation introduced above, we have*

$$\boldsymbol{\nu}^{(m+n)} = \boldsymbol{\nu}^{(m)}\mathbf{P}_n = \boldsymbol{\nu}^{(m)}\mathbf{P}^n, \qquad \textit{for all } n \in \mathbb{N}, m \in \mathbb{N}_0.$$

Hence

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)}\mathbf{P}_n = \boldsymbol{\nu}^{(0)}\mathbf{P}^n, \qquad \textit{for all } n \in \mathbb{N}.$$

Proof. For any $n \in \mathbb{N}, m \in \mathbb{N}_0$ and for any $j \in E$, we have

$$\begin{aligned}\nu_j^{(m+n)} &= \mathrm{P}(X_{m+n} = j) = \sum_{i\in E} \mathrm{P}(X_{m+n} = j | X_m = i)\mathrm{P}(X_m = i) \\ &= \sum_{i\in E} p_{ij}(n)\nu_i^{(m)} = \sum_{i\in E} \nu_i^{(m)} p_{ij}(n).\end{aligned}$$

By the CK equations we know that $\mathbf{P}_n = \mathbf{P}^n$ hence we can conclude that $\boldsymbol{\nu}^{(m+n)} = \boldsymbol{\nu}^{(m)}\mathbf{P}_n = \boldsymbol{\nu}^{(m)}\mathbf{P}^n$. □

We have seen that the marginal distributions of the Markov chain are determined by the initial distribution $\boldsymbol{\nu}^{(0)}$ and the transition matrix $\mathbf{P}$. Actually, an even stronger result holds.

Theorem 3.3.4. *Let $X = \{X_n\}_{n\in\mathbb{N}_0}$ denote a Markov chain on a countable state space E. Then its initial distribution $\boldsymbol{\nu}^{(0)}$ and its transition matrix $\mathbf{P}$ determine all the finite dimensional distributions of the Markov chain, i.e. for all $0 \leq n_1 < n_2 < \cdots < n_{k-1} < n_k$ ($n_i \in \mathbb{N}_0, i = 1, \ldots, k$), $k \in \mathbb{N}$ and states $x_1, \ldots, x_k \in E$ we have*

$$\begin{aligned}\mathrm{P}(X_{n_1} = x_1, X_{n_2} = x_2, \ldots, X_{n_k} = x_k) &= (\boldsymbol{\nu}^{(0)}\mathbf{P}^{n_1})_{x_1}(\mathbf{P}^{n_2-n_1})_{x_1x_2} \cdots (\mathbf{P}^{n_k-n_{k-1}})_{x_{k-1}x_k} \\ &= (\boldsymbol{\nu}^{(0)}\mathbf{P}^{n_1})_{x_1} p_{x_1x_2}(n_2 - n_1) \cdots p_{x_{k-1}x_k}(n_k - n_{k-1}).\end{aligned}$$

Proof. The result follows from applying the multiplication rule for intersections of events, see Theorem 2.3.1, and the Markov property and time-homogeneity of the Markov chain and finally Theorem 3.3.3:

$$\begin{aligned}&\mathrm{P}(X_{n_k} = x_k, X_{n_{k-1}} = x_{k-1}, \ldots, X_{n_2} = x_2, X_{n_1} = x_1) \\ &= \mathrm{P}(X_{n_k} = x_k | X_{n_{k-1}} = x_{k-1}, \ldots, X_{n_2} = x_2, X_{n_1} = x_1)\times \\ &\quad \mathrm{P}(X_{n_{k-1}} = x_{k-1} | X_{n_{k-2}} = x_{k-2}, \ldots, X_{n_2} = x_2, X_{n_1} = x_1)\times \\ &\quad \cdots \mathrm{P}(X_{n_2} = x_2 | X_{n_1} = x_1)\mathrm{P}(X_{n_1} = x_1) \\ &= \mathrm{P}(X_{n_k} = x_k | X_{n_{k-1}} = x_{k-1})\mathrm{P}(X_{n_{k-1}} = x_{k-1} | X_{n_{k-2}} = x_{k-2}) \cdots \mathrm{P}(X_{n_2} = x_2 | X_{n_1} = x_1)\mathrm{P}(X_{n_1} = x_1) \\ &= (\mathbf{P}^{n_k-n_{k-1}})_{x_{k-1}x_k} \cdots (\mathbf{P}^{n_2-n_1})_{x_1x_2}(\boldsymbol{\nu}^{(0)}\mathbf{P}^{n_1})_{x_1}.\end{aligned}$$

□

The dynamics of a time-homogeneous Markov chain are determined by the initial probability mass function $\boldsymbol{\nu}^{(0)}$ and the transition matrix $\mathbf{P}$.

Remark 3.3.5. *Note that the CK equations are* necessary *for the Markov property, but they are* not sufficient. *This is related to the fact that pairwise independence of random variables is weaker than independence. You can find an example of a stochastic process, which satisfies the CK equations, but which is not a Markov chain in Grimmett & Stirzaker (2001*b*, p. 218–219).*

Example 3.3.6. *Let us revisit Example 3.1.6. Here* $E = \{1, 2, 3\}$ *where 1=rain, 2=sun, 3=fog.*

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}.$$

Suppose that $\boldsymbol{\nu}^{(0)} = (0.5, 0.5, 0)$. *We would like to find* $\boldsymbol{\nu}^{(1)}$ *and* $\boldsymbol{\nu}^{(2)}$.

Applying Theorem 3.3.3 leads to

$$\boldsymbol{\nu}^{(1)} = \boldsymbol{\nu}^{(0)}\mathbf{P} = (0.5, 0.5, 0) \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} = (0.5, 0.125, 0.375),$$

and

$$\begin{aligned} \boldsymbol{\nu}^{(2)} = \boldsymbol{\nu}^{(0)}\mathbf{P}^2 &= (0.5, 0.5, 0) \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} \\ &= (0.5, 0.5, 0) \begin{pmatrix} 0.4375 & 0.1875 & 0.375 \\ 0.375 & 0.25 & 0.375 \\ 0.375 & 0.1875 & 0.4375 \end{pmatrix} \\ &= (0.40625, 0.21875, 0.375). \end{aligned}$$

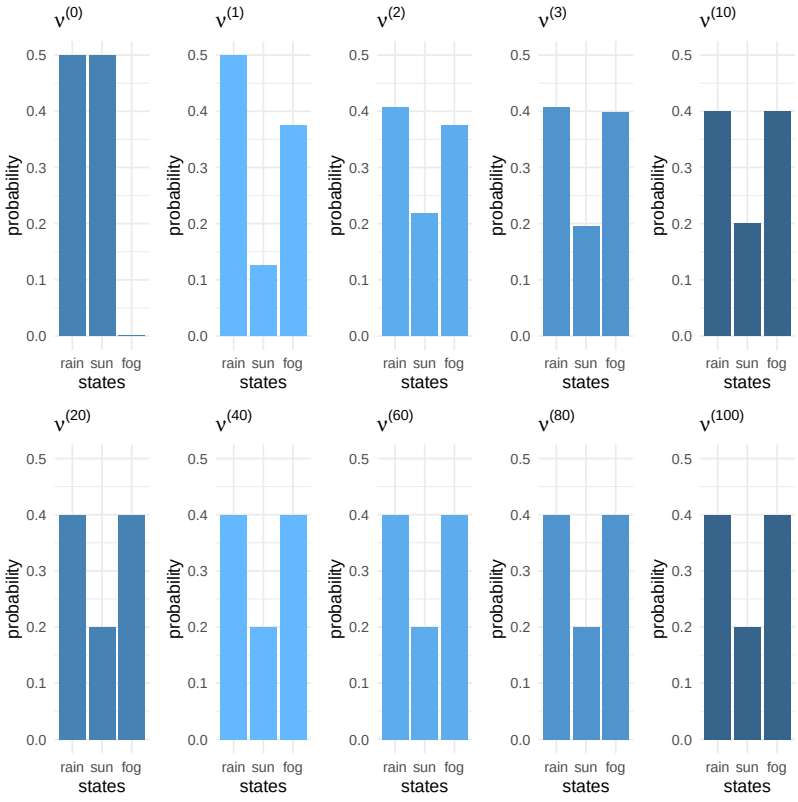


Figure 3.3: Marginal distribution $\boldsymbol{\nu}^{(n)}$ of the weather Markov chain, see Example 3.3.6, at times $n = 0, 1, 2, 3, 10$ in the first row and at times $n = 20, 40, 60, 80, 100$ in the second row.

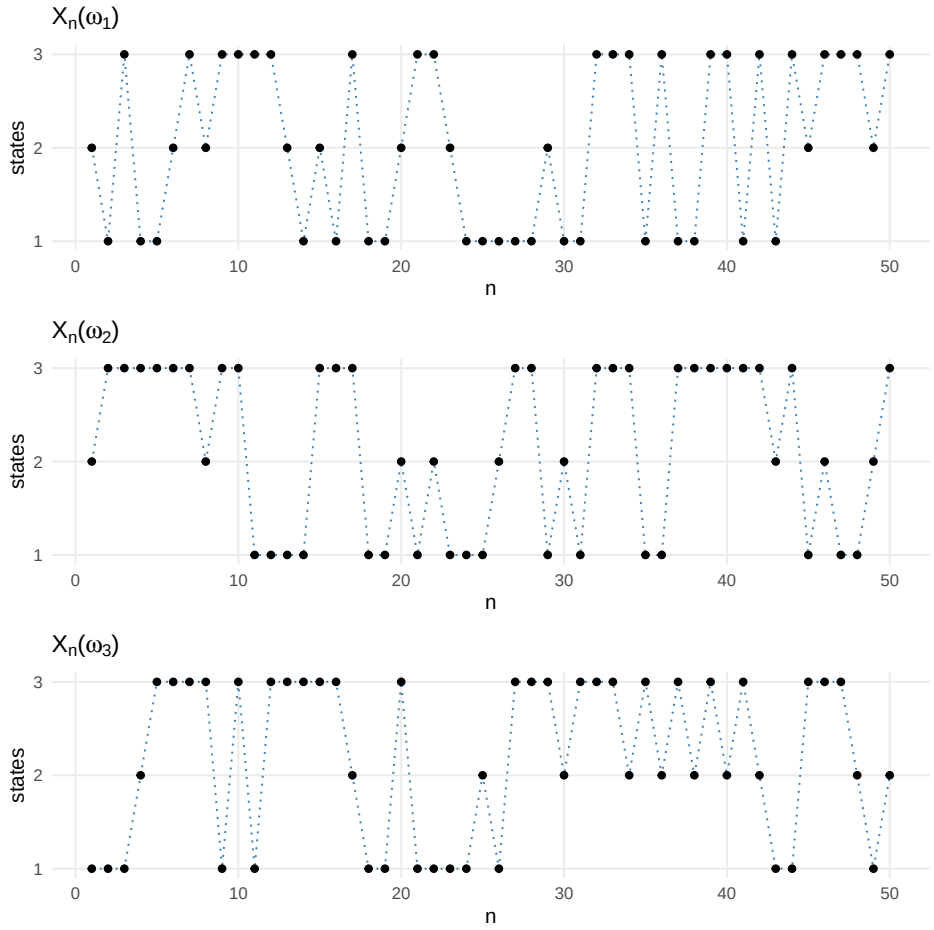


Figure 3.4: Three simulated sample paths $(X_n(\omega_i))_{n\in\{0,1,\ldots,50\}}$, for $i = 1, 2, 3$, of the weather Markov chain, see Example 3.3.6.

Example 3.3.7. *Suppose that $E = \mathbb{Z}$. A famous Markov chain is the simple random walk:*

$$p_{ij} = \begin{cases} p & \text{if } j = i+1 \\ 1-p & \text{if } j = i-1 \\ 0 & \text{o/w} \end{cases}$$

for $p \in (0, 1)$. Find the n-step transition probabilities.

Note that the simple random walk can be written as the sum

$$X_n = \sum_{i=0}^{n} Y_i,$$

where $Y_1, Y_2, \ldots$ are independent random variables taking the values -1, 1 with probabilities $(1-p)$ and p, respectively. Also $X_0 = Y_0$ denotes the initial value.

We want to find

$$p_{ij}(n) = \mathrm{P}(X_n = j | X_0 = i).$$

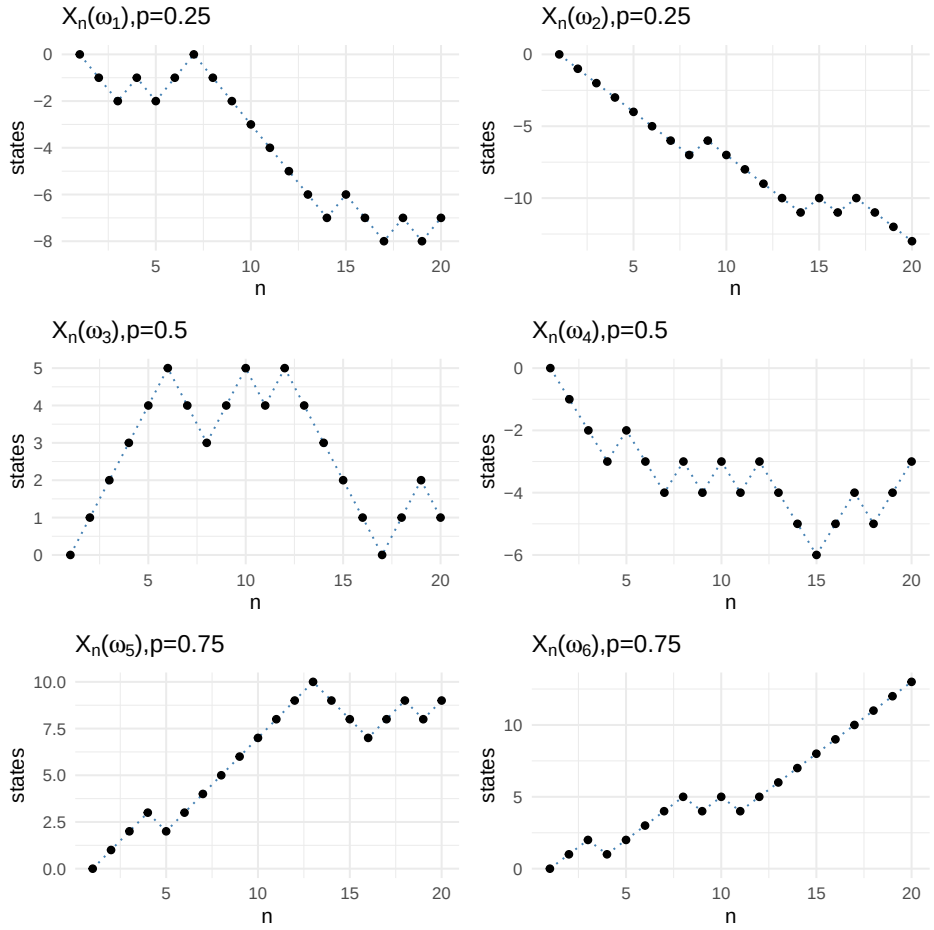


Figure 3.5: Six simulated sample paths $(X_n(\omega_i))_{n\in\{0,1,\ldots,20\}}$, for $i = 1, 2, 3, 4, 5, 6$, of a simple random walk with initial value $X_0 \equiv 0$. The parameter p is chosen as 0.25, 0.5, 0.75 in the first, second and third row, respectively, see Example 3.3.7.

In order to get from i to j in n steps, we could go up u times and down d times. Note that we require

$$n = u + d, \qquad i + u - d = j.$$

Solving for u and d we have

$$u = \frac{1}{2}(n - i + j), \qquad d = \frac{1}{2}(n - j + i) \quad \text{for } u, d \geq 0.$$

There are $\binom{n}{u}$ possibilities of going up u steps, hence, we have

$$p_{ij}(n) = \mathrm{P}(X_n = j | X_0 = i) = \binom{n}{u} p^u (1-p)^d$$
$$= \binom{n}{\frac{1}{2}(n-i+j)} p^{\frac{1}{2}(n-i+j)} (1-p)^{\frac{1}{2}(n-j+i)},$$

if $n - i + j$ is even and $p_{ij}(n) = 0$ otherwise.

End of lecture 3. .....

We now discuss a variety of properties of Markov chains which will allow us to solve a number of interesting questions associated to Markov chains.

We always denote the discrete-time, time-homogeneous Markov chain by $X = \{X_n\}_{n\in\mathbb{N}_0}$ and its countable state space by E unless stated otherwise.

3.4 First passage/hitting times

Definition 3.4.1. *We define the **first passage/hitting time** of X for state $j \in E$ as*

$$T_j = \min\{n \in \mathbb{N} : X_n = j\}.$$

If $X_n \neq j$ for all $n \in \mathbb{N}$, then we set $T_j = \infty$.

The first passage time T_j describes the first time (counting from 1 and not from 0!) that the chain ever visits the state j. Note that, for $n \in \mathbb{N}$, we have

$$\{T_j = n\} = \{X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j\},$$

which implies that T_j is a **stopping time**, since $T_j : \Omega \to \mathbb{N} \cup \{\infty\}$ and the event $\{T_j = n\}$ only depends on $X_1, X_2, \dots, X_n$, for $n \in \mathbb{N}$.

Definition 3.4.2. *For $i, j \in E, n \in \mathbb{N}$ we define the **first passage probability***

$$f_{ij}(n) = \mathrm{P}(T_j = n | X_0 = i) = \mathrm{P}(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j | X_0 = i),$$

which is the probability that the first time we visit state j, is at time n, given we started in state i at time 0.

Remark 3.4.3. *We note that we define $f_{ij}(0) = 0$ for all $i, j \in E$. Also, $f_{ij}(1) = p_{ij}$ for all $i, j \in E$.*

Definition 3.4.4. *We define*

$$f_{ij} = \mathrm{P}(T_j < \infty | X_0 = i).$$

For $i \neq j$, f_{ij} is the probability that the chain ever *visits j, starting at i; for $i = j$, f_{ii} is the probability that the chain* ever *returns to i, starting at i. We often call f_{ii} the **return probability**.*

Proposition 3.4.5. *For all $i, j \in E$, we have*

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n).$$

Proof. We use the law of total probability with additional conditioning, see Theorem 2.3.6, with partition given by $\{\{T_j = n\}, n \in \mathbb{N} \cup \{\infty\}\}$:

$$\begin{aligned} f_{ij} &= \mathrm{P}(T_j < \infty | X_0 = i) \\ &= \sum_{n=1}^{\infty} \mathrm{P}(T_j < \infty | T_j = n, X_0 = i) \mathrm{P}(T_j = n | X_0 = i) \\ &\quad + \mathrm{P}(T_j < \infty | T_j = \infty, X_0 = i) \mathrm{P}(T_j = \infty | X_0 = i) \\ &= \sum_{n=1}^{\infty} \mathrm{P}(T_j = n | X_0 = i) = \sum_{n=1}^{\infty} f_{ij}(n). \end{aligned}$$

□

Example 3.4.6. *Let $E = \{1, 2\}$ and*

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

We want to find f_{11} and f_{22}. We note that

$$f_{11}(1) = p_{11} = \frac{1}{2}, f_{11}(2) = \frac{1}{2}\frac{1}{4}, f_{11}(3) = \frac{1}{2}\frac{3}{4}\frac{1}{4}, \dots, \text{i.e.} f_{11}(n) = \frac{1}{2}\left(\frac{3}{4}\right)^{n-2}\frac{1}{4}, \text{ for } n \geq 2.$$

Hence

$$f_{11} = \sum_{n=1}^{\infty} f_{11}(n) = \frac{1}{2} + \frac{1}{8}\sum_{n=2}\left(\frac{3}{4}\right)^{n-2} = \frac{1}{2} + \frac{1}{8}\sum_{n=0}\left(\frac{3}{4}\right)^{n} = \frac{1}{2} + \frac{1}{8}\frac{1}{1-\frac{3}{4}} = 1.$$

Similarly

$$f_{22}(1) = p_{22} = \frac{3}{4}, f_{22}(2) = \frac{1}{4}\frac{1}{2}, f_{22}(3) = \frac{1}{4}\frac{1}{2}\frac{1}{2}, \ldots, \text{i.e.} f_{22}(n) = \frac{1}{4}\left(\frac{1}{2}\right)^{n-2}\frac{1}{2}, \text{ for } n \geq 2.$$

Hence

$$f_{22} = \sum_{n=1}^{\infty} f_{22}(n) = \frac{3}{4} + \frac{1}{8}\sum_{n=2}\left(\frac{1}{2}\right)^{n-2} = \frac{3}{4} + \frac{1}{8}\sum_{n=0}\left(\frac{1}{2}\right)^{n} = \frac{3}{4} + \frac{1}{8}\frac{1}{1-\frac{1}{2}} = 1.$$

Lemma 3.4.7. *Show that for all $i, j \in E, n \in \mathbb{N}$, we have*

$$p_{ij}(n) = \sum_{l=0}^{n} f_{ij}(l)p_{jj}(n-l) = \sum_{l=1}^{n} f_{ij}(l)p_{jj}(n-l).$$

Proof. We note that $f_{ij}(0) = 0$ for all $i, j \in E$, hence the second identity holds trivially. For $n = 1$, note that $p_{ij}(1) = p_{ij} = f_{ij}(1)p_{jj}(0) = f_{ij}(1)$.

For the general case, we consider the partition $\{\{T_j = l\}, l \in \mathbb{N} \cup \{\infty\}\}$. Then, using the law of total probability (LTP) with additional conditioning and the Markov property:

$$\begin{aligned}
p_{ij}(n) &= \mathrm{P}(X_n = j | X_0 = i) \\
&\overset{\text{LTP}}{=} \sum_{l=1}^{n} \mathrm{P}(X_n = j | T_j = l, X_0 = i)\mathrm{P}(T_j = l | X_0 = i) \\
&= \sum_{l=1}^{n} \mathrm{P}(X_n = j | X_l = j, X_{l-1} \neq j, \ldots, X_1 \neq j, X_0 = i)\mathrm{P}(T_j = l | X_0 = i) \\
&\overset{\text{Markov}}{=} \sum_{l=1}^{n} \mathrm{P}(X_n = j | X_l = j)\mathrm{P}(T_j = l | X_0 = i) \\
&= \sum_{l=1}^{n} p_{jj}(n-l)f_{ij}(l) \overset{f_{ij}(0)=0}{=} \sum_{l=0}^{n} f_{ij}(l)p_{jj}(n-l).
\end{aligned}$$

□

3.5 Recurrence and transience

Definition 3.5.1. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain on a countable state-space E. A state $j \in E$ is **recurrent** if*

$$\mathrm{P}(X_n = j \text{ for some } n \in \mathbb{N} | X_0 = j) = 1$$

*that is, the probability of returning to j, starting from j is 1. If the probability is less than 1, the state j is **transient**.*

I.e. a state can either be recurrent or transient. Recurrence is of interest, if we want to answer questions about eventual returns to given states of the Markov chain.

We note that

$$\{X_n = j \text{ for some } n \in \mathbb{N}\} = \{\exists n \in \mathbb{N} : X_n = j\} = \{T_j < \infty\}$$

Hence, we note that state j is recurrent if $f_{jj} = 1$ and it is transient if $f_{jj} < 1$. We can now formulate a very useful result for checking whether a given state is recurrent or transient.

Theorem 3.5.2.

1. *$j \in E$ is recurrent if and only if $\sum_{n=1}^{\infty} p_{jj}(n) = \infty$.*
2. *$j \in E$ is transient if and only if $\sum_{n=1}^{\infty} p_{jj}(n) < \infty$.*

The two statements in the above theorem are equivalent, so one only needs to prove one of them. We will provide the proof together with some related results, see Kirkwood (2015, p. 66-71). Let N_j denote the number of periods that the chain is in state j:

$$N_j = \sum_{n=0}^{\infty} I_n^{(j)}, \qquad \text{where} \qquad I_n^{(j)} = I_{\{X_n = j\}} = \begin{cases} 1, & \text{if } X_n = j, \\ 0, & \text{if } X_n \neq j, \end{cases}$$

Theorem 3.5.3. *Let $j \in E$ denote a transient state. Then*

1. *$\mathrm{P}(N_j = n | X_0 = j) = f_{jj}^{n-1}(1 - f_{jj})$ for $n \in \mathbb{N}$ (geometric distribution with parameter f_{jj}),*
2. *Let $i \neq j$, then*

$$\mathrm{P}(N_j = n | X_0 = i) = \begin{cases} 1 - f_{ij}, & \text{if } n = 0, \\ f_{ij} f_{jj}^{n-1}(1 - f_{jj}), & \text{if } n \in \mathbb{N}. \end{cases}$$

Proof. 1. In order to get $n \in \mathbb{N}$ visits to state j provided that $X_0 = j$, the chain needs to re-visit state j (after the initial visit at time 0) $n - 1$ times, which happens with probability f_{jj} each, hence f_{jj}^{n-1} is the probability of $n - 1$ return visits, and then there must be no further return to state j afterwards which happens with probability $1 - f_{jj}$. Hence $\mathrm{P}(N_j = n | X_0 = j) = f_{jj}^{n-1}(1 - f_{jj})$ for $n \in \mathbb{N}$.

2. Case $n = 0$: The probability that the chain never visits state j given that it starts in i is given by $1 - f_{ij}$. Hence $\mathrm{P}(N_j = 0 | X_0 = i) = 1 - f_{ij}$.

   Case $n \in \mathbb{N}$: The first visit to state j starting from i happens with probability f_{ij}, then there need to be $n-1$ return visits, which happen with probability f_{jj}^{n-1}, and there must be no further return to state j afterwards which happens with probability $1 - f_{jj}$. Hence $\mathrm{P}(N_j = n | X_0 = i) = f_{ij} f_{jj}^{n-1}(1 - f_{jj})$ for $n \in \mathbb{N}$.

□

End of lecture 4. .....

Corollary 3.5.4. *Let $j \in E$ denote a transient state. Then*

1. $$\mathrm{E}(N_j | X_0 = j) = \frac{1}{1 - f_{jj}}. \tag{3.5.1}$$
2. *For $i \neq j$, we have*

$$\mathrm{E}(N_j | X_0 = i) = \frac{f_{ij}}{1 - f_{jj}}.$$

Proof. 1. We will use the following property of the geometric series: Let $q \in \mathbb{R}$ with $|q| < 1$. Set

$$f(q) := \sum_{n=0}^{\infty} q^n = \frac{1}{1 - q}.$$

Then

$$f'(q) = \sum_{n=0}^{\infty} n q^{n-1} = \sum_{n=1}^{\infty} n q^{n-1} = \frac{1}{(1 - q)^2}.$$

Applying this result to $q = f_{jj} < 1$ (since j is transient) leads to

$$\mathrm{E}(N_j|X_0 = j) = \sum_{n=0}^{\infty} n\mathrm{P}(N_j|X_0 = j) = \sum_{n=1}^{\infty} n f_{jj}^{n-1}(1 - f_{jj}) = \frac{1}{1 - f_{jj}}.$$

2. For $i \neq j$, we have

$$\begin{aligned}\mathrm{E}(N_j|X_0 = i) &= \sum_{n=0}^{\infty} n\mathrm{P}(N_j = n|X_0 = i) = \sum_{n=1}^{\infty} n f_{ij} f_{jj}^{n-1}(1 - f_{jj}) \\ &= f_{ij}(1 - f_{jj}) \sum_{n=1}^{\infty} n f_{jj}^{n-1} = \frac{f_{ij}}{1 - f_{jj}}.\end{aligned}$$

□

Theorem 3.5.5. *Given that $X_0 = j$, the expected number of visits to state j is given by*

$$\mathrm{E}(N_j|X_0 = j) = \sum_{n=0}^{\infty} p_{jj}(n), \tag{3.5.2}$$

where the infinite series might diverge to ∞.

Proof. Recall that the so-called Tonelli's theorem says that we are allowed to interchange the expectation and infinite sum provided that all summands are nonnegative (even if we obtain the value ∞).

Then

$$\begin{aligned}\mathrm{E}(N_j|X_0 = j) &= \mathrm{E}\left(\sum_{n=0}^{\infty} I_n^{(j)}|X_0 = j\right) = \sum_{n=0}^{\infty} \mathrm{E}\left(I_n^{(j)}|X_0 = j\right) \\ &= \sum_{n=0}^{\infty} \mathrm{P}(X_n = j|X_0 = j) = \sum_{n=0}^{\infty} p_{jj}(n).\end{aligned}$$

since

$$\begin{aligned}\mathrm{E}(I_n^{(j)}|X_0 = j) &= 0 \times \mathrm{P}(I_n^{(j)} = 0|X_0 = j) + 1 \times \mathrm{P}(I_n^{(j)} = 1|X_0 = j) \\ &= \mathrm{P}(X_n = j|X_0 = j).\end{aligned}$$

□

We can now provide the proof for Theorem 3.5.2.

Proof of Theorem 3.5.2. We need to prove two directions:

1. " j transient $\Rightarrow \sum_{n=1}^{\infty} p_{jj}(n) < \infty$": Suppose that j is transient. Then we have, using equations (3.5.1) and (3.5.2),

$$\mathrm{E}\left(N_j|X_0 = j\right) = \sum_{n=0}^{\infty} p_{jj}(n) = \frac{1}{(1 - f_{jj})} < \infty$$

since $f_{jj} < 1$.

2. " $\sum_{n=1}^{\infty} p_{jj}(n) < \infty \Rightarrow j$ transient ": Conversely, suppose that $\sum_{n=0}^{\infty} p_{jj}(n) < \infty$. Then, conditional on $X_0 = j$, N_j is a nonnegative random variable with finite (conditional) mean and hence must be finite. That implies that starting from state j, the chain returns to state j only finitely many times. Hence, there is a positive probability, that starting from state j, the chain never returns to j. I.e. $1 - f_{jj} > 0$. Hence $f_{jj} < 1$, hence j is transient.

□

These results lead us to

Corollary 3.5.6. *If $j \in E$ is transient then $p_{ij}(n) \to 0$ as $n \to \infty$ for all $i \in E$.*

Proof. If $j \in E$ is transient, then $\sum_{n=1}^{\infty} p_{jj}(n) < \infty$. Hence $p_{jj}(n) \to 0$ as $n \to \infty$.

According to Lemma 3.4.7, we have for all $i, j \in E, n \in \mathbb{N}$:

$$p_{ij}(n) = \sum_{l=0}^{n} f_{ij}(l) p_{jj}(n-l) = \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l).$$

Hence

$$\begin{aligned}
\sum_{n=0}^{\infty} p_{ij}(n) &= p_{ij}(0) + \sum_{n=1}^{\infty} p_{ij}(n) \\
&= p_{ij}(0) + \sum_{n=1}^{\infty} \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l) \\
&= \delta_{ij} + \sum_{l=1}^{\infty} f_{ij}(l) \sum_{n=l}^{\infty} p_{jj}(n-l) \\
&= \delta_{ij} + \sum_{l=1}^{\infty} f_{ij}(l) \sum_{m=0}^{\infty} p_{jj}(m) \\
&= \delta_{ij} + \sum_{l=0}^{\infty} f_{ij}(l) \sum_{m=0}^{\infty} p_{jj}(m) \\
&= \delta_{ij} + f_{ij} \sum_{m=0}^{\infty} p_{jj}(m) \\
&\leq \delta_{ij} + \sum_{l=0}^{\infty} p_{jj}(l) < \infty,
\end{aligned}$$

where we used that $f_{ij} \leq 1$. The necessary condition for the convergence of the infinite series is $p_{ij}(n) \to 0$ as $n \to \infty$. □

3.5.1 Mean recurrence time, null and positive recurrence

Recall that we define the **first passage/hitting time** of X as

$$T_j = \min\{n \in \mathbb{N} : X_n = j\}, \qquad \text{for } j \in E.$$

If $X_n \neq j$ for all $n \in \mathbb{N}$, then we set $T_j = \infty$.

Definition 3.5.7. *The **mean recurrence time** μ_i of a state $i \in E$ is defined as $\mu_i = \mathrm{E}[T_i | X_0 = i]$.*

Theorem 3.5.8. *Let $i \in E$. We have $\mathrm{P}(T_i = \infty | X_0 = i) > 0$ if and only if i is transient; in that case $\mu_i = \mathrm{E}[T_i | X_0 = i] = \infty$.*

Proof. We have

$$\begin{aligned}
&\mathrm{P}(T_i = \infty | X_0 = i) = \mathrm{P}(X_n \neq i \text{ for all } n \in \mathbb{N} | X_0 = i) \\
&= 1 - \mathrm{P}(X_n = i \text{ for some } n \in \mathbb{N} | X_0 = i) > 0 \\
\Leftrightarrow &\mathrm{P}(X_n = i \text{ for some } n \in \mathbb{N} | X_0 = i) < 1 \Leftrightarrow i \text{ is transient.}
\end{aligned}$$

Further, if i is transient, then

$$\mu_i = \mathrm{E}(T_i|X_0 = i) = \sum_{n=1}^{\infty} n\mathrm{P}(T_i = n|X_0 = i) + \infty\mathrm{P}(T_i = \infty|X_0 = i) = \infty.$$

□

Theorem 3.5.9. *For a recurrent state $i \in E$, we have*

$$\mu_i = \mathrm{E}[T_i|X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}(n),$$

which can be finite or infinite.

Proof. For a recurrent state $i \in E$, we have $\mathrm{P}(T_i = \infty|X_0 = i) = 0$. Hence

$$\mu_i = \mathrm{E}[T_i|X_0 = i] = \sum_{n=1}^{\infty} n\mathrm{P}(T_i = n|X_0 = i) = \sum_{n=1}^{\infty} n f_{ii}(n).$$

□

Summarising we can say that

$$\mu_i = \mathrm{E}[T_i|X_0 = i] = \begin{cases} \sum_{n=1}^{\infty} n f_{ii}(n) & \text{if } i \text{ is recurrent} \\ \infty & \text{if } i \text{ is transient.} \end{cases}$$

Definition 3.5.10. *A recurrent state $i \in E$ is called **null** if $\mu_i = \infty$ and **positive** (or non-null) if $\mu_i < \infty$.*

Note that we will later show that, if the state space is finite, i.e. $\mathrm{card}(E) = K < \infty$, then all recurrent states are positive. I.e. null recurrent states can only appear in the case when $\mathrm{card}(E) = K = \infty$.

Theorem 3.5.11. *A recurrent state $i \in E$ is null iff $p_{ii}(n) \to 0$ as $n \to \infty$; if this holds then $p_{ji}(n) \to 0$ as $n \to \infty$ for all $j \in E$.*

Proof. See Grimmett & Stirzaker (2001*b*, p. 222, 232). □

Example 3.5.12. *Let us re-visit Example 3.4.6. Here we have $E = \{1, 2\}$ and*

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}.$$

We want to find the mean recurrence times μ_1 and μ_2.

We already showed that $f_{11}(1) = \frac{1}{2}, f_{11}(n) = \frac{1}{8}\left(\frac{3}{4}\right)^{n-2}$, for $n \geq 2$, hence $f_{11} = 1$, and $f_{22}(1) = p_{22} = \frac{3}{4}, f_{22}(n) = \frac{1}{8}\left(\frac{1}{2}\right)^{n-2}$, for $n \geq 2$, Hence $f_{22} = 1$.

Recall that for $|q| < 1 : \sum_{n=0}^{\infty} q^n = (1-q)^{-1}, \sum_{n=1}^{\infty} nq^{n-1} = (1-q)^{-2}$.

$$\mu_1 = \sum_{n=1}^{\infty} n f_{11}(n) = \frac{1}{2} + \frac{1}{8}\sum_{n=2}^{\infty} n\left(\frac{3}{4}\right)^{n-2} = \frac{1}{2} + \frac{1}{8}\frac{4}{3}\left[\sum_{n=1}^{\infty} n\left(\frac{3}{4}\right)^{n-1} - 1\right]$$
$$= \frac{1}{2} + \frac{1}{6}\left[\left(1 - \frac{3}{4}\right)^{-2} - 1\right] = \frac{1}{2} + \frac{16}{6} - \frac{1}{6} = 3,$$

$$\mu_2 = \sum_{n=1}^{\infty} n f_{22}(n) = \frac{3}{4} + \frac{1}{8}\sum_{n=2}^{\infty} n\left(\frac{1}{2}\right)^{n-2} = \frac{3}{4} + \frac{1}{8}2\left[\sum_{n=2}^{\infty} n\left(\frac{1}{2}\right)^{n-1} - 1\right]$$
$$= \frac{3}{4} + \frac{1}{4}\left[\left(1 - \frac{1}{2}\right)^{-2} - 1\right] = \frac{3}{4} + 1 - \frac{1}{4} = \frac{3}{2}.$$

End of lecture 5. .....

3.5.2 Generating functions for $p_{ij}(n)$ and $f_{ij}(n)$ [Reading material]

Next we study the so-called generating functions of $p_{ij}(n)$ and $f_{ij}(n)$. Recall that you have studied so-called probability generating functions in your Y1 probability course.

Definition 3.5.13 (Probability generating function (p.g.f.))**.** *Let X denote a discrete random variable with $\mathrm{Im}X \subseteq \mathbb{N} \cup \{0\}$. We denote by*

$$\mathcal{S}_X = \left\{ s \in \mathbb{R} : \sum_{x=0}^{\infty} |s|^x \mathrm{P}(X = x) < \infty \right\}.$$

*Then the **probability generating function** (pgf) of X is defined as the function $G_X : \mathcal{S}_X \to \mathbb{R}$ given by*

$$G_X(s) = \mathrm{E}(s^X) = \sum_{x=0}^{\infty} s^x \mathrm{P}(X = x).$$

We observe that the pgf is well-defined for $|s| \leq 1$ since

$$\sum_{x=0}^{\infty} |s|^x \mathrm{P}(X = x) \leq \sum_{x=0}^{\infty} \mathrm{P}(X = x) = 1 < \infty.$$

Also, $G_X(0) = \mathrm{P}(X = 0)$ and $G_X(1) = 1$. More generally, we can define *generating functions* for any sequence $(a_n)_n \in \mathbb{N}_0$ of real numbers:

Definition 3.5.14. *For a sequence $(a_n)_{n\in\mathbb{N}_0}$ of real numbers, we define its **generating function***

$$G_{(a_n)}(s) := a_0 + a_1 s + a_2 s^2 + \cdots = \sum_{n=0}^{\infty} a_n s^n, \tag{3.5.3}$$

for all s for which $\sum_{n=0}^{\infty} |a_n||s|^n < \infty$.

One can show that a generating function of a real sequence specifies this sequence uniquely. I.e., given a sequence $(a_n)_{n\in\mathbb{N}_0}$, one can compute its generating function 3.5.3. Conversely, given a generating function G, one can Taylor-expand it around 0:

$$G(s) = a_0 + a_1 s + a_2 s^2 + \cdots,$$

for small s, which is a unique series expansion which characterises $(a_n)_n \in \mathbb{N}_0$ uniquely.

Let us recall Abel's theorem, which we will be applying later:

Theorem 3.5.15. *If $a_n \geq 0$ for all $n \in \mathbb{N}_0$ and $G_{(a_n)}(s) = \sum_{n=0}^{\infty} a_n s^n$ is finite for $|s| < 1$, then*

$$\lim_{s\uparrow 1} G_{(a_n)}(s) = \sum_{n=0}^{\infty} a_n,$$

whether the sum is finite of equals $+\infty$.

Note that $s \uparrow 1$ means that we are taking the left limit in 1, which is sometimes also denoted by $s \to 1-$.

Definition 3.5.16. *For $i, j \in E$, we define the generating functions*

$$G_{(p_{ij}(n))}(s) = \sum_{n=0}^{\infty} p_{ij}(n) s^n,$$

$$G_{(f_{ij}(n))}(s) = \sum_{n=0}^{\infty} f_{ij}(n) s^n,$$

for $|s| < 1$.

Theorem 3.5.17. *For $i, j \in E$, $|s| < 1$, we have*

$$G_{(p_{ij}(n))}(s) = \delta_{ij} + G_{(f_{ij}(n))}(s) G_{(p_{jj}(n))}(s),$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Proof. From Lemma 3.4.7, we know that

$$p_{ij}(n) = \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l).$$

Multiply both sides by s^n and take the sum from $n = 1$ to ∞:

$$\sum_{n=1}^{\infty} s^n p_{ij}(n) = \sum_{n=1}^{\infty} s^n \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l).$$

For the left hand side, we get

$$\sum_{n=1}^{\infty} s^n p_{ij}(n) = -p_{ij}(0) + \sum_{n=0}^{\infty} s^n p_{ij}(n) = -\delta_{ij} + G_{(p_{ij}(n))}(s).$$

For the right hand side, we get by changing the order of summation ($1 \leq l \leq n < \infty$):

$$\begin{aligned}
&\sum_{n=1}^{\infty} s^n \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l) = \sum_{n=1}^{\infty} \sum_{l=1}^{n} s^l f_{ij}(l) s^{n-l} p_{jj}(n-l) \\
&= \sum_{l=1}^{\infty} s^l f_{ij}(l) \sum_{n=l}^{\infty} s^{n-l} p_{jj}(n-l) = \sum_{l=1}^{\infty} s^l f_{ij}(l) \sum_{m=0}^{\infty} s^m p_{jj}(m) \\
&= G_{(f_{ij}(n))}(s) G_{(p_{jj}(n))}(s)
\end{aligned}$$

Combining the left and right hand side leads to

$$-\delta_{ij} + G_{(p_{ij}(n))}(s) = G_{(f_{ij}(n))}(s) G_{(p_{jj}(n))}(s),$$

which is equivalent to

$$G_{(p_{ij}(n))}(s) = \delta_{ij} + G_{(f_{ij}(n))}(s) G_{(p_{jj}(n))}(s).$$

□

We now provide an alternative proof of Theorem 3.5.2 using generating functions.

Alternative proof of Theorem 3.5.2 using generating functions. Recall

$$G_{(p_{jj}(n))}(s) = \sum_{n=0}^{\infty} p_{jj}(n) s^n, \qquad G_{(f_{jj}(n))}(s) = \sum_{n=0}^{\infty} f_{jj}(n) s^n, \qquad \text{for } |s| < 1.$$

Hence, taking the left limit when s tends to 1 (since $|s| < 1$) and applying Abel's theorem (Theorem 3.5.15):

$$\lim_{s \uparrow 1} G_{(p_{jj}(n))}(s) = \sum_{n=0}^{\infty} p_{jj}(n), \qquad \lim_{s \uparrow 1} G_{(f_{jj}(n))}(s) = \sum_{n=0}^{\infty} f_{jj}(n) = f_{jj}.$$

From Theorem 3.5.17 we deduce for $|s| < 1$:

$$G_{(p_{jj}(n))}(s) = 1 + G_{(f_{jj}(n))}(s) G_{(p_{jj}(n))}(s) \Leftrightarrow G_{(p_{jj}(n))}(s) = \frac{1}{1 - G_{(f_{jj}(n))}(s)}.$$

Hence, taking the left limit when s tends to 1 leads to

$$\sum_{n=0}^{\infty} p_{jj}(n) = \frac{1}{1-f_{jj}},$$

which could be of the form $\infty = \infty$. Note that f_{jj} is a probability and hence in $[0,1]$. We can now conclude:

1. j is recurrent $\Leftrightarrow f_{jj} = 1 \Leftrightarrow \sum_{n=0}^{\infty} p_{jj}(n) = \infty$,
2. j is transient $\Leftrightarrow f_{jj} < 1 \Leftrightarrow \sum_{n=0}^{\infty} p_{jj}(n) < \infty$,

□

3.5.3 Example: Null recurrence/transience of a simple random walk [Reading material]

We return to the example of a simple random walk, see Example 3.3.7. We want to show that the state 0 is null recurrent for $p = 0.5$ and transient otherwise.

From Example 3.3.7 we know that

$$p_{00}(n) = \binom{n}{\frac{n}{2}} (p(1-p))^{n/2},$$

for even n and $p_{00}(n) = 0$ otherwise.

We would like to find the generating function, for $|s| < 1$,

$$G_{(f_{00}(n))}(s) = \sum_{n=0}^{\infty} s^n f_{00}(n), \qquad \frac{d}{ds} G_{(f_{00}(n))}(s) = \sum_{n=1}^{\infty} n s^{n-1} f_{00}(n).$$

Applying Abel's theorem, leads to

$$\lim_{s\uparrow 1} G_{(f_{00}(n))}(s) = \lim_{s\uparrow 1} \sum_{n=0}^{\infty} s^n f_{00}(n) = \sum_{n=0}^{\infty} f_{00}(n) = f_{00},$$
$$\lim_{s\uparrow 1} \frac{d}{ds} G_{(f_{00}(n))}(s) = \lim_{s\uparrow 1} \sum_{n=1}^{\infty} n s^{n-1} f_{00}(n) = \sum_{n=1}^{\infty} n f_{00}(n) = \mu_0.$$

For the latter, note that the derivative of a power series has the same radius of convergence as the original power series. Hence Abel's theorem is applicable here. We use Theorem 3.5.17 to relate the generating functions of $(f_{00}(n))$ and $(p_{00}(n))$: For $|s| < 1$, we have

$$G_{(p_{00}(n))}(s) = 1 + G_{(f_{00}(n))}(s) G_{(p_{00}(n))}(s).$$

Hence

$$G_{(f_{00}(n))}(s) = 1 - \frac{1}{G_{(p_{00}(n))}(s)},$$

where, for $|s| < 1$,

$$\begin{aligned} G_{(p_{00}(n))}(s) &= \sum_{n=0}^{\infty} s^n p_{00}(n) = \sum_{n=0}^{\infty} \binom{2n}{n} (p(1-p))^n s^{2n} \\ &= \sum_{n=0}^{\infty} \binom{2n}{n} (p(1-p)s^2)^n. \end{aligned}$$

Note that

$$\sum_{k=0}^{\infty} \binom{2k}{k} x^k = \frac{1}{\sqrt{1-4x}}, \qquad \text{for } |x| < \frac{1}{4}.$$

Hence

$$G_{(p_{00}(n))}(s) = \frac{1}{\sqrt{1-4p(1-p)s^2}},$$

and

$$G_{(f_{00}(n))}(s) = 1 - \frac{1}{G_{(p_{00}(n))}(s)} = 1 - \sqrt{1-4p(1-p)s^2}.$$

Taking the limit leads to

$$f_{00} = \lim_{s\uparrow 1} G_{(f_{00}(n))}(s) = 1 - \sqrt{1-4p(1-p)} = 1 - 2\left|p - \frac{1}{2}\right|.$$

Hence $f_{00} = 1$ (and hence 0 is recurrent) for $p = \frac{1}{2}$, and $f_{00} < 1$ (and hence 0 is transient) for $p \neq \frac{1}{2}$. Next we focus on the case when $p = \frac{1}{2}$ and show that 0 is null recurrent. Note that

$$G_{(f_{00}(n))}(s) = 1 - \sqrt{1-s^2},$$

and hence

$$\frac{d}{ds} G_{(f_{00}(n))}(s) = \frac{1}{2}(1-s^2)^{-1/2} 2s = \frac{s}{\sqrt{1-s^2}},$$

and

$$\mu_0 = \lim_{s\uparrow 1} \frac{d}{ds} G_{(f_{00}(n))}(s) = \infty,$$

hence 0 is indeed null recurrent.

Exercise 3.5.18. *Show that*

$$\sum_{n=0}^{\infty} \binom{2n}{n} x^n = \frac{1}{\sqrt{1-4x}}, \qquad \text{for } |x| < \frac{1}{4}. \tag{3.5.4}$$

Note that this is the generating function of the sequence $\left(\binom{2n}{n}\right)_{n\in\mathbb{N}_0}$. *Hint: Proceed in four steps:*

1. *Show that*

$$\binom{2(n+1)}{n+1} = \frac{4n+2}{n+1}\binom{2n}{n}, n \in \mathbb{N}_0$$

2. *Show that*

$$\binom{-\frac{1}{2}}{n+1} = (-1)\frac{n+\frac{1}{2}}{n+1}\binom{-\frac{1}{2}}{n}, n \in \mathbb{N}_0,$$

   by using the definition of the general binomial coefficient.

3. *Show that*

$$\binom{2n}{n} = (-1)^n 4^n \binom{-\frac{1}{2}}{n}, n \in \mathbb{N}_0.$$

4. *Use the generalise binomial theorem to show* (3.5.4).

Remark 3.5.19. *We have shown that the state 0 in a symmetric simple random walk on $E = \mathbb{Z}$ is (null) recurrent. We will soon show that this is true for all other states in E. If one considers a symmetric random walk on the plane $\mathbb{Z}^2$ where one can either move up, down, left or right with probability $\frac{1}{4}$ each, one can show that such a random walk (often called the* drunkard's walk*) is also recurrent. However, a symmetric random walk in $\mathbb{Z}^3$ is transient. Hence there is the famous quote by the mathematician Shizuo Kakutani saying,*

> *"A drunk man will find his way home, but a drunk bird may get lost forever."*

3.6 Aperiodicity and ergodicity

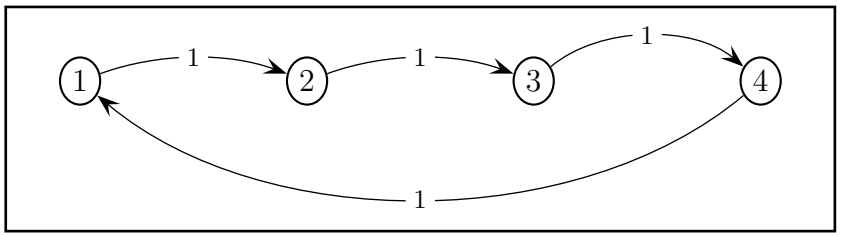
Definition 3.6.1. *The period of a state i is defined by*

$$d(i) = gcd\{n : p_{ii}(n) > 0\}$$

*the greatest common divisor of the epochs at which return is possible. If $d(i) > 1$ then the state is **periodic** otherwise it is **aperiodic**.*

Example 3.6.2. *Consider the Markov chain with state space $E = \{1, 2, 3, 4\}$ and transition matrix*

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$



We can easily see that for $i \in \{1, 2, 3, 4\}$ we have $p_{ii}(n) = 1 > 0$ for $n \in \{4, 8, 12, \dots\}$. Hence $d(i) = 4$.

Exercise 3.6.3. *Suppose that $i \in E$ is a periodic state with period $d(i) = gcd\{n \in \mathbb{N} : p_{ii}(n) > 0\} > 1$. Do we always have that $p_{ii}(d(i)) > 0$? No! Counterexample: Consider a Markov chain with state space $E = \{1, 2, 3, 4\}$ and transition matrix*

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}.$$

Then $A_1 := \{n \in \mathbb{N} : p_{11}(n) > 0\} = \{4, 6, 8, 10, \dots\}$ and $d(1) = gcd(A_1) = 2$. However, $p_{11}(2) = 0$.

Definition 3.6.4. *A state is **ergodic** if it is positive recurrent and aperiodic.*

3.7 Communicating classes

In this section we study how we can divide the state space of a Markov chain into equivalence classes, where the states in each class share important properties.

Definition 3.7.1. *1. We say that state j is **accessible** from state i, written $i \to j$, if the chain may ever visit state j, with positive probability, starting from i. In other words, $i \to j$ if there exist $m \in \mathbb{N}_0$ such that $p_{ij}(m) > 0$.*

*2. i and j **communicate** if $i \to j$ and $j \to i$, written $i \leftrightarrow j$.*

Clearly, if $i \neq j$, $i \to j$ iff $f_{ij} > 0$.

Theorem 3.7.2. *The concept of communication is an equivalence relation.*

Proof. 1. Reflexivity ($i \leftrightarrow i$): Note that we have $p_{ii}(0) = 1$ since

$$p_{ij}(0) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

2. Symmetry (if $i \leftrightarrow j$, then $j \leftrightarrow i$): This follows directly from the definition.

3. Transitivity (if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$): $i \leftrightarrow j$ and $j \leftrightarrow k$ imply that there exist integers $n, m \geq 0$ such that $p_{ij}(n) > 0$ and $p_{jk}(m) > 0$. Hence (using the CK equations and the positivity of the transition probabilities)

$$p_{ik}(n+m) = \sum_{l \in E} p_{il}(n) p_{lk}(m) \geq p_{ij}(n) p_{jk}(m) > 0 \quad \Rightarrow i \to k.$$

Similarly, one can show that $i \leftarrow k$.

□

Remark 3.7.3.
- *The totality of states can be partitioned into equivalence classes, which are called **communicating classes**. Note that the states in an equivalence class are those which communicate with each other.*
- *Note that it may be possible starting in one equivalence class to enter another class with positive probability. In that case, we could not return to the initial class (else the classes would form a single equivalence class).*
- *These findings will be formalised in Theorem 3.7.9.*

We will now show that states which communicate with each other share properties such as recurrence, transience and periodicity.

Theorem 3.7.4. *If $i \leftrightarrow j$ then*

1. *i and j have the same period*
2. *i is transient if and only if j is transient*
3. *i is recurrent if and only if j is recurrent.*
4. *i is null recurrent if and only if j is null recurrent.*

Proof. We only show the recurrence and transience: Let $i \leftrightarrow j$.
Then there exist integers $n, m \geq 0$ such that

$$p_{ij}(n) > 0, \qquad p_{ji}(m) > 0.$$

For any integer $l \geq 0$, we have (using the CK equations twice):

$$p_{jj}((m+l)+n) = \sum_{k \in E} p_{jk}(m+l) p_{kj}(n) \geq p_{ji}(m+l) p_{ij}(n),$$
$$p_{ji}(m+l) = \sum_{k \in E} p_{jk}(m) p_{ki}(l) \geq p_{ji}(m) p_{ii}(l).$$

Hence

$$p_{jj}(m+l+n) \geq p_{ji}(m)p_{ii}(l)p_{ij}(n),$$

where $p_{ji}(m) > 0, p_{ij}(n) > 0$.

Then

$$\sum_{l=1}^{\infty} p_{jj}(l) \geq \sum_{l=1}^{\infty} p_{jj}(m+l+n) \geq p_{ji}(m)p_{ij}(n) \sum_{l=1}^{\infty} p_{ii}(l)$$

Case 1: Suppose that i is recurrent. Then

$$\sum_{l=1}^{\infty} p_{jj}(l) \geq \sum_{l=1}^{\infty} p_{jj}(m+l+n) \geq p_{ji}(m)p_{ij}(n) \sum_{l=1}^{\infty} p_{ii}(l) = \infty.$$

Hence, since $p_{ji}(m) > 0, p_{ij}(n) > 0$, $\sum_{l=1}^{\infty} p_{jj}(l) = \infty$ which is equivalent to j being recurrent by Theorem 3.5.2.

Case 2: Suppose that j is transient. Then

$$\infty > \sum_{l=1}^{\infty} p_{jj}(l) \geq \sum_{l=1}^{\infty} p_{jj}(m+l+n) \geq p_{ji}(m)p_{ij}(n) \sum_{l=1}^{\infty} p_{ii}(l).$$

Hence, since $p_{ji}(m) > 0, p_{ij}(n) > 0$, $\sum_{l=1}^{\infty} p_{ii}(l) < \infty$ which is equivalent to i being transient. □

See Grimmett & Stirzaker (2001*b*, p. 224) for the complete proof.

Definition 3.7.5. *A set of states C is*

1. ***closed*** *if it is impossible to leave the class, i.e. for all $i \in C$, $j \notin C$ we have $p_{ij} = 0$.*
2. ***irreducible*** *if all states in the set communicate with each other, i.e. $i \leftrightarrow j$ for all $i, j \in C$.*

What the definition tells us, is once we enter a closed set, we never leave; if a closed set only contains one state, i.e. $C = \{i\}$ for an $i \in E$, then i is called **absorbing**.

An irreducible set is aperiodic (or null recurrent etc) if all the states in C have this property; thanks to Theorem 3.7.4, this makes sense.

Importantly, if the entire state-space E is irreducible, then we say that the **Markov chain is irreducible**.

Theorem 3.7.6. *Let C denote an closed communicating class. Then the transition matrix $\mathbf{P}$ restricted to C is stochastic.*

Often we write $\mathbf{P}(C)$ (or $\mathbf{P}_C$) for the restriction of $\mathbf{P}$ to C.

Proof. Since all elements of $\mathbf{P}$ are non-negative, this is also true for $\mathbf{P}(C)$. Further, we know that $\sum_{j \in E} p_{ij} = 1$ for all $i \in E$. Let $i \in C$, then $p_{il} = 0$ for $l \notin C$. Hence, we have

$$1 = \sum_{j \in E} p_{ij} = \sum_{j \in C} p_{ij} + \sum_{j \in E \setminus C} p_{ij} = \sum_{j \in C} p_{ij}.$$

□

Example 3.7.7. *Let $E = \{1, 2, 3\}$ and*

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

Find the communicating classes and determine whether they are closed. Write down the transition matrices restricted to the communicating classes. We have two communicating classes: $C_1 = \{1, 2\}, C_2 = \{3\}$. We observe that C_1 is closed and C_2 is not closed. Further, we find that

$$\mathbf{P}(C_1) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad \mathbf{P}(C_2) = \begin{pmatrix} \frac{1}{2} \end{pmatrix}.$$

We note that $\mathbf{P}(C_1)$ is a stochastic matrix (since it is associated with a closed class), whereas $\mathbf{P}(C_2)$ is not a stochastic matrix.

3.7.1 The decomposition theorem

Theorem 3.7.8. *Let C denote a communicating class consisting of recurrent states. Then C is closed.*

Proof. We prove this statement by contradiction: Suppose there exists an $i \in C$, $j \notin C$, such that $p_{ij} > 0$. Hence, $i \to j$, but $j \not\to i$. Then

$$\mathrm{P}(X_n \neq i \text{ for all } n \in \mathbb{N} | X_0 = i) \geq \mathrm{P}(X_1 = j | X_0 = i) = p_{ij} > 0,$$

which implies that

$$\mathrm{P}(X_n = i \text{ for some } n \in \mathbb{N} | X_0 = i) = 1 - \mathrm{P}(X_n \neq i \text{ for all } n \in \mathbb{N} | X_0 = i) < 1.$$

This is a contradiction to the recurrence of i.

□

Theorem 3.7.9. *The state-space E can be partitioned uniquely into*

$$E = T \cup \left(\bigcup_i C_i \right),$$

where T is the set of transient states, and the C_i are irreducible, closed sets of recurrent states.

Proof. From the properties of equivalence relations, we know that the equivalence classes of $\leftrightarrow$ partition the state space uniquely. From Theorem 3.7.4, we know that transience and recurrence are class properties. Hence it only remains to show that the recurrent equivalence classes are closed, and this follows from Theorem 3.7.8. □

Remark 3.7.10. *Note that in the decomposition theorem above T is not assumed to be a* communicating class*. More precisely, in general we can denote by $T_1, T_2, \ldots$ the transient communicating classes, then $T = \cup_i T_i$ is the collection of all transient states of the Markov chain.*

This result helps us to understand what is going on, in terms of the Markov chain. If we start the chain in any of the C_i then the chain never leaves, and, effectively this is the state-space. On the other hand, if the chain starts in the transient set, the chain either stays there forever, or moves, and gets absorbed into a closed class. Note, however, that in the case when $\mathrm{card}(E) < \infty$, the chain will not be able to stay in transient states forever, but will eventually move to a closed class.

End of lecture 6. .....

Theorem 3.7.11. *Let $K < \infty$. Then at least one state is recurrent and all recurrent states are positive.*

Proof. Part 1: Suppose that all states are transient; then according to Corollary 3.5.6 $\lim_{n \to +\infty} p_{ij}(n) = 0$ for all i. Since the transition matrix is stochastic, we have $\sum_{j \in E} p_{ij}(n) = 1$. Since the state space is finite, we can interchange limit operations and finite summation: Take the limit through the summation sign leads the following contradiction:

$$1 = \lim_{n \to \infty} \sum_{j \in E} p_{ij}(n) = \sum_{j \in E} \lim_{n \to \infty} p_{ij}(n) = 0.$$

Part 2: Suppose there is a non-empty communicating class consisting of null recurrent states denoted by C_{i^*}. Then C_{i^*} is closed and $\mathbf{P}(C_{i^*})$ is stochastic. Hence, for all $i \in C_{i^*}$, we have

$$1 = \sum_{k \in E} p_{ik}(n) = \sum_{k \in C_{i^*}} p_{ik}(n).$$

Since all states in the class C_{i^*} are null recurrent, we have $\lim_{n\to\infty} p_{ik}(n) = 0$ for all $k \in C_{i*}$ according to Theorem 3.5.11. As before, we take the limit through the summation sign to obtain the contradiction

$$1 = \lim_{n\to\infty} \sum_{k \in C_{i^*}} p_{ik}(n) = \sum_{k \in C_{i^*}} \lim_{n\to\infty} p_{ik}(n) = 0.$$

□

Theorem 3.7.12. *Let C be a communicating class which is finite (i.e. $\mathrm{card}(C) < \infty$) and closed. Then all states in C are positive recurrent.*

Proof. We can argue exactly as in the proof of Theorem 3.7.11: Since C is closed $\mathbf{P}(C)$ is stochastic. Hence, for all $i \in C$, we have

$$1 = \sum_{k \in E} p_{ik}(n) = \sum_{k \in C} p_{ik}(n).$$

Suppose that all states in C are transient. Then $\lim_{n\to\infty} p_{ik}(n) = 0$ for all $k \in C$ according to Corollary 3.5.6. As before, we take the limit through the summation sign to obtain the contradiction

$$1 = \lim_{n\to\infty} \sum_{k \in C} p_{ik}(n) = \sum_{k \in C} \lim_{n\to\infty} p_{ik}(n) = 0.$$

Hence all states in C must be recurrent, and in fact positive, since we have already shown that null recurrent states do not exist on a finite state space. □

3.7.2 Class properties

Type of class	Finite	Infinite
Closed	positive recurrent	positive recurrent null recurrent transient
Not closed	transient	transient

We can use the results listed in the table above to classify the states of a Markov chain. We will demonstrate this in the following example.

Exercise 3.7.13. *Suppose we have a Markov chain with state space $E = \{1, 2, 3, 4\}$ and transition matrix*

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Draw a transition diagram an find the communicating classes! Determine whether the classes are (positive) recurrent or transient.

Solution. The transition diagram is given by:

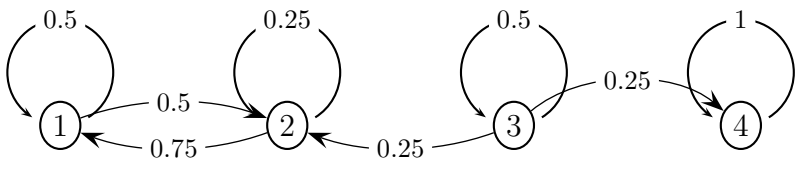


Figure 3.6: Transition diagram for the Markov chain described in Exercise 3.7.13

We have three communicating classes:

1. $C_1 = \{1, 2\}$ is finite and closed, hence positive recurrent;
2. $E_1 = \{3\}$ is not closed, hence transient;
3. $C_2 = \{4\}$ is finite and closed, hence positive recurrent, and, in fact, absorbing.

□

3.8 Application: The gambler's ruin problem

3.8.1 The problem and the results

Now we study a very famous example: The Gambler's ruin problem.

Let $N \geq 2$ be an integer. Consider a gambler with an initial fortune of $i \in \{0, 1, \ldots, N\}$. At each play of the game, the gambler has the

- probability p of winning one unit;
- probability $q = 1 - p$ of losing one unit;
- Assume that successive games are independent.

What is the probability, that if the gambler starts with i units, that the gambler's fortune will reach N before reaching 0?

Let X_n denote the gambler's fortune at time n. Then $\{X_n\}_{n \in \mathbb{N}_0}$ is a Markov chain with transition probabilities

$$\begin{aligned} p_{00} &= p_{NN} = 1, \\ p_{i(i+1)} &= p = 1 - p_{i(i-1)}, \qquad i = 1, 2, \ldots, N-1. \end{aligned}$$

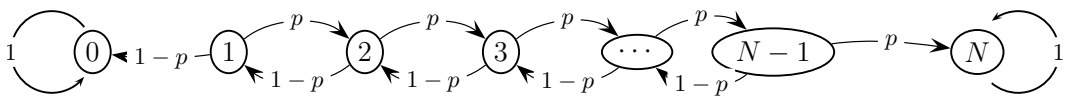


Figure 3.7: Transition diagram for the gambler's ruin problem.

This Markov chain has three communicating classes:

$$C_1 = \{0\}, T_1 = \{1, 2, \ldots, N-1\}, C_2 = \{N\},$$

where the classes C_1 and C_2) are positive recurrent (since they are finite and closed) and the class T_1 is transient (since it is not closed).

A transient class on a finite state space will only be visited finitely many times. Hence after some finite amount of time, the gambler will either win N or be ruined.

Note the Markov chain describing the gambler's fortune is a simple random walk with two absorbing barriers, one at 0 and one at N.

For $i = 0, 1, \ldots, N$, define the first time X visits state i when we consider all possible time points in $\mathbb{N}_0$:

$$V_i = \min\{n \in \mathbb{N}_0 : X_n = i\}.$$

[Note that this is different from the first passage time T_i which considers the minimum over the time points in $\mathbb{N}$ rather than $\mathbb{N}_0$.]

We are interested in the event that the gambler's fortune reaches N before reaching 0, i.e. $\{V_N < V_0\}$.

We define

$$h_i = h_i(N) = \mathrm{P}(V_N < V_0 | X_0 = i),$$

which is the conditional probability of the gambler's fortune reaching N before going bankrupt when starting with i units.

Note that $1 - h_i(N)$ denotes the corresponding conditional probability of the gambler's ruin.

From the definition of V_i, we deduce that

$$\begin{aligned} h_0 &= h_0(N) = \mathrm{P}(V_N < V_0 | X_0 = 0) = 0, \\ h_N &= h_N(N) = \mathrm{P}(V_N < V_0 | X_0 = N) = 1. \end{aligned}$$

We would like to find $h_i(N)$ for $0 < i < N$. In order to tackle this problem, we carry out a so-called *first step analysis*, where we condition on the outcome of the first game. Define the events

$$\begin{aligned} F &:= \text{gambler wins first game}, & \mathrm{P}(F) &= p, \\ F^c &= \text{gambler loses first game}, & \mathrm{P}(F^c) &= q = 1 - p. \end{aligned}$$

If he wins, then he has $i + 1$ otherwise he has $i - 1$.

By conditioning on the outcome of the initial game and applying the law of total probability (for conditional probabilities), we get, for $0 < i < N$,

$$\begin{aligned} h_i &= \mathrm{P}(V_N < V_0 | X_0 = i) \\ &= \mathrm{P}(V_N < V_0 | \{X_0 = i\} \cap F)\mathrm{P}(F | X_0 = i) + \mathrm{P}(V_N < V_0 | \{X_0 = i\} \cap F^c)\mathrm{P}(F^c | X_0 = i) \\ &= \mathrm{P}(V_N < V_0 | X_0 = i, X_1 = i + 1)\mathrm{P}(F) + \mathrm{P}(V_N < V_0 | X_0 = i, X_1 = i - 1)\mathrm{P}(F^c) \\ &= h_{i+1}p + h_{i-1}(1 - p), \end{aligned}$$

where we used that the events F and $\{X_0 = i\}$ are independent and that X is a homogeneous Markov chain. I.e. we have derived the recurrence relation

$$h_i = h_{i+1}p + h_{i-1}q, \qquad i = 1, 2, \ldots, N - 1.$$

We get the following result:

Theorem 3.8.1. *In the gambler's ruin problem described above, we have*

$$h_i = h_i(N) = \begin{cases} \frac{1-(q/p)^i}{1-(q/p)^N}, & \text{if } p \neq \frac{1}{2}, \\ \frac{i}{N} & \text{if } p = \frac{1}{2}. \end{cases}$$

We depict simulated sample paths of the gambler's fortune in Figure 3.8. Moreover, we plot the probabilities $h_i(N)$ for the symmetric case in Figure 3.9 and for the asymmetric case in Figure 3.10.

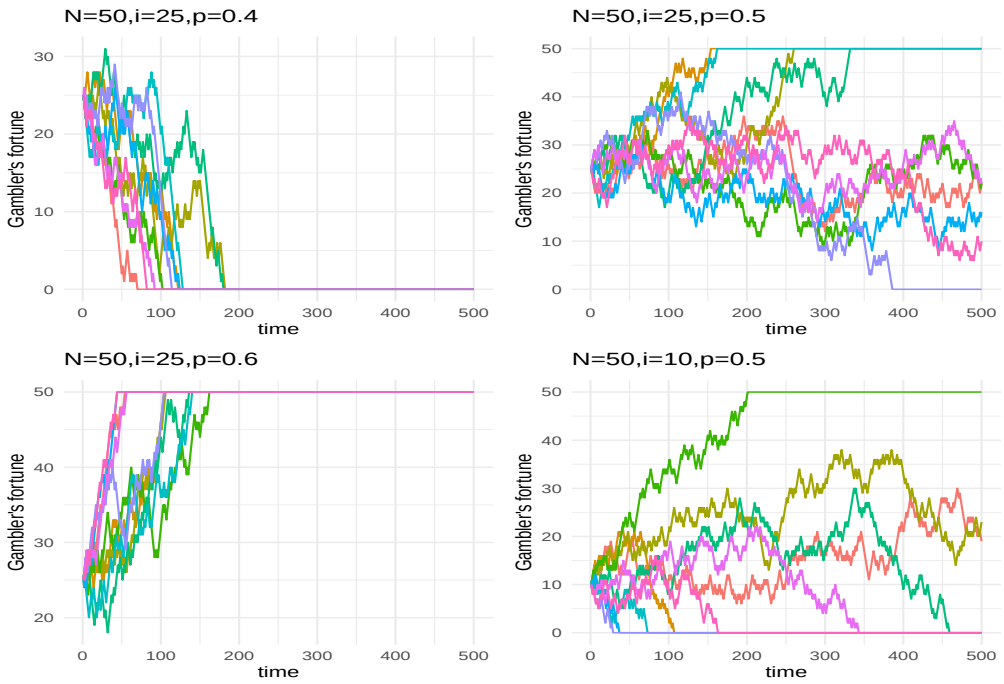


Figure 3.8: Each of the four pictures depicts ten sample paths of the gambler's fortune recorded at time points $0, 1, \ldots, 500$ for $N = 50$. In the first column we set the initial wealth to $i = 25$ and consider two choices for $p \in \{0.4, 0.6\}$. In the second column, we set $p = 0.5$ and vary the initial wealth $i \in \{25, 10\}$.

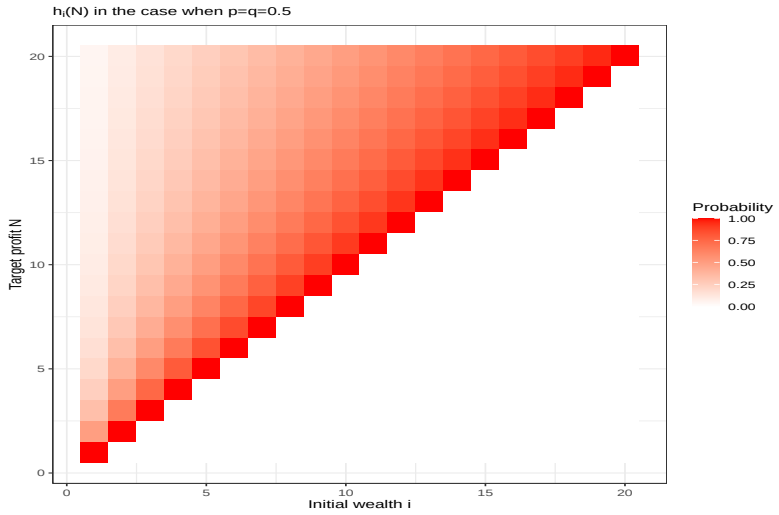


Figure 3.9: We consider the symmetric case when $p = q = \frac{1}{2}$ and depict the probability $h_i(N) = \frac{i}{N}$, for $i, N \in \{1, \ldots, 20\}$.

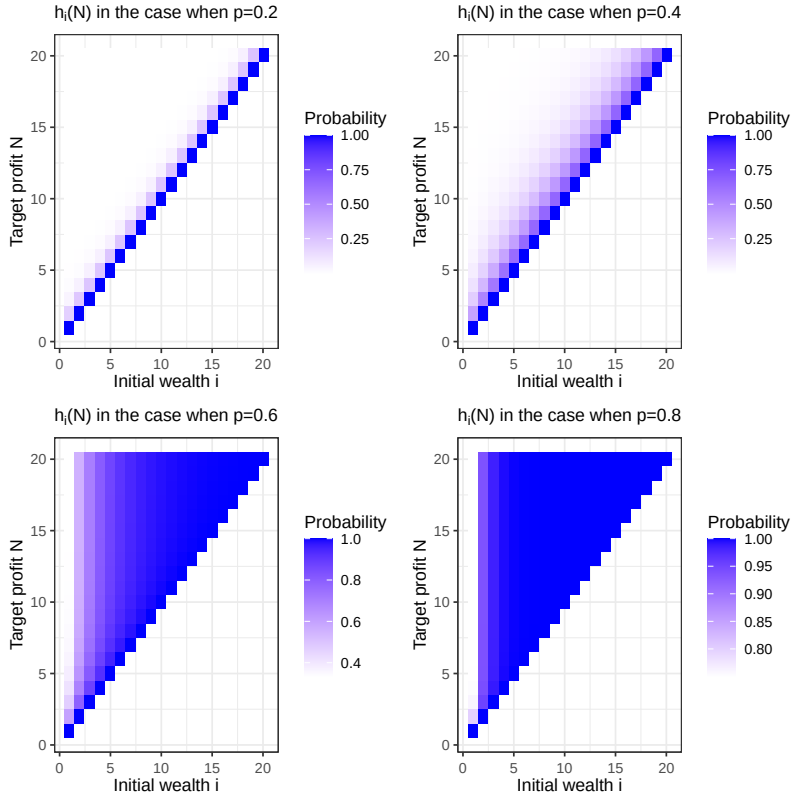


Figure 3.10: We consider the asymmetric case when $p \neq q$ and depict the probability $h_i(N) = \frac{1-(q/p)^i}{1-(q/p)^N}$, for $i, N \in \{1, \ldots, 20\}$.

Let us study the limit as $N \to \infty$.

Theorem 3.8.2. *In the gambler's ruin problem described above, we have*

$$\lim_{N\to\infty} h_i(N) = h_i(\infty) = \begin{cases} 1-(q/p)^i, & \text{if } p > 1/2, \\ 0 & \text{if } p \leq 1/2, \end{cases}$$

Proof. If $p > \frac{1}{2}$, then $\frac{q}{p} < 1 \Rightarrow \lim_{N\to\infty} \left(\frac{q}{p}\right)^N = 0$.

If $p < \frac{1}{2}$, then $\frac{q}{p} > 1 \Rightarrow \lim_{N\to\infty} \left(\frac{q}{p}\right)^N = \infty$. □

Note that $h_i(\infty)$ describes the probability that the gambler will obtain an infinitely large fortune if he is allowed to play forever unless ruined. We find that if $p > 0.5$ (i.e. the gamble has to be strictly better than a fair game), there is a positive probability that the gambler's fortune will increase indefinitely. However, if $p \leq 0.5$, then the gambler will be ruined with probability 1 (assuming he plays against an infinitely rich adversary).

3.8.2 Proof

We still need to prove Theorem 3.8.1. We will show two approaches.

Method 1

We need to solve the difference equation

$$h_i = h_{i+1}p + h_{i-1}q, \qquad i = 1, 2, \ldots, N-1.$$

We guess a solution to the difference equation. In particular, let us try the choice $h_i = cx^i$. Then

$$cx^i = cpx^{i+1} + cqx^{i-1}.$$

Hence we get the auxiliary/characteristic equations:

$$x = px^2 + q \Leftrightarrow 0 = px^2 - x + q,$$

The solutions are $x = q/p$ and $x = 1$. We need to distinguish the two cases when we have two distinct roots, i.e. when $p \neq q$, and the case when we have one root, i.e. when $p = q = \frac{1}{2}$.

The general solution is given by

$$h_i = \begin{cases} c_1 \left(\frac{q}{p}\right)^i + c_2, & \text{if } p \neq \frac{1}{2}, \\ c_1 i + c_2 & \text{if } p = \frac{1}{2}, \end{cases}$$

for constants c_1, c_2 and for $i = 0, 1, \ldots, N$.

Next we use the boundary conditions $h_0 = 0$ and $h_N = 1$ to find the constants c_1, c_2.

For $p \neq q$: $h_i = c_1 \left(\frac{q}{p}\right)^i + c_2$, we find

$$0 = h_0 = c_1 + c_2 \Leftrightarrow c_1 = -c_2,$$
$$1 = h_N = c_1 \left(\frac{q}{p}\right)^N + c_2 = c_1 \left[\left(\frac{q}{p}\right)^N - 1\right] \Leftrightarrow c_1 = \left[\left(\frac{q}{p}\right)^N - 1\right]^{-1}.$$

Hence, we have

$$c_1 = \left[\left(\frac{q}{p}\right)^N - 1\right]^{-1}, \qquad c_2 = \left[1 - \left(\frac{q}{p}\right)^N\right]^{-1}.$$

For $p = q = \frac{1}{2}$: $h_i = c_1 i + c_2$, we have

$$0 = h_0 = c_1 0 + c_2 \Leftrightarrow c_2 = 0,$$
$$1 = h_N = c_1 N + c_2 = c_1 N \Leftrightarrow c_1 = \frac{1}{N}.$$

This concludes the first proof.

Method 2

Next we show how we can solve the recurrence equation directly. For $i = 1, 2 \ldots, N-1$ we have

$$h_i = h_{i+1} p + h_{i-1} q.$$

Since $p + q = 1$, the above equation is equivalent to

$$ph_i + qh_i = h_{i+1}p + h_{i-1}q \Leftrightarrow q(h_i - h_{i-1}) = p(h_{i+1} - h_i) \Leftrightarrow (h_{i+1} - h_i) = \frac{p}{q}(h_i - h_{i-1}).$$

Recall the boundary conditions $h_0 = 0, h_N = 1$. Hence we get

$$h_2 - h_1 = \frac{q}{p}(h_1 - h_0) = \frac{q}{p} h_1,$$
$$h_3 - h_2 = \frac{q}{p}(h_2 - h_1) = \left(\frac{q}{p}\right)^2 h_1, \ldots,$$
$$h_{k+1} - h_k = \frac{q}{p}(h_k - h_{k-1}) = \left(\frac{q}{p}\right)^k h_1,$$

for $k = 1, \ldots, N-1$.

Next we sum the left and right hand side over all $k = 1, \ldots, i$ for $i = 1, \ldots, N-1$ and get, using results for telescoping sums,

$$\sum_{k=1}^{i}(h_{k+1} - h_k) = h_{i+1} - h_1 = \sum_{k=1}^{i}\left(\frac{q}{p}\right)^k h_1 \Leftrightarrow h_{i+1} = \sum_{k=0}^{i}\left(\frac{q}{p}\right)^k h_1,$$

where we note that the last equation trivially also holds for $i = 0$.
Case 1: $p \neq q$: We use the finite geometric series expansion:

$$h_{i+1} = \sum_{k=0}^{i}\left(\frac{q}{p}\right)^k h_1 = h_1 \frac{1 - \left(\frac{q}{p}\right)^{i+1}}{1 - \frac{q}{p}}.$$

Next choose $i = N-1$ and use the boundary condition $h_N = 1$, which leads to

$$1 = h_N = h_1 \frac{1 - \left(\frac{q}{p}\right)^N}{1 - \frac{q}{p}} \Leftrightarrow h_1 = \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}.$$

Hence

$$h_{i+1} = \frac{1 - \left(\frac{q}{p}\right)^{i+1}}{1 - \frac{q}{p}} \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N} = \frac{1 - \left(\frac{q}{p}\right)^{i+1}}{1 - \left(\frac{q}{p}\right)^N},$$

for $i = 0, 1, \ldots, N-1$. Including the boundary cases, we have

$$h_i = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}, \qquad \text{for} \quad i = 0, 1, \ldots, N.$$

Case 2: $p = q = \frac{1}{2}$: Here we have

$$1 - h_1 = \sum_{i=1}^{N-1} h_1 = (N-1)h_1 \Leftrightarrow h_1 = \frac{1}{N}.$$

Hence, for all $i = 0, \ldots, N$, we have $h_i = \frac{i}{N}$.

End of lecture 7. .....

3.9 Stationarity

We will (finally) start to use the theory that we have been building up on Markov chains. The question we look to answer here is the following: 'what can we say about the probabilistic behaviour of the chain; is there a 'stationary' behaviour?' To answer this question, we begin with the concept of a stationary distribution.

Definition 3.9.1. *1. A row vector $\boldsymbol{\lambda}$ is called a **distribution** on E if*

$$\forall\, j \in E, \lambda_j \geq 0, \quad \text{and} \quad \sum_{j \in E} \lambda_j = 1.$$

2. *A row vector $\boldsymbol{\lambda}$ with non-negative entries is called **invariant** for the transition matrix $\mathbf{P}$ if*

$$\lambda \mathbf{P} = \lambda.$$

3. *A row vector $\boldsymbol{\pi}$ is a **invariant/stationary/equilibrium distribution** of a Markov chain on E with transition matrix $\mathbf{P}$ if:*

   (a) *it is a distribution: for each $j \in E$, $\pi_j \geq 0$ and $\sum_{j\in E} \pi_j = 1$.*

   (b) *it is invariant: $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, that is, $\forall\, j \in E$, $\pi_j = \sum_{i\in E} \pi_i p_{ij}$.*

The term stationarity is used for the following reason:

$$\boldsymbol{\pi}\mathbf{P}^2 = (\boldsymbol{\pi}\mathbf{P})\mathbf{P} = \boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$$

that is

$$\boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi} \tag{3.9.1}$$

for any $n \in \mathbb{N}$. Also, if $\boldsymbol{\nu}^{(0)} = \boldsymbol{\pi}$, then using Theorem 3.3.3 and (3.9.1) we get that

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)}\mathbf{P}^n = \boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi} \qquad \forall\, n \in \mathbb{N}_0.$$

That is to say, if $\boldsymbol{\pi}$ is the initial distribution of the Markov chain, then the marginal distribution for any subsequent time instant is also $\boldsymbol{\pi}$.

3.9.1 Stationarity distribution for irreducible Markov chains

We formulate a very important theorem, which we will prove in detail.

Theorem 3.9.2. *An irreducible chain has a stationary distribution $\boldsymbol{\pi}$ if and only if all the states are positive recurrent; in this case $\boldsymbol{\pi}$ is the unique stationary distribution of the chain and is given by $\pi_i = \mu_i^{-1}$ for each $i \in E$ and where μ_i is the mean recurrence time.*

We will divide the proof into several parts and formulate intermediate lemmas to structure the presentation of the rather long proof.

Lemma 3.9.3. *For a Markov chain X, we have for all $j \in E$ and for all $n, m \in \mathbb{N}$*

$$f_{jj}(m+n) = \sum_{i\in E, i\neq j} l_{ji}(m) f_{ij}(n), \tag{3.9.2}$$

where

$$l_{ji}(n) = \mathrm{P}(X_n = i, T_j \geq n | X_0 = j), \qquad i \neq j, \tag{3.9.3}$$

*denotes the **probability that the chain reaches state i in n steps without intermediate return to its starting point j**.*

Proof. Let $j \in E$ and $n, m \in \mathbb{N}$. Then, using the law of total probability,

$$\begin{aligned}
f_{jj}(m+n) &= \mathrm{P}(T_j = m+n | X_0 = j) = \mathrm{P}(T_j = m+n, X_0 = j)[\mathrm{P}(X_0 = j)]^{-1} \\
&= \mathrm{P}(X_{m+n} = j, X_{m+n-1} \neq j, \ldots, X_1 \neq j, X_0 = j)[\mathrm{P}(X_0 = j)]^{-1} \\
&= \sum_{i\in E, i\neq j} \mathrm{P}(X_{m+n} = j, X_{m+n-1} \neq j, \ldots, X_1 \neq j, X_0 = j, X_m = i)[\mathrm{P}(X_0 = j)]^{-1} \\
&= \sum_{i\in E, i\neq j} \mathrm{P}(X_{m+n} = j, X_{m+n-1} \neq j, \ldots, X_{m+1} \neq j | X_m = i, X_{m-1} \neq j, \ldots X_1 \neq j, X_0 = j) \\
&\qquad \mathrm{P}(X_m = i, X_{m-1} \neq j, \ldots X_1 \neq j, X_0 = j)[\mathrm{P}(X_0 = j)]^{-1}
\end{aligned}$$

$$\begin{aligned}
&\overset{\text{Markov}}{=} \sum_{i\in E, i\neq j} \mathrm{P}(X_{m+n}=j, X_{m+n-1}\neq j,\dots,X_{m+1}\neq j|X_m=i)\\
&\qquad \mathrm{P}(X_m=i, X_{m-1}\neq j,\dots X_1\neq j|X_0=j)\\
&\overset{\text{time-hom.}}{=} \sum_{i\in E, i\neq j} \mathrm{P}(X_n=j, X_{n-1}\neq j,\dots,X_1\neq j|X_0=i)\\
&\qquad \mathrm{P}(X_m=i, X_{m-1}\neq j,\dots X_1\neq j|X_0=j)\\
&\overset{\text{def. of } T_j}{=} \sum_{i\in E, i\neq j} \mathrm{P}(T_j=n|X_0=i)\mathrm{P}(X_m=i, T_j\geq m|X_0=j)\\
&= \sum_{i\in E, i\neq j} f_{ij}(n)l_{ji}(m).
\end{aligned}$$

□

Since the $l_{ji}(m)$ and $f_{ij}(n)$ are probabilities and hence non-negative we get the following result.

Corollary 3.9.4. *For a Markov chain X, we have for all $i,j\in E, i\neq j$ and for all $n,m\in\mathbb{N}$*

$$f_{jj}(m+n)\geq l_{ji}(m)f_{ij}(n). \tag{3.9.4}$$

We formulate a recursive formula for $l_{ji}(n)$ which will be useful in the proof of the main theorem.

Lemma 3.9.5. *Let $i\neq j$. Then $l_{ji}(1)=p_{ji}$, and for integers $n\geq 2$,*

$$l_{ji}(n)=\sum_{r\in E: r\neq j} p_{ri}l_{jr}(n-1). \tag{3.9.5}$$

Proof. Let $i\neq j$. Clearly, $l_{ji}(1)=p_{ji}$, and for $n\geq 2$, by the law of total probability and the Markov property, we get

$$\begin{aligned}
l_{ji}(n) &= \sum_{r\in E: r\neq j} \mathrm{P}(X_n=i, X_{n-1}=r, T_j\geq n|X_0=j)\\
&= \sum_{r\in E: r\neq j} \mathrm{P}(X_n=i|X_{n-1}=r, T_j\geq n, X_0=j)\mathrm{P}(X_{n-1}=r, T_j\geq n|X_0=j)\\
&= \sum_{r\in E: r\neq j} \mathrm{P}(X_n=i|X_{n-1}=r)\mathrm{P}(X_{n-1}=r, T_j\geq n-1|X_0=j)\\
&= \sum_{r\in E: r\neq j} p_{ri}l_{jr}(n-1).
\end{aligned}$$

□

Construction of an invariant vector for irreducible, recurrent chains:

Fix a state $j\in E$ and define for any $i\in E$ by $N_i(j)$ the **number of visits to the state i before visiting state j** (when counting from time $n=1$ onwards), i.e.

$$N_i(j)=\sum_{n=1}^{\infty}\mathbb{I}_{\{X_n=i\}\cap\{T_j\geq n\}}=\sum_{n=1}^{T_j}\mathbb{I}_{\{X_n=i\}}.$$

Clearly, $N_j(j)=1$. Note that

$$\sum_{i\in E}N_i(j)=\sum_{i\in E}\sum_{n=1}^{\infty}\mathbb{I}_{\{X_n=i\}}\mathbb{I}_{\{T_j\geq n\}}$$

$$\overset{\text{Tonelli}}{=} \sum_{n=1}^{\infty} \mathbb{I}_{\{T_j \geq n\}} \sum_{i \in E} \mathbb{I}_{\{X_n = i\}} = \sum_{n=1}^{\infty} \mathbb{I}_{\{T_j \geq n\}} = \sum_{n=1}^{T_j} 1 = T_j, \tag{3.9.6}$$

where we used Tonelli's theorem to justify swapping the sums for non-negative addends.

Next we define $\rho_i(j)$ to be the **expected number of visits to the state i between two successive visits to state j**, i.e.

$$\rho_i(j) = \mathrm{E}[N_i(j)|X_0 = j], \tag{3.9.7}$$

where $\rho_j(j) = 1$. From the definition of the mean recurrence time and equation 3.9.6 we deduce that

$$\begin{aligned} \mu_j = \mathrm{E}(T_j|X_0 = j) &= \mathrm{E}\left[\sum_{i \in E} N_i(j)|X_0 = j\right] \\ &\overset{\text{Tonelli}}{=} \sum_{i \in E} \mathrm{E}\left[N_i(j)|X_0 = j\right] = \sum_{i \in E} \rho_i(j). \end{aligned} \tag{3.9.8}$$

We write $\boldsymbol{\rho}(j)$ for the the row vector consisting of elements $\rho_i(j)$ for $i \in E$ with $\mu_j = \sum_{i \in E} \rho_i(j)$. Recall that, if the chain is positive recurrent, then $\mu_j < \infty$ and if it is null recurrent, then $\mu_j = \infty$.

We can now prove that the elements $\rho_i(j)$ are finite and $\boldsymbol{\rho}(j)$ is invariant for $\mathbf{P}$:

Lemma 3.9.6. *For any state $j \in E$ of an irreducible, recurrent chain, the vector $\boldsymbol{\rho}(j)$ satisfies $\rho_i(j) < \infty$ for all i, and furthermore $\boldsymbol{\rho}(j) = \boldsymbol{\rho}(j)\mathbf{P}$.*

Proof. First we prove that $\rho_i(j) < \infty$ for all $i \neq j$ (recall that $\rho_j(j) = 1$).

Applying Tonelli's theorem and using the fact that the conditional expectation of an indicator variable is equal to the conditional probability that the indicator variable is equal to one, we get

$$\begin{aligned} \rho_i(j) = \mathrm{E}[N_i(j)|X_0 = j] &= \mathrm{E}\left[\sum_{n=1}^{\infty} \mathbb{I}_{\{X_n = i\} \cap \{T_j \geq n\}} \middle| X_0 = j\right] \\ &\overset{\text{Tonelli}}{=} \sum_{n=1}^{\infty} \mathrm{E}\left[\mathbb{I}_{\{X_n = i\} \cap \{T_j \geq n\}} \middle| X_0 = j\right] = \sum_{n=1}^{\infty} \mathrm{P}(X_n = i, T_j \geq n | X_0 = j) \\ &= \sum_{n=1}^{\infty} l_{ji}(n). \end{aligned}$$

Since the chain is irreducible, there exists an $n^* \in \mathbb{N}$ such that $f_{ij}(n^*) > 0$. Hence, by equation (3.9.4) we have, for all $m \in \mathbb{N}$,

$$f_{jj}(m + n^*) \geq l_{ji}(m) f_{ij}(n^*) \Leftrightarrow l_{ji}(m) \leq \frac{f_{jj}(m + n^*)}{f_{ij}(n^*)},$$

Hence

$$\begin{aligned} \rho_i(j) = \sum_{m=1}^{\infty} l_{ji}(m) &\leq \frac{1}{f_{ij}(n^*)} \sum_{m=1}^{\infty} f_{jj}(m + n^*) \\ &\leq \frac{1}{f_{ij}(n^*)} \sum_{m=1}^{\infty} f_{jj}(m) = \frac{1}{f_{ij}(n^*)} f_{jj} = \frac{1}{f_{ij}(n^*)} < \infty. \end{aligned}$$

Next, we prove that $\boldsymbol{\rho}(j) = \boldsymbol{\rho}(j)\mathbf{P}$, which is equivalent to showing that for all $i \in E$ we have $\rho_i(j) = \sum_{r \in E} \rho_r(j) p_{ri}$. Recall that

$$\rho_i(j) = \sum_{n=1}^{\infty} l_{ji}(n).$$

Also, by Lemma 3.9.5, we have $l_{ji}(1) = p_{ji}$, and for $n \geq 2$, we get

$$l_{ji}(n) = \sum_{r \in E: r \neq j} p_{ri} l_{jr}(n-1).$$

Hence, for all $i \in E$, we have

$$\begin{aligned}
\rho_i(j) &= \sum_{n=1}^{\infty} l_{ji}(n) = l_{ji}(1) + \sum_{n=2}^{\infty} l_{ji}(n) \\
&= p_{ji} + \sum_{n=2}^{\infty} \sum_{r \in E: r \neq j} p_{ri} l_{jr}(n-1) \\
&\overset{\rho_j(j)=1, \text{Tonelli}}{=} \rho_j(j) p_{ji} + \sum_{r \in E: r \neq j} p_{ri} \sum_{n=2}^{\infty} l_{jr}(n-1) \\
&= \rho_j(j) p_{ji} + \sum_{r \in E: r \neq j} p_{ri} \sum_{n=1}^{\infty} l_{jr}(n) \\
&= \rho_j(j) p_{ji} + \sum_{r \in E: r \neq j} p_{ri} \rho_r(j) = \sum_{r \in E} \rho_r(j) p_{ri},
\end{aligned}$$

which concludes the proof. □

End of lecture 8. .....

Lemma 3.9.7. *Every irreducible, positive recurrent chain has a stationary distribution.*

Proof. From Lemma 3.9.6 we get the representation result $\boldsymbol{\rho}(j) = \boldsymbol{\rho}(j)\mathbf{P}$. From equation (3.9.8) we know that $\mu_j = \sum_{i \in E} \rho_i(j)$, which is clearly nonnegative. Also the μ_j are finite for any positive recurrent chain.

Define

$$\pi_i := \frac{\rho_i(j)}{\mu_j}.$$

Then $\pi_i \geq 0$ for all i and $\sum_{i \in E} \pi_i = 1$ and $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, hence $\boldsymbol{\pi}$ is a stationary distribution. □

We can now summarise our findings in the following theorem.

Theorem 3.9.8. *If the chain is irreducible and recurrent, then there exists a positive root* $\mathbf{x}$ *of the equation* $\mathbf{x} = \mathbf{x}\mathbf{P}$*, which is unique up to a multiplicative constant. Moreover, the chain is positive recurrent if* $\sum_i x_i < \infty$ *and null if* $\sum_i x_i = \infty$*.*

Proof. The existence of the root is an immediate consequence of Lemma 3.9.6. This root is always non–negative and can in fact to be taken strictly positive (this can be shown using similar arguments as the ones used in the context of equation (3.9.9) below). The proof of the uniqueness is left as an exercise, see Exercise 2- 16. □

Lemma 3.9.9. *Let* T *be a nonnegative integer-valued random variable on a probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *and let* $A \in \mathcal{F}$ *be an event with* $\mathrm{P}(A) > 0$*. Show that*

$$\mathrm{E}(T|A) = \sum_{n=1}^{\infty} \mathrm{P}(T \geq n|A).$$

Proof. The proof is left as an exercise, see Exercise 2- 17. □

We are now in a position to prove Theorem 3.9.2.
Recall that we already proved that any irreducible, positive recurrent chain has a stationary distribution. Hence, we need to show:

- If an irreducible chain has a stationary distribution, then the chain is positive recurrent.
- $\pi_i = 1/\mu_i$.

Note that the uniqueness of the stationary distribution will follow from Theorem 3.9.8.

Proof of Theorem 3.9.2. Suppose that $\boldsymbol{\pi}$ is the stationary distribution of the chain. Assume there exists a transient state. Since the chain is irreducible that implies that all states are transient. If all states are transient then $p_{ij}(n) \to 0$, as $n \to \infty$, for all i, j, by Corollary 3.5.6. Since $\boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi}$, for any j

$$\pi_j = \lim_{n\to\infty} \pi_j = \lim_{n\to\infty} \sum_{i\in E} \pi_i p_{ij}(n) \overset{\text{DOM}}{=} \sum_{i\in E} \pi_i \lim_{n\to\infty} p_{ij}(n) = 0,$$

thus, π could not be a stationary vector (see Definition 3.9.1); all states are recurrent. This follows from switching the order of summation and limits using the dominated convergence theorem. Recall Theorem 2.8.2:

Theorem (Dominated Convergence Theorem)**.** *Let $\mathcal{I}$ denote a countable index set. If $\sum_{i\in\mathcal{I}} a_i(n)$ is an absolutely convergent series for all $n \in \mathbb{N}$ such that*

1. *for all $i \in \mathcal{I}$ the limit $\lim_{n\to\infty} a_i(n) = a_i$ exists,*
2. *there exists a sequence $(b_i)_{i\in\mathcal{I}}$, such that $b_i \geq 0$ for all $i \in \mathcal{I}$ and $\sum_{i\in\mathcal{I}} b_i < \infty$ such that for all $n, i : |a_i(n)| \leq b_i$.*

Then $\sum_{i\in\mathcal{I}} |a_i| < \infty$ and

$$\sum_{i\in\mathcal{I}} a_i = \sum_{i\in\mathcal{I}} \lim_{n\to\infty} a_i(n) = \lim_{n\to\infty} \sum_{i\in\mathcal{I}} a_i(n).$$

Here we have $a_i(n) = \pi_i p_{ij}(n)$. Clearly, $\sum_i a_i(n)$ is absolutely convergent for all n since $\sum_i |\pi_i p_{ij}(n)| = \sum_i \pi_i p_{ij}(n) = \pi_j \leq 1 < \infty$.
Also $\lim_{n\to\infty} a_i(n) = 0 =: a_i$ for all i.
Next, $|a_i(n)| = \pi_i p_{ij}(n) \leq \pi_i =: b_i \geq 0$ and $\sum_i b_i = \sum_i \pi_i = 1 < \infty$. Applying Theorem 2.8.2 concludes the proof.
Now, we show that the existence of $\boldsymbol{\pi}$ implies that all states are *positive* (recurrent) and that $\pi_i = \mu_i^{-1}$ for each i. Suppose that $X_0 \sim \boldsymbol{\pi}$ (i.e. $\mathrm{P}(X_0 = i) = \pi_i$ for each i), using Lemma 3.9.9,

$$\begin{aligned}\pi_j\mu_j &= \mathrm{P}(X_0 = j)\mathrm{E}(T_j|X_0 = j) = \sum_{n=1}^{\infty} \mathrm{P}(T_j \geq n|X_0 = j)\mathrm{P}(X_0 = j)\\ &= \sum_{n=1}^{\infty} \mathrm{P}(T_j \geq n, X_0 = j).\end{aligned}$$

But, $\mathrm{P}(T_j \geq 1, X_0 = j) = \mathrm{P}(X_0 = j)$ (since $T_j \geq 1$ by definition). Then for $n \geq 2$, we have

$$\begin{aligned}&\mathrm{P}(T_j \geq n, X_0 = j) = \mathrm{P}(X_0 = j, X_m \neq j, 1 \leq m \leq n-1)\\ &= \mathrm{P}(X_m \neq j, 1 \leq m \leq n-1) - \mathrm{P}(X_m \neq j, 0 \leq m \leq n-1)\\ &= \mathrm{P}(X_m \neq j, 0 \leq m \leq n-2) - \mathrm{P}(X_m \neq j, 0 \leq m \leq n-1)\\ &= a_{n-2} - a_{n-1},\end{aligned}$$

where we have used homogeneity and define

$$a_n = \mathrm{P}(X_m \neq j, 0 \leq m \leq n).$$

Note that we have used the law of total probability again:

$$\mathrm{P}(X_m \neq j, 1 \leq m \leq n-1) = \mathrm{P}(X_0 = j, X_m \neq j, 1 \leq m \leq n-1) + \mathrm{P}(X_0 \neq j, X_m \neq j, 1 \leq m \leq n-1),$$

hence

$$\mathrm{P}(X_0 = j, X_m \neq j, 1 \leq m \leq n-1) = \mathrm{P}(X_m \neq j, 1 \leq m \leq n-1) - \mathrm{P}(X_0 \neq j, X_m \neq j, 1 \leq m \leq n-1).$$

Then, summing over n (telescoping sum!)

$$\begin{aligned}\pi_j \mu_j &= \mathrm{P}(X_0 = j) + \sum_{n=2}^{\infty} (a_{n-2} - a_{n-1}) \\ &= \mathrm{P}(X_0 = j) + \mathrm{P}(X_0 \neq j) - \lim_{n\to\infty} a_n \\ &= 1 - \lim_{n\to\infty} a_n.\end{aligned}$$

However,

$$\lim_{n\to\infty} a_n = \mathrm{P}(X_m \neq j, \ \forall m) = 0$$

by recurrence of j. That is, $\pi_j^{-1} = \mu_j$ if $\pi_j > 0$.

To see that $\pi_j > 0$ for all j, suppose the converse; then

$$0 = \pi_j = \sum_{i \in E} \pi_i p_{ij}(n) \geq \pi_i p_{ij}(n), \tag{3.9.9}$$

for all i, n, yielding that $\pi_i = 0$ whenever $i \to j$. However, the chain is irreducible, so that $\pi_i = 0$ for each i - a contradiction to the fact that $\boldsymbol{\pi}$ is a stationary vector. Thus $\mu_i < \infty$ and all states are positive.

To finish, if $\boldsymbol{\pi}$ exists then it is unique and all states are positive recurrent. Conversely, if the states of the chain are positive recurrent then the chain has a stationary distribution from Lemma 3.9.6. □

Remark 3.9.10. *Note that Theorem 3.9.2 provides a very useful criterion for checking whether an irreducible chain is positive recurrent: You just have to look for a stationary distribution!*

Note that the stationary distribution is a left eigenvector of the transition matrix.

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}.$$

Example 3.9.11. *Let $E = \{1, 2\}$ and the transition matrix is given by*

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{pmatrix}.$$

Find the stationary distribution. We need to find a solution to the following equation

$$(\pi_1, \pi_2) = (\pi_1, \pi_2)\mathbf{P}.$$

We get the following system of equations:

$$\pi_1 = \frac{1}{2}\pi_1 + \frac{1}{4}\pi_2, \qquad \pi_2 = \frac{1}{2}\pi_1 + \frac{3}{4}\pi_2.$$

Also, we can use the 'probability condition' that $\pi_1 + \pi_2 = 1$. We obtain $(\pi_1, \pi_2) = \left(\frac{1}{3}, \frac{2}{3}\right)$. Compute the mean recurrent times. $\mu_1 = 1/\pi_1 = 3, \qquad \mu_2 = 1/\pi_2 = 3/2$.

3.9.2 Limiting distribution

Definition 3.9.12. *A distribution $\boldsymbol{\pi}$ is the **limiting distribution** of a discrete-time Markov chain if, for all states $i, j \in E$, we have*

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

An important question to answer is, when is the limiting distribution (as the time parameter goes to ∞) also the stationary distribution?

A problem arises when the chain is periodic:

Example 3.9.13. *Let $E = \{1, 2\}$ and transition matrix $\mathbf{P}$ with $p_{12} = p_{21} = 1, p_{11} = p_{22} = 0$. Then the Markov chain is irreducible and periodic with $d(1) = d(2) = 2$. Then we get that*

$$\mathbf{P}^{2n} = I_{2\times 2}, \qquad \mathbf{P}^{2n+1} = \mathbf{P}, \qquad \forall n \in \mathbb{N}_0.$$

Hence $\mathbf{P}^n$ does not converge as $n \to \infty$. So this Markov chain does not have a limiting distribution. However, it has a unique stationary distribution given by $\boldsymbol{\pi} = (0.5, 0.5)$.

Theorem 3.9.14. *For an irreducible aperiodic chain we have*

$$\lim_{n\to\infty} p_{ij}(n) = \frac{1}{\mu_j}.$$

The proof is quite long and can be seen in Grimmett & Stirzaker (2001*b*, p. 232–235). We conclude this section with some important remarks:

Remark 3.9.15.

- *If the irreducible chain is transient or null recurrent, then $\mu_j = \infty$ and*

$$\lim_{n\to\infty} p_{ij}(n) = 0, \quad \forall\, i, j \in E.$$

- *If the chain is irreducible aperiodic and positive recurrent then we have:*

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j = \frac{1}{\mu_j} \quad \forall i, j \in E,$$

  where $\boldsymbol{\pi}$ is the unique stationary distribution.

- *The limiting distribution of an irreducible aperiodic chain does not depend on the starting point $(X_0 = i)$/the initial distribution $\boldsymbol{\nu}^{(0)}$, but forgets its origin, hence we have by Theorems 3.3.3 and 2.8.2,*

$$\begin{aligned}\lim_{n\to\infty} \mathrm{P}(X_n = j) &= \lim_{n\to\infty} \nu_j^{(n)} \\ &= \lim_{n\to\infty} \sum_{i\in E} \mathrm{P}(X_0 = i) p_{ij}(n) = \lim_{n\to\infty} \sum_{i\in E} \nu_i^{(0)} p_{ij}(n) \\ &\overset{\text{DOM}}{=} \sum_{i\in E} \nu_i^{(0)} \lim_{n\to\infty} p_{ij}(n) = \sum_{i\in E} \nu_i^{(0)} \frac{1}{\mu_j} = \frac{1}{\mu_j}.\end{aligned}$$

3.9.3 Ergodic theorem

Now we formulate the ergodic theorem which is concerned with limiting behaviour of averages over time.

Theorem 3.9.16 (Ergodic Theorem)**.** *Suppose we are given an irreducible Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$ with state space E. Let μ_i denote the mean recurrence time to state $i \in E$ and let*

$$V_i(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k = i\}},$$

denote the number of visits to i before n. *Then* $V_i(n)/n$ *denotes the proportion of time before* n *spent in state* i. *Then*

$$\mathrm{P}\left(\frac{V_i(n)}{n} \to \frac{1}{\mu_i}, \qquad \text{as } n \to \infty\right) = 1.$$

See Norris (1998, Chapter 1.10) for a proof.

Altogether, we get the following results. If the chain is irreducible and positive recurrent, then $V_i(n)/n \to \pi_i$ (the unique stationary distribution) as $n \to \infty$. If it is irreducible and null recurrent or transient, we have $V_i(n)/n \to 0$ as $n \to \infty$.

Summary: Properties of irreducible Markov chains

There are three kinds of **irreducible** Markov chains:

1. **Positive recurrent**
   (a) Stationary distribution $\boldsymbol{\pi}$ exists.
   (b) Stationary distribution is unique.
   (c) All mean recurrence times are finite and $\mu_i = \frac{1}{\pi_i}$
   (d) $V_i(n)/n \to \pi_i$ as $(n \to \infty)$, where $V_i(n)/n$ denotes the proportion of time before n spent in state i.
   (e) If the chain is aperiodic, then
   $$\lim_{n\to\infty} \mathrm{P}(X_n = i) = \pi_i, \quad \forall\, i \in E.$$
2. **Null recurrent**
   (a) Recurrent, but all mean recurrence times are infinite.
   (b) No stationary distribution exists.
   (c) $V_i(n)/n \to 0$ as $(n \to \infty)$
   (d)
   $$\lim_{n\to\infty} \mathrm{P}(X_n = i) = 0, \quad \forall\, i \in E.$$
3. **Transient**
   (a) Any particular state is eventually never visited.
   (b) No stationary distribution exists.
   (c) $V_i(n)/n \to 0$ as $(n \to \infty)$
   (d)
   $$\lim_{n\to\infty} \mathrm{P}(X_n = i) = 0, \quad \forall\, i \in E.$$

3.9.4 Properties of the elements of a stationary distribution associated with transient or null-recurrent states

Before moving on to Markov chains on a finite state space, we formulate an important result which can be derived using the same arguments as in the proof of Theorem 3.9.2, where we applied the dominated convergence theorem.

Theorem 3.9.17. *Let* X *denote a time-homogeneous Markov chain on a countable state space* E. *If* $\boldsymbol{\pi}$ *is a stationary distribution of this Markov chain and a state* $i \in E$ *is either transient or null-recurrent, then* $\pi_i = 0$.

We note that the theorem above does not claim the existence of a stationary distribution for general Markov chains, but states that if there is one (or possibly many) stationary distribution(s) of the Markov chain, then the elements of the stationary distribution associated with transient and null-recurrent states are equal to 0.

This is a very useful result, since it means that, in practice, when computing stationary distributions, we can focus on the states which are positive recurrent and set the remaining elements of the stationary distribution to 0 (provided it exists).

Proof. Suppose that $\boldsymbol{\pi}$ is a stationary distribution of the chain. Assume that state $j \in E$ is either transient or null-recurrent. Then $p_{ij}(n) \to 0$, as $n \to \infty$, for all i, by Corollary 3.5.6 and Theorem 3.5.11. Since $\boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi}$, we have

$$\pi_j = \lim_{n\to\infty} \pi_j = \lim_{n\to\infty} \sum_{i\in E} \pi_i p_{ij}(n) \overset{\text{DOM}}{=} \sum_{i\in E} \pi_i \lim_{n\to\infty} p_{ij}(n) = 0.$$

This follows from switching the order of summation and limits using the dominated convergence theorem (DOM, Theorem 2.8.2). For the DOM, note that we can set $a_i(n) = \pi_i p_{ij}(n)$. Clearly, $\sum_i a_i(n)$ is absolutely convergent for all n since $\sum_i |\pi_i p_{ij}(n)| = \sum_i \pi_i p_{ij}(n) = \pi_j \leq 1 < \infty$. Also $\lim_{n\to\infty} a_i(n) = 0 =: a_i$ for all i. Next, $|a_i(n)| = \pi_i p_{ij}(n) \leq \pi_i =: b_i \geq 0$ and $\sum_i b_i = \sum_i \pi_i = 1 < \infty$.

□

Remark 3.9.18. *We note that in the case of a finite state-space, the proof of Theorem 3.9.17 simplifies. In that case, null-recurrent states do not exist, so only transient states need to be considered. We are allowed to interchange the limit and the finite sum in the above proof, without the need to appeal to the dominated convergence theorem.*

End of lecture 9. .....

3.9.5 Existence of a stationary distribution on a finite state space

So far, we have studied general Markov chains in discrete time with a countable state space. As mentioned before, the state space E could be any countably infinite set, e.g. $E = \mathbb{N}$. There are however, many examples where the state space is indeed *finite*, i.e. $K = \text{card}(E) = |E| < \infty$, e.g. $E = \{1, 2, 3\}$ or $E = \{\text{sunny, rainy}\}$ etc.. In that case, the theory simplifies and we obtain some very nice results which are very useful in applications.

The main results we will establish in this subsection are the following ones:

Existence: A discrete-time Markov chain on a finite state space always has (at least) one stationary distribution.

Uniqueness: Every Markov chain with a finite state space has a unique stationary distribution unless the chain has two or more closed communicating classes.

We already proved that on a finite state space there is at least one positive recurrent class.

Theorem 3.9.19. *If the state space is finite, then there is at least one positive recurrent communicating class.*

Proof. This is an immediate consequence of Theorems 3.7.11, 3.7.8, 3.7.12.

□

We can now formulate an important result.

Theorem 3.9.20. *Suppose we have a finite state space. The stationary distribution $\boldsymbol{\pi}$ for a transition matrix $\mathbf{P}$ is unique if and only if there is a unique closed communicating class.*

Proof of Theorem 3.9.20. First suppose there is a unique closed communicating class C. Write $\mathbf{P}(C)$ for the restriction of the matrix $\mathbf{P}$ to the states in C. From Theorem 3.7.6, we concluded that $\mathbf{P}(C)$ is a stochastic matrix, and from Theorem 3.7.11 we obtain that all states in C are positive recurrent, hence we can apply Theorem 3.9.2 to conclude that there exists a unique stationary distribution $\boldsymbol{\pi}^C_{1:|C|}$ for $\mathbf{P}(C)$. Let $\boldsymbol{\pi}$ be a stationary distribution on E satisfying $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. Theorem 3.9.17 tells us that $\pi_i = 0$ whenever $i \notin C$. Hence $\boldsymbol{\pi}$ is supported on C only.

Consequently, for $i \in C$, we have

$$\pi_i = \sum_{j\in E} \pi_j p_{ji} = \sum_{j\in C} \pi_j p_{ji} = \sum_{j\in C} \pi_j p^C_{ji},$$

where $\boldsymbol{\pi}$ restricted to C is stationary for $\mathbf{P}(C)$. Since the stationary distribution for $\mathbf{P}(C)$ is unique, we get $\pi_i = \pi^C_i$ for all $i \in C$. Altogether we have

$$\pi_i = \begin{cases} \pi^C_i, & \text{if } i \in C, \\ 0, & \text{if } i \notin C. \end{cases}, \tag{3.9.10}$$

and the solution $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ is unique.

Now suppose that there is a unique stationary distribution and *two* distinct closed communicating classes for $\mathbf{P}$, say C_1 and C_2. Clearly, the restriction of $\mathbf{P}$ to each of these classes is irreducible. Therefore, for each $i = 1, 2$ there exists a distribution $\boldsymbol{\pi}^{(i)}$ supported on C_i which is stationary for $\mathbf{P}(C_i)$. We can see (check it!) that each $\boldsymbol{\pi}^{(i)}$ is stationary for $\mathbf{P}$. Hence $\mathbf{P}$ has a stationary distribution, but it is not unique. □

The arguments presented in the above proof, allow us to deduce the following useful result. Consider a Markov chain with a finite state space and at least two closed classes. Then every stationary distribution can be represented as a linear combination of the stationary distributions associated with the closed classes and extended to the whole space.

Corollary 3.9.21. *Consider a Markov chain with a finite state space and $N \geq 2$ closed classes. Let C_i denote the closed classes of the Markov chain and we denote by $\boldsymbol{\pi}^{(i)}$ the stationary distribution associated with class C_i using the construction*

$$\pi^{(i)}_j = \begin{cases} \pi^{C_i}_j, & \textit{if } j \in C_i, \\ 0, & \textit{if } j \notin C_i. \end{cases}, \tag{3.9.11}$$

Then every stationary distribution of the Markov chain can be represented as

$$\sum_{i=1}^{N} w_i \boldsymbol{\pi}^{(i)},$$

for weights $w_i \geq 0$, $\sum_{i=1}^n w_i = 1$.

Finding the stationary distributions for Markov chains with a finite state space

Suppose we have a Markov chain with a **finite state space**. Then:

- A stationary distribution always exists.
- The stationary distribution is unique. $\Leftrightarrow$ There is a unique closed communicating class. $\Leftrightarrow$ There is a unique positive recurrent communicating class.

If you need to find a stationary distribution, proceed as follows:

- Find all (N, say) closed communicating classes C_i (e.g. by looking at the transition diagram or by examining the transition matrix $\mathbf{P}$).

- For each closed communicating class C_i, $i = 1, 2, \ldots$, you need to solve a system of equations. I.e. if C_i is such a closed communicating class, let $\boldsymbol{\pi}^{C_i}$ denote a $\mathrm{card}(C_i)$-dimensional row vector with non-negative entries. Solve $\boldsymbol{\pi}^{C_i}\mathbf{P}(C_i) = \boldsymbol{\pi}^{C_i}$ such that all elements of $\boldsymbol{\pi}^{C_i}$ are non-negative and sum up to 1.
- One possible stationary distribution is then given by the row vector $\boldsymbol{\pi}^{(i)}$ which consists of the corresponding elements of the vectors $\boldsymbol{\pi}^{C_i}$ and of zeros corresponding to the transient states, see the construction (3.9.11). You need to be careful to get the order of the elements right. We often study "nicely blocked Markov chains" in this course, but that does not need to be the case in a real application!
- In a final step, you can represent all possible stationary distributions by

$$\sum_{i=1}^{N} w_i \boldsymbol{\pi}^{(i)},$$

  for weights $w_i \geq 0$, $\sum_{i=1}^{n} w_i = 1$.
- You might also want to check, that if you only found *one* closed class, the above conditions should lead to a *unique* stationary distribution. If you still have some free parameters, then there has to be a mistake in your calculations!

As homework, try using the strategy described above to answer the following exam question.

Exercise 3.9.22 (Exam question 2020)**.** *Consider a discrete-time homogeneous Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$ with state space $E = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and transition matrix given by*

$$\mathbf{P} = \begin{pmatrix} \frac{1}{4} & 0 & \frac{3}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

1. *Draw the transition diagram.*
2. *Specify the communicating classes and determine whether they are transient, null recurrent or positive recurrent. Please note that you need to justify your answers.*
3. *Find all stationary distributions.*

Solution. 1. The transition diagram is given by

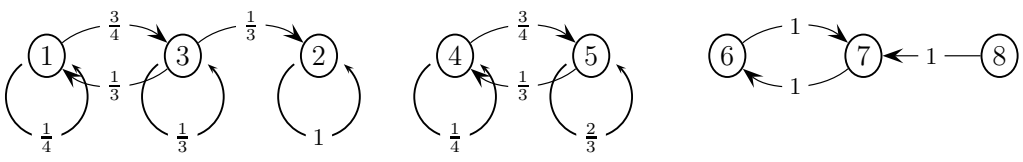


Figure 3.11: Transition diagram for the Markov chain described in Exercise 3.9.22

2. We have a finite state space which can be divided into five communicating classes: The classes $T_1 = \{1, 3\}, T_2 = \{8\}$ are not closed and hence transient.

   The classes $C_1 = \{2\}, C_2 = \{4, 5\}, C_3 = \{6, 7\}$ are finite and closed and hence positive recurrent.

3. This Markov chain does not have a unique stationary distribution $\boldsymbol{\pi}$ since we have three closed communicating classes. For the transient states we know from the lectures that $\pi_i = 0$ for $i = 1, 3, 8$.
For the positive recurrent states, we solve $\pi_2 \cdot 1 = \pi_2$, $(\pi_4, \pi_5) = (\pi_4, \pi_5) \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$ and $(\pi_6, \pi_7) = (\pi_6, \pi_7) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, which leads to $\pi_2 = \pi_2$, $\pi_5 = \frac{9}{4}\pi_4$ and $\pi_6 = \pi_7$.
There are various ways of representing all possible stationary distributions (only one is needed!), e.g.:
   - $\boldsymbol{\pi} = (0, \pi_2, 0, \pi_4, \frac{9}{4}\pi_4, \pi_6, \pi_6, 0)$ for all $\pi_2, \pi_4, \pi_6 \geq 0$ with $\pi_2 + \frac{13}{4}\pi_4 + 2\pi_6 = 1$,
   - $\boldsymbol{\pi} = (0, \pi_2, 0, \frac{4}{9}\pi_5, \pi_5, \pi_6, \pi_6, 0)$ for all $\pi_2, \pi_5, \pi_6 \geq 0$ with $\pi_2 + \frac{13}{9}\pi_5 + 2\pi_6 = 1$,
   - $\boldsymbol{\pi} = a(0, 1, 0, 0, 0, 0, 0, 0) + b(0, 0, 0, \frac{4}{13}, \frac{9}{13}, 0, 0, 0) + c(0, 0, 0, 0, 0, \frac{1}{2}, \frac{1}{2}, 0)$ for all $a, b, c \geq 0$ with $a + b + c = 1$.

□

3.9.6 Limiting distributions on a finite state space

We have already discussed that limiting distributions might not always exist. We can show, however, that if there is a limiting distribution on a finite state space, then the limiting distribution is a also a stationary distribution.

Theorem 3.9.23. *Let $K = |E| < \infty$. Suppose for some $i \in E$ that*

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j, \quad \forall j \in E.$$

Then $\boldsymbol{\pi}$ is a stationary distribution.

Proof. Homework, see Exercise 2- 20. □

End of lecture 10. .....

3.10 Time reversibility

An interesting concept in the study of Markov chains is that of time reversibility. The idea is to reverse the time scale of the Markov chain; such a concept, as we will see in a moment, is very useful for constructing Markov chains with a pre-specified stationary distribution. This is important, for example for Markov chain Monte Carlo (MCMC) algorithms.

Define an irreducible, positive recurrent Markov chain $\{X_n\}_{n\in\{0,1,\ldots,N\}}$ for an $N \in \mathbb{N}$. We assume that $\boldsymbol{\pi}$ is the stationary distribution, and $\mathbf{P}$ is the transition matrix, and that for any $n \in \{0, 1, \ldots, N\}$ the marginal distribution $\boldsymbol{\nu}^{(n)}$ is equal to $\boldsymbol{\pi}$. The reversed chain is defined to be, for any $n \in \{0, 1, \ldots, N\}$

$$Y_n = X_{N-n}.$$

Theorem 3.10.1. *The sequence Y is a Markov chain which satisfies*

$$\mathrm{P}(Y_{n+1} = j | Y_n = i) = \frac{\pi_j}{\pi_i} p_{ji}.$$

Proof.

$$\begin{aligned} &\mathrm{P}(Y_{n+1} = i_{n+1} | Y_n = i_n, Y_{n-1} = i_{n-1}, \ldots, Y_0 = i_0) \\ &= \frac{\mathrm{P}(Y_k = i_k, 0 \leq k \leq n+1)}{\mathrm{P}(Y_k = i_k, 0 \leq k \leq n)} \end{aligned}$$

$$= \frac{\mathrm{P}(X_{N-k} = i_k, 0 \leq k \leq n+1)}{\mathrm{P}(X_{N-k} = i_k, 0 \leq k \leq n)}.$$

Now we apply Bayes theorem and the Markov property to deduce that

$$\begin{aligned}
&\mathrm{P}(X_{N-k} = i_k, 0 \leq k \leq n+1) \\
&= \mathrm{P}(X_N = i_0 | X_{N-k} = i_k, 1 \leq k \leq n+1)\mathrm{P}(X_{N-k} = i_k, 1 \leq k \leq n+1) \\
&= \mathrm{P}(X_N = i_0 | X_{N-1} = i_1)\mathrm{P}(X_{N-k} = i_k, 1 \leq k \leq n+1) \\
&= \mathrm{P}(X_N = i_0 | X_{N-1} = i_1)\mathrm{P}(X_{N-1} = i_1 | X_{N-2} = i_2) \cdots \mathrm{P}(X_{N-n} = i_n | X_{N-n-1} = i_{n+1}) \\
&\quad \mathrm{P}(X_{N-n-1} = i_{n+1}) \\
&= \pi_{i_{n+1}} p_{i_{n+1} i_n} \cdots p_{i_1 i_0}.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathrm{P}(Y_{n+1} = i_{n+1} | Y_n = i_n, Y_{n-1} = i_{n-1}, \ldots, Y_0 = i_0) &= \frac{\pi_{i_{n+1}} p_{i_{n+1} i_n} \cdots p_{i_1 i_0}}{\pi_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0}} \\
&= \frac{\pi_{i_{n+1}} p_{i_{n+1} i_n}}{\pi_{i_n}}.
\end{aligned}$$

Similarly, we get that

$$\begin{aligned}
&\mathrm{P}(Y_{n+1} = i_{n+1} | Y_n = i_n) = \frac{\mathrm{P}(Y_{n+1} = i_{n+1}, Y_n = i_n)}{\mathrm{P}(Y_n = i_n)} = \frac{\mathrm{P}(X_{N-n-1} = i_{n+1}, X_{N-n} = i_n)}{\mathrm{P}(X_{N-n} = i_n)} \\
&= \frac{\mathrm{P}(X_{N-n} = i_n | X_{N-n-1} = i_{n+1})\mathrm{P}(X_{N-n-1} = i_{n+1})}{\mathrm{P}(X_{N-n} = i_n)} = \frac{\pi_{i_{n+1}} p_{i_{n+1} i_n}}{\pi_{i_n}}.
\end{aligned}$$

So overall we have shown that for any $n \in \mathbb{N}$ and for any states $i_0, \ldots, i_{n+1} \in E$ we have that

$$\begin{aligned}
&\mathrm{P}(Y_{n+1} = i_{n+1} | Y_n = i_n, Y_{n-1} = i_{n-1}, \ldots, Y_0 = i_0) = \mathrm{P}(Y_{n+1} = i_{n+1} | Y_n = i_n) \\
&= \frac{\pi_{i_{n+1}} p_{i_{n+1} i_n}}{\pi_{i_n}},
\end{aligned}$$

which completes the proof.

□

Definition 3.10.2. *Let $X = \{X_n : n \in \{0, 1, \ldots, N\}\}$ be an irreducible Markov chain with stationary distribution $\boldsymbol{\pi}$ and the marginal distributions are given by $\boldsymbol{\nu}^{(n)} = \boldsymbol{\pi}$ for all $n \in \{0, 1, \ldots, N\}$. The Markov chain X is called **time-reversible** if the transition matrices of X and its time-reversal Y are the same.*

Theorem 3.10.3. *$\{X_n\}_{n \in \{0,1,\ldots,N\}}$ is time–reversible if and only if for any $i, j \in E$*

$$\pi_i p_{ij} = \pi_j p_{ji}. \tag{3.10.1}$$

Note that the condition (3.10.1) is often referred to as **detailed–balance**.

Proof. Let Q be the transition matrix of $\{Y_n\}_{n \in \{0,1,\ldots,N\}}$. Then from the above arguments, we have

$$q_{ij} = p_{ji} \frac{\pi_j}{\pi_i}$$

thus $q_{ij} = p_{ij}$ iff (3.10.1) holds. □

Theorem 3.10.4. *For an irreducible chain, if there exist a probability vector $\boldsymbol{\pi}$ such that (3.10.1) holds, for any $i, j \in E$, then the chain is time–reversible (once it is in its stationary regime) and positive recurrent, with stationary distribution $\boldsymbol{\pi}$.*

Proof. Given the detailed balance condition and any $j \in E$, we have

$$\sum_{i \in E} \pi_i p_{ij} = \sum_{i \in E} \pi_j p_{ji} = \pi_j \sum_{i \in E} p_{ji} = \pi_j$$

thus $\boldsymbol{\pi}$ is stationary. The remainder of the result follows from Theorem 3.9.2. □

Essentially, the result tells us, if we want to construct a chain with stationary distribution $\boldsymbol{\pi}$, then one way is through the detailed balance condition.

Remark 3.10.5. *Note that it is possible to extend the definition of time reversibility to an infinite time set* $\{0, 1, 2, \dots\}$*, or even to a doubly–infinite time set* $\{\dots, -2, -1, 0, 1, 2, \dots\}$.

Exercise 3.10.6. *Let* $\{X_n\}_{n \in \mathbb{N}_0}$ *denote a Markov chain with state space* $E = \{1, 2, 3\}$ *with transition matrix*

$$\mathbf{P} = \begin{pmatrix} 0 & p & 1-p \\ 1-p & 0 & p \\ p & 1-p & 0 \end{pmatrix}, \qquad \textit{for } 0 < p < 1.$$

Is the Markov chain reversible?

Solution. The Markov chain is irreducible, with finite state space. Hence there is a unique stationary distribution, which is given by $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$. [Since the transition matrix is doubly-stochastic, the uniform distribution is the stationary distribution.]

Now we check the detailed balance equations: $\pi_i p_{ij} = \pi_j p_{ji}$. Here we need $\frac{1}{3} p_{ij} = \frac{1}{3} p_{ji}$, i.e. $p_{ij} = p_{ji}$, for any $i, j \in \{1, 2, 3\}$. These equations only hold if and only if $p = 1 - p \Leftrightarrow p = 1/2$. So, the chain is reversible if and only if $p = 1/2$. □

Chapter 4

Properties of the exponential distribution

In this chapter we will discuss various important properties of the exponential distribution which will play a central role in our study of Poisson processes and, more generally, continuous-time Markov chains.

4.1 Definition and basic properties

Let us briefly recall the definition of the exponential distribution.

Definition 4.1.1 (Exponential distribution)**.** *A continuous random variable X is said to have the* exponential distribution *with parameter $\lambda > 0$, i.e. $X \sim \mathrm{Exp}(\lambda)$, if its density function is given by*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Its cumulative distribution function is given by

$$F_X(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1 - e^{-\lambda x}, & \text{if } x > 0. \end{cases}$$

We observe that the so-called survival function of the exponential distribution is given by

$$\mathrm{P}(X > x) = \begin{cases} 1, & \text{if } x \leq 0, \\ e^{-\lambda x}, & \text{if } x > 0. \end{cases}$$

The probability density function and cumulative distribution function of an exponential variable with various choices of the rate parameter λ are depicted in Figure 4.1.

Theorem 4.1.2. *Let $X \sim \mathrm{Exp}(\lambda)$ for $\lambda > 0$. Then*

1. $\mathrm{E}(X) = \frac{1}{\lambda}$.
2. $\lambda X \sim \mathrm{Exp}(1)$.

Proof. 1. Using integration by parts or the Gamma function, we deduce that

$$\mathrm{E}(X) = \int_0^\infty x\lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^\infty x\lambda e^{-\lambda x} \lambda dx = \frac{1}{\lambda}\Gamma(2) = \frac{1}{\lambda}.$$

2. Let $x < 0$, then $\mathrm{P}(\lambda X \leq x) = 0$, for $x \geq 0$, we have

$$\mathrm{P}(\lambda X \leq x) = \mathrm{P}(X \leq x/\lambda) = F_X(x/\lambda) = 1 - e^{-x},$$

which is the cdf of an $\mathrm{Exp}(1)$ random variable.

□

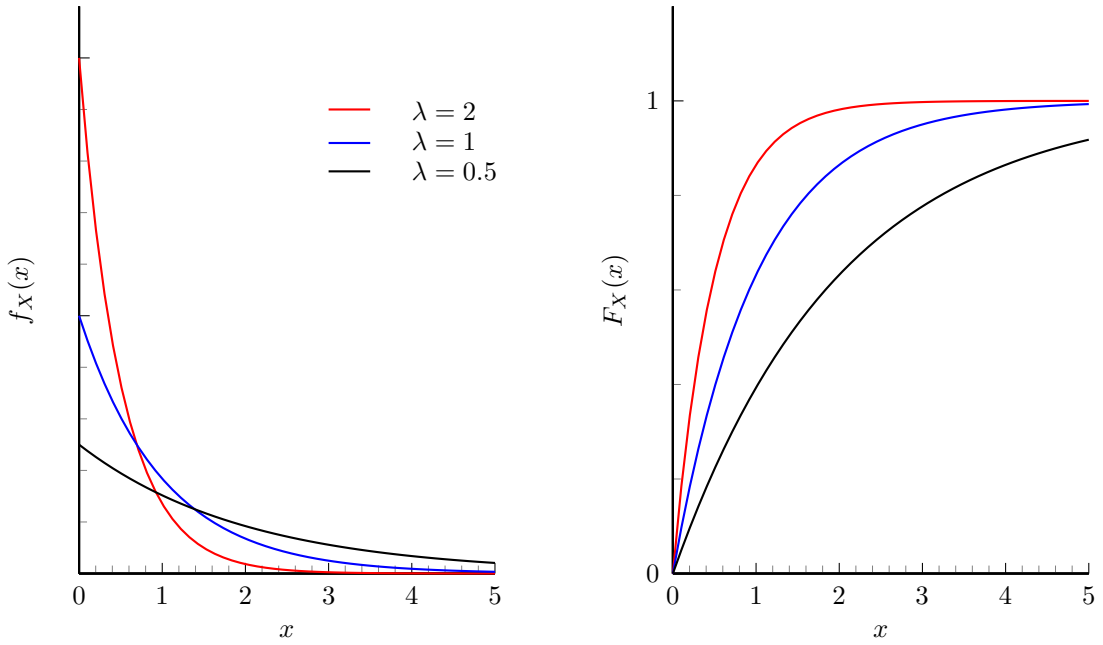


Figure 4.1: Plot of the p.d.f. (left) and the c.d.f. (right) of a random variable $X \sim \text{Exp}(\lambda)$ for $\lambda \in \{0.5, 1, 2\}$.

Theorem 4.1.3. *Let $n \in \mathbb{N}$ and $\lambda > 0$. Consider independent and identically distributed random variables $H_i \sim \text{Exp}(\lambda)$ for $i = 1, \ldots, n$. Let $J_n := \sum_{i=1}^n H_i$. Then J_n follows the $\text{Gamma}(n, \lambda)$ distribution, i.e.*

$$f_{J_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, \quad t > 0.$$

Proof. There are a few ways to prove this, but the simplest, is to use the uniqueness of the Laplace transform.

Let $u > 0$. By definition

$$\text{E}(e^{-uJ_n}) = \text{E}(e^{-u\sum_{i=1}^n H_i}) \overset{\text{independence}}{=} \prod_{i=1}^n \text{E}(e^{-uH_i}) \overset{\text{identical distribution}}{=} [\text{E}(e^{-uH_1})]^n.$$

For $u > 0$, we have

$$\text{E}(e^{-uH_1}) = \int_0^\infty e^{-ux} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda + u},$$

and hence

$$\text{E}(e^{-uJ_n}) = \left(\frac{\lambda}{\lambda + u}\right)^n,$$

which is the Laplace transform of a $\text{Gamma}(n, \lambda)$ random variable.

To see this, note that if $Y \sim \text{Gamma}(n, \lambda)$, then for $u > 0$,

$$\text{E}(e^{-uY}) = \int_0^\infty e^{-uy} \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy$$

$$
\begin{aligned}
&= \frac{\lambda^n}{\Gamma(n)} \int_0^\infty e^{-(u+\lambda)y} y^{n-1} dy \\
&= \frac{\lambda^n}{\Gamma(n)} \frac{1}{(u+\lambda)^n} \int_0^\infty e^{-(u+\lambda)y} [(u+\lambda)y]^{n-1} (u+\lambda) dy \\
&= \frac{\lambda^n}{\Gamma(n)} \frac{1}{(u+\lambda)^n} \int_0^\infty e^{-z} z^{n-1} dz \\
&= \frac{\lambda^n}{\Gamma(n)} \frac{1}{(u+\lambda)^n} \Gamma(n) = \frac{\lambda^n}{(u+\lambda)^n}.
\end{aligned}
$$

□

End of lecture 11.

We state some additional useful properties of the exponential distribution, which are fundamental for continuous-time Markov chains.

Theorem 4.1.4. *Let $n \in \mathbb{N}$ and $\lambda_1, \dots, \lambda_n > 0$. Consider independent random variables $H_i \sim \text{Exp}(\lambda_i)$ for $i = 1, \dots, n$. Let $H := \min\{H_1, \dots, H_n\}$. Then*

1. *$H \sim \text{Exp}(\sum_{i=1}^n \lambda_i)$.*
2. *For any $k = 1, \dots, n$, $\text{P}(H = H_k) = \lambda_k / (\sum_{i=1}^n \lambda_i)$.*

Proof. See Exercise 3- 22 on the problem sheet. □

The result in Theorem 4.1.4 can be strengthened to the case of countably many random variables, see Norris (1998, p.72).

Theorem 4.1.5. *Consider a countable index set E (later this will be the state space) and $\{H_i : i \in E\}$ independent random variables with $H_i \sim \text{Exp}(\lambda_i)$ for all $i \in E$. Suppose that $\sum_{i \in E} \lambda_i < \infty$ and set $H := \inf_{i \in E} H_i$.*

Then the infimum is attained at a unique random value I of E with probability 1. Moreover, H and I are independent, with $H \sim \text{Exp}(\sum_{i \in E} \lambda_i)$ and $\text{P}(I = i) = \lambda_i / \sum_{k \in E} \lambda_k$.

Proof. We set $I = i$ if $H_i < H_j$ for all $j \neq i$, otherwise let I be undefined. Then, using the continuous law of total probability, see (2.7.2),

$$
\begin{aligned}
&\text{P}(I = i, H \geq y) = \text{P}(H_i \geq y, H_j > H_i \, \forall j \neq i) \\
&= \int_0^\infty \text{P}(H_i \geq y, H_j > H_i \, \forall j \neq i | H_i = x_i) f_{H_i}(x_i) dx_i \\
&\overset{(H_i) \text{ indep.}}{=} \int_y^\infty \text{P}(H_j > x_i \, \forall j \neq i) f_{H_i}(x_i) dx_i \\
&\overset{(H_i) \text{ indep.}}{=} \int_y^\infty \prod_{j \in E, j \neq i} e^{-\lambda_j x_i} \lambda_i e^{-\lambda_i x_i} dx_i \\
&= \frac{\lambda_i}{\sum_{k \in E} \lambda_k} e^{-\sum_{k \in E} \lambda_k y}.
\end{aligned}
$$

Hence $\text{P}(I = i \text{ for some } i) = \sum_{i=1}^\infty \frac{\lambda_i}{\sum_{k \in E} \lambda_k} = 1$ and H and I have the claimed joint distribution. □

Exercise 4.1.6. *Suppose that blue and red cars arrive at a petrol station. Let X denote the waiting time to the arrival of the next red car and assume that $X \sim \text{Exp}(\lambda_X)$. Also, let Y denote the waiting time to the arrival of the next blue car and assume that $Y \sim \text{Exp}(\lambda_Y)$. What is the probability that a red car arrives before a blue?*

Solution to Exercise 4.1.6. Using the (continuous) law of total probability:

$$\begin{aligned}
\mathrm{P}(X<Y) &= \int_{-\infty}^{\infty} \mathrm{P}(X<Y|Y=y) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} \mathrm{P}(X<y|Y=y) f_Y(y) dy \\
&\overset{X,Y \text{ independent}}{=} \int_{-\infty}^{\infty} \mathrm{P}(X<y) f_Y(y) dy \\
&= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy \\
&= \int_0^{\infty} \left[\int_0^y \lambda_X e^{-\lambda_X x} dx \right] \lambda_Y e^{-\lambda_Y y} dy \\
&= \frac{\lambda_X}{\lambda_X + \lambda_Y}.
\end{aligned}$$

□

Remark 4.1.7. *Note that, rather than doing the computations in Exercise 4.1.6 above, we could have applied Theorem 4.1.4 to conclude that*

$$\mathrm{P}(X<Y) = \mathrm{P}(\min\{X,Y\} = X) = \frac{\lambda_X}{\lambda_X + \lambda_Y}.$$

4.2 Lack of memory property

Theorem 4.2.1 (Lack of memory property)**.** *A continuous random variable $X : \Omega \to (0,\infty)$ has an exponential distribution if and only if it has the lack of memory property:*

$$\mathrm{P}(X > x+y | X > x) = \mathrm{P}(X > y), \qquad \forall x, y > 0.$$

Proof [Reading material.]
Suppose $X \sim \mathrm{Exp}(\lambda)$, then we have for all $x, y > 0$,

$$\begin{aligned}
\mathrm{P}(X > x+y | X > x) &= \frac{\mathrm{P}(X > x+y, X > x)}{\mathrm{P}(X > x)} \\
&= \frac{\mathrm{P}(X > x+y)}{\mathrm{P}(X > x)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} \\
&= e^{-\lambda y} = \mathrm{P}(X > y).
\end{aligned}$$

Now suppose that X has the lack of memory property whenever $\mathrm{P}(X > x) > 0$. Set $G(x) = \mathrm{P}(X > x)$, then G is continuous (since X is continuous) and monotonically decreasing. Moreover, for all $x, y > 0$,

$$\begin{aligned}
G(x+y) &= \mathrm{P}(X > x+y) = \mathrm{P}(X > x+y | X > x)\mathrm{P}(X > x) \\
&\overset{\text{lack of mem.}}{=} \mathrm{P}(X > y)\mathrm{P}(X > x) = G(x)G(y).
\end{aligned} \tag{4.2.1}$$

Using equation (4.2.1) and induction, we can then show that

$$G(2) = G(1+1) = [G(1)]^2,$$

and

$$G(n) = [G(1)]^n.$$

Also, since X is assumed to be positive, i.e. $X > 0$, there exists an $m \in \mathbb{N}$ such that $G(1/m) = \mathrm{P}(X > 1/m) > 0$.

$$G(1) = G\left(\frac{1}{m} + \cdots + \frac{1}{m}\right) = \left[G\left(\frac{1}{m}\right)\right]^m > 0,$$

hence

$$G\left(\frac{1}{m}\right) = [G(1)]^{1/m}.$$

So

$$G\left(\frac{n}{m}\right) = G\left(\frac{1}{m} + \cdots + \frac{1}{m}\right) = \left[G\left(\frac{1}{m}\right)\right]^n = [G(1)]^{n/m}.$$

Hence, we obtain, by the same arguments, that, for all positive rational numbers $x = \frac{n}{m}$,

$$G(x) = G\left(\frac{n}{m}\right) = [G(1)]^{n/m} = [G(1)]^x.$$

Hence, for rational $x > 0$

$$\mathrm{P}(X > x) = G(x) = [G(1)]^x = e^{x\log(G(1))} = e^{-\lambda x}, \text{ for } \lambda = -\log(G(1)).$$

Since $G(1)$ is a probability, we note that $\lambda = -\log(G(1)) > 0$. Now, consider an irrational positive number $x > 0$ and any rational numbers u, v such that $v \leq x \leq u$. Since G is non-increasing

$$G(u) \leq G(x) \leq G(v) \Leftrightarrow [G(1)]^u \leq G(x) \leq [G(1)]^v.$$

Now we take the limits (through the rational numbers) as $v \uparrow x$ and $u \downarrow x$ to conclude that, for all $x > 0$

$$\mathrm{P}(X > x) = G(x) = [G(1)]^x = e^{x\log(G(1))} = e^{-\lambda x}, \text{ for } \lambda = -\log(G(1)).$$

Hence G is the survival function of the exponential distribution with parameter λ, which concludes the proof. □

Remark 4.2.2. *Note that the continuity assumption in the above theorem can be dropped and one can show the following stronger result: A random variable $X : \Omega \to (0, \infty)$ has an exponential distribution if and only if it has the lack of memory property:*

$$\mathrm{P}(X > x + y | X > x) = \mathrm{P}(X > y), \qquad \forall x, y > 0.$$

This can be proved using a very mild modification of the above proof, see e.g. (Norris 1998, p. 70-71) or Nelsen (1987) for details.

4.3 Criterion for the convergence/divergence of an infinite sum of independent exponentially distributed random variables

We will now study a criterion for the convergence/divergence of an infinite sum of independent exponentially distributed random variables. This result will be fundamental when we study birth processes, which are special cases of continuous-time Markov chains. In particular, we would like to know when such processes *explode*. In particular, we will describe how explosion relates to the convergence/divergence of the sum of the expected inter-arrival times of such processes.

Theorem 4.3.1. *Consider a sequence of independent random variables $H_i \sim \mathrm{Exp}(\lambda_i)$, for $0 < \lambda_i < \infty$ for all $i \in \mathbb{N}$ and let $J_\infty = \sum_{i=1}^{\infty} H_i$. Then:*

1. *If* $\sum_{i=1}^{\infty} \frac{1}{\lambda_i} < \infty$, *then* $\mathrm{P}(J_\infty < \infty) = 1$;
2. *If* $\sum_{i=1}^{\infty} \frac{1}{\lambda_i} = \infty$, *then* $\mathrm{P}(J_\infty = \infty) = 1$.

Proof of Theorem 4.3.1 (1.) Suppose that $\sum_{i=1}^{\infty} \frac{1}{\lambda_i} < \infty$. Let $J_n = \sum_{i=1}^{n} H_i$. Note that $J_n \uparrow J_\infty$, i.e. J_n monotonically increases to J_∞ as $n \to \infty$. Hence by the monotone convergence theorem (MON), see Theorem 2.8.3, we can interchange the limit and expectation and get

$$\mathrm{E}(J_\infty) = \mathrm{E}\left(\lim_{n\to\infty} \sum_{i=1}^{n} H_i\right) \overset{\text{MON}}{=} \lim_{n\to\infty} \sum_{i=1}^{n} \mathrm{E}(H_i) = \sum_{i=1}^{\infty} \frac{1}{\lambda_i} \overset{\text{by assumption}}{<} \infty.$$

$\mathrm{E}(J_\infty) < \infty$ implies that $\mathrm{P}(J_\infty < \infty) = 1$.

To see this note that, using the continuity property of the probability measure (Theorem 2.5.2) and the Markov inequality, we have

$$\begin{aligned}
\mathrm{P}(J_\infty = \infty) &= \mathrm{P}\left(\bigcap_{K=1}^{\infty} \{J_\infty \geq K\}\right) = \mathrm{P}\left(\lim_{N\to\infty} \bigcap_{K=1}^{N} \{J_\infty \geq K\}\right) \\
&= \mathrm{P}\left(\lim_{N\to\infty} \{J_\infty \geq N\}\right) \\
&\overset{\text{Theorem 2.5.2}}{=} \lim_{N\to\infty} \mathrm{P}\left(J_\infty \geq N\right) \\
&\overset{\text{Markov inequality}}{\leq} \lim_{N\to\infty} \frac{\mathrm{E}(J_\infty)}{N} = 0.
\end{aligned}$$

Alternatively, we could write,

$$\begin{aligned}
\infty > \mathrm{E}(J_\infty) &= \mathrm{E}(J_\infty | J_\infty < \infty)\mathrm{P}(J_\infty < \infty) + \mathrm{E}(J_\infty | J_\infty = \infty)\mathrm{P}(J_\infty = \infty) \\
&\overset{J_\infty > 0}{\geq} \mathrm{E}(J_\infty | J_\infty = \infty)\mathrm{P}(J_\infty = \infty),
\end{aligned}$$

which implies that $\mathrm{P}(J_\infty = \infty) = 0$.

[Note that $\mathrm{E}(J_\infty) = \infty$ *does not* imply that $\mathrm{P}(J_\infty = \infty) > 0$.]

□

Proof of Theorem 4.3.1 (2.) We show that $\mathrm{E}\left(\exp(-J_\infty)\right) = 0$ since this will imply $\mathrm{P}(J_\infty = \infty) = 1$.

We can apply the monotone convergence theorem for decreasing sequences, see Theorem 2.8.3, use the result for the Laplace transform $\mathrm{E}(\exp(-H_i)) = (1 + 1/\lambda_i)^{-1}$, and the independence of H_i, to deduce that

$$\begin{aligned}
\mathrm{E}\left(\exp(-J_\infty)\right) &= \mathrm{E}\left(\prod_{i=1}^{\infty} \exp(-H_i)\right) \overset{\text{MON2}}{=} \lim_{n\to\infty} \mathrm{E}\left(\prod_{i=1}^{n} \exp(-H_i)\right) \\
&\overset{(H_i) \text{ independent}}{=} \lim_{n\to\infty} \prod_{i=1}^{n} \mathrm{E}\left(\exp(-H_i)\right) \\
&\overset{\text{use Laplace transform of } H_i}{=} \prod_{i=1}^{\infty} \frac{1}{1 + 1/\lambda_i}.
\end{aligned}$$

Taking logs, we get

$$-\log\left(\mathrm{E}\left(\exp(-J_\infty)\right)\right) = \sum_{i=1}^{\infty} \log\left(1 + \frac{1}{\lambda_i}\right). \tag{4.3.1}$$

We will now show that $-\log\left(\mathrm{E}\left(\exp(-J_\infty)\right)\right) = \infty$, which implies that $\mathrm{E}\left(\exp(-J_\infty)\right) = 0$. So we can conclude that $\mathrm{P}(J_\infty = \infty) = 1$.

Consider $\sum_{i=1}^{\infty} \log\left(1 + \frac{1}{\lambda_i}\right)$. Two cases are possible:

- Either $\lambda_i \leq 1$ for infinitely many i, in which case $\log(1+1/\lambda_i) \geq \log(2)$ for each such i and the sum in (4.3.1) diverges,
- or $\lambda_i \leq 1$ for only finitely many i. Note that if $\lambda_i \geq 1$, then $\log(1+1/\lambda_i) \geq \log(2)\frac{1}{\lambda_i}$. Since the series $\sum_{i=0}^{\infty} \frac{1}{\lambda_i}$ diverges, the sum $\sum_{i=0}^{\infty} \frac{1}{\lambda_i}\mathbb{I}_{\{\lambda_i \geq 1\}}$, which is obtained by omitting finitely many terms, must also diverge. Hence the sum (4.3.1) diverges, too.

Note that

$$\mathbb{I}_{\{\lambda_i \geq 1\}} = \begin{cases} 1, & \text{if } \lambda_i \geq 1, \\ 0, & \text{otherwise.} \end{cases}$$

In the proof above we used the fact that if $\lambda_i \geq 1$, then $\log(1+1/\lambda_i) \geq \log(2)\frac{1}{\lambda_i}$. This can be checked using standard methods from analysis:

Lemma 4.3.2. *For $x \geq 1$, we have*

$$\log\left(1+\frac{1}{x}\right) \geq \log(2)\frac{1}{x}. \tag{4.3.2}$$

Recall the following inequality for the logarithm:

$$\log(1+x) > \frac{x}{x+1}, \qquad \text{for} \quad x > -1. \tag{4.3.3}$$

Proof of Lemma 4.3.2. Note that proving inequality (4.3.2) is equivalent to showing that

$$f(x) := x\log\left(1+\frac{1}{x}\right) - \log(2) \geq 0, \quad \text{for } x \geq 1.$$

We note that $f(1) = 0$ and

$$f'(x) = -\frac{1}{1+x} + \log\left(\frac{1+x}{x}\right) > -\frac{1}{1+x} + \frac{\frac{1}{x}}{\frac{1}{x}+1} = 0,$$

for $x \geq 1$, where we used (4.3.3), which implies that $f(1) = 0$ and then is monotonically increasing for $x \geq 1$, which concludes the proof. □

Alternatively, one could argue as follows: Inequality (4.3.3) implies that for $\lambda_i \geq 1$, we have

$$\log\left(1+\frac{1}{\lambda_i}\right) > \frac{1/\lambda_i}{1/\lambda_i+1} = \frac{1}{1+\lambda_i} \geq \frac{1}{2\lambda_i}.$$

The simpler inequality above would also lead to a divergent series.

□

End of lecture 12. .....

Chapter 5

Poisson processes

We will now start our discussion of continuous-time stochastic processes, where, as in the discrete-time case, we will (mainly) focus on processes taking values in a countable state space. We will start with some general remarks before introducing Poisson processes formally.

5.1 Remarks on continuous-time stochastic processes on a countable state space

The following summary is based on Norris (1998, p.67–70).

As in discrete-time, we denote by E a countable set (the state space). We recall that a continuous-time stochastic process denoted by $X = (X_t)_{t \geq 0}$ with values in E is a collection of random variables. How can we characterise the *law*, i.e. the probabilistic behaviour, of such a process? For instance, we might be interested in computing probabilities such as $\mathrm{P}(X_t = i)$, $\mathrm{P}(X_{t_0} = i_0, \ldots, X_{t_n} = i_n)$ or $\mathrm{P}(X_t = i \text{ for some } t)$.

We have the sigma-additivity property of the probability measure that for disjoint (A_i), the probability of the *countable* union satisfies

$$\mathrm{P}(\cup_i A_i) = \sum_i \mathrm{P}(A_i).$$

However, as soon as the union is not countable such as $\cup_{t \geq 0} A_t$, the sigma-additivity property is not applicable.

In order to overcome this problem, we typically work with **right-continuous processes**. One can show that any event depending on a right-continuous process can be determined from its **finite-dimensional distributions**, which are the probabilities

$$\mathrm{P}(X_{t_0} = i_0, \ldots, X_{t_n} = i_n),$$

for $n \in \mathbb{N}_0, 0 \leq t_0 \leq t_1 \leq \cdots \leq t_n$ and $i_0, \ldots, i_n \in E$.

A path/realisation $t \mapsto X_t(\omega)$ of a right continuous process on a countable state space resembles a step-function, i.e. it stays constant for some time before jumping to a new state. More precisely, there are three possible scenarios:

1. The path has infinitely many jumps, but only finitely many in any finite time interval $[0, t]$

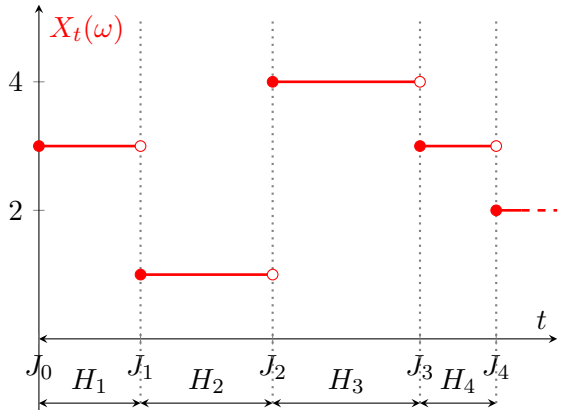


Figure 5.1: Scenario 1: One realisation of a continuous-time process with infinitely many jumps, but only finitely many on any finite interval.

2. Absorption: The path has only finitely many jumps and gets absorbed in one state where it stays forever:

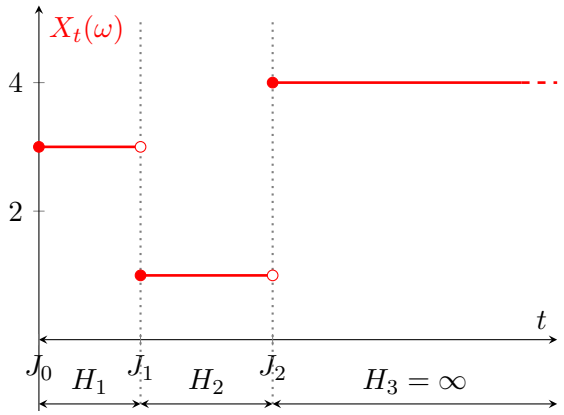


Figure 5.2: Scenario 2: One realisation of a continuous-time process with absorption.

3. Explosion: The path has infinitely many jumps in a finite time interval $[0, t]$, i.e. the process explodes. Afterwards the process starts again and might explode again or it might not.

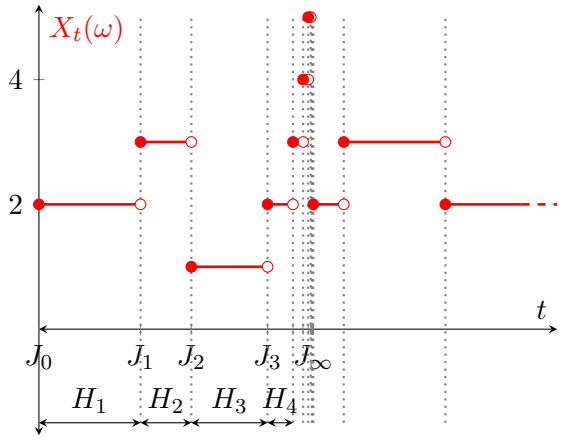


Figure 5.3: Scenario 3: One realisation fo a continuous-time process with explosion.

We call $J_0, J_1, \ldots$ the **jump times** of $X = \{X_t\}_{t\geq 0}$ and $H_1, H_2, \ldots$ the **holding times**. They can be derived from X as follows:

$$J_0 = 0, \qquad J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\}, \quad n \in \mathbb{N}_0,$$

where $\inf \emptyset = \infty$. Moreover, for $n \in \mathbb{N}$,

$$H_n = \begin{cases} J_n - J_{n-1}, & \text{if } J_{n-1} < \infty, \\ \infty, & \text{otherwise.} \end{cases}$$

Due to right-continuity we have that $H_n > 0$ for all $n \in \mathbb{N}$. If $J_{n+1} = \infty$ for some n, then we define $X_\infty := X_{J_n}$, i.e. we set it to the final value of the chain, otherwise X_∞ is undefined.

Note that

$$J_n = \sum_{i=1}^{n} H_i.$$

The **(first) explosion time** is defined as

$$J_\infty := \sup_{n\in\mathbb{N}_0} J_n = \sum_{n=1}^{\infty} H_n.$$

We can define the **jump process** associated with $X = (X_t)_{t\geq 0}$, or **jump chain** if X is a Markov chain, as the discrete-time process $(Z_n)_{n\in\mathbb{N}_0}$ with $Z_n := X_{J_n}$. (Note that you will sometimes see the notation $Z_n := X_{J_n+}$ which indicates that we are working with the right limit if right-continuity is not explicitly assumed.) The jump chain is just the sequence of all the values X takes up to explosion.

We are typically not interested in what happens to the process after explosion, but consider **minimal processes**: We can extend the state space E by adding the state ∞, say, and set $X_t = \infty$ if $t > J_\infty$. The term "minimal" here means that we are looking at a process which is active only for a minimal time, since its activities cease after the time of explosion.

We remark that a minimal process can be constructed from its holding times and jump process. In particular, this enables us to compute probabilities associated with $X = (X_t)_{t\geq 0}$ via countable unions. For instance,

$$\mathrm{P}(X_t = i) = \sum_{n=0}^{\infty} \mathrm{P}(Z_n = i, J_n \leq t < J_{n+1}),$$

and

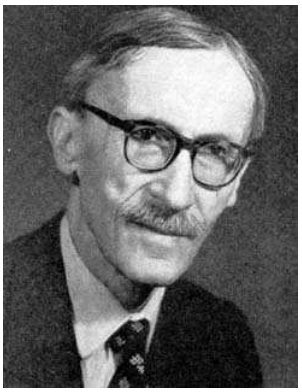
$$\mathrm{P}(X_t = i \text{ for some } t \in [0, \infty)) = \mathrm{P}(Z_n = i \text{ for some } n \in \mathbb{N}_0).$$

This will turn out to be very useful when proving properties for continuous-time Markov chains, since we will often relate them to the properties of the corresponding jump chain, where we have already derived many important results.

5.2 Introduction to Poisson processes

After having studied Markov chains in discrete time, we want to study Markov chains in continuous time. The general theory will be introduced in the next chapter. Here we start off with one particular example of a Markov process in continuous-time, the *Poisson process*. As a digression, the Poisson process (as well as Brownian motion) is a *Lévy* process.

(a)



(b)

Figure 5.4: The Poisson process is named after Siméon-Denis Poisson, see Figure (5.4a), a French mathematician, geometer, and physicist who lived from 21 June 1781 to 25 April 1840. Lévy processes are named after Paul Pierre Lévy, see Figure (5.4b), a French mathematician who lived from 15.09.1886 to 15.12.1971.

Poisson processes are the most basic form of *continuous-time* stochastic processes. Informally, we have a process that, starting at zero, *counts* events that occur during some time period; a realisation of the process is displayed in the following figure.

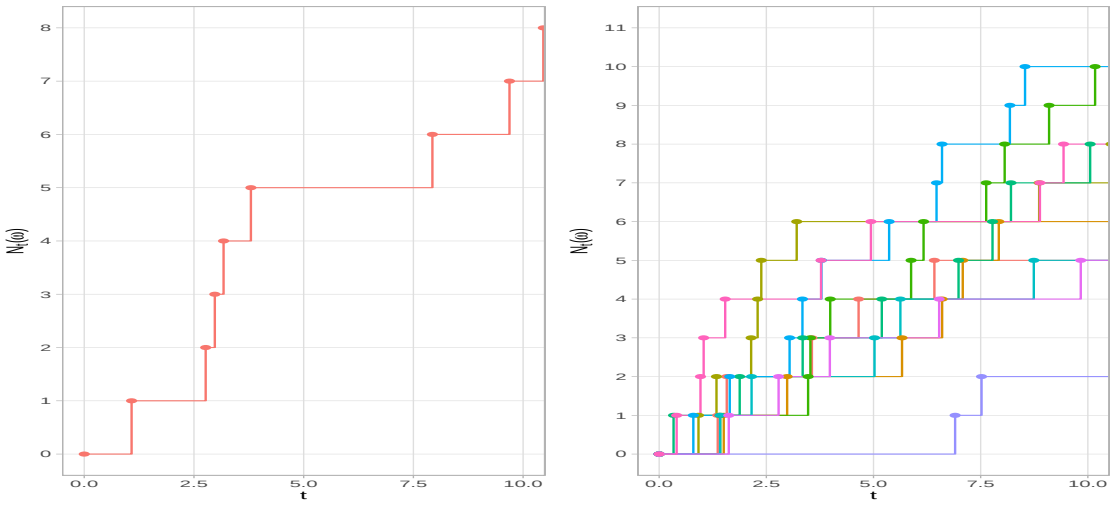


Figure 5.5: One realisation $(N_t(\omega))_{t\in[0,10]}$ (left) and ten realisations (for ten different choices of ω) (right) of a Poisson process with rate $\lambda = 0.7$.

Such a process is very useful in practice: it can be used as a model for earthquakes, queues, traffic etc. More interestingly, it can be combined with more complex processes to describe: jumps in the value of a stock and a process for mutations on a genealogical tree.

During this chapter we will prove a variety of properties associated to a Poisson process including:

- The number of events that occur in some interval $[0, t]$ is a Poisson random variable of rate λt.
- The time to the next event is independent of all previous times and is exponentially distributed of parameter λ.
- The time to the n^{th} event is a Gamma random variable.

One of the key points to remember throughout is that a Poisson process is *not* a Poisson random variable: this sounds obvious but many students confuse this issue.

There will be many extensions/aspects to (the basic) Poisson processes including:

- Thinning
- Non-Homogeneity
- Birth Processes

These ideas are important extensions of Poisson processes.

5.3 Some definitions

We begin by introducing the notion of a counting process. This will help to define the basic idea of the Poisson process. A first rather informal definition is given as follows.

A first definition: *A stochastic process $\{N_t\}_{t\geq 0}$ is said to be a counting process if N_t represents the total number of 'events' that have occurred up to time t.*

That is a counting process has the following properties:

1. $N_0 = 0$,
2. $\forall t \geq 0$, $N_t \in \mathbb{N}_0$,
3. If $0 \leq s < t$, $N_s \leq N_t$.

4. For $s < t$, $N_t - N_s$ equals the number of events that occur in the time interval $(s, t]$.
5. The process is piecewise constant and has upward jumps of size 1 (i.e. $N_t - N_{t-} \in \{0, 1\}$).

Note that $N_{t-} = \lim_{s \uparrow t} N_s$, which is the left limit at time t.
You can formalise the above definition in terms of the arrival times of events:

Definition 5.3.1. *Let $(J_n)_{n \in \mathbb{N}_0}$ be a strictly increasing sequence of positive random variables with $J_0 = 0$ almost surely. The process $\{N_t\}_{t \geq 0}$ defined by*

$$N_t = \sum_{n=1}^{\infty} \mathbb{I}_{\{J_n \leq t\}},$$

*which takes values in $\mathbb{N}_0$ is called the **counting process associated to the sequence** $(J_n)_{n \in \mathbb{N}_0}$.*

Recall that

$$\mathbb{I}_{\{J_n \leq t\}} = \mathbb{I}_{\{w \in \Omega : J_n(w) \leq t\}}(\omega) = \begin{cases} 1, & \text{if } J_n(\omega) \leq t, \\ 0, & \text{if } J_n(\omega) > t. \end{cases}$$

You can interpret J_n as the (random) time at which the nth event occurs or, equivalently, as the nth jump time.

Note that we typically add further assumptions, e.g. we are interested in stochastic processes which have *independent* and/or *stationary* increments.

5.3.1 Poisson process: First definition

We would like to give a first definition of a Poisson process. For this, we need to recall the so-called $o(\cdot)$ notation:

A function, f, is $o(\delta)$ if

$$\lim_{\delta \downarrow 0} \frac{f(\delta)}{\delta} = 0.$$

Example 5.3.2.
1. *Show that the function $f(x) = x^2$ is $o(\delta)$.*
2. *Show that if $f(\cdot)$ and $g(\cdot)$ are $o(\delta)$, then so is $f(\cdot) + g(\cdot)$.*
3. *Show that if $f(\cdot)$ is $o(\delta)$ and $c \in \mathbb{R}$, then $cf(\cdot)$ is $o(\delta)$.*

Definition 5.3.3. *A **Poisson process** $\{N_t\}_{t \geq 0}$ of rate $\lambda > 0$ is a non-decreasing stochastic process with values in $\mathbb{N}_0$ satisfying:*

1. $N_0 = 0$[1].
2. *The* increments *are* independent*, that is, given any choice $n \in \mathbb{N}$ and $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$, the random variables $N_{t_0}, N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.*
3. *The* increments *are* stationary*: Given any two distinct times $0 \leq s < t$ and for any $k \in \mathbb{N}_0$:*

$$\mathrm{P}(N_t - N_s = k) = \mathrm{P}(N_{t-s} = k).$$

4. *There is a 'single arrival', i.e. for any $t \geq 0$ and $\delta > 0, \delta \to 0$:*

$$\mathrm{P}(N_{t+\delta} - N_t = 1) = \lambda\delta + o(\delta),$$
$$\mathrm{P}(N_{t+\delta} - N_t \geq 2) = o(\delta),$$

[1] Technically, we only require that $N_0 = 0$ almost surely, i.e. $\mathrm{P}(N_0 = 0) = 1$.

A simple interpretation of the conditions:

- Condition (1) means that the process starts at 0.
- Condition (2) means that the increase of the number of events, in disjoint intervals of time:

$$[0, t_0], (t_0, t_1], \ldots, (t_{n-1} - t_n]$$

  are independent.
- Condition (3) means that the probability law is not affected by translation of the time parameter.
- Informally, condition (4) means that in an infinitesimal period of time there is either one or no event.

Also note that the single arrival property implies that

$$\mathrm{P}(N_{t+\delta} - N_t = 0) = 1 - \lambda\delta + o(\delta).$$

Note that a Poisson process is a counting process.

5.3.2 Poisson process: Second definition

If that definition is a little unclear, let us consider a second definition.

Definition 5.3.4. *A **Poisson process** $\{N_t\}_{t\geq 0}$ of rate $\lambda > 0$ is a stochastic process with values in $\mathbb{N}_0$ satisfying:*

1. $N_0 = 0$.
2. *The increments are independent, that is, given any choice $n \in \mathbb{N}$ and $0 \leq t_0 < t_1 < t_2 < \cdots < t_n$, the random variables $N_{t_0}, N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, N_{t_3} - N_{t_2}, \ldots, N_{t_n} - N_{t_{n-1}}$ are independent.*
3. *The* increments *are* stationary*: Given any two distinct times $0 \leq s < t$ and for any $k \in \mathbb{N}_0$:*

$$\mathrm{P}(N_t - N_s = k) = \mathrm{P}(N_{t-s} = k).$$

4. *For any $t \geq 0$, $N_t \sim \mathrm{Poi}(\lambda t)$, i.e. for all $t \geq 0$ and for all $k \in \mathbb{N}_0$ we have*

$$\mathrm{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

This definition is a little more concrete, as the probability distribution of the increments of the process is now explicitly given. In most rigorous probability work, the first definition is, essentially, a by-product of the definition of a Lévy process. However, perhaps, here, the second helps us to understand what is happening.

We note that in the second definition we do not need to add the assumption that the process is non-decreasing since this is implied by conditions 3.) and 4.).

Note that, in the definition above, we could combine the 3.) and 4.) conditions and assume the equivalent condition that, for any $0 \leq s < t$, $k \in \mathbb{N}_0$, we have

$$\mathrm{P}(N_t - N_s = k) = \frac{(\lambda(t-s))^k e^{-\lambda(t-s)}}{k!}.$$

That is, the number of events in $[s, t]$ is a Poisson random variable, of mean $\lambda(t - s)$.

5.3.3 Right-continuous modification

Note that when we have two stochastic processes $\{X_t\}_{t\geq 0}$ and $\{Y_t\}_{t\geq 0}$, we say that X is a *modification* of Y if

$$X_t = Y_t, \text{almost surely for each } t \geq 0,$$

i.e.

$$\mathrm{P}(X_t = Y_t) = 1, \text{for each } t \geq 0.$$

One can show that for each Poisson process there exists a *unique modification* which is *càdlàg* and which is also a Poisson process.

The term "càdlàg"comes from the French expression: *continue à droite, limitée à gauche*, which means *right continuous with left limits*.

Throughout the course, we always work with the càdlàg modification of a Poisson process.

In fact one can show that for each Lévy process there exists a *unique modification* which is *càdlàg* and which is also a Lévy process.

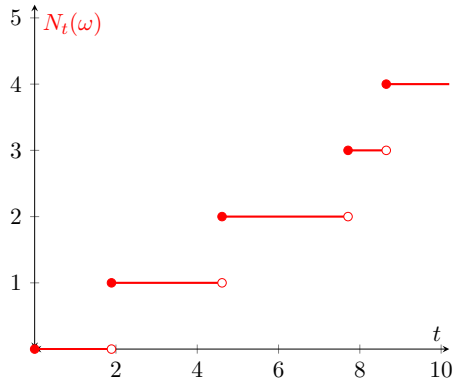


Figure 5.6: A right-continuous path of a Poisson process $(N_t(\omega))_{t\in[0,10]}$.

Remark 5.3.5. *We note that the jump chain of the Poisson process is given by* $Z = (Z_n)_{n\in\mathbb{N}_0}$*, where* $Z_n = n$*, for* $n \in \mathbb{N}_0$.

5.3.4 Equivalence of definitions

Clearly, we cannot have two definitions for a process that do not coincide. We have the first main result of the chapter.

Theorem 5.3.6. *Definitions 5.3.3 and 5.3.4 are equivalent.*

End of lecture 13. .....

Proof that Definition 5.3.3 implies Definition 5.3.4

We will introduce two methods for proving that Definition 5.3.3 implies Definition 5.3.4, the other direction will be left as an exercise.

The first method uses the Laplace transform of a Poisson random variable. Recall that, for a random variable X with discrete support X, the Laplace transform is, for $u > 0$

$$\mathcal{L}_X(u) = \mathrm{E}[e^{-uX}] = \sum_{x\in\mathsf{X}} e^{-ux}\mathrm{P}(X = x).$$

Let us derive the Laplace transform of a Poisson random variable first.

Lemma 5.3.7. *The Laplace transform of a Poisson random variable of mean λt (i.e. $X \sim \text{Poi}(\lambda t)$) for $\lambda > 0, t > 0$ is given by*

$$\mathcal{L}_X(u) = \exp\{\lambda t[e^{-u} - 1]\}, \qquad \forall u > 0.$$

Proof. For $u > 0$, we have

$$\begin{aligned}
\mathcal{L}_X(u) = \mathrm{E}[e^{-uX}] &= \sum_{x=0}^{\infty} e^{-ux}\mathrm{P}(X = x) \\
&= \sum_{x=0}^{\infty} e^{-ux}\frac{(\lambda t)^x}{x!}\exp(-\lambda t) \\
&= \exp(-\lambda t)\sum_{x=0}^{\infty}\frac{(e^{-u}\lambda t)^x}{x!} \\
&= \exp(-\lambda t)\exp\left(e^{-u}\lambda t\right) = \exp\left(\lambda t(e^{-u} - 1)\right).
\end{aligned}$$

□

We prove the direction that Definition 5.3.3 implies Definition 5.3.4. We note that conditions (1), (2) and (3) are identical in both definitions, so they are trivially satisfied. Hence we only need to show that condition (4) holds. I.e. we need to prove that $N_t \sim \text{Poi}(\lambda t)$.

Proof of Theorem 5.3.6 (using Laplace transforms). We begin by deriving a differential equation for $\mathcal{L}_N$ as follows. For $\delta > 0$, $t \geq 0$ and for $u > 0$,

$$\begin{aligned}
\mathcal{L}_N(t+\delta, u) &:= \mathrm{E}[e^{-uN_{t+\delta}}] \qquad \text{(multiply by 1 inside E}(\cdot)\text{)} \\
&= \mathrm{E}[e^{-u[N_{t+\delta}-N_t]}e^{-uN_t}] \qquad \text{(use independent incr.)} \\
&= \mathrm{E}[e^{-u[N_{t+\delta}-N_t]}]\mathbb{E}[e^{-uN_t}] \qquad \text{(use stationary incr.)} \\
&= \mathrm{E}[e^{-uN_\delta}]\mathcal{L}_N(t,u). \qquad (5.3.1)
\end{aligned}$$

The third line follows via the independent increments property and the last by the stationarity.

Now consider

$$\begin{aligned}
\mathrm{E}[e^{-uN_\delta}] &= \sum_{x=0}^{\infty} e^{-ux}\mathrm{P}(N_\delta = x) \\
&= e^{-u\cdot 0}\mathrm{P}(N_\delta = 0) + e^{-u}\mathrm{P}(N_\delta = 1) + \sum_{x=2}^{\infty} e^{-ux}\mathrm{P}(N_\delta = x)
\end{aligned}$$

Recall the single-arrival property: $\mathrm{P}(N_{t+\delta} - N_t = 0) = 1 - \lambda\delta + o(\delta)$ and $\mathrm{P}(N_{t+\delta} - N_t = 1) = \lambda\delta + o(\delta)$, $\mathrm{P}(N_{t+\delta} - N_t \geq 2) = o(\delta)$. Also, for $u > 0$:

$$0 \leq \sum_{x=2}^{\infty} e^{-ux}\mathrm{P}(N_\delta = x) < \sum_{x=2}^{\infty}\mathrm{P}(N_\delta = x) = \mathrm{P}(N_\delta \geq 2) = o(\delta).$$

Hence:

$$\begin{aligned}
\mathrm{E}[e^{-uN_\delta}] &= 1\cdot(1 - \lambda\delta + o(\delta)) + e^{-u}(\lambda\delta + o(\delta)) + o(\delta) \\
&= 1 - \lambda\delta + e^{-u}\lambda\delta + o(\delta). \qquad (5.3.2)
\end{aligned}$$

Combining (5.3.1) and (5.3.2) yields

$$\mathcal{L}_N(t+\delta, u) = \mathcal{L}_N(t,u)[1 - \lambda\delta + e^{-u}\lambda\delta] + o(\delta)$$

then it follows that

$$\mathcal{L}_N(t+\delta,u) - \mathcal{L}_N(t,u) = \mathcal{L}_N(t,u)\lambda\delta[-1+e^{-u}] + o(\delta),$$

then

$$\frac{\mathcal{L}_N(t+\delta,u) - \mathcal{L}_N(t,u)}{\delta} = \mathcal{L}_N(t,u)\lambda[e^{-u}-1] + \frac{o(\delta)}{\delta}$$

taking limits as $\delta \downarrow 0$ yields

$$\frac{\partial \mathcal{L}_N(t,u)}{\partial t} = \mathcal{L}_N(t,u)\lambda[e^{-u}-1]$$

that is

$$\frac{\partial \mathcal{L}_N(t,u)}{\partial t}\frac{1}{\mathcal{L}_N(t,u)} = \lambda[e^{-u}-1].$$

Since $\mathcal{L}_N(0,u) = \mathrm{E}(e^{-uN_0}) = \mathrm{E}(e^{-u\cdot 0}) = 1$, and integrating both sides w.r.t t we obtain

$$\log[\mathcal{L}_N(t,u)] = \lambda t[e^{-u}-1]$$

i.e.

$$\mathcal{L}_N(t,u) = \exp\{\lambda t[e^{-u}-1]\}$$

this is the Laplace transform of a Poisson random variable of mean λt. That is (due to the uniqueness property of Laplace transforms) N_t is a Poisson random variable, as specified in (3) of Definition 2.2.4. Note that we have shown that $N_t \sim \mathrm{Poi}(\lambda t)$. □

Remark 5.3.8. *In the proof above, we showed that the Laplace transform of N_t (when considered as a function in t) is differentiable from the right. Strictly speaking, we should check the continuity and differentiability from the left and the right. These results follow easily from our computations above. Please convince yourself that this is true!*

Proof of Theorem 5.3.6 (using forward equations). An alternative way is via the *forward equations*. It is very important to understand this concept. Again, we prove the direction that Definition 5.3.3 implies Definition 5.3.4.

Define, for $n \in \mathbb{N}_0, t \geq 0$

$$p_n(t) = \mathrm{P}(N_t = n).$$

Using the properties of Definition 3.2.3, it must be that the probabilities will coincide with that of a Poisson random variable. Let $n = 0, t \geq 0, \delta > 0$, then

$$\begin{aligned}
p_0(t+\delta) &= \mathrm{P}(N_{t+\delta} = 0) = \mathrm{P}(\text{no event in } [0, t+\delta]) \\
&= \mathrm{P}(\text{no event in} [0,t] \text{ and no event in} (t, t+\delta]) \\
&= \mathrm{P}(N_t = 0, N_{t+\delta} - N_t = 0) \\
&\overset{\text{ind. incr.}}{=} \mathrm{P}(N_t = 0)\mathrm{P}(N_{t+\delta} - N_t = 0) \\
&\overset{\text{stat. incr.}}{=} \mathrm{P}(N_t = 0)\mathrm{P}(N_\delta = 0) \\
&\overset{\text{singl. arrival}}{=} p_0(t)[1 - \lambda\delta + o(\delta)].
\end{aligned}$$

Hence we have

$$\frac{p_0(t+\delta) - p_0(t)}{\delta} = -\lambda p_0(t) + \frac{o(\delta)}{\delta}.$$

Letting $\delta \downarrow 0$, we get

$$\frac{dp_0(t)}{dt} = -\lambda p_0(t),$$

with $p_0(0) = \mathrm{P}(N_0 = 0) = 1$. Using the same approach as above, it clearly follows that

$$p_0(t) = e^{-\lambda t}.$$

For $n \in \mathbb{N}, t \geq 0, \delta > 0$, we have

$$\begin{aligned} p_n(t+\delta) &= \mathrm{P}(N_{t+\delta} = n) \\ &= \sum_{k=0}^{n} \mathrm{P}(N_{t+\delta} = n | N_t = k)\mathrm{P}(N_t = k) \quad \text{(Law of total probability)} \end{aligned}$$

Note that

$$\begin{aligned} \mathrm{P}(N_{t+\delta} = n | N_t = k) &= \mathrm{P}(N_{t+\delta} - N_t = n - k | N_t = k) \\ &\overset{\text{indep. incr.}}{=} \mathrm{P}(N_{t+\delta} - N_t = n - k) \\ &\overset{\text{stat. incr.}}{=} \mathrm{P}(N_\delta - N_0 = n - k) \\ &\overset{N_0=0}{=} \mathrm{P}(N_\delta = n - k) \\ &\overset{\text{single arrival}}{=} \begin{cases} o(\delta) & k = 0, 1, \ldots, n-2, \\ \lambda\delta + o(\delta), & k = n-1, \\ 1 - \lambda\delta + o(\delta), & k = n. \end{cases} \end{aligned}$$

Hence

$$\begin{aligned} p_n(t+\delta) &= \sum_{k=0}^{n} \mathrm{P}(N_{t+\delta} = n | N_t = k)\mathrm{P}(N_t = k) \\ &= \sum_{k=0}^{n-2} o(\delta)\mathrm{P}(N_t = k) + (\lambda\delta + o(\delta))\mathrm{P}(N_t = n-1) + (1 - \lambda\delta + o(\delta))\mathrm{P}(N_t = n) \\ &= (1 - \lambda\delta)p_n(t) + \lambda\delta p_{n-1}(t) + o(\delta). \end{aligned}$$

Hence we have

$$\frac{p_n(t+\delta) - p_n(t)}{\delta} = -\lambda p_n(t) + \lambda p_{n-1}(t) + \frac{o(\delta)}{\delta}.$$

Letting $\delta \downarrow 0$ we have

$$\frac{dp_n(t)}{dt} = -\lambda p_n(t) + \lambda p_{n-1}(t).$$

The probabilities can then be obtained by induction. Let $n = 1$, then we have the ODE

$$\frac{dp_1(t)}{dt} + \lambda p_1(t) = \lambda e^{-\lambda t}.$$

Recall to solve the (1-d, positive x) ODE

$$\frac{df}{dx} + \alpha(x)f(x) = g(x),$$

we have

$$f(x) = \frac{\int_0^x g(u)M(u)du + C}{M(x)},$$

where M is the integrating factor

$$M(x) = \exp\left(\int_0^x \alpha(u)du\right).$$

In the case when α and g are continuous functions (on an open interval), we also know that the solution to the ODE is unique.

In our case, the integrating factor is

$$M(t) = e^{\lambda t}$$

and the solution is

$$p_1(t) = \frac{\int_0^t \lambda ds + C}{e^{\lambda t}}$$

since $p_1(0) = \mathrm{P}(N_0 = 1) = 0$ we have the solution

$$p_1(t) = \lambda t e^{-\lambda t}.$$

We can easily complete a proof by induction, solving the ODE as above and using the induction hypothesis:

$$p_n(t) = \frac{(\lambda t)^n}{n!}e^{-\lambda t}$$

(exercise).

Hence we have shown the required property. □

Remark 5.3.9. *As before, in the proof above, we showed that the pmf of N_t (when considered as a function in t) is differentiable from the right. Strictly speaking, we should check the continuity and differentiability from the left and the right. These results follow easily from our computations above. Please convince yourself that this is true and possibly consult (Norris 1998, p. 76-77) if you get stuck.*

Proof that Definition 5.3.4 implies Definition 5.3.3

It remains to prove that Definition 5.3.4 implies Definition 5.3.3, which we leave as an exercise.

Exercise 5.3.10. *Prove that Definition 5.3.4 implies Definition 5.3.3.*

It is important that you try this exercise yourself! After you have completed it, you can compare your proof with the following model solution:

Solution to Exercise 5.3.10. We need to check the four conditions in Definition 5.3.3:
Conditions (1), (2) and (3): Conditions (1), (2) and (3) are trivially satisfied. Hence we only have to show that there is a single arrival.
Single arrival: We apply conditions (3) and (4) of Definition 5.3.4. For $k \in \mathbb{N}_0$, $t \geq 0$ and $\delta > 0$:

$$\mathrm{P}(N_{t+\delta} - N_t = k) = \frac{1}{k!}(\lambda\delta)^k e^{-\lambda\delta}.$$

Recall the Taylor series expansion of the exponential function:

$$e^{-\lambda\delta} = \sum_{n=0}^{\infty} \frac{(-\lambda\delta)^n}{n!} = 1 + \sum_{n=1}^{\infty} \frac{(-\lambda\delta)^n}{n!}.$$

The case $k = 1$: Hence for $k = 1$, we have

$$\mathrm{P}(N_{t+\delta} - N_t = 1) = \lambda\delta e^{-\lambda\delta} = \lambda\delta\left(1 + \sum_{n=1}^{\infty}\frac{(-\lambda\delta)^n}{n!}\right) = \lambda\delta + o(\delta),$$

since

$$\lim_{\delta\to 0}\frac{\lambda\delta\sum_{n=1}^{\infty}\frac{(-\lambda\delta)^n}{n!}}{\delta} = \lim_{\delta\to 0}\lambda\sum_{n=1}^{\infty}\frac{(-\lambda\delta)^n}{n!} = 0.$$

The case $k \geq 2$: Also, we have

$$\mathrm{P}(N_{t+\delta} - N_t \geq 2) = o(\delta),$$

since

$$\lim_{\delta\to 0}\frac{\mathrm{P}(N_{t+\delta} - N_t \geq 2)}{\delta} = \lim_{\delta\to 0}\frac{\sum_{k=2}^{\infty}\mathrm{P}(N_{t+\delta} - N_t = k)}{\delta} = \lim_{\delta\to 0}\frac{\sum_{k=2}^{\infty}\frac{1}{k!}(\lambda\delta)^k e^{-\lambda\delta}}{\delta}$$
$$= \lim_{\delta\to 0}\sum_{k=2}^{\infty}\frac{1}{k!}\lambda^k\delta^{(k-1)}e^{-\lambda\delta} = 0.$$

Alternatively, you could argue as follows:

$$\mathrm{P}(N_{t+\delta} - N_t = 0) = \exp(-\lambda\delta) = 1 - \lambda\delta + \sum_{n=2}^{\infty}\frac{(-\lambda\delta)^n}{n!} = 1 - \lambda\delta + o(\delta),$$

since

$$\lim_{\delta\to 0}\frac{1}{\delta}\sum_{n=2}^{\infty}\frac{(-\lambda\delta)^n}{n!} = \lim_{\delta\to 0}\sum_{n=2}^{\infty}\frac{(-\lambda)^n\delta^{n-1}}{n!}.$$

Then:

$$\mathrm{P}(N_{t+\delta} - N_t \geq 2) = 1 - \mathrm{P}(N_{t+\delta} - N_t < 2) = 1 - \mathrm{P}(N_{t+\delta} - N_t = 0) - \mathrm{P}(N_{t+\delta} - N_t = 1)$$
$$= 1 - (1 - \lambda\delta + o(\delta)) - (\lambda\delta + o(\delta)) = o(\delta).$$

□

End of lecture 14.

5.4 Some properties of Poisson processes

Now that we have a definition of our stochastic process, let us consider some properties of it.

5.4.1 Inter-arrival time distribution

We have a process that counts events. A natural question is then: 'What is the time between events?'. To help answer this question, we derive the inter-arrival time distribution. That is, the distribution of the time to the next event.

Theorem 5.4.1. *Let $\{N_t\}_{t\geq 0}$ be a Poisson process of rate $\lambda > 0$. Then the inter-arrival times are independently and identically distributed exponential random variables with rate parameter λ.*

Recall, that a continuous random variable taking non-negative values has the lack-of-memory property if and only if it follows the exponential distribution. We will re-visit this result on the problem sheet. Intuitively speaking, the fact that the inter-arrival times are independent and exponentially distributed means that the Poisson process has no memory and restarts itself every time an event occurs. We will come back to that concept when we study general Markov processes in continuous time in the next chapter.

Proof of Theorem 5.4.1. Let $(H_1, \dots, H_n) := H_{1:n}$ be the inter-arrival times for the first n events. Now consider, for $t > 0$

$$\begin{aligned} \mathrm{P}(H_1 > t) &= \mathrm{P}(\text{no events in}[0,t]) = \mathrm{P}(N_t = 0) \\ &= e^{-\lambda t}. \end{aligned}$$

$\mathrm{P}(H_1 > t)$ is sometimes called the *survival function* of H_1. We can now easily compute the cumulative distribution function of H_1:

$$F_{H_1}(t) = \mathrm{P}(H_1 \leq t) = 1 - \mathrm{P}(H_1 > t) = 1 - e^{-\lambda t}.$$

Hence the density function is:

$$f_{H_1}(t) = \lambda e^{-\lambda t},$$

which we recognise as an exponential density function, of rate λ, i.e. $H_1 \sim \mathrm{Exp}(\lambda)$.

Consider the second inter-arrival time, for $t > 0$, by the continuous law of total probability, we have

$$\mathrm{P}(H_2 > t) = \int_0^\infty \mathrm{P}(H_2 > t | H_1 = t_1) f_{H_1}(t_1) dt_1.$$

$$\begin{aligned} \mathrm{P}(H_2 > t | H_1 = t_1) &= \mathrm{P}(\text{no events in } (t_1, t_1 + t] | H_1 = t_1) \\ &= \mathrm{P}(N_{t_1+t} - N_{t_1} = 0 | H_1 = t_1) \\ &\overset{\text{indep. incr.}}{=} \mathrm{P}(N_{t_1+t} - N_{t_1} = 0) \\ &\overset{\text{stat. incr.}}{=} \mathrm{P}(N_t = 0) \\ &= e^{-\lambda t}. \end{aligned}$$

I.e. H_2 is independent of H_1 and

$$\begin{aligned} \mathrm{P}(H_2 > t) &= \int_0^\infty \mathrm{P}(H_2 > t | H_1 = t_1) f_{H_1}(t_1) dt_1 \\ &= e^{-\lambda t} \int_0^\infty f_{H_1}(t_1) dt_1 = e^{-\lambda t}. \end{aligned}$$

That is, the random variable H_2 is exponentially distributed with parameter λ, i.e. $H_2 \sim \mathrm{Exp}(\lambda)$.

Here we have used the independent and stationary increment property of the Poisson process (also note that we consider inter-arrival times, so we consider the number of events in the interval $(t_1, t_1 + t]$)[2].

This construction can be repeated for any $n \in \mathbb{N}$ with $n \geq 2$ (conditioning on $H_{1:n-1}$). In particular, set $T = t_1 + \cdots + t_{n-1}$. Then for $t > 0$

$$\begin{aligned} &\mathrm{P}(H_n > t | H_1 = t_1, \dots, H_{n-1} = t_{n-1}) \\ &= \mathrm{P}(\text{no events in } (T, T+t] | H_1 = t_1, \dots, H_{n-1} = t_{n-1}) \\ &= \mathrm{P}(N_{T+t} - N_T = 0 | H_1 = t_1, \dots, H_{n-1} = t_{n-1}) \\ &\overset{\text{indep. incr.}}{=} \mathrm{P}(N_{T+t} - N_T = 0) \\ &\overset{\text{stat. incr.}}{=} \mathrm{P}(N_t = 0) = e^{-\lambda t}. \end{aligned}$$

Using induction on n leads the result.

□

[2]Note that we can consider the interpret $\mathrm{P}(H_2 > t | H_1 = t_1)$ as $\mathrm{P}(H_2 > t | H_1 = t_1) = \lim_{\epsilon \downarrow 0} \mathrm{P}(H_2 > t | H_1 \in [t_1 - \epsilon, t_1]) = \lim_{\epsilon \downarrow 0} \mathrm{P}(H_2 > t | N_{t_1} = 1, N_{(t_1-\epsilon)-} = 0) = \lim_{\epsilon \downarrow 0} \mathrm{P}(H_2 > t | N_{t_1} - N_{(t_1-\epsilon)-} = 1, N_{(t_1-\epsilon)-} - N_0 = 0)$.

5.4.2 Time to the n^{th} event

Let $\{N_t\}_{t\geq 0}$ be a Poisson process of rate $\lambda > 0$. Define $J_0 = 0$, and, for $n \in \mathbb{N}$,

$$J_n = \sum_{i=1}^{n} H_i,$$

denotes the time to the n^{th} event (also called the time of the n^{th} jump).

Theorem 5.4.2. *Then, for any $n \in \mathbb{N}$, the time to the n^{th} event J_n follows a $\text{Gamma}(n, \lambda)$ distribution, i.e. its density is given by*

$$f_{J_n}(t) = \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, \quad t > 0.$$

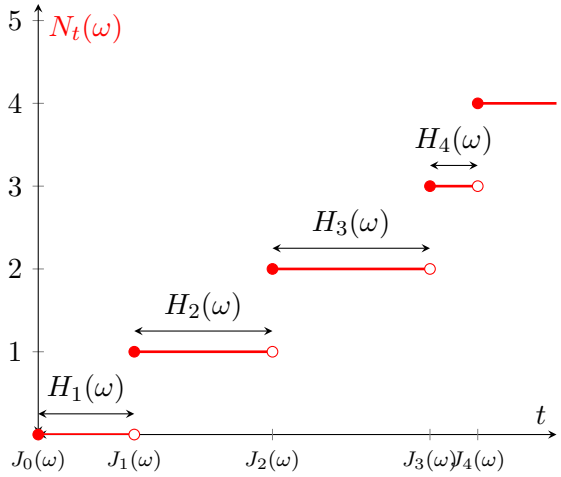


Figure 5.7: Sample path of a Poisson process with corresponding realisations of the inter-arrival times H_1, H_2, H_3, H_4 and jump times J_0, J_1, J_2, J_3, J_4.

Proof. This follows directly from Theorem 4.1.3. □

An alternative proof is based upon considering the quantity $\text{P}(J_n \leq t)$, and using the properties of the Poisson process.

Alternative proof of Theorem 5.4.2. Note that, for $t > 0$, $n \in \mathbb{N}_0$,

$$J_n \leq t \Leftrightarrow N_t \geq n.$$

Hence

$$F_{J_n}(t) = \text{P}(J_n \leq t) = \text{P}(N_t \geq n) = \sum_{k=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Now you only need to differentiate with respect to t to obtain the density f_{J_n}: In particular, we have[3]

$$f_{J_n}(t) = \frac{d}{dt} \sum_{k=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \sum_{k=n}^{\infty} \left(e^{-\lambda t}(-\lambda) \frac{(\lambda t)^k}{k!} + e^{-\lambda t} \frac{(\lambda t)^{k-1} \lambda}{(k-1)!} \right)$$

[3]The interchange of the derivative and the infinite sequence $\frac{d}{dt} \sum_{k=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \sum_{k=n}^{\infty} \frac{d}{dt} e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ can be justified by noting that the latter sum converges uniformly.

$$
\begin{aligned}
&= e^{-\lambda t}\lambda\left(\sum_{k=n}^{\infty}(-1)\frac{(\lambda t)^k}{k!}+\frac{(\lambda t)^{k-1}}{(k-1)!}\right)\\
&= e^{-\lambda t}\lambda\left(-\sum_{k=n}^{\infty}\frac{(\lambda t)^k}{k!}+\sum_{k=n-1}^{\infty}\frac{(\lambda t)^k}{k!}\right)\\
&= e^{-\lambda t}\frac{\lambda^n t^{n-1}}{(n-1)!} = e^{-\lambda t}\frac{\lambda^n t^{n-1}}{\Gamma(n)}, \quad \forall t>0,
\end{aligned}
$$

which is the density of a $\text{Gamma}(n,\lambda)$ random variable.

□

Example 5.4.3. *You call a telephone hot-line, and 'service' occurs according to a Poisson process of rate λ per minute. You are told that you are the n^{th} customer in line ($n \geq 1$):*

1. *How long, on average, will you have to wait to be served?*
2. *What is the probability that you have to wait longer than 1 hour?*

1. The mean of the time to the n^{th} event is:

$$\frac{n}{\lambda}.$$

2. The probability that you have to wait more than 1 hour is

$$\int_{60}^{\infty}\frac{\lambda^n}{\Gamma(n)}t^{n-1}e^{-\lambda t}dt.$$

This integral is not available analytically and needs to be approximated numerically.

5.4.3 Poisson process: Third definition

The properties we just derived can actually be used to *define* a Poisson process, which leads us to a third definition of a Poisson process

Definition 5.4.4. *A **Poisson process** $\{N_t\}_{t\geq 0}$ of rate $\lambda > 0$ is a stochastic process with values in $\mathbb{N}_0$ defined as follows:*

1. *Let $H_1, H_2, \ldots$ denote independent and identically exponentially distributed random variables with parameter $\lambda > 0$.*
2. *Let $J_0 = 0$ and $J_n = \sum_{i=1}^{n} H_i$*
3. *Define*

$$N_t = \sup\{n \in \mathbb{N}_0 : J_n \leq t\}, \quad \forall t \geq 0.$$

Theorem 5.4.5. *Definitions 5.3.3, 5.3.4 and 5.4.4 are equivalent.*

Proof. The derivations in the previous section (Theorem 5.4.1) contain the proof that a Poisson process according to Definition 5.3.4 is also a Poisson process according to Definition 5.4.4. It remains to show that Definition 5.4.4 implies Definition 5.3.4, which is left as an exercise, see Exercise 3- 25 on the problem sheet. □

5.4.4 Conditional distribution of the arrival times

Let J_n denote the time of the nth event. We now derive the conditional distribution of $J_1, \ldots, J_n$ given that $N_t = n$ (i.e. that n events have occurred on an interval $[0, t]$).

Theorem 5.4.6. *Let $\{N_t\}_{t\geq 0}$ be a Poisson process of rate $\lambda > 0$. Then for any $n \in \mathbb{N}$, $t > 0$ the conditional density of $(J_1, \ldots, J_n)$, given $N_t = n$ is given by:*

$$f_{(J_1,\ldots,J_n)}(t_1, \ldots, t_n | N_t = n) = \begin{cases} \frac{n!}{t^n} & \text{if } 0 < t_1 < \cdots < t_n \leq t, \\ 0, & \text{otherwise.} \end{cases}$$

I.e. the arrival times conditional on $N_t = n$ have the same joint distribution as the order statistics corresponding to n independent random variables uniformly distributed on the interval $[0, t]$.

Proof. We prove the case: $n = 1$. I.e. we know that one event has happened on an interval $[0, t]$. Given this information, what is the distribution of the time (J_1) at which the first event occurred?

Let $t_1 \leq t$. Then

$$\begin{aligned} \mathrm{P}(J_1 \leq t_1 | N_t = 1) &= \frac{\mathrm{P}(J_1 \leq t_1, N_t = 1)}{\mathrm{P}(N_t = 1)} \\ &= \frac{\mathrm{P}(1 \text{ event in } [0, t_1], 0 \text{ events in } (t_1, t])}{\mathrm{P}(N_t = 1)} \\ &= \frac{\mathrm{P}(N_{t_1} = 1, N_t - N_{t_1} = 0)}{\mathrm{P}(N_t = 1)} \\ &\overset{\text{indep. incr.}}{=} \frac{\mathrm{P}(N_{t_1} = 1)\mathrm{P}(N_t - N_{t_1} = 0)}{\mathrm{P}(N_t = 1)} \\ &\overset{\text{stat. incr.}}{=} \frac{\mathrm{P}(N_{t_1} = 1)\mathrm{P}(N_{t-t_1} = 0)}{\mathrm{P}(N_t = 1)} \\ &= \frac{\lambda t_1 e^{-\lambda t_1} \, e^{-\lambda(t-t_1)}}{e^{-\lambda t}\lambda t} = \frac{t_1}{t}. \end{aligned}$$

Hence we see that the time of the first event, given that there has been one event in $[0, t]$, is uniformly distributed over $[0, t]$.

Suppose that $0 < t_1 < \cdots < t_n$. Then we have

$$\begin{aligned} &\mathrm{P}(J_1 \in [0, t_1], J_2 \in (t_1, t_2], \ldots, J_n \in (t_{n-1}, t_n] | N_t = n) \\ &\mathrm{P}(J_1 \leq t_1, t_1 < J_2 \leq t_2, \ldots, t_{n-1} < J_n \leq t_n | N_t = n) \\ &= \frac{\mathrm{P}(J_1 \leq t_1, t_1 < J_2 \leq t_2 \ldots, t_{n-1} < J_n \leq t_n, N_t = n)}{\mathrm{P}(N_t = n)} \\ &= \frac{n! e^{\lambda t}}{(\lambda t)^n} \mathrm{P}(N_{[0,t_1]} = 1, \ldots, N_{(t_{n-1},t_n]} = 1, N_{(t_n,t]} = 0), \end{aligned}$$

where we have written the number of events upon an interval $(t_i, t_{i+1}]$ for $i = 1, \ldots, n-1$ as $N_{(t_i,t_{i+1}]}$. Note that $N_{(t_i,t_{i+1}]} = N_{t_{i+1}} - N_{t_i}$ and $N_{(t_n,t]} = N_t - N_{t_n}$.

Using the independent increments property of a Poisson process we obtain

$$\begin{aligned} &\mathrm{P}(J_1 \leq t_1, t_1 < J_2 \leq t_2, \ldots, t_{n-1} < J_n \leq t_n | N_t = n) \\ &= \frac{n! e^{\lambda t}}{(\lambda t)^n} \lambda t_1 e^{-\lambda t_1} \times \cdots \times \lambda [t_n - t_{n-1}] e^{-\lambda [t_n - t_{n-1}]} e^{-\lambda [t - t_n]} \\ &= \frac{n! e^{\lambda t}}{(\lambda t)^n} \left[\prod_{i=1}^{n} \lambda (t_i - t_{i-1}) \right] \exp \left[-\lambda \sum_{i=1}^{n+1} (t_i - t_{i-1}) \right], \end{aligned}$$

where $t_0 = 0, t_{n+1} = t$. Thus it clearly follows that

$$\mathrm{P}(J_1 \leq t_1, t_1 < J_2 \leq t_2, \ldots, t_{n-1} < J_n \leq t_n | N_t = n) = \frac{n!}{t^n} \prod_{i=1}^{n} (t_i - t_{i-1}).$$

Computing the partial derivative w.r.t t_n then t_{n-1} etc. gives the desired result.

To see that, define

$$G(t_1, \ldots, t_n) := \mathrm{P}(J_1 \leq t_1, t_1 < J_2 \leq t_2, \ldots, t_{n-1} < J_n \leq t_n | N_t = n) = \frac{n!}{t^n} \prod_{i=1}^{n} (t_i - t_{i-1}),$$

which is differentiable. To simplify the exposition, let us write $f(t_1, \ldots, t_n) := f_{(J_1,\ldots,J_n)}(t_1, \ldots, t_n | N_t = n)$ and consider the representation

$$G(t_1, \ldots, t_n) = \int_0^{t_1} \int_{t_1}^{t_2} \cdots \int_{t_{n-1}}^{t_n} f(x_1, \ldots, x_n) dx_n \cdots dx_1.$$

Then

$$\begin{aligned}
\frac{\partial G(t_1, \ldots, t_n)}{\partial t_n} &= \int_0^{t_1} \int_{t_1}^{t_2} \cdots \int_{t_{n-2}}^{t_{n-1}} f(x_1, \ldots, x_{n-1}, t_n) dx_{n-1} \cdots dx_1, \\
\frac{\partial^2 G(t_1, \ldots, t_n)}{\partial t_n \partial t_{n-1}} &= \int_0^{t_1} \int_{t_1}^{t_2} \cdots \int_{t_{n-3}}^{t_{n-2}} f(x_1, \ldots, x_{n-2}, t_{n-1}, t_n) dx_{n-2} \cdots dx_1, \ldots, \\
\frac{\partial^n G(t_1, \ldots, t_n)}{\partial t_n \cdots \partial t_1} &= f(t_1, \ldots, , t_n).
\end{aligned}$$

So we get indeed the conditional density function from partially differentiating the function G. □

Remark 5.4.7. *The above theorem says that conditional on the fact that n events have occurred in $[0, t]$, the times $J_1, \ldots, J_n$ at which events occur when considered as unordered random variables are independently and uniformly distributed on $[0, t]$.*

Exercise 5.4.8. *Show that the expectation of the k^{th} value ($1 \leq k \leq n$) of n uniformly distributed order statistics[4] on $[0, t]$ is*

$$\frac{tk}{n+1}.$$

Exercise 5.4.9. *Individuals arrive at a train station according to a Poisson process of rate λ per-unit time. The train departs at time t; what is the expected time that all the individuals (arriving in $(0, t)$) have to wait?*

Solution to Exercise 5.4.9. The problem asks us to calculate $\mathrm{E}[\sum_{i=1}^{N_t} [t - J_i]]$. Conditioning upon $N_t = n$ we have that

$$\begin{aligned}
\mathrm{E}\left[\sum_{i=1}^{n} [t - J_i] \middle| N_t = n \right] &= nt - \sum_{i=1}^{n} \mathrm{E}[J_i | N_t = n] = nt - \sum_{i=1}^{n} \frac{it}{n+1} \\
&= nt - \frac{t}{n+1} \sum_{i=1}^{n} i = nt - \frac{t}{n+1} \frac{n(n+1)}{2} = nt - \frac{nt}{2},
\end{aligned}$$

where we have used Theorem 5.4.6 and Exercise 5.4.8. Thus we conclude

$$\mathrm{E}\left[\sum_{i=1}^{N_t} [t - J_i] \right] = \frac{t}{2} \mathrm{E}[N_t],$$

[4] Probably you have studied order statistics in Y2

that is,

$$\mathrm{E}\left[\sum_{i=1}^{N_t}[t-J_i]\right] = \frac{\lambda t^2}{2}.$$

□

End of lecture 15. .....

5.5 Some extensions to Poisson processes

5.5.1 Superposition

Suppose now, that we are given two independent Poisson processes $\{N_t^{(1)}\}_{t\geq 0}$ and $\{N_t^{(2)}\}_{t\geq 0}$ (of rates $\lambda_1 > 0$ and $\lambda_2 > 0$), and we define a new stochastic process

$$N_t = N_t^{(1)} + N_t^{(2)}.$$

Exercise 5.5.1. *Show that $\{N_t\}_{t\geq 0}$ is a Poisson process of rate $\lambda_1 + \lambda_2$.*

Proof. See Exercise 3- 26 on the problem sheet. □

More generally, we have the following result.

Theorem 5.5.2. *Given n independent Poisson processes $\{N_t^{(1)}\}_{t\geq 0}, \ldots, \{N_t^{(n)}\}_{t\geq 0}$ with respective rates $\lambda_1, \ldots, \lambda_n > 0$, define*

$$N_t = \sum_{i=1}^{n} N_t^{(i)}, \quad \text{for } t \geq 0.$$

*Then $\{N_t\}_{t\geq 0}$ is a Poisson process with rate $\lambda = \sum_{i=1}^n \lambda_i$ and is called a **superposition of Poisson processes**.*

We revisit Example 4.1.6.

Exercise 5.5.3. *Suppose that blue cars arrive at a petrol station according to a Poisson process of rate λ_Y and red cars arrive, independently, according to a Poisson process of rate λ_X. What is the probability that N cars arrive in $[0,t]$?*

Solution to Exercise 5.5.3. Here, $N_t \sim \mathrm{Poi}((\lambda_X + \lambda_Y)t)$. I.e.

$$\mathrm{P}(N_t = N) = \frac{[(\lambda_X + \lambda_Y)t]^N}{N!} e^{-(\lambda_X+\lambda_Y)t}.$$

□

5.5.2 Thinning

Now, suppose in the context of Example 5.5.3, that we know that all cars arrive according to a Poisson process (of rate λ), but, that we are only interested in the process, of (say) green cars, which are observed independently of the Poisson process, with probability $p \in (0,1)$; what can we say about this process? Writing $\{N_t^g\}$ as this process, we have the following result.

Exercise 5.5.4. *Show that $\{N_t^g\}$ is a Poisson process of rate λp.*

Proof. See Exercise 3- 28 on the a problem sheet. □

More generally, we have the following result.

Theorem 5.5.5. *Let $\{N_t\}_{t\geq 0}$ denote a Poisson process with rate $\lambda > 0$. Assume that each arrival, independent of other arrivals, is marked as a* type k *event with probability p_k, for $k = 1, \ldots, n$, where $\sum_{i=1}^{n} p_i = 1$. Let $N_t^{(k)}$ denote the number of type k events in $[0, t]$. Then $\{N_t^{(k)}\}_{t\geq 0}$ is a Poisson process with rate λp_k, and the processes*

$$\{N_t^{(1)}\}_{t\geq 0}, \ldots, \{N_t^{(n)}\}_{t\geq 0}$$

are independent. Each process is called a ***thinned Poisson process****.*

We illustrate the concept of thinning of a Poisson process in the Figure 5.8. Here we first simulated one sample path (which is the same as a realisation) of a Poisson process $\{N_t\}_{t\in[0,10]}$ of rate $\lambda = 2$. We then thinned the process using $p_1 = 0.4$ and $p_2 = 1 - p_1 = 0.6$. We obtain two new (independent) thinned Poisson processes: the Poisson process $\{N_t^{(1)}\}_{t\in[0,10]}$ with rate $\lambda p_1 = 0.8$ and the Poisson process $\{N_t^{(2)}\}_{t\in[0,10]}$ with rate $\lambda p_2 = 1.2$. We note that $N_t = N_t^{(1)} + N_t^{(2)}$, for all $t \in [0, 10]$.

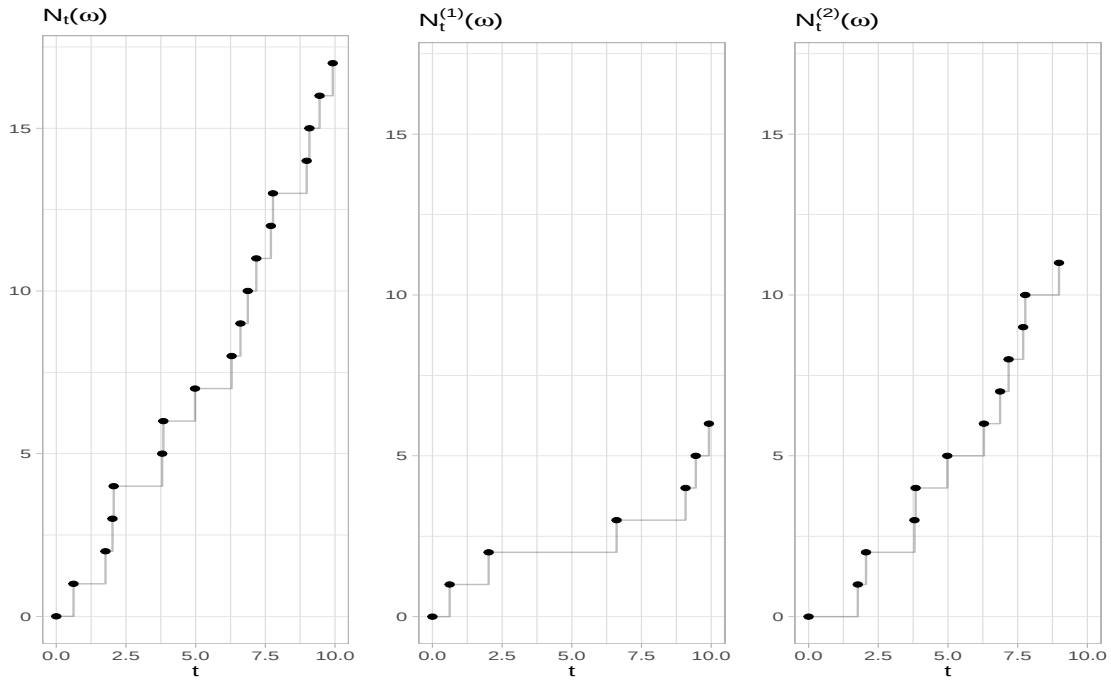


Figure 5.8: Left: Sample path of a Poisson process $\{N_t(\omega)\}_{t\in[0,10]}$ with rate $\lambda = 2$. We thin the Poisson process (with probabilities $p_1 = 0.4, p_2 = 0.6$) and split it into two parts such that $N_t(\omega) = N_t^{(1)}(\omega) + N_t^{(2)}(\omega)$. Middle: Sample path of the thinned Poisson process $\{N_t^{(1)}(\omega)\}_{t\in[0,10]}$ with rate $\lambda p_1 = 0.8$. Right: Sample path of the thinned Poisson process $\{N_t^{(2)}(\omega)\}_{t\in[0,10]}$ with rate $\lambda p_2 = 1.2$.

5.5.3 Non-homogeneous Poisson processes

Definition 5.5.6. *Let $\lambda : [0, \infty) \mapsto (0, \infty)$ denote a non-negative and locally integrable function, called the* ***intensity function****. A non-decreasing stochastic process $N = \{N_t\}_{t\geq 0}$ with values in $\mathbb{N}_0$ is called a* ***non-homogeneous Poisson process*** *with intensity function $(\lambda(t))_{t\geq 0}$ if it satisfies the following properties:*

1. $N_0 = 0$.
2. *N has independent increments.*

3. *'Single arrival' property: For $t \geq 0$, $\delta > 0$:*

$$\begin{aligned} \mathrm{P}(N_{t+\delta} - N_t = 1) &= \lambda(t)\delta + o(\delta), \\ \mathrm{P}(N_{t+\delta} - N_t \geq 2) &= o(\delta), \end{aligned}$$

that is, the rate/intensity is now dependent upon the time parameter.

Also note that the single arrival property implies that

$$\mathrm{P}(N_{t+\delta} - N_t = 0) = 1 - \lambda(t)\delta + o(\delta).$$

We would like to derive the marginal distribution of N_t. More specifically, we would like to prove the following result.

Theorem 5.5.7. *Let $N = \{N_t\}_{t\geq 0}$ denote a non-homogeneous Poisson process with continuous intensity function $(\lambda(t))_{t\geq 0}$. Then*

$$N_t \sim \mathrm{Poi}(m(t)), \qquad \text{where} \qquad m(t) = \int_0^t \lambda(s)ds,$$

i.e., for all $t \geq 0$ and $n \in \mathbb{N}_0$,

$$\mathrm{P}(N_t = n) = \frac{[m(t)]^n}{n!} e^{-m(t)}.$$

Remark 5.5.8. *Please note that, in the Poisson distribution above, you need to work with the integrated intensity function $m(t)$ and **not** with the intensity function $\lambda(t)$ directly. This is a common source for errors in calculations. Hence, if you would like to compute probabilities involving non-homogeneous Poisson processes, make sure that you integrate the intensity function over the time interval of interest first before "plugging" it into the Poisson probability mass function.*

Proof of Theorem 5.5.7. We prove the result using forward equations and induction in n. First we derive the forward equations. Define, for $n \in \mathbb{N}_0$, $t \geq 0$

$$p_n(t) = \mathrm{P}(N_t = n).$$

We want to show that, for all $t \geq 0$ and $n \in \mathbb{N}_0$,

$$\mathrm{P}(N_t = n) = \frac{[m(t)]^n}{n!} e^{-m(t)}. \tag{5.5.1}$$

We check the base case $n = 0$. Then, for $t \geq 0, \delta > 0$,

$$\begin{aligned} p_0(t+\delta) = \mathrm{P}(N_{t+\delta} = 0) &= \mathrm{P}(N_t = 0, N_{t+\delta} - N_t = 0) \\ &\overset{\text{ind. incr.}}{=} \mathrm{P}(N_t = 0)\mathrm{P}(N_{t+\delta} - N_t = 0) \\ &\overset{\text{single arrival}}{=} p_0(t)[1 - \lambda(t)\delta + o(\delta)]. \end{aligned}$$

Hence we have

$$\frac{p_0(t+\delta) - p_0(t)}{\delta} = -\lambda(t)p_0(t) + \frac{o(\delta)}{\delta}.$$

Letting $\delta \downarrow 0$ we get

$$\frac{dp_0(t)}{dt} = -\lambda(t)p_0(t),$$

with boundary condition $p_0(0) = \mathrm{P}(N_0 = 0) = 1$.

Hence, we obtain

$$p_0(t) = \exp\left(-\int_0^t \lambda(s)ds\right) = \frac{[m(t)]^0}{0!}e^{-m(t)}.$$

Next, we consider the case when $n \in \mathbb{N}$. For $t \geq 0, \delta > 0$, we have

$$\begin{aligned} p_n(t+\delta) &= \mathrm{P}(N_{t+\delta} = n) \\ &= \sum_{k=0}^{n} \mathrm{P}(N_{t+\delta} = n | N_t = k)\mathrm{P}(N_t = k) \quad \text{(Law of total probability)} \end{aligned}$$

Note that

$$\begin{aligned} \mathrm{P}(N_{t+\delta} = n | N_t = k) &= \mathrm{P}(N_{t+\delta} - N_t = n - k | N_t = k) \\ &\overset{\text{indep. incr.}}{=} \mathrm{P}(N_{t+\delta} - N_t = n - k) \\ &\overset{\text{single arrival}}{=} \begin{cases} o(\delta) & k = 0, 1, \ldots, n-2, \\ \lambda(t)\delta + o(\delta), & k = n-1, \\ 1 - \lambda(t)\delta + o(\delta), & k = n. \end{cases} \end{aligned}$$

Hence

$$\begin{aligned} p_n(t+\delta) &= \sum_{k=0}^{n} \mathrm{P}(N_{t+\delta} = n | N_t = k)\mathrm{P}(N_t = k) \\ &= \sum_{k=0}^{n-2} o(\delta)\mathrm{P}(N_t = k) + (\lambda(t)\delta + o(\delta))\mathrm{P}(N_t = n-1) \\ &\quad + (1 - \lambda(t)\delta + o(\delta))\mathrm{P}(N_t = n) \\ &= (1 - \lambda(t)\delta)p_n(t) + \lambda(t)\delta p_{n-1}(t) + o(\delta). \end{aligned}$$

Hence we have

$$\frac{p_n(t+\delta) - p_n(t)}{\delta} = -\lambda(t)p_n(t) + \lambda(t)p_{n-1}(t) + \frac{o(\delta)}{\delta}.$$

Letting $\delta \downarrow 0$ we have

$$\frac{dp_n(t)}{dt} = -\lambda(t)p_n(t) + \lambda(t)p_{n-1}(t). \tag{5.5.2}$$

If we define $p_{-1} = 0$, then equation (5.5.2) describes the forward equations for all $n \in \mathbb{N}_0$.

Now we do the induction step. Suppose that (5.5.1) holds for an $n \in \mathbb{N}_0$, then we want to show that it also holds for $n+1$.

From the forward equations, we get for $n+1$:

$$\frac{dp_{n+1}(t)}{dt} = -\lambda(t)p_{n+1}(t) + \lambda(t)p_n(t).$$

From the induction hypotheses we get

$$p_n(t) = \frac{[m(t)]^n}{n!}e^{-m(t)},$$

and hence

$$\frac{dp_{n+1}(t)}{dt} + \lambda(t)p_{n+1}(t) = \lambda(t)\frac{[m(t)]^n}{n!}e^{-m(t)} =: g(t).$$

Recall that in order to solve the (one-dimensional) ordinary differential equation

$$\frac{df}{dx} + \alpha(x)f(x) = g(x), \qquad x > 0,$$

we have

$$f(x) = \frac{\int_0^x g(u)M(u)du + C}{M(x)},$$

where M is the integrating factor

$$M(x) = \exp\left(\int_0^x \alpha(u)du\right).$$

Here the integrating factor is given by

$$M(x) = \exp\left(\int_0^x \lambda(u)du\right) = \exp(m(x)).$$

Then

$$\begin{aligned} p_{n+1}(t) &= \left(\int_0^t g(u)M(u)du + C\right) M(t)^{-1} \\ &= \left(\int_0^t g(u)M(u)du\right) M(t)^{-1} \\ &= \int_0^t \lambda(u)\frac{[m(u)]^n}{n!}du\, e^{-m(t)}. \end{aligned}$$

Since $p_{n+1}(0) = \mathrm{P}(N_0 = n+1) = 0$, we have that $C = 0$.

I.e. it remains to show that

$$\int_0^t \lambda(u)\frac{[m(u)]^n}{n!}du = \frac{1}{(n+1)!}\left(\int_0^t \lambda(s)ds\right)^{n+1}.$$

This is an application of the chain rule: Define

$$f(u) := \frac{1}{n!}u^n, \qquad m(u) = \int_0^u \lambda(s)ds.$$

Note that $m'(u) = \lambda(u)$. Let

$$F(t) := \int_0^t f(u)du = \frac{1}{(n+1)!}t^{n+1}.$$

Then, since $m(0) = 0$,

$$\int_0^t f(m(u))m'(u)du = F(m(u))|_0^t = F(m(t)) = \frac{1}{(n+1)!}[m(t)]^{n+1},$$

which concludes the proof. □

Remark 5.5.9. *We note that we have again focused on right-differentiability in our proof above. The case of considering left-limits follows similarly. We added the assumption that the intensity function is continuous to avoid complications when considering the left and right-limits of the intensity function λ (which are identical for continuous functions).*

Exercise 5.5.10. *Derive the distribution of the increment $N_t - N_s$, for $0 \leq s < t$. Does a non–homogeneous Poisson process have stationary increments?*

Solution to Exercise 5.5.10. Let $0 \leq s < t$. We have shown that $N_t \sim \text{Poi}(m(t))$. Also, $N_s \sim \text{Poi}(m(s))$. Observe that

$$N_t = (N_t - N_s) + (N_s - N_0),$$

since $N_0 = 0$. Now, use the Laplace transform. We know that for $u > 0$

$$\text{E}\left(\exp(-uN_t)\right) = \exp(m(t)(\exp(-u) - 1)).$$

Also,

$$\text{E}\left(\exp(-uN_s)\right) = \exp(m(s)(\exp(-u) - 1)).$$

Using the independence increment property, we get

$$\text{E}\left(\exp(-uN_t)\right) = \text{E}\left(\exp(-u(N_t - N_s + N_s - N_0))\right) = \text{E}\left(\exp(-u(N_t - N_s))\right)\text{E}\left(\exp(-u(N_s - N_0))\right)$$

Hence

$$\text{E}\left(\exp(-uN_t)\right)\left[\text{E}\left(\exp(-u(N_s - N_0))\right)\right]^{-1} = \text{E}\left(\exp(-u(N_t - N_s))\right)$$

Now we only have to plug in the results for the Laplace transform of N_t, and N_s and we get

$$\begin{aligned}\text{E}\left(\exp(-u(N_t - N_s))\right) &= \exp(m(t)(\exp(-u) - 1))\exp(-m(s)(\exp(-u) - 1)) \\ &= \exp((m(t) - m(s))(\exp(-u) - 1)),\end{aligned}$$

which is the Laplace transform of a Poisson random variable with rate

$$m(t) - m(s) = \int_s^t \lambda(u)du.$$

Hence, we see that the increments are generally **not** stationary. □

Exercise 5.5.11. *Revise the material on* conditional *distribution, mass, density and conditional expectation from your first and second year probability courses. E.g. you can read Grimmett & Stirzaker (2001*b*, p. 67–68 (Section 3.7) and p. 104–106 (Section 4.6)).*

End of lecture 16. .....

5.5.4 Compound Poisson processes

An interesting extension of the Poisson process (which is also a Lévy process) is called the compound Poisson process.

Definition 5.5.12. *Let $\{N_t\}_{t\geq 0}$ be a Poisson process of rate $\lambda > 0$. In addition, let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables, that are independent of $\{N_t\}_{t\geq 0}$. Then the process $\{S_t\}_{t\geq 0}$ with*

$$S_t = \sum_{i=1}^{N_t} Y_i, \qquad t \geq 0,$$

*is a **compound Poisson process**.*

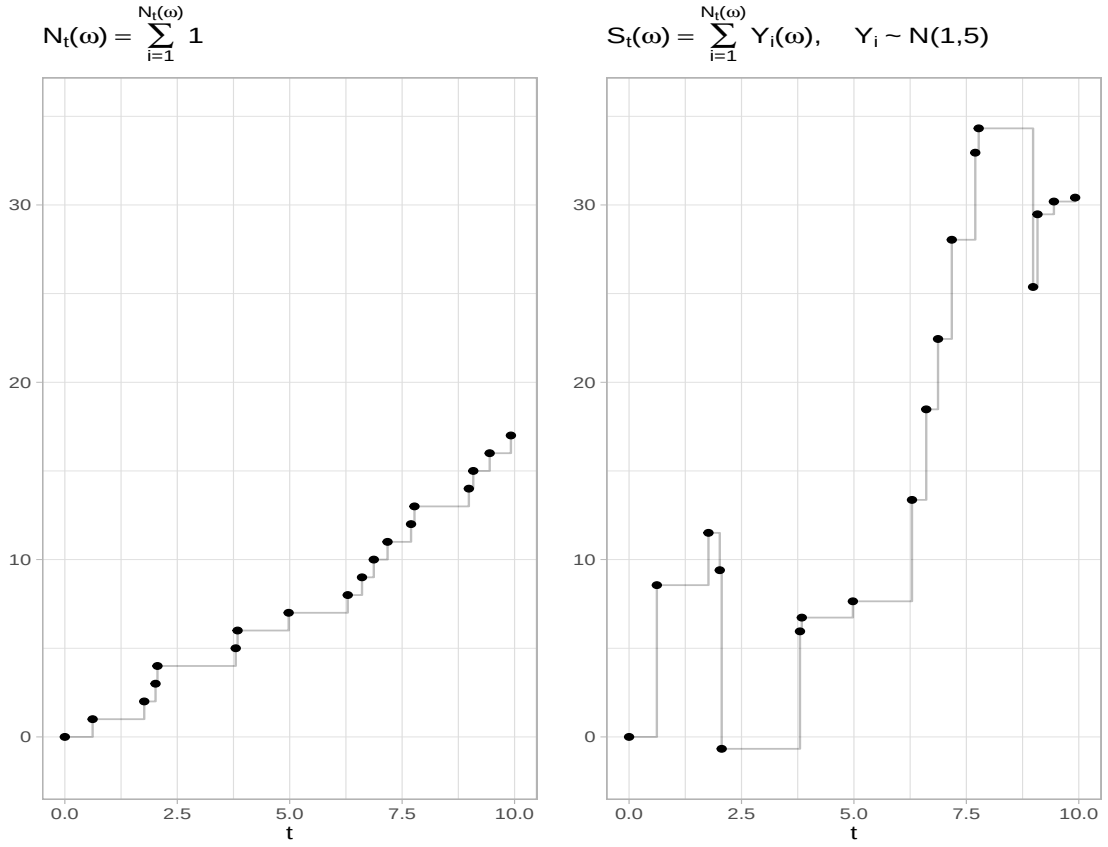


Figure 5.9: Left: Sample path of a Poisson process $\{N_t(\omega)\}_{t\in[0,10]}$ with rate $\lambda = 2$. Right: Sample path of a compound Poisson process $\{S_t(\omega)\}_{t\in[0,10]}$ with $S_t(\omega) = \sum_{i=1}^{N_t(\omega)} Y_i(\omega)$. We use the same realisation of N as in the left picture and consider jump sizes $Y_i \sim \mathrm{N}(1,5)$. Note that a compound Poisson process is not in general a counting process and it can take values in $\mathbb{R}$ (if Y_i take values in $\mathbb{R}$).

Theorem 5.5.13. *Let $\{S_t\}_{t\geq 0}$ denote a compound Poisson process as defined in Definition 5.5.12. Then for $t \geq 0$,*

$$\mathrm{E}(S_t) = \lambda t \mathrm{E}(Y_1), \qquad \mathrm{Var}(S_t) = \lambda t \mathrm{E}(Y_1^2).$$

Proof. When proving results for the compound Poisson process, we typically condition on N_t in the first step. Here we will use the law of total expectation:

Using the law of total expectation, we have

$$\mathrm{E}(S_t) = \sum_{n=0}^{\infty} \mathrm{E}(S_t|N_t = n)\mathrm{P}(N_t = n),$$

where for $n \in \mathbb{N}_0$

$$\mathrm{E}(S_t|N_t = n) = \mathrm{E}\left(\sum_{i=1}^{N_t} Y_i|N_t = n\right) = \mathrm{E}\left(\sum_{i=1}^{n} Y_i|N_t = n\right)$$
$$\overset{(Y_i),N \text{ indep.}}{=} \mathrm{E}\left(\sum_{i=1}^{n} Y_i\right) \overset{(Y_i) \text{ identically distr.}}{=} n\mathrm{E}(Y_1),$$

Hence

$$\mathrm{E}(S_t) = \sum_{n=0}^{\infty} \mathrm{E}(S_t|N_t = n)\mathrm{P}(N_t = n) = \sum_{n=0}^{\infty} n\mathrm{E}(Y_1)\mathrm{P}(N_t = n)$$

$$= \sum_{n=0}^{\infty} nP(N_t = n)E(Y_1) = E(N_t)E(Y_1) \overset{N_t \sim \text{Poi}(\lambda t)}{=} \lambda t E(Y_1).$$

Note that we can shorten the proof slightly, by using the short-hand notation for the law of total expectation, where we write

$$E(S_t) = E[E(S_t|N_t)]$$

instead of the long version

$$E(S_t) = \sum_{n=0}^{\infty} E(S_t|N_t = n)P(N_t = n).$$

Since we have shown that $E(S_t|N_t = n) = nE(Y_1)$ for all $n \in \mathbb{N}_0$, we can conclude that

$$E(S_t|N_t) = N_t E(Y_1),$$

and then, by the linearity of the expectation,

$$E(S_t) = E[E(S_t|N_t)] = E[N_t E(Y_1)] = E[N_t]E(Y_1) = \lambda t E(Y_1).$$

Similarly, for the variance, we use the *law of total variance*, which (using the short-hand notation) reads as follows:

$$\text{Var}(S_t) = \text{Var}\left(\text{E}\left(S_t|\, N_t\right)\right) + \text{E}\left(\text{Var}\left(S_t|\, N_t\right)\right). \tag{5.5.3}$$

(*If formula* (5.5.3) *is new to you, then prove it at home as an exercise!*

Let us compute the conditional variance first: For all $n \in \mathbb{N}_0$, we have

$$\begin{aligned}\text{Var}(S_t|N_t = n) &= \text{Var}\left(\sum_{i=1}^{N_t} Y_i|N_t = n\right) = \text{Var}\left(\sum_{i=1}^{n} Y_i|N_t = n\right) \\ &\overset{(Y_i),N \text{ indep.}}{=} \text{Var}\left(\sum_{i=1}^{n} Y_i\right) \overset{(Y_i) \text{ ind. \& identically distr.}}{=} n\text{Var}(Y_1).\end{aligned}$$

Then

$$\begin{aligned}\text{E}\left(\text{Var}\left(S_t|\, N_t\right)\right) &= \sum_{n=0}^{\infty} \text{Var}(S_t|N_t = n)\text{P}(N_t = n) \\ &= \sum_{n=0}^{\infty} n\text{Var}(Y_1)\text{P}(N_t = n) \\ &= \text{Var}(Y_1)\sum_{n=0}^{\infty} n\text{P}(N_t = n) = \text{Var}(Y_1)\text{E}(N_t) \\ &\overset{N_t \sim \text{Poi}(\lambda t)}{=} \lambda t \text{Var}(Y_1).\end{aligned}$$

Alternatively, we can argue that, since $\text{Var}(S_t|N_t = n) = n\text{Var}(Y_1)$ for all $n \in \mathbb{N}_0$, we have

$$\text{Var}(S_t|N_t) = N_t \text{Var}(Y_1),$$

and, hence,

$$\text{E}\left(\text{Var}\left(S_t|\, N_t\right)\right) = \text{E}[N_t \text{Var}(Y_1)] = \text{E}[N_t]\text{Var}(Y_1) = \lambda t \text{Var}(Y_1).$$

Also, using the short-hand proof only, we have

$$\mathrm{Var}(\mathrm{E}(S_t|N_t)) = \mathrm{Var}(N_t\mathrm{E}(Y_1)) = \mathrm{Var}(N_t)(\mathrm{E}(Y_1))^2 = \lambda t(\mathrm{E}(Y_1))^2.$$

Recall that if you take constants (in our case $\mathrm{E}(Y_1)$) out of the variance, you need to square them!

Using formula (5.5.3), we get

$$\mathrm{Var}(S_t) = \lambda t\mathrm{Var}(Y_1) + \lambda t(\mathrm{E}(Y_1))^2 = \lambda t\mathrm{E}(Y_1^2).$$

□

Exercise 5.5.14. *Suppose that asteroids fall to the earth according to a Poisson process of rate $\lambda > 0$. In addition, and independent of the arrival of the asteroid (and other asteroids), the asteroid will cause a human fatality with probability $p \in (0,1)$. Let S_t denote number of human fatalities at time t. Find the probability generating function (pgf) of S_t, and, using the pgf, show that the expected number of human fatalities is λtp.*

Solution to Exercise 5.5.14 . From the problem, a convenient model is

$$S_t = \sum_{i=1}^{N_t} Y_i,$$

where each Y_i is a Bernoulli random variable with parameter p.

By the law of total expectation we have, for $t \geq 0$,

$$G_{S_t}(u) = \mathrm{E}(u^{S_t}) = \mathrm{E}[\mathrm{E}(u^{S_t}|N_t)],$$

where we have for all $n \in \mathbb{N}_0$

$$\begin{aligned}
\mathrm{E}(u^{S_t}|N_t = n) &= \mathrm{E}\left(u^{\sum_{i=1}^{N_t} Y_i}\middle| N_t = n\right) = \mathrm{E}\left(u^{\sum_{i=1}^{n} Y_i}\middle| N_t = n\right) \\
&\overset{(Y_i),N \text{ indep.}}{=} \mathrm{E}\left(u^{\sum_{i=1}^{n} Y_i}\right) \\
&\overset{(Y_i) \text{ independent}}{=} \prod_{i=1}^{n} \mathrm{E}(u^{Y_1}) \\
&\overset{(Y_i) \text{ identically distr.}}{=} \left[\mathrm{E}(u^{Y_1})\right]^n = \left[G_{Y_1}(u)\right]^n.
\end{aligned}$$

Hence

$$\mathrm{E}(u^{S_t}|N_t) = \left[G_{Y_1}(u)\right]^{N_t}$$

and

$$G_{S_t}(u) = \mathrm{E}[\mathrm{E}(u^{S_t}|N_t)] = \mathrm{E}[\left[G_{Y_1}(u)\right]^{N_t}] = G_{N_t}(G_{Y_1}(u)).$$

Next we need to derive the pgfs of Y_1 and N_t (or recall them from Y1!): We have

$$\begin{aligned}
G_{Y_1}(u) &= \mathrm{E}(u^{Y_1}) = u^0\mathrm{P}(Y_1 = 0) + u^1\mathrm{P}(Y_1 = 1) \\
&= 1 - p + up = 1 + p(u-1) =: z,
\end{aligned}$$

and

$$G_{N_t}(z) = \sum_{n=0}^{\infty} z^n \mathrm{P}(N_t = n) = \sum_{n=0}^{\infty} z^n \frac{(\lambda t)^n}{n!} e^{-\lambda t} = \sum_{n=0}^{\infty} \frac{(z\lambda t)^n}{n!} e^{-\lambda t}$$

$$= \exp(\lambda t(z-1)).$$

Hence

$$G_{S_t}(u) = G_{N_t}(G_{Y_1}(u)) = \exp(\lambda t(z-1)) = \exp(\lambda tp(u-1)).$$

We deduce that S_t follows the Poisson distribution with parameter λtp. Also,

$$\mathrm{E}(S_t) = \frac{d}{du} G_{S_t}(u)\Big|_{u=1} = e^{\lambda tp(u-1)}\lambda tp\Big|_{u=1} = \lambda tp.$$

□

Remark 5.5.15. *In the exercise above, we could have worked with a thinned Poisson process instead of the compound Poisson process with Bernoulli jumps. Both processes are equivalent.*

Whilst this application is perhaps a little 'unrealistic', compound Poisson processes are used in a variety of important applications in insurance and finance (for example). In financial applications, they are often used in stochastic volatility models, to help reflect jumps in a volatility (standard deviation of financial instruments) process. We will study an example from insurance mathematics in the following.

5.6 The Cramér-Lundberg model in insurance mathematics

The compound Poisson process is often used in insurance mathematics to model the total amount of insurance claims. Let us study the Cramér-Lundberg model, which can be regarded as the basic insurance risk model.

Note that you can find more details in the excellent textbooks Embrechts et al. (1997) and Mikosch (2009).

Definition 5.6.1. *The **Cramér-Lundberg model** is given by the following five conditions.*

1. *The* claim size process *is denoted by $Y = (Y_k)_{k\in\mathbb{N}}$, where the Y_k denote positive i.i.d. random variables with finite mean $\mu = \mathrm{E}(Y_1)$, and variance $\sigma^2 = \mathrm{Var}(Y_1) \leq \infty$.*
2. *The* claim times *occur at the random instants of time*
$$0 < J_1 < J_2 < \cdots \ a.s..$$
3. *The* claim arrival process *is denoted by*
$$N_t = \sup\{n \in \mathbb{N} : J_n \leq t\}, \quad t \geq 0,$$
which is the number of claims in the interval $[0,t]$. (Note that $\sup\emptyset := 0$).
4. *The* inter-arrival times *are denoted by*
$$H_1 = J_1,\ H_k = J_k - J_{k-1},\ k = 2,3,\ldots,$$
and are independent and exponentially distributed with parameter λ.
5. *The sequences (Y_k) and (H_k) are independent of each other.*

Exercise 5.6.2. *Convince yourself that the process $\{N\}_{t\geq 0}$ defined above is a equivalent to the Poisson process defined in the (third) Definition 5.4.4.*

Definition 5.6.3. *The **total claim amount** is defined as the process $(S_t)_{t\geq 0}$ satisfying*

$$S_t = \begin{cases} \sum_{i=1}^{N_t} Y_i, & N_t > 0, \\ 0, & N_t = 0. \end{cases}$$

We observe that the total claim amount is modelled as a compound Poisson process.
We can derive the *total claim amount distribution*.

Theorem 5.6.4. *The total claim amount distribution is given by*

$$\mathrm{P}(S_t \leq x) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathrm{P}\left(\sum_{i=1}^{n} Y_i \leq x \right), \qquad x \geq 0,\ t \geq 0,$$

and $\mathrm{P}(S_t \leq x) = 0$ *for* $x < 0$.

Proof. Let $x \geq 0$, then

$$\begin{aligned}
\mathrm{P}(S_t \leq x) &\overset{\text{Law of total prob.}}{=} \sum_{n=0}^{\infty} \mathrm{P}(S_t \leq x, N_t = n) \\
&= \sum_{n=0}^{\infty} \mathrm{P}\left(\sum_{i=1}^{n} Y_i \leq x, N_t = n \right) \\
&\overset{(Y_i), N \text{ independent}}{=} \sum_{n=0}^{\infty} \mathrm{P}\left(\sum_{i=1}^{n} Y_i \leq x \right) \mathrm{P}(N_t = n) \\
&\overset{N_t \sim \mathrm{Poi}(\lambda t)}{=} \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathrm{P}\left(\sum_{i=1}^{n} Y_i \leq x \right).
\end{aligned}$$

□

Definition 5.6.5. *The* ***risk process*** $\{U_t\}_{t \geq 0}$ *is defined as*

$$U_t = u + ct - S_t, \qquad t \geq 0,$$

where $u \geq 0$ *stands for the* ***initial capital*** *and* $c > 0$ *denotes the* ***premium income rate***.

Exercise 5.6.6. *Draw a sample path of the risk process* U*!*

Now we can define the ruin probability.

Definition 5.6.7.

1. *The* ***ruin probability in finite time*** *is given by*

$$\psi(u, T) = \mathrm{P}(U_t < 0 \text{ for some } t \leq T),\ 0 < T < \infty, u \geq 0.$$

2. *The* ***ruin probability in infinite time*** *is given by*

$$\psi(u) := \psi(u, \infty), u \geq 0.$$

We can derive a useful result:

Theorem 5.6.8.

$$\mathrm{E}(U_t) = u + ct - \lambda t \mu = u + (c - \lambda \mu)t.$$

We can use the above result in order to come up with a first guess on how to choose the premium rate c: Note that we wish to choose c such that the ruin probability $\psi(u, T)$ (for given u and T) is "small".
A minimal requirement when choosing the premium could be

$$c > \lambda \mu,$$

which is often referred to as the **net profit condition**.

It implies that the risk process has positive mean (for all $t \geq 0$), i.e. the premium income is sufficiently high to cover the claim payments. Also

$$\lim_{t\to\infty} \frac{\mathrm{E}(U_t)}{t} = c - \lambda\mu > 0.$$

One can show that if the net profit condition is not satisfied, then ruin is certain (i.e. the ruin probability is 1) in the Cramér-Lundberg model.

We illustrate the importance of the net profit condition in Figure 5.10. Here we simulate 10 sample paths of the risk process U when the net profit condition is not satisfied (left) and when it is satisfied (right). We observe a strikingly different behaviour of the sample paths, where ruin happens for all 10 sample paths considered in the time interval of interest (on the left), whereas we do not observe a ruin on the right hand side when the net profit condition is satisfied. We note that a simulation with only 10 paths is very small and you should typically consider many more paths before drawing any conclusions!

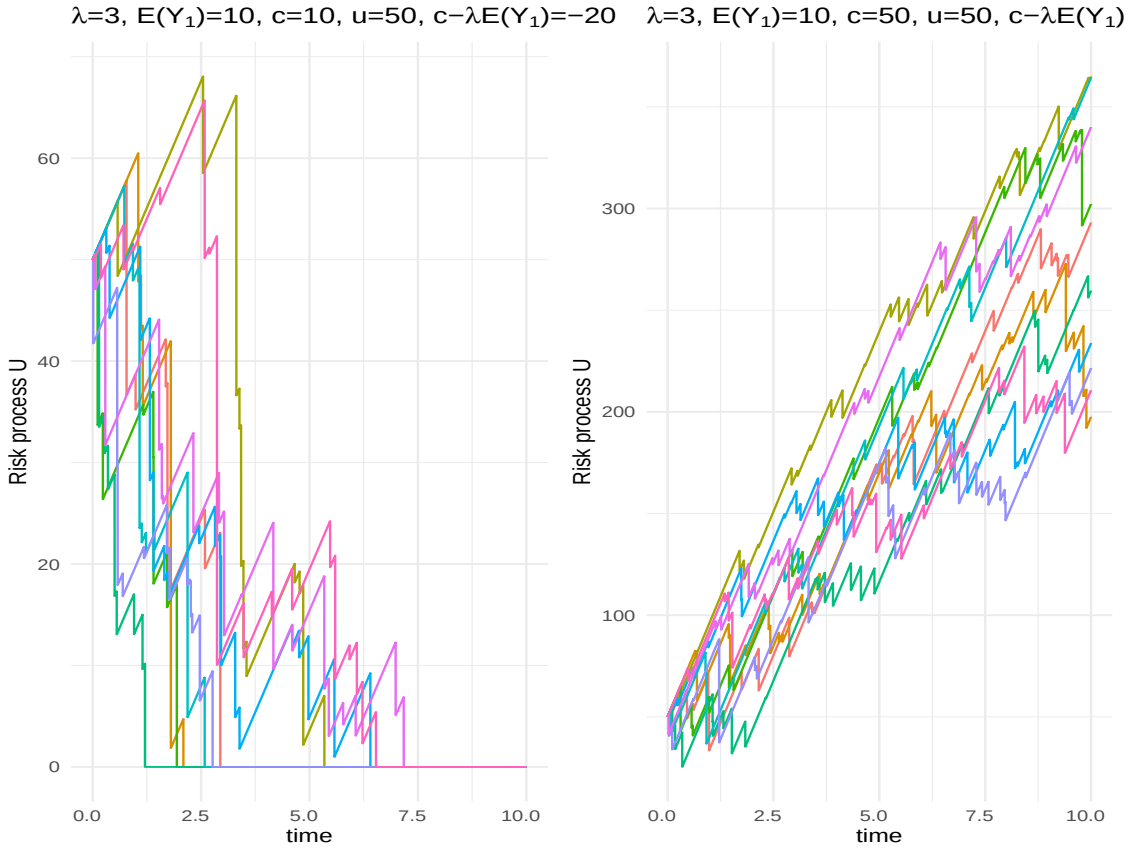


Figure 5.10: Each picture depicts 10 realisations of the risk process $U_t = u + ct - S_t$, where in both cases we have that $\lambda = 3$, $\mathrm{E}(Y_1) = 10$ (where the Y_i follow an exponential distribution), $u = 50$. The premium income rate varies and is set to $c = 10$ on the left and to $c = 50$ on the right. The corresponding net profits are given by -20 and 20, respectively.

5.7 The coalescent process

To finish the chapter, we will look at a genuine application of the Poisson process, which is used in population genetics. Due to the complexities of such models, we will look at the most basic ideas, but hopefully, it is clear that Poisson processes play a vital role in stochastic modelling of real processes.

5.7.1 Problem

- We are given a collection of n individuals, for which we have observed (for example) a DNA sequence from that individual.
- A DNA sequence is a collection of letters A, C, T and G, for simplicity, we assume that only one letter (base) is observed.
- The coalescent process helps to provide a genealogical tree representation of this data. That is, a tree like structure that, probabilistically represents the history of the individuals backward in time, where individuals join together (coalesce) until there is only one individual: the most recent common ancestor.

A coalescent tree can be seen in Figure 5.11, which also describes mutation (although we do not consider this below).

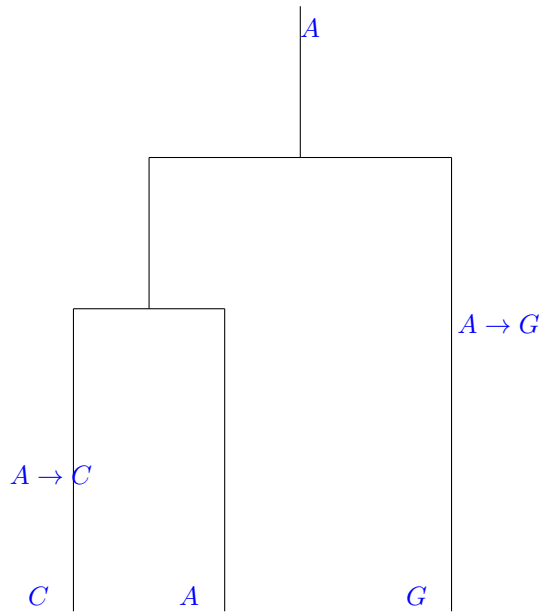


Figure 5.11: A Coalescent graph. The letters denote the types of the three observed chromosomes. Going up the figure (backward in time), the points where the graph join are coalescent events and the arrows denote a mutation of the type of a chromosome to another (forward in time).

5.7.2 Process

We will describe a very water-downed version of the coalescent process see for example the article Nordborg (2000), and also Nordborg (2004), for a detailed introduction.

- At the beginning of the process there are n ($n \in \mathbb{N}, n \geq 2$) individuals (all of the same DNA base), and
- each pair of individuals coalesce according to an (independent) Poisson process of rate 1.
- Since there are

$$\binom{n}{2}$$

pairs, the time to the first coalescent event is an exponential random variable of rate $\binom{n}{2}$, since we are considering the minimum of $\binom{n}{2}$ independent Exp(1)-distributed random variables, see Theorem 4.1.4.

- At the first event, two individuals are picked uniformly at random and combined.
- This continues until there is only one individual (the most recent common ancestor);
- hence there are $n-1$ coalescent events.
- The model, assumes all individuals have the same DNA base, so clearly another mechanism is needed for real data - a mutation process.
- Note that in this process the number of individuals decreases, and is our first example of a *death* process.

5.7.3 Time to the most recent common ancestor

The time to the most recent common ancestor, i.e. the height of the tree, can be estimated by

$$\mathrm{E}\left(\sum_{k=1}^{n-1} H_k\right), \quad \text{for} \quad n \in \mathbb{N}, n \geq 2,$$

where H_k is the time to k^{th} coalescence.

Since

$$H_k \sim \mathrm{Exp}\left(\binom{n-(k-1)}{2}\right) \qquad \Rightarrow \qquad \mathrm{E}(H_k) = \left(\binom{n-(k-1)}{2}\right)^{-1},$$

it follows that

$$\begin{aligned}
\mathrm{E}\left(\sum_{k=1}^{n-1} H_k\right) &= \sum_{k=1}^{n-1} \mathrm{E}(H_k) \\
&= \sum_{k=1}^{n-1} \left(\frac{(n-k+1)!}{(n-k-1)!2!}\right)^{-1} = \sum_{k=1}^{n-1} \frac{2(n-k-1)!}{(n-k+1)!} \\
&= \sum_{k=1}^{n-1} \frac{2}{(n-k+1)(n-k)} = \sum_{k=1}^{n-1} \frac{2}{k(k+1)}.
\end{aligned}$$

Note that

$$\sum_{k=1}^{n-1} \frac{2}{k(k+1)} = 2\left(1 - \frac{1}{n}\right).$$

Further, since $H_{n-1} \sim \mathrm{Exp}(\binom{2}{2})$, we find that

$$\mathrm{E}(H_{n-1}) = 1.$$

We interpret this finding as follows: The expected time during which there are only two branches is greater than half of the expected total tree height.

End of lecture 17.

Chapter 6

Continuous-time Markov chains

We now look at Markov chains in continuous time, i.e. $\mathcal{T} = [0, \infty)$, but on a countable state-space E, where often $E \subset \mathbb{Z}$ with $K = \text{Card}(E)$. As in the discrete case, we could in fact choose any finite or countably infinite set as the state space. These processes are much more complex than in discrete time (in particular if $K = \infty$), and it will take much more sophisticated machinery to deal with such processes rigorously. As a result, we will sometimes only discuss the intuition behind some proofs rather than all technical details. This chapter is based on Grimmett & Stirzaker (2001*b*, Chapters 6.9 & 6.11).

6.1 Some definitions

We begin with a basic definition. Throughout, we assume that E is a countable state space.

Definition 6.1.1. *A continuous-time process $\{X_t\}_{t\in[0,\infty)}$ satisfies the **Markov property** if*

$$\mathrm{P}(X_{t_n} = j | X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}) = \mathrm{P}(X_{t_n} = j | X_{t_{n-1}} = i_{n-1})$$

for all $j, i_1, \ldots, i_{n-1} \in E$ and for any sequence $0 \leq t_1 < \cdots < t_n < \infty$ of times (with $n \in \mathbb{N}$).

Comparing with the definition of Markov chains in discrete time, we can see that the main modification is with the inclusion of the process at a finite number of times. If we think, intuitively, the process is a path and hence we consider the dependence upon the path on a finite number of points up-to t_{n-1}; there is a technical way to describe this dependence through *filtrations*, again, we do not explore this definition, and we restrict ourselves to **finite** dimensional behaviour.

In discrete-time, we looked at the mechanics of the chain via the transition matrix. However, in continuous-time, there is no direct analogue; there is no notion of unit time. The way out is to use the idea of the **generator**, and we now look to introduce this concept.

Definition 6.1.2. *The **transition probability** $p_{ij}(s,t)$ is, for $s \leq t$, $i, j \in E$*

$$p_{ij}(s,t) = \mathrm{P}(X_t = j | X_s = i)$$

*and, in addition, the chain is **homogeneous** if*

$$p_{ij}(s,t) = p_{ij}(0, t-s)$$

writing $p_{ij}(t-s) = p_{ij}(s,t)$ in this case.

> **From herein, it is assumed that the chain is homogeneous and the probabilities are continuous in t.**

Let $\mathbf{P}_t = (p_{ij}(t)))$. Then we have the following result:

Theorem 6.1.3. *The family $\{\mathbf{P}_t : t \geq 0\}$ is a **stochastic semigroup**; that is, it satisfies*

1. $\mathbf{P}_0 = I_{K\times K}$, *the identity*
2. $\mathbf{P}_t$ *is stochastic, that is* $\mathbf{P}_t$ *has non-negative entries with rows summing to 1.*
3. *the Chapman-Kolmogorov equations hold:* $\mathbf{P}_{s+t} = \mathbf{P}_s\mathbf{P}_t$ *for all* $s, t \geq 0$.

Proof. 1. Part 1 follows since for $i, j \in E$, $p_{ii}(0) = 1$ and $p_{ij}(0) = 0$ for $i \neq j$.

2. Then (using the law of total probability), for all $i \in E$,

$$\sum_{j\in E} p_{ij}(t) = \sum_{j\in E} \frac{\mathrm{P}(X_t = j, X_0 = i)}{\mathrm{P}(X_0 = i)} = \frac{\mathrm{P}(X_0 = i)}{\mathrm{P}(X_0 = i)} = 1.$$

3. Using the law of total probability (with additional conditioning) and the Markov property, we have, for all $i, j \in E, s, t \geq 0$,

$$\begin{aligned} p_{ij}(s+t) &= \mathrm{P}(X_{s+t} = j|X_0 = i) \\ &= \sum_{k\in E} \mathrm{P}(X_{s+t} = j|X_s = k, X_0 = i)\mathrm{P}(X_s = k|X_0 = i) \\ &= \sum_{k\in E} \mathrm{P}(X_{s+t} = j|X_s = k)\mathrm{P}(X_s = k|X_0 = i) \\ &= \sum_{k\in E} p_{ik}(s)p_{kj}(t). \end{aligned}$$

□

As in the discrete-time case, the evolution of the Markov chain is specified by the stochastic semigroup $\{\mathbf{P}_t\}$ and the distribution of X_0.

Warning: We will *not* study the general theory of continuous-time Markov chains in this course in detail, but rather focus on some applications. Hence, we only sketch some important results in the following without giving all technical conditions and rigorous proofs!

We have not yet defined the generator, but note that much of the transition dynamics of the Markov chain can be expressed in terms of the semigroup. The continuity assumption can be expressed as follows:

Definition 6.1.4. *The semigroup* $\{\mathbf{P}_t\}$ *is called* ***standard*** *if*

$$\lim_{t\downarrow 0} \mathbf{P}_t = \mathbf{I} \quad (= \mathbf{P_0}),$$

where $\mathbf{I} = \mathbf{I}_{K\times K}$ *denotes the* $K \times K$*-dimensional identity matrix.*

Note that a semigroup is standard if and only if its elements $p_{ij}(t)$ are continuous functions in t, cf. Exercise 4- 35. (Recall: A function f is continuous in $y \in \mathbb{R}$ if, $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\forall x \in \mathbb{R}$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \epsilon$).

In the following, we only consider Markov chains with standard semigroups of transition probabilities.

Exercise 6.1.5. *Show that a Poisson process* $\{N_t\}_{t\geq 0}$ *with rate* $\lambda > 0$ *is a Markov chain in continuous time.*

Solution to Exercise 6.1.5. This result is an immediate consequence of the independent increment property. To see this, note that for any $j, i_1, \ldots, i_{n-1} \in E$ and for any sequence $0 \leq t_1 < \cdots < t_n < \infty$ of times (with $n \in \mathbb{N}$), we have

$$\begin{aligned} &\mathrm{P}(N_{t_n} = j|N_{t_1} = i_1, \ldots, N_{t_{n-1}} = i_{n-1}) \\ &= \frac{\mathrm{P}(N_{t_n} = j, N_{t_{n-1}} = i_{n-1}, \ldots, N_{t_1} = i_1)}{\mathrm{P}(N_{t_{n-1}} = i_{n-1}, \ldots, N_{t_1} = i_1)} \end{aligned}$$

$$= \frac{\mathrm{P}(N_{t_n} - N_{t_{n-1}} = j - i_{n-1}, N_{t_{n-1}} - N_{t_{n-2}} = i_{n-1} - i_{n-2}, \ldots, N_{t_2} - N_{t_1} = i_2 - i_1, N_{t_1} = i_1)}{\mathrm{P}(N_{t_{n-1}} - N_{t_{n-2}} = i_{n-1} - i_{n-2}, \ldots, N_{t_2} - N_{t_1} = i_2 - i_1, N_{t_1} = i_1)}$$
$$\overset{\text{indep. incr.}}{=} \frac{\mathrm{P}(N_{t_n} - N_{t_{n-1}} = j - i_{n-1})\mathrm{P}(N_{t_{n-1}} - N_{t_{n-2}} = i_{n-1} - i_{n-2}) \cdots \mathrm{P}(N_{t_2} - N_{t_1} = i_2 - i_1)\mathrm{P}(N_{t_1} = i_1)}{\mathrm{P}(N_{t_{n-1}} - N_{t_{n-2}} = i_{n-1} - i_{n-2}) \cdots \mathrm{P}(N_{t_2} - N_{t_1} = i_2 - i_1)\mathrm{P}(N_{t_1} = i_1)}$$
$$= \mathrm{P}(N_{t_n} - N_{t_{n-1}} = j - i_{n-1}).$$

Also,

$$\mathrm{P}(N_{t_n} = j | N_{t_{n-1}} = i_{n-1}) = \frac{\mathrm{P}(N_{t_n} = j, N_{t_{n-1}} = i_{n-1})}{\mathrm{P}(N_{t_{n-1}} = i_{n-1})} = \frac{\mathrm{P}(N_{t_n} - N_{t_{n-1}} = j - i_{n-1}, N_{t_{n-1}} = i_{n-1})}{\mathrm{P}(N_{t_{n-1}} = i_{n-1})}$$
$$\overset{\text{indep. incr.}}{=} \frac{\mathrm{P}(N_{t_n} - N_{t_{n-1}} = j - i_{n-1})\mathrm{P}(N_{t_{n-1}} = i_{n-1})}{\mathrm{P}(N_{t_{n-1}} = i_{n-1})} = \mathrm{P}(N_{t_n} - N_{t_{n-1}} = j - i_{n-1}).$$

Hence, the Markov condition is satisfied. □

Example 6.1.6. *Let us consider a Poisson process $\{N_t\}_{t\geq 0}$ with rate $\lambda > 0$. Let $i, j \in \mathbb{N}_0$. The transition probabilities are given by $p_{ij}(t) = 0$ for $i > j$, and for $i \leq j$ by*

$$p_{ij}(t) = \mathrm{P}(N_{s+t} = j | N_s = i) = \mathrm{P}(N_{s+t} - N_s = j - i)$$
$$\overset{\text{stat. incr.}}{=} \mathrm{P}(N_t = j - i) = \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda(j-i)}.$$

I.e. the stochastic semigroup of the Poisson process with rate λ is given by

$$\mathbf{P}_t = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & \cdots \end{array} \\ \begin{pmatrix} e^{-\lambda t} & (\lambda t)e^{-\lambda t} & (\lambda t)^2 e^{-\lambda t}/2 & \cdots \\ 0 & e^{-\lambda t} & (\lambda t)e^{-\lambda t} & \vdots \\ 0 & 0 & e^{-\lambda t} & \ddots \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix} \end{array}.$$

6.2 Holding times and alarm clocks

6.2.1 Holding times

We will now introduce the concept of holding times and show that they are exponentially distributed. Suppose $\{X_t\}_{t\geq 0}$ is a continuous-time homogeneous Markov chain.

Suppose that $t \geq 0$ and, for $i \in E$, we have $X_t = i$. Given $X_t = i$, define

$$H_{|i} = \inf\{s \geq 0 : X_{t+s} \neq i\}$$

to be the **holding time at state** i, that is the length of time that a continuous-time Markov chain started in state i stays in state i before transitioning to a new state. Note that the holding time does not depend on t since we work under the time-homogeneity assumption. I.e.

$$\inf\{s \geq 0 : X_{t+s} \neq i\} | X_t = i \overset{\text{distr.}}{=} \inf\{s \geq 0 : X_s \neq i\} | X_0 = i.$$

Remark 6.2.1. *Note that we denote by H_i the ith holding time (in time) and by $H_{|i}$ the holding time associated with state i which are not in general the same objects. E.g. for a Poisson process (starting at 0), we would have that $H_1 = H_{|0}, H_2 = H_{|1}, \ldots$, for a birth process starting in k say, i.e. $N_0 = k$, we would have, $H_1 = H_{|k}, H_2 = H_{|k+1}, \ldots$. For a general Markov chain we have for $n \in \mathbb{N}$, given that $X_{J_{n-1}+} = i$ (which is the right-limit of X at time J_{n-1}), that $H_n = H_{|X_{J_{n-1}+}} = H_{|i}$ for $i \in \mathbb{N}$, i.e. we associate*

$$H_n | X_{J_{n-1}+} = H_{|X_{J_{n-1}+}}.$$

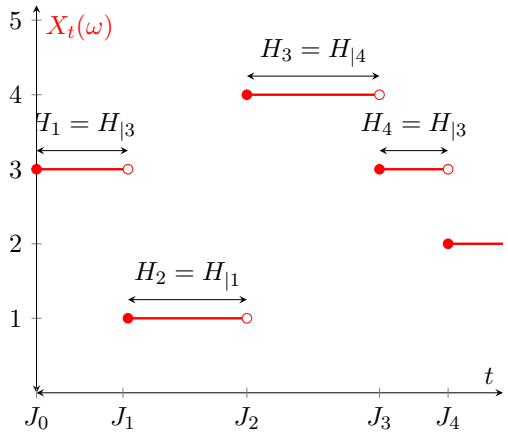


Figure 6.1: This picture clarifies the notation we use for holding times: We depict a path of a Markov chain with its corresponding holding times and jump times (where we suppress the ωs to make the picture easier to read).

Theorem 6.2.2. *The holding time $H_{|i}$, for $i \in E$, follows an exponential distribution.*

Proof. By time homogeneity assume without loss of generality that $X_0 = i$. Note that for $x, y \geq 0$, we have

$$\{H_{|i} > x\} \overset{\text{distr.}}{=} \{X_t = i, \text{ for } 0 \leq t \leq x\},$$
$$\{H_{|i} > x + y\} \overset{\text{distr.}}{=} \{X_t = i, \text{ for } 0 \leq t \leq x + y\}.$$

Then

$$\begin{aligned}
&\mathrm{P}(H_{|i} > x + y | H_{|i} > x) \\
&= \mathrm{P}(X_t = i, \text{ for } 0 \leq t \leq x + y | X_t = i, \text{ for } 0 \leq t \leq x) \\
&= \frac{\mathrm{P}(X_t = i, \text{ for } 0 \leq t \leq x, X_t = i, \text{ for } x < t \leq x + y)}{\mathrm{P}(X_t = i, \text{ for } 0 \leq t \leq x)} \\
&= \mathrm{P}(X_t = i, \text{for } x < t \leq x + y | X_t = i, \text{ for } 0 \leq t \leq x) \\
&\overset{\text{Markov}}{=} \mathrm{P}(X_t = i, \text{for } x < t \leq x + y | X_x = i) \\
&\overset{\text{time-hom.}}{=} \mathrm{P}(X_t = i, \text{for } 0 < t \leq y | X_0 = i) = \mathrm{P}(H_{|i} > y).
\end{aligned}$$

Hence $H_{|i}$ has a continuous distribution satisfying the lack of memory property, hence it follows the exponential distribution.

□

End of lecture 18. .....

6.2.2 Describing the evolution of a Markov chain using exponential holding times

Let us now describe how a continuous-time Markov chain evolves in time using the concept of exponential alarm clocks. The following presentation follows closely Dobrow (2016, p.269–272).

For each $i \in E$, let us denote by $q_i > 0$, the parameter of the exponential distribution associated with $H_{|i}$. We shall assume that $0 < q_i < \infty$.

Technically, the boundary cases could be considered, but we will not do so in the following: If $q_i = 0$, then that would mean that i is an **absorbing** state, i.e. when i gets visited, the process stays there forever. If $q_i = \infty$, then i would be an **instantaneous** state, meaning that the process would leave i immediately after visiting it. This would allow for infinitely many transitions in a finite time interval and hence would lead to explosion.

Assuming that the Markov chain is neither absorbing nor explosive, then its evolution in time can be described as follows.

- Suppose the process starts in state i.
- It stays in i for an exponentially distributed length of time with parameter q_i, i.e. the mean length of stay in state i is $\mathrm{E}(H_{|i}) = 1/q_i$ time units.
- Then it transitions into a state $j \neq i$, with some probability which we denote by p_{ij}^Z.
- It stays in j for an exponentially distributed length of time with parameter q_j, i.e. the mean length of stay in state j is $\mathrm{E}(H_{|j}) = 1/q_j$ time units.
- Then it transitions into a state $k \neq j$, with some probability which we denote by p_{jk}^Z.

If we ignore time, we see a sequence $Z_0, Z_1, Z_2, \ldots$, where Z_n denotes the nth state visited by the continuous-time Markov chain X. As mentioned earlier $Z = (Z_n)_{n\in\mathbb{N}_0}$ denotes the **jump chain**, which is sometimes also called the **embedded chain** associated with X. The jump chain Z is a discrete-time Markov chain with transition matrix $\mathbf{P}^Z = (p_{ij}^Z)_{i,j\in E}$, where the diagonal elements are all zero, i.e. $p_{ii}^Z = 0$ for all $i \in E$.

[In the case of i being an absorbing state, we would have that $p_{ii}^Z = 1$.]

Example 6.2.3. *Let us consider a Poisson process with rate $\lambda > 0$. Here, $q_i = \lambda$ for all $i \in \mathbb{N}_0$. The jump chain is given by $Z_n = n$ for $n \in \mathbb{N}_0$ and the transition matrix of the jump chain is given by*

$$\mathbf{P}^Z = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & \cdots \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & \vdots \\ 0 & 0 & 0 & 1 \ddots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix} \end{array}.$$

6.2.3 Describing the evolution of a Markov chain using exponential alarm clocks and transition rates

Another way of describing the evolution of continuous-time Markov chains is by specifying **transition rates** between states and using the concept of **exponential alarm clocks**:

- For each state $i \in E$, we denote by n_i the number of states which can be reached from state i.
- We associate n_i independent, exponential alarm clocks with rates q_{ij} provided that state j can be reached from state i.
- When the chain first visits state i, all n_i exponential alarm clocks will be set simultaneously.
- The first alarm clock which rings, determines which state the chain transitions to. Suppose the clock with rate q_{ij} rings first, then the chain moves to state j.
- As soon as state j has been reached, we set the n_j independent exponential alarm clocks associated with state j and continue as before.

We call the q_{ij} the **transition rates** of the process.

We will now describe how we can derive the holding time parameters and the transition probabilities of the embedded chain from the transition rates.

- Let $i \neq j$ and let $q_{ij} > 0$ denote the transition rates when state j can be reached from state i.
- Let $i \neq j$, set $q_{ij} = 0$ if j cannot be reached from i.
- Also, we set $q_{ii} = 0$ for all $i \in E$.
- Suppose the process starts in state i. We set the n_i exponential alarm clocks and wait for the first one to ring.
- According to Theorems 4.1.4 and 4.1.5, the minimum/infimum of these exponential alarm clocks follows an exponential distribution with rate

$$q_i = \sum_{j \in E} q_{ij}.$$

- The probability that the chain moves from i to j is the probability that the minimum/infimum of the exponential waiting times until the first alarm clocks rings is equal to the waiting time until the alarm clocks associated with rate q_{ij} rings. According to Theorems 4.1.4 and 4.1.5, this probability is equal to

$$\frac{q_{ij}}{q_i}.$$

- Hence, the transition probabilities of the embedded chain Z are given by

$$p_{ij}^Z = \frac{q_{ij}}{q_i}.$$

In the above construction, we assumed again that $0 < q_i < \infty$. In the case of $q_i = 0$ (i.e. when i is an absorbing state), then $p_{ii}^Z = 1$.

Similar to the transition diagrams for discrete-time Markov chains, we can draw transition diagrams using the transition rates of the continuous-time Markov chains:

Example 6.2.4. *Consider the three-state Markov chain with $E = \{1, 2, 3\}$ and transition rates and holding time parameters given by*

$$(q_1, q_2, q_3) = (q_{12} + q_{13}, q_{21} + q_{23}, q_{31} + q_{32}).$$

We assume that $q_i \neq 0$ for all $i = 1, 2, 3$.

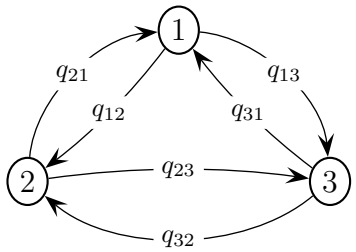


Figure 6.2: Transition diagram displaying the transition rates of a three state continuous-time Markov chain.

The transition matrix of the embedded chain is given by

$$\mathbf{P}^Z = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} 0 & q_{12}/q_1 & q_{13}/q_1 \\ q_{21}/q_2 & 0 & q_{23}/q_2 \\ q_{31}/q_3 & q_{32}/q_3 & 0 \end{pmatrix} \end{array}.$$

6.3 The generator

We will now link the notion of transition rates to the concept of a generator for continuous-time Markov chains, see Dobrow (2016, p. 273-275).

Recall that the derivative of a function describes its rate of change. In the case of continuous-time Markov chains the derivative of the matrix of the transition probabilities is hence closely linked to the transition rates as we shall see in the following.

Recall that, as in discrete time, we also have that $\mathbf{P}_0 = \mathbf{I}$, which is the $\mathrm{card}(E) \times \mathrm{card}(E)$-identity matrix.

Definition 6.3.1. *The generator $\mathbf{G} = (g_{ij})_{i,j\in E}$ of the Markov chain with stochastic semigroup $\mathbf{P}_t$ is defined as the $\mathrm{card}(E) \times \mathrm{card}(E)$-matrix given by*

$$\mathbf{G} := \lim_{\delta \downarrow 0} \frac{1}{\delta}[\mathbf{P}_\delta - \mathbf{I}] = \lim_{\delta \downarrow 0} \frac{1}{\delta}[\mathbf{P}_\delta - \mathbf{P}_0],$$

that is, $\mathbf{P}_t$ is differentiable at $t = 0$.

We will now provide some heuristic arguments, to link the elements of the generator to the transition rates we studied earlier. Please note that the following discussion is **not** completely rigorous.

Suppose that $X_t = i$, then the instantaneous transition rate of hitting state $j \neq i$ is given by

$$\begin{aligned} &\lim_{\delta \downarrow 0} \frac{\mathrm{E}(\text{number of transitions to } j \text{ in } (t, t+\delta] | X_t = i)}{\delta} \\ &\overset{(\star)}{=} \lim_{\delta \downarrow 0} \frac{\mathrm{P}(X_{t+\delta} = j | X_t = i)}{\delta} = \lim_{\delta \downarrow 0} \frac{\mathrm{P}(X_\delta = j | X_0 = i)}{\delta} \\ &= \lim_{\delta \downarrow 0} \frac{p_{ij}(\delta)}{\delta} = \lim_{\delta \downarrow 0} \frac{p_{ij}(\delta) - p_{ij}(0)}{\delta} = p'_{ij}(0), \end{aligned}$$

where equality $(\star)$ can be justified as follows: If δ is very small, then the number of transitions in a time interval of length δ is either 0 or 1. The $p'_{ij}(0)$ instantaneous transition rates are identical to the transition rates q_{ij} we introduced above. I.e.

$$g_{ij} = q_{ij} = p'_{ij}(0).$$

Note that we can also say that, for $i \neq j$,

$$p_{ij}(\delta) \approx g_{ij}\delta.$$

Also, we have, for all $i \in E$,

$$\begin{aligned} g_{ii} = p'_{ii}(0) &= \lim_{\delta \downarrow 0} \frac{p_{ii}(\delta) - p_{ii}(0)}{\delta} = \lim_{\delta \downarrow 0} \frac{p_{ii}(\delta) - 1}{\delta} \\ &= \lim_{\delta \downarrow 0} \frac{-\sum_{j\in E, j\neq i} p_{ij}(\delta)}{\delta} \overset{(\star\star)}{=} -\sum_{j\in E, j\neq i} \lim_{\delta \downarrow 0} \frac{p_{ij}(\delta)}{\delta} = -\sum_{j\in E, j\neq i} p'_{ij}(0) \\ &= -\sum_{j\in E, j\neq i} q_{ij} = -\sum_{j\in E} q_{ij} = -q_i, \end{aligned}$$

where we assume for our heuristic discussion that the interchange of limit and (potentially infinite) sum in $(\star\star)$ is valid. Note that we can also conclude that

$$p_{ii}(\delta) \approx 1 + g_{ii}\delta.$$

6.3.1 Transition probabilities of the associated jump chain

We can now derive the transition probabilities of the embedded/jump chain and express them in terms of the generator:

If $X_t = i$ it stays there for an exponentially distributed time with rate $-g_{ii} = q_i$ (assuming $q_i \neq 0$), and then moves to some other state j. The probability that the chain jumps to $j \neq i$ is $-g_{ij}/g_{ii}$. To see this, note that we have for $\delta > 0$ and $i \neq j$ that

$$\begin{aligned}
&\mathrm{P}(X_{t+\delta} = j | X_t = i, X_{t+\delta} \neq i) \\
&= \frac{\mathrm{P}(X_{t+\delta} = j, X_t = i, X_{t+\delta} \neq i)}{\mathrm{P}(X_t = i, X_{t+\delta} \neq i)} \\
&= \frac{\mathrm{P}(X_{t+\delta} = j, X_t = i)\mathrm{P}(X_t = i)}{\mathrm{P}(X_t = i, X_{t+\delta} \neq i)\mathrm{P}(X_t = i)} \\
&= \frac{\mathrm{P}(X_{t+\delta} = j | X_t = i)}{\mathrm{P}(X_{t+\delta} \neq i | X_t = i)} = \frac{p_{ij}(\delta)}{1 - p_{ii}(\delta)} \to -\frac{g_{ij}}{g_{ii}} \qquad \text{as } \delta \downarrow 0.
\end{aligned}$$

I.e., for $i \neq j$,

$$p^Z_{ij} = -\frac{g_{ij}}{g_{ii}} = \frac{q_{ij}}{q_i},$$

which is equivalent to

$$q_{ij} = q_i p^Z_{ij}.$$

Note that for a discrete-time Markov chain, the transition matrix together with the initial distribution characterises the probabilistic properties of the process completely. In continuous-time, the generator plays the role of the transition matrix and together with the initial distribution characterises the probabilistic properties of the process completely. However, it is important not to confuse these two. While the transition matrix for a discrete-time chain is a stochastic matrix consisting of (1-step) transition probabilities, the generator is not a stochastic matrix (it has negative elements on the diagonal) and its row sums are (typically) equal to 0 and not to 1. Also, transition rates are not probabilities, so they can take any positive value and are not bounded between 0 and 1.

End of lecture 19.

6.4 The forward and backward equations

We have seen that it is possible to find $\mathbf{G}$, given $\{\mathbf{P}_t\}$, by setting $\mathbf{G} = \mathbf{P}'_0$, but the converse is also usually true. The matrix of transition functions $\{\mathbf{P}_t\}$ can be derived for a given generator $\mathbf{G}$ by solving either Kolmogorov's forward or backward equations (subject to regularity conditions):

Theorem 6.4.1. *Subject to regularity conditions, see Section 6.9, a continuous-time Markov chain with stochastic semigroup $\{\mathbf{P}_t\}$ and generator $\mathbf{G}$ satisfies the* ***Kolmogorov forward equation***

$$\mathbf{P}'_t = \mathbf{P}_t\mathbf{G},$$

*and the **Kolmogorov backward equation***

$$\mathbf{P}_t' = \mathbf{G}\mathbf{P}_t,$$

for all $t \geq 0$.

Proof. Let $t \geq 0$ and $\delta > 0$. Using the Chapman-Kolmogorov equations, we know that $\mathbf{P}_{t+\delta} = \mathbf{P}_t\mathbf{P}_\delta$. Hence

$$\frac{\mathbf{P}_{t+\delta} - \mathbf{P}_t}{\delta} = \frac{\mathbf{P}_t\mathbf{P}_\delta - \mathbf{P}_t}{\delta} = \frac{\mathbf{P}_t(\mathbf{P}_\delta - \mathbf{I})}{\delta} = \mathbf{P}_t\frac{(\mathbf{P}_\delta - \mathbf{P}_0)}{\delta}.$$

Hence, taking the limit $\delta \downarrow 0$ on both sides leads to

$$\mathbf{P}_t' = \lim_{\delta\downarrow 0}\frac{\mathbf{P}_{t+\delta} - \mathbf{P}_t}{\delta} = \mathbf{P}_t\mathbf{G}.$$

Also, using the Chapman-Kolmogorov equations again, we know that also $\mathbf{P}_{t+\delta} = \mathbf{P}_\delta\mathbf{P}_t$. Hence

$$\frac{\mathbf{P}_{t+\delta} - \mathbf{P}_t}{\delta} = \frac{\mathbf{P}_\delta\mathbf{P}_t - \mathbf{P}_t}{\delta} = \frac{(\mathbf{P}_\delta - \mathbf{I})\mathbf{P}_t}{\delta} = \frac{(\mathbf{P}_\delta - \mathbf{P}_0)}{\delta}\mathbf{P}_t.$$

Hence, taking the limit $\delta \downarrow 0$ on both sides leads to

$$\mathbf{P}_t' = \lim_{\delta\downarrow 0}\frac{\mathbf{P}_{t+\delta} - \mathbf{P}_t}{\delta} = \mathbf{G}\mathbf{P}_t.$$

□

As a result, we are able to express the semigroup in terms of the generator.

6.4.1 Matrix exponentials

It is often (but not always!) the case that the differential equations with boundary condition $\mathbf{P}_0 = \mathbf{I}$ can be solved uniquely, with a solution of the form

$$\mathbf{P}_t = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbf{G}^n. \tag{6.4.1}$$

Note that we have powers of matrices here and $\mathbf{G}^0 = \mathbf{I}$.

We can express (6.4.1) as

$$\mathbf{P}_t = e^{t\mathbf{G}},$$

where $e^{\mathbf{A}}$ is the abbreviation for $\sum_{n=0}^{\infty}\frac{1}{n!}\mathbf{A}^n$ for square matrices $\mathbf{A}$.

Hence

$$\mathbf{P}_t = e^{t\mathbf{G}} = \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbf{G}^n = \mathbf{I} + t\mathbf{G} + \frac{t^2}{2}\mathbf{G}^2 + \frac{t^3}{6}\mathbf{G}^3 + \cdots$$

It can be tricky in practice and computationally challenging to compute such matrix exponentials. The situation is significantly easier, if we can diagonalise $\mathbf{G}$ and hence diagonalise $e^{t\mathbf{G}}$.

Recall that $K = \text{card}(E)$, hence $\mathbf{G}$ is a $K \times K$-matrix. Suppose that $\mathbf{G}$ is diagonalisable with

$$\mathbf{G} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1},$$

where $\mathbf{D} = \text{diag}(\lambda_1, \ldots, \lambda_K)$ is a $K \times K$-diagonal matrix with the eigenvalues of $\mathbf{G}$ on the diagonal and the columns of the invertible matrix $\mathbf{S}$ contain the corresponding eigenvectors. Then, for any $n \in \mathbb{N}_0$, we have

$$\mathbf{G}^n = (\mathbf{S}\mathbf{D}\mathbf{S}^{-1})^n = \mathbf{S}\mathbf{D}^n\mathbf{S}^{-1} = \mathbf{S}\text{diag}(\lambda_1^n, \ldots, \lambda_K^n)\mathbf{S}^{-1},$$

and hence

$$\begin{aligned}\mathbf{P}_t = e^{t\mathbf{G}} &= \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{S}\text{diag}(\lambda_1^n, \ldots, \lambda_K^n)\mathbf{S}^{-1} \\ &= \mathbf{S}\sum_{n=0}^{\infty} \frac{t^n}{n!}\text{diag}(\lambda_1^n, \ldots, \lambda_K^n)\mathbf{S}^{-1} \\ &= \mathbf{S}\text{diag}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_1^n, \ldots, \sum_{n=0}^{\infty} \frac{t^n}{n!}\lambda_K^n\right)\mathbf{S}^{-1} \\ &= \mathbf{S}\text{diag}\left(e^{\lambda_1 t}, \ldots, e^{\lambda_K t}\right)\mathbf{S}^{-1},\end{aligned}$$

which is typically much easier to compute in practice.

Example 6.4.2. *Let $E = \{1, 2\}$, $\alpha, \beta \in \mathbb{R}^+$. There are two equivalent ways to describe the chain:*

1. *X has a generator*

$$G = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}.$$

2. *Use the holding times: If the chain is in state 1 (resp. 2), it stays there for an exponential time of parameter α (resp. β) before jumping to 2 (resp. 1).*

The forward equations take the form

$$p_{11}'(t) = -\alpha p_{11}(t) + \beta p_{12}(t)$$

and the system of equations may be solved to yield the transition probabilities.

6.5 Irreducibility, stationarity and limiting distribution

We also extend the notions of irreducibility, stationarity and the limiting distribution.

We can define accessibility, communication and irreducibility as in the discrete case.

Definition 6.5.1. *The chain is **irreducible** if for any $i, j \in E$ we have $p_{ij}(t) > 0$ for some t.*

We have the following result:

Theorem 6.5.2. *If $p_{ij}(t) > 0$, for some $t > 0$, then $p_{ij}(t) > 0$ for all $t > 0$.*

Proof. See Dobrow (2016, p. 285) for an explanation of the intuition behind the proof of this result and (Norris 1998, p. 111) for the proof. □

This result implies in particular that, in a continuous-time setting, the concept of periodicity is no longer relevant and is hence not considered/defined since all states are essentially "aperiodic".

Recall that a distribution is defined as a row vector with non-negative elements which add up to 1.

Definition 6.5.3. *A distribution $\boldsymbol{\pi}$ is the **limiting distribution** of a continuous-time Markov chain if, for all states $i, j \in E$, we have*

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j$$

Definition 6.5.4. *A distribution $\boldsymbol{\pi}$ is a **stationary distribution** if $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ for all $t \geq 0$.*

One can show, that if there exists a limiting distribution, then it is also a stationary distribution (as we discussed in the discrete-time case (for a finite state space)). The converse does not hold in general, but depends on the class structure of the Markov chain.

To find the stationary distribution, in discrete-time, we solved the vector equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. There is a similar situation in continuous-time, but, there is another way, through the generator.

Theorem 6.5.5. *Subject to regularity conditions, we have $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ for all $t \geq 0$ if and only if $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$.*

Proof. A sketch proof. Using (6.4.1) and $\mathbf{G}^0 = \mathbf{I}$:

$$\begin{aligned}
\boldsymbol{\pi}\mathbf{G} = \mathbf{0} &\Leftrightarrow \boldsymbol{\pi}\mathbf{G}^n = \mathbf{0} \quad \text{for all } n \in \mathbb{N} \\
&\Leftrightarrow \sum_{n=1}^{\infty} \frac{t^n}{n!}\boldsymbol{\pi}\mathbf{G}^n = \mathbf{0} \quad \text{for all } t \geq 0 \\
&\Leftrightarrow \boldsymbol{\pi}\sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n = \boldsymbol{\pi} \quad \text{for all } t \geq 0 \\
&\Leftrightarrow \boldsymbol{\pi}\mathbf{P}_t = \boldsymbol{\pi} \quad \text{for all } t \geq 0.
\end{aligned}$$

□

This helps us to find stationary distributions, given their existence.

As in discrete time, if $\boldsymbol{\nu}^{(t)}$ is the marginal distribution of X_t, then we have

$$\boldsymbol{\nu}^{(t)} = \boldsymbol{\nu}^{(0)}\mathbf{P}_t.$$

We finish the section with the ergodic theorem.

Theorem 6.5.6. *Let X be an irreducible Markov chain with a standard semigroup $\{\mathbf{P}_t\}$ of transition probabilities.*

1. *If there exists a stationary distribution $\boldsymbol{\pi}$ then it is unique and for any $i, j \in E$*

$$\lim_{t\to+\infty} p_{ij}(t) = \pi_j.$$

2. *If there is no stationary distribution then*

$$\lim_{t\to+\infty} p_{ij}(t) = 0$$

for all $i, j \in E$.

Note that this theorem holds exactly as stated. We did not skip any conditions here!

Proof. A sketch proof: Fix $\delta > 0$ and define $Y_n := X_{\delta n}$. Then one can show that $\{Y_n\}$ is an irreducible aperiodic discrete-time Markov chain, which we call *skeleton*. If Y is positive recurrent, then it has a unique stationary distribution $\boldsymbol{\pi}^{(\delta)}$ and

$$p_{ij}(n\delta) = \mathrm{P}(Y_n = j | Y_0 = i) \to \pi_j^{(\delta)}, \qquad \text{as } n \to \infty,$$

otherwise $p_{ij}(n\delta) \to 0$ as $n \to \infty$. Apply this argument to two *rational values* δ_1, δ_2: Then the sequences $\{n\delta_1 : n \in \mathbb{N}_0\}, \{n\delta_2 : n \in \mathbb{N}_0\}$ have infinitely many points in common and hence $\boldsymbol{\pi}^{(\delta_1)} = \boldsymbol{\pi}^{(\delta_2)}$ in the positive recurrent case. Hence the limit exists along all sequences $\{n\delta : n \in \mathbb{N}_0\}$ of times with rational δ. Next use the continuity of the transition semigroup to fill in the gaps!

□

End of lecture 20.

6.6 Jump chain and explosion

We have already discussed the concept of jump chains (also called embedded chains) associated with a continuous-time Markov chain on a countable state space E. We briefly review the notion of such jump chains again.

Subject to regularity conditions not stated here, we can construct the jump chain Z from a continous-time Markov chain X as follows:

- Let J_n denote the nth change in value of the chain X and set $J_0 = 0$.
- The values $Z_n = X_{J_n+}$ of X (i.e. the values right after the jump, i.e. the right limit) form a discrete-time Markov chain $Z = \{Z_n\}_{n\in\mathbb{N}_0}$.
- The transition matrix of Z is denoted by $\mathbf{P}^Z$ and satisfies
  - $p^Z_{ij} = g_{ij}/g_i$ if $g_i := -g_{ii} > 0$,
  - if $g_i = 0$, then the chain gets absorbed in state i once it gets there for the first time.
- If $Z_n = j$, then the holding time $H_{n+1} = J_{n+1} - J_n = H_{|j}$ has exponential distribution with parameter g_j.
- The chain Z is called the **jump chain of** X.

Let us look at the converse to the above statement: Suppose $Z = \{Z_n\}_{n\in\mathbb{N}_0}$ is a discrete-time Markov chain on a countable state space E. We want to find a continuous-time Markov chain, which has Z as its jump chain. Many such chains X exist!

- Let $\mathbf{P}^Z$ denote the transition matrix of the discrete-time Markov chain Z taking values in E. Assume $p^Z_{ii} = 0$ for all $i \in E$. This assumption is not very important. It only accounts for the fact that you cannot see jumps from any state i to itself in continuous time!
- For $i \in E$, let g_i denote non-negative constants. Define

$$g_{ij} = \begin{cases} g_i p^Z_{ij}, & \text{if } i \neq j, \\ -g_i & \text{if } i = j. \end{cases}$$

The construction of the continuous-time Markov chain $X = \{X_t\}_{t\geq 0}$ is done as follows:

- Set $X_0 = Z_0$.
- After a holding time $H_1 = H_{|Z_0} \sim \text{Exp}(g_{Z_0})$ the process jumps to state Z_1.
- After a holding time $H_2 = H_{|Z_1} \sim \text{Exp}(g_{Z_1})$ the process jumps to state Z_3, etc...
- More formally: Conditional on the values Z_n of the chain Z, let $H_1, H_2, \ldots$ be independent random variables with exponential distribution $H_i \sim \text{Exp}(g_{Z_{i-1}})$, $i = 1, 2, \ldots$. Set $J_n = H_1 + \cdots + H_n$.
- Then define

$$X_t = \begin{cases} Z_n, & \text{if } J_n \leq t < J_{n+1} \text{ for some } n, \\ \infty, & \text{otherwise, i.e. if } J_\infty \leq t. \end{cases}$$

- Note that the special state ∞ has been added in case the chain explodes.

Recall that $J_\infty = \lim_{n\to\infty} J_n$. J_∞ is called explosion time and we say that the chain explodes if

$$\mathrm{P}(J_\infty < \infty) > 0.$$

One can show that:

- X is a continuous-time Markov chain with state space $E \cup \{\infty\}$.
- The matrix $\mathbf{G}$ is the generator of X (up to the explosion time).
- Z is the jump chain of X.

It is possible to define the process X in different ways at time of explosion. Note that in the case of a finite state space $|E| = K < \infty$ it is not that difficult to prove the above properties. Things get much more complicated in the case of an infinite state space.

Note that the chain X constructed before is called *minimal*, since it is "active" for a minimal interval of time. Next we study conditions which ensure that the process does not explode.

Theorem 6.6.1. *The chain X constructed above does not explodes if any of the following three conditions hold.*

1. *The state space E is finite.*
2. $\sup_{i \in E} g_i < \infty$.
3. $X_0 = i$ *where i is a recurrent state for the jump chain Z.*

Proof. Clearly (1.) implies (2.). Hence we only need to check conditions (2.) and (3.).

We start with condition (2.).

- Suppose that $g_i < \gamma < \infty$ for all i.
- For the nth holding time we have $H_n \sim \text{Exp}(g_{Z_{n-1}})$.
- Clearly, if $g_{Z_{n-1}} > 0$, then $V_n = g_{Z_{n-1}} H_n \sim \text{Exp}(1)$, see Theorem 4.1.2.
- If $g_{Z_{n-1}} = 0$, then $H_n = \infty$ almost surely.
- Hence

$$\gamma J_\infty = \begin{cases} \infty, & \text{if } g_{Z_{n-1}} = 0 \text{ for some } n, \\ \sum_{n=1}^{\infty} \gamma H_n \geq \sum_{n=1}^{\infty} V_n, & \text{otherwise.} \end{cases}$$

- Similarly to the proof of Theorem 4.3.1 one can then show that the sum is almost surely infinite and hence there is no explosion.

Now assume that condition (3.) holds.

- If $g_i = 0$, then $X_t = i$ for all t, and there is nothing to prove!
- The case $g_i > 0$ is more interesting. We know that $Z_0 = i$ and i is a recurrent state for Z. Hence Z visits i infinitely many times at times $N_0 < N_1 < \cdots$ say.
- Then

$$g_i J_\infty \geq \sum_{j=0}^{\infty} g_i H_{|Z_{N_j}}, \text{ where } H_{|Z_{N_j}} \sim \text{Exp}(g_i), \quad \forall j \in \mathbb{N}_0.$$

- Again, as in the proof of Theorem 4.3.1 one can then show that the sum is almost surely infinite and hence there is no explosion.

□

6.7 Birth processes

In the previous chapter, we have looked at Poisson processes as well as various extensions. However, an important extension is the *birth process*. This is a stochastic (counting) process, which describes the arrivals of individuals in a more general way than a Poisson process.

Definition 6.7.1. *A* birth process *with intensities $\lambda_0, \lambda_1, \dots \geq 0$ is a stochastic process $\{N_t\}_{t\geq 0}$ with values in $\mathbb{N}_0$ such that*

1. *Non-decreasing process: $N_0 \geq 0$; if $s < t$, then $N_s \leq N_t$,*
2. *There is a 'single arrival', i.e. the infinitesimal transition probabilities are for $t \geq 0$, $\delta > 0$, $n, m \in \mathbb{N}_0$:*
$$\mathrm{P}(N_{t+\delta} = n + m | N_t = n) = \begin{cases} 1 - \lambda_n \delta + o(\delta) & \text{if } m = 0 \\ \lambda_n \delta + o(\delta) & \text{if } m = 1 \\ o(\delta) & \text{if } m > 1, \end{cases}$$
3. *Conditionally independent increments: Let $s < t$, then conditional on the value of N_s, the increment $N_t - N_s$ is independent of all arrivals prior to s.*

Note that by *conditionally independent increments*, we mean that for $0 \leq s < t$, conditional on the value of N_s, the increment $N_t - N_s$ is independent of all arrivals prior to s. I.e. for $k, l, x(r) \in \{0, 1, 2, \dots\}$ for $0 \leq r < s$, we have

$$\mathrm{P}(N_t - N_s = k | N_s = l, N_r = x(r) \text{ for } 0 \leq r < s) = \mathrm{P}(N_t - N_s = k | N_s = l).$$

Note that a birth process is a continuous-time Markov chain. In fact, the Markov property is an immediate consequence of the conditionally independent increments. (You can briefly check this yourself as an exercise!)

A Poisson process is a a special case of a birth process (with $\lambda_n = \lambda$ for all $n \in \mathbb{N}_0$). In the case of a (general) birth process, the birth rates depend on the current state of the process.

Example 6.7.2. *A simple birth process is a model with $\lambda_n = n\lambda$. This models the growth of a population, in which each individual may give birth to a new one with rate $\lambda > 0$; no deaths occur. Then, for $t \geq 0$, $\delta > 0$, $n, m \in \mathbb{N}_0$ (and $m \leq n$):*

$$\begin{aligned} \mathrm{P}(N_{t+\delta} = n + m | N_t = n) &= \binom{n}{m} (\lambda\delta)^m (1 - \lambda\delta)^{n-m} + o(\delta) \\ &= \begin{cases} (1-\lambda\delta)^n + o(\delta) & = \sum_{i=0}^{n} \binom{n}{i} (-\lambda\delta)^i + o(\delta) & = 1 - n\lambda\delta + o(\delta), & \text{if } m = 0 \\ n\lambda\delta(1-\lambda\delta)^{n-1} + o(\delta) & = n\lambda\delta \sum_{i=0}^{n-1} \binom{n-1}{i} (-\lambda\delta)^i & = n\lambda\delta + o(\delta), & \text{if } m = 1 \\ & & o(\delta), & \text{if } m > 1. \end{cases} \end{aligned}$$

Example 6.7.3. *A simple birth with immigration. This is a model with $\lambda_n = n\lambda + \nu$, with $\nu > 0$. Here each individual can give birth, but there is a constant rate of immigration.*

Example 6.7.4. *Let us study the generator of a birth process. For $i, j \in \mathbb{N}_0$, we have $p_{i(i+1)}(\delta) = \lambda_i \delta + o(\delta)$ and $p_{ii}(\delta) = 1 - \lambda_i \delta + o(\delta)$.*

Hence

$$g_{ii} = \lim_{\delta \downarrow 0} \frac{p_{ii}(\delta) - 1}{\delta} = -\lambda_i, \quad g_{i,i+1} = \lim_{\delta \downarrow 0} \frac{p_{i(i+1)}(\delta)}{\delta} = \lambda_i$$

and otherwise $g_{ij} = 0$, if $i > j$ or $j > i + 1$. That is

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

End of lecture 21.

6.7.1 The forward and backward equations

We will now derive the forward and backward equations for birth processes.

Let $\{N_t\}$ be a birth process with positive intensities $\lambda_0, \ldots$. Recall that the transition probabilities are defined as

$$p_{ij}(t) = \mathrm{P}(N_{t+s} = j | N_s = i) = \mathrm{P}(N_t = j | N_0 = i), \quad \text{for } i, j \in E, \tag{6.7.1}$$

(note that this is time-homogeneity).

Theorem 6.7.5. *For $i, j \in E$, $i < j$, $t \geq 0$, the forward equations of a birth process are given by*

$$\frac{dp_{ij}(t)}{dt} = -\lambda_j p_{ij}(t) + \lambda_{j-1} p_{i,j-1}(t),$$

with $\lambda_{-1} = 0$, and the backward equations are given by

$$\frac{dp_{ij}(t)}{dt} = -\lambda_i p_{ij}(t) + \lambda_i p_{i+1,j}(t).$$

In both cases, the boundary condition is given by $p_{ij}(0) = \delta_{ij}$.

Recall that δ_{ij} denotes the Kronecker delta with $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Proof. Let $i, j \in E, i < j$, and set $\lambda_{-1} = 0$. Let $t \geq 0, \delta > 0$.

We consider the **forward equations** first. Then, by the Chapman-Kolmogorov equations and the single-arrival property

$$\begin{aligned} p_{ij}(t+\delta) &= \sum_{l \in E} p_{il}(t) p_{lj}(\delta) \\ &= p_{i,j-1}(t) p_{j-1,j}(\delta) + p_{ij}(t) p_{jj}(\delta) + o(\delta) \\ &= p_{i,j-1}(t) \lambda_{j-1} \delta + p_{ij}(t)(1 - \lambda_j \delta) + o(\delta). \end{aligned}$$

Then rearranging and taking the limit $\delta \downarrow 0$ it follows

$$\frac{dp_{ij}(t)}{dt} = -\lambda_j p_{ij}(t) + \lambda_{j-1} p_{i,j-1}(t),$$

with the boundary condition $p_{ij}(0) = \delta_{ij}$.

The **backward equations** may be derived in a similar fashion. The only difference is that we apply the Chapman-Kolmogorov equations "the other way round":

$$\begin{aligned} p_{ij}(t+\delta) &= \sum_{l \in E} p_{il}(\delta) p_{lj}(t) \\ &= p_{ii}(\delta) p_{ij}(t) + p_{i,i+1}(\delta) p_{i+1,j}(t) + o(\delta) \\ &= (1 - \lambda_i \delta) p_{ij}(t) + \lambda_i \delta p_{i+1,j}(t) + o(\delta). \end{aligned}$$

Rearranging, and taking $\delta \downarrow 0$ as before we obtain

$$\frac{dp_{ij}(t)}{dt} = -\lambda_i p_{ij}(t) + \lambda_i p_{i+1,j}(t),$$

with the boundary condition $p_{ij}(0) = \delta_{ij}$.

□

As we can see, we have two ODEs which are satisfied by the transition probabilities. The solutions of these ODEs give us the transition probabilities; this helps us to answer questions about the size of the population.

We conclude the section with an important result:

Theorem 6.7.6. *Let $\{N_t\}_{t \geq 0}$ be a birth process of positive intensities $\lambda_0, \lambda_1, \ldots$. Then the forward equations have a unique solution, which satisfies the backward equations.*

Proof. See Grimmett & Stirzaker (2001*b*, p. 251). □

6.7.2 Explosion of a birth process

If the rate $\lambda_0, \lambda_1, \dots$ in the birth process increase too quickly, it may happen that infinitely many individuals are born in finite time. Such a phenomenon is called *explosion*. More formally, we have the following definition.

Definition 6.7.7. *Let $J_0, J_1, \dots$ denote the jump times of a birth process N*

$$J_0 = 0, \qquad J_{n+1} = \inf\{t \geq J_n : N_t \neq N_{J_n}\}, \quad n \in \mathbb{N}_0.$$

Further, let $H_1, H_2, \dots$ denote the corresponding holding times. As before, we write

$$J_\infty = \lim_{n\to\infty} J_n = \sum_{i=1}^{\infty} H_i.$$

Then we say that explosion of the birth process N is possible if

$$\mathrm{P}(J_\infty < \infty) > 0.$$

Theorem 6.7.8. *Let N be a birth process started from $k \in \mathbb{N}_0$, with rates $\lambda_k, \lambda_{k+1}, \dots > 0$. Then:*

1. *If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} < \infty$, then $\mathrm{P}(J_\infty < \infty) = 1$, i.e. explosion occurs with probability 1;*
2. *If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} = \infty$, then $\mathrm{P}(J_\infty = \infty) = 1$, i.e. the probability that explosion occurs is 0.*

Theorem 6.7.8 follows immediately from the arguments used in the proof of Theorem 4.3.1 since the holding times are exponentially distributed with $H_{|i} \sim \mathrm{Exp}(\lambda_i)$ for all $i \geq k$ and, conditionally on $N_0 = k$, independent.

6.8 Birth-death processes

Recall the definition of a birth process. This is a non-decreasing Markov chain for which the probability of moving from n to $n+1$ in $(t, t+\delta)$ is $\lambda_n\delta + o(\delta)$. More realistic models for population growth incorporate death also. Suppose we are given the following process $\{X_t\}_{t\geq 0}$:

1. $\{X_t\}_{t\geq 0}$ is Markov chain on $E = \mathbb{N}_0$
2. The infinitesimal transition probabilities are (for $t \geq 0$, $\delta > 0$, $n \in \mathbb{N}_0, m \in \mathbb{Z}$):

$$\mathrm{P}(X_{t+\delta} = n + m | X_t = n) = \begin{cases} 1 - (\lambda_n + \mu_n)\delta + o(\delta), & \text{if } m = 0, \\ \lambda_n\delta + o(\delta) & \text{if } m = 1 \\ \mu_n\delta + o(\delta) & \text{if } m = -1 \\ o(\delta) & \text{if } |m| > 1 \end{cases}$$

3. The birth rates $\lambda_0, \lambda_1, \dots$ and the death rates $\mu_0, \mu_1, \dots$ satisfy

$$\lambda_i \geq 0 \quad \mu_i \geq 0 \quad \mu_0 = 0.$$

Then the process is called a **birth-death process**.

The generator is given by

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

The transition probabilities can be calculated using the birth and death rates; although these can be very complicated.

However, it is often of interest (and easier) to look at the asymptotic behaviour of the process. Suppose that $\mu_i, \lambda_i > 0$ for each i where the rates make sense. Then using the claim $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$;

$$\begin{aligned} -\lambda_0\pi_0 + \mu_1\pi_1 &= 0 \\ \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} &= 0 \quad n \geq 1. \end{aligned}$$

Example 6.8.1. *Show, using induction*

$$\pi_n = \frac{\lambda_0 \times \cdots \times \lambda_{n-1}}{\mu_1 \times \cdots \times \mu_n}\pi_0$$

for any $n \in \mathbb{N}$.

Such a vector $\boldsymbol{\pi}$ is a stationary distribution if and only if $\sum_n \pi_n = 1$; that is

$$\sum_{n=0}^{\infty} \frac{\lambda_0 \times \cdots \times \lambda_{n-1}}{\mu_1 \times \cdots \times \mu_n} < +\infty \tag{6.8.1}$$

with the first term ($n = 0$) defined to be 1, i.e. $\lambda_0\lambda_{-1}/\mu_1\mu_0 := 1$. Given this condition, it follows

$$\pi_0 = \left(\sum_{n=0}^{\infty} \frac{\lambda_0 \times \cdots \times \lambda_{n-1}}{\mu_1 \times \cdots \times \mu_n}\right)^{-1}.$$

By Theorem 6.5.6 we have that the process settles into equilibrium if and only if (6.8.1) holds; i.e. that the birth rates are not too large relative to the death rates.

Example 6.8.2 (**Simple death with immigration**). *Suppose that we have a continuous-time Markov chain* $\{X_t\}$ *such that* $X_0 = I$. *In this population of individuals, there is no reproduction (i.e. birth) but new individuals migrate into the population according to a Poisson process of rate* $\lambda \in \mathbb{R}_+$. *Each individual may die in* $(t, t+\delta)$ *(where* $\delta > 0$*) with probability* $\mu\delta + o(\delta)$, $\mu > 0$. *The transition probabilities are* $(i, j \in \mathbb{N}_0)$*:*

$$p_{ij}(\delta) = \mathrm{P}(X_{t+\delta} = j | X_t = i) = \begin{cases} \mathrm{P}(j - i \text{ arrivals, no deaths}) + o(\delta) & \text{if } j \geq i \\ \mathrm{P}(i - j \text{ deaths, no arrivals}) + o(\delta) & \text{if } j < i \end{cases}$$

since the probability of two or more changes in $(t, t+\delta)$ *is* $o(\delta)$. *As a result*

$$\begin{aligned} p_{i,i+1}(\delta) &= \lambda\delta(1-\mu\delta)^i + o(\delta) = \lambda\delta + o(\delta) \\ p_{i,i-1}(\delta) &= (1-\lambda\delta)\binom{i}{1}\mu\delta(1-\mu\delta)^{i-1} + o(\delta) = i\mu\delta + o(\delta), \\ p_{ii}(\delta) &= 1 - (\lambda + i\mu)\delta + o(\delta), \\ p_{ij}(\delta) &= o(\delta) \quad \textit{if } |j-i| > 1, \end{aligned}$$

which is birth-death process with parameters $\lambda_n = \lambda$, $\mu_n = n\mu$.

We study a simple birth-death process in the following in more detail.

Example 6.8.3 (**Simple birth-death process**). *We are given a biological system, where organisms give birth and die, independently. Suppose, that in* $(t, t+\delta)$*, each individual alive gives birth with probability* $\lambda\delta + o(\delta)$ *and dies with probability* $\mu\delta + o(\delta)$*; in other words* $X_t = n$*, the number of organisms in the system, evolves by increasing by 1 with probability* $\lambda n\delta + o(\delta)$ *and decreases by 1 with probability* $\mu n\delta + o(\delta)$. *The initial population, at time 0, is size* n_0. *Suppose that* $\lambda \neq \mu$. *Using the forward equations, show that the probability generating function of* X_t *is*

$$G(s,t) = \mathrm{E}(s^{X_t}) = \left\{\frac{\mu(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}}{\lambda(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}}\right\}^{n_0}.$$

If $\lambda \neq \mu$, what is the probability that extinction has occurred at, or before time t?

Let $\mathrm{P}(X_t = n) = p_n(t)$ and use the convention that $p_{-1}(t) \equiv 0$, then the forward equations for any $n \in \mathbb{N}_0$ are

$$p_n'(t) = -n(\lambda+\mu)p_n(t) + (n+1)\mu p_{n+1}(t) + (n-1)\lambda p_{n-1}(t),$$

with boundary condition $p_{n_0}(0) = 1$. By definition

$$G(s,t) = \mathrm{E}(s^{X_t}) = \sum_{n=0}^{\infty} s^n p_n(t),$$

and note (assuming we can interchange the derivative and infinite series)

$$\frac{\partial G(s,t)}{\partial t} = \sum_{n=0}^{\infty} s^n p_n'(t),$$

and

$$\frac{\partial G(s,t)}{\partial s} = \sum_{n=0}^{\infty} n s^{n-1} p_n(t).$$

Thus multiplying the nth forward equations by s^n, on both sides, and summing over n yields

$$\frac{\partial G(s,t)}{\partial t} = -s(\lambda+\mu)\sum_{n=0}^{\infty} n s^{n-1} p_n(t) + \mu\sum_{n=0}^{\infty}(n+1)s^n p_{n+1}(t) + \lambda\sum_{n=1}^{\infty}(n-1)s^n p_{n-1}(t).$$

Clearly the first expression on the R.H.S. is

$$-(\lambda+\mu)s\frac{\partial G(s,t)}{\partial s}.$$

The second is

$$\mu\sum_{n=0}^{\infty}(n+1)s^n p_{n+1}(t) = \mu\sum_{n=1}^{\infty} n s^{n-1} p_n(t) = \mu\frac{\partial G(s,t)}{\partial s};$$

and the third is

$$\lambda s^2\frac{\partial G(s,t)}{\partial s}.$$

Thus we need to solve the PDE

$$\frac{\partial G(s,t)}{\partial t} = (\lambda s - \mu)(s-1)\frac{\partial G(s,t)}{\partial s}.$$

Either you solve this PDE using standard methods, or you show that the given solution satisfies the PDE!

Let $\rho := \lambda/\mu$ (recall $\lambda \neq \mu$). The extinction probability is given by

$$\eta(t) = \mathrm{P}(X_t = 0) = G(0,t) = \left\{\frac{\mu - \mu e^{-(\lambda-\mu)t}}{\lambda - \mu e^{-(\lambda-\mu)t}}\right\}^{n_0}.$$

Then, as $t \to \infty$,

$$\eta(t) \to \begin{cases} 1, & \text{if } \rho \leq 1, \\ \rho^{-n_0}, & \text{if } \rho > 1. \end{cases}$$

6.9 Uniform semigroups [Reading material]

Remark 6.9.1. *If you would like to learn more about the technical details we skipped in our heuristic proofs, then please read Chapter 6.10 on "Uniform semigroups" in Grimmett & Stirzaker (2001*b*)!*

Definition 6.9.2. *A semigroup* $\{\mathbf{P}_t\}$ *is called* uniform *if* $\mathbf{P}_t \to \mathbf{I}$ *uniformly as* $t \downarrow 0$*, i.e.*

$$p_{ii}(t) \to 1 \quad \text{as } t \downarrow 0, \quad \text{uniformly in } i \in E. \tag{6.9.1}$$

Since $p_{ij}(t) \leq 1 - p_{ii}(t)$, equation (6.9.1) implies that $p_{ij}(t) \to 0$ for $i \neq j$. Clearly, a uniform semigroup is standard. The converse statement does not hold in general, but is true when the state space is finite. One can show the following result.

Theorem 6.9.3. *The semigroup* $\{\mathbf{P}_t\}$ *is uniform if and only if* $\sup_i(-g_{ii}) < \infty$.

Using the stronger condition we can now formulate the precise result on the forward and backward equations:

Theorem 6.9.4. *If* $\{\mathbf{P}_t\}$ *is a uniform semigroup with generator* $\mathbf{G}$*, then it is the unique solution to both the forward equation* $\mathbf{P}_t' = \mathbf{P}_t\mathbf{G}$ *and the backward equation* $\mathbf{P}_t' = \mathbf{G}\mathbf{P}_t$*, subject to the boundary condition* $\mathbf{P}_0 = \mathbf{I}$*. Moreover,*

$$\mathbf{P}_t = e^{t\mathbf{G}} \qquad \text{and} \qquad \mathbf{G}\mathbf{1}' = \mathbf{0}'.$$

In the statement above $\mathbf{0}$ and $\mathbf{1}$ denote row vectors consisting of 0s and 1s, respectively.

End of lecture 22.

Chapter 7

Brownian motion

7.1 Introduction

Brownian motion is one of the most important building blocks of stochastic processes and is applied routinely in many areas including: statistics, finance, economics, physics, chemistry and many more. Due to the complexity of the mathematics, we will not see a rigorous construction of Brownian motion. For those students interested in learning more, the reference Karatzas & Shreve (1991) is a good place to start. Brownian motion, can be thought of as a random walk in continuous time and on continuous state-spaces. It is an example of **Markov process**, a **diffusion process** and is also a **martingale**. This chapter is quite short, and as noted above, it is not possible to fully understand Brownian motion on the basis of what we study here. However, it will give you a flavour of some of the nice properties a Brownian motion has. This chapter is based on Ross (2010, Chapter 10).

7.2 From random walk to Brownian motion

Let us study a (*discrete–time*) symmetric simple random walk. Such a process is a Markov chain (X_n) on $E = \mathbb{Z}$ with transition probabilities

$$p_{ij} = \begin{cases} 0.5, & \text{if } j = i+1 \\ 0.5, & \text{if } j = i-1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i, j \in \mathbb{Z}.$$

Then

$$X_n = \sum_{i=0}^{n} Y_i,$$

where the Y_i ($i = 1, 2, \dots$) are assumed to be independent with Y_i taking values in $\{-1, 1\}$ and Y_0 denotes the (deterministic) initial value. Here we choose $Y_0 = 0$. By definition, the increments are i.i.d.: $X_n - X_{n-1} = Y_n$, in particular, $X_n = X_{n-1} + Y_n$.

A simulated path of a symmetric simple random walk is depicted in Figure 7.1.

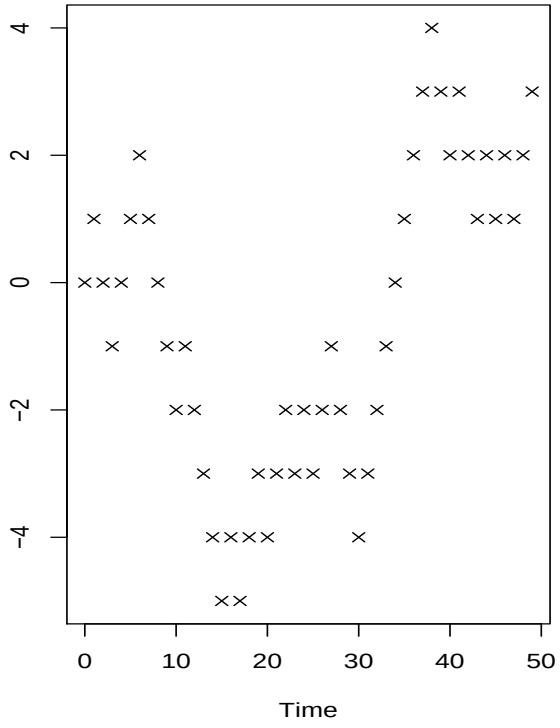


Figure 7.1: Simulated path of a symmetric simple random walk.

Exercise 7.2.1. *Compute the mean and the variance of the symmetric simple random walk* X_n*. (Recall that we chose* $Y_0 = 0$*.)*

Solution to Exercise 7.2.1 . Clearly, X_n can only take values in $\{-n, -(n-1), \ldots, (n-1), n\}$. Then, for any $i \in \mathbb{N}$,

$$\mathrm{E}(Y_i) = 0.5 \cdot 1 + 0.5 \cdot (-1) = 0,$$
$$\mathrm{Var}(Y_i) = \mathrm{E}(Y_i^2) = 0.5 \cdot 1^2 + 0.5 \cdot (-1)^2 = 1.$$

Hence $\mathrm{E}(X_n) = 0$ and $\mathrm{Var}(X_n) = n$. □

7.2.1 Modes of convergence in distribution, Slutsky's theorem and the CLT

We briefly recall the following modes of convergence, which have been studied in Y2.

Definition 7.2.2 (Convergence in probability)**.** *A sequence of random variables* $X_1, X_2, \ldots$ converges in probability *to* X*, written* $X_n \xrightarrow{\mathrm{P}} X$ *if, for each* $\epsilon > 0$

$$\lim_{n\to\infty} \mathrm{P}(\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}) = \lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \epsilon) = 0.$$

Definition 7.2.3 (Convergence in distribution)**.** *Let the cumulative distribution function of* X_n *and* X *be denoted by* F_n *and* F*, respectively. Then,* X_n converges in distribution/weakly to X*, written* $X_n \xrightarrow{d} X$ *if,*

$$\lim_{n\to\infty} F_n(x) = F(x), \qquad \textit{for every continuity point } x \textit{ of } F(x).$$

We state without proof the famous Slutsky's theorem.

Theorem 7.2.4 (Slutsky's theorem)**.** *Suppose that* $X_n \xrightarrow{d} X, A_n \xrightarrow{\mathrm{P}} a$*, and* $B_n \xrightarrow{\mathrm{P}} b$*, where* a *and* b *are (deterministic) constants. Then*

$$A_n X_n + B_n \xrightarrow{d} aX + b.$$

Recall the central limit theorem:

Theorem 7.2.5. *Let* $Z_1, Z_2, \ldots$ *be a sequence of independent, identically distributed random variables, each with finite mean* μ *and finite variance* σ^2*. Then the distribution of*

$$\frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} Z_i - n\mu\right)$$

tends to the standard normal distribution as $n \to \infty$*.*

From the CLT, we immediately get that

$$\frac{X_n}{\sqrt{n}} \xrightarrow{d} N(0,1).$$

Now define

$$B_t^{(n)} = \frac{X_{\lfloor nt \rfloor}}{\sqrt{n}} = \frac{X_k}{\sqrt{n}} = \frac{\sqrt{\lfloor nt \rfloor}}{\sqrt{n}} \frac{X_k}{\sqrt{k}}, \quad \text{for } k = \lfloor nt \rfloor, k \leq nt < k+1.$$

Note that $\frac{\sqrt{\lfloor nt \rfloor}}{\sqrt{n}} \to \sqrt{t}$ as $n \to \infty$. Hence, by Slutsky's Theorem, we have

$$\frac{X_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} N(0,t), \text{ as } n \to \infty.$$

Note that $\xrightarrow{d}$ denotes convergence in law/in distribution/weak convergence. Also, $\lfloor nt \rfloor$ denotes the largest integer less than or equal to nt.

In fact, we get an even stronger result: The rescaled symmetric simple random walk converges in distribution to the Brownian motion, i.e. for $t \geq 0$

$$\frac{X_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} B_t, \qquad \text{and more generally} \qquad \frac{X_{\lfloor n\cdot \rfloor}}{\sqrt{n}} \xrightarrow{d} B_\cdot,$$

as $n \to \infty$.

The latter result is due to Donsker's Theorem.

We depict the paths of a rescaled random walk for $n \in \{10, 100, 1000\}$ in Figure 7.2.

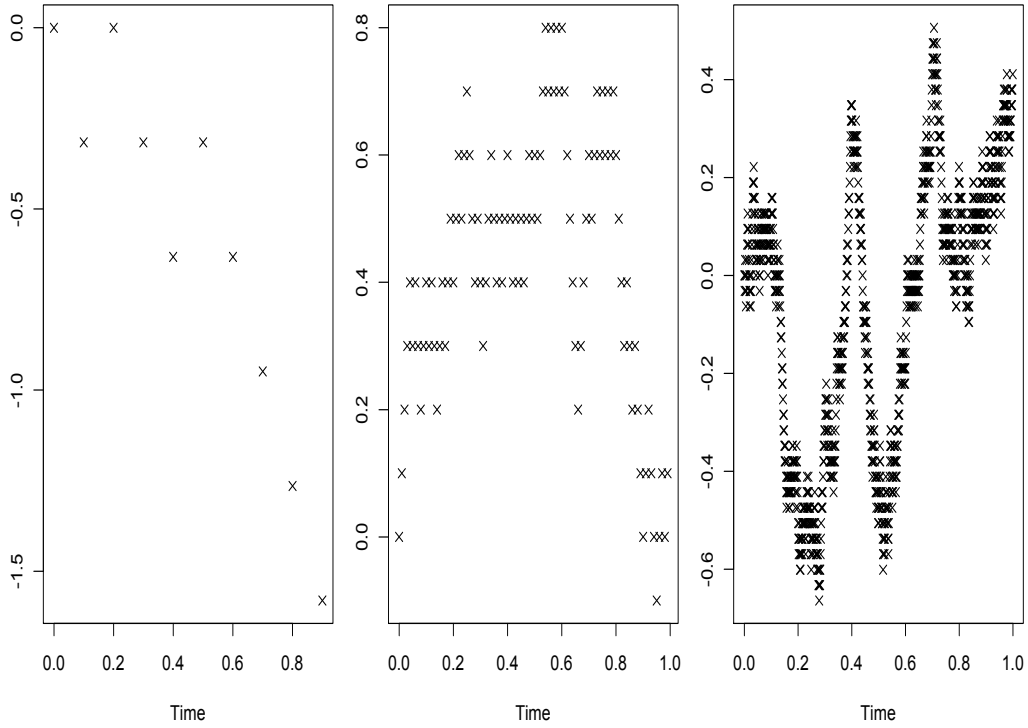


Figure 7.2: Simulation of a rescaled random walk for $n \in \{10, 100, 1000\}$.

7.3 Brownian motion

Definition 7.3.1. *A real-valued stochastic process $B = \{B_t\}_{t\geq 0}$ is a **standard Brownian motion** if*

1. *$B_0 = 0$ almost surely;*
2. *B has independent increments;*
3. *B has stationary increments;*
4. *The increments are Gaussian, for $0 \leq s < t$*

$$B_t - B_s \sim \mathrm{N}(0, (t-s));$$

5. *The sample paths are almost surely continuous, i.e. the function $t \mapsto B_t$ is almost surely continuous in t.*

Definition 7.3.2. *Let $B = \{B_t\}_{t\geq 0}$ denote a standard Brownian motion. The stochastic process $Y = \{Y_t\}_{t\geq 0}$ defined by*

$$Y_t = \sigma B_t + \mu t, \qquad \text{for all } t \geq 0,$$

*is called a **Brownian motion with drift parameter $\mu \in \mathbb{R}$ and variance parameter σ^2**, where $\sigma > 0$.*

Note that for $0 \leq s < t$, $Y_t - Y_s \sim \mathrm{N}(\mu(t-s), \sigma^2(t-s))$.

7.4 Historical background

- The Scottish botanist Robert **Brown** (1773 –1858) observed jittery motion of pollens in water (1828)
- The Danish astronomer, actuary and mathematician Thorvald N. **Thiele** (1838 –1910) described the mathematics behind Brownian motion (1880)
- The French mathematician Louis J. A. **Bachelier** (1870 –1946) introduced the Brownian motion in Finance (1900)
- The German (later Swiss and American) physicist Albert **Einstein** (1879 – 1955) studied the Brownian motion in physics (1905)
- The American mathematician Norbert **Wiener** (1894 –1964) was the first one to prove the existence of the Brownian motion (1923) (which is a deep mathematical result!)

7.5 Finite dimensional distributions and transition densities

We remark on some ideas of stochastic processes. In general, and in continuous time, we have an underlying probability space $(\Omega, \mathcal{F}, \mathrm{P})$, and for $\omega \in \Omega$, $X_t(\omega)$ is a map $X : \mathcal{T} \times \Omega \to E$. (In this chapter we choose $\mathcal{T} = [0, \infty)$.) Intrinsically, this means that, when constructing the underlying probability space, we need to think about probability measures on paths, that is on an infinite number of random variables. It is simpler (and technically possible due what is known as the Kolmogorov existence theorem) to consider the *finite dimensional distributions* (FDDs):

$$\mathrm{P}(X_{t_1} \leq x_1, \cdots, X_{t_n} \leq x_n)$$

with $0 \leq t_1 < \cdots < t_n$. This is the joint distribution of a finite collection of random variables from the process. As stated above, it is possible (under the Kolmogorov consistency conditions) to construct a (unique) probability space defining only the FDDs. Throughout, we mention only the FDDs, but it should be noted that there is a much deeper theory than what is presented here.

Let $B = \{B_t\}_{t \geq 0}$ denote a standard Brownian motion. Let $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \cdots < t_n$. We want to find the joint density function of

$$B_{t_1}, B_{t_2}, \ldots, B_{t_n}.$$

Note that the set of equalities

$$B_{t_1} = x_1, B_{t_2} = x_2, \ldots, B_{t_n} = x_n,$$

for $x_1, \ldots, x_n \in \mathbb{R}$ is equivalent to

$$B_{t_1} = x_1, B_{t_2} - B_{t_1} = x_2 - x_1, \ldots, B_{t_n} - B_{t_{n-1}} = x_n - x_{n-1}.$$

We can us the transformation formula for multivariate densities (see the Y1 lecture notes for the bivariate case) and the fact that B has independent, stationary, Gaussian increments, to deduce that

$$f_{(B_{t_1}, \ldots, B_{t_n})}(x_1, \ldots, x_n) = f_{B_{t_1}}(x_1) f_{B_{t_2} - B_{t_1}}(x_2 - x_1) \cdots f_{B_{t_n} - B_{t_{n-1}}}(x_n - x_{n-1}).$$

$$f_{(B_{t_1}, \ldots, B_{t_n})}(x_1, \ldots, x_n) = \frac{\exp\left(-\frac{1}{2}\left\{\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1} + \cdots + \frac{(x_n - x_{n-1})^2}{t_n - t_{n-1}}\right\}\right)}{\sqrt{(2\pi)^n (t_1 (t_2 - t_1) \cdots (t_n - t_{n-1})}}$$

From this equation, we can compute any desired probabilities. We spend some time discussing the transition densities of Brownian motion.

E.g. suppose we want to compute the conditional density of B_{t+s} given that $B_s = x$ for $0 \leq s, t$. Then

$$f_{B_{t+s}|B_s}(y|x) = \frac{f_{B_s,B_{t+s}}(x,y)}{f_{B_s}(x)} = \frac{f_{B_s}(x) f_{B_{t+s}-B_s}(y-x)}{f_{B_s}(x)}$$
$$= \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}(y-x)^2\right).$$

In order to shorten the notation, we write

$$p_t(y|x) := p_t(x,y) := \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}(y-x)^2\right),$$

which is called the *Gauss kernel*, or sometimes the *heat kernel*.

A very interesting property of the density (called the Gauss/heat kernel) is that it provides a probabilistic interpretation to a solution of a partial differential equation:

Theorem 7.5.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying some additional regularity conditions. Then the unique (continuous) solution $u_t(x)$ to the initial value problem*

$$\frac{\partial}{\partial t} u_t(x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} u_t(x)$$
$$u_0(x) = f(x)$$

is given by

$$u_t(x) = \mathrm{E}[f(W_t^x)] = \int_{-\infty}^{\infty} p_t(x,y) f(y) dy$$

where $\{W_t^x\}$ is a Brownian motion started at x.

The PDE is called the heat equation. The interesting point is that the solution of the PDE is now written as an expectation w.r.t. a Brownian motion, which gives a numerical scheme to approximate its solution (very important in high dimensions, where finite difference methods do not work well).

A verification of the result can be yielded by taking partial derivatives of the expectation w.r.t t and then showing that

$$\frac{\partial p_t(x,y)}{\partial t} = \frac{1}{2} \frac{\partial^2 p_t(x,y)}{\partial x^2}.$$

Exercise 7.5.2. *Verify the result in Theorem 7.5.1.*

End of lecture 23. .....

7.6 Symmetries and scaling laws

We look at the following result:

Proposition 7.6.1. *Let $\{B_t\}_{t\geq 0}$ be a standard Brownian motion. Then each of the following processes is also a standard Brownian motion:*

$$\{-B_t\}_{t\geq 0} \qquad \text{[Reflection]}$$
$$\{B_{t+s} - B_s\}_{t\geq 0} \qquad \text{for } s \geq 0, \quad \text{[Translation]}$$
$$\{aB_{t/a^2}\}_{t\geq 0} \qquad \text{for } a > 0, \quad \text{[Rescaling]}$$
$$\{tB_{1/t}\}_{t\geq 0} \qquad \text{[Inversion]}.$$

7.6.1 Some remarks

Now these results have some important implications. First we look at the maximum and minimum processes:

$$M_t^+ := \max\{B_s : 0 \leq s \leq t\}$$
$$M_t^- := \min\{B_s : 0 \leq s \leq t\}$$

These are well-defined, because the Brownian motion has **continuous** paths, and continuous functions always attain their maximal and minimal values on **compact intervals**. Now observe that if the path B_t is replaced by its reflection $-B_t$ then the maximum and the minimum are interchanged and negated. But since $-B_t$ is again a Brownian motion, it follows that M_t^+ and $-M_t^-$ have the same distribution:

$$M_t^+ \stackrel{d}{=} -M_t^-,$$

where our notation means 'equal in distribution'. The property that $\{aB_{t/a^2}\}$ is again a Brownian motion is called the **Brownian scaling property** (Exercise: Prove it!). It is perhaps the most useful elementary tool in the study of BM. As a first example, consider its implications for the distributions of the maximum random variables M_t^+. Fix $a > 0$, and define

$$B_t^* := aB_{t/a^2}$$
$$M_t^{+,*} := \max_{0\leq s\leq t} B_s^*$$
$$= aM_{t/a^2}^+$$

By the Brownian scaling property, B_t^* is a standard Brownian motion, and so the random variable $M_t^{+,*}$ has the same distribution as M_t^+. Therefore,

$$M_t^+ \stackrel{d}{=} aM_{t/a^2}^+.$$

On first sight, this relation appears rather harmless. However, it can be shown, it implies that the **sample paths** of a Brownian motion are with probability one, **nowhere differentiable**.

7.7 The reflection property and first-passage times

We now look at some well-known properties of Brownian motion.

Proposition 7.7.1. *Let $x > 0$ then*

$$\mathrm{P}(M_t^+ \geq x) = 2\mathrm{P}(B_t > x) = 2 - 2\Phi(x/\sqrt{t}),$$

where Φ is the normal c.d.f.

Before we prove this result we give the following fact, which is not proved (due to the complexity of the proof, interested students can check Billingsley (2012).)

Let $x > 0$ and

$$\tau := \min\{s : B_s \geq x\}.$$

Then

$$B_t'' = \begin{cases} B_t & \text{if} \quad t \leq \tau \\ x - (B_t - x) & \text{if} \quad t > \tau \end{cases}$$

is a Brownian motion. This is termed the **reflection principle**. Formally τ is a *stopping-time*.

Note that the new path B_t'' is obtained from B_t by reflection: For $t > \tau$ we reflect B_t about the horizontal line at height $x > 0$.

Proof. Let $\tau'' = \min\{t : B''_t \geq x\}$, which is the same as τ (due to the definition of B''_t). Now we have that

$$\mathrm{P}(M_t^+ \geq x) = \mathrm{P}(\tau \leq t)$$

which simply means that the probability that the maximum of the Brownian motion, on $[0, t]$, exceeding x, is the same as the probability that the first time the Brownian motion is greater or equal to x is less than t. Then using the law of total probability, we have

$$\begin{aligned}\mathrm{P}(\tau \leq t) &= \mathrm{P}(\tau \leq t, B_t \leq x) + \mathrm{P}(\tau \leq t, B_t \geq x) \\ &= \mathrm{P}(\tau'' \leq t, B''_t \leq x) + \mathrm{P}(\tau \leq t, B_t \geq x),\end{aligned}$$

where we have used the equivalence of τ'' and τ and the fact that B''_t is a Brownian motion. Now conditional on $\tau'' \leq t$ (i.e. $\tau \leq t$) we have that

$$B''_t \leq x \iff 2x - B_t \leq x \iff B_t \geq x.$$

Hence

$$\begin{aligned}\mathrm{P}(\tau \leq t) &= \mathrm{P}(\tau'' \leq t, B_t \geq x) + \mathrm{P}(\tau \leq t, B_t \geq x) \\ &= \mathrm{P}(\tau \leq t, B_t \geq x) + \mathrm{P}(\tau \leq t, B_t \geq x).\end{aligned}$$

Then

$$\mathrm{P}(\tau \leq t, B_t \geq x) = \mathrm{P}(B_t \geq x)$$

due to the intermediate value theorem (recall B_t is continuous, $B_0 = 0$, and $x > 0$). Hence

$$\mathrm{P}(M_t^+ \geq x) = 2\mathrm{P}(B_t \geq x) = 2\left[1 - \Phi\left(\frac{x}{\sqrt{t}}\right)\right].$$

□

During the course of this proof we have seen that:

$$\mathrm{P}(\tau \leq t, B_t \geq x) = \mathrm{P}(\tau \leq t, B_t \leq x)$$

which can be interpreted as, if we hit the level x, the probability that the Brownian motion is above the level, is exactly the same as being below the level. This is, essentially the argument behind the proof of the reflection principle, which we have used here.

An obvious corollary of this result is that

$$p_\tau(t) = \frac{x}{\sqrt{2\pi t^3}} \exp\left\{-\frac{x^2}{2t}\right\}$$

which is the density function of the **first hitting time** of a Brownian motion.

7.8 A model for asset prices

A rather well-known model for describing the movement of an asset price $\{S_t\}_{0 \leq t \leq T}$, $S_t \in \mathbb{R}^+$ is as follows

$$S_t = S_0 \exp\{(\mu - \sigma^2/2)t + \sigma B_t\}$$

where S_0 is the initial value of the underlying (e.g. the value of the stock at the start of trading), $\mu \in \mathbb{R}$ is the risk-free interest rate and σ is the volatility (i.e. the instantaneous standard deviation of the stock). This

process is known as **geometric Brownian motion**, and is the solution of what is known as a stochastic differential equation.

In practice, this model is routinely used in most, if not all, major investment banks and hedge funds, to calculate what is known as *option prices* and hedge financial derivatives (essentially, for firms to protect themselves from losing money in financial contracts).

From the point of view of statistics and econometrics, it is well-known that **this model does not fit the stylized features of financial returns data**. That is, real financial data does not follow the dynamic above; this is because in practice the **volatility of asset prices is typically not constant**, and often responds to a variety of market conditions (for example, September 2008, the volatility of many banking stocks was very high, due to the collapse of the investment bank Lehman Brothers).

We typically observe **time-varying volatility clusters**. We demonstrate this by looking at daily data from the S&P500 before, during and after the financial crisis.

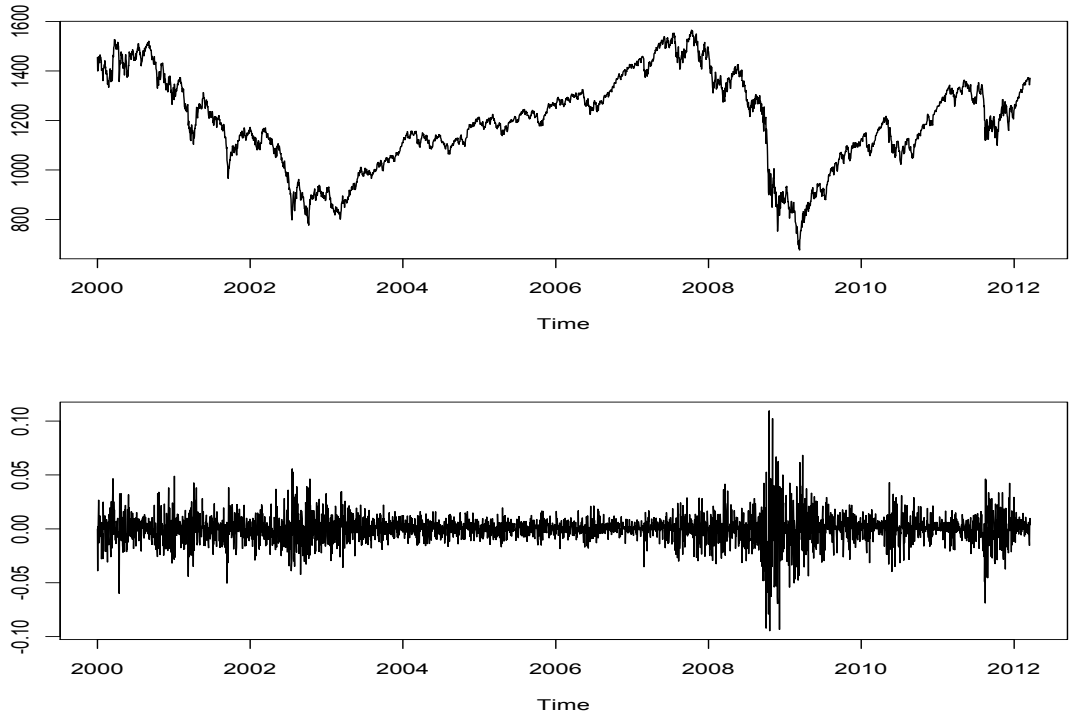


Figure 7.3: The daily index and log returns of the S&P 500 index between 3rd January 2000 to 9th March 2012. The data are based upon the adjusted closing prices.

This has yielded much academic and industrial research into cases (which goes back to at least the late 1970s) where σ is a stochastic process, e.g:

$$S_t = S_0 \exp\left\{\left(\mu t - \frac{1}{2}\int_0^t \sigma_s^2 ds\right) + \int_0^t \sigma_s dB_s\right\}$$
$$\sigma_t = \sigma_0 \exp\{\gamma t + \eta W_t\}$$

where W_t is an independent Brownian motion; such a model is termed a stochastic volatility model and there are a great variety of these. (The model for σ given above is in fact not a good one since we would typically require the volatility to be a mean-reverting process.)

- **Stochastic volatility** is a key concept in financial (and in many other!) applications.
- It poses many probabilistic as well as statistical challenges, e.g.:

  - How can stochastic integrals with respect to Brownian motion be defined?
  - How can we estimate stochastic volatility?

- Stochastic volatility can be studied both in continuous and in discrete time.
- Those of you who have taken the time series course, might have heard of the famous autoregressive conditional heteroskedastic **ARCH** model, which was developed by Robert Engel (1982), who won the Nobel Prize in Economics in 2003 "for methods of analyzing economic time series with time-varying volatility (ARCH)". This class of models has later been generalised to **GARCH** models by Tim Bollerslev (1986).

End of lecture 24.

Bibliography

Billingsley, P. (2012), *Probability and measure*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

Dobrow, R. P. (2016), *Introduction to stochastic processes with* R, John Wiley & Sons, Inc., Hoboken, NJ.

Durrett, R. (2016), *Essentials of stochastic processes*, Springer Texts in Statistics, Springer, Cham. Third edition.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling extremal events (for insurance and finance)*, Vol. 33 of *Applications of Mathematics (New York)*, Springer-Verlag, Berlin. Corrected Fourth Printing 2003.

Grimmett, G. R. & Stirzaker, D. R. (2001*a*), *One Thousand Exercises in Probability*, Oxford University Press, New York.

Grimmett, G. R. & Stirzaker, D. R. (2001*b*), *Probability and random processes*, third edn, Oxford University Press, New York.

Kallenberg, O. (2002), *Foundations of modern probability*, Probability and its Applications (New York), second edn, Springer-Verlag, New York.

Karatzas, I. & Shreve, S. E. (1991), *Brownian motion and stochastic calculus*, Vol. 113 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York.

Kirkwood, J. R. (2015), *Markov processes*, Advances in Applied Mathematics, CRC Press, Boca Raton, FL.

Mikosch, T. (2009), *Non-Life Insurance Mathematics: An Introduction with the Poisson Process*, 2 edn, Springer-Verlag, Berlin, Heidelberg, Germany.

Nelsen, R. B. (1987), 'Consequences of the memoryless property for random variables', *The American Mathematical Monthly* **94**(10), 981–984.

Nordborg, M. (2000), Coalescent theory. Preprint.

Nordborg, M. (2004), *Coalescent Theory*, American Cancer Society, chapter 20.

Norris, J. R. (1998), *Markov chains*, Vol. 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge. Reprint of 1997 original.

Pinsky, M. A. & Karlin, S. (2011), *An introduction to stochastic modeling*, fourth edn, Elsevier/Academic Press, Amsterdam.

Ross, S. M. (2010), *Introduction to probability models*, tenth edn, Harcourt/Academic Press, Burlington, MA.

Shiryaev, A. N. (1996), *Probability*, Vol. 95 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York. Translated from the first (1980) Russian edition by R. P. Boas.

Williams, D. (1991), *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge.