MATH96052/MATH97083 Applied Probability

Autumn 2019

Prof. Almut E. D. Veraart
Room 551 Huxley Building
Imperial College London
E-Mail: a.veraart@imperial.ac.uk

Department of Mathematics
Imperial College London
180 Queens Gate, London, SW7 2AZ

Last Update: Friday 27th September, 2019

Abstract

This course aims to give students an understanding of the basics of stochastic processes. The theory of different kinds of processes will be described, and will be illustrated by applications in several areas. The groundwork will be laid for further deep work, especially in such areas as genetics, finance, industrial applications, and medicine. Revision of basic ideas of probability.

Important discrete and continuous probability distributions. Random processes: Poisson processes and their properties; Superposition, thinning of Poisson processes; Non-homogeneous, compound, and doubly stochastic Poisson processes. Autocorrelation functions. Probability 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. Random walks. Gamblers ruin. Branching processes and their properties. 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.

Contents

Chapter 0

Course information

Course content and references

Lecture 1

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 continuously or discretely. In particular, we will focus upon three broad areas:

- Markov chains in discrete time
- Markov chains in continuous time (incl. Poisson, Birth-death processes)

- This course is based on material by Dr. Ajay Jasra, who taught the course in 2010 and 2011 and has been further revised subsequently.
- I set the following summer exams: 2012, 2013, 2014, 2015, 2017, 2018 (but not the ones in 2016 and 2019)
- The notes are designed so that it is possible to understand applied probability, however, at some points, there are missing definitions (which are filled in during lectures). At numerous points there are exercises in the lecture notes, and it is quite important to attempt them. (The model solutions are provided in the lecture notes, but you are not supposed to read them before you have tried to solve the problem yourself!)
- The course material will be available on-line through *Blackboard Learn*, where you will need to sign up for this course: `https://bb.imperial.ac.uk/`
- The lectures and problem classes will be recorded via Panopto, but you are strongly encouraged to attend all sessions in person and to actively participate in the lectures by asking or answering questions.
- **All parts of the lecture notes and of the coursework are potentially examinable.**

References: The recommended reference for this course is:

Grimmett, G. & Stirzaker, D. *Probability and Random Processes* (2001). Third edition, Oxford University Press: Oxford. [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.
Solutions to the exercises in the book can be found in:

Grimmett, G. & Stirzaker, D. *One Thousand Exercises in Probability* (2001). Oxford University Press: Oxford. [Grimmett & Stirzaker (2001*a*)]

Other very good textbooks include:
Norris, J. R. *Markov chains* (1998). Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press: Cambridge. [Norris (1998)]

Ross, S. *Introduction to Probability Models* (2010). Academic Press: London. [Ross (2010)]

Pinsky, M. and Karlin, S. *An Introduction to Stochastic Modeling* (2011). Fourth Edition. Academic Press. [Pinsky & Karlin (2011)]

For very motivated students, the references:
Billingsley, P. *Probability and Measure* (1995). Wiley: New York. [Billingsley (2012)]
Kallenberg, O. *Foundations of modern probability* (2002). Second edition. Probability and its Applications. Springer-Verlag, New York. [Kallenberg (2002)]
Shiryaev, A. *Probability* (1996). Springer: New York. [Shiryaev (1996)]
Williams, D. *Probability with Martingales* (1991). Cambridge University Press: Cambridge.[Williams (1991)]
might be of interest, but note that they are *very* advanced.

Assessment and feedback

Progress tests: There are **two** in-class progress tests. Each test accounts for 5% of the course mark.

- First progress test: Tuesday 29 October 2019.
- Second progress test: Thursday 28 November 2019. (Dates to be confirmed on Blackboard.)

Exam (worth 90%): 2 hour May/June exam for 3rd year students comprised of 4 questions. 2.5 hour May/June exam for 4th year students comprised of 5 questions. There will be **no** formula sheet in the exam, and it is expected that you will know standard properties of densities (and their functional form).

Office Hour: Please see Blackboard. Room 551, Huxley Building.

If you have any questions regarding the course material, please come to my office hour or ask me before/after the lectures. I cannot answer longer questions by email.

Problem sheets and classes: There will be non-assessed problem sheets, to be handed out during the course.

- There will be two **non-assessed feedback problem sheets** which you can submit individually or in groups. The submission deadlines are specified in the lecture plan on Blackboard. They will be marked for feedback only, they do not count towards your final course mark.
- In addition, there will be six regular problem sheets which will be discussed in six dedicated problem classes. It is recommended that these problems are attempted beforehand.

You will obtain feedback through:

- Marked progress tests,
- Marked (but non-assessed) feedback problem sheets,
- Non-assessed coursework,
- One-to-one meeting with me in my office hour.

You can give me feedback about the course by:

- sending me an email or talking to me before/after the lectures or in my office hour,
- contacting one of our course representatives.

Chapter 1

Preliminaries

1.1 The probability space

Before we begin the course, there are number of concepts that need to be known. You should be familiar with the material in this section from your first and second year probability and statistics courses. 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!

Recall that probability theory starts with a *probability space* which consists of a triple $(\Omega, \mathcal{F}, \mathbb{P})$, 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).
- $\mathbb{P}$ is a probability measure which 'measures' the probability of each set $A \in \mathcal{F}$.

Let us revise the underlying concepts briefly.

Let Ω be a set.

Definition 1.1.1. *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 1.1.2. *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 1.1.3. *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 1.1.4. *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 1.1.5. *The smallest $\sigma-$algebra associated with Ω is the collection $\mathcal{F} = \sigma(\Omega) = \{\emptyset, \Omega\}$.*

Example 1.1.6. *Let A be any subset of Ω. Then $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$.*

Example 1.1.7. *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 1.1.8. *A* probability measure $\mathbb{P}$ *on the measurable space $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ satisfying the following three conditions:*

1. $\mathbb{P}(\emptyset) = 0$;
2. $\mathbb{P}(\Omega) = 1$;
3. *For any collection $\{A_i : i = 1, 2, \dots\}$ of disjoint (i.e. $A_i \cap A_j = \emptyset$ for all pairs i, j satisfying $i \neq j$) and $\mathcal{F}$-measurable subsets of Ω, we have*

$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i).$$

Example 1.1.9. *Suppose that $\Omega = \{0, 1\}$, then it can be established that $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$. If*

$$\mathbb{P}(\{0\}) = \mathbb{P}(\{1\}) = 0.5$$

then we have defined a probability space for the uniform *distribution on $\{0, 1\}$.*

Summarising, we re-iterate that throughout the course, we always assume that we are on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, 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;
- $\mathbb{P}$ is a probability measure on the measurable space $(\Omega, \mathcal{F})$.

1.2 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 $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

Addition law $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

Conditional probability If $\mathbb{P}(B) \neq 0$, we define $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

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!

Definition 1. A finite or countably infinite family of events $\{A_i : i = 1, 2, \dots\}$ is called a *partition* of the set Ω if

$$A_i \cap A_j = \emptyset \quad \text{when } i \neq j, \quad \text{and} \quad \bigcup_i A_i = \Omega.$$

Recall the **law of total probability**, which we will be using many, many times throughout this course.

Theorem 2 (Law of total probability)**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space. Let $\{A_i : i = 1, 2, \dots\}$ be a partition of Ω with $A_i \in \mathcal{F}$ and $\mathbb{P}(A_i) > 0$ for all i. Then we have*

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B \cap A_i) = \sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

Example 1.2.1. *Given two events $A, B \in \mathcal{F}$ such that $0 < \mathbb{P}(B) < 1$, we have*

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

Definition 1.2.2. *Events A and B are* independent *if*

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

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

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

for all finite subsets J of I.

1.3 Random variables and stochastic processes

Let us recall further important definitions:

Definition 1.3.1. *A* random variable *is a map $X : \Omega \to \mathbb{R}$, such that for any $A \in \mathcal{B}(\mathbb{R})$, $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}$. Such a function is said to be $\mathcal{F}$-measurable.*

Here $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.
You can picture a $\mathcal{F}$-measurable function X as follows:

$$\Omega \xrightarrow{X} \mathbb{R}$$
$$\mathcal{F} \xleftarrow{X^{-1}} \mathcal{B}(\mathbb{R}).$$

Note that a random variable X, say, defines a σ–algebra

$$\sigma(X) = \{X^{-1}(A), A \in \mathcal{B}(\mathbb{R})\} \subset \mathcal{F},$$

which is called the *σ–algebra generated by X*.

Definition 1.3.2. *The random variable is called* discrete *if it takes values in some countable subset $\{x_1, x_2, \dots, \}$, only, of $\mathbb{R}$. The discrete random variable has* (probability) mass function $f : \mathbb{R} \mapsto [0, 1]$ *given by $f(x) = \mathbb{P}(X = x)$.*

Definition 1.3.3. *The random variable X is called* continuous *if its distribution function can be expressed as*

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)du, \qquad x \in \mathbb{R},$$

for some integrable function $f : \mathbb{R} \mapsto [0, \infty)$ called the (probability) density function *of X.*

Recall that in Definition 1.2.2 we defined independent events. Next, we want to define the concept of independence of random variables.

Definition 1.3.4. *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: A family $\{X_i : i \in I\}$ of discrete random variables is independent *if and only if*

$$\mathbb{P}(X_i = x_i \textit{ for all } i \in J) = \prod_{i \in J} \mathbb{P}(X_i = x_i),$$

for all sets $\{x_i : i \in I\}$ and for all finite subsets J of 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.).

Definition 1.3.5. *Random variables X and Y are called* independent *if*

$$\{X \leq x\} \textit{ and } \{Y \leq y\} \textit{ are independent events for all } x, y \in \mathbb{R}.$$

More generally, we have

Definition 1.3.6 (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:*

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

Definition 1.3.7. *A stochastic process $(X_t)_{t\in\mathcal{T}}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is a collection of $\mathbb{R}$–valued random variables. I.e. $X : \Omega \times \mathcal{T} \to \mathbb{R}$, where $\mathcal{T}$ is some time domain (e.g. $\mathcal{T} = [0, T]$ or $\mathcal{T} = [0, \infty)$).*

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

Chapter 2

Discrete–time Markov chains

Lecture 2

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, 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.

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.

2.1 The Markov property

We begin by giving the basic set-up. Let $X_0, X_1, \ldots$ be a sequence of random variables, each variable taking some value in a state-space $E \subseteq \mathbb{Z}$ (indeed E can be any countably infinite set). We write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Here $\{X_n\}_{n\in\mathbb{N}_0}$ is a discrete time stochastic process (note the time parameter is countably infinite).

Definition 2.1.1. *A discrete-time stochastic process $\{X_n\}_{n\in\mathbb{N}_0}$ on E is a **Markov chain** if it satisfies the **Markov condition***

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1})$$

for all $n \in \mathbb{N}$ and for all $x_0, \ldots, x_{n-1}, x_n \in E$.

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

An important point to note is that X_n is *not* independent of (say) X_{n-2}. That is, for $n \geq 2$

$$\begin{aligned}
& \mathbb{P}(X_n = x_n, X_{n-2} = x_{n-2}) \\
= & \sum_{x_{n-1}\in E} \mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}) \\
= & \sum_{x_{n-1}\in E} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}) \mathbb{P}(X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}) \\
= & \sum_{x_{n-1}\in E} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \mathbb{P}(X_{n-1} = x_{n-1} | X_{n-2} = x_{n-2}) \mathbb{P}(X_{n-2} = x_{n-2}),
\end{aligned}$$

where we used the law of total probability, the definition of the conditional probability and the Markov property. Note that the expression above does not factorise into the two marginal probabilities of X_{n-2} and X_n.

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

Definition 2.1.2. 1. *The Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$ is* time-homogeneous *if*

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

for every $n \geq 0$, $i, j \in E$.

2. *In addition, let $K = \text{Card}(E) = |E|$. The* transition matrix $\mathbf{P} = (p_{ij})$ *is the $K \times K$ matrix of transition probabilities*

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

We finish this section with the following theorem:

Theorem 2.1.3. *The transition matrix $\mathbf{P}$ is a* stochastic matrix, *i.e.:*

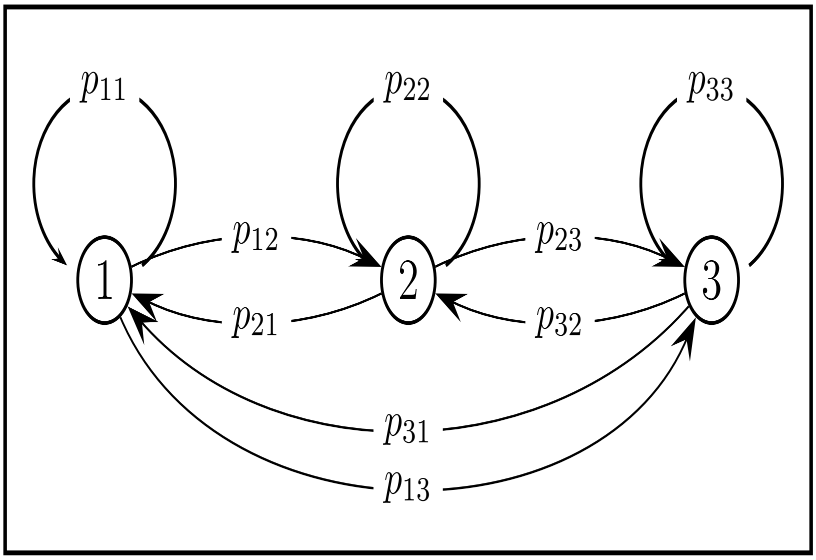
1. $\mathbf{P}$ *has non-negative entries, $p_{ij} \geq 0$ for all $i, j \in E$*
2. $\mathbf{P}$ *has row sums equal to 1; $\sum_{j\in E} p_{ij} = 1$ for all $i \in E$.*

Proof. Exercise! □

Remark 2.1.4. • *Note that throughout this section, we always assume that the state space E either **finite** (i.e. $K < \infty$) or **countably infinite** (with $K = \infty$).*

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 2.1.5. *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*



2.2 The 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 2.2.1. *Let* $n \geq 1$. *The* $n-$*step transition matrix* $\mathbf{P}_n = (p_{ij}(n))$ *is the matrix of* $n-$*step transition probabilities*

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

with $m \geq 0$.

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.

Exercise 2.2.2. *Before we continue, consider three random variables* X_1, X_2, X_3. *Show that*

$$\mathbb{P}(X_1 = x_1, X_2 = x_2 | X_3 = x_3) = \mathbb{P}(X_1 = x_1 | X_2 = x_2, X_3 = x_3)\mathbb{P}(X_2 = x_2 | X_3 = x_3).$$

Remark 2.2.3. *For a discrete Markov chain $(X_n)_{n\geq 0}$ on the state space E, we have*

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

for $m \geq 1$ and for all $x_{n+m}, x_n, \ldots, x_0 \in E$ (see Problem sheet 1).

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

Theorem 2.2.4. *Let $m \geq 0$, $n \geq 1$. 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 2.2.5. *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 2.2.4.

For any $i, j \in E$ and integers $m \geq 0, n \geq 1$ we have

$$\begin{aligned}
p_{ij}(m+n) &= \mathbb{P}(X_{m+n} = j | X_0 = i) = \frac{\mathbb{P}(X_{m+n} = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \quad \text{[Apply the law of total probability]} \\
&= \sum_{l \in E} \frac{\mathbb{P}(X_{m+n} = j, X_m = l, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{l \in E} \frac{\mathbb{P}(X_{m+n} = j, X_m = l, X_0 = i)}{\mathbb{P}(X_m = l, X_0 = i)} \frac{\mathbb{P}(X_m = l, X_0 = i)}{\mathbb{P}(X_0 = i)} \quad \text{[Use the def. of cond. prob.]} \\
&= \sum_{l \in E} \mathbb{P}(X_{m+n} = j | X_m = l, X_0 = i) \mathbb{P}(X_m = l | X_0 = i) \quad \text{[Apply the Markov property]} \\
&= \sum_{l \in E} \mathbb{P}(X_{m+n} = j | X_m = l) \mathbb{P}(X_m = l | X_0 = i) \\
&= \sum_{l \in E} p_{lj}(n) p_{il}(m) = \sum_{l \in E} p_{il}(m) p_{lj}(n).
\end{aligned}$$

□

The proof that $\mathbf{P}_n = \mathbf{P}^n$ is left as an exercise.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 3

Example 2.2.6. *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 $p_{00}(4) = 0.5749$.

2.3 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, at time n of the process. Let

$$\nu_i^{(n)} = \mathbb{P}(X_n = i), \qquad i \in E, n \geq 0,$$

(this is the probability mass function of X_n) and for $K = card(E) = |E|$ let

$$\boldsymbol{\nu}^{(n)} = \left(\nu_1^{(n)}, \ldots, \nu_K^{(n)}\right),$$

be the $K-$dimensional row vector of probabilities for the chain at time n.

Theorem 2.3.1. *With the notation introduced above, we have*

$$\boldsymbol{\nu}^{(m+n)} = \boldsymbol{\nu}^{(m)}\mathbf{P}_n, \qquad \textit{for } n \geq 1, m \geq 0.$$

and hence

$$\boldsymbol{\nu}^{(n)} = \boldsymbol{\nu}^{(0)}\mathbf{P}_n, \qquad \textit{for } n \geq 1.$$

Proof. For any $j \in E$, we have

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

□

It is then concluded that:

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

Remark 2.3.2. *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).

Exercise 2.3.3. *Suppose that $E = \mathbb{Z}$. A famous (and sometimes useful) 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 $\mathbb{P}(X_n = j | X_0 = i)$.

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

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

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

2.4 Properties of Markov chains

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.

2.4.1 Recurrence and transience

We begin with the concept of recurrence:

Definition 2.4.1. *Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain on a state-space E. A state $i \in E$ is* ***recurrent*** *if*

$$\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$$

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

Recurrence is of interest, if we want to answer questions about eventual returns to given states of the Markov chain.

Next we introduce the idea of the **first passage time** for $i, j \in E$:

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

with $i \neq j$. This is the probability that the **first time** we visit state j, is at time n.
Likewise, we can define

$$f_{ii}(n) = \mathbb{P}(X_n = i, X_{n-1} \neq i, \ldots, X_1 \neq i | X_0 = i),$$

which is the probability that the **first time** we return to state i, is at time n.
Also, write

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

which is the probability that the chain *ever* visits j, starting at i. What condition, on f_{jj} is needed for state j to be recurrent? Clearly, we need $f_{jj} = 1$.

Exercise 2.4.2. *Show that $f_{ii}(1) = p_{ii}$ and that we have the recursion*

$$p_{ii}(n) = \sum_{l=0}^{n} f_{ii}(l) p_{ii}(n-l), \qquad n \geq 1, (f_{ii}(0) := 0 \text{ for all } i).$$

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 4

We start off with the case $n = 1$. Here we have $p_{ii}(0) = 1$ and $f_{ii}(0) = 0$. Also, for $n = 1$, the first visit is equal to the nth visit, we can write $p_{ii}(1) = f_{ii}(1)$ and

$$\sum_{l=0}^{1} f_{ii}(l) p_{ii}(1-l) = \underbrace{f_{ii}(0)}_{=0} p_{ii}(1) + f_{ii}(1) \underbrace{p_{ii}(0)}_{=1} = f_{ii}(1).$$

Altogether we have

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

In the general case, we argue as follows. Consider all possible cases where the process starts at $X_0 = i$ and also $X_n = i$. The first return to state i occurs at the lth transition. Now define the disjoint events (for $l = 1, \dots, n$) A_l that the first visit to i (after time 0) takes place at time l, i.e.

$$A_l := \{X_l = i, X_r \neq i, \text{for } 1 \leq r < l\}.$$

Then (using the law of total probability):

$$\mathbb{P}(X_n = i | X_0 = i) = \sum_{l=1}^{n} \mathbb{P}(\{X_n = i\} \cap A_l | \{X_0 = i\}).$$

Now, we use the Markov property to deduce that

$$\begin{aligned}
\mathbb{P}(\{X_n = i\} \cap A_l | \{X_0 = i\}) &= \mathbb{P}(\{X_n = i\} | A_l \cap \{X_0 = i\}) \frac{\mathbb{P}(A_l \cap \{X_0 = i\})}{\mathbb{P}(\{X_0 = i\})} \\
&= \mathbb{P}(\{X_n = i\} | A_l \cap \{X_0 = i\}) \mathbb{P}(A_l | \{X_0 = i\}) \\
&= \mathbb{P}(\{X_n = i\} | \{X_l = i\}) \mathbb{P}(A_l | \{X_0 = i\}) \\
&= p_{ii}(n-l) f_{ii}(l).
\end{aligned}$$

So, altogether, we have

$$\mathbb{P}(X_n = i | X_0 = i) = p_{ii}(n) = \sum_{l=1}^{n} p_{ii}(n-l) f_{ii}(l).$$

From the simple exercise, we see that:

A state $j \in E$ is recurrent if and only if, the $f_{jj} = 1$ holds. Likewise, a state $j \in E$ is transient if and only if $f_{jj} < 1$.
In general, this can be very difficult to calculate, so let us consider a condition on the $n-$step transition probabilities.

Theorem 2.4.3. *We have:*

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* $\sum_{n=1}^{\infty} p_{jj}(n) < \infty$.

Proof.

Suppose the Markov chain is currently in state $j \in E$ and j is recurrent, then with probability 1 it will return to state j. By the Markov property, we know that the process starts over again when it reenters j, and we know that with probability 1, it will reenter state j again (and again and again...). Hence, the process will in fact reenter the recurrent state j infinitely often.

" j transient $\Rightarrow \sum_{n=1}^{\infty} p_{jj}(n) < \infty$":
Suppose now that j is transient. Then, each time when the chain visits j, there is a positive probability $(1 - f_{jj})$ that it will never again enter state j. Let M denote the number of periods that the chain is in state j:

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

If the chain starts at j, then the probability that the chain will be in state j exactly n times (for $n \geq 1$) is given by

$$\mathbb{P}(M = n | X_0 = j) = f_{jj}^{n-1}(1 - f_{jj}),$$

This is the probability mass function of a geometric distribution, which has (finite) mean

$$\mathbb{E}(M|X_0 = j) = \sum_{n=1}^{\infty} n f_{jj}^{n-1}(1 - f_{jj}) = \frac{(1 - f_{jj})}{(1 - f_{jj})^2} = \frac{1}{(1 - f_{jj})} < \infty,$$

We used the following property of the geometric series: Let $q \in \mathbb{R}$ with $|q| < 1$. Set $f(q) := \sum_{n=0}^{\infty} q^n = (1-q)^{-1}$. Then $f'(q) = \sum_{n=0}^{\infty} nq^{n-1} = \sum_{n=1}^{\infty} nq^{n-1} = (1-q)^{-2}$.

Then we have

$$\infty > \frac{1}{(1 - f_{jj})} = \mathbb{E}\left(M|X_0 = j\right) = \sum_{n=0}^{\infty} \mathbb{E}(I_n|X_0 = j) = \sum_{n=0}^{\infty} \mathbb{P}(X_n = j|X_0 = j)$$
$$= \sum_{n=0}^{\infty} p_{jj}(n).$$

" j transient $\Leftarrow \sum_{n=1}^{\infty} p_{jj}(n) < \infty$":
Conversely, suppose that $\sum_{n=0}^{\infty} p_{jj}(n) < \infty$. Then M is a random variable with finite mean, thus M 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.

" j recurrent $\Leftrightarrow \sum_{n=1}^{\infty} p_{jj}(n) = \infty$":
Similarly, we conclude that the state j is recurrent if and only if , starting in state j, the expected number of returns to state j is infinite. □

These results lead us to

Corollary 2.4.4. *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 $n \to \infty$. Similarly as in Exercise 2.4.2, we have for any $i \in E$ that

$$p_{ij}(n) = \sum_{l=0}^{n} f_{ij}(n-l)p_{jj}(l), \qquad n \geq 1, f_{ij}(0) := 0, \text{for all } i.$$

Note that $p_{ij}(0) = 0$ for $i \neq j$. Then

$$\begin{aligned}\sum_{n=0}^{\infty} p_{ij}(n) &= \sum_{n=0}^{\infty}\sum_{l=0}^{n} f_{ij}(n-l)p_{jj}(l) = \sum_{l=0}^{\infty} p_{jj}(l)\sum_{n=l}^{\infty} f_{ij}(n-l) \\ &= \sum_{l=0}^{\infty} p_{jj}(l)\underbrace{\sum_{n=0}^{\infty} f_{ij}(n)}_{=f_{ij}} \leq \sum_{l=0}^{\infty} p_{jj}(l) < \infty.\end{aligned}$$

The necessary condition for the convergence of the infinite series is $p_{ij}(n) \to 0$ as $n \to \infty$. □

We can see that a state is either recurrent or transient.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

First hitting times and mean recurrence time

Lecture 5

An alternative and important way to classify the states of a Markov chain is through the first hitting time

$$T_i = \inf\{n \geq 1 : X_n = i\}, \qquad \text{for } i \in E,$$

where we use the convention that $\inf\{\emptyset\} = \infty$.

Definition 2.4.5. *The **mean recurrence time** μ_i of a state $i \in E$ is defined as $\mu_i = \mathbb{E}[T_i|X_0 = i]$.*

Note that $\mathbb{P}(T_i = \infty|X_0 = i) > 0$ if and only if i is transient; in that case $\mu_i = \mathbb{E}[T_i|X_0 = i] = \infty$.
For a recurrent state $i \in E$, we have

$$\begin{aligned}\mu_i = \mathbb{E}[T_i|X_0 = i] &= \sum_{n=1}^{\infty} n\mathbb{P}(T_i = n|X_0 = i) \\ &= \sum_{n=1}^{\infty} n\mathbb{P}(X_n = i, X_{n-1} \neq i, \cdots, X_1 \neq i|X_0 = i) = \sum_{n=1}^{\infty} nf_{ii}(n),\end{aligned}$$

which can be finite or infinite.
Summarising we can say that

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

Definition 2.4.6. *A recurrent state $i \in E$ is called **null** if $\mu_i = \infty$ and **positive** (or non-null) if $\mu_i < \infty$.*

An important result, proved in Grimmett & Stirzaker (2001*b*, p. 222, 232) is:

Theorem 2.4.7. *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 $j \in E$.*

2.4.2 Aperiodicity and ergodicity

Definition 2.4.8. *The period of a state i is defined by*

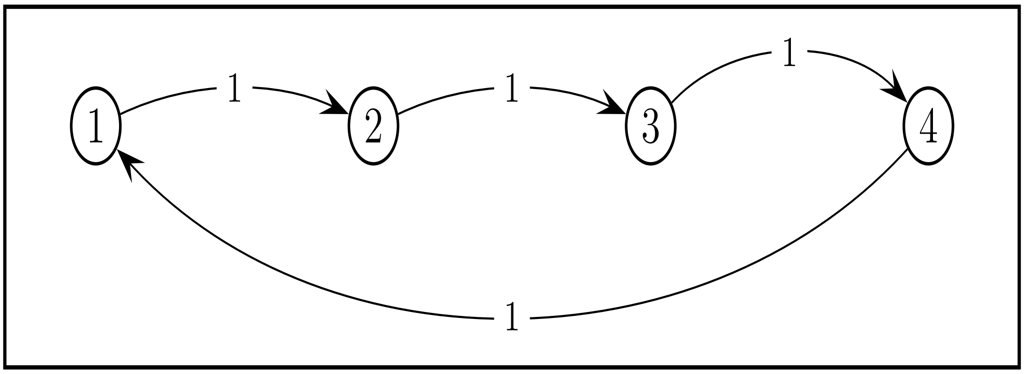
$$d(i) = gcd\{n : p_{ii}(n) > 0\}$$

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****.*

Definition 2.4.9. *A state is* ***ergodic*** *if it is positive, recurrent and aperiodic.*

Example 2.4.10. *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 $d_i = 4$.

2.5 Communicating classes

We now look at the way various states are inter-related.

Definition 2.5.1.
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 $m \geq 0$ 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 2.5.2. *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)

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

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

□

Remark 2.5.3.

- *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 formalized in Theorem 2.5.6.*

We have the following Theorem:

Theorem 2.5.4. *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: Let $i \leftrightarrow j$ and let i be recurrent. 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

$$p_{jj}(m+l+n) \geq \underbrace{p_{ji}(m)}_{>0} p_{ii}(l) \underbrace{p_{ij}(n)}_{>0}.$$

Then

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

Now, we only need to apply Theorem 2.4.3 to finish the proof. □

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

Definition 2.5.5. *A set of states C is*

1. ***closed*** *if $p_{ij} = 0$ for all $i \in C$, $j \notin C$*
2. ***irreducible*** *if $i \leftrightarrow j$ 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 2.5.4, this makes sense. Importantly, if the entire state-space E is irreducible, then we say that the **Markov chain is irreducible**.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

2.5.1 The decomposition theorem

Lecture 6

An important result is:

Theorem 2.5.6. *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. Let $C_1, C_2, \ldots$ denote the recurrent equivalence classes of $\leftrightarrow$. We only need to show that each C_r is closed. Suppose there exists an $i \in C_r$, $j \notin C_r$, such that $p_{ij} > 0$. Hence, $i \to j$, but $j \not\to i$. Then

$$\mathbb{P}(X_n \neq i \text{ for all } n \geq 1 | X_0 = i) \geq \mathbb{P}(X_1 = j | X_0 = i) = p_{ij} > 0,$$

which contradicts that i is recurrent. □

Remark 2.5.7. *Note that in the decomposition theorem above T is not assumed to be a* communicating class*. More precisely, T is the collection of all transient states of the Markov chain (which do not need to communicate with each other). In general, we can denote by $T_1, T_2, \ldots$ the transient communicating classes, then $T = \cup_i T_i$.*

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 set.

We finish the subsection with the interesting result:

Lemma 2.5.8. *Let $K < \infty$. Then at least one state is recurrent and all recurrent states are positive.*

Proof. Suppose that all states are transient; then according to Corollary 2.4.4 $\lim_{n\to+\infty} p_{ij}(n) = 0$ for all i. Since the transition matrix is stochastic, we have $\sum_{j\in E} p_{ij}(n) = 1$. Then take the limit through the summation sign to obtain

$$\lim_{n\to+\infty} \sum_j p_{ij}(n) = 0.$$

However, this is a contradiction as the L.H.S must be 1. The second part is proved similarly. □

Note: The proof of the second part goes as follows: Assume there exists a null recurrent state $j \in E = T \cup (\bigcup_i C_i)$. I.e. there exists an i^* such that $j \in C_{i^*}$. Since C_{i^*} is a closed, recurrent communicating class with one null recurrent state, we know that in fact *all* states in C_{i^*} are null recurrent. Also, all states in C_{i^*} are essential (since they are recurrent). Note now that a transition matrix restricted to an essential class is stochastic.

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

According to Theorem 2.4.7 we know that $p_{jj}(n) \to 0$ as $n \to \infty$ and also $p_{ij}(n) \to 0$ as $n \to \infty$ for all $i \in E$. Since all states in the class C_{i^*} are null recurrent, we get in fact $p_{ik}(n) \to 0$ as $n \to \infty$ for all $i \in E$ and also for all $k \in C_{i^*}$. As before, we take the limit through the summation sign to obtain

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

which is a contradiction.

We conclude this subsection with some remarks.

Remark 2.5.9.

- *Every recurrent class is closed.*
- *Every finite closed class is recurrent and also positive.*

If you would like to read the proofs of the properties stated above, then have a look at Chapter 1 of the excellent textbook Norris (1998).

2.5.2 Class properties

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

2.6 Application: Gambler's ruin

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

Consider a gambler who at each play of the game has

- 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?
(We assume that $N \in \mathbb{N}$.)

Let X_n denote the gambler's fortune at time n. Then $\{X_n\}_{n\in\{0,1,2,\dots\}}$ is a Markov chain with transition probabilities

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

How many classes does this Markov chain have? Three classes: $\{0\}, \{1, 2, \dots, N-1\}, \{N\}$,the first and third being positive recurrent and the second transient.
Recall that a transient class will only be visited finitely many times. Hence after some finite amount of time, the gambler will either win N or go broke.

For $i = 0, 1, \dots, N$, define the event

$$W_i := \text{gambler's fortune will eventually reach } N \text{ if he starts with } i \text{ units},$$

and

$$h_i = \mathbb{P}(W_i) = \mathbb{P}(X_n = N \text{ for some } n \geq 0 | X_0 = i).$$

Define the event $F :=$ gambler wins first game. Clearly, $\mathbb{P}(F) = p$. Also $F^c =$ gambler loses first game, $\mathbb{P}(F^c) = q = 1 - p$. If he wins, then he has $i + 1$ otherwise he has $i - 1$. By conditioning on the outcome of the initial game we get the recurrence relation

$$h_i = \mathbb{P}(W_i) = \mathbb{P}(W_i|F)\mathbb{P}(F) + \mathbb{P}(W_i|F^c)\mathbb{P}(F^c) = h_{i+1}p + h_{i-1}q, \qquad i = 1, 2, \dots, N-1.$$

We need to solve the difference equation

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

Try $h_i = cx^i$. 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 $x = 1$. The general solution is given by

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

for constants c_1, c_2.

Using the boundary conditions $h_0 = 0$ and $h_N = 1$, we get

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

As $N \to \infty$, we have

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

So if $p > 0.5$, there is a positive probability that the gambler's fortune will increase indefinitely. However, if $p \leq 0.5$, then the gambler will go broke with probability 1 (assuming he plays against an infinitely rich adversary).

Note that I do not expect you to be able to solve general difference equations in the exam, however, I do expect you to be able to reproduce the solution to the gambler's ruin problem!

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

2.7 Stationarity

Lecture 7

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 2.7.1. *A vector $\boldsymbol{\pi}$ is a stationary distribution of a Markov chain $\{X_n\}_{n \in \mathbb{N}_0}$ on E if:*

1. *for each $j \in E$, $\pi_j \geq 0$ and $\sum_{j \in E} \pi_j = 1$.*
2. *$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, that is, for each $j \in E$, $\pi_j = \sum_{i \in E} \pi_i p_{ij}$.*

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

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

for any $n \in \mathbb{N}$. 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 $\boldsymbol{\pi}$.

Exercise 2.7.2. *Assume X_0 has distribution $\boldsymbol{\pi}$. Show that also X_n has distribution $\boldsymbol{\pi}$ for all $n \in \mathbb{N}$.* Hint:

Use Theorem 2.3.1.

Our objective is to study the limiting behaviour of the Markov chain. We start with a result for Markov chains with a finite state space.

Theorem 2.7.3. *Let $K = |E| < \infty$. Suppose for some $i \in E$ that*

$$p_{ij}(n) \to \pi_j \text{ as } n \to \infty \text{ for all } j \in E.$$

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

Proof. Homework, see problem sheet! □

Note that sometimes we refer to the stationary distribution also as *invariant* distribution or as *equilibrium*. Now we turn to the case of a not necessarily finite state space.

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

Theorem 2.7.4. *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.*

In order to prove the theorem, we need an intermediate lemma.

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

$$\rho_i(j) = \mathbb{E}[N_i(j)|X_0 = j], \qquad \text{where } N_i(j) = \sum_{n=1}^{\infty} \mathbb{I}_{\{i\}\cap\{T_j\geq n\}}(X_n, \mathbf{X}_{1:n}), \tag{2.7.1}$$

where T_j is the first time to return to state j. (Hence $N_j(j) = 1$ and $\rho_j(j) = 1$). Then

$$\rho_i(j) = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i, T_j \geq n | X_0 = j).$$

Write $\boldsymbol{\rho}(j)$ as the vector of these quantities. Since $T_j = \sum_{i\in E} N_i(j)$, we find, taking (conditional) expectations that

$$\mu_j = \mathbb{E}[T_j|X_0 = j] = \sum_{i\in E} \mathbb{E}[N_i(j)|X_0 = j] = \sum_{i\in E} \rho_i(j). \tag{2.7.2}$$

Lemma 2.7.5. *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 $\boldsymbol{\rho}(j) = \boldsymbol{\rho}(j)\mathbf{P}$.*

Note that the sum in the definition of $N_i(j)$ in formula (2.7.1) has to be understood in the following sense: For each $n \in \mathbb{N}$, we have $\mathbf{X}_{1:n} := (X_1, \ldots, X_n)$ and

$$\mathbb{I}_{\{i\}\cap\{T_j\geq n\}}(X_n, \mathbf{X}_{1:n}) = \begin{cases} 1, & \text{if } X_n = i \text{ and } T_j \geq n, \\ 0, & \text{otherwise.} \end{cases}$$

Recall the notion of the first passage time $f_{ij}(n)$!

Proof.

First we prove that $\rho_i(j) < \infty$ for all $i \neq j$. Let

$$l_{ji}(n) = \mathbb{P}(X_n = i, T_j \geq n | X_0 = j),$$

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

$$f_{jj}(m+n) \geq l_{ji}(m) f_{ij}(n), \tag{2.7.3}$$

since the first return takes place after $n+m$ steps if
(1.) $X_m = i$, (2.) there is no return to j prior to time m, (3.) the next subsequent visit to j takes place after another n steps.
Since the chain is irreducible, there exists an n^* such that $f_{ij}(n^*) > 0$. Using (2.7.3), we get

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

then

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

Next, we prove $\boldsymbol{\rho}(j) = \boldsymbol{\rho}(j)\mathbf{P}$. As before,

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

Also, $l_{ji}(1) = p_{ji}$, and for $n \geq 2$:

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

Then, by summing over n, we get

$$\rho_i(j) = \underbrace{1}_{=\rho_j(j)} p_{ji} + \sum_{r:r\neq j} \underbrace{\left(\sum_{n=2}^{\infty} l_{jr}(n-1) \right)}_{=\rho_r(j)} p_{ri} = \sum_r \rho_r(j) p_{ri},$$

which concludes the proof. □

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 8

Lemma 2.7.6. *Every positive recurrent, irreducible chain has a stationary distribution.*

Proof. From Lemma 2.7.5 we get the representation result $\boldsymbol{\rho}(j) = \boldsymbol{\rho}(j)\mathbf{P}$. From equation (2.7.2) 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. □

Theorem 2.7.7. *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 2.7.5. This root is always non–negative and can in fact to be taken strictly positive. The proof of the uniqueness is left as an exercise! □

You can find the solution to the above theorem in Grimmett & Stirzaker (2001*a*, p. 305, Problem 6.15.7).

Exercise 2.7.8. *A useful result used below is as follows. Let X be a non-negative, integer valued random variable. Show that*

$$\mathbb{E}[X] = \sum_{n=0}^{\infty} \mathbb{P}(X > n) = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n). \tag{2.7.4}$$

You can check your solution with the following calculation:

Solution to Exercise 2.7.8.

$$\begin{aligned}\mathbb{E}[X] &= \sum_{n=0}^{\infty} n\mathbb{P}(X = n) = \sum_{n=0}^{\infty}\sum_{m=0}^{n-1} \mathbb{P}(X = n) = \sum_{m=0}^{\infty}\sum_{n=m+1}^{\infty} \mathbb{P}(X = n)\\ &= \sum_{m=1}^{\infty} \mathbb{P}(X \geq m) = \sum_{m=0}^{\infty} \mathbb{P}(X > m).\end{aligned}$$

□

We are now in a position to prove Theorem 2.7.4. 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 2.7.7.

Proof of Theorem 2.7.4.

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 2.4.4. Since $\boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi}$, for any j

$$\pi_j = \sum_{i \in E} \pi_i p_{ij}(n) \to 0 \quad n \to \infty$$

thus, π could not be a stationary vector (see Definition 2.7.1); all states are recurrent. This follows from switching the order of summation and limits using the dominated convergence theorem.

Theorem 2.7.9 (Dominated Convergence Theorem). *If $\sum_i a_i(n)$ is an absolutely convergent series for all $n \in \mathbb{N}$ such that*

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

Then $\sum_i |a_i| < \infty$ and

$$\sum_i a_i = \sum_i \lim_{n\to\infty} a_i(n) = \lim_{n\to\infty} \sum_i a_i(n).$$

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 2.7.9 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. $\mathbb{P}(X_0 = i) = \pi_i$ for each i), using (2.7.4),

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

But, $\mathbb{P}(T_j \geq 1, X_0 = j) = \mathbb{P}(X_0 = j)$ (since $T_j \geq 1$ by definition) and for $n \geq 2$

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

where we have used homogeneity.

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

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

hence

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

We define $a_n = \mathbb{P}(X_m \neq j, 0 \leq m \leq n)$. Then, summing over n (telescoping sum!)

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

However,

$$a_n \to \mathbb{P}(X_m \neq j, \ \forall m) = 0$$

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

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

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 2.7.5.

□

Remark 2.7.10. *Note that Theorem 2.7.4 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 2.7.11. *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!

$$\mu_1 = 1/\pi_1 = 3, \qquad \mu_2 = 1/\pi_2 = 3/2.$$

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 9

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

Theorem 2.7.12. *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 2.7.13.
- *If the irreducible chain is transient or null recurrent, then $p_{ij}(n) \to 0$ for all $i, j \in E$ since $\mu_j = \infty$.*
- *If the chain is irreducible aperiodic and positive recurrent then we have:*

$$\lim_{n\to+\infty} p_{ij}(n) = \pi_j = 1/\mu_j \text{ for all } i, j \in E,$$

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

- *The limiting distribution of an irreducible aperiodic chain does not depend on the starting point ($X_0 = i$), but forgets its origin, hence we have*

$$\mathbb{P}(X_n = j) = \sum_{i\in E} p_{ij}(n)\mathbb{P}(X_0 = i) \to \frac{1}{\mu_j}, \text{ as } n \to \infty.$$

2.7.1 Existence of a stationary distribution on a finite state space

So far, we have studied general Markov chains in discrete time with a discrete 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. $K = Card(E) = |E| < \infty$, 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.

Definition 2.7.14. *A state $i \in E$ is* ***essential*** *if for all j such that $i \to j$ it is also true that $j \to i$. A state $i \in E$ is* ***inessential*** *if it is not essential.*

Lemma 2.7.15. *If i is an essential state and $i \to j$, then j is essential.*

Proof. Choose any $l \in E$ such that $j \to l$. We know that $i \to j$, hence we have $i \to l$. Since i is essential, we have $l \to i$ and also $i \to j$, which implies $l \to j$. □

Remark 2.7.16. *The previous lemma implies that all states in a single communicating class are either all essential or inessential.*

Remark 2.7.17. *The definition above is equivalent to saying: A state is essential if it belongs to a closed communicating class. Also, a state is essential, if it communicates with every state which is accessible from it.*

Note that one can show that all inessential states are transient. This implies that all recurrent states are essential! I.e.

$$\begin{aligned} \text{inessential} &\Longrightarrow \text{transient} \\ \text{recurrent} &\Longrightarrow \text{essential.} \end{aligned}$$

Note here that essential states need not be recurrent. For example, when we consider the asymmetric random walk on $E = \mathbb{Z}$ (with $0 < p < 1$, $p \neq 0.5$), then we have an irreducible Markov chain (which implies in particular that all states are essential), which is transient.
However, on a **finite** state space (i.e. when $K < \infty$) the definition of a transient state coincides with that of an inessential state. Also, the definition of a recurrent state coincides with the one of an essential state, i.e.

$$\begin{aligned} \text{inessential} &\Longleftrightarrow \text{transient} \\ \text{recurrent} &\Longleftrightarrow \text{essential.} \end{aligned}$$

Lemma 2.7.18. *Every finite chain has at least one essential class.*

Proof. We can inductively define a sequence $(i_0, i_1, \dots)$ as follows: Fix an arbitrary initial state i_0. For $n \geq 1$, given $(i_0, i_1, \dots, i_{n-1})$, if i_{n-1} is essential, stop. Otherwise we need to proceed... Recall that an essential state is a state in a closed communicating class. Hence, if state i_{n-1} is inessential, it is in a communicating class, which is not closed. Hence, we can find a state i_n such that $i_{n-1} \to i_n$ but $i_n \not\to i_{n-1}$. There can be no repeated states in the sequence since if $m < n$ and $i_n \to i_m$, then $i_n \to i_{n-1}$ which would be a contradiction. Since the state space is finite and since the sequence cannot repeat elements, the sequence must terminate in a state, which is in a closed communicating class, i.e. in an essential state. □

Remark 2.7.19. *Note that according to Lemma 2.5.8, every finite Markov chain has at least one recurrent state. Then, Remark 2.5.9 states that every recurrent class is closed. Hence, combining these results, we get that every finite Markov chain has at least one closed communicating class. This implies the existence of an essential class. So there was actually no need to prove Lemma 2.7.18, since the result is already implied by our previous results!*

Lemma 2.7.20. *Let C denote an essential 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

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

□

Theorem 2.7.21. *Suppose we have a finite state space. If $\boldsymbol{\pi}$ is stationary for the transition matrix $\mathbf{P}$, then $\pi_i = 0$ for all inessential states i.*

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 10

Proof.

Let C denote an essential communicating class and let $\mathbf{P}$ denote the transition matrix for the entire chain (not just for the restriction to C). Define

$$\boldsymbol{\pi}\mathbf{P}(C) := \sum_{j \in C} (\boldsymbol{\pi}\mathbf{P})_j.$$

Note that

$$(\boldsymbol{\pi}\mathbf{P})_j = \sum_{l \in C} \pi_l p_{lj} + \sum_{l \notin C} \pi_l p_{lj}.$$

Since both sums are finite we can interchange the order of summation and we obtain

$$\begin{aligned}\boldsymbol{\pi}\mathbf{P}(C) &= \sum_{j \in C} (\boldsymbol{\pi}\mathbf{P})_j = \sum_{j \in C} \left(\sum_{l \in C} \pi_l p_{lj} + \sum_{l \notin C} \pi_l p_{lj} \right) \\ &= \sum_{l \in C} \pi_l \sum_{j \in C} p_{lj} + \sum_{l \notin C} \pi_l \sum_{j \in C} p_{lj}.\end{aligned}$$

Recall that for $l \in C$ we have $\sum_{j \in C} p_{lj} = 1$. Hence

$$\boldsymbol{\pi}\mathbf{P}(C) = \sum_{l \in C} \pi_l + \sum_{l \notin C} \sum_{j \in C} \pi_l p_{lj}.$$

Define $\boldsymbol{\pi}(C) := \sum_{l \in C} \pi_l$. Since $\boldsymbol{\pi}$ is a stationary distribution, we have that $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, and hence also

$$\boldsymbol{\pi}\mathbf{P}(C) = \boldsymbol{\pi}(C).$$

This implies that

$$\sum_{l \notin C} \sum_{j \in C} \pi_l p_{lj} = 0.$$

The sum consists of only nonnegative elements and is therefore only zero if all elements are zero, i.e.

$$\pi_l p_{lj} = 0, \qquad \text{for all } l \notin C, j \in C. \tag{2.7.5}$$

Suppose now that l_0 is inessential. According to the proof of Lemma 2.7.18 we can find a sequence of states $l_0, l_1, l_2, \ldots, l_r$ satisfying $p_{l_{i-1} l_i} > 0$ for $i = 1, \ldots, r$; the states $l_0, l_1, l_2, \ldots, l_{r-1}$ are inessential and $l_r \in C$ where C is an essential communicating class. We know that $p_{l_{r-1} l_r} > 0$, but also $\pi_{l_{r-1}} p_{l_{r-1} l_r} = 0$ (from (2.7.5)). Hence $\pi_{l_{r-1}} = 0$. We can now carry out an induction backwards where we argue that if $\pi_{l_i} = 0$, then

$$0 = \pi_{l_i} = \sum_{j \in E} \pi_j p_{j l_i},$$

which implies $\pi_j p_{j l_i} = 0$ and $\pi_j = 0$ for $j \in \{l_{r-1}, l_{r-2}, \ldots, l_0\}$. We conclude that $\pi_{l_0} = 0$.

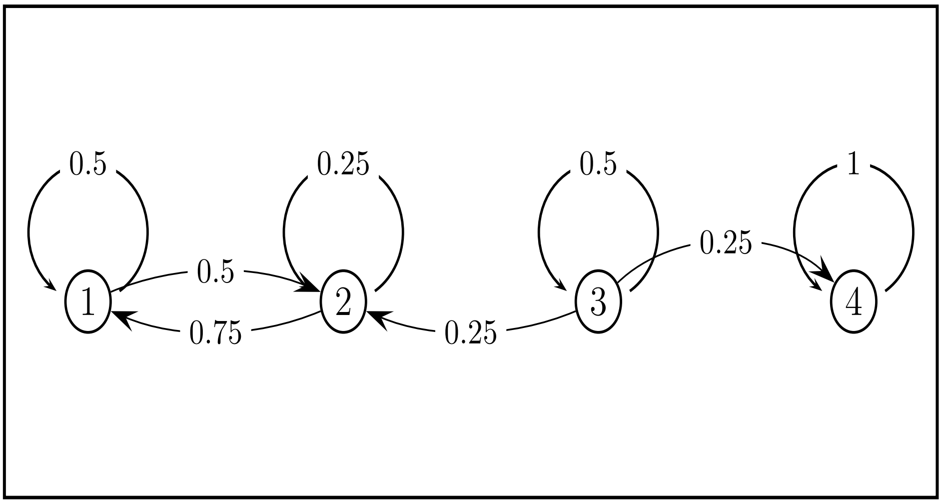
□

Exercise 2.7.22. *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:



We have three communicating classes:

1. $C_1 = \{1, 2\}$ is (positive) recurrent, closed;
2. $E_1 = \{3\}$ is transient, not closed.
3. $C_2 = \{4\}$ is (positive) recurrent, closed, absorbing.

We can now formulate an important result.

Theorem 2.7.23. *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 essential communicating class.*

Proof of Theorem 2.7.23.
First suppose there is a unique essential communicating class C. Write $\mathbf{P}_C$ for the *restriction of the matrix $\mathbf{P}$ to the states in C*. Suppose $i \in C$ and $p_{ij} > 0$. Since i is essential and $i \to j$, it must also hold that $j \to i$, hence $j \in C$.
So we know that $\mathbf{P}_C$ is a transition matrix which has to be irreducible on C (not necessarily on the entire state space!). From Lemma 2.5.8 we obtain that all states in C are positive recurrent, hence we can apply Theorem 2.7.4 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 2.7.21 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

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

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

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

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

Now suppose that there is a unique stationary distribution and *two* distinct essential 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.

□

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

2.7.2 Ergodic theorem

Lecture 11

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

Theorem 2.7.24 (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

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

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 $n \to \infty$.

2.8 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 is also $\boldsymbol{\pi}$. The reversed chain is defined to be, for any $n \in \{0,1,\ldots,N\}$

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

Theorem 2.8.1. *The sequence Y is a Markov chain which satisfies*

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

Proof.

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

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

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

Hence

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

Similarly, we get that

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

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

$$\mathbb{P}(Y_{n+1} = i_{n+1} | Y_n = i_n, Y_{n-1} = i_{n-1}, \ldots, Y_0 = i_0) = \mathbb{P}(Y_{n+1} = i_{n+1} | Y_n = i_n)$$

$$= \frac{\pi_{i_{n+1}} p_{i_{n+1} i_n}}{\pi_{i_n}},$$

which completes the proof.

□

Definition 2.8.2. *Let $X = \{X_n : n \in \{0, 1, \ldots, N\}\}$ be an irreducible Markov chain such that X_n has a stationary distribution $\boldsymbol{\pi}$ 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 2.8.3. *$\{X_n\}_{n \in \{0,1,\ldots,N\}}$ is time–reversible if and only if for any $i, j \in E$*

$$\pi_i p_{ij} = \pi_j p_{ji}. \tag{2.8.1}$$

Note that the condition (2.8.1) 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 (2.8.1) holds. □

Theorem 2.8.4. *For an irreducible chain, if there exist a probability vector $\boldsymbol{\pi}$ such that (2.8.1) 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 2.7.4. □

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

Exercise 2.8.6. *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 $\boldsymbol{\pi} = (1/3, 1/3, 1/3)$. Now we check the detailed balance equations: $\pi_i p_{ij} = \pi_j p_{ji}$. Here we need

$$\frac{1}{3} p_{ij} = \frac{1}{3} p_{ji}, \quad \text{i.e.} \quad p_{ij} = p_{ji}$$

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 $p = 1/2$.

2.9 Summary

2.9.1 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
   $$\mathbb{P}(X_n = i) \to \pi_i \text{ for all } i \in E \text{ as } n \to \infty.$$
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)
   $$\mathbb{P}(X_n = i) \to 0 \text{ for all } i \in E \text{ as } n \to \infty.$$
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)
   $$\mathbb{P}(X_n = i) \to 0 \text{ for all } i \in E \text{ as } n \to \infty.$$

2.9.2 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 essential communicating class. $\Leftrightarrow$ There is a unique closed communicating class.

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

- Find all 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 $Card(C_i)$-dimensional row vector with non-negative entries. Solve $\boldsymbol{\pi}_{C_i}\mathbf{P}(C_i) = \boldsymbol{\pi}_{C_i}$.
- One possible stationary distribution is then given by the row vector $\boldsymbol{\pi}$ which consists of the corresponding elements of the vectors $\boldsymbol{\pi}_{C_i}$, for $i = 1, 2, \ldots$ and of zeros corresponding to the inessential (transient) states. 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, derive the conditions needed to ensure that $\boldsymbol{\pi}$ has non-negative entries and the elements sum up to 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!

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

Chapter 3

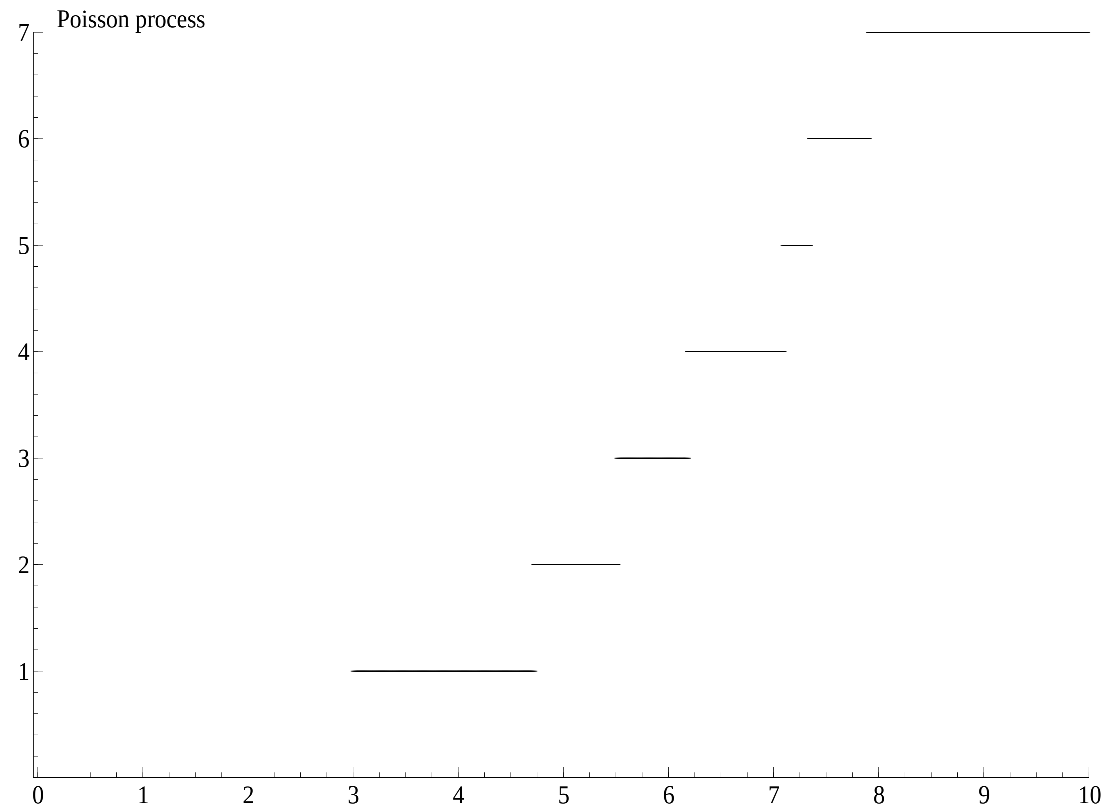
Poisson processes

Lecture 12

3.1 Introduction

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*.

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.



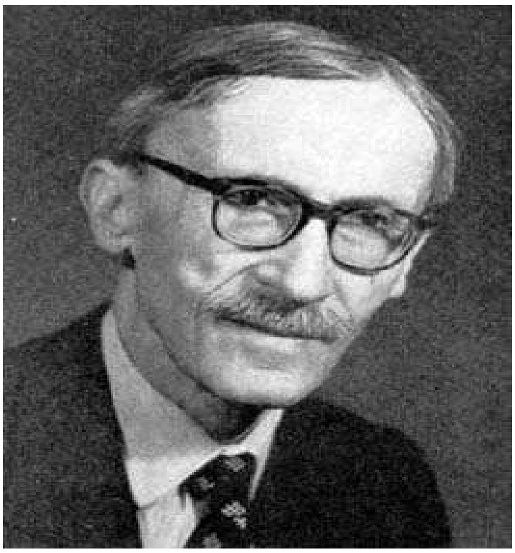
A realisation of a Poisson process on $[0, 10]$.

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
- A process for mutations on a genealogical tree.

Siméon-Denis Poisson (21 June 1781- 25 April 1840) was a French mathematician, geometer, and physicist.

As a digression, the Poisson process (as well as Brownian motion) is a *Lévy* process, and we will see some common features in the definition of Poisson processes and Brownian motion.
Paul Pierre Lévy (15.09.1886 - 15.12.1971) was a French mathematician.



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

These ideas are important extensions of Poisson processes.

3.2 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. $N_t - N_{t-} \in \{0, 1\}$).

We use the following notation throughout the lecture notes: $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$ are the non-negative integers.

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 3.2.1. *Let $(T_n)_{n\geq 0}$ be a strictly increasing sequence of positive random variables with $T_0 = 0$ almost surely. The process $\{N_t\}_{t\geq 0}$ defined by*

$$N_t = \sum_{n\geq 1} \mathbb{I}_{\{t\geq T_n\}},$$

which takes values in $\mathbb{N}_0$ is called the counting process associated to the sequence $(T_n)_{n\geq 0}$.

Recall that

$$\mathbb{I}_{\{t\geq T_n\}} = \begin{cases} 1, & \text{if } t \geq T_n(\omega), \\ 0, & \text{if } t < T_n(\omega). \end{cases}$$

You can interpret T_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. Our notation is such that the stochastic process is written in a $\{\bullet\}$ notation and the subscript is the time parameter. N_t is the realized value of the process at time t.

3.2.1 Poisson process: First definition

We give a first definition of a Poisson process.
An important concept to recall here is the $o(\cdot)$ notation: A function, f, is $o(\delta)$ if

$$\lim_{\delta \downarrow 0} \frac{f(\delta)}{\delta} = 0.$$

Exercise 3.2.2. *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 3.2.3. *A* Poisson process $\{N_t\}_{t\geq 0}$ *of rate $\lambda > 0$ is a stochastic process with values in $\mathbb{N}_0$ satisfying:*

1. *$N_0 = 0$ (technically this only needs to hold with probability 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 $k \in \mathbb{N}_0$:*

$$\mathbb{P}(\{N_t - N_s = k\}) = \mathbb{P}(\{N_{t-s} = k\}).$$

4. *There is a 'single arrival', i.e. for any $t \geq 0$ and $\delta > 0$:*

$$\begin{aligned}\mathbb{P}(\{N_{t+\delta} - N_t = 1\}) &= \lambda\delta + o(\delta)\\ \mathbb{P}(\{N_{t+\delta} - N_t \geq 2\}) &= o(\delta)\end{aligned}$$

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.

For simplicity, we now denote $\mathbb{P}(\{N_t = k\}) = \mathbb{P}(N_t = k)$.

Also note that the single arrival property implies that

$$\mathbb{P}(\{N_{t+\delta} - N_t = 0\}) = 1 - \lambda\delta + o(\delta).$$

Note that a Poisson process is a counting process.

3.2.2 Poisson process: Second definition

If that definition is a little unclear, let us consider a second definition.

Definition 3.2.4. *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. *For any* $0 \leq s < t$*,* $k \in \mathbb{N}_0$ *we have*

$$\mathbb{P}(N_t - N_s = k) = \frac{(\lambda(t-s))^k e^{-\lambda(t-s)}}{k!}.$$

That is, the number of events in $[s,t]$ *is a Poisson random variable, of mean* $\lambda(t-s)$.

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.

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.

$$P(X_t = Y_t) = 1, \text{for each } t \geq 0.$$

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.

3.2.3 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 3.2.5. *Definition 3.2.3 implies 3.2.4.*

We are to show that Definition 3.2.3 implies Definition 3.2.4.

We will use the Laplace transform to verify this result. Recall that, for a random variable X with discrete support X, the Laplace transform is, for $u > 0$

$$\mathcal{L}_X(u) = \mathbb{E}[e^{-uX}] = \sum_{x\in\mathsf{X}} e^{-ux}\mathbb{P}(X = x).$$

Let us derive the Laplace transform of a Poisson random variable first.

Lemma 3.2.6. *The Laplace transform of a Poisson random variable of mean λt (i.e. $X \sim Poi(\lambda t)$) is given by*

$$\mathcal{L}_X(u) = \exp\{\lambda t[e^{-u} - 1]\}.$$

Proof.

$$\begin{aligned}
\mathcal{L}_X(u) = \mathbb{E}[e^{-uX}] &= \sum_{x=0}^{\infty} e^{-ux}\mathbb{P}(X = x) \\
&= \sum_{x=0}^{\infty} e^{-ux}\frac{(\lambda t)^x}{x!}\exp(-\lambda t) \\
&= \exp(-\lambda t)\sum_{x=0}^{\infty}\frac{(e^{-u}\lambda t)^x}{x!} \\
&= \exp(-\lambda t)\exp\left(e^{-u}\lambda t\right) = \exp\left(\lambda t(e^{-u} - 1)\right).
\end{aligned}$$

□

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ Lecture 13

In the following we study two different proofs of Theorem 3.2.5.

Proof of Theorem 3.2.5 (using Laplace transforms).

Recall that we need to show that the number of events in $[s, t]$ (for $0 \leq s < t$) is a Poisson random variable of mean $\lambda(t - s)$. Due to fact that we have stationary increments, it is in fact sufficient to prove that $N_t \sim Poi(\lambda t)$.
We begin by deriving a differential equation for $\mathcal{L}_N$ as follows. For $\delta > 0$, $t \geq 0$ and for $u > 0$,

$$\begin{aligned}
\mathcal{L}_N(t+\delta, u) &:= \mathbb{E}[e^{-uN_{t+\delta}}] \qquad \text{(multiply by 1 inside } \mathbb{E}(\cdot)) \\
&= \mathbb{E}[e^{-u[N_{t+\delta}-N_t]}e^{-uN_t}] \qquad \text{(use independent incr.)} \\
&= \mathbb{E}[e^{-u[N_{t+\delta}-N_t]}]\mathbb{E}[e^{-uN_t}] \qquad \text{(use stationary incr.)} \\
&= \mathbb{E}[e^{-uN_\delta}]\mathcal{L}_N(t, u). \qquad (3.2.1)
\end{aligned}$$

The third line follows via the independent increments property and the last by the stationarity.

Now consider

$$\begin{aligned}
\mathbb{E}[e^{-uN_\delta}] &= \sum_{x=0}^{\infty} e^{-ux}\mathbb{P}(N_\delta = x) \\
&= e^{-u\cdot 0}\mathbb{P}(N_\delta = 0) + e^{-u}\mathbb{P}(N_\delta = 1) + \sum_{x=2}^{\infty} e^{-ux}\mathbb{P}(N_\delta = x)
\end{aligned}$$

Recall the single-arrival property: $\mathbb{P}(\{N_{t+\delta} - N_t = 0\}) = 1 - \lambda\delta + o(\delta)$ and $\mathbb{P}(\{N_{t+\delta} - N_t = 1\}) = \lambda\delta + o(\delta)$, $\mathbb{P}(\{N_{t+\delta} - N_t \geq 2\}) = o(\delta)$. Also, for $u > 0$:

$$0 \leq \sum_{x=2}^{\infty} e^{-ux}\mathbb{P}(N_\delta = x) < \sum_{x=2}^{\infty}\mathbb{P}(N_\delta = x) = \mathbb{P}(N_\delta \geq 2) = o(\delta).$$

Hence:

$$\begin{aligned}
\mathbb{E}[e^{-uN_\delta}] &= 1\cdot(1 - \lambda\delta + o(\delta)) + e^{-u}(\lambda\delta + o(\delta)) + o(\delta) \\
&= 1 - \lambda\delta + e^{-u}\lambda\delta + o(\delta). \qquad (3.2.2)
\end{aligned}$$

Combining (3.2.1) and (3.2.2) 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 $\mathcal{L}_N(0, u) = \mathbb{E}(e^{-uN_0}) = \mathbb{E}(e^{-u\cdot 0}) = 1$, 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 $N_t \sim Poi(\lambda t)$. Now recall that we know that N has stationary increments, hence we can conclude that for $s < t$ $N_t - N_s$ has the same distribution as N_{t-s}, and by combining the two results, we get

$$N_t - N_s \sim Poi(\lambda(t - s)).$$

□

Proof of Theorem 3.2.5 (using forward equations). An alternative way is via the *forward equations*. It is very important to understand this concept.

Define, for $n \in \mathbb{N}_0 (= \mathbb{N} \cup \{0\})$.

$$p_n(t) = \mathbb{P}(N_t = n).$$

Using the properties of Definition 3.2.3, it must be that the probabilities will coincide with that of a Poisson random variable. Let $n = 0$, then

$$\begin{aligned} p_0(t+\delta) &= \mathbb{P}(N_{t+\delta} = 0) = \mathbb{P}(\text{no event in}[0, t+\delta]) \\ &= \mathbb{P}(\text{no event in}[0, t] \text{ and no event in}(t, t+\delta]) \end{aligned}$$

Now apply the independent increments property, then

$$\begin{aligned} p_0(t+\delta) &= \mathbb{P}(\text{no event in}[0, t])\mathbb{P}(\text{no event in}(t, t+\delta]) \\ &= p_0(t)[1 - \lambda\delta + o(\delta)], \end{aligned} \tag{3.2.3}$$

which is simply the probability that no events occurred up-to time t multiplied by the probability that no event occurs in $(t, t+\delta]$.

Returning to (3.2.3), 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 $p_0(0) = \mathbb{P}(N_0 = 0) = 1$. Using the same approach as above, it clearly follows that

$$p_0(t) = e^{-\lambda t}.$$

For $n \geq 1$ (i.e. $n \in \mathbb{N}$) we have

$$\begin{aligned}
p_n(t+\delta) &= \mathbb{P}(N_{t+\delta} = n) \\
&= \sum_{k=0}^{\infty} \mathbb{P}(N_{t+\delta} = n | N_t = k) \mathbb{P}(N_t = k) \quad \text{(Law of total probability)} \\
&= \sum_{k=0}^{\infty} \mathbb{P}((n-k) \text{ events in } (t, t+\delta]) \mathbb{P}(N_t = k) \\
&= \mathbb{P}(1 \text{ event in } (t, t+\delta]) \mathbb{P}(N_t = n-1) \\
&\quad + \mathbb{P}(0 \text{ events in } (t, t+\delta]) \mathbb{P}(N_t = n) + o(\delta) \\
&= p_{n-1}(t) \lambda \delta + p_n(t)(1 - \lambda \delta) + o(\delta) \\
&= p_n(t)(1 - \lambda \delta) + p_{n-1}(t) \lambda \delta + o(\delta).
\end{aligned}$$

Re-arranging and 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 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 $p_1(0) = \mathbb{P}(N_0 = 1) = 0$ 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 3.2.7. *Definition 3.2.4 also implies 3.2.3. I.e. the two Definitions are equivalent.*

Exercise 3.2.8. *Prove Remark 3.2.7.*

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 3.2.8. We need to check the four conditions in Definition 3.2.3:
Conditions (1) and (2): Conditions (1) and (2) are trivially satisfied. Hence we only have to show that the increments are stationary and that there is a single arrival.

Stationarity: Given any two distinct times $0 \leq s < t$ and for any $k \in \mathbb{N}_0$ we have
$$\mathbb{P}(N_t - N_s = k) = \frac{(\lambda(t-s))^k e^{-\lambda(t-s)}}{k!} = \mathbb{P}(N_{t-s} = k).$$
Single arrival: We need to apply condition (3) of Definition 3.2.4. For $k \in \mathbb{N}_0$, $t \geq 0$ and $\delta > 0$:
$$\mathbb{P}(N_{t+\delta} - N_t = k) = \frac{1}{k!}(\lambda\delta)^k e^{-\lambda\delta}.$$
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
$$\mathbb{P}(N_{t+\delta} - N_t = 1) = \lambda\delta e^{-\lambda\delta} = \lambda\delta\left(1 + \sum_{n=1}^{\infty} \frac{(-\lambda\delta)^n}{n!}\right) = \lambda\delta + o(\delta),$$
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
$$\mathbb{P}(N_{t+\delta} - N_t \geq 2) = o(\delta),$$

since

$$\lim_{\delta\to 0}\frac{\mathbb{P}(N_{t+\delta}-N_t\geq 2)}{\delta}=\lim_{\delta\to 0}\frac{\sum_{k=2}^{\infty}\mathbb{P}(N_{t+\delta}-N_t=k)}{\delta}=\lim_{\delta\to 0}\frac{\sum_{k=2}^{\infty}\frac{1}{k!}(\lambda\delta)^k e^{-\lambda\delta}}{\delta}$$
$$=\lim_{\delta\to 0}\sum_{k=2}^{\infty}\frac{1}{k!}\lambda^k\delta^{(k-1)}e^{-\lambda\delta}=0.$$

Alternatively, you could argue as follows:

$$\mathbb{P}(N_{t+\delta}-N_t=0)=\exp(-\lambda\delta)=1-\lambda\delta+\sum_{n=2}^{\infty}\frac{(-\lambda\delta)^n}{n!}=1-\lambda\delta+o(\delta),$$

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:

$$\mathbb{P}(N_{t+\delta}-N_t\geq 2)=1-\mathbb{P}(N_{t+\delta}-N_t<2)=1-\mathbb{P}(N_{t+\delta}-N_t=0)-\mathbb{P}(N_{t+\delta}-N_t=1)$$
$$=1-(1-\lambda\delta+o(\delta))-(\lambda\delta+o(\delta))=o(\delta).$$

□

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

3.3 Some properties of Poisson processes

Lecture 14

Now that we have a definition of our stochastic process, let us consider some properties of it.

3.3.1 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. The derivation is rather simple.

Let $(X_1,\ldots,X_n):=X_{1:n}$ be the inter-arrival times for the first n events. Now consider, for $t>0$

$$\begin{aligned}\mathbb{P}(X_1>t) &= \mathbb{P}(\text{no events in}[0,t])=\mathbb{P}(N_t=0)\\ &= e^{-\lambda t}.\end{aligned}$$

$\mathbb{P}(X_1>t)$ is sometimes called the *survival function* of X_1. We can now easily compute the cumulative distribution function of X_1:

$$F_{X_1}(t)=\mathbb{P}(X_1\leq t)=1-\mathbb{P}(X_1>t)=1-e^{-\lambda t}.$$

Hence the density function is:

$$f_{X_1}(t)=\lambda e^{-\lambda t},$$

which we recognise as an exponential density function, of rate λ ($\mathcal{E}xp(\lambda)$).

Before we study the second inter-arrival time, recall the following fact:

Let X, Y denote 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

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)}, \text{ for any } x, \text{ such that } f_X(x) > 0.$$

Also, the *conditional distribution function of Y given $X = x$* is defined as

$$F_{Y|X=x}(y|x) = \mathbb{P}(Y \leq y|X = x) = \int_{-\infty}^{y} f_{Y|X}(v|x)dv, \text{ for any } x, \text{ such that } f_X(x) > 0.$$

Also, recall that the conditional expectation satisfies

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \int_{-\infty}^{\infty} \mathbb{E}(Y|X = x)f_X(x)dx.$$

This implies that for any event A say, we have for $Y = \mathbb{I}_A$ that

$$\mathbb{P}(A) = \mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{P}(A|X = x)f_X(x)dx,$$

which can be viewed as the continuous version of the law of total probability.

Consider the second inter-arrival time, for $t > 0$

$$\begin{aligned}
\mathbb{P}(X_2 > t|X_1 = t_1) &= \mathbb{P}\left(\text{no events in } (t_1, t_1 + t]\middle| X_1 = t_1\right) \\
&= \mathbb{P}(N_{t_1+t} - N_{t_1} = 0|X_1 = t_1) \qquad \text{(indep . incr.)} \\
&= \mathbb{P}(N_{t_1+t} - N_{t_1} = 0) \qquad \text{(use stationary incr.)} \\
&= \mathbb{P}(N_t = 0) \\
&= e^{-\lambda t}.
\end{aligned}$$

I.e.

$$\mathbb{P}(X_2 > t) = \int_0^{\infty} \mathbb{P}(X_2 > t|X_1 = t_1)f_{X_1}(t_1)dt_1 = e^{-\lambda t}\int_0^{\infty} f_{X_1}(t_1)dt_1 = e^{-\lambda t}.$$

That is, independently of X_1, the random variable X_2 is exponentially distributed with parameter λ ($X_2 \sim \mathcal{E}xp(\lambda)$).

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 $(t_1, t_1 + t]$).

This construction can be repeated for any $n \in \mathbb{N}$ with $n \geq 2$ (conditioning on $X_{1:n-1}$). In particular,

$$\mathbb{P}(X_n > t|X_1 = t_1, \ldots, X_{n-1} = t_{n-1}) = \mathbb{P}\left(\text{no events in } (T, T + t]\right),$$

where $T = t_1 + \cdots + t_{n-1}$. Using induction on n, we can establish that:

Theorem 3.3.1. *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.*

Intuitively speaking that 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.

3.3.2 Time to the n^{th} event

A simple corollary to the above result is as follows.

Corollary 3.3.2. *Let $\{N_t\}_{t\geq 0}$ be a Poisson process of rate $\lambda > 0$. Then for any $n \geq 1$ the time to the n^{th} event follows a Gamma (n, λ) distribution:*

$$\begin{aligned} T_n &:= \sum_{i=1}^{n} X_i \\ f_{T_n}(t) &= \frac{\lambda^n}{\Gamma(n)} t^{n-1} e^{-\lambda t}, \quad t \geq 0. \end{aligned}$$

Proof. There are a few ways to prove this, but the simplest, is to use the uniqueness of the moment generating function.

Let $u \in \mathbb{R}$. By definition

$$\begin{aligned} \mathbb{E}[e^{uT_n}] &= \mathbb{E}[e^{u(\sum_{i=1}^{n} X_i)}] \qquad \text{use independence} \\ &= \prod_{i=1}^{n} \mathbb{E}[e^{uX_i}] \qquad \text{use identical distribution} \\ &= (\mathbb{E}[e^{uX}])^n \end{aligned}$$

where $X \sim \mathcal{E}xp(\lambda)$.

The second and third lines follow by independence and the identical distributions of the X_i.

Since

$$\mathbb{E}[e^{uX}] = \frac{\lambda}{\lambda - u}$$

for $u < \lambda$, it follows

$$\mathbb{E}[e^{uT_n}] = \left(\frac{\lambda}{\lambda - u}\right)^n$$

which is the moment generating function of a Gamma(n, λ) distribution (exercise: check this). □

Exercise 3.3.3. *An alternative proof is based upon considering the quantity $\mathbb{P}(T_n \leq t)$, and using the properties of the Poisson process. Give a different proof, using this idea.*

Solution: Note that

$$T_n \leq t \Leftrightarrow N_t \geq n.$$

Hence

$$F_{T_n}(t) = \mathbb{P}(T_n \leq t) = \mathbb{P}(N_t \geq n) = \sum_{k=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Now you only need to differentiate with respect to t to obtain the density f_{T_n}:

In particular, we have

$$\begin{aligned} f_{T_n}(t) = \frac{d}{dt} \sum_{k=n}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} &= \sum_{k=n}^{\infty} \left(e^{-\lambda t} (-\lambda) \frac{(\lambda t)^k}{k!} + e^{-\lambda t} \frac{(\lambda t)^{k-1} \lambda}{(k-1)!} \right) \\ &= e^{-\lambda t} \lambda \left(\sum_{k=n}^{\infty} (-1) \frac{(\lambda t)^k}{k!} + \frac{(\lambda t)^{k-1}}{(k-1)!} \right) \\ &= e^{-\lambda t} \lambda \left(-\sum_{k=n}^{\infty} \frac{(\lambda t)^k}{k!} + \sum_{k=n-1}^{\infty} \frac{(\lambda t)^k}{k!} \right) \end{aligned}$$

$$= e^{-\lambda t}\frac{\lambda^n t^{n-1}}{(n-1)!}, \qquad \text{for all } t \geq 0.$$

Example 3.3.4. *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.

3.3.3 Conditional distribution of the arrival times

Let T_n denote the time of the nth event. We now derive the conditional distribution of $T_1, \ldots, T_n$ given that $N_t = n$ (i.e. that n events have occurred on an interval $[0, t]$). Recall that the indicator of a set A is written $\mathbb{I}_A(x)$ and is one if $x \in A$ and 0 otherwise.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 (T_1) at which the first event occurred? Let $t_1 \leq t$. Then

$$\begin{aligned}
\mathbb{P}(T_1 \leq t_1 | N_t = 1) &= \frac{\mathbb{P}(T_1 \leq t_1, N_t = 1)}{\mathbb{P}(N_t = 1)} \\
&= \frac{\mathbb{P}(1 \text{ event in } [0, t_1], 0 \text{ events in } (t_1, t])}{\mathbb{P}(N_t = 1)} \\
&= \frac{\mathbb{P}(1 \text{ event in } [0, t_1])\mathbb{P}(0 \text{ events in } (t_1, t])}{\mathbb{P}(N_t = 1)} \\
&= \frac{e^{-\lambda t_1}\lambda t_1 e^{-\lambda(t-t_1)}}{e^{-\lambda t}\lambda t} = \frac{t_1}{t}.
\end{aligned}$$

Hence we see that the time of the first event, given that there has been one event in $[0, t]$, is uniformly distributed over $[0, t]$.

Theorem 3.3.5. *Let $\{N_t\}_{t \geq 0}$ be a Poisson process of rate $\lambda > 0$. Then for any $n > 0$, $0 < t$ the conditional density of $T_{1:n} = (T_1, \ldots, T_n)$, given $N_t = n$ is for $t_{1:n} = (t_1, \ldots, t_n)$:*

$$f_{T_{1:n}}(t_{1:n} | N_t = n) = \frac{n!}{t^n} \mathbb{I}_{S_{n,t}}(t_{1:n})$$

where

$$S_{n,t} = \{x_{1:n} | 0 < x_1 < \cdots < x_n \leq t\}.$$

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 $[0, t]$.

Note that

$$f_{T_{1:n}}(t_{1:n} | N_t = n) = \frac{n!}{t^n} \mathbb{I}_{S_{n,t}}(t_{1:n})$$

means that

$$f_{T_{1:n}}(t_{1:n} | N_t = n) = f(t_1, \ldots, t_n | N_t = n) = \begin{cases} \frac{n!}{t^n} & \text{if } 0 < t_1 < \cdots < t_n \leq t, \\ 0, & \text{otherwise.} \end{cases}$$

We do not prove the more general result (for $n > 1$) in the lecture. You can find a detailed proof for the more general case in Mikosch (2009, p. 24) (note that we have online access to the book via the library).

Remark 3.3.6. *The above theorem says that conditional on the fact that n events have occurred in $[0, t]$, the times $T_1, \ldots, T_n$ at which events occur when considered as unordered random variables are independently and uniformly distributed on $[0, t]$.*

Exercise 3.3.7. *Show that the expectation of the k^{th} value ($1 \leq k \leq n$) of n uniformly distributed order statistics*[1] *on $[0, t]$ is*

$$\frac{tk}{n+1}.$$

Exercise 3.3.8. *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 3.3.8. The problem asks us to calculate $\mathbb{E}[\sum_{i=1}^{N_t}[t - T_i]]$. Conditioning upon $N_t = n$ we have that

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{n}[t - T_i]\middle| N_t = n\right] &= nt - \sum_{i=1}^{n} \mathbb{E}[T_i | N_t = n] = nt - \sum_{i=1}^{n} \frac{it}{n+1} \\ &= nt - \frac{t}{n+1}\sum_{i=1}^{n} i = nt - \frac{t}{n+1}\frac{n(n+1)}{2} = nt - \frac{nt}{2} \end{aligned}$$

where we have used Theorem 3.3.5 and Exercise 3.3.7. Thus we conclude

$$\mathbb{E}\left[\sum_{i=1}^{N_t}[t - T_i]\right] = \frac{t}{2}\mathbb{E}[N_t]$$

that is,

$$\mathbb{E}\left[\sum_{i=1}^{N_t}[t - T_i]\right] = \frac{\lambda t^2}{2}.$$

□

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

3.4 Some extensions to Poisson processes

Lecture 15

3.4.1 Superposition

Suppose now, that we are given two independent Poisson processes $\{N_t^1\}$ and $\{N_t^2\}$ (of rates $\lambda_1 > 0$ and $\lambda_2 > 0$), and we define a new stochastic process

$$\widetilde{N}_t = N_t^1 + N_t^2.$$

Two natural questions arise: why? and what is this new process?

Proposition 3.4.1. *$\{\widetilde{N}_t\}$ is a Poisson process of rate $\lambda_1 + \lambda_2$.*

Proof. Exercise (part of a problem sheet). □

[1] Probably you have studied order statistics in M1S or M2S1/2

Example 3.4.2. *Suppose that blue cars arrive at a petrol station according to a Poisson process of rate λ_1 and red cars arrive, independently, according to a Poisson process of rate λ_2. What is the probability that N cars arrive in $[0,t]$? What is the probability that a red car arrives before a blue?*

Example 3.4.3. *Returning to Example 3.4.2, we clearly have that $\widetilde{N}_t \sim \mathcal{P}ois((\lambda_1+\lambda_2)t)$.*
To answer the second we want

$$\mathbb{P}(\text{red before blue}) = \mathbb{P}(X < Y)$$

where X is the time to the next red car ($X \sim \mathcal{E}xp(\lambda_2)$), and Y is the time to the next blue ($Y \sim \mathcal{E}xp(\lambda_1)$). Then it follows that

$$\begin{aligned}
\mathbb{P}(X<Y) &= \int_{-\infty}^{\infty} \mathbb{P}(X<Y|Y=y) f_Y(y)dy \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X<y|Y=y) f_Y(y)dy \quad (\textit{use independence}) \\
&= \int_{-\infty}^{\infty} \mathbb{P}(X<y) f_Y(y)dy \\
&= \int_{-\infty}^{\infty} F_X(y) f_Y(y)dy \\
&= \int_0^{\infty} \left[\int_0^y \lambda_2 e^{-\lambda_2 x} dx\right] \lambda_1 e^{-\lambda_1 y} dy.
\end{aligned}$$

Standard integration and manipulations lead us to

$$\mathbb{P}(X<Y) = \frac{\lambda_2}{\lambda_1+\lambda_2}.$$

3.4.2 Thinning

Now, suppose in the context of Example 3.4.2, 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.

Proposition 3.4.4. *$\{N_t^g\}$ is a Poisson process of rate λp.*

Proof. Exercise (part of a problem sheet). □

3.4.3 Non-homogeneous Poisson processes

Definition 3.4.5. *A non–homogeneous Poisson process with (non-negative)* intensity function *$(\lambda(t))$ is a stochastic process $N = \{N_t\}_{t\geq 0}$ with values in $\mathbb{N}_0$, which satisfies the following properties:*

1. $N_0 = 0$.
2. *N has independent increments.*
3. *'Single arrival' property: For $t \geq 0$, $\delta > 0$:*

$$\begin{aligned}\mathbb{P}(N_{t+\delta} - N_t = 1) &= \lambda(t)\delta + o(\delta),\\ \mathbb{P}(N_{t+\delta} - N_t \geq 2) &= o(\delta),\end{aligned}$$

that is, the rate/intensity is now dependent upon the time parameter.

From herein we assume that the intensity function, λ, is an integrable function.

Also note that the single arrival property implies that

$$\mathbb{P}(N_{t+\delta} - N_t = 0) = 1 - \lambda(t)\delta + o(\delta).$$

Let us derive the distribution of N_t. In order to do that, we consider the forward equations:

Define, for $n \in \mathbb{N}_0$, $t \geq 0$

$$p_n(t) = \mathbb{P}(N_t = n).$$

Let $n = 0$, then

$$\begin{aligned}p_0(t+\delta) &= \mathbb{P}(N_{t+\delta} = 0) = \mathbb{P}(\text{no event in}[0, t+\delta])\\ &= \mathbb{P}(\text{no event in}[0,t] \text{ and no event in}(t, t+\delta])\\ &= \mathbb{P}\Big(\text{no event in}(t, t+\delta]\Big)\mathbb{P}\Big(\text{no event in}[0,t]\Big),\end{aligned}$$

where we applied the independent increments property, then

$$p_0(t+\delta) = [1 - \lambda(t)\delta + o(\delta)]p_0(t).$$

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 $p_0(0) = \mathbb{P}(N_0 = 0) = 1$.

Hence, we obtain

$$p_0(t) = \exp\left(-\int_0^t \lambda(s)ds\right).$$

For $n \geq 1$ (i.e. $n \in \mathbb{N}$) we have

$$\begin{aligned} p_n(t+\delta) &= \mathbb{P}(N_{t+\delta} = n) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(N_{t+\delta} = n | N_t = k)\mathbb{P}(N_t = k) \quad \text{(Law of total probability)} \\ &= \sum_{k=0}^{\infty} \mathbb{P}((n-k) \text{ events in } (t, t+\delta])\mathbb{P}(N_t = k) \\ &= \mathbb{P}(1 \text{ event in } (t, t+\delta])\mathbb{P}(N_t = n-1) \\ &\quad + \mathbb{P}(0 \text{ events in } (t, t+\delta])\mathbb{P}(N_t = n) + o(\delta) \\ &= p_{n-1}(t)\lambda(t)\delta + p_n(t)(1-\lambda(t)\delta) + o(\delta) \\ &= p_n(t)(1-\lambda(t)\delta) + p_{n-1}(t)\lambda(t)\delta + o(\delta). \end{aligned}$$

Re-arranging and letting $\delta \downarrow 0$ we have

$$\frac{dp_n(t)}{dt} = -\lambda(t)p_n(t) + \lambda(t)p_{n-1}(t).$$

The probabilities can then be obtained by induction. Let $n = 1$, then we have the ODE

$$\frac{dp_1(t)}{dt} + \lambda(t)p_1(t) = \lambda(t)\exp\left\{ -\int_0^t \lambda(s)ds \right\}.$$

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 our case, the integrating factor is

$$M(t) = \exp\left\{ \int_0^t \lambda(u)du \right\}.$$

Since $p_1(0) = \mathbb{P}(N_0 = 1) = 0$, the solution is

$$p_1(t) = \left[\int_0^t \lambda(s)ds \right] \exp\left\{ -\int_0^t \lambda(s)ds \right\}.$$

Now letting $n = 2$ we obtain (in the same manner)

$$p_2(t) = \int_0^t \lambda(s) \int_0^s \lambda(u) du ds \ \exp\Big\{ - \int_0^t \lambda(s) ds \Big\}.$$

By interchanging the order of integration we have

$$\int_0^t \lambda(s) \int_0^s \lambda(u) du ds = \int_0^t \lambda(u) \int_u^t \lambda(s) ds du \left(= \int_0^t \lambda(s) \int_s^t \lambda(u) du ds \right).$$

Thus it follows that

$$\begin{aligned} 2\int_0^t \lambda(s) \int_0^s \lambda(u) du ds &= \int_0^t \lambda(s) \int_0^s \lambda(u) du ds + \int_0^t \lambda(s) \int_s^t \lambda(u) du ds \\ &= \int_0^t \lambda(s) \Big[\int_0^s \lambda(u) du + \int_s^t \lambda(u) du \Big] ds \\ &= \int_0^t \lambda(s) \int_0^t \lambda(u) du ds \\ &= \Big[\int_0^t \lambda(s) ds \Big]^2. \end{aligned}$$

Thus, for $n = 2$

$$p_2(t) = \Big[\int_0^t \lambda(s) ds \Big]^2 \frac{1}{2} \exp\Big\{ - \int_0^t \lambda(s) ds \Big\}.$$

Then using an inductive proof (exercise, complete the proof) it follows that

$$p_n(t) = \Big[\int_0^t \lambda(s) ds \Big]^n \frac{1}{n!} \exp\Big\{ - \int_0^t \lambda(s) ds \Big\}.$$

Hence N_t has Poisson distribution with parameter $m(t) := \int_0^t \lambda(s) ds$.

Note that in order to complete the proof by induction you can argue as follows: Assume the induction hypothesis holds for n. 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

$$p_n(t) = \Big[\int_0^t \lambda(s) ds \Big]^n \frac{1}{n!} \exp\Big\{ - \int_0^t \lambda(s) ds \Big\},$$

and hence

$$\frac{dp_{n+1}(t)}{dt} + \lambda(t) p_{n+1}(t) = \lambda(t) \Big[\int_0^t \lambda(s) ds \Big]^n \frac{1}{n!} \exp\Big\{ - \int_0^t \lambda(s) ds \Big\} =: g(t).$$

Using the integrating factor again, we have

$$M(x) = \exp\left(\int_0^x \lambda(u) du \right).$$

Then

$$p_{n+1}(t) = \left(\int_0^t g(u) M(u) du + C \right) M(t)^{-1}.$$

Note that the initial condition is $p_{n+1}(0) = \mathbb{P}(N_0 = n+1) = 0$, hence $C = 0$. Then

$$\begin{aligned} p_{n+1}(t) &= \left(\int_0^t g(u) M(u) du \right) M(t)^{-1} \\ &= \int_0^t \lambda(u) \left(\int_0^u \lambda(s) ds \right)^n \frac{1}{n!} du \; \exp\left(- \int_0^t \lambda(s) ds \right). \end{aligned}$$

I.e. we need to show that

$$\int_0^t \lambda(u) \left(\int_0^u \lambda(s) ds \right)^n \frac{1}{n!} du = \frac{1}{(n+1)!} \left(\int_0^t \lambda(s) ds \right)^{n+1}.$$

This is an application of the chain rule: Define

$$f(u) := \frac{1}{n!} u^n, \qquad g(u) := \int_0^u \lambda(s) ds.$$

Note that $g'(u) = \lambda(u)$. Let

$$F(t) := \int_0^t f(u) du = \frac{1}{(n+1)!} t^{n+1}.$$

Then

$$\int_0^t f(g(u)) g'(u) du = F(g(u))|_0^t = F(g(t)) - F(g(0)).$$

Altogether we have

$$\begin{aligned} \int_0^t f(g(u)) g'(u) du &= \int_0^t \lambda(u) \left(\int_0^u \lambda(s) ds \right)^n \frac{1}{n!} du \\ &= F(g(t)) - F(g(0)) = \frac{1}{(n+1)!} \left(\int_0^t \lambda(s) ds \right)^{n+1}, \end{aligned}$$

which concludes the proof.

Exercise 3.4.6. *Derive the distribution of the increment $N_t - N_s$, for $0 \le s < t$. Does a non–homogeneous Poisson process have stationary increments?*

Solution to Exercise 3.4.6. Let $0 \le s < t$. We have shown that $N_t \sim Poi(m(t))$. Also, $N_s \sim Poi(m(s))$. 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 $u > 0$

$$\mathbb{E}\left(\exp(-uN_t)\right) = \exp(m(t)(\exp(-u) - 1)).$$

Also,

$$\mathbb{E}\left(\exp(-uN_s)\right) = \exp(m(s)(\exp(-u) - 1)).$$

Using the independence increment property, we get

$$\mathbb{E}\left(\exp(-uN_t)\right) = \mathbb{E}\left(\exp(-u(N_t - N_s + N_s - N_0))\right) = \mathbb{E}\left(\exp(-u(N_t - N_s))\right) \mathbb{E}\left(\exp(-u(N_s - N_0))\right)$$

Hence

$$\mathbb{E}\left(\exp(-uN_t)\right) \left[\mathbb{E}\left(\exp(-u(N_s - N_0))\right)\right]^{-1} = \mathbb{E}\left(\exp(-u(N_t - N_s))\right)$$

Now we only have to plug in the results for the Laplace transform of N_t, and N_s and we get

$$\mathbb{E}\left(\exp(-u(N_t - N_s))\right) = \exp(m(t)(\exp(-u) - 1)) \exp(-m(s)(\exp(-u) - 1))$$

$$= \exp((m(t) - m(s))(\exp(-u) - 1)),$$

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 3.4.7. *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)).*

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

3.4.4 Compound Poisson processes

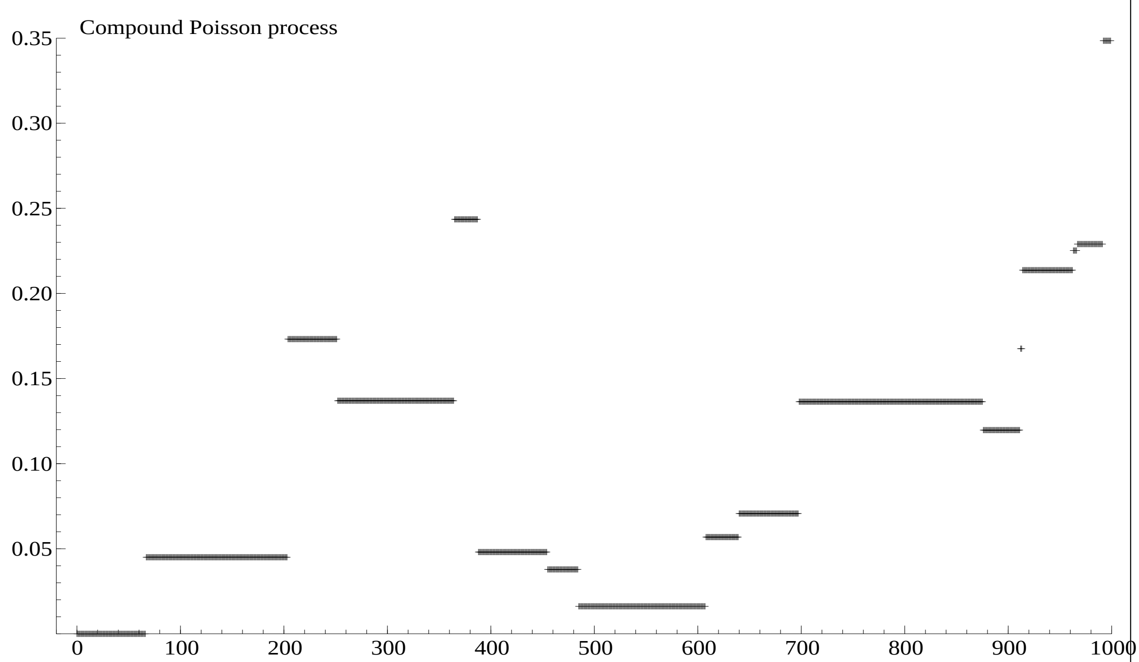
Lecture 16

An interesting extension of the Poisson process (which is also a Lévy process) is called the compound Poisson process.

Definition 3.4.8. *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 $\{N_t\}$. Then*

$$X_t = \sum_{i=1}^{N_t} Y_i$$

is a compound Poisson process.



A realisation of a compound Poisson process with normally distributed jump size.

Lemma 3.4.9. *Let $\{X_t\}_{t\geq 0}$ denote a compound Poisson process as defined in Definition 3.4.8. Then for $t \geq 0$, $\mathbb{E}(X_t) = \lambda t \mathbb{E}(Y_i)$ and $\mathbb{V}ar(X_t) = \lambda t \mathbb{E}(Y_i^2)$.*

Note that for $n \in \{0, 1, 2, \dots\}$

$$\mathbb{E}(X_t|N_t = n) = \mathbb{E}\left(\sum_{i=1}^{N_t} Y_i|N_t = n\right) = \mathbb{E}\left(\sum_{i=1}^{n} Y_i|N_t = n\right) = \mathbb{E}\left(\sum_{i=1}^{n} Y_i\right) = n\mathbb{E}(Y_i),$$

where we used the fact that (Y_i) and N are independent and further that the (Y_i) are i.i.d.. Hence, we get

$$\mathbb{E}(X_t|N_t) = N_t\mathbb{E}(Y_i).$$

Using the properties of the conditional expectation, we get

$$\mathbb{E}(X_t) = \mathbb{E}(\mathbb{E}(X_t|N_t)) = \mathbb{E}(N_t\mathbb{E}(Y_i)) = \mathbb{E}(Y_i)\mathbb{E}(N_t) = \lambda t\mathbb{E}(Y_i).$$

For the variance, we use

$$\mathbb{V}ar(X_t) = \mathbb{V}ar\left(\mathbb{E}\left(X_t|\, N_t\right)\right) + \mathbb{E}\left(\mathbb{V}ar\left(X_t|\, N_t\right)\right). \tag{3.4.1}$$

(*If formula* (3.4.1) *is new to you, then prove it at home as an exercise!*
Let us compute the conditional variance first:

$$\begin{aligned}\mathbb{V}ar(X_t|N_t = n) &= \mathbb{V}ar\left(\sum_{i=1}^{N_t} Y_i|N_t = n\right) = \mathbb{V}ar\left(\sum_{i=1}^{n} Y_i|N_t = n\right)\\ &= \mathbb{V}ar\left(\sum_{i=1}^{n} Y_i\right) = n\mathbb{V}ar(Y_i),\end{aligned}$$

where we used the fact that (Y_i) and N are independent and further that the (Y_i) are i.i.d.. Hence

$$\mathbb{V}ar(X_t|N_t) = N_t\mathbb{V}ar(Y_i).$$

Then we have

$$\mathbb{E}(\mathbb{V}ar(X_t|N_t)) = \mathbb{E}(N_t\mathbb{V}ar(Y_i)) = \mathbb{E}(N_t)\mathbb{V}ar(Y_i) = \lambda t\mathbb{V}ar(Y_i),$$

and

$$\mathbb{V}ar(\mathbb{E}(X_t|N_t)) = \mathbb{V}ar(N_t\mathbb{E}(Y_i)) = \mathbb{V}ar(N_t)(\mathbb{E}(Y_i))^2 = \lambda t(\mathbb{E}(Y_i))^2.$$

Using formula (3.4.1), we get

$$\mathbb{V}ar(X_t) = \lambda t\mathbb{V}ar(Y_i) + \lambda t(\mathbb{E}(Y_i))^2 = \lambda t\mathbb{E}(Y_i^2).$$

Example 3.4.10.

Suppose that asteroids fall to the earth according to a Poisson process of rate $\lambda \in \mathbb{R}_+$. In addition, and independent of the arrival of the asteroid (and other asteroids), the asteroid will cause a human fatality with probability $p \in (0, 1)$. Find the probability generating function (pgf) of the number of human fatalities at time t (X_t), and, using the pgf, show that the expected number is λtp. From the problem, a convenient model is

$$X_t = \sum_{i=1}^{N_t} Y_i$$

where each Y_i is a Bernoulli random variable.

Then we have, for $t \in [0, T]$

$$\begin{aligned} G_{X_t}(s) &= \mathbb{E}[s^{X_t}] \\ &= \mathbb{E}[s^{\sum_{i=1}^{N_t} Y_i}] \\ &= \mathbb{E}\left[\mathbb{E}\left(s^{\sum_{i=1}^{N_t} Y_i}\Big| N_t\right)\right], \end{aligned}$$

by the property of the conditional expectation. Note that due to the independence of the (Y_i) and N, we have

$$\begin{aligned} \mathbb{E}\left(s^{\sum_{i=1}^{N_t} Y_i}\Big| N_t = n\right) &= \mathbb{E}\left(s^{\sum_{i=1}^{n} Y_i}\Big| N_t = n\right) = \mathbb{E}\left(s^{\sum_{i=1}^{n} Y_i}\right) \\ &= \prod_{i=1}^{n} \mathbb{E}(s^{Y_i}) = \left(\mathbb{E}(s^{Y_i})\right)^n, \end{aligned}$$

since the (Y_i) are i.i.d.. Hence

$$G_{X_t}(s) = \mathbb{E}\Bigg[\mathbb{E}[s^{Y_i}]^{N_t}\Bigg] = \mathbb{E}[(1 - p + ps)^{N_t}] = \mathbb{E}[z^{N_t}],$$

where $z = (1 - p + ps) = 1 + p(s - 1)$.

Recall that the pgf of a Poisson random variable (of rate λ) is given by

$$G_n(s) = \exp\{\lambda[s - 1]\}.$$

Here we have that $N_t \sim Poi(\lambda t)$; it follows that

$$G_{X_t}(s) = \exp\{\lambda t[z - 1]\} = \exp\{\lambda t p[s - 1]\}.$$

Differentiating this function w.r.t s and setting $s = 1$ completes the exercise.

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.

3.5 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 3.5.1.

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* $\mu = \mathbb{E}(Y_1)$*, and variance* $\sigma^2 = \mathbb{V}ar(Y_1) \leq \infty$*.*
2. *The* claim times *occur at the random instants of time*
$$0 < T_1 < T_2 < \cdots \ a.s..$$
3. *The* claim arrival process *is denoted by*
$$N_t = \sup\{n \geq 1 : T_n \leq t\}, \quad t \geq 0,$$
which is the number of claims in the interval $[0, t]$*. (Note that* $\sup\emptyset := 0$*).*
4. *The* inter-arrival times *are denoted by*
$$X_1 = T_1,\ X_k = T_k - T_{k-1},\ k = 2, 3, \ldots,$$
and are exponentially distributed with parameter λ*.*
5. *The sequences* (Y_k) *and* (X_k) *are independent of each other.*

Note that the process (N_t) is a (homogeneous) Poisson process with intensity $\lambda > 0$.

Definition 3.5.2. *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 can derive the *total claim amount distribution*.

Theorem 3.5.3. *The total claim amount distribution is given by*
$$\mathbb{P}(S_t \leq x) \quad = \quad \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathbb{P}\left(\sum_{i=1}^{n} Y_i \leq x\right), \qquad x \geq 0,\ t \geq 0.$$

Proof.
$$\begin{aligned}
& \mathbb{P}(S_t \leq x) \\
= \ & \mathbb{P}(S_t \leq x | N_t > 0)\mathbb{P}(N_t > 0) + \mathbb{P}(S_t \leq x | N_t = 0)\mathbb{P}(N_t = 0) \\
= \ & \mathbb{P}(\sum_{i=1}^{N_t} Y_i \leq x | N_t > 0)\mathbb{P}(N_t > 0) + \mathbb{P}(0 \leq x | N_t = 0)\mathbb{P}(N_t = 0) \\
= \ & \sum_{n=1}^{\infty} \mathbb{P}\left(\sum_{i=1}^{n} Y_i \leq x | N_t = n\right)\mathbb{P}(N_t = n) + \mathbb{P}(N_t = 0) \\
= \ & \sum_{n=0}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathbb{P}\left(\sum_{i=1}^{n} Y_i \leq x\right).
\end{aligned}$$

□

A slightly shorter proof is given as follows (where we work with the definition that an empty sum is equal to 0):

$$\mathbb{P}(S_t \leq x) = \sum_{n=0}^{\infty} \mathbb{P}(S_t \leq x, N_t = n) = \sum_{n=0}^{\infty} \mathbb{P}(\sum_{i=1}^{n} Y_i \leq x, N_t = n) = \sum_{n=0}^{\infty} \mathbb{P}(\sum_{i=1}^{n} Y_i \leq x)\mathbb{P}(N_t = n)$$
$$= \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \mathbb{P}\left(\sum_{i=1}^{n} Y_i \leq x\right)$$

Definition 3.5.4. *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 3.5.5. *Draw a sample path of the risk process U!*

Now we can define the ruin probability.

Definition 3.5.6. 1. *The* ruin probability in finite time *is given by*

$$\psi(u, T) = \mathbb{P}(U_t < 0 \, \textit{for some } t \leq T), \; 0 < T < \infty, u \geq 0.$$

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 3.5.7.

$$\mathbb{E}(U_t) = u + ct - \lambda t \mu.$$

Proof. Exercise! □

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

$$\lim_{t \to \infty} \frac{\mathbb{E}(U_t)}{t} = c - \lambda \mu > 0.$$

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.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

Chapter 4

Continuous–time Markov chains

Lecture 17

We now look at Markov chains in continuous time, i.e. $\mathcal{T} = [0, \infty)$, but on a discrete state–space $E \subset \mathbb{Z}$ with $K = \text{Card}(E)$. As in the discrete case, we could in fact choose any 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 rigourously. As a result, we cannot be, completely, mathematically accurate. None-the-less we proceed with such processes.

This chapter is based on Grimmett & Stirzaker (2001*b*, Chapters 6.9 & 6.11).

4.1 Some definitions

We begin with a basic definition.

Definition 4.1.1. *A continuous-time process $\{X_t\}_{t\in[0,\infty)}$ satisfies the **Markov property** if*

$$\mathbb{P}(X_{t_n} = j | X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}) = \mathbb{P}(X_{t_n} = j | X_{t_{n-1}} = i_{n-1})$$

for all $j, i_1, \dots, i_{n-1} \in E$ and for any sequence $0 \leq t_1 < \cdots < t_n < \infty$ of times (with $n > 1$).

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.

We will see, in the next chapter, that much of continuous time stochastic processes are based upon this notion.

Exercise 4.1.2. *Show that a Poisson process is a Markov chain in continuous-time.*

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 4.1.3. *The* ***transition probability*** $p_{ij}(s,t)$ *is, for* $s \leq t$, $i, j \in E$

$$p_{ij}(s,t) = \mathbb{P}(X_t = j | X_s = i)$$

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 4.1.4. *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:* $\mathbf{P}_{s+t} = \mathbf{P}_s \mathbf{P}_t$.

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. Let $\mathbf{1}$ denote the column vector of ones. Then (using the law of total probability)

$$(\mathbf{P}_t \mathbf{1})_i = \sum_{j \in E} p_{ij}(t) = \sum_{j \in E} \frac{\mathbb{P}(X_t = j, X_0 = i)}{\mathbb{P}(X_0 = i)} = \frac{\mathbb{P}(X_0 = i)}{\mathbb{P}(X_0 = i)} = 1.$$

3. Using the law of total probability and the Markov property, we have

$$\begin{aligned}
p_{ij}(s+t) &= \mathbb{P}(X_{s+t} = j | X_0 = i) = \frac{\mathbb{P}(X_{s+t} = j, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{k \in E} \frac{\mathbb{P}(X_{s+t} = j, X_s = k, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{k \in E} \frac{\mathbb{P}(X_{s+t} = j, X_s = k, X_0 = i)}{\mathbb{P}(X_s = k, X_0 = i)} \frac{\mathbb{P}(X_s = k, X_0 = i)}{\mathbb{P}(X_0 = i)} \\
&= \sum_{k \in E} \mathbb{P}(X_{s+t} = j | X_s = k, X_0 = i) \mathbb{P}(X_s = k | X_0 = i) \\
&= \sum_{k \in E} \mathbb{P}(X_{s+t} = j | X_s = k) \mathbb{P}(X_s = k | X_0 = i) \\
&= \sum_{k \in E} p_{ik}(s) p_{kj}(t).
\end{aligned}$$

□

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 4.1.5. *The semigroup $\{\mathbf{P}_t\}$ is called **standard** if*

$$\lim_{t \downarrow 0} \mathbf{P}_t = \mathbf{I},$$

where $\mathbf{I} = \mathbf{I}_{K \times K}$ denotes the $K \times K$-dimensional identity matrix.

That is to say, that a semigroup is standard if and only if its elements $p_{ij}(t)$ are continuous functions in t. (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.

4.2 The generator

We are now in a position to start describing the generator of the process.

Suppose that at time t, $X_t = i$. What can happen over a small time interval $(t, t + \delta)$?

- Nothing happens with probability $p_{ii}(\delta) + o(\delta)$. (Here the error term allows for the possibility that the chain leaves i and returns to i in the small time interval).
- The chain makes exactly one transition and moves from state i to a new state j with probability $p_{ij}(\delta) + o(\delta)$
- We assume that the probability of two or more transitions is $o(\delta)$;

The above properties can be proved, but we do not do so.

It can also be shown that there exist constants $\{g_{ij} : i, j \in E\}$ such that

$$p_{ij}(\delta) \approx g_{ij}\delta \quad i \neq j,$$

i.e. $p_{ij}(\delta)$ is approximately linear in δ for small δ, and

$$p_{ii}(\delta) \approx 1 + g_{ii}\delta.$$

It is clear that $g_{ij} \geq 0$ if $i \neq j$ and $g_{ii} \leq 0$.
The Matrix $\mathbf{G} = (g_{ij})$ is the **generator**. It thus follows that

- the probability of no transition in $(t, t + \delta)$ is

$$1 + g_{ii}\delta + o(\delta)$$

- and the probability of a transition from i to j is

$$g_{ij}\delta + o(\delta).$$

We hope that $\sum_{j\in E} p_{ij}(t) = 1$, so

$$1 = \sum_{j\in E} p_{ij}(\delta) \approx 1 + \delta \sum_{j\in E} g_{ij}$$

yielding

$$\sum_{j\in E} g_{ij} = 0 \tag{4.2.1}$$

for every $i \in E$. Note that this does not always occur, there are some chains for which (4.2.1) does not hold.

Remark 4.2.1. *Just a little note in case you decide to study various textbooks on continuous–time Markov chains: Note that the generator is often called $\mathbf{Q}$ (rather than $\mathbf{G}$) in many textbooks. Then the matrix is just called the Q-matrix rather than the generator. But it is the same object!*

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

4.3 The forward and backward equations

Lecture 18

We now return to the forward and backward equations. Typically the generator is defined as

$$\mathbf{G} := \lim_{\delta\downarrow 0} \frac{1}{\delta}[\mathbf{P}_\delta - \mathbf{I}]$$

that is, $\mathbf{P}_t$ is differentiable at $t = 0$. It is possible to find $\mathbf{G}$, given $\{\mathbf{P}_t\}$, but the converse is also usually true. Let $X_0 = i$, and condition $X_{t+\delta}$ on X_t:

$$\begin{aligned} p_{ij}(t+\delta) &= \sum_{l\in E} p_{il}(t)p_{lj}(\delta) \qquad \text{by Chapman-Kolmogorov} \\ &\approx p_{ij}(t)[1 + g_{jj}\delta] + \sum_{l\in E: l\neq j} p_{il}(t)g_{lj}\delta \\ &= p_{ij}(t) + \sum_{l\in E} p_{il}(t)g_{lj}\delta \end{aligned}$$

then rearranging

$$\frac{1}{\delta}[p_{ij}(t+\delta) - p_{ij}(t)] \approx \sum_{l\in E} p_{il}(t)g_{lj} = (\mathbf{P}_t\mathbf{G})_{ij}.$$

Letting $\delta \downarrow 0$ we yield the **forward equations**

$$p'_{ij}(t) = \sum_{l\in E} p_{il}(t)g_{lj}, \qquad \text{for all } i, j \in E,$$

i.e.

$$\mathbf{P}'_t = \mathbf{P}_t\mathbf{G}.$$

(Note that we did not mention all technical assumptions needed for the forward equations to hold!)

We can also derive the *backward equations*, using much the same argument, by conditioning $X_{t+\delta}$ on X_δ:

$$\begin{aligned} p_{ij}(t+\delta) &= \sum_l p_{il}(\delta)p_{lj}(t) \qquad \text{by Chapman-Kolmogorov} \\ &\approx (1+g_{ii}\delta)p_{ij}(t) + \sum_{l\in E, l\neq i} g_{il}\delta p_{lj}(t) \\ &= p_{ij}(t) + \sum_{l\in E} g_{il}\delta p_{lj}(t). \end{aligned}$$

Hence

$$\frac{1}{\delta}(p_{ij}(t+\delta) - p_{ij}(t)) \approx \sum_{l\in E} g_{il}p_{lj}(t).$$

yielding the **backward equations**

$$p'_{ij}(t) = \sum_l g_{il}p_{lj}(t), \qquad \text{for all } i,j \in E,$$

$$\text{i.e.} \qquad \mathbf{P}'_t = \mathbf{G}\mathbf{P}_t.$$

(Note that we did not mention all technical assumptions needed for the backward equations to hold!)

(Note that we will derive backward equations for birth processes (a particular example of a continuous-time Markov chain) in the next lecture. Then we will discuss in detail the difference in deriving forward versus backward equations.)

As a result, we are able to express the semigroup in terms of the generator. It is often the case that the differential equations with boundary condition $\mathbf{P}_0 = \mathbf{I}$ can be solved uniquely, with a solution of the form

$$\mathbf{P}_t = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n. \tag{4.3.1}$$

Note that we have powers of matrices here and $\mathbf{G}^0 = \mathbf{I}$.
We can express (4.3.1) 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 $\mathbf{A}$.

4.4 Holding times and other properties

We now look at some properties of continuous time Markov chains.

We begin by looking at holding times.
Suppose that for $i \in E$ we have $X_t = i$. For $t \geq 0$, let

$$H_i = \inf\{s \geq 0 : X_{t+s} \neq i | X_t = i\} \overset{\text{law}}{=} \inf\{s \geq 0 : X_s \neq i | X_0 = i\}$$

be the further time until the chain changes it state. H_i is called a *holding time*. Note that it does not depend on t since we work under the homogeneity assumption. Then we have the following claim (subject to regularity conditions):

H_i is exponentially distributed with paramter $-g_{ii}$.

Remark 4.4.1. *The proof of the above claim is based on the lack of memory property of the exponential distribution. Note that the exponential distribution is the* only *continuous distribution which has the lack of memory property.*

For discrete random variables one can show that the geometric distribution has the lack of memory property.

Exercise 4.4.2. *Prove that if a continuous distribution has the lack of memory property, then it has to be the exponential distribution. (If you get stuck, you can find the proof in Grimmett & Stirzaker (2001*a*, p. 210–211).)*

Proof.

By time homogeneity assume without loss of generality that $X_0 = i$. Note that for $x, y \geq 0$, we have

$$\{H_i > x\} = \{X_t = i, \text{ for } 0 \leq t \leq x\},$$
$$\{H_i > x + y\} = \{X_t = i, \text{ for } 0 \leq t \leq x + y\}.$$

Then

$$\begin{aligned}
&\mathbb{P}(H_i > x + y | H_i > x) \\
&= \mathbb{P}(X_t = i, \text{ for } 0 \leq t \leq x + y | X_t = i, \text{ for } 0 \leq t \leq x) \\
&= \frac{\mathbb{P}(X_t = i, \text{ for } 0 \leq t \leq x, X_t = i, \text{ for } x < t \leq x + y)}{\mathbb{P}(X_t = i, \text{ for } 0 \leq t \leq x)} \\
&= \mathbb{P}(X_t = i, \text{ for } x < t \leq x + y | X_t = i, \text{ for } 0 \leq t \leq x) \\
&\overset{\text{Markov}}{=} \mathbb{P}(X_t = i, \text{ for } x < t \leq x + y | X_x = i) \\
&\overset{\text{time-hom.}}{=} \mathbb{P}(X_t = i, \text{ for } 0 < t \leq y | X_0 = i) = \mathbb{P}(H_i > y).
\end{aligned}$$

Hence H_i has a continuous distribution satisfying the lack of memory property, hence it follows the exponential distribution. I.e. for $x \geq 0$ we have the following cdf of H_i: $F_{H_i}(x) = 1 - e^{-\lambda_i x}$, where one can show that $\lambda_i = -h_i(1) = -g_{ii}$ where $h_i(x) = \log(\mathbb{P}(H_i > x))$. [We skip the details of deriving the rate parameter.]

□

Thus, if $X_t = i$ it stays there for an exponentially distributed time, and then moves to some other state j. We have our next claim (subject to regularity conditions):

The probability that the chain jumps to $j \neq i$ is $-g_{ij}/g_{ii}$.

Proof. We have for $\delta > 0$ and $i \neq j$ that

$$\begin{aligned}
&\mathbb{P}(X_{t+\delta} = j | X_t = i, X_{t+\delta} \neq i) \\
&= \frac{\mathbb{P}(X_{t+\delta} = j, X_t = i, X_{t+\delta} \neq i)}{\mathbb{P}(X_t = i, X_{t+\delta} \neq i)} \\
&= \frac{\mathbb{P}(X_{t+\delta} = j, X_t = i)\mathbb{P}(X_t = i)}{\mathbb{P}(X_t = i, X_{t+\delta} \neq i)\mathbb{P}(X_t = i)} \\
&= \frac{\mathbb{P}(X_{t+\delta} = j | X_t = i)}{\mathbb{P}(X_{t+\delta} \neq i | X_t = i)} = \frac{p_{ij}(\delta)}{1 - p_{ii}(\delta)} \to -\frac{g_{ij}}{g_{ii}} \qquad \text{as } \delta \downarrow 0.
\end{aligned}$$

□

Example 4.4.3. *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.

We also extend the notions of irreducibility, stationarity and the limiting distribution. Note that it can be proved that either $p_{ij}(t) = 0$ (for all t) or $p_{ij}(t) > 0$ for all $t > 0$.

Definition 4.4.4. *The chain is **irreducible** if for any $i, j \in E$ we have $p_{ij}(t) > 0$ for some t.*

Definition 4.4.5. *The vector $\boldsymbol{\pi}$ is the **stationary distribution** if $\pi_i \geq 0$, $\sum_{j \in E} \pi_j = 1$ and $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ for all $t \geq 0$.*

As in discrete time, if $\boldsymbol{\nu}^{(t)}$ is the marginal distribution of X_t, then we have

$$\boldsymbol{\nu}^{(t)} = \boldsymbol{\nu}^{(0)}\mathbf{P}_t.$$

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. We claim that (subject to regularity conditions)

$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ *for all t if and only if* $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$.

Proof. A <u>sketch</u> proof. Using (4.3.1) and $\mathbf{G}^0 = \mathbf{I}$:

$$\begin{aligned} \boldsymbol{\pi}\mathbf{G} = \mathbf{0} \quad &\Leftrightarrow \quad \boldsymbol{\pi}\mathbf{G}^n = \mathbf{0} \quad \text{for all } n \geq 1 \\ &\Leftrightarrow \quad \sum_{n=1}^{\infty} \frac{t^n}{n!}\boldsymbol{\pi}\mathbf{G}^n = \mathbf{0} \quad \text{for all } t \\ &\Leftrightarrow \quad \boldsymbol{\pi}\sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n = \boldsymbol{\pi} \quad \text{for all } t \\ &\Leftrightarrow \quad \boldsymbol{\pi}\mathbf{P}_t = \boldsymbol{\pi} \quad \text{for all } t. \end{aligned}$$

□

This helps us to find stationary distributions, given their existence.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 19

We finish the section with the ergodic theorem.

Theorem 4.4.6. *Let X be an irreducible Markov chain with a standard semigroup $\{P_t\}$ 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 $p_{ij}(t) \to 0$ as $t \to \infty$, 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

$$p_{ij}(n\delta) = \mathbb{P}(Y_n = j | Y_0 = i) \to \pi_j^{(\delta)}, \qquad \text{as } n \to \infty,$$

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 $\{n\delta_1 : n \geq 0\}, \{n\delta_2 : n \geq 0\}$ 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 $\{n\delta : n \geq 0\}$ of times with rational δ. Next use the continuity of the transition semigroup to fill in the gaps! □

4.5 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 4.5.1. *A* birth process *with intensities $\lambda_0, \lambda_1, \dots$ is a stochastic process $\{N_t\}_{t \geq 0}$ with values in $\mathbb{N}_0$ such that*

1. *$\{N_t\}_{t \geq 0}$ is a Markov chain on $E = \mathbb{N}_0$.*
2. *There is a 'single arrival', i.e. the infinitesimal transition probabilities are for $t \geq 0$, $\delta > 0$, $n, m \in \mathbb{N}_0$:*
$$\mathbb{P}(N_{t+\delta} = n + m | N_t = n) = \begin{cases} 1 - \lambda_n \delta + o(\delta) & \text{if } m = 0 \\ \lambda_n \delta + o(\delta) & \text{if } m = 1 \\ o(\delta) & \text{o/w} \end{cases}$$
3. *The birth rates satisfy $\lambda_i \geq 0$ for all $i \in E$.*

Note that the Markov property implies that a birth process has *conditionally independent increments*, i.e. 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. This means that for $k, l, x(r) \in \{0, 1, 2, \dots\}$ for $0 \leq r < s$, we have

$$\mathbb{P}(N_t - N_s = k | N_s = l, N_r = x(r) \text{ for } 0 \leq r < s) = \mathbb{P}(N_t - N_s = k | N_s = l).$$

As in the case of a Poisson process, the birth process is also an increasing process meaning that whenever $0 \leq s \leq t$, then $N_s \leq N_t$.

An obvious example of a birth process is the Poisson process; $\lambda_n = \lambda$ for any $n \geq 0$.

Example 4.5.2. *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; no deaths occur.*

Example 4.5.3. *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 4.5.4. *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 $p_{ii}(\delta) = 1 - \lambda_i\delta + o(\delta)$. We have*

$$g_{ii} = (p_{ii}(\delta) - 1)/\delta = -\lambda_i, \quad g_{i,i+1} = p_{i(i+1)}(\delta)/\delta = \lambda_i$$

and otherwise $g_{ij} = 0$, if $i > j$ or $j > i + 1$. That is

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

4.5.1 The forward and backward equations

We will now derive the forward and backward equations for birth processes. I will explain, at the end, why it is useful.

Let $\{N_t\}$ be a birth process with positive intensities $\lambda_0, \ldots$. Define the transition probabilities

$$p_{ij}(t) = \mathbb{P}(N_{t+s} = j | N_s = i) = \mathbb{P}(N_t = j | N_0 = i) \tag{4.5.1}$$

(note that this is stationary w.r.t t).

Exercise 4.5.5. *Compute the probabilities*

- $\mathbb{P}(N_{t+s} = j | N_s = i)$ *and*
- $\mathbb{P}(N_t = j | N_0 = i)$,

based on the 'single arrival property' assuming that $\{N_s = i\}$ and $\{N_0 = i\}$, respectively, are true and that t is "small"!

Now we can obtain the **forward equations** for a birth process ($i, j \in E, j > i$):

$$p_{ij}(t + \delta) = p_{ij}(t)[1 - \lambda_j\delta] + p_{i,j-1}(t)\lambda_{j-1}\delta + o(\delta)$$

with $\lambda_{-1} = 0$.

That is, we have

$$\begin{aligned} p_{ij}(t+\delta) &= \mathbb{P}(N_{t+\delta}=j|N_0=i) \\ &= \sum_{l\in E}\mathbb{P}(N_{t+\delta}=j|N_t=l,N_0=i)\mathbb{P}(N_t=l|N_0=i) \quad \text{(by Law of total probability)} \\ &= \sum_{l\in E}\mathbb{P}(N_{t+\delta}=j|N_t=l)\mathbb{P}(N_t=l|N_0=i) \quad \text{(by Markov property)} \\ &= \sum_{l\in E}\mathbb{P}(N_{t+\delta}=j|N_t=l)p_{il}(t) \end{aligned}$$

From the definition of the birth process, we get

$$\mathbb{P}(N_{t+\delta}=l+(j-l)|N_t=l)=\begin{cases} 1-\lambda_j\delta+o(\delta) & \textit{if } j=l \\ \lambda_{j-1}\delta+o(\delta) & \textit{if } j-1=l \\ o(\delta) & \textit{o/w} \end{cases}.$$

Hence

$$p_{ij}(t+\delta) = p_{ij}(t)[1-\lambda_j\delta]+p_{i,j-1}(t)\lambda_{j-1}\delta+o(\delta).$$

Note that we used the fact that we can only be in the state where there are $j-1$ individuals, and there is a birth, or we have jumped from i to j already and there is no birth.

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 $p_{ij}(0)=\delta_{ij}$.

Recall that δ_{ij} denotes the Kronecker delta.

The **backward equations** may be derived in a similar fashion. The difference here, is that instead of considering a transition at time δ in the past, we think about the transition at time t in the past

$$\begin{aligned} p_{ij}(t+\delta) &= \mathbb{P}(N_{t+\delta}=j|N_0=i) \\ &= \sum_{l}\mathbb{P}(N_{t+\delta}=j|N_\delta=l,N_0=i)\mathbb{P}(N_\delta=l|N_0=i) \\ &= \sum_{l}\mathbb{P}(N_{t+\delta}=j|N_\delta=l)\mathbb{P}(N_\delta=l|N_0=i) \\ &= \sum_{l}p_{lj}(t)\mathbb{P}(N_\delta=l|N_0=i). \end{aligned}$$

From the definition of the birth process, we get

$$\mathbb{P}(N_\delta=i+(l-i)|N_0=i)=\begin{cases} 1-\lambda_i\delta+o(\delta) & \textit{if } i=l \\ \lambda_i\delta+o(\delta) & \textit{if } i+1=l \\ o(\delta) & \textit{o/w} \end{cases}.$$

Hence

$$p_{ij}(t+\delta) = p_{ij}(t)[1-\lambda_i\delta]+p_{i+1,j}(t)\lambda_i\delta+o(\delta),$$

where we consider the two possible transitions that are not $o(\delta)$.

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 $p_{ij}(0) = \delta_{ij}$.

So, we found the forward equations:

$$\frac{dp_{ij}(t)}{dt} = -\lambda_j p_{ij}(t) + \lambda_{j-1} p_{i,j-1}(t)$$

with $\lambda_{-1} = 0$., and the backward equations:

$$\frac{dp_{ij}(t)}{dt} = -\lambda_i p_{ij}(t) + \lambda_i p_{i+1,j}(t)$$

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 (which we consider below) gives 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 4.5.6. *Let $\{N_t\}_{t\geq 0}$ be a birth process of positive intensities $\lambda_0, \ldots$. Then the forward equations have a unique solution, which satisfies the backward equations.*

Proof. See Grimmett & Stirzaker (2001*b*, p. 251). □

Example 4.5.7.
Suppose that we are given a birth process with rates $\lambda_n = n\lambda$ and $N_0 = I$. Find the probability that $N_t = k$, $k \geq I$. To solve this problem, we use an inductive proof, via the forward equations. Recall the forward equation for $j \geq i$:

$$p'_{ij}(t) = -\lambda_j p_{ij}(t) + p_{i,j-1}(t)\lambda_{j-1}.$$

Since $N_0 = I$, we set $i := I$. Let $j = I$, then the forward equations give:

$$p'_{II}(t) = -I\lambda p_{II}(t), \quad p_{II}(0) = 1,$$

(since $p_{I(I-1)}(t) = 0$). The solution of the ODE is easily seen to be

$$p_{II}(t) = e^{-I\lambda t}.$$

Now, let $j = I + 1$, which gives the ODE

$$p'_{I(I+1)}(t) = -(I+1)\lambda p_{I(I+1)}(t) + I\lambda e^{-I\lambda t}.$$

Using the integrating factor approach, coupled with the initial condition $p_{I(I+1)}(0) = 0$ we obtain

$$p_{I(I+1)}(t) = Ie^{-\lambda It}(1 - e^{-\lambda t}).$$

After some extra manipulations, try the induction hypothesis

$$p_{Ij}(t) = \binom{j-1}{I-1} e^{-\lambda It}(1 - e^{-\lambda t})^{j-I} \quad j \geq I.$$

Letting $j = j + 1$ and solving the resulting ODE (as above) yields the above formula as the transition probability $p_{Ij}(t)$.

Exercise 4.5.8. *Show that $\mathbb{E}[N_t] = Ie^{\lambda t}$ and $\mathbb{V}ar[N_t] = Ie^{2\lambda t}(1 - e^{-\lambda t})$.*

Solution to Exercise 4.5.8. Note that we were given the information that $N_0 = I$, hence $\mathbb{P}(N_0 = I) = 1$. Hence for $j \geq I$:

$$p_j(t) := \mathbb{P}(N_t = j) = \mathbb{P}(N_t = j|N_0 = I)\mathbb{P}(N_0 = I) = p_{Ij}(t) = \binom{j-1}{I-1} e^{-\lambda I t}(1 - e^{-\lambda t})^{j-I}.$$

Let $p := e^{-\lambda t}$. Then

$$p_j(t) = \binom{j-1}{I-1} p^I (1-p)^{j-I}.$$

This is the probability mass function of a random variable X with negative binomial distribution with parameters I and p. Recall that the negative binomial distribution describes the waiting time for the Ith success and can be written as the sum of i.i.d. geometric random variables. I.e. let $G_i \sim Geometric(p)$ i.i.d.. Then

$$\mathbb{P}(G_i = j) = p(1-p)^{j-1}, \text{ for } i = 1, 2, \ldots,$$

which is the waiting time for the ith success. Now, $X = \sum_{i=1}^{I} G_i$. Recall that the mean and the variance of the geometric distribution are given by $1/p$ and $(1-p)p^{-2}$. Then you get immediately the result for the mean and variance of the negative Binomial distribution.
Alternatively, you can compute e.g. the moment generating function:

$$M_X(\theta) := \mathbb{E}(e^{\theta X}) = \sum_{j=I}^{\infty} e^{\theta j} \binom{j-1}{I-1} p^I (1-p)^{j-I}.$$

Now replace j by $j + I$ in the sum, and you get

$$M_X(\theta) = \sum_{j=0}^{\infty} e^{\theta(j+I)} \binom{j+I-1}{I-1} p^I (1-p)^j.$$

Note that

$$\binom{j+I-1}{I-1} = \frac{(j+I-1)!}{(I-1)!j!} = \binom{j+I-1}{j}.$$

Then

$$M_X(\theta) = \sum_{j=0}^{\infty} e^{\theta(j+I)} \binom{j+I-1}{j} p^I (1-p)^j = e^{\theta I} p^I \sum_{j=0}^{\infty} e^{\theta j} \binom{j+I-1}{j} (1-p)^j.$$

Recall the following identity for the binomial coefficient

$$\binom{j+I-1}{j} = (-1)^j \binom{-I}{j},$$

which leads to

$$M_X(\theta) = e^{\theta I} p^I \sum_{j=0}^{\infty} e^{\theta j} (-1)^j \binom{-I}{j} (1-p)^j = e^{\theta I} p^I \sum_{j=0}^{\infty} \binom{-I}{j} (e^{\theta}(p-1))^j = e^{\theta I} p^I (1 + e^{\theta}(p-1))^{-I}$$
$$= \left(\frac{e^{\theta} p}{1 - e^{\theta}(1-p)} \right)^I$$

where we used the binomial theorem. Now you only have to evaluate the first and second derivative at $\theta = 0$ and you obtain the mean and the second moment and hence the variance. We compute the first derivative here, the second derivative is left as an exercise.

$$M_X'(\theta) = I \left(\frac{e^{\theta} p}{1 - e^{\theta}(1-p)} \right)^{I-1} \frac{(1 - e^{\theta}(1-p))e^{\theta} p - \left(e^{\theta} p(-e^{\theta}(1-p))\right)}{(1 - e^{\theta}(1-p))^2}.$$

Setting $\theta = 0$, we obtain

$$\mathbb{E}(X) = M_X'(\theta)|_{\theta=0} = I \underbrace{\left(\frac{e^0 p}{1 - e^0(1-p)}\right)^{I-1}}_{=1} \frac{(1 - e^0(1-p))e^0 p - \left(e^0 p(-e^0(1-p))\right)}{(1 - e^0(1-p))^2}$$

$$= I \frac{p^2 - p^2 + p}{p^2} = I \frac{1}{p} = I e^{\lambda t}.$$

□

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

4.5.2 Explosion of a birth process

Lecture 20

If the rate $\lambda_0, \lambda_1, \ldots$ 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 4.5.9. *Let $T_1, T_2, \ldots$ denote the event/arrival times of a birth process N (i.e. $T_n = \inf\{t \geq 0 : N_t = n\}$) and $X_1, X_2, \ldots$ the corresponding inter-arrival times. Let the limit of the arrival times be denoted by*

$$T_\infty = \lim_{n\to\infty} T_n.$$

Then we say that explosion of the birth process N is possible if

$$\mathbb{P}(T_\infty < \infty) > 0.$$

Note that since

$$T_n = \sum_{i=1}^{n} X_i, \quad \text{we also have that} \quad T_\infty = \lim_{n\to\infty} T_n = \sum_{i=1}^{\infty} X_i.$$

Theorem 4.5.10. *Let N be a birth process started from $k \in \mathbb{N}_0$. Then:*

1. *If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} < \infty$, then $\mathbb{P}(T_\infty < \infty) = 1$, i.e. explosion occurs with probability 1;*
2. *If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} = \infty$, then $\mathbb{P}(T_\infty = \infty) = 1$, i.e. the probability that explosion occurs is 0.*

Theorem 4.5.10 follows immediately from the Lemma below. (Convince yourself that this is indeed true!)

Lemma 4.5.11. *Let $T_\infty = \sum_{i=k+1}^{\infty} X_i$ for independent random variables $X_i \sim \mathcal{E}xp(\lambda_{i-1})$ for $0 < \lambda_i < \infty$ for all i and for $k \in \mathbb{N}_0$. Then:*

1. *If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} < \infty$, then $\mathbb{P}(T_\infty < \infty) = 1$;*
2. *If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} = \infty$, then $\mathbb{P}(T_\infty = \infty) = 1$.*

W.l.o.g. we assume in the following that $k = 0$ (otherwise we could just shift the index).

Proof of Lemma 4.5.11 (1.)

Recall that the X_i are independent $\mathcal{E}xp(\lambda_{i-1})$ random variables and

$$T_\infty = \sum_{i=1}^{\infty} X_i.$$

Recall that Tonelli's says that you can interchange order of integration, countable summation and expectation whenever the integrand/summands/random variables are *nonnegative*.
Taking expectation:

$$\begin{aligned} \mathbb{E}(T_\infty) &= \mathbb{E}\left(\sum_{i=1}^{\infty} X_i\right) \qquad \text{apply Tonelli's theorem} \\ &= \sum_{i=1}^{\infty} \mathbb{E}(X_i) = \sum_{i=1}^{\infty} \frac{1}{\lambda_{i-1}} = \sum_{i=0}^{\infty} \frac{1}{\lambda_i}. \end{aligned}$$

Now use the assumption that the series is finite, to deduce that $\mathbb{E}(T_\infty) < \infty$.
This implies that $\mathbb{P}(T_\infty < \infty) = 1$.

Note that $\mathbb{E}(T_\infty) = \infty$ *does not* imply that $\mathbb{P}(T_\infty = \infty) > 0$.

□

Proof of Lemma 4.5.11 (2.)

We show that $\mathbb{E}(\exp(-T_\infty)) = 0$ since this will imply $\mathbb{P}(T_\infty = \infty) = 1$.
Using monotonicity of the T_n, dominated convergence, and independence of X_i, we get

$$\begin{aligned} \mathbb{E}(\exp(-T_\infty)) &= \mathbb{E}\left(\prod_{i=1}^{\infty} \exp(-X_i)\right) = \lim_{n\to\infty} \mathbb{E}\left(\prod_{i=1}^{n} \exp(-X_i)\right) \\ &= \lim_{n\to\infty} \prod_{i=1}^{n} \mathbb{E}(\exp(-X_i)). \end{aligned}$$

Recall the mgf of an exponential rv:

$$\mathbb{E}(\exp(-X_i)) = \frac{1}{1 + 1/\lambda_{i-1}}.$$

Therefore

$$\mathbb{E}(\exp(-T_\infty)) = \prod_{i=1}^{\infty} \frac{1}{1 + 1/\lambda_{i-1}} = \prod_{i=0}^{\infty} \left(1 + \frac{1}{\lambda_i}\right)^{-1}.$$

Taking logs, we get

$$-\log\left(\mathbb{E}(\exp(-T_\infty))\right) = \sum_{i=0}^{\infty} \log\left(1 + \frac{1}{\lambda_i}\right). \tag{4.5.2}$$

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 (4.5.2) 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 (4.5.2) diverges, too.

Hence, $-\log(\mathbb{E}(\exp(-T_\infty))) = \infty$, which implies that $\mathbb{E}(\exp(-T_\infty)) = 0$. So we can conclude that $\mathbb{P}(T_\infty = \infty) = 1$.

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. Alternatively, we can recall the following inequality for the logarithm: $\log(1 + x) > \frac{x}{x+1}$ for $x > -1$. In our case, this 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 harmonic series, which diverges.

$\square$

4.6 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 $E = \{0, 1, \dots\}$
2. The infinitesimal transition probabilities are (for $t \geq 0$, $\delta > 0$, $n, m \in \mathbb{N}_0$):

$$\mathbb{P}(X_{t+\delta} = n + m | X_t = n) = \begin{cases} 1 - (\lambda_n + \mu_n)\delta + o(\delta), & \textit{if } m = 0, \\ \lambda_n\delta + o(\delta) & \textit{if } m = 1 \\ \mu_n\delta + o(\delta) & \textit{if } m = -1 \\ o(\delta) & \textit{if} |m| > 1 \end{cases}$$

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

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

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 $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$;

$$\begin{aligned} -\lambda_0\pi_0 + \mu_1\pi_1 &= 0 \\ \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} &= 0 \quad n \geq 1. \end{aligned}$$

Exercise 4.6.1. *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

$$\sum_{n=0}^{\infty} \frac{\lambda_0 \times \cdots \times \lambda_{n-1}}{\mu_1 \times \cdots \times \mu_n} < +\infty \tag{4.6.1}$$

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 4.4.6 that the process settles into equilibrium if and only if (4.6.1) holds; i.e. that the birth rates are not too large relative to the death rates.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 21

Example 4.6.2 (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 $(i, j \in \mathbb{N}_0)$:*

$$p_{ij}(\delta) = \mathbb{P}(X_{t+\delta} = j | X_t = i) = \begin{cases} \mathbb{P}(j-i \text{ arrivals, no deaths}) + o(\delta) & \text{if } j \geq i \\ \mathbb{P}(i-j \text{ deaths, no arrivals}) + o(\delta) & \text{if } j < i \end{cases}$$

since the probability of two or more changes in $(t, t+\delta)$ is $o(\delta)$. As a result

$$\begin{aligned} p_{i,i+1}(\delta) &= \lambda\delta(1-\mu\delta)^i + o(\delta) = \lambda\delta + o(\delta) \\ p_{i,i-1}(\delta) &= (1-\lambda\delta)(1-\mu\delta)^{i-1} \underbrace{\binom{i}{1}}_{=i} \mu\delta + o(\delta) = i\mu\delta + o(\delta), \\ p_{ii}(\delta) &= 1 - (\lambda + i\mu)\delta + o(\delta), \\ p_{ij}(\delta) &= o(\delta) \quad \text{if } |j-i| > 1 \end{aligned}$$

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 4.6.3 (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 $N_t = n$, 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 N_t is

$$G(s,t) = \mathbb{E}(s^{N_t}) = \left\{ \frac{\mu(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}}{\lambda(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}} \right\}^{n_0}.$$

If $\lambda \neq \mu$, what is the probability that extinction has occurred at, or before time t?

Let $\mathbb{P}(N_t = n) = p_n(t)$ and use the convention that $p_{-1}(t) \equiv 0$, then the forward equations for any $n \geq 0$ are

$$p_n(t+\delta) = p_n(t)(1 - n(\lambda+\mu)\delta) + p_{n+1}(t)(n+1)\mu\delta + p_{n-1}(t)(n-1)\lambda\delta + o(\delta)$$

that is,

$$p_n'(t) = -n(\lambda+\mu)p_n(t) + (n+1)\mu p_{n+1}(t) + (n-1)\lambda p_{n-1}(t).$$

and $p_{n_0}(0) = 1$. By definition

$$G(s,t) = \mathbb{E}(s^{N_t}) = \sum_{n=0}^{\infty} s^n p_n(t)$$

and note

$$\frac{\partial G(s,t)}{\partial t} = \sum_{n=0}^{\infty} s^n p_n'(t)$$

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

$$\eta(t) = \mathbb{P}(N_t = 0) = G(0,t) = \left\{ \frac{\mu - \mu e^{-(\lambda-\mu)t}}{\lambda - \mu e^{-(\lambda-\mu)t}} \right\}^{n_0}.$$

Then, as $t \to \infty$,

$$\eta(t) \to \begin{cases} 1, & \text{if } \rho \leq 1, \\ \rho^{-n_0}, & \text{if } \rho > 1. \end{cases}$$

4.7 Jump chain and explosion

An alternative approach to continuous–time Markov chains is to focus on its changes of state at the times of jumps. Subject to regularity conditions not stated here, one could argue as follows:

- Let T_n denote the nth change in value of the chain X and set $T_0 = 0$.
- The values $Z_n = X(T_n+)$ 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 $T_{n+1} - T_n$ has exponential distribution with parameter g_j.
- The chain Z is called the **jump chain of** X.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆ Lecture 22

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. Find a continuous-time Markov chain, which has Z as its jump chain!
Many such chains X exist!

- Let E be a countable state space.
- 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 $H_0 \sim Exp(g_{Z_0})$ the process jumps to state Z_1.
- After a holding time $H_1 \sim Exp(g_{Z_1})$ the process jumps to state Z_2, etc...
- More formally: Conditional on the values Z_n of the chain Z, let $H_0, H_1, \ldots$ be independent random variables with exponential distribution $H_i \sim Exp(g_{Z_i})$, $i = 0, 1, \ldots$. Set $T_n = H_0 + \cdots + H_{n-1}$.
- Then define
$$X_t = \begin{cases} Z_n, & \text{if } T_n \leq t < T_{n+1} \text{ for some } n, \\ \infty, & \text{otherwise, i.e. if } T_\infty \leq t. \end{cases}$$
- Note that the special state ∞ has been added in case the chain explodes.

Recall that $T_\infty = \lim_{n\to\infty} T_n$. T_∞ is called explosion time and we say that the chain explodes if

$$\mathbb{P}(T_\infty < \infty) > 0.$$

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 4.7.1. *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 $H_n \sim Exp(g_{Z_n})$.
- Clearly, if $g_{Z_n} > 0$, then $V_n = g_{Z_n} H_n \sim Exp(1)$ (Exercise!).
- If $g_{Z_n} = 0$, then $H_n = \infty$ almost surely.
- Hence
$$\gamma T_\infty = \begin{cases} \infty, & \text{if } g_{Z_n} = 0 \text{ for some } n, \\ \sum_{n=1}^\infty \gamma H_n \geq \sum_{n=1}^\infty V_n, & \text{otherwise.} \end{cases}$$
- Similarly to the proof of Lemma 4.5.11 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 time $N_0 < N_1 < \cdots$ say.
- Then
$$g_i T_\infty \geq \sum_{j=0}^\infty g_i H_{N_j}, \text{ where } H_{N_j} \sim Exp(g_i) \quad \forall j.$$
- Again, as in the proof of Lemma 4.5.11 one can then show that the sum is almost surely infinite and hence there is no explosion.

□

4.8 Uniform semigroups

Remark 4.8.1. *If you would like to learn more about the technical details we skipped in our heuristic proofs, then read Chapter 6.10 on "Uniform semigroups" in Grimmett & Stirzaker (2001*b*)!*

Definition 4.8.2. *A semigroup* $\{\mathbf{P}_t\}$ *is called* uniform *if* $\mathbf{P}_t \to \mathbf{I}$ *uniformly as* $t \downarrow 0$, *i.e.*

$$p_{ii}(t) \to 1 \quad \text{as } t \downarrow 0, \quad \text{uniformly in } i \in E. \tag{4.8.1}$$

Since $p_{ij}(t) \leq 1 - p_{ii}(t)$, equation (4.8.1) 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 4.8.3. *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 4.8.4. *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,*

$$\mathbf{P}_t = e^{t\mathbf{G}} \qquad \textit{and} \qquad \mathbf{G}\mathbf{1}' = \mathbf{0}'.$$

In the statement above $\mathbf{0}$ and $\mathbf{1}$ denote row vectors consisting of 0s and 1s, respectively.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

Bibliography

Billingsley, P. (2012), *Probability and measure*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, NJ.

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.

Mikosch, T. (2009), *Non-Life Insurance Mathematics: An Introduction with the Poisson Process*, 2 edn, Springer-Verlag, Berlin, Heidelberg, Germany.

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.