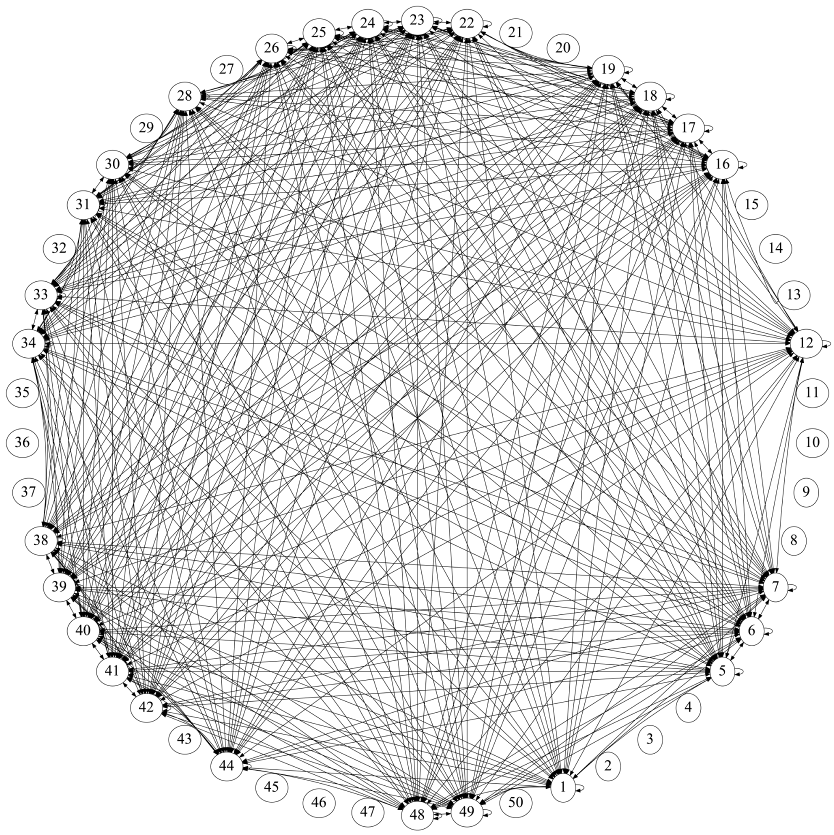
Part B
Applied Probability

Julien Berestycki
Based on notes by Matthias Winkel
Last updated January 2020
Department of Statistics
University of Oxford



MT 2020

Part B
Applied Probability

16 lectures MT 2020

Aims

This course is intended to show the power and range of probability by considering real examples in which probabilistic modelling is inescapable and useful. Theory will be developed as required to deal with the examples.

Synopsis

Poisson processes and birth processes. Continuous-time Markov chains. Transition rates, jump chains and holding times. Forward and backward equations. Class structure, hitting times and absorption probabilities. Recurrence and transience. Invariant distributions and limiting behaviour. Time reversal.

Renewal theory. Limit theorems: strong law of large numbers, strong law and central limit theorem of renewal theory, elementary renewal theorem, renewal theorem, key renewal theorem. Excess life, inspection paradox.

Applications in areas such as: queues and queueing networks – M/M/s queue, Erlang's formula, queues in tandem and networks of queues, M/G/1 and G/M/1 queues; insurance ruin models; epidemic models; applications in applied sciences.

Reading

- J.R. Norris, *Markov chains*, Cambridge University Press (1997)
- G.R. Grimmett, and D.R. Stirzaker, *Probability and Random Processes*, 3rd edition, Oxford University Press (2001)
- G.R. Grimmett, and D.R. Stirzaker, *One Thousand Exercises in Probability*, Oxford University Press (2001)
- S.M. Ross, *Introduction to Probability Models*, 4th edition or later editions, Academic Press (1989+)
- D.R. Stirzaker, *Elementary Probability*, 1st edition or later editions, Cambridge University Press (1994, 2003)

Contents

This course is, in the first place, a course for 3rd year undergraduates who did Part A Probability in their 2nd year. Other students such as MSc students are welcome, but should note the prerequisites of the course. These are essentially an introductory course in probability *not* based on measure theory. It will be an advantage if this included the central aspects of discrete-time Markov chains. This will be relevant by the time we get to Lecture 5 in week 3.

This is a mathematics course. The name "Applied probability" suggests that we apply probability. However, there is more to it. This course is about "probability" and "applications" or application-driven probability theory. In particular, it is not just Part A Probability that we apply, but also further probability building on Part A. Effectively, we will be spending a fair share of our time developing theory so that we can analyse certain examples and applications.

Lecture 1

Introduction: Poisson processes, generalisations and applications

Reading: Part A Probability; Grimmett-Stirzaker 6.1, 6.8 up to (10)
Further reading: Ross 4.1, 5.3; Norris Introduction, 1.1, 2.4

The aim of Lecture 1 is to give a brief overview of the course. To do this at an appropriate level, we begin with a review of Poisson processes, which were treated at the end of the Part A course. The material most relevant to us is again included here, and some more is on the first assignment sheet.

For the rest of the course, let $\mathbb{N} = \{0, 1, 2, \ldots\}$ denote the natural numbers including zero. Apart from very few exceptions, all stochastic processes that we consider in this course will have state space $\mathbb{N}$ (or a subset of $\mathbb{N}$). Specifically, we have in mind that we are counting, and studying the evolution of, numbers of people in a population, affected by a disease, of a certain genetic type, in a queue, etc. or just balls in an urn, bacteria in a dish, numbers of claim-free years for a motor insurance, the wealth of a gambler, numbers of defective items in a production line.

However, most results in the theory of Markov chains will be treated for any *countable*, i.e. finite or countably infinite, state space $\mathbb{S}$. This does not pose any complications as compared with $\mathbb{N}$, since we can always enumerate all states in $\mathbb{S}$ and hence give them labels in $\mathbb{N}$. Important examples are $\mathbb{Z}$, $\mathbb{N}^2$ and finite sets such as $\{\text{bachelor}, \text{married}, \text{divorced}, \text{widowed}\}$ or a set of colours, car makes, universities, shops, or indeed sets like $\{0, 1\}^n$. For uncountable state spaces, however, several technicalities arise that are beyond the scope of this course, at least in any generality – we will naturally come across a few examples of Markov processes in $\mathbb{R}$ towards the end of the course.

1.1 Poisson processes

There are many ways to define Poisson processes. We will use the following definition. We write $Z \sim \text{Exp}(\lambda)$ to say "Z is an exponentially distributed random variable with probability density function $\lambda e^{-\lambda t}$, $t \geq 0$", for some $\lambda > 0$.

Definition 1 Let $Z_n \sim \text{Exp}(\lambda)$, $n \geq 1$, independent, for some $\lambda > 0$. Let $T_n = Z_1 + \cdots + Z_n$, $n \geq 1$. Then the process $X = (X_t, t \geq 0)$ defined by

$$X_t = \#\{n \geq 1 \colon T_n \leq t\}, \qquad t \geq 0,$$

is called *Poisson process with rate* λ, abbreviated $\text{PP}(\lambda)$. For $k \in \mathbb{N}$, the process

$$Y_t = k + X_t \qquad t \geq 0,$$

is called the *Poisson process with rate* λstarted from k. Unless specified otherwise, a $\text{PP}(\lambda)$ is always assumed to start from 0.

Think of T_n as arrival times of customers (arranged in increasing order). Then X_t is counting the numbers of arrivals up to time t for all $t \geq 0$ and we study the evolution of this counting process. Instead of customers, one might be counting particles detected by a Geiger counter or cars driving through St. Giles, etc. Something more on the link and the important distinction between real observations (cars in St. Giles) and mathematical models (Poisson process) will be included in Lecture 2. For the moment we have a mathematical model, well specified in the language of probability theory. Starting from a simple sequence of independent random variables Z_n, $n \geq 0$, we have defined a more complex object $(X_t, t \geq 0)$, that we call the Poisson process.

1.2 The Markov property

Discrete time Markov chains:

Let $\mathbb{S}$ be a countable state space, typically $\mathbb{S} = \mathbb{N}$. Let $\Pi = (\pi_{rs})_{r,s\in\mathbb{S}}$ be a Markov transition matrix on $\mathbb{S}$. An $\mathbb{S}$-valued stochastic process $M = (M_n, n \geq 0)$ is called a *discrete time Markov chain with transition matrix* Π *starting from* $i_0 \in \mathbb{S}$ if for all $n \geq 1$ and $i_1, \ldots, i_n \in \mathbb{S}$

$$\mathbb{P}(M_1 = i_1, \ldots, M_n = i_n) = \prod_{j=1}^{n} \pi_{i_{j-1}, i_j}.$$

It is often convenient to capture the initial state by writing $\mathbb{P}_{i_0}$ instead of $\mathbb{P}$. We then say that M is starting from i_0 *under* $\mathbb{P}_{i_0}$. We will use notation such as M, $(M_n, n \geq 0)$ and $(M_n)_{n\geq 0}$ interchangeably. Markov chains have the Markov property, which can be stated in several useful ways:

- For all paths $i_0, \ldots, i_{n+1} \in \mathbb{S}$ of positive probability $\mathbb{P}(M_0 = i_0, \ldots, M_n = i_n) > 0$, we have
$$\mathbb{P}(M_{n+1} = i_{n+1} | M_0 = i_0, \ldots, M_n = i_n) = \mathbb{P}(M_{n+1} = i_{n+1} | M_n = i_n) = \pi_{i_n, i_{n+1}}.$$
- For all $k \in \mathbb{S}$ and events $\{(M_j)_{0\leq j\leq n} \in A\}$ and $\{(M_{n+m})_{m\geq 0} \in B\}$, we have: if $\mathbb{P}(M_n = k, (M_j)_{0\leq j\leq n} \in A) > 0$, then
$$\mathbb{P}((M_{n+m})_{m\geq 0} \in B | M_n = k, (M_j)_{0\leq j\leq n} \in A) = \mathbb{P}((M_{n+m})_{m\geq 0} \in B | M_n = k) = \mathbb{P}_k(M \in B).$$
- The processes $(M_j)_{0\leq j\leq n}$ and $(M_{n+m})_{m\geq 0}$ are conditionally independent given $M_n = k$, for all $k \in \mathbb{S}$. Furthermore, given $M_n = k$, the process $(M_{n+m})_{m\geq 0}$ is a Markov chain with transition matrix Π starting from k.

Informally: no matter how we got to a state, the future behaviour of the chain is as if we were starting a new chain from that state. This is one reason why it is vital to study Markov chains not starting from one initial state but from any state in the state space $\mathbb{S}$.

The Markov property for Poisson processes:

Proposition 2 *Let $X = (X_t)_{t\geq 0}$ be a Poisson process of rate λ started from 0. Fix $t \geq 0$, then conditionally on $X_t = k$, the processes $(X_r)_{0\leq r\leq t}$ and $(X_{t+s})_{s\geq 0}$ are independent and $(X_{t+s})_{s\geq 0}$ is a Poisson process of rate λ started from k.*

Remark 3

1. *We could, equivalently, have said that $(X_{t+s} - X_t)_{s\geq 0}$ is a Poisson process of rate λ started from 0 independent of $(X_r)_{0\leq r\leq t}$.*
2. *This property will be later generalized to the strong Markov property.*

We will need the following facts about Exponential random variables for the proof:

Lemma 4 *(Extended memoryless property) Let $E \sim Exp(\lambda)$ and let $x, y > 0$. Then*

$$\mathbb{P}(E > x + y | E > y) = \mathbb{P}(E > x) = e^{-\lambda x}.$$

Let $L \geq 0$ be a non-negative random variable independent of E. Then

$$\mathbb{P}(E > x + L | E > L) = \mathbb{P}(E > x) = e^{-\lambda x}$$

so conditionnally on $E > L$ we have $E - L \sim Exp(\lambda)$ and is independent of L.

See p.s.# 1 for the proof.

Proof: [Proof of Porposition 2] Define $\tilde{X}_s = X_{t+s}$ and remember that we are working conditionally on $X_t = k$. Clearly, $\tilde{X}_0 = k$ and $\tilde{X}$ is a right-continuous, integer valued, increasing process. Let us write $\tilde{Z}_1, \tilde{Z}_2, \ldots$ for the holding times of $\tilde{X}$ (see next section for why this is well defined). Observe that

$$\tilde{Z}_1 = Z_{k+1} - (t - T_k), \qquad \tilde{Z}_n = Z_{k+n}, n \geq 2.$$

Furthermore,

$$\{X_t = k\} = \{T_k \leq t < T_{k+1}\} = \{T_k \leq t\} \cap \{Z_{k+1} > t - T_k\}.$$

By the extended memoryless property, conditionally on $Z_{k+1} > t - T_k$ we have that $\tilde{Z}_1 = Z_{k+1} - (t - T_k) \sim \text{Exp}(\lambda)$ indpendently of T_k. Since $Z_{k+2}, Z_{k+3}, \ldots$ are i.i.d. $\text{Exp}(\lambda)$ independent of $Z_1, Z_2, \ldots Z_k$ we see that $\tilde{Z}_1, \tilde{Z}_2, \ldots$ are i.i.d. $\text{Exp}(\lambda)$ independent of $Z_1, Z_2, \ldots Z_k$. Since, given $X_t = k$

$$X_r = \#\{1 \leq n \leq k : \sum_{i=1}^{n} Z_i \leq r\}, \qquad r \leq t$$

$$\tilde{X}_s = \#\{n \geq 1 : \sum_{i=1}^{n} \tilde{Z}_i \leq s\}, \qquad s \geq 0$$

we see that $(X_r)_{r \leq t}$ and $(\tilde{X}_s)_{s \geq 0}$ are conditionally independent and that $(\tilde{X}_s)_{s \geq 0}$ is a $\text{PP}(\lambda)$ started from k. □

Markov models (stochastic processes that have the Markov property and that model real-life evolutions, natural or man-made) are useful in a wide range of applications, e.g. price processes in Mathematical Finance, evolution of genetic material in Mathematical Biology, evolutions of particles in space in Mathematical Physics. The Markov property is a property that makes the model somewhat simple (not easy, but it could be much less tractable). We will develop tools that support this statement.

1.3 Properties and characterization

Let us collect some properties that, apart from some technical details (to do with handling uncountably many random variables), can serve as an alternative definition of the Poisson process.

Proposition 5 A process $X \sim \text{PP}(\lambda)$ has the following properties:

(i) $X_t \sim \text{Poi}(\lambda t)$ for all $t \geq 0$, where $\text{Poi}(\lambda t)$ refers to the Poisson distribution with mean λt.

(ii) X has independent increments, i.e. for all $t_0 \leq \cdots \leq t_n$, the random variables $X_{t_j} - X_{t_{j-1}}$, $1 \leq j \leq n$, are independent.

(iii) X has stationary increments, i.e. $X_{t+s} - X_t \sim X_s$ for all $t \geq 0$, $s \geq 0$, where $\sim$ means "has the same distribution as".

Proof: Start with (i). We are going to use that $T_n = \sum_{i=1}^n Z_i$ is a sum of n independent $\text{Exp}(\lambda)$ r.v.'s and thus $T_n \sim \Gamma(n, \lambda)$ with density $f_{T_n}(s) = \lambda^n s^{n-1} e^{-\lambda s}/(n-1)!$. Hence, since they are independent, T_n, Z_{n+1} have joint density

$$f_{T_n, Z_{n+1}}(s, z) = \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!} \lambda e^{-\lambda z}$$

and we get

$$\begin{aligned}
\mathbb{P}(X_t = n) &= \mathbb{P}(T_n \le t, Z_{n+!} \ge t - T_n) \\
&= \int_0^t \int_{t-s}^\infty f_{T_n, Z_{n+1}}(s, z) ds dz \\
&= \int_0^t \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!} e^{-\lambda(t-s)} ds \\
&= \int_0^t \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!} e^{-\lambda(t-s)} ds \\
&= \frac{e^{-\lambda t}}{(n-1)!} \int_0^t s^{n-1} ds \\
&= \frac{e^{-\lambda t} (\lambda t)^n}{n!}.
\end{aligned}$$

For (ii), Take $n \ge 1$, $0 = t_0 \le t_1 \le \ldots \le t_n < \infty$ and $i_1, \ldots, i_n \in \mathbb{N}$. Define $A_k = \{X_{t_k} - X_{t_{k-1}} = i_k\}$. Then, using the Markov property

$$\begin{aligned}
\mathbb{P}(\cap_{k=1}^n A_k) &= \mathbb{P}(A_n | \cap_{k=1}^{n-1} A_k) \mathbb{P}(\cap_{k=1}^{n-1} A_k) \\
&= \mathbb{P}(X_{t_n} = i_n + \sum_{k=1}^{n-1} i_k | X_{t_{n-1}} = \sum_{k=1}^{n-1} i_k, X_{t_{n-2}} = \sum_{k=1}^{n-2} i_k, \ldots, X_{t_1} = i_1) \mathbb{P}(\cap_{k=1}^{n-1} A_k) \\
&= \mathbb{P}(X_{t_n} = i_n + \sum_{k=1}^{n-1} i_k | X_{t_{n-1}} = \sum_{k=1}^{n-1} i_k) \mathbb{P}(\cap_{k=1}^{n-1} A_k) \\
&= \mathbb{P}(X_{t_n} - X_{t_{n-1}} = i_n) \mathbb{P}(\cap_{k=1}^{n-1} A_k) \\
&= \mathbb{P}(A_n) \mathbb{P}(\cap_{k=1}^{n-1} A_k) \\
&= \prod_{k=1}^n \mathbb{P}(A_k)
\end{aligned}$$

by induction. This proves (ii). (iii) is a direct consequence of the fact that $(X_{t+s} - X_t)_{s \ge 0}$ is a $\text{PP}(\lambda)$. □

Remark 6 *In fact, we will see that if X is a right-continuous random process, then X is a Poisson process with rate λ if and only if it satisfies properties (1),(2) and (3) of Proposition 5.*

1.4 Continuous-time random processes

We have been very careful so far to define a Poisson process $(X_t, t \ge 0)$ entirely through the sequence of holding times $(Z_n)_{n \ge 1}$. That is because working with infinite sequences of random variables does not create (too many) measure-theoretical difficulties.

However, we can't always use this trick when working with more general random processes. In that case, it is useful to observe that $(X_t, t \ge 0)$ is not just an uncountable family of (dependent!) random variables but indeed that $t \mapsto X_t$ is a random right-continuous function. This point of view is very useful since it is the formal justification for pictures of "typical realisations" of X.

Let $\mathbb{S}$ be a countable set. A *continuous-time random process*

$$(X_t)_{t\geq 0} = (X_t : 0 \leq t < \infty)$$

with values in $\mathbb{S}$ is a family of random variables $X_t : \Omega \mapsto \mathbb{S}$.

When we talk about the *law* of the process X, we mean a probability measure that should, in principle, be able to handle any event associated with X. We want to be able to deal with quantities such as $\mathbb{P}(X_t = i)$ or $\mathbb{P}(X_{t_0} = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n)$, or $\mathbb{P}(X_t = i, \text{for some } t > 0)$. There are subtleties in this problem not present in the discrete-time case. They arise because

$$\mathbb{P}\left(\cup_n A_n\right) = \sum_n \mathbb{P}(A_n),$$

whereas for an uncountable union $\cup_{t\geq 0} A_t$ no such rule exists.

To avoid these subtleties as far as possible we only consider processes $(X_t)_{t\geq 0}$ which are **right-continuous**. Since our state-space $\mathbb{S}$ is countable this means that for all $\omega \in \Omega$ and $t \geq 0$ there exists $\epsilon > 0$ such that

$$X_s(\omega) = X_t(\omega) \qquad \text{for } t \leq s \leq t + \epsilon.$$

By a standard result of measure theory (see section 6.6 in Norris) the law of a right-continuous process is completely determined by its *finite-dimensional distributions*, that is from the probabilities

$$\mathbb{P}(X_{t_0} = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n)$$

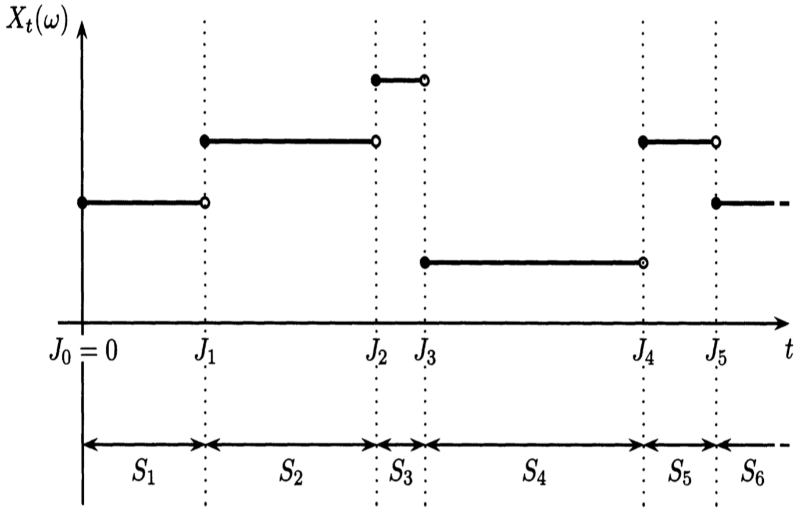
for $n \geq 0, 0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$ and $i_0, \ldots, i_n \in \mathbb{S}$.

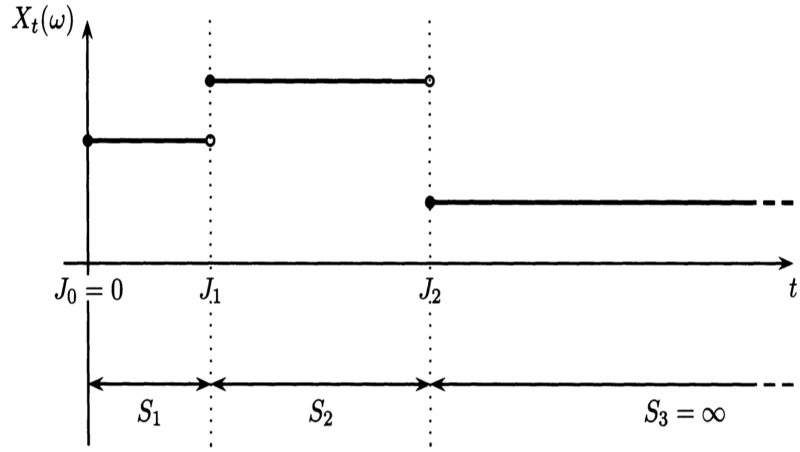
Example 7

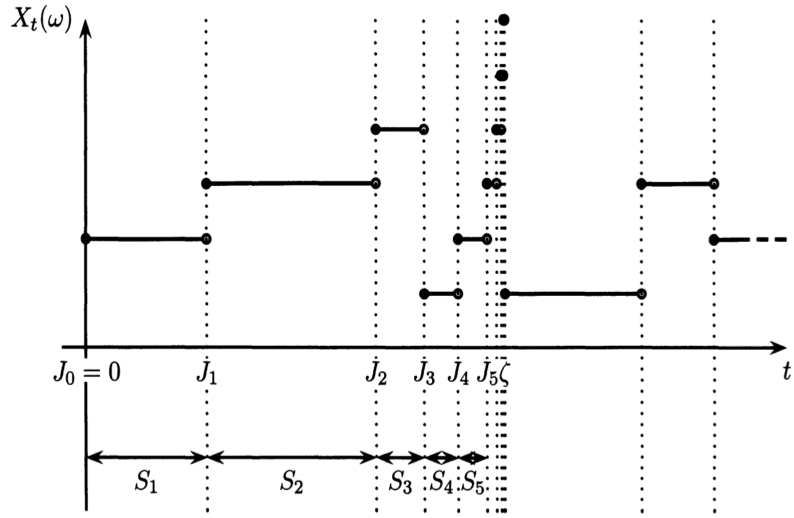
$$\mathbb{P}(X_t = i, \textit{for some } t > 0) = 1 - \lim_{n\to\infty} \sum_{j_1,\ldots,j_n \neq i} \mathbb{P}(X_{q_1} = j_1, \ldots, X_{q_n} = j_n)$$

where $q_1, q_2, \ldots$ is an enumeration of the rationals.

Every path $t \mapsto X_t(\omega)$ of a right-continuous process must remain constant for a while in each new state, so there are three possibilities for the sort of paths we get. In the first case the process makes infinitely many jumps but only finitely many on any time-interval $[0, t]$.







The second case is where the path makes finitely many jumps and then becomes stuck in some state forever.

The last case is when the process makes infinitely many jumps in finite time. In this case, after the explosion time ζ the process can start-up again; it may explode again or it may not.

Coming back to our statement of the Markov property for a Poisson process $(X_t)_{t\geq 0}$, we are really treating each of $(X_r)_{r\leq t}$ and $(X_{t+s})_{s\geq 0}$ as a random variable in its own right (with values in the sets of integer-valued right-continuous paths) rather than a collection of random variables.

An example of an *event* for such a random variable is

$$\{(X_r)_{r\geq 0} \in A\}$$

where

$$A = \{ \text{ right-continuous functions } f : [0,t] \mapsto \mathbb{S} \text{ such that } f(r) \leq 2 \text{ for all } r \leq t\}.$$

Of course, this is usually written simply as

$$\{X_r \leq 2 \text{ for } 0 \leq r \leq t\}.$$

When we say that for a Poisson process $(X_t)_{t\geq 0}$ the variables $(X_r)_{r\leq t}$ and $(X_{t+s})_{s\geq 0}$ are conditionally independent given $X_t = k$, we mean that for all measurable sets of paths A and B

$$\mathbb{P}((X_{t+s})_{s\geq 0} \in B, (X_r)_{r\leq t} \in A | X_t = k) = \mathbb{P}((X_{t+s})_{s\geq 0} \in B | X_t = k)\mathbb{P}((X_r)_{r\leq t} \in A | X_t = k).$$

Moreover, since the finite dimensional distributions characterise such processes, this is equivalent to having

$$\begin{aligned}&\mathbb{P}(X_{r_1} = x_1, \ldots, X_{r_m} = x_m, X_{t+s_1} = y_1, \ldots, X_{t+s_n} = y_n | X_t = k)\\ &\qquad = \mathbb{P}(X_{r_1} = x_1, \ldots, X_{r_m} = x_m | X_t = k)\mathbb{P}(X_{t+s_1} = y_1, \ldots, X_{t+s_n} = y_n | X_t = k)\end{aligned}$$

for all $n, m, r_1 \leq \ldots \leq r_m, s_1 \leq \ldots \leq s_n, x_1 \leq \ldots \leq x_m \leq k \leq y_1 \leq \ldots \leq y_n$.

1.5 Brief summary of the course

Two generalisations of the Poisson process and several applications make up this course.

- The Markov property of Proposition **??**(i)-(ii) can be used as a starting point to a bigger class of processes, so-called *continuous-time Markov chains*. They are analogues of discrete-time Markov chains, and they are often better adapted to applications. On the other hand, new aspects arise that did not arise in discrete time, and connections between the two will be studied. Roughly, the first half of this course is concerned with continuous-time Markov chains. Our main reference book will be Norris's book on Markov Chains.

- The Poisson process is the prototype of a counting process. For the Poisson process, "everything" can be calculated explicitly. In practice, though, this is often only helpful as a first approximation. E.g. in insurance applications, the Poisson process is used as a model to count claim arrivals. However, there is empirical evidence that inter-arrival times are neither exponentially distributed nor independent nor identically distributed. The second approximation is to relax exponentiality of inter-arrival times but to keep their independence and identical distribution. These counting processes are called *renewal processes*. Since exact calculations are often impossible or not helpful, the most important results of renewal theory are limiting results. Our main reference will be Chapter 10 of Grimmett and Stirzaker's book on Probability and Random Processes.

- Many applications that we discuss are in queueing theory. The easiest, so-called $M/M/1$ queue consists of a server and customers arriving according to a Poisson process. Independently of the arrival times, each customer has an exponential service time for which he will occupy the server, when it is his turn. When the server is busy, customers queue until being served. Everything has been designed so that the queue length is a continuous-time Markov chain, and various quantities can be studied or calculated (equilibrium distribution, lengths of idle periods, waiting time distributions etc.). More complicated queues arise if the Poisson process is replaced by a renewal process or the exponential service time distribution by any other distribution. There are also systems with $k = 2, 3, \ldots, \infty$ servers. Abstract queueing systems can be applied in telecommunications, computing networks, etc.

- Some other applications include insurance ruin models and the propagation of diseases.

Lecture 2

Simple birth processes and explosion

Reading: Norris 2.2-2.3, 2.5; Grimmett-Stirzaker 6.8 (11),(18)-(20)

In this lecture we introduce birth processes as a generalisation of the Poisson process. This definition and similar definitions (also for Markov chains) answer the following questions: given the current state is m, how does the process behave in the future, and (for the purpose of an inductive description) how does it depend on the past? (Answer to the last bit: conditionally independent given the current state, for certain processes this can be expressed in terms of genuine independence).

2.1 Definition and example of a simple birth process

If we use the Poisson process as a model for a growing population, we assume that new members are born at the same rate irrespective of what the size of the population is. This is often not realistic. We would rather expect this rate to increase with size (more births in larger populations). Some saturation effects may occur, as well.

Also, the use of the Poisson process as a counting process of alpha particle emissions of a decaying radioactive substance becomes questionable when the half-life time is short. We may prefer a model where rates decrease with the number of emissions.

Here is a definition that gives ample modelling choice for these and other examples.

Definition 8 (Simple birth process) A random process $(X_t)_{t\geq 0}$ of the form $X_t = k + \#\{n \geq 1\colon Z_1 + \cdots + Z_n \leq t\}$ is called a *simple birth process of rates* $(\lambda_n)_{n\geq 0}$ starting from $X_0 = k \in \mathbb{N}$, if the *inter-arrival times* Z_j, $j \geq 1$, are independent with $Z_j \sim \text{Exp}(\lambda_{k+j-1})$, $j \geq 1$. We also refer to $X = (X_t, t \geq 0)$ as a $(k, (\lambda_n)_{n\geq 0})$-birth process.

Note that the parameter λ_n is attached to state n. The so-called holding time of X in state n has an $\text{Exp}(\lambda_n)$ distribution, even if the chain starts in state $k > 0$. The first holding time $Z_1 \sim \text{Exp}(\lambda_k)$. Hence, when the chain first reach the state n it waits for an $\text{Exp}(\lambda_n)$ distributed time and then jumps to $n + 1$.

"Simple" refers to the fact that no two births occur at the same time, which one would call "multiple" births. Multiple birth processes can be studied as well, and given certain additional assumptions, these are also examples of continuous-time Markov chains.

Example 9 Consider a population in which each individual gives birth after an exponential time of parameter λ, independently and repeatedly. Denote by X_t the population size at time $t \geq 0$ and suppose that $X_0 = 1$. Let us show that $(X_t, t \geq 0)$ is a simple birth process.

Clearly, the first birth occurs at $Z_1 \sim \text{Exp}(\lambda)$. For the second birth, two $\text{Exp}(\lambda)$ times ("clocks") compete, independently of Z_1. We can study the inter-birth times Z_n, $n \geq 2$, inductively, using parts of the **theory of competing exponentials**

Lemma 10 (Theory of competing exponentials-1) *Let* $E_1, \ldots, E_n$ *be* n *independent* $\sim \text{Exp}(\lambda)$ *random variables. Define* $W := \min\{E_1, \ldots, E_n\}$. *Then* $W \sim \text{Exp}(n\lambda)$. *Furthermore, the* $n - 1$ *residual times* $O_i = E_i - W$ *are independent* $\text{Exp}(\lambda)$, *and are also independent of* W *(a form of the* **lack of memory property of the exponential distribution).**

Proof: See p.s. #1. □

Here is the induction step for our population model: *if the population size is* n, with n (residual) times to birth independent of $Z_1, \ldots, Z_{n-1}$, the next birth occurs after $Z_n \sim \text{Exp}(n\lambda)$, by the *theory of competing exponentials*. By *lack of memory*, there are $n-1$ independent $\text{Exp}(\lambda)$ residual times to birth. With two more independent $\text{Exp}(\lambda)$ times for the new individual and the individual who has just given birth, the induction proceeds with new population size $n+1$.

Since $X_t = 1 + \#\{n \geq 1\colon Z_1 + \cdots + Z_n \leq t\}$, with $Z_j \sim \text{Exp}(j\lambda)$, the process $(X_t, t \geq 0)$ is a simple birth process with $\lambda_n = n\lambda$, a $(1, (n\lambda)_{n\geq 0})$-birth process.

Example 11 In the setting of the previous example, what is $m(t) = \mathbb{E}(X_t)$? We will condition on $T_1 = Z_1$ as in Example 139. We can write $X_{T_1+s} = X_s^{(1)} + X_s^{(2)}$, where $X_s^{(1)}$ and $X_s^{(2)}$ are separate counts of the numbers of descendants of the two individuals after time T_1. By model assumption, $(X_s^{(1)}, s \geq 0)$ and $(X_s^{(2)}, s \geq 0)$ are independent, independent of T_1 and distributed as $(X_s, s \geq 0)$. Therefore, by Proposition 135(c) in the Appendix

$$
\begin{aligned}
m(u) \;=\; \mathbb{E}(X_u) &= \int_0^\infty \lambda e^{-\lambda t}\mathbb{E}(X_u|T_1 = t)dt \\
&= \int_0^u \lambda e^{-\lambda t}\mathbb{E}(X_u|T_1 = t)dt + e^{-\lambda u} \\
&= \int_0^u \lambda e^{-\lambda t}\mathbb{E}(X_{T_1+(u-T_1)}|T_1 = t)dt + e^{-\lambda u} \\
&= \int_0^u \lambda e^{-\lambda t}\mathbb{E}\left(X_{u-T_1}^{(1)} + X_{u-T_1}^{(2)}\,\middle|\, T_1 = t\right)dt + e^{-\lambda u} \\
&= \int_0^u \lambda e^{-\lambda t}\mathbb{E}\left(X_{u-t}^{(1)} + X_{u-t}^{(2)}\right)dt + e^{-\lambda u} \\
&= \int_0^u \lambda e^{-\lambda t}2m(u-t)dt + e^{-\lambda u}
\end{aligned}
$$

also using that $X_u = X_0 = 1$ if $T_1 > u$, the population split $X_{T_1+s} = X_s^{(1)} + X_s^{(2)}$ noted above and Fact 138. Setting $r = u - t$, we can proceed as in Example 139:

$$e^{\lambda u}m(u) = 1 + 2\lambda\int_0^u e^{\lambda r}m(r)dr.$$

differentiated yields

$$m'(u) = \lambda m(u)$$

so the mean population size grows exponentially, and $X_0 = 1$ gives $m(0) = \mathbb{E}(X_0) = 1$, so that

$$\mathbb{E}(X_t) = m(t) = e^{\lambda t}, \qquad t \geq 0.$$

2.2 The explosion phenomenon for simple birth processes

If the rates $(\lambda_n)_{n\geq 0}$ increase too quickly, it may happen that "infinitely many individuals are born in finite time". We call this phenomenon *explosion* and say X *explodes*. Formally, we can express the possibility of explosion by $\mathbb{P}(T_\infty < \infty) > 0$, where $T_\infty = \lim_{n\to\infty} T_n = \sum_{j\geq 1} Z_j$. It is easy to see that this can indeed happen, with probability 1 in fact: by Tonelli's theorem,

$$\mathbb{E}(T_\infty) = \mathbb{E}\left(\sum_{j=0}^\infty Z_j\right) = \sum_{j=0}^\infty \mathbb{E}(Z_j) = \sum_{j=0}^\infty \frac{1}{\lambda_{k+j}}; \qquad (1)$$

so this expectation is finite *if the series of inverse birth rates is summable* (avoid this when modelling!), and then we have $\mathbb{P}(T_\infty < \infty) = 1$, i.e. explosion with probability 1.

Since such birth processes are hardly useful for applications, it will be more useful to have a criterion under which explosion does *not* happen, i.e. under which $\mathbb{P}(T_\infty = \infty) = 1$.

Remember that it is not a valid argument to say that this is ridiculous for the application we are modelling and hence cannot occur in our model. We have to check whether it can occur under the model assumptions. And if it does occur and is ridiculous for the application, it means that the model is not a good model for the application. For simple birth processes, we have the following necessary and sufficient condition.

Proposition 12 *Let X be a $(k, (\lambda_n)_{n\ge0})$-birth process. Then*

$$X \text{ does not explode} \iff \mathbb{P}(T_\infty = \infty) = 1 \iff \mathbb{P}(T_\infty = \infty) > 0 \iff \sum_{m=k}^{\infty} \frac{1}{\lambda_m} = \infty.$$

Proof: The first equivalence is the definition and the others are logically equivalent to

$$\mathbb{P}(T_\infty < \infty) > 0 \iff \mathbb{P}(T_\infty < \infty) = 1 \iff \sum_{m=k}^{\infty} \frac{1}{\lambda_m} < \infty.$$

As noted above, the calculation in (1) gives $\sum_{m\ge k} 1/\lambda_m < \infty \Rightarrow \mathbb{P}(T_\infty < \infty) = 1$. Trivially $\mathbb{P}(T_\infty < \infty) = 1 \Rightarrow \mathbb{P}(T_\infty < \infty) > 0$. It remains to show $\sum_{m\ge k} 1/\lambda_m = \infty \Rightarrow \mathbb{P}(T_\infty = \infty) = 1$. In the following we use extended functions $\exp\colon [-\infty, \infty] \to [0, \infty]$ and $\log\colon [0, \infty] \to [-\infty, \infty]$. Using dominated convergence, continuity of exp and log, and the independence of Z_j, $j \ge 0$,

$$\begin{aligned}
-\log \mathbb{E}\left(\exp(-T_\infty)\right) &= -\log \mathbb{E}\left(\exp\left(-\lim_{n\to\infty} \sum_{j=0}^{n-1} Z_j\right)\right) = -\lim_{n\to\infty} \log \mathbb{E}\left(\exp\left(-\sum_{j=0}^{n-1} Z_j\right)\right) \\
&= -\lim_{n\to\infty} \log \mathbb{E}\left(\prod_{j=0}^{n-1} e^{-Z_j}\right) = -\lim_{n\to\infty} \log \prod_{j=0}^{n-1} \mathbb{E}(e^{-Z_j}) = -\sum_{j=0}^{\infty} \log \mathbb{E}\left(e^{-Z_j}\right) \\
&= \sum_{m=k}^{\infty} \log\left(1 + \frac{1}{\lambda_m}\right) \ge \begin{cases} \displaystyle\sum_{i=1}^{\infty} \log\left(1 + \frac{1}{\lambda_{n_i}}\right) & \text{if } \lambda_{n_i} \le 1 \text{ infinitely often.} \\ \displaystyle\sum_{n=n_0}^{\infty} \log\left(1 + \frac{1}{\lambda_n}\right) & \text{if } \lambda_n > 1 \text{ for all } n \ge n_0 \ge k, \end{cases}
\end{aligned}$$

In both cases, the series is infinite, as we see using, respectively, $\log(1 + 1/\lambda_{n_i}) \ge \log(2)$ and $\log(1 + 1/\lambda_n) \ge \log(2)/\lambda_n$. Hence, $\sum_{m\ge k} 1/\lambda_m = \infty \Rightarrow \mathbb{E}(e^{-T_\infty}) = 0 \Rightarrow \mathbb{P}(T_\infty = \infty) = 1$. □

We have not explicitly specified X after T_∞ if $T_\infty < \infty$. In a "population size model", $X_t = \infty$ for all $t \ge T_\infty$ is a reasonable convention. Formally, this means that X is a process in $\overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. This process is called the *minimal process*. It is "active" on a minimal time interval. We will show the Markov property for minimal processes. It can also be shown that there are other ways to specify X after explosion that preserve the Markov property. The next natural thing to do is to start afresh independently after explosion. Any such process is then called *non-minimal*.

Example 13 In the setting of Example 9, we have $\lambda_m = m\lambda$, hence $\sum_{m\ge k} 1/\lambda_m = \infty$. By Proposition 12, a $(k, (m\lambda)_{m\ge0})$-birth process X does not explode. We knew this already since we showed in Example 11 that $\mathbb{E}(X_t) < \infty$ for all $t \ge 0$, and if X did explode, we would have $\mathbb{P}(X_t = \infty) = \mathbb{P}(T_\infty \le t) > 0$, at least for t sufficiently large (actually for all $t > 0$).

Lecture 3

The Markov property of simple birth processes

Reading: Norris 2.4-2.5; Grimmett-Stirzaker 6.8 (21)-(25)

In this lecture we discuss in detail the Markov property for birth processes. We will use the Markov property for Poisson processes in the construction and analysis of continuous-time Markov chains.

3.1 Statements of the Markov property

Proposition 14 (Markov property) *Let X be a $(k, (\lambda_n)_{n\geq 0})$-birth process, $t \geq 0$ and $\ell \geq k$. Then the processes $(X_r)_{r\leq t}$ and $(X_{t+s})_{s\geq 0}$ are conditionally independent given $X_t = \ell$, and $\widetilde{X} := (X_{t+s})_{s\geq 0}$ is an $(\ell, (\lambda_n)_{n\geq 0})$-birth process. We use the notation*

$$(X_r)_{r\leq t} \coprod_{X_t=\ell} (X_{t+s})_{s\geq 0} =: \widetilde{X} \sim (\ell, (\lambda_n)_{n\geq 0})\text{-birth process.}$$

To apply the Markov property, recall from our discussion of the Markov propertuy of Poisson processes (see also Fact 137 in the appendix) that the statement of conditional independence is actually equivalent to the following statement. For all $0 \leq r_1 < \cdots < r_n \leq t$, $0 \leq s_1 < \cdots < s_m$, $i_1, \ldots, i_n \in \mathbb{N}$, $j_1, \ldots, j_m \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{P}(X_{r_1} = i_1, \ldots, X_{r_n} = i_n, X_{t+s_1} = j_1, \ldots, X_{t+s_m} = j_m | X_t = \ell) \\ &\qquad = \mathbb{P}(X_{r_1} = i_1, \ldots, X_{r_n} = i_n | X_t = \ell)\mathbb{P}(\widetilde{X}_{s_1} = j_1, \ldots, \widetilde{X}_{s_n} = j_n).\end{aligned}$$

The most useful case is often $m = n = 1$:

$$\mathbb{P}(X_r = i, X_{t+s} = j | X_t = \ell) = \mathbb{P}(X_r = i | X_t = \ell)\mathbb{P}(\widetilde{X}_s = j).$$

In words, we can express the Markov property as "past and future are (conditionally) independent given the present", which we can reformulate as "the past is irrelevant for the future, provided we know the present".

Mathematically, we can reformulate using an argument most economically written, as follows. The conditional independence statement in the Markov property is about three events

$$\text{past } E = \{(X_r)_{r\leq t} \in A\}, \quad \text{future } F = \{(X_{t+s})_{s\geq 0} \in B\} \quad \text{and present } C = \{X_t = \ell\}$$

and states $\mathbb{P}(E \cap F | C) = \mathbb{P}(E|C)\mathbb{P}(F|C)$. This is a special property that does not hold for general events E, F and C. If $\mathbb{P}(E \cap C) > 0$, we can always use the definition of conditional probabilities to obtain

$$\mathbb{P}(E \cap F | C) = \frac{\mathbb{P}(E \cap F \cap C)}{\mathbb{P}(C)} = \frac{\mathbb{P}(F | E \cap C)\mathbb{P}(E \cap C)}{\mathbb{P}(C)} = \mathbb{P}(F | E \cap C)\mathbb{P}(E|C),$$

so that, by comparison, we here have $\mathbb{P}(F|E \cap C) = \mathbb{P}(F|C)$, and we deduce

Corollary 15 (Markov property, alternative formulation) *For all $t \geq 0$, $\ell \geq k$ and sets of paths A and B with $\mathbb{P}(X_t = \ell, (X_r)_{r \leq t} \in A) > 0$, we have*

$$\mathbb{P}((X_{t+s})_{s\geq 0} \in B | X_t = \ell, (X_r)_{r\leq t} \in A) = \mathbb{P}((X_{t+s})_{s\geq 0} \in B | X_t = \ell) = \mathbb{P}((\widetilde{X}_s)_{s\geq 0} \in B),$$

where $\widetilde{X}$ is an $(\ell, (\lambda_n)_{n\geq 0})$-birth process.

In fact, the condition $\mathbb{P}(X_t = \ell, (X_r)_{r\leq t} \in A) > 0$ can often be waived, if the conditional probabilities can still be defined via an approximation by events of positive probability. This is a delicate statement, but it is useful for illustration:

Example 16 Recall in Example 11 we conditioned a $(1, (n\lambda)_{n\geq 0})$-birth process X on T_1 and carefully showed that $\mathbb{E}(X_u | T_1 = t) = 2\mathbb{E}(X_{u-t})$ for all $t \in (0, u)$ using the independence of T_1 and the descendants of the two individuals at time T_1. Note that $T_1 = \inf\{t \geq 0\colon X_t = 2\}$. This means that

$$\{T_1 = t\} = \{X_r = 1, r < t; X_t = 2\} \subset \{X_t = 2\}.$$

If we ignore the fact $\mathbb{P}(T_1 = t) = 0$, the Markov property yields for all $n \geq 0$

$$\mathbb{P}(X_u = n | T_1 = t) = \mathbb{P}(X_u = n | X_r = 1, r < t; X_t = 2) = \mathbb{P}(X_u = n | X_t = 2) = \mathbb{P}(\widetilde{X}_{u-t} = n),$$

where $\widetilde{X}$ is a $(2, (n\lambda)_{n\geq 0})$-birth process. This also gives

$$\mathbb{E}(X_u | T_1 = t) = \sum_{n\in\mathbb{N}} n\mathbb{P}(X_u = n | T = t) = \sum_{n\in\mathbb{N}} n\mathbb{P}(\widetilde{X}_{u-t} = n) = \mathbb{E}(\widetilde{X}_{u-t}) = 2\mathbb{E}(X_{u-t}),$$

since, by model assumption, the families of two individuals evolve completely independently like separate populations starting from one individual each.

3.2 Proof of the Markov property

Proof: The general case $k \geq 0$ follows from the special case $k = 0$ since for X is a $(k, (\lambda_n)_{n\geq 0})$-birth process if and only if $(X_t - k)_{t\geq 0}$ is a $(0, (\lambda_{k+n})_{n\geq 0})$-birth process, and the Markov property of one implies the Markov property of the other.

We now assume $k = 0$. On $\{X_t = \ell\} = \{T_\ell \leq t < T_{\ell+1}\}$ we have

$$\begin{aligned}\widetilde{X}_s &:= X_{t+s} = \#\left\{n \geq 1\colon \sum_{j=0}^{n-1} Z_j \leq t + s\right\} = \ell + \#\left\{n \geq \ell + 1\colon \sum_{j=0}^{n-1} Z_j - t \leq s\right\} \\ &= \ell + \#\left\{m \geq 1\colon \sum_{j=0}^{m-1} \widetilde{Z}_j \leq s\right\},\end{aligned}$$

where $\widetilde{Z}_j = Z_{\ell+j}$, $j \geq 1$, and $\widetilde{Z}_0 = T_{\ell+1} - t$. Therefore, $\widetilde{X}$ has the structure of a birth process starting from ℓ, since given $X_t = \ell$, the $\widetilde{Z}_j \sim \text{Exp}(\lambda_{k+j})$, $j \geq 1$, are independent. For $j = 0$ note that

$$\mathbb{P}(\widetilde{Z}_0 > z | X_t = \ell) = \mathbb{P}(Z_\ell > (t - T_\ell) + z | Z_\ell > t - T_\ell \geq 0) = \mathbb{P}(Z_\ell > z)$$

where we applied the lack of memory property of Z_ℓ to the *independent* threshold $t - T_\ell$. This actually requires a slightly more thorough explanation, since we are dealing with repeated conditioning (first $X_t = \ell$, then $Z_\ell > t - T_\ell$), but we leave this to the reader and only point out that the key result that we need is Lemma 4 below.

This shows that $\widetilde{X}$ is an $(\ell, (\lambda_n)_{n\geq 0}$-birth process. To complete the proof we also have to establish conditional independence from $(X_r)_{r\leq t}$ given $X_t = \ell$. Again by the lack of memory property we see that $\widetilde{Z}_0$ (as well as the $\widetilde{Z}_j$, $j \geq 1$) are conditionally independent of $Z_0, \ldots, Z_{\ell-1}$ given $X_t = \ell$, and the assertion follows. We refer to the optional Exercise A.2.7 for details. □

3.3 The strong Markov property

The Markov property means that at whichever fixed time we inspect our process, information about the past is not relevant for its future behaviour. If we think of an example of radioactive decay, this would allow us to describe the behaviour of the emission process, say, after the experiment has been running for 2 hours. Alternatively, we may wish to reinspect after 1000 emissions. This time is random, but we can certainly carry out any action we wish at the time of the 1000th emission. Such times are called *stopping times.*

We will mainly be interested in stopping times of the form

$$T_{\{n\}} = \inf\{t \geq 0 : X_t = n\} \quad \text{or} \quad T_C = \inf\{t \geq 0 : X_t \in C\}$$

for $n \in \mathbb{N}$ or $C \subset \mathbb{N}$ (or later $C \subset \mathbb{S}$, a countable state space).

Proposition and Fact 17 (Strong Markov property) (i) *Let X be a $(k, (\lambda_n)_{n\geq 0})$-birth process and $T \geq 0$ a stopping time. Then for all $\ell \in \mathbb{N}$ with $\mathbb{P}(X_T = \ell) > 0$, we have that $(X_r)_{r\leq T}$ and $(X_{T+s})_{s\geq 0}$ are conditionally independent given $X_T = \ell$, and the conditional distribution of $(X_{T+s})_{s\geq 0}$ is a $(\ell, (\lambda_n)_{n\geq 0}$-birth process.*

(ii) *Let $X \sim \text{PP}(\lambda)$. Then $(X_{T+s} - X_T)_{s\geq 0} \sim \text{PP}(\lambda)$ starting from 0, independent of T and of $(X_r)_{r\leq T}$, provided that we have one of the following:*

(a) *$T = S$ for a stopping time S,*

(b) *or $T = R$ for a random time R independent of X,*

(c) *or $T = \min(R, S)$ for a stopping time S and an independent time R.*

The proof of the strong Markov property (in full generality) is beyond the scope of this course, but we will use the result from time to time. See the Appendix of Norris's book for details. The proof of the strong Markov property for first hitting times T_i is straightforward since then $\mathbb{P}(X_{T_i} = i) = 1$, so the only relevant statement is for $\ell = i$, and $(X_r)_{r\leq T_i}$ can be expressed in terms of Z_j, $0 \leq j \leq i - 1$, and $(X_{T+s} - X_T)_{s\geq 0}$ in terms of Z_j, $j \geq i$. Specifically, conditioning on $\{X_{T_i} = i\}$ is like not conditioning at all, because this event has probability 1. Furthermore, it is enough to establish independence of holding times, because "can be expressed in terms of" is actually "is a measurable function G of", where the first function e.g. goes from $[0, \infty)^i$ to a space

$$\widetilde{\mathbb{X}} = \{f : [0, t] \to \mathbb{N},\ f \text{ right-continuous}, t \geq 0\}.$$

Now, there is a general result saying that if $G \colon \mathbb{A} \to \mathbb{X}$ and $H \colon \mathbb{B} \to \mathbb{Y}$ are (measureable) functions, and A and B are independent random variables in $\mathbb{A}$ and $\mathbb{B}$, then $G(A)$ and $H(B)$ are also independent. To prove this using our definition of independence, just note that for all $E \subset \mathbb{X}$ and $F \subset \mathbb{Y}$ (measurable), we have

$$\begin{aligned}
\mathbb{P}(G(A) \in E, H(B) \in F) &= \mathbb{P}(A \in G^{-1}(E), B \in H^{-1}(F)) \\
&= \mathbb{P}(A \in G^{-1}(E))\mathbb{P}(B \in H^{-1}(F)) \\
&= \mathbb{P}(G(A) \in E)\mathbb{P}(H(B) \in F).
\end{aligned}$$

A more formal definition of stopping times is as follows.

Definition 18 (Stopping time) A random time T taking values in $[0, \infty]$ is called a stopping time for a continuous-time process $X = (X_t)_{t\geq 0}$ if, for all $t \geq 0$, the event $\{T \leq t\}$ can be expressed (in a measurable way) in terms of $(X_t)_{r\leq t}$.

This definition makes sense for processes X that are not simple birth processes. In the next lecture we have several independent processes on the same time scale. The easiest example is two independent birth processes $X = (X^{(1)}, X^{(2)})$ modelling e.g. two populations that we observe simultaneously.

Example 19 1. Let X be a simple birth process starting from $X_0 = 0$. Then for all $i \geq 1$, $T_i = \inf\{t \geq 0 : X_t = i\}$ is a stopping time since $\{T \leq t\} = \{\exists s \leq t : X_s = i\} = \{X_t \geq i\}$ (the latter equality uses the property that birth processes do not decrease; thus, strictly speaking, this equality is to mean that the two events differ by sets of probability zero in the sense that we write $E = F$ if $\mathbb{P}(E \setminus F) = \mathbb{P}(F \setminus E) = 0$). T_i is called the *first hitting time* of i. Clearly, for X modelling a Geiger counter and $i = 1000$, we are in the situation of our motivating example.

2. Let X be a simple birth process. Then for $\varepsilon > 0$, the random time $T_\varepsilon = \inf\{T_i \geq T_1 : T_i - T_{i-1} < \varepsilon\}$, i.e. the first time that two births have occurred within time at most ε of one another, is a stopping time.

In general, the first time that something happens, or that several things have happened successively, is a stopping time. It is essential that we don't have to look ahead to decide. In particular, the last time that something happens, e.g. the last birth time before time t, is not a stopping time, and the statement of the strong Markov property is usually wrong for such times.

Lecture 4

Continuous-time Markov chains

Reading: Norris 2.1, 2.6
Further reading: Grimmett-Stirzaker 6.9; Ross 6.1-6.3; Norris 2.9

In this lecture, we generalise the notion of a birth process to allow deaths and other transitions. We will allow, in principle, transitions between any two states in a countable state space $\mathbb{S}$, just as for discrete-time Markov chains.

Continuous-time Markov chains are similar in many respects to discrete-time Markov chains, but there are also important differences. Roughly, we will spend Lectures 5 and 6 exploring the differences and tools to handle these, then we will turn to similarities in Lectures 7 and 8.

4.1 Definition and terminology

A (ν, Π)-Markov chain $(M_n, n \geq 0)$ is a discrete-time Markov chain on a countable state space $\mathbb{S}$ with initial distribution $\nu = (\nu_i)_{i \in \mathbb{S}}$ and transition matrix $\Pi = (\pi_{i,j})_{i,j \in \mathbb{S}}$. Its distribution is given by

$$\mathbb{P}(M_0 = i_0, \ldots, M_n = i_n) = \nu_{i_0} \pi_{i_0,i_1} \cdots \pi_{i_{n-1},i_n}, \qquad i_0, \ldots, i_n \in \mathbb{S}, n \geq 1.$$

Recall also that for $E \sim \text{Exp}(1)$ and $\lambda \in (0, \infty)$, we have $Z = E/\lambda \sim \text{Exp}(\lambda)$. We will extend the family of exponential distributions and use the convention $Z = E/\lambda = \infty$ when $\lambda = 0$.

We are going to see three equivalent definitions of a continuous-time Markov chain with countable state space $\mathbb{S}$. The first one is that it is a discrete-time Markov chain which spends a random amount of time in each state.

Definition 20 (Jump-chain and holding time definition) Let $(M_n)_{n \geq 0}$ be a (ν, Π)-Markov chain. Let $\lambda_i \geq 0$, $i \in \mathbb{S}$, and $(Z_n)_{n \geq 0}$ such that conditionally given $M_0 = i_0, \ldots, M_n = i_n$, we have $Z_j \sim \text{Exp}(\lambda_{i_j})$, $0 \leq j \leq n$, independent, for all $i_0, \ldots, i_n \in \mathbb{S}$ and $n \geq 0$. Define

$$X_t = \begin{cases} M_n, & T_n \leq t < T_{n+1}, n \geq 0, \\ \infty, & T_\infty \leq t < \infty, \qquad \text{if } T_\infty < \infty, \end{cases}$$

where $T_0 = 0$, $T_n = Z_0 + \cdots + Z_{n-1}$, $n \geq 1$. Then $(X_t)_{t \geq 0}$ is called *minimal continuous-time Markov chain with initial distribution ν, jump probabilities $(\pi_{ij})_{i,j \in \mathbb{S}}$ and holding rates $(\lambda_i)_{i \in \mathbb{S}}$*.

Remark 21

1. *In other words, when in state $i \in \mathbb{S}$ the chain waits for an $\sim$Exp(λ_i) time and then jumps to a new state chosen according to the transition probabilities $\Pi = (p_{i,j})$.*
2. *The qualifier "minimal" is only relevant when $\mathbb{P}(T_\infty < \infty) > 0$, i.e. when X* can explode. *We will essentially always work with continuous-time Markov chains with $\mathbb{P}(T_\infty < \infty) = 0$, i.e. $\mathbb{P}(T_\infty = \infty) = 1$. We studied the explosion phenomenon in Lecture 3 for the special case of a birth process. A "non-minimal" continuous-time Markov chain would be one that is not absorbed in a state ∞ at time T_∞, but may continue to evolve in $\mathbb{S}$ after T_∞, if $T_\infty < \infty$. Clearly, this requires further specification – this is beyond this course in any generality. We will include the qualifier "minimal" whenever we make general statements that may fail for non-minimal chains. We will omit the qualifier "minimal" when we already know that $\mathbb{P}(T_\infty = \infty) = 1$.*

3. *We may write $Z_n \sim \text{Exp}(\lambda_{M_n})$ given M_n as short-hand for $Z_n \sim \text{Exp}(\lambda_i)$ conditionally given $M_n = i$, for all $i \in \mathbb{S}$. Similarly, Definition 20 says that $Z_n \sim \text{Exp}(\lambda_{M_n})$, $n \geq 0$, are "conditionally independent given $(M_n)_{n\geq 0}$". This is informal, since while $\mathbb{S}$ is countable, the space $\mathbb{S}^{\mathbb{N}}$ of sequences in $\mathbb{S}$ where $(M_n)_{n\geq 0}$ takes its values, is not countable, and Definition 136 in the Annex of conditional independence does not apply. Definition 136 in the Annex does apply to finite sequences in $\mathbb{S}$, and gives rigorous meaning to Definition 20.*

Example 22 (Simple birth processes) For $(k, (\lambda_n)_{n\geq 0})$-birth processes, we have $M_n = k+n$ deterministic, i.e. $\nu_k = 1$, $\pi_{i,i+1} = 1$. Conditional independence of the Z_n follows since they are unconditionally independent to start with, and $\text{Exp}(\lambda_{M_n}) = \text{Exp}(\lambda_{k+n})$ is the unconditional distribution of Z_n. Hence, simple birth processes are minimal continuous-time Markov chains.

Observe that there is no loss of generality in supposing that $\pi_{i,i} \in \{0, 1\}$ for all $i \in \mathbb{S}$, and that $\lambda_i = 0$ if and only if $\pi_{i,i} = 1$. Indeed, we can always modify Π and (λ) without changing the behaviour of the chain to have those properties:

- If, $\pi_{i,i} \in (0, 1)$, then not all "jump times" T_n are jump times. More precisely, starting from $M_0 = i$ we have $G = \inf\{n \geq 1\colon M_n \neq i\} \sim \text{geom}(1 - \pi_{i,i})$ and $Z_0 + \cdots + Z_{G-1} \sim \text{Exp}(\pi_{i,i}\lambda_i)$, so we can define $\widetilde{\lambda}_i = \pi_{i,i}\lambda_i$ and $\widetilde{\pi}_{i,i} = 0$, $\widetilde{\pi}_{i,j} = \pi_{i,j}/(1 - \pi_{i,i})$, $j \neq i$,
- if $\lambda_i = 0$ while $\pi_{i,i} \neq 1$, we can define $\widetilde{\pi}_{i,i} = 1$ and $\widetilde{\pi}_{i,j} = 0$, $j \neq i$

It is customary to represent the transition probabilities $\pi_{i,j}$ and the holding rates λ_i in a single matrix, called the Q-matrix $Q = (q_{i,j})_{i,j\in\mathbb{S}}$, as follows.

Definition 23 (Q-matrix) A Q-matrix on $\mathbb{S}$ is any matrix $Q = (q_{i,j} : i, j \in \mathbb{S})$ which satisfies the following conditions

1. $0 \leq -q_{i,i} < \infty$ for all i;
2. $q_{i,j} \geq 0$ for all $i, j \in \mathbb{S}$
3. $\sum_{j\in\mathbb{S}} q_{i,j} = 0$ for all $i \in \mathbb{S}$.

Definition 24 (Q-matrix and continuous-time Markov chains) *Given jump probabilities $(\pi_{ij})_{i,j\in\mathbb{S}}$ and holding rates $(\lambda_i)_{i\in\mathbb{S}}$ satisfying $\pi_{i,i} \in \{0, 1\}$ and $\lambda_i = 0 \iff \pi_{i,i} = 1$, the matrix Q defined by*

$$q_{i,j} = \lambda_i \pi_{i,j}, \quad j \in \mathbb{S}, j \neq i \qquad \textit{and} \qquad q_{i,i} = -\lambda_i, \quad i \in \mathbb{S}$$

is a Q-matrix. Conversely, we can express $(\lambda_i)_{i\in\mathbb{S}}$ and Π in terms of Q, as follows:

$$\lambda_i = -q_{i,i}, \qquad \pi_{i,i} = \begin{cases} 0 & \textit{if } \lambda_i > 0, \\ 1 & \textit{if } \lambda_i = 0, \end{cases} \qquad \pi_{i,j} = \begin{cases} q_{i,j}/\lambda_i & \textit{if } \lambda_i > 0, \\ 0 & \textit{if } \lambda_i = 0. \end{cases}$$

A (minimal) continuous-time Markov chain X with initial distribution ν, jump probabilities $(\pi_{ij})_{i,j\in\mathbb{S}}$ and holding rates $(\lambda_i)_{i\in\mathbb{S}}$ satisfying $\pi_{i,i} \in \{0, 1\}$ and $\lambda_i = 0 \iff \pi_{i,i} = 1$ is called a (ν, Q)-Markov chain.

Remark 25 We have $q_{i,i} = -\sum_{j\in\mathbb{S}\colon j\neq i} q_{i,j}$ for each $i \in \mathbb{S}$, since either $\sum_{j\in\mathbb{S}\colon j\neq i} \pi_{i,j} = 1$ or $\lambda_i = 0$. As a consequence, the row sums of a Q-matrix vanish.

Often, a continuous-time Markov chain will start from a fixed state i_0. We use notation

$$\nu_i = \delta_{i,i_0} = \begin{cases} 1 & i = i_0 \\ 0 & i \neq i_0 \end{cases} \qquad \text{or short} \qquad \nu = \delta_{i_0},$$

where δ_{i,i_0} with two indices, here i and i_0, is called the Kronecker delta, while δ_{i_0} as a distribution only charging one point, here i_0, is called the Dirac delta.

Example 26 (Simple birth processes) For a $(k, (\lambda_n)_{n\geq 0})$-birth processes, we obtain

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

A $(k, (\lambda_n)_{n\geq 0})$-birth process is a (δ_k, Q)-Markov chain.

4.2 Second construction

Firs we need a key result about sequences of Exponential random variables and their minimum: Recall that $Z = E/\lambda \sim \text{Exp}(\lambda)$ for $E \sim \text{Exp}(1)$. If $\lambda = 0$, this means that $\mathbb{P}(Z = \infty) = 1$.

Lemma 27 (Theory of competing exponentials) *Let $\alpha_j \geq 0$, $j \geq 0$, with $\lambda := \sum_{j\geq 0} \alpha_j < \infty$. Let $C_j \sim \text{Exp}(\alpha_j)$, $j \geq 0$. Then $W = \min\{C_j, j \geq 0\} \sim \text{Exp}(\lambda)$. If $\lambda > 0$, let $M = j$ if $W = C_j$. Then M is independent of W with distribution $\mathbb{P}(M = j) = \alpha_j/\lambda$. Furthermore, conditionally given $M = j$, the residual times $C_k - W$ are independent $\text{Exp}(\alpha_k)$, $k \neq j$.*

Proof: First separately

$$\mathbb{P}(W > t) = \mathbb{P}(C_j > t, j \geq 0) = \prod_{j=0}^{\infty} \mathbb{P}(C_j > t) = \exp\left(-\sum_{j=0}^{\infty} q_{i,j} t\right) = \exp(-\lambda t).$$

so $W \sim \text{Exp}(\lambda)$. Similarly $V_j = \min\{C_k, k \neq j\} \sim \text{Exp}(\lambda - \alpha_j)$.

$$\mathbb{P}(M = j) = \mathbb{P}(C_j < V_j) = \frac{\alpha_j}{\alpha_j + (\lambda - \alpha_j)} = \alpha_j/\lambda.$$

For independence and residual times, we refer to the argument in Exercise A.1.5, which extends to the present setting, using appropriate vector notation for the overshoots $C_k - W$. □

Proposition 28 (Jump rates construction) *Start with an initial state $X_0 = Y_0$ with distribution ν. Let $(S_n^i,: n \geq 1, i \in \mathbb{S})$ be a collection of independent Exponential variables of parameter 1. Then inductively for $n = 0, 1, 2, \ldots$, if $Y_n = i$ we set*

$$Z_{n+1}^j = S_{n+1}^j / q_{i,j} \text{ for } j \neq i,$$
$$Z_{n+1} = \inf_{j\neq i} Z_{n+1}^j,$$
$$Y_{n+1} = \begin{cases} j \text{ if } Z_{n+1} = Z_{n+1}^j, \\ i \text{ if } Z_{n+1} = \infty. \end{cases}$$

For $n = 1, 2, \ldots$ set $T_n = Z_1 + Z_2 + \ldots + Z_n$ and for $t \geq 0$ let

$$X_t = \begin{cases} Y_n \text{ if } T_n \leq t < T_{n+1} \text{ for some } n \\ \infty \text{ otherwise.} \end{cases}$$

Then $(X_t : t \geq 0)$ is a (ν, Q)-Markov chain.

Proof: It is a straightforward consequence of Lemma 27 that $(Y_n, n \geq 0)$ is a discrete time (ν, Π) Markov chain. Then conditionally on $(Y_n, n \geq 0)$, the same Lemma shows that the holding times Z_n are independent exponential with parameter λ_{Y_n}. □

4.3 Third construction

Our third and final construction of a Markov chain with generator matrix Q and initial distribution and ν is based on the Poisson process. Imagine the state-space $\mathbb{S}$ as a labyrinth of chambers and passages, each passage shut off by a single door which opens briefly from time to time to allow you through in one direction only. Suppose the door giving access to chamber j from chamber i opens at the jump times of a Poisson process of rate $q_{i,j}$ and you take every chance to move that you can, then you will perform a Markov chain with Q-matrix Q. In more mathematical terms,

Proposition 29 (Poisson processes construction) *Start with an initial state $X_0 = Y_0$ with distribution ν, and with a family of independent Poisson processes $\{(N_t^{ij}, t \geq 0), i, j \in \mathbb{S}, i \neq j\}$ having respective rates $q_{i,j}$. Then set $T_0 = 0$ and define inductively for $n = 0, 1, 2, \ldots$*

$$T_{n+1} = \inf\{t > T_n : N_t^{Y_n j} \neq N_{T_n}^{Y_n j} \textit{ for some } j \neq Y_n$$

and

$$Y_{n+1} = \begin{cases} j \textit{ if } T_{n+1} < \infty \textit{ and } N_{T_n}^{Y_n j} \neq N_{T_{n+1}}^{Y_n j}. \end{cases}$$

For $t \geq 0$ let

$$X_t = \begin{cases} Y_n \textit{ if } T_n \leq t < T_{n+1} \textit{ for some } n \\ \infty \textit{ otherwise.} \end{cases}$$

Then $(X_t : t \geq 0)$ is a (ν, Q)-Markov chain.

Proof: We have to check that the process defined here has the correct jump chain, holding times and dependence structure. Clearly $X_0 = M_0$ has the right initial distribution. Given $M_0 = i_0$, the first jump occurs at the first time at which one of the Poisson processes $N^{i_0 j}$, $j \neq i_0$, has its first jump. This time is a minimum of independent exponential random variables of parameters $q_{i_0 j}$, $T_1 = \inf\{T_1^{i_0 j}, j \neq i_0\}$. By Lemma 27, (M_0, Z_0, M_1) have the distribution as required, and the post-T_1 Poisson processes $(N_{T_1+s}^{i_0 j} - N_{T_1}^{i_0 j})_{s \geq 0}$ are independent of (M_0, Z_0, M_1). It is easy to see that the post-T_1 Poisson processes $(N_{T_1+s}^{ij} - N_{T_1}^{ij})_{s \geq 0}$ for $i \neq i_0$, $j \neq i$, are also independent Poisson processes since T_1 is independent of these N^{ij}. Inductively, assuming we have $(M_0, Z_0, \ldots, M_{n-1}, Z_{n-1}, M_n)$, as required, and independent $(N_{T_n+s}^{ij} - N_{T_n}^{ij})_{s \geq 0}$, the same argument applies to give $(M_0, Z_0, \ldots, M_n, Z_n, M_{n+1})$ independent of $(N_{T_{n+1}+s}^{ij} - N_{T_{n+1}}^{ij})_{s \geq 0}$. This completes the induction step and hence the proof. □

4.4 Markov property for continuous-time chains

Corollary 30 (Markov property and strong Markov property) *Let X be a (ν, Q)-Markov chain, $t \geq 0$ a fixed time and $\ell \in \mathbb{S}$. Then*

$$(X_r)_{r \leq t} \coprod_{X_t = \ell} (X_{t+s})_{s \geq 0} =: \widetilde{X} \sim (\delta_\ell, Q)\textit{-Markov chain.}$$

Let T be a stopping time and $\ell \in \mathbb{S}$. Then

$$(X_r)_{r \leq T} \coprod_{X_T = \ell} (X_{T+s})_{s \geq 0} =: \widetilde{X} \sim (\delta_\ell, Q)\textit{-Markov chain.}$$

Proof: The post-t Poisson processes $\widetilde{N}^{ij} = (N_{t+s}^{ij} - N_t^{ij})_{s \geq 0}$ are Poisson processes independent of the pre-t Poisson processes $(N_r^{ij})_{0 \leq r \leq t}$. The pre-$t$ process $(X_r)_{r \leq t}$ is a function of X_0 and of the pre-t Poisson processes, while the post-t process $\widetilde{X}$ is constructed from $\widetilde{X}_0 = X_t$ and $(\widetilde{N}^{ij})_{i,j \in \mathbb{S}}$ just as X is constructed from X_0 and $(N^{ij})_{i,j \in \mathbb{S}}$ in Proposition 29. Hence, we have conditional independence given $X_t = \ell$, as required, and since $\widetilde{X}_0 \sim \delta_\ell$ conditionally given $X_t = \ell$, the process $\widetilde{X}$ is a (δ_ℓ, Q)-Markov chain conditionally given $X_t = \ell$.

For $T = T_n$, $n \geq 1$, and for first passage times such as $H_\ell = \inf\{t \geq T_1 \colon X_t = \ell\}$ and nth passage times of ℓ, $n \geq 1$, this proof can be adapted in combination with an induction in n and the strong Markov property of N^{ij}. The general case is beyond the scope of this course. □

4.5 Examples: M/M/1 and M/M/s queues

In applications, the model description is usually of the following form. There is an initial distribution ν on a state space $\mathbb{S}$. When in state $i \in \mathbb{S}$, there is a set of "neighbouring" states into which transitions are possible, and for each such state $j \in \mathbb{S}$, we have one or more independent exponential clocks with total rate $q_{i,j}$ that trigger a transition into j. We can build a Q-matrix $Q = (q_{i,j})_{i,j \in \mathbb{S}}$ directly from the rates of the exponential clocks, also setting $q_{i,i} = -\sum_{j \in \mathbb{S} \colon j \neq i} q_{i,j}$. To show that the process is a (ν, Q)-Markov chain, we have to show that *the system starts afresh* in the new state after each transition, from i to j, say. What this means is that the model assumptions and the theory of competing exponentials yield a full set of clocks (residual or new) for the new state j, which is conditionally independent of the past given the present state.

This part of the proof is different for each model and therefore needs to be repeated for each model. The formalisation to actually establish the joint distributions as required for Definition 20 is rather technical and always essentially the same. Since it adds little insight into the model (but some more insight into arguments involving conditional independence), we will only provide this argument once (postponed to the non-examinable Section 4.6).

Example 31 (M/M/1 queue) Let us denote by X_t the number of customers in a single-server queueing system at time $t \geq 0$, including any customer currently being served, where we assume that new customers arrive according to $\mathrm{PP}(\lambda)$, and that service times are independent $\mathrm{Exp}(\mu)$.

Given a queue size of $i \geq 1$, two transitions are possible. If a customer arrives (at rate λ), then X increases to $i+1$. If the customer being served leaves (at rate μ), then X decreases to $i-1$. Given a queue size of $i = 0$, only the former can happen. This amounts to a Q-matrix

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -\mu-\lambda & \lambda & 0 & \cdots \\ 0 & \mu & -\mu-\lambda & \lambda & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

If $X_0 = M_0 \sim \nu$, then X is indeed a (ν, Q)-Markov chain, by the following induction step. Given state $M_n = i_n \geq 2$ and two independent $\mathrm{Exp}(\lambda)$ and $\mathrm{Exp}(\mu)$ clocks ticking, the theory of competing exponential clocks (Exercise A.1.5 or Lemma 27) shows that the system starts afresh in $i_n + 1$ (respectively $i_n - 1$) with the residual (respectively new) $\mathrm{Exp}(\mu)$ service clock and the new (respectively residual) $\mathrm{Exp}(\lambda)$ inter-arrival clock independent of the past. The case $i_n = 1$ is similar, except that for $i_n - 1 = 0$ no new service clock is set up. Also, for $i_n = 0$, no service clock is ticking so the transition is always to $i_n + 1$. This completes the induction step.

Example 32 (M/M/s queue) If there are $s \geq 1$ servers in the system, the rate at which customers leave is s-fold, provided there are at least s customers. We obtain the Q-matrix

$$\begin{pmatrix} -\lambda & \lambda & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \mu & -\mu-\lambda & \lambda & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 2\mu & -2\mu-\lambda & \lambda & \ddots & 0 & 0 & 0 & \ddots \\ 0 & 0 & 3\mu & -3\mu-\lambda & \ddots & 0 & 0 & 0 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & \ddots & -s\mu-\lambda & \lambda & 0 & \ddots \\ 0 & 0 & 0 & 0 & \ddots & s\mu & -s\mu-\lambda & \lambda & \ddots \\ 0 & 0 & 0 & 0 & \ddots & 0 & s\mu & -s\mu-\lambda & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

which is better written as $q_{i,i+1} = \lambda$, $i \geq 0$; $q_{i,i-1} = i\mu$, $1 \leq i \leq s$; $q_{i,i-1} = s\mu$, $i \geq s$; $q_{i,i} = -i\mu - \lambda$, $0 \leq i \leq s$; $q_{ii} = -s\mu - \lambda$, $i \geq s$; $q_{ij} = 0$ otherwise. A slight variation of the argument for Example 31 shows that the M/M/s queue is a continuous-time Markov chain.

4.6 Checking joint distributions in Definition 20

This section is non-examinable. We demonstrate how to formalise the induction to fully prove that M/M/1-queues (or other processes) are (ν, Q)-Markov chains in the sense of Definitions 20 and 24.

Induction hypothesis: $(M_0, \ldots, M_n) \sim (\nu, \Pi)$-Markov chain and $Z_j \sim \text{Exp}(\lambda_{i_j})$ independent conditionally given $M_0 = i_0, \ldots, M_n = i_n$, with times $A_n \sim \text{Exp}(\lambda)$ to next arrival after T_n and $B_n \sim \text{Exp}(\mu)$ to next completed service after T_n (if $i_n \geq 1$) independent conditionally given $M_n = i_n$.

Suppose $i_n \geq 2$. By induction hypothesis, model specification and theory of competing exponentials,

$$\begin{aligned} &\mathbb{P}(M_0 = i_0, Z_0 > z_0, M_1 = i_1, \ldots, Z_{n-1} > z_{n-1}, M_n = i_n, Z_n > z_n, M_{n+1} = i_n + 1, A_{n+1} > s, B_{n+1} > t) \\ &= \mathbb{P}(M_0 = i_0, Z_0 > z_0, M_1 = i_1, \ldots, Z_{n-1} > z_{n-1}, M_n = i_n)\mathbb{P}(A_{n+1} > s, B_n - t > A_n > z_n | M_n = i_n) \\ &= \left(\nu_{i_0} e^{-\lambda_{i_0} z_0} \pi_{i_0,i_1} \cdots e^{-\lambda_{i_{n-1}} z_{n-1}} \pi_{i_{n-1},i_n}\right) \left(\pi_{i_n,i_n+1} e^{-(\lambda+\mu) z_n} e^{-\lambda s} e^{-\mu t}\right). \end{aligned}$$

Arguments for $M_{n+1} = i_n - 1$, and for $i_n = 1$ and $i_n = 0$ are similar. This completes the induction step.

Lecture 5

Transition probabilities

Reading: Norris 2.8, 3.1
Further reading: Grimmett-Stirzaker 6.8 (12)-(17), 6.9; Ross 6.4; Norris 2.7, 2.10

In this lecture we establish transition matrices $P(t)$, $t \geq 0$, for continuous-time Markov chains. This family of matrices gives the analogues of n-step transition matrices $\Pi^n = (\pi_{ij}^{(n)})_{i,j\in\mathbb{S}}$, $n \geq 0$, for discrete-time Markov chains.

5.1 The semigroup property of transition matrices

As a consequence of the Markov property of continuous-time Markov chains, the probabilities $\mathbb{P}(X_{t+s} = j | X_t = i)$ do not depend on t. We define

$$p_{ij}(s) := \mathbb{P}(X_{t+s} = j | X_t = i) \quad \text{and} \quad P(s) = (p_{ij}(s))_{i,j\in\mathbb{S}},$$

the time-s transition probabilities and time-s transition matrix.

Example 33 For a Poisson process with rate λ, we have for $n \geq 0$ and $j \geq i \geq 0$

$$p_{i,i+n}(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \qquad \text{so that} \qquad p_{ij}(t) = \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t}$$

by Remark 5. For fixed $t \geq 0$ and $i \geq 0$, these are Poi(λt) probabilities, shifted by i.

Proposition 34 *$\{P(t), t \geq 0\}$ is a semigroup. Specifically, we have $P(t)P(s) = P(t+s)$ in the sense of matrix multiplication for all $t, s \geq 0$, and $P(0) = I$, the identity matrix.*

Proof: Just note that for all $i, k \in \mathbb{S}$

$$\begin{aligned} p_{ik}(t+s) &= \sum_{j\in\mathbb{S}} \mathbb{P}(X_{t+s} = k, X_t = j | X_0 = i) \\ &= \sum_{j\in\mathbb{S}} \mathbb{P}(X_t = j | X_0 = i)\mathbb{P}(X_{t+s} = k | X_t = j, X_0 = i) = \sum_{j\in\mathbb{S}} p_{ij}(t)p_{jk}(s) \end{aligned}$$

where we applied the Markov property. □

If $t \mapsto P(t)$ was a real-valued function, the functional equation $P(t)P(s) = P(t+s), P(0) = 1$ would have solutions $P(t) = e^{tA}$ for some real A. In the matrix-valued case, when $\mathbb{S}$ is finite, this is still true, if we define $e^{tA} = \sum_{n\geq 0} t^n A^n/n!$ where A^n is a matrix power, and the series and multiplication by scalars are taken component-wise. Furthermore, component-wise differentiation yields $P'(t) = P(t)A$. We will see that the matrix A is in fact the Q-matrix $A = Q$.

But first, we note that we can express finite-dimensional marginal distributions in terms of transition probabilities:

Corollary 35 *For a (ν, Q)-Markov chain X and any $0 = t_0 < t_1 < \cdots < t_n$ and $i_0, \ldots, i_n \in \mathbb{S}$, we have*

$$\mathbb{P}(X_{t_0} = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = \nu_{i_0} \prod_{j=1}^{n} p_{i_{j-1},i_j}(t_j - t_{j-1}).$$

It will be useful to use notation $\mathbb{P}_i(\cdot)$ if we have a fixed initial state $X_0 = i$.

5.2 Backward equations

Proposition 36 *The transition matrices $(P(t))_{t\geq 0}$ of a minimal (ν, Q)-Markov chain satisfy the backward equation*

$$P'(t) = QP(t)$$

with initial condition $P(0) = I$, the identity matrix.

Given Q, $(P(t))_{t\geq 0}$ is the unique solution to the backward equation if $\mathbb{P}(T_\infty = \infty) = 1$, in particular if $\mathbb{S}$ is finite.

Furthermore, if $\mathbb{P}(T_\infty < \infty) > 0$, then $(P(t))_{t\geq 0}$ is the minimal nonnegative solution in the sense that all other nonnegative solutions $(\widetilde{P}(t))_{t\geq 0}$ satisfy $\widetilde{p}_{ik}(t) \geq p_{ik}(t)$ for all $i, k \in \mathbb{S}$, $t \geq 0$.

Proof: We first show that $(P(t))_{t\geq 0}$ solves $P'(t) = QP(t)$, i.e. for all $i, k \in \mathbb{S}$, $t \geq 0$

$$p'_{ik}(t) = \sum_{j\in\mathbb{S}} q_{ij} p_{jk}(t).$$

We start by a one-step analysis (using the strong Markov property at the first jump time T_1, or directly identifying the structure of the post-T_1 process) to get

$$\begin{aligned} p_{ik}(t) &= \mathbb{P}_i(X_t = k) = \int_0^\infty \mathbb{P}_i(X_t = k | T_1 = s)\lambda_i e^{-\lambda_i s} ds \\ &= \delta_{ik} e^{-\lambda_i t} + \int_0^t \sum_{j\in\mathbb{S}} \mathbb{P}_i(X_t = k, X_s = j | T_1 = s)\lambda_i e^{-\lambda_i s} ds \\ &= \delta_{ik} e^{-\lambda_i t} + \int_0^t \sum_{j\in\mathbb{S}} \mathbb{P}_i(X_t = k | X_s = j, T_1 = s)\mathbb{P}_i(X_s = j | T_1 = s)\lambda_i e^{-\lambda_i s} ds \\ &= \delta_{ik} e^{-\lambda_i t} + \int_0^t \sum_{j\in\mathbb{S}:\, j\neq i} p_{jk}(t-s)\pi_{ij}\lambda_i e^{-\lambda_i s} ds \\ &= \delta_{ik} e^{-\lambda_i t} + \int_0^t \sum_{j\in\mathbb{S}:\, j\neq i} q_{ij} p_{jk}(u) e^{-\lambda_i (t-u)} du, \end{aligned} \qquad (1)$$

i.e.

$$e^{\lambda_i t} p_{ik}(t) = \delta_{ik} + \int_0^t \sum_{j\in\mathbb{S}:\, j\neq i} q_{ij} p_{jk}(u) e^{\lambda_i u} du.$$

Clearly this implies that p_{ij} is differentiable and we obtain

$$e^{\lambda_i t} p'_{ik}(t) + \lambda_i e^{\lambda_i t} p_{ik}(t) = \sum_{j\in\mathbb{S}:\, j\neq i} q_{ij} p_{jk}(t) e^{\lambda_i t},$$

which after cancellation of $e^{\lambda_i t}$ and by $\lambda_i = -q_{ii}$ is what we require.

The rest of this proof is non-examinable. Suppose now, we have another non-negative solution $\widetilde{p}_{ij}(t)$. Then, by integration, $\widetilde{p}_{ij}(t)$ also satisfies the integral equations (1) (the δ_{ik} come from the initial conditions). Trivially

$$T_0 = 0 \quad \Rightarrow \quad \mathbb{P}_i(X_t = k, t < T_0) = 0 \leq \widetilde{p}_{ik}(t) \quad \text{for all } i, k \in \mathbb{S} \text{ and } t \geq 0.$$

If for some $n \in \mathbb{N}$

$$\mathbb{P}_i(X_t = k, t < T_n) \leq \widetilde{p}_{ik}(t) \quad \text{for all } i, k \in \mathbb{S} \text{ and } t \geq 0,$$

then as above

$$\begin{aligned}\mathbb{P}_i(X_t = k, t < T_{n+1}) &= e^{-q_i t}\delta_{ik} + \int_0^t \sum_{j\in\mathbb{S}:\, j\neq i} q_{ij}\mathbb{P}_j(X_u = k, u < T_n)e^{-\lambda_i(t-u)}du \\ &\leq e^{-q_i t}\delta_{ik} + \int_0^t \sum_{j\in\mathbb{S}:\, j\neq i} q_{ij}\widetilde{p}_{jk}(u)e^{-\lambda_i(t-u)}du = \widetilde{p}_{ik}(t)\end{aligned}$$

and therefore

$$p_{ik}(t) = \lim_{n\to\infty} \mathbb{P}_i(X_t = k, t < T_n) \leq \widetilde{p}_{ik}(t)$$

as required. We conclude that $p_{ik}(t)$ is the minimal non-negative solution to the backward equation.

Finally, if $\mathbb{P}(X_t \in \mathbb{S}) = \mathbb{P}(T_\infty > t) = 1$ for all $t \geq 0$, i.e. if $\mathbb{P}(T_\infty = \infty) = 1$, then the solution is unique since any strictly larger solution would have $\sum_{j\in\mathbb{S}} \widetilde{p}_{ij}(t) > \sum_{j\in\mathbb{S}} p_{ij}(t) = 1$ for some $i \in \mathbb{S}$ and $t \geq 0$, so that $(\widetilde{P}(t))_{t\geq 0} = (\widetilde{p}_{ij}(t))_{t\geq 0}$ is not a family of transition matrices. If $\mathbb{S}$ is finite, then $\overline{\lambda} = \max\{\lambda_i, i \in \mathbb{S}\}$ is finite, and the construction from $Z_n = E_n/\lambda_{M_n}$ from $(M_n, n \geq 0)$ and independent $E_n \sim \text{Exp}(1)$, $n \geq 0$, shows that $T_n \geq (E_0 + \cdots + E_{n-1})/\overline{\lambda} \to \infty$ a.s., by the Strong Law of Large Numbers (stated in Part A, see also Lecture 9 here). □

"Other nonnegative solutions" are related to non-minimal extensions of explosive Markov chains.

Corollary 37 *Let Q be a Q-matrix. Then the backward equation*

$$P'(t) = QP(t), \quad P(0) = I$$

has a minimal non-negative solution $(P(t), t \geq 0)$. This solution forms a semigroup

$$P(s)P(t) = P(s+t) \qquad \textit{for all } s, t \geq 0.$$

5.3 Forward equations

Proposition 38 *The transition matrices $(P(t))_{t\geq 0}$ of a minimal (ν, Q)-Markov chain satisfy the forward equation*

$$P'(t) = P(t)Q$$

with initial condition $P(0) = I$, the identity matrix. Given Q, $(P(t))_{t\geq 0}$ is the unique solution to the forward equation if $\mathbb{P}(T_\infty < \infty) = 0$, in particular if $\mathbb{S}$ is finite.

Furthermore, if $\mathbb{P}(T_\infty < \infty) > 0$, then $(P(t))_{t\geq 0}$ is the minimal nonnegative solution in the sense that any other nonnegative solutions $(\widetilde{P}(t))_{t\geq 0}$ satisfy $\widetilde{p}_{ik}(t) \geq p_{ik}(t)$ for all $i, k \in \mathbb{S}$, $t \geq 0$.

The proof is very similar to the proof we used for the bacward equation, but this time, instead of conditioning on the first jump, we condition on the last jump.

Proof: For the case $\mathbb{S}$ finite, see Exercise A.4.3 for one argument. If $\mathbb{S}$ is infinite, another argument, in fact a variation of the argument of Proposition 36 works, conditioning on the *last* jump before t. Since this is not a stopping time, the Markov property does not apply and certain calculations have to be done "by hand". See Norris Theorem 2.8.6. for details (non-examinable). □

Remark 39 Transition semigroups and the Markov property can form the basis for a definition of continuous-time Markov chains. In order to obtain a definition that is equivalent to Definition 20, we can say that a (ν, Q)-Markov chain is a process such that $t \mapsto X_t$ is right-continuous in $\mathbb{S}$ and such that

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = p_{i_n, i_{n+1}}(t_{n+1} - t_n)$$

for all $0 = t_0 < t_1 < \cdots < t_{n+1}$ and $i_0, \ldots, i_{n+1} \in \mathbb{S}$, where $P(t)$ satisfies the forward equations. See Norris 2.8.

5.4 Examples

Solving forward or backward equations can be done by a computer, since they are just systems of linear ordinary differential equations. The forward equations have the distinct advantage that the unknown $p_{i,\cdot}(\cdot)$ for fixed i appears on both sides of the equation, and (in the non-explosive case), we know that summing transition probabilities $p_{i,j}(t)$ over j gives 1. This "extra equation" can simplify the calculations. As always, the situation is particularly nice for the Poisson process:

Example 40 (Poisson process, forward equations) For the Poisson process $q_{i,i+1} = \lambda$, $q_{i,i} = -\lambda$, $q_{i,j} = 0$ otherwise, hence we have forward equations

$$\begin{aligned} p'_{i,i}(t) &= -p_{i,i}(t)\lambda, && i \in \mathbb{N} \\ p'_{i,i+n}(t) &= p_{i,i+n-1}(t)\lambda - p_{i,i+n}(t)\lambda, && i \in \mathbb{N}, n \geq 1 \end{aligned}$$

and it is easily seen inductively (fix i and proceed $n = 0, 1, 2, \ldots$) that Poisson probabilities

$$p_{i,i+n}(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

are the unique solutions under boundary conditions $p_{i,i+n}(0) = \delta_{n,0}$. Writing $i+n$ rather than j is convenient because of the stationarity of increments in this special case of the Poisson process.

Example 41 (Poisson process, backward equations) Alternatively, we may consider the backward equations for the Poisson process

$$\begin{aligned} p'_{j,j}(t) &= -\lambda p_{j,j}(t), && j \in \mathbb{N} \\ p'_{i,j}(t) &= \lambda p_{i+1,j}(t) - \lambda p_{i,j}(t), && i, j \in \mathbb{N}, j \geq i+1 \end{aligned}$$

and solve inductively (fix j and proceed $i = j, j-1, \ldots, 0$).

We have seen an easier way to derive the Poisson transition probabilities in Remark 5. The link between the two ways is revealed by the passage to probability generating functions

$$G_i(z,t) = \mathbb{E}_i\left(z^{X_t}\right)$$

which then have to satisfy differential equations

$$\frac{\partial}{\partial t} G_i(z,t) = \sum_{n=0}^{\infty} z^{i+n} p'_{i,i+n}(t) = \lambda(z-1)G_i(z,t), \qquad G_i(z,0) = \mathbb{E}_i(z^{X_0}) = z^i.$$

Solutions for these equations are obvious. In general, if we have G_i sufficiently smooth in t and z, we can derive from differential equations for probability generating functions differential equations for moments $m_i(t) = \mathbb{E}_i(X_t) = \frac{\partial}{\partial z} G_i(z,t)\big|_{z=1-}$ that yield here

$$m'_i(t) = \frac{\partial}{\partial z}\frac{\partial}{\partial t} G_i(z,t)\Bigg|_{z=1-} = \lambda\, G_i(z,t)|_{z=1-} = \lambda, \qquad m_i(0) = \mathbb{E}_i(X_0) = i.$$

Often (even in this case), this can be solved more easily than the differential equation for probability generating functions. Together with a similar equation for the variance, we can capture what are often the two most important distributional features of a model.

Lecture 6

The class structure of continuous-time Markov chains

Reading: Norris 3.2-3.5
Further reading: Grimmett-Stirzaker 6.9

In this lecture, we introduce for continuous-time chains the notions of irreducibility and positive recurrence that will be needed for the convergence theorems in Lecture 8.

6.1 Communicating classes and irreducibility

We define the class structure characteristics as for discrete-time Markov chains. We emphasize that from now on we deal only with *minimal* Markov chains, those that die after explosion.

Definition 42 Let X be a continuous-time Markov chain.

(a) We say that $i \in \mathbb{S}$ *leads to* $j \in \mathbb{S}$ and write $i \to j$ if

$$\mathbb{P}_i(X_t = j \text{ for some } t \geq 0) = \mathbb{P}_i(T_{\{j\}} < \infty) > 0, \quad \text{where } T_{\{j\}} = \inf\{t \geq 0 : X_t = j\}.$$

(b) We say $i \in \mathbb{S}$ *communicates with* $j \in \mathbb{S}$ and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

(c) An equivalence class of the equivalence relation $\leftrightarrow$ on $\mathbb{S}$ is called a *(communicating) class.*

(d) A class $A \subset \mathbb{S}$ is *closed* if there is no $i \in A$, $j \in \mathbb{S} - A$ with $i \to j$, i.e. X cannot leave A.

(e) i is an *absorbing state* if $\{i\}$ is a closed class.

(f) X is *irreducible* if $\mathbb{S}$ is a communicating class (the only one).

In the following we denote by $M = (M_n)_{n\geq 0}$ the jump chain and by $(Z_n)_{n\geq 0}$ the holding times that we used in the definition of a continuous-time Markov chain $X = (X_t)_{t\geq 0}$, in Definition 20.

Proposition 43 *Let X be a minimal (i.e. $T_\infty = \infty$ or $X_t = \infty$ for $t \geq T_\infty$) continuous-time Markov chain. For $i, j \in \mathbb{S}$, $i \neq j$, the following are equivalent*

(i) $i \to j$ *for X.*

(ii) $i \to j$ *for the jump chain M.*

(iii) *There is a sequence $(i_0, \ldots, i_n)$, $i_k \in \mathbb{S}$, from $i_0 = i$ to $i_n = j$ such that* $\displaystyle\prod_{k=0}^{n-1} q_{i_k, i_{k+1}} > 0$.

(iv) $p_{i,j}(t) > 0$ *for all $t > 0$.*

(v) $p_{i,j}(t) > 0$ *for some $t > 0$.*

Proof: Implications (iv)⇒(v)⇒(i)⇒(ii) are straightforward.

(ii)⇒(iii): From the discrete-time theory, we know that $i \to j$ for M implies that there is a path $(i_0, \ldots, i_n)$ from i to j with

$$\prod_{k=0}^{n-1} \pi_{i_k,i_{k+1}} > 0, \quad \text{hence} \quad \prod_{k=0}^{n-1} q_{i_k,i_{k+1}} = \prod_{k=0}^{n-1} \pi_{i_k,i_{k+1}} \lambda_{i_k} > 0$$

since $\lambda_m = 0$ if and only if $\pi_{m,m} = 1$.

(iii)⇒(iv) If $q_{i,j} > 0$, then we can get a lower bound for $p_{i,j}(t)$ by only allowing one transition in $[0, t]$ by

$$\begin{aligned} p_{i,j}(t) &\geq \mathbb{P}_i(Z_0 \leq t, M_1 = j, Z_1 > t) \\ &= \mathbb{P}_i(Z_0 \leq t)\mathbb{P}_i(M_1 = j)\mathbb{P}(Z_1 > t | M_1 = j) \\ &= (1 - e^{-\lambda_i t})\pi_{i,j} e^{-\lambda_j t} > 0 \end{aligned}$$

for all $t > 0$, hence in general for the path $(i_0, \ldots, i_n)$ given by (iii)

$$\begin{aligned} p_{i,j}(t) &= \mathbb{P}_i(X_t = j) \geq \mathbb{P}_i(X_{kt/n} = i_k \text{ for all } k = 1, \ldots, n) \\ &= \prod_{k=0}^{n-1} p_{i_k,i_{k+1}}(t/n) > 0 \end{aligned}$$

for all $t > 0$. For the last equality, we used the Markov property which implies that for all $m = 1, \ldots, n$

$$\begin{aligned} \mathbb{P}(X_{mt/n} = i_m | X_{kt/n} = i_k \text{ for all } k = 0, \ldots, m-1) &= \mathbb{P}(X_{mt/n} = i_m | X_{(m-1)t/n} = i_{m-1}) \\ &= p_{i_{m-1},i_m}(t/n). \end{aligned}$$

□

Condition (iv) shows that the situation is simpler than in discrete-time where it may be possible to reach a state, but only after a certain length of time, and then only periodically.

Example 44 (M/M/1 queue) The M/M/1 queue with $q_{i,i+1} = \lambda > 0$ and $q_{i+1,i} = \mu > 0$, $i \geq 0$, is irreducible: for all $m > n \geq 0$, we have $q_{m,m-1} \cdots q_{n+1,n} = \mu^{m-n} > 0$ and $q_{n,n+1} \cdots q_{m-1,m} = \lambda^{m-n} > 0$ and Proposition 43 yields $m \leftrightarrow n$.

Example 45 (Simple birth-and-death process) The simple birth and death process with

$$q_{i,i+1} = i\lambda, \quad q_{i,i-1} = i\mu, \quad q_{i,i} = -(\lambda + \mu)i, \quad q_{i,j} = 0 \text{ otherwise}, \qquad i \geq 0,$$

has 0 as an absorbing state, while $\mathbb{N} \setminus \{0\}$ is an open communicating class by the same argument as in Example 44.

6.2 Recurrence and transience, positive and null recurrence

Definition 46 Let X be a continuous-time Markov chain.

(a) $i \in \mathbb{S}$ is called *recurrent* if $\mathbb{P}_i(\{t \geq 0 \colon X_t = i\} \text{ is unbounded}) = 1$.

(b) $i \in \mathbb{S}$ is called *transient* if $\mathbb{P}_i(\{t \geq 0 \colon X_t = i\} \text{ is unbounded}) = 0$.

X is called *recurrent (transient)* if all states $i \in \mathbb{S}$ are recurrent (transient) for X.

Remark 47 If X can explode starting from i and if X is a minimal continuous-time Markov chain, then i is certainly not recurrent. Hence, whenever we assume minimality and recurrence, this implies that X is non-explosive.

We denote by $N_i = \inf\{n \geq 1 : M_n = i\}$ the *first passage time of M to state i*, and by

$$H_i = T_{N_i} = \inf\{t \geq T_1 : X_t = i\},$$

the *first passage time of X to state i*. Note that we require the chain to do at least one jump. This is to force X to leave i first if $X_0 = i$. We also define the *successive passage times* $N_i^{(1)} = N_i$ and $N_i^{(m)} = \inf\{n > N_i^{(m-1)} : M_n = i\}$, $m \geq 2$, and $H_i^{(m)} = T_{N_i^{(m)}}$, $m \geq 1$.

Proposition 48 *$i \in \mathbb{S}$ is recurrent (transient) for a minimal continuous-time Markov chain X if and only if i is recurrent (transient) for the jump chain M.*

Proof: Suppose, i is recurrent for the jump chain M, i.e. M visits i infinitely often, at steps $(N_i^{(m)})_{m\geq 1}$. If we denote by $1_{\{X_0=i\}}$ the random variable that is 1 if $X_0 = i$ and 0 otherwise, the total amount of time that X spends at i is

$$Z_0 1_{\{X_0=i\}} + \sum_{m=1}^{\infty} Z_{N_i^{(m)}} = \infty$$

with probability 1 by the argument for Proposition 12 (convergent and divergent sums of independent exponential variables) since $Z_{N_i^{(m)}} \sim \text{Exp}(\lambda_i)$ and the sum of their (identical!) inverse parameters is infinite. In particular $\{t \geq 0 : X_t = i\}$ must be unbounded with probability 1.

Suppose, i is transient for the jump chain M, then there is a last step $L < \infty$ away from i and $\{t \geq 0 : X_t = i\} \subset [0, T_L)$ is bounded with probability 1.

The inverse implications are now obvious since i can only be either recurrent or transient for M and we constructed all minimal continuous-time Markov chains from jump chains. □

From this result and the analogous properties for discrete-time Markov chains, we deduce

Corollary 49 *Every state $i \in \mathbb{S}$ is either recurrent or transient for X.*

Recall that a class property is a property of states that either all states in a (communicating) class have or all states in a (communicating) class don't have.

Corollary 50 *Recurrence and transience are class properties.*

Proof: If i is recurrent and $i \leftrightarrow j$, for X, then i is recurrent and $i \leftrightarrow j$ for M. From discrete-time Markov chain theory, we know that j is recurrent for M. Therefore j is recurrent for X.

The proof for transience is similar. □

Proposition 51 *For any $i \in \mathbb{S}$ the following are equivalent:*

(i) *i is recurrent for X.*

(ii) *$\lambda_i = 0$ or $\mathbb{P}_i(H_i < \infty) = 1$.*

(iii) $\displaystyle\int_0^\infty p_{ii}(t)dt = \infty.$

Proof: (iii)⇒(ii): One can deduce this from the corresponding discrete-time result, but we give a direct argument here. Assume $\lambda_i > 0$ and $h_i = \mathbb{P}_i(H_i = \infty) > 0$. Then, the strong Markov property at $H_i^{(m)}$ states that, given $H_i^{(m)} < \infty$, the post-$H_i^{(m)}$ process $X^{(m+1)} = (X_{H_i^{(m)}+t})_{t\geq 0}$ is distributed as X and independent of the pre-$H_i^{(m)}$ process. By independent trials (success being not to return to i), the total number G of visits of X to i must have a geometric distribution with parameter h_i, and is independent of the $\mathrm{Exp}(\lambda_i)$-holding times in i. Therefore, the total time spent in i is

$$\sum_{m=0}^{G-1} Z_{N_i^{(m)}} \sim \mathrm{Exp}(h_i\lambda_i).$$

With notation $1_{\{X_t=i\}} = 1$ if $X_t = i$ and $1_{\{X_t=i\}} = 0$ otherwise, we obtain by Tonelli's theorem

$$\int_0^\infty p_{ii}(t)dt = \int_0^\infty \mathbb{E}_i(1_{\{X_t=i\}})dt = \mathbb{E}_i\left(\int_0^\infty 1_{\{X_t=i\}}dt\right) = \mathbb{E}_i\left(\sum_{m=0}^{G-1} Z_{N_i^{(m)}}\right) = \frac{1}{h_i\lambda_i} < \infty.$$

The other implications can be established using similar arguments. □

As in the discrete-time case, there is a link between recurrence and the existence of invariant distributions. More precisely, recurrence is strictly weaker. The stronger notion required for the existence of invariant distributions is positive recurrence:

Definition 52 A recurrent state $i \in \mathbb{S}$ is called *positive recurrent* if either $\lambda_i = 0$ or $m_i := \mathbb{E}_i(H_i) < \infty$. Otherwise, we call i null recurrent.

Proposition 53 *Positive recurrence is a class property.*

Proof: Suppose $\mathbb{E}_i(H_i) < \infty$ and $i \to j$. Then $p = \mathbb{P}_i(H_j < H_i) > 0$ and

$$p\mathbb{E}_i(H_i|H_j < H_i) + (1-p)\mathbb{E}_i(H_i|H_j > H_i) = \mathbb{E}_i(H_i) < \infty.$$

Hence $\mathbb{E}_i(H_j|H_j < H_i) < \mathbb{E}_i(H_i|H_j < H_i) < \infty$. By independent trials,

$$\mathbb{E}_i(H_j) = (1/p - 1)\mathbb{E}_i(H_i|H_j > H_i) + \mathbb{E}_i(H_j|H_j < H_i) < \infty.$$

By the Strong Markov property at H_j, also $\mathbb{E}_j(H_i) \le \mathbb{E}_i(H_i|H_j < H_i) < \infty$. Finally, $\mathbb{E}_j(H_j) \le \mathbb{E}_j(H_i) + \mathbb{E}_i(H_j) < \infty$, as required. □

Positive recurrence for a continuous-time Markov chain and its jump chain are not equivalent.

Example 54 (M/M/1 queue) The M/M/1 queue with $\lambda > 0$ and $\mu > 0$ is positive recurrent if $\lambda < \mu$, null recurrent if $\lambda = \mu$ and transient if $\lambda > \mu$.

Specifically, $\lambda > \mu$ means that customers arrive at a higher rate than they leave. Intuitively, this means that $X_t \to \infty$ (this can be shown by comparison of the jump chain with a simple random walk with up probability $\lambda/(\lambda+\mu) > 1/2$). As a consequence, $L_i = \sup\{t \ge 0 : X_t = i\} < \infty$ for all $i \in \mathbb{N}$, and since $\{t \ge 0 : X_t = i\} \subset [0, L_i]$, we deduce that i is transient.

$\lambda < \mu$ means that customers arrive at a slower rate than they can leave. Intuitively, this means that X_t will return to zero infinitely often. The mean of the return time can be estimated by comparison of the jump chain with a simple random walk with up probability $\lambda/(\lambda+\mu) < 1/2$:

$$\begin{aligned}\mathbb{E}_0(H_0) &= \mathbb{E}\left(\sum_{k=0}^{N_0-1} Z_k\right) = \sum_{n=2}^{\infty}\mathbb{P}(N_0 = n)\mathbb{E}\left(\sum_{k=0}^{n-1} Z_k \,\middle|\, N_0 = n\right) \\ &= \frac{1}{\lambda} + \sum_{n=2}^{\infty}\mathbb{P}(N_0 = n)\mathbb{E}\left(\sum_{k=1}^{n-1} Y_k\right) \\ &= \frac{1}{\lambda} + \sum_{n=2}^{\infty}\mathbb{P}(N_0 = n)\frac{n-1}{\lambda+\mu} = \frac{1}{\lambda} + \frac{\mathbb{E}_0(N_0) - 1}{\lambda+\mu} < \infty,\end{aligned}$$

where $Y_1, Y_2, \ldots \sim \mathrm{Exp}(\lambda+\mu)$. Therefore, 0 is positive recurrent. Since positive recurrence is a class property, all states are positive recurrent.

For $\lambda = \mu$, the same argument shows that 0 is null-recurrent, by comparison with simple symmetric random walk.

Note in each case, that the jump chain is not a simple random walk, but coincides with a simple random walk until it hits zero. This is enough to calculate $\mathbb{E}_0(N_0)$.

Lecture 7

Invariant distributions and time reversal

Reading: Norris 3.5, 3.7
Further reading: Grimmett-Stirzaker 6.9; Ross 6.5

In Lecture 7 we studied the class structure of continuous-time Markov chains. We can summarize the findings by saying that the state space can be decomposed into (disjoint) communicating classes

$$\mathbb{S} = \bigcup_{m\in I_1} \mathcal{T}_m \cup \bigcup_{m\in I_2} \mathcal{N}_m \cup \bigcup_{m\in I_3} \mathcal{P}_m,$$

for countable index sets I_1, I_2 and I_3, where the (states in) $\mathcal{P}_m$ are positive recurrent, hence closed, the $\mathcal{N}_m$ null recurrent, hence closed, and the $\mathcal{T}_m$ transient, open or closed. This is similar for the jump chain, but while transience holds for a class of the continuous-time chain if and only if it holds for the jump chain, this equivalence fails for positive recurrence, in general.

To understand equilibrium behaviour, we can look at each recurrent class separately. The complete picture can then be put together from its pieces on the separate classes. This is relevant in some applications, but not for the majority, and not for those most relevant to us. We therefore focus mainly on the case where we have only one class. We called this case "irreducible". The reason for this name is that we cannot further "reduce" the state space without changing the transition probabilities. We will further focus on the positive recurrent case.

7.1 Invariant distributions

Note that for an initial distribution ν on $\mathbb{S}$, $X_0 \sim \nu$ we have

$$\mathbb{P}(X_t = j) = \sum_{i\in\mathbb{S}} \mathbb{P}(X_0 = i)\mathbb{P}(X_t = j|X_0 = i) = (\nu P(t))_j$$

where $\nu P(t)$ is the product of a row vector ν with the matrix $P(t)$, and we extract the jth component of the resulting row vector.

Definition 55 A distribution ξ on $\mathbb{S}$ is called *invariant* for a continuous-time Markov chain if $\xi P(t) = \xi$ for all $t \geq 0$.

If $X_0 \sim \xi$ for an invariant distribution ξ, then $X_t \sim \xi$ for all $t \geq 0$. We therefore also say that ξ is a *stationary distribution* of X, and we then refer to X as a *stationary Markov chain.*

Proposition 56 *In the non-explosive case, in particular when $\mathbb{S}$ is finite, ξ is invariant if and only if $\xi Q = 0$.*

Proof: We only prove the case of finite $\mathbb{S}$. If $\xi P(t) = \xi$ for all $t \geq 0$, then by the forward equation

$$\xi Q = \xi P(t)Q = \xi P'(t) = \xi \lim_{h\to 0} \frac{P(t+h) - P(t)}{h} = \lim_{h\to 0} \frac{\xi P(t+h) - \xi P(t)}{h} = 0.$$

If $\xi Q = 0$, we have

$$\xi P(t) = \xi P(0) + \xi \int_0^t P'(s)ds = \xi + \int_0^t \xi Q P(s)ds = \xi$$

by the backward equation. Here, also the integration is understood component-wise.

Interchanging limits/integrals and matrix multiplication is justified since $\mathbb{S}$ is finite. For the (non-explosive) infinite case, see Norris Theorem 3.5.5. □

In the explosive case, the equation $\xi Q = 0$ may have solutions that are not invariant, see Norris Example 3.5.4.

Proposition 57 *For a vector ξ let $\eta_i = \lambda_i \xi_i$, $i \in \mathbb{S}$. Then $\xi Q = 0$ if and only if $\eta \Pi = \eta$.*

Proof: We have $(\eta(\Pi - I))_j = \sum_{i \in \mathbb{S}} \xi_i \lambda_i (\pi_{ij} - \delta_{ij}) = \sum_{i \in \mathbb{S}} \xi_i q_{ij} = (\xi Q)_j$. for all $j \in \mathbb{S}$. □

Note that the entries of ξ and η do not have to sum to 1. However, if ξ they are positive and sum to something finite, they can be normalised to give stationary distributions, e.g. always when $\mathbb{S}$ is finite.

Proposition 58 *An irreducible (minimal) non-explosive continuous-time Markov chain is positive recurrent if and only if it has an invariant distribution. An invariant distribution ξ can then be given by*

$$\xi_i = \frac{1}{m_i \lambda_i}, \qquad i \in \mathbb{S},$$

where $m_i = \mathbb{E}_i(H_i)$ is the mean return time to i and $\lambda_i = -q_{ii}$ the holding rate in i.

The analogous result for discrete chains holds and gives $\eta_i = 1/\mathbb{E}_i(N_i)$ as invariant distribution. The further factor λ_i occurs because a chain in stationarity is likely to be found in i if the return time is short and the holding time is long; both observations are reflected through the inverse proportionality to m_i and λ_i, respectively.

Proof: When $\mathbb{S}$ is finite, irreducibility implies positive recurrence, since $p = \inf_{j \in \mathbb{S}} \mathbb{P}_j(H_i \leq 1) \geq p_{j,i}(1) > 0$ and by the Markov property at integer times, repeated trials with success probabilities bounded below by p get us from any j to i in a random time with expectation bounded by $1/p$, hence $m_i \leq 1/\lambda_i + 1/p < \infty$. When $\mathbb{S}$ is finite, there is an invariant distribution by the previous proposition, since this holds for the jump chain. We will deduce the form of the invariant distribution from the Convergence theorem and a version of the Ergodic theorem in Assignment 5.5(d). When $\mathbb{S}$ is finite, this will complete the proof.

When $\mathbb{S}$ is infinite, we still need the equivalence of positive recurrence and the existence of an invariant distribution to prove the Convergence Theorem. See Norris Theorem 3.5.3 for a direct proof. □

Example 59 Consider the M/M/1 queue of Example 31. The equations $\xi Q = 0$ are given by

$$-\lambda \xi_0 + \mu \xi_1 = 0, \quad \lambda \xi_{i-1} - (\lambda + \mu) \xi_i + \mu \xi_{i+1} = 0, \quad i \geq 1.$$

This system of linear equations (for the unknowns ξ_i, $i \in \mathbb{N}$) has a probability function as its solution if and only if $\lambda < \mu$. It is given by the geometric probabilities

$$\xi_i = \left(\frac{\lambda}{\mu}\right)^i \left(1 - \frac{\lambda}{\mu}\right), \quad i \in \mathbb{N}.$$

By Proposition 58, we can calculate $\mathbb{E}_i(H_i) = m_i = 1/(\lambda_i \xi_i)$. In particular, for $i = 0$, we have the length of a full cycle beginning and ending with an empty queue. Since the initial empty period has average length $1/\lambda$, the busy period has length

$$\mathbb{E}_0(H_0) - 1/\lambda = \frac{1}{\lambda\,(1 - \lambda/\mu)} = \frac{1}{\mu - \lambda}.$$

Note that this tends to infinity as $\lambda \uparrow \mu$.

7.2 Detailed balance equations and time reversal

Proposition 60 *Consider the Q-matrix Q of a non-explosive Markov chain. If the* detailed balance equations

$$\xi_i q_{ij} = \xi_j q_{ji}, \qquad i, j \in \mathbb{S},$$

have a solution $\xi = (\xi_i)_{i\in\mathbb{S}}$, then ξ is a stationary distribution.

Proof: Let ξ be such that all detailed balance equations hold. Then fix $j \in \mathbb{S}$ and sum the equations over $i \in \mathbb{S}$ to get

$$(\xi Q)_j = \sum_{i\in\mathbb{S}} \xi_i q_{ij} = \sum_{i\in\mathbb{S}} \xi_j q_{ji} = \xi_j \sum_{i\in\mathbb{S}} q_{ji} = 0$$

since the row sums of any Q-matrix vanish (see Remark 25). Therefore $\xi Q = 0$, as required. □

Note that (in the case of finite $\#\mathbb{S} = n$), while $\xi Q = 0$ is a set of as many equations as unknowns, n, the detailed balance equations form a set of $n(n-1)/2$ different equations for n unknowns, so one would not expect solutions, in general. However, if the Q-matrix is sparse, i.e. contains lots of zeros, corresponding equations will be automatically satisfied, and these are the cases where we will successfully apply detailed balance equations.

The class of continuous-time Markov chains for which the detailed balance equations have solutions can be studied further. They also arise naturally in the context of time reversal, a tool that may seem of little practical relevance, since our world lives forward in time, but sometimes it is useful to model by a random process an unknown past. Sometimes, one can identify a duality relationships between two different processes, both forward in time that reveals that the behaviour of one is the same as the behaviour of the time reversal of the other. This can allow to translate known results for one into interesting new results for the other.

Proposition 61 *Let X be an irreducible positive recurrent (minimal) continuous-time Markov chain with Q-matrix Q and starting from the invariant distribution ξ. Let $t > 0$ be a fixed time and $\hat{X}_s = X_{t-s-}$. Then the process $\hat{X}$ is a continuous-time Markov chain with Q-matrix $\hat{Q}$ given by $\xi_j \hat{q}_{ji} = \xi_i q_{ij}$.*

Proof: First note that $\hat{Q}$ has the properties of a Q-matrix in being non-negative off the diagonal and satisfying

$$\sum_{i\in\mathbb{S}} \hat{q}_{ji} = \sum_{i\in\mathbb{S}} \frac{\xi_i}{\xi_j} q_{ij} = \frac{1}{\xi_j}(\xi Q)_j = 0$$

by the invariance of ξ. Similarly, we define $\xi_j \hat{p}_{ji}(t) = \xi_i p_{ij}(t)$ and see that $\hat{P}(t)$ have the properties of transition matrices. In fact the transposed forward equation $P'(t) = P(t)Q$ yields $\hat{P}'(t) = \hat{Q}\hat{P}(t)$, the backward equation for $\hat{P}(t)$. Now $\hat{X}$ is a continuous-time Markov chain with transition probabilities $\hat{P}(t)$ since

$$\begin{aligned}\mathbb{P}_\xi(\hat{X}_{t_0} = i_0, \ldots, \hat{X}_{t_n} = i_n) &= \mathbb{P}_\xi(X_{t-t_n} = i_n, \ldots, X_{t-t_0} = i_0) \\ &= \xi_{i_n} \prod_{k=1}^{n} p_{i_k, i_{k-1}}(t_k - t_{k-1}) \\ &= \xi_{i_0} \prod_{k=1}^{n} \hat{p}_{i_{k-1}, i_k}(t_k - t_{k-1}).\end{aligned}$$

From this we can deduce the Markov property. More importantly, the finite-dimensional distributions of $\hat{X}$ are the ones of a continuous-time Markov chain with transition matrices $\hat{P}(t)$. Together with right-continuity, Remark 39 implies that $\hat{X}$ is a Markov chain with Q-matrix $\hat{Q}$. □

If $\hat{Q} = Q$, X is called *reversible*. It is evident from the definition of $\hat{Q}$ that ξ then satisfies the *detailed balance equations* $\xi_i q_{ij} = \xi_j q_{ji}$, $i, j \in \mathbb{S}$.

7.3 Simple birth-and-death processes and Erlang's formula

Consider the general simple birth-and-death process with birth rates $q_{i,i+1} = \beta_i$ and death rates $q_{i,i-1} = \gamma_i$, $q_{ii} = -\beta_i - \gamma_i$, $i \in \mathbb{N}$, all other entries zero, and also $\gamma_0 = 0$. We recognise simple birth processes and queueing models as special cases.

To calculate invariant distributions, we solve the balance equations $\xi Q = 0$, i.e.

$$\xi_1\gamma_1 - \xi_0\beta_0 = 0 \qquad \text{and} \qquad \xi_{n+1}\gamma_{n+1} - \xi_n(\beta_n + \gamma_n) + \xi_{n-1}\beta_{n-1} = 0, \quad n \geq 1$$

or more easily the detailed balance equations

$$\xi_i\beta_i = \xi_{i+1}\gamma_{i+1}, \qquad i \geq 0.$$

giving

$$\xi_n = \frac{\beta_{n-1}\cdots\beta_0}{\gamma_n\cdots\gamma_1}\xi_0, \qquad n \geq 0,$$

where ξ_0 is determined by the normalisation requirement of ξ to be a probability function, i.e.

$$\xi_0 = \frac{1}{S} \quad \text{where} \quad S = 1 + \sum_{n=1}^{\infty} \frac{\beta_{n-1}\cdots\beta_0}{\gamma_n\cdots\gamma_1}$$

provided S is finite. Assuming non-explosion, we have found an invariant distribution, and we will see it is unique since the state space is irreducible. (We can get a sufficient condition for non-explosion, by ignoring deaths to find a simple birth process with rates $\lambda_n = \beta_n$, for which the explosion criterion applies.)

If S is infinite there does not exist an invariant distribution. This cannot be deduced directly from the detailed balance equations but for general simple birth-and-death processes here it is easy to see inductively that $\xi Q = 0$ is equivalent to the detailed balance equations (the first equation is the same, the other balance equations are differences of consecutive detailed balance equations). The fact that there is no probability measure such that ξ such that $\xi Q = 0$ does not necessarily mean there is explosion in finite time, infact $S = \infty$ for all simple birth processes since they model growing populations and cannot be in equilibrium. By Proposition 58, it means that X is then null recurrent or transient. On the other hand, if $\beta_0 = 0$ as in many population models, then the invariant distribution is concentrated in 0, i.e. $\xi_0 = 1$, $\xi_n = 0$ for all $n \geq 1$.

Many special cases can be given more explicitly.

Example 62 If $\beta_n = \lambda$, $n \geq 0$, $\gamma_n = n\mu$, we get

$$\xi_n = \frac{(\lambda/\mu)^n}{n!}e^{-\lambda/\mu}, \qquad n \geq 0,$$

Poisson probabilities. What is this model? We can give two different interpretations both of which tie in with models that we have studied. First, as a population model, $\beta_n = \lambda$ means that arrivals occur according to a Poisson process, this can model immigration; $\gamma_n = n\mu$ is obtained from as many Exp(μ) clocks as individuals in the population, i.e. independent Exp(μ) lifetimes for all individuals, an **immigration-death model**. Second, as a queueing model with arrivals according to a Poisson process, each individual leaves the system after an Exp(μ) time, no matter how many other people are in the system – this can be obtained from infinitely many servers working at rate μ, an **M/M/∞-queue**.

Example 63 (Erlang's formula) If $\beta_n = \lambda$, $0 \leq n \leq s-1$, and $\gamma_n = n\mu$, we obtain a model for a telephone exchange with s lines, where calls arrive according to PP(λ) and have independent Exp(μ) durations. Of particular interest is the probability of having all lines busy. A slight variant of the formulas in the previous example yields

$$\xi_s = \frac{(\lambda/\mu)^s}{s!} \Bigg/ \sum_{n=0}^{s} \frac{(\lambda/\mu)^n}{n!} .$$

This is known as Erlang's formula.

Lecture 8

Convergence theorems

Reading: Norris 3.6, 3.8; Grimmett-Stirzaker 7.2
Further reading: Williams "Probability with Martingales" 7.2; Grimmett-Stirzaker 7.1, 7.3-7.5

The Convergence Theorem and the Ergodic Theorem for continuous-time Markov chains, which we will discuss in this lecture, are two very different statements of convergence to a stationary distribution. We begin this lecture by reviewing modes of convergence. We also prove the Strong Law of Large Numbers.

8.1 Modes of convergence

Definition 64 Let X_n, $n \geq 1$, and X be real-valued random variables. Then we define

1. $X_n \to X$ *in probability*, if for all $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$.
2. $X_n \to X$ *in distribution*, if $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$ as $n \to \infty$, for all $x \in \mathbb{R}$ at which $x \mapsto \mathbb{P}(X \leq x)$ is continuous.
3. $X_n \to X$ *in* L^1, if $\mathbb{E}(|X_n|) < \infty$ for all $n \geq 1$ and $\mathbb{E}(|X_n - X|) \to 0$ as $n \to \infty$.
4. $X_n \to X$ *almost surely* (a.s.), if $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$.

The following result from Part A will be useful.

Proposition 65 *The following implications hold*

$$\begin{array}{ccc} X_n \to X \text{ almost surely} & & \\ \Downarrow & & \\ X_n \to X \text{ in probability} & \Rightarrow & X_n \to X \text{ in distribution} \\ \Uparrow & & \\ X_n \to X \text{ in } L^1 & \Rightarrow & \mathbb{E}(X_n) \to \mathbb{E}(X) \end{array}$$

No other implications hold in general.

For $\mathbb{N}$-valued random variables X and X_n, $n \geq 1$, convergence $X_n \to X$ in distribution is equivalent to

$$\mathbb{P}(X_n = j) \to \mathbb{P}(X = j), \qquad \text{as } n \to \infty, \text{ for all } j \in \mathbb{N},$$

as $\mathbb{P}(X_n = j) = \mathbb{P}(X_n \leq j+1/2) - \mathbb{P}(X_n \leq j-1/2)$ and $\mathbb{P}(X_n \leq x) = \mathbb{P}(X_n = 0) + \cdots + \mathbb{P}(X_n = \lfloor x \rfloor)$. It is therefore natural to define for $\mathbb{S}$-valued random variables $X_n \to X$ in distribution

$$\mathbb{P}(X_n = j) \to \mathbb{P}(X = j), \qquad \text{as } n \to \infty, \text{ for all } j \in \mathbb{S}.$$

8.2 The convergence theorem and the ergodic theorem

The Convergence theorem for Markov chains is of central importance in applications since is it often assumed that "a system is in equilibrium". The convergence theorem is a justification for this assumption, since it means that a system must only be running long enough to be (approximately) in equilibrium. Recall that an "invariant distribution" ($\xi P(t) = \xi$ for all $t \geq 0$) is also called "stationary distribution" (since $X_0 \sim \xi \Rightarrow X_t \sim \xi$ for all $t \geq 0$). We will now see that ξ is also "equilibrium distribution", i.e. $X_t \sim \xi$ approximately for large t even if $X_0 \sim \nu \neq \xi$.

Theorem 66 *Let $X = (X_t)_{t\geq 0}$ be a (minimal) irreducible positive-recurrent continuous-time Markov chain, $X_0 \sim \nu$, and ξ an invariant distribution, then*

$$\mathbb{P}(X_t = j) \to \xi_j \qquad \text{as } t \to \infty \text{ for all } j \in \mathbb{S}.$$

This result can be deduced from the Convergence theorem for discrete-time Markov chains by looking at the processes $Z_n^{(h)} = X_{nh}$, $n \geq 0$, for each fixed $h > 0$. The process $Z^{(h)}$ is easily seen to be a discrete-time Markov chains with transition matrix $P(h)$.

However, it is more instructive to see a (very elegant) direct argument, using the coupling method in continuous time.

Sketch of proof: Let $X \sim (\nu, Q)$-Markov chain and $Y \sim (\xi, Q)$-Markov chain independent. Choose $i \in \mathbb{S}$ and define $T = \inf\{t \geq 0 : (X_t, Y_t) = (i, i)\}$ the time they first meet (in i, to simplify the argument). A third process is constructed

$$\widetilde{X}_t = \begin{cases} X_t & \text{if } t < T, \\ Y_t & \text{if } t \geq T. \end{cases}$$

The following three steps complete the proof:

1. $\mathbb{P}(T < \infty) = 1$ – this is because $\eta_{i,j} = \xi_i \xi_j$, $(i, j) \in \mathbb{S}^2$, is invariant for (X, Y), which implies positive recurrence of (X, Y), by Proposition 58, and by irreducibility expected hitting times from any fixed starting point are finite (see the proof of Proposition 53), and hence hitting times from any random starting point are finite with probability 1;

2. $\widetilde{X} \sim (\nu, Q)$-Markov chain just as X;

3. $|\mathbb{P}(X_t = j) - \xi_j| = \left|\mathbb{E}(1_{\{\widetilde{X}_t = j\}}) - \mathbb{E}(1_{\{Y_t = j\}})\right| \leq \mathbb{E}(1_{\{T>t\}}) = \mathbb{P}(T > t) \to 0.$ □

Corollary 67 *A (minimal) positive recurrent irreducible continuous-time Markov chain has a unique stationary distribution.*

Proof: For ν and ξ stationary and $j \in \mathbb{S}$, we have $\nu_j = \mathbb{P}(X_t = j) \to \xi_j$ as $t \to \infty$, so $\nu = \xi$. □

Theorem 68 (Ergodic theorem) *In the setting of Theorem 66, $X_0 \sim \nu$*

$$\mathbb{P}\left(\frac{1}{t}\int_0^t 1_{\{X_s = i\}} ds \to \xi_i \ \text{ as } t \to \infty\right) = 1$$

Proof: A proof using renewal theory is in Exercise A.5.4. □

We interpret this as follows. For any initial distribution, the long-term proportions of time spent in any state i approaches the invariant probability for this state. This result establishes a time-average analogue for the distributional average of Theorem 66. This is of practical importance, since it allows us to *observe* the invariant distribution by looking at time proportions over a long period of time. If we tried to observe the stationary distribution using Theorem 66, we would need many independent observations of the same system at a large time t to estimate ξ, using the Strong Law of Large Numbers, see Example 74.

Example 69 For the M/M/1 queue, we calculated in Example 59 that the stationary probability of the idle state is $\xi_0 = 1 - \lambda/\mu$. By Theorem 66, this is the probability for a customer to find the server idle if the system has been running for a long time. By Theorem 68, this is the asymptotic proportion of time that the server is idle (which in some applications means the server is not earning any money).

8.3 The Strong Law of Large Numbers

Key to the proof of Ergodic theorems is the Strong Law of Large Numbers, which was stated, but not proved in Part A.

Theorem 70 *Let $(Y_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(|Y_1|) < \infty$ and $\mathbb{E}(Y_1) = \mu$. Let $S_n = Y_1 + \cdots + Y_n$, $n \geq 1$. Then*

$$\frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i \to \mu \qquad \textit{almost surely.}$$

The proof for the general result is lengthy (or requires techniques beyond this course), but under the extra moment condition $\mathbb{E}(Y_1^4) < \infty$ there is a nice proof, which ultimately exploits an argument we employed in the proof of the explosion criterion, Proposition 12: if $\sum \mathbb{E}(Z_j) < \infty$, then $\sum Z_j < \infty$ with probability 1.

Lemma 71 *In the situation of Theorem 70, there is a constant $K < \infty$ such that for all $n \geq 0$*

$$\mathbb{E}((S_n - n\mu)^4) \leq 4\mathbb{E}(\widetilde{Y}_1^4)n^2 \,,$$

where $\widetilde{Y}_k = Y_k - \mu$.

Proof: Let $\widetilde{S}_n = \widetilde{Y}_1 + \cdots + \widetilde{Y}_n = S_n - n\mu$. Then

$$\mathbb{E}\left(\widetilde{S}_n^4\right) = \mathbb{E}\left(\left(\sum_{i=1}^{n}\widetilde{Y}_i\right)^4\right) = \mathbb{E}\left(\sum_{i,j,k,\ell=1}^{n}\widetilde{Y}_i\widetilde{Y}_j\widetilde{Y}_k\widetilde{Y}_\ell\right) = n\mathbb{E}(\widetilde{Y}_1^4) + 6\binom{n}{2}\mathbb{E}(\widetilde{Y}_1^2\widetilde{Y}_2^2)$$

by expanding the fourth power and noting that most terms vanish such as

$$\mathbb{E}(\widetilde{Y}_1\widetilde{Y}_2^3) = \mathbb{E}(\widetilde{Y}_1)\mathbb{E}(\widetilde{Y}_2^3) = 0.$$

Finally observe that $4\max\{\mathbb{E}(\widetilde{Y}_1^4), (\mathbb{E}(\widetilde{Y}_1^2))^2\} = 4\mathbb{E}(\widetilde{Y}_1^4)$ by Jensen's inequality and $n \leq n^2$ for $n \geq 0$. □

Proof of Theorem 70 when $\mathbb{E}(Y_1^4) < \infty$: By the lemma,

$$\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^4\right) \leq \mathbb{E}(\widetilde{Y}_1^4)n^{-2} \leq Kn^{-2}\infty\,.$$

for some $K > 0$ independent of n. Now, by Tonelli's theorem,

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\left(\frac{S_n}{n} - \mu\right)^4\right) = \sum_{n=1}^{\infty}\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^4\right) < \infty \quad \Rightarrow \quad \sum_{n=1}^{\infty}\left(\frac{S_n}{n} - \mu\right)^4 < \infty \quad \text{a.s.}$$

But if a series converges, the underlying sequence converges to zero, and so

$$\left(\frac{S_n}{n} - \mu\right)^4 \to 0 \quad \text{almost surely} \qquad \Rightarrow \qquad \frac{S_n}{n} \to \mu \quad \text{almost surely.}$$

□

8.4 Examples

Example 72 (Poisson process) For $X \sim \mathrm{PP}(\lambda)$, we have arrival times $T_n = Z_0 + \cdots + Z_{n-1}$ for $Z_i \sim \mathrm{Exp}(\lambda)$ independent. The Strong Law of Large Numbers yields

$$\frac{T_n}{n} \to \frac{1}{\lambda} \qquad \text{almost surely, as } n \to \infty.$$

Since X itself has stationary independent increments, we also have

$$\frac{X_n}{n} = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_{i-1}) \to \mathbb{E}(X_1) = \lambda \qquad \text{almost surely, as } n \to \infty.$$

Furthermore, we can use the sandwich theorem and algebra of limits to deduce

$$\frac{[t]}{t}\frac{X_{[t]}}{[t]} \le \frac{X_t}{t} \le \frac{X_{[t]+1}}{[t]+1}\frac{[t]+1}{t} \qquad \Rightarrow \qquad \frac{X_t}{t} \to \lambda \quad \text{almost surely, as } t \to \infty.$$

Example 73 (Return times of Markov chains) For a $M \sim (\nu, \Pi)$-Markov chain positive recurrent, recall notation

$$N_i = N_i^{(1)} = \inf\{n > 0 : M_n = i\}, \quad N_i^{(m+1)} = \inf\{n > N_i^{(m)} : M_n = i\}, \quad m \in \mathbb{N},$$

for the successive return times to $i \in \mathbb{S}$. By the strong Markov property, the random variables $N_i^{(m+1)} - N_i^{(m)}$, $m \ge 1$, are independent and identically distributed. If we define $N_i^{(0)} = 0$ and $\nu = \delta_i$, then this holds for all $m \ge 0$. The Strong Law of Large Number yields

$$\frac{N_i^{(m)}}{m} \to \mathbb{E}_i(N_i) \qquad \text{almost surely, as } m \to \infty.$$

Similarly, for a positive recurrent (ν, Q)-Markov chain X, for

$$H_i = H_i^{(1)} = \inf\{t \ge T_1 : X_t = i\}, \quad H_i^{(m)} = T_{N_i^{(m)}}, m \in \mathbb{N},$$

we get

$$\frac{H_i^{(m)}}{m} \to \mathbb{E}_i(H_i) = m_i \qquad \text{almost surely, as } m \to \infty.$$

Example 74 (Empirical distributions) If $(Y_n)_{n\ge1}$ is an infinite sample (independent and identically distributed random variables) from a discrete distribution ν on $\mathbb{S}$, then the random variables $B_n^{(i)} = 1_{\{Y_n = i\}}$, $n \ge 1$, are also independent and identically distributed for each fixed $i \in \mathbb{S}$, as functions of independent variables. The Strong Law of Large Numbers yields

$$\nu_i^{(n)} = \frac{\#\{k = 1, \ldots, n : Y_k = i\}}{n} = \frac{B_1^{(i)} + \cdots + B_n^{(i)}}{n} \to \mathbb{E}(B_1^{(i)}) = \mathbb{P}(Y_1 = i) = \nu_i$$

almost surely, as $n \to \infty$. The probability function $\nu^{(n)}$ is called empirical distribution. It lists relative frequencies in the sample and, for a specific realisation, can serve as an approximation of the true distribution. In applications of statistics, it is the sample distribution associated with a population distribution. The result that empirical distributions converge to the true distribution, is true uniformly in i and in higher generality, it is usually referred to as the Glivenko-Cantelli theorem.

Remark 75 (Discrete ergodic theorem) If $(M_n)_{n\ge0}$ is a positive-recurrent (ν, Π)- Markov chain, the Ergodic Theorem is a statement very similar to the example of empirical distributions

$$\frac{\#\{k = 0, \ldots, n-1 : M_k = i\}}{n} \to \eta_i \qquad \text{almost surely, as } n \to \infty,$$

for a stationary distribution η, but of course, the M_n, $n \ge 0$, are not independent (in general), and only identically distributed if $\nu = \eta$. We need to work a bit harder to deduce the Ergodic Theorem from the Strong Law of Large Numbers.

Lecture 9

Renewal processes and equations

Reading: Grimmett-Stirzaker 10.1-10.2; Ross 7.1-7.3

We introduce renewal processes $(X_t, t \geq 0)$ and study the asymptotic behaviour of X_t.

9.1 Motivation and definition

The Poisson process (and all continuous-time Markov chains) have exponentially distributed holding times, which possess the lack of memory property. In practice, this assumption is often not reasonable.

Example 76 Suppose that you count the changing of batteries for an electrical device. Given that the battery has been in use for time t, is its residual lifetime distributed as its total lifetime? We would assume this, if we were modelling with a Poisson process.

We may wish to replace the exponential distribution by other distributions, e.g. one that cannot take arbitrarily large values or, for other applications, one that can produce clustering effects (many short holding times separated by significantly longer ones). Here is the analogue of a Poisson process when we drop the requirement for $\mathrm{Exp}(\lambda)$ inter-arrival distribution:

Definition 77 Let Z_n, $n \geq 0$, be independent identically distributed and positive random variables. Let $T_n = Z_0 + \cdots + Z_{n-1}$, $n \geq 1$. Then the process $X = (X_t, t \geq 0)$ defined by

$$X_t = \#\{n \geq 1 : T_n \leq t\}$$

is called a *renewal process*. The common distribution of Z_n, $n \geq 0$, is called the *inter-arrival distribution* or the *inter-renewal distribution*.

Renewal processes also arise naturally in the study of continuous-time Markov chains.

Example 78 If $(Y_t)_{t\geq 0}$ is a continuous-time Markov chain with $Y_0 = i$, then $Z_n = H_i^{(n+1)} - H_i^{(n)}$, the times between successive returns to i by Y, are independent and identically distributed (by the strong Markov property). The associated counting process

$$X_t = \#\{n \geq 1 : H_i^{(n)} \leq t\}, \qquad t \geq 0,$$

counting the number of visits to i up to time t, $t \geq 0$, is thus a renewal process.

9.2 The renewal function

Definition 79 The function $m\colon [0,\infty) \to [0,\infty)$, $m(t) := \mathbb{E}(X_t)$, is called the *renewal function.*

For $Z_n \sim \mathrm{Exp}(\lambda)$ we have $X_t \sim \mathrm{Poi}(\lambda t)$ and $m(t) = \mathbb{E}(X_t) = \lambda t$. In general, the renewal function is one of the central objects in renewal theory.

To calculate the renewal function for general renewal processes, we investigate the distribution of X_t. For any counting process $(X_t, t \geq 0)$, we have

$$X_t = k \iff T_k \leq t < T_{k+1},$$

so that we can express

$$\mathbb{P}(X_t = k) = \mathbb{P}(T_k \le t < T_{k+1}) = \mathbb{P}(T_k \le t) - \mathbb{P}(T_{k+1} \le t)$$

in terms of the distributions of $T_k = Z_0 + \cdots + Z_{k-1}$, $k \ge 1$.

If Z_i has a probability density function, we can express the density of T_k as a convolution power. Specifically, recall that for two nonnegative independent continuous random variables Z and T with densities f and g, the random variable $Z + T$ has density

$$(f * g)(u) = \int_0^u f(u-t)g(t)dt, \quad u \ge 0.$$

It is not hard to check that the convolution product is symmetric, associative and distributes over sums of functions. While the first two of these properties follow from $Z + T = T + Z$ and $(Z + T) + V = Z + (T + V)$ for associated random variables, the third property has no such meaning, since sums of densities are no longer probability densities. However, the definition of the convolution product makes sense for general nonnegative integrable functions and distributivity follows straight from the definition. We can define convolution powers $f^{*(1)} = f$ and $f^{*(k+1)} = f * f^{*(k)}$, $k \ge 1$. Then

$$\mathbb{P}(T_k \le t) = \int_0^t f_{T_k}(s)ds = \int_0^t f^{*(k)}(s)ds,$$

if Z_n, $n \ge 0$, are continuously distributed with density f.

Proposition 80 *Let X be a renewal process with inter-renewal density f. Then*

$$m(t) = \int_0^t \sum_{k=1}^{\infty} f^{*(k)}(s)ds.$$

Proof: By Tonelli's Theorem,

$$\int_0^t \sum_{k=1}^{\infty} f^{*(k)}(s)ds = \sum_{k=1}^{\infty} \int_0^t f^{*(k)}(s)ds = \sum_{k=1}^{\infty} \mathbb{P}(T_k \le t) = \sum_{k=1}^{\infty} \mathbb{P}(X_t \ge k) = \mathbb{E}(X_t) = m(t).$$

□

9.3 The renewal equation

For continuous-time Markov chains, conditioning on the first transition time was a powerful tool. We can do this here and get what is called the *renewal equation.*

Proposition 81 *Let X be a renewal process with inter-renewal density f. Then $m(t) = \mathbb{E}(X_t)$ is the unique (locally bounded) solution of*

$$m(t) = F(t) + \int_0^t m(t-s)f(s)ds, \qquad \text{i.e. } m = F + f * m,$$

where $F(t) = \int_0^t f(s)ds = \mathbb{P}(Z_1 \le t)$.

Proof: Conditioning on the first arrival will involve the process $\widetilde{X}_u := X_{T_1+u}$, $u \ge 0$. Note that $\widetilde{X}_0 = 1$ and that $(\widetilde{X}_u - 1, u \ge 0)$ is a renewal process with inter-renewal times $\widetilde{Z}_n = Z_{n+1}$, $n \ge 0$, independent of T_1. Therefore

$$m(t) = \mathbb{E}(X_t) = \int_0^{\infty} f(s)\mathbb{E}(X_t|T_1 = s)ds = \int_0^t f(s)\mathbb{E}(\widetilde{X}_{t-s})ds = F(t) + \int_0^t f(s)m(t-s)ds\,,$$

where in the last equality we used $\widetilde{X}_{t-s}$ has the same distribution as $1 + X_{t-s}$. For uniqueness, suppose that also $\ell = F + f * \ell$, then $\alpha = \ell - m$ is locally bounded and satisfies $\alpha = f * \alpha = \alpha * f$. Iteration gives $\alpha = \alpha * f^{*(k)}$ for all $k \geq 1$ and, summing over k gives for the right hand side something finite:

$$\begin{aligned} \left| \left(\sum_{k=1}^{\infty} \alpha * f^{*(k)} \right)(t) \right| &= \left| \left(\alpha * \sum_{k=1}^{\infty} f^{*(k)} \right)(t) \right| = \left| \left(\alpha * m' \right)(t) \right| \\ &= \left| \int_0^t \alpha(t-s) m'(s) ds \right| \leq \left(\sup_{u \in [0,t]} |\alpha(u)| \right) m(t) < \infty \end{aligned}$$

but the left-hand side is infinite unless $\alpha(t) = 0$. Therefore $\ell(t) = m(t)$, for all $t \geq 0$. □

Remark 82 We can write the solution of the renewal equation as $m = F + F * \sum_{k \geq 1} f^{*(k)}$. Indeed, we check that the right-hand side $\ell := F + F * \sum_{k \geq 1} f^{*(k)}$ satisfies the renewal equation:

$$F + f * \ell = F + F * f + F * \sum_{j=2}^{\infty} f^{*(j)} = F + F * \sum_{k=1}^{\infty} f^{*(k)} = \ell,$$

just using properties of the convolution product. By the uniqueness part of Proposition 81, we conclude that $\ell = m$.

Unlike Poisson processes, general renewal processes do not have a linear renewal function, but it will be asymptotically linear (Elementary Renewal Theorem, as we will see). In fact, renewal functions are in one-to-one correspondence with inter-renewal distributions – we do not prove this, but it should not be too surprising given that $m = F + f * m$ is almost symmetric in f and m, and is symmetric after differentiation (we omit the details), which gives $m' = f + f * m'$.

9.4 Strong Law and Central Limit Theorem of renewal theory

Theorem 83 (Strong Law of renewal theory) *Let X be a renewal process with mean inter-renewal time $\mu = \mathbb{E}(Z_1) \in (0, \infty)$. Then*

$$\frac{X_t}{t} \to \frac{1}{\mu} \qquad \textit{almost surely, as } t \to \infty.$$

Proof: Note that X is constant on $[T_n, T_{n+1})$ for all $n \geq 0$, and therefore constant on $[T_{X_t}, T_{X_t+1}) \ni t$. Therefore, for all $t \geq T_1$,

$$\frac{T_{X_t}}{X_t} \leq \frac{t}{X_t} < \frac{T_{X_t+1}}{X_t} = \frac{T_{X_t+1}}{X_t+1} \frac{X_t+1}{X_t}.$$

Now $\mathbb{P}(X_t \to \infty) = 1$, since $X_\infty \leq n \iff T_{n+1} = \infty$ which is absurd, since $T_{n+1} = Z_0 + \cdots + Z_n$ is a finite sum of finite random variables. Therefore, we conclude from the Strong Law of Large Numbers for T_n, that

$$\frac{T_{X_t}}{X_t} \to \mu \qquad \text{almost surely, as } t \to \infty.$$

Therefore, if $X_t \to \infty$ and $T_n/n \to \mu$, then by the sandwich theorem

$$\mu \leq \lim_{t \to \infty} \frac{t}{X_t} \leq \mu\,,$$

but this means $\mathbb{P}(X_t/t \to 1/\mu) \geq \mathbb{P}(X_t \to \infty, T_n/n \to \mu) = 1$, as required. □

Of course, there is also a Weak Law of Renewal Theory, which now follows, because almost sure convergence implies convergence in probability. A direct proof for convergence in probability is more difficult, because ε-deviations of X_t/t are not quite the same as ε-deviations from T_{X_t}/X_t.

Theorem 84 (Central Limit Theorem of Renewal Theory) *Let X be a renewal process whose inter-renewal times Z_n, $n \geq 0$, have finite variance $\sigma^2 = \mathrm{Var}(Z_1) \in (0, \infty)$ and $\mu = \mathbb{E}(Z_1)$. Then*

$$\frac{X_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \to \mathrm{Normal}(0,1) \qquad \textit{in distribution, as } t \to \infty.$$

A rough proof is not difficult, the details a little harder and left as Exercise A.5.3.

9.5 The elementary renewal theorem

Theorem 85 *Let X be a renewal process with mean inter-renewal times μ and $m(t) = \mathbb{E}(X_t)$. Then*

$$\frac{m(t)}{t} = \frac{\mathbb{E}(X_t)}{t} \to \frac{1}{\mu} \qquad \textit{as } t \to \infty$$

This does not follow from the Strong Law of Renewal Theory since almost sure convergence does not imply convergence of means (cf. Proposition 65, see Exercise A.5.2(a)). The proof is not examinable, but the statement of the following lemma is instructive:

Lemma 86 *For a renewal process X with arrival times $(T_n, n \geq 1)$, we have*

$$\mathbb{E}(T_{X_t+1}) = \mu(m(t) + 1), \qquad \textit{where } m(t) = \mathbb{E}(X_t) \textit{ and } \mu = \mathbb{E}(T_1).$$

This may not be surprising, because T_{X_t+1} is the sum of $X_t + 1$ inter-renewal times, each with mean μ. Taking expectations, we may expect $m(t) + 1$ times μ. However, we have $\mathbb{E}(Z_{X_t}) > \mathbb{E}(Z_1)$ even for $X \sim \mathrm{PP}(\lambda)$, by the lack of memory property, so the lemma can only hold if $\mathbb{E}(T_{X_t}) < \mu m(t)$.

Proof of Lemma 86: Let us do a one-step analysis on the quantity of interest $g(t) = \mathbb{E}(T_{X_t+1})$:

$$g(t) = \int_0^\infty \mathbb{E}(T_{X_t+1} | T_1 = s) f(s) ds = \int_0^t \big(s + \mathbb{E}(T_{X_{t-s}+1})\big) f(s) ds + \int_t^\infty s f(s) ds = \mu + (g * f)(t).$$

This is almost the renewal equation. In fact, $g_1(t) = g(t)/\mu - 1$ satisfies the renewal equation

$$g_1(t) = \frac{1}{\mu} \int_0^t g(t-s) f(s) ds = \int_0^t (g_1(t-s) + 1) f(s) ds = F(t) + (g_1 * f)(t),$$

and, by Proposition 81, $g_1(t) = m(t)$, i.e. $g(t) = \mu(1 + m(t))$ as required. □

Proof of Theorem 85: Clearly $t < \mathbb{E}(T_{X_t+1}) = \mu(m(t) + 1)$ gives the lower bound $\liminf_{t\to\infty} \frac{m(t)}{t} \geq \frac{1}{\mu}$. For the upper bound we use a truncation argument and introduce the renewal process $\widetilde{X}$ associated with

$$\widetilde{Z}_j = Z_j \wedge a = \begin{cases} Z_j & \text{if } Z_j < a, \\ a & \text{if } Z_j \geq a, \end{cases} \qquad j \geq 0.$$

Then $\widetilde{Z}_j \leq Z_j$ for all $j \geq 0$ implies $\widetilde{X}_t \geq X_t$ for all $t \geq 0$, hence $\widetilde{m}(t) \geq m(t)$. By the lemma,

$$t \geq \mathbb{E}(\widetilde{T}_{\widetilde{X}_t}) = \mathbb{E}(\widetilde{T}_{\widetilde{X}_t+1}) - \mathbb{E}(\widetilde{Z}_{\widetilde{X}_t}) = \widetilde{\mu}(\widetilde{m}(t) + 1) - \mathbb{E}(\widetilde{Z}_{\widetilde{X}_t}) \geq \widetilde{\mu}(m(t) + 1) - a.$$

Therefore

$$\frac{m(t)}{t} \leq \frac{1}{\widetilde{\mu}} + \frac{a - \widetilde{\mu}}{\widetilde{\mu} t} \qquad \Rightarrow \qquad \limsup_{t\to\infty} \frac{m(t)}{t} \leq \frac{1}{\widetilde{\mu}} \qquad \Rightarrow \qquad \limsup_{t\to\infty} \frac{m(t)}{t} \leq \frac{1}{\mu},$$

since $\widetilde{\mu} = \mathbb{E}(\widetilde{Z}_1) = \mathbb{E}(Z_1 \wedge a) \to \mathbb{E}(Z_1) = \mu$ a $a \to \infty$, by monotone convergence. □

Note that truncation was necessary to get $\mathbb{E}(\widetilde{Z}_{\widetilde{X}_t}) \leq a$. It would have been enough if we had $\mathbb{E}(Z_{X_t}) = \mathbb{E}(Z_1) = \mu$, but as noted above, this is *not* true. The method of truncation can also be used to prove the Strong Law of Large Numbers without finite fourth moment, just assuming $\mathbb{E}(|Y_1|) < \infty$.

Lecture 10

Excess life and stationarity

Reading: Grimmett-Stirzaker 10.3-10.4; Ross 7.7

So far, we have studied the asymptotic behaviour of one-dimensional distributions of a renewal process X, as $t \to \infty$:

$$\frac{X_t}{t} \to \frac{1}{\mu} \quad \text{almost surely}, \qquad \frac{\mathbb{E}(X_t)}{t} \to \frac{1}{\mu}, \qquad \frac{X_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \to \text{Normal}(0,1) \quad \text{in distribution.}$$

For Poisson processes we also studied finite-dimensional marginal distributions and joint distributions of

$$X_t, X_{t+s} - X_t, \ldots \qquad \text{stationary, independent increments.}$$

In this lecture, we study the increments of renewal processes.

10.1 The renewal property, age and excess life

To begin with, let us study the post-t process for a renewal process X, i.e. $(X_{t+s} - X_t)_{s\geq 0}$. For fixed t, this is not a renewal process as defined so far. We first consider renewal times T_i, $i \geq 0$.

Proposition 87 *Let X be a renewal process and $T_i = \inf\{t \geq 0 : X_t = i\}$ the ith renewal time for some $i \geq 1$. Then $(X_r)_{r\leq T_i}$ and $\widetilde{X} = (X_{T_i+s} - X_{T_i})_{s\geq 0}$ are independent and $\widetilde{X}$ has the same distribution as X.*

Proof: The proof is the same as (actually easier than) the proof of the strong Markov property of birth processes at birth times T_i, $i \geq 1$, cf. Exercise A.2.8(b). Specifically, the inter-renewal times $\widetilde{Z}_n = Z_{i+n}$, $n \geq 0$, are independent of $Z_0, \ldots, Z_{i-1}$. Functions of independent random variables are independent. Here, the pre-T_i-process $(X_r)_{r\leq T_i}$ is a function of $Z_0, \ldots, Z_{i-1}$, while the post-T_i-process $\widetilde{X}$ is constructed from $\widetilde{Z}_n$, $n \geq 0$, just as X is constructed from Z_n, $n \geq 0$. □

Here are two examples to see how this breaks down when T_i is replaced by a fixed time $t > 0$.

Example 88 If the inter-renewal times are constant, say $\mathbb{P}(Z_n = 3) = 1$, then $\widetilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ has a first arrival time $\widetilde{Z}_0$ with $\mathbb{P}(\widetilde{Z}_0 = 3 - t) = 1$, for $0 \leq t < 3$.

The second example shows that also the independence of the pre-t and post-t processes fails.

Example 89 Let $\mathbb{P}(Z_n = 1) = 0.7$, $\mathbb{P}(Z_n = 2) = 0.2$ and $\mathbb{P}(Z_n = 19) = 0.1$ for all $n \geq 0$. Then

$$\mu = \mathbb{E}(Z_n) = 0.7 \times 1 + 0.2 \times 2 + 0.1 \times 19 = 3.$$

Consider a renewal process X with this inter-renewal distribution, and let $t = 2$. Denoting the post-t process by $\widetilde{X} = (X_{t+s} - X_t)_{s\geq 0}$, with inter-renewal times $\widetilde{Z}_n$, $n \geq 0$, we have

- If $X_2 = 1$ and $X_1 = 1$ then $Z_0 = 1$ and $Z_1 \geq 2$. Hence $\mathbb{P}(\widetilde{Z}_0 = 1 | X_2 = 1, X_1 = 1) = \mathbb{P}(Z_1 = 2 | Z_1 \geq 2) = 0.2/0.3 = 2/3$;
- If $X_2 = 1$ and $X_1 = 0$ then $Z_0 = 2$. Hence $\mathbb{P}(\widetilde{Z}_0 = 1 | X_2 = 1, X_1 = 0) = 0.7$;

Therefore, $\widetilde{Z}_0$ is not conditionally independent of X_1 given $X_2 = 1$, so $\widetilde{X}$ depends on $(X_r)_{r\leq 2}$, even conditionally given $X_2 = 1$. Similar calculations of $\mathbb{P}(\widetilde{Z}_0 = m_0, \ldots, \widetilde{Z}_k = m_k | X_2 = 1, X_1 = i)$, however, do show that $\widetilde{Z}_n$, $n \geq 1$, are independent of $(X_r)_{r\leq 2}$ and of $\widetilde{Z}_0$.

In the following it is useful to use vocabulary associated with a technical component (e.g. a light bulb) that fails at the times of a renewal process X, and each time is replaced immediately by a new component. The Z_n, $n \geq 0$, in the definition of the renewal process are then just the lifetimes of independent identical components. Let us place the observations in Example 89 in a general context. The *age* $A_t = t - T_{X_t}$ of the component at time $t \geq 0$, may contain information about the *excess lifetime* (residual lifetime) $E_t = T_{X_t+1} - t$ of the component at time t, which is the first renewal time $\widetilde{Z}_0$ of the post-t process $\widetilde{X} = (X_{t+s} - X_t)_{s\geq 0}$. In particular, A_t and E_t are not independent, in general, and also the distribution of E_t depends on t. Further components $\widetilde{Z}_n$, $n \geq 1$, are independent identical components.

This motivates the definition of "delayed" renewal processes, where the first renewal time is different from the other inter-renewal times. In other words, the typical renewal behaviour is delayed until the first renewal time. Our main application will be that Z_0 is just part of an inter-renewal time, but it will be useful to have a general definition:

Definition 90 Let Z_n, $n \geq 1$, be independent and identically distributed inter-renewal times and Z_0 independent with a possibly *different* distribution. Then the associated counting process

$$X_t = \#\{n \geq 1 \colon Z_0 + \cdots + Z_{n-1} \leq t\}, \qquad t \geq 0,$$

is called a *delayed renewal process.* We refer to the distribution of Z_0 as the *delay distribution.* A renewal process, where Z_0 has the same distribution as Z_n, $n \geq 1$, is called *undelayed.*

As a corollary to Proposition 87, we have:

Corollary 91 (Renewal property) *Let X be a delayed renewal process, $i \geq 1$, and $T_i = \inf\{t \geq 0 : X_t = i\}$ the ith renewal time. Then $(X_r)_{r\leq T_i}$ and $\widetilde{X} = (X_{T_i+s} - X_{T_i})_{s\geq 0}$ are independent and $\widetilde{X}$ is an undelayed renewal process with the given inter-renewal distribution.*

Just as the Markov property holds at more general stopping times, the renewal property also holds at more general stopping times T, provided that $\mathbb{P}(T \in \{T_n, n \geq 0\}) = 1$. A key example is T_{X_t+1}, the next renewal time after time $t \geq 0$. As with the general strong Markov property, we will not prove the general result, but we prove:

Corollary 92 *Let X be a delayed renewal process and $T = T_{X_t+1}$ the next renewal time after $t \geq 0$. Then $(X_r)_{r\leq T}$ and $\widetilde{X} = (X_{T+s} - X_T)_{s\geq 0}$ are independent and $\widetilde{X}$ is an undelayed renewal process with the given inter-renewal distribution.*

Proof: This proof is not examinable. On $\{X_t = \ell\} = \{Z_0 + \cdots + Z_{\ell-1} \leq t < Z_0 + \cdots + Z_\ell\}$, we have $\widetilde{Z}_n = Z_{\ell+1+n}$, $n \geq 0$, independent and identically distributed and independent of $Z_0, \ldots, Z_\ell$, hence conditionally independent of $(X_r, r \leq T)$ given $X_t = \ell$. In particular, the conditional distribution of $\widetilde{Z}_n$, $n \geq 0$, given $X_t = \ell$ does not depend on ℓ. By Exercise A.1.11, $(X_r)_{r\leq T}$ is independent of $\widetilde{Z}_n$, $n \geq 0$, hence of $\widetilde{X}$. □

Proposition 93 *Given a (possibly delayed) renewal process X, for every $t \geq 0$, the post-t process $\widetilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ is a delayed renewal process with $\widetilde{Z}_0 = E_t$.*

Proof: We apply the renewal property of Corollary 92 to the renewal time $T = T_{X_t+1}$, the first renewal time after t. This establishes that $\widetilde{Z}_n$, $n \geq 1$, are independent identically distributed inter-renewal times, independent from the past, in particular from $\widetilde{Z}_0 = T_{X_t+1} - t = E_t$. Therefore, $\widetilde{X}$ is a delayed renewal process. □

Remark 94 Note that we make no further statement about the dependence of $\widetilde{X}$ on the pre-t process in Proposition 93. We have seen in Example 89 that independence fails, in general.

10.2 Size-biased distributions and stationarity

In this section, we will show that for each inter-renewal distribution on $(0, \infty)$, there is a special delay distribution of Z_0 for which the delayed renewal process X with the given inter-renewal distribution has stationary increments, i.e. such that the distribution of $X_{t+s} - X_t$ is independent of $t \geq 0$, for each $s \geq 0$.

Proposition 95 *Let X be a delayed renewal process. Then the distribution of the excess lifetime $E_t = T_{X_t+1} - t$ does not depend on $t \geq 0$ if and only if X has stationary increments.*

Proof: By Proposition 93, the post-t process $\widetilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ is a delayed renewal process with $\widetilde{Z}_0 = E_t$. If the delay distribution of $\widetilde{X}$ does not depend on t, then the distribution of $\widetilde{X}$ does not depend on t. In particular, the distribution of $\widetilde{X}_s = X_{t+s} - X_t$ does not depend on t. Conversely, if X has stationary increments, $\mathbb{P}(E_t \leq s) = \mathbb{P}(X_{t+s} - X_t \geq 1)$ does not dependent on $t \geq 0$, for all $s \geq 0$. Hence the distribution of E_t does not depend on $t \geq 0$. $\square$

It can be shown that $E = (E_t, t \geq 0)$ is a Markov process on the uncountable state space $\mathbb{S} = (0, \infty)$, so X has stationary increments if and only if E is a stationary Markov process. We will study an example of a discrete inter-renewal distribution, where the process $(E_n, n \geq 0)$ is a discrete-time Markov chain in a countable state space $\mathbb{S} \subseteq \mathbb{N}\backslash\{0\}$. This example will motivate the definition of discrete size-biased distributions, which have natural analogues in continuous size-biased distributions. We will use size-biased distributions to specify the (unique) special delay distribution, which finally gives rise to a delayed renewal process with stationary increments, by Proposition 95.

Example 96 In Example 89, consider $E_n = T_{X_n+1} - n \in \mathbb{N} \setminus \{0\}$, $n \geq 0$. It is not hard to see that $(E_n, n \geq 0)$ is a discrete-time Markov chain on $\{1, 2, \ldots, 19\}$ with transition probabilities

$$\pi_{i,i-1} = 1, \quad i \geq 2, \qquad \pi_{1,j} = \mathbb{P}(Z_1 = j), \quad j \in \{1, 2, 19\},$$

which is clearly irreducible and positive recurrent since $\mathbb{E}_1(H_1) = \mathbb{E}(Z_1) < \infty$, so it has a unique stationary distribution. This stationary distribution can be calculated in a general setting of $\mathbb{N}$-valued inter-renewal times with finite mean, by solving $\eta\Pi = \eta$, see Exercise A.6.5.

Let us here present a heuristic argument to find η. The ergodic theorem for discrete-time Markov chains establishes the stationary distribution as asymptotic proportions. Let us first look at the simpler question of what proportion of times $n \in \mathbb{N}$ fall into inter-renewal times of length 19. On average, 10 renewals give 7x1, 2x2 and 1x19 and cover 30 time units, of which

- 19 of the $n \in \{0, \ldots, 29\}$ fall into the 1 inter-renewal time of length 19, i.e. 19x1,
- 2 fall into each of the 2 inter-renewal times of length 2, i.e. 2x2,
- 1 falls into each of the 7 inter-renewal times of length 1, i.e. 1x7.

The proportions describe the size-biased distribution associated with the inter-renewal distribution, with probabilities 0.1, 0.2 and 0.7 weighted by the sizes 19, 2 and 1, respectively. At a renewal, E_n jumps to a new inter-renewal time and then descends to 1, so that of the 30 steps,

- 1 step each is in state $19, \ldots, 3$,
- 1+2 steps are in state 2,
- 1+2+7 steps are in state 1.

The proportion in each state reflects the number of inter-renewal times greater than the state. A random variable with a distribution according to these proportions can be thought of as a uniformly distributed fraction of an inter-renewal time with the size-biased distribution.

Variations of the size-biasing effect appear under the names of *inspection paradox* and *waiting time paradox* on Assignment Sheet 4.

Definition 97 For a probability function p on $\mathbb{N}$ with $\mu = \sum_{n\geq 0} np_n \in (0,\infty)$ we associate the *size-biased probability function*

$$p_n^{\rm sb} = \frac{np_n}{\mu}, \qquad n \geq 1.$$

With a probability density function f on $(0,\infty)$ with $\mu = \int_0^\infty tf(t)dt < \infty$, we associate the *size-biased probability density function*

$$f^{\rm sb}(z) = \frac{zf(z)}{\mu}, \qquad z \in (0,\infty).$$

We say that a random variable L with probability function $p^{\rm sb}$, respectively with probability density function $f^{\rm sb}$, has the *(associated) size-biased distribution.*

Example 98 Let p be a probability function to model the number of children in a random family of a given population. Ask a random child "How many children are in your family"? The answer has distribution $p^{\rm sb}$, since a family with n children is n times more likely to be sampled.

Proposition 99 *Let X be a delayed renewal process with any inter-renewal distribution, and so that $Z_0 \sim LU$ for a random variable L with the associated size-biased distribution and $U \sim$* Unif$(0,1)$ *independent. Then X has stationary increments, and $E_t \sim LU$ for all $t \geq 0$.*

Proof: A proof is in the optional Exercise A.6.6. □

Example 100 For the inter-arrival distribution Exp(λ) of PP(λ), we have $\mu = 1/\lambda$ and get

$$f^{\rm sb}(z) = \frac{zf(z)}{\mu} = \lambda^2 ze^{-\lambda z}, \qquad z \geq 0,$$

the probability density function of a Gamma$(2,\lambda)$ distribution, the distribution of the sum of two independent Exp(λ) random variables. It was an exercise in Part A to show that $LU \sim$ Exp(λ). This confirms that PP(λ) has stationary increments when $Z_0 \sim$ Exp(λ).

In general, we can now calculate the distribution of a uniform proportion of a random variable L with the size-biased distribution: in the continuous case, for independent $U \sim$ Unif$(0,1)$,

$$\mathbb{P}(LU > y) = \int_0^1 \int_{y/u}^\infty \frac{z}{\mu} f(z)dzdu = \int_y^\infty \int_{y/z}^1 du \frac{z}{\mu} f(z)dz = \frac{1}{\mu}\int_y^\infty (z-y)f(z)dz, \quad y \in (0,\infty),$$

and we just differentiate to get

$$f_0(y) := -\frac{d}{dy}\mathbb{P}(LU > y) = \frac{1}{\mu}yf(y) + \frac{1}{\mu}\overline{F}(y) - \frac{1}{\mu}yf(y) = \frac{1}{\mu}\overline{F}(y), \qquad y \in (0,\infty).$$

Example 96 is for $\mathbb{N}$-valued inter-renewal times, where the renewal process $(X_t, t \geq 0)$ has integer renewal times, so it made sense to consider the discretised processes $(X_n, n \geq 0)$ and $(E_n, n \geq 0)$. For continuous inter-renewal distributions, we can still calculate the asymptotic proportion of time that the excess life E_s exceeds a threshold $y \in (0,\infty)$. Here is a rigorous calculation for an undelayed renewal process X with mean inter-renewal time $\mu \in (0,\infty)$:

$$\frac{1}{T_k}\int_0^{T_k} 1_{\{E_s > y\}} ds = \frac{k}{T_k}\,\frac{1}{k}\sum_{j=0}^{k-1}(Z_j - y)1_{\{Z_j > y\}}$$

allows the application of the Strong Law of Large Numbers twice, to independent and identically distributed inter-renewal times Z_j, $j \geq 0$, with partial sums $T_k = Z_0 + \cdots + Z_{k-1}$, $k \geq 1$, and to independent and identically distributed $Y_j = (Z_j - y)1_{\{Z_j > y\}}$, $j \geq 0$, to obtain almost sure convergence as $k \to \infty$ to

$$\frac{1}{\mathbb{E}(Z_1)}\mathbb{E}(Y_1) = \frac{1}{\mu}\int_y^\infty (z-y)f(z)dz = \mathbb{P}(LU > y).$$

If we had developed a theory of $(0,\infty)$-valued Markov process including an ergodic theorem, which applies to $(E_t, t \geq 0)$, this calculation would identify the distribution of LU as a stationary distribution of $(E_t, t \geq 0)$, and would hence prove Proposition 99.

Lecture 11

Convergence to equilibrium – renewal theorems

Reading: Grimmett-Stirzaker 10.4

For an irreducible positive recurrent Markov chain $(X_t, t \geq 0)$, there is a stationary distribution ξ. If $X_0 \sim \xi$, then $X_t \sim X_0 \sim \xi$ for all $t \geq 0$. Furthermore, X_t converges to this stationary distribution as $t \to \infty$, when $X_0 \sim \nu$ for any initial distribution ν.

For any inter-renewal distribution with finite mean, there is a delay distribution for which the renewal process $(X_t, t \geq 0)$ has *stationary increments*. With this delay distribution $X_{t+s} - X_t \sim X_s$ for all $t \geq 0$. We will see today that $X_{t+s} - X_t$ converges to this stationary increment distribution as $t \to \infty$, for any delay distribution.

For discrete-time Markov chains, the convergence theorem assumes aperiodicity. We will discuss a similar condition for lattice-valued inter-renewal times.

11.1 Convergence to equilibrium

Theorems in this section and the next are stated separately for two cases. These are essentially for discrete and continuous inter-renewal distributions, respectively. More precisely, we will distinguish lattice-valued and non-lattice-valued distributions. The standard terminology in this context is "arithmetic" and "non-arithmetic".

Definition 101 A positive random variable Z (and its distribution) is called non-arithmetic if

$$\{r > 0 : \mathbb{P}(Z \in r\mathbb{N}) = 1\} = \varnothing.$$

Otherwise, it is called arithmetic, and d-arithmetic, if $d = \sup\{r > 0 : \mathbb{P}(Z \in r\mathbb{N}) = 1\}$.

Example 102

- All continuous random variables are non-arithmetic.
- All integer-valued random variables Z are arithmetic, and 1-arithmetic if furthermore $\mathbb{P}(Z \in k\mathbb{N}) < 1$ for all $k \geq 2$.
- Let $B \sim \text{Binomial}(1, p)$. Then $X = 1 + B$ is 1-arithmetic and $Y = \sqrt{2}(1 + B)$ is $\sqrt{2}$-arithmetic. $Z = 1 + aB$ is non-arithmetic if and only if $a \geq 0$ is irrational. If $a = p/q$ is rational, then Z is $1/q$-arithmetic, provided that $p, q \in \mathbb{N}$, $q \geq 1$, with $\gcd(p, q) = 1$.

We will apply this concept to inter-renewal times. We will formulate results for non-arithmetic and 1-arithmetic inter-renewal times. These are the two most relevant cases. The 1-arithmetic case can be easily generalised to d-arithmetic inter-renewal times.

In the sequel we will use the following notation. Let X be a renewal process, which may have any delay distribution (or X may be undelayed). We will assume that its inter-renewal times have finite mean $\mu = \mathbb{E}(Z_1)$. Let L be a *size-biased inter-renewal time*, i.e. L has the size-biased distribution associated with the inter-renewal distribution of Z_1, and let $U \sim \text{Unif}(0, 1)$ be independent. We saw in Proposition 99 that $\widetilde{Z}_0 = LU$ is such that a renewal process $\widetilde{X}$ with delay LU has stationary increments. While this holds for non-arithmetic and arithmetic inter-renewal times, this process only features in the following convergence theorems in the non-arithmetic case.

In the 1-arithmetic case with probability function p_n, $n \geq 1$, we will meet a discretised version $\widehat{Z}_0$ of $\widetilde{Z}_0$, which is uniformly distributed on $\{1, \ldots, L\}$. More precisely,

$$\mathbb{P}(\widehat{Z}_0 = k) = \sum_{\ell=k}^{\infty} \mathbb{P}(L = \ell)\frac{1}{\ell} = \sum_{\ell=k}^{\infty} \frac{\ell p_\ell}{\mu}\frac{1}{\ell} = \frac{1}{\mu}\mathbb{P}(Z_1 \geq k), \qquad k \geq 1. \tag{1}$$

It is easy to see that a renewal process $\widehat{X} = (\widehat{X}_t, t \geq 0)$ with delay $\widehat{Z}_0$ has stationary increments $X_{n+s} - X_n$, $n \geq 0$, for all $s > 0$.

Recall definitions $A_t = t - T_{X_t}$ and $E_t = T_{X_t+1} - t$ for the age and excess life of the component in use at time $t \geq 0$. We obtain convergence for increments, ages and excess lives:

Theorem 103 (Convergence in distribution) *Let X be a (possibly delayed) renewal process and assume that the inter-renewal times have finite mean μ.*

(a) *If the inter-renewal distribution is non-arithmetic, then*

$$X_{t+s} - X_t \to \widetilde{X}_s \qquad \text{in distribution, as } t \to \infty,$$

where $\widetilde{X}$ is an associated stationary renewal process.

Also $(A_t, E_t) \to (L(1-U), LU)$ in distribution, as $t \to \infty$, where L is a size-biased inter-renewal time, and $U \sim \mathrm{Unif}(0,1)$ is independent of L.

(b) *If the inter-renewal distribution is 1-arithmetic, then*

$$X_{n+s} - X_n \to \widehat{X_s} \qquad \text{in distribution, as } n \to \infty.$$

where $\widehat{X}$ is an associated delayed renewal process with $\widehat{Z}_0$ as in (1). Also, $(A_n, E_n) \to (L - \widehat{Z}_0, \widehat{Z}_0)$ in distribution, as $n \to \infty$.

We will give a sketch of the coupling proof for the arithmetic case below.

11.2 Renewal theorems

The renewal theorems are analogues of Theorem 103 for the convergence of certain moments. The renewal theorem itself concerns means. It is a refinement of the Elementary Renewal Theorem to increments.

Theorem 104 (Renewal theorem) *Let X be a (possibly delayed) renewal process having inter-renewal times with finite mean μ and renewal function $m(t) = \mathbb{E}(X_t)$.*

(a) *If the inter-renewal times are non-arithmetic, then for all $h \geq 0$*

$$m(t+h) - m(t) \to \frac{h}{\mu} \qquad \text{as } t \to \infty.$$

(b) *If the inter-renewal times are 1-arithmetic, then for all $h \in \mathbb{N}$*

$$m(t+h) - m(t) \to \frac{h}{\mu} \qquad \text{as } t \to \infty.$$

As a generalisation that is often useful in applications, there is the key renewal theorem:

Theorem 105 (Key renewal theorem) *Let X be a renewal process with continuous inter-renewal distribution and $m(t) = \mathbb{E}(X_t)$. If $g : [0,\infty) \to [0,\infty)$ is Riemann integrable, then*

$$(g * m')(t) = \int_0^t g(t-x)m'(x)dx \to \frac{1}{\mu}\int_0^\infty g(x)dx \qquad \text{as } t \to \infty.$$

The integrability condition on g is somewhat delicate (direct Riemann integrability on $[0,\infty)$ is the precise technical term that many authors use), but we skip the details. The restrictions on the inter-renewal distribution can also be relaxed (at the cost of some limits through discrete lattices in the d-arithmetic case), but the convolution with m' needs to be rephrased, essentially by replacing $m'(x)dx$ by an integral with respect to the Stieltjes measure associated with m. Even the case $\mu = \infty$ can be shown to correspond to zero limits. We will not use such generalisations.

Example 106 If we take $g(x) = g_h(x) = 1_{[0,h]}(x)$ in the Key renewal theorem, we get

$$\int_0^t g(t-x)m'(x)dx = \int_{t-h}^t m'(x)dx = m(t) - m(t-h)$$

and this allows to deduce the renewal theorem from the Key renewal theorem.

The renewal theorem and the key renewal theorem should be thought of as results where time windows $[t-h,t]$ or $[t,t+h]$ are sent to infinity, and a stationary picture is obtained in the limit. In the case of the renewal theorem, we are only looking at the mean of an increment. In the key renewal theorem, we can consider other quantities related to the mean behaviour in a window. We will see that moments of excess lifetimes fall into this category. Note e.g. that $\{E_t \leq h\}$ is a quantity only depending on X in the time window $[t,t+h]$.

The key renewal theorem can be deduced from the renewal theorem by approximating g by step functions.

11.3 The coupling proofs

This section is not examinable. Suppose that X is a renewal process with integer-valued inter-renewal times, and suppose that $\mathbb{P}(Z_1 \in d\mathbb{N}) < 1$ for all $d \geq 2$, i.e. suppose that Z_1 is 1-arithmetic. Let $\widehat{X}$ be a renewal process with delay as in (1). It will be useful to discretise $\widehat{X}$ to integer times: $(\widehat{X}_n, n \geq 0)$ has stationary increments.

Suppose that X and $\widehat{X}$ are independent. Define $N = \inf\{n \geq 1 : T_n = \widehat{T}_n\}$ the index of the first simultaneous renewal (same index for both, to simplify the argument). A third process is constructed

$$\overline{X}_t = \begin{cases} X_t & \text{if } t < T_N, \\ \widehat{X}_t & \text{if } t \geq T_N, \end{cases} \qquad \text{with } \overline{Z}_j = \begin{cases} Z_j & \text{if } j < N, \\ \widehat{Z}_j & \text{if } j \geq N. \end{cases}$$

The following three steps complete the essence of the proofs:

1. $\mathbb{P}(N < \infty) = 1$ – this is because N is the first time that the random walk $S_n = T_n - \widehat{T}_n$ hits 0. While S is not a simple random walk, it can be shown that it is recurrent (hence hits 0 in finite time) since $\mathbb{E}(Z_1 - \widehat{Z}_1) = 0$.
2. $\overline{X}$ is a renewal process with the same distribution as X.
3. $\sup_B \left|\mathbb{P}((\overline{X}_{n+s} - \overline{X}_n)_{s\geq 0} \in B) - \mathbb{P}(\widehat{X} \in B)\right| \leq \mathbb{P}(T_N > n) \to 0$, as $n \to \infty$, where the supremum is over all (suitably measurable) $B \subset \{f : [0,\infty) \to \mathbb{N}\}$.

This shows convergence of the distribution of $(X_{t+s} - X_t)_{s\geq 0}$ to the distribution of $\widehat{X}$ in *total variation*, which is stronger than convergence in distribution. In particular, for $B_{s,k} = \{f : [0,\infty) \to \mathbb{N} : f(s) = k\}$ we have $\{X_{n+s} - X_n = k\} = \{(X_{n+u} - X_n)_{u\geq 0} \in B_{s,k}\}$ and conclude the proof of Theorem 103(b), as far as convergence in distribution of increments are concerned.

For the convergence of excess lives, we note that $\{E_t \le r\} = \{X_{t+r} - X_t \ge 1\}$, while joint distributions with ages can be obtained $\{A_t \le a, E_t \le r\} = \{X_t - X_{t-a} \ge 1, X_{t+r} - X_t \ge 1\}$ using two consecutive increments of $(X_{t-a+u} - X_{t-a})_{u\ge 0}$.

For the Renewal Theorem, Theorem 104(b), another argument, beyond the convergence in 3., is needed to see that the means of $X_{t+s} - X_t$ converge as $t \to \infty$.

For the non-arithmetic case, the proof is harder, since $N = \infty$ for N as defined above, and times $N_\varepsilon = \inf\{n \ge 1 : |T_n - \widehat{T}_n| < \varepsilon\}$ do not achieve a perfect coupling.

There is also an alternative proof using renewal equations.

11.4 Example

Let us investigate the asymptotics of $\mathbb{E}(E_t^r)$, as $t \to \infty$. We condition on the last renewal time before t, distinguishing the events that this is the kth renewal time with density $f^{*(k)}$, $k \ge 1$. Since in the kth case, we have $E_t = T_{k+1} - t = Z_k - (t - T_k)$, we get

$$\begin{aligned}
\mathbb{E}(E_t^r) &= \mathbb{E}\left(E_t^r \sum_{k=0}^{\infty} 1_{\{X_t = k\}}\right) \\
&= \mathbb{E}(((Z_0 - t)^+)^r) + \sum_{k\ge 1} \int_0^t \mathbb{E}(((Z_k - (t-x))^+)^r) f^{*(k)}(x)dx \\
&= \mathbb{E}(((Z_0 - t)^+)^r) + \int_0^t \mathbb{E}(((Z_1 - (t-x))^+)^r) m'(x)dx.
\end{aligned}$$

The first term tends to 0 as $t \to \infty$. For the second term, let us write $g(y) = \mathbb{E}(((Z_1 - y)^+)^r)$. This is a clearly a nonnegative Riemann-integrable function of y if $\mathbb{E}(Z_1^{r+1}) < \infty$ (see below). The Key Renewal Theorem gives

$$\mathbb{E}(E_t^r) = \mathbb{E}(((Z_0 - t)^+)^r) + (g * m')(t) \to \frac{1}{\mu}\int_0^\infty g(y)dy, \qquad \text{as } t \to \infty.$$

We can now calculate

$$\begin{aligned}
\frac{1}{\mu}\int_0^\infty g(y)dy &= \frac{1}{\mu}\int_0^\infty \mathbb{E}(((Z_1 - x)^+)^r)dx \\
&= \frac{1}{\mu}\int_0^\infty \int_x^\infty (z-x)^r f(z)dzdx \\
&= \frac{1}{\mu}\int_0^\infty \int_0^\infty y^r f(y+x)dydx \\
&= \frac{1}{\mu}\int_0^\infty y^r \int_0^\infty f(y+x)dxdy \\
&= \frac{1}{\mu}\int_0^\infty y^r \overline{F}(y)dy \\
&= \frac{\mathbb{E}(Z_1^{r+1})}{(r+1)\mu},
\end{aligned}$$

where the last step is by integration by parts. It is now easy to check that, in fact, these are the moments of the limit distribution LU for the excess life E_t.

Lecture 12

Ruin theory

Reading: Ross 7.10; CT4 Unit 6
Further reading: Norris 5.3

The last four lectures will be on applications. We will apply the general theory developed for Markov chains and renewal processes to more specific processes that arise in certain applications. This will go a bit beyond the scale of the examples that have complemented the theory so far.

12.1 The insurance ruin model

Insurance companies deal with large numbers of insurance policies at risk. They are grouped according to type and various other factors into so-called portfolios. Let us focus on such a portfolio and model the associated claim processes, the claim sizes and the reserve process. We make the following assumptions.

- Claims arrive according to a Poisson process $X = (X_t, t \geq 0)$ with rate $\lambda > 0$.
- Claim amounts A_j, $j \geq 1$, are positive, independent identically distributed, independent of X, with probability density function $k(a)$, $a > 0$, and finite mean $\mu = \mathbb{E}(A_1) < \infty$.
- The insurance company provides an initial reserve of $x \geq 0$ money units.
- Premiums are paid continuously at constant rate c generating a linear premium income accumulating to ct at time t.
- Premiums are added to the reserve, while claim amounts are deducted. We mostly assume $c > \lambda\mu$ to have more premium income than claim outgo, on average.
- We ignore all expenses and other influences.

In this setting, we define the following objects of interest:

- the aggregate claims process $C = (C_t, t \geq 0)$, where $C_t = \sum_{n=1}^{X_t} A_n$, $t \geq 0$;
- the reserve process $R = (R_t, t \geq 0)$, where $R_t = x + ct - C_t$, $t \geq 0$;
- the time $T_{\text{ruin}} = \inf\{t \geq 0 \colon R_t < 0\}$ of technical ruin;
- the probability $\psi(x) = \mathbb{P}_x(T_{\text{ruin}} < \infty)$, as a function of $R_0 = x \geq 0$.

12.2 Aggregate claims and reserve processes

Proposition 107 *The processes C and R have stationary independent increments. Their moment generating functions are given by*

$$\mathbb{E}(e^{\gamma C_t}) = \exp\left(\lambda t \left(\mathbb{E}(e^{\gamma A_1} - 1)\right)\right) = \exp\left(\lambda t \int_0^\infty (e^{\gamma a} - 1)k(a)da\right)$$

and

$$\mathbb{E}(e^{\beta R_t}) = \exp\left(\beta x + \beta ct - \lambda t \left(1 - \mathbb{E}(e^{-\beta A_1})\right)\right) = \exp\left(\beta x + \beta ct - \lambda t \int_0^\infty (1 - e^{-\beta a})k(a)da\right).$$

Proof: We first calculate the moment generating function of C_t by conditioning on X_t:

$$\begin{aligned}\mathbb{E}(e^{\gamma C_t}) &= \mathbb{E}\left(\exp\left(\gamma\sum_{j=1}^{X_t}A_j\right)\right)\\ &= \sum_{n=0}^{\infty}\mathbb{E}\left(\exp\left(\gamma\sum_{j=1}^{n}A_j\right)\right)\mathbb{P}(X_t=n)\\ &= \sum_{n=0}^{\infty}\left(\mathbb{E}\left(e^{\gamma A_1}\right)\right)^n\mathbb{P}(X_t=n)\\ &= \exp\left(\lambda t\left(\mathbb{E}\left(e^{\gamma A_1}\right)-1\right)\right),\end{aligned}$$

which in the case where A_1 has a density k, also gives the second formula required. The same calculation for the joint moment generating function $\mathbb{E}(e^{-\gamma C_t-\delta(C_{t+s}-C_t)})$ of C_t and $C_{t+s}-C_t$, conditioning on X_t and on $X_{t+s}-X_t$, and similarly for more increments, yields stationarity and independence of increments (only using the stationarity and independence of increments of X, and the independence of the A_j, $j\geq 1$).

The statements for R follow easily since $R_t=x+ct-C_t$, $t\geq 0$, setting $\gamma=-\beta$. □

We can use the moment generating function to calculate moments and the stationary independent increments give us independent identically distributed random variables to which the Strong Law of Large Numbers applies:

Example 108 We differentiate the moment generating functions at zero to obtain

$$\mathbb{E}(C_t)=\frac{\partial}{\partial\gamma}\exp\left\{\lambda t\left(\mathbb{E}\left(e^{\gamma A_1}\right)-1\right)\right\}\bigg|_{\gamma=0}=\lambda t\,\frac{\partial}{\partial\gamma}\mathbb{E}\left(e^{\gamma A_1}\right)\bigg|_{\gamma=0}=\lambda t\mu.$$

and $\mathbb{E}(R_t)=x+ct-\lambda t\mu=x+(c-\lambda\mu)t$. Now the Strong Law of Large Numbers can be applied to the independent and identically distributed increments $Z_j=R_j-R_{j-1}$, $j\geq 1$ to give

$$\frac{R_n}{n}=\frac{x}{n}+\frac{1}{n}\sum_{j=1}^{n}Z_j\to\mathbb{E}(Z_1)=c-\lambda\mu>0\qquad\text{a.s., as } n\to\infty.$$

This confirms that $c>\lambda\mu$ means that, in a long-term average, there is more premium income than claim outgo. In particular, this implies that $R_n\to\infty$ a.s. as $n\to\infty$. As in Example 72 (Strong Law for Poisson processes), this does not directly imply that $R_t/t\to c-\lambda\mu$ a.s. nor that $R_t\to\infty$ a.s. as $t\to\infty$. It is conceivable that between integers, the reserve process takes larger and larger negative values. But this does not actually happen, and we will show that $R_t\to\infty$ a.s. as $t\to\infty$.

There are other random walks that are also embedded in the reserve process, and we shall now use one of these to get a first asymptotic result about ruin probabilities:

Example 109 Consider the process at claim times $W_n=R_{T_n}$, $n\geq 0$, where T_n, $n\geq 1$, are the arrival times of the Poisson process X, and $T_0=0$. Now

$$W_j-W_{j-1}=R_{T_j}-R_{T_{j-1}}=c(T_j-T_{j-1})-A_j,\qquad j\geq 1,$$

are also independent identically distributed increments with $\mathbb{E}(W_j-W_{j-1})=c/\lambda-\mu$, and the Strong Law of Large Numbers yields

$$\frac{W_n}{n}=\frac{x}{n}+\frac{1}{n}\sum_{j=1}^{n}(W_j-W_{j-1})\to c/\lambda-\mu\qquad\text{a.s. as } n\to\infty.$$

Again, we conclude $W_n \to \infty$, provided that $c > \lambda\mu$, but note that W_n are the local minima of R, and indeed

$$T_{X_t+1} > t \geq T_{X_t} \quad \Rightarrow \quad R_t \geq R_{T_{X_t}} = W_{X_t} \to \infty \quad \text{a.s. as } t \to \infty \quad \text{since } X_t \to \infty.$$

But $R_t \to \infty$ implies that $R_t^0 := ct - C_t = R_t - x \to \infty$ a.s., and so

$$I_\infty := \inf\{R_t^0 : 0 \leq t < \infty\} > -\infty \qquad \text{a.s..}$$

As a consequence, $c > \lambda\mu$ yields

$$\begin{aligned}\psi(\infty) := \lim_{x\to\infty} \psi(x) &= \lim_{x\to\infty} \mathbb{P}_x(R_t < 0 \text{ for some } t \geq 0) \\ &= \lim_{x\to\infty} \mathbb{P}(R_t^0 < -x \text{ for some } t \geq 0) = \lim_{x\to\infty} \mathbb{P}(I_\infty < -x) = 0.\end{aligned}$$

In other words, for any $\varepsilon > 0$, there is an initial capital $x > 0$, for which the ruin probability $\psi(x)$ is at most ε.

12.3 Ruin probabilities

We now turn to studying the ruin probabilities $\psi(x)$ for finite x, as a function of $x \geq 0$.

Proposition 110 *Let $c > \lambda\mu$. Then the ruin probabilities $\psi(x)$ satisfy the renewal-type equation*

$$\psi = g + \psi * f, \qquad i.e. \qquad \psi(x) = g(x) + \int_0^x \psi(x-y)f(y)dy, \qquad x \geq 0,$$

where

$$f(y) = \frac{\lambda}{c}\overline{K}(y) = \frac{\lambda}{c}\int_y^\infty k(x)dx \qquad \text{and} \qquad g(x) = \frac{\lambda\mu}{c}\overline{K}_0(x) = \frac{\lambda}{c}\int_x^\infty \overline{K}(y)dy.$$

using notation $\overline{K}$ and $\overline{K}_0$ associated with k in analogy with tail integrals $\overline{F}$ and $\overline{F}_0$ associated with f and f_0 at the end of Lecture 11.

Proof: We first note that the functions f and g satisfy

$$f'(y) = -\frac{\lambda}{c}k(y), \quad f(0) = \frac{\lambda}{c}, \qquad g'(x) = -f(x).$$

We condition on $T_1 \sim \text{Exp}(\lambda)$ and on A_1, which has probability density function k, to obtain

$$\psi(x) = \int_0^\infty \int_0^\infty \psi(x + ct - a)k(a)da\lambda e^{-\lambda t}dt = \int_x^\infty \frac{\lambda}{c}e^{-(s-x)\lambda/c}\int_0^\infty \psi(s-a)k(a)dads$$

since $(R_{T_1+s}, s \geq 0)$ is a reserve process starting from $x + cT_1 - A_1$ evolving independently of T_1 and A_1, if $R_0 = x$, and where we use the convention that $\psi(x) = 1$ for $x < 0$.

We multiply by $e^{-\lambda x/c}$, differentiate with respect to x and cancel $e^{-\lambda x/c}$ to find

$$\psi'(x) = \frac{\lambda}{c}\psi(x) - \frac{\lambda}{c}\int_0^\infty \psi(x-a)k(a)da = \frac{\lambda}{c}\psi(x) - \frac{\lambda}{c}\int_0^x \psi(x-a)k(a)da - \frac{\lambda}{c}\overline{K}(x).$$

We use notation f and g to write this as

$$\psi'(x) = g'(x) + \psi(x)f(0) + \int_0^x \psi(a)f'(x-a)da,$$

where the last two terms are the derivative of $(\psi * f)(x)$. Integration yields $\psi = g + \psi * f + b$ for some $b \in \mathbb{R}$. Finally, we have $\psi(\infty) = 0$, so we investigate $x \to \infty$ in

$$g(x) + (\psi * f)(x) = \frac{\lambda}{c}\int_x^\infty \overline{K}(y)dy + \int_0^x \psi(x-y)f(y)dy \to 0,$$

using $\mu = \int_0^\infty \overline{K}(y)dy < \infty$ and dominated convergence, so that $b = 0$, as required. $\square$

Example 111 In the setting of the proposition, we can calculate $\psi(0) = g(0) = \lambda\mu/c$. In particular, there is positive probability to avoid ruin when starting from zero initial reserve. In other words, $x \mapsto \psi(x)$ jumps at $x = 0$ from $\psi(0-) = 1$ to $\psi(0) = \psi(0+) = \lambda\mu/c < 1$.

Corollary 112 *If $c > \lambda\mu$, then ψ is given by*

$$\psi = g + g * m' \qquad \textit{where} \quad m'(y) = \sum_{n=1}^{\infty} f^{*(n)}(y).$$

Proof: This is an application of Exercise A.6.4(a)-(b), the general solution of renewal equations. Note that f is not a probability density for $\lambda\mu < c$, but the results (and arguments) are valid for nonnegative f with $\int_0^\infty f(y)dy \leq 1$. □

Lecture 13

Asymptotics of ruin probabilities and queueing models

In this lecture, we will use the Key Renewal Theorem to study the asymptotic behaviour of ruin probabilities. We will also begin the second area of applications by studying departure times in queueing models.

13.1 Asymptotic behaviour of ruin probabilities

In Proposition 110, we showed that ruin probabilities satisfy a renewal-type equation. However, we did not identify a corresponding renewal process. Indeed, the framework was such that the function f taking the role of the inter-renewal density does not integrate to 1, but to $\mu\lambda/c < 1$. In fact, we can associate a *defective renewal process* that only counts a geometrically distributed number of renewals. Each time there is probability $1 - \mu\lambda/c$ that there are no further renewals.

Let us look at the reserve process $R_t^0 = ct - C_t$ in the case $c = \lambda\mu$. The process R^0 is defined in terms of a Poisson process X (or indeed its independent inter-arrival times $Z_n \sim \text{Exp}(\lambda)$, $n \geq 0$) and independent and identically distributed claim amounts A_n, $n \geq 1$. In this setting, $f(y) = \frac{1}{\mu}\mathbb{P}(A_1 > y)$, $y > 0$, is a probability density function. Consider $J_0 := 0$ and the times of new minima of R^0

$$J_m := \inf\{t \geq J_{m-1} \colon R_t^0 < R_{J_{m-1}}^0\}, \quad m \geq 1.$$

Lemma 113 *For each $m \geq 1$, the post-J_m process $\widetilde{R}^0 = (R_{J_m+s}^0 - R_{J_m}^0, s \geq 0)$ is independent of $(R_t^0, 0 \leq t \leq R_{J_m})$ and has the same distribution as R^0.*

Proof: This proof is not examinable. We argue as in the proof of Corollary 92 that the inter-arrival times and here also associated claim sizes of $\widetilde{R}^0$, are conditionally independent from $(R_t^0, 0 \leq t \leq R_{J_m})$ given $X_{J_m} = \ell$, and indeed the conditional distribution of $\widetilde{R}^0$ is the distribution of R^0 and does not depend on $\ell \geq 0$. By Exercise A.1.11, independence and unconditional distribution follow. □

As a consequence, the process

$$N_s = \#\{m \geq 1 \colon -R_{J_m} \leq s\}, \qquad s \geq 0,$$

that is counting new minima up to depth s is a renewal process, with inter-renewal times $R_{J_m} - R_{J_{m+1}}$, $m \geq 0$. Each inter-renewal time is a partial claim size, since the reserve process attains new minima by jumping from above the old minimum to any new minimum. In fact, it turns out to be a uniform proportion LU of a size-biased claim size L, where $U \sim \text{Unif}((0,1))$. Intuitively, this is, because big claims are more likely to exceed the previous minimal reserve level, hence size-biased L, but the previous level will only be exceeded by a fraction LU, since R will not be at its minimum when the claim arrives. By calculations at the end of Lecture 12, LU has indeed probability density function $\frac{1}{\mu}\int_x^\infty k(a)da$, which is the $f(x)$ of Proposition 110, when considering $c = \lambda\mu$ instead of $c > \lambda\mu$. We will not prove that new minima indeed exceed old minima by a random amount distributed as LU, but turn to the case $c > \lambda\mu$ now.

When $c > \lambda\mu$, there will only be a finite number of claims that exceed the previous minimal reserve level since now $R_t^0 \to \infty$ a.s., and N_s remains constant for depths s exceeding $-I_\infty$, where $I_\infty = \inf\{R_t^0, t \geq 0\}$ is the global minimum of R^0.

To conclude, we specify the precise tail behaviour of $\psi(x)$ as $x \to \infty$.

Proposition 114 *Assume that $c > \lambda\mu$ and that there is $\alpha > 0$ such that*

$$1 = \int_0^\infty e^{\alpha y} f(y)dy \qquad \textit{and} \qquad C := \int_0^\infty e^{\alpha y} g(y)dy < \infty.$$

in the notation of Proposition 110. Then

$$\psi(x) \sim Ce^{-\alpha x} \qquad \textit{as } x \to \infty.$$

Proof: Define a probability density function $\hat{f}(y) = e^{\alpha y} f(y)$, and $\hat{g}(y) = e^{\alpha y} g(y)$ and $\hat{\psi}(x) = e^{\alpha x}\psi(x)$. Then $\hat{\psi}(x)$ satisfies

$$\hat{\psi}(x) = \hat{g}(x) + \int_0^x \hat{\psi}(x-y)\hat{f}(y)dy.$$

The solution (obtained as in Corollary 112) converges by the key renewal theorem:

$$\hat{\psi}(x) = \hat{g}(x) + (\hat{g} * \hat{m}')(x) \to \frac{1}{\hat{\mu}} \int_0^\infty \hat{g}(y)dy = C \quad \text{as } x \to \infty, \qquad \text{where } \hat{m}'(x) = \sum_{n \geq 1} \hat{f}^{*(n)}(x).$$

Note that we assume that $\hat{g}$ is integrable. [The technical conditions for direct Riemann integrability are easy to check since $\hat{g}(x)$ is a product of $e^{\alpha x}$ and the decreasing function $g(x) = \int_x^\infty f(y)dy$, $x > 0$.] □

Example 115 If $A_n \sim \text{Exp}(1/\mu)$, then in the notation of Proposition 110

$$g(x) = \frac{\lambda\mu}{c} e^{-x/\mu} \qquad \text{and} \qquad f(y) = \frac{\lambda}{c} e^{-y/\mu}$$

so that the renewal equation becomes

$$e^{x/\mu}\psi(x) = \frac{\lambda\mu}{c} + \frac{\lambda}{c} \int_0^x \psi(y)e^{y/\mu}dy.$$

In particular $\psi(0) = \lambda\mu/c$. After differentiation and cancellation

$$\psi'(x) = \left(\frac{\lambda}{c} - \frac{1}{\mu}\right)\psi(x) \qquad \Rightarrow \qquad \psi(x) = \frac{\lambda\mu}{c} \exp\left\{-\frac{c-\lambda\mu}{c\mu}x\right\}.$$

13.2 Some simple finite-state-space models

Example 116 (Sickness-death) In health insurance, the following model arises. Let $\mathbb{S} = \{H, S, \Delta\}$ consist of the states healthy, sick and dead. Clearly, Δ is absorbing. All other transitions are possible, at different rates. Under the assumption of full recovery after sickness, the state of health of the insured can be modelled by a continuous-time Markov chain.

Example 117 (Multiple decrement model) A life assurance often pays benefits not only upon death but also when a critical illness or certain losses of limbs, sensory losses or other disability are suffered. The assurance is not usually terminated upon such an event.

Example 118 (Marital status) Marital status has a non-negligible effect for various insurance types. The state space is $\mathbb{S} = \{B, M, D, W, \Delta\}$ to model bachelor, married, divorced, widowed, dead. Not all direct transitions are possible.

Example 119 (No claims discount) In car insurance and some other types of general insurances, you get a discount on your premium depending on the number of years without (or at most one) claim. This gives rise to a whole range of models, e.g. $\mathbb{S} = \{0\%, 20\%, 40\%, 50\%, 60\%\}$.

In all these examples, the exponential holding times are not particularly realistic. There are usually costs associated either with the transitions or with the states. Also, estimation of transition rates is of importance. A lot of data are available and sophisticated methods have been developed.

13.3 Summary results for M/M/1 queues

Consider a single-server queueing system in which customers arrive according to a Poisson process of rate λ and service times are independent $\mathrm{Exp}(\mu)$. Let X_t denote the length of the queue at time t including any customer that is currently served. This is the setting of Exercise A.4.3 and A.4.4 and from there we recall the following results:

- An invariant distribution exists if and only if $\lambda < \mu$, and is given by

$$\xi_n = (\lambda/\mu)^n(1-\lambda/\mu) = \rho^n(1-\rho), \qquad n \geq 0.$$

  where $\rho = \lambda/\mu$ is called the *traffic intensity*. Cleary $\lambda < \mu \iff \rho < 1$. By the ergodic theorem, the server is busy a (long-term) proportion ρ of the time.

- The embedded "jump chain" $(M_n)_{n\geq 0}$, $M_n = X_{T_n}$, has a different invariant distribution $\eta \neq \xi$. In fact, ξ puts more weight into the state 0 than η.

- During any $\mathrm{Exp}(\mu)$ service time, a $\mathrm{geom}(\lambda/(\lambda+\mu))$ number of customers arrives.

13.4 M/M/1 queues and the departure process

Define $D_0 = 0$ and successive departure times

$$D_{n+1} = \inf\{t > D_n : X_t - X_{t-} = -1\} \qquad n \geq 0.$$

Let us study the process $V_n = X_{D_n}$, $n \geq 0$, i.e. the process of queue lengths after departures. By the lack of memory property of $\mathrm{Exp}(\lambda)$, the geometric random variables N_n, $n \geq 1$, that record the number of new customers between D_{n-1} and D_n, are independent. Therefore, $(V_n)_{n\geq 0}$ is a Markov chain, with transition probabilities

$$d_{k,k-1+m} = \left(\frac{\lambda}{\lambda+\mu}\right)^m \frac{\mu}{\lambda+\mu}, \qquad k \geq 1, m \geq 0.$$

For $k = 0$, we get $d_{0,m} = d_{1,m}$, $m \geq 0$, since the next service only begins when a new customer enters the system.

Proposition 120 *The invariant distribution of the discrete-time Markov chain V is ξ.*

Proof: A simple calculation shows that with $\rho = \lambda/\mu$ and $q = \lambda/(\lambda+\mu)$

$$\sum_{k\in\mathbb{N}} \xi_k d_{k,n} = \xi_0 d_{0,n} + \sum_{k=1}^{n+1} \xi_k d_{k,n} = (1-\rho)q^n(1-q) + (1-\rho)(1-q)q^{n+1}\sum_{k=1}^{n+1}\left(\frac{\rho}{q}\right)^k = \xi_n,$$

after bringing the partial geometric progression into closed form and carrying out appropriate cancellations. □

Note that the conditional distribution of $D_{n+1} - D_n$ given $V_n = k$ is the distribution of a typical service time $G \sim \mathrm{Exp}(\mu)$ if $k \geq 1$, but the distribution of $Y + G$, where $Y \sim \mathrm{Exp}(\lambda)$ is a typical inter-arrival time, if $k = 0$, since we have to wait for a new customer and his service. We can also calculate the *unconditional* distribution of $D_{n+1} - D_n$, at least if V is in equilibrium.

Proposition 121 *If X (and hence V) is in equilibrium, then the $D_{n+1} - D_n$ are independent $\mathrm{Exp}(\lambda)$ distributed.*

Proof: Let us first study D_1. We can calculate its moment generating function by Proposition 135 a), conditioning on V_0, which has the stationary distribution ξ:

$$\begin{aligned}
\mathbb{E}(e^{\gamma D_1}) &= \mathbb{E}(e^{\gamma D_1}|V_0=0)\mathbb{P}(V_0=0)+\sum_{k=1}^{\infty}\mathbb{E}(e^{\gamma D_1}|V_0=k)\mathbb{P}(V_0=k)\\
&= \frac{\lambda}{\lambda-\gamma}\frac{\mu}{\mu-\gamma}\left(1-\frac{\lambda}{\mu}\right)+\frac{\mu}{\mu-\gamma}\frac{\lambda}{\mu}\\
&= \frac{\lambda}{\mu-\gamma}\frac{\mu-\lambda+\lambda-\gamma}{\lambda-\gamma}=\frac{\lambda}{\lambda-\gamma}
\end{aligned}$$

and identify the $\mathrm{Exp}(\lambda)$ distribution.

For independence of V_1 and D_1 we have to extend the above calculation and check that

$$\mathbb{E}(e^{\gamma D_1}\alpha^{V_1})=\frac{\lambda}{\lambda-\gamma}\frac{\mu-\lambda}{\mu-\alpha\lambda},$$

because the second ratio is the probability generating function of the $\mathrm{geom}(\lambda/\mu)$ stationary distribution ξ. To do this, condition on $V_0\sim\xi$ and then on D_1:

$$\mathbb{E}(e^{\gamma D_1}\alpha^{V_1})=\sum_{k=0}^{\infty}\xi_k\mathbb{E}(e^{\gamma D_1}\alpha^{V_1}|V_0=k)$$

and use the fact that given $V_1=k\geq 1$, $V_1=k+N_1-1$, where $N_1\sim\mathrm{Poi}(\lambda x)$ conditionally given $D_1=x$, because N_1 is counting Poisson arrivals in an interval of length $D_1=x$:

$$\begin{aligned}
\mathbb{E}(e^{\gamma D_1}\alpha^{V_1}|V_0=k) &= \alpha^{k-1}\int_0^{\infty}\mathbb{E}(e^{\gamma D_1}\alpha^{N_1}|V_0=k,D_1=x)f_{D_1}(x)dx\\
&= \alpha^{k-1}\int_0^{\infty}e^{\gamma x}\exp\{-\lambda x(1-\alpha)\}f_{D_1}(x)dx\\
&= \alpha^{k-1}\mathbb{E}(e^{(\gamma-\lambda(1-\alpha)D_1)})=\alpha^{k-1}\frac{\mu}{\mu-\gamma+\lambda(1-\alpha)}.
\end{aligned}$$

For $k=0$, we get the same expression without α^{k-1} and with a factor $\lambda/(\lambda-\gamma)$, because $D_1=Y+G$, where no arrivals occur during Y, and N_1 is counting those during $G\sim\mathrm{Exp}(\mu)$. Putting things together, we get

$$\mathbb{E}(e^{\gamma D_1}\alpha^{V_1})=(1-\rho)\left(\frac{\lambda}{\lambda-\gamma}+\frac{\rho}{1-\rho\alpha}\right)\frac{\mu}{\mu-\gamma+\lambda(1-\alpha)},$$

which simplifies to the expression claimed.

Now an induction shows $D_{n+1}-D_n\sim\mathrm{Exp}(\lambda)$, and they are independent, because the strong Markov property at D_n makes the system start afresh conditionally independently of the past given V_n. Since $D_1,\ldots,D_n-D_{n-1}$ are independent of V_n, they are then also independent of the whole post-D_n process. □

The argument is very subtle, because the post-D_n process is actually not independent of the whole pre-D_n process, just of the departure times. The result, however, is not surprising since we know that X is reversible, and the departure times of X are the arrival times of the time-reversed process, which form a Poisson process of rate λ.

In the same way, we can study $A_0=0$ and successive arrival times

$$A_{n+1}=\inf\{t>A_n:X_t-X_{t-}=1\},\qquad n\geq 0.$$

Clearly, these also have $\mathrm{Exp}(\lambda)$ increments, since the arrival process is a Poisson process with rate λ. We study X_{A_t} in the next lecture in a more general setting.

Lecture 14

M/G/1 and G/M/1 queues

Reading: Norris 5.2.7-5.2.8; Grimmett-Stirzaker 11.1; 11.3-11.4; Ross 8.5, 8.7
Further reading: Grimmett-Stirzaker 11.5-11.6

The M/M/1 queue is the simplest queueing model. We have seen modifications with more than 1 server. These were all continuous-time Markov chains. It was always the exponential distribution that described inter-arrival times as well as service times. In practice, this assumption is often unrealistic. If we keep exponential distributions for either inter-arrival times or service times, but allow more general distributions for the other, the model can still be handled using Markov techniques that we have developed.

We call an *M/G/1 queue* a queue length process with *M*arkovian arrivals (Poisson process of rate λ), a *G*eneral service time distribution (we also use G for a random variable with this general distribution on $(0,\infty)$), and *1* server.

We call a *G/M/1 queue* a queue length process with a *G*eneral inter-arrival distribution and *M*arkovian service times (exponential with rate parameter μ), and *1* server.

There are other queues that have names in this formalism. We have seen M/M/s queues (Example 32) and M/M/∞ queues (Example 62).

14.1 Stationarity in M/G/1 queues

An M/G/1 queue has independent and identically distributed service times with any distributions on $(0,\infty)$, but independent Exp(λ) inter-arrival times. Let X_t be the queue length at time t. X is *not* a continuous-time Markov chain, since the service distribution does not have the lack of memory property (unless it is exponential which brings us back to M/M/1). This means that after an arrival, we have a residual service distribution that is not just the distribution of a full service. However, after departures, we have exponential residual inter-arrival distributions:

Proposition 122 *The process of queue lengths $V_n = X_{D_n}$ at successive departure times D_n, $n \geq 0$, is a Markov chain with transition probabilities*

$$d_{k,k-1+m} = \mathbb{E}\left(\frac{(\lambda G)^m}{m!}e^{-\lambda G}\right), \qquad k \geq 1, m \geq 0,$$

and $d_{0,m} = d_{1,m}$, $m \geq 0$. Here G is a (generic) service time.

Proof: The proof is not hard since we recognise the ingredients. Given $G = t$ the number N of arrivals during the service times has a Poisson distribution with parameter λt. Therefore, if G has density g

$$\begin{aligned}\mathbb{P}(N = m) &= \int_0^\infty \mathbb{P}(N = m | G = t) g(t) dt \\ &= \int_0^\infty \frac{(\lambda t)^m}{m!} e^{-\lambda t} g(t) dt \\ &= \mathbb{E}\left(\frac{(\lambda G)^m}{m!}e^{-\lambda G}\right).\end{aligned}$$

If G is discrete, a similar argument works. The rest of the proof is the same as for M/M/1 queues (cf. the discussion before Proposition 120). In particular, when the departing customer leaves an empty system behind, there has to be an arrival before the next service time starts. □

For the M/M/1 queue, we defined the traffic intensity $\rho = \lambda/\mu$, in terms of the arrival rate $\lambda = 1/\mathbb{E}(Y)$ and the (potential) service rate $\mu = 1/\mathbb{E}(G)$ for a generic inter-arrival time $Y \sim \text{Exp}(\lambda)$ and service time $G \sim \text{Exp}(\mu)$. We say "potential" service rate, because in the queueing system, the server may have idle periods (empty system), during which there is no service. Indeed, a main reason to consider traffic intensities is to describe whether there are idle periods, i.e. whether the queue length is a recurrent process.

If G is not exponential, we can interpret "service rate" as *asymptotic* rate. Consider a renewal process N with inter-renewal times distributed as G. By the strong law of renewal theory $N_t/t \to 1/\mathbb{E}(G)$. It is therefore natural, for the M/G/1 queue, to define the traffic intensity as $\rho = \lambda\mathbb{E}(G)$.

Proposition 123 *Let $\rho = \lambda\mathbb{E}(G)$ be the traffic intensity of an M/G/1 queue. If $\rho < 1$, then V has a unique invariant distribution ξ. This ξ has probability generating function*

$$\sum_{k=0}^{\infty} \xi_k s^k = (1-\rho)(1-s)\frac{1}{1 - s/\mathbb{E}(e^{\lambda(s-1)G})}.$$

Proof: By the uniqueness theorem for probability generating functions,

$$\xi_j = \sum_{i=0}^{j+1} \xi_i d_{i,j} \iff \phi(s) := \sum_{j=0}^{\infty} \xi_j s^j = \sum_{j=0}^{\infty}\sum_{i=0}^{j+1} \xi_i d_{i,j} s^j.$$

To calculate the right-hand side, first note that

$$\sum_{m=0}^{\infty} d_{k+1,k+m} s^m = \sum_{m=0}^{\infty} \mathbb{E}\left(\frac{(s\lambda G)^m}{m!} e^{-\lambda G}\right) = \mathbb{E}(e^{(s-1)\lambda G}).$$

Then

$$\begin{aligned}
\sum_{j=0}^{\infty}\sum_{i=0}^{j+1} \xi_i d_{i,j} s^j &= \sum_{j=0}^{\infty} \xi_0 d_{0,j} s^j + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \xi_{k+1} d_{k+1,k+m} s^{k+m} \\
&= \mathbb{E}(e^{(s-1)\lambda G})\left(\xi_0 + \sum_{k=0}^{\infty} \xi_{k+1} s^k\right) \\
&= \mathbb{E}(e^{(s-1)\lambda G}) s^{-1}\left(\phi(s) - \xi_0(1-s)\right),
\end{aligned}$$

and this equals $\phi(s)$ if and only if

$$\phi(s) = \frac{\xi_0(1-s)}{1 - s/\mathbb{E}(e^{(s-1)\lambda G})}.$$

For ϕ to be a probability generating function, we need $\phi(1) = 1$, and de l'Hôpital's rule yields $\xi_0 = 1 - \rho$. □

14.2 Waiting times in M/G/1 queues

An important quantity in queueing theory is the waiting time of a customer. Here we have to be specific about the service discipline. We will assume throughout that customers queue and are served in their order of arrival. This discipline is called FIFO (First In First Out). Other disciplines like LIFO (Last In First Out) with or without interruption of current service can also be studied.

Clearly, under the FIFO discipline, the waiting time of a given customer depends on the service times of customers in the queue when he arrives. Also, all customers X_{D_n} in the system when a customer leaves at time D_n, have arrived during his waiting time W_n and service times G_n.

Proposition 124 *If the system has been running for a long time before time* 0*, i.e.* $X_0 \sim \xi$*, then the waiting time* W_n *of any customer* $n \geq 1$ *has distribution given by*

$$\mathbb{E}(e^{\gamma W_n}) = \frac{(1-\rho)\gamma}{\lambda + \gamma - \lambda \mathbb{E}(e^{\gamma G})}.$$

Proof: We have not established equilibrium of X at the arrival times of customers, but at departure times, so we argue from the time when a customer leaves. Due to the FIFO discipline, he will leave behind all those customers that arrived during his waiting time W and his service time G. Given $T := W + G = t$, their number N has a Poisson distribution with parameter λt so that

$$\begin{aligned}\mathbb{E}(s^N) &= \int_0^\infty \mathbb{E}(s^N | T = t) f_T(t) dt = \int_0^\infty e^{\lambda t(s-1)} f_T(t) dt \\ &= \mathbb{E}(e^{\lambda T(s-1)}) = \mathbb{E}(e^{\lambda(s-1)W})\mathbb{E}(e^{\lambda(s-1)G}).\end{aligned}$$

By stationarity, $N \sim \xi$, and we take $\mathbb{E}(s^N)$ from Proposition 123. We deduce the required formula by setting $\gamma = \lambda(s-1)$ and by solving for $\mathbb{E}(e^{\gamma W})$. □

Corollary 125 *In the special case of M/M/1, the distribution of* W *is given by*

$$\mathbb{P}(W = 0) = 1 - \rho \qquad \text{and} \qquad \mathbb{P}(W > w) = \rho e^{-(\mu-\lambda)w}, \quad w \geq 0.$$

Proof: We calculate the moment generating function of the proposed distribution

$$e^{\gamma 0}(1-\rho) + \int_0^\infty e^{\gamma t}\rho(\mu - \lambda)e^{-(\mu-\lambda)t} dt = \frac{\mu - \lambda}{\mu} + \frac{\lambda}{\mu}\frac{\mu-\lambda}{\mu - \lambda - \gamma} = \frac{\mu-\lambda}{\mu}\frac{\mu - \gamma}{\mu - \lambda - \gamma}.$$

From the preceding proposition we get for our special case

$$\mathbb{E}(e^{\gamma W}) = \frac{\gamma(\mu - \lambda)/\mu}{\lambda + \gamma - \lambda\mu/(\mu - \gamma)} = \frac{\mu - \lambda}{\mu}\frac{(\mu - \gamma)\gamma}{(\lambda + \gamma)(\mu - \gamma) - \lambda\mu}$$

and we see that the two are equal. We conclude by the Uniqueness Theorem for moment generating functions. □

14.3 G/M/1 queues

For G/M/1 queues, the arrival process is a renewal process. Clearly, by the renewal property and by the lack of memory property of the service times, the queue length process X starts afresh after each arrival, i.e. $\widetilde{U}_n = X_{A_n}$, $n \geq 0$, is a Markov chain on $\{1, 2, 3, \ldots\}$, where A_n is the nth arrival time. It is actually more natural to consider the Markov chain $U_n = \widetilde{U}_n - 1 = X_{A_n -}$ on $\mathbb{N}$.

It can be shown that for M/M/1 queues the invariant distribution of U is the same as the invariant distribution of V and of X. For general G/M/1 queues we get

Proposition 126 *Let* $\rho = 1/(\mu\mathbb{E}(A_1))$ *be the traffic intensity. If* $\rho < 1$*, then* U *has a unique invariant distribution given by*

$$\xi_k = (1 - q)q^k, \qquad k \in \mathbb{N},$$

where q *is the smallest positive root of* $q = \mathbb{E}(e^{\mu(q-1)A_1})$.

Proof: First note that given an inter-arrival time $Y = y$, a Poi(μy) number of customers are served, so U has transition probabilities

$$a_{i,i+1-j} = \mathbb{E}\left(\frac{(\mu Y)^j}{j!}e^{-\mu Y}\right), \quad j = 0,\ldots,i; \quad a_{i,0} = 1 - \sum_{j=0}^{i} a_{i,i+1-j}.$$

Now for any geometric ξ, we get, for $k \geq 1$, from Tonelli's theorem,

$$\begin{aligned}\sum_{i=k-1}^{\infty} \xi_i a_{ik} &= \sum_{j=0}^{\infty} \xi_{j+k-1} a_{j+k-1,j} \\ &= \sum_{j=0}^{\infty} (1-q)q^{j+k-1}\mathbb{E}\left(\frac{(\mu Y)^j}{j!}e^{-\mu Y}\right) \\ &= (1-q)q^{k-1}\mathbb{E}\left(e^{-\mu Y(1-q)}\right),\end{aligned}$$

and clearly this equals $\xi_k = (1-q)q^k$ if and only if $q = \mathbb{E}(e^{\mu(q-1)Y}) =: f(q)$, as required. Note that both sides are continuously differentiable on $[0,1)$ and on $[0,1]$ if and only if limits $q \uparrow 1$ are finite, $f(0) > 0$, $f(1) = 1$ and $f'(1) = \mathbb{E}(\mu Y) = 1/\rho$, so there is a solution if $\rho < 1$, since then $f(1-\varepsilon) < 1-\varepsilon$ for ε small enough. The solution is unique, since there is at most one stationary distribution for the irreducible Markov chain U. The case $k = 0$ can be checked by a similar computation, so ξ is indeed a stationary distribution. □

Proposition 127 *The waiting time W of a customer arriving in equilibrium has distribution*

$$\mathbb{P}(W = 0) = 1 - q, \qquad \mathbb{P}(W > w) = qe^{-\mu(1-q)w}, \quad w \geq 0.$$

Proof: In equilibrium, an arriving customer finds a number $N \sim \xi$ of customers in the queue in front of him, each with a service of $G_j \sim \text{Exp}(\mu)$. Clearly $\mathbb{P}(W = 0) = \xi_0 = 1 - q$. Also since the conditional distribution of N given $N \geq 1$ is geometric with parameter q and geometric sums of exponential random variables are exponential, we have that W given $N \geq 1$ is exponential with parameter $\mu(1-q)$. □

Alternatively, we can write this proof in formulas as a calculation of $\mathbb{P}(W > y)$ by conditioning on N and by using Tonelli's theorem to interchange summation and integration:

$$\begin{aligned}\mathbb{P}(W > w) &= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{P}(W > w|N = n) \\ &= 0 + \sum_{n=1}^{\infty} q^n(1-q)\int_w^{\infty} \frac{\mu^n}{(n-1)!}x^{n-1}e^{-\mu x}dx \\ &= \int_w^{\infty} e^{-\mu x}q\mu(1-q)\sum_{n=1}^{\infty}\frac{\mu^{n-1}}{(n-1)!}q^{n-1}x^{n-1}dx \\ &= q\int_w^{\infty} \mu(1-q)\exp\{-\mu x + \mu q x\}dx = q\exp\{-\mu(1-q)y\},\end{aligned}$$

where we used that the sum of n independent identically exponentially distributed random variables is Gamma distributed.

Lecture 15

Queueing networks and some further applications and generalisations

Reading: Norris 5.2.1-5.2.6; Grimmett-Stirzaker 11.2, 11.7; Ross 6.6, 8.4

In this lecture we study tandem queues and queueing networks to demonstrate how methods developed for general countable state space $\mathbb{S}$ apply to state spaces such as $\mathbb{N}^2$ and $\mathbb{N}^m$, $m \geq 3$. The lecture notes conclude by giving a summary of the course and pointers for further study.

15.1 Tandem queues

The simplest non-trivial network of queues is a so-called tandem system that consists of two queues with one server each, having independent Exp(μ_1) and Exp(μ_2) service times, respectively. Customers join the first queue according to a Poisson process of rate λ, and on completing service immediately enter the second queue. We denote by $X_t^{(1)}$ the length of the first queue at time $t \geq 0$ and by $X_t^{(2)}$ the length of the second queue at time $t \geq 0$.

Proposition 128 *The queue length process $X = (X^{(1)}, X^{(2)})$ is a continuous-time Markov chain with state space $\mathbb{S} = \mathbb{N}^2$ and non-zero transition rates*

$$q_{(i,j),(i+1,j)} = \lambda, \quad q_{(i+1,j),(i,j+1)} = \mu_1, \quad q_{(i,j+1),(i,j)} = \mu_2, \qquad i,j \in \mathbb{N}.$$

Proof: Just note that in state $(i+1, j+1)$, three exponential clocks are ticking, that lead to transitions at rates as described. Similarly, there are fewer clocks for $(0, j+1)$, $(i+1, 0)$ and $(0, 0)$ since one or both servers are idle. The lack of memory property makes the process start afresh after each transition. A standard inductive argument completes the proof. □

Proposition 121 yields that the departure process of the first queue, which is now also the arrival process of the second queue, is a Poisson process with rate λ, provided that the first queue is in equilibrium. This can be achieved if $\lambda < \mu_1$.

Proposition 129 *X is positive recurrent if and only if $\rho_1 := \lambda/\mu_1 < 1$ and $\rho_2 := \lambda/\mu_2 < 1$. The unique stationary distribution is then given by*

$$\xi_{(i,j)} = \rho_1^i(1-\rho_1)\rho_2^j(1-\rho_2)$$

i.e. in equilibrium, the lengths of the two queues at any fixed time are independent.

Proof: As shown in Example 54, $\rho_1 \geq 1$ violates positive recurrence, and expected return times for X and $X^{(1)}$ satisfy $m_{(0,0)} \geq m_0^{(1)} = \infty$. If $\rho_1 < 1$ and $X^{(1)}$ is in equilibrium, then by Proposition 121, the arrival process for the second queue is a Poisson process at rate λ, and $\rho_2 \geq 1$ would violate positive recurrence of $X^{(2)}$. Specifically, if we assume $m_{0,0} < \infty$, then we get the contradiction $\infty = m_0^{(2)} \leq m_{(0,0)} < \infty$.

If $\rho_1 < 1$ and $\rho_2 < 1$, ξ as given in the statement of the proposition is an invariant distribution, it is easily checked that the $(i+1, j+1)$ entry of $\xi Q = 0$ holds:

$$\xi_{(i,j+1)}q_{(i,j+1),(i+1,j+1)} + \xi_{(i+2,j)}q_{(i+2,j),(i+1,j+1)} + \xi_{(i+1,j+2)}q_{(i+1,j+2),(i+1,j+1)} \\ + \xi_{(i+1,j+1)}q_{(i+1,j+1),(i+1,j+1)} = 0$$

for $i,j \in \mathbb{N}$, and similar equations for states $(0,j+1)$, $(i+1,0)$ and $(0,0)$. It is unique since X is clearly irreducible (we can find paths between any two states in $\mathbb{N}^2$). □

We stressed that queue lengths are independent *at fixed times*. In fact, they are not independent in a stronger sense, e.g. $(X_s^{(1)}, X_t^{(1)})$ and $(X_s^{(2)}, X_t^{(2)})$ for $s < t$ turn out to be dependent. More specifically, consider $X_s^{(1)} - X_t^{(2)} = n$ for big n, then it is easy to see that $0 < \mathbb{P}(X_t^{(2)} = 0 | X_s^{(1)} - X_t^{(1)} = n) \to 0$ as $n \to \infty$, since it is increasingly unlikely that n customers are served by server 2 during $[s,t]$.

15.2 Closed and open migration networks

More general queueing systems are obtained by allowing customers to move in a system of m single-server queues according to a Markov chain on $\{1,\ldots,m\}$. For a single customer, no queues ever occur, since he is simply served where he goes. If there are r customers in the system with no new customers arriving and no existing customers departing, the system is called a *closed migration network*. If at some (or all) queues, also new customers arrive according to a Poisson process, or at some (or all) queues, customers served may leave the system, the system is called an *open migration network*.

The tandem queue is an open migration network with $m = 2$, where new customers only arrive at the first queue and existing customers only leave the system after service from the second server. The Markov chain is deterministic and sends each customer from state 1 to state 2: $\pi_{12} = 1$. Customers then go into an absorbing exit state 0, say, $\pi_{2,0} = 1$, $\pi_{0,0} = 1$. The following are two results that can be shown using similar methods to those used for Propositions 128 and 129.

Proposition 130 *If service times are independent* $\mathrm{Exp}(\mu_k)$ *at server* $k \in \{1,\ldots,m\}$*, arrivals occur according to independent Poisson processes of rates* λ_k, $k = 1,\ldots,m$*, and departures are modelled by transitions to another server or an exit state* 0*, according to transition probabilities* $\pi_{k,\ell}$*, then the queue-lengths process* $X = (X^{(1)},\ldots,X^{(m)})$ *is a continuous-time Markov chain. Its non-zero transition rates are*

$$q_{x,x+e_k} = \lambda_k, \qquad q_{x,x-e_k+e_\ell} = \mu_k \pi_{k\ell}, \qquad q_{x,x-e_k} = \mu_k \pi_{k0}$$

for all $k,\ell \in \{1,\ldots,m\}$, $x = (x_1,\ldots,x_m) \in \mathbb{N}^m$ *such that* $x_k \geq 1$ *for the latter two, and where* $e_k = (0,\ldots,0,1,0,\ldots,0)$ *is the* k*th unit vector.*

Proposition 131 *Suppose* $X = (X^{(1)},\ldots,X^{(m)})$ *models a closed migration network with irreducible migration chain. Then the following hold.*

(i) *The total number of customers* $X_t^{(1)} + \cdots + X_t^{(m)}$ *remains constant over time* $t \geq 0$.

(ii) *If* $X_0^{(1)} + \cdots + X_0^{(m)} = 1$*, then the* continuous-time migration chain $K_t = k$ *if* $X_t^{(k)} = 1$ *has as holding rate in* k *the service rate* μ_k *of the* k*th server,* $k \in \{1,\ldots,m\}$*, and* K *possesses a unique stationary distribution* η *on* $\{1,\ldots,m\}$.

(iii) *For any* $r = X_0^{(1)} + \cdots + X_0^{(2)} \geq 1$*, the process* X *has the unique invariant distribution*

$$\xi_x = B_r \prod_{k=1}^m \eta_k^{x_k}, \qquad \text{for all } x \in \mathbb{N}^m \text{ such that } x_1 + \cdots + x_m = r,$$

where η *is the invariant distribution of the continuous-time migration chain of* (ii)*, and* B_r *is a normalising constant.*

ξ has product form on each communicating class $C_r = \{x \in \mathbb{N}^m : x_1 + \cdots + x_m = r\}$, but the queue lengths at servers $k = 1,\ldots,m$ under the stationary distribution are *not* independent, not even at fixed times in stationarity, since the admissible x-values are constrained by $x_1 + \cdots + x_m = r$.

15.3 Summary of the course

This course is about stochastic process models $X = (X_t)_{t\geq 0}$ in continuous time and (mostly) a discrete state space $\mathbb{S}$, often $\mathbb{N}$. Applications include those where X describes

- counts of births, atoms, bacteria, visits, trials, arrivals, departures, insurance claims, etc.,
- the size of a population, the number of buses in service, the length of a queue,

and others. Important is the structure, the transition mechanism in the real world that we wish to model by X. Memory plays an important role. We use the following mathematical concepts

- Independence (of individuals in population models, of counts over disjoint time intervals, of service times, of different ingredients to a model etc.),
- Markov property (lack of memory, exponential holding times; past irrelevant for the future except for the current state),
- Renewal property (information on previous states irrelevant, but duration in state relevant; "Markov property" at transition times),
- Stationarity, equilibrium (behaviour homogeneous in time; for Markov chains, invariant marginal distribution; for renewal processes, stationary increments).

Once we have a model X for the real world process, which satisfies such properties, we study it under the model assumptions. We study

- different descriptions of X (jump chain - holding times, transition probabilities - forward-backward equations, infinitesimal behaviour),
- convergence to equilibrium (invariant distributions, convergence of transition probabilities, ergodic theorem; strong law and CLT of renewal theory, renewal theorems),
- hitting times, excess life, return times, waiting times, ruin probabilities,
- odd/undesirable(?) behaviour (explosion, transience, arithmetic inter-renewal times).

Techniques:

- conditioning, often on the first jump (one-step analysis),
- functions of independent random variables are independent,
- detailed balance equations,
- algebra of limits for almost sure convergence,
- renewal equations and Key renewal theorem.

The following sections contain some natural generalisations of concepts studied in this course. They are, of course, non-examinable.

15.4 Duration-dependent transition rates

Renewal processes can be thought of as duration-dependent transition rates. If the inter-renewal distribution is not exponential, then (at least some) residual distributions will not be the same as the full inter-renewal distribution, but we can still express, say for Z with density f that

$$\mathbb{P}(Z-t>s|Z>t)=\frac{\mathbb{P}(Z>t+s)}{\mathbb{P}(Z>t)}\qquad\text{and}\qquad f_{Z-t|Z>t}(s)=\frac{f(t+s)}{\mathbb{P}(Z>t)}.$$

If we define $\lambda(t)=\dfrac{f(t)}{\mathbb{P}(Z>t)}=\dfrac{-\overline{F}'(t)}{\overline{F}(t)}$, where $\overline{F}(t)=\mathbb{P}(Z>t)$ and in particular $\overline{F}(0)=1$, we can write

$$\overline{F}(t)=\exp\left\{-\int_0^t\lambda(s)ds\right\}\qquad\text{and}\qquad f(t)=\lambda(t)\exp\left\{-\int_0^t\lambda(s)ds\right\}.$$

We can then also express the residual distributions in terms of $\lambda(s)$

$$\mathbb{P}(Z-t>s|Z>t)=\exp\left\{-\int_t^{t+s}\lambda(r)dr\right\}.$$

$\lambda(t)$ can be interpreted as the instantaneous arrival rate time t after the previous arrival. Similarly, we can use this idea in Markov models and split a holding rate $\lambda_i(d)$ depending on the duration d of the current visit to state i into transition rates $\lambda_i(d)=\sum_{j\neq i}q_{ij}(d)$. This will be explored in SB3b Statistical Lifetime-Models, where Z is usually modelling a lifetime. The context is the estimation and statistical comparison of life-time distributions and related models.

15.5 Spatial Poisson processes

In the case of Poisson counts, one can also look at intensity functions on $\mathbb{R}^2$ or $\mathbb{R}^d$ and look at "arrivals" as random points in the plane.

$$N([0,t]\times[0,z])=X(t,z)\sim\text{Poi}\left(\int_0^t\int_0^z\lambda(s,y)dyds\right)$$

and such that counts in disjoint rectangles are independent Poisson variables.

15.6 Markov processes in uncountable state spaces ($\mathbb{R}$ or $\mathbb{R}^d$)

We have come across some processes for which we could have proved a Markov property, the age process $(A_t)_{t\geq 0}$ of a renewal process, the excess process $(E_t)_{t\geq 0}$ of a renewal process, but also the processes $(C_t)_{t\geq 0}$ and $(R_t)_{t\geq 0}$ with stationary independent increments that arose in insurance ruin by combining Poisson arrival times with jump sizes. A systematic study of such Markov processes in $\mathbb{R}$ is technically much harder, although many ideas and results transfer from our countable state space model.

Diffusion processes as a special class of such Markov processes are studied in a Finance context in B8.3 Mathematics of Financial Derivatives and they appear from a more theoretical perspective in C8.1 Stochastic Differential Equations. Some examples also feature in MS4 Stochastic Models for Mathematical Genetics.

One nice class of processes are processes with stationary independent increments, so-called Lévy processes. General Lévy processes can be built from three ingredients. The first is a deterministic linear drift. The second is a Brownian motion B, which features strongly in B8.2 Continuous Martingales, as well as B8.3 Mathematics of Financial Derivatives. The third is a compound Poisson process C such as the claim size process C in ruin theory. Then any process $X_t=\mu t+\sigma B_t+C_t$ is a Lévy process and any Lévy process is of this form, if C is allowed to contain a component that is a (martingale limit of) a compound Poisson processes. In fact, C may have infinitely many jumps in a finite interval, that can be described by a spatial Poisson process on $[0,\infty)\times\mathbb{R}^*$.

15.7 Stationary processes

We have come across stationary Markov chains and stationary increments of other processes. Stationarity is a concept that can be studied separately. In our examples, the dependence structure of processes was simple: independent increments, or Markovian dependence, independent holding times etc. More complicated dependence structures may be studied.

Appendix A

Conditioning and stochastic modelling

Reading: Grimmett-Stirzaker 3.7, 4.6
Further reading: Grimmett-Stirzaker 4.7; CT4 Unit 1

This lecture consolidates the ideas of conditioning and modelling. Along the way, we explain the full meaning of statements such as the Markov properties of Lecture 1.

A.1 Modelling of events

As in the Prelims and Part A courses, random variables are defined as functions on a *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the *sample space*, $\mathcal{F}$ is a collection of subsets of Ω called *events*, the *probability measure* $\mathbb{P}$ assigns a probability in $[0, 1]$ to each event. A probability space satisfies

- $\Omega \in \mathcal{F}$; • $A \in \mathcal{F} \Rightarrow A^c = \Omega \setminus A \in \mathcal{F}$; • $A_n \in \mathcal{F}, n \geq 1, \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.
- $\mathbb{P}(\Omega) = 1$; • $A_n \in \mathcal{F}, n \geq 1$, disjoint $\Rightarrow \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Random variables $X: \Omega \to \mathbb{X}$, where $\mathbb{X}$ is typically either a subset of $\mathbb{R}$ or any countable set $\mathbb{S}$, but can also be a space of functions such as $\mathbb{X} = \{f: [0, \infty) \to \mathbb{S} \text{ right-continuous}\}$, are such that

- $\{X \in B\} := X^{-1}(B) = \{\omega \in \Omega: X(\omega) \in B\} \in \mathcal{F}$ for all (measurable) $B \subseteq \mathbb{X}$.

This course is not based on measure theory, and in fact we will only occasionally have to refer back to these properties for clarity of argument.

Modelling means specifying a mathematical model for a real-world phenomenon. *Stochastic modelling* include some randomness, i.e. some real-world events are assigned probabilities or some real-world observables are assigned distributions. At first, real-world events can be named e.g. $A_1 =$"the die shows an even number" and $A_2 =$"the first customer arrives before 10am". A stochastic model identifies such an event as a subset of a sample space Ω and assigns probabilities. We seem to be able to write down some probabilities directly without much sophistication ($\mathbb{P}(A_1) = 0.5$? still making implicit assumptions about the fairness of the die and the conduct of the experiment). Others require less obvious specification of a stochastic model ($\mathbb{P}(A_2) =$?).

Hardly any real situations involve genuine randomness. It is rather our incomplete perception/information that makes us think there was randomness. Nevertheless, assuming a specific random model to inform our decision-making can be very helpful and lead to decisions that are sensible/good/beneficial in some sense.

Mathematical models always make assumptions and reflect reality only partially. The following situation is quite common: the better a model represents reality, the more complicated it is to analyse. There is a trade-off here. In any case, we must base all our calculations on the model specification, the model assumptions. Translating reality into models is at least partly a non-mathematical task. Analysing a model is purely mathematical.

Models have to be consistent, i.e. they must not contain contradictions. This statement may seem obvious, but the point is that not all contradictions are immediately apparent. There are models that have undesirable features that cannot be easily removed, least by postulating the contrary. E.g., you may wish to specify a model for customer arrival where arrival counts over disjoint time intervals are independent, arrival counts over time intervals of equal lengths have the same distribution (cf. Remark 5 (ii)-(iii)), and times between two arrivals have a non-exponential distribution. Well, such a model does not exist (we won't prove this statement now, it's a bit too hard at this stage). On the other hand, within a consistent model, all properties that were not specified in the model assumptions have to be derived from these. Otherwise it must be assumed that the model may not have the property.

Suppose we are told that a shop opens at 9.30am, and on average, there are 10 customers per hour. One model could be to assume that a customer arrives exactly every six minutes. Another model could be to assume customers arrive according to a Poisson process at rate $\lambda = 10$ (time unit=1 hour). Whichever model we use, we can "calculate" $\mathbb{P}(A_2)$, and it is not the same in the two models, so we should reflect this in our notation. Since A_2 does not really change from one model to the other, it had better be $\mathbb{P}$ that changes, and we may wish to write $\widetilde{\mathbb{P}}$ for the second model. The probability measure $\mathbb{P}$ should be thought of as defining the randomness. Similarly, we can express dependence on a parameter by $\mathbb{P}^{(\lambda)}$, dependence on an initial value by $\mathbb{P}_k$. Informally, for a Poisson process model, we set $\mathbb{P}_k(A) := \mathbb{P}(A|X_0 = k)$ for all events A (formally, this should make us wonder whether $\mathbb{P}(X_0 = k) > 0$, and in fact, we first define $\mathbb{P}_k$ and could then write $\mathbb{P}(A|X_0 = k) := \mathbb{P}_k(A)$ as a long-hand notation).

Aside: Technically, we cannot in general call all subsets of Ω events if Ω is uncountable, but we will not worry about this, since it is hard to find examples of *non-measurable sets*. ω should be thought of as a scenario, a realisation of all the randomness, which we typically express in terms of random variables $X(\omega)$. What matters are (joint!) distributions of random variables, not usually the precise form of $(\Omega, \mathcal{F}, \mathbb{P})$. It is important, though, that $(\Omega, \mathcal{F}, \mathbb{P})$ exists for all our purposes to make sure that the random objects we study exist. We will assume that all our random variables can be defined as (measurable) functions on some $(\Omega, \mathcal{F}, \mathbb{P})$. This existence can be proved for all our purposes, using measure theory.

In fact, when we express complicated families of random variables such as a Poisson process $(X_t)_{t\geq 0}$ in terms of a countable family $(Z_n)_{n\geq 0}$ of independent random variables, we do this for two reasons. The intuitive reason may be apparent: countable families of independent variables are conceptually easier than uncountable families of dependent variables. The formal reason is that a result in measure theory says that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which we can define countable families of independent variables whereas any more general result for uncountable families or dependent variables requires additional assumptions or other caveats.

It is very useful to think about random variables Z_n as functions $Z_n(\omega)$, because it immediately makes sense to define a Poisson process $X_t(\omega)$ as in Definition 1, by defining new functions in terms of old functions. A certain class of probability problems can be solved by applying analytic *rules* to calculations involving functions of random variables (transformation formula for densities, expectation of a function of a random variable in terms of its density or probability function, etc.). Here we are dealing more explicitly with random variables and events themselves, operating on them directly.

This course is not based on measure theory, but you should be aware that some of the proofs are only mathematically complete if based on measure theory. Ideally, this only means that we apply an easily stated result from measure theory that is intuitive enough to believe without proof. In a few cases, however, the gap is more serious. We will identify technicalities, but without drawing attention away from the probabilistic arguments that we develop in this course and that are useful for applications.

B8.1 Martingales Through Measure Theory provides as pleasant an introduction to measure theory as can be given. That course nicely complements this course in providing the formal basis for probability theory in general and hence for this course in particular. However, it is by no means a co-requisite, and when we do refer to this course, it is likely to be to material that has not yet been covered there. Williams' Probability with Martingales is the recommended book reference.

A.2 Conditional probabilities, densities and expectations

Conditional probabilities were introduced in Prelims as

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

for events $A, B \subseteq \Omega$, where we require $\mathbb{P}(A) > 0$.

Example 132 Let X be a Poisson process. Then

$$\mathbb{P}(X_t = k + j | X_s = k) = \frac{\mathbb{P}(X_t - X_s = j, X_s = k)}{\mathbb{P}(X_s = k)} = \mathbb{P}(X_t - X_s = j) = \mathbb{P}(X_{t-s} = j),$$

by the independence and stationarity of increments, Remark 5 (ii)-(iii).

Conditional densities were introduced in Part A as

$$f_{S|T}(s|t) = f_{S|T=t}(s) = \frac{f_{S,T}(s,t)}{f_T(t)}.$$

Example 133 Let X be a Poisson process. Then, for $t > s$,

$$f_{T_2|T_1=s}(t) = \frac{f_{T_1,T_2}(s,t)}{f_{T_1}(s)} = \frac{f_{Z_0,Z_1}(s,t-s)}{f_{Z_0}(s)} = \frac{f_{Z_0}(s) f_{Z_1}(t-s)}{f_{Z_0}(s)} = f_{Z_1}(t-s) = \lambda e^{-\lambda(t-s)},$$

by the transformation formula for bivariate densities to relate f_{T_1,T_2} to f_{Z_0,Z_1}, and independence of Z_0 and Z_1.

Conditioning has to do with available information. In the real world, we often observe a process with time. When we have set up a stochastic model, e.g. a Poisson process with a known parameter $\lambda > 0$. (If we don't know λ, we should estimate λ and update estimates as we observe the real-world process, but we do not worry about this in this course.) It is instructive to think of updating the stochastic process by its realisation in the real world as time evolves. If the first arrival takes a long time to happen, this gives us information about the second arrival time T_2, simply since $T_2 = T_1 + Z_1 > T_1$. When we eventually observe $T_1 = s$, the conditional density of T_2 given $T_1 = s$ takes into account this observation and captures the remaining stochastic properties of T_2. The result of the formal calculation to derive the conditional density is in agreement with the intuition that if $T_1 = s$, $T_2 = T_1 + Z_1$ ought to have the distribution of Z_1 shifted by s.

Example 134 Conditional probabilities and conditional densities are compatible in that

$$\mathbb{P}(S \in B | T = t) = \int_B f_{S|T=t}(s) ds = \lim_{\varepsilon \downarrow 0} \mathbb{P}(S \in B | t \le T \le t + \varepsilon),$$

provided only that the distribution of (S, T) is sufficiently smooth. To see this, when $(s, t) \mapsto f_{S,T}(s, t)$ is sufficiently smooth (e.g. right-continuous in the t-variable and bounded), we check that for all intervals $B = (a, b)$

$$\mathbb{P}(S \in B | t \le T \le t + \varepsilon) = \frac{\mathbb{P}(S \in B, t \le T \le t + \varepsilon)}{\mathbb{P}(t \le T \le t + \varepsilon)} = \frac{\frac{1}{\varepsilon} \int_t^{t+\varepsilon} \int_B f_{S,T}(s, u) ds du}{\frac{1}{\varepsilon} \mathbb{P}(t \le T \le t + \varepsilon)}$$

and under the smoothness condition (by dominated convergence, Tonelli's theorem etc., see Lecture 3 for statements of these), this tends to

$$\frac{\int_B f_{S,T}(s,t) ds}{f_T(t)} = \int_B f_{S|T=t}(s) ds = \mathbb{P}(S \in B | T = t).$$

Similarly, we can also define the following for discrete X and continuous T, when the limits exist:

$$\mathbb{P}(X=k|T=t)=\lim_{\varepsilon\downarrow 0}\mathbb{P}(X=k|t\le T\le t+\varepsilon) \quad \text{and} \quad f_{T|X=k}(t)=\lim_{\varepsilon\downarrow 0}\frac{1}{\varepsilon}\mathbb{P}(t\le T\le t+\varepsilon \mid X=k).$$

One can define conditional expectations in analogy with unconditional expections, e.g. in the latter case by

$$\mathbb{E}(X|T=t)=\sum_{j=0}^{\infty} j\mathbb{P}(X=j|T=t).$$

Proposition 135 (a) *If X and Y are (dependent) discrete random variables in $\mathbb{N}$, then*

$$\mathbb{E}(X)=\sum_{n=0}^{\infty}\mathbb{E}(X|Y=n)\mathbb{P}(Y=n).$$

(b) *If X and T are jointly continuous random variables in $(0,\infty)$ or*

(c) *if X is discrete and T is continuous, and if T has a right-continuous density, then*

$$\mathbb{E}(X)=\int_0^{\infty}\mathbb{E}(X|T=t)f_T(t)dt.$$

Proof: (c) We start at the right-hand side

$$\int_0^{\infty}\mathbb{E}(X|T=t)f_T(t)dt=\int_0^{\infty}\sum_{j=0}^{\infty} j\mathbb{P}(X=j|T=t)f_T(t)dt$$

and calculate

$$\begin{aligned}\mathbb{P}(X=j|T=t) &= \lim_{\varepsilon\downarrow 0}\frac{\mathbb{P}(X=j,t\le T\le t+\varepsilon)}{\mathbb{P}(t\le T\le t+\varepsilon)}\\ &= \lim_{\varepsilon\downarrow 0}\frac{\frac{1}{\varepsilon}\mathbb{P}(t\le T\le t+\varepsilon|X=j)\mathbb{P}(X=j)}{\frac{1}{\varepsilon}\mathbb{P}(t\le T\le t+\varepsilon)}\\ &= \frac{f_{T|X=j}(t)\mathbb{P}(X=j)}{f_T(t)}\end{aligned}$$

so that we get on the right-hand side

$$\int_0^{\infty}\sum_{j=0}^{\infty} j\mathbb{P}(X=j|T=t)f_T(t)dt=\sum_{j=0}^{\infty} j\mathbb{P}(X=j)\int_0^{\infty} f_{T|X=j}(t)dt=\mathbb{E}(X)$$

after interchanging summation and integration. This is justified by Tonnelli's theorem that we state in Lecture 3.

(b) is similar to (c).

(a) is more elementary and left to the reader. □

Statement and argument hold for left-continuous densities and approximations from the left, as well. For continuous densities, one can also approximate $\{T=t\}$ by $\{t-\varepsilon\le T\le t+\varepsilon\}$ (for $\varepsilon<t$, and normalisation by 2ε, as adequate).

Recall that we formulated the Markov property of the Poisson process as

$$\mathbb{P}((X_{t+s})_{s\ge 0}\in B|X_t=k,(X_r)_{r\le t}\in A)=\mathbb{P}_k((X_{t+s})_{s\ge 0}\in B)$$

for all events $\{(X_r)_{r\le t}\in A\}$ such that $\mathbb{P}(X_t=k,(X_r)_{r\le t}\in A)>0$, and $\{(X_{t+u})_{u\ge 0}\in B\}$. For certain sets A with zero probability, this can still be established by approximation.

A.3 Independence and conditional independence

Recall that independence of two random variables is defined as follows. Two discrete random variables X and Y are independent if

$$\mathbb{P}(X=j,Y=k)=\mathbb{P}(X=j)\mathbb{P}(Y=k) \qquad \text{for all } j,k\in\mathbb{S}.$$

Two jointly continuous random variables S and T are independent if their joint density factorises, i.e. if

$$f_{S,T}(s,t)=f_S(s)f_T(t) \qquad \text{for all } s,t\in\mathbb{R}, \text{ where } f_S(s)=\int_{\mathbb{R}} f_{S,T}(s,t)dt.$$

Recall also (or check) that this is equivalent, in both cases, to

$$\mathbb{P}(S\le s,T\le t)=\mathbb{P}(S\le s)\mathbb{P}(T\le t) \qquad \text{for all } s,t\in\mathbb{R}.$$

In fact, it is also equivalent to

$$\mathbb{P}(S\in A,T\in B)=\mathbb{P}(S\in B)\mathbb{P}(T\in B) \qquad \text{for all (measurable) } A,B\subset\mathbb{R},$$

and we define more generally:

Definition 136 Let X and Y be two random variables with values in any, possibly different spaces $\mathbb{X}$ and $\mathbb{Y}$. Then we call X and Y independent if

$$\mathbb{P}(X\in A,Y\in B)=\mathbb{P}(X\in A)\mathbb{P}(Y\in B) \qquad \text{for all (measurable) } A\subset\mathbb{X} \text{ and } B\subset\mathbb{Y}.$$

We call X and Y conditionally independent given a third random variable Z if for all $z\in\mathbb{S}$ (if Z has values in $\mathbb{S}$) or $z\in[0,\infty)$ (if Z has values in $[0,\infty)$),

$$\mathbb{P}(X\in A,Y\in B|Z=z)=\mathbb{P}(X\in A|Z=z)\mathbb{P}(Y\in B|Z=z).$$

Remark and Fact 137[1] Conditional independence is in many ways like ordinary (unconditional) independence. E.g., if X is discrete, it suffices to consider $A=\{x\}$, $x\in\mathbb{X}$. If X is real-valued, it suffices to consider $A=(-\infty,x]$, $x\in\mathbb{R}$. If X is bivariate, it suffices to consider all A of the form $A=A_1\times A_2$.

If $X=(X_r)_{r\le t}$, it suffices to consider $A=\{X_{r_1}=x_1,\ldots,X_{r_n}=x_n\}$ for all $0\le r_1<\cdots<r_n\le t$, $x_1,\ldots,x_n\in\mathbb{S}$, $n\ge 1$. This is how Proposition **??**(ii) can be interpreted, applied and proved.

We conclude with a fact that may seem obvious, but does not follow immediately from the definitions. Also the approximation argument only gives some special cases.

Fact 138 *Let X be any random variable, and T a $[0,\infty)$-valued random variable with right-continuous density. Then, for all (measurable) $f\colon\mathbb{X}\times[0,\infty)\to[0,\infty)$ and $t\ge 0$, we have*

$$\mathbb{E}(f(X,T)|T=t)=\mathbb{E}(f(X,t)|T=t).$$

Furthermore, if X and T are independent and $g:\mathbb{X}\to[0,\infty)$ (measurable), $t\ge 0$, we have

$$\mathbb{E}(g(X)|T=t)=\mathbb{E}(g(X)).$$

If X takes values in $[0,\infty)$ also, example for f are e.g. $f(x,t)=1_{\{x+t>s\}}$, where $1_{\{x+t>s\}}:=1$ if $x+t>s$ and $1_{\{x+t>s\}}:=0$ otherwise; or $f(x,t)=e^{\lambda(x+t)}$ in which case the statements are

$$\mathbb{P}(X+T>s|T=t)=\mathbb{P}(X+t>s|T=t) \text{ and } \mathbb{E}(e^{\lambda(X+T)}|T=t)=e^{\lambda t}\mathbb{E}(e^{\lambda X}|T=t),$$

and the condition $\{T=t\}$ can be removed on the right-hand sides if X and T are independent. This can be shown by the approximation argument.

The analogue of Fact 138 for discrete T is elementary.

[1] Facts are theorems that we cannot fully prove in this course. Note also that there is a grey zone between theorems/propositions and facts, since partial proofs of facts or full proofs of theorems/propositions sometimes appear on assignment sheets, in the main or optional parts.

A.4 Method: One-step analysis, conditioning on T_1

The main conditioning method in this course is to condition on the first event. In the case of a discrete-time Markov chain this is the value after the first step. In the case of a Poisson process (or simple birth process or renewal process, as studied later), this is the time of the first arrival. In the case of a continuous-time Markov chain, it will be a combination of the two.

Example 139 Let $X \sim \mathrm{PP}(\lambda)$ and $m(u) = \mathbb{E}(X_u)$, $u \geq 0$. Then by Proposition 135(c),

$$m(u) = \mathbb{E}(X_u) = \int_0^\infty \mathbb{E}(X_u | T_1 = t) f_{T_1}(t) dt = \int_0^u \mathbb{E}(1 + \widetilde{X}_{u-T_1} | T_1 = t) \lambda e^{-\lambda t} dt,$$

where $\widetilde{X}_s = X_{T_1+s} - 1$ is a $\mathrm{PP}(\lambda)$ independent of $T_1 = Z_0$, since $\widetilde{X}_s = \#\{n \geq 1 \colon \widetilde{T}_n \leq s\}$, where $\widetilde{Z}_n = Z_{n+1} \sim \mathrm{Exp}(\lambda)$, $n \geq 0$, are independent, and independent of Z_0. By Fact 138, this yields

$$m(u) = \int_0^u (1 + \mathbb{E}(X_{u-t})) \lambda e^{-\lambda t} dt = 1 - e^{-\lambda u} + \int_0^u m(r) \lambda e^{-\lambda(u-r)} dr.$$

If we multiply this by $e^{\lambda u}$, differentiate and cancel $e^{\lambda u}$ again, we find $m'(u) = \lambda$. Since also $m(0) = \mathbb{E}(X_0) = 0$, we obtain $m(t) = \lambda t$ for all $t \geq 0$, using a quite different argument from Remark 5. The real power of this argument will be revealed when applied for processes other than the Poisson process, for which many stronger tools yield stronger results.

A.5 $[0, \infty]$-valued random variables and some useful theorems

In the next section, we will study $T_\infty := \lim_{n\to\infty} T_n = \sum_{n\geq 0} Z_n$. In general, this is not a finite random variable, but may take values in $[0, \infty]$, infinity included. Recall that a function $F \colon [0, \infty) \to [0, 1]$ is a cumulative distribution for a distribution on $[0, \infty)$ if and only if F is right-continuous increasing with $F(\infty-) := \lim_{t\to\infty} F(t) = 1$. We drop the requirement $F(\infty-) = 1$:

Definition 140 A right-continuous increasing function $F \colon [0, \infty) \to [0, 1]$ with $F(\infty-) \leq 1$ is called a *cumulative distribution function*. A function $T \colon \Omega \to [0, \infty]$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a $[0, \infty]$-*valued random variable* if $\{T \leq t\} \in \mathcal{F}$ for all $t \in [0, \infty)$. Its cumulative distribution function is $F_T(t) = \mathbb{P}(T \leq t)$, $t \in [0, \infty)$. Its *expectation* is defined as

$$\mathbb{E}(T) := \int_0^\infty \mathbb{P}(T > t) dt = \int_0^\infty (1 - F_T(t)) dt.$$

We have $\mathbb{P}(T \leq t) \to \mathbb{P}(T < \infty) = F(\infty-)$ and $\mathbb{P}(T = \infty) = 1 - \mathbb{P}(T < \infty) = 1 - F(\infty-)$. Note that $\mathbb{E}(T) = \infty$ whenever $\mathbb{P}(T = \infty) > 0$ since then $\mathbb{P}(T > t) \geq \mathbb{P}(T = \infty) > 0$ makes the integral infinite. Recall that $\mathbb{E}(T) = \infty$ is also possible if $\mathbb{P}(T = \infty) = 0$, e.g. if $\mathbb{P}(T > t) = 1/t$, $t \geq 1$, so that $\mathbb{E}(T) = \int_1^\infty (1/t) dt = \infty$. Note also that, if T is $\mathbb{N} \cup \{\infty\}$-valued, the definition of $\mathbb{E}(T)$ implies

$$\mathbb{E}(T) = \sum_{0 \leq n \leq \infty} n \mathbb{P}(T = n) = \infty \mathbb{P}(T = \infty) + \sum_{0 \leq n < \infty} n \mathbb{P}(T = n)$$

and if F_T is differentiable with $F_T' = f_T$ on $(0, \infty)$ then

$$\mathbb{E}(T) = \infty \mathbb{P}(T = \infty) + \int_{(0,\infty)} t f_T(t) dt\,.$$

Here, we use the convention that $\infty \times p = \infty$ for $p > 0$, and $\infty \times 0 = 0$ to include the classical case $\mathbb{P}(T < \infty) = 1$.

The following is a result from measure theory that we will not prove here, but that we will refer to when we interchange certain sums, expectations and integrals.

Fact 141 (Tonelli) *The order of integration, countable summation and expectation can be interchanged whenever the integrand/summands/random variables are* nonnegative. *Specifically,*

$$\mathbb{E}\left(\sum_{n=0}^{\infty} X_n\right) = \sum_{n=0}^{\infty} \mathbb{E}(X_n), \qquad \int_0^\infty \sum_{n=0}^{\infty} f_n(x)dx = \sum_{n=0}^{\infty} \int_0^\infty f_n(x)dx$$

$$\int_0^\infty \int_0^x f(x,y)dydx = \int_0^\infty \int_y^\infty f(x,y)dxdy,$$

also when the integrals, sums, random variables or expectations are infinite.

There were already two applications of this, one each in Lectures 1 and 2. You may now wish to consider again the arguments of Remark 5 and Proposition 135.

Interchanging limits is more delicate, and there are monotone and dominated convergence for this purpose. In this course we will only interchange limits when this is justified by monotone or dominated convergence, but we do not have the time to prove these. Here are statements.

Fact 142 (Monotone convergence) *Expectations of an* increasing *sequence of* nonnegative *random variables Y_n converge $\lim_{n\to\infty} \mathbb{E}(Y_n) = \mathbb{E}(\lim_{n\to\infty} Y_n)$, also when the limits or expectations or random variables are infinite.*

Fact 143 (Dominated convergence) *Integrals of a pointwise convergent sequence of functions $f_n \to f$ converge if $|f_n| \le g$ for an integrable function g. Expectations of a pointwise convergent sequence of random variables $X_n \to X$ converge if $|X_n| \le Y$ for an integrable random variable Y, i.e.*

$$\int g(x)dx < \infty \quad \Rightarrow \quad \lim_{n\to\infty} \int f_n(x)dx = \int \lim_{n\to\infty} f_n(x)dx = \int f(x)dx$$

$$\mathbb{E}(Y) < \infty \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}\left(\lim_{n\to\infty} X_n\right) = \mathbb{E}(X).$$

For proofs of these (or equivalent or more general statements), we refer to Part A Integration and B8.1 Martingales Through Measure Theory.