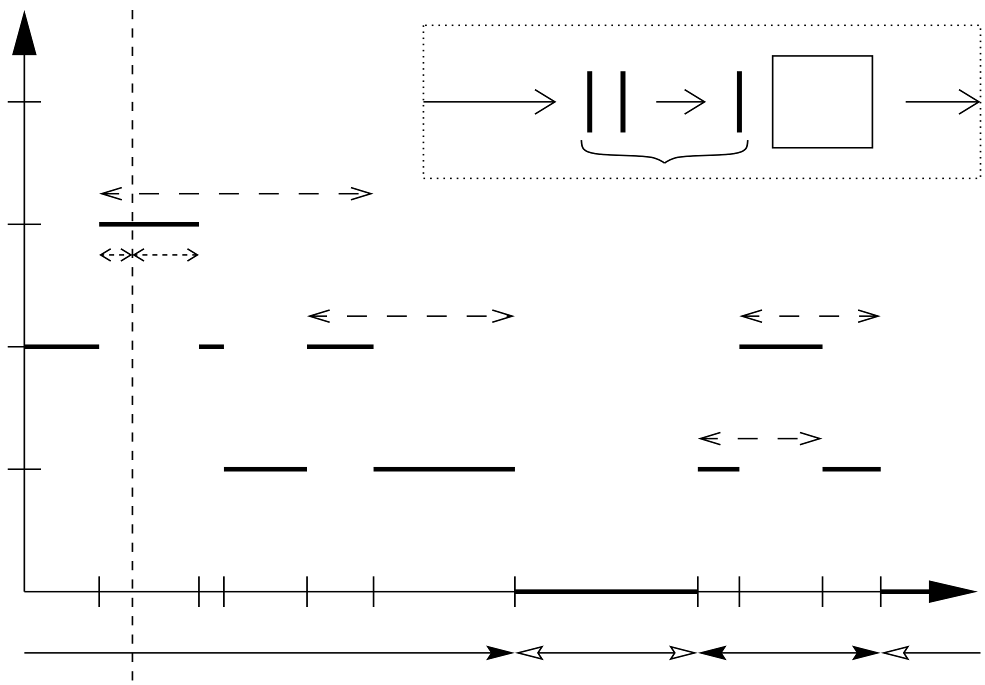
PART B
APPLIED PROBABILITY

Matthias Winkel
Department of Statistics
University of Oxford



MT 2007

Part B
Applied Probability

16 lectures MT 2007

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.

Applications of Markov chains 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.

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

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)
- D.R. Stirzaker, *Elementary Probability*, Cambridge University Press (1994)
- S.M. Ross, *Introduction to Probability Models*, 4th edition, Academic Press (1989)

Contents

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

This course is, in the first place, a course for 3rd year undergraduates who did Part A Probability in their second year. Other students from various M.Sc.'s are welcome as long as they are aware of 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, by the time we get to Lecture 5 in week 3.

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 that were treated at the very end of the Part A syllabus. The parts most relevant to us in today's lecture are again included here, and some more material is on the first assignment sheet.

This is a mathematics course. "Applied probability" means that we apply probability, but not so much Part A Probability but further probability building on Part A and not covered there, so effectively, we will be spending a lot of our time developing theory as required for certain examples and applications.

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 thereof). However, most results in the theory of Markov chains will be treated for any *countable* state space $\mathbb{S}$, which does not pose any complications as compared with $\mathbb{N}$, since one can always enumerate all states in $\mathbb{S}$ and hence give them labels in $\mathbb{N}$. For uncountable state spaces, however, several technicalities arise that are beyond the scope of this course, at least in any reasonable 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 choose an elementary definition that happens to be the most illustrative since it allows to draw pictures straight away.

Definition 1 Let $(Z_n)_{n\geq 0}$ be a sequence of independent exponential random variables $Z_n \sim Exp(\lambda)$ for a parameter (inverse mean) $\lambda \in (0,\infty)$, $T_0 = 0$, $T_n = \sum_{k=0}^{n-1} Z_k$, $n \geq 1$. Then the process $X = (X_t)_{t\geq 0}$ defined by

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

is called *Poisson process with rate* λ.

Note that $(X_t)_{t\geq 0}$ is not just a family of (dependent!) random variables but indeed $t \mapsto X_t$ is a random right-continuous function. This view is very useful since it is the formal justification for pictures of "typical realisations" of X.

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 Poisson process.

Let us collect some properties that, apart from some technical details, can serve as an alternative definition of the Poisson process.

Remark 2 A Poisson process X with rate λ has the following properties

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

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

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

To justify (i), calculate

$$\begin{aligned}
\mathbb{E}\left(q^{X_t}\right) &= \sum_{n=0}^{\infty} q^n \mathbb{P}(X_t = n) = \sum_{n=0}^{\infty} q^n \mathbb{P}(T_n \leq t, T_{n+1} > t) \\
&= \sum_{n=0}^{\infty} q^n \left(\mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t)\right) = 1 - \sum_{j=1}^{\infty} q^{j-1}(1-q)\mathbb{P}(T_j \leq t) \\
&= 1 - \int_0^t \sum_{j=1}^{\infty} q^{j-1}(1-q)\frac{\lambda^j}{(j-1)!} z^{j-1} e^{-\lambda z} dz \\
&= 1 - \int_0^t (1-q)\lambda e^{-\lambda z + \lambda q z} dz = e^{-\lambda t(1-q)},
\end{aligned}$$

where we used the well-known fact that T_n as a sum of independent $Exp(\lambda)$-variables has a $Gamma(n,\lambda)$ distribution. It is now easily checked that this is the probability generating function of the Poisson distribution with parameter λt. We conclude by the Uniqueness Theorem for probability generating functions. Note that we interchanged summation and integration. This may be justified by an expansion of $e^{-\lambda z}$ into a power series and using uniform convergence of power series twice. We will see another justification in Lecture 3.

(ii)-(iii) can be derived from the following Proposition 3, see also Lecture 4.

1.2 The Markov property

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 $\mathbb{S}$. For every $s_0\in\mathbb{S}$ this specifies the distribution of a Markov chain $(M_n)_{n\geq 0}$ starting from s_0 (under $\mathbb{P}_{s_0}$, say), by

$$\mathbb{P}_{s_0}(M_1=s_1,\ldots,M_n=s_n)=\mathbb{P}(M_1=s_1,\ldots,M_n=s_n|M_0=s_0)=\prod_{j=1}^{n}\pi_{s_{j-1},s_j}$$

We say that $(M_n)_{n\geq 0}$ is a Markov chain with transition matrix Π starting from s_0. There are several formulations of the Markov property:

- For all paths $s_0,\ldots,s_{n+1}\in\mathbb{S}$ of positive probability, we have

$$\mathbb{P}(M_{n+1}=s_{n+1}|M_0=s_0,\ldots,M_n=s_n)=\mathbb{P}(M_{n+1}=s_{n+1}|M_n=s_n)=\pi_{s_n,s_{n+1}}$$

- For all $s\in\mathbb{S}$ and events $\{(M_0,\ldots,M_n)\in A\}$ and $\{(M_n,M_{n+1},\ldots)\in B\}$, we have: if $\mathbb{P}(M_n=s,(M_j)_{0\leq j\leq n}\in A)>0$, then

$$\mathbb{P}((M_{n+k})_{k\geq 0}\in B|M_n=s,(M_j)_{0\leq j\leq n}\in A)=\mathbb{P}((M_{n+k})_{k\geq 0}\in B|M_n=s).$$

- $(M_j)_{0\leq j\leq n}$ and $(M_{n+k})_{k\geq 0}$ are conditionally independent given $M_n=s$, for all $s\in\mathbb{S}$. Furthermore, given $M_n=s$, $(M_{n+k})_{k\geq 0}$ is a Markov chain with transition matrix Π starting from s.

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.

In analogy, we will here study Poisson processes X starting from initial states $X_0=k\in\mathbb{N}$ (under $\mathbb{P}_k$), by which we just mean that we consider $X_t=k+\widetilde{X}_t$, $t\geq 0$, where $\widetilde{X}$ is a Poisson process starting from 0 as in the above definition.

Proposition 3 (Markov property) *Let X be a Poisson process with rate λ starting from 0 and $t\geq 0$ a fixed time. Then the following hold.*

(i) *For all $k\in\mathbb{N}$ and events $\{(X_r)_{r\leq t}\in A\}$ and $\{(X_{t+s})_{s\geq 0}\in B\}$, we have: if $\mathbb{P}(X_t=k,(X_r)_{r\leq t}\in A)>0$, then*

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

(ii) *Given* $X_t = k$, $(X_r)_{r\le t}$ *and* $(X_{t+s})_{s\ge 0}$ *are independent, and the conditional distribution of* $(X_{t+s})_{s\ge 0}$ *is that of a Poisson process with rate* λ *starting from* k.

(iii) $(X_{t+s} - X_t)_{s\ge 0}$ *is a Poisson process with rate* λ *starting from* 0*, independent of* $(X_r)_{r\le t}$.

We will prove a more general Proposition 17 in Lecture 4. Also, in Lecture 2, we will revise and push further the notion of conditioning. For this lecture we content ourselves with the formulation of the Markov property and proceed to the overview of the course.

Markov models (models that have the Markov property) 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 Brief summary of the course

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

- The Markov property of Proposition 3(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 for the arrival of claims. 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 keep their independence and identical distribution. This class of counting processes is 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 consist 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. If 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 times by any other distribution. There are also systems with $k = 2, 3, \ldots, \infty$ servers. The abstract queueing systems can be more concretely applied in telecommunication, computing networks, etc.

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

Lecture 2

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 preparing a more varied range of applications and a less mechanical use of probability than what was the focus of Part A. Along the way, we explain the meaning of statements such as the Markov properties of Lecture 1.

2.1 Modelling of events

Much of probability theory is about events and probabilities of events. Informally, this is an easy concept. Events like $A_1 =$"the die shows an even number" and $A_2 =$"the first customer arrives before 10am" make perfect sense in real situations. When it comes to assigning probabilities, things are less clear. We seem to be able to write down some ($\mathbb{P}(A_1) = 0.5$?) probabilities directly without much sophistication (still making implicit assumptions about the fairness of the die and the conduct of the experiment). Others ($\mathbb{P}(A_2)$) definitely require a mathematical model.

Hardly any real situations involve genuine randomness. It is rather our incomplete perception/information that makes us think there was randomness. In fact, assuming a specific random model in 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. Quite commonly, we have the following phenomenon: 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 a non-mathematical task. Analysing a model is purely mathematical. Models have to be consistent, i.e. not contain contradictions. This statement may seem superfluous, but 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 2 ii)-iii)), and times between two arrivals have a nonexponential 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 say, that a customer arrives exactly every six minutes. Another model could be to say, customers arrive according to a Poisson process at rate $\lambda = 10$ (time unit=1 hour). Whichever model you prefer, fact is, you can "calculate" $\mathbb{P}(A_2)$, and it is not the same in the two models, so we should reflect this in our notation. We don't want it to be A_2 that changes, so it must be $\mathbb{P}$, and we may wish to write $\tilde{\mathbb{P}}$ for the second model. $\mathbb{P}$ 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 you should wonder whether $\mathbb{P}(X_0 = k) > 0$).

Aside: Formally, there is a way to define random variables as functions $X_t : \Omega \to \mathbb{N}$, $Z_n : \Omega \to [0, \infty)$ etc. $\mathbb{P}_k$ can then be defined as a *measure* on Ω for all k, and this measure is compatible with our distributional assumptions which claim that

$$\text{the probability that } X_t = k + j \text{ is } \frac{(\lambda t)^j}{j!} e^{-\lambda t}$$

in that

$$\mathbb{P}_k(X_t = j + k) = \mathbb{P}_k(\{\omega \in \Omega : X_t(\omega) = j + k\}) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}.$$

In the mathematical sense, the set $A_{j+k} := \{X_t = j+k\} := \{\omega \in \Omega : X_t(\omega) = j+k\} \subset \Omega$ is called an *event*. Technically[1], we cannot in general call all subsets of Ω events if Ω is uncountable, but we will not worry about this, since it is very hard to find examples of *non-measurable sets*. ω should be thought of as a scenario, a realisation of all the randomness. Ω collects the possibilities and $\mathbb{P}$ tells us how likely each event is to occur. If we denote the set of all events by $\mathcal{A}$, then $(\Omega, \mathcal{A}, \mathbb{P})$ is called a *stochastic basis*. Its form is usually irrelevant. It is important that it 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 Ω. 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 $(Z_n)_{n \geq 1}$ of independent random variables, we do this for two reasons. The first should be apparent: countable families of independent variables are conceptually easier than uncountable families of dependent variables. The second is that a result in measure theory says that there exists a stochastic basis 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.

[1]The remainder of this paragraph is in a smaller font. This means (now and whenever something is in small font), that it can be skipped on first reading, and the reader may or may not want to get back to it at a later stage.

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. When learning probability, it is usual to first apply analytic *rules* to calculate distributions of functions of random variables (transformation formula for densities, expectation of a function of a random variable in terms of its density or probability mass function). 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 a result from measure theory that is intuitive enough to believe without proof. In a few cases, however, the gap is more serious. Every effort will be made to point out technicalities, but without drawing attention away from the probabilistic arguments that constitute this course and that are useful for applications.

B10a 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 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.

2.2 Conditional probabilities, densities and expectations

Conditional probabilities were introduced in Part A (or even Mods) as

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)},$$

where we require $\mathbb{P}(A) > 0$.

Example 4 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 2 (ii)-(iii).

Conditional densities were introduced as

$$f_{S|T}(s|t) = f_{S|T=t}(s) = \frac{f_{S,T}(s,t)}{f_T(t)}.$$

Example 5 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. Many models are stochastic only because the detailed deterministic structure is too complex. E.g. the counting process of natural catastrophies (hurricanes, earthquakes, volcanic eruptions, spring tide) etc. is genuinely deterministic. Since we cannot observe and model precisely weather, tectonic movements etc. it is much more fruitful to write down a stochastic model, e.g. a Poisson process, as a first approximation. We observe this process over time, and we can update the stochastic process by its realisation. Suppose we know the value of the intensity parameter $\lambda \in (0, \infty)$. (If we don't, any update will lead to a new estimate of λ, but we do not worry about this here). 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 6 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 f_T is right-continuous. To see this, when $f_{S,T}$ is sufficiently smooth, write 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 etc.), 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

$$\mathbb{P}(X = k | T = t) = \lim_{\varepsilon \downarrow 0} \mathbb{P}(X = k | t \le T \le t + \varepsilon)$$

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 7 *a) If X and Y are (dependent) discrete random variables in $\mathbb{N}$, then*

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} \mathbb{E}(X | Y = n) \mathbb{P}(Y = n).$$

If b) 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

$$\int_0^\infty \sum_{j=0}^\infty j\mathbb{P}(X=j|T=t)f_T(t)dt = \sum_{j=0}^\infty j\mathbb{P}(X=j)\int_0^\infty f_{T|X=j}(t)dt = E(X)$$

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

$$\mathbb{P}((X_{t+u})_{u\ge 0}\in B|X_t=k,(X_r)_{r\le t}\in A) = \mathbb{P}_k((X_{t+u})_{u\ge 0}\in B)$$

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.

2.3 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

$$\mathbb{P}(S \leq s, T \leq t) = \mathbb{P}(S \leq s)\mathbb{P}(T \leq t) \qquad \text{for all } s, t \in \mathbb{R}.$$

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 8 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 9 [2] Condional independence is in many ways like ordinary (unconditional) independence. E.g., if X is discrete, it suffices to check the condition for $A = \{x\}$, $x \in \mathbb{X}$. If Y is real-valued, it suffices to consider $B = (-\infty, y]$, $y \in \mathbb{R}$. If Y is bivariate, it suffices to consider all B of the form $B = B_1 \times B_2$.

If $\mathbb{X} = \{f : [0, t] \to \mathbb{N} \text{ right-continuous}\}$, it suffices to consider $A = \{f \in \mathbb{X} : f(r_1) = n_1, \ldots, f(r_m) = n_m\}$ for all $0 \leq r_1 < \ldots < r_m \leq t$ and $n_1, \ldots, n_m \in \mathbb{N}$. This is the basis for the meaning of Proposition 3(ii).

We conclude by a fact that may seem obvious, but does not follow immediately from the definition. Also the approximation argument only gives some special cases.

Fact 10 *Let X be any random variable, and T a real-valued random variable with right-continuous density. Then, for all (measurable) $f : \mathbb{X} \times [0, \infty) \to [0, \infty)$, 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) 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 10 for discrete T is elementary.

[2] Facts are results that are true, but that we cannot prove in this course. Note also that there is a grey zone between facts and propositions, since proofs or partial proofs sometimes appear on assignment sheets, in the main or optional parts.

Lecture 3

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 in analogy with our definition of Poisson processes. The common description (also for Markov chains) is always as follows: given the current state is m, what is the next part of the evolution, 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, but this can here be expressed in terms of genuine independence).

3.1 Definition and an example

If we consider the Poisson process as a model for a growing population, it is not always sensible to assume that new members are born at the same rate regardless what the size of the population. You would rather expect this rate to increase with size (more births in larger populations), although some saturation effects may make sense as well.

Also, if the Poisson process is used as a counting process of alpha particle emissions of a decaying radioactive substance, it makes sense to assume that the rate is decreasing with the number of emissions, particularly if the half-life time is short.

Definition 11 A stochastic process $X = (X_t)_{t\geq 0}$ of the form

$$X_t = k + \#\left\{n \geq 1 : \sum_{j=0}^{n-1} Z_j \leq t\right\}$$

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, $j \geq 0$, are independent exponential random variables with parameters λ_{k+j}, $j \geq 0$. X is called a $(k, (\lambda_n)_{n\geq 0})$-birth process.

Note that the parameter λ_n is attached with height n. The so-called holding time at height n has an $Exp(\lambda_n)$ distribution.

"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, form examples of continuous-time Markov chains, like simple birth processes do, as we will see soon.

Example 12 Consider a population in which each individual gives birth after an exponential time of parameter λ, all independently and repeatedly.

If n individuals are present, then the first birth will occur after an exponential time of parameter $n\lambda$. Then we have $n+1$ individuals and, by the lack of memory property (see Assignment 1), the process begins afresh with $n+1$ exponential clocks independent of the previous evolution ($n-1$ clocks still running with residual $Exp(\lambda)$ times and 2 new clocks from the individual that has just given birth and the new-born individual). By induction, the size of the population performs a birth process with rates $\lambda_n = n\lambda$.

Let X_t denote the number of individuals at time t and suppose $X_0 = 1$. Write T for the time of the first birth. Then by Proposition 7c)

$$\mathbb{E}(X_t) = \int_0^\infty \lambda e^{-\lambda u}\mathbb{E}(X_t|T=u)du = \int_0^t \lambda e^{-\lambda u}\mathbb{E}(X_t|T=u)du + e^{-\lambda t}$$

since $X_t = X_0 = 1$ if $T > t$.

Put $\mu(t) = \mathbb{E}(X_t)$, then for $0 < u < t$, intuitively $\mathbb{E}(X_t|T=u) = 2\mu(t-u)$ since from time u, the two individuals perform independent birth processes of the same type and we are interested in their population sizes $t-u$ time units later. We will investigate a more formal argument later, when we have the Markov property at our disposal.

Now

$$\mu(t) = \int_0^t 2\lambda e^{-\lambda u}\mu(t-u)du + e^{-\lambda t}$$

and setting $r = t-u$

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

Differentiating we obtain

$$\mu'(t) = \lambda\mu(t)$$

so the mean population size grows exponentially, and $X_0 = 1$ implies

$$\mathbb{E}(X_t) = \mu(t) = e^{\lambda t}.$$

3.2 Tonelli's Theorem, monotone and dominated convergence

The following is a result from measure theory that we cannot prove or dwell on here.

Fact 13 (Tonelli) *You may interchange order of integration, countable summation and expectation whenever the integrand/summands/random variables are* nonnegative, *e.g.*

$$\mathbb{E}\left(\sum_{n\geq 0} X_n\right) = \sum_{n\geq 0}\mathbb{E}(X_n), \qquad \int_0^\infty \sum_{n\geq 0} f_n(x)dx = \sum_{n\geq 0}\int_0^\infty f_n(x)dx$$

$$\int_0^\infty\int_0^x f(x,y)dydx = \int_0^\infty\int_y^\infty f(x,y)dxdy.$$

There were already two applications of this, one each in Lectures 1 and 2. The focus was on other parts of the argument, but you may wish to consider the justification of Remark 2 and Proposition 7 again now.

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 work out the details. Here are indicative statements.

Fact 14 (Monotone convergence) *Integrals (expectations) of an increasing sequence of nonnegative functions (random variables Y_n) converge (in the sense that* $\lim_{n\to\infty}\mathbb{E}(Y_n) = \mathbb{E}(\lim_{n\to\infty} Y_n)$*).*

Fact 15 (Dominated convergence) *Integrals (expectations) of a pointwise convergent sequence of functions $f_n \to f$ (applied to a random variable) converge, if the sequence $|f_n| \le g$ is dominated by an integrable function g (function which when applied to the random variable has finite expectation), i.e.*

$$\int g(x)dx < \infty \quad \Rightarrow \quad \lim_{n\to\infty}\int f_n(x)dx = \int \lim_{n\to\infty} f_n(x)dx$$
$$\mathbb{E}(g(X)) < \infty \quad \Rightarrow \quad \lim_{n\to\infty}\mathbb{E}(f_n(X)) = \mathbb{E}(\lim_{n\to\infty} f_n(X)).$$

We refer to B10a Martingales Through Measure Theory for those of you who follow that course. In any case, our working hypothesis is that, in practice, we may interchange all limits that come up in this course, but also know that these two theorems are required for formal justification, and refer to them.

3.3 Explosion

If the rates $(\lambda_n)_{n\in\mathbb{N}}$ increase too quickly, it may happen that infinitely many individuals are born in finite time (as with deterministic inter-birth times). We call this phenomenon *explosion*. Formally, we can express the possibility of explosion by $\mathbb{P}(T_\infty < \infty) > 0$ where $T_\infty = \lim_{n\to\infty} T_n$. 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 16 *Let X be a $(k, (\lambda_n)_{n\ge 0})$-birth process. Then*

$$\mathbb{P}(T_\infty < \infty) > 0 \quad \text{if and only if} \quad \sum_{m=k}^{\infty} \frac{1}{\lambda_m} < \infty.$$

Furthermore, in this case $\mathbb{P}(T_\infty < \infty) = 1$.

Proof: Note that

$$\begin{aligned}\mathbb{E}(T_\infty) &= \mathbb{E}\left(\sum_{n=0}^{\infty} Z_n\right) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(Z_n) \\ &= \sum_{m=k}^{\infty} \frac{1}{\lambda_m}\end{aligned}$$

where Tonelli's Theorem allows us to interchange summation and expectation. Therefore, if the series is finite, then $\mathbb{E}(T_\infty) < \infty$ implies $\mathbb{P}(T_\infty < \infty) = 1$.

Since, in general, $\mathbb{P}(T_\infty < \infty)$ does not imply $\mathbb{E}(T_\infty) < \infty$, we are not yet done for the converse. However, (using monotone convergence), and also the independence of the Z_n, we can calculate

$$\begin{aligned}-\log \mathbb{E}\left(e^{-T_\infty}\right) &= -\sum_{n=0}^{\infty} \log \mathbb{E}\left(e^{-Z_n}\right) \\ &= \sum_{m=k}^{\infty} \log\left(1 + \frac{1}{\lambda_m}\right).\end{aligned}$$

Either this latter sum is greater than $\log(2) \sum_{n=n_0}^{\infty} \frac{1}{\lambda_n}$ if $\lambda_n \geq 1$ for $n \geq n_0$, by linear interpolation and concavity of the logarithm. Or otherwise a restriction to any subsequence $\lambda_{n_k} \leq 1$ shows that the sum is infinite as each of these summands contributes at least $\log(2)$.

Therefore, if the series diverges, then $\mathbb{E}(e^{-T_\infty}) = 0$, i.e. $\mathbb{P}(T_\infty = \infty) = 1$. □

Note that we have not explicitly specified what happens after T_∞ if $T_\infty < \infty$. With a population size model in mind, $X_t = \infty$ for all $t \geq T_\infty$ is a reasonable convention. Formally, this means that X is a process in $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$. This process is often called the *minimal process*, since 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 after explosion. Such a process is then called *non-minimal.*

Lecture 4

Birth processes and the Markov property

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.

4.1 Statements of the Markov property

Proposition 17 (Markov property) *Let X be a simple birth process with rates $(\lambda_n)_{n\geq 0}$ starting from $k \geq 0$, and $t \geq 0$ a fixed time, $\ell \geq k$ a fixed height. Then given $X_t = \ell$, $(X_r)_{r\leq t}$ and $(X_{t+s})_{s\geq 0}$ are conditionally independent, and the conditional distribution of $(X_{t+s})_{s\geq 0}$ is that of a simple birth process with rates $(\lambda_n)_{n\geq 0}$ starting from k, we use notation*

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

Example 18 In formulas, we can write (and apply)

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

where $\widetilde{X}$ is an $(\ell, (\lambda_n)_{n\geq 0})$-birth process and A, B arbitrary sets of paths, e.g. $A = \{f : [0,t] \to \mathbb{N} : f(r_j) = n_j, j = 1, \ldots, m\}$ for some $0 \leq r_1 < \ldots < r_m \leq t$, $n_j \in \mathbb{N}$. Note that

$$(X_r)_{r\leq t} \in A \iff X_{r_j} = n_j \text{ for all } j = 1, \ldots, m.$$

Therefore, in particular

$$\begin{aligned}&\mathbb{P}(X_{r_1} = n_1, \ldots, X_{r_m} = n_m, X_{t+s_1} = p_1, \ldots, X_{t+s_h} = p_h | X_t = \ell) \\ &\quad = \mathbb{P}(X_{r_1} = n_1, \ldots, X_{r_m} = n_m | X_t = \ell)\mathbb{P}(\widetilde{X}_{s_1} = p_1, \ldots, \widetilde{X}_{s_h} = p_h)\end{aligned}$$

The Markov property as formulated in Proposition 17 (or Example 18) says that "past and future are (conditionally) independent given the present". We can say this in a rather different way as "the past is irrelevant for the future, only the present matters". Let us derive this reformulation using the elementary rules of conditional probability: the Markov property is about three events

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

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. In fact, by the definition of conditional probabilities, we always have

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

so that, by comparison, if $\mathbb{P}(E \cap C) > 0$, we here have $\mathbb{P}(F|E \cap C) = \mathbb{P}(F|C)$, and we deduce

Corollary 19 (Markov property, alternative formulation) *For all $t \ge 0$, $\ell \ge k$ and sets of paths A and B with $\mathbb{P}(X_t = \ell, (X_r)_{r\le t} \in A) > 0$, we have*

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

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

In fact, the condition $\mathbb{P}(X_t = \ell, (X_r)_{r\le t} \in A) > 0$ can often be waived, if the conditional probabilities can still be defined via an approximation by events of positive probability. This is a very informal statement since different approximations could, in principle, give rise to different values for such conditional probabilities. This is, in fact, very subtle, and uniqueness cannot be achieved in a strong sense, but some exceptional points usually occur. Formally, one would introduce versions of conditional probabilities, and sometimes, a nice version can be identified. For us, this is just a technical hurdle that we do not attempt. If we did attempt it, we would find enough continuity (or right-continuity), and see that we usually integrate conditional probabilities so that any non-uniqueness would not affect our final answers. See B10a Martingales Through Measure Theory for details.

4.2 Application of the Markov property

We make (almost) rigorous the intuitive argument used in Example 12 to justify $\mathbb{E}(X_u|T = t) = 2\mathbb{E}(X_{u-t})$, where $u \ge t$, X is a $(1, (n\lambda)_{n\ge 0})$-birth process and $T = \inf\{t \ge 0 : X_t = 2\}$. In fact, we now have

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

so that for all $n \ge 0$

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

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

$$\begin{aligned}\mathbb{E}(X_u|T=t) &= \sum_{n\in\mathbb{N}} n\mathbb{P}(X_u=n|T=t)\\ &= \sum_{n\in\mathbb{N}} n\mathbb{P}(X_{u-t}=n|X_0=2) = \mathbb{E}(\widetilde{X}_{u-t}) = 2\mathbb{E}(X_{u-t}),\end{aligned}$$

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

Remark: The event $\{T=t\}$ has probability zero ($\mathbb{P}(T=t)=0$ as T is exponentially distributed). Therefore any conditional probabilities given $T=t$ have to be approximated (using $\{t-\varepsilon<T\leq t\}$) in order to justify the application of the Markov property. We will not go through the details of this in this course and leave this to the insisting reader. Our focus is on the probabilistic argument that the application of the Markov property constitutes here.

4.3 Proof of the Markov property

Proof: Assume for simplicity that $X_0=0$. The general case can then be deduced. On $\{X_t=k\}=\{T_k\leq t<T_{k+1}\}$ we have

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

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

$$\mathbb{P}(\tilde{Z}_0>z|X_t=k) = \mathbb{P}(Z_k>(t-T_k)+z|Z_k>t-T_k\geq 0) = \mathbb{P}(Z_k>z)$$

where we applied the lack of memory property of Z_k to the *independent* threshold $t-T_k$. This actually requires a slightly more thorough explanation since we are dealing with repeated conditioning (first $X_t=k$, then $Z_k>t-T_k$), but the key result that we need is

Lemma 20 (Exercise A.1.4(b)) *The lack of memory property of the exponential distribution holds at independent thresholds, i.e. for $Z\sim Exp(\lambda)$ and L a random variable independent of Z, the following holds: given $Z>L$, $Z-L\sim Exp(\lambda)$ and $Z-L$ is conditionally independent of L.*

The proof is now completed again by the lack of memory property to see that $\tilde{Z}_0$ (as well as the $\tilde{Z}_j$, $j\geq 1$) is conditionally independent of $Z_0,\ldots,Z_{k-1}$ given $X_t=k$, and the assertion follows. □

4.4 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 only look at 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 21 (Strong Markov property) *(i) Let X be a simple birth process with rates $(\lambda_n)_{n\geq 0}$ and $T \geq 0$ a stopping time. Then for all $k \in \mathbb{N}$ with $\mathbb{P}(X_T = k) > 0$, we have given $T < \infty$ and $X_T = k$, $(X_r)_{r\leq T}$ and $(X_{T+s})_{s\geq 0}$ are independent, and the conditional distribution of $(X_{T+s})_{s\geq 0}$ is a simple birth process with rates $(\lambda_n)_{n\geq 0}$ starting from k.*

(ii) In the special case where X is a Poisson process and $\mathbb{P}(T < \infty) = 1$, $(X_{T+s} - X_T)_{s\geq 0}$ is a Poisson process starting from 0 independent of T and $(X_r)_{r\leq T}$.

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. Note however, that the proof the strong Markov property for first hitting times T_i is not so hard since then $\mathbb{P}(X_{T_i} = i) = 1$, so the only relevant statement is for $k = i$, 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, $j \geq i$. In fact, this is not just a sketch of a proof, but a complete proof, if we make two more remarks. First, conditioning on $\{X_{T_i} = i\}$ is like not conditioning at all, because this event has probability 1. Second, it is really enough to establish independence of holding times, because "can be expressed in terms of" is actually "is a function G of", where the first function e.g. goes from $[0, \infty)^i$ to the space

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

Now, we have a general result saying that if $G : \mathbb{A} \to \mathbb{X}$ and $H : \mathbb{B} \to \mathbb{Y}$ 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 22 (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 $(X_t)_{r \leq t}$.

Of course, you should think of X as an $\mathbb{N}$-valued simple birth process for the moment, but you will appreciate that this definition makes sense in much more generality. In our next lecture we will see examples where we are observing 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 23 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.

4.5 The Markov property for Poisson processes

In Proposition 3, we reformulated the Markov property of the Poisson process in terms of genuine independence rather than just conditional independence. Let $(X_t)_{t \geq 0}$ be a Poisson process with rate $\lambda > 0$, i.e. a $(0, (\lambda_n)_{n \geq 0})$-birth process with $\lambda_n = \lambda$ for all $n \geq 0$, then

$$(X_r)_{r \leq t} \coprod (X_{t+s} - X_t)_{s \geq 0} \sim (X_s)_{s \geq 0}.$$

Proof: By the version of the Markov property given in Proposition 17, we have for all $\ell \geq 0$, that

$$(X_r)_{r \leq t} \coprod_{X_t = \ell} (X_{t+s})_{s \geq 0} \sim (\ell + X_s)_{s \geq 0},$$

also since $(\ell + X_s)_{s \geq 0}$ is a Poisson process starting from ℓ. Specifically, it is not hard to see that for a general $(k, (\lambda_n)_{n \geq 0})$-birth process, the process $(\ell + X_s)_{s \geq 0}$ is a $(k + \ell, (\lambda_{n-\ell})_{n \geq 0}$-birth process (for any choice of λ_n, $-\ell \leq n \leq 0$), and here $\lambda_n = \lambda$ for all n. Clearly also

$$(X_r)_{r \leq t} \coprod_{X_t = \ell} (X_{t+s} - \ell)_{s \geq 0} \sim (X_s)_{s \geq 0}.$$

However, conditionally given $X_t = \ell$, we have $X_{t+s} - \ell = X_{t+s} - X_t$ for all $s \geq 0$, so that

$$(X_r)_{r\leq t} \coprod_{X_t=\ell} (X_{t+s} - X_t)_{s\geq 0} \sim (X_s)_{s\geq 0}.$$

Since the distribution of $(X_{t+s} - X_t)_{s\geq 0}$ (the distribution of a Poisson process of rate λ starting from 0) does not depend on $\ell \in \mathbb{N}$, we can now apply Exercise A.1.5 to conclude that this conditional independence is in fact genuine independence, and the common conditional distribution of $(X_{t+s} - X_t)_{s\geq 0}$ given $X_t = \ell$, $\ell \in \mathbb{N}$, is in fact also the unconditional distribution.

More precisely, we apply the statement of the exercise with $N = X_t$, $Z = (X_{t+s} - X_t)_{s\geq 0}$ and $M = (X_r)_{r\leq t}$. Strictly speaking, Z and M are here random variables taking values in *suitable* function spaces, but this does not pose any problems from a measure-theoretic perspective, once "suitable" has been defined, but this definition of "suitable" is again beyond the scope of this course. □

We can now also establish our assertions in Remark 2, namely X has stationary increments since

$$(X_{t+s} - X_t)_{s\geq 0} \sim (X_s)_{s\geq 0} \quad \Rightarrow \quad X_{t+u} - X_t \sim X_u \quad \text{for all } u \geq 0.$$

The subtlety of this statement is that we have for all $B \subset \{f : [0,\infty) \to \mathbb{N}\}$ (measurable)

$$\mathbb{P}((X_{t+s} - X_t)_{s\geq 0} \in B) = \mathbb{P}((X_s)_{s\geq 0} \in B),$$

and for all $n \geq 0$ and $B = B_{u,n} = \{f : [0,\infty) \to \mathbb{N} : f(u) = n\}$ this gives $\mathbb{P}(X_{t+u} - X_t = n) = \mathbb{P}(X_u = n)$ as required. Independence of two increments X_t and $X_{t+u} - X_t$ also follows directly. An careful inductive argument allows to extend this to the required independence of m successive increments $X_{t_j} - X_{t_{j-1}}$, $j = 1, \ldots, m$, because

$$\begin{aligned}
&\mathbb{P}(X_{t_{m+1}} - X_{t_m} = n_{m+1}, X_{t_m} - X_{t_{m-1}} = n_m, \ldots, X_{t_1} - X_{t_0} = n_1) \\
&\quad = \mathbb{P}(X_{t_{m+1}} - X_{t_m} = n_{m+1})\mathbb{P}(X_{t_m} - X_{t_{m-1}} = n_m, \ldots, X_{t_1} - X_{t_0} = n_1)
\end{aligned}$$

follows from the Markov property at $t = t_m$.

An alternative way to establish Proposition 3 is to give a direct proof of the stationarity and independence of increments in Remark 2, based on the lack of memory property, and deduce the Markov property from there.

Lecture 5

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, actually transitions between any two states in a state space $\mathbb{S}$, just as for discrete-time Markov chains and using the notion of a discrete-time Markov chain.

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

5.1 Definition and terminology

Definition 24 Let $(M_n)_{n\geq 0}$ be a discrete-time Markov chain on $\mathbb{S}$ with transition probabilities π_{ij}, $i,j \in \mathbb{S}$. Let $(Z_n)_{n\geq 0}$ be a sequence of conditionally independent exponential random variables with conditional distributions $Exp(\lambda_{M_n})$ given $(M_n)_{n\geq 0}$, where $\lambda_i \in (0,\infty)$, $i \in \mathbb{S}$. Then the process $X = (X_t)_{t\geq 0}$ defined by

$$X_t = M_n, \qquad T_n \leq t < T_{n+1}, n \geq 0, \qquad X_t = \infty, \quad T_\infty \leq t < \infty,$$

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

Usually, $T_\infty = \infty$, but the explosion phenomenon studied in Lecture 3 for the special case of a birth process has to be taken into account in a general definition. This is the so-called jump-chain holding-time definition of continuous-time Markov chains. There are others, and we will point these out when we have established relevant connections.

Here $Z_n \sim Exp(\lambda_{M_n})$ given M_n is short for $Z_n \sim Exp(\lambda_k)$ conditionally given $M_n = k$, for all $k \in \mathbb{S}$ and conditional independence given $(M_n)_{n\geq 0}$ means that for all $m \geq 0$, $Z_0, \ldots, Z_m$ are conditionally independent given $M_0 = k_0, \ldots, M_m = k_m$.

Example 25 (Birth processes) For $(k, (\lambda_n)_{n\geq 0})$-birth processes, we have $M_n = k + n$ deterministic, i.e. $\pi_{i,i+1} = 1$. Conditional independence of Z_n given M_n is independence, and $Exp(\lambda_{M_n}) = Exp(\lambda_{k+n})$ is the unconditional distribution of Z_k.

The representation of the distribution of X by $(\pi_{ij})_{i,j\in\mathbb{S}}$ and $(\lambda_i)_{i\in\mathbb{S}}$ is unique if we assume furthermore $\pi_{ii} \in \{0,1\}$ so that M either jumps straight away or remains in a given state forever, and by setting $\lambda_i = 0$ if $\pi_{ii} = 1$. This eliminates the possibility of the discrete chain proposing a "jump" from state i to itself.

It is customary to represent the transition probabilities π_{ij} and the holding rates λ_i in a single matrix, called the Q-matrix, as follows. Define for $i \neq j$

$$q_{ij} = \lambda_i \pi_{ij} \qquad \text{and} \qquad q_{ii} = -\lambda_i$$

Remark 26 $q_{ii} = -\sum_{j\neq i} q_{ij}$, since either $\sum_{j\neq i} \pi_{ij} = 1$ or $\lambda_i = 0$. As a consequence, the row sums of a Q-matrix vanish.

X is then also referred to as a *continuous-time Markov chain with Q-matrix Q*, a (k,Q)-Markov chain if starting from $X_0 = M_0 = k$.

Example 25 (continued) For birth processes, we obtain

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

As with discrete-time chains it is sometimes useful to specify an initial distribution for X_0 that we call ν, i.e. we let $\nu_i = \mathbb{P}(X_0 = i)$, $i \in \mathbb{S}$. Such a continuous-time Markov chain will be referred to as a (ν, Q)-Markov chain. Often

$$\nu_i = \delta_{ii_0} = \begin{cases} 1 & i = i_0 \\ 0 & i \neq i_0 \end{cases} \qquad \text{or short} \qquad \nu = \delta_{i_0}$$

δ_{ii_0} as a number derived from two arguments, here i and i_0, is called Kronecker delta, δ_{i_0} as a distribution only charging one point, here i_0, is called Dirac delta.

As an example of how an initial distribution can arise in practice, consider the number of customers arriving before a shop opens at time $t = 0$. As this number typically varies from day to day, it is natural to model it by a random variable, and to specify its distribution.

5.2 Construction

Defining lots of conditional distributions for infinite families of random variables requires some care in a measure-theoretic context. Also outside the measure-theoretic context, it is conceptually easier to express complicated random objects such as continuous-time Markov chains, in terms of a countable family of independent random variables. We have already done this for Poisson processes and can therefore use independent Poisson processes as building blocks. This leads to the following maze construction of a continuous-time Markov chain. It is the second appearance of the theory of competing exponentials and nicely illustrates the evolution of continuous-time Markov chains:

Proposition 27 *Let $M_0 \sim \nu$ and $(N_t^{ij})_{t\geq 0}$, $i,j \in \mathbb{S}$, $i \neq j$, independent Poisson processes with rates q_{ij}. Then define $T_0 = 0$ and for $n \geq 0$*

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

and

$$M_{n+1} = j \text{ if } T_{n+1} < \infty \text{ and } N_{T_{n+1}}^{M_n j} \neq N_{T_n}^{M_n j}.$$

Then

$$X_t = M_n, \qquad T_n \leq t < T_{n+1}, n \geq 0, \qquad X_t = \infty, \quad T_\infty \leq t < \infty,$$

is a (ν, Q)-Markov chain.

Think of the state space $\mathbb{S}$ as a maze where $q_{ij} > 0$ signifies that there is a gate from state $i \in \mathbb{S}$ to state $j \in \mathbb{S}$. Each gate (i,j) opens at the event times of a Poisson process N^{ij}. If after a given number of transitions the current state is i, then the next jump time of X is when the next gate leading away from i opens. If this gate leads from i to j, then j is the new state for X. Think of each Poisson process as a clock that rings at its event times. A ringing clock here corresponds to a gate opening instantaneously (i.e. immediately closing afterwards).

Proof: We have to check that the process defined here has the correct jump chain, holding times and dependence structure. Clearly $M_0 = X_0$ has the right starting distribution. Given $M_0 = i$, the first jump occurs at the first time at which one of the Poisson processes N^{ij}, $j \neq i$, has its first jump. This time is a minimum of independent exponential random variables of parameters q_{ij}, $T_1 = \inf\{T_1^{ij}, j \neq i\}$ for which

$$\mathbb{P}(T_1 > t) = \mathbb{P}(T_1^{ij} > t \text{ for all } j \neq i) = \prod_{j\neq i} \mathbb{P}(T_1^{ij} > t) = \exp\left\{-t \sum_{j \neq i} q_{ij}\right\} = e^{-\lambda_i t},$$

i.e. $Z_0 = T_1 \sim Exp(\lambda_{M_0})$ given M_0. Furthermore

$$\mathbb{P}(T_1 = T_1^{ij}) = \mathbb{P}(T_1^{ij} < \inf\{T_1^{ik}, k \notin \{i,j\}\}) = \frac{q_{ij}}{\lambda_i} = \pi_{ij}.$$

For independence, the second and inductively all further holding times and transitions, we apply the strong Markov property of the Poisson processes (Fact 21, or a combination of the lack of memory property at minima of exponential variables T_1^{ij} and at an independent exponential variable T_1 for N^{kj}, $k \neq i$, as on assignment sheet 1) to see that the post-T_1 Poisson processes $(N_{T_1+s}^{ij} - N_{T_1})_{s\geq 0}$ are Poisson processes themselves, and therefore the previous argument completes the induction step and hence the proof. T_1 is a stopping time since

$$\{T_1 \leq t\} = \bigcap_{j\neq i} \{T_1^{ij} \leq t\}$$

and the latter were expressed in terms of $(N_r^{ij})_{r\leq t}$, $j \neq i$, respectively, in Example 23. □

Corollary 28 (Markov property) *Let X be a (ν, Q)-Markov chain and $t \geq 0$ a fixed time. Then given $X_t = k$, $(X_r)_{r \leq t}$ and $(X_{t+s})_{s \geq 0}$ are independent, and the conditional distribution of $(X_{t+s})_{s \geq 0}$ is that of a (k, Q)-Markov chain.*

Proof: The post-t Poisson processes $(N^{ij}_{t+s} - N^{ij}_t)_{s \geq 0}$ are themselves Poisson processes, independent of the pre-t Poisson processes $(N^{ij}_r)_{0 \leq r \leq t}$. The post-$t$ behaviour of X only depends on X_t and the post-t Poisson processes. If we condition on $\{X_t = k\}$, then clearly $(X_{t+s})_{s \geq 0}$ is starting from k. □

Continuous-time Markov chains also have the strong Markov property. We leave the formulation to the reader. Its proof is beyond the scope of this course.

5.3 M/M/1 and M/M/s queues

Example 29 (M/M/1 queue) Let us model by X_t the number of customers in a single-server queueing system at time $t \geq 0$, including any customer currently being served. We assume that new customers arrive according to a Poisson process with rate λ, and that service times are independent $Exp(\mu)$ distributed.

Given a queue size of n, two transitions are possible. If a customer arrives (at rate λ), X increases to $n+1$. If the customer being served leaves (at rate μ), the X decreases to $n-1$. If no customer is in the system, only the former can happen. This amounts to a Q-matrix

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

X is indeed a continuous-time Markov chain, since given state $n \geq 1$ ($n = 0$) and two (one) independent clocks $Exp(\lambda)$ and $Exp(\mu)$ (unless $n = 0$) ticking, the theory of competing exponential clocks (Exercise A.1.2) shows that the system starts afresh with the residual clock and the new clock (except $n = 1$ and transition to 0) exponential and independent of the past, and the induction proceeds.

Example 30 (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 in the system.

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 maybe better represented by $q_{i,i+1} = \lambda$ for all $i \geq 0$, $q_{i,i-1} = i\mu$ for $1 \leq i \leq s$, $q_{i,i-1} = s\mu$ for $i \geq s$, $q_{ii} = -i\mu - \lambda$ for $0 \leq i \leq s$, $q_{ii} = -s\mu - \lambda$ for $i \geq s$, and $q_{ij} = 0$ otherwise. A slight variation of the argument for Exercise 29 shows that the M/M/s queue is a continuous-time Markov chain.

Lecture 6

Transition semigroups

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 are 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. While we will continue to use the Q-matrix to specify the distribution of a continuous-time Markov chain, transition matrices $P(t)$ give some of the most important probabilities related to a continuous-time Markov chain, but they are available explicitly only in a limited range of examples.

6.1 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 denote by

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

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

Example 31 For a Poisson process with rate λ, we have for $j \geq i$ or $n \geq 0$

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

by Remark 2. For fixed $t \geq 0$ and $i \geq 0$, these are $Poi(\lambda t)$ probabilities, shifted by i.

Proposition 32 *$(P(t))_{t\geq 0}$ is a semigroup, i.e. for all $t, s \geq 0$ we have $P(t)P(s) = P(t+s)$ in the sense of matrix multiplication, 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. □

We will remind ourselves of a fixed initial state $X_0 = i$ by writing $\mathbb{P}(\cdot|X_0 = i)$ or $\mathbb{P}_i(\cdot)$.

6.2 Backward equations

The following result is useful to calculate transition probabilities.

Proposition 33 *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.

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

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

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

i.e.

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

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

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

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

Suppose now, we have another non-negative solution $\tilde{p}_{ij}(t)$. Then, by integration, $\tilde{p}_{ij}(t)$ also satisfies the above integral equations (the δ_{ik} come from the initial conditions). Trivially

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

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

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

then as above

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

and therefore

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

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

Note that non-minimal solutions that satisfy the conditions for transition matrices can only exist if $\sum_{k \in \mathbb{S}} p_{ik}(t) < 1$ for some $i \in \mathbb{S}$ and $t \geq 0$, i.e. the continuous-time Markov chain must be explosive in the sense that $\mathbb{P}(T_\infty < \infty) > 0$, and then $p_{i\infty}(t) = 1 - \sum_{k \in \mathbb{S}} p_{ik}(t) > 0$.

6.3 Forward equations

Proposition 34 *If $\mathbb{S}$ is finite, then the transition matrices $(P(t))_{t \geq 0}$ of a (ν, Q)-Markov chain satisfy the forward equation*

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

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

Proof: See Assignment question A.3.5. □

Fact 35 *If $\mathbb{S}$ is infinite, then the statement of the proposition still holds for minimal (ν, Q)-Markov chains.*

Furthermore, $P(t)$ is the minimal nonnegative solution.

The proof of the proposition can be adapted under a uniformity assumption. This assumption will be sufficient for most practical purposes, but the general case is best proved by conditioning on the *last* jump before t. Since this is not a stopping time, the Markov property does not apply and calculations have to be done by hand, which is quite technical, see Norris 2.8.

In fact, both forward and backward equations admit unique solutions if the corresponding continuous-time Markov chain does not explode. This is the case in all practically relevant situations. The non-uniqueness arises since Markovian extensions of explosive chains other than the minimal extension that we consider, will also have transition semigroups that satisfy the backward equations.

Remark 36 Transition semigroups and the Markov property can form the basis for a definition of continuous-time Markov chains. In order to match our definition, we could say that a (ν, Q)-Markov chain is a process with *right-continuous sample paths* in $\mathbb{S}$ 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 $0 \le t_0 \le t_1 \le \ldots \le t_{n+1}$ and $i_0, \ldots, i_{n+1} \in \mathbb{S}$, where $P(t)$ satisfies the forward equations. See Norris 2.8.

6.4 Example

Example 37 For the Poisson process $q_{i,i+1} = \lambda$, $q_{ii} = -\lambda$, $q_{ij} = 0$ otherwise, hence we have forward equations

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

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 solutions. Writing $i + n$ rather than j is convenient because of the stationarity of increments in this special case of the Poisson process. Alternatively, we may consider the backward equations

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

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 2. 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

$$m_i(t) = \mathbb{E}_i(X_t) = \left.\frac{\partial}{\partial z} G_i(z, t)\right|_{z=1-}$$

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 the two most important distributional features of a model.

6.5 Matrix exponentials

It is tempting to say that the differential equation $P'(t) = QP(t)$, $P(0) = I$, has as its unique solution $P(t) = e^{tQ}$, that the same is true for $P'(t) = P(t)Q$, $P(0) = I$, and that the functional equation $P(s)P(t) = P(s+t)$ also has as its solutions precisely $P(t) = e^{tQ}$ for some Q. Remember, however that Q is a matrix, and, in general, the state space is countably infinite. Therefore, we have to define e^{tQ} in the first place, and to do this, we could use the (minimal or unique) solutions to the differential equations. Another, more direct, possibility is the exponential series

$$e^{tQ} := \sum_{n\geq 0} \frac{(tQ)^n}{n!}$$

where tQ is scalar multiplication, i.e. multiplication of each entry of Q by the scalar t, and $(tQ)^n$ is an n-fold matrix product. Let us focus on a finite state space $\mathbb{S}$, so that the only limiting procedure is the series over $n \geq 0$. It is natural to consider a series of matrices as a matrix consisting of the series of the corresponding entries. In fact, this works in full generality, as long as $\mathbb{S}$ is finite, see Norris 2.10.

For infinite state space, this is much harder, since every entry in a matrix product is then already a limiting quantity and one will either need uniform control over entries or use operator norms to make sense of the series of matrices. The limited benefits from such a theory are not worth setting up the technical apparatus in our context.

Lecture 7

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.

7.1 Communicating classes and irreducibility

We define the class structure characteristics as for discrete-time Markov chains.

Definition 38 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 *communicates with* j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

(c) We say $A \subset \mathbb{S}$ is a *communicating class* if it is an equivalence class for the equivalence relation $\leftrightarrow$ on $\mathbb{S}$, i.e. if for all $i, j \in A$ we have $i \leftrightarrow j$ and A is maximal with this property (for all $k \in \mathbb{S} - A$, $i \in A$ at most one of $i \to k$, $k \to i$ holds).

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

(e) We say that i is an *absorbing state* if $\{i\}$ is closed.

(f) We say that X is *irreducible* if $\mathbb{S}$ is (the only) communicating class.

In the following we denote by $M = (M_n)_{n\geq 0}$ the jump chain, $(Z_n)_{n\geq 0}$ the holding times that we used in the construction of $X = (X_t)_{t\geq 0}$.

Proposition 39 *Let X be a minimal 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* M.

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

(iv) $p_{ij}(t)>0$ *for all* $t>0$.

(v) $p_{ij}(t)>0$ *for some* $t>0$.

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

(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

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

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

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

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

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

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

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.

7.2 Recurrence and transience

Definition 40 Let X be a continuous-time Markov chain.

(a) $i \in \mathbb{S}$ is called *recurrent* if

$$\mathbb{P}_i(\{t \geq 0 : X_t = i\} \text{ is unbounded}) = 1.$$

(b) $i \in \mathbb{S}$ is called *transient* if

$$\mathbb{P}_i(\{t \geq 0 : X_t = i\} \text{ is bounded}) = 1.$$

Note that if X can explode starting from i and if X is a minimal continuous-time Markov chain, then i is certainly not recurrent.

Recall that $N_i = \inf\{n \geq 1 : M_n = i\}$ is called the first passage time of M to state i. We define

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

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 by $N_i^{(1)} = N_i$ and $N_i^{(m+1)} = \inf\{n > N_i^{(m)} : M_n = i\}$, $m \geq 1$, for M, and

$$H_i^{(m)} = T_{N_i^{(m)}},$$

$m \geq 1$, for X.

Proposition 41 *$i \in \mathbb{S}$ 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

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

with probability 1 by the argument for Proposition 16 (convergent and divergent sums of independent exponential variables) since $Z_{N_m} \sim Exp(\lambda_i)$ 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. $\square$

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

Corollary 42 *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 43 *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 44 *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. Now the total number G of visits of X to i must have a geometric distribution with parameter h_i since $\mathbb{P}_i(G = 1) = h_i$ and $\mathbb{P}_i(G = m|G \geq m) = h_i$, $m \geq 2$. Therefore, the total time spent in i is

$$\sum_{m=0}^{G-1} Z_{N_i^{(m)}} \sim Exp(h_i\lambda_i), \qquad \text{cf. Solution to Exercise A.2.6.}$$

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

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

The other implications can be established using similar arguments. □

7.3 Positive and null recurrence

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 is positive recurrence:

Definition 45 A 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.

Fact 46 *Positive recurrence is a class property.*

7.4 Examples

Example 47 The M/M/1 queue with $\lambda > 0$ and $\mu > 0$ is irreducible since 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 39 yields $m \leftrightarrow n$.

$\lambda > \mu$ 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, $L_i = \sup\{t \geq 0 : X_t = i\} < \infty$ for all $i \in \mathbb{N}$, and since $\{t \geq 0 : X_t = i\} \subset [0, L_i]$, 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 $Y_1, Y_2, \ldots \sim Exp(\lambda + \mu)$. 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 $\mathbb{E}_0(N_0)$.

Example 48 Let $\lambda \geq 0$ and $\mu \geq 0$. Consider a simple birth and death process with Q-matrix $Q = (q_{nm})_{n,m \in \mathbb{N}}$, where $q_{nn} = -n(\lambda + \mu)$, $q_{n,n+1} = n\lambda$, $q_{n,n-1} = n\mu$, $q_{nm} = 0$ otherwise.

- If $\mu = 0$ and $\lambda = 0$, then $Q \equiv 0$, all states are absorbing, so the communicating classes are $\{n\}$, $n \in \mathbb{N}$. They are all closed and positive recurrent.

- If $\mu = 0$ and $\lambda > 0$, then 0 is still absorbing since $q_{00} = 0$, but otherwise $n \to m$ if and only if $1 \leq n \leq m$. Again, the communicating classes are $\{n\}$, $n \in \mathbb{N}$, $\{0\}$ is closed and positive recurrent, but $\{n\}$ is open and transient for all $n \geq 1$, since the process will not return after the $Exp(n\lambda)$ holding time.

- If $\mu > 0$ and $\lambda = 0$, then $\{0\}$ is still absorbing, $\{n\}$ is an open transient class.

- If $\mu > 0$ and $\lambda > 0$, then $\{0\}$ is still absorbing, $\{1, 2, \ldots\}$ is an open and transient communicating class. It can be shown that the process when starting from $i \geq 1$ will be absorbed in $\{0\}$ if $\lambda \leq \mu$ and that it will do so with a probability in $(0, 1)$ if $\lambda > \mu$.

Lecture 8

Convergence to equilibrium

Reading: Norris 3.5-3.8
Further reading: Grimmett-Stirzaker 6.9; Ross 6.5-6.6

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

$$\mathbb{S} = \bigcup_{m\geq 1} \mathcal{R}_m \cup \bigcup_{m\geq 1} \mathcal{T}_m,$$

where the (states in) $\mathcal{R}_m$ are recurrent, hence closed, and the $\mathcal{T}_m$ are transient, whether closed or not. This is the same as for discrete-time Markov chains, in fact equivalent to the decomposition for the associated jump chain. Furthermore, each recurrent class is either positive recurrent or null recurrent.

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

8.1 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 49 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 we take an invariant distribution ξ as initial distribution, then $X_t \sim \xi$ for all $t \geq 0$. We then say that X is in equilibrium.

Proposition 50 *If $\mathbb{S}$ is finite, then ξ is invariant if and only if $\xi Q = 0$.*

Proof: 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

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

where we applied the backward equation. Here, also the integration is understood componentwise.

Interchanging limits/integrals and matrix multiplication is justified since $\mathbb{S}$ is finite. □

Fact 51 *If $\mathbb{S}$ is infinite, Q is a Q-matrix and $(P(t))_{t \geq 0}$ are the transition matrices of the* minimal *continuous-time Markov chain associated with Q-matrix Q. Then $\xi Q = 0$ if and only if $\xi P(t) = \xi$ for all $t \geq 0$.*

As a consequence, ξ can then only exist if X is non-explosive in the sense that $\mathbb{P}(T_\infty = \infty) = 1$.

Fact 52 *An irreducible (minimal) 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 proof is quite technical and does not give further intuition. 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. Since this is a key result for both Convergence Theorem and Ergodic Theorem, the diligent reader may want to refer to Norris Theorem 3.5.3.

Example 53 Consider the M/M/1 queue of Example 29. 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 mass 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 Fact 52, 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 $\lambda \uparrow \mu$.

8.2 Convergence to equilibrium

The convergence theorem 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.

Theorem 54 *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 result for discrete-time Markov chains by looking at the processes $Z_n^{(h)} = X_{nh}$ that are easily seen to be Markov chains with transition matrices $P(h)$.

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

Sketch of proof: Let X be the continuous-time Markov chain starting according to ν, Y an independent continuous-time Markov chain with the same Q-matrix, but starting from the invariant distribution ξ. 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 $\widetilde{X}_t = X_t$, $t < T$ (following the ν-chain before T), $\widetilde{X}_t = Y_t$, $t \geq T$ (following the ξ-chain after T). The following three steps complete the proof:

1. the meeting time T is finite with probability 1 (this is because $\eta_{ij} = \xi_i\xi_j$ is stationary distribution for the bivariate process (X, Y) and existence of stationary distribution implies positive recurrence of X, Y (by Fact 52);

2. the third chain $\widetilde{X}$ has the same distribution as the ν-chain X;

3. the third chain (which eventually coincides with the ξ-chain Y) is asymptotically in equilibrium in the sense of the convergence statement in Theorem 54.

□

Note that we obtain the uniqueness of the invariant distribution as a consequence since also the marginal distribution of a Markov chain starting from a second invariant distribution would remain invariant and converge to ξ.

Theorem 55 (Ergodic theorem) *In the setting of Theorem 54, $X_0 \sim \nu$*

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

Proof: A proof using renewal theory is in assignment question 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 spatial average of Theorem 54. This is of great 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 54, we would need many independent observations of the same system at a large time t to estimate ξ.

8.3 Detailed balance equations and time reversal

Proposition 56 *Consider a Q-matrix Q. 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 26). 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 57 *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

$$\sum_{i \neq j} \hat{q}_{ji} = \sum_{i \neq j} \frac{\xi_i}{\xi_j} q_{ij} = -q_{jj} = -\hat{q}_{jj}$$

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 the path structure, Remark 36 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* $\xi_i q_{ij} = \xi_j q_{ji}$, $i, j \in \mathbb{S}$.

8.4 Erlang's formula

Example 58 Consider the birth-death process with birth rates $q_{i,i+1} = \lambda_i$ and death rates $q_{i,i-1} = \mu_i$, $q_{ii} = -\lambda_i - \mu_i$, $i \in \mathbb{N}$, all other entries zero (and also $\mu_0 = 0$). (This is standard notation for this type of process, but note that $\lambda_i = q_{i,i+1}$, we will *not* use earlier notation $\lambda_i = -q_{ii}$).

We recognise birth processes and queueing systems as special cases.

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

$$\xi_1\mu_1 - \xi_0\lambda_0 = 0 \qquad \text{and} \qquad \xi_{n+1}\mu_{n+1} - \xi_n(\lambda_n + \mu_n) + \xi_{n-1}\lambda_{n-1} = 0, \quad n \geq 1$$

or more easily the detailed balance equations

$$\xi_i\lambda_i = \xi_{i+1}\mu_{i+1}.$$

giving

$$\xi_n = \frac{\lambda_{n-1}\dots\lambda_0}{\mu_n\dots\mu_1}\xi_0$$

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

$$\xi_0 = \frac{1}{S} \quad \text{where} \quad S = 1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\dots\lambda_0}{\mu_n\dots\mu_1}$$

provided S is finite.

If S is infinite, then there does not exist an invariant distribution. This cannot be deduced from the detailed balance equations but can be argued directly by showing that $\xi Q = 0$ does not have a solution. It does not necessarily mean explosion in finite time, but includes all simple birth processes since they model growing populations and cannot be in equilibrium. By Fact 52, it means that X is then null recurrent or transient.

On the other hand, if $\lambda_0 = 0$ 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. E.g., if $\lambda_n = \lambda$, $n \geq 0$, $\mu_n = n\mu$, we get

$$\xi_n = \frac{(\lambda/\mu)^n}{n!}e^{-\lambda/\mu}.$$

You recognise the 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, $\lambda_n = \lambda$ means that arrivals occur according to a Poisson process, this can model immigration; $\mu_n = n\mu$ is obtained from as many $Exp(\mu)$ clocks as individuals in the population, i.e. independent $Exp(\mu)$ lifetimes for all individuals. Second, as a queueing model with arrivals according to a Poisson process, each individual leaves the system after an $Exp(\mu)$ time, no matter how many other people are in the system – this can be obtained from infinitely many servers working at rate μ.

Lecture 9

The Strong Law of Large Numbers

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

With the Convergence Theorem (Theorem 54) and the Ergodic Theorem (Theorem 55) we have two very different statements of convergence of something to a stationary distribution. We are looking at a recurrent Markov chain $(X_t)_{t\geq 0}$, i.e. one that visits *every* state at arbitrarily large times, so clearly X_t itself does not converge, as $t \to \infty$. In this lecture, we look more closely at the different types of convergence and develop methods to show the so-called almost sure convergence, of which the statement of the Ergodic Theorem is an example.

9.1 Modes of convergence

Definition 59 Let X_n, $n \geq 1$, and X be 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 $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$.

Almost sure convergence is the notion that we will study in more detail here. It helps to consider random variables as functions $X_n : \Omega \to \mathbb{R}$ on a sample space Ω, or at least as functions of a common, typically infinite, family of independent random variables. What is different here from previous parts of the course (except for the Ergodic Theorem, which we yet have to inspect more thoroughly), is that we want to calculate probabilities that fundamentally depend on an infinite number of random variables. So far, we have been able to revert to events depending on only finitely many random variables by conditioning. This will not work here.

Let us start by recalling the definition of convergence of sequences, as $n \to \infty$,

$$x_n \to x \qquad \iff \qquad \forall_{m\ge 1}\exists_{n_m\ge 1}\forall_{n\ge n_m}|x_n - x| < 1/m.$$

If we want to consider all sequences $(x_n)_{n\ge 1}$ of possible values of the random variables $(X_n)_{n\ge 1}$, then

$$n_m = \inf\{k \ge 1 : \forall_{n\ge k}|x_n - x| < 1/m\} \in \mathbb{N} \cup \{\infty\}$$

will vary as a function of the sequence $(x_n)_{n\ge 1}$, and so it will become a random variable

$$N_m = \inf\{k \ge 1 : \forall_{n\ge k}|X_n - X| < 1/m\} \in \mathbb{N} \cup \{\infty\}$$

as a function of $(X_n)_{n\ge 1}$. This definition of N_m permits us to write

$$\mathbb{P}(X_n \to X) = \mathbb{P}(\forall_{m\ge 1}N_m < \infty).$$

This will occasionally help, when we are given almost sure convergence, but is not much use when we want to prove almost sure convergence. To prove almost sure convergence, we can transform as follows

$$\begin{aligned} & \mathbb{P}(X_n \to X) = \mathbb{P}(\forall_{m\ge 1}\exists_{N\ge 1}\forall_{n\ge N}|X_n - X| < 1/m) = 1 \\ \iff \quad & \mathbb{P}(\exists_{m\ge 1}\forall_{N\ge 1}\exists_{n\ge N}|X_n - X| \ge 1/m) = 0. \end{aligned}$$

We are used to events such as $A_{m,n} = \{|X_n - X| \ge 1/m\}$, and we understand events as subsets of Ω, or loosely identify this event as set of all $((x_k)_{k\ge 1}, x)$ for which $|x_n - x| \ge 1/m$. This is useful, because we can now translate $\exists_{m\ge 1}\forall_{N\ge 1}\exists n \ge N$ into set operations and write

$$\mathbb{P}(\cup_{m\ge 1} \cap_{N\ge 1} \cup_{n\ge N} A_{m,n}) = 0.$$

This event can only have zero probability if all events $\cap_{N\ge 1} \cup_{n\ge N} A_{m,n}$, $m \ge 1$, have zero probability (formally, this follows from the sigma-additivity of the measure $\mathbb{P}$). The Borel-Cantelli lemma will give a criterion for this.

Proposition 60 *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.

Proof: Most of this is Part A material. Some counterexamples are on Assignment 5. It remains to prove that almost sure convergence implies convergence in probability. Suppose, $X_n \to X$ almost surely, then the above considerations yield $\mathbb{P}(\forall_{m\ge 1}N_m < \infty) = 1$, i.e. $\mathbb{P}(N_k < \infty) \ge \mathbb{P}(\forall_{m\ge 1}N_m < \infty) = 1$ for all $k \ge 1$.

Now fix $\varepsilon > 0$. Choose $m \ge 1$ such that $1/m < \varepsilon$. Then clearly $|X_n - X| > \varepsilon > 1/m$ implies $N_m > n$ so that

$$\mathbb{P}(|X_n - X| > \varepsilon) \le \mathbb{P}(N_m > n) \to \mathbb{P}(N_m = \infty) = 0,$$

as $n \to \infty$, for any $\varepsilon > 0$. Therefore, $X_n \to X$ in probability. □

9.2 The first Borel-Cantelli lemma

Let us now work on a sample space Ω. It is safe to think of $\Omega = \mathbb{R}^{\mathbb{N}} \times \mathbb{R}$ and $\omega \in \Omega$ as $\omega = ((x_n)_{n\geq 1}, x)$ as the set of possible outcomes for an infinite family of random variables (and a limiting variable).

The Borel-Cantelli lemmas are useful to prove almost sure results. Particularly limiting results often require certain events to happen infinitely often (i.o.) or only a finite number of times. Logically, this can be expressed as follows. Consider events $A_n \subset \Omega$, $n \geq 1$. Then

$$\omega \in A_n \text{ i.o.} \quad \Longleftrightarrow \quad \forall_{n\geq 1}\exists_{m\geq n} \quad \omega \in A_m \quad \Longleftrightarrow \quad \omega \in \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m$$

Lemma 61 (Borel-Cantelli (first lemma)) *Let $A = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m$ be the event that infinitely many of the events A_n occur. Then*

$$\sum_{n\geq 1} \mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \mathbb{P}(A) = 0$$

Proof: We have that $A \subset \bigcup_{m\geq n} A_m$ for all $n \geq 1$, and so

$$\mathbb{P}(A) \leq \mathbb{P}\left(\bigcup_{m\geq n} A_m\right) \leq \sum_{m\geq n} \mathbb{P}(A_m) \to 0 \qquad \text{as } n \to \infty$$

whenever $\sum_{n\geq 1} \mathbb{P}(A_n) < \infty$. □

9.3 The Strong Law of Large Numbers

Theorem 62 *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed (iid) random variables with $\mathbb{E}(X_1^4) < \infty$ and $\mathbb{E}(X_1) = \mu$. Then*

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

Fact 63 *Theorem 62 remains valid without the assumption $\mathbb{E}(X_1^4) < \infty$, just assuming $\mathbb{E}(|X_1|) < \infty$.*

The proof for the general result is hard, but under the extra moment condition $\mathbb{E}(X_1^4) < \infty$ there is a nice proof.

Lemma 64 *In the situation of Theorem 62, there is a constant $K < \infty$ such that for all $n \geq 0$*

$$\mathbb{E}((S_n - n\mu)^4) \leq Kn^2.$$

Proof: Let $Z_k = X_k - \mu$ and $T_n = Z_1 + \ldots + Z_n = S_n - n\mu$. Then

$$\mathbb{E}(T_n^4) = \mathbb{E}\left(\left(\sum_{i=1}^{n} Z_i\right)^4\right) = n\mathbb{E}(Z_1^4) + 3n(n-1)\mathbb{E}(Z_1^2 Z_2^2) \leq Kn^2$$

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

$$\mathbb{E}(Z_1 Z_2^3) = \mathbb{E}(Z_1)\mathbb{E}(Z_2^3) = 0.$$

K was chosen appropriately, say $K = 4\max\{\mathbb{E}(Z_1^4), (\mathbb{E}(Z_1^2))^2\}$. □

Proof of Theorem 62: By the lemma,

$$\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^4\right) \leq Kn^{-2}$$

Now, by Tonelli's theorem,

$$\mathbb{E}\left(\sum_{n\geq 1}\left(\frac{S_n}{n} - \mu\right)^4\right) = \sum_{n\geq 0}\mathbb{E}\left(\left(\frac{S_n}{n} - \mu\right)^4\right) < \infty \quad \Rightarrow \quad \sum_{n\geq 1}\left(\frac{S_n}{n} - \mu\right)^4 < \infty \quad \text{a.s.}$$

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

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

□

This proof did not use the Borel-Cantelli lemma, but we can also conclude by the Borel-Cantelli lemma:

Proof of Theorem 62: We know by Markov's inequality that

$$\mathbb{P}\left(\frac{1}{n}|S_n - n\mu| \geq n^{-\gamma}\right) \leq \frac{\mathbb{E}((S_n/n - \mu)^4)}{n^{-4\gamma}} = Kn^{-2+4\gamma}.$$

Define for $\gamma \in (0, 1/4)$

$$A_n = \left\{\frac{1}{n}|S_n - n\mu| \geq n^{-\gamma}\right\} \quad \Rightarrow \quad \sum_{n\geq 1}\mathbb{P}(A_n) < \infty \quad \Rightarrow \quad \mathbb{P}(A) = 0$$

by the first Borel-Cantelli lemma, where $A = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_n$. But now, event A^c happens if and only if

$$\exists_N \forall_{n\geq N} \left|\frac{S_n}{n} - \mu\right| < n^{-\gamma} \qquad \Rightarrow \qquad \frac{S_n}{n} \to \mu.$$

□

9.4 The second Borel-Cantelli lemma

We won't need the second Borel-Cantelli lemma in this course, but include it for completeness.

Lemma 65 (Borel-Cantelli (second lemma)) *Let* $A = \bigcap_{n\geq 1}\bigcup_{m\geq n} A_n$ *be the event that infinitely many of the events* A_n *occur. Then*

$$\sum_{n\geq 1}\mathbb{P}(A_n) = \infty \text{ and } (A_n)_{n\geq 1} \text{ independent} \quad \Rightarrow \quad \mathbb{P}(A) = 1.$$

Proof: The conclusion is equivalent to $\mathbb{P}(A^c) = 0$. By de Morgan's laws

$$A^c = \bigcup_{n\geq 1}\bigcap_{m\geq n} A_m^c.$$

However,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{m\geq n} A_m^c\right) &= \lim_{r\to\infty}\mathbb{P}\left(\bigcap_{m=n}^{r} A_m^c\right)\\ &= \prod_{m\geq n}(1-\mathbb{P}(A_m)) \leq \prod_{m\geq n}\exp(-\mathbb{P}(A_m)) = \exp\left(-\sum_{m\geq n}\mathbb{P}(A_m)\right) = 0\end{aligned}$$

whenever $\sum_{n\geq 1}\mathbb{P}(A_n) = \infty$. Thus

$$\mathbb{P}(A^c) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{m\geq n} A_m^c\right) = 0.$$

□

As a technical detail: to justify some of the limiting probabilities, we use "continuity of $\mathbb{P}$" along increasing and decreasing sequences of events, that follows from the sigma-additivity of $\mathbb{P}$, cf. Grimmett-Stirzaker, Lemma 1.3.(5).

9.5 Examples

Example 66 (Arrival times in Poisson process) A Poisson process has independent and identically distributed inter-arrival times $(Z_n)_{n\geq 0}$ with $Z_n \sim Exp(\lambda)$. We denoted the partial sums (arrival times) by $T_n = Z_0 + \ldots + Z_{n-1}$. The Strong Law of Large Numbers yields

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

Example 67 (Return times of Markov chains) For a positive-recurrent discrete-time Markov chain we denoted by

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

the successive return times to 0. By the strong Markov property, the random variables $N_i^{(m+1)} - N_i^{(m)}$, $m \geq 1$ are independent and identically distributed. If we define $N_i^{(0)} = 0$ and start from i, then this holds for $m \geq 0$. 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, in continuous time, for

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

we get

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

Example 68 (Empirical distributions) If $(Y_n)_{n\geq 1}$ is an infinite sample (independent and identically distributed random variables) from a discrete distribution ν on $\mathbb{S}$, then the random variables $B_n^{(i)} = 1_{\{Y_n=i\}}$, $n \geq 1$, 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

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

almost surely, as $n \to \infty$. The probability mass 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 69 (Discrete ergodic theorem) If $(M_n)_{n\geq 0}$ is a positive-recurrent discrete-time Markov chain, the Ergodic Theorem is a statement very similar to the example of empirical distributions

$$\frac{\#\{k = 0, \ldots, n-1 : M_k = i\}}{n} \to \mathbb{P}_\eta(M_0 = i) = \eta_i \qquad \text{almost surely, as } n \to \infty,$$

for a stationary distribution η, but of course, the M_n, $n \geq 0$, are not independent (in general). Therefore, we need to work a bit harder to deduce the Ergodic Theorem from the Strong Law of Large Numbers.

Lecture 10

Renewal processes and equations

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

10.1 Motivation and definition

So far, the topic has been continuous-time Markov chains, and we've introduced them as discrete-time Markov chains with exponential holding times. In this setting we have a theory very much similar to the discrete-time theory, with independence of future and past given the present (Markov property), transition probabilities, invariant distributions, class structure, convergence to equilibrium, ergodic theorem, time reversal, detailed balance etc. A few odd features can occur, mainly due to explosion.

These parallels are due to the exponential holding times and their lack of memory property which is the key to the Markov property in continuous time. In practice, this assumption is often not reasonable.

Example 70 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). We started the discussion of continuous-time Markov chains with birth processes as generalised Poisson processes. Similarly, we start here generalising the Poisson process to have non-exponential but independent identically distributed inter-arrival times.

Definition 71 Let $(Z_n)_{n\geq 0}$ be a sequence of independent identically distributed positive random variables, $T_n = \sum_{k=0}^{n-1} Z_k$, $n \geq 1$, the partial sums. 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 *inter-arrival distribution*.

Example 72 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

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

counting the visits to i is thus a renewal process.

10.2 The renewal function

Definition 73 The function $t \mapsto m(t) := \mathbb{E}(X_t)$ is called the *renewal function.*

It plays an important role in renewal theory. Remember that for $Z_n \sim Exp(\lambda)$ we had $X_t \sim Poi(\lambda t)$ and in particular $m(t) = \mathbb{E}(X_t) = \lambda t$.

To calculate the renewal function for general renewal processes, we should investigate the distribution of X_t. Note that, as for birth processes,

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

so that we can express

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

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

Recall that for two independent continuous random variables S and T with densities f and g, the random variable $S + T$ has density

$$(f * g)(u) = \int_{-\infty}^{\infty} f(u-t)g(t)dt, \quad u \in \mathbb{R},$$

the *convolution (product)* of f and g, and if $S \geq 0$ and $T \geq 0$, then

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

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 translate as $S + T = T + S$ and $(S + T) + U = S + (T + U)$ 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 we will meet other relevant examples soon. We can define convolution powers $f^{*(1)} = f$ and $f^{*(k+1)} = f * f^{*(k)}$, $k \geq 1$. Then

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

if Z_n, $n \geq 0$, are continuous with density f.

Proposition 74 *Let X be a renewal process with interarrival density f. Then $m(t) = \mathbb{E}(X_t)$ is differentiable in the weak sense that it is the integral function of*

$$m'(s) := \sum_{k=1}^{\infty} f^{*(k)}(s)$$

Lemma 75 *Let X be an $\mathbb{N}$-valued random variable. Then $\mathbb{E}(X) = \sum_{k\geq 1} \mathbb{P}(X \geq k)$.*

Proof: We use Tonelli's Theorem

$$\sum_{k\geq 1} \mathbb{P}(X \geq k) = \sum_{k\geq 1} \sum_{j\geq k} \mathbb{P}(X = j) = \sum_{j\geq 1} \sum_{k=1}^{j} \mathbb{P}(X = j) = \sum_{j\geq 0} j\mathbb{P}(X = j) = \mathbb{E}(X).$$

□

Proof of Proposition 74: Let us integrate $\sum_{k=1}^{\infty} f^{*(k)}(s)$ using Tonelli's Theorem

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

□

10.3 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 76 *Let X be a renewal process with interarrival density f. Then $m(t) = \mathbb{E}(X_t)$ is the unique (locally bounded) solution of*

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

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

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

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

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:

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

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

Example 77 We can express m as follows: $m = F + F * \sum_{k\geq 1} f^{*(k)}$. Indeed, we check that $\ell = F + F * \sum_{k\geq 1} f^{*(k)}$ satisfies the renewal equation:

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

just using properties of the convolution product. By Proposition 76, $\ell = m$.

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 interarrival 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. Unlike the Poisson process, increments of general renewal processes are not stationary (unless we change the distribution of Z_0 in a clever way, as we will see) nor independent. Some of the important results in renewal theory are asymptotic results.

These asymptotic results will, in particular, allow us to prove the Ergodic Theorem for Markov chains.

10.4 Strong Law and Central Limit Theorem of renewal theory

Theorem 78 (Strong Law of renewal theory) *Let X be a renewal process with mean interarrival time $\mu \in (0, \infty)$. 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, as soon as $X_t > 0$,

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

Now $\mathbb{P}(X_t \to \infty) = 1$, since $X_\infty \leq n \iff T_{n+1} = \infty$ which is absurd, since $T_{n+1} = Z_0 + \ldots + Z_n$ 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

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

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. □

Try to do this proof for convergence in probability. The nasty ε expressions are not very useful in this context, and the proof is very much harder. But we can now deduce a corresponding Weak Law of Renewal Theory, because almost sure convergence implies convergence in probability.

We also have a Central Limit Theorem:

Theorem 79 (Central Limit Theorem of Renewal Theory) *Let $X = (X_t)_{t\geq 0}$ be a renewal process whose interarrival times $(Y_n)_{n\geq 0}$ satisfy $0 < \sigma^2 = Var(Y_1) < \infty$ and $\mu = \mathbb{E}(Y_1)$. Then*

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

The proof is not difficult and left as an exercise on Assignment 5.

10.5 The elementary renewal theorem

Theorem 80 *Let X be a renewal process with mean interarrival times μ and $m(t) = \mathbb{E}(X_t)$. Then*

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

Note that this does not follow easily from the strong law of renewal theory since almost sure convergence does not imply convergence of means (cf. Proposition 60, see also the counter example on Assignment 5). In fact, the proof is longer and not examinable: we start with a lemma.

Lemma 81 *For a renewal process X with arrival times $(T_n)_{n\geq 1}$, we have*

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

This ought to be true, because T_{X_t+1} is the sum of X_t+1 interarrival times, each with mean μ. Taking expectations, we should get $m(t)+1$ times μ. However, if we condition on X_t we have to know the distribution of the residual interarrival time after t, but without lack of memory, we are stuck.

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

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

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 76, $g_1(t)=m(t)$, i.e. $g(t)=\mu(1+m(t))$ as required. □

Proof of Theorem 80: Clearly $t<\mathbb{E}(T_{X_t+1})=\mu(m(t)+1)$ gives the lower bound

$$\liminf_{t\to\infty}\frac{m(t)}{t}\ge\frac{1}{\mu}.$$

For the upper bound we use a truncation argument and introduce

$$\widetilde{Z}_j=Z_j\wedge a=\begin{cases} Z_j & \text{if } Z_j<a \\ a & \text{if } Z_j\ge a\end{cases}$$

with associated renewal process $\widetilde{X}$. $\widetilde{Z}_j\le Z_j$ for all $j\ge 0$ implies $\widetilde{X}_t\ge X_t$ for all $t\ge 0$, hence $\widetilde{m}(t)\ge m(t)$. Putting things together, we get from the lemma again

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

Therefore

$$\frac{m(t)}{t}\le\frac{1}{\widetilde{\mu}}+\frac{a-\widetilde{\mu}}{\widetilde{\mu}t}$$

so that

$$\limsup_{t\to\infty}\frac{m(t)}{t}\le\frac{1}{\widetilde{\mu}}$$

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

$$\limsup_{t\to\infty}\frac{m(t)}{t}\le\frac{1}{\mu}.$$

□

Note that truncation was necessary to get $\mathbb{E}(\widetilde{Z}_{\widetilde{X}_t})\le a$. It would have been enough if we had $\mathbb{E}(Z_{X_t})=\mathbb{E}(Z_1)=\mu$, but this is *not* true. Look at the Poisson process as an example. We know that the residual lifetime has already mean $\mu=1/\lambda$, but there is also the part of Z_{X_t} before time t. We will explore this in Lecture 11 when we discuss residual lifetimes in renewal theory.

Lecture 11

Excess life and stationarity

Reading: Grimmett-Stirzaker 10.3-10.4; Ross 7.7

So far, we have studied the behaviour of one-dimensional marginals X_t:

$$\frac{X_t}{t} \to \frac{1}{\mu}, \qquad \frac{\mathbb{E}(X_t)}{t} \to \frac{1}{\mu}, \qquad \frac{X_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \to \mathcal{N}(0,1).$$

For Poisson processes we also studied finite-dimensional marginals and described the joint distributions of

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

In this lecture, we will make some progress with such a programme for renewal processes.

11.1 The renewal property

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 in the strict sense, but for certain random $t = T$ we have

Proposition 82 *Let X be a renewal process, $T_i = \inf\{t \geq 0 : X_t = i\}$. Then $(X_r)_{r\leq T_i}$ and $(X_{T_i+s} - X_{T_i})_{s\geq 0}$ are independent and $(X_{T_i+s} - X_{T_i})_{s\geq 0}$ 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 T_i, cf. Exercise A.2.3(a). We identify the interarrival times $\widetilde{Z}_n = Z_{i+n}$, $n \geq 0$, independent of $Z_0, \ldots, Z_{i-1}$. Here, as for the special case of Poisson processes, we describe $\widetilde{X} = (X_{T_i+s} - X_{T_i})_{s\geq 0}$ rather than $(X_{T_i+s})_{s\geq 0}$ since the former is a renewal process (in the strict sense) whereas the latter starts at i. □

Here are two simple examples that show why this cannot be extended to fixed times $t = T$.

Example 83 If the interarrival times are constant, say $\mathbb{P}(Z_n = 3) = 1$, then $\tilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ 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 84 Suppose, the interarrival times can take three values, say $\mathbb{P}(Z_n = 1) = 0.7$, $\mathbb{P}(Z_n = 2) = 0.2$ and $\mathbb{P}(Z_n = 19) = 0.1$. Note

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

Let us investigate a potential renewal property at time $t = 2$. Denoting $\widetilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ with holding times $\widetilde{Z}_n$, $n \geq 0$, we have

(1) $X_2 = 2$ implies $X_1 = 1$, $Z_0 = Z_1 = 1$, and we get $\mathbb{P}(\widetilde{Z}_0 = j | X_2 = 2) = \mathbb{P}(Z_2 = j)$;

(2) $X_2 = 1$ and $X_1 = 1$ implies $Z_0 = 1$ and $Z_1 \geq 2$, we get $\mathbb{P}(\widetilde{Z}_0 = 1 | X_2 = 1, X_1 = 1) = \mathbb{P}(Z_1 = 2 | Z_1 \geq 2) = 0.2/0.3 = 0.\overline{6}$;

(3) $X_2 = 1$ and $X_1 = 0$ implies $Z_0 = 2$, we get $\mathbb{P}(\widetilde{Z}_0 = 1 | X_2 = 1, X_1 = 0) = 0.7$;

(4) $X_2 = 0$ implies $Z_0 = 19$, we get $\mathbb{P}(\widetilde{Z}_0 = 17 | X_2 = 0) = 1$.

From (1) and (4) we obtain that $\widetilde{Z}_0$ is not independent of X_2 and hence $\widetilde{X}$ depends on $(X_r)_{r\leq 2}$.

From (2) and (3) we obtain that $\widetilde{Z}_0$ is not conditionally independent of X_1 given $X_2 = 1$, so $\widetilde{X}$ depends on $(X_r)_{r\leq 2}$ even conditionally given $X_2 = 1$.

In general, the situation is more involved, but you can imagine that knowledge of the *age* $A_t = t - T_{X_t}$ at time t, gives you information about the *excess lifetime* (residual lifetime) $E_t = T_{X_t+1} - t$ which is the first arrival time $\widetilde{Z}_0$ of the post-t process. In particular, A_t and E_t are not independent, and also the distribution of E_t depends on t. Note that if the distribution of E_t did not depend on A_t, we would get (something very close to, not exactly, since A_t is random) the lack of memory property and expect that Z_j is exponential.

However, this is all about $\widetilde{Z}_0$, we have not met any problems for the following interarrival times. Therefore, we define "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 interpretation will be that Z_0 is just part of an inter-renewal time, but we define a more general concept within which we develop the idea of partial renewal times, which will show some effects that seem paradoxial at first sight.

Definition 85 Let $(Z_n)_{n\geq 1}$ be a family of independent and identically distributed interarrival times and Z_0 an independent interarrival time with possibly *different* distribution. Then the associated counting process X with interarrival times $(Z_n)_{n\geq 0}$ is called a *delayed renewal process*.

As a corollary to Proposition 82, we obtain.

Corollary 86 (Renewal property) *Proposition 82 also remains true for* delayed *renewal processes. In this case, the post-T_i process will be* undelayed*.*

Just as the Markov property holds at more general stopping times, the renewal property also holds at more general stopping times, provided that they take values in the set $\{T_n, n \geq 0\}$ of renewal times. Here is a *heuristic* argument that can be made precise: we apply Proposition 82 conditionally on $T = T_i$ and get conditional independence and the conditional distribution of the post-T process given $T = T_i$. We see that the conditional distribution of the post-T process does not depend on i and apply Exercise A.1.5 to deduce the unconditional distribution of the post-T process and unconditional independence.

In particular, the renewal property does not apply to fixed $T = t$, but it does apply, e.g., to the first arrival (renewal) after time t, which is T_{X_t+1}. Actually, this special case of the renewal property can be proved directly by conditioning on X_t.

Proposition 87 *Given a (possibly delayed) renewal process X, for every $t \geq 0$, the post-t process $\widetilde{Z} = (X_{t+s} - X_t)_{s\geq 0}$ is a delayed renewal process with $\widetilde{Z}_0 = E_t$.*

Remark 88 Note that we make no statement about the dependence on the pre-t process.

Proof: We apply the renewal property to the renewal time T_{X_t+1}, the first renewal time after t. This establishes that $(\widetilde{Z}_n)_{n\geq 1}$ are independent identically distributed interarrival times, independent from the past, in particular from $\widetilde{Z}_0 = E_t$. $\widetilde{X}$ therefore has the structure of a delayed renewal process. □

11.2 Size-biased picks and stationarity

An important step towards further limit results will be to choose a good initial delay distribution. We cannot achieve independence of increments, but we will show that we can achieve stationarity of increments.

Proposition 89 *Let X be a delayed renewal process. If the distribution of the excesses E_t does not depend on $t \geq 0$, then X has stationary increments.*

Proof: By Proposition 87, 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 does not depend on t, then the distribution of $\widetilde{X}$ does not depend on t. In particular, the distribution of $\widetilde{X}_s = X_{t+s} - X_t$, an increment of X, does not depend on t. □

In fact,the converse is true, as well. Also, it can be shown that E is a Markov process on the uncountable state space $\mathbb{S} = [0, \infty)$, and we are looking for its invariant distribution.

Since we have not identified the invariant distribution, we should look at E_t for large t for a non-delayed renewal process since equilibrium (for Markov chains at least) is established asymptotically. We may either appeal to the Convergence Theorem and look at the distribution of E_t, say $\mathbb{P}(E_t > x)$ for large t, or we may appeal to the Ergodic Theorem and look at the proportion of $s \in [0, t]$ for which $E_s > x$.

Example 90 Let us continue the discussion of Example 84 and look at E_t for large t. First we notice that the fractional parts $\{E_t\}$ have a deterministic sawtooth behaviour forever since Z_n, $n \geq 0$, only take integer values. So, look at E_n for large n. 1. What is the probability that n falls into a long interarrival time? Or 2. What proportion of ns fall into a long interarrival time? 10%?

In fact, E_n is a discrete-time Markov chain on $\{1, 2, \ldots, 19\}$ with transition probabilities

$$\pi_{i+1,i} = 1, \quad i \geq 1, \qquad \pi_{1,j} = \mathbb{P}(Z_1 = j),$$

which is clearly irreducible and positive recurrent if $\mathbb{E}_1(H_1) = \mathbb{E}(Z_1) < \infty$, so it has a unique stationary distribution. This stationary distribution can be calculated in a general setting of $\mathbb{N}$-valued interarrival times, see Assignment 6. Let us here work on the intuition: The answer to the first question is essentially the probability of being in a state $3, 4, \ldots, 19$ under the stationary distribution (Convergence Theorem for $\mathbb{P}(E_n = j)$), whereas the answer to the second question is essentially the probability of being in a state $3, 4, \ldots, 19$ under the stationary distribution (Ergodic Theorem for $n^{-1}\#\{k = 0, \ldots, n-1 : E_k = j\}$).

In a typical period of 30 time units, we will have 10 arrivals, of which two will be separated by one long interarrival time. More precisely, on average 11 ns out of 30 will fall into small interarrival times, and 19 will fall into the long interarrival time. This is called a size-biased pick since 19 is the size of the long interarrival time. 1 and 2 are the sizes of the short interarrival times, this is weighted with their seven resp. two times higher proportion of occurence.

Furthermore, of these 19 that fall into the long interarrival time, one each has an excess of k for $k = 1, \ldots, 19$, i.e. we may expect the split of the current interarrival time into age and excess to be uniformly distributed over the interarrival time.

Definition 91 With a probability mass function p on $\mathbb{N}$ with $\mu = \sum_{n \geq 0} np_n < \infty$ we associate the size-biased pick distribution p^{sb} as

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

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

$$f_{sb}(z) = \frac{zf(z)}{\mu}.$$

Example 90 is for an $\mathbb{N}$-valued interarrival distribution, where the renewal process $(X_t)_{t\geq 0}$ will only see renewals at integer times, so it makes sense to consider the discretised process $(X_n)_{n\geq 0}$ as well as a discretised process $(E_n)_{n\geq 0}$ of excesses at integer times. The story with continuous interarrival times is completely analogous but does not tie in with the theory of discrete-time Markov chains, rather with $[0,\infty)$-valued Markov processes. We can still calculate

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

using the Strong Law of Large Numbers on $(Z_j)_{j\geq 0}$ and $Y_j=(Z_j-y)1_{\{Z_j>y\}}$, $j\geq 0$, to obtain

$$\frac{k}{\sum_{j=0}^{k-1}Z_j}\frac{\sum_{j=0}^{k-1}Y_j}{k}\to\frac{\mathbb{E}(Y_1)}{\mathbb{E}(Z_1)}=\frac{1}{\mathbb{E}(Z_1)}\int_y^\infty(z-y)f(z)dz$$

as (survival function $\mathbb{P}(\widetilde{Z}_0>y)$ of the) proposed stationary distribution for $(E_t)_{t\geq 0}$.

It is not hard to show that if L has the size-biased pick distribution and U is uniform on $[0,1]$ then LU has density

$$f_0(y)=\frac{1}{\mu}\bar{F}(y).$$

Just check

$$\begin{aligned}\mathbb{P}(LU>x) &= \int_0^1\mathbb{P}(L>x/u|U=u)du=\int_0^1\int_{x/u}^\infty\frac{z}{\mu}f(z)dzdu=\int_x^\infty\int_{x/z}^1 du\frac{z}{\mu}f(z)dz\\ &= \frac{1}{\mu}\int_x^\infty(z-x)f(z)dz\end{aligned}$$

and differentiate to get

$$f_0(x)=-\frac{d}{dx}\mathbb{P}(LU>x)=\frac{1}{\mu}xf(x)+\frac{1}{\mu}\overline{F}(x)-\frac{1}{\mu}xf(x)=\frac{1}{\mu}\overline{F}(x).$$

If there is uniqueness of stationary distribution, Ergodic Theorem, etc. for the Markov process $(E_t)_{t\geq 0}$, we obtain the following result.

Proposition 92 *Let X be a delayed renewal process where $\mathbb{P}(Z_n>y)=\bar{F}(y)$, $n\geq 1$, and $\mathbb{P}(Z_0>y)=\bar{F}_0(y)$. Then X has stationary increments, and $E_t\sim LU$ where $L\sim F_{sb}$ and $U\sim U(0,1)$ independent, for all $t\geq 0$.*

Proof: A proof within the framework of this course is on Assignment 6. □

Example 93 Clearly, for the $Exp(\lambda)$ distribution as inter-arrival distribution, we get $f_0(x)=\lambda e^{-\lambda x}$ also. Note, however, that

$$f^{sb}(x)=\frac{xf(x)}{\mu}=\lambda^2xe^{-\lambda x},\qquad x\geq 0,$$

is the probability density function of a $Gamma(2,\lambda)$ distribution, the distribution of the sum of two independent $Exp(\lambda)$ random variables.

Lecture 12

Convergence to equilibrium – renewal theorems

Reading: Grimmett-Stirzaker 10.4

Equilibrium for renewal processes X (increasing processes!) cannot be understood in the same way as for Markov chains. We rather want increment processes $X_{t+s} - X_t$, $t \geq 0$, to form a stationary processes for all fixed s. Note that we look at $X_{t+s} - X_t$ as a process in t, so we are looking at all increments over a fixed interval length. This is not the post-t process that takes $X_{t+s} - X_t$ as a process in s.

12.1 Convergence to equilibrium

We will have to treat the discrete and continuous cases separately in most of the sequel. More precisely, periodicity is an issue for discrete interarrival times. The general notion needed is that of arithmetic and non-arithmetic distributions:

Definition 94 A nonnegative random variable Z (and its distribution) are called d-arithmetic if $\mathbb{P}(Z \in d\mathbb{N}) = 1$, and d is maximal with this property. If Z is not d-arithmetic for any $d > 0$, it is called non-arithmetic.

We think of Z as an interarrival time. All continuous distributions are non-arithmetic, but there are others, which will not be relevant for us. We will mostly focus on this case. Our second focus is the 1-arithmetic case, i.e. the integer-valued case, where additionally $\mathbb{P}(Z \in d\mathbb{N}) < 1$ for all $d \geq 2$. It is not difficult to formulate corresponding results for d-arithmetic interarrival distributions and also for non-arithmetic ones, once an appropriate definition of a size-biased distribution is found.

Fact 95 (Convergence in distribution) *Let X be a (possibly delayed) renewal process having interarrival times with finite mean μ.*

(a) *If the interarrival distribution is continuous, then*

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

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

Also $(A_t, E_t) \to (L(1-U), LU)$ *in distribution, as* $t \to \infty$ *where* L *has the size-biased density* f_{sb} *and* $U \sim Unif(0,1)$ *is independent of* L.

(b) *If the interarrival distribution is integer-valued and* 1*-arithmetic, then*

$$X_{n+s} - X_n \to \hat{X}_s \qquad \textit{in distribution, as } n \to \infty.$$

where $\hat{X}$ *is an associated delayed renewal process, not with continuous first arrival* LU*, but with a discrete version: let* L *have size-biased probability mass function* p_{sb} *and* $\hat{Z}_0$ *conditionally uniform on* $\{1, \ldots, L\}$*. Also,* $(A_n, E_n) \to (L - \hat{Z}_0, \hat{Z}_0)$ *in distribution, as* $n \to \infty$.

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

12.2 Renewal theorems

The renewal theorems are now extensions of Fact 95 to convergence of certain moments. The renewal theorem itself concerns means. It is a refinement of the Elementary Renewal Theorem to increments, i.e. a second-order result.

Fact 96 (Renewal theorem) *Let* X *be a (possibly delayed) renewal process having interarrival times with finite mean* μ *and renewal function* $m(t) = \mathbb{E}(X_t)$.

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

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

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

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

As a generalisation that is often useful in applications, we mention a special case of the key renewal theorem:

Fact 97 (Key renewal theorem) *Let* X *be a renewal process with continuous interarrival time distribution and* $m(t) = \mathbb{E}(X_t)$*. If* $g : [0,\infty) \to [0,\infty)$ *is integrable (over* $[0,\infty)$ *in the limiting sense) and non-increasing, then*

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

There are generalisations that allow bigger classes of g (directly Riemann-integrable functions) and put no restrictions on the interarrival distribution (at the cost of some limits through discrete lattices in the d-arithmetic case). Even an infinite mean can be shown to correspond to zero limits.

Note that for $g_h(x) = 1_{[0,h]}(x)$, 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 leads to the renewal theorem. The renewal theorem and the key renewal theorem should be thought of as results where time windows 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. E.g., moments of excess lifetimes fall into this category. Note the general scheme that e.g. $\{E_t \leq h\}$ is a quantity only depending on a time window of size h.

Vice versa, the key renewal theorem can be deduced from the renewal theorem by approximating g by step functions.

12.3 The coupling proofs

Suppose now that X is a renewal process with integer-valued interarrival 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 that is not itself stationary, but that is such that $(\widehat{X}_n)_{n\geq 0}$ has stationary increments and such that $\widehat{X}$ has also a 1-arithmetic first arrival time. This can be achieved by choosing L with the size-biased pick distribution p^{sb}, which is 1-arithmetic, and choosing $\widehat{Z}_0$ conditionally uniform on $\{1, \ldots, L\}$.

We want to couple these two independent processes X and $\widehat{X}$. Define $N = \inf\{n \geq 1 : T_n = \widehat{T}_n\}$, the first index of simultaneous arrivals. Note that we require not only that arrivals happen at the same time, but that the index is the same. We can show that $\mathbb{P}(N < \infty) = 1$ by invoking a result on centered random walks. In fact N is the first time, the random walk $S_n = T_n - \widehat{T}_n$ hits 0. S is not a simple random walk, but it can be shown that it is recurrent (hence hits 0 in finite time) since $\mathbb{E}(Z_1 - \widehat{Z}_1) = 0$.

Now the idea is similar to the one in the coupling proof of the convergence theorem for Markov chains. We construct a new process

$$\overline{Z}_n = \begin{cases} Z_n & \text{if } n < N \\ \widehat{Z}_n & \text{if } n \geq N \end{cases}$$

so that

$$\overline{X}_t = \begin{cases} X_t & \text{if } t < T_N \\ \widehat{X}_t & \text{if } t \geq T_N \end{cases}$$

Now $\overline{X} \sim X$ and we get, in fact

$$\begin{aligned}\mathbb{P}((\overline{X}_{t+s}-\overline{X}_t)_{s\geq 0}\in A) &= \mathbb{P}(T_N\leq t,(\widehat{X}_{t+s}-\widehat{X}_t)_{s\geq 0}\in A)+\mathbb{P}(T_N>t,(\overline{X}_{t+s}-\overline{X}_t)_{s\geq 0}\in A)\\ &\to \mathbb{P}((\widehat{X}_{t+s}-\widehat{X}_t)_{s\geq 0}\in A),\end{aligned}$$

for all suitable $A\subset\{f:[0,\infty)\to\mathbb{N}\}$, since $\mathbb{P}(T_N>t)\to 0$.

This shows convergence of the distribution of $(X_{t+s}-X_t)_{s\geq 0}$ to the distribution of $\tilde{X}$ in so-called total variation, which is actually stronger than convergence in distribution as we claim. In particular, we can now take sets $A=\{f:[0,\infty)\to\mathbb{N}:f(s)=k\}$ to conclude.

For the renewal theorem, we deduce from this stronger form of convergence, that the means of $X_{t+s}-X_t$ converge as $t\to\infty$.

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

There is also a proof using renewal equations.

12.4 Example

Let us investigate the asymptotics of $\mathbb{E}(E_t^r)$. We condition on the last arrival time before t and that it is the kth arrival

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

Let us write $h(y)=\mathbb{E}(((Z_1-y)^+)^r)$. This is a clearly a nonnegative nonincreasing function of y and can be seen to be integrable if and only if $\mathbb{E}(Z_1^{r+1})<\infty$ (see below). The Key Renewal Theorem gives

$$\mathbb{E}(E_t^r)\to\frac{1}{\mu}\int_0^\infty h(y)dy=\frac{\mathbb{E}(Z_1^{r+1})}{\mu(r+1)}$$

since

$$\mathbb{E}(((Z_1-x)^+)^r)=\int_x^\infty(z-x)^rf(z)dz=\int_0^\infty y^rf(y+x)dy$$

and hence

$$\begin{aligned}\int_0^\infty\mathbb{E}(((Z_1-x)^+)^r)dx &= \int_0^\infty y^r\int_0^\infty f(y+x)dxdy\\ &= \int_0^\infty y^r\bar{F}(y)dy=\frac{\mathbb{E}(Z_1^{r+1})}{r+1}.\end{aligned}$$

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 13

M/M/1 queues and queueing networks

Reading: Norris 5.2.1-5.2.6; Grimmett-Stirzaker 11.2, 11.7; Ross 6.6, 8.4

Consider a single-server queueing system in which customers arrive according to a Poisson process of rate λ and service times are independent $Exp(\mu)$. 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.4.2 and from there we recall that

- 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.

- ξ_n can be best obtained by solving the detailed balance equations. By Proposition 57, X is reversible in equilibrium.

- The embedded "jump chain" $(M_n)_{n\geq 0}$, $M_n = X_{T_n}$, has a different invariant distribution $\eta \neq \xi$ since the holding times are $Exp(\lambda+\mu)$ everywhere except in 0, where they are $Exp(\lambda)$, hence rather longer, so that X spends "more time" in 0 than M. Hence η puts higher weight on 0, again by the ergodic theorem, now in discrete time. Let us state more explicitly the two ergodic theorems. They assert that we can obtain the invariant distributions as almost sure limits as $n \to \infty$

  $$\frac{1}{T_n}\int_0^{T_n} 1_{\{X_s = i\}} ds = \frac{1}{T_n}\sum_{k=0}^{n-1} Z_k 1_{\{M_k = i\}} \quad \to \quad \xi_i$$

  $$\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{M_k = i\}} \quad \to \quad \eta_i,$$

  for all $i \geq 0$, actually in the first case more generally, as $t \to \infty$ where t replaces the special choice $t = T_n$. Note how the holding times change the proportions as weights in the sums, $T_n = Z_0 + \ldots + Z_{n-1}$ being just the sum of the weights.

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

13.1 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 $Exp(\lambda)$, 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 98 *V has invariant distribution 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 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 $G \sim Exp(\mu)$ if $k \geq 1$ and the distribution of $Y + G$, where $Y \sim Exp(\lambda)$ is a typical interarrival 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 99 *If X (and hence V) is in equilibrium, then the $D_{n+1} - D_n$ are independent $Exp(\lambda)$ distributed.*

Proof: Let us first study D_1. We can calculate its moment generating function by Proposition 7 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 $Exp(\lambda)$ 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 $geom(\lambda/\mu)$ 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 $N_1 \sim Poi(\lambda x)$ 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 $G \sim Exp(\mu)$. 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 $D_{n+1}-D_n \sim Exp(\lambda)$, 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 $Exp(\lambda)$ 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.

13.2 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 $Exp(\mu_1)$ and $Exp(\mu_2)$ 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. Denote by $X_t^{(1)}$ the length of the first queue at time t and by $X_t^{(2)}$ the length of the second queue at time t.

Proposition 100 *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. Standard reasoning completes the proof. □

Proposition 99 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 queue is in equilibrium. This can be achieved if $\lambda < \mu_1$.

Proposition 101 *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 Exercise A.4.3, $\rho_1 \geq 1$ would prevent equilibrium for $X^{(1)}$, and expected return times for X and $X^{(1)}$ then clearly satisfy $m_{(0,0)} \geq m_0^{(1)} = \infty$. If $\rho_1 < 1$ and $X^{(1)}$ is in equilibrium, then by Proposition 99, the arrival process for the second queue is a Poisson process at rate λ, and $\rho_2 \geq 1$ would prevent equilibrium for $X^{(2)}$. 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:

$$\begin{aligned} \xi_{(i,j+1)} q_{(i,j+1),(i+1,j+1)} + \xi_{(i+2,j)} q_{(i+2,j),(i+1,j+1)} + \xi_{(i+1,j+2)} q_{(i+1,j+2),(i+1,j+1)} & \\ + \xi_{(i+1,j+1)} q_{(i+1,j+1),(i+1,j+1)} &= 0 \end{aligned}$$

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 at least n customers will then have been served by server 2 also.

13.3 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 or 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, and 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, $\pi_{2,0} = 1$, $\pi_{0,0} = 1$.

Fact 102 *If service times are independent $Exp(\mu_k)$ 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 additional state 0, according to transition probabilities $\pi_{k,\ell}$, then the queue-lengths process $X = (X^{(1)}, \ldots, X^{(m)})$ is well-defined and a continuous-time Markov chain. Its transition rates can be given as*

$$q_{x,x+e_k} = \lambda_k, \qquad q_{x,x-e_k+e_\ell} = \mu_k \pi_{k\ell}, \qquad q_{x,x-e_k} = \mu_k \pi_{k0}$$

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, $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the kth unit vector.

Fact 103 *If $X = (X^{(1)}, \ldots, X^{(m)})$ models a closed migration network with irreducible migration chain, then the total number of customers $X_t^{(1)} + \ldots + X_t^{(2)}$ remains constant over time, and for any such constant r, say, X has a unique invariant distribution given by*

$$\xi_x = B_r \prod_{k=1}^{m} \eta_k^{x_k}, \qquad \text{for all } x \in \mathbb{N}^m \text{ such that } x_1 + \ldots + x_m = r,$$

where η is the invariant distribution of the continuous-time migration chain and B_r is a normalising constant.

Note that ξ has a product form, but the queue lengths at servers $k = 1, \ldots, m$ under the stationary distribution are *not* independent, since the admissible x-values are constrained by $x_1 + \ldots + x_m = r$.

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 how it can be applied/modified in queueing networks, with several servers etc. These were all continuous-time Markov chains. It was always the exponential distribution that described interarrival times as well as service times. In practice, this assumption is often unrealistic. If we keep exponential distributions for either interarrival 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 *M/G/1* queue a queue 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 *G/M/1* queue a queue with a *G*eneral interarrival 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 30), and also M/M/∞ (queues with an infinite number of servers) – this model is the same as the immigration-death model that we formulated at the end of Example 58.

14.1 M/G/1 queues

An M/G/1 queue has independent and identically distributed service times with any distributions on $(0,\infty)$, but independent $Exp(\lambda)$ interarrival 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 nasty residual service distribution. However, after departures, we have exponential residual interarrival distributions:

Proposition 104 *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 98). 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 interarrival time $Y \sim Exp(\lambda)$ and service time $G \sim Exp(\mu)$. 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 interrenewal 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 105 *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*

$$\sum_{k\in\mathbb{N}} \xi_k s^k = (1-\rho)(1-s)\frac{1}{1 - s/\mathbb{E}(e^{\lambda(s-1)G})}.$$

Proof: We define ξ via its probability generating function

$$\phi(s) = \sum_{k\in\mathbb{N}} \xi_k s^k := (1-\rho)(1-s)\frac{1}{1 - s/\mathbb{E}(e^{\lambda(s-1)G})}$$

and note that $\xi_0 = \phi(0) = 1 - \rho$. To identify ξ as solution of

$$\xi_j = \sum_{i=0}^{j+1} \xi_i d_{i,j}, \qquad j \geq 0,$$

we can check the corresponding equality of probability generating functions. The probability generating function of the left-hand side is $\phi(s)$. To calculate the probability generating function of the right-hand side, calculate first

$$\sum_{m\in\mathbb{N}} d_{k+1,k+m} s^m = \sum_{m\in\mathbb{N}} \mathbb{E}\left(\frac{(s\lambda G)^m}{m!} e^{-\lambda G}\right) = \mathbb{E}(e^{(s-1)\lambda G}).$$

and then we have to check that the following sum is equal to $\phi(s)$:

$$\begin{aligned}
\sum_{j\in\mathbb{N}} \sum_{i=0}^{j+1} \xi_i d_{i,j} s^j &= \sum_{j\in\mathbb{N}} \xi_0 d_{0,j} s^j + \sum_{k\in\mathbb{N}} \sum_{m\in\mathbb{N}} \xi_{k+1} d_{k+1,k+m} s^{k+m} \\
&= \mathbb{E}(e^{(s-1)\lambda G}) \left(\xi_0 + \sum_{k\in\mathbb{N}}^{\infty} \xi_{k+1} s^k\right) \\
&= \mathbb{E}(e^{(s-1)\lambda G}) s^{-1} \left(\phi(s) - (1-\rho)(1-s)\right),
\end{aligned}$$

but this follows using the definition of $\phi(s)$. This completes the proof since uniqueness follows from the irreducibility of V. □

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. Similarly, all customers in the system when a given customer leaves, have arrived during his waiting and service times.

Proposition 106 *If X is such that V is in equilibrium, then the waiting time of any customer has distribution given by*

$$\mathbb{E}(e^{\gamma W}) = \frac{(1-\rho)\gamma}{\lambda + \gamma - \lambda \mathbb{E}(e^{\gamma G})}.$$

Proof: Unfortunately, we have not established equilibrium of X at the arrival times of customers. Therefore, we have to 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

$$\begin{aligned}\mathbb{E}(s^N) &= \int_0^\infty \mathbb{E}(s^N|T=t)f_T(t)dt = \int_0^\infty e^{\lambda t(s-1)}f_T(t)dt \\ &= \mathbb{E}(e^{\lambda T(s-1)}) = \mathbb{E}(e^{\lambda(s-1)W})\mathbb{E}(e^{\lambda(s-1)G}).\end{aligned}$$

From Proposition 105 we take $\mathbb{E}(s^N)$, and putting $\gamma = \lambda(s-1)$, we deduce the formula required by rearrangement. □

Corollary 107 *In the special case of M/M/1, the distribution of W is given by*

$$\mathbb{P}(W=0) = 1-\rho \qquad \textit{and} \qquad \mathbb{P}(W>w) = \rho e^{-(\mu-\lambda)w}, \quad w \geq 0.$$

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 108 *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 interarrival time $Y = y$, a $Poi(\mu y)$ 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 109 *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 $G_j \sim Exp(\mu)$. 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.

$$\begin{aligned}\mathbb{P}(W > w) &= \sum_{n=0}^{\infty} \mathbb{P}(N = n)\mathbb{P}(W > w|N = n) \\ &= 0 + \sum_{n=1}^{\infty} q^n(1-q)\int_w^{\infty} \frac{\mu^n}{(n-1)!}x^{n-1}e^{-\mu x}dx \\ &= \int_w^{\infty} e^{-\mu x}q\mu(1-q)\sum_{n=1}^{\infty}\frac{\mu^{n-1}}{(n-1)!}q^{n-1}x^{n-1}dx \\ &= q\int_w^{\infty} \mu(1-q)\exp\{-\mu x + \mu q x\}dx = q\exp\{-\mu(1-q)y\},\end{aligned}$$

where we used that the sum of n independent identically exponentially distributed random variables is Gamma distributed.

Lecture 15

Markov models for insurance

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

15.1 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 $(X_t)_{t\geq 0}$ with rate λ.
- Claim amounts $(A_j)_{j\geq 1}$ are positive, independent of the arrival process and identically distributed with common probability density function $k(a)$, $a > 0$, and mean $\mu = \mathbb{E}(A_1)$.
- The insurance company provides an initial reserve of $u \geq 0$ money units.
- Premiums are paid continuously at constant rate c generating a linear premium income accumulating to ct at time t. We 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 $C_t = \displaystyle\sum_{n=1}^{X_t} A_n$, $t \geq 0$.
- The reserve process $R_t = u + ct - C_t$, $t \geq 0$.
- The ruin probability $\psi(u) = \mathbb{P}_u(R_t < 0 \text{ for some } t \geq 0)$, as a function of $R_0 = u \geq 0$.

15.2 Aggregate claims and reserve processes

Proposition 110 *C and R have stationary independent increments. Their moment generating functions are given by*

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

and

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

Proof: First calculate the moment generating function of C_t:

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

which in the case where A_1 has a density k, gives the formula required. The same calculation for the joint moment generating function of C_t and $C_{t+s} - C_t$, or 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 with $\beta = -\gamma$. □

The moment generating function is useful to calculate moments.

Example 111 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 $\mathbb{E}(R_t) = u + ct - \lambda t\mu = u + (c - \lambda\mu)t$. Note that the Strong Law of Large Numbers, applied to increments $Z_n = R_n - R_{n-1}$ yields

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

confirming our claim that $c > \lambda\mu$ means that, on average, there is more premium income than claim outgo. In particular, this implies $R_n \to \infty$ a.s. as $n \to \infty$. Does this imply that $\inf\{R_s : 0 \leq s < \infty\} > -\infty$? No. It is conceivable that between integers, the reserve process takes much smaller values. But we will show that $\inf\{R_s : 0 \leq s < \infty\} > -\infty$.

There are other random walks that are embedded in the reserve process:

Example 112 Consider the process at claim times $W_n = R_{T_n}$, $n \geq 0$, where $(T_n)_{n\geq 0}$ are the event times of the Poisson process (and $T_0 = 0$). Now

$$W_{n+1} - W_n = R_{T_{n+1}} - R_{T_n} = c(T_{n+1} - T_n) - A_{n+1}, \qquad n \geq 0,$$

are also independent identically distributed increments with $\mathbb{E}(W_{n+1} - W_n) = c/\lambda - \mu > 0$, and the Strong Law of Large Numbers yields

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

Again, we conclude $W_n \to \infty$, but note that W_n are the local minima of R, so

$$I_\infty := \inf\{R_t : 0 \leq t < \infty\} = \inf\{W_n, n \geq 0\} > -\infty.$$

As a consequence, if we denote $R_t^0 = ct - C_t$ with associated I_∞^0, then

$$\psi(u) = \mathbb{P}_u(R_t < 0 \text{ for some } t \geq 0) = \mathbb{P}(I_\infty^0 \leq -u) \to \mathbb{P}(I_\infty^0 = -\infty) = 0$$

as $u \to \infty$, but this is then $\psi(\infty) = 0$.

15.3 Ruin probabilities

We now turn to studying the ruin probabilities $\psi(u)$, $u \geq 0$.

Proposition 113 *The ruin probabilities* $\psi(u)$ *satisfy the renewal equation*

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

where

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

Proof: Condition on $T_1 \sim Exp(\lambda)$ and $A_1 \sim k(a)$ to obtain

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

where we use the convention that $\psi(x) = 1$ for $x < 0$.

Differentiation w.r.t. x yields

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

Note that we also have a terminal condition $\psi(\infty) = 0$. With this terminal condition, this integro-differential equation has a unique solution. It therefore suffices to check that any solution of the renewal equation also solves the integro-differential equation.

For the renewal equation, we only sketch the argument since the technical details would distract from the main steps: note that differentiation (we skip the details for differentiation under the integral sign!) yields (setting $s = x - y$ in the convolution integral)

$$\begin{aligned}\psi'(x) &= g'(x) + \psi(x)f(0) + \int_0^x \psi(s)f'(x-s)ds \\ &= -\frac{\lambda}{c}\overline{K}(x) + \frac{\lambda}{c}\psi(x) - \frac{\lambda}{c}\int_0^x \psi(x-a)k(a)da,\end{aligned}$$

and note also that, with the convention $\psi(x) = 1$ for $x < 0$, we can write the renewal equation as

$$\psi(x) = \int_0^\infty \psi(x-y)f(y)dy,$$

where f is a nonnegative function with $\int_0^\infty f(y)dy = \lambda/c < 1$, so for any nonnegative solution $\psi \geq 0$, $\psi(x)$ is less than an average of ψ on $(-\infty, x]$, and hence ψ is decreasing (this requires a bit more care), so $\psi(\infty)$ exists with (by monotone convergence using $\psi(x-y) \downarrow \psi(\infty)$ as $x \to \infty$)

$$\psi(\infty) = \lim_{x\to\infty}\int_0^\infty \psi(x-y)f(y)dy = \int_0^\infty \psi(\infty)f(y)dy = \frac{\lambda}{c}\psi(\infty) \quad \Rightarrow \quad \psi(\infty) = 0.$$

□

Example 114 We can calculate $\psi(0) = g(0) = \lambda\mu/c$. In particular, zero initial reserve does not entail ruin with probability 1. In other words, ψ jumps at $u = 0$ from $\psi(0-) = 1$ to $\psi(0) = \psi(0+) = \lambda\mu/c < 1$.

Corollary 115 *If $\lambda\mu \leq c$, then ψ is given by*

$$\psi(x) = g(x) + \int_0^x g(x-y)u(y)dy \qquad \textit{where} \quad u(y) = \sum_{n\geq 1} f^{*(n)}(y).$$

Proof: This is an application of Exercise A.6.2(c), the general solution of the renewal equation. Note that f is not a probability density for $\lambda\mu < c$, but the results (and arguments) are valid for nonnegative f with $\int_0^\infty f(y)dy \leq 1$. □

Where is the renewal process? For $\lambda\mu < c$, there is no renewal process with interarrival density f in the strict sense, since f is not a probability density function. One can associate a defective renewal process that only counts a geometric number of points, and the best way to motivate this is by looking at $\lambda\mu = c$, where the situation is nicer. It can be shown that the renewal process is counting new minima of R or R^0, not in the time parameterisation of R^0, but in the height variable, i.e.

$$Y_h = \#\{n \geq 1 : W_n \in [-h, 0] \text{ and } W_n = \min\{W_0, \ldots, W_n\}\}, \qquad h \geq 0,$$

is a renewal process. Note that f is the distribution of LU where L is a size-biased claim and $U \sim Unif(0,1)$. 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.

So what happens if $\lambda\mu < c$? There will only be a finite number of claims that exceed the previous minimal reserve level since now $R_t \to \infty$, and Y remains constant for any lower levels of h.

This is not very explicit. To conclude, let us derive more explicit estimates of ψ.

Proposition 116 *Assume that there is $\alpha > 0$ such that*

$$1 = \int_0^\infty e^{\alpha y} f(y) dy = \frac{\lambda}{c} \int_0^\infty e^{\alpha y} \bar{K}(y) dy.$$

Then there is a constant $C > 0$ such that

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

Proof: Define a probability density function $\hat{f}(y) = e^{\alpha y} f(y)$, and $\hat{g}(y) = e^{\alpha y} g(y)$ and $\hat{\psi}(x) = e^{\alpha y}\psi(x)$. 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 115) converges by the key renewal theorem:

$$\hat{\psi}(x) = \hat{g}(x) + \int_0^x \hat{g}(x - y)\hat{u}(y)dy \to \frac{1}{\hat{\mu}} \int_0^\infty \hat{g}(y)dy =: C \qquad \text{as } x \to \infty,$$

where

$$\hat{u}(x) = \sum_{n \geq 1} \hat{f}^{*(n)}(x).$$

Note that $\hat{g}$ is not necessarily non-increasing, but it can be checked that it is integrable, and a version of the key renewal theorem still applies. □

Example 117 If $A_n \sim Exp(1/\mu)$, then in the notation of Proposition 113

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

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

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

15.4 Some simple finite-state-space models

Example 118 (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 119 (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 120 (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 121 (No claims discount) In automobile and some other 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.

Lecture 16

Conclusions and Perspectives

16.1 Summary of the course

This course was 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 can be added. Important is the real structure, the real transition mechanism that we wish to model by X. Memory plays an important role. We distinguish

- 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)
- Independence (of individuals in population models, of counts over disjoint time intervals, etc.)

Once we are happy that such conditions are met, we have a model X for the real process, and we study it under our 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, recurrence times, waiting times, ruin probabilities

- odd behaviour (explosion, transience, arithmetic interarrival times)

Techniques

- Conditioning, one-step analysis
- detailed balance equations
- algebra of limits for almost sure convergence

16.2 Duration-dependent transition rates

Renewal processes can be thought of as duration-dependent transition rates. If the interarrival distribution is not exponential, then (at least some) residual distributions will not be the same as the full interarrival 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

$$\lambda(t)=\frac{f(t)}{\mathbb{P}(Z>t)}=\frac{-\overline{F}'(t)}{\overline{F}(t)},$$

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 $\lambda_i(d)=\sum_{j\neq i}q_{ij}(d)$.

16.3 Time-dependent transition rates

A different type of varying transition intensities is obtained if we make the rates $\lambda(t)$ dependent on global time t. Here, the time passed in a given state is irrelevant, but only the actual time matters. This is useful to model seasonal effects. E.g. the intensity of road accidents may be considered higher in winter than in summer. So, a Poisson process to model this could have intensity $\lambda(t)=\lambda_0+\lambda_1\cos(2\pi t)$. This can also be built into birth- and death-rates of population models, insurance models etc.

16.4 Spatial 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.

$$N([0,t] \times [0,z]) = X(t,z) \sim Poi\left(\int_0^t \int_0^z \lambda(s,y)dyds\right)$$

and such that counts in disjoint rectangles are independent Poisson variables.

16.5 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 B10b, in the context of Stochastic Differential Equations in a Part C course, and in a Genetics context in another Part C course.

16.6 Lévy processes

A particularly nice class of processes are processes with stationary independent increments, so-called Lévy processes. If you have learned or are about to learn about Brownian motion in another course (B10b), then you know most Lévy processes since a general Lévy process can be written as $X_t = \mu t + \sigma B_t + C_t$ where μ is a drift coefficient, σ a scale parameter for the Brownian motion process B, and C is a limit of compound Poisson processes like the claim size process above. In fact, C may have infinitely many jumps in a finite interval, that are summable in some sense, but not necessarily absolutely summable, but these jumps can bedescribed by a family of independent Poisson processes with associated independent jump heights, in fact a Poisson measure on $[0,\infty) \times \mathbb{R}^*$ with intensity function ...

There is a Part C course on Lévy processes in Finance.

16.7 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.

16.8 4th year dissertation projects

Other than this, if you wish to study any of the stochastic processes or the applications more deeply in your fourth year, there are several people in the Statistics Department and the Mathematical Institute who would be willing to supervise you. Think about this, maybe over the Christmas vacations since the Easter vacations are close to the exams. Dissertation projects for Maths&Stats students are arranged before the summer to ensure that every student obtains a project well in advance, you can start working on your project during the summer. For Maths students dissertation projects are optional. Please get in touch with me or other potential supervisors in Hilary term, if you wish to discuss your possibilities. I am also happy to direct you to other people.