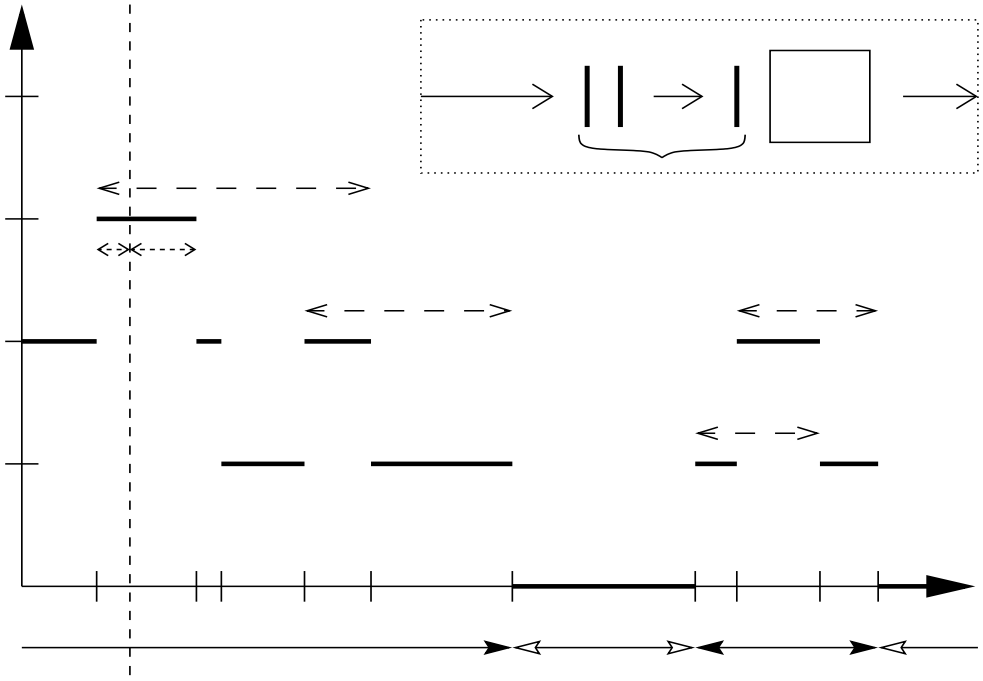
Part B
Applied Probability

Matthias Winkel[1]
University of Oxford



MT 2004

[1]Departmental lecturer (Institute of Actuaries and Aon Lecturer in Statistics) at the Department of Statistics, University of Oxford

Part B
Applied Probability

16 lectures MT 2004

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

1 Introduction: Poisson processes, generalisations, applications **1**
- 1.1 Poisson processes . . . 1
- 1.2 The Markov property . . . 3
- 1.3 Brief summary of the course . . . 4

2 Stochastic modelling and conditioning **5**
- 2.1 Modelling of events . . . 5
- 2.2 Conditional probabilities, densities and expectations . . . 7
- 2.3 Independence and conditional independence . . . 9

3 Birth processes and explosion **11**
- 3.1 Definition and an example . . . 11
- 3.2 Tonelli's Theorem, monotone and dominated convergence . . . 13
- 3.3 Explosion . . . 14

4 The Markov property **15**
- 4.1 The Markov property for birth processes . . . 15
- 4.2 The strong Markov property . . . 16
- 4.3 General continuous-time Markov chains . . . 17

5 Transition probabilities of continuous-time Markov chains **19**
- 5.1 The Markov property . . . 19
- 5.2 M/M/1 and M/M/s queues . . . 21
- 5.3 Transition semigroups . . . 22

6 Forward and backward equations **23**
- 6.1 Backward equations . . . 23
- 6.2 Forward equations . . . 24
- 6.3 Examples . . . 25
- 6.4 Invariant distributions and equilibrium . . . 26

7 Class properties of continuous-time Markov chains **27**
- 7.1 Class structure . . . 27
- 7.2 Recurrence and transience . . . 28
- 7.3 Positive and null recurrence . . . 30

8 Equilibrium and time reversal **31**
8.1 Recurrence and positive recurrence revisited 31
8.2 Convergence to equilibrium 32
8.3 Time reversal 33
8.4 Erlang's formula 34

9 The Strong Law of Large Numbers **35**
9.1 Motivation 35
9.2 Modes of convergence 36
9.3 Borel-Cantelli lemmas 37
9.4 Strong law of large numbers 38

10 Three limit theorems for renewal processes **39**
10.1 The (classical) Strong Law of Large Numbers 39
10.2 Strong Law and Central Limit Theorem of renewal theory 40
10.3 The elementary renewal theorem 41

11 Excess life and stationarity **43**
11.1 The renewal property 43

12 Convergence to equilibrium – renewal theorems **45**
12.1 Size-biased picks and stationarity 45
12.2 Convergence to equilibrium 46
12.3 Renewal theorems 47
12.4 The coupling proofs 48

13 M/M/1 queues and queueing networks **49**
13.1 M/M/1 queues and the departure process 49
13.2 Tandem queues 51
13.3 Closed and open migration networks 52

14 M/G/1 and G/M/1 queues **53**
14.1 M/G/1 queues 53
14.2 Waiting times in M/G/1 queues 54
14.3 G/M/1 queues 55

15 Markov models for insurance **57**
15.1 Insurance ruin 57
15.2 Some simple finite-state-space models 60

16 Conclusions and perspectives **61**
16.1 Summary of the course 61
16.2 The exam 62
16.3 Duration-dependent transition rates 62
16.4 Time-dependent transition rates 63
16.5 Spatial processes 63
16.6 Markov processes in uncountable state spaces ($\mathbb{R}$ or $\mathbb{R}^d$) 63
16.7 Lévy processes 64
16.8 Stationary processes 64
16.9 Internships, 4th year projects 64

A Assignments **I**
A.1 Revision, conditioning and Poisson processes III
A.2 Poisson processes and birth processes V
A.3 Continuous-time Markov chains VII
A.4 Continuous-time Markov chains II IX
A.5 Renewal theory I XI
A.6 Renewal theory II, Queues XIII

B Solutions **XV**
B.1 Revision, conditioning and Poisson processes XV
B.2 Poisson processes and birth processes XX
B.3 Continuous-time Markov chains XXV
B.4 Continuous-time Markov chains II XXIX
B.5 Renewal theory I XXXIV
B.6 Renewal theory II, Queues XXXIX

C Specimen and old exam questions **XLV**
C.1 Specimen exam questions XLV
C.2 Some old exam questions XLVII

Lecture 1

Introduction: Poisson processes, generalisations and applications

This course is, in the first place, a course for 3rd year undergraduates who did Part A Probability last year. Other students from various M.Sc.'s are welcome, but they should be aware of certain prerequisites of the course. These are essentially an introductory course in probability *not* based on measure theory, but also a good knowledge of the central aspects of discrete-time Markov chains is very useful.

The aim of today's lecture 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 on the Part A syllabus but I am aware that they were not treated very thoroughly. Therefore, I include the parts most relevant to us in today's lecture, and some more material is on the first assignment sheet.

I should also like to insist that 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 that you have not seen before, so effectively, we will be spending most 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 at most *countable* state space $\mathbb{S}$, which does not pose any complications as compared with $\mathbb{N}$. 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 $(Z_n)_{n\geq 1}$ be a sequence of independent exponential random variables $Z_n \sim Exp(\lambda)$ for a parameter (inverse mean) $\lambda \in (0,\infty)$, $T_0 = 0$, $T_n = \sum_{k=1}^n Z_k$, $n \geq 1$, the associated *random walk*. Then the process $X = (X_t)_{t\geq 0}$ defined by

$$X_t = \sum_{n=1}^{\infty} 1_{\{T_n \leq t\}} = \#\{n \geq 1 : T_n \leq t\}$$

is called *Poisson process with rate* λ. Here

$$\left[B_t^{(n)} := \right] 1_{\{T_n \leq t\}} := \begin{cases} 1 & \text{if } T_n \leq t \\ 0 & \text{if } T_n > t \end{cases}$$

can be considered as a Bernoulli random variable with success ($1_{\{T_n\leq t\}} = 1$) probability $\mathbb{P}(T_n \leq t)$.

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. I'll say something more on the link and the important distinction between real observations (cars in St. Giles) and mathematical models (Poisson process) 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 $(Z_n)_{n\geq 1}$ we have defined a more complex object $(X_t)_{t\geq 0}$, that we call Poisson process.

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

Remark 1 A Poisson process X with rate λ has the following properties

i) $X_t \sim Poi(\lambda t)$ for all $t \geq 0$.

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), just calculate

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

and it is easily checked that this is the probability generating function of the Poisson distribution with parameter λt. 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.

ii)-iii) can be derived from the following Proposition 1.

1.2 The Markov property

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), e.g. by

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

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 $\mathbb{P}(M_n = s, (M_k)_{k\leq n} \in A) > 0$, then

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

- $(M_k)_{0\leq k\leq n}$ 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 $\tilde{X}$ starting from initial states $\tilde{X}_0 = k \in \mathbb{N}$, by which we just mean that we consider $\tilde{X}_t = k + X_t$, $t \geq 0$, where X is a Poisson process starting from 0 as in the above definition.

Proposition 1 (Markov property) *Let X be a Poisson process with rate λ 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 $X_t = k$, $(X_r)_{r\leq t}$ and $(X_{t+s})_{s\geq 0}$ are independent, and the conditional distribution of $(X_{t+s})_{s\geq 0}$ is the same as the distribution of X given $X_0 = k$.

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

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

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 1ii) 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.

- The Poisson process is the prototype of a counting process. For the Poisson process, you can calculate "everything" explicitly. In practice, though, this is often only helpful as a first approximation. E.g. in insurance, the Poisson process is used for the arrival of claims. However, there is empirical evidence that inter-arrival times are neither exponential 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.

- Many applications that we discuss are in queueing theory. The easiest, so-called $M/M/1$ queue consists of a server and customers arriving according to a Poisson process. Independently of the arrival times, each customer has an exponential service time for which he will occupy the server, when it is his turn. 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 $s = 2, 3, \ldots, \infty$ 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

Stochastic modelling and conditioning

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 probabilities directly without much sophistication. Others 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 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 lenghts have the same distribution (cf. Remark 1 ii)-iii)), and times between two arrivals have a nonexponential distribution. Well, such a model does not exist. All properties of a model that were not specified in the model assumptions have to be derived from these. Otherwise you have to assume that your model may not have the property.

Suppose we are told that the 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 $\mathbb{P}$, so 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, 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, we cannot 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 events are 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) function on Ω. This existence can be proved for all our purposes, using measure theory.

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 $X_t(omega)$ by composing functions. Previously, you have mostly applied 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 themselves.

This course is not based on measure theory, but you should be aware that some of the proofs are only mathematically sound if based on measure theory. Ideally, this only means that we apply a result from measure theory that is intuitive enough for you to believe. In a few cases, however, the gap is more serious. Every effort will be made to point out technicalities, but without drawing too much of your attention away from the probabilistic arguments that I want you to appreciate and apply elsewhere.

Part B Martingales through measure theory or Section c Martingales and Stochastic Analysis provide pleasant introductions to measure theory. If you feel uneasy about lack of precision and mathematical formalism, then I recommend this course. I'm also open to discussions relating this course to the measure-theoretic framework after lectures.

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 1 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 1 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 2 Let X be a Poisson process. Then, for $t > s$,

$$f_{T_2|T_1=s}(t) = \frac{f_{T_1,T_2}(s,t)}{f_{T_1}(s)} = \frac{f_{Z_1,Z_2}(s,t-s)}{f_{Z_1}(s)} = f_{Z_2}(t-s),$$

by the transformation formula for bivariate densities to relate f_{T_1,T_2} to f_{Z_1,Z_2}, and independence of Z_1 and Z_2.

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 $T_2 = T_1 + Z_2 > T_1$. 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 $T_1 = s$, $T_2 = T_1 + Z_2$ ought to have the distribution of Z_2 shifted by s.

Example 3 Conditional probabilities and conditional densities are compatible in that

$$\mathbb{P}(S \in B|T = t) = \int_B f_{S|T=t}(s)ds = \lim_{\varepsilon \downarrow 0} \mathbb{P}(S \in B|t \leq T \leq t + \varepsilon)$$

provided only that f_T is right-continuous. To see this, write the right hand side as

$$\mathbb{P}(S \in B|t \leq T \leq t + \varepsilon) = \frac{\mathbb{P}(S \in B, t \leq T \leq t + \varepsilon)}{\mathbb{P}(t \leq T \leq t + \varepsilon)} = \frac{\frac{1}{\varepsilon}\int_t^{t+\varepsilon} \int_B f_{S,T}(s, u)dsdu}{\frac{1}{\varepsilon}\mathbb{P}(t \leq T \leq t + \varepsilon)}$$

and under the continuity condition (and by dominated convergence), 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

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

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 2 *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

$$\begin{aligned}
\mathbb{P}(X = j|T = t) &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}(X = j, t \leq T \leq t + \varepsilon)}{\mathbb{P}(t \leq T \leq t + \varepsilon)} \\
&= \lim_{\varepsilon \downarrow 0} \frac{\frac{1}{\varepsilon}\mathbb{P}(t \leq T \leq t + \varepsilon|X = j)\mathbb{P}(X = j)}{\frac{1}{\varepsilon}\mathbb{P}(t \leq T \leq t + \varepsilon)} \\
&= \frac{f_{T|X=j}(t)\mathbb{P}(X = j)}{f_T(t)}
\end{aligned}$$

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 $\{t-\varepsilon \leq T \leq t+\varepsilon\}$ (for $\varepsilon < t$, and normalisation by 2ε, as adequate).

Recall that we formulated the Markov property of the Poisson process as

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

for all events $\{(X_r)_{r\leq t} \in A\}$ such that $\mathbb{P}(X_t = k, (X_r)_{r\leq t} \in A) > 0$, and $\{(X_{t+u})_{u\geq 0} \in B\}$. For certain sets A with zero probability, this can still be established by approximation.

2.3 Independence and conditional independence

You have defined independence of two random variables 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

$$f_{S,T}(s,t) = f_S(s)f_T(t) \qquad \text{for all } s,t \in \mathbb{R}, \text{ where } f_S(s) = \int_{\mathbb{R}} f_{S,T}(s,t)dt.$$

You may also know (or can easily 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 2 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 $[0,\infty)$),

$$\mathbb{P}(X \in A, Y \in B | Z = z) = \mathbb{P}(X \in A | Z = z)\mathbb{P}(Y \in B | Z = z).$$

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 1 *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. $f(x, t) = 1_{\{x+t>s\}}$ 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 can be removed if X and T are independent. This can be shown by the approximation argument.

Lecture 3

Birth processes and explosion

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 3 *(i)* A stochastic process $X = (X_t)_{t\geq 0}$ of the form

$$X_t = X_0 + \#\left\{n \geq 1 : \sum_{j=1}^{n} Z_j \leq t\right\}$$

is called a *simple birth process of rates* $(\lambda_n)_{n\geq 0}$ starting from $k \in \mathbb{N}$, if $X_0 = k$ and the *inter-arrival times* Z_j, $j \geq 1$, are independent exponential random variables with parameters λ_{k+j-1}, $j \geq 1$.

(ii) A process of the form in (i) is called a birth process with initial distribution given by $p_k = \mathbb{P}(X_0 = k)$, $k \in \mathbb{N}$, if for all $k \in \mathbb{N}$ with $p_k > 0$, given $X_0 = k$, X is a simple birth process starting from $k \in \mathbb{N}$.

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.

Recall our notation $\mathbb{P}_k = \mathbb{P}(\cdot|X_0 = k)$. Rather than specifying conditional distributions (under $\mathbb{P}$), we can also say, under $\mathbb{P}_k$, the Z_j are independent exponential random variables with parameters λ_{k+j}, and have the starting point $X_0 = k$ fixed for the model $\mathbb{P}_k$.

"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.

We will not use the second part of this definition yet. However, it is vital to point out that, in general, the distribution of a Markov process (e.g. a Markov chain) is determined by an intial distribution and the transition mechanism. Here, we chose to describe the transition mechanism by the exponential holding times Z_n, $n \geq 1$, whose distributions depend on the initial state $k \in \mathbb{N}$.

Example 4 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. Thus 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 2c)

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

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

Put $\mu(t) = \mathbb{E}(X_t)$, then for $0 < s < t$, intuitively $\mathbb{E}(X_t|T=s) = 2\mu(t-s)$ since from time s, the two individuals perform independent birth processes of the same type and we are interested in their population sizes $t-s$ time units later. Formally, we go back to the structure of X in terms of arrival times $T_1 = T$, $T_n = Z_1 + \ldots + Z_n$ and calculate

$$\begin{aligned}
\mathbb{E}(X_t|T_1=s) &= \mathbb{E}\left(1 + \sum_{k=1}^{\infty} 1_{\{T_n \leq t\}}|T_1 = s\right) \\
&= 1 + \sum_{k=1}^{\infty} \mathbb{P}(T_n \leq t|T_1 = s) \\
&= 1 + \sum_{k=1}^{\infty} \mathbb{P}(T_n - T_1 \leq t-s|T_1 = s) \\
&= 1 + \sum_{k=1}^{\infty} \mathbb{P}(T_n - T_1 \leq t-s) \\
&= \mathbb{E}(X_{t-s}|X_0 = 2) = 2\mathbb{E}(X_{t-s}|X_0 = 1) = 2m(t-s).
\end{aligned}$$

The key step is where we replace $T_n \leq t$ by $T_n - T_1 \leq t - s$, which needs further justification. This is Fact 1, but can be checked directly by the approximation argument.

The whole argument could also be based on $(X_s)_{s\geq 0}$ rather than $(T_n)_{n\geq 0}$, but this would require the Markov property which we have not yet established.

Now

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

and setting $r = t - s$

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

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 2 (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 1 and Proposition 2 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 3 (Monotone convergence) *Integrals (expectations) of an increasing sequence of nonnegative functions (applied to a random variable) converge.*

Fact 4 (Dominated convergence) *Integrals (expectations) of a pointwise convergent sequence of functions $f_n \to f$ (applied to a random variable) converge, if the sequence $|f_n| \leq g$ 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 Part B 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.

3.3 Explosion

If the rates 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 $T_\infty = \lim_{n\to\infty} T_n$. For simple birth processes, we have the following necessary and sufficient condition.

Proposition 3 *Let X be a birth process with rates $(\lambda_n)_{n\geq 0}$. Then*

$$\mathbb{P}(T_\infty < \infty) > 0 \quad \textit{if and only if} \quad \sum_{n=0}^{\infty} \frac{1}{\lambda_n} < \infty.$$

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

Proof: Note that

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

where Tonelli's Theorem allows us to interchange summation and expectation. Therefore, if the series is finite, then $\mathbb{E}(T_\infty) < \infty$ implies $\mathbb{P}(T_\infty < \infty) = 1$. The other implication does not hold, in general. However, (using monotone convergence), and also the independence of the Z_n, we can calculate

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

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

The Markov property

4.1 The Markov property for birth processes

Proposition 4 (Markov property) *Let X be a simple birth process with rates $(\lambda_n)_{n\geq 0}$ starting from $\ell \geq 0$, and $t \geq 0$ a fixed time, $k \geq \ell$ a fixed height. 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 simple birth process with rates $(\lambda_n)_{n\geq 0}$ starting from k.*

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

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

$\tilde{Z}_j = Z_{k+j}$, $j \geq 2$, and $\tilde{Z}_1 = T_{k+1} - t$. 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. For $j = 1$ note that

$$\mathbb{P}(\tilde{Z}_1 > z | X_t = k) = \mathbb{P}(Z_{k+1} > (t - T_k) + z | Z_{k+1} > t - T_k \geq 0) = \mathbb{P}(Z_{k+1} > z)$$

where we applied the lack of memory property of Z_{k+1} to the *independent* threshold $t - T_k$. This actually requires a more thorough explanation since we are dealing with repeated conditioning but we just apply

Fact 5 *The lack of memory property of the exponential distribution holds at independent thresholds, i.e. given $Z \sim Exp(\lambda)$ and L a random variable independent of Z, then given $Z > L$, $Z - L \sim Exp(\lambda)$, and $Z - L$ is conditionally independent of Z given $Z > L$.*

The proof is now completed again by the lack of memory property to see that $\tilde{Z}_1$ (as well as the $\tilde{Z}_j$, $j \geq 2$) is conditionally independent of $Z_1, \ldots, Z_k$ given $X_t = k$ and the assertion follows. □

Proposition 1 for the special case of the Poisson process gave a second version of the Markov property, that we can deduce now. From Proposition 4 we know that the conditional distribution of $(X_{t+s} - X_t)_{s\geq 0}$ given $X_t = k$ is that of a Poisson process, and in particular does not depend on k, hence $(X_{t+s} - X_t)_{s\geq 0}$ is independent of X_t. Since we have shown that it is conditionally independent of $(X_r)_{r\leq t}$ we conclude unconditional independence.

Here we successively applied parts (a) and (b) of Exercise A.1.3, where the (b) part should read (M, N) independent of Z. 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 approved of, but this is beyond the scope of this course.

An alternative way to establish Proposition 1 is to give a direct proof of the stationarity and independence of increments in Proposition 1, based on the lack of memory property, and deduce the Markov property from there.

Neither approach is nice enough to be examinable in its mathematical rigour, but the general ideas used may be helpful.

4.2 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*. A more formal definition is as follows.

Definition 4 (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 written in terms of $(X_t)_{r\leq t}$.

Of course, you should think of X as a 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 5 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 $\{T \leq t\} = \{\exists s \leq t : X_s = i\} = \{X_t \geq i\}$. 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, provided we don't have to look ahead to decide.

Proposition 5 (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 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 $(\lambda_n)_{n\geq 0}$ given $X_0 = k$.*

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 independent of T and $(X_r)_{r\leq T}$.

Proof: 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. □

Note however, that the proof the strong Markov property for first hitting times T_i is easy since then $\mathbb{P}(X_{T_i} = i) = 1$, $(X_r)_{r\leq T_i}$ can be expressed in terms of Y_j, $j \leq i-1$, and $(X_{T+s} - X_T)_{s\geq 0}$ in terms of Y_j, $j \geq i$.

4.3 General continuous-time Markov chains

Simple birth processes are special in that they run through each state only once in a prescribed order. Such a behaviour would be rather boring (deterministic!) for a discrete-time Markov chain. However, if we combine the two, we get (almost) the most general (minimal) continuous-time Markov chain.

Definition 5 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

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

is called *(minimal) continuous-time Markov chain with jump probabilities $(\pi_{ij})_{i,j\in\mathbb{S}}$ and holding rates $(\lambda_i)_{i\in\mathbb{S}}$.*

Lecture 5

Transition probabilities of continuous Markov chains

5.1 The Markov property

Definition 5′ 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

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

is called *(minimal) continuous-time Markov chain with jump probabilities* $(\pi_{ij})_{i,j\in\mathbb{S}}$ *and holding rates* $(\lambda_i)_{i\in\mathbb{S}}$.

Clearly, 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 changing to $\lambda_i = 0$ if $\pi_{ii} = 1$.

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 $i \neq j$

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

X is then also referred to as a *continuous-time Markov chain with Q-matrix Q.*

It is also customary to specify an initial distribution for M_0 that we call ν, i.e. we choose $\nu_i = \mathbb{P}(M_0 = i)$, $i \in \mathbb{S}$. 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 an initial distribution consider (a distribution of) the number of customers arriving before a shop opens at time $t = 0$.

Proposition 6 *Let $(N_t^{ij})_{t\geq 0}$, $i,j \in \mathbb{S}$, $i \neq j$, be independent Poisson processes with rates q_{ij}. Let $M_0 \sim \nu$ and define 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 (minimal) continuous-time Markov chain with Q-matrix Q.

Think of the state space $\mathbb{S}$ as a maze where $q_{ij} > 0$ signifies that there is a gate between states $i, j \in \mathbb{S}$. Each gate opens at the times of a Poisson process. If after a given number of transitions the current state is i, then the next jump time is when the next gate leading away from i opens. If this gate leads from i to j, then j is the new state. Think of each Poisson process as a clock that rings at arrivals. A ringing clock here corresponds to a gate opening (and 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. $Y_0 = T_1 \sim Exp(\lambda_{M_0})$ 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 (Proposition 5) to see that the post-T_1 Poisson processes 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 5. □

Corollary 1 (Markov property) *Let X be a continuous-time Markov chain with jump chain M and holding rates $(\lambda_i)_{i\in\mathbb{S}}$ 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 the distibution of X given $X_0 = k$.*

Proof: The post-t Poisson processes are themselves Poisson processes, independent of the pre-t Poisson processes. The post-t behaviour of X only depends on X_t and the post-t Poisson processes. □

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.

5.2 M/M/1 and M/M/s queues

Example 6 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. This means that given a queue size of n, two transitions are possible. If a customer arrives (at rate λ), the 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

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

Example 7 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.

5.3 Transition semigroups

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

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

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

Example 8 For a Poisson process with rate λ, we have

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

by Proposition 1. For fixed $t \geq 0$ and $i \geq 0$, these are $Poi(\lambda t)$ probabilities, shifted by i.

Proposition 7 *$(P(t))_{t \geq 0}$ 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.*

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. □

Lecture 6

Forward and backward equations

6.1 Backward equations

The following result is useful to calculate transition probabilities.

Proposition 8 *$(P(t))_{t\geq 0}$ is the minimal solution of the so-called backward equation*

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

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

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

$$\mathbb{P}_i(X_t = k, t < T_0) = 0 \leq \tilde{p}_{ik}(t).$$

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

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

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. □

6.2 Forward equations

Fact 6 *The minimal non-negative solution $(P(t))_{t \geq 0}$ of the backward equation is also the minimal non-negative solution of the forward equation*

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

This will be proved under a uniformity assumption on the next exercise sheet. This assumption will be sufficient for most practical purposes.

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. These are the non-minimal solutions.

6.3 Examples

Example 9 For the Poisson process $q_{i,i+1} = \lambda$, $q_{ii} = -\lambda$, $q_{ij} = 0$ otherwise, hence we have forward equations

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

and it is easily seen that Poisson probabilities

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

are solutions. Alternatively, we may consider the backward equations

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

Writing $i+n$ rather than j is convenient because of the stationarity of increments in this special case of the Poisson process. We have seen an easier way to derive the Poisson transition probabilities in Proposition 1. The link between the two is by passing to probability generating functions

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

which then have to satisfy differential equations

$$G'_i(t) = \sum_{n=0}^{\infty} z^{i+n} p'_{i,i+n}(t) = z^i \lambda (z-1) G_i(t).$$

Solutions for these equations are obvious. In general, if we have G_i sufficiently smooth in t and z, we can derive from differential equations for probability generating functions differential equations for moments

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

that yield here

$$m'_i(t) = \frac{\partial}{\partial z} G'_i(t) \bigg|_{z=1-} = \lambda \, G_i(t)|_{z=1-} = \lambda.$$

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 capture the two most important distributional features.

6.4 Invariant distributions and equilibrium

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 6 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 9 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. □

If time permits, we will extend this result to infinite $\mathbb{S}$ later, under suitable assumptions, when we discuss convergence to equilibrium.

Lecture 7

Class properties of continuous-time Markov chains

7.1 Class structure

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

Definition 7 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 $\mathbb{P}_i(X_t = j \text{ for some } t \geq 0) > 0$.

(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. for all $i, j \in A$ we have $i \leftrightarrow j$ and A is maximal with this property, i.e. 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 10 *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 path $(i_0 \ldots, i_n)$, $i_j \in \mathbb{S}$, *from* $i_0 = i$ *to* $i_n = j$ *such that* $\prod_{j=1}^{n} q_{i_{j-1}, i_j} > 0$.

(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

$$\prod_{j=1}^{n} \pi_{i_{j-1},i_j} > 0, \quad \text{hence} \quad \prod_{j=1}^{n} \pi_{i_{j-1},i_j} \lambda_{i_{j-1}} > 0$$

since $\lambda_k = 0$ if and only if $\pi_{kk} = 1$.

(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)

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

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 8 Let X be a (minimal) 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

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

Note that if X can explode starting from i 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 11 *$i \in \mathbb{S}$ 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 $(N_i^{(m)})_{m \geq 1}$. Then 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 3 (explosion for birth processes) 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 $N < \infty$ away from i and

$$\{t \geq 0 : X_t = i\} \subset [0, T_N)$$

is bounded with probability 1.

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

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

Corollary 2 *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 3 *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. □

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.

Definition 9 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 7 *Positive recurrence is a class property.*

Lecture 8

Equilibrium and time reversal

As in discrete time, the state space of a general continuous-time Markov chain can be decomposed into 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. 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. To be precise, recurrence is not implied by irreducibility, so we will be concerned with irreducible recurrent Markov chains.

8.1 Recurrence and positive recurrence revisited

Proposition 12 *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) $\int_0^\infty p_{ii}(t)dt = \infty$.

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

$$\sum_{m=1}^{G} Z_{N_i^{(m-1)}} \sim Exp(h_i\lambda_i), \qquad \text{cf. Solution to Exercise A.2.4.}$$

Now, by Tonelli's theorem,

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

The other implications can be established using similar arguments, see Assignment 4. □

Fact 8 *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*

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

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.

Fact 9 *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*

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

We shall not provide the details here, but point out the general principles of proof. 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 direct argument, and a very nice argument, too, using the coupling method.

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 following the ν-chain before T and the ξ-chain afterwards. The following statements can be shown

1. the meeting time is finite with probability 1;
2. the third chain has the same distribution as the ν-chain;
3. the third chain (which eventually coincides with the ξ-chain) is asymptotically in equilibrium.

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.

Fact 10 (Ergodic theorem) *In the setting of Fact 9, $X_0 \sim \nu$*

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

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 Fact 9. 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 Fact 9, we would need many independent observations of the same system at a large time t to estimate ξ.

8.3 Time reversal

Proposition 13 *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, a measure theory "fact" 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* $\xi_i q_{ij} = \xi_j q_{ji}$, and as in discrete time, they are easier to handle than $\xi Q = 0$. In fact, whenever the detailed balance equations have a solution, this is an invariant distribution as is easily seen summing the equations over over j.

8.4 Erlang's formula

Example 10 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 $\mu_0 = 0$).

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

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

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

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

provided S is finite.

If S is infinite, then there does not exist an invariant distribution. This 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 8, 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. 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.

Lecture 9

The Strong Law of Large Numbers

9.1 Motivation

So far, we've been concerned with 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), (stationary) transition probabilities, invariant distributions, class structure, convergence to equilibrium, ergodic theorem, time reversal, detailed balance etc. A few odd features may occur (mainly due to explosion), but they seemed not too relevant for applications. An annoying feature is that transition probabilities often cannot be calculated explicitly.

All the good properties are due to the exponential holding times. In practice, this assumption is usually not met, and we may wish to replace the exponential distribution by other distributions, e.g. that cannot take arbitrarily large values or produce clustering effects. 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 inter-arrival times.

Definition 10 Let $(Y_n)_{n\geq 0}$ be a sequence of independent identically distributed positive random variables, $T_n = \sum_{k=0}^{n-1} Y_k$, the associated random walk. Then the process $X = (X_t)_{t\geq 0}$ defined by

$$X_t = \sum_{n=1}^{\infty} 1_{\{T_n \leq t\}} = \#\{n \geq 1 : T_n \leq t\}$$

is called a *renewal process*. The distribution of Y_1 is called inter-arrival distribution.

Example 11 If $(Z_t)_{t\geq 0}$ is a continuous-time Markov chain with $Z_0 = i$, then $Y_k = H_i^{(k+1)} - H_i^{(k)}$, the times between successive hitting of i by Z, are independent and identically distributed (by the strong Markov property). The associated counting process X counting the visits to i is thus a renewal process.

Remember, that for $Y_n \sim Exp(\lambda)$ we had $X_t \sim Poi(\lambda t)$ and in particular $m(t) = \mathbb{E}(X_t) = \lambda t$. For general renewal processes, we will see that $m(t)$ will no longer be linear, but only asymptotically linear (elementary renewal theorem). In general, we will not be able to understand the marginal distributions, let alone increment distributions, since increments will no longer be stationary (unless we change the distribution of Y_0 in a clever way). Here we will restrict our attention to asymptotic results that describe average behaviour and the deviations of X_t from $m(t)$ in a central limit theorem fashion. We will also look at the evolution of increment means (renewal theorem).

When talking about asymptotic mean behaviour, there are different ways to make precise statements. We will first discuss different modes of convergence, with an emphasis on almost sure convergence, as this concept is new to you. Then we will establish the strong law of large numbers. The following lectures will move more explicitly into renewal theory by studying the results indicated above.

9.2 Modes of convergence

Definition 11 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* if $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$.

Fact 11 *The following implications hold*

$$\begin{array}{l} X_n \to X \text{ almost surely} \\ \qquad \Downarrow \\ X_n \to X \text{ in probability} \quad \Rightarrow X_n \to X \text{ in distribution} \\ \qquad \Uparrow \\ X_n \to X \text{ in } L^1 \end{array}$$

No other implications hold in general.

We prove that almost sure convergence implies convergence in probability, as follows. Suppose, $X_n \to X$ almost surely, then for almost all $\omega \in \Omega$, i.e. all $\omega \in \Omega_0$ for some $\Omega_0 \subset \Omega$ with $\mathbb{P}(\Omega_0) = 1$,

$$X_n(\omega) \to X(\omega), \quad \text{i.e.} \quad \forall_{\varepsilon>0} \exists_{N^{(\varepsilon)}(\omega)<\infty} \forall_{n \geq N^{(\varepsilon)}(\omega)} |X_n(\omega) - X(\omega)| < \infty.$$

We define $N^{(\varepsilon)}(\omega) = \infty$ for all $\omega \in \Omega - \Omega_0$. Note that then $N^{(\varepsilon)} : \Omega \to \bar{\mathbb{N}}$ is a random variable that is finite on Ω_0, i.e. $\mathbb{P}(N^{(\varepsilon)} < \infty) = \mathbb{P}(\Omega_0) = 1$. Furthermore, clearly $|X_n(\omega) - X(\omega)| > \varepsilon$ implies $N^{(\varepsilon)}(\omega) > n$ so that

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

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

You will see that convergence in probability does not imply almost sure convergence, as an example on the next assignment sheet.

9.3 Borel-Cantelli lemmas

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. Call the (successive) events A_n, $n \geq 1$. Then

$$\omega \in A_n \text{ i.o.} \iff \forall_{n\geq 1}\exists_{m\geq n}\omega \in A_n \iff \omega \in \bigcap_{n\geq 1}\bigcup_{m\geq n} A_m =: \limsup_{n\to\infty} A_n.$$

Theorem 1 (Borel-Cantelli lemmas) *Let $A = \limsup_{n\to\infty} A_n$ be the event that infinitely many of the events A_n occur. Then*

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

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

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

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

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

The conclusion in the second statement 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,

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

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 σ-additivity of $\mathbb{P}$, cf. Grimmett-Stirzaker, Lemma 1.3.(5).

9.4 Strong law of large numbers

Theorem 2 *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*

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

Fact 12 *Theorem 2 remains valid without the assumption $\mathbb{E}(X_1^4) < \infty$, just assuming $\mathbb{E}(|X_1|) < \infty$.*

The quickest proof of this general result proceeds via a truncation, that we will see in a less technical argument when we prove the elementary renewal theorem.

Lecture 10

Three limit theorems for renewal processes

10.1 The (classical) Strong Law of Large Numbers

Theorem 2 *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*

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

Lemma 1 *In the situation of Theorem 2, 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 2: Denote

$$S_n = \sum_{i=1}^{n} X_i, \qquad n \geq 0.$$

We know by Markov's inequality that

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

If we define

$$A_n = \left\{ \frac{1}{n} |S_n - n\mu| \geq n^{-\gamma} \right\}$$

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

$$\sum_{n \geq 1} \mathbb{P}(A_n) < \infty \qquad \Rightarrow \qquad \mathbb{P}(A_n \text{ i.o.}) = 0$$

by (the easy part of) the Borel-Cantelli lemma. However

$$\Omega_0 = \{A_n \text{ i.o.}\}^c = \left\{ \exists_N \forall_{n \geq N} \left| \frac{S_n}{n} - \mu \right| < n^{-\gamma} \right\}$$

and for all $\omega \in \Omega_0$

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

□

10.2 Strong Law and Central Limit Theorem of renewal theory

Theorem 3 (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 $X_t \to \infty$ a.s., as $t \to \infty$, since $X_\infty \leq n \iff T_{n+1} = \infty$ which is a.s. impossible, i.e. happens with probability 0. 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, for all $\omega \in \{X_t \to \infty\} \cap \{T_n/n \to \mu\}$

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

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. We rather deduce a weak law for renewal processes from the strong law.

Theorem 4 (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*

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

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

10.3 The elementary renewal theorem

Theorem 5 *Let X be a renewal process with mean interarrival times μ and $m(t) = \mathbb{E}(X_t)$. Then*

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

Note that this does not follow easily from the strong law of renewal theory since almost sure convergence does not imply convergence in L^1 (cf. Fact 11), not even convergence of means, as an example on Assignment 5 shows. In fact, the proof is longer, and we start with a lemma, that is of independent interest. Recall that a random variable M is a stopping time with respect to the sequence $(T_n)_{n\geq 0}$ of random variables if, for all $m \geq 1$, the event $\{M \leq m\}$ is determined by $T_0, \ldots, T_m$.

Example 12 $M = X_t + 1$ is a stopping time for $(T_n)_{n\geq 0}$ since

$$\{M \leq m\} = \{X_t \leq m-1\} = \{T_m > t\}.$$

$\tilde{M} = X_t$ is not a stopping time for $(X_n)_{n\geq 0}$ since we need to know T_{m+1} to determine $\{\tilde{M} \leq m\}$.

Lemma 2 (Wald's equation) *Let $(T_n)_{n\geq 0}$ be the arrival times of a renewal process and $Z_n = T_n - T_{n-1}$ have finite mean, let M be a stopping time with respect to $(T_n)_{n\geq 0}$ satisfying $\mathbb{E}(M) < \infty$. Then*

$$\mathbb{E}\left(\sum_{k=1}^{M} Z_k\right) = \mathbb{E}\left(T_M\right) = \mathbb{E}(Z_1)\mathbb{E}(M)$$

Proof: The calculation is simple. Just note where we use the stopping time property that implies that Z_k is independent of $\{M \geq k\} = \{M \leq k-1\}^c$

$$\begin{aligned}\mathbb{E}\left(\sum_{k=1}^{M} Z_k\right) &= \mathbb{E}\left(\sum_{k=1}^{\infty} Z_k 1_{\{M\geq k\}}\right) = \sum_{k=1}^{\infty} \mathbb{E}\left(Z_k 1_{\{M\geq k\}}\right) \\ &= \sum_{k=1}^{\infty} \mathbb{E}(Z_k)\mathbb{P}(M \geq k) = \mathbb{E}(Z_1)\sum_{k=1}^{\infty}\mathbb{P}(M\geq k) = \mathbb{E}(Z_1)\mathbb{E}(M)\end{aligned}$$

where the formula for $\mathbb{E}(M)$is easily derived. Cf. Solution to Assignment A.3.2 for the proof of the corresponding formula $\mathbb{E}(T) = \int_0^\infty \mathbb{P}(T > t)dt$ for continuous random variables. □

Proof of Theorem 5: From the proof of the strong law we take $t < T_{X_t+1}$, hence by Wald's equation

$$t < \mathbb{E}(T_{X_t+1}) = \mathbb{E}(Y_1)\mathbb{E}(X_t+1) = \mu(m(t)+1) \qquad \Rightarrow \qquad \frac{m(t)}{t} > \frac{1}{\mu} - \frac{1}{t}$$

and therefore we get the lower bound

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

For the upper bound we use a truncation argument and introduce

$$\tilde{Y}_j = Y_j \wedge a = \begin{cases} Y_j & \text{if } Y_j < a \\ a & \text{if } Y_j \geq a \end{cases}$$

with associated renewal process $\tilde{X}$. $\tilde{Y}_j \leq Y_j$ for all $j \geq 0$ implies $\tilde{X}_t \geq X_t$ for all $t \geq 0$, hence $\tilde{m}(t) \geq m(t)$. Putting things together, we get from Wald's equation again

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

Therefore

$$\frac{m(t)}{t} \leq \frac{1}{\tilde{\mu}} + \frac{a-\tilde{\mu}}{\tilde{\mu}t}$$

so that

$$\limsup_{t\to\infty} \frac{m(t)}{t} \leq \frac{1}{\tilde{\mu}}$$

Now $\tilde{\mu} = \mathbb{E}(\tilde{Y}_1) = \mathbb{E}(Y_1 \wedge a) \to \mathbb{E}(Y_1) = \mu$ a $a \to \infty$ (by monotone convergence). Therefore

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

□

Note that truncation was necessary to get $\mathbb{E}(\tilde{Y}_{\tilde{X}_t+1}) \leq a$. It would have been enough if we had $\mathbb{E}(Y_{\tilde{X}_t+1}) = \mathbb{E}(Y_1) = \mu$, 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 $Y_{\tilde{X}_t+1}$ before time t. We will further explore this next week when we discuss residual lifetimes in renewal theory.

Lecture 11

Excess life and stationarity

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

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

The strength of Poisson processes was that we could study finite-dimensional marginals, i.e. describe the joint distributions of

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

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 14 *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.2.6. We identify the interarrival times. Here, as for the special case of Poisson processes, we describe $(X_{T_i+s} - X_{T_i})_{s \geq 0}$ rather than $(X_{T_i+s})_{s \geq 0}$ since the former has a distribution independent of i whereas the latter starts at i. □

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

Example 13 If the interarrival times are constant, say $\mathbb{P}(Y_i = 3) = 1$, then $\tilde{X} = (X_{t+s} - X_t)_{s \geq 0}$ has a first arrival time $\tilde{Y}_0$ with $\mathbb{P}(\tilde{Y}_0 = 3 - t) = 1$, for $0 \leq t < 3$.

The second example also shows that the independence from the pre-t process fails.

Example 14 Suppose, the interarrival times can take two values, say $\mathbb{P}(Y_i = 1) = 0.9$, $\mathbb{P}(Y_i = 21) = 0.1$. Note $\mathbb{E}(Y_i) = 3$. Then for $t = 2$ and $\tilde{X} = (X_{t+s} - X_t)_{s\geq 0}$, clearly,

- given $X_2 = 0$, we must have $\tilde{Y}_0 = 19$,
- given $X_2 = 2$, $\tilde{Y}_0$ has the same distribution as Y_0, because then $2 = T_2$,
- given $X_2 = 1$, the knowledge of X_1 is vital since
  (a) given also $X_1 = 1$, $\tilde{Y}_0 = 20$
  (b) given also $X_1 = 0$, $\tilde{Y}_0$ has the same distribution as Y_0, because $2 = T_1$.

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 of the post-t process. In particular, the two are not independent, and also the distribution of E_t depends on t.

Definition 12 Let $(Y_n)_{n\geq 1}$ be a family of independent and identically distributed interarrival times and Y_0 an independent interarrival time with possibly different distribution. Then the associated counting process X with interarrival times $(Y_n)_{n\geq 0}$ is called a *delayed renewal process.*

Fact 13 (Renewal property) *Proposition 14 remains true if T_i is replaced by a stopping times T, that satisfies $\mathbb{P}(T \in \{T_i : i \geq 1\}) = 1$.*

Proposition 14 also remains true for delayed renewal processes. In this case, the post-T process will be undelayed.

Here is a *heuristic* argument that can be made precise: we apply Proposition 14 conditional 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.1.3 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.

Proposition 15 *Given a (possibly delayed) renewal process X, for every $t \geq 0$, the post-t process $\tilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ is a delayed renewal process with $\tilde{Y}_0 = E_t$.*

Proof: We apply the renewal property, Fact 13, to the renewal time T_{X_t+1} which we have shown to be a stopping time. This establishes that $(\tilde{Y}_n)_{n\geq 1}$ are independent identically distributed interarrival times, independent from the past, in particular from $\tilde{Y}_0 = E_t$. $\tilde{X}$ therefore has the structure of a delayed renewal process. □

Lecture 12

Convergence to equilibrium – renewal theorems

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 be stationary processes for all 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 looks at $X_{t+s} - X_t$ as a process in s.

12.1 Size-biased picks and stationarity

An important trick for further limit results will be to choose a good initial delay distribution. It is clear that we cannot achieve independence of increments, but we will show that we can achieve stationarity of increments.

Proposition 16 *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: The distribution of the post-t process $\tilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ does not depend on t if the distribution of $E_t = \tilde{Y}_0$ does not. □

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

In any case, not knowing 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.

Example 14 (continued) Let us look at E_t for large t. First we notice that the fractional parts $\{E_t\}$ have a deterministic sawtooth behaviour forever since Y_i only take integer values. So, look at E_n for large n. What is the probability that n falls into a long interarrival time? Or, what proportion of ns fall into a long interarrival time? 10%?

In a typical period of 30 time units, we will have 10 arrivals, of which two will be separated by one long interarrival time. Therefore, 9 ns out of 30 will fall into small

interarrival times, and 21 will fall into the long interarrival time. This is called a size-biased pick since 21 is the size of the long interarrival time. 1 is the size of the short interarrival time, this is weighted with its nine times higher proportion of occurence.

Furthermore, of these 21 that fall into the long interarrival time, one each has an excess of k for $k = 1, \ldots, 21$, i.e. we expect the split of the current interarrival time into age and excess to be uniformly distributed over the interarrival time.

Definition 13 Given a lifetime (or interarrival) distribution, with survival function $\bar{F}$ and with finite mean μ, we define the associated *size-biased pick distribution* as

$$\bar{F}_{sb}(y) = \frac{1}{\mu}\int_y^\infty \bar{F}(z)dz + \frac{1}{\mu}y\bar{F}(y), \qquad y \geq 0.$$

If F has density f, a partial integration (or writing $\bar{F}$ as integrals and using Tonelli) shows that F^{sb} has density

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

This is what we understand intuitively by a size-biased pick. If F has probability mass function p_n, $n \in \mathbb{N}$, then, F^{sb} has probability mass function

$$p_n^{sb} = \frac{n}{\mu}p_n.$$

Also, 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

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

Fact 14 *Let X be a delayed renewal process where $\mathbb{P}(Y_n > y) = \bar{F}(y)$, $n \geq 1$, and $\mathbb{P}(Y_0 > y) = \bar{F}_0(y)$. Then X has stationary increments, and $E_t \sim LU$ where $L \sim F_{sb}$ and $U \sim U(0, 1)$, for all $t \geq 0$.*

12.2 Convergence to equilibrium

Definition 14 A nonnegative random variable Y and its distribution are called d-arithmetic if $\mathbb{P}(Y \in d\mathbb{N}) = 1$, and d is maximal with this property. If Y and its distribution is not d-arithmetic for any $d > 0$, it is called non-arithmetic.

We think of Y as an interarrival distribution. All continuous distributions are non-arithmetic. We will mostly focus on this case. Our second focus is the 1-arithmetic case. It is not difficult to formulate corresponding results for d-arithmetic interarrival distributions.

Fact 15 (Convergence in distribution) *Let X be a (possibly delayed) renewal process having interarrival times with finite mean μ.*

(a) If the interarrival distribution is non-arithmetic then

$$(X_{t+s} - X_t)_{s\geq 0} \to \tilde{X} \qquad \text{in distribution, as } t \to \infty,$$

where $\tilde{X}$ is the associated stationary renewal process. In particular,

$$X_{t+h} - X_t \to \tilde{X}_h \qquad \text{in distribution, as } t \to \infty.$$

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

(b) If the interarrival distribution is 1-arithmetic then

$$(X_{n+s} - X_n)_{s\geq 0} \to \hat{X} \qquad \text{in distribution, as } n \to \infty.$$

where $\hat{X}$ is an associated delayed renewal process, not with first arrival distribution F_0, but a discrete version: let L be according to F_{sb} and $\hat{Y}_0$ conditionally uniform on $\{1, \ldots, L\}$. Also, for all $\alpha \in [0,1)$, $(A_{n+\alpha}, E_{n+\alpha}) \to (\hat{A}_\alpha, \hat{E}_\alpha)$ in distribution, as $n \to \infty$.

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

12.3 Renewal theorems

The renewal theorems are now extensions of Fact 15 to convergence of certain moments. The renewal theorem itself concerns means:

Fact 16 (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 $h \geq 0$

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

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

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

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

Fact 17 (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*

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

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. On the assignment sheet, it will be seen that 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.

12.4 The coupling proofs

Suppose now that X is a renewal process with 1-arithmetic interarrival time distribution. Let $\hat{X}$ be a renewal process that is not stationary in the sense above, but such that $(\hat{X}_n)_{n\geq 0}$ is stationary and $\hat{X}$ has also a 1-arithmetic first arrival time. This can be achieved by choosing L as the size-biased pick F_0 that is a 1-arithmetic distribution, and choosing $\hat{Y}_0$ conditionally uniform on $\{1, \ldots, L\}$.

We want to couple these two independent processes. Define $N = \inf\{n \geq 1 : T_n = \hat{T}_n\}$, 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 $S_n = T_n - \hat{T}_n$ 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 $\mathbb{E}(Y_1 - \hat{Y}_1) = 0$.

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

$$\bar{Y}_n = \begin{cases} Y_n & \text{if } n < N \\ \hat{Y}_n & \text{if } n \geq N \end{cases}$$

so that the same relationship holds between $\bar{T}$, T and $\hat{T}$, and indeed between $\bar{X}$, X and $\hat{X}$, where in the latter case the switching time is T_N rather than N.

Now $\hat{X} \sim X$ and $\{(\hat{X}_{t+s} - \hat{X}_t)_{s\geq 0} \neq (\tilde{X}_{t+s} - \tilde{X}_t)_{s\geq 0}\} = \{t < T_N\}$, and $\mathbb{P}(t < T_N) \to 0$ as $t \to \infty$. This shows so-called convergence of the distribution of $(X_{t+s} - X_t)_{s\geq 0}$ to the distribution of $\tilde{X}$ in total variation, which is actually stronger than convergence in distribution as we claim above.

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

Lecture 13

M/M/1 queues and queueing networks

13.1 M/M/1 queues and the departure process

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 13, 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

$$\begin{aligned} \frac{1}{T_n}\int_0^{T_n} 1_{\{X_s=i\}}ds = \frac{1}{T_n}\sum_{k=0}^{n-1} Y_k 1_{\{M_k=i\}} &\to \xi_i \\ \frac{1}{n}\sum_{k=0}^{n-1} 1_{\{M_k=i\}} &\to \eta_i \end{aligned}$$

  as $n \to \infty$, 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, $T_n = Y_0 + \ldots + Y_{n-1}$ being just the sum of the weights.

- During any service period, a $geom(\lambda/(\lambda+\mu))$ number of new customers arrive.

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

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

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

Proposition 17 *V has invariant distribution ξ.*

Proof: A simple calculation shows that

$$\sum_{k\in\mathbb{N}} \xi_k d_{k,n} = \xi_0 d_{0,n} + \sum_{k=1}^{n+1} \xi_k d_{k,n} = \ldots = \xi_n.$$

□

Note that the conditional distribution of $D_{n+1} - D_n$ given V_n is the distribution of a typical service time $G \sim Exp(\mu)$ if $V_n = k \geq 1$ and the distribution of $G + Y$, where $Y \sim Exp(\lambda)$ is a typical interarrival time, if $V_n = 0$ 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 18 *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 2 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

$$\mathbb{P}(D_1 > y, V_1 = m) = \sum_{k=0}^{m+1} \mathbb{P}(D_1 > y, V_1 = m | V_0 = k)\xi_k = \ldots = e^{-\lambda y}\xi_m.$$

Now an induction shows $D_{n+1} - D_n \sim Exp(\lambda)$, and they are independent. □

This result 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

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

We do this 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.

Proposition 19 *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. □

Proposition 18 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 20 *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.4.2, $\rho_1 \geq 1$ would prevent equilibrium for $X^{(1)}$. If $\rho_1 < 1$ and $X^{(1)}$ is in equilibrium, then by Proposition 18, the arrival process for the second queue is a Poisson process at rate λ, and $\rho_2 \geq 1$ would prevent equilibrium for $X^{(2)}$. If $\rho_1 < 1$ and $\rho_2 < 1$, ξ as given in the statement of the proposition is an invariant distribution:

$$\xi_{(i,j+1)}q_{(i,j+1),(i+1,j+1)} + \xi_{(i+2,j)}q_{(i+2,j),(i+1,j+1)} + \xi_{(i+1,j+2)}q_{(i+1,j+2),(i+1,j+1)} = \xi_{(i+1,j+1)}$$

for $i, j \in \mathbb{N}$, and similar equations for $\xi_{(0,j+1)}$, $\xi_{(i+1,0)}$ and $\xi_{(0,0)}$ are easily checked. It is unique since X is clearly irreducible (we can find paths between any two states in $\mathbb{N}^2$). □

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: $\pi_{12} = 1$.

Fact 18 *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 an additional state 0, 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 19 *If X 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, there is 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.

Lecture 14

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

14.1 M/G/1 queues

An M/G/1 queue has independent and identically distributed service 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. Vice versa, however, we have nice residual interarrival distributions after a departure:

Proposition 21 *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 simple. You 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. In particular, when the departing customer leaves an empty system behind, there has to be an arrival, before the next service time starts. □

Proposition 22 *Let $\rho = \lambda\mathbb{E}(G)$ be the traffic intensity. If $\rho < 1$, then V has a unique invariant distribution ξ that is determined by*

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

Proof: To identify ξ as the 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. We define

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

and note that $\xi_0 = \phi(0) = 1 - \rho$. We have to check that 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 $\phi(s)$:

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

This equals $\phi(s)$ can be rearranged to give the formula that defines $\phi(s)$. This completes the proof since uniqueness follows from the obvious irreducibility. □

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 customer depends on the number of customers in the queue when he arrives.

Proposition 23 *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 22 we take $\mathbb{E}(s^N)$, and putting $\gamma = \lambda(s-1)$, we deduce the formula required by rearrangement. □

Corollary 4 *In the special case of M/M/1, the distribution of W is given by*

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

Proof: We calculate the moment generating function of this distribution to be

$$\frac{\mu-\lambda}{\mu} + \frac{\lambda}{\mu}\frac{\mu-\lambda}{\mu-\lambda-\gamma} = \frac{\mu-\lambda}{\mu}\frac{\mu-\gamma}{\mu-\lambda-\gamma}.$$

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. □

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. $U_n = X_{A_n}$, $n \geq 0$, is a Markov chain, where A_n is the nth arrival time.

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

Fact 20 *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 $q = \mathbb{E}(e^{\mu(q-1)A_1})$.

Proposition 24 *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, we have that W given $N \geq 1$ is exponential with parameter $\mu(1-q)$. □

Lecture 15

Markov models for insurance

15.1 Insurance ruin

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 $(A_n)_{n\geq 1}$ are positive, independent of the arrival process and identically distributed with common density $k(a)$, $a > 0$, and mean $\mu = \mathbb{E}(A_n)$.
- The insurance company risks an initial amount 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 claim process $C_t = \sum_{n=1}^{X_t} A_n$, $t \geq 0$.
- The reserve process $R_t = u + ct - C_t$, $t \geq 0$.
- The ruin probability $\psi(u) = \mathbb{P}(R_t < 0 \text{ for some } t \geq 0)$, $u \geq 0$.

Proposition 25 *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

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

Proof: First calculate the moment generating function of C_t:

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

which in the case where A_1 has a density, 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 $(A_n)_{n\geq 1}$).

The statements for R follow easily. □

We now turn to studying the ruin probabilities ψ.

Proposition 26 *ψ satisfies the renewal equation*

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

where

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

Proof: The arguments are standard. We only sketch. By a one-step analysis we obtain

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

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

We differentiate w.r.t. x, cancel the exponential terms, and integrate x again, from 0 to t to get, after suitable partial integrations

$$\psi(t) = \frac{\lambda}{c}\int_t^\infty \bar{K}(y) dy + \frac{\lambda}{c}\int_0^t \psi(t-y)\bar{K}(y) dy$$

as required. □

Corollary 5 *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). Note that f is not a probability density but the results hold for $\int_0^\infty f(y)dy \leq 1$. □

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

Proposition 27 *Assume that there is $R > 0$ such that*

$$1 = \int_0^\infty e^{Ry} f(y)dy = \frac{\lambda}{c}\int_0^\infty e^{Ry}\bar{K}(y)dy = \frac{\lambda\mu}{c} M_{K_0}(R).$$

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

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

Proof: Define $\hat{f}(y) = e^{Ry}f(y)$, $\hat{g}(y) = e^{Ry}g(y)$ and $\hat{\psi}(x) = e^{Ry}\psi(x)$. Then $\hat{\psi}(x)$ satisfies

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

The solution converges by the key renewal theorem to

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

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 more general version of the key renewal theorem still applies. □

Example 15 If $A_n \sim Exp(1/\mu)$, then in the notation of Proposition 26

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

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

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

15.2 Some simple finite-state-space models

Example 16 (Sickness-death) In health insurance, the following model arises. Let $\mathbb{S} = \{H, S, \Delta\}$ consist of the states healthy, sick and dead. Clearly, D 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 17 (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 18 (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 19 (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
- simple deterministic arguments "ω by ω" using almost sure convergence results

16.2 The exam

4 questions on 16 hours of lectures, 6 assignment sheets

1 question each on (4 lectures + 2 assignment sheets)

birth processes, cont-time Markov chains, renewal theory

1 question on last 4 lectures (actually 13-15), builing heavily on previous material

cf. specimen paper

previous exam papers are relevant as long as continuous-time Markov chains are concerned. Maybe old papers contain renewal theory. Some generalisations were not covered in this course. Discrete-time Markov chains are now a5.

Notes: do you have good notes? I might provide lecture notes for the first half of the course, since there are too few copies of Norris' book in the college libraries. Otherwise, I can only repeat my encouragement to read a book alongside the notes.

16.3 Duration-dependent transition rates

Renewal processes can be thought of as duration-dependent transition rates. If we take the infinitesimal viewpoint of Exercise A.3.3, where

$$p_{ij}(h) = \delta_{ij} + q_{ij}h + o(h)$$

then we can represent renewal processes by infinitesimal transition rates

$$\mathbb{P}_i(X_{d+h} = j | T_1 > d) = \delta_{ij} + q_{ij}(d)h + o(h)$$

where $q_{i,i+1}(d) = \lambda(d)$, $t \geq 0$, is obtained from

$$\mathbb{P}_i(X_{d+h} = i+1 | T_1 > d) = \mathbb{P}_i(T_1 \leq d+h | T_1 > d) = \frac{\bar{F}(d) - \bar{F}(d+h)}{\bar{F}(d)} = \frac{f(d)}{\bar{F}(d)}h + o(h),$$

i.e. $\lambda(d) = f(d)/\bar{F}(d)$, provided the interarrival distribution is continuous. Note that we can then express $\bar{F}$ in terms of $\lambda(\cdot)$ as

$$\bar{F}(d) = \exp\left\{ -\int_d^\infty \lambda(x)dx \right\}$$

since $\lambda(d) = f(d)/\bar{F}(d)$ can be read as a differential equation

$$\bar{F}'(d) = -\lambda(d)\bar{F}(d), \qquad \bar{F}(0) = 1,$$

that we solve, e.g. by integration of $(\log \bar{F}(d))' = -\lambda(d)$.

$\lambda(d)$ can be thought of as duration-dependent transition rate. In general, $\lambda_{ij}(d)$ may, of course, depend on the states i and j of the transition, as well.

The processes are no longer Markovian themselves, but the bivariate process (X_t, A_t) where A_t is the "age", the time passed since the last transition, is a Markov process in $\mathbb{S} \times [0, \infty)$. A theory can be built from this observation.

As mentioned earlier, such models are useful in insurance models where the length of an illness, the lifetime or the times between increase in no-claims discount are markedly non-exponential, but this is only one field of applications. Also other applications studied earlier such as population models, queueing systems etc. can be studied in this context.

16.4 Time-dependent transition rates

A different type of varying transition intensities is obtained if we make the rates dependent on time t, i.e.

$$\mathbb{P}(X_{t+h} = j | X_t = i) = \delta_{ij} + q_{ij}(t)h + o(h)$$

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 $\lambda(t) = \lambda_0 + \lambda_1 \sin(2\pi t)$. This can also be built into birth- and death-rates of population models, insurance models etc.

16.5 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.

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

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

16.6 Markov processes in uncountable state spaces ($\mathbb{R}$ or $\mathbb{R}^d$)

We have come across some processes for which we could have proved a Markov property, the age process $(A_t)_{t\geq 0}$ of a renewal process, the excess process $(E_t)_{t\geq 0}$ of a renewal process, but also the processes $(C_t)_{t\geq 0}$ and $(R_t)_{t\geq 0}$ with stationary independent increments that arose in insurance ruin by combining Poisson arrival times with jump sizes. A systematic study of such Markov processes in $\mathbb{R}$ is technically much harder, although many ideas and results transfer from our countable state space model.

16.7 Lévy processes

A particularly nice class of processes are processes with stationary independent increments, so-called Lévy processes. If you have learned about Brownian motion in another course, 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 ...

Brownian motion with drift is the only Lévy process with continuous sample paths. It will also be studied in b10 Stochastic Processes next term.

16.8 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. Time series are an example to be studied in b10 Stochastic Processes.

16.9 Internships, 4th year projects

My position funded by the Institute of Actuaries and the insurance company Aon. My research is in stochastic processes.

Particularly for people in the 4-year degree course. If anybody is interested in insurance applications, we have had a request from a Danish insurance company who collaborate with Royal Sun Alliance. They have a research center in Copenhagen and would like to offer summer internships that can lead into 4th year projects. The department of Statistics has also contact to Watson-Wyatt, and Mercer have shown interest in stronger links. We may not be able to arrange internships, but possibly find a contact person who may advise you in your applications.

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. If a project is arranged before the summer, you can start working on it in the summer and have more time for your courses in the year. 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.