SB3.1 Applied Probability

Christina Goldschmidt
Department of Statistics, University of Oxford

Hilary Term 2021
(Version of January 17, 2021)

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 list

- J.R. Norris, *Markov Chains*, Cambridge University Press (1997)
- G.R. Grimmett, and D.R. Stirzaker, *Probability and Random Processes*, 3rd edition, Oxford University Press (2001)
- G.R. Grimmett, and D.R. Stirzaker, *One Thousand Exercises in Probability*, Oxford University Press (2001)
- S.M. Ross, *Introduction to Probability Models*, 12th edition, Academic Press (2019)
- D.R. Stirzaker, *Elementary Probability*, Cambridge University Press (1994)

Pre-requisites

Part A Probability is essential.

Acknowledgements

The original version of this course was written by Matthias Winkel, and recently updated by Julien Berestycki. I am very grateful to them both. Any errors are my responsibility.

1 Introduction

This course is about modelling real-world phenomena using stochastic processes. A *stochastic process* is a random quantity which evolves in time, i.e. a collection $\{X_t : t \in T\}$ of random variables indexed by an ordered time-set T. The set T could be discrete or continuous, and it is often helpful to think of the process as a random function $t \mapsto X_t$. In general, X_t could take values in some metric space, although in this course will will focus on the case of a countable state-space, often the natural numbers. In general, the elements of $\{X_t : t \in T\}$ will be *dependent*, and that this considerably complicates the task of studying them.

What sort of real-world phenomena might one seek to model using such a process?

1.1 Examples

1. Population size
The number of individuals in a population of some animal or plant species: individuals live for some length of time, during which they give birth to children, before dying. How does the population evolve? Can it go extinct?

2. Epidemics
The number of people suffering from an infectious disease: one might want to model the number of susceptible individuals, the number infected, and the number who have recovered or died. Under what circumstances will the epidemic take off and infect a large proportion of the population, and when will it peter out without infecting many people?

3. Queues
In a supermarket, people queue at different check-outs, waiting until they reach the front of the queue to be served. Alternatively, in the post office, people wait in a single queue until one of several servers becomes free. How long will a customer wait to be served? How long is a busy period?

4. Insurance ruin
An insurance company is paid a regular stream of premium income by its customers, and claims of different sizes arrive through time. The premium is set sufficiently high that the company typically makes a profit, but there is still some chance that a large claim will arrive which causes it to go bankrupt. How likely is that?

5. Stock prices
The price of stocks in a company change (essentially) continuously over the course of time and are influenced by many complex factors. Can we make predictions about the behaviour of a stock price? How much should one charge for an option on the stock?

In all of these examples, it makes sense to use a model containing randomness. The first four examples have continuous time and discrete state-spaces, and will be treated in this course. Example 5 has both continuous time and continuous space. The mathematical framework needed to deal with such processes is developed in B8.2 Continuous Martingales and Stochastic Calculus, and the modelling aspects are addressed in B8.3 Mathematical Models of Financial Derivatives.

In order to develop sensible models, we will first need to develop a certain amount of theory. This course very much follows on from the Part A Probability course, and so you will probably find it helpful to go back to your notes for that course and re-familiarise yourself with it.

1.2 A note about rigour

B8.1 Probability, Measure and Martingales is not a pre-requisite for this course so, in particular, we will not focus on measure-theoretic issues. However, any properly rigorous treatment of probability requires measure theory, and so there will be some places where we will make appeal to standard results from measure theory. If you have not seen these before, and would like to know more, the Appendix to the book *Markov chains* by James Norris is a good place to start. Those who have attended B8.1 and are looking to understand some of these issues more thoroughly should go to B8.2 Continuous Martingales and Stochastic Calculus.

2 Poisson processes

In this section, we will recap some material about Poisson processes from Part A Probability and set things up in a way that will be useful for the rest of the course.

2.1 Building blocks

We will use the notation $\mathbb{N} := \{0, 1, 2, \ldots\}$.

We begin by recalling the definition of two probability distributions which will play a key role in this course.

A discrete random variable X has the *Poisson distribution* with mean $\lambda \geq 0$ (we will write $\mathrm{Po}(\lambda)$) if

$$\mathbb{P}(X = n) = \frac{e^{-\lambda}\lambda^n}{n!}, \quad n \geq 0.$$

A continuous random variable T has the *exponential distribution* with parameter $\lambda \geq 0$ (we will write $\mathrm{Exp}(\lambda)$) if it has density

$$f(t) = \lambda e^{-\lambda t}, \quad t \geq 0.$$

In particular, $\mathbb{P}(T > t) = e^{-\lambda t}$ and $\mathbb{E}[T] = 1/\lambda$.

2.2 Poisson processes

Definition 2.1. *A random process $X = (X_t)_{t \geq 0}$ is a* counting process *if it takes values in $\mathbb{N}$ and $X_s \leq X_t$ whenever $s \leq t$.*

Poisson processes are important examples of counting processes.

Definition 2.2 (Holding-time definition). *Let $(Z_n)_{n \geq 1}$ be a sequence of i.i.d. Exp(λ) random variables, for some $\lambda \in (0, \infty)$. Set $T_0 = 0$ and, for $n \geq 1$, $T_n = \sum_{k=1}^{n} Z_k$. Define*

$$X_t = \#\{n \geq 1 : T_n \leq t\}, \quad t \geq 0,$$

so that, in particular, $X_0 = 0$. Then the process $(X_t)_{t \geq 0}$ is called a Poisson process of rate λ *(PP(λ) for short). The random variables $Z_1, Z_2, \ldots$ are called* holding times *or* inter-arrival *times.*

We may think of $T_1, T_2, \ldots$ as the arrival times of customers at a shop, or cars driving along St Giles', or particles detected by a Geiger counter. Then X_t gives the number of customers, cars or particles which have arrived by time t. (Of course, this is a mathematical model, and we have said nothing about how *well* it might model these phenomena! We will come back to this issue.)

[picture]

Note that we have drawn the process as a (random) right-continuous function $[0, \infty) \to \mathbb{N}$ given by $t \mapsto X_t$. (Recall that right-continuous means that $f(y) \to f(x)$ as $y \downarrow x$, for all $x \in [0, \infty)$.)

Observe that $\{X_t : t \geq 0\}$ is a *dependent* collection of random variables. For example, we have $\mathbb{P}(X_{3.6} = 0) = \exp(-3.6\lambda) > 0$ but $\mathbb{P}(X_{3.6} = 0 | X_{3.5} = 3) = 0$.

2.3 Markov property

Let $\mathbb{S}$ be a countable state-space. Recall from Part A Probability that a process $(Y_n)_{n \geq 0}$ is a discrete-time Markov chain with state-space $\mathbb{S}$ if it satisfies the Markov property:

$$\mathbb{P}(Y_n = y_n | Y_0 = y_0, Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}) = \mathbb{P}(Y_n = y_n | Y_{n-1} = y_{n-1})$$

for all $n \geq 1$ and all $y_0, y_1, \ldots, y_n \in \mathbb{S}$. Assuming time-homogeneity (i.e. that $\mathbb{P}(Y_n = j | Y_{n-1} = i)$ does not depend on n), the distribution of $(Y_n)_{n \geq 0}$ is entirely specified by an *initial distribution* $\mu = (\mu_i)_{i \in \mathbb{S}}$ and a *transition matrix* $P = (p_{ij})_{i,j \in \mathbb{S}}$, where $\mu_i = \mathbb{P}(Y_0 = i)$ and $p_{ij} = \mathbb{P}(Y_n = j | Y_{n-1} = i)$. Then

$$\mathbb{P}(Y_0 = y_0, Y_1 = y_1, \ldots, Y_n = y_n) = \mu_{y_0} p_{y_0 y_1} \ldots p_{y_{n-1} y_n}.$$

Suppose that $(Y_n)_{n \geq 0}$ is a Markov chain with transition matrix P started from some state $i \in \mathbb{S}$ i.e.

$$\mu_k = \delta_{ik} = \begin{cases} 1 & \text{if } k = i \\ 0 & \text{otherwise} \end{cases}$$

(we write $\mu = \delta_i$ for short). Then an equivalent formulation of the Markov property is that $(Y_k)_{0 \leq k \leq n}$ and $(Y_k)_{k \geq n}$ are conditionally independent given $Y_n = j$, for any $j \in \mathbb{S}$. Moreover, given $Y_n = j$, $(Y_{n+k})_{k \geq 0}$ is a Markov chain with transition matrix P started from j.

It turns out that a Markov property also holds for the Poisson process. By a *Poisson process started from* k, we mean a process $(X_t)_{t \geq 0}$ such that $X_t = k + \tilde{X}_t$ where $(\tilde{X}_t)_{t \geq 0}$ is a Poisson process started from 0, as in our earlier definition.

Theorem 2.3 (Markov property)**.** *Let $X = (X_t)_{t \geq 0}$ be a Poisson process of rate λ started from 0. Fix $t \geq 0$. Then, given $X_t = k$, $(X_r)_{r \leq t}$ and $(X_{t+s})_{s \geq 0}$ are independent and $(X_{t+s})_{s \geq 0}$ is a Poisson process of rate λ started from k.*

Remark 2.4. *We could, equivalently, have said that $(X_{t+s} - X_t)_{s \geq 0}$ is a Poisson process of rate λ started from 0 independent of $(X_r)_{0 \leq r \leq t}$. But the version stated above will generalise better.*

The key to this result is the following property of the exponential distribution.

Lemma 2.5 (Memoryless property)**.** *Let $E \sim Exp(\lambda)$. Then for all $x, y \geq 0$,*

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

Proof. This is a straightforward calculation:

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

Lemma 2.6 (Extended memoryless property)**.** *Suppose $E \sim Exp(\lambda)$ and that $L \geq 0$ is a random variable independent of E. Then, given $E > L$, $E - L$ is conditionally independent of L and*

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

Proof. See Problem Sheet 1. □

Proof of Theorem 2.3. Set $\tilde{X}_s = X_{t+s}$. Conditional on $X_t = k$, the holding times of $(\tilde{X}_s)_{s\geq 0}$ are $\tilde{Z}_1, \tilde{Z}_2, \ldots$ where

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

Note that $Z_1, Z_2, \ldots \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$. We have

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

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

Since, given $X_t = k$,

$$X_r = \#\left\{1 \leq n \leq k : \sum_{i=1}^{n} Z_i \leq r\right\}, \quad r \leq t$$
$$\tilde{X}_s = k + \#\left\{n \geq 1 : \sum_{i=1}^{n} \tilde{Z}_i \leq s\right\}, \quad s \geq 0$$

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

Remark 2.7. *As you will see on Problem Sheet 1, the exponential distribution is the only continuous distribution to possess the memoryless property. So if we had trying to build our counting process with any other holding times, we would not have obtained the Markov property. We will come back to this idea when we study renewal processes.*

2.4 Dealing with processes in continuous time

There are various technical subtleties associated with working in continuous time which are not present in discrete time. These stem from the fact that $\{X_t : t \geq 0\}$ is an *uncountable* collection of random variables and, in principle, we have problems making sense of probabilities of unions and intersections of uncountably many events. For example, we might want to calculate

$$\mathbb{P}\left(X_t = i \text{ for some } t \in [0, \infty)\right) = \mathbb{P}\left(\cup_{t\geq 0}\{X_t = i\}\right),$$

where the union on the right-hand side is uncountable. Fortunately, there is a result from measure theory which says that the distribution of a right-continuous process $(X_t)_{t\geq 0}$ with values in a countable state-space $\mathbb{S}$ is entirely determined by its *finite-dimensional distributions:*

$$\mathbb{P}\left(X_{t_1} = x_1, X_{t_2} = x_2, \ldots, X_{t_n} = x_n\right)$$

for $n \geq 1$, $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$ and $x_1, x_2, \ldots, x_n \in \mathbb{S}$. To give an idea of why this works, for our example, if we assume right-continuity then we can say that

$$\mathbb{P}\left(X_t = i \text{ for some } t \in [0, \infty)\right) = 1 - \lim_{n\to\infty} \sum_{j_1, j_2, \ldots, j_n \neq i} \mathbb{P}\left(X_{q_1} = j_1, \ldots, X_{q_n} = j_n\right),$$

where $q_1, q_2, \ldots$ is an enumeration of the rationals, and the right-hand side then only concerns countably many events. (Since this is a course in applied probability, we won't go further into the details here.) So we shall always consider right-continuous processes and, when necessary, make appeal to this result.

Coming back to our statement of the Markov property for a Poisson process $(X_t)_{t\geq 0}$, we are really treating each of $(X_r)_{r\leq t}$ and $(X_{t+s})_{s\geq 0}$ as a random variable in its own right, rather than as a collection of a random variables. Indeed, each of these *can* be thought of as a random variable taking values in the space of right-continuous integer-valued functions (with domains $[0,t]$ and $[0,\infty)$ respectively).

What might an event look like for such a random variable? An example is

$$\{(X_r)_{r\leq t} \in A\},$$

where

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

Of course, we would usually write this more simply as

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

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

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

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

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

for all n, m, $r_1 \leq \ldots \leq r_m$, $s_1 \leq \ldots \leq s_n$, $x_1, \ldots, x_m, y_1, \ldots, y_n \in \mathbb{S}$.

2.5 Alternative Poisson process definitions

Proposition 2.8 (Transition probability definition). *A right-continuous integer-valued process $X = (X_t)_{t\geq 0}$ started from 0 is a Poisson process of rate λ if and only if it has the following properties:*

1. *$X_t \sim Po(\lambda t)$ for all $t \geq 0$.*
2. *X has independent increments i.e. for any sequence of times $0 = t_0 \leq t_1 \leq \ldots \leq t_n < \infty$, the random variables*
$$\{X_{t_k} - X_{t_{k-1}}, 1 \leq k \leq n\}$$
are independent.
3. *X has stationary increments, i.e. for all $s, t \geq 0$,*
$$X_{t+s} - X_t \overset{d}{=} X_s - X_0 = X_s.$$

Proof. First suppose that X is a Poisson process. We prove that the three properties hold.

1. We have $\mathbb{P}(X_t = 0) = \mathbb{P}(T_1 > t) = e^{-\lambda t}$. For $n \geq 1$, we have

$$\begin{aligned}\mathbb{P}(X_t = n) &= \mathbb{P}(T_n \leq t, T_{n+1} > t) \\ &= \mathbb{P}(T_n \leq t) - \mathbb{P}(T_{n+1} \leq t).\end{aligned}$$

Recall that a sum of n independent $\mathrm{Exp}(\lambda)$ random variables has $\mathrm{Gamma}(n, \lambda)$ distribution, with density $\lambda^n x^{n-1} e^{-\lambda x}/(n-1)!$, $x \geq 0$. So $T_n \sim \mathrm{Gamma}(n, \lambda)$ and $T_{n+1} \sim \mathrm{Gamma}(n+1, \lambda)$. Hence, for $n \geq 1$,

$$\mathbb{P}(X_t = n) = \int_0^t \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} dx - \int_0^t \frac{\lambda^{n+1} x^n e^{-\lambda x}}{n!} dx$$

and integrating the first term by parts gives

$$\begin{aligned}&= \left[\frac{\lambda^n x^n e^{-\lambda x}}{n!}\right]_0^t + \int_0^t \frac{\lambda^{n+1} x^n e^{-\lambda x}}{n!} dx - \int_0^t \frac{\lambda^{n+1} x^n e^{-\lambda x}}{n!} dx \\ &= \frac{e^{-\lambda t}(\lambda t)^n}{n!},\end{aligned}$$

which implies that $X_t \sim \mathrm{Po}(\lambda t)$.

2. We proceed by induction on n. The statement is trivial for $n = 1$. Let $i_1, i_2, \ldots, i_n \in \mathbb{N}$. Then

$$\begin{aligned}&\mathbb{P}\left(\bigcap_{k=1}^{n} \{X_{t_k} - X_{t_{k-1}} = i_k\}\right) \\ &= \mathbb{P}\left(X_{t_n} - X_{t_{n-1}} = i_n \;\middle|\; \bigcap_{k=1}^{n-1} \{X_{t_k} - X_{t_{k-1}} = i_k\}\right) \mathbb{P}\left(\bigcap_{k=1}^{n-1} \{X_{t_k} - X_{t_{k-1}} = i_k\}\right).\end{aligned}$$

Recall that $t_0 = 0$. Now, the first term equals

$$\begin{aligned}&\mathbb{P}\left(X_{t_n} = \sum_{k=1}^{n-1} i_k + i_n \;\middle|\; X_{t_{n-1}} = \sum_{k=1}^{n-1} i_k, X_{t_{n-2}} = \sum_{k=1}^{n-2} i_k, \ldots, X_{t_1} = i_1\right) \\ &= \mathbb{P}\left(X_{t_n} = \sum_{k=1}^{n-1} i_k + i_n \;\middle|\; X_{t_{n-1}} = \sum_{k=1}^{n-1} i_k\right) \quad \text{by the Markov property applied at time } t_{n-1} \\ &= \mathbb{P}\left(X_{t_n} - X_{t_{n-1}} = i_n\right),\end{aligned}$$

since $(X_{t_{n-1}+s} - X_{t_{n-1}})_{s \geq 0}$ is another Poisson process, independent of $(X_r)_{r \leq t_{n-1}}$. By induction,

$$\mathbb{P}\left(\bigcap_{k=1}^{n} \{X_{t_k} - X_{t_{k-1}} = i_k\}\right) = \prod_{k=1}^{n} \mathbb{P}\left(X_{t_k} - X_{t_{k-1}} = i_k\right),$$

and so $\left\{X_{t_k} - X_{t_{k-1}}, 1 \leq k \leq n\right\}$ are independent.

3. This follows directly from the fact that $(X_{t+s} - X_t)_{s \geq 0}$ is also a Poisson process of rate λ.

Suppose now that X is a right-continuous integer-valued process satisfying the three conditions. Then for $0 = t_0 \le t_1 \le \ldots \le t_n$ and $k_0 = 0, k_1, \ldots, k_n \in \mathbb{N}$, we have

$$\begin{aligned}
&\mathbb{P}\left(X_{t_1} = k_1, \ldots, X_{t_n} = k_n\right) \\
&= \mathbb{P}\left(X_{t_1} = k_1\right)\mathbb{P}\left(X_{t_2} - X_{t_1} = k_2 - k_1\right)\ldots\mathbb{P}\left(X_{t_n} - X_{t_{n-1}} = k_n - k_{n-1}\right) \\
&= \prod_{i=1}^{n} \frac{e^{-\lambda(t_i - t_{i-1})}(\lambda(t_i - t_{i-1}))^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} \\
&= e^{-\lambda t_n}\lambda^{k_n}\prod_{i=1}^{n}\frac{(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}.
\end{aligned}$$

But since the finite-dimensional distributions characterise the distribution of such a process, we must have $X \sim PP(\lambda)$. □

There is a third characterisation of a Poisson process which will also play an important role in this course.

Proposition 2.9 (Infinitesimal definition)**.** *Suppose $X = (X_t)_{t\ge 0}$ is a right-continuous integer-valued increasing process started from 0. Then X is a Poisson process of rate $\lambda > 0$ if and only if it has independent increments and, as $h \downarrow 0$, uniformly in t,*

$$\mathbb{P}\left(X_{t+h} - X_t = 0\right) = 1 - \lambda h + o(h), \qquad \mathbb{P}\left(X_{t+h} - X_t = 1\right) = \lambda h + o(h). \tag{1}$$

Proof. We have already shown that a Poisson process has independent increments. The rest of the proof of the "only if" part follows from Problem Sheet 1.

Let us turn to the "if" part. The given conditions imply, in particular, that for $k \ge 2$,

$$\mathbb{P}\left(X_{t+h} - X_t = k\right) = o(h),$$

uniformly in $t \ge 0$. Set $p_k(t) = \mathbb{P}(X_t = k)$ for $k \ge 0$. We will derive a system of differential equations satisfied by $(p_k(t))_{k\ge 0}$ and demonstrate that it has unique solution given by the Poisson probability mass function.

To this end, note that by the law of total probability, for $t \ge 0$ and $k \ge 1$,

$$\begin{aligned}
p_k(t+h) &= \sum_{i=0}^{k}\mathbb{P}\left(X_{t+h} - X_t = i\right)\mathbb{P}\left(X_t = k - i\right) \\
&= (1 - \lambda h + o(h))p_k(t) + (\lambda h + o(h))p_{k-1}(t) + o(h).
\end{aligned}$$

So

$$\frac{p_k(t+h) - p_k(t)}{h} = \lambda(p_{k-1}(t) - p_k(t)) + O(h).$$

Substituting $s = t - h$, we also obtain that for $s \ge h$ and $k \ge 1$,

$$\frac{p_k(s) - p_k(s-h)}{h} = \lambda(p_{k-1}(s-h) - p_k(s-h)) + O(h).$$

Letting $h \downarrow 0$, we see that $p_k(t)$ is continuous and differentiable with

$$p_k'(t) = \lambda(p_{k-1}(t) - p_k(t)).$$

Similarly, we see that

$$p_0'(t) = -\lambda p_0(t).$$

Now note that we have $X_0 = 0$ and so the initial conditions are $p_0(0) = 1$, $p_k(0) = 0$, $k \geq 2$. It is clear that the unique solution to the last differential equation is $p_0(t) = e^{-\lambda t}$. But then we can solve the remaining differential equations inductively, each time with an exponential integrating factor (note that they are linear, and so have unique solutions), to obtain

$$p_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \quad k \geq 0.$$

Now observe that $(X_{t+s} - X_t)_{s \geq 0}$ also satisfies (1) and so we deduce that $X_{t+s} - X_t \sim \mathrm{Po}(\lambda t)$. It then follows from Proposition 2.8 that X is a Poisson process of rate λ. □

The system of differential equations appearing in the proof are known as the *forward equations*, which we will come across in much greater generality when we come to continuous-time Markov chains.

2.6 Further properties

Theorem 2.10. *Let $X = (X_t)_{t \geq 0}$ be a Poisson process. Then conditional on $\{X_t = n\}$ the jump times $T_1, T_2, \ldots, T_n$ have the same distribution as an ordered sample of size n from the uniform distribution on* $[0, t]$.

Proof. First note that if $U_1, U_2, \ldots, U_n$ are i.i.d. $\mathrm{U}[0, t]$, then

$$\mathbb{P}(U_1 \leq U_2 \leq \ldots \leq U_n) = \frac{1}{n!},$$

as there are $n!$ possible orderings, all equally likely. The joint density of an ordered sample of size n from the uniform distribution on $[0, t]$ is the same as the joint density of $(U_1, U_2, \ldots, U_n)$ conditioned on $\{U_1 \leq U_2 \leq \ldots \leq U_n\}$, which is

$$\frac{t^{-n}\mathbb{1}_{\{0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t\}}}{1/n!} = \frac{n!\mathbb{1}_{\{0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t\}}}{t^n}.$$

Now the holding times $Z_1, \ldots, Z_{n+1}$ are i.i.d. $\mathrm{Exp}(\lambda)$, with joint density

$$\lambda^{n+1} \exp(-\lambda(z_1 + \cdots + z_n)), \quad z_1, \ldots, z_{n+1} \in \mathbb{R}_+.$$

Changing variable to $T_1 = Z_1, T_2 = Z_1 + Z_2, \ldots, T_{n+1} = \sum_{i=1}^{n+1} Z_i$, and noting that the Jacobian is 1, we get

$$\lambda^{n+1} \exp(-\lambda t_{n+1}), \quad 0 \leq t_1 \leq t_2 \leq \ldots \leq t_{n+1}.$$

So for $A \subseteq [0, t]^n$,

$$\begin{aligned}
\mathbb{P}((T_1, \ldots, T_n) \in A | X_t = n) &= \frac{\mathbb{P}((T_1, \ldots, T_n) \in A, T_{n+1} > t)}{\mathbb{P}(X_t = n)} \\
&= \frac{\lambda^{n+1} \int_{(t_1,\ldots,t_n) \in A} \int_t^\infty \exp(-\lambda t_{n+1}) dt_{n+1} \mathbb{1}_{\{0 \leq t_1 \leq \ldots \leq t_n \leq t\}} dt_1 \ldots dt_n}{e^{-\lambda t}(\lambda t)^n / n!} \\
&= \frac{n!}{t^n} \int_{(t_1,\ldots,t_n) \in A} \mathbb{1}_{\{0 \leq t_1 \leq \ldots \leq t_n \leq t\}} dt_1 \ldots dt_n,
\end{aligned}$$

as required. □

We recall some useful facts which were proved in Part A Probability. It is a good exercise to try to reprove them for yourself. (Which of the three definitions is easiest to work with in each case?)

Theorem 2.11 (Superposition of Poisson processes)**.** *Let* $X = (X_t)_{t \geq 0}$ *and* $Y = (Y_t)_{t \geq 0}$ *be independent Poisson processes of rates* λ *and* μ *respectively. Let* $Z_t = X_t + Y_t$*. Then* $Z = (Z_t)_{t \geq 0}$ *is a Poisson process of rate* $\lambda + \mu$*.*

Theorem 2.12 (Thinning of Poisson processes)**.** *Let* Z *be a Poisson process of rate* λ *and let* $p \in [0, 1]$*. Mark each point of the process independently with probability* p*. Let* X *be the counting process of marked points, and let* Y *be the counting process of unmarked points. Then* X *is a Poisson process of rate* λp*,* Y *is a Poisson process of rate* $\lambda(1 - p)$ *and* X *and* Y *are independent.*

2.7 Summary of the rest of the course

The theoretical content of this course is concerned with two generalisations of the Poisson process.

- The construction of the Poisson process in terms of exponential holding times and the resulting Markov property can be considerably generalised. This is done essentially by allowing different parameters for the holding times in different states and allowing jumps, which instead of always being $+1$, are random and depend on the current state. This gives the class of *continuous-time Markov chains*. We will spend roughly the first half of the course studying continuous-time Markov chains. Our main reference will be *Markov Chains* by Norris.

- The Poisson process is the prototype of a counting process. Many quantities can be explicitly calculated for it. However, in applications, exponential inter-arrival times may not be a appropriate, for example when modelling the arrival of insurance claims. If we relax the assumption of exponentiality of the inter-arrival times (but keep their independence and identical distribution) we obtain the class of counting processes called *renewal processes*. Since exact calculations are often impossible or not helpful in this context, the most important results of renewal theory are limiting results. Our main reference will be Chapter 10 of *Probability and Random Processes* by Grimmett and Stirzaker.

We will also spend a lot of time on applications.

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

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

3 Birth processes

Suppose we want to model a growing population in continuous time. If new individuals are born (or arrive) at a constant rate, then we could use a Poisson process. If, however, the birth rate depends on the number of individuals present, the Poisson process will not be a good model.

Definition 3.1. *Let $(\lambda_n)_{n\geq 0}$ be a sequence such that $0 \leq \lambda_n < \infty$ for all $n \geq 0$. Fix $k \in \mathbb{N}$ and let $Z_1, Z_2, \ldots$ be independent random variables such that $Z_n \sim Exp(\lambda_{k+n-1})$ for $n \geq 1$. Then the process $(X_t)_{t\geq 0}$ defined by*

$$X_t = k + \# \left\{ n \geq 1 : \sum_{i=1}^{n} Z_i \leq t \right\},$$

is called a simple birth process*, started from k.*

Remark 3.2. *Note that X is a counting process which, when it first reaches state n, waits a length of time distributed as $Exp(\lambda_n)$ and then jumps to $n+1$. "Simple" refers to the fact that no two births occur at the same time.*

Proposition 3.3 (Competing exponentials)**.** *Let $E_1, E_2, \ldots, E_n$ be independent and identically distributed Exp(λ) random variables, thought of as the times until n alarm clocks ring. Then*

$$M := \min\{E_1, E_2, \ldots, E_n\},$$

the time until the first clock rings, has $Exp(n\lambda)$ distribution. Let K be the index of the first clock to ring. Then K is uniformly distributed on $\{1, 2, \ldots, n\}$ and conditionally on $K = k$, the random variables M and $\{E_j - M : j \neq k\}$ are independent and $E_j - M \sim Exp(\lambda)$.

Proof. This is a special case of a question on Problem Sheet 1. □

3.1 Example: the Yule process

Consider a population in which each individual gives birth after an Exp(λ) time, independently and repeatedly. If n individuals are present then each waits an Exp(λ) time until it gives birth. So the first birth occurs after an Exp($n\lambda$) time. Then we have $n+1$ individuals and, by Proposition 3.3, the process begins afresh:

- the $n-1$ individuals which didn't reproduce must each wait a further Exp(λ) time;
- the individual which did reproduce gets a new Exp(λ) time until it next gives birth;
- so does the individual which was born.

So the size of the population performs a simple birth process with rates $\lambda_n = n\lambda$, $n \geq 1$. This is often known as a *Yule process* of rate λ.

Suppose $Y_0 = 1$ and let Y_t be the number of individuals alive at time $t > 0$. Let $m(t) := \mathbb{E}[Y_t]$.

Proposition 3.4. *We have $m(t) = e^{\lambda t}$, for $t \geq 0$.*

Proof. Write T for the time of the first birth, i.e. $T = \inf\{t \geq 0 : Y_t = 2\}$. Notice that after the first birth has occurred, by construction we have two independent copies of the original Yule process. Now, let us split the expectation according to whether T has occurred by time t or not:

$$m(t) = \mathbb{E}[Y_t] = \mathbb{E}\left[Y_t \mathbb{1}_{\{T \le t\}}\right] + \mathbb{E}\left[Y_t \mathbb{1}_{\{T > t\}}\right]$$
$$= \int_0^t \mathbb{E}[Y_t | T = u] \lambda e^{-\lambda u} du + \mathbb{P}(T > t),$$

since $Y_t = 1$ on the event $\{T > t\}$. Now, if $T = u$ then $(Y_{s+u})_{s \ge 0}$ evolves as the sum of two independent copies of the original Yule process. So we have $\mathbb{E}[Y_t | T = u] = 2\mathbb{E}[Y_{t-u}] = 2m(t-u)$. Putting this together, we obtain

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

We need to solve this integral equation. Changing variable in the integral to $s = t - u$ gives

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

and so

$$e^{\lambda t} m(t) = 2\lambda \int_0^t e^{\lambda s} m(s) ds + 1.$$

Differentiating in t, we obtain

$$\lambda e^{\lambda t} m(t) + e^{\lambda t} m'(t) = 2\lambda e^{\lambda t} m(t)$$

which is equivalent to

$$m'(t) = \lambda m(t).$$

Since $m(0) = \mathbb{E}[Y_0] = 1$, we obtain $m(t) = e^{\lambda t}$. □

So, on average, the population size grows exponentially with rate λ. You will see on Problem Sheet 1 that, in fact, the population size also grows exponentially in an almost sure sense.

3.2 Markov property

Like Poisson processes, simple birth processes have the Markov property.

Proposition 3.5. *Let X be a simple birth process with rates $(\lambda_n)_{n \ge 0}$ started from $X_0 = k$. Fix $t \ge 0$ and $i \ge k$. Then, given $X_t = i$, $(X_r)_{r \le t}$ and $(X_{t+s})_{s \ge 0}$ are conditionally independent, and the conditional distribution of $(X_{t+s})_{s \ge 0}$ is that of a simple birth process with rates $(\lambda_n)_{n \ge 0}$ started from i.*

Proof. This proceeds in exactly the same way as for the Poisson process, just replacing λ by λ_n for the appropriate n. □

3.3 Explosion

There are two phenomena which may arise in the setting of birth processes which cannot happen for a Poisson process. The first is simple to understand: if it happens that $\lambda_0, \ldots, \lambda_{n-1} > 0$ but $\lambda_n = 0$ for some $n \ge 0$ then if $X_0 \le n$, the birth process will eventually get stuck at population size n. (We interpret an Exp(0) as being almost surely infinite.) This is not completely absurd from the modelling perspective: consider the situation where there is only a finite amount of space or resources for the population, beyond which point it is impossible for individuals to reproduce. We have actually already come across this phenomenon, *absorption*, in the setting of discrete-time Markov chains.

The other phenomenon is *explosion*: if the rates λ_n increase too quickly, it may happen that infinitely many individuals are born in finite time. Note that this would not be a desirable feature of a model: real-world populations do not become infinite! So it is useful from the modelling perspective to know for which models explosion cannot occur, and restrict our attention to those.

Definition 3.6. *Consider a simple birth process X with rates $(\lambda_n)_{n\geq 0}$ started from $k \in \mathbb{N}$, and let $T_n = \inf\{t \geq 0 : X_t = k+n\}$ for $n \geq 1$. Let $T_\infty = \lim_{n\to\infty} T_n = \sum_{i=1}^{\infty} Z_i$ (where we allow ∞ as a possible value for the limit). Then we say* explosion is possible *if $\mathbb{P}(T_\infty < \infty) > 0$.*

There turns out to be a simple criterion for whether explosion is possible.

Theorem 3.7. *Let X be a simple birth process started from k.*

(a) If $\sum_{i=k}^{\infty} \frac{1}{\lambda_i} < \infty$ then $\mathbb{P}(T_\infty < \infty) = 1$ i.e. explosion occurs with probability 1.

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

Proof. Without loss of generality, we shall suppose that $k = 0$ (otherwise, simply shift the indices).

(a) We have

$$\mathbb{E}[T_\infty] = \mathbb{E}\left[\sum_{i=1}^{\infty} Z_i\right] = \sum_{i=1}^{\infty} \mathbb{E}[Z_i] = \sum_{i=1}^{\infty} \frac{1}{\lambda_i},$$

where we may interchange the sum and expectation by Tonelli's theorem. Since the series is finite, $\mathbb{E}[T_\infty] < \infty$, which implies that $\mathbb{P}(T_\infty < \infty) = 1$.

(b) Note that $\mathbb{P}(T_\infty < \infty) = 0$ iff $\mathbb{P}(T_\infty = \infty) = 1$, and that the latter is implied by $\mathbb{E}\left[e^{-T_\infty}\right] = 0$. Now for any n,

$$\mathbb{E}\left[e^{-T_\infty}\right] \leq \mathbb{E}\left[\exp\left(-\sum_{i=1}^{n} Z_i\right)\right] = \prod_{i=1}^{n} \mathbb{E}\left[\exp(-Z_i)\right]$$

by independence of the holding times. Each term in the product is the moment generating function $\mathbb{E}[\exp(\theta Z_i)]$ evaluated at $\theta = -1$. So

$$\mathbb{E}\left[\exp(-Z_{i+1})\right] = \frac{\lambda_i}{\lambda_i + 1} = \left(1 + \frac{1}{\lambda_i}\right)^{-1}.$$

Taking logs, we obtain

$$-\log \mathbb{E}\left[e^{-T_\infty}\right] \geq \sum_{i=0}^{n-1} \log\left(1 + \frac{1}{\lambda_i}\right)$$

and since this holds for each n we get

$$-\log \mathbb{E}\left[e^{-T_\infty}\right] \geq \sum_{i=0}^{\infty} \log\left(1 + \frac{1}{\lambda_i}\right).$$

Now, if $\lambda_i \leq 1$ for infinitely many values of i then clearly the right-hand side is infinite. On the other hand, if $\lambda_i > 1$ for all sufficiently large i, say $i \geq I$, then we have $\log(1+1/\lambda_i) \geq \log(2)/\lambda_i$ for all $i \geq I$ and then

$$-\log \mathbb{E}\left[e^{-T_\infty}\right] \geq \log(2) \sum_{i\geq I} \frac{1}{\lambda_i}.$$

Since we have only omitted finitely many terms from the sum $\sum_{i=0}^{\infty} 1/\lambda_i$, the right-hand side must also be infinite, from which it follows that $\mathbb{E}\left[e^{-T_\infty}\right] = 0$, as desired. □

Note that this theorem boils down to the fact that a sum of independent exponential random variables is finite if and only if its mean is finite. This is *not true* in general: finiteness of the expectation of a random variable implies finiteness of the random variable, but the converse is false.

3.4 Quick recap: branching processes

In Prelims Probability, we saw a different model for a growing population, *in discrete time*: a branching process. In that setting, we have a population in which each individual lives for a unit time and, just before dying, gives birth to a random number of children, distributed according to the *offspring distribution*. Different individuals then reproduce independently in the same manner. Let P_n be the size of the population in generation n. Then $(P_n)_{n\geq 0}$ is a *branching process*. Let N be a random variable with the offspring distribution. Suppose that $G(s) = \mathbb{E}\left[s^N\right]$ is the probability generating function and $\mathbb{E}[N] = \mu < \infty$. Let us assume $P_0 = 1$. Then we saw in Prelims that

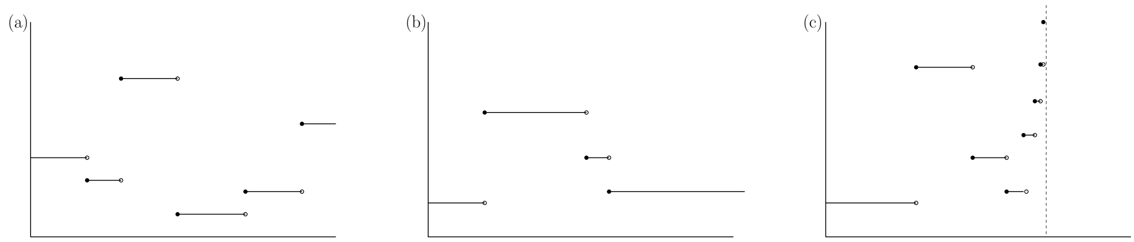
$$\mathbb{E}\left[s^{P_n}\right] = G^{(n)}(s)$$

(the n-fold composition of G with itself) and that the probability q of extinction is given by the minimal non-negative solution of $s = G(s)$. We also saw that $q = 1$ if $\mu \leq 1$, while $q < 1$ if $\mu > 1$. You will see a continuous-time version of a branching process on Problem Sheet 1.

4 Continuous-time Markov chains: basic theory

4.1 Right-continuous processes

Let $\mathbb{S}$ be a countable state-space. A right-continuous process $X = (X_t)_{t\geq 0}$ taking values in $\mathbb{S}$ must remain for a while in each state, and there are three possible behaviours:



Let $T_0 = 0$ and $T_{n+1} = \inf\{t \geq T_n : X_t \neq X_{T_n}\}$, $n \geq 1$. Set $Z_n = T_n - T_{n-1}$, $n \geq 1$. Then $T_0, T_1, T_2, \ldots$ are the *jump-times* of X and $Z_1, Z_2, \ldots$ are its *holding times*. In case (b), we have $T_n = \infty$ for some n. In case (c), $T_\infty := \lim_{n\to\infty} T_n < \infty$. If the process explodes, we will adjoin an extra state, ∞, to the state space, and always set $X_t = \infty$ for $t > T_\infty$. A process which is set to ∞ after any explosion is called *minimal* because it is active for the smallest possible time. Now let $Y_n = X_{T_n}$, $n \geq 0$ be the sequence of successive states taken by the process. The discrete-time process $(Y_n)_{n\geq 0}$ is called the *jump chain*.

We will be interested in processes $(X_t)_{t\geq 0}$ which possess the Markov property. It turns out that these can be totally determined by specifying the distributions of the holding times and jump chain.

4.2 Jump chain and holding times

Important information will be summarised by a special matrix.

Definition 4.1. *A* Q-matrix *or* generator *is a matrix $Q = (q_{ij})_{i,j\in\mathbb{S}}$ such that*

(i) $0 \le -q_{ii} < \infty$ *for all* $i \in \mathbb{S}$ *(negative diagonal entries)*

(ii) $q_{ij} \ge 0$ *for all* $i \ne j$ *(non-negative off-diagonal entries)*

(iii) $\sum_{j\in\mathbb{S}} q_{ij} = 0$ *(zero row sums).*

Write $q_i := -q_{ii}$ *and note that we also have* $q_i = \sum_{j\ne i} q_{ij}$.

Recall that a stochastic matrix has all non-negative entries and rows which sum to 1. We derive a stochastic matrix $\Pi = (\pi_{ij})_{i,j\in\mathbb{S}}$ from Q as follows:

$$\pi_{ij} := \begin{cases} q_{ij}/q_i & \text{if } j \ne i, q_i \ne 0 \\ 0 & \text{if } j = i, q_i \ne 0 \\ 0 & \text{if } j \ne i, q_i = 0 \\ 1 & \text{if } j = i, q_i = 0. \end{cases}$$

Informally, a continuous-time Markov chain is a process which, whenever it is in state $i \in \mathbb{S}$, waits an $\mathrm{Exp}(q_i)$ time and then jumps to a different state, chosen to be j with probability π_{ij}.

We will make use of the fact (proved on Problem Sheet 1) that if $E \sim \mathrm{Exp}(1)$ then $E/\lambda \sim \mathrm{Exp}(\lambda)$.

Definition 4.2 (Jump chain/holding time definition)**.** *A minimal right-continuous process* $(X_t)_{t\ge0}$ *is a continuous-time Markov chain with initial distribution* ν *and Q-matrix* Q *if*

- $(Y_n)_{n\ge0}$ *is a discrete-time Markov chain with initial distribution* ν *and transition matrix* Π*;*
- *conditional on* $Y_0 = i_0, Y_1 = i_1, \ldots, Y_{n-1} = i_{n-1}$*, the holding times* $Z_1, Z_2, \ldots, Z_n$ *are independent exponential random variables with parameters* $q_{i_0}, q_{i_1}, \ldots, q_{i_{n-1}}$ *respectively.*

We will write $X \sim \mathit{Markov}(\nu, Q)$.

We think of the initial distribution $\nu = (\nu_i)_{i\in\mathbb{S}}$ as a row vector. In the case where the chain starts in some fixed state $k \in \mathbb{S}$, we have $\nu = \delta_k$.

We can *construct* such a process by taking $(Y_n)_{n\ge0}$ to be a discrete-time Markov chain with initial distribution ν and transition matrix Π and taking $E_1, E_2, \ldots$ to be i.i.d. $\mathrm{Exp}(1)$ random variables. Then for $i \ge 1$ set $Z_i = E_i/q_{Y_{i-1}}$, $T_0 = 0$, $T_n = \sum_{i=1}^n Z_i$ for $n \ge 1$ and let

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

Then $(X_t)_{t\ge0}$ satisfies the conditions of the definition.

Remark 4.3. *When the process jumps, it never jumps to itself. If* $q_i = 0$ *for some* $i \in \mathbb{S}$ *then* i *is an absorbing state: if* X *ever hits* i*, it stays there forever. This is encoded in the jump-chain, exceptionally, as a "phantom jump" back to* i.

Example 4.4. *A simple birth process with birth rates $(\lambda_n)_{n\geq 0}$ started from k. We have $\nu_i = \delta_{ik}$, $q_{i\,i+1} = \lambda_i$, $i \in \mathbb{N}$, $q_{ij} = 0$ for $j \neq i, i+1$ i.e.*

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

The jump chain is deterministic: $Y_n = k + n$.

The memoryless property of the exponential distribution translates into the following key fact.

Proposition 4.5 (Competing exponentials)**.** *Let I be a finite or countably infinite index-set. Let $\{E_i : i \in I\}$ be independent random variables such that $E_i \sim Exp(\lambda_i)$ for $i \in I$, where $\lambda_i \geq 0$ and $\sum_{i\in I} \lambda_i < \infty$. Then*

$$M := \inf_{i\in I} E_i \sim Exp\left(\sum_{i\in I} \lambda_i\right)$$

and

$$\mathbb{P}\left(E_k < \inf_{i\neq k} E_i\right) = \frac{\lambda_k}{\sum_{i\in I}\lambda_i}.$$

It follows that the infimum is attained at a (random) index K such that

$$\mathbb{P}(K = k) = \frac{\lambda_k}{\sum_{i\in I}\lambda_i}, \quad k \in I.$$

Moreover, conditionally on $K = k$, the random variables $\{E_j - M : j \neq k\}$ are independent with $E_j - M \sim Exp(\lambda_j)$.

Proof. See Problem Sheet 1. □

Let $Y_0 \sim \nu$ and, for $i, j \in \mathbb{S}$ such that $i \neq j$, let $(N_{ij}(t))_{t\geq 0} \sim \mathrm{PP}(q_{ij})$, independently for distinct pairs i, j and independent of Y_0. Define $T_0 = 0$ and, inductively for $n \geq 0$,

$$T_{n+1} = \inf\{t > T_n : N_{Y_n j}(t) \neq N_{Y_n j}(T_n), \text{ for some } j \neq Y_n\}$$

and

$$Y_{n+1} = j \text{ if } T_{n+1} < \infty \text{ and } N_{Y_n j}(T_{n+1}) \neq N_{Y_n j}(T_n).$$

Then define

$$X_t = \begin{cases} Y_n & \text{if } T_n \leq t < T_{n+1} \text{ for some } n \geq 0 \\ \infty & \text{otherwise.} \end{cases}$$

Proposition 4.6. *We have that $(X_t)_{t\geq 0} \sim Markov(\nu, Q)$.*

Proof. We need to check that X has the correct jump chain, holding times and dependence structure.

Clearly, $X_0 = Y_0 \sim \nu$. Given $Y_0 = i$, the first jump occurs at the first time one of the Poisson processes N_{ij}, $j \neq i$ has a jump. But this is the minimum of independent exponentials with parameters $q_{i,j}, j \neq i$. Since $q_i = \sum_{j\neq i} q_{ij} < \infty$, we have $T_1 \sim \mathrm{Exp}(q_i)$. Moreover, the minimum is attained by the Poisson process N_{ij} with probability $q_{ij}/q_i = \pi_{ij}$.

By the memoryless property, $(N_{kl}(T_1+s)-N_{kl}(T_1))_{s\geq 0}$ is a new PP(q_{kl}) for each $k,l\in S$, $k\neq l$, independent of $\{N_{kl}(t), 0\leq t\leq T_1, k,l\in S\}$. Hence, the second holding time and jump are independent of the first. By induction, the same is true for subsequent holding times and jumps. $\square$

Remark 4.7. *This construction makes it clear that q_{ij} is the* rate *of going from i to j, $i\neq j$. Moreover, $q_i=-q_{ii}$ is the rate at which the chain leaves the state i.*

Proposition 4.8 (Markov property)**.** *Let $X\sim$ Markov(ν, Q) and let $t\geq 0$ be 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 Markov(δ_k, Q).*

Proof. This follows straightforwardly from the last construction and the Markov property of the Poisson processes involved. $\square$

4.3 Stopping times and the strong Markov property

The Markov property tells us that for a fixed time $t\geq 0$, conditional on $X_t=k$, the process after time t begins afresh from k. It is also very useful to be able to say something about the way the process evolves after certain *random* times.

Definition 4.9. *A random time T taking values in $[0,\infty]$ is a* stopping time *for a process $(X_t)_{t\geq 0}$ if the event $\{T\leq t\}$ depends only on $(X_s)_{0\leq s\leq t}$.*

Intuitively, we can tell from looking at the process up to time t whether T has occurred or not. In other words, if asked to stop at time T, you know when to stop.

Example 4.10. *(a) Let $T=\inf\{t\geq 0: X_t=k\}$ for some fixed k. Then $\{T\leq t\}=\{\exists 0\leq s\leq t: X_s=k\}$ and so T is a stopping time.*

(b) Let $\tilde{T}=\sup\{t\geq 0: X_t=k\}$ for some fixed k. In general, we cannot tell just from looking at $(X_s)_{0\leq s\leq t}$ whether we have hit k for the last time or not. So $\tilde{T}$ is not a stopping time.

(c) Let $T=\inf\{t\geq 0: X_t=10\}-1$ Then

$$\{T\leq t\}=\{\exists\, 0\leq s\leq t+1: X_s=k\},$$

which clearly depends on $(X_s)_{0\leq s\leq t+1}$. So T is not a stopping time.

Remark 4.11. *Note that a stopping time can take the value ∞. For example, if $(X_t)_{t\geq 0}$ is a continuous-time Markov chain, it is perfectly possible that for a fixed state $k\in\mathbb{S}$ we never hit k. In that case, the set $\{t\geq 0: X_t=k\}$ is empty, and so has infimum ∞.*

Consider Example 4.10 (c). We cannot expect the process started from T to look like a new copy of the original process because we now it *must* hit 10 at time $T+1$. On the other hand, for T a stopping time, a continuous-time Markov chain started from T *is* a new continuous-time Markov chain.

Theorem 4.12 (Strong Markov property)**.** *Let $X\sim$ Markov(ν, Q) and let T be a stopping time. Then for all $k\in\mathbb{S}$ such that $\mathbb{P}(X_T=k)>0$, given $T<\infty$ and $X_T=k$, we have that $(X_r)_{r\leq t}$ and $(X_{T+s})_{s\geq 0}$ are independent. Moreover, the conditional distribution of $(X_{T+s})_{s\geq 0}$ is Markov(δ_k, Q).*

The proof is beyond the scope of this course, but can be found in Section 6.5 of *Markov Chains* by Norris.

4.4 Transition semigroups

By the Markov property, $\mathbb{P}(X_{t+s}=j|X_t=i)$ does not depend on t. Write

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

and $P(s)=(p_{ij}(s))_{ij\in\mathbb{S}}$ as a matrix.

Example 4.13. *For a PP(λ),*

$$p_{i\ i+n}=\frac{(\lambda s)^n e^{-\lambda s}}{n!},\quad n\geq 0,\ i\geq 0.$$

Proposition 4.14. $(P(t))_{t\geq 0}$ *is a* semigroup *i.e.* $P(0)=I$ *and for all* $s,t\geq 0$, $P(t+s)=P(t)P(s)$.

Proof. $P(0)=I$ is obvious. For all $i,k\in\mathbb{S}$,

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

□

When we have a fixed initial state $X_0=i$, we will write $\mathbb{P}(\cdot|X_0=i)$ or $\mathbb{P}_i(\cdot)$.

4.5 The backward and forward equations

Theorem 4.15. *The transition matrices* $(P(t))_{t\geq 0}$ *of a minimal* (ν,Q)*-CTMC satisfy the* backward equation*:*

$$P'(t)=QP(t),\quad t\geq 0,\tag{2}$$

where we differentiate each entry of the matrix. Moreover, with the initial condition $P(0)=I$*, we have that* $P(t)$ *is the minimal non-negative solution to this system of differential equations i.e. any other solution* $(\tilde{P}(t))_{t\geq 0}$ *has*

$$\tilde{p}_{ij}(t)\geq p_{ij}(t),\quad \textit{for all } i,j\in\mathbb{S}.$$

Proof. Condition on the time of the first jump, T_1:

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

So

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

and differentiating gives

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

Cancelling the exponentials, using the fact that $q_i = -q_{ii}$ and rearranging gives (2).

Suppose now we have another non-negative solution $\tilde{p}_{ij}(t)$. Then reversing the last few steps we must have

$$\tilde{p}_{ik}(t) = \delta_{ik} e^{-q_i t} + \int_0^t \sum_{j \neq i} q_{ij} \tilde{p}_{jk}(u) e^{-q_i (t-u)} du. \tag{3}$$

Now note that $p_{ik})(t) = \lim_{n \to \infty} \mathbb{P}_i(X_t = k, t < T_n)$. We will compare $\tilde{p}_{ik}(t)$ and $\mathbb{P}_i(X_t = k, t < T_n)$. First since $T_0 = 0$ and $\tilde{p}_{ik}(t) \geq 0$, we have

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

We now proceed by induction. Suppose that for some $n \in \mathbb{N}$ and all $i, k \in \mathbb{S}$,

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

Then we have, by the same argument as before,

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

by (3). Hence, by induction, $\mathbb{P}_i(X_t = k, t < T_n) \leq \tilde{p}_{ik}(t)$ for all $n \in \mathbb{N}$. Taking the limit as $n \to \infty$, we obtain $p_{ik}(t) \leq \tilde{p}_{ik}(t)$ for all $i, k \in \mathbb{S}$. □

Remark 4.16. *The condition $\sum_{j \in \mathbb{S}} p_{ij}(t) = 1$ for all $i \in \mathbb{S}$ and all $t \geq 0$ is sufficient to give uniqueness of the solution to the backward equation. So in order to have non-uniqueness we must have $\sum_{j \in \mathbb{S}} p_{ij}(t) < 1$ for some i i.e. the chain must explode in the sense that $\mathbb{P}(T_\infty < \infty) > 0$.*

Theorem 4.17. *The transition matrices $(P(t))_{t \geq 0}$ of a minimal continuous-time Markov chain with initial distribution ν and Q-matrix Q satisfy the* forward equation*:*

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

Moreover, with the initial condition $P(0) = I$, $(P(t))_{t \geq 0}$ is the minimal non-negative solution to this system of equations.

Proof. For finite state-space: see Problem Sheet 2. The proof for infinite state-space is beyond the scope of the course; see Section 2.8 of *Markov Chains* by Norris. □

We will deal almost exclusively with non-explosive continuous-time Markov chains for which uniqueness of the solution to the backward and forward equations is guaranteed.

4.6 Matrix exponentials

Suppose, for the moment, that the state-space $\mathbb{S}$ is finite so that $P(t)$ is an $N \times N$ matrix for some N and each $t \geq 0$. Consider the backward equation,

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

and the forward equation,

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

both with initial condition $P(0) = I$. Recall that we also have the semigroup property: $P(t + s) = P(s)P(t)$. In view of all this, seems natural to want to write $P(t) = e^{tQ}$, although we are, of course, dealing with matrices and so have to be a bit careful. The correct way to define the right-hand side is via a series expansion:

$$e^{tQ} = \sum_{k=0}^{\infty} \frac{t^k}{k!} Q^k.$$

It can be shown (see Norris *Markov Chains* Section 2.10) that this series converges in an appropriate sense for all $t \geq 0$ and that, moreover, it satisfies the forward and backward equations (which have unique solutions for $\mathbb{S}$ finite). This is a very useful view-point but one which we can only easily make use of for finite $\mathbb{S}$. For infinite $\mathbb{S}$, the matrices are of infinite size and so it is much harder to make sense of such exponentials.

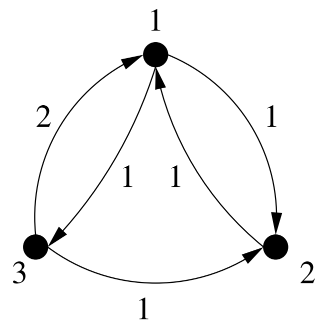
4.7 Finding transition probabilities in finite systems

In many cases, the forward and backward equations are not at all easy to solve. However, we have seen some cases where they are straightforward (for example, for a Poisson process.) In finite state-spaces, there is a reasonably general technique which you can employ. (The worked example which follows is adapted from Example 2.1.3 of *Markov Chains* by Norris.)

Consider the continuous-time Markov chain with Q-matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

To see what is going on, a diagram helps:



What is $p_{11}(t)$? Since we have a finite state-space, we can write $P(t) = e^{tQ} = \sum_{k=0}^{\infty} \frac{t^k Q^k}{k!}$. We can diagonalize Q as $Q = U \Lambda U^{-1}$, where Λ is a diagonal matrix,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Then $Q^k = U\Lambda U^{-1} U\Lambda U^{-1} \dots U\Lambda U^{-1} = U\Lambda^k U^{-1}$ and so

$$\begin{aligned}
P(t) &= \sum_{k=0}^{\infty} \frac{t^k}{k!} U\Lambda^k U^{-1} = U \left(\sum_{k=0}^{\infty} \frac{t^k}{k!} \Lambda^k \right) U^{-1} \\
&= U \begin{pmatrix} \sum_{k=0}^{\infty} \frac{(\lambda_1 t)^k}{k!} & 0 & 0 \\ 0 & \sum_{k=0}^{\infty} \frac{(\lambda_2 t)^k}{k!} & 0 \\ 0 & 0 & \sum_{k=0}^{\infty} \frac{(\lambda_3 t)^k}{k!} \end{pmatrix} U^{-1} = U \begin{pmatrix} e^{\lambda_1 t} & 0 & 0 \\ 0 & e^{\lambda_2 t} & 0 \\ 0 & 0 & e^{\lambda_3 t} \end{pmatrix} U^{-1}.
\end{aligned}$$

It follows that there exist constants α, β, γ such that

$$p_{11}(t) = \alpha e^{\lambda_1 t} + \beta e^{\lambda_2 t} + \gamma e^{\lambda_3 t}.$$

In our example, Q has eigenvalues $0, -2, -4$, so that $p_{11}(t) = \alpha + \beta e^{-2t} + \gamma e^{-4t}$. To find α, β, γ, note that $p_{11}(0) = 1$ and so

$$\alpha + \beta + \gamma = 1. \tag{4}$$

The backward equation gives $P'(0) = Q$ and so $p'_{11}(0) = -2$. Hence,

$$-2\beta - 4\gamma = -2. \tag{5}$$

Applying the backward equation twice gives $P''(0) = Q^2$ and so $p''_{11}(0) = 7$. Hence,

$$4\beta + 16\gamma = 7. \tag{6}$$

We have three equations in three unknowns and so can solve to obtain

$$p_{11}(t) = \frac{3}{8} + \frac{1}{4} e^{-2t} + \frac{3}{8} e^{-4t}.$$

5 Properties of continuous-time Markov chains

Many aspects of the behaviour of continuous-time Markov chains can be deduced from corresponding facts for the jump chain.

5.1 Class structure

Definition 5.1. *Let X be a continuous-time Markov chain.*

(a) *We say i* leads to *j 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) *$A \subseteq \mathbb{S}$ is a* communicating class *if for all $i, j \in A$, we have $i \leftrightarrow j$ and for all $k \in \mathbb{S} \setminus A$, at most one of $i \to k$ and $k \to i$ holds. Note that the communicating classes partition $\mathbb{S}$.*

(d) *A is a* closed class *if the chain cannot leave A i.e. there are no $i \in A, j \in \mathbb{S} \setminus A$ such that $i \to j$.*

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

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

Class structure is inherited from the jump chain.

Proposition 5.2. *Let X be a minimal continuous-time Markov chain with jump chain Y. For $i, j \in \mathbb{S}$, $i \neq j$, the following are equivalent:*

(i) $i \to j$ for X

(ii) $i \to j$ for Y

(iii) there exists $n \geq 1$ and state $i_0 = i, i_1, \ldots, i_{n-1}, i_n = j$ such that $\prod_{k=0}^{n-1} q_{i_k\, i_{k+1}} > 0$

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

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

Proof. The implications (iv) ⇒ (v) ⇒ (i) ⇒ (ii) are clear.

(ii) ⇒ (iii): In discrete time, $i \to j$ for Y implies

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

for some states $i_0 = i, i_1, \ldots, i_{n-1}, i_n = j$, and so

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

since $q_i = 0$ if and only if $\pi_{ii} = 1$.

(iii) ⇒ (iv): If $q_{ij} > 0$ then

$$\begin{aligned} p_{ij}(t) &\geq \mathbb{P}_i(Z_1 \leq t, Z_2 > t, Y_1 = j) \\ &= \mathbb{P}_i(Z_1 \leq t)\mathbb{P}_i(Y_1 = j)\mathbb{P}\left(Z_2 > t | Y_1 = j\right) \\ &= (1 - e^{-q_i t})\pi_{ij} e^{-q_j t} > 0 \end{aligned} \tag{7}$$

for all $t > 0$. We do not necessarily have $q_{ij} > 0$. But if $q_{ij} = 0$, for the path $(i_0, i_1, \ldots, i_n)$ given by (iii) we have $q_{i_k\, i_{k+1}} > 0$ for all $0 \leq k \leq n-1$ and then

$$p_{ij}(t) \geq \mathbb{P}_i(X_{t/n} = i_1, X_{2t/n} = i_2, \ldots, X_t = i_n) = \prod_{k=0}^{n-1} p_{i_k\, i_{k+1}}(t/n) > 0$$

for all $t > 0$ by (7). □

Remark 5.3. *It is not possible to have periodic behaviour in a continuous-time Markov chain, even if the underlying jump-chain is periodic.*

5.2 Hitting probabilities

Suppose that $(X_t)_{t\geq 0}$ is a continuous-time Markov chain with Q-matrix Q and let $(Y_n)_{n\geq 0}$ be its jump-chain (with transition matrix Π). Define the *first hitting time* of a set $A \subseteq S$ by

$$\tau_A^X := \inf\{t \geq 0 : X_t \in A\}.$$

This is a stopping time. Also let

$$\tau_A^Y := \inf\{n \geq 0 : Y_n \in A\},$$

the equivalent quantity for the jump-chain. Then

$$\{\tau_A^X < \infty\} = \{\tau_A^Y < \infty\}$$

and so the *hitting probability* $h_i^A := \mathbb{P}_i(\tau_A^X < \infty)$ is equal to $\mathbb{P}_i(\tau_A^Y < \infty)$ i.e. hitting probabilities are the same for the jump-chain and the original chain. If A is a closed class, h_i^A is called the *absorption probability.*

We know (from Part A Probability) that the hitting probabilities $(h_i^A, i \in S)$ of the set A are the *minimal non-negative solution* to the following system of equations:

$$\begin{cases} h_i^A = 1 & \text{if } i \in A \\ h_i^A = \sum_{j \in S} \pi_{ij} h_j^A & \text{if } i \notin A. \end{cases}$$

(See your Part A notes, or Theorem 1.3.2 of Norris. The proof, once again, essentially involves conditioning on which state we jump to first.) Using the definition of Π in terms of entries of Q, we see that this can be re-phrased directly as the minimal non-negative solution to

$$\begin{cases} h_i^A = 1 & \text{if } i \in A \\ \sum_{j \in S} q_{ij} h_j^A = 0 & \text{if } i \notin A. \end{cases}$$

5.3 Recurrence and transience

Recall that for a discrete-time Markov chain, recurrence of a state i means that we come back infinitely often to i, and transience means we eventually leave i forever.

Definition 5.4. *Let X be a continuous-time Markov chain. We say that $\{t \geq 0 : X_t = i\}$ is* bounded *if there exists M such that $X_t \neq i$ for all $t > M$ and* unbounded *otherwise.*

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

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

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

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

Note that if X can explode started from i and X is minimal then i must be transient.

Let $H_i = \inf\{t \geq T_1 : X_t = i\}$ be the *first passage time to i* (note that we force the chain to make at least one jump, so that if $X_0 = i$, X must leave and then come back.)

Proposition 5.5. *$i \in \mathbb{S}$ is recurrent (transient) for a minimal continuous-time Markov chain X iff it is recurrent (transient) for the jump chain Y.*

Proof. Suppose i is transient for Y. Then if $X_0 = i$, $N = \sup\{n \geq 0 : Y_n = i\} < \infty$. So

$$\{t \geq 0 : X_t = i\} \subseteq [0, T_{N+1}),$$

which is finite since T_{N+1} is a sum of finitely many exponential random variables.

Suppose now that i is recurrent for Y. Then if $X_0 = i$, there exists an infinite sequence $N_1 \le N_2 \le \ldots$ of times such that $Y_{N_k} = i$. Then the time spent at i by X is bounded below by

$$\sum_{k=1}^{\infty} Z_{N_k},$$

where $Z_{N_k} \sim \text{Exp}(q_i)$ for all $k \ge 1$. But we saw in the proof of the explosion criterion for a birth process that a sum $\sum_{k=1}^{\infty} E_k$ where $E_1, E_2, \ldots$ are independent and $E_k \sim \text{Exp}(\lambda_k)$ is finite iff $\sum_{k=1}^{\infty} 1/\lambda_k < \infty$. Here, $\lambda_k = q_i$ for all $k \ge 1$ and so $\sum_{k=1}^{\infty} Z_{N_k} = \infty$ with probability 1. Hence, X spends an unbounded amount of time at i.

Since i must be either recurrent or transient for Y, the result follows. □

Corollary 5.6. *Every state $i \in \mathbb{S}$ is either recurrent or transient for X. Moreover, recurrence and transience are class properties.*

Proof. This follows immediately from the corresponding results for the jump chain. □

Recall that for a discrete-time Markov chain Y with transition matrix Π and first passage time

$$H_i^Y = \inf\{n \ge 1 : Y_n = i\},$$

i is recurrent iff $\mathbb{P}_i\left(H_i^Y < \infty\right) = 1$ iff $\sum_{i=0}^{\infty} \pi_{ii}^{(n)} = \infty$.

Theorem 5.7. *For any state $i \in \mathbb{S}$, the following are equivalent:*

(i) i *is recurrent*

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

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

Proof. If $q_i = 0$, X cannot leave i and so i is recurrent. Also in that case $p_{ii}(t) = 1$ for all $t > 0$ and so $\int_0^\infty p_{ii}(t)dt = \infty$.

Now suppose that $q_i > 0$. Then i is recurrent iff it is recurrent for the jump chain, which is equivalent to

$$\mathbb{P}_i\left(H_i^Y < \infty\right) = 1 \quad \text{and} \quad \sum_{n=0}^{\infty} \pi_{ii}^{(n)} = \infty.$$

Now $\mathbb{P}_i(H_i < \infty) = \mathbb{P}_i\left(H_i^Y < \infty\right)$. Moreover,

$$\begin{aligned}
\int_0^\infty p_{ii}(t)dt &= \int_0^\infty \mathbb{P}_i(X_t = i)\, dt \\
&= \int_0^\infty \mathbb{E}_i\left[\mathbb{1}_{\{X_t = i\}}\right] dt \\
&= \mathbb{E}_i\left[\int_0^\infty \mathbb{1}_{\{X_t = i\}} dt\right] \quad \text{by Tonelli's theorem} \\
&= \mathbb{E}_i\left[\sum_{n=0}^{\infty} Z_{n+1} \mathbb{1}_{\{Y_n = i\}}\right] \\
&= \sum_{n=0}^{\infty} \mathbb{E}_i\left[Z_{n+1} | Y_n = i\right] \mathbb{P}_i(Y_n = i) \quad \text{by Tonelli's theorem} \\
&= \frac{1}{q_i} \sum_{n=0}^{\infty} \pi_{ii}^{(n)}.
\end{aligned}$$

The result follows, since $q_i > 0$. □

5.4 Examples

A birth-and-death process

Consider a population in which each individual gives birth after an Exp(λ) time, independently and repeatedly, and has a lifetime which is distributed as Exp(μ), independently of the births and of the other individuals. Let $X_0 = 1$ and let X_t be the number of individuals in the population at time t. Because everything is built out of competing exponentials, $(X_t)_{t\geq 0}$ evolves as a continuous-time Markov chain with state-space $\mathbb{N}$ and Q-matrix

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \cdots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \cdots \\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & 0 & \cdots \\ 0 & 0 & 3\mu & -3(\lambda+\mu) & 3\lambda & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{pmatrix}.$$

There are two communicating classes: $\{0\}$ which is absorbing, and $\{1, 2, \ldots\}$ which is open. So the chain is clearly transient. The jump chain is a simple random walk with up probability $\lambda/(\lambda+\mu)$ and down probability $\mu/(\lambda+\mu)$, absorbed at 0. So the question of absorption at 0 is precisely the gambler's ruin problem. From Part A Probability, we know that the probability of absorption at 0 started from 1 is 1 if $\lambda \leq \mu$ and μ/λ if $\lambda > \mu$.

Suppose we want to know the total number of individuals that are ever born. (This will be finite if $\lambda \leq \mu$ and may be infinite if $\lambda > \mu$.) Let N be the number of children of the initial individual. This individual has a lifetime $L \sim$ Exp(μ) and, given L, we have $N \sim$ Po(λL). So for $n \geq 0$,

$$\begin{aligned} \mathbb{P}(N = n) = \int_0^\infty \mu e^{-\mu x} \mathbb{P}(\text{Po}(\lambda x) = n)\, dx &= \int_0^\infty \mu e^{-\mu x} \frac{e^{-\lambda x}(\lambda x)^n}{n!} dx \\ &= \lambda^n \mu \int_0^\infty \frac{x^n}{n!} e^{-(\lambda+\mu)x} dx = \left(\frac{\lambda}{\lambda+\mu}\right)^n \frac{\mu}{\lambda+\mu}. \end{aligned}$$

So $N \sim$ Geometric($\mu/(\lambda+\mu)$). Now observe that each of these children itself has an independent number of children with the same distribution, and so on. In other words, if we think in terms of genealogy, we have a branching process. The offspring distribution is the distribution of N. Write $G_N(s) = \mathbb{E}\left[s^N\right]$ for its probability generating function, and note that

$$G_N(s) = \frac{\mu}{\lambda + \mu - \lambda s}.$$

Let Z be the total number of individuals who are ever born. Then note that Z must have the same distribution as $1 + \sum_{i=1}^N Z_i$, where $Z_1, Z_2, \ldots$ are i.i.d. copies of Z, since we start with a single individual, and each of its N children is the original progenitor of a new independent branching process with the same distribution. So the probability generating function G_Z of Z must satisfy

$$G_Z(s) = \mathbb{E}\left[s^{1+\sum_{i=1}^N Z_i}\right] = s\mathbb{E}\left[G_Z(s)^N\right] = sG_N(G_Z(s)) = \frac{\mu s}{\lambda + \mu - \lambda G_Z(s)}.$$

Rearranging gives

$$\lambda G_Z(s)^2 - (\lambda+\mu) G_Z(s) + \mu s = 0$$

with possible solutions

$$\frac{\lambda + \mu \pm \sqrt{(\lambda+\mu)^2 - 4\lambda\mu s}}{2\lambda}.$$

Since $G_Z(s)$ is an increasing function of s, we must take the $-$ root, to obtain

$$G_Z(s) = \frac{\lambda + \mu - \sqrt{(\lambda+\mu)^2 - 4\lambda\mu s}}{2\lambda}.$$

Expanding the series, for all λ and μ we obtain

$$\mathbb{P}(Z = n) = \frac{1}{2(2n-1)} \binom{2n}{n} \left(\frac{\lambda}{\lambda+\mu}\right)^{n-1} \left(\frac{\mu}{\lambda+\mu}\right)^{n}, \quad n \geq 1, \tag{8}$$

and if $\lambda > \mu$ we also have $\mathbb{P}(Z = \infty) = 1 - G_Z(1) = 1 - \mu/\lambda$.

You can also check that the extinction probability for the branching process, which is the minimal non-negative solution to the equation $s = G_N(s)$ is, in this case, the minimal non-negative solution to $(\lambda s - \mu)(s-1) = 0$, which is indeed 1 if $\lambda \geq \mu$ and μ/λ if $\lambda > \mu$.

The M/M/1 queue

Suppose customers arrive at a bank according to a Poisson process of rate λ. There is a single server and each customer is served for a length of time distributed as $\mathrm{Exp}(\mu)$, independently for different customers. If someone is already being served when a customer arrives, they join the back of the queue and wait their turn.

Let X_t be the number of people in the queue at time t, including the person being served. Since the holding times are constructed out of competing exponential random variables, $(X_t)_{t\geq 0}$ is a continuous-time Markov chain (indeed, a birth-and-death process) with Q-matrix

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

The state-space is clearly irreducible. We will investigate recurrence and transience on Problem Sheet 2.

6 Application: a stochastic epidemic

Suppose we want to model the spread of a disease in a population of size N. We will consider an idealised model in which individuals can have one of three states: susceptible (S), infected (and infectious) (I) or recovered (R). We make the following assumptions:

- only susceptible individuals may become infected;
- after having been infectious for some time, an individual recovers and becomes immune, or dies.

In particular, an individual may only make two moves: from S to I or from I to R. For this reason, the model we now describe is often known as an *SIR model*.

We assume that all individuals come into close contact randomly and independently at a common rate λ, whether or not they are infected. Close contact between an infected individual and a susceptible individual results in the susceptible individual becoming infected. Individuals remain infectious for an $\mathrm{Exp}(\gamma)$ amount of time before recovering, independently of other individuals. In particular, we can model the dynamics of the epidemic as a continuous-time Markov chain $(S_t, I_t, R_t)_{t\geq 0}$ with (finite) state-space $\{(s,i,r) : s,i,r \geq 0, s+i+r = N\}$. We take the initial state to be $(N-m, m, 0)$ for some $1 \le m \le N$ (i.e. there are m infected initially individuals and everyone else is susceptible) and transition rates

$$q_{(s,i,r)\,(s-1,i+1,r)} = \lambda s i,$$
$$q_{(s,i,r)\,(s,i-1,r+1)} = \gamma i,$$

with all other off-diagonal entries of the Q-matrix taken to be 0.

Obviously this model is unrealistic in several ways, but we can already learn something from it. More sophisticated versions of it are used in practice.

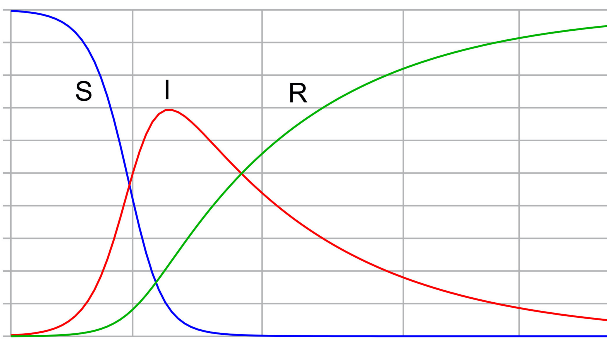
It is clear that the Markov chain is transient. We let $T = \inf\{t \ge 0 : I_t = 0\}$ be the absorption time. A quantity of great interest is then the terminal state R_T; in other words, how many people were ever infected in the course of the epidemic?

6.1 A deterministic approximation

Usually we are interested in large population size N. Let's take $I_0 = \lfloor N\epsilon \rfloor$ for some $\epsilon \in (0,1)$. It's reasonable to suppose that $\lambda = \beta/N$. It turns out that, in that case, we have that the proportions $(S_t/N, I_t/N, R_t/N)_{t\ge 0}$ behave approximately like the solution $(s(t), i(t), r(t))_{t\ge 0}$ to the following system of differential equations:

$$s'(t) = -\beta s(t) i(t)$$
$$i'(t) = (\beta s(t) - \gamma) i(t)$$
$$r'(t) = \gamma i(t),$$

with $s(0) = 1-\epsilon$, $i(0) = \epsilon$ and $r(0) = 0$. (Note that we do not, in fact, need to track the proportion $r(t)$ of recovered individuals separately, since $r(t) = 1 - s(t) - i(t)$ at time $t \ge 0$.) We will not prove this here, but instead try to get an idea of its consequences. It is straightforward to see that $s(t)$ is monotone decreasing to some value $s(\infty)$ and $r(t)$ is monotone increasing to some value $r(\infty)$. If $\beta(1-\epsilon) > \gamma$ then $i(t)$ initially increases and then eventually decreases to $i(\infty) = 0$.



The SIR model with $s(0) = 0.997, i(0) = 0.03$, $\beta = 0.4$ and $\gamma = 0.04$. Picture by Klaus-Dieter Keller, CC0, https://commons.wikimedia.org/w/index.php?curid=77633956

On the other hand, if $\beta(1-\epsilon) < \gamma$ then $i(t)$ is simply decreasing to 0, and the epidemic never "takes off". If we take $i(0) = \epsilon$ close to 0, as is natural, we see that the critical value separating the two scenarios is $\rho := \beta/\gamma = 1$. The quantity ρ (often called "R_0") is called the *basic reproduction number*, and is interpreted as the average number of new infections caused by a single infectious individual. When $\rho > 1$, the epidemic takes off and affects a large number of people; if $\rho < 1$, the epidemic remains relatively small.

In either case, R_T/N should look approximately like $r(\infty)$. Dividing the first differential equation by the third, we see that

$$\frac{ds}{dr} = -\rho s,$$

which implies that $s(t) = (1-\epsilon)e^{-\rho r(t)}$. Since $s(\infty) = 1 - r(\infty)$, we see that $r(\infty)$ solves

$$1 - r(\infty) = (1-\epsilon)e^{-\rho r(\infty)}.$$

6.2 Stochastic approximation by a birth-and-death process

Deterministic approximations are, however, not sufficient to capture all the possible behaviours of interest. Suppose we start with a single infectious individual. Then, even if ρ is much larger than 1, it is clearly possible that the epidemic dies out quickly.

Let us return to our original Markov chain model, and consider what happens close to the start if $I_0 = 1$, $\lambda = \beta/N$ and N is large. It is difficult to make precise distributional computations, but it turns out that we can make a useful *stochastic* approximation. (Let us now ignore the recovered individuals, since we know we can deduce their number from $R_t = N - S_t - I_t$.) Then

$$\begin{aligned} q_{(s,i)\,(s-1,i+1)} &= \beta\frac{s}{N}i \\ q_{(s,i)\,(s,i-1)} &= \gamma i. \end{aligned}$$

As long as $\frac{S_t}{N} \approx 1$ then these are approximately the transition rates of a birth-and-death process; indeed, the down-rate is the same, and the up-rate is *bounded above* by the up-rate of the birth-and-death process. We can use this to make a comparison between the two. Let us define a new continuous-time Markov chain $(S_t, I_t, G_t)_{t\geq 0}$ with transition rates

$$\begin{aligned} q_{(s,i,g)\,(s-1,i+1,g)} &= \beta\frac{s}{N}i \\ q_{(s,i,g)\,(s,i,g+1)} &= \beta\left(1-\frac{s}{N}\right)i + \beta g \\ q_{(s,i,g)\,(s,i-1,g)} &= \gamma i \\ q_{(s,i,g)\,(s,i,g-1)} &= \gamma g. \end{aligned}$$

Start from $S_0 = N-1, I_0 = 1, G_0 = 0$. The quantity $(G_t)_{t\geq 0}$ doesn't have any meaning in the epidemic model – we can think of it simply as an immigration of ghost individuals into the population (at a slightly complicated rate), each of which thereafter reproduces at rate β and dies at rate γ, without interacting with the other individuals. It's straightforward to check that $(S_t, I_t)_{t\geq 0}$ is still evolving according to the SIR model, that $(G_t)_{t\geq 0}$ is non-negative, and that $(I_t + G_t)_{t\geq 0}$ evolves exactly like the birth-and-death process in Section 5.4, where an individual gives birth at rate β and dies at rate γ. In particular, the absorption time $T = \inf\{t \geq 0 : I_t = 0\}$ for the epidemic is always smaller than the absorption time $T' = \inf\{t \geq 0 : I_t + G_t = 0\}$ for the birth-and-death process. The distribution of T' is explicit (see Problem Sheet 2). Moreover, R_T is bounded above by the total number of Z individuals that are ever born in the birth-and-death process, whose distribution we calculated at (8).

7 Convergence to equilibrium for continuous-time Markov chains

To understand equilibrium behaviour, we must consider the communicating classes of a continuous-time Markov chain separately. So without loss of generality we will restrict attention here to the case of irreducible chains.

7.1 Invariant distributions

Note that if $X_0 \sim \nu$,

$$\mathbb{P}(X_t = j) = \sum_{i \in \mathbb{S}} \nu_i p_{ij}(t) = (\nu P(t))_j.$$

Definition 7.1. *A distribution ξ on $\mathbb{S}$ is* invariant *for a continuous-time Markov chain if*

$$\xi P(t) = \xi \quad \text{for all } t \geq 0.$$

If we take $X_0 \sim \xi$ then $X_t \sim \xi$ for all $t \geq 0$ and we say that X is in equilibrium.

Theorem 7.2. *Suppose that Q is a Q-matrix and $(P(t))_{t \geq 0}$ are the transition matrices of the associated minimal continuous-time Markov chain. Then ξ is invariant iff $\xi Q = 0$.*

Proof. We give the proof for finite $\mathbb{S}$.

Suppose first $\xi P(t) = \xi$ for all $t \geq 0$. Then

$$\begin{aligned}
\xi Q = \xi P(t) Q = \xi P'(t) \quad &\text{by the forward equation} \\
= \xi \lim_{h \to 0} \frac{P(t+h) - P(t)}{h}& \\
= \lim_{h \to 0} \frac{\xi P(t+h) - \xi P(t)}{h}& = 0.
\end{aligned}$$

If $\xi Q = 0$, we have

$$\begin{aligned}
\xi P(t) &= \xi P(0) + \xi \int_0^t P'(s) ds \\
&= \xi + \int_0^t \xi Q P(s) ds \quad \text{by the backward equation} \\
&= \xi.
\end{aligned}$$

(We used the finiteness of $\mathbb{S}$ when interchanging limits/integrals and matrix multiplication. The argument for countably infinite $\mathbb{S}$ is more complicated and beyond the scope of this course.) ☐

Definition 7.3. *Recall that that $H_i = \inf\{t \geq T_1 : X_t = i\}$ is the first passage time to i. A state $i \in \mathbb{S}$ is* positive recurrent *if either $q_i = 0$ or $m_i = \mathbb{E}_i[H_i] < \infty$. Otherwise, a recurrent state i is* null recurrent.

Theorem 7.4. *Let Q be an irreducible and non-explosive Q-matrix. The following are equivalent:*

(i) every state is positive recurrent

(ii) some state i is positive recurrent

(iii) Q has an invariant distribution ξ which satisfies $\xi_i = \frac{1}{m_i q_i}$, $i \in \mathbb{S}$.

Proof omitted (see Norris Theorem 3.5.3).

7.2 Convergence to equilibrium

This is of central importance in applications.

Theorem 7.5. *Let $X = (X_t)_{t\geq 0} \sim$ Markov(ν, Q) be a minimal irreducible positive recurrent continuous-time Markov chain and let ξ be an invariant distribution. Then*

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

Proof. Let $X^{(1)} \sim \text{Markov}(\nu, Q)$ and, independently, let $X^{(2)} \sim \text{Markov}(\xi, Q)$, so that $X_t^{(2)} \sim \xi$ for all $t \geq 0$. Note that $(X_t^{(1)}, X_t^{(2)})_{t\geq 0}$ is a continuous-time Markov chain on the state-space $\mathbb{S} \times \mathbb{S}$. Let $T = \inf\{t \geq 0 : X_t^{(1)} = X_t^{(2)}\}$. The bivariate chain has invariant distribution $\eta_{(i,j)} = \xi_i \xi_j$ for $(i, j) \in \mathbb{S} \times \mathbb{S}$. Thus it is positive recurrent, and in particular recurrent, which implies that $\mathbb{P}(T < \infty) = 1$. Now set

$$X_t = \begin{cases} X_t^{(1)} & t < T \\ X_t^{(2)} & t \geq T. \end{cases}$$

Then by the strong Markov property, $(X_t)_{t\geq 0} \sim \text{Markov}(\nu, Q)$. But then

$$\begin{aligned} \mathbb{P}(X_t = j) &= \mathbb{P}\left(X_t^{(1)} = j, T > t\right) + \mathbb{P}\left(X_t^{(2)} = j, T \leq t\right) \\ &= \mathbb{P}\left(X_t^{(1)} = j, T > t\right) + \xi_j \mathbb{P}(T \leq t). \end{aligned}$$

Since $\mathbb{P}(T < \infty) = 1$ we get $\mathbb{P}(T > t) \to 0$ as $t \to \infty$ and so the right-hand side converges to ξ_j, as desired. □

Remark 7.6. *It follows that the invariant distribution is unique.*

It is also the case that the long-run average proportion of time we spend in a state i converges *almost surely* to ξ_i.

Theorem 7.7 (Ergodic theorem)**.** *Let $X = (X_t)_{t\geq 0} \sim$ Markov(ν, Q) be a minimal irreducible positive recurrent continuous-time Markov chain and let ξ be an invariant distribution. Then*

$$\frac{1}{t} \int_0^t \mathbb{1}_{\{X_s = i\}} ds \to \xi_i \quad \text{almost surely}$$

as $t \to \infty$.

Proof. We will prove a more general result for renewal processes later, and deduce this from it. □

As a consequence, we can *estimate* the invariant distribution (in the statistical sense) by looking at the proportion of time spent in each state over a long period of time.

7.3 Reversibility and detailed balance

Let us consider for the moment an irreducible *discrete-time* Markov chain $Y = (Y_n)_{n\geq 0}$ with transition matrix Π and invariant distribution μ, started in equilibrium. Fix n and define

$$\hat{Y}_m := Y_{n-m}.$$

We call $(\hat{Y}_m)_{0\leq m\leq n}$ the *time-reversal* of $(Y_m)_{0\leq m\leq n}$.

Let $\hat{\pi}_{ij} := \mu_j \pi_{ji} / \mu_i$ and set $\hat{\Pi} = (\hat{\pi}_{ij})_{i,j\in\mathbb{S}}$.

Proposition 7.8. *$(\hat{Y}_m)_{0\le m\le n}$ is a discrete-time Markov chain with initial distribution μ and transition matrix $\hat{\Pi}$. Moreover, μ is invariant for $\hat{\Pi}$.*

Proof. First note that

$$\sum_{j\in\mathbb{S}} \hat{\pi}_{ij} = \sum_{j\in\mathbb{S}} \mu_j\pi_{ji}/\mu_i = \mu_i/\mu_i = 1$$

by stationarity of μ for Π. So $\hat{\Pi}$ is a stochastic matrix.

For $i_0, i_1, \dots, i_n \in \mathbb{S}$,

$$\begin{aligned}\mathbb{P}\left(\hat{Y}_0 = i_0, \hat{Y}_1 = i_1, \dots, \hat{Y}_n = i_n\right) &= \mathbb{P}\left(Y_0 = i_n, Y_1 = i_{n-1}, \dots, Y_n = i_0\right) \\ &= \mu_{i_n}\pi_{i_n i_{n-1}} \dots \pi_{i_1 i_0} \\ &= \mu_{i_0}\hat{\pi}_{i_0 i_1} \dots \hat{\pi}_{i_{n-1} i_n}.\end{aligned}$$

Finally, we have

$$\sum_{i\in\mathbb{S}} \mu_i\hat{\pi}_{ij} = \sum_{i\in\mathbb{S}} \mu_j\pi_{ji} = \mu_j$$

and so μ is indeed invariant for $\hat{\Pi}$. □

This result means that we can, in fact, make sense of an *eternal* stationary version of the Markov chain with time indexed by all of $\mathbb{Z}$: take $Y_0 \sim \mu$, then let $(Y_n)_{n\ge 0}$ be the chain run forwards in time with transition matrix Π and let $(Y_n)_{n\le 0}$ be the chain run backwards in time with transition matrix $\hat{\Pi}$. Putting these together gives $(Y_n)_{n\in\mathbb{Z}}$ which is such that $Y_n \sim \mu$ for all $n \in \mathbb{Z}$, $\mathbb{P}(Y_{n+1} = j | Y_n = i) = \pi_{ij}$ for every $n \in \mathbb{Z}$ and $\mathbb{P}(Y_{n-1} = j | Y_n = i) = \hat{\pi}_{ij}$ for every $n \in \mathbb{Z}$.

If it happens that $\hat{\Pi} = \Pi$ then we say that Y is *reversible*, since then the time-reversal has the same distribution as the forward chain. This is the case if and only if

$$\mu_i\pi_{ij} = \mu_j\pi_{ji} \quad \text{for all } i, j \in \mathbb{S}. \tag{9}$$

The equations (9) are known as the *detailed balance equations* and if they hold we say Π and μ are *in detailed balance.*

Lemma 7.9. *Suppose that μ and Π are in detailed balance. Then μ is invariant for Π.*

Proof. Summing (9) in i gives

$$\sum_{i\in\mathbb{S}} \mu_i\pi_{ij} = \mu_j \sum_{i\in\mathbb{S}} \pi_{ji} = \mu_j,$$

as required. □

This is a very useful result because it is often easier to solve the detailed balance equations for μ than $\mu = \mu\Pi$. However, it is perfectly possible that an invariant distribution exists (for example, if the chain has a finite state-space and is irreducible we know that this must be the case) but that there is no solution to the detailed balance equations.

There are some situations where it is easy to exclude the possibility of reversibility, and in those cases trying to solve the detailed balance equations to find an invariant distribution is clearly a bad strategy!

Example: random walk on a path with reflecting barriers. Fix $p \in (0, 1)$. Suppose we have a random walk on $\{0, 1, \dots, N\}$ such that $\pi_{i\,i+1} = p$ and $\pi_{i\,i-1} = 1 - p$, $1 \le i \le N - 1$, $\pi_{01} = p$, $\pi_{00} = 1 - p$, $\pi_{NN-1} = 1 - p$, $\pi_{NN} = p$. Then the detailed balance equations are

$$\mu_i\pi_{i\ i+1} = \mu_{i+1}\pi_{i+1\ i} \quad \text{for } 0 \le i \le N - 1$$

i.e.

$$\mu_{i+1} = \frac{p}{1-p}\mu_i \quad \text{for } 0 \le i \le N-1.$$

Then $\mu_i = \left(\frac{p}{1-p}\right)^i \mu_0$ for $0 \le i \le N$ solves the detailed balance equations, and taking $\mu_0 = \frac{(1-2p)(1-p)^N}{(1-p)^{N+1}-p^{N+1}}$ gives a distribution. So the chain is reversible.

Compare with the more complicated system of equations $\mu\Pi = \mu$:

$$\begin{aligned}
\mu_0(1-p) + \mu_1(1-p) &= \mu_0 \\
\mu_0 p + \mu_2(1-p) &= \mu_1 \\
\mu_1 p + \mu_3(1-p) &= \mu_2 \\
&\vdots \\
\mu_{N-2} p + \mu_N(1-p) &= \mu_{N-1} \\
\mu_{N-1} p + \mu_N(1-p) &= \mu_N.
\end{aligned}$$

Example: the frog on an infinite ladder (from Part A Probability Problem Sheet 4). A frog jumps on an infinite ladder. At each jump, with probability $1-p$ he jumps up one step, while with probability p he slips off and falls all the way to the bottom. The stationary distribution is $\mu_i = p(1-p)^i$, $i \ge 0$. Now note that from state 0, the frog can only move to 0 or to 1 (we have $\pi_{00} = p$ and $\pi_{01} = 1-p$). However, running backwards in time, the frog can jump from 0 to any element of $\mathbb{N}$. So clearly the chain cannot be reversible: we can tell the direction of time by observing the behaviour of the chain. You can indeed check that the detailed balance equations do not have a solution.

It will probably not surprise you to learn that there are continuous-time analogues of these ideas. Let $X = (X_t)_{t\ge0}$ be a continuous-time Markov chain. Fix a time t and define the *time-reversal* $\hat{X} = (\hat{X}_s)_{0\le s\le t}$ by

$$\hat{X}_s = X_{(t-s)-} := \lim_{r\uparrow t-s} X_r.$$

We take the value *just before* $t-s$ in order to make $\hat{X}$ right-continuous.

Theorem 7.10. *Let X be an irreducible positive-recurrent minimal continuous-time Markov chain with Q-matrix Q started from its invariant distribution ξ. Then $\hat{X} = (\hat{X}_s)_{0\le s\le t}$ is a continuous-time Markov chain with Q-matrix $\hat{Q}$ such that*

$$\hat{q}_{ij} = \xi_j q_{ji}/\xi_i, \quad i, j \in \mathbb{S}.$$

Proof. Let us first check that $\hat{Q}$ is a Q-matrix: it has non-negative off-diagonal entries and for $i \in \mathbb{S}$, by the invariance of ξ,

$$\sum_{j\in\mathbb{S}} \hat{q}_{ij} = \sum_{j\in\mathbb{S}} \frac{\xi_j q_{ji}}{\xi_i} = \frac{1}{\xi_i}\sum_{j\in\mathbb{S}} \xi_j q_{ji} = 0.$$

Now $(P(t))_{t\ge0}$ is the minimal non-negative solution to the forward equation

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

Define $\hat{P}(t)$ by

$$\xi_i \hat{p}_{ij}(t) = \xi_j p_{ji}(t).$$

It is straightforward to check that $(\hat{P}(t))_{t\geq 0}$ are transition matrices. Moreover,

$$\begin{aligned}\hat{p}'_{ij}(t) = \frac{\xi_j}{\xi_i} p'_{ji}(t) &= \frac{\xi_j}{\xi_i} \sum_{k\in\mathbb{S}} p_{jk}(t) q_{ki} \quad \text{by the forward equation} \\ &= \frac{\xi_j}{\xi_i} \sum_{k\in\mathbb{S}} \hat{p}_{kj}(t) \frac{\xi_k}{\xi_j} \hat{q}_{ik} \frac{\xi_i}{\xi_k} = \sum_{k\in\mathbb{S}} \hat{q}_{ik} \hat{p}_{kj}(t),\end{aligned}$$

i.e. $\hat{P}'(t) = \hat{Q}\hat{P}(t)$. Hence $(\hat{P}(t))_{t\geq 0}$ solves the backward equation for $\hat{Q}$. (It is clear that $\hat{P}(0) = I$.) It remains to prove that $(\hat{P}(t))_{t\geq 0}$ are the transition matrices of $\hat{X}$. For $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq t$,

$$\begin{aligned}\mathbb{P}\left(\hat{X}_{t_1} = i_1, \ldots, \hat{X}_{t_n} = i_n\right) &= \mathbb{P}\left(X_{(t-t_1)-} = i_1, \ldots, X_{(t-t_n)-} = i_n\right) \\ &= \xi_{i_n} p_{i_n\, i_{n-1}}(t_n - t_{n-1}) \ldots p_{i_2\, i_1}(t_2 - t_1) \\ &= \xi_{i_0} \hat{p}_{i_1\, i_2}(t_2 - t_1) \ldots \hat{p}_{i_{n-1}\, i_n}(t_n - t_{n-1}).\end{aligned}$$

(For the second equality, we used that the transition probabilities are continuous.) So the finite-dimensional distributions are correct and hence $\hat{X} \sim \text{Markov}(\xi, \hat{Q})$. □

Again, we can define an eternal stationary version of the chain, $(X_t)_{t\in\mathbb{R}}$.

If $\hat{Q} = Q$ then $(\hat{X}_s)_{0\leq s\leq t}$ has the same distribution as $(X_s)_{0\leq s\leq t}$ and we say that $(X_t)_{t\geq 0}$ is *reversible*. This happens if and only if

$$\xi_i q_{ij} = \xi_j q_{ji} \quad \text{for all } i, j \in \mathbb{S}. \tag{10}$$

These equations are again known as the *detailed balance equations* and if ξ is a solution we say that ξ and Q are *in detailed balance.*

Lemma 7.11. *Suppose that Q and ξ are in detailed balance. Then ξ is invariant.*

Proof. As in the discrete case, sum (10) over $i \in \mathbb{S}$ to show that $\xi Q = 0$. □

8 Application: queueing theory

The theory of queues originated with the study of calls at a telephone exchange. It is now a large and well-developed area of Applied Probability. Here, we will talk about some of the basic models and their properties, using the theory that we have developed earlier in the course. Most of what follows will be worked examples. We will make the following general assumptions:

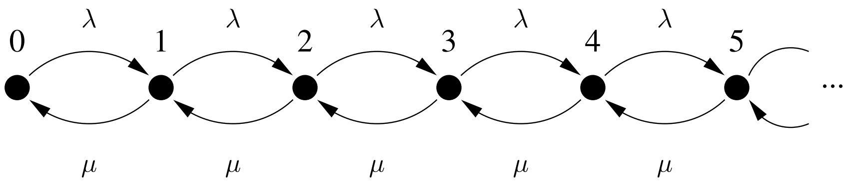
- Inter-arrival times are i.i.d. r.v.'s.
- Arriving customers join the end of the queue and are served in the order they arrive.
- Service times are i.i.d. r.v.'s which do not depend on the arriving stream of customers.

The queueing processes we will study are

- **M/M/s**: **M**emoryless inter-arrival times, **M**emoryless service times, **s** servers.
- **M/G/1**: **M**emoryless inter-arrival times, **G**eneral service times, **1** server
- **G/M/1**: **G**eneral inter-arrival times, **M**emoryless service times, **1** server.

8.1 The M/M/1 queue

Customers arrive according to a Poisson process of rate λ and service times are $\text{Exp}(\mu)$. Write $\rho = \lambda/\mu$, and call it the *traffic intensity*. X_t is the number of customers in the queue at time t, including the customer being served, if there is one. $X = (X_t)_{t\geq 0}$ is irreducible and non-explosive. It is a birth-and-death process:



We saw earlier that

- if $\rho \leq 1$, X is recurrent
- if $\rho > 1$, X is transient.

If $\rho < 1$, an invariant distribution ξ exists and is given by

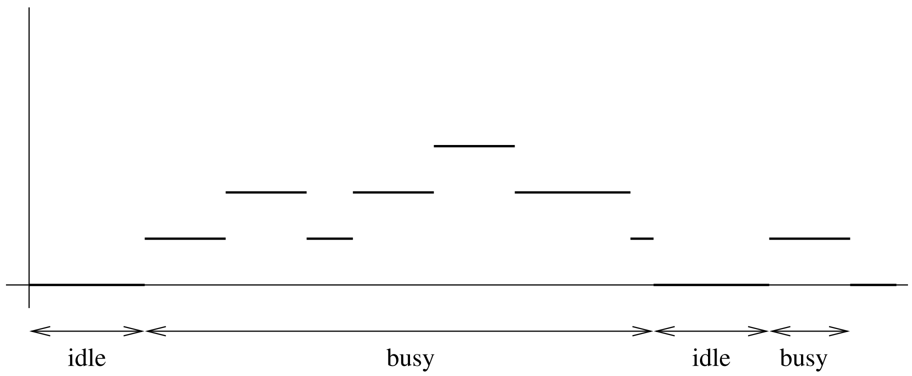
$$\xi_n = \rho^n(1-\rho), \quad n \geq 0.$$

If $\rho = 1$, no invariant distribution exists. To see this, note that the jump-chain is the modulus of a simple symmetric random walk on $\mathbb{Z}$, which is null recurrent. So X cannot be positive recurrent here, and so cannot possess an invariant distribution.

We will focus on the case $\rho < 1$.

8.1.1 Busy and idle periods

A *busy period* is a time-interval $[r, s)$ such that $X_t \geq 1$ for all $t \in [r, s)$, $X_{r-} = 0$ and $X_s = 0$ (i.e. the queue is empty just before and just after the interval). An *idle period* is a time-interval $[r, s)$ such that $X_t = 0$ for all $t \in [r, s)$, $X_{r-} \geq 1$ and $X_s \geq 1$ (i.e. the queue is non-empty just before and just after the interval).



By the ergodic theorem, the long-run proportion of time for which the server is idle is

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s=0\}} ds = \xi_0 = 1-\rho \text{ a.s.}$$

Likewise, the long-run proportion of time for which the server is busy is

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X_s\geq 1\}} ds = \rho \text{ a.s.}$$

Write $H_0 = \inf\{t \geq T_1 : X_t = 0\}$ for the first passage time to 0. Then we know that $m_0 := \mathbb{E}_0[H_0]$ satisfies

$$\xi_0 = \frac{1}{m_0 q_0},$$

where q_0 is the rate of leaving state 0. Since $\xi_0 = 1-\rho$ and $q_0 = \lambda$, we get

$$m_0 = \frac{1}{\lambda(1-\rho)}.$$

But the server isn't busy until a customer has arrived and so the mean length of a busy period is

$$\mathbb{E}_0[H_0] - \mathbb{E}_0[T_1] = \frac{1}{\lambda(1-\rho)} - \frac{1}{\lambda} = \frac{1}{\mu-\lambda}.$$

8.1.2 The departure process and Burke's theorem

Continue to assume $\rho < 1$. Let

$$\begin{aligned} A_s &= \text{the number of customers who have arrived by time } s \\ D_s &= \text{the number of customers who have departed by time } s. \end{aligned}$$

Recall that for a function $f : [0,\infty) \to [0,\infty)$, we write $f(t-) = \lim_{s\uparrow t} f(s)$ for the *left limit* of f at t. Then we could also have written

$$\begin{aligned} A_s &= \#\{0 \leq r \leq s : X_r - X_{r-} = +1\} \\ D_s &= \#\{0 \leq r \leq s : X_r - X_{r-} = -1\}. \end{aligned}$$

By assumption, $(A_s)_{s\geq 0} \sim \mathrm{PP}(\lambda)$. What can we say about $(D_s)_{s\geq 0}$? Note that

$$X_s = X_0 + A_s - D_s.$$

It seems that the distribution of $(D_s)_{s\geq 0}$ will be complicated to describe. But we will be able to do so, using a very beautiful and powerful application of reversibility.

Recall that $\xi_n = \rho^n(1-\rho)$, $n \geq 0$ is the invariant distribution, where $\rho = \lambda/\mu$. For $n \geq 0$,

$$\xi_n q_{n,n+1} = \rho^n(1-\rho)\lambda = \rho^{n+1}(1-\rho)\mu = \xi_{n+1} q_{n+1,n}.$$

So ξ and Q are in detailed balance. It follows that $(X_t)_{t\geq 0}$ is reversible in equilibrium.

So suppose that the queue is in equilibrium (i.e. $X_0 \sim \xi$). Fix $t > 0$ and let $(\hat{X}_s)_{0\leq s\leq t}$ be the time-reversal of $(X_s)_{0\leq s\leq t}$. Then $(\hat{X}_s)_{0\leq s\leq t} \stackrel{d}{=} (X_s)_{0\leq s\leq t}$. Define

$$\begin{aligned} \hat{A}_s &= \#\{0 \leq r \leq s : \hat{X}_r - \hat{X}_{r-} = +1\} \\ \hat{D}_s &= \#\{0 \leq r \leq s : \hat{X}_r - \hat{X}_{r-} = -1\}. \end{aligned}$$

Since $(\hat{X}_s)_{0\leq s\leq t} \stackrel{d}{=} (X_s)_{0\leq s\leq t}$ we must also have

$$(\hat{A}_s)_{0\leq s\leq t} \stackrel{d}{=} (A_s)_{0\leq s\leq t}.$$

So $(\hat{A}_s)_{0\le s\le t} \sim \mathrm{PP}(\lambda)$. Notice that the times of the jumps of a Poisson process in $[0,t]$ are uniformly distributed. Jumps of $(D_s)_{0\le s\le t}$ correspond to jumps of $(\hat{A}_s)_{0\le s\le t}$ and so, using the symmetry of the Poisson process,

$$(D_s)_{0\le s\le t} \stackrel{d}{=} (\hat{A}_s)_{0\le s\le t}.$$

It follows that the departure process $(D_s)_{0\le s\le t}$ is a Poisson process of rate λ for any $t > 0$. This somewhat surprising result is known as *Burke's theorem.*

Remark 8.1. *One might expect that departures had more to do with the parameter μ than λ. The point is that $(A_s)_{0\le s\le t}$ and $(D_s)_{0\le s\le t}$ are* dependent *and the link between them is provided by the equilibrium distribution ξ.*

8.2 Tandem queues

Suppose now we have two single-server queues with independent service times of lengths $\mathrm{Exp}(\mu_1)$ and $\mathrm{Exp}(\mu_2)$ respectively. Customers arrive at the first queue as a Poisson process of rate λ. When they have been served by the first server, they proceed directly to the second queue.

Let

$$\begin{aligned} X_t^{(1)} &= \text{ length of the first queue at time } t \\ X_t^{(2)} &= \text{ length of the second queue at time } t. \end{aligned}$$

Then $X = (X^{(1)}, X^{(2)})$ is a continuous-time Markov chain with transition rates

$$q_{(i,j),(i+1,j)} = \lambda, \quad q_{(i+1,j),(i,j+1)} = \mu_1, \quad q_{(i,j+1),(i,j)} = \mu_2, \quad i,j \in \mathbb{N}.$$

Proposition 8.2. $X = (X^{(1)}, X^{(2)})$ *is positive recurrent iff* $\rho_1 := \lambda/\mu_1 < 1$ *and* $\rho_2 := \lambda/\mu_2 < 1$. *The unique stationary distribution is then given by*

$$\xi_{(i,j)} = \rho_1^i(1-\rho_1)\rho_2^j(1-\rho_2), \quad i,j \in \mathbb{N}.$$

This implies that, in equilibrium, the length of the two queues at any fixed time t *are independent.*

Proof. Let $m_0^{(1)}$ be the expected return time to 0 for $X^{(1)}$. Let $m_{(0,0)}$ be the expected return time to $(0,0)$ for X. Then $m_{(0,0)} \ge m_0^{(1)}$. But if $\rho_1 \ge 1$ then $m_0^{(1)} = \infty$ and so we cannot have positive recurrence for X. So suppose that $\rho_1 < 1$. If $X^{(1)}$ is in equilibrium, its departure process is a $\mathrm{PP}(\lambda)$. But this is the arrival process for the second queue. If we were to have $\rho_2 \ge 1$, we could not obtain equilibrium for the second queue. So we must have $\rho_2 < 1$ also. Finally, it is straightforward to check that ξ as specified satisfies $\xi Q = 0$.

Since $\xi_{(i,j)}$ is the product of the invariant probabilities of being in states i and j for the first and second queues respectively, we have independence of the queue lengths in equilibrium for any fixed time. $\square$

8.3 Networks of queues

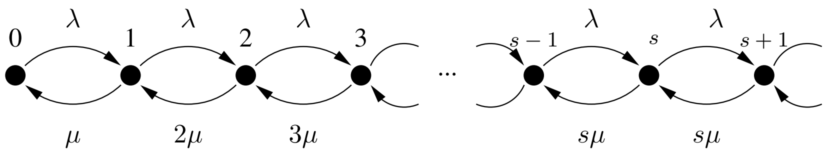
Suppose now we have m single-server queues, labelled by $\{1, 2, \ldots, m\}$. Customers move around the network according to a Markov chain on $\{1, 2, \ldots, m\}$. If there are r customers in total, new customers cannot enter the system and existing customers cannot leave, then the system is called a *closed migration network*. If, on the other hand, new customers can arrive at some of the queues according to a Poisson process, and customers can leave after service at some of the queues, then the system is called a *open migration network*. The tandem queue is an open migration network with $m = 2$, where new customers can only arrive at the first queue and existing customers can only leave after service at the second server. The Markov chain is deterministic and just sends a customer leaving queue 1 straight to queue 2. Customers leaving the system go to an absorbing exit state, which we can label 0.

Take an open or closed migration network. Suppose service times are $\text{Exp}(\mu_k)$ at server k, $k \in \{1, 2, \ldots, m\}$ and new arrivals at queue k form a $\text{PP}(\lambda_k)$, $k \in \{1, 2, \ldots, m\}$. Suppose departures are always to another queue, or to an exit state denoted 0, according to transition probabilities $\pi_{i,j}$. Then the process of queue lengths at each of the m servers, $X = (X^{(1)}, X^{(2)}, \ldots, X^{(m)})$, is a continuous-time Markov chain on $\mathbb{N}^m$.

8.4 The M/M/s queue

Now suppose that instead of a single server we have s servers. Arrivals form a $\text{PP}(\lambda)$ and service times are $\text{Exp}(\mu)$. When $1 \leq k \leq s$ servers are occupied, the first service is completed at rate $k\mu$. The length of the queue forms a birth-and-death process with rates

$$\begin{aligned} q_{i,i+1} &= \lambda, && \text{for all } i \geq 0, \\ q_{k,k-1} &= k\mu && \text{for } 1 \leq k \leq s, \\ q_{k.k-1} &= s\mu && \text{for } k > s. \end{aligned}$$



The chain is transient if $\lambda > s\mu$ and recurrent otherwise. It is positive recurrent if $\lambda < s\mu$. Then the unique invariant distribution is most easily found by solving the detailed balance equations to obtain

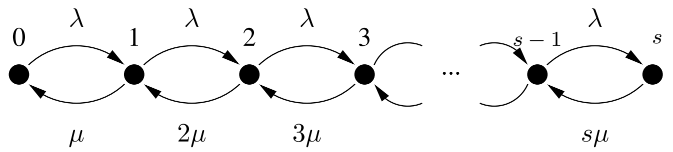
$$\xi_i = \begin{cases} \xi_0 \frac{\rho^i}{i!} & 1 \leq i \leq s, \\ \xi_0 \frac{\rho^i}{s^{i-s} s!} & i > s, \end{cases}$$

where ξ_0 is the constant which gives $\sum_{i=0}^{\infty} \xi_i = 1$ and $\rho = \lambda/\mu$.

Note that if $s = \infty$ (i.e. no-one ever has to wait), the queue is always positive recurrent and has $\text{Poisson}(\rho)$ invariant distribution.

8.5 The telephone exchange

Calls arrive at the exchange as a Poisson process of rate λ. A call lasts an $\text{Exp}(\mu)$ time (independently of everything else). There are s telephone lines, and if all s lines are busy then arriving calls are lost. (This is effectively an $M/M/s$ queue but where customers are turned away if all the servers are busy.) The number of busy lines evolves as a continuous-time Markov chain with state-space $\{0, 1, 2, \ldots, s\}$.



Since the state-space is finite and irreducible, there must exist an invariant measure. Solving the detailed balance equations gives

$$\xi_i = \frac{\rho^i}{i!} \Big/ \sum_{j=0}^{s} \frac{\rho^j}{j!}, \quad 0 \le i \le s.$$

In consequence, in stationarity, the probability that all of the lines are busy is

$$\xi_s = \frac{\rho^s}{s!} \Big/ \sum_{j=0}^{s} \frac{\rho^j}{j!}.$$

This is called *Erlang's formula.*

8.6 The M/G/1 queue

Here, the arrival process is still a Poisson process of rate λ, but the service times are i.i.d. with a general distribution on $(0, \infty)$. Let X_t be the length of the queue at time t. Since the service times are no longer memoryless, $(X_t)_{t\ge0}$ is no longer a Markov chain. In particular, after an arrival, we have a residual amount of service time whose distribution we do not know. To get around this problem, consider the length of the queue just after a departure. Then we know both the distribution of time time until the next arrival (Exp(λ)) and the time until the next departure (a full service time).

Let $D_0 = 0$ and, for $n \ge 1$, D_n be the time of the nth departure in $(X_t)_{t\ge0}$ i.e.

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

Let $V_n = X_{D_n}$.

Proposition 8.3. *Let G have the service time distribution. Then $(V_n)_{n\ge0}$ is a continuous-time Markov chain with transition probabilities*

$$d_{k,k-1+m} = \mathbb{E}\left[e^{-\lambda G}\frac{(\lambda G)^m}{m!}\right], \quad k \ge 1, \quad m \ge 0$$

and $d_{0,m} = d_{1,m}$, $m \ge 0$ (since no departures can occur when there are no customers).

Proof. Suppose that G has density g. Given that a service time is of length t, a Po(λt) number customers arrive during that time. So the number N of customers arriving in a generic service interval (of length G) has distribution given by

$$\begin{aligned}\mathbb{P}(N = m) &= \int_0^\infty \mathbb{P}(N = m | G = t)\, g(t)dt \\ &= \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^m}{m!} g(t)dt \\ &= \mathbb{E}\left[e^{-\lambda G}\frac{(\lambda G)^m}{m!}\right].\end{aligned}$$

So, if we had $k \geq 1$ customers in the queue at the beginning of the service time, we have $k-1+m$ afterwards with the above probability.

Let $G_1, G_2, \ldots$ be the service times of successive customers. These are independent. Let $N_1, N_2, \ldots$ be the numbers of customers arriving in each of these service intervals. Then $N_1, N_2, \ldots$ are independent because they are the numbers of Poisson arrivals in disjoint intervals whose lengths are independent.

If the queue is empty after a departure, then the next event is an arrival with probability 1. The next departure will occur at the end of that customer's service time. So $d_{0,m}$ is the same as $d_{1,m}$. It follows that $(V_n)_{n\geq 0}$ is a Markov chain with the claimed transition probabilities. □

Let G have moment generating function $\psi(\theta) = \mathbb{E}\left[e^{\theta G}\right]$. Then, since $G \geq 0$, this is finite for all $\theta \leq 0$. Now define the *traffic intensity* to be $\rho = \lambda \mathbb{E}[G]$.

Proposition 8.4. *If $\rho < 1$ then $V = (V_n)_{n\geq 0}$ has a unique invariant distribution ξ whose probability generating function ϕ is given by*

$$\phi(s) = \sum_{k=0}^{\infty} \xi_k s^k = (1-\rho)(1-s)\frac{\psi(\lambda(s-1))}{\psi(\lambda(s-1)) - s}.$$

Proof. We need to check that ξ with p.g.f. ϕ is a solution to

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

It is sufficient to check that the p.g.f.'s of the left- and right-hand sides are equal. Uniqueness will then follow from the irreducibility of V.

The left-hand side has p.g.f. $\phi(s)$. For the right-hand side, first note that for any $k \geq 0$,

$$\begin{aligned}\sum_{m=0}^{\infty} d_{k+1,k+m} s^m &= \sum_{m=0}^{\infty} \mathbb{E}\left[e^{-\lambda G}\frac{(\lambda s G)^m}{m!}\right] \\ &= \mathbb{E}\left[e^{(s-1)\lambda G}\right] = \psi(\lambda(s-1)).\end{aligned}$$

So then

$$\begin{aligned}\sum_{j=0}^{\infty}\left(\sum_{i=0}^{j+1} \xi_i d_{i,j}\right) s^j &= \xi_0 \sum_{j=0}^{\infty} d_{0,j} s^j + \sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \xi_{k+1} d_{k+1,k+m} s^{k+m} \\ &= \xi_0 \sum_{j=0}^{\infty} d_{1,j} s^j + \sum_{k=0}^{\infty} \xi_{k+1} s^k \sum_{m=0}^{\infty} d_{k+1,k+m} s^m \\ &= \left(\xi_0 + \sum_{k=0}^{\infty} \xi_{k+1} s^k\right) \psi(\lambda(s-1)) \\ &= \left(\xi_0 + s^{-1}\sum_{j=0}^{\infty} \xi_j s^j - \xi_0 s^{-1}\right) \psi(\lambda(s-1)).\end{aligned}$$

Since $\xi_0 = \phi(0) = (1-\rho)$, we obtain

$$s^{-1}(\phi(s) - (1-\rho)(1-s))\psi(\lambda(s-1)) = \phi(s),$$

by definition of $\phi(s)$. □

How long must a customer wait to be served in a M/G/1 queue? Let us suppose that X is such that the departure process $(V_n)_{n\geq 0}$ is in equilibrium. Let Q be the total time the customer spends in the queue, W his time waiting to be served and G his service time, so that $Q = W + G$.

Given that $Q = x$, we know that a $\mathrm{Po}(\lambda x)$ number of customers arrive during the time from our customer's arrival until his departure. Call this number N. Moreover, this is the queue length just after he leaves. Since the departure process is in equilibrium, N must have distribution ξ and hence p.g.f. ϕ.

In particular,

$$\phi(s) = \int_0^\infty \mathbb{E}\left[s^N | Q = x\right] f_Q(x)dx = \int_0^\infty e^{\lambda x(s-1)} f_Q(x)dx,$$

since $\mathrm{Po}(\lambda x)$ has p.g.f. $e^{\lambda x(s-1)}$. The last expression is equal to

$$\mathbb{E}\left[e^{\lambda Q(s-1)}\right] = \mathbb{E}\left[e^{\lambda W(s-1)}\right] \mathbb{E}\left[e^{\lambda G(s-1)}\right],$$

since W and Q are independent.

Recall that $\mathbb{E}\left[e^{\theta G}\right] = \psi(\theta)$. Setting $\theta = \lambda(s-1)$, we have

$$\phi(s) = \phi(\theta/\lambda + 1) = \mathbb{E}\left[e^{\theta W}\right] \psi(\theta).$$

Now

$$\phi(\theta/\lambda + 1) = \frac{(1-\rho)\theta\psi(\theta)}{\theta + \lambda - \lambda\psi(\theta)}.$$

Hence, putting things together,

$$\mathbb{E}\left[e^{\theta W}\right] = \frac{\phi(s)}{\psi(\theta)} = \frac{(1-\rho)\theta}{\theta + \lambda - \lambda\psi(\theta)}.$$

In other words, if we know the moment generating function of the service-time distribution, we can calculate the moment generating function of the length of time a customer has to wait to be served (assuming the departure process is in equilibrium).

8.7 The G/M/1 queue

For the G/M/1 queue, the arrival process is no longer a Poisson process. Service times are $\mathrm{Exp}(\mu)$. Let A denote a generic inter-arrival time and suppose that $\mathbb{E}[A] < \infty$.

Let U_n be the number of *other* customers in the queue at the time of the nth arrival.

Proposition 8.5. *$(U_n)_{n\geq 0}$ is a Markov chain on $\mathbb{N}$ with transition probabilities*

$$a_{i,i-j+1} = \mathbb{E}\left[e^{-\mu A}\frac{(\mu A)^j}{j!}\right], \quad 0 \leq j \leq i, \quad a_{i,0} = 1 - \sum_{j=0}^{i} a_{i,i-j+1}.$$

Proof. Consider the number of customers who are served between two arrival times i.e. in a length of time with the same distribution as A. Suppose that just after the first arrival there are $i+1$ people in the queue (i.e. $U_n = i$). Conditional on $A = a$, the departures occur as a Poisson process of rate μ, run for time a and stopped if it ever hits $i+1$ (since no more than $i+1$ people can depart). So the number of people who depart has the same distribution as $\max\{P, i+1\}$ where $P \sim \mathrm{Po}(\mu a)$.

We get

- $\max\{P, i+1\} = j$ with probability $e^{-\mu a}(\mu a)^j/j!$ for $0 \le j \le i$;
- $\max\{P, i+1\} = i+1$ with probability $1 - \sum_{j=0}^{i} e^{-\mu a}(\mu a)^j/j!$.

Hence,

$$a_{i,i-j+1} = \int_0^\infty e^{-\mu a}\frac{(a\mu)^j}{j!} f_A(a)da = \mathbb{E}\left[e^{-\mu A}\frac{(\mu A)^j}{j!}\right],$$
$$a_{i,0} = 1 - \sum_{j=0}^{i} a_{i,i-j+1}.$$

Finally, because the inter-arrival times are independent and the service times are memoryless, the number of departures in a particular inter-arrival period is independent of previous numbers of departures, given the queue length at the beginning of the interval. So $(U_n)_{n\ge0}$ is a Markov chain. □

We define the *traffic intensity* to be $\rho = 1/(\mu\mathbb{E}[A])$.

Proposition 8.6. *If $\rho < 1$ then $(U_n)_{n\ge0}$ has a unique invariant distribution given by*

$$\xi_k = (1-q)q^k, \quad k \ge 0,$$

where q is the smallest positive solution of the equation $s = \mathbb{E}\left[e^{\mu(s-1)A}\right]$.

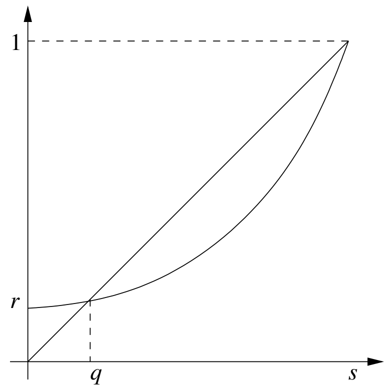
Proof. We show first that ξ as defined in the statement is invariant. Fix $k \ge 1$. Then

$$\begin{aligned}\sum_{i=k-1}^{\infty} \xi_i a_{i,k} &= \sum_{j=0}^{\infty} \xi_{j+k-1} a_{j+k-1,j} \\ &= \sum_{j=0}^{\infty} (1-q)q^{j+k-1}\mathbb{E}\left[e^{-\mu A}\frac{(\mu A)^j}{j!}\right] \\ &= (1-q)q^{k-1}\mathbb{E}\left[e^{-\mu A}\sum_{j=0}^{\infty}\frac{(\mu q A)^j}{j!}\right] \\ &= (1-q)q^{k-1}\mathbb{E}\left[e^{\mu(q-1)A}\right] \\ &= (1-q)q^k,\end{aligned}$$

since $q = \mathbb{E}\left[e^{\mu(q-1)A}\right]$. $\xi_0 = 1-q$ is implied since ξ must be a distribution.

Since the Markov chain is irreducible, there can be at most one invariant distribution. Note that $s = 1$ is always a solution to $s = \mathbb{E}\left[e^{\mu(s-1)A}\right]$ but if $q = 1$ then $\xi_k = 0$ for all $k \ge 0$ which is not a distribution. So we need to find a solution in $(0,1)$; if we can do so then it must unique since otherwise there would be at least 2 invariant distributions, which is impossible.

Now $\phi(s) := \mathbb{E}\left[e^{\mu(s-1)A}\right]$ is differentiable in $[0,1)$ and $\lim_{s\uparrow1}\phi'(s) = \mathbb{E}[\mu A] = 1/\rho$. Since $\rho < 1$, ϕ has gradient > 1 at 1 and so, for sufficiently small $\epsilon > 0$, $\phi(1-\epsilon) < 1-\epsilon$. Since $r = \phi(0) > 0$, the graph of $\phi(s)$ must intersect with that of $f(s) = s$ somewhere in $(0,1)$.



So $\xi_k = (1-q)q^k$, $k \geq 0$ must be the invariant distribution. □

Note that this is a Geometric distribution.

How long must a customer wait to be served in a G/M/1 queue? Suppose that the Markov chain $(U_n)_{n\geq 0}$, which gives the number of customers present in the queue just before arrival times, is in equilibrium. Let W be the length of time an arriving customer waits until he is served. He arrives to find $N \sim \xi$ customers ahead of him in the queue.

If $N = 0$, he does not have to wait at all, so

$$\mathbb{P}(W = 0) = \xi_0 = 1 - q.$$

If $N \geq 1$, each of the other customers has an $\text{Exp}(\mu)$ service time. So he must wait an amount of time distributed as $\sum_{i=1}^N E_i$ where $E_1, E_2, \ldots \sim \text{Exp}(\mu)$ independently of one another and N.

We find the distribution of $\sum_{i=1}^N E_i$ by calculating its moment generating function. Recall that $\mathbb{E}\left[e^{\theta E_1}\right] = \mu/(\mu - \theta)$, so

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\theta \sum_{i=1}^N E_i\right) \mathbb{1}_{\{N\geq 1\}}\right] &= \sum_{k=1}^\infty (1-q)q^k \mathbb{E}\left[\exp\left(\theta \sum_{i=1}^k E_i\right)\right] \\
&= \sum_{k=1}^\infty (1-q)q^k \left(\frac{\mu}{\mu - \theta}\right)^k \\
&= q\frac{(1-q)\mu}{(1-q)\mu - \theta}.
\end{aligned}$$

Since $\mathbb{P}(N \geq 1) = q$, this means that, conditional on $N \geq 1$, we get a $\text{Exp}((1-q)\mu)$ waiting time. So,

$$\mathbb{P}(W = 0) = 1 - q, \quad \mathbb{P}(W > w) = qe^{-\mu(1-q)w}, \quad w \geq 0.$$

9 Renewal theory

Renewal processes are counting processes which generalise the Poisson process. We use them for modelling in circumstances where exponential inter-arrival times are inappropriate.

Definition 9.1. *Let $Z_1, Z_2, \ldots$ be i.i.d. strictly positive random variables with distribution function F (i.e. $F(t) = \mathbb{P}(Z_1 \leq t)$). Let $T_0 = 0$ and $T_n = \sum_{k=1}^n Z_k$ for $n \geq 1$. Then the process $X = (X_t)_{t\geq 0}$ defined by*

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

is called a renewal process *and F is referred to as the* inter-arrival distribution.

Example: return times in a Markov chain.
Let $(Y_t)_{t\geq 0}$ be a continuous-time Markov chain with $Y_0 = i$. Let $H_i^{(1)} = \inf\{t > T_1 : Y_t = i\}$ be the first passage time to i and define $H_i^{(2)}, H_i^{(2)}, \dots$ to be the successive return times to i. Set $H_i^{(0)} = 0$. By the strong Markov property, the random variables $\{H_i^{(n+1)} - H_i^{(n)}, n \geq 0\}$ are i.i.d. and so

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

is a renewal process (which counts the number of visits to i by time t).

9.1 The renewal function

We will investigate various aspects of the distribution of X_t for a general renewal process. First, note that since

$$X_t \geq k \quad \text{iff} \quad T_k \leq t$$

we have

$$\mathbb{P}\left(X_t \geq k\right) = \mathbb{P}\left(T_k \leq t\right)$$

and

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

So we want to investigate the distribution of a sum of i.i.d. random variables.

Lemma 9.2. *Suppose that S and T are independent strictly positive random variables having distribution functions F and G respectively and density functions f and g respectively. Then $S + T$ has distribution function*

$$\mathbb{P}\left(S + T \leq t\right) = \int_0^t F(t-u)g(u)du = \int_0^t G(t-u)f(u)du$$

and density

$$\int_0^t f(t-u)g(u)du = \int_0^t g(t-u)f(u)du.$$

Proof. S and T have joint density $f(x)g(y)$. So

$$\begin{aligned}\mathbb{P}\left(S + T \leq t\right) &= \int_0^t \int_0^{t-y} f(x)g(y)dxdy = \int_0^t F(t-y)g(y)dy \\ &= \int_0^t \int_0^{t-x} f(x)g(y)dydx = \int_0^t G(t-x)f(x)dx.\end{aligned}$$

Differentiating using Leibniz's rule gives density

$$\int_0^t f(t-y)g(y)dy = \int_0^t g(t-x)f(x)dx.$$

□

For functions f and g, define the *convolution product*

$$(f \star g)(t) = \int_{-\infty}^{\infty} f(t-u)g(u)du$$

and note that, by change of variable, $f \star g = g \star f$. Write $f^{\star(k)}$ for $\underbrace{f \star f \star \cdots \star f}_{k}$.

Recall that $F(t) = \mathbb{P}(Z_1 \le t) = \mathbb{P}(T_1 \le t)$ and write $F_k(t) := \mathbb{P}(T_k \le t)$. Then

$$F_{k+1}(t) = (F_k \star f)(t) = (F \star f^{\star(k)})(t) \quad \text{for } k \ge 1.$$

Let

$$m(t) = \mathbb{E}[X_t].$$

We refer to $m(t)$ as the *renewal function.*

Proposition 9.3. *Let X be a renewal process with inter-arrival density f. Then*

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

Proof. We have

$$X_t = \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k \le t\}}$$

and so the result follows by taking expectations and using Tonelli's theorem. □

Proposition 9.4. *Let X be a renewal process with inter-arrival density f. Then $m(t)$ satisfies the* renewal equation

$$m(t) = F(t) + \int_0^t m(t-s)f(s)ds$$

(i.e. $m = F + m \star f$). Moreover, m is the unique solution to the renewal equation which is bounded on finite intervals (i.e. $\sup_{t\in[0,K]} |m(t)| < \infty$ for all $K < \infty$).

Proof. Let $\tilde{X}_s = X_{T_1+s} - 1$ for $s \ge 0$. Note that $\tilde{X}_0 = 0$ and that $(\tilde{X}_s)_{s\ge0}$ is a renewal process with inter-arrival times $\tilde{Z}_n = Z_{n+1}$, $n \ge 1$, independent of T_1.

We condition on T_1:

$$\mathbb{E}[X_t] = \int_0^\infty \mathbb{E}[X_t | T_1 = s] f(s)ds.$$

If $s > t$ then $\mathbb{E}[X_t|T_1 = s] = 0$. If $s \le t$ then

$$\mathbb{E}[X_t|T_1 = s] = 1 + \mathbb{E}\left[\tilde{X}_{t-s}\right] = 1 + m(t-s).$$

So

$$m(t) = \int_0^t (1 + m(t-s))f(s)ds = F(t) + (m \star f)(t).$$

Now suppose, for a contradiction, that ℓ is another solution which is bounded on finite intervals. Let $\alpha = \ell - m$. Then α is also bounded on finite intervals and $\alpha = F + \ell \star f - (F + m \star f) = (\ell - m) \star f = \alpha \star f$. Iterating gives $\alpha = \alpha \star f^{\star(k)}$ for all $k \ge 1$. Now, by linearity, we have

$$\sum_{k=1}^{\infty} \left(\alpha \star f^{\star(k)}\right) = \alpha \star \sum_{k=1}^{\infty} f^{\star(k)} = \alpha \star m'.$$

So

$$\left| \left(\sum_{k=1}^{\infty} \alpha \star f^{\star(k)} \right)(t) \right| \le \int_0^t \alpha(t-s)m'(s)ds \le m(t) \sup_{u\in[0,t]} |\alpha(u)| < \infty.$$

But since all of the terms in the sum on the left-hand side are equal, the sum must be infinite, unless $\alpha \equiv 0$. □

9.2 Limit theorems

By Problem Sheet 1, we have a strong law of large numbers for $X \sim \mathrm{PP}(\lambda)$. A version of this result holds in general for renewal processes.

Theorem 9.5 (Strong law of renewal theory)**.** *Let X be a renewal process with $\mathbb{E}[Z_1] = \mu \in (0, \infty)$. Then*

$$\frac{X_t}{t} \to \frac{1}{\mu} \quad \text{a.s. as } t \to \infty.$$

Proof. By the strong law of large numbers,

$$\frac{T_n}{n} = \frac{1}{n}\sum_{k=1}^{n} Z_k \to \mu \quad \text{a.s. as } n \to \infty. \tag{11}$$

Now note that

$$T_{X_t} \le t < T_{X_t+1}$$

and so

$$\frac{1}{T_{X_t+1}} < \frac{1}{t} \le \frac{1}{T_{X_t}}$$

and

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

Now note that we must have $\mathbb{P}(X_t \to \infty \text{ as } t \to \infty) = 1$ since if $\lim_{t\to\infty} X_t \le m$ then $T_{m+1} = \infty$ which is impossible since T_{m+1} is a finite sum of finite random variables. So by (11),

$$\frac{X_t}{T_{X_t}} \to \frac{1}{\mu} \quad \text{a.s.} \quad \text{and} \quad \frac{X_t+1}{T_{X_t+1}} \to \frac{1}{\mu} \quad \text{a.s.}$$

Clearly $X_t/(X_t+1) \to 1$ a.s. as $t \to \infty$. The result follows by sandwiching. □

Theorem 9.6 (Central limit theorem of renewal theory)**.** *Let $X = (X_t)_{t\ge 0}$ be a renewal process whose inter-arrival times $Z_1, Z_2, \ldots$ satisfy $\mu = \mathbb{E}[Z_1]$ and $0 < \sigma^2 = \mathrm{var}(Z_1) < \infty$. Then*

$$\frac{X_t - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \xrightarrow{d} N(0,1)$$

as $t \to \infty$.

Proof. See Problem Sheet 4. □

Theorem 9.7 (The elementary renewal theorem)**.** *Let X be a renewal process with mean inter-arrival time μ and $m(t) = \mathbb{E}[X_t]$. Then*

$$\frac{m(t)}{t} \to \mu$$

as $t \to \infty$.

Note that this *does not* follow immediately from the Strong law since convergence almost surely does not imply convergence of means.

The proof is non-examinable. We need a lemma.

Lemma 9.8. *For a renewal process X with arrival times $(T_n)_{n\geq 1}$, we have*

$$\mathbb{E}\left[T_{X_t+1}\right] = \mu(m(t)+1),$$

where $m(t) = \mathbb{E}\left[X_t\right]$ and $\mu = \mathbb{E}\left[T_1\right]$.

This ought to be true, because T_{X_t+1} is the sum of X_t+1 inter-arrival times, each with mean μ. Taking expectations, we should get $(m(t)+1)\mu$. However, if we condition on X_t we have to know the distribution of the residual inter-arrival time after t but, without the memoryless property, it's not clear how to do this.

Proof. We do a one-step analysis of $g(t) = \mathbb{E}\left[T_{X_t+1}\right]$:

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

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

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

Since we know that $m(t)$ is the unique solution to the renewal equation which is bounded on finite intervals, $h(t) = m(t)$, i.e. $g(t) = \mu(1+m(t))$, as required. □

Proof of the elementary renewal theorem. We certainly have $t < T_{X_t+1}$ and so $t < \mathbb{E}\left[T_{X_t+1}\right] = \mu(m(t)+1)$ gives the lower bound

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

For the upper bound we use a truncation argument: introduce

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

with associated renewal process $\widetilde{X}$. $\widetilde{Z}_j \leq Z_j$ for all $j \geq 0$ implies $\widetilde{X}_t \geq X_t$ for all $t \geq 0$ and so $\widetilde{m}(t) \geq m(t)$. We can apply the lemma again to obtain

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

Therefore,

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

so that

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

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

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

which completes the proof. □

Remark 9.9. *Note that truncation was necessary to get $\mathbb{E}\left[\widetilde{Z}_{\widetilde{X}_t+1}\right] \leq a$. It would have been enough to have had $\mathbb{E}\left[Z_{X_t+1}\right] = \mathbb{E}\left[Z_1\right] = \mu$, but this is* not true. *Consider at the Poisson process as an example. We know that the residual lifetime already has mean $\mu = 1/\lambda$, but there is also the part of Z_{X_t+1} before time t to take care of.*

9.3 The renewal property

We cannot expect the Markov property to hold for a renewal process which does not have exponential inter-arrival times.

Example.
Suppose that $\mathbb{P}(Z_1 = 1) = \mathbb{P}(Z_1 = 3) = 1/2$ and take $t = 1$. Let $\tilde{X}_s = X_{t+s} - X_t$ and consider $\tilde{Z}_1$, the first inter-arrival time of $\tilde{X}$. Then

- If $Z_1 = 1$ and $Z_2 = 1$ then $\tilde{Z}_1 = 1$.
- If $Z_1 = 1$ and $Z_2 = 3$ then $\tilde{Z}_1 = 3$.
- If $Z_1 = 3$ then $\tilde{Z}_1 = 2$.

So $\mathbb{P}\left(\tilde{Z}_1 = 1\right) = 1/4$, $\mathbb{P}\left(\tilde{Z}_1 = 2\right) = 1/2$ and $\mathbb{P}\left(\tilde{Z}_1 = 3\right) = 1/4$. Self-evidently $\tilde{Z}_1$ does not have the same distribution as Z_1.

Suppose now we consider the distribution of $\tilde{Z}_1$ conditional on X_1. If $X_1 = 0$ we know that $Z_1 = 3$ and so $\tilde{Z}_1 = 2$ i.e. $\mathbb{P}\left(\tilde{Z}_1 = 2 | X_1 = 0\right) = 1$. So the distribution of $\tilde{Z}_1$ depends on $(X_r)_{r\leq 1}$ and so $(\tilde{X}_s)_{s\geq 0}$ is *not* independent of $(X_r)_{r\leq 1}$.

If we want a version of the Markov property, we have to look at the jump times.

Proposition 9.10 (The renewal property). *Let X be a renewal process. Fix $i \geq 1$ and let $T_i = \inf\{t \geq 0 : X_t = i\}$. 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. For $n \geq 1$, set $\tilde{Z}_n = Z_{i+n}$. Then $\tilde{Z}_1, \tilde{Z}_2, \ldots$ are i.i.d. and independent of $Z_1, \ldots, Z_i$. Write $\tilde{X}_s = X_{T_i+s} - X_{T_i}$. Then $\tilde{X}_0 = 0$ and $\tilde{X}$ is a counting process with inter-arrival times $\tilde{Z}_1, \tilde{Z}_2, \ldots$. So $\tilde{X}$ has the same distribution as X and is independent of $(X_r)_{r\leq T_i}$. □

Let us define the *age process* $(A_t)_{t\geq 0}$ by

$$A_t = t - T_{X_t}$$

(this is the time since the last arrival) and the *excess lifetime process* $(E_t)_{t\geq 0}$ by

$$E_t = T_{X_t+1} - t$$

(the time until the next arrival). In particular, if we fix $t \geq 0$ and consider $\tilde{X}_s = X_{t+s} - X_t$ then $\tilde{Z}_1 = E_t$. In general, unless we have something like the memoryless property, A_t and E_t will be dependent, and the distribution of E_t will depend on t. However, this only causes problems for $\tilde{Z}_1$: subsequent inter-arrival times of $\tilde{X}$ are unaffected. This motivates the following definition.

Definition 9.11. *Let $Z_2, Z_3, \ldots$ be i.i.d. strictly positive random variables and let Z_1 be an independent strictly positive random variable with a (possibly) different distribution. Then the associated counting process $X = (X_t)_{t\geq 0}$ with*

$$X_t = \#\left\{n \geq 1 : \sum_{k=1}^{n} Z_k \leq t\right\}$$

is called a delayed renewal process.

Remark 9.12. *(a) The renewal property remains true for delayed renewal processes, and in this case the post-T_i process is an* undelayed *renewal process.*

(b) *It can also be shown that the renewal property holds at more general stopping times, provided that they take values only in the set of jump times. So, for example, $T = \inf\{s \geq t : X_s = 10\}$ wouldn't qualify (since we might be at 10 at time t) but T_{X_t+1} (the next jump after time t) would qualify.*

Proposition 9.13. *Given a (possibly delayed) renewal process X, for each $t \geq 0$ the process $\tilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ is a delayed renewal process with $\tilde{Z}_1 = E_t$.*

Proof. Apply the renewal property at T_{X_t+1}. This establishes that the inter-arrival times $\tilde{Z}_2, \tilde{Z}_3, \ldots$ of $\tilde{X}$ are i.i.d. and independent of $(X_r)_{r\leq T_{X_t+1}}$. So, in particular, $\tilde{Z}_2, \tilde{Z}_3, \ldots$ are independent of $\tilde{Z}_1 = T_{X_t+1} - t = E_t$. □

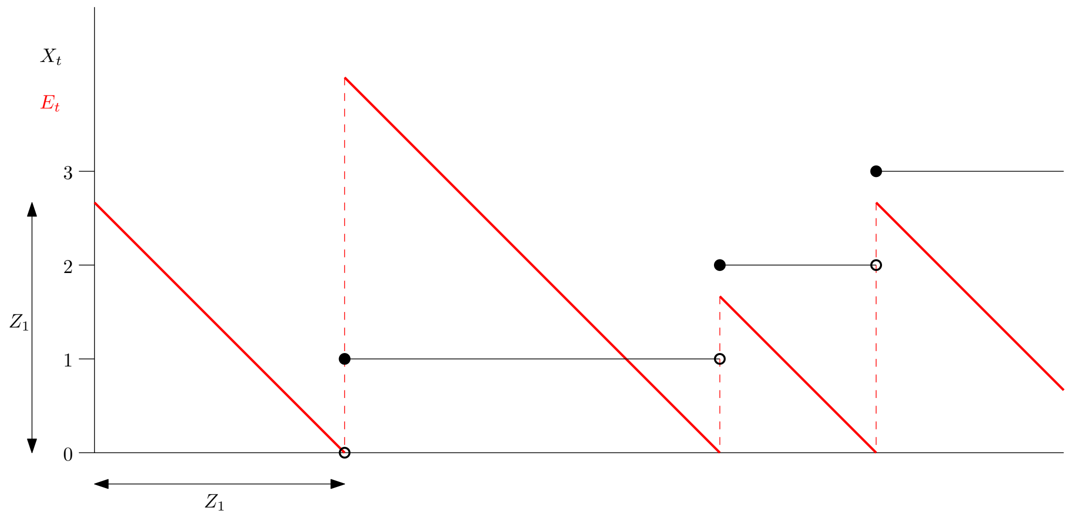
9.4 Size-biasing, stationarity and renewal theorems

The previous discussion shows that it we cannot hope to obtain independent increments for a renewal process. But by picking a good delay distribution for Z_1, it is possible that we can achieve stationarity of increments.

Proposition 9.14. *Let X be a delayed renewal process such that the distribution of the excess life E_t does not depend on $t \geq 0$. Then X has stationary increments i.e. $X_{t+s} - X_t \stackrel{d}{=} X_s$ for all $s, t \geq 0$.*

Proof. By the previous proposition, $\tilde{X} = (X_{t+s} - X_t)_{s\geq 0}$ is a delayed renewal process with $\tilde{Z}_1 = E_t$. If the distribution of $\tilde{Z}_1$ does not depend on t then nor does the distribution of $\tilde{X}$. □

$(E_t)_{t\geq 0}$ is a process taking values in $[0, \infty)$:



It in fact turns out that $(E_t)_{t\geq 0}$ is a Markov process with a continuous state-space. Intuitively, this is because whenever it hits 0, it jumps up by a random amount with distribution F, independent of everything which has gone before. When $(E_t)_{t\geq 0}$ is away from 0, on the other hand, it just decreases deterministically and linearly at rate 1. Such continuous-state Markov processes are beyond the scope of this course, so we will not prove this. If we wish the distribution of E_t to be independent of t then we should be looking for an invariant distribution for it. We can do this in the case where the inter-arrival times are discrete random variables.

Example: discrete renewal process
Suppose that the inter-arrival distribution is discrete, so that Z_1 is integer-valued. Assume that $\mathbb{P}(Z_1 = 0) = 0$, $\mathbb{E}[Z_1] < \infty$ and $\gcd\{i \geq 1 : \mathbb{P}(Z_1 = i) > 0\} = 1$.

Consider the renewal process $(X_t)t \geq 0$ at discrete times (we will write $(X_n)_{n\geq 0}$). Let E_n be the time until the next renewal at step n. Then $(E_n)_{n\geq 0}$ is a discrete-time Markov chain with transition probabilities

$$\begin{aligned}\pi_{i\,i-1} &= 1, \quad i \geq 2, \\ \pi_{1\,k} &= \mathbb{P}(Z_1 = k), \quad k \geq 1.\end{aligned}$$

(Checking this forms part of a question on Problem Sheet 4.) You can check that the chain is irreducible and positive recurrent (is it reversible?), so has a unique stationary distribution η. Moreover, it is aperiodic and so, by the theorem on convergence to equilibrium,

$$E_n \xrightarrow{d} \eta \quad \text{as } n \to \infty.$$

We don't have the theoretical tools to do the same in continuous time. But we can identify what the stationary distribution should be by finding $\lim_{t\to\infty} \mathbb{P}(E_t > x)$.

Proposition 9.15. *Suppose that Z_1 has density f and mean $\mu = \mathbb{E}[Z_1] < \infty$. Then*

$$\mathbb{P}(E_t > x) \to \frac{1}{\mu}\int_x^\infty (y - x) f(y) dy.$$

Proof. See Problem Sheet 4. □

Before going any further, we will interpret the limiting probability distribution.

Definition 9.16.

- *Suppose $(p_n)_{n\geq 0}$ is a probability distribution on $\mathbb{N}$ with mean $\mu = \sum_{k=1}^\infty kp_k < \infty$. Then the* size-biased distribution $(\hat{p}_n)_{n\geq 0}$ *is given by*

$$\hat{p}_n = \frac{np_n}{\mu}, \quad n \geq 0.$$

- *If f is a probability density function on $[0, \infty)$ with $\mu = \int_0^\infty tf(t)dt < \infty$ then the* size-biased distribution *has density $\hat{f}$ given by*

$$\hat{f}(t) = \frac{tf(t)}{\mu}, \quad t \geq 0.$$

Lemma 9.17. *Suppose that F is a distribution on $[0, \infty)$ with density f and mean $\mu < \infty$. Then if L has the size-biased distribution and $U \sim U[0, 1]$ independently of L,*

$$\mathbb{P}(LU > x) = \frac{1}{\mu}\int_x^\infty (y - x) f(y) dy,$$

and LU has density

$$g(x) = \frac{1}{\mu}(1 - F(x)), \quad x \geq 0.$$

Proof. Conditioning on the value of L, we obtain

$$\begin{aligned}\mathbb{P}(LU > x) &= \int_x^\infty \mathbb{P}(U > x/y)\frac{yf(y)}{\mu}dy \\ &= \int_x^\infty \left(1 - \frac{x}{y}\right)\frac{yf(y)}{\mu}dy = \frac{1}{\mu}\int_x^\infty (y - x) f(y) dy.\end{aligned}$$

The density is obtained by differentiating. □

Why should this give the stationary distribution for E_t? Imagine inspecting the process at some large time t. What is the distribution of the inter-arrival interval we fall into? We are more likely to fall into a longer interval. So, in fact, we fall into a size-biased inter-arrival interval. Moreover, intuitively, we are equally likely to fall anywhere in the interval. So the excess life, which is the time until the end of the interval, should have the same distribution as LU.

Theorem 9.18. *Let X be a delayed renewal process with inter-arrival times $Z_1, Z_2, \ldots$ such that $Z_2, Z_3, \ldots$ have density f and $Z_1 \stackrel{d}{=} LU$, where L has density $\hat{f}$ and $U \sim U[0,1]$. Then $E_t \stackrel{d}{=} LU$ for all $t \geq 0$ and, moreover, X has stationary increments.*

Proof. See Problem Sheet 4. □

There is an accompanying notion of convergence to equilibrium for a renewal process. Such a process is increasing, so we cannot hope to obtain that X_t converges in distribution as $t \to \infty$. But we can instead *fix* s and consider $X_{t+s} - X_t$ as $t \to \infty$.

Theorem 9.19. *Let X be a (possibly delayed) renewal process having a continuous inter-arrival distribution of finite mean μ. Then*

$$X_{t+s} - X_t \xrightarrow{d} \tilde{X}_s \quad \text{as } t \to \infty,$$

where $\tilde{X} = (\tilde{X}_s)_{s \geq 0}$ is the associated stationary *renewal process.*

Moreover,

$$(A_t, E_t) \xrightarrow{d} (L(1-U), LU)$$

as $t \to \infty$, where L has density $\hat{f}$ and $U \sim U[0,1]$, independently of L.

Recall that $m(t) = \mathbb{E}[X_t]$ is the renewal function. The corresponding result for the expectation of an increment of length s is the renewal theorem.

Theorem 9.20 (Renewal theorem)**.** *Under the conditions of the previous theorem, for all $s \geq 0$,*

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

The following result is often useful in applications.

Theorem 9.21 (Key renewal theorem)**.** *Let X be a renewal process with continuous inter-arrival distribution of mean $\mu < \infty$. If $h : [0,\infty) \to [0,\infty)$ is integrable and non-increasing,*

$$(h * m')(t) = \int_0^t h(t-x)m'(x)dx \to \frac{1}{\mu}\int_0^\infty h(x)dx$$

as $t \to \infty$.

10 Application: the insurance ruin model

Insurance companies deal with large numbers of policies. These policies are grouped according to type and other factors into *portfolios*. We will concentrate on such a portfolio and model the associated process of arrivals of claims, the claim sizes and the reserve process.

We make the following assumptions:

- Claims arrive according to a Poisson process $(X_t)_{t \geq 0}$ of rate λ.

- Claim amounts $(A_j)_{j\geq 0}$ are positive, independent of the arrival process and are i.i.d. with density $f(x), x > 0$ and mean $\mu = \mathbb{E}[A_1] < \infty$.
- The insurance company provides an initial reserve of $u \geq 0$ units of money.
- Premiums are paid continuously at a constant rate c, so that the accumulated premium income by time t is ct. Let $\rho = \lambda\mu/c$. We will assume that $\rho < 1$, which says that on average we have more premium income coming in than claims going out.

We ignore all expenses and other influences. We will be interested in

- The aggregate claims process, $(C_t)_{t\geq 0}$, given by $A_t = \sum_{n=1}^{X_t} C_n$;
- The reserve process, $(R_t)_{t\geq 0}$, given by $R_t = u + ct - A_t$;
- The ruin probability, $r(u) = \mathbb{P}_u(R_t < 0 \text{ for some } t \geq 0)$ as a function of the initial reserve $R_0 = u$.

10.1 Aggregate claims and reserve processes

Proposition 10.1. *The process $R = (R_t)_{t\geq 0}$ has stationary independent increments. Its moment generating function is given by*

$$\mathbb{E}\left[e^{\theta R_t}\right] = \exp\left(\theta u + \theta ct - \lambda t \int_0^\infty (1 - e^{-\theta a}) f(x) dx\right).$$

Proof. Let $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$. Then

$$(A_{t_1} - A_{t_0}, A_{t_2} - A_{t_1}, \ldots, A_{t_n} - A_{t_{n-1}}) = \left(\sum_{n=X_{t_0}+1}^{X_{t_1}} C_n, \sum_{n=X_{t_1}+1}^{X_{t_2}} C_n, \ldots, \sum_{n=X_{t_{n-1}}+1}^{X_{t_n}} C_n\right),$$

which are clearly independent. Moreover,

$$A_{t+s} - A_t = \sum_{n=X_t+1}^{X_{t+s}} C_n \stackrel{d}{=} \sum_{n=1}^{X_s} C_n,$$

since $X_{t+s} - X_t \stackrel{d}{=} X_s$ and $C_1, C_2, \ldots$ are i.i.d. So A has stationary independent increments. The same properties for R follow because $R_t = u + ct - A_t$. We have

$$\begin{aligned}
\mathbb{E}\left[e^{\theta A_t}\right] &= \mathbb{E}\left[\exp\left(\theta \sum_{j=1}^{X_t} C_j\right)\right] \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_t = n)\, \mathbb{E}\left[\exp\left(\theta \sum_{j=1}^{n} C_j\right)\right] \\
&= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \mathbb{E}\left[e^{\theta C_1}\right]^n \\
&= \exp\left(\lambda t \left(\mathbb{E}\left[e^{\theta C_1}\right] - 1\right)\right) \\
&= \exp\left(\lambda t \int_0^\infty (e^{\theta x} - 1) f(x) dx\right).
\end{aligned}$$

Furthermore,

$$\mathbb{E}\left[e^{\theta R_t}\right] = \mathbb{E}\left[e^{\theta(u+ct-A_t)}\right] = e^{\theta u+\theta ct}\mathbb{E}\left[e^{-\theta A_t}\right]$$

from which the result follows. $\square$

On average, how much will have been paid out on claims which arrive before time t? We can differentiate the moment generating function to get the mean of A_t:

$$\begin{aligned}\mathbb{E}[A_t] = \frac{\partial}{\partial\theta}\mathbb{E}\left[e^{\theta A_t}\right]\Big|_{\theta=0} &= \frac{\partial}{\partial\theta}\exp\left(\lambda t\left(\mathbb{E}\left[e^{\theta C_1}\right]-1\right)\right)\Big|_{\theta=0} \\ &= \lambda t\frac{\partial}{\partial\theta}\mathbb{E}\left[e^{\theta C_1}\right]\Big|_{\theta=0} \\ &= \lambda t\mu.\end{aligned}$$

It follows that $\mathbb{E}[R_t] = u + (c-\lambda\mu)t$ which is positive and increasing since $c > \lambda\mu$. Note that since R has stationary independent increments, we may apply the strong law of large numbers to see that

$$\frac{R_n}{n} = \frac{u}{n} + c - \frac{1}{n}\sum_{i=1}^{n}(A_i - A_{i-1}) \to c - \lambda\mu \quad \text{a.s.}$$

In particular, this implies that $R_t \to \infty$ a.s. as $t \to \infty$. However, this doesn't tell us anything about the probability that R_t ever hits 0 i.e. the probability of ruin.

10.2 Ruin probabilities

Recall that $r(u) = \mathbb{P}_u(R_t < 0 \text{ for some } t \geq 0)$, where u is the initial reserve (i.e. $R_0 = u$). Let $\bar{F}(x) = \int_x^\infty f(y)dy = \mathbb{P}(C_1 > x)$.

Proposition 10.2. *The ruin probabilities satisfy the (renewal-type) equation*

$$r(x) = \frac{\lambda}{c}\int_x^\infty \bar{F}(y)dy + \frac{\lambda}{c}\int_0^x r(x-y)\bar{F}(y)dy. \tag{12}$$

Sketch proof. Condition on the time T_1 of the first claim arrival and C_1 the first claim amount. Note that $T_1 \sim \text{Exp}(\lambda)$ and C_1 has density f, so

$$\begin{aligned}r(x) &= \int_0^\infty\int_0^\infty r(x+ct-y)f(y)dy\lambda e^{-\lambda t}dt \\ &= \int_x^\infty \frac{\lambda}{c}e^{-(s-x)\lambda/c}\int_0^\infty r(s-y)f(y)dyds,\end{aligned}$$

changing variable with $s = x + ct$. Differentiating, we obtain

$$\begin{aligned}r'(x) &= \frac{\lambda}{c}r(x) - \frac{\lambda}{c}\int_0^\infty r(x-y)f(y)dy \\ &= \frac{\lambda}{c}r(x) - \frac{\lambda}{c}\int_0^x r(x-y)f(y)dy - \frac{\lambda}{c}\int_x^\infty f(y)dy.\end{aligned}$$

Now note that we also have the terminal condition $r(\infty) = 0$. With this condition, it turns out that the above integro-differential equation has a unique solution. It, thus, suffices to show that any solution of the renewal equation (12) also solves the integro-differential equation and has $r(\infty) = 0$.

Write the convolution product in (12) with the variables interchanged, i.e.

$$r(x) = \frac{\lambda}{c}\int_x^\infty \bar{F}(y)dy + \frac{\lambda}{c}\int_0^x r(y)\bar{F}(x-y)dy.$$

Differentiating, we get

$$\begin{aligned} r'(x) &= -\frac{\lambda}{c}\bar{F}(x) + \frac{\lambda}{c}r(x)\bar{F}(0) - \frac{\lambda}{c}\int_0^x r(y)f(x-y)dy \\ &= -\frac{\lambda}{c}\int_x^\infty f(y)dy + \frac{\lambda}{c}r(x) - \frac{\lambda}{c}\int_0^x r(x-y)f(y)dy, \end{aligned}$$

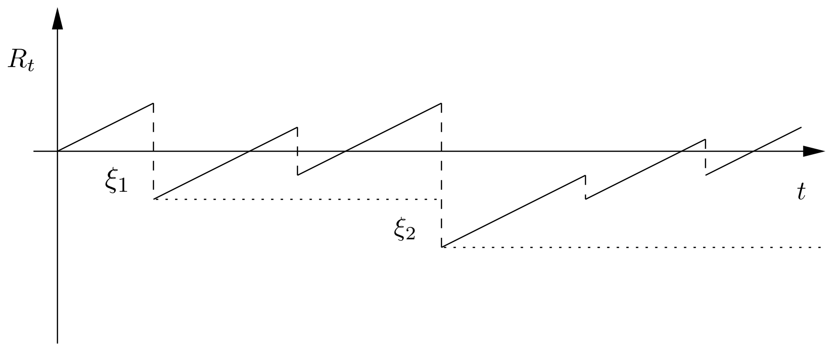
which is what we wanted. The condition $r(\infty) = 0$ can be checked with a little care. □

Note that $r(0) = \frac{\lambda}{c}\int_0^\infty \bar{F}(y)dy = \frac{\lambda\mu}{c} = \rho$, so that even with 0 initial reserve, the company is not certain to be ruined.

Where is the renewal process? Imagine we start with 0 reserve but don't stop when $(R_t)_{t\geq 0}$ goes negative. The process $(R_t)_{t\geq 0}$ has stationary independent increments. This holds not only at deterministic times but also at the stopping times defined by $\tau_0 = 0$ and, for $i \geq 0$, $\tau_{i+1} := \inf\{t \geq \tau_i : R_t < R_{\tau_i}\}$. These are the successive times at which $(R_t)_{t\geq 0}$ hits a new minimum. Since $R_t \to \infty$ a.s. as $t \to \infty$, there is a last time that $(R_t)_{t\geq 0}$ hits its past-minimum. So for some $n < \infty$, $\tau_n = \infty$. Now, if $\tau_i < \infty$, the process

$$(R_{\tau_i+s} - R_{\tau_i})_{s\geq 0}$$

is independent of $(R_t)_{0\leq t\leq \tau_i}$ and has the same distribution as $(R_t)_{t\geq 0}$. Every time we hit a new low, we have a probability ρ of hitting a lower point later (since this is just like starting at 0 and asking if $(R_t)_{t\geq 0}$ goes negative). The ith time that the process goes below its last minimum, assuming that it does (i.e. that $\tau_i < \infty$), it overshoots by an amount $\xi_i = R_{\tau_{i-1}} - R_{\tau_i}$, $i \geq 1$:



Moreover, $\xi_1, \xi_2, \ldots$ are i.i.d. Since we have independent trials with probability $1-\rho$ of never hitting a new low, eventually the process will reach its lowest point and never come back. So $(R_t)_{t\geq 0}$ hits a geometric number T of new lows before never going lower:

$$\mathbb{P}(T = n) = \mathbb{P}(\tau_n < \infty, \tau_{n+1} = \infty) = \rho^n(1-\rho), \quad n \geq 0.$$

Since hitting 0 started from u is the same as hitting $-u$ started from 0, we can express $r(u)$ as follows:

$$r(u) = \mathbb{P}\left(\sum_{i=1}^{T}\xi_i > u\right) \tag{13}$$

The idea is that $\xi_1, \xi_2, \ldots$ are the inter-arrival times of a renewal process, but that we only use the first T of them.

It turns out that the right distribution for $\xi_1, \xi_2, \ldots$ is the distribution of LU, where L has the size-biased density $\hat{f}$ associated with f, and $U \sim \mathrm{U}[0,1]$ independently. Intuitively, it takes a larger-than-average claim to go below the previous minimum level, and it turns out that the correct distribution is the size-biased distribution. Moreover, the overshoot is the portion which goes below the past-minimum; this turns out to be a uniformly distributed fraction of the claim amount.

Let us now check that $r(u)$ defined as in (13) satisfies (12). Suppose $u \geq 0$. Firstly note that if $T = 0$ then the sum is empty and so does not exceed u. Conditioning on the value of T, then, we obtain

$$r(u) = \sum_{n=1}^{\infty} (1-\rho)\rho^n \mathbb{P}\left(\sum_{i=1}^{n} \xi_i > u\right) = \sum_{n=1}^{\infty} (1-\rho)\rho^n (1 - G_n(u)),$$

where G is the distribution function of ξ_1 and G_n is the distribution function of $\sum_{i=1}^{n} \xi_i$. Let g be the density corresponding to G. Then we have

$$r(u) = \rho - (1-\rho) \sum_{n=1}^{\infty} \rho^n G_n(u). \tag{14}$$

Now note that g is the density of LU and so

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

Taking the convolution product of (14) with $\frac{\lambda}{c}\bar{F}$, we get

$$\begin{aligned}
\frac{\lambda}{c}\int_0^x r(y)\bar{F}(x-y)dy &= \int_0^x \rho^2 g(x-y)dy - \int_0^x (1-\rho)\sum_{n=1}^{\infty} \rho^{n+1} G_n(y) g(x-y)dy \\
&= \rho^2 G(x) - (1-\rho)\sum_{n=1}^{\infty} \rho^{n+1} G_{n+1}(x) \\
&= \rho G(x) - (1-\rho)\sum_{n=1}^{\infty} \rho^n G_n(x) \\
&= \frac{\lambda}{c}\int_0^x \bar{F}(y)dy - (1-\rho)\sum_{n=1}^{\infty} \rho^n G_n(x) \\
&= \frac{\lambda}{c}\int_0^x \bar{F}(y)dy + r(x) - \rho \\
&= r(x) - \frac{\lambda}{c}\int_x^{\infty} \bar{F}(y)dy.
\end{aligned}$$

Rearranging, we get (12), as claimed.